Technical requirements

Digitization and creation of METS-ALTO metadata for monographs

Author: zarko.stojkovic@bnl.etat.lu Version: 1 Date: 03/12/14

Table of contents

1	Inti 1.1	ntroduction Purpose of the project			
	1.2	Vol	ume indication	. 5	
	1.3	Stat	te of the monographs	. 5	
	1.3	.1	Transport	. 5	
	1.3.2		Storage at the service provider	. 5	
	1.3.3		Handling	. 6	
	1.4	Pro	ject goals	.7	
	1.4	.1	Digitization	.7	
	1.4.2		Metadata	. 8	
	1.4	.3	Languages	. 8	
2	Del	iver	ables	. 8	
3	Dig	itiza	ntion	.9	
	3. I Dolini	Ger	ieral	. 9	
				. 7	
).Z	JCa		10	
	د.د د د	ופוט זי	TIEE files in true color	11	
	3.3.1			11	
	3.3.Z		Plack and white DNC	11	
	3.3.3			11 12	
	3.3.4		PDF Tites	12	
	د.د م	.)		13	
	3.3 2.3	.0	Naming and location of the image and metadate files	17	
4	3.3	./ TC	Naming and location of the image and metadata mes	17	
4	MLIS 4.1 General		neral	18	
	4.2	Stru	ucture of a METS file	19	
	4.2	.1	Sections	19	
	4.2	.2	Pointers	19	
	4.3	MET	۲S profile	19	
5	ALT	ALTO			
	5.1 Ger		neral	20	
	5.2	5.2 Structure of an ALTO file			
6	Guide for formatting the information			23 23	
	6.2	Stru	ucturing elements for monographs	23	

	6.2.1 Ba	sic structure for monographs	24
	6.2.2 FF	ONT section	25
	6.2.2.1	TITLE_SECTION	26
	6.2.2.2	STATEMENT_SECTION	32
	6.2.2.3	FRONTISPIECE	36
	6.2.2.4	COVER_SECTION	39
	6.2.2.5	BASTARD_TITLE_SECTION	47
	6.2.2.6	TABLE_OF_CONTENTS	51
	6.2.2.7	INDEX	57
	6.2.2.8	PREFACE	64
	6.2.2.9	BIBLIOGRAPHY	76
	6.2.2.10	D ADVERTISEMENT	78
	6.2.3 M	AIN section	81
	6.2.3.1	TABLE_OF_CONTENTS	82
	6.2.3.2	INDEX	82
	6.2.3.3	CHAPTER	82
	6.2.3.4	APPENDIX	84
	6.2.3.5	ADVERISEMENT	89
	6.2.4 BA	ACK section	89
	6.2.4.1	STATEMENT_SECTION	90
	6.2.4.2	COVER_SECTION	90
	6.2.4.3	TABLE_OF_CONTENTS	90
	6.2.4.4	INDEX	90
	6.2.4.5	BIBLIOGRAPHY	90
	6.2.4.6	APPENDIX	90
	6.2.4.7	ADVERTISEMENT	90
	6.3 Specia	ıl cases	91
	6.3.1 Se	ction breaks (separators images) and ornaments	91
	6.3.2 Hy	/mnals	95
	6.3.3 Ve	erso-recto multi language	96
	6.4 More	examples	97
7	Schemas a	and validation	98
	7.1 Schem		98
	1.2 Valida	tion of MEIS and ALIU files	99
	7.3 Verific	cations and manual corrections	. 100
	7.3.1 Ge		. 100
	7.3.2 Ite	ems to be verified	. 100

Bibliothèque nationale de Luxembourg

	7.3.3	E	stimated number of manual checks	100	
8 9	Volun MFTS	netr and	ic statistics	101 104	
, 10	10 Ouality reports and control				
10	.1	Ćon	npleteness of the collection	104	
10	.2	Dat	a integrity	105	
10	.3	Qua	ality control of the manual corrected titles	105	
10	.4	Seg	mentation quality control	107	
	10.4.	1	Blocking errors	107	
	10.4.2	2	Major errors	108	
	10.4.3	3	Minor errors	108	
11	Inven	tory	excel sheet	110	

Table of appendixes

Appendix A - File locations and naming	111
Appendix B - METS & ALTO Requirements Specification	115
Appendix C - Color codes	173
Appendix D - Examples of how to structure a monograph	177

1 Introduction

1.1 Purpose of the project

The purpose of the project is the digitization and creation of METS (Metadata Encoding and Transmission Standard) / ALTO (Analyzed Layout and Text Object) metadata for 465 monographs following the instructions of the present specification document.

1.2 Volume indication

Number of monographs: 465 Number of pages: ~50500 Page format: DIN A5-A3. Some of the books contain fold-out maps. These maps can be up to A1 and have to be scanned folded out.

Number of chapters, images, maps, etc. to detect: ~67000 Number of elements to be corrected manually: ~26000

For more details, see chapter 8 "Volume statistics".

1.3 State of the monographs

The service provider is responsible for transport, storage and proper handling of the provided monographs. Monographs' security must be guaranteed at all times.

1.3.1 Transport

The monographs will be put in numbered boxes. When returning the boxes, the monographs should be in their respective box.

1.3.2 Storage at the service provider

It is essential to store the monographs in a clean place and under suitable climatic conditions. Ideal storage consists of a relative humidity of 50% and a temperature of 18° C. Direct sunlight should be avoided. The monographs must be stored horizontally flat, without superposing too many monographs to avoid that their weight cause deformation or mechanical damage.

1.3.3 Handling

Paper

The paper is generally fragile and must be handled with care to avoid damaging it further. Some pages have defects such as folds, rips or a bunch of pages are stuck together. The following list explains what to do for what type of defect:

Folds:

If the folds can be unfolded without forcing, the page has to be scanned unfolded. If it's not possible to unfold the page without ripping it, than the page has to be scanned as-is.

Rips:

For the rips or damages, the page should be scanned as-is.

Stuck pages:

If the pages cannot be un-stuck without damaging them, the monograph shouldn't be digitized and should be put in a separate box for sending back to the BnL.

Uncut pages:

The set of monographs may contain monographs with uncut pages. If such a monograph is detected, it should put in a separate box for sending back to the BnL.

Loose pages:

If loose pages are found, they should be digitized at the position where they are found. If the position cannot be defined, these pages should be put at the end of the monograph and digitized in this position.

External pages or material:

If external pages or material that doesn't belong to the monograph is found, it should be put in a separate box for sending back to the BnL.

Monographs with stuck or uncut pages can be put in the same box. External material should be put in a different box. These boxes should be sent to the BnL after the digitization process.

If some other damage type than here described is found, the service provider should contact the BnL for further instructions.

Binding

Monographs that have a wide opening angle can be digitized as-is. Some monographs can't be opened at 180 degrees and must be digitized at smaller angles.

1.4 Project goals

The digitized monographs will be made public on the internet. The added value of this whole project is to allow the use of cutting-edge technologies that allow the end-user to search through the structure and full text of the monographs. Digital access to the monograph will also allow preserving the fragile sometime unique originals. The data produced in this process (images and metadata) meets the demands of the long-term digital preservation.

1.4.1 Digitization

Files from the digitization process are:

- TIFF images (24 bit true color, not compressed) intended for archiving and long-term preservation
- PNG images (2 colors, black RGB 0:0:0 and white RGB 255:255:255) for web presentation
- PDF files of smaller size for printing and downloading by the end user. The PDF files must be based on the TIFF images (24 bit true color). One PDF per page and one PDF of the whole monograph must be created. Creating bookmarks is required for the PDF of the whole monograph. If the monograph has a table of contents, the bookmark in the PDF must be based on it. This means that at least all the chapters and subchapters in the table of contents of the monograph must be present in the bookmark of the PDF.

1.4.2 Metadata

Metadata entry is based on digital images. The METS format will be used to model the logical structure allowing the search through the monograph for pages, chapters, etc. It will be also used for managing the technical metadata allowing long-term preservation. The ALTO metadata will be used to describe the physical structure of each page and its full text, created by an OCR process (Optical Character Recognition). The combination of METS and ALTO metadata will allow to search through the full text of the monograph and redirect the user to the pages and sections he was searching for.

1.4.3 Languages

All the documents concerning project planning and management or manuals are to be written in English. Meetings are held in Luxemburgish, French, English or German.

2 Deliverables

- 1 TIFF per page, 24 bits true color, 300 dpi, "Original Optimized Scans"
- 1 PNG per page, black and white, 300 dpi
- 1 PDF per page containing a small picture, 300 dpi, with hidden full-text
- 1 PDF of the whole monograph, small pictures, 300 dpi with hidden fulltext and navigational structure
- 1 METS file per monograph, structured with help of MODS (Metadata Object Description Schema) schema, MIX (Metadata for Images in XML Standard), Monograph (see chapter 6)
- 1 ALTO file per page, containing the layout and the OCRed text
- 1 "batch.xml" file, describing the files of the digital monograph

3 Digitization

3.1 General

While digitization is the step preceding the creation of METS and ALTO metadata, it is essential to note that these processes cannot be entirely separated from each other. There are many interdependencies between the image files and the metadata.

Examples (not exhaustive)

- METS metadata files refer to ALTO and image files.
- The creation of whole monograph PDF requires the integration of the following elements:
 - Full-text from OCR process
 - Data from the METS file
 - Digitized images
 - Bookmarks from the METS data structure
 - Manually corrected text
- A manual correction of the generated text from the OCR process is required for all the titles, the authors, images-, tables- and map captions as well as their authors. The corrected text appears in the METS files as well as in the ALTO and PDF files.

Delivery of the images and metadata

All the digital files are to be copied to an external hard disk provided by the service provider. They must be tested for their good working condition and formatted in NTFS.

Final deliveries should regroup complete monographs on hard drives. If a monograph requires more space that there is on a disc, the whole monograph is stored on another disc.

Each disc must be accompanied by an identical backup copy on a second disc.

All the XML files (METS, ALTO, batch.xml) must be encoded in UTF-8.

3.2 Scan station

The scan station must have a horizontal and vertical true optical resolution of 300 dpi (dots per inch) without interpolation.

Some monographs are bound in a very tight manner and some are so fragile that they cannot be opened enough to be digitized spread on a flat support. They should be scanned with an opening of less than 180 degrees, sometimes between 90 and 110 degrees. These volumes can be scanned on a device similar to this one



3.3 Digital images

3.3.1 TIFF files in true color

Format: not compressed TIFF, 24 bit true color

All the pages must be digitized, including front and back cover and blank pages.

3.3.2 Optimizations

After the digitization process, the images have to undergo certain basic transformations so that they accurately represent the information on the pages. These transformed images are referred to as "Original Optimized Scan". The transformations that have to be made are the following:

- Alignment of pages. If digital images were scanned obliquely, they have to be reoriented the right way.
- Correction of the curvature of the page. Images should have a flat appearance.
- Cropping of the digital image to show only the original page. The cropping is mainly used to remove the black borders. However the original page dimension must be respected.
- The color tone of the images must be as accurate as possible. On the ColorChecker chart, all the tones must be distinguishable.
- The illumination of the page must be uniform.

Additional optimizations can be suggested by the service provider but they have to be accepted by the BnL prior to the beginning of work.

3.3.3 Black and white PNG

Format: PNG, 2 colors, black (RGB 0:0:0) and white (RGB 255:255:255)

The PNG files should have the same size and resolution as the TIFF images. The coordinates in the ALTO files have to correspond to the TIFF and the PNG files.

The text and the graphical elements such as illustrations or maps should be as legible as possible. The aim is to provide these digitized images to the public as a lower size alternative to the higher-quality TIFF images.

The service provider should describe the method used for generating the PNG images.

3.3.4 PDF files

Two types of PDF/A-2 files must be created. They are intended to be downloaded by the end-user and optionally printed. The PDFs are based on a lighter image format like JPG or JP2000 but in 300 dpi resolution for printing in original format.

PDF containing one page

- It contains one JPG image of the page in 300 dpi with hidden plain-text but without bookmarks.

PDF containing the whole monograph

- It contains all the pages of the monograph with hidden plain-text and bookmarks.

The service provider may suggest an advanced compression method for PDFs.

The following example shows a PDF. The three illustrations show the different levels of navigation via bookmarks.

First level bookmarks:



Navigation example:



Bookmark example referring to a chapter:

🔁 Making Cheese.pdf (SECURED) - Adobe Reader				
File Edit View Window Help *				
🗁 Open 🛛 🧔 🔁 🏠 🔛	→ ↓			
Bookmarks	filtered through a porcelain filter it will leave a gelatinous mass in the filter. This is the casein; or, if skim milk be revolved for a long time in a separator bowl, a layer of casein will be de-			
PrefACE Chapter I. THE CONSTITUTION OF	posited on the walls of the bowl. Casein is dissolved in solutions of borax, sodium phosphate, and alkalis. It is used commer- cially as a sizing for paper.			
MILK. MILK. 1. PURPOSE OF MILK. 2. COMPOSITION. 3. MAN'S USE OF MILK. 4. ALBUMINOIDS. 5. CASEIN. 6. ALBUMEN. 7. ALBUMOSE. 8. ASH. 9. MILK SUGAR. 10. FAT. 11. IN EMULSION. 10. CASEMBRE OF	 6. ALBUMEN. By referring to the preceding tables (2) it will be seen that the casein does not constitute all of the protein of milk. When milk has coagulated by rennet the casein is precipitated. If the whey be heated to 180° F. another precipitate will be thrown down. This is the albumen. It is much like the white of an egg which is coagulated by heat. It is in solution until the heat precipitates it. It probably accounts for part of the burnt taste of boiled milk. Albumen cannot be incorporated in Cheddar cheese without giving the conditions of sour cheese. 7. ALBUMOSE. The albumose is not coagulated by rennet heat. It is derived for the submose. 			

Туре	Default label
CHAPTER	Chapitre sans titre
TABLE	Tableau
MAP	Carte
ILLUSTRATION	Illustration
TABLE_OF_CONTENTS	Table des matières
PREFACE	Préface
BIBLIOGRAPHY	Bibliographie
APPENDIX	Appendice
ADVERTISEMENT	Publicité [X] page [Y]
	(Where [X] is the sequence number of the
	advertisement on page [Y])
INDEX	Index
CORRECTIONS	Errata
FRONTISPIECE	Frontispice
BOOKPLATE	Ex-libris

The bookmarks should be composed of the following elements:

3.3.5 Special cases for digitization

Sometimes monographs contain folded pages. In general these are maps or diagrams. They have to be unfolded and thus have a larger format than the monograph itself. In some cases these pages are detached from the monograph, in other cases these pages are glued or bound to the monograph.

The folded pages must be digitized unfolded. If the format is not rectangular, black background is used to mark places where there is no page. After digitizing the page, it should be folded in the same way as it was delivered and put in the same place if it's detached from the monograph. The orientation of the unfolded page should be so that any caption or title is horizontal from left to right.

In the example below, the verso is rotated.

Example:





3.3.6 OCR

The recognized OCR text must be encoded in UTF-8.

<u>Note</u>: If the original font is Gothic type, please do not use Unicode codes to represent the Gothic letters. Instead use the corresponding normal Latin letters. The goal is not to show the original Gothic letter type, but to represent everything in modern Latin letters.

Description of the collection:

- Languages: German, French, Luxemburgish, Dutch and Latin and books with several languages like French/German, French/German/Luxemburgish.
- Fonts: Gothic and non-Gothic fonts together with many fancy fonts for ads or first letter of a chapter.
- The character-sets to be used are German, French, Luxemburgish, Dutch and Latin, depending on the processed language.
- The special Luxemburgish character-set is "äàâ ç ëéêè îi ôö üùû ß ÄÀÂ Ç ËÉÊÈ ÎI ÔÖ ÜÙÛ"

The OCR must follow the structure of the chapters and sections.

The supplier is asked to do his best effort for OCR. No OCR error rate is stated by the BnL. However the supplier is invited to describe his OCR process and explain clearly the process and the used OCR engines (including software versions).

3.3.7 Naming and location of the image and metadata files

See appendix A: File locations.

4 METS

4.1 General

The METS schema (Metadata Encoding and Transmission Standard), which allows the exchange of digital documents among heritage institutions was created at the initiative of the Digital Library Federation (DLF) and is a particular implementation of the OAIS reference model (Open Archival Information System). The library of Congress in the United States (<u>http://www.loc.gov/standards/mets</u>) is currently the maintenance agency for the METS schema.

METS is an XML schema for document creation, or more precisely digital objects (one digital object corresponds to a METS file). A digital object can be simple or complex (one or more digital files, one or more file formats, by its internal and external structure).

As part of this project, one METS file will be created per monograph. The METS file will connect

- The digitized TIFF files
- The PDF files of every page
- The ALTO files with the contents to be detected within the OCR process

Inside the METS file, the MODS.XSD scheme will be used to describe bibliographic data, the MIX.XSD schema for describing the administrative metadata of the TIFF files.

An extraction of the METS file must be validated against the BNL-MONOGRAPH_V1.0.XSD schema to ensure logical data structure. This validation is described in more details in chapter 7.

4.2 Structure of a METS file

4.2.1 Sections

A METS file is structured into seven sections, which may include one or more groups of metadata.

In this project, the elements contained in each METS file are:

- Header: information on the METS document
- Descriptive Metadata: descriptive metadata, external with links or embedded in the document. We record MODS data describing bibliographic data in this section.
- Administrative Metadata: administrative metadata, external or embedded. We use this section to describe the technical metadata using the MIX scheme.
- FileSection: list of files that make up the object (id, format, url)
- Structural Map: Here is the link between the image, the ALTO file and the logical structure of the monograph.

4.2.2 Pointers

METS uses a system of pointers to relate metadata elements and files between them. METS offers several types:

- xlink: to point to a block outside the METS document
- filepointer: internal to the METS document and can point to a file identifier
- area: integrated in the structural map and allows to point to portions of files

4.3 METS profile

METS profiles are intended to describe a class of METS documents.

They give instruction to authors of METS documents and programmers to create valid METS documents. The METS profile is not intended for machine processing. It serves as documentation.

5 ALTO

5.1 General

ALTO is a standard from the European project METAe to represent the physical layout of a document, the position of words, etc...

It is used to store information about the layout and content of any printed document. It is therefore particularly suited to represent the OCR results. As part of the project, ALTO is used together with METS. One ALTO file will be created per scanned page.

5.2 Structure of an ALTO file

ALTO is an XML file used to store layout information and the recognized text in the OCR phase. The METS file saves the metadata and the structural information while the ALTO file saves the contents of the page and his physical shape.

The ALTO file contains a tag "STYLES" that describes the different styles (paragraphs and fonts) that are used. The "LAYOUT" tag describes the content of the page. A page is divided into regions (Print space, left margin, right margin, top margin and bottom margin).

The following illustration represents the main tags that are used by ALTO.



And here is sample of the alto.xsd schema:



6 Guide for formatting the information

6.1 General

One of the aims of the project is to create a hierarchical structure of the information contained in the pages of the monograph to allow optimal search and navigation. This structure will be modeled within the METS file in the structMap section (logical part).

The METS profile is very general (mets_profile_bnl_monograph.xml). As part of this project the schema BNL-MONOGRAPH_V1.0.XSD will be used and it can be downloaded from the following URL: http://downloads.bnl.lu/schemas/bnl-monograph_v1.0.xsd

<u>Warning</u>: To ensure that the hierarchical structure described by the structMap (locigal map) meets the permitted structures defined in the BNL-MONOGRAPH_V1.0.XSD scheme, it is required from the service provider that each of the strucMap tags are validated against this schema. However, BNL-MONOGRAPH_V1.0.XSD is not directly included in the METS schema. Consequently, a preprocessing (using regular expressions for example) is required in order to validate the contents of the section against the schema.

This validation is described in more detail in Chapter 7.

6.2 Structuring elements for monographs

The monographs are structured according to the BNL-MONOGRAPH_V1.0.XSD schema. All the elements of the schema are not used in this project. However, the following elements must be identified:

- Title of the work
- Authors
- Table of contents
- Chapters
- Illustrations / Photos / Images / Maps (if possible linked to the corresponding paragraph or chapter)
- Tables
- Prefaces
- Bibliographies
- Footnotes
- Advertisement
- Appendixes
- Statements
- Indexes
- Marginalia
- Corrections
- Table of contents



- Bastard title sections
- Title sections
- Frontispieces
- Bookplates

The chapters are composed of paragraphs.

The following sections describe how to detect these elements, how to structure and describe them using the BNL-MONOGRAPH_V1.0.XSD schema.

6.2.1 Basic structure for monographs

Each monograph is described by the following structure:



The monograph is composed of three main sections. The FRONT, MAIN and BACK section. For our project each section appears once.

These sections are described below.

6.2.2 FRONT section

The FRONT section is composed of the following elements (extract of the BNL-MONOGRAPH_V1.0.XSD):



6.2.2.1 TITLE_SECTION

The TITLE_SECTION repeats the title and author as printed on the cover or spine. It is composed of the following elements:



In our case the following elements are to be detected. The colors used in the example below are shown before the name of the element.

- Blue color: TITLE_SECTION
- Orange color: TITLE_OF_WORK
- Red color: CREATOR_OF_WORK
- Green color: PUBLISHING_STMT
- Black color: STATEMENT
- CONTENT
 - Violet color: ILLUSTRATION
 - o BOOKPLATE

NOTE:

- PUBLISHING_STMT is a block of text that contains information about the publication, publisher, date of publication, place of publication, etc.
- STATEMENT is a block of text that cannot be defined as a TITLE_OF_WORK, CREATOR_OF_WORK, PUBLISHING_STMT or any other CONTENT element.

Here are some examples of TITLE_SECTIONs and how they should be tagged:







NOTE:

Comments written by hand don't have to be structured

- If the TITLE_PAGE is followed by some PUBLISHING_STMT or STATEMENT on the next page, the PUBLISHING_STMT resp. the STATEMENT should be added to the TITLE_PAGE.





6.2.2.2 STATEMENT_SECTION

If there is some block of text on a single page before the first chapter begins and is not part of the TITLE_SECTION, COVER_SECTION or any other section from FRONT, then the block of text should be put in the STATEMENT_SECTION.



In our case the following elements are to be detected. The colors used in the example below are shown before the name of the element.

- TABLE
- Green color: PUBLISHING_STMT
- PARAGRAPH
 - Light blue color: TEXT
 - o MARGINALIA
- MAP
- ILLUSTRATION
- BOOKPLATE
- SUBHEADLINE
- Yellow color: HEADLINE
- Red color: AUTHOR

NOTE:

A block of text should be tagged as <PARAGRAPH><TEXT> and not <TEXTBLOCK>.

Here are some examples of STATEMENT_SECTIONs and how they should be tagged:





NOTE:

If the block of text contains the word "Extrait", then it should be tagged as PUBLISHING_STMT.



6.2.2.3 FRONTISPIECE

A frontispiece in books generally refers to a decorative or informative illustration facing a book's title page, being the verso opposite the recto title page. While some books depict thematic elements, other books feature the author's portrait as the frontispiece.



In our case the following elements are to be detected. The colors used in the example below are shown before the name of the element.

- ILLUSTRATION
 - Violet color: IMAGE
 - Black color: CAPTION
 - Red color: AUTHOR

Here are some examples of FRONTISPIECEs and how they should be tagged: (Left side: FRONTISPIECE; right side: Light blue color TITLE_SECTION)
Technical requirements







6.2.2.4 COVER_SECTION

The outside of a book which wraps around the text and protects the pages is the cover. Hardcover books are those bound in cloth or leather over boards. Softcover or paperback books have covers without boards usually made of heavy paper or other flexible material.



- Green color: PUBLISHING_STMT
- Black color: PARAGRAPH
 - TEXT
- ILLUSTRATION
 - Violet color: IMAGE



- Pink color: CAPTION
- AUTHOR
- BOOKPLATE
 - Yellow color: IMAGE
 - CAPTION
 - AUTHOR
- SUBHEADLINE
- Orange color: HEADLINE
- Red color: AUTHOR

Here are some examples of COVER_SECTIONs and how they should be tagged:





In the following COVER_SECTION, the IMAGE (violet) is overlapped with HEADLINE (orange), AUTHOR (red), STATEMENT (black) and PUBLISHING_STMT (green).



The following monograph has 2 COVER_SECTIONs:





The following Monograph has a front cover with some patterns and no text or images.

NOTE:

If a monograph begins with a blank page or a patterned page with no text or illustrations, the page should not appear in the logical structure. Anyway the page must be scanned and put in the physical structure. The same rule applies to any blank pages in the monograph whether in FRONT, MAIN or BACK section in the monograph.



The following example shows a BOOKPLATE in a COVER_SECTION on the verso side of the outside cover.



6.2.2.5 BASTARD_TITLE_SECTION

The bastard title only contains the title of work (as opposed to the author, publisher etc. found on the full title page) and usually precedes the title page.



- SUBHEADLINE
- Orange color: HEADLINE

Here are some examples of BASTARD_TITLE_SECTIONs and how they should be tagged:







6.2.2.6 TABLE_OF_CONTENTS

This is a list of chapter headings, and nested subheadings, together with their respective page numbers. It can include all FRONT-matter items, together with chapters in the MAIN-matter and BACK-matter.



- Green color: TABLE_OF_CONTENTS
 - Orange color: CAPTION
 - Yellow color: DATA
 - Light blue color: TEXTBLOCK
 - Dark blue color: PAGE_NUMBER
 - Violet color: FOOTNOTE
 - Red color: DATA
 - Black color: ITEM_ID
 - Pink color: TEXTBLOCK
 - Green color: TABLE_OF_CONTENTS

If the table of contents is organized according to the principle of at least one line per reference, it is necessary that each reference is in its own TEXTBLOCK. If the page number is present, it must be recognized and placed in the PAGE_NUMBER element directly after the block of text (TEXTBLOCK) to which it refers.

	Innansverzeichnis.
Erstes Kapitel. —	- Das Kloster Marienthal
Zweites Kapitel. –	- Maße und Münzen des dreizehnten Jahrhunderts · · · · · · · · 7
	§ 1. — Maíse
	§ 2. — Münzen
Drittes Kapitel	- Erwerbstitel des Klosters
	§ 1. — Mitgift der Schwestern
	§ 2. – Begräbnisstätten und Anniversarien
	§ 3. — Patronatsrechte
	§ 4. — Inkorporationen
	§ 5. — Kauf und Tausch
Viertes Kapitel	- Besitzungen und Herkunft derselben ; Länder, auf die sie sich verteilen
	§ 4. — Grofsherzogtum Luxemburg
	§ 2. — Rheinpreußen
	§ 3. — Deutsch-Lothringen
	§ 4. — Frankreich
	§ 5. — Belgien
Fünftes Kapitel	- Verwaltung der Güter
Ashang Ver	alabaia dan Vinnahaman dan Vinatana ing Ling tara

In the following example the table of contents is spread over 2 pages:

TABLE	
DES MATIÈRES.	
	Pages.
réface	1
es indications thérapeutiques des eaux de Mondorf. Introduction	n. 5
laladies scrofuleuses et lymphatiques.	17
Observation I	21
Observation II Tumeur ganglionnaire	26
Observation III Engorgement des ganglions sous-maxillaire	s. 28
Observation IV Scrofules Anémie profonde Goître	29
Observation V Otorrhée Engorgement ganglionnaire	31
Observation VI Tumeur blanche du genou	33
Observation VII Tumeur blanche du coude droit Car	fie .
et fistules osseuses	34
Observation VIII Otorrhée purulente; osteite strumeuse	de
l'oreille	37
Observation IX Adénite scrofuleuse	43
Observation X Tumeur blanche de l'articulation tibio-féme	0
rale	. 44
Observation XI Carie du tibia	47
hlorose. Anémie	49
Observation XII. — Chloro-anémie. — Dysménorrhée	
Observation XIII. — Aménorrhée. — Chlorose	58
Observation XIV Chlorose Hystérie	59
Observation XV Ménorrhagie Troubles digestifs grave	55.
- Anémie profonde	61
Observation XVI. — Chlorose, — Oedème	64
laladies des organes digestifs. — Dyspepsie	67
Observation XVII. — Dyspepsie simple,	88
Observation XVIII Dyspepsie Migraine	89
Observation XIX. — Dyspepsie pituiteuse	91
Observation XX Dyspepsie flatulente Vertige stomacs	d.
Hypochondrie	92

Observation XXL Constinution	0
Valadies du faie - Istère jannisse	10
Observation XXII - Istère simple - Eréquentes régidires	10
Observation XXIII Ictore chroniona	10
Congestion chronique on Engangement du faie	10
Observation XXIV Congestion chronique du foie - Dysnen-	(a co
sie, anémie	11
Observation XXV Congestion du foie Ictère	11
Observation XXVI Congestion du foie Calcula biliaires	44
Observation XXVII Congestion chronique du foie Ané-	
mie Dyspepsie Hypochondrie.	14
Maladies de l'utérus	12
Observation XXVIII Congestion utérine Leucorrhée	
Chlorose Dyspepsie	13
Observation XXIX Engorgement et ulcérations du col uté-	
rin Dysménorrhée et aménorrhée Hystérie	13
Observation XXX Engorgement, ramollissement et ulcérations	
fongueuses du col utérin Abaissement Antéversion	
Ménorrhagie, - Etat nerveux	15
Maladies de la peau	14
Observation XXXI Eczéma chronique Chlorose	14
Observation XXXII Acné	15
Observation XXXIII Impetigo strumeux	15
Goutte et rhumatisme,	115
Observation XXXIV	16
Observation XXXV Goutte chronique	17
Observation XXXVI Rhumatisme articulaire chronique	
Observation VVVVV	17
Observation XXXVII Rhumatisme articulaire chronique	17
trophia at paralucio Arthrite chronique localisée Amyo-	
Observation XXXIX	17
Observation XXXX - Lombago chronique.	17
culaira	No. 10
Observation XXXXI Phone	18
thoracique gancha	
Observation XXXXII D	18
musculaire - Paralysie rhumatismale Atrophie	1
De la gravelle,	18
Observation XXXXIII Green II	18
Observation XXXXIV - Gravelle urique.	19
Gravelle urique	19

If the table of contents contains more than one caption, in other words it contains sub-levels, it should be considered as a sub table of contents or

TABLE_OF_CONTENTS inside a TABLE_OF_CONTENTS. The following example shows the principle TABLE_OF_CONTENTS inside of a TABLE_OF_CONTENTS and how the different elements should be tagged:

(frfter Weil _]	Neutliche Gelänge.
Statt Sta.	geminie Geimige.
l. Sieder für die verschiedene	n Zeiten des Kirchenjahres.
Seite	Seite
1 91dpent, 1-7 1	10. Bfinaften. 70-73 93
2. Beibnachten. 8-18 10	11. 93redigtlieder. 74-75 . 98
3. Name Sefus. 19-24 . 23	12. Dreifaltigkeit. 76-79 . 101
4. Erscheinung des Berrn.	13. Fronleichnam. 80—91 . 106
25-26	14. Kommuniongesänge :
5. Septuagesima. 1) 27—39. 33	a) Vor der hl. Kommunion
6. Fastenzeit. 40-56 47	92-96
7 Oftern, 57-63	b) glach ber hi. Rommunion
8. In der Bittwoche. 64-66. 84	97-101
9. Christinimeliabri.07-09 85	15. Dety gefu. 102-104 105
II. Sieder gur bi	. Mutter Dottes,
6 a) Pobaciánae, 105-117, 136	lieder zu Maria, Muter
h)Bittaefänge. 118-121. 152	Jefu, Tröfterin der Be-
7. Suxemburger Wallfahrts-	trübten. 122-134 156
III. Oisbar han ban	Lucely und Beiligen
m. Steber bon ben	or Olm (Taba sing) Olhottala
8. Sum hi. Erzengel Ulichael	überhaunt 144 101
	04 Sum bl. Mortinus, 145, 199
136_138 180	95 Qum hl. Donatus, 146, 193
0 Qum bl. Sob. 23apt. 139. 185	96, Sum bl 9Billibrord, 147, 194
1 Qum hl. Sofeph. 140-142, 186	27. Sum bl. Gebaftian. 148. 196
9. 91m Fefte der bl. 2[poite]	28. Sum bl. Allopfius. 149. 197
	09. Sur bl. Barbara, 150, 198
Betrus u. Baulus, 143. 190	and the second s





6.2.2.7 INDEX

An index is a list of words or phrases and associated page numbers to where useful material relating to that heading can be found in a monograph. It defines a relation of 1 (term/name) to N (references).

Indexes can be nested.



- Red color: CAPTION
- Brown color: DATA
 - Pink color: TEXTBLOCK
 - Green color: PAGENUMBER
 - ITEM_CAPTION
 - AUTHOR
- FOOTNOTE
- Black color: INDEX

Here is an example of nested indexes over multiple pages:



Page 1

Page 2



Page 3



The following example shows a table of contents followed by an index over multiple pages.

NOTE:

If the heading begins with "Index" and there is only one page number per DATA element and the numbering is incremental, then the section should be a TABLE_OF_CONTENTS and not an INDEX.

INDLA.	
	Page.
Avant-Propos.	N I
I Tableaux	9
II. Aquarelles, Sénias et Encres de Chine	88
III. Dessins	97
IV. Pastels.	99
V. Gravures.	100
VI. Peintures sur porcelaine	101
VIII. Mosaiques	102
Y Saulatura	104
X Mergueterie	108
XI. Curiosités	109
XII. Photographies	110
Bakhuysen (H. van de Sande)	10
Baron (Henry)	II
Bodinier (Guillaume)	13
Bouterbak (J.)	14
Braseassat (Raymond)	14
Braun (Adolphe)	110
Brakalancamp (Quirun yan)	10
orekesencamp reamyn vanj	88
Calame (Alexandre)	17
Calame (Alexandre)	
Calame (Alexandre) Callow (William) Canal (Antoine) Canelle (Jean van de)	19
Calame (Alexandre) Callow (William) Canal (Antoine) Capelle (Jean van de). Cattermole (G.)	19 89
Callame (Alexandre) Callow (William) Canal (Antoine) Capelle (Jean van de) Cattermole (G.) Chaponnière	19 89 104
Callame (Alexandre) Callow (William) Canal (Antoine) Capelle (Jean van de) Cattermole (G.) Chaponnière Charlet	19 89 104 90

-	Pare.
C	odde (Pierre)
C	remer
D	agnan (Isidore) 20
D	e Bray (Salomon) 21
P	ecamps (Alexandre) 22, 911
10	elacroix (Eugene)
H	elaroche (Paul)
H	ow (Gerard)
n	nyarger (Emmonuel)
15	steve (Ranhaë)
F	elu (Charles)
D	ischer
F	ovatier
E	ratin
G	31 81
G	rapet (Marius)
G	rosclaude (Louis) 33
G	udin (Théodore) 85
H	agen (Jean van der)
H	eyden (Jean van der).
÷	ornung (Joseph)
i.	
h	ahannot (Alfred) 49
J	ohannot (Tony)
K	evser (Nicaise De)
K	irsch (Mathien) 44
1E	ækkæk (Corneille) 45
L	anglacé 102
Ľ	asinsky (Adolphe)
L	ucas (D.)
110	are (Antoine)
	algonnian (France)
	orta
N	eer (Eglon van der)
N	encini 107
N	von
O	s (Jean van)
P	rud'hon (Pierre Paul)
P	vnacker (Adam)
IR	egny (A. de)
R	ondé (Philippe)
R	omeyn (Guillaume van)
IS	aint-Jean

	444		
	- 115 -		
mail .			
Schendel (Pierre van)	12 - 1 N F & L P 2 1 2 1	el la secondada	Page.
Schnetz (Victor)			63
Slingelandt (Pierre van)			64
Spændonck (Corneille va	m)		
Steen (Jean)			67
Steuben (Charles de)			69
Stry (Abraham van)			71
Stry (Jacob van)			
Téniers (David)			
Trimolet (Anthelme)			
Velde (Adrien van de).			36, 78
Verbeckhoven (Joseph)			
Vernet (Horace)			
Waldown (Antoino)			82, 99
Wonvormans (Philippo)			02
wouvermans (rmmppe)			
			S. Street Courters
	·		

The following example shows an INDEX with incremental numbering but no heading "Index".

LISTE DES	GRAVURĖS	
1. Le Grand Duc Guillaume	28. Le Deivelselter	52
de Nassau	29. Vue de Diekirch	53
2. La Grande Duchesse Ma-	30. La Súre	54
rie de Bragance 4	31. Château de Brandenbourg	55
3. Luxembourg : Le Bock et	32. Château de Bourscheid	57
la Vallée de l'Alzette . 7	33. Pont sur la Sùre	59
4. La Princesse héritière Ma-	34. Teufelsloch	61
rie-Adelaïde 8	35. La Súre	62
5. Le Hallerbach II	36. Esch-sur-Sûre,	64
6. Türkenkopf	37. Château de Clervaux	68
7. La Wolfsschlucht 15	38. Geierslay	70
8. Mondorf-les-Bains Le	39. Château de Vianden	73
Pavillon	40. Echternach, place du Mar-	
p. Rabenlay 17	chė	75
o. Schiessentumpel	41. Panorama d'Echternach .	76
II. Aesbachtal	42. Louis XIV	77
2. Halsbachtal-Zigeunerlay. 20	43. De Man an d'Fra op der Le	79
3. Panorama de la ville de	44. Monument d'Igel	81
Luxembourg 22	15. Mosaïque de Nennig	85
A. La Cathédrale	46. Le parc de Mondorf - les-	
5 Le Parc	Bains	80
6. L'Erenz Noire	47. Le Pavillon principal des	
7 Moulin à eau	Bains	90
8 Moulin dans le Müllerthal	48. Les Bains de Mondorf	92
a Hallarbach	49. La Hohllay	94
o, Esch-sur-Alzette	50. Pavillon dans le parc	
21. Rocher dit « Predigtstuhl» 38	d'Echternach.	96
22 Charles Martel	51. La Basilique d'Echternach	97
23. Château de Berg	52. Le Dingstuhl	99
A Larochette	53. Château de Beaufort	OI
25 Château de Meysembourg	54. RouteEchternach-Berdorf	03
26 Château de Hoblenfels	5. Müllerthal-Vogelsmühle .	06
27 Ruines du château de Bran-	56. Pavillon de Troosknep-	-
denhourg	chen	00
denbourg		

6.2.2.8 PREFACE

A preface generally covers the story of how the book came into being, or how the idea for the book was developed. This is often followed by thanks and acknowledgments to people who were helpful to the author during the time of writing. Often the preface is written by somebody else than the author.



- Brown color: PREFACE
 - Orange color: TITLE
 - SUBTITLE
 - Red color: AUTHOR
 - Yellow color: BODY
 - Dark-blue color: BODYCONTENT
 - TEXTBLOCK
 - Light-blue color: PARAGRAPH
 - FOOTNOTE
 - DATA
 - ITEM_ID

- TEXTBLOCK
- MARGINALIA
- TABLE
- MAP
- >Violet color: ILLUSTRATION
- TABLE
- CHAPTER

This non-exhaustive list of headings should be used to detect a PREFACE section:

- Vorrede
- Vorword
- Préface
- Praefatio
- Avertissement
- Avant-propos
- Vorbetrachtung
- Vorbemerkung
- Vorerinnerung
- Quelques remarques préliminaires
- Avant-propos historique
- Vorwort des Verfassers
- ...

NOTE:

PREFACE should only be used in the FRONT-section. If a heading from the list above is detected somewhere else (MAIN or BACK), it should be marked as CHAPTER section.

The following example shows a preface over multiple pages.



PRÉFACE 9 se donnait tout entier, est timidement raturée, pour faire place à des expressions plus froides et plus voilées. Les écrits de Charles Kayser n'ont point passé par le crible de cette critique réflexe, et je les donne tels qu'il les a confiés au papier, dans ses moments de loisir et de rêverie. Nous verrons donc l'auteur penser et agir d'une facon naturelle, sans contrainte, nous verrons dans les replis les plus cachés de son cœur, nous surprendrons les opérations les plus intimes de son esprit et, de la sorte, il nous sera aisé de fixer son caractère et d'augurer de ce qu'il cût pu devenir, si la mort impitoyable ne l'avait arraché à ses occupations littéraires. Mais, avant de procéder à la critique de ses écrits, parlons un peu de l'auteur lui-même. Ce sera fait en quelques lignes, car ceux auxquels principalement est destiné ce petit livre, c'est-àdire les parents et les amis du défunt, ont parfaitement connaissance de la vie et du caractère de l'auteur; d'ailleurs, que pourrait-on parler longuement d'une existence de vingt ans, paisible et obscure? Charles Kayser fut un jeune-homme doué de rares qualités. Si l'antique adage : de mortuis

nihil nisi bene, a quelque valeur, je crois qu'il

PRÉFACE 15	\$
	-
bourg, à travers l'enchantement mystérieux de	88
nuits d'été, pendant qu'au loin on entendait le	es
accords de la musique militaire, et que les étoile	s,
du fond de l'immensité bleue, nous regardaien	t,
derrière le branchage frémissant des grands arbre	s,
clignant leurs paupières dorées et éternelles! Qu	ie
d'idées riantes ou mélancoliques, tristes ou gaie	s,
folles ou sages ne naissaient point dans nos têt	es 🕯
pendant ces longues promenades, et quell	es
expressions ne trouvaient-elles pas dans no	os
bouches!	
Tag conserving contennes dans co netit liv	re
no sont on somme que la reproduction réfléch	ie
de ses conversions jotées au vent Le lecter	Ir
remanuere hermoure d'originalité un langas	20
y remarquera beaucoup u originante, un imge	re
coulant, elegant et pittoresque. Si ron y rencone	
partois queique tournure qui semple suraintee e	és
etrange, il ne laut point l'attribuer aux difficult	17
que l'auteur aurait eues a prier sa pensee at	10
exigences d'une langue qui n'etait pas sa langu	nt
maternelle; bien au contraire : ces tournures sol	08
dues à la prédilection qu'il professait pour le	it
maîtres de la vieille litterature française; il sava	al
bien que c'est là que La Fontaine, Mohere, Fasca	11,
Saint-Simon, Paul-Louis Courrier, Victor Hug	0,
Jules Janin, Hégésippe Moreau, Sainte-Deuv	0,







	PRÉFACE	19
Dann Zum Gestä	aber stieg ich ruh'gen Muths und gerne kleinen Haus hinab, zu neuem Handeln rkt in meinen einfachstillen Wänden.	
Pour	la fin ic rends attentif à une	nièce
rouvée d lui sembl l'outre-to l veut q our où l	lans un repli du carnet de Charl e être une parole de consolation adi ombe à sa mère chérie, pièce par lac qu'elle se console en attendant le le fils reverra sa mère et la mèr	es et ressée quelle grand e son
ils; la v	oici:	
	Und stiller wird die Welt, Und stiller werden Wald und Flur, Vom bleichen Mond erhellt.	
	Ich lehne hier am Buchenstamm,	
	Auf unbewohnter Höh',	
1	Der Wind rauscht durch die Blätter hin Mich fasst ein tiefes Weh.	
		12
	Wo ist o Herz, wohin ist nun	
	Die Freude und die Lust?	
	Was pochst du jetst so ungestüm, So wild in meiner Brust?	
	Was ist's, das mir das Auge trübt?	
	Was ist's, das mir das Auge trübt? Das mich beim Nah'n der Nacht	
	Was ist's, das mir das Auge trübt? Das mich beim Nah'n der Nacht An dieser öden Stätte hier	


The following example shows a preface which spreads over 2 pages.



action dans un nombre restreint d'états morbides, où j'ai la certitude qu'elles peuvent rendre les plus grands services. Aussi je me crois autorisé à dire que j'offre aux praticiens et au public un guide sincère et loya!, malgré les imperfections et les lacunes que la critique pourra facilement y signaler. Je m'estimerai heureux si, grâce aux indications et aux renseignements que je leur fournis, je puis rendre quelques services aux nombreuses victimes des maladies chroniques.

3 -

Mondorf, le 1er septembre 1869.

MARCHAL, D.-M. P.

6.2.2.9 BIBLIOGRAPHY

A bibliography is a list of source materials (usually books) that are used or consulted in the preparation of a work.



In our case the following elements are to be detected. The colors used in the example below are shown before the name of the element.

- Dark blue color: BIBLIOGRAPHY
 - Orange color: CAPTION
 - Yellow color: DATA
 - Green color: TEXTBLOCK
 - PAGENUMBER
 - ITEM_CAPTION
 - Red color: AUTHOR
 - PARAGRAPH
 - Light blue color: TEXT



MARGINALIA

- ILLUSTRATION
 - Violet color: IMAGE
 - Black color: CAPTION
 - AUTHOR

The following example shows a BIBLIOGRAPHY.



6.2.2.10 ADVERTISEMENT

An advertisement is a notice or an announcement promoting a product, a service, an event or publicizing a job vacancy.



The ADVERTISMENT element is final and is to be used every time an advertisement is detected.

In the examples that follow, the ADVERTISEMENT section is colored in violet.

The following example shows a CHAPTER section followed by 2 ADVERTISMENTs.



The following example is a full page ad.



This example shows multiple ads on a page.



6.2.3 MAIN section

The MAIN section is composed of the following elements:



6.2.3.1 TABLE_OF_CONTENTS

See chapter 6.2.2.6 TABLE_OF_CONTENTS

A table of contents can appear in the MAIN section if for example the MAIN section is divided into parts (Part 1, Part 2, Part 3, etc) regrouping chapters and each part has its own table of contents.

6.2.3.2 INDEX

See chapter 6.2.2.7 INDEX

The same principle as in TABLE_OF_CONTENTS (chapter 6.2.3.1) can be applied here. If a monograph is divided into parts, each part can have its own INDEX.

6.2.3.3 CHAPTER

A chapter is one of the main divisions of a book. It can be numbered or titled or both.



In our case the following elements are to be detected.

- TITLE
- SUBTITLE
- AUTHOR
- HEADING_TEXT
- BODY
 - BODY_CONTENT
 - TEXTBLOCK
 - PARAGRAPH
 - FOOTNOTE
 - TEXT
 - MARGINALIA
 - TABLE
 - o MAP
 - **o** ILLUSTRATION
 - o **TABLE**
- TABLE_OF_CONTENTS
- INDEX
- CHAPTER
- APPENDIX

6.2.3.4 APPENDIX

An appendix, in general, is an addition required to be made to a monograph by its author subsequent to its printing or publication. It may explain inconsistencies or expand the existing work or otherwise explain or update the information found in the main work



In our case the following elements are to be detected. The colors used in the example below are shown before the name of the element.

- Yellow color: APPENDIX
 - Orange color: TITLE
 - SUBTITLE
 - AUTHOR
 - HEADING_TEXT
 - o BODY
 - BODY_CONTENT
 - TEXTBLOCK
 - Light blue color: PARAGRAPH
 - FOOTNOTE
 - DATA
 - Brown color: ITEM_ID
 - Black color: TEXTBLOCK

- MARGINALIA
- TABLE
- MAP
- ILLUSTRATION
- TABLE
- TABLE_OF_CONTENTS
- o INDEX
- Green color: CHAPTER
 - Dark blue color: TITLE
 - BODY
 - BODY_CONTENT
 - Light blue color: PARAGRAPH
- APPENDIX

Note:

Normally an APPENDIX section is in the back section of a monograph. If a level 1 chapter comes after the APPENDIX then the appendix is put into the previous level 1 chapter.

```
<FRONT> ... </FRONT>
<MAIN>
<CHAPTER> ... </CHAPTER>
<CHAPTER>
<CHAPTER>
<APPENDIX></APPENDIX>
</CHAPTER>
<CHAPTER>
<CHAPTER></CHAPTER>
</MAIN>
<BACK> ... </BACK>
```

The following example shows an APPENDIX with chapters.





60 ANNEXES Luxemburger voll gesichert und geachtet bleiben werden, 3º dass die deutschen Truppen eiserne Disziplin zu halten gewöhnt sind, 4. dass alle Leistungen bar entschädigt werden. Ich vertraue auf den Gerechtigkeitssinn des luxemburgischen Volkes, dass es sich der Einsicht nicht verschliessen wird, dass Seine Majestät nur dem unvermeidlichen Zwange folgend und veranlasst durch die Nichtachtung der Neutralität seitens Frankreichs den Einmarsch der Truppen in Luxemburg befohlen haben und erwarte unter nochmaliger Betonung der oben gegebenen Garantien, dass das luxemburgische Volk und seine Regierung durch ihre Haltung die den deutschen Truppen gestellte Aufgabe nicht erschweren werden. TULFF VON TSCHEPE UND WEIDENBACH. Kommandierender General des preussischen VIII. Armee korps. TRADUCTION Tous les efforts tentés par Sa Majesté, notre 1. La traduction publiée par M. G. WAMPACH (le Dossier de la Gaerre, t. II, p. 104) est celle qui figurait en regard même du texte allemand. Celle que nous donnons sera mieux comprise, croyons-nous, des lecteurs français, tout en étant plus conforme à l'original rédigé en allemand.

6.2.3.5 ADVERISEMENT

See chapter 6.2.2.10 ADVERTISEMENT

6.2.4 BACK section

The BACK section is composed of the following elements:



6.2.4.1 STATEMENT_SECTION

See Chapter 6.2.2.2 STATEMENT_SECTION

6.2.4.2 COVER_SECTION

See Chapter 6.2.2.4 COVER_SECTIONs

6.2.4.3 TABLE_OF_CONTENTS

See Chapter 6.2.2.6 TABLE_OF_CONTENTS

6.2.4.4 INDEX

See Chapter 6.2.2. INDEX

6.2.4.5 BIBLIOGRAPHY

See Chapter 6.2.2.9 BIBLIOGRAPHY

6.2.4.6 APPENDIX

See Chapter 6.2.3.4 APPENDIX

6.2.4.7 ADVERTISEMENT

See Chapter 6.2.2.10 ADVERTISEMENT

6.3 Special cases

6.3.1 Section breaks (separators images) and ornaments

Sections or paragraphs are visually separated from each other with a section break, typically consisting of extra space between the sections and sometimes accompanied by an asterism, a horizontal rule, fleurons, or by other ornamental symbols.

The simple ornaments like horizontal lines or asterisms should not be tagged as an ILLUSTRATION. Here are some examples of these separators:





Nort über die Geschichte des Klosters im dreizennten enden üblichen Maße und Münzen wird das Vers ektüre erleichtern.

Quellen.

laire du Prieuré de Marienthal (Publ.). Luxembourg atus, tituli S. Agathes, in Longuiono. Augustæ Trevi

Consolatrix afflictorum.

In allen Nöthen eilet das Kind zum Mutterschooß, On the other hand, there are some more complex ornaments that should be tagged as an ILLUSTRATION. Here are some examples of these ornaments:



Schliesslich trat Baron Fagell hervor und erklärte uter Stimme, dass die Leichenfeierlichkeiten des J en Heinrich beendigt seien.



vous, ce sera pour vous murmurer doucement. »Calme tes pleurs, maman, le bon Dieu m'aime tant!.... Papa, calme tes pleurs!«....







6.3.2 Hymnals

A hymnal is a collection of hymns, i.e. religious songs, usually in the form of a book.

Here is an example of how they should be structured.

Orange colors are <CHAPTER> <TITLES>. Violet colors are music partitions structured as <CHAPTER> <BODY> <ILLUSTRATION>. Light blue colors are the song text structured as <CHAPTER><BODY><BODYCONTENT><PARAGRAPH><TEXT>.



6.3.3 Verso-recto multi language

This special case is applied if a monograph is written in 2 languages, one language on the verso page and the other on the recto page. In this case handle all the verso pages first, then all the recto pages. Because the ILLUSTRATIONS are not directly linked to the text, regroup them at the end of the monograph. The same rule applies if other structural elements are found (e.g. TABLES or MAPS). In these CHAPTERS, use the following TITLE:

- "Illustrations" for ILLUSTRATION
- "Tables" for TABLE
- "Cartes" for MAP

```
<FRONT>...</FRONT>
<MAIN>
      <CHAPTER>Chapitre 1</CHAPTER> (verso page)
      <CHAPTER>Chapitre 2</CHAPTER> (verso page)
      <CHAPTER>Chapitre 3</CHAPTER> (verso page)
      <CHAPTER>...</CHAPTER> (verso page)
      <CHAPTER>Kapitel 1</CHAPTER> (recto page)
      <CHAPTER>Kapitel 2</CHAPTER>(recto page)
      <CHAPTER>Kapitel 3</CHAPTER>(recto page)
      <CHAPTER>...</CHAPTER>(recto page)
      <CHAPTER>
            <TITLE>Illustrations</TITLE>
            <BODY>
                  <ILLUSTRATION>[caption of illustration 1]</ILLUSTRATION>
                  <ILLUSTRATION>[caption of illustration 2]</ILLUSTRATION>
                  <ILLUSTRATION>...</ILLUSTRATION>
            </BODY>
      </CHAPTER>
      <CHAPTER>
            <TITLE>Tableaux</TITLE>
            <BODY>
                  <TABLE>[caption of table 1]</TABLE>
                  <TABLE>[caption of table 2]</TABLE>
                  <TABLE>...</TABLE>
            </BODY>
      </CHAPTER>
      <CHAPTER>
            <TITLE>Cartes</TITLE>
            <BODY>
                  <MAP>[caption of map 1]</MAP>
                  <MAP>[caption of map 2]</MAP>
                  <MAP>...</MAP>
            </BODY>
      </CHAPTER>
</MAIN>
<BACK>...</BACK>
```

6.4 More examples

For more examples of complete monographs and how they should be modelled see "Appendix D - Examples of how to structure a monograph".

7 Schemas and validation

7.1 Schemas

The following schemas are to be used for this project. They can be downloaded at: <a href="http://donwloads.bnl.etat.lu/schemas/<FILE_NAME">http://donwloads.bnl.etat.lu/schemas/<FILE_NAME

FILE_NAME	Description	
mets_profile_bnl_monograph.xml	METS profile to use	
mets.xsd	METS schema (Metadata Encoding &	
	Transmission Standard)	
	Documentation:	
	http://www.loc.gov/standards/mets)	
mix.xsd	MIX schema (Metadata for still images)	
	Documentation:	
	http://www.loc.gov/standards/mix	
mods.xsd	MODS schema (Metadata Object Description	
	Ticker)	
	Documentation:	
	http://www.loc.gov/standards/mods/	
bnl-monograph_v1.0.xsd	This schema shows the structure of the	
	section <i>logical structMap</i> of the METS file for	
	monographs.	
alto.xsd	ALTO schema	
marc.xsd	The MARC formats are standards for the	
	representation and communication of	
	bibliographic and related information in	
	machine-readable form.	
samples.zip	Example of a METS file with ALTO files for a	
	different project with a different logical	
	structMap (newspaper)	

7.2 Validation of METS and ALTO files

METS and ALTO files must be validated against their schemas.

Further validation is required for each METS file. The service provider must ensure that the structMap tag (logical part), part of the METS files, is built according to the rules defined formally in bnl-monograph_v1.0.xsd. This can be achieved through:

- 1. Processing the content of the structMap tag (logical part)
- 2. Validation of the results of step 1. against the schema bnlmonograph_v1.0.xsd.

The purpose of the transformation is to remove the <div> tag (contained in structMap) and their attributes, with the exception of the "TYPE" attribute. The value of this attribute will be transformed into a tag.

Example of processing:

```
<div ID="DIVL1" TYPE="Monograph" LABEL="Aus Luxemburgs Vergangenheit und
Gegenwart"> -> <Monograph>
```

The following XSL stylesheet does this kind of transformation:

```
<xsl:stylesheet version="2.0"</pre>
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
      <xsl:output method="xml" encoding="utf-8" indent="yes"/>
      <xsl:output indent="yes"/>
      <xsl:strip-space elements="*"/>
      <xsl:template match="/"><xsl:apply-templates</pre>
select="/*:mets/*:structMap[@TYPE='LOGICAL']"/></xsl:template>
      <xsl:template match="*:div">
            <xsl:element name="{@TYPE}">
                  <xsl:apply-templates/>
            </xsl:element>
      </xsl:template>
      <xsl:template match="*:fptr">
            <fptr><area/></fptr>
      </xsl:template>
</xsl:stylesheet>
```

7.3 Verifications and manual corrections

7.3.1 General

The results of the OCR stage include errors in the automatically recognized text. To create a usable navigation system, it is necessary to ensure that a number of items have to be saved correctly. This is important for the construction of a navigation system within a PDF file containing a monograph, and for the construction of index within an application.

Items for which the result of the OCR is to be verified and corrected manually are described in the following section. For these items we require a reliability rate of 99,8 %.

The result of these corrections should be reflected in the METS, ALTO and PDF files.

7.3.2 Items to be verified

To ensure quality of the delivered files, the BnL requests that the service provider checks and corrects if necessary the following items:

- 1. The title and author of the title section
- 2. The titles of chapters, appendixes, bibliographies, prefaces, tables of contents
- 3. The captions of lists, illustrations, maps, tables, indexes
- 4. The authors of illustrations, maps, tables

7.3.3 Estimated number of manual checks

The BnL believes that the collection requires 26000 verifications of titles and, if necessary, manual correction of them. The next chapter explains the methodology used to determine this number.

8 Volumetric statistics

This chapter describes the methodology used to estimate the total number of sections and other structural elements described in Chapter 6. It also describes the number of titles, authors, captions, etc., to be checked manually. In the following chapter, all the elements listed in section 7.3.2 are called "title".

The set for calculating the statistics is composed of 53 monographs. In the 53 monographs we had 8251 pages. In the 8251 pages we have 4021 titles. Titles were manually counted.

The formula used to count the total number of manual correction in set is as follows:

[sum of titles] / [sum of pages] => [titles per page] 4021 / 8251 = 0,487334 titles per page

To estimate the total number of pages to be digitized, we take the page field from our bibliographical database and add them together to get the total page number to be digitized. But, because the page number from our bibliographical database corresponds to the last printed page number of the monograph, we can say that this number is not the actual page count of the monograph. To correct this number, we add 10% to the page number from our database.

```
[page number from database] * 1,1 = [page count]
```

To get the total page count for all the monographs we simply add the [page count] for every monograph together which gives us the [total page count].

```
[total page count] in number 50479 total pages
```

With the [total page count] and the [titles per page] we can now estimate the total number of titles to be checked manually in the provided set of monographs for digitalization:

> [titles per page] * [total page count] 0,487334 * 50479 = 24600

Each estimated number can only be approximate. Some monographs have surely been poorly estimated.

To equalize these errors, we finally added a margin of 5%.

Sections	Estimated number of	Estimated number with
	sections to be detected	added margin
Chapters	17910	18800
Appendixes	87	92
Images	11350	11900
Image captions	9700	11000
Image authors	164	172
Tables	1480	1550
Table caption	810	850
Table authors	8	9
Maps	280	300
Map captions	79	83
Map authors	40	42
Advertisements	1900	2000
Footnotes	10820	11400
Statements	273	287
Indexes	52	54
Marginalia	4740	5000
Corrections	42	44

The following table shows the full results of this calculation:

The following table shows how many manual corrections are to be done:

	Estimated number of titles to be corrected	Estimated number with added margin
Titles to be corrected manually	24600	26000

Another method is used to calculate the sections that occur only once in the whole monograph, e.g. the table of contents, the title section, the preface etc...

The set for calculating the statistic was composed of 53 monographs. If we take the example of table of contents, in the 53 monographs there are in total 24 tables of contents. To get an average we divide the total number of monographs with the total number of tables of contents:

[total number of monograph in the set] / [total number of table of contents] = [average number of table of contents per monograph]

With the [average number of table of contents per monograph] and the total number of monographs to be digitized we can calculate the approximate number of tables of contents present:

[average number of table of contents per monograph] * [total count of monographs to be digitized] = [approximate number of tables of contents]

Sections	Estimated number of	Estimated number with
	Sections to be detected	
Prefaces	80	84
Table of contents	215	225
Bastard title sections	70	74
Title sections	425	445
Frontispieces	35	37
Bibliographies	27	29
Bookplates	35	37

9 METS and ALTO files in detail

A detailed description of METS and ALTO (structure, required elements, default values ...) can be found in Appendix B.

10 Quality reports and control

Upon delivery of data by the service provider, the following three quality reports must always be provided with the discs.

A delivery may be refused after checking for one or more of the following reasons:

- Non complete collection An incomplete collection cannot be accepted. A delivery that contains a portion of a monograph that would be too large to be delivered whole.
- Non-compliant data If elements of this specification are not met in the data, the delivery is refused.
- Quality level of corrected titles or structure are not reached If in the samples describes in Chapter 10.3 and 10.4 the quality level is not reached for one of the two elements (or both), the delivery is refused.

10.1 Completeness of the collection

An Excel inventory file will be delivered to the service provider upon delivery of physical monographs for digitization. This inventory was created by the BnL to facilitate the work of the service provider.

If differences are found between the inventory provided by the BnL and the paper version, the service provider must return a corrected inventory to the BnL. If no difference is found, an explanatory note saying "Correct inventory" must go with the delivery.

10.2 Data integrity

A check should be performed by the service provider to ensure that the technical points required in the specifications are met 100%. These include:

- Validation of XML files by their respective schemas
- Validation of the TYPE tag from the logical structure with the schema bnl-monograph_v1.0.xsd
- The naming rules for files
- The naming rules for the different values automatically inserted into the METS file
- The values to be inserted for each monograph in the METS file
- The validity of image files (TIFF and PNG)
- The validity of PDF files

A specific quality report of data integrity is not required, but a description of the process to accomplish these controls is required.

10.3 Quality control of the manual corrected titles

The quality level for manual corrections of titles and captions of images is 99.8%. For simplicity, in the rest of this chapter, the term "title" will be used for the following elements: "titles and images, lists, tables, maps, bibliographies, indexes, table of contents and author names". These are all the items listed in section 7.3.2. The service provider has to check by sampling a subset of the titles in order to verify the obtained level of quality. The sample must be done on all titles, regardless of the monograph in which they are located and sections to which they belong. The sample should be at least two titles per thousand titles that were to be corrected (2 %).

A quality report is to be delivered listing the checked titles, the number of characters in each of these titles as well as the number of erroneous characters in this title.

The method used to calculate the final percentage is as follows:

- 1. Calculation of the total number of characters such as the title appears in the LABEL section in the logical structure. Spaces are included. This number is ORIG.
- 2. Calculation of the number of insertion of a character, deletion of a character or replacement of a character with another, needed to get the title LABEL to fully correct title. This number is ERRORS.
- 3. If ERRORS is greater than ORIG, then we put ERRORS equals to ORIG.
- 4. Calculate (ORIG ERRORS) / ORIG * 100.

Example: Title of a chapter:

Titel IV. - Berfchiebene Beftimmungen.

Text entered by the operator in the LABEL logical structure: Litle IV. - Berswiedene BeAimmungen.
So the number of ORIG characters is 36.
The corrected text should be: Title IV. - Verschiedene Bestimmungen. 38

To get from the first title to the corrected one, it is necessary to:

- 1. Replace "L" by "T" for the first letter
- 2. Replace "B" by "V" for the 13th letter
- 3. Replace "w" by "c" for the 17th letter
- 4. Add "h" after the 17th letter
- 5. Replace "A" by "s" for the 27th letter
- 6. Add "t" after the 27th letter

So the number of errors is 6. Finally, the percentage of accuracy of this individual title is (36 - 6) / 36 * 100 = 83.3 %

For all the checked titles in a sample, the total should be first on the total number of characters in all the checked titles and the total number of errors.

The format of this quality report should be an Excel or OpenOffice.org Calc file that contains all the information required to trace back the calculation. For example:

Title	ORIG	ERRORS	Image of title
Acht	14	1	9 11 Nuestatan
Dredigten			Agt preotyten
gehaten	7	1	astrattan
in der Fest-	54		gegutten
Oktav zu			in der Best-detan en Ehren der Trösterin der Betrühten
Ehren der			the work channes an challend and have a set
Trösterin der			in der
Betrübten			Bathebrole zu Luremburg 1878
in der	6	0	
Rathedrale	15	15 (16)	
1878			

Total	96	17
Percentage		78.29%

In the case "Rathedrale 1878", we can see that ERRORS is greater that ORIG. There are 15 characters missing " zu Luxembourg " (spaces count as well) and the first letter "R" which should be a "K".

10.4 Segmentation quality control

The segmentation into sections must be controlled by sampling and a quality report must be delivered. The verification of the structure must be made on a subset consisting of complete monographs. The sample size must be at least 5 monographs per thousand monographs livered (5 ‰). For simplicity, in the rest of this chapter, the word "chapter" will be used for chapters, illustrations, pictures, advertisements, maps, table of contents, bibliographies, and indexes. These are the elements to detect as described in Chapters 6.2 and 6.3.

The quality report must contain all the elements needed to track the calculation of the score. The format of the report must contain at least the identifiers of considered monographs and the subtotals by monographs of all the error types.

The quality score of segmentation is divided into 3 parts:

- a) The blocking errors
- b) The minor errors
- c) The major errors

In the sample, no blocking error must be found. Categories b) and c) have a maximum with respect to a "standard book" of 160 pages. The maximum will be proportionally recalculated if an edition contains more than 160 pages. The table below shows these maximums:

Page numbers	Maximum major errors	Maximum minor errors
1-32	1	4
33-64	2	8
65-96	3	12
97-128	4	16
129-160	5	20

It should be noted that a single error of an operator may very well result in several different errors.

10.4.1 Blocking errors

Blocking errors are the most serious and if such an error is found in the sample, the delivery will be rejected. There are 2 cases of blocking errors:



- 1. Part of a page is not segmented A portion of the page with chapters, images, advertisements or any other element, is not structured. This does not apply for the running title.
- 2. Wrong page order If the pages are in the wrong order, all chapters that continue between the pages are affected and the table of contents is in the wrong order.

10.4.2 Major errors

Major errors are errors that have a large impact on the consistency of the chapters. There are 2 cases of major errors:

1. A chapter has extra blocks

This error occurs when a page block (text or picture) is added to a element with which it has no relationship. This is classified as a major error for 2 reasons. The first is that the user can cut out a chapter, based on the structural information. In this case, there will be blocks without relationship or the system will deliver text blocks the user doesn't want. The second reason is that an automated analysis of the content will encounter problems in classifying the titles of chapters. NB: This error does not apply to advertisements which have extra blocks.

- Chapters not segmented into sub-chapters
 If there are further titles inside of the chapter and they are not
 recognized as such, an important loss of structural information occurs.
- 3. Wrong classification

This happens when the wrong tag is used to segment a part of a page. Errors like a CHAPTER, which has been classified wrongly as a TABLE_OF_CONTENTS, or a MAP, which has been classified as a TABLE, enter into this category.

10.4.3 Minor errors

There are 7 categories of minor errors. These are errors that do not have a major impact on the consistency of the structure but, if they are numerous, can confuse the reader.

- 1. A chapter has missing blocks
- 2. Wrong title / caption

The wrong title is attached to a chapter or the wrong caption is attached to an illustration, map or table. This is independent of the manual correction of the titles which is checked separately.


3. Missing chapter title / caption

A chapter that has a title, respectively an illustration or a table that has a caption on paper but not in the structure.

4. Chapter at wrong level

A sub-chapter that should be part of a hierarchy is all alone, or a subchapter that should be alone is attached to a chapter with which it has no relationship.

5. Chapter not recognized

A title is not recognized as such and therefore the paragraphs that follow the chapter are not recognized. This also applies to illustrations, tables, footnotes, bibliographies, table of contents and indexed that are not recognized as such.

- 6. Advertisement has extra blocks Unlike chapters, if an advertisement has extra blocks, this is recognized as a minor error.
- 7. Wrong order of chapters

The reading direction of a monograph is from top left to bottom right. If this reading direction is not respected, this error occurs. An instance of error is counted for each pair of items that must be exchanged in order to find the correct reading order.

11 Inventory excel sheet

The following table shows all the fields that are in the inventory sheet:

Tag	Description of tag
BOX-ID	ID of the box containing the monograph
[BIBREC_SYS_NUM]	System number of the bibliographic record
[ITEM_barcode]	Barcode of the physical document
[BIBREC_CALL_NUM]	Call number of the physical document
[languageTerm]	Language of the document
[BIBREC_100a-1]	First name: Personal name
[BIBREC_100a-2]	Family name: Personal name (NR)
[BIBREC_245a]	The full title of the work
[BIBREC_245b]	Subtitle of the work
[BIBREC_260b]	Name of publisher, distributor, etc.
[BIBREC_260c]	Date of publication, distribution, etc.
[BIBREC_300a]	Number of pages (from the bibliographic record)
[BIBREC_300a_ref]	Reformatted number of pages
[BIBREC_300c]	Physical dimensions of the monograph

Appendix A

Author: zarko.stojkovic@bnl.etat.lu Version: 1 Date: 03/12/14

Appendix A: File locations and naming 112

Appendix A: File locations and naming

We suppose that for every monograph, files (tif, alto, pdf and png files) are contained in the following locations

(See Chapter 11 from "Technical requirements")

[BIBREC_SYS_NUM]/ (containing 1 pdf file and 1 mets file for the whole monograph) [BIBREC_SYS_NUM]/tif (containing 1 tif file for every page of the monograph) [BIBREC_SYS_NUM]/alto (containing 1 alto file for every page of the monograph) [BIBREC_SYS_NUM]/pdf (containing 1 pdf file for every page of the monograph) [BIBREC_SYS_NUM]/png (containing 1 png file for every page of the monograph)

[BIBREC_SYS_NUM] corresponds to the system number in the ILS of the electronic version of the digitized book that will be included in the excel file delivered by the BnL.

Example system number: "000239390"

The files contained in the directories must have filenames corresponding to the following pattern:

tif files: [ddddd].tif alto files: [ddddd].xml single page pdf files: [ddddd].pdf png files: [ddddd].png

where [ddddd] corresponds to a 5 digit page counter (filled with zeros) starting at 1 for every issue.

The METS file must have the filename [BIBREC_SYS_NUM]-mets.xml and be contained in the root folder for the monograph.

The "complete monograph"-.pdf file must have the filename [BIBREC_SYS_NUM].pdf and be contained in the root folder for the monograph.

Example:

Suppose the monograph "Heinrich, Prinz der Niederlande, Statthalter S.M. des Königs-Grossherzogs im Grossherzogtum Luxemburg" with [BIBREC_SYS_NUM] = "000239390". Directories and files would then have to be as follows:

/000239390/ 000239390-mets.xml 000239390.pdf /000239390/tif/ 00001.tif 00002.tif 00003.tif 00004.tif /000239390/alto/ 00001-alto.xml 00002-alto.xml 00003-alto.xml 00004-alto.xml /000239390/pdf/ 00001.pdf 00002.pdf 00003.pdf 00004.pdf /000239390/png/ 00001.png 00002.png 00003.png 00004.png

Page intentionally left blank

Appendix B

METS & ALTO Requirements Specification

Author: zarko.stojkovic@bnl.etat.lu Version: 1 Date: 03/12/14

Table of contents

G N D 1	eneral otes c ownloa Gei	once ads nera	erning XML (METS) code included in this document	18 18 18 20
Z	<m 2.1</m 	ets> Ref	: METS Root element	20 20
	2.2	Rec	juirements	21
	2.3	Con	nplete <mets> root element example:1</mets>	21
3	<m 3.1</m 	ets⊦ Ref	Idr>: METS Header element1 erence document (from METS documentation)1	22 22
	3.2	Rec	juirements	22
	3.3	Con	nplete <metshdr> example:</metshdr>	23
4	<dr 4.1</dr 	ndSe Ref	ec>: Descriptive Metadata Section1 erence documentation1	24 24
	4.1	.1	from METS documentation	24
	4.1	.2	from MODS documentation	24
	4.2	Rec	juirements1	27
	4.2	.1	Requirements for all <dmdsec> elements</dmdsec>	27
	4.2	.2	Requirements for 1 st <dmdsec> element</dmdsec>	28
	4.2	.3	Requirements for 2 nd <dmdsec> element</dmdsec>	30
	4.2	.4	Requirements for the 3rd <dmdsec> elements</dmdsec>	31
	4.2	.5	Other <dmdsec> elements1</dmdsec>	31
	4	1.2.5	.1 <dmdsec> elements for chapters1</dmdsec>	31
	4 C	1.2.5 hart	5.2 <dmdsec> elements in monographs: illustrations, maps, tables, cs, diagrams</dmdsec>	34
5 <amd 5.1 R</amd 		ndSe Ref	ec>: Administrative Metadata Section1 erence documentation1	36 36
	5.1	.1	from METS documentation1	36
	5.1	.2	from MIX documentation	37
	5.2	NIS	O Data Dictionary: Technical Metadata for Digital Still Images 14	43
	5.3	Tag	explanations	44
6	The 6.1	e <fi Ref</fi 	leSec> elelemt	47 47
6.2 Tag explanation		explanation	49	
7	The 7.1	e <st Ref</st 	ructMap> elements	53 53
	7.2	[PH	YSICAL_STRUCTMAP]1	56

	7.3	[LOGICAL_STRUCTMAP]	159
	7.3	1 Introduction	159
	7.3	2 Requirements	159
	7	3.2.1 General requirements	159
7.3.2		3.2.2 Referencing descriptive metadata	160
	7	3.2.3 LABEL attribute	160
8	Gei	eral requirements	162
9	Doc	umentation: ALTO files	162
	9.1	Reference documentation	162
9.2 Structure of ALTO Files		162	
	9.3	9.3 MeasurementUnit	
9.4 TextStyles		163	
	9.5 ParagraphStyles		164
	9.6 Attributes of a Page Element		164
9.7 Page Areas		165	
9.8 The structure of each of the Page Area (PageSpace) Elements		166	
	9.9	Additional Attributes of the Tags	167
1(0 Ref	erencing ALTO files from METS	168
1	1 ALTO Example		
12	2 File batch.xml: requirements		

General

The digitized collection of the monographs is described and structured by a set of METS and ALTO documents. Because METS is a very flexible XML schema, this document describes further requirements which must be followed by the METS and ALTO documents describing the collection.

Notes concerning XML (METS) code included in this document

- [BOLD tags] are to be filled by the service provider as described
- Values in *Italic* are be included as they appear in this document ("hard-coded values")

Downloads

The following files are required and available for download at: <a href="http://downloads.bnl.lu/schemas/<FILENAME">http://downloads.bnl.lu/schemas/<FILENAME>

FILENAME	Description
mets_profile_bnl_monograph.xml	METS profile
mets.xsd	METS Schema (Metadata encoding & transmission standard)
	Documentation: <u>http://www.loc.gov/standards/mets</u>
mix.xsd	MIX Schema (Metadata for still images
	Documentation: http://www.loc.gov/standards/mix
mods.xsd	MODS Schema (Metadata object description schema)
	Documentation: <u>http://www.loc.gov/standards/mods/</u>
bnl-monograph_v1.0.xsd	Important: This schema describes the structure of the logical
	structMap of the METS file for the monographs.
alto.xsd	ALTO Schema
MARC21slim_bnl.xsd	MARC XML Schema, slightly adapted for BnL need with
	relaxed conditions.
samples.zip	Example of a METS file with ALTO files for a different project
	with a different logical structMap (newspaper)



Section one - METS files

1 General requirements

One METS file must exist for every monograph. The filename for the METS file must be [BIBREC_SYS_NUM]-METS.xml where [BIBREC_SYS_NUM] is the system number of the bibliographic record of the monograph.

Each monograph has its own title. In remainder of this document, we'll use these placeholders:

Tag	Description of tag
[BIBREC_245a]	The full title of the work
[BIBREC_245b]	Subtitle of the work
[BIBREC_100a-1]	First name: Personal name
[BIBREC_100a-2]	Family name: Personal name (NR)
[BIBREC_260b]	Name of publisher, distributor, etc.
[BIBREC_260c]	Date of publication, distribution, etc.
[languageTerm]	Language of the monograph
[ITEM_barcode]	Barcode of the monograph
[BIBREC_SYS_NUM]	System number of the bibliographic record

2 <mets>: METS Root element

2.1 Reference documentation (from METS documentation)

<u>Note:</u> External documentation is included in this document for reference. Please note that some of the information appearing in those descriptions might not be relevant (attributes that we don't require etc...)

Element	Description of element and attributes
<mets></mets>	METS: Metadata Encoding and Transmission Standard.
	METS is intended to provide a standardized XML format for transmission of complex digital library objects between systems. As such, it can be seen as filling a role similar to that defined for the Submission Information Package (SIP), Archival Information Package (AIP) and Dissemination Information Package (DIP) in the Reference Model for an Open Archival Information System.
	A METS document consists of seven possible subsidiary sections: metsHdr (METS document header), dmdSec (descriptive metadata section), amdSec (administrative metadata section), fileGrp (file inventory group), structLink (structural map linking), structMap (structural map) and behaviorSec (behaviors section). It also has five possible attributes: 1. ID (an XML ID):
	 OBJID: a primary identifier assigned to the METS document; LABEL: a title/text string identifying the document for users; TYPE: a type for the object, e.g., book, journal, stereograph, etc.; and PROFILE: the registered profile to which this METS document conforms. METS registry information is available from the Library of Congress at http://www.loc.gov/mets.



2.2 Requirements

The METS Root element must have the following structure:

```
<mets xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/METS/"
xsi:schemaLocation="http://www.loc.gov/METS/
http://downloads.bnl.lu/schemas/mets.xsd http://www.loc.gov/MARC21/slim
http://downloads.bnl.lu/schemas/MARC21slim_bnl.xsd"
xmlns:MODS="http://www.loc.gov/mods/v3"
xmlns:mix="http://www.loc.gov/mix/"
xmlns:xlink="http://www.loc.gov/MARC21/slim"
xmlns:MARC="http://www.loc.gov/MARC21/slim"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
TYPE="Monograph"
LABEL="[BIBREC 245a]">
```

2.3 Complete <mets> root element example:

```
<mets xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xmlns="http://www.loc.gov/METS/"

xsi:schemaLocation="http://www.loc.gov/METS/

http://downloads.bnl.lu/schemas/mets.xsd http://www.loc.gov/MARC21/slim

http://downloads.bnl.lu/schemas/MARC21slim_bnl.xsd "

xmlns:MODS="http://www.loc.gov/mods/v3"

xmlns:mix="http://www.loc.gov/mix/"

xmlns:xlink="http://www.loc.gov/mix/"

xmlns:xlink="http://www.loc.gov/MARC21/slim"

xmlns:MARC="http://www.loc.gov/MARC21/slim"

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

TYPE="Monograph"

LABEL="Making Cheese">
```

3 <metsHdr>: METS Header element

3.1 Reference document (from METS documentation)

<u>Note:</u> External documentation is included in this document for reference. Please note that some of the information appearing in those descriptions might not be relevant (attributes that we don't require etc...)

Element	Description of element and attributes
<metshdr></metshdr>	metsHdr: METS Header.
	Like a TEI Header, the METS Header element records metadata about the METS document itself (not the digital library object that the METS document encodes). It has two possible subsidiary elements, agent (document agent) and altRecordID. (alternative Record ID). It also has the following four attributes: 1. ID (an XML ID); 2. CREATEDATE: the date/time the METS document was created; 3. LASTMODDATE: the date/time the METS document was last modified; 4. RECORDSTATUS: a string indicating the status of the METS document, to be used mainly for internal processing purposes.
<agent></agent>	agent: METS agent.
	 The agent element allows for various parties and their roles with respect to the METS document to be recorded. It has five attributes: ID (an XML ID); ROLE: one of 7 set roles with respect to the document, CREATOR, EDITOR, ARCHIVIST, PRESERVATION, DISSEMINATOR, CUSTODIAN and IPOWNER or the value OTHER to indicate a non-set role; OTHERROLE: a string to specify a non-set role if OTHER is indicated in the ROLE attribute; TYPE: either the set values of INDIVIDUAL agent or ORGANIZATION, or the value OTHER to indicate a non-set value; and OTHERTYPE: a string to indicate the type of agent if a value of OTHER is indicated in the TYPE attribute.

3.2 Requirements

The first element in every METS file is <metsHdr> element.

It must have the following structure:

```
<metsHdr CREATEDATE="[CREATE_DATE]" LASTMODDATE="[LASTMOD_DATE]">
        <agent ROLE="CREATOR" [TYPE]>
        <name>[NAME]</name>
        </agent>
        </metsHdr>
```

Tag description (values must be inserted by supplier)

Tag	Description of tag
[CREATE_DATE]	The date/time the METS document was created;
	Format: YYYY-MM-DDTHH:MM:SS
[LASTMOD_DATE]	The date/time the METS document was modified;
	Format: YYYY-MM-DDTHH:MM:SS
[TYPE]	We assume that the METS files are created by software. If this is the case, replace [TYPE] by:
	TYPE="OTHER" OTHERTYPE="SOFTWARE"
[NAME]	The name of the creator agent (the name of the software)

3.3 Complete <metsHdr> example:

4 <dmdSec>: Descriptive Metadata Section

4.1 Reference documentation ...

<u>Note:</u> External documentation is included in this document for reference. Please note that some of the information appearing in those descriptions might not be relevant (attributes that we don't require etc...)

4.1.1 ... from METS documentation

Element	Description of element attributes
<dmdsec></dmdsec>	dmdSec: Description Metadata Section. Section records all of the descriptive metadata for all items in the METS object (including both structural map divs and descriptive metadata for data files).
	Can be either included in the METS hub document (mdWrap) or via an identifier/locator (mdRef), a la Warwick Framework. Multiple elements are allowed so that descriptive metadata be recorded for each separate item within the METS object.
<mdwrap></mdwrap>	 mdWrap: metadata wrapper. The mdWrap element is a generic element used throughout the METS schema to allow the encoder to place arbitrary metadata conforming to other standards/schema within a METS document. The included metadata can either be encoded in XML, in which case it may be placed directly within the mdWrap element, or it can be Base64 encoded, and placed within a subsidiary binData element. The mdWrap element can have the following attributes: ID: an XML ID for this element MIMETYPE: the MIME type for the metadata contained in the element MDType: the type of metadata contained (e.g., MODS) OTHERMDTYPE: a string indicating an alternative MDTYPE when the MDTYPE value is set to "OTHER." LABEL: a label to display to the viewer of the METS document identifying the metadata.
<xmldata></xmldata>	A wrapper to contain Base64 encoded metadata

4.1.2 ... from MODS documentation

Element	Description of element and attributres
<mods></mods>	MODS root element
<titleinfo></titleinfo>	Definition: A word, phrase, character, or group of characters, normally appearing in a resource, that names it or the work contained in it.
<language></language>	Definition: a designation of the language in which the content of a resource is expressed.

-

<languageterm></languageterm>	 "languageTerm" contains the language(s) of the content of the resource. It may be expressed in textual or coded form. If in coded form, the source of the code is contained in the value of the authority attribute. If there is more than one representation of the same language, <languageterm> is repeated. If the content of the resource is in more than one language, <language> is repeated.</language></languageterm> Attributes: type - This attribute may be used with the following values: <i>text</i> - This value is used to express language in a textual form. <i>code</i>- This value is used to express language in a coded form. authority - This attribute may contain the following values: <i>iso639-2b</i> - A bibliographic language code from ISO 639-2 (Codes for the representation of names of languages. <i>rfc3066</i> - A language identifier as specified by the Internet Best Current Practice specification RFC3066. This value includes an ISO 639-1 (alpha-2) or ISO 639-2 (alpha-3) language code with optional secondary subtags.
<identifier></identifier>	Definition: a unique standard number or code that distinctively identifies a resource.
<origininfo></origininfo>	Definition: Information about the origin of the resource, including place of origin or publication, publisher/originator, and dates associated with the resource
dateCreated	Definition: the date of creation of the resource
<datelssued></datelssued>	Definition: the date that the resource was published, released or issued.
	<i>Date Attributes.</i> Certain date attributes may be applied to some MODS elements, as indicated in the schema. They are defined below:
	• encoding - The following values are used with the encoding
	 w3cdtf - This value is used for the profile of ISO 8601 that specifies the following date pattern: YYYY-MM-DD.
	 point - If no point attribute is specified, date is assumed to be a single date. The following values are used with the point attribute: start - This value is used for the first date of a range (or a single date, if used). end - This value is used for the end date of a range.
	 keyDate - The following value is used with the keyDate attribute: yes - This value is used so that a particular date may be distinguished among several dates. Thus for example, when sorting MODS records by date, a date with keyDate="yes" would be the date to sort on. It should occur only for one date at most in a given record.
	attribute:

	 approximate - This value is used to identify a date that may not be exact, but is approximated, such as "ca. 1972". inferred - This value is used to identify a date that has not been transcribed directly from a resource, such as "[not before 1852]". questionable - This value is used to identify a questionable date for a resource, such as "1972?".
<title></title>	Definition: A word, phrase, character, or group of characters that constitutes the chief title of a resource (i.e. the title normally used when citing the resource).
<subtitle></subtitle>	Definition: A word, phrase, character, or group of characters that contains the remainder of the title information after the title proper
<partnumber></partnumber>	Definition: "partNumber" is used for a part or section number of a title
<name></name>	 Definition: The name of a person, organization, or event (conference, meeting, etc.) associated in some way with the resource. Attribute type - This attribute indicates the type of name. The following values may be used with it: personal corporate conference
<namepart></namepart>	"namePart" includes each part of the name that is parsed Attribute • type - The following values may be used with the type attribute: • date • family • given • termsOfAddress
<role></role>	Definition: A term(s) that designates the relationship (role) of the entity recorded in name in relation to the resource. Application: "role" is a wrapper element that may contain a value in coded or textual form.
<roleterm></roleterm>	"roleTerm" contains the textual or coded form of a relator/role.

4.2 Requirements

4.2.1 Requirements for all <dmdSec> elements

- The descriptive metadata is to be included in the METS hub document (via mdWrap NOT mdRef see reference for <dmdSec> above).
- The <mdWrap> element must contain the metadata in *MODS* form (Metadata Object Description Schema v.3). Thus, the <mdWrap> element will have the attribute MDTYPE="MODS"
- To guarantee that the MODS data included is valid, the content of every <mdWrap>'s <xmlData> element must be validated against the schema: http://downloads.bnl.lu/schemas/mods.xsd

Note: About MODS

The Library of Congress' Network Development and MARC Standards Office, with interested experts, has developed a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. As an XML schema, the "Metadata Object Description Schema" (MODS) is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records.

4.2.2 Requirements for 1st <dmdSec> element

The first <dmdSec> element has ID = "MODSMD_PRINT" (bibliographic metadata of the printed version).

```
<dmdSec ID="MODSMD PRINT">
  <mdWrap MIMETYPE="text/xml" MDTYPE="MODS" LABEL="Bibliographic meta-data
of the printed version">
    <xmlData>
       <MODS:mods>
         <MODS:titleInfo ID="MODSMD PRINT TI1" xml:lang="gr">
            <MODS:title>[BIBREC 245a]</MODS:title>
            <MODS:subTitle>[BIBREC 245b]</MODS:subTitle>
         </MODS:titleInfo>
         <MODS:name ID="MODSMD PRINT N1" type="personal">
            <MODS:namePart type="given">[BIBREC 100a-1]</MODS:namePart>
            <MODS:namePart type="family">[BIBREC 100a-2]</MODS:namePart>
            <MODS:role>
              <MODS:roleTerm>author</MODS:roleTerm>
            </MODS:role>
         </MODS:name>
         <MODS:language>
            <MODS:languageTerm type="code"
authority="rfc3066">[languageTerm]</MODS:languageTerm>
         </MODS:language>
         <MODS:originInfo>
            <MODS:publisher>[BIBREC 260b]</MODS:publisher>
            <MODS:dateIssued point="start" keyDate="yes"
qualifier="approximate">[BIBREC 260c]</MODS:dateIssued>
         </MODS:originInfo>
         <MODS:identifier type="local">[ITEM barcode]</MODS:identifier>
       </MODS:mods>
    </xmlData>
  </md\rap>
</dmdSec>
```

<u>Note:</u> *languageTerm* is the language present in the printed version of the monograph. The possible values for languageTerm are:

- "de" for German
- "fr" for French
- "la" for Latin
- "lb" for Luxemburgish
- "en" for English
- "nl" for Dutch
- "mul" for multiple languages

If a "mul" value is found, the service provider must detect all the languages from the monograph and create one languageTerm entry per language detected. Example: Heinrich, Prinz der Niederlande, Statthalter S.M. des Königs-Grossherzogs im Grossherzogtum Luxemburg : eine biographische Skizze / 1879

Sys:000239390

```
<dmdSec ID="MODSMD PRINT">
  <mdWrap MIMETYPE="text/xml" MDTYPE="MODS" LABEL="Bibliographic meta-data
of the printed version">
  <xmlData>
    <MODS:mods>
       <MODS:titleInfo ID="MODSMD PRINT TI1" xml:lang="gr">
         <MODS:title>Heinrich, Prinz der Niederlande, Statthalter S.M. des
Königs-Grossherzogs im Grossherzogtum Luxemburg</MODS:title>
         <MODS:subTitle>eine biographische Skizze</MODS:subTitle>
       </MODS:titleInfo>
       <MODS:name ID="MODSMD PRINT N1" type="personal">
         <MODS:namePart type="given">Jean-Venceslas-
Charles</MODS:namePart>
         <MODS:namePart type="family">Arendt</MODS:namePart>
         <MODS:role>
            <MODS:roleTerm>author</MODS:roleTerm>
         </MODS:role>
       </MODS:name>
       <MODS:language>
         <MODS:languageTerm type="code"
authority="rfc3066">gr</MODS:languageTerm>
       </MODS:language>
       <MODS:originInfo>
         <MODS:publisher>V. Bück</MODS:publisher>
         <MODS:dateIssued point="start" keyDate="yes"
qualifier="approximate">1879</MODS:dateIssued>
       </MODS:originInfo>
       <MODS:identifier type="local">KM 0652922</MODS:identifier>
    </MODS:mods>
  </xmlData>
  </mdWrap>
</dmdSec>
```



4.2.3 Requirements for 2nd <dmdSec> element

The second <dmdSec> element has ID = "MODSMD_ELEC" (bibliographic metadata of the electronic version)

Depending on the monograph, the structure could look like this:

```
<dmdSec ID="MODSMD ELEC">
  <mdWrap MIMETYPE="text/xml" MDTYPE="MODS" LABEL="Bibliographic meta-data
of the electronic version">
    <xmlData>
       <MODS:mods>
         <MODS:originInfo>
           <MODS:dateCreated point="start" keyDate="yes"
qualifier="approximate">[DATE CREATED]</MODS:dateCreated>
            <MODS:publisher>Bibliothèque nationale de
Luxembourg</MODS:publisher>
         </MODS:originInfo>
         <MODS:titleInfo ID="MODSMD ELEC TI1" xml:lang="gr">
            <MODS:title>[BIBREC 245a] </MODS:title>
            <MODS:subTitle>[BIBREC 245b]</MODS:subTitle>
         </MODS:titleInfo>
         <MODS:name ID="MODSMD ELEC N1" type="personal">
            <MODS:namePart type="given">[BIBREC 100a-1]</MODS:namePart>
            <MODS:namePart type="family">[BIBREC 100a-2]</MODS:namePart>
            <MODS:role>
              <MODS:roleTerm>author</MODS:roleTerm>
            </MODS:role>
         </MODS:name>
         <MODS:language>
            <MODS:languageTerm type="code"
authority="rfc3066">[languageTerm]</MODS:languageTerm>
         </MODS:language>
         <MODS:identifier type="local">[BIBREC SYS NUM]</MODS:identifier>
       </MODS:mods>
    </xmlData>
  </mdWrap>
</dmdSec>
```

Tag description (values to be inserted by supplier)

4.2.4 Requirements for the 3rd <dmdSec> elements

The 3rd <dmdSec> element has ID = "MARCMD_ALEPHSYNC" (bibliographic metadata of the electronic version).

Depending on the monograph, the structure could look like this:

```
<dmdSec ID="MARCMD_ALEPHSYNC">
	<mdWrap MIMETYPE="text/xml" MDTYPE="MARC" LABEL="Bibliographic meta-data
	synched from Aleph">
	<xmlData>
	<MARC:record>
	<MARC:controlfield tag="001">
		[BIBREC_SYS_NUM]
		</MARC:controlfield>
	</MARC:record>
	</xmlData>
	</mdWrap>
</dmdSec>
```

4.2.5 Other <dmdSec> elements

The 2 first <dmdSec> elements have been described above. After these 2 elements come a number of <dmdSec> elements with structure and content dependent on the structure of the printed monograph.

Every chapter, illustration, map, table, chart and diagram must be described by an individual <dmdSec> element. The author(s) must be included in the dmdSec for the corresponding element.

<u>Note:</u> Please refer to *Technical note* (chapter 6) for a definition of layout analysis terms like *chapter*, *illustration*, *table*, etc.

4.2.5.1 <dmdSec> elements for chapters

For every chapter in the monograph, a <dmdSec> element with attribute ID = "MODSMD_CHAPTER[d]" ([d] being a counter starting at "1" in every METS file and increased by one with every chapter described) is to be added to the METS file.

This element must have the following structure:

```
<dmdSec ID="[DMD_SEC_ID]">
  <mdWrap MIMETYPE="text/xml" MDTYPE="MODS" LABEL="Bibliographic meta-data
of chapter">
  <xmlData>
    <MODS:mods>
    [TITLE_INFO]
    [AUTHOR_INFO]
    [CHAPTER_LANGUAGE]
    </MODS:mods>
    </xmlData>
    </mdWrap>
</dmdSec>
```

Tag description (values to be inserted by supplier)

Tag	Description of tag
[DMD_SEC_ID]	Value must be "MODSMD_CHAPTER[d]" where [d] is a
	counter starting for every METS file at 1, and
	incremented by one with every chapter described
	within the current monograph. E.g., the 10th chapter
	discovered in a given monograph will have
	[DMD_SEC_ID] = "MODSMD_CHAPTER10"
[TITLE_INFO]	[TITLE_INFO] must be replaced by, depending on the type and number of printed headlines for this chapter, one or more <mods:titleinfo> elements.</mods:titleinfo>
	 For a title composed of a single headline, replace [TITLE_INFO] by a single <mods:titleinfo> element:</mods:titleinfo>
	<mods:titleinfo <br="" id="[TITLE_INFO_ID]">xml:lang="[TITLE_LANGUAGE]"> [HEADLINE] </mods:titleinfo>
	 For a title composed of a main headline and a sub- headline, replace [TITLE_INFO] by a single <mods:titleinfo> element:</mods:titleinfo>
	<mods:titleinfo <br="" id="[TITLE_INFO_ID]">xml:lang="[TITLE_LANGUAGE]"> [HEADLINE] [SUB_HEADLINE] </mods:titleinfo>
	Value must be "[DMD_SEC_ID1_TI[d1" where [d1 is a
	<pre>counter starting at 1 and incremented with every <mods:titleinfo> element created for the current chapter</mods:titleinfo></pre>

	E.g., the second <mods:titleinfo> element of the 10th</mods:titleinfo>	
	chapter will have [TITLE_INFO_ID] =	
	"MODSMD_CHAPTER10_TI2"	
[TITLE_LANGUAGE]	Replace by the language code for the language of the	
	title. German: "de", French: "fr", Luxembourgish:	
	"lb", English: "en".	
[HEADLINE]	<mods:title> element holding headline of the article</mods:title>	
	e.g.: <mods:title> Heinrich, Prinz der</mods:title>	
	Niederlande, Statthalter S.M. des Königs-	
	Grossherzogs im Grossherzogtum Luxemburg	
	Important: If the headline starts with an article like	
	"Der" "Die" "Das" "Le" "La" "Les" "Un" "Une"	
	"Des" "The" etc. then $[HE\Delta D]$ INF] must be replaced	
	by a <mods:nonsort> element, as well as the</mods:nonsort>	
	<mods:title> element</mods:title>	
	e.g.: <mods:nonsort>Die</mods:nonsort>	
	<mods:title> Die Besitzungen des Priorates</mods:title>	
	Marienthal während des ersten Jahrhunderts	
	<pre>seines Bestehens</pre>	
[SUB_HEADLINE]	<pre><mods:subtitle> element holding sub-headline of the</mods:subtitle></pre>	
	article	
	e.g. : <mods:subl'itle> eine biographische Skizze</mods:subl'itle>	
	Important: If the sub-headline starts with an article like	
	"Der". "Die". "Das". "Le". "La". "Les". "Un". "Une".	
	"Des", "The", etc, then [SUB_HEADLINE] must be	
	replaced by a <mods:nonsort> element, as well as the</mods:nonsort>	
	<mods:title> element</mods:title>	
[AUTHOR_INFO]	Optional - Must be included if the author of an article is	
	determinable	
	<mods:name> element holding the article's owner.</mods:name>	
	Must have the following structures	
	MUST have the following structure:	
	<pre><mods:namepart type="given">[AUTHOR FIRST</mods:namepart></pre>	
	NAME]	
	<mods:namepart type="family"></mods:namepart>	
	[AUTHOR_SECOND_NAME]	
	<pre><mods:roleterm>aut</mods:roleterm></pre>	

	<pre>where [AUTHOR_INFO_ID] is to be replaced by "[DMD_SEC_ID]_N[d]", [d] being a counter starting at 1 and incremented with every <mods:name> element created for the current chapter (in case there are more authors) E.g. for the 12th chapter in a monograph, the <mods:name> element for the author of the chapter will have ID= "MODSMD_CHAPTER12_N1". If there is a second author, it will have ID = "MODSMD_CHAPTER12_N2".</mods:name></mods:name></pre>
[CHAPTER_LANGUAGE]	<pre>Replace by: <mods:language><mods:languageterm type="code" authority="rfc3066">[lang_code]uageTerm> </mods:languageterm </mods:language></pre>
	where [lang_code] is the language code for the language appearing in the text. German: "de", French: "fr", Luxembourgish: "lb", English: "en", Latin: "la", Dutch: "nl". If more than one different language appears in the article, then [CHAPTER_LANGUAGE] must be replaced by one <mods:language> element for each of them</mods:language>

4.2.5.2 <dmdSec> elements in monographs: illustrations, maps, tables, charts, diagrams

For each of these elements in the monograph, if it has caption and/or and author a <dmdSec> element with the following attributes IDs must be added to the METS file:

Element	[DMD_SEC_ID]	[LABEL]
illustration	MODSMD_PICT[d]	Bibliographic meta-data of picture
map	MODSMD_MAP[d]	Bibliographic meta-data of map
table	MODSMD_TABLE[d]	Bibliographic meta-data of table
chart / diagram	MODSMD_CHARTDIAG[d]	Bibliographic meta-data of
		chartdiagram

Where **[d]** is a counter starting for every METS file at 1 and incremented by one with every element of the same type described within the current monograph.

This element must have the following structure:

```
<dmdSec ID="[DMD_SEC_ID]">
<mdWrap MIMETYPE="text/xml" MDTYPE="MODS" LABEL="[LABEL]">
<xmlData>
<MODS:mods>
[ELEMENT_TITLE_INFO]
[AUTHOR_INFO]
</MODS:mods>
</xmlData>
</mdWrap>
</dmdSec>
```

Tag description (values to be inserted by supplier)

Tag	Description of tag
[DMD_SEC_ID]	Value must be from the table above.
	E.g., the 10th picture discovered in a given monograph
	will have:
	[DMD_SEC_ID] = "MODSMD_PICT10"
	and the 5th table will have:
	[DMD_SEC_ID] = "MODSMD_TABLE5"
[LABEL]	The value must be from the table above, depending on
	the type of structural element it is.
	E.g.
	<mdwrap <="" mdtype="MODS" mimetype="text/xml" th=""></mdwrap>
	LABEL="Bibliographic meta-data of
	chartdiagram">
[ELEMENT_TITLE_INFO]	[PICTURE_TITLE_INFO] describes the captions
	belonging to a picture if available.
	See [TITLE_INFO] in Chapter " <dmdsec> elements for</dmdsec>
	chapters" above for details concerning the composition
	of [TITLE_INFO]. The same requirements as described
	for [TITLE_INFO] are valid for the construction of
	[PICTURE_TITLE_INFO]
[AUTHOR_INFO]	[AUTHOR_INFO] has to be included if the author of the
	element can be determined. For the format, please see
	" <dmdsec> elements for chapters" above</dmdsec>

Attention: Either the caption or the Author has to be present. For elements of type illustration, map, table or chart_diagram that have no caption and no author, no <dmdSec> element should be created.

<u>Note:</u> At this stage, we are still considering the hierarchy as *flat* - we are not concerned with the actual hierarchy yet (will be covered in <structMap> further below) - BUT it is important that descriptive metadata for all the chapters and other structural elements appearing in the monograph are defined at this stage.

5 <amdSec>: Administrative Metadata Section

5.1 Reference documentation

<u>Note:</u> External documentation is included in this document for reference. Please note that some of the information appearing in those descriptions might not be relevant (attributes that we don't require etc...)

5.1.1 ... from METS documentation

Element	Description of element and attributes
<amdsec></amdsec>	amdSec: Administrative Metadata Section.
	This section records all of the administrative metadata for all items in the METS object (including structural map divs, data files, descriptive metadata sections and adminstrative metadata sections themselves), and is divided into four subsections: techMD (technical metadata), rightsMD (intellectual property rights metadata), sourceMD (analog/digital source metadata), and digiprovMD (digital provenance metadata). Each of these subsections follows the mdSecType model, so that they can either include metadata within the METS hub document (mdWrap) or reference it via an identifier/locator (mdRef). Multiple techMD, rightsMD, sourceMD and digiprovMD elements are allowed so that administrative metadata can be recorded for
	each separate item within the METS object.
	The administrative metadata section consists of four possible subsidiary sections: techMD (technical metadata for text/image/audio/video files), rightsMD (intellectual property rights metadata), sourceMD (analog/digital source metadata), and digiprovMD (digital provenance metadata, that is, the history of migrations/translations performed on a digital library object from its original digital capture/encoding). amdSecType has a single attribute, ID (XML ID).
<tecnmd></tecnmd>	technu: technical metadata.
	The techMD element provides a wrapper around a generic metadata section, which should contain technical metadata regarding a file or files. It has a single attribute, ID, which file/fileGrp elements can use to reference the technical metadata that applies to them.
<mdwrap></mdwrap>	mdWrap: metadata wrapper.

	 The mdWrap element is a generic element used throughout the METS schema to allow the encoder to place arbitrary metadata conforming to other standards/schema within a METS document. The included metadata can either be encoded in XML, in which case it may be placed directly within the mdWrap element, or it can be Base64 encoded, and placed within a subsidiary binData element. The mdWrap element can have the following attributes: ID: an XML ID for this element MIMETYPE: the MIME type for the metadata contained in the element
	• MDType: the type of metadata contained (e.g., MARC, EAD, etc.)
	 OTHERMDTYPE: a string indicating an alternative MDTYPE when the MDTYPE value is set to "OTHER."
	 LABEL: a label to display to the viewer of the METS document identifying the metadata
<pre><vmldata></vmldata></pre>	A wrapper to contain Base64 encoded metadata

5.1.2 ... from MIX documentation

Element	Description of element and attributes
<mix:basicimageparameters></mix:basicimageparameters>	The items in this section are fundamental to the reconstruction of the digital file as a viewable image on electronically interfaced displays. It makes no presumption about the rendered or spatial accuracy of the displayed image, only that a reasonably appearing image can be reconstructed.
<mix:mimetype></mix:mimetype>	Definition: Designation of the Multipurpose Internet Mail Extensions (MIME) type associated with the image data. The values listed below represent MIME types for digital still image formats commonly used in library and museum digital reformatting initiatives. The x- convention is used to construct an unofficial type for any image format lacking a formally registered MIME type. See http://www.mime- types.com/ for an up-to-date list of formally registered MIME types.
<mix:byteorder></mix:byteorder>	Definition: Designates the byte order in multi-byte numbers are stored. Virtually all computer architectures are byte addressable. The bytes of a multi-byte data value can be stored in memory in different orders. "Little_endian" means that the low-order byte of the number is stored in memory at the lowest address, and the high-order byte at the highest address. Big_endian" means that the high-order byte of the number is stored in memory at the lowest address, and the low-order byte at the highest address. Values can be: "big-endian" "little-endian"

 Store the image data. Values above are drawn from TIFF 6.0 specification (p17) though institutions are encouraged to devise a local enumerated list to allow for the addition of new values as technology changes. This data element allow for the designation of subelements in order to record the level of compression applied. See CompressionLevel <mix:compressionscheme></mix:compressionscheme> 1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoding.
 store the image data. Values above are drawn norm nerror. specification (p117) though institutions are encouraged to devise a local enumerated list to allow for the addition of new values as technology changes. This data element allow for the designation of subelements in order to record the level of compression applied. See CompressionLevel <mix:compressionscheme></mix:compressionscheme> 1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoding.
 specification (p117) though institutions are encouraged to devise a local enumerated list to allow for the addition of new values as technology changes. This data element allow for the designation of subelements in order to record the level of compression applied. See CompressionLevel <mix:compressionscheme></mix:compressionscheme> 1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoding.
devise a local enumerated list to allow for the addition of new values as technology changes. This data element allow for the designation of subelements in order to record the level of compression applied. See CompressionLevel <mix:compressionscheme>1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoding.</mix:compressionscheme>
new values as technology changes. This data element allow for the designation of subelements in order to record the level of compression applied. See CompressionLevel <mix:compressionscheme>1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoding.</mix:compressionscheme>
for the designation of subelements in order to record the level of compression applied. See CompressionLevel <mix:compressionscheme>1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoding.</mix:compressionscheme>
level of compression applied. See CompressionLevel <mix:compressionscheme>1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoder</mix:compressionscheme>
<mix:compressionscheme> 1 = No compression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final–June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encodered</mix:compressionscheme>
 Instruction pression, but pack data into bytes as tightly as possible, leaving no unused bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final–June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run length encoding.
bits (except at the end of a row). The component values are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
are stored as an array of type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
type BYTE. Each scan line (row) is padded to the next BYTE boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
boundary. 2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
2 = CCITT Group 3 1-Dimensional Modified Huffman run length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
length encoding. See TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
TIFF 6.0 Specification Final—June 3, 1992 18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
18 Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
Section 10 for a description of Modified Huffman Compression. 32773 = PackBits compression, a simple byte-oriented run
Compression. 32773 = PackBits compression, a simple byte-oriented run
32773 = PackBits compression, a simple byte-oriented run
32//3 = PackBits compression, a simple byte-oriented run
longth ach and
tength scheme.
<pre><mix:colorspace> The color space of the image data.</mix:colorspace></pre>
Tag = 262 (106.H)
Type = SHORT
N = 1
$\Omega = WhitelsZero, For hilevel and gravscale images: \Omega is$
imaged as white 2**RitePorSample_1 is imaged as black
This is the normal value for Compression 2
This is the normal value for Compression=2.
1 = BlackIsZero. For bilevel and grayscale images: 0 is
imaged as black. 2**BitsPerSample-1 is imaged as white. If
this value is specified for Compression=2, the image should
display and print reversed.
2 = RGB. In the RGB model, a color is described as a
combination of the three primarycolors of light (red. green
and blue) in particular concentrations. For each of the
three components. O correspond minimum interstities and
2**Dispersion and the second s
2 Dissel - 1 represents maximum intensity. This a
RGB value of (0,0,0) represents black, and (255,255,255)
represents white, assuming 8-bit components. For
PlanarConfiguration = 1, the components are stored in the
indicated order: first Red, then Green, then Blue. For
PlanarConfiguration = 2. the StripOffsets for the component
planes are stored in the indicated order: first the Red
component plane StripOffsets, then the Green plane
StripOffcots, then the Blue plane StripOffcots
sulponsets, then the blue plane sulponsets.
3 = Palette color. In this model, a color is described with a
single component. The value of the component is used as
an index into the red, green and blue curves in the
ColorMap field to retrieve an RGB triplet that defines the
color. When PhotometricInterpretation=3 is used, ColorMap
must be present and SamplesPerPixel must be 1.
4 = Transparency Mask This means that the image is used
to define an irregularly shaped region of another image in

	the same TIFF file. SamplesPerPixel and BitsPerSample must be 1. PackBits compression is recommended. The 1- bits define the interior of the region; the 0-bits define the exterior of the region.
<mix:imageidentifier></mix:imageidentifier>	Definition: A location qualifier to be used in conjunction with 2.2.1 ImageIdentifier. Persistent identifier required at prime object level, optional at all other levels. This identifier must be unique within the local system. To facilitate file sharing or interoperability with other systems, ImageIdentifierLocation may be added to designate the system or application within which the identifier is unique.
<mix:filesize></mix:filesize>	Definition: Extent of image in number of bytes. The file size must record the number of bytes as provided by the system. Do not attempt to record file sizes in terms of KB, MB or other notation.
<mix:orientation></mix:orientation>	Definition: Designates the orientation of the image, with respect to the placement of its rows (ImageWidth) and columns (ImageLength), as it was saved to disk. * normal" is defined as follows: when opened, the top (0th) row of pixels corresponds to the visual top of the image, and the first (0th) column of pixels on left corresponds to the visual left-hand side of the image. Consult TIFF for additional values referring to mirrored images. (Note that TIFF/EP supports only five values, which are proposed above as the finite list of enumerated type values.) This field is to be used to record only the orientation of the image, not the orientation of the source to the device (e.g., camera) used to capture the image. Possible values: • 1 = normal* • 3 = normal rotated 180° • 6 = normal rotated cw 90°
<mix:displayorientation></mix:displayorientation>	 Y= UNKNOWN Definition: Designates the orientation in which the image should be presented to a conventional monitor with a 3:2 aspect ratio. This value is important to record when the preferred orientation of the image sent to a 3:2 aspect ratio computer monitor is different from Orientation. While Orientation refers to the placement of pixels in the digital image file, DisplayOrientation refers to the preferred orientation in which to display the content (text, picture, table, etc.) within the file. This field will likely become obsolete when standard" delivery applications, such as web browsers, incorporate an image rotation tool. Possible values: 0 = portrait
<mix:imagecreation></mix:imagecreation>	 1 = landscape This section can best be described as descriptive technical metadata. While it provides no quantitative information, per se, it can provide critical information with respect to the logistics and administrative conditions surrounding digital image data capture. Frequently, simple interrogation of these fields offers valuable diagnostics about the image creation step as well as those of

	subsequent image generations. This metadata block documents selected, irreversible attributes of the analog- to-digital conversion process that may be used for future
	quality assessment of the image data. By definition, image creation occurs only once. See, 5.1 Image Processing for
	fields to record digital-to-digital conversion processes.
<mix:sourcetype></mix:sourcetype>	Definition: This field specifies the medium of the analogue source material scanned to create a digital still image.
<mix:sourceid></mix:sourceid>	Definition: A unique identifier for a descriptive record of
	the source of image. Notes: Link to existing data record for
<pre>cmix:lmageBroducers</pre>	Definition: Identifies the organization level producer(s) of
	the image Notes: Identifies the organization-level producer
	of the 'file/bitstream,' i.e., the scanned image, transcribed text, audio file, etc.
<mix:hostcomputer></mix:hostcomputer>	Definition: Computer and/or operating system in use at the
	time of image creation. Notes: The definition for this multi-
	layered data element can be interpreted narrowly, as in
	TIFF (see above), or broadly as in Cedars, which states: This
	element contains information about the operating
	ingest including information on relevant bardware and
	operating systems together with the software products
	that would have been required in order to use it.? Use OS
	and OSVersion to record specifics about the operating
	system.
<mix:operatingsystem></mix:operatingsystem>	Operating system in use at the time of image creation.
	Values (Examples): Windows,
	Mac,
	Unix,
	Solaris,
<mix:osversion></mix:osversion>	Keulial
	creation.
	Values (Examples):
	2000 (e.g., Windows 2000)
	NT SECOND
	X (e.g., Mac OS X) V (e.g., Unix System V)
<mix:devicesource></mix:devicesource>	Definition: Classification of device used to create the image
	data.
	Values (Examples):
	transmission scanner
	reflection print scanner
	digital still camera
<pre>cmix:ScapperManufacturer></pre>	Sull HOIII VIDEO
	the image.
<mix:scannermodel></mix:scannermodel>	Definition: The model name of the scanner used to create
	the image.
<mix:scannermodelnumber></mix:scannermodelnumber>	Definition: The model number of the scanner used to create
<mix scannermodelserialno=""></mix>	Definition: The serial number of the scapper used to create
	the image.

<mix:scanningsoftware></mix:scanningsoftware>	Definition: The name of the capture software used to create the image
<mix:scanningsoftwareversionno></mix:scanningsoftwareversionno>	Definition: The version number of the capture software
cmix: DixolSizo>	Used to create the image.
	scanner.
<mix:xphysscanresolution></mix:xphysscanresolution>	Definition: This specifies the physical scanning resolution of
	the device, in meters, recording the x (width) direction.
	Usage Note: This is NOT the interpolated resolution of the
	final output data.
<mix:xphysscanresolution></mix:xphysscanresolution>	Definition: This specifies the physical scanning resolution of
	Usage Note: This is NOT the interpolated resolution of the
	final output data.
<mix:imageperformanceassessment></mix:imageperformanceassessment>	The operative principle in this section is to maintain the
	attributes of the image inherent to its quality. The title
	Performance Assessment has both a present and future
	context: these elements serve as metrics to assess the
	of preservation techniques, particularly migration (future
	use). Sub-sections 4.1 Spatial Metrics and 4.2 Energetics are
	meant as high-level quantitative measures of imaging
	performance. Sub-section 4.3 Targets is meant to
	complement the former by providing low-level
	benchmarking quantification of the absolute imaging
	information in this latter section should be closely tied to
	sanctioned imaging performance standards when available.
	In the absence of such standards, de-facto standards are
	appropriate. To a large extent, the image of any source can
	be linked backed to that source with appropriate capture
	documentation and benchmarking targets. While the
	original source characteristics are not unequivocally
	can, in principle, occur. The high level metrics of sub-
	sections 4.1 and 4.2 can provide nominal recovery of the
	original source characteristics. Detailed imaging
	performance information in Section 4.3, if properly
	documented, is a reliable thread to more accurate source
	Definition: This specifies the width of the digital image i e
	horizontal or X dimension, in pixels, Notes: The image
	width may be the shorter or longer dimension of the image,
	depending upon the orientation of the camera or scanner
	during image capture. For multiple-resolution image file
	formats, value shall specify the highest resolution. This
	Source X dimension is given in inches and
	SamplingFrequencyUnit = 2. Formula to calculate
	XSamplingFrequency:
	XSamplingFrequency = ImageWidth/Source_Xdimension
<mix:imagelength></mix:imagelength>	Definition: specifies the length of the digital image, i.e.
	vertical or Y dimension, in pixels. Notes: The image length
	depending upon the orientation of the camera or scapper
	during image capture. For multiple-resolution image file
	formats, value shall specify the highest resolution. This
	field may be used to calculate YSamplingFrequency when

Source_Ydimension is given in inches and
SamplingFrequencyUnit = 2. Formula to calculate
YSamplingFrequency:
YSamplingFrequency = ImageLength/Source_Ydimension

<u>Note:</u> In this section, the term "image" refers to the digitized pages (.TIF files), not the illustrations appearing in the monograph.

The technical metadata for each digitized image (.tif) making up the monograph is described by an individual <amdSec> element with the following structure:

Tag description (values to be inserted by supplier)

Tag	Description of tag
[AMD_SEC_ID]	Value must be "IMGPARAM[ddddd]" where [ddddd] is a counter starting at 00001 for every monograph, and incremented by 1 for every new image of the current monograph.
	For example, for the second image of the current monograph, [AMD_SEC_ID] = IMGPARAM00002
[TECHMD_ID]	<pre>"[AMD_SEC_ID]TECHMD" For example, [TECHMD_ID] = "IMGPARAM00002TECHMD" for the <techmd> element contained in an <amdsec> element with [AMD_SEC_ID] = IMGPARAM00002</amdsec></techmd></pre>

5.2 NISO Data Dictionary: Technical Metadata for Digital Still Images

The Library of Congress' Network Development and MARC Standards Office, in partnership with the NISO Technical Metadata for Digital Still Images Standards Committee and other interested experts, has developed an XML schema for a set of technical data elements required to manage digital image collections. The schema provides a format for interchange and/or storage of the data specified in the NISO Draft Standard Data Dictionary: Technical Metadata for Digital Still Images (Version 1.2). This schema is currently in draft status and is being referred to as "NISO Metadata for Images in XML (NISO MIX)". MIX is expressed using the XML schema language of the World Wide Web Consortium. MIX is maintained for NISO by the Network Development and MARC Standards Office of the Library of Congress.

The following mix data elements have to be captured. Other tags that are defined in the mix schema (*http://downloads.bnl.lu/standards/mix.xsd*) can be added by the service provider if necessary.

We suppose that images are scanned with the following specifications:

TIFF / uncompressed / 300 dpi / 24 bits true color / "Original Optimized Scans"

There must be 1 <amdSec> element for each TIFF file making up the monograph.

[NISO_MIX_DATA]

<mix:basicimageparameters></mix:basicimageparameters>		
<mix:format></mix:format>		
<mix:mimetype><i>image/tiff</i></mix:mimetype>		
<mix:byteorder><i>little-endian</i></mix:byteorder>		
<mix:compression></mix:compression>		
<mix:compressionscheme>1</mix:compressionscheme>		
<mix:photometricinterpretation></mix:photometricinterpretation>		
<mix:colorspace>1</mix:colorspace>		
<mix:file></mix:file>		
<mix:imageidentifier imageidentifierlocation="</td"></mix:imageidentifier>		
[IMAGE_IDENTIFIER_LOCATION]>[IMAGE_IDENTIFIER_FILENAME]		
tifier>		
<mix:filesize>[FILESIZE]</mix:filesize>		
<mix:orientation>[ORIENTATION]</mix:orientation>		
<mix:displayorientation>[DISPLAY_ORIENTATION]</mix:displayorientation>		
<mix:imagecreation></mix:imagecreation>		
<mix:sourcetype>Monograph</mix:sourcetype>		
<mix:sourceid>[BARCODE]</mix:sourceid>		
<mix:imageproducer>[ORGANIZATION_PRODUCER]</mix:imageproducer>		
<mix:host></mix:host>		
<mix:hostcomputer>[HOST_COMPUTER]</mix:hostcomputer>		
<mix:operatingsystem>[OS]</mix:operatingsystem>		



5.3 Tag explanations

Tag	Description of tag
[IMAGE_IDENTIFIER_LOCATION]	Replace by "[BIBREC_SYS_NUM]/tif" where
	[BIBREC_SYS_NUM] will correspond to the
	system number of the bibliographic record.
	This will be the same for all <amdsec> elements in a particular METS file.</amdsec>
	Example:
	For the monograph with the title "Heinrich,
	Prinz der Niederlande, Statthalter S.M. des
	Königs-Grossherzogs im Grossherzogtum
	Luxemburg" [IMAGE_IDENTIFIER_LOCATION] =
	"000239390/tif" (for every <amdsec> element</amdsec>
-----------------------------	---
	in the METS file describing that monograph).
[IMAGE_IDENTIFIER_FILENAME]	Replace by "[ddddd].tif" where [ddddd]
	corresponds to a 5 digit page counter (filled
	with zeros) starting at 1 for every monograph.
	Example:
	For monograph with title "Heinrich, Prinz der Niederlande, Statthalter S.M. des Königs- Grossherzogs im Grossherzogtum Luxemburg", the first <amdsec> element will have [IMAGE_IDENTIFIER_FILENAME] = "00001.tif", the second one [IMAGE_IDENTIFIER_FILENAME] = "00002.tif", etc (every .tif file in [IMAGE_IDENTIFIER_LOCATION] is referenced once - each file in an individual <amdsec> element)</amdsec></amdsec>
	Definition: Extent of image in number of bytes
	"1" if orientation - normal
	1 II offentation = normat.
	Refer to definition of <mix:orientation> in 4.3 Element description above</mix:orientation>
[DISPLAY_ORIENTATION]	"0" if Portrait: "1" if Landscape.
	Refer to definition of <mix:displayorientation></mix:displayorientation>
	in 4.3 Element description above
[BARCODE]	The physical volume where the monograph was
	scanned from. These are indicated in the Excel
	inventory sheets under [ITEM_BARCODE].
[ORGANIZATION_PRODUCER]	Definition: Identifies the organization-level producer(s) of the image
	producer(s) of the image.
	Notes: Identifies the organization-level
	producer of the 'file/bitstream,' i.e., the
	scanned image, transcribed text, audio file,
	etc.
[HOST-COMPUTER]	Computer and/or operating system in use at the time of image creation.
	This element contains information about the operating environment of the original digital object at the time of ingest, including information on relevant hardware and operating systems, together with the software products that would have been required in order to use it.

[OS]	Operating system in use at the time of image
	creation.
	Values (Examples):
	Windows,
	Mac,
	Unix,
	Solaris,
	Redhat
[US_VERSION]	Version of the operating system in use at the
	time of image creation.
	Values (Examples):
	Values (Examples). 2000 (e.g. Windows 2000)
	NT
	X (e.g. Mac OS X)
	V (e.g., Mac OS X)
	Definition: Classification of device used to
	create the image data
	create the image data.
	Values (Examples):
	transmission scanner
	reflection print scanner
	digital still camera
	still from video
[SCANNER_MANUFACTURER]	Definition: The manufacturer of the scanner
	used to create the image.
[SCANNER_MODEL_NAME]	Definition: The model name of the scanner used
	to create the image.
[SCANNER_MODEL_NUMBER]	Definition: The model number of the scanner
	used to create the image.
[SERIAL_NUMBER]	Definition: The serial number of the scanner
	used to create the image.
[SCANNING_SOFTWARE]	Definition: The name of the capture software
	used to create the image.
[SCANNING_SOFTWARE_VERSION]	Definition: The version number of the capture
	software used to create the image.
[SCANNER_PIXEL_SIZE]	Definition: This specifies the pixel size, in
	meters, of the scanner.
[IMAGE_WIDTH]	Width of the digital image, i.e. horizontal or X
	dimension, in pixels
[IMAGE_LENGIH]	Length of the digital image, i.e. vertical or Y
	aimension, in pixels
	The identifier of the original image from which
	unis image is a part. The value of this identifier
	is the ID of the original image in the
	mesec/meurp/me tag.

6 The <fileSec> elelemt

6.1 Reference documentation (From METS documentation)

<u>Note:</u> External documentation is included in this document for reference. Please note that some of the information appearing in those descriptions might not be relevant (attributes that we don't require etc...)

<filesec></filesec>	fileSec: Content File Section.
	The content file section records information regarding all of the files which comprise the digital library object.
<filegrp></filegrp>	fileGrp: File Group.
	The file group is used to cluster all of the digital files composing a digital library object in a hierarchical arrangement (fileGrp is recursively defined to enable the creation of the hierarchy).
	Any file group may contain zero or more file elements.
	File elements in turn can contain one or more FLocat elements (a pointer to a file containing content for this object) and/or a FContent element (the contents of the file, in either XML or Base64 encoding).
	A fileGrp element may have the following attributes:
	 ID: an XML ID for the element VERSDATE: date this version/fileGrp of the digital object was created. ADMID: IDREFs to administrative metadata sections in the METS document that correspond with all files in this file group; USE: a string to indicate the intended use of files within this file group (e.g., master, reference, thumbnails for image files).
<file></file>	The file element provides access to content files for a METS document. A file element may contain one or more FLocat elements, which provide pointers to a content file, and/or an FContent element, which wraps an encoded version of the file. Note that ALL FLocat and FContent elements underneath a singel file element should identify/contain identical copies of a single file. The file element has the following attributes:

1. ID: an XML ID for the element;
2. MIMETYPE: the MIME type for the file;
3. SEQ: an integer indicating the sequence of this file relative
to the others in its file group; 4. SIZE: the size of the file in
bytes;
5. CREATED: the date of creation for the file;
6. CHECKSUM: a checksum value for the included file;
7. CHECKSUMTYPE: the type of checksum contained in the
CHECKSUM attribute;
8. OWNERID: a primary identifier assigned to the file by its
owner;
9. ADMID: IDREFS to administrative metadata sections in the
METS document that correspond with this file;
10. DMDID: IDREFS to descriptive metadata sections in the
METS document that correspond with this file;
11. GROUPID: an identifier that establishes a correspondence
between this file and files in other file groups. Typically, this
will be used to associate a master file in one file group with
derivative files in other file groups;
12. USE: a string indicated the intended use of this file (e.g.,
master, reference, thumbnail for image files).
12. USE: a string indicated the intended use of this file (e.g., master, reference, thumbnail for image files).

We require the <fileSec> element to contain 3 <fileGrp> elements

- one file group for all the images contained in the monograph, including those created by any special treatments, as described in *Technical requirements Chapter 3.5.5*) [ID="IMGGRP"]
- one file group for the ALTO files (containing OCR text) [ID="ALTOGRP"]
- one file group for the PDF files (1 per page) [ID="PDFGRP"]
- one file group for the PNG files (1 per page) [ID="reference image"]

The <fileSec> element must therefore have the following structure

```
<fileSec>
<fileGrp ID="IMGGRP" USE="Images">
[IMG_FILES]
</fileGrp>
<fileGrp ID="ALTOGRP" USE="Text">
[ALTO_FILES]
</fileGrp>
<fileGrp ID="PDFGRP" USE="PDF">
[PDF_FILES]
</fileGrp>
<fileGrp ID="PNGGRP" USE="reference image">
[PNG_FILES]
</fileGrp>
</fileGrp>
</fileSec>
```

6.2 Tag explanation

Please see "Appendix A: File locations and naming" to understand the file locations referred to in the following table.

Tag	Definition of tag
[IMG_FILES]	Must contain one <file> element for every one of the images of</file>
	the monograph.
	These <file> elements must have the following structure:</file>
	<file< th=""></file<>
	ID="[IMG_ID]" CDEAMED-"[DAME]"
	ADMID="[ADM ID]"
	MIMETYPE="image/tif"
	SEQ="[SEQ]"
	GROUPID="[GROUPID]"
	CHECKSUM="[CHECKSUM]" CHECKSUMTYDE="SHA512"
	SIZE="[SIZE]"
	<flocat loctype="URL" xlink:href="[HREF]"></flocat>
	Example:
	<file <="" created="2006-03-16T14:16:01" id="IMG00001" th=""></file>
	ADMID="IMGPARAM00001" MIMETYPE="image/tif" SEQ="1"
	GROUPID="1"
	CHECKSUM="A268D1BD5F992910337D45110806153A170F907E84E23BA
	342D72A991D071E3D6C0BF49" CHECKSUMTYPE="SHA512"
	SIZE="13253616">
	<flocat <="" loctype="URL" th=""></flocat>
	<pre>link:href="file://./tif/00001.tif"/></pre>
[IMG ID]	Replace by "IMG[ddddd]" where [ddddd] corresponds to a 5
	digit counter (filled with zeros) starting at 1 for every
	monograph.
[DATE]	The date of creation for the file - format: yyyy-mm-
	ddThh:mm:ss
[ADM_ID]	IDREFS to administrative metadata sections ([ADM_SEC_ID] in
	<admsec>) in the METS document that correspond with this</admsec>
	file;
	e.g. "IMGPARAM00001"
[SEQ]	An integer indicating the sequence of this file relative to the
_	others in its file group;
	A counter starting at 1 for every monograph and incremented
	by 1 for every image in the <filegrp> element.</filegrp>

[GROUPID]	An integer indicating the sequence of this file relative to the
	ALTO. PDF and IMG files. so that the IMG, ALTO, PDF and PNG
	files corresponding to page 1 have GROUPID="1". IMG, ALTO,
	PDF and PNG files corresponding to page 2 have GROUPID="2"
	etc.
	A counter starting at 1 for eveny monograph and incremented
	by 1 for every file in the <filegrp> element.</filegrp>
[CHECKSUM]	SHA512-Checksum of the file
[FILESIZE]	The size of the file in bytes
[HREF]	Location of the file relative to the METS. See Appendix A: File
_	Locations for details.
	Events for MCCDD - Say the METS is in directory (000220200)
	Example for IMGGRP: Say the MEIS IS IN directory / 000239390/
	and we refer to the image for page 2, then $[11x_{E1}] =$
	Example for ALTOGRP: If we refer to the ALTO for page 3, then
	[HREF] = "file://./alto/00003.xml"
	Example for PDECPP. If we refer to the PDE for page 1, then
	Example for PDFGKF. II we refer to the PDF for page 1, then $\mathbf{FPF1} = \mathbf{Ff1} + \mathbf$
	Example for PNGGRP: If we refer to the PNG for page 7, then
	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf"</pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the measurement. These ufiles elements must have the following.</file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure:</file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure:</file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file< pre=""></file<></file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file <="" id="[ALTO_ID]" mimetype="text/yml" pre=""></file></file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file <="" checksum="[CHECKSUM]" id="[ALTO_ID]" mimetype="text/xml" pre=""></file></file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file <="" checksum="[CHECKSUM]" checksumtype="SHA512" id="[ALTO_ID]" mimetype="text/xml" onumpe="SHA512" pre=""></file></file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: </file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file checksum="[CHECKSUM]" checksumtype="SHA512" groupid="[GROUPID]" id="[ALTO_ID]" mimetype="text/xml" size="[FILESIZE]"> <flocat loctype="URL" xlink:href="[HREF]"></flocat></file></file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file checksum="[CHECKSUM]" checksumtype="SHA512" groupid="[GROUPID]" id="[ALTO_ID]" mimetype="text/xml" size="[FILESIZE]"> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file></file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file checksum="SHA512" groupid="[GROUPID]" id="[ALTO_ID]" mimetype="text/xml" size="[FILESIZE]"> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file> Example:</file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: </file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file checksum="[GROUPID]" id="[ALTO_ID]" mimetype="text/xml" size="[FILESIZE]"> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file> Example: <file .="" 00007.pdf"="" <file="" checksum="51F67A67ED05A60B27432D436D6A2FFFD4c578D8FD4BD9F</pre></th></tr><tr><th>[ALTO_FILES]</th><th><pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = " contain="" file:="" id="ALTO0001" mimetype="text/xml" must="" one="" png=""> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: <file checksum="[GROUPID]" id="[ALTO_ID]" mimetype="text/xml" size="[FILESIZE]"> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file> Example: <file <="" checksum="51F67A67ED05A60B27432D436D6A2FFFD4C578D8FD4ED9F F25A9DD29F7B5EB8A0A67FA93C18554A08B0E51B4B9477E8C30A47F2D 92644EDBD327CE9242160070" checksumtype="SHA512" id="ALTO00001" mimetype="text/xml" pre=""></file></file></file></file></file></pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: </file> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file> Example: </pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: </file> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file> Example: </pre>
[ALTO_FILES]	<pre>Example for PNGGRP: If we refer to the PNG for page 7, then [HREF] = "file://./png/00007.pdf" Must contain one <file> element for every page (ALTO file) of the monograph. These <file> elements must have the following structure: </file> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file> Example: <flocat loctype="URL" xlink:href="file://./alto/00001.xml"></flocat> </pre>

[ALTO_ID]	Replace by "ALTO[ddddd]" where [ddddd] corresponds to a 5
	digit counter (filled with zeros) starting at 1 for every
	monograph.
[PDF_FILES]	Must contain one <file> element for every page (pdf) of the monograph. These <file> elements must have the following</file></file>
	structure:
	<file ID="[PDF_ID]" MIMETYPE="text/pdf" CHECKSUM="[CHECKSUM]" CHECKSUMTYPE="SHA512" GROUPID="[GROUPID]" SIZE="[FILESIZE]"> <flocat loctype="URL" xlink:href="[HREF]"></flocat> </file
	Example:
	<file checksum="
BD4FDE65F0AE8F3BC2A4A53B786900A5EC981EEFB6C878013BA3FC8B7
F0A29685C9009425C9F801EB8935984757234709FB584AD790F928CCA
3C79A9FFA2BD10" checksumtype="SHA512" id="PDF00001" mimetype="text/pdf" size="1993828"> <flocat <="" loctype="URL" th=""></flocat></file>
	<pre>xlink:href=file://./pdf/00001.pdf/> </pre>
[PDF ID]	Replace by "PDF[ddddd]" where [ddddd] corresponds to a 5
[]	digit counter (filled with zeros) starting at 1 for every
	monograph.
[PNG_FILES]	Must contain one <file> element for every page in PNG format</file>
	of the monograph. These <file> elements must have the</file>
	following structure:
	<file ID="[PNG ID]"</file
	MIMETYPE="image/png"
	CHECKSUM="[CHECKSUM]"
	GROUPID="[GROUPID]"
	SIZE="[FILESIZE]">
	<pre><flocat loctype="URL" xlink:href="[HREF]"></flocat> </pre>
	Example:
	<file <br="" id="PNG00001" mimetype="image/png">CHECKSUM="9BE383FDDCDE3CE18377B6B4520A748E1165AC24540E20A 5D3CE3D97F22CB39F225DE14DF0F9CBD0579B26EA8ACE536091C34BA9 360EB4B2E4209D11DE716FB2" CHECKSUMTYPE="SHA512" SIZE="63351"> <flocat <="" loctype="URL" th=""></flocat></file>
	<pre>xlink:href="file://./png/00001.pdf/"> </pre>



[PNG_ID]	Replace by "PNG[ddddd]" where [ddddd] corresponds to a 5
	digit counter (filled with zeros) starting at 1 for every
	monograph.

7 The <structMap> elements

7.1 Reference documentation (From METS documentation)

<u>Note:</u> External documentation is included in this document for reference. Please note that some of the information appearing in those descriptions might not be relevant (attributes that we don't require etc...)

<structmap></structmap>	The structural map is the heart of a METS document, defining the hierarchical arrangement of a primary source document which has been digitized. This hierarchy is encoded as a tree of 'div' elements. Any given 'div' can point to another METS document via the 'mptr' element, or to a single file, to a group of files, or to segments of individual files or groups of files through the 'fptr' and subsidiary elements.
	elements to point to segments of individual files. 'mptr' is not used.
<div></div>	 The METS standard represents a document structurally as a series of nested div elements, that is, as a hierarchy (e.g., a book, which is composed of chapters, which are composed of subchapters, which are composed of text). Every <div> node in the structural map hierarchy may be connected (via subsidiary mptr or fptr elements) to content files which represent that div's portion of the whole document.</div> The div element has the following attributes: ID (an XML ID); ORDER: an integer representation of this div's order among its siblings (e.g., its sequence); ORDERLABEL: a string representation of this div's order among its siblings (e.g., "xii"), or a non-integer native numbering system. It is presumed that this value will still be machine-actionable (e.g., supports a page 'go to' function), and is not a replacement/substitute for the LABEL attribute. LABEL: a string label to describe this div to an end user viewing the document, as per a table of contents entry (NB: a div LABEL should be specific to its level in the structural map. In the case of a book with chapters, the book div LABELS should have the individual chapter titles, rather than having the chapter div LABELs combine both book title and chapter title).

	 NB: to clarify the differences between ORDER, ORDERLABEL, and LABEL, imagine a text with 10 roman numbered pages followed by 10 arabic numbered pages. Page iii would have an ORDER of "3", an ORDERLABEL of "iii" and a LABEL of "Page iii", while page 3 would have an ORDER of "13", an ORDERLABEL of "3" and a LABEL of "Page 3". 5. DMDID: a set of IDREFs to descriptive metadata sections within this METS document applicable to this
	 div. 6. ADMID: a set of IDREFS to administrative metadata sections within this METS document applicable to this div. 7. TYPE: a type of division (e.g., chapter, article, page, etc.).
<fptr></fptr>	 fptr: File Pointer. The fptr element associates a div element with content files that represent that div. It can either point to a file directly itself, via the FILEID attribute, or it can do more complex links to content via the subsidiary area, par and seq elements. The fptr element can have the following attributes: ID: an XML ID for this element; FILEID: an IDREF to a file element which corresponds with the div containing this ftpr.
<par> / <seq></seq></par>	 par: Parallel files. The par element should be used to link a div to a set of content files when those files should be played/displayed in unison to deliver the content to the user. A par element has two possible subsidiary elements, which should be used in different cases. In cases where each bytestream to be played in parallel can fit in a single file, you should use subsidiary area elements within the par element to point to those files. However, in some cases, bytestreams which should be played in parallel are too large to fit in a single file (high quality multi-track audio, or video). In those cases, you should use subsidiary seq elements, where each seq contains the files comprising a particular bytestream in the order they should be played back. Par has the following attributes: ID: an XML ID for this element. Seq: Sequential files.

	files that must be played or displayed sequentially to manifest a block of digital content. This might be the case, for example, if the parent <div> element represented a logical division, such as a diary entry, that spanned multiple pages of a diary and, hence, multiple page image files. In this case, a <seq> element would aggregate multiple, sequentially arranged <area/> elements, each of which pointed to one of the image files that must be presented sequentially to manifest the entire diary entry. If the diary entry started in the middle of a page, then the first <area/> element (representing the page on which the diary entry starts) might be further qualified, via its SHAPE and COORDS attributes, to specify the specific, pertinent area of the associated image file.</seq></div>
<area/>	 The area element provides for more sophisticated linking between a div element and content files representing that div, be they text, image, audio, or video files. An area element can link a div to a point within a file, to a one-dimension segment of a file (e.g., text segment, image line, audio/video clip), or a two-dimensional section of a file (e.g, subsection of an image, or a subsection of the video display of a video file. The area element has no content, and the following attributes: FILEID: an IDREF to the file element being pointed to by the div; BEGIN: a beginning location in a referenced file; BETYPE: the syntax used in specifying the BEGIN and END attributes (byte offset, IDREF value, SMPTE time code, SMIL time value, MIDI time code, a simple time code of the form HH:MM:SS, or a TCF time code);

<u>Requirement:</u> Every METS file must have two <structMap> elements - one of TYPE="PHYSICAL" and one of TYPE="LOGICAL".

Hence, the following structure must appear in every METS file:

```
<structMap LABEL="Physical Structure" TYPE="PHYSICAL">
  [PHYSICAL_STRUCTMAP]
</structMap>
<structMap LABEL="Logical Structure" TYPE="LOGICAL">
  [LOGICAL_STRUCTMAP]
</structMap>
```

7.2 [PHYSICAL_STRUCTMAP]

The physical <structMap> describes the physical composition of the original work (it describes the physical sequence and image page linking of the document).

The physical <structMap> will be used for a page turning browse interface. However, no structural information (chapters, paragraphs etc. exceeding one page) is reflected here. It contains one top-level <div> element representing the document. This contains the attribute DMDID="MODSMD_ELEC MODSMD_PRINT MODSMD_00001" for referencing the global document metadata.

This <div> then contains one second-level <div> for every physical page of the monograph.

The IDs for the different <div> elements in the physical <structMap> element must have the format "DIVP[d]" where [d] is a counter starting at 1 for every monograph (the top-level <div> will have ID="DIVP1") and incremented by one every time a new <div> is defined the (the second-level <div>'s for the individual pages will have ID="DIVP2", "DIVP3"...)

```
<structMap LABEL="Physical Structure" TYPE="PHYSICAL">
<div ID="DIVP1" DMDID="MODSMD_ELEC MODSMD_PRINT
MODSMD_00001" LABEL="Heinrich, Prinz der Niederlande,
Statthalter S.M. des Königs-Grossherzogs im Grossherzogtum
Luxemburg" TYPE="Monograph">
[SECOND_LEVEL_DIVS]
</div>
</structMap>
```

[SECOND_LEVEL_DIVS]

On the next level, each <div> element represents one physical page. These <div>'s contain parallel <fptr> nodes pointing to the related .tif (image), ALTO, and .pdf files described in section 5. The <fileSec> element.

The following attributes are used for the described purpose:

- attribute "TYPE" contains the value "PAGE"
- attribute "ORDER" contains the automatically incremented values starting at '1'. It reflects the physical sequence of images.
- attribute "LABEL" contains the page number as it is printed on this particular page. If no page number is printed on the page, the value of "LABEL" should be the page number within the page sequence (i.e. the same as ORDERLABEL)
- attribute "ORDERLABEL" contains the page number within the page sequence. It is filled automatically for pages without printed page number.

Take the example of a monograph with 12 pages and there are no page numbers. The pages should have the following attributes:

<div ID="DIVP2" ORDER="1" TYPE="PAGE" LABEL="1" ORDERLABEL="1">
<div ID="DIVP2" ORDER="2" TYPE="PAGE" LABEL="2" ORDERLABEL="2">
<div ID="DIVP2" ORDER="3" TYPE="PAGE" LABEL="3" ORDERLABEL="3">
<div ID="DIVP2" ORDER="4" TYPE="PAGE" LABEL="4" ORDERLABEL="4">
<div ID="DIVP2" ORDER="5" TYPE="PAGE" LABEL="5" ORDERLABEL="5">
<div ID="DIVP2" ORDER="6" TYPE="PAGE" LABEL="6" ORDERLABEL="6">
<div ID="DIVP2" ORDER="6" TYPE="PAGE" LABEL="7" ORDERLABEL="6">
<div ID="DIVP2" ORDER="8" TYPE="PAGE" LABEL="7" ORDERLABEL="6">
<div ID="DIVP2" ORDER="7" TYPE="PAGE" LABEL="7" ORDERLABEL="7">
<div ID="DIVP2" ORDER="8" TYPE="PAGE" LABEL="7" ORDERLABEL="7">
<div ID="DIVP2" ORDER="8" TYPE="PAGE" LABEL="10" ORDERLABEL="10">
<div ID="DIVP2" ORDER="10" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
<div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
<div ID="DIVP2" ORDER="12" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
<div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
<div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="12" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="12" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="12" TYPE="PAGE" LABEL="12" ORDERLABEL="12">
</div ID="DIVP2" ORDER="12" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="11" ORDERLABEL="11">
</div ID="DIVP2" ORDER="11" TYPE="PAGE" LABEL="12" ORDERLABE

Each of the <div> elements describing the physical pages holds a <fptr> element linking to the image, pdf and alto files. To link to these 3 files "in parallel", we use a <par> element, holding three individual <area> elements to link to the 3 files respectively.

If a parent <div> element contains more than one child <div> element, then each of the child <div> elements must have an attribute ORDER=[ORDER_NUMBER] where [ORDER_NUMBER] is a counter starting at "1" for every first child <div> element and incremented by one with every new child <div> element.

Example physical structMap for a 4-page monograph:

```
<structMap LABEL="Physical Structure" TYPE="PHYSICAL">
  <div ID="DIVP1" DMDID="MODSMD ELEC MODSMD PRINT" LABEL="Heinrich,</pre>
  Prinz der Niederlande, Statthalter S.M. des Königs-Grossherzogs im
  Grossherzogtum Luxemburg " TYPE="Monograph">
    <div ID="DIVP2" ORDER="1" TYPE="PAGE" LABEL="1">
      <fptr>
         <par>
           <area FILEID="IMG00001"/>
           <area FILEID="PDF00001"/>
           <area FILEID="ALTO00001"/>
           <area FILEID="PNG00001"/>
         </par>
      </fptr>
    </div>
    <div ID="DIVP3" ORDER="2" TYPE="PAGE" LABEL="2">
      <fptr>
         <par>
           <area FILEID="IMG00002"/>
           <area FILEID="PDF00002"/>
           <area FILEID="ALTO00002"/>
           <area FILEID="PNG00002"/>
         </par>
      </fptr>
    </div>
    <div ID="DIVP4" ORDER="3" TYPE="PAGE" LABEL="3">
      <fptr>
         <par>
           <area FILEID="IMG00003"/>
           <area FILEID="PDF00003"/>
           <area FILEID="ALTO00003"/>
           <area FILEID="PNG00003"/>
         </par>
      </fptr>
    </div>
    <div ID="DIVP5" ORDER="4" TYPE="PAGE" LABEL="4">
      <fptr>
         <par>
           <area FILEID="IMG00004"/>
           <area FILEID="PDF00004"/>
           <area FILEID="ALTO00004"/>
           <area FILEID="PNG00004"/>
         </par>
      </fptr>
    </div>
  </div>
</structMap>
```

7.3 [LOGICAL_STRUCTMAP]

7.3.1 Introduction

Note: Instructions and requirements, which are essential for understanding how the hierarchy of a monograph must be defined, are explained in *Technical requirements* - *Chapter 6*.

The logical <structMap> describes the intellectual and logical structure of the document.

METS describes this logical structure as a hierarchical tree of <div> elements, each representing a whole monograph, or a chapter, illustration, image caption, heading, subheading, paragraph etc...

The purpose of such a structured hierarchy is to allow full text search to be applied within particular structure elements (for example article headings) and also within certain zone types such as captions of illustrations or headlines.

7.3.2 Requirements

7.3.2.1 General requirements

Every <div> has an ID. The IDs for the different <div> elements in the logical <structMap> element must have the format "DIVL[d]" where [d] is a counter starting at 1 for every monograph (the top-level <div> will have ID="DIVL1") and incremented every time a new lower-level <div> is defined the (the second-level <div> will have ID="DIVL2", and then the following ID="DIVL3", "DIVL4"...)

Each <div> element has an attribute TYPE which describes what part of an monograph the <div> refers to. Thus, attribute "TYPE" can have values such as "MONOGRAPH", "HEADLINE", "BODY_CONTENT", "PARAGRAPH", "CHAPTER", "TABLE" etc. To simplify the handling of the data (e.g. loading data in a presentation system) these values must be chosen from controlled vocabulary - allowed types are defined by the monograph schemas.

If a parent <div> element contains more than one child <div> elements, then each of the child <div> elements must have an attribute ORDER=[ORDER_NUMBER] where [ORDER_NUMBER] is a counter starting at "1" for every first child <div> element and incremented by one with every new child <div> element. Inside the logical structMap, the <seq> container must be used, not the <par> container for <div> elements.

7.3.2.2 Referencing descriptive metadata

To make the link between the hierarchical elements and the descriptive metadata of these elements (the <dmdSec> elements defined earlier on in the METS file), every <div> element with either:

Monographs:

- TYPE = VOLUME
- TYPE = CHAPTER
- TYPE = MAP*
- TYPE = TABLE*
- TYPE = ILLUSTRAION*

must have an attribute DMDID=[DMD_IDS] where [DMD_IDS] is to be replaced by the ID(s) for the <dmdSec> elements describing the volume, chapter, illustration that the <div> is about.

*) For MAP, TABLE, ILLUSTRATION and CHART_DIAGRAM, the following rule applies: If the element has no caption and no author, then it should not have a corresponding <dmdSec>, and so should have no label.

For example, the second-level <div> (with TYPE="VOLUME") must have attribute DMDID="MODSMD_PRINT MODSMD_ELEC".

7.3.2.3 LABEL attribute

The top-level <div> (TYPE="Monograph") and the second level <div> (TYPE="VOLUME") must have an attribute LABEL with the same value as the LABEL attribute for the <METS> root element.



Section two - ALTO Files

8 General requirements

An ALTO file must be created for every physical page in the collection

9 Documentation: ALTO files

9.1 Reference documentation

The following information was obtained from <u>http://www.ccs-gmbh.com/alto/technical.html</u>

The reference starts from an entey of the type:

```
<area BETYPE="IDREF" FILEID="ALTO00001" BEGIN="P1_TB00023" />
```

Withing the sturctmap

The FILEID refers to the following structure within the file group

```
<fileGrp ID="ALTOGRP">
<file ID="ALTO00001" MIMETYPE="text/xml">
<FLocat LOCTYPE="URL" xlink:href="file://./samplebook-
ALTO/samplebook-ALTO00001.xml" />
</file>
```

the BEGIN attribute then points into the alto file itself.

9.2 Structure of ALTO Files

The ALTO file consists of three major sections:

Description Styles Layout

The Description section contains metadata about the ALTO file itself and processing information on how the file was created. The Styles section contains the text and paragraph styles with their individual descriptions:

TextStyle has font descriptions ParagraphStyle has paragraph descriptions, e.g. alignment information The Layout section contains the content information. It is subdivided into Pages. A page consists of margins and printspace, all of those are non-intersection rectangular areas within the page area. Each of these can contain any number of objects like lines, images or textblocks and more. A textblock is divided into textlines and those are divided furthermore in strings and spaces.

The global structure of the ALTO file is as follows:

alto Description MeasurementUnit sourceImageInformation Processing Styles TextStyle TextStyle TextStyle MaragraphStyle ParagraphStyle ParagraphStyle MaragraphStyle MaragraphStyle MaragraphStyle

9.3 MeasurementUnit

The measurement unit must be "mm10". This is one tenth of a millimeter.

9.4 TextStyles

TextStyles have no content. The attributes are

BottomMargin PrintSpace

FONTFAMILY FONTSIZE FONTCOLOR FONTWEIGHT FONTSTYLE FONTPITCH FRONTCHARSET

UNDERLINED

Only FONTFAMILY and FONTSIZE are required.

9.5 ParagraphStyles

Paragraph styles have no content. The attributes are:

Name	with one of the values
ALIGN	Left
	Right
	Center
	Block
LEFT	Numeric
RIGHT	Numeric
LINESPACE	Numeric
FIRSTLINE	Numeric

9.6 Attributes of a Page Element

PAGECLASS STYLEREFS HEIGHT WIDTH PHYSICAL_IMG_NR PRINTED_IMG_NR QUALITY (OK, Damaged, Missing) POSITION (Left, Right, Foldout, Single) PROCESSING (A link to processing information)

9.7 Page Areas

Each page is divided into different areas (TopMargin, LeftMargin, RightMargin, BottomMargin and PrintSpace).

The positions are given as HPOS, VPOS, WIDTH and HEIGHT.

e.g. <TopMargin ID="P34_TM00001" HPOS="0" VPOS="0" WIDTH="2073" HEIGHT="298" />

TopMargin	The area between the top line of print and the upper edge of the leaf. It may contain page number, running title or a complete page header
LeftMargin	The left margin of a page. May contain margin notes.
RightMargin	The right margin of a page. May contain margin notes.
BottomMargin	The area between the bottom line of letterpress or writing and
	the bottom edge of the leaf. It may contain a page number, a
	signature number or a catch word.
PrintSpace	Rectangle surrounding the printed area of a page. Page number
	and running title are not part of the print space.



The position of the margins on a page is illustrated in this page.

9.8 The structure of each of the Page Area (PageSpace) Elements

The page area elements have the attributes:

HPOS	Horizontal position upper/left conrner (1/10 mm)
VPOS	Vertical position upper/left conrner (1/10 mm)
WIDTH	Width (1/10 mm)
HEIGHT	Height (1/10 mm)
ROTATION	In deg. as floating point number (optional)

Each page area may contain any number of elements. Those elements are one of the following:

TextBlock	A block of text
ComposedBlock	A block that consist of other blocks
Illustration	A picture of image
GraphicalElement	A graphic used to separate blocks. Mostly a line or a rectangle

Each of them may have the following attributes:

ID	Unique ID
STYLEREFS	Reference for text or paragraph styles
HPOS	Horizontal position upper/left corner (1/10 mm)
VPOS	Vertical position upper/left corner (1/10 mm)
WIDTH	Width (1/10 mm)
HEIGHT	Height (1/10 mm)
ROTATION	In deg. as floating point number (optional)
IDNEXT	Reference to the next element relating to the reading order

If the shape of the element is not rectangular an element SHAPRE might be added:

<Shape> <Polygon POINTS="x,y x,y x,y x,y x,y..." /> </Shape>

Polygons are coded as x,y x,y ... with different coordinate pairs separated by spaces.

Circles and ellipses are, although allowed in principle, not supported by docWORKS. Instead, such shapes are represented as polygons with sufficient accuracy.

A TextBlock is divided into lines and those are divided into strings, spaces and hyphens:

TextBlock

TextLine String SP String SP ... TextLine ...

Meanings of those tags

Tag	Description
TextLine	Line of text
String	A single word
SP	White space
HYP	Hyphenation

9.9 Additional Attributes of the Tags

TextBlock	language	
String	CONTENT	String content (word)
	SUBS_TYPE	HypPart 1: If content is the first part of a hyphenated word, applies only for the last word of a line if it is hyphenated.
		HypPart 2: If content is the second part of a hyphenated word, applies only for the first word of a line if it is hyphenated
	SUBS_CONTENT	Complete content of a hyphenated word
	WC	Word Confidence: Confidence level of the OCR results for this string. A float value between 0 (unsure) and 1 (confident)
	сс	Confidence level of each character in that string, A list of number between 0 (confident) and 9 (unsure) for each character

	STYLEREFS	Text style used for this string, if it is different from the parent text block style
	STYLE	Any combination of font style (italics, bold,)
	ALTERNATIVE	(element) Any number of alternative string to be used
Illustration	ТҮРЕ	A user defined description of the type of the illustration
	FILEID	A link to a separation file that contains just the illustration
ComposedBlock	TYPE	A user defined description of the type of the composed block
	FILEID	A link to a separation file that contains just the composed block

10 Referencing ALTO files from METS

In a METS file's logical <structMap> element, at the level where the <div>s can no longer be decomposed because the finest granularity is reached, the link to the right position in the right ALTO file has to be made. (See 6.3.3 Logical <structMap>: An example in Section One)

This is done by using a <fptr> element which holds an <area> element with the following structure:

```
<fptr>
<area BETYPE="IDREF" FILEID="[ALTO_ID]" BEGIN="[ALTO_BEGIN_IDREF]" />
</ftpr>
```

where **[ALTO_ID]** is the ID attribute given to the corresponding <file> element in the <fileGrp ID="ALTOGRP" USE="Text"> grouping element of <fileSec>

and **[ALTO_BEGIN_IDREF]** is the ID of the element (defined in that ALTO file) which corresponds to <area>.

If more than one different ALTO element contain the data corresponding to a single bottom-level <div> element of the logical structMap, then all those ALTO elements have to be referenced. This is usually achieved by declaring an <area> element for each ALTO element that must be referenced, and encapsulating those <area> elements in a <seq> element.

Example:

11 ALTO Example

Some examples of ALTO files are available for download from http://downloads.bnl.lu/schemas/samples.zip

Section Three - The batch.xml file

12 File batch.xml: requirements

Along with all the METS, ALTO, and image files, we require the delivery of one XML file called *batch.xml*, whose purpose it is to provide an overview of all the monographs making up the complete set that had to digitized.

This file must link to every single METS file, as well as to every one of the "complete monograph"-.PDF files in the set, and have the following structure:

```
<?xml version="1.0" encoding="UTF-8"?>
<batch xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<document>
<title>[BIBREC_245a]</title>
<monographNumber>[BIBREC_SYS_NUM]</monographNumber>
<monograph_path>[MONOGRAPH_PATH]</monograph_path>
<files>
<mets>[METS_FILE_LOCATION]</mets>
<pdf>[MONOGRAPH_PDF_LOCATION]</pdf>
</files>
</document>
.....
</document>
</batch>
```

where

- [BIBREC_245a] is the full title of the monograph
- [BIBREC_SYS_NUM] is the system number of the monograph.
- [MONOGRAPH_PATH] is the root path for that monograph (the directory that the METS and "complete monograph"-.PDF files are contained in) relative to the root path for the batch.
- [METS_FILE_LOCATION] is the filename of the METS file for that monograph
- [MONOGRAPH_PDF_LOCATION] is the filename of the .pdf file containing all the pages of the monograph

Please use Unix-style relative paths.



Example:

Appendix C

Author: zarko.stojkovic@bnl.etat.lu Version: 1 Date: 03/12/14

Append	ix C: Color codes	174
1.1	COVER_SECTION	174
1.2	TITLE_SECTION	174
1.3	STATEMENT_SECTION	174
1.4	FRONTISPIECE	174
1.5	BASTARD_TITLE_SECTION	175
1.6	TABLE_OF_CONTENTS	175
1.7	INDEX	175
1.8	PREFACE	175
1.9	BIBLIOGRAPHY	176
1.10	ADVERTISEMENT	176
1.11	APPENDIX	176
1.12	CHAPTER	176

Appendix C: Color codes

1.1 COVER_SECTION

PUBLISHING_STMT	Green
PARAGRAPH	Black
ILLUSTRATION > IMAGE	Violet
ILLUSTRATION > CAPTION	Pink
ILLUSTRATION > AUTHOR	Red
BOOKPLATE	Yellow
HEADLINE	Orange
SUBHEADLINE	Light blue
AUTHOR	Red

1.2 TITLE_SECTION

TITLE_OF_WORK	Orange
CREATOR_OF_WORK	Red
PUBLISHING_STMT	Green
STATEMENT	Black
ILLUSTRATION	Violet
BOOKPLATE	Yellow

1.3 STATEMENT_SECTION

PUBLISHING_STMT	Green
PARAGRAPH	Light blue
HEADLINE	Yellow
AUTHOR	Red

1.4 FRONTISPIECE

IMAGE	Violet
CAPTION	Black
AUTHOR	Red

1.5 BASTARD_TITLE_SECTION

ŀ	IEADLINE	Orange

1.6 TABLE_OF_CONTENTS

TABLE_OF_CONTENTS	Green
CAPTION	Orange
DATA	Yellow
TEXTBLOCK	Light blue
PAGE_NUMBER	Dark blue
FOOTNOTE	Violet
FOOTNOTE > DATA	Red
FOOTNOTE > DATA > ITEM_ID	Black
FOOTNOTE > DATA > TEXTBLOCK	Pink

1.7 INDEX

INDEX	Black
CAPTION	Red
DATA	Brown
TEXTBLOCK	Pink
PAGE_NUMBER	Green

1.8 PREFACE

PREFACE	Brown
TITLE	Orange
AUTHOR	Red
BODY	Yellow
BODYCONTENT	Dark blue
PARAGRAPH	Light blue
ILLUSTRATION	Violet

1.9 BIBLIOGRAPHY

BIBLIOGRAPHY	Dark blue
CAPTION	Orange
DATA	Yellow
TEXTBLOCK	Green
AUTHOR	Red
PARAGRAPH > TEXT	Light blue
ILLUSTRATION > IMAGE	Violet
ILLUSTRATION > CAPTION	Black

1.10 ADVERTISEMENT

ADVERTISEMENT	Violet

1.11 APPENDIX

APPENDIX	Yellow
TITLE	Orange
PARAGRAPH	Light blue
FOOTNOTE > DATA > ITEM_ID	Brown
FOOTNOTE > DATA > TEXTBLOCK	Black
CHAPTER	Green
TITLE	Dark blue
CHAPTER > BODY > BODY_CONTENT >	Light blue
PARAGRAPH	

1.12 CHAPTER

TITLE	Orange
SUBTITLE	Dark blue
AUTHOR	Red
BODY_CONTENT > PARAGRAPH	Light blue
FOOTNOTE	Red square
FOOTNOTE > DATA > ITEM_ID	Brown
FOOTNOTE > DATA > TEXTBLOCK	Black
ILLUSTRATION > IMAGE	Violet
ILLUSTRATION > CAPTION	Black
ILLUSTRATION > AUTHOR	Red
TABLE > IMAGE	Yellow
TABLE > CAPTION	Black
TABLE > AUTHOR	Red

Appendix D

Examples of how to structure a monograph

Author: zarko.stojkovic@bnl.etat.lu Version: 1 Date: 03/12/14

Table of contents

Introduction	
Color codes	
Example 1: Mitteilungen der Kommission für landw. Versuche	
Example 2: Die Kleinstatdt. Eine kulturpsychologische Studie	
Example 3: Les croix de pierre du Grand-Duché de Luxembourg	203
Example 4: Die wahre biblische Religion	
	Introduction Color codes Example 1: <i>Mitteilungen der Kommission für landw. Versuche</i> Example 2: Die Kleinstatdt. Eine kulturpsychologische Studie Example 3: Les croix de pierre du Grand-Duché de Luxembourg Example 4: Die wahre biblische Religion

1 Introduction

This document explains by color drawings how to structure a monograph.

2 Color codes

The color codes that are used can be found in Appendix B: Color codes

3 Example 1: Mitteilungen der Kommission für landw. Versuche

Here we are in the FRONT matter.

This is the COVER_SECTION. It contains an ILLUSTRATION, a HEADLINE, a PUBLISHING_STMT, and some TEXTBLOCKs.



Then comes the TITLE_SECTION with an ILLUSTRATION, TITLE_OF_WORK, a PUBLISHING_STMT and two STATEMENTs.


Then the PREFACE with some PARAGRAPH TEXT.



At this point, the first CHAPTER starts. It contains only the TITLE of the chapter.



On the next page, a sub-chapter begins. It contains a TITLE and some PARAGRAPH TEXT



This next page contains a sub-chapter of the CHAPTER before with a TABLE, a CAPTION of TABLE and an AUTHOR of TABLE

L.		urbe en öße	bili- nur- ihre eife	19e- 3fr. fr.	
Australian Collins .	Bemerkunge	Per Serfuch mu anf drei Parzell von je 10 ar Gr ausgeführt. För bie Rental färberechnung u tärberechnung u tärberechnung u 1913 gettenbe im Sa 1913 gettenbe im Sa nommen: Rartoffeln, 5 fr. 40%iges Ratifals,1 Ratkfidefforff, 30			
ate.	Eewinn pro ha in fr gegenüber einer einfei- figen Thomas- mehldüngung. fr	176	80	iniu di	
orte: Up t) d	Geldwert der Ernke nach Aldzug der Koften für die Spesial- düngung. fr	1001	905	825	
hmboden. S	Geldwerf der Ernte pro ha. fr	1115	086	825	
teiniger Eel	Ernfemenge pro ha in Dz.	223	196	165	
Bodenart: S	Roffen der Spezial- düngung pro ha. fr	75 39 { 114	22	negative Training	
ater are wrapping a appropriate	Besondere Düngung pro ha neben der üblichen Thomas- mehldüngung.	250 kg Kalkfiickfioff 300 kg 40%,01900 Ralifalz	250 kg Ralkflickfloff	Shne befondere Düngung (Nur die orfsübliche Thomas- mehldüngung).	

The next page continues with another TABLE without a CAPTION or an AUTHOR. After the TABLE comes some PARARAGRAPH TEXT.

		•				
	- 1	29 —				
en node of the line		Geldwert der Ernte pro ha nach Abzug der Kosten für die Düngung.				
Vodenarf und Gegend.	Angebaute Sorte.	bei Solldüngung.	bei Düngung nur mif Kali und Phos- phorjäure.	bei Düngung nur mit Phosphor- fäure.		
Leichte Gandböden des Gutlandes.	Magnum bonum	fr 1300	fr 1195	tr 1.010		
	Magnum bonum	1580	1295	1020		
Schwere Böden des Gutlandes.	Up to date und Wohltmann	1854	1599	1218		
In den befferen Rartoffelböden des Gutlandes.	Magnum bonum	1700	1595	1390		
	Magnum bonum	1790	1575	1300		
Böden des Öslings.	Up to date und Wohltmann	1742	1495	1258		

Objchon die Ernteerträge, besonders was "Magnum bonum" anbetrifft, in diesem Jahre ziemlich gering ausgefallen find, hat sich die Spezialdüngung dennoch sehr gut bezahlt gemacht. Aller= dings hatten wir in diesem Jahre ganz besonders hohe Kartoffel= preise, wodurch eine auffallend hohe Rentabilität troch der gerin= gen Erntemengen sich ergab. Aber selbst bei den niedrigen Kar= toffelpreisen des Jahres 1913, in welchem nur halb so hohe Preise als 1914 bezahlt wurden, wäre troch der geringen Ernte eine Beidüngung mit Kalisalz und Stickstoffdünger noch sehr lohnend gewesen.

Aus den eingegangenen Berichten wurde festgestellt, daß verschiedene Bersuchsansteller diesen Kartoffeldüngungsversuch in einem mit Stallmist reichlich gedüngten Felde ausgesührt hatten.

Hierdurch war die erste Parzelle (Volldüngung) zu stark mit Stickstoff gedüngt worden, während die zweite Parzelle in

This page contains some PARAGRAPH TEXT and an IMAGE

	5. Als durchichnittliche Düngermengen pro ha kann r
anfi	ihren:
	a) 600 kg Thomasment leves zweite Juhr bber, 1049 die
	h) 600 kg Soinit indes zweite Jahr ober zweckmäßiger
	ka Roinit jährlich.
	c) 150 kg eines der genannten Stickstoffdünger jedes d
	Jahr, falls als Erjatz hierfür Kompost oder Jauche 1
	in ausreichender Menge gegeben wird.
Los	Die Kosten dieser rationellen Volldüngung betragen pro
had	a) für Thomasmehl pro Jahr 12 Fr.
1994	b) für Kainit pro Jahr 12 Fr.
	c) fur Stickstoffdungung pro Jahr,
	auts maji mit scompoli ober Julicie
1.	Jusammen 39 Fr.
14.9	
	es die Selber, mit ichneren follerninge nicht zu galig

2.22

The next page contains some PARAGRAPH TEXT and a FOOTNOTE with an ITEM_ID and a DATA TEXTBLOCK element.

Eine Samenmischung als feststehendes Rezept kann natür- lich nicht gegeben werden, da dabei die Art des Bodens, Lage usw. mit berücksichtigt werden müssen. Bir raten daher allen Landwirten an, sich bei einer Neuanlage an die Uckerbauverwal - tung zu wenden, welche auf Grund der von der
verlugsfelderkommtiftion gefammerten Ets fahrungen jederzeit zuverläffige Angaber über zweckmäßige Grasmifchungen koftenlos übermittelt. Allgemein jei jedoch noch hervorgehoben, daß die Saat= menge pro ha einfchließlich Weißkleebei= mifchung etwa 30 kg betragen muß. Die Koften einer folchen Grasmifchung belaufen fich auf etwa 50 bis 70 Fr. pro ha. Die nachftehenden Zufammenstellungen geben einige An= haltspunkte über Grasmifchungen: ¹)
Für schwere Böden:
A. Grafer:Englifdes Rangras15-18 kgBiefenfdwingel4-7 "Rifpengras3-4 "Rammgras2-3 "Roter Edwingel1-3 "25-35 kgB. Rleearten:
Beißklee 5 kg

Here the page starts with some PARAGRAPH TEXT followed by a new sub-CHAPTER followed by some PARAGRAPH TEXT.

Daß unzufriedenstellende Resultate mit einseitiger Mais= oder Gerstenschrotfütterung erzielt werden, ist jedem, der einigermaßen mit den Fütterungsbedürfnissen vertraut ist, ohne weiteres einleuchtend.

82

XI. Demonstrationen, Versuche und Musteranlagen mit Maschinen, Geräten und baulichen Einrichtungen.

Die Kommiffion für landwirtschaftliche Bersuche hat stets den Neuerungen auf dem Gebiete des Maschinen = und Ge = rätewesens sowie der baulichen Einrichtungen besondere Beachtung geschenkt und war immer bestrebt, dieselben auf ihre Brauchbarkeit zu prüfen und das Wertvolle den Land= wirten möglichst zugänglich zu machen.

Ju diefem Zwecke werden die zweckmäßig erscheinenden Maschinen, Geräte und andern Einrichtungen angekauft und den Lokalvereinen leihweise zur Versügung gestellt. Auf diese Weise werden Neuerungen in den verschiedensten Gegenden des Landes und unter den mannigsaltigsten Verhältnissen von unsern Land= wirten auf ihre Brauchbarkeit und Zweckmäßigkeit erprobt.

So wurden in den letzten Jahren Versuche u. Demonstrationen veranstaltet mit Spritmaschinen zur Desinfektion der Stallungen, zur Bekämpfung der Blatt= krankheiten der Kartoffel, der Obstbaumschäd= linge und ber Sommergetreide unkräuter (Sederich und Ackersenf), ferner mit neuen Bflug=, Eggen= und Rultivatorkonstruktionen, mit Rartoffelsor= tier=, Kartoffellege= und Kartoffelerntema= schinen, mit Drillmaschinen, Schrotmühlen, Trockenfütterungsapparaten, Selbsttränk = vorrichtungen, Kleetrockengerüften, Ablade= vorrichtungen, Bentilationsanlagen für Stal= lungen, Krippenbauten, Feldscheunen, Schutz= hütten für Biehweidenanlagen, Stockrobema= schinen, sowie mit Dörrapparaten, Abbeerma= schinen, Obst= und Traubenmühlen, Keltern und verschiedenen kleinern für die Landwirt=

Then comes the TABLE_OF_CONTENTS

Inhaltsverzeichnis.	
worf	eute.
ebnisse der Düngungsversuche	11
A. Düngungsversuche zu Kartoffeln	14
B. " Runkelrüben	31
C. " " Salmfrüchten	38
D. " Stehmeiden und Futterflächen .	44
A Sortenanhaumerinche zu Rartoffeln	51
B. Runkelrüben	53
C. """Safer	55
D. " " Beizen und Roggen	58
E. Anbauversuche zur Neuansaat von Viehweiden	59
Schädlinge : I. Versuchsergebnisse über die Bekämpfung von Hedruck, Ackersenf und ähnlichen Unkräutern	65
II. Versuche mit gebeiztem Saatweizen gegen Brand.	67
III. Bekämpfung der Krautkrankheit der Kartoffeln	69
IV. Versuche zur Bekämpfung der Schnecken in den jungen Gaaten	69
schiedene Versuche:	
I. Düngungsversuche mit unmagnetischem Conventor-	73
itaub	74
II. Dungungsverjuche mit Pytonomy	74
IV Rhosphate Bernard	74
V. Serjuchsergebniffe mit Gründüngungspflanzen	75
VI. Düngungsversuche zu Samenrunkelrüben	75
VII. Berfuchsergebniffe über das Abblättern der Runkelrüben	78
VIII. Impfpersuche mit Nifragin	170
IX. Berjuche zur Förderung des Garten- und Gemufebaues.	19
X. Schweinemastversuch mit automatigier Ltoutenan- ferung, ausgeführt im Herbste 1913 auf der Staats- domäne Givenich	80
XI. Demonstrationen, Versuche und Musteranlagen mit	82
Badauhantan	87
CODDERRATION.	93

And after the TABLE_OF_CONTENTS comes a last CHAPTER with some PARAGRAPH TEXT



4 Example 2: Die Kleinstatdt. Eine kulturpsychologische Studie

The first page is a COVER_SECTION with a HEADLINE, an AUTHOR, a PUBLISHING_STMT and a STATEMENT_SECTION.



Then comes a TITLE_SECTION with a TITLE_OF_WORK and a STATEMENT.



After the TITLE_SECTION comes the first CHAPTER with some PARAGRAPH TEXT.



- 5 -

lege ich biefem Brundfehler meiner Darstellung keine wesentliche Bedeutung bei. Die Soziologie mag sich noch so missenschaftlich gebärden, ihr Tatsachengebiet schneidet zu tief ins persönliche Erleben ein, als daß das forschende und wer= tende Individuum von dem gefühlsmäßigen Ber= hältnisse, das es nun einmal zum Dargestellten hat, restlos abstrahieren könnte. Ich halte es deshalb für ehrlicher und nützlicher, anstatt mit wissenschaftlicher Pose das individuelle Erleben zu verhüllen und zu entschuldigen, frei heraus. - fo parador das auch klingen mag, — dem fühlenden und wollenden Menschen sein Recht zu geben und einmal das Umgekehrte wahr zu machen: daß die Exaktheit auf beiden Seiten fein kann und daß der Impressionismus manches zu enthüllen vermag, was der behutsamen In= duktion nie aufgeht.

So will ich denn nicht auf Wissenschaftlichkeit Anspruch erheben; ich bin sogar überzeugt, daß bei Problemen wie das vorliegende, die Wiffen= schaft und die Wissenschaftlichkeit so gut wie gar nichts vermögen. Wenn ich darauf halten müßte, meiner Darstellung irgend einen bestimmten Charakter zu geben, wäre es mir leicht, das Banze eine Mischung von empirischer Sozial= psychologie und Sozialethik zu nennen. Ich mache auf diefen Charakter auch keinen Anspruch und ließ mich in manchen Zeilen zu sehr von einem rein künftlerischen Wohlgefallen an den Erscheinungen selbst und an deren Formulie= rung leiten, um für diefe Arbeit irgend eine Be= beutung erzwingen zu können, die über die einer ehrlichen und rein fubjektiven philosophisch=

Here starts a new CHAPTER with some PARAGRAPH TEXT. a tetsdatofissa a Grundlegung. 2 tieg retto m In der Gesellschafts= und Bölkerpsychologie kann man felten eine Erscheinung unbedingt auf eine Raufalreihe, auf einen bestimmten Ent= stehungsgrund zurückführen. Und wenn man das auch kann, so ist doch die spezifische Indivi= bualität biefer Erscheinung wieder von so vielen Faktoren abhängig, daß der Soziologe in seinen Erwägungen am klügsten verfährt, wenn er sich mit der Umschreibung und Wertung der verschie= denen vorhandenen Möglichkeiten begnügt. Benn wir mit einigermaßen Sicherheit auf die Frage: Wie entsteht eine Kleinstadt? ant= worten wollen, bleibt uns nichts anderes übrig, als diefe Bildung in ihrem Verhältnis zu andern Bildungen zu betrachten; aus den verschiedenen Fähen, die wir auffinden, ergibt sich bann von selbst vie mögliche Rausalreihe, an deren Ende die Bildung: "Kleinstadt" steht. So find in einer Hinsicht manche Kleinstädte stehen gebliebene Großstädte; das will sagen: Stäbte, die in frühern Perioden eine Bedeutung hatten, die, abgesehen von Umfang und Ginwoh= nerzahl, irgendwie der entsprach, bie in unserm Gegenwartsleben unsere Großstädte haben. In Belgien, unter anderem speziell in ben Flandern, ist die Städteanlage= und =verteilung seit dem Ausgang des Mittelalters ganz verschoben und alte Zentren, die innerlich und äußerlich Bultur= träger= und verbreiter waren, sind zu Markt=

meisten prinzipiellen Arbeiten in der Soziologie, so sehlen auch hier die Borarbeiten.

- 10 -

Aus der Genefis der Kleinftadt ergibt sich nach dem Gesagten schon von seldst, daß sie keine seste Form, kein gesellschaftliches Urelement ist; sie ist eine übergamgsersche dei nung, bleidt vielleicht in alle Zeiten als solche bestehen und nimmt einen Charakter an, der scheindar unveränderlich ist, kann aber trotzem nicht als abgeschlossenes, sestgegliedertes Kulturganzes betrachtet werden. Ihr schlt mithin vor allem die Aus= prägung, die Bodenständigkeit, der einheitliche Charakter; sie hat in allem die Eigenschaften einer Zwitterersche in ung and es bleidt uns nachzuweisen, daß jede ihrer Kußerungen von diesem Grundcharakter bestimmt wird.

De mare von größtem Interesse das Gervel einer Rieinstodt ingenbrielcher Art an dieler Stelle genate zu verfolgen und ankoulich dar justellen, oder damit märe so viel Raum verfan das in einer fummarischen Monogrophie wie die das in einer fummarischen Monogrophie wie die gescheende, ein Mihverhältnis zwischen den ein zelnen Teilen einfründe. Ebenso interessan



une die genet iner Ateinstod

abstration aber fich zur Erohftabt ausnachfen. Diefe Ubergangewegesse fich aus erohortlich lehrreich – wie übrigens alle Ubergangeprozesse – und es til bedauerlich, daß wir nicht mehr difterliche Wenographien von Ortfchaften beden, difterliche Wenographien von Ortfchaften beden, dittailiche Wenographien von Ortfchaften beden, difterliche Wenographien von Ortfchaften beden, difterliche Wenographien von Berthovisten von Here again starts a new CHAPTER.

Der äussere habitus.

11 -

Die Boraussetzung einer Harmonie zwischen inneren Wesenskräften und äußever Erscheinung ist heute so allgemein angenommen, daß es banal wäre, für unseren speziellen Zweck die Ersistenz eines solchen Verhältnisses zu beweissen. Auf die= seines solchen Verhältnisses zu beweissen. seine Berhältnisses zu beweissen. Auf die= seines solchen Verhältnisses zu beweissen. nen Zeit, daß sie die äußeren Formen alles menschlichen Seins und Schaffens in Hinsicht auf die zugrundeliegenden Regungen der Gemein= schaftsseele werten lernte.

In keiner ihrer ungerungen gibt die Klein= ftadt ihren Zwittercharakter fo zu erkennen, wie in ihrer baulichen Anlage, Häufern und Gär= ten, Straßen und Plätzen. Es gibt wenige Dörfer, die fich in diefer Beziehung gegen bie Ein= heitlichkeit versündigen, und wenn man folche findet, so rührt das daher, daß sie im Begriffe find, sich zur Kleinstadt auszuwachsen oder daß Bestrebungen auftauchen, die zu manchen in der Rleinstadt geltenden parallel laufen. Die bau= liche Kultur eines Dorfes vollzieht sich gewöhn= lich in einer bestimmten Tradition, und diese Tra= dition ist, von den wenigen seltenen Ausnahme= fällen abgesehen, der jeweilige Ausdruck für die Bultur dieser Ortschaft, jedenfalls insofern sie Wohn= und Ausdruckskultur ist. Diese Tradition kann also verschiedenartig gewertet werden, obschon divekte Häßlichkeit von vorweherein aus= geschlossen ist; ein Abweichen von ber Tradition kann wiederum eine verschiedenartige Wertung

schon vor so was zunück. Und in dieser Ge= dankenordnung bürfte man auch an einem alten schönen Plaz nichts ändern. Was tut man je= doch? Man baut alles Mögliche hinein und ba= van; man reißt ab und verputzt; wo schatten= spendende Roßkastanien ihre Blütenkandelaber in der Sonne erglänzen ließen, erhebt sich einer jener prozigen Kaften, an dem alles Fassade ift und der hintere Teil einigermaßen zweckmäßig aussieht, weil dort die obligatorische Bedürfnis= anstalt angebracht wurde. Ich kenne für ein neues öffentliches oder unöffentliches Bebäude schon aus andern Rücksichten heraus keine un= geeignetere Stelle als einen alten Plat; benn die alten Kleinstadtplätze sind räumlich meist so beschränkt, daß sie nach der "modernen" Ber= änderung aussehen wie ein blondes kleines Mäbchen, dem man einen Küraffierhelm aufgestülpt hat. Jst das neue Bauwerk dabei noch sehens= wert — und das ist selten genug — so muß man sich schon auf den Rücken legen, wenn es auf die Augen wirken foll.

15

Genau dassselbe geschieht mit Plätzen, die keine eigentlichen öffentlichen Plätze find. Jedes Dorf hat ursprünglich Baumwiesen, die wie das grüne Herz der Ortschaft aussehen. Wer seine Kindheit auf dem Dorfe verbrachte, weiß was solche Baumwiesen allein schon für die Kinder wert sind; und wer als ausgewachsener Mensch das Dorf besucht, hat hier einen rechten Augentrost. Mit der Wiese aber ist's vorbei, wenn das Dorf sich zur Kleinstadt auswächst. Die Talmispekulation, die alles Miserable des Großstadtterroinwuchers an sich hat, nur nicht bessen A new CHAPTER starts again.



jpeziell wegen diefer Kleinigkeiten — ebenjo fehr über die Achfel anfieht, wie er es mit dem Land= bewohner tut.

37 -

bie speite titto aangi and fahroose, and ohne atti

Was ich bis dahin alles über Philisterium er= zählt habe, was ich an den Zuständen der Klein= stadt geißelte, gilt in meinen Augen nicht als übertrieben. Es wäre jedoch ungerecht, wenn ich nicht den Revers der Medaille zeigen würde. Und alles in allem hat die Kleinstadt so viele Annehmlichkeiten, daß es begreiflich ist, wenn ihr auch heute noch ernsthafte Lobredner erstehen und wenn vollwertige Menschen im Kleinstadtleben Befriedigung finden. Eine Tatjache, die jeden einseitigen Ankläger aufklären muß ift die, daß es jo viele Dichter der Kleinstadt gibt. Ich meine hiermit nicht diejenigen Darsteller, die sich schöp= ferisch mit den spezifischen Lebensproblemen der Rleinstadt beschäftigen, sondern die starken und wahrhaften Künstler, welche die eigenartige Poe= fie der kleinen Städte erfaßt haben und in fatten Farben miedergeben.

Wenn wir dahin gelangt find, die Kleinstadt als Zwittererscheinung, als Mittellage zwischen Großstadt und Dorf aufzufaffen, wenn wir von dieser Stellung alle Unvollkommenheit des Rleinstadtlebens ableiteten, so verlangt die ele= mentarste Gerechtigkeit, daß wir die aus dieser Ronstatation sich ergebenden Werte nicht ver= schweigen. So bietet das Provinzstädtchen mehr als das Dorf in allem, was man Lebenskomsort nennt. Man liegt an einer oder an mehreren Eisenbahnlinien; für Wasser in Hause ist gesorgt Here starts a new CHAPTER with a TITLE and a SUBTITLE.



The monograph finishes with a blank cover page.



5 Example 3: Les croix de pierre du Grand-Duché de Luxembourg

The first page is a COVER_SECTION with a HEADLINE, an AUTHOR, a PUBLISHING_STMT, some TEXTBLOCK and an ILLUSTRATION



Then comes a BASTARD_TITLE_SECTION with a HEADLINE.



Then the TITLE_SECTION starts with a PUBLISHING_STMT.



Followed by another TITLE_SECTION page containing the TITLE_OF_WORK, the AUTHOR, the PUBLISHING_STMT and some STATEMENT.



The first CHAPTER starts here.



Saints, ou par des anges en adoration, ou par deux soldats porteurs de la lance et de l'éponge imbibée de fiel. Il y a d'autres variantes. La Crucifixion est accompagnée d'emblêmes et d'accessoires divers, têtes d'ange, instruments de la Passion, ciboires, ostensoirs, etc., placés à divers endroits dans l'édicule, sur le chapiteau ou au sommet du fût.

Au bas du fût se trouvent assez fréquemment la figure d'un saint, le patron du donateur ou de la localité sans doute ; par exemple saint Nicolas, dont le nom est très répandu dans le pays, saint Wendel ou Wendelin, etc. ; ou bien un sujet complémentaire, tel que le Christ descendu de la Croix et couché sur les genoux de Marie, la tête du Christ imprimée sur le voile de sainte

werdange, et dont il a fait la description détaillée, reproduite dans l'article ci-dessus de 1850.

Nous devons d'utiles renseignements à ce dernier, ainsi qu'à notre confrère, M. Jos. Schwartz.

Ces croix sont différentes aussi bien par leurs formes que par leur destination de celles signalées par H. SCHUERMANS dans son étude sur les anciens chemins et mo uments dans les Hautes Fagues, et dont les unes servaient de délimitation, les autres étaient commémoratives de quelque accident. Outre ces croix de pierre, de autres étaient commémoratives de quelque accident. Outre ces croix de pierre, de dimensions restreintes, (deux pieds de hauteur au plus), il en signale de plus grandes en bois, hautes de 3 ou 4 mètres, méthodiquement placées de distance en distance « pour l'adresse des passants ». V. Bulletin des commissions royales d'Art et d'Ar-chéologie, tomes, X, XXIV et XXV ; v. spécialement t, XXV (1886) p. 184. Les croix du Grand Duché ont dans l'aspect général une analogie éloignée avec certaines croix de cimetière et autres de la fin de l'époque ogivale, reproduites par de Caumont, Abécédaire d'Archéologie, t. II, p. 708 et suivantes, et par VIOLLET-LE-DUC, Dictionnaire raisonné de l'architecture française, t. IV, au mot Croix. L'on connaît les croix et les calvaires de Bretagne. V. Voyage pilloresque et roman-tiene deux France, nar I. TAVLOR, CH. NODER et ALPH. DE CALUAUX.

tique dans l'ancienne France, par J. TAYLOR, CH. NODIER et ALPH. DE CAILLAUX,

Paris, 1845.
A comparer encore avec les « Croix d'Évangélisation » de la Moravie. V. un article de A. FRANTZ, initulé Alle Steinkreuze und Kreuzsteine in Mähren, dans les Mitheilungen der K. K. Central-Commission zur Erforschung und Erhaltung der Kunst-und bistorischen Denkmale, Vienne, 1893, t. XIX, 2^e livraison, p. 106.
II a existé aussi un grand nombre de croix et il en subsiste encore en Angleterre. V. Encyclopedia of architecture by JOSEPH GWILT. Londres, Longmans, 1867, édit, revue par WYATT-PAPWORTH, à la page 1183, article Cross. V. aussi Encyclopedia Britannica, 9e édition, vol. VI, Edimbourg, 1877, au même mot. On y distingue des croix de marché, croix de cimetière, croix des chemins. Beaucoup sont ou étaient de véritables monuments, empruntant le nom de croix à leur couronnement. Telle est celle de Charing Cross, reproduction moderne d'une des neuf croix de la reine Eléonore. Telle est aussi la City Gross de Winchester, dont nous donnons plus Join la description.

loin la description.

A SUB-CHAPTER starts here. Some FOOTNOTES are included in this page.







- 13

15. Esch sur l'Alzette, (marché).

Dans un édicule à contours déchiquetés, le Christ sur la Croix, entre deux anges agenouillés, dont l'un présente le calice. L'édicule est porté sur un pilastre que décorent la tête du Sauveur dans un mascaron, et au-dessous la lance et la pique avec l'éponge. Sur le pied, inscription : Errichtet auf Kosten Jacobs Gobels, 1827. Le socle est revêtu d'une plaque en fonte avec ornements.

16. Eyschen.

Trois croix, représentant le Christ entre les deux larrons, élevées sur un soubassement continu. Ce sont des croix simples sans couronnement. Celle du milieu, peu heureusement remplacée en 1856, figurait le Christ. Les deux autres ont cette particularité bizarre que les bras, au lieu d'être horizontaux, sont inclinés dans la direction de la croix centrale. L'un des larrons dirige son regard vers le Sauveur, l'autre s'en détourne ; leurs corps se tordent dans des positions tourmentées et avec une expression de souffrance d'un rendu étrange.

17. Même village, dans une chapelle.

La sculpture formant le couronnement d'une ancienne colonne a été recueillie et dressée sur une table d'autel. Elle représente l'Elévation du Christ en croix ; un bourreau enfonce un clou dans la main droite du Sauveur ; dans le fond les saintes Femmes, un personnage porteur d'un étendard, et un cavalier. Le tout est curieusement doré et polychromé.

18. Même village.

Sur un pilastre, cartouche représentant Jésus enfant, conduit par deux personnages (la Vierge et Ste Anne ?); au-dessus, l'image symbolique de la sainte Trinité,

Certaines de ces croix avaient une destination spéciale, révélée par leurs inscriptions.

* *

La Croix blanche de Grevenmacher⁴ a vraisemblablement été adossée à un cimetière.

1 Voir ci-dessus nº 3.



A SUB-CHAPTER starts here.





6 Example 4: Die wahre biblische Religion

The first page is a COVER_SECTION with a HEADLINE, a SUBHEADLINE and a PUBLISHING_STMT.


Then comes a TITLE_SECTION containing a TITLE_OF_WORK, an AUTHOR and a PUBLISHING_STMT.



Followed by some STATEMENT.



The first CHAPTER starts here. The CHAPTER has no TITLE.



7 Diefer Thatsache, daß man ihr so entschieden miß= trauet, ben Beweis entnehmen, daß Biele wirklich ben Krallen bes Satans verfallen find und bag fie nicht eifrig genug bestrebt fein tann, die Jugend Die= fen Belialsföhnen zu entziehen? 3war finden wir es etwas übertrieben, wenn unfer Apostolischer Bifar fcmerglich ausruft: "Wir fegnen und werben ver= "flucht, wir bulden und werden verfolgt, wir beten "und werden geläftert; wir find wie der Ausfehricht "ber Welt geworden und wie ein Schmußlappen für "Jedermann; ja wir werden um Chrifti willen auf "ben Tod gedrängt den ganzen Tag und find ge= "achtet wie Schlachtschaafe" - benn in einer Beit, wo anderthalb Millionen Gläubige von ihren Bir= ten geführt, bas Symbol ber Einen wahren Rirche ungestört verehren konnten, in einer Beit, wo unter ben Augen unferes Apostolischen Bifars fich jährlich breißigtaufend Gläubige versammeln, um den ,,boch= feierlichen Triumphzug mit dem Gnadenbilde unferer himmlischen Serrinn und Mutter Mariä" zu begeben, einem Bifchof, bem unfer protestantischer Rönig=Groß= berzog fo großartig alles Berlangte, fogar bie vollige Freiheit zur Ernennung ber Dbern und Leb= rer des Seminars (alfo wohl ohne Allerhöchft fich auch nur bas Bestätigungsrecht ber Lehrer vorzube= halten) bewilligt bat, in einer folchen Zeit und unter einem folchen König scheint es uns übel angebracht,

9 Klerus ober biefes und jenes Bifchofs zu fprechen, wer aber bamit bas Räthfel gelöft zu haben glaubt, täuscht sich gewaltig und begehet in unfern Augen eine wirfliche Todfünde. Warum follte benn ber Klerus oder diefer und jener Bifchof, deffen Leben ohnehin aller weltlichen Freuden entbehrt, fo febr auf's herrichen halten? Warum follte ber Klerus, bie einzige weltliche - wenn man sie eine weltliche nennen darf - Freude, die ihm erlaubt ift, noch von fich weisen und es nicht vorziehen, ein geliebter als ein gefürchteter hirte zu fein? Bewirkt ja die Liebe auch den Geborfam, mabrend Furcht nur ben Schein deffelben hervorbringt? Leidenschaften, wie Serrichfucht u. dgl., tonnen fich wohl eines Indivi= buums, nie aber eines gangen Standes bemächtigen. 200 ein ganzer Stand einmüthig und offen, wie die fatholifche Geiftlichfeit unferer Tage, auftritt, ba muß etwas viel Höheres, als gewöhnliche Leidenschaften, ba muß es eine göttliche 3dee fein, für die fie in ben Rampf tritt. Es ift mahr; noch vor we= nigen Jahren schienen folche Kämpfe, wie wir fie jest feben, unmöglich. Uber muß bas gerade für Die Serrichsucht bes jetigen Klerus beweifen? Rann es nicht auch von einer unverantwortlichen Schlaff= beit des frühern Zeugniß ablegen, der feine Pflicht, bem Rechte der Kirche und bem geiftigen Wohl ber Gläubigen nichts zu vergeben, verfannt hatte? Wenn 1**

Here starts a new CHAPTER with a TITLE "I" and a SUBTITLE "Das katholische Dogma."



21 fie bem Menschen vermöge feiner irdischen, wenn auch noch von eigentlicher Sünde reinen natur ichon möglich ift, fondern in einer folchen, wie fie nur möglich wird, wenn Gott dem Menschen noch böhere als reinmenschliche Unlagen, wenn er ihm überna= türliche, übermenschliche Geiftesträfte ichenft. Abam, lautet ber Ausbrud bes Dogma's, befaß ein ju feinen naturgaben bingugegebenes Ge= fcent ber göttlichen Gnade. Denn ber Gottes= Dienft, die Seligfeit, behauptet der Ratholif, fann boch nur in ber Gemeinschaft mit Gott bestehen. Wie kann aber die endliche, wenn auch noch fo fün= denreine Natur, das Unendliche erfaffen ? 3wischen Endlichem und Unendlichem, zwischen Gott und Mensch, bestehet an sich eine folche Kluft, bag ber Mensch nie Gott erreichen fann, wenn Gott nicht durch ein wunderbares Gnadengeschent, burch Mittheilung eines höhern Geiftes, es dem Menfchen möglich gemacht hätte, jene Kluft zu überfpringen. Der Mensch muß alfo zu feiner menschlichen natur noch von Anfang an eine höhere natur als Gnabengeschent Gottes befeffen haben. Durch bie erfte Sünde bes Menschen wurde nun nicht bloß bie ganze natur bem Berderben preisge= geben und ber Mensch bem leiblichen Tobe, fondern mit biesem ftarb er auch ben geiftigen, indem er jenes göttliche Gnadengeschent verlor und obendrein fein

49 gebührt ihnen, vom Tode erlöft zu werden. Jeder wird baber feinen urfprünglichen Leib wieder erhalten, und nachdem ber herr bie Gläubigen von ben Ungläubigen geschieden und biefe bem ewigen Höllenfeuer überliefert, werden jene mit bem Berrn in bas himmlische Jerufalem einziehen, bort mit bem herrn in ewiger Seligfeit bie Siegesfreuden feiner Rirche genießen. Das ift bas, wenn man will, herrliche Gebäube ber fatholischen Lehre. Ein Stein paßt ba auf ben andern, ift eng an ben andern gefugt. Alles ift aus Einem Ouf, Alles aus Einer 2Burgel bervorge= gangen. Giebt man biefe zu, fo ift man fchlechter= bings genöthigt, das Ganze als Wahrheit anzunehmen. Es ftehet baber Denjenigen unferer Beit, bie es mit Niemandem verderben wollen, bie einerfeits ber Aufflärung und Bildung bes Jahrhunderts bulbigen, anderseits sich boch für katholisch gläubige Chriften ausgeben, ich fage, es ftehet biefen schlecht an, bas, was ihnen am tatholischen System nicht zufagt, als nicht zum achten Ratholizismus gehörig, als nur im Intereffe einer berrichfüchtigen Priefterichaft erfunden zu verschreien. Coon barin, baß fie von einer herrichfuchtigen Priefterichaft fprechen, - man bedenke biefe Sünde gegen ben beiligen Geift! - beweisen fie, daß fie ber tatholischen Rirche nicht mehr angehören. Der fatholische Priefter fann zwars 3



Here starts a new CHAPTER with the TITLE "II" and a SUBTITLE "Die welthistorische Stellung der katholischen Kirche"









The monograph finishes with a COVER_SECTION containing some TEXTBLOCKs.

