

ECP-2008-DILI-518001

BHL-Europe

Catalogue of content holder requirements
(quality, quantity, accessibility, standards,
specifications of content and metadata)

Deliverable number	<i>D2.1</i>
Dissemination level	<i>Restricted</i>
Delivery date	<i>10 August 2009</i>
Status	<i>Final</i>
Author	<i>Bernard Scaife</i>



eContentplus

This project is funded under the *eContentplus* programme¹,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.

Table of contents

0	DOCUMENT HISTORY.....	3
0.1	<i>Contributors</i>	3
0.2	<i>Revision History</i>	3
0.3	<i>Reviewers and Approvals</i>	4
0.4	<i>Distribution</i>	4
1	PURPOSE.....	5
2	BACKGROUND AND CONTEXT.....	5
2.1	<i>Project approach</i>	5
3	IMAGING REQUIREMENTS.....	5
3.1	<i>Basic principles</i>	5
3.2	<i>Resolution, bit-depth and file format</i>	5
3.3	<i>Post-processing</i>	6
4	METADATA REQUIREMENTS.....	6
4.1	<i>Granularity</i>	6
4.2	<i>Intellectual Property Metadata</i>	6
4.3	<i>Title level</i>	6
4.4	<i>Item level</i>	8
4.5	<i>Page level</i>	9
4.6	<i>Creator</i>	10
4.7	<i>Rights</i>	10
5	FILE SUBMISSION GUIDELINES.....	11
5.1	<i>Files to be supplied</i>	11
5.2	<i>Delivery method</i>	12
5.3	<i>Worked Example</i>	12

0 Document History

0.1 Contributors

A discussion about the specifications was open on the BHL wiki and the following persons provided input that was used for the present document.

Person	Partner
Henning Scholz	MfN
Bernard Scaife	NHM
Nancy Chillingworth	NHM
Richard Sipek	NMP
Walter Koch	AIT
Francisco Welter-Schultes	UGOE
Tom Gilissen	NAT
Patricia Mergen	RMCA
Antonio Valdecasas	CSIC
Graham Hardy	RBGE
Suzanne Pilsk	SIL
Chris Freeland	MOBOT
Kimmo Koskinen	UH-Viikki
Uwe Mueller	UBER

0.2 Revision History

Revision Date	Author	Version	Change Reference & Summary
24 June 2009	Bernard Scaife	0.1	Initial draft version for internal review
30 June 2009	Bernard Scaife	0.2	Updated draft following internal review
6 July 2009	Judith Daehne*	0.2	Notes and questions for further revision
10 July 2009	Bernard Scaife	0.3	Updated following review by JD and PMG/TMB
5 August 2009	Bernard Scaife	0.4	Updated following further internal review
10 August 2009	Bernard Scaife	1.0	Final version after approval of PSG

* Judith Daehne is working for the Virtuelle Fachbibliothek Biologie (Virtual Library Biology) at University Library Frankfurt/Main.

0.3 Reviewers and Approvals

This document requires the following reviews and approvals. Signed approved forms are kept in the project file.

Name	Position	Signature on approval	Date	Version
Henning Scholz	BHL-Europe PCO			1.0

0.4 Distribution

This document has been distributed to:

Group	Date of issue	Version
BHL-Europe consortia	25 June 2009	0.1
Henning Scholz	30 June 2009	0.2
BHL-Europe consortia	10 July 2009	0.3
BHL-Europe consortia	6 August 2009	0.4

1 Purpose

To provide a minimum specification for the supply of metadata and digital image files to the BHL-Europe project. In some cases, a preferred standard that exceeds the minimum is stated, but it is not obligatory to follow this.

2 Background and Context

BHL-Europe sites have been scanning for some time as have external non-BHL institutions who may have relevant material to contribute. Before we devise a method to ingest this material, we need to decide on what is considered the minimum acceptable technical criteria for assets which may be relevant to the project. This is deliverable 2.1 in Work Package 2 of the BHL-Europe project.

2.1 Project approach

The project lifecycle follows four main stages as outlined below. Each stage is considered complete on sign-off of key deliverables.

Stage 1: Requirements Definition Stage

The primary output of this stage is this Project Initiation Document, which captures all of the information required to begin blueprinting the detailed specification of the products outlined in the 'Scope' section of this document.

3 Imaging Requirements

3.1 Basic principles

In line with the BHL, BHL-Europe has endorsed the recommendation to follow, in part, the Digital Library Federation's (DLF) "Benchmark for Faithful Digital Reproductions of Monographs and Serials".¹

3.2 Resolution, bit-depth and file format

Type	Resolution (minimum)	Resolution (preferred)	Bit-depth	File format
Bitonal	300dpi	600dpi	1-bit	TIFF
Grayscale	300dpi	600dpi	8-bit	uncompressed TIFF or lossless compressed image (e.g. LZW, JPEG2000 [* .jp2]).

¹ (currently available at <http://www.diglib.org/standards/bmarkfin.htm> with an archived copy at <https://bhl.wikispaces.com/file/view/bmarkfin.pdf> on the BHL Wiki

Color	300dpi	400dpi	24-bit	uncompressed TIFF, or lossless compressed images (e.g. LZW, JPEG2000 [*.jp2]).
-------	--------	--------	--------	--

NOTE: BHL will accept lossy files. In the case of JPEG2000, files with a compression level of 85% are acceptable.

3.3 Post-processing

Master images should be supplied without post-processing. These should be uncropped and where possible, the colour profile used during the scanning workflow should be embedded. These will be retained for storage and preservation purposes.

Additionally, cropped versions should be made available for embedding within the page-turner software wherever possible.

4 Metadata Requirements

4.1 Granularity

Metadata is to be recorded at three levels of granularity – Title, Item & Page. In addition, metadata for the Creator(s) of the title should be supplied.

4.2 Intellectual Property Metadata

It is also necessary to supply metadata relating to intellectual property issues pertaining to a title or an item. Such metadata will include the copyright status of a work (whether it is subject to copyright protection or in the public domain). Where a work is protected by copyright and permission has been sought from the right holder(s), partners must supply details of the license terms under which it is being made available through the BHL-Europe site.

It is desirable that licence terms are not infinitely variable but selected from a range of Open Access licensing solutions, most likely, Creative Commons licences. There may be exceptions whereby rights holders are unhappy to license a work under such terms and in these cases partners will need to supply accurate licensing information in the metadata of the object.

BHL is currently in the process of standardizing these intellectual property statements. More information on this and where it should be indicated in the metadata will be forthcoming and updated in this document.

4.3 Title level

4.3.1 Definition

A “Title” is described as the bibliographic authority for a single or multiple scanned Item(s).

4.3.2 Description

Name	Obligation	Occurrence	Description	MARC21 Mapping guidance ¹
Category	Required	Once	Monograph OR Serial component part	Leader pos 7
Sub-category	Optional	Repeatable	Free text e.g. Atlas, Biography, Encyclopedia, Scientific journal, popular journal	
Title	Required	Once	The title of the title digitized. Either plain text or a concatenation of relevant MARC 245 subfields.	245 \$a and \$c
Abbreviation	Optional	Repeatable	Standard abbreviation of the title.	210
Creator	Required	Repeatable	People or organizations responsible for creation of title.	100, 700
Description	Optional	Once	A description of the title.	500-599, except 506, 530, 540, 546
Publisher	Required	Once	The person or organization responsible for publishing the original object. A concatenation of relevant MARC 260 subfields.	260 \$b
Edition	Required – monographs	Once	The edition of the book being scanned	250 \$a and \$b
Start Date Published	Optional (serials)	Once	Year title began publication.	
End Date Published	Optional (serials)	Once	Year title ended publication.	
Date Indicated on Title Page	Required (monographs)	Once	Date indicated on title page	260\$c
Local Library Identifier	Optional	Once	Unique identifier for title in scanning library's catalog. (Allows for later resolution with ILS record)	
OCLC Number	Optional	Repeatable	OCLC Number(s) for original title.	035
ISSN	Optional	Repeatable	ISSN(s) for original title, used for serials/journals.	022
ISBN	Optional	Repeatable	ISBN(s) for original title, used for monographs.	020
Call Number	Optional	Repeatable	Shelf location of book in scanning library.	050, 090
Subject	Optional	Repeatable	Subjects assigned to work, as indicated in subfields within MARC 650.	650
Language	Optional	Repeatable	3 letter language code taken from bits 35-37 in MARC control field 008 or see list at http://www.loc.gov/standards/iso639-2/ISO-639-2_utf-8.txt	008/35-37, 041
Date Created	Required	Once	Date digital object was created or scanned.	

¹ This is not an exclusive set of MARC tags that may be applicable but intended as a guide.

Date Last Modified	Required	Once	Date digital object was last updated.	
Foreign Keys (TropicosID, IPNI PubID, TL2, BPH)	Optional	Repeatable	Identifiers for titles as recognized by popular indices.	

4.4 Item level

4.4.1 Definition

An “Item” is described as the physical bound object scanned within a Title.

4.4.2 Description

Name	Obligation	Occurrence	Description
Barcode	Optional	Once	The barcode affixed to the physical object scanned. Note: A library may have more than one copy of an item and this is where the correct copy is identified
Sequence	Required	Once	If a multivolume monograph or serial, the sort order for this item relative to the other objects scanned in this title.
Local Library Identifier	Required	Once	Unique identifier for item scanned, from scanning library catalogue. Note: Use same as local library identifier if this information is not easily accessible
Start Volume	Required – Serials (unless Start Date present) Optional – Monographs	Once	Beginning volume number of bound object.
End Volume	Optional	Once	Ending volume number of bound object.
Start Date	Required – Serials (unless Start Volume present) Optional – monographs	Once	Beginning year of bound object.
End Date	Optional	Once	Ending year of bound object.
Language	Optional	Repeatable	3 letter language code, if different from Language of associated Title.
Scanning Institution	Required	Once	Institution responsible for scanning.

Scanning Contributor	Required	Repeatable	Institution(s) assisting with scanning.
Scanning Sponsor	Optional	Repeatable	Institution sponsoring scanning.
Embedded Rights	Required only if they differ from title rights.	Once	Only required where relating to a specific item.
Date Created	Required	Once	Date item scanned.
Date Last Modified	Required	Once	Date item modified.

4.5 Page level

4.5.1 Definition

A “Page” is described as a single image file contained within a scanned Item.

4.5.2 Description

Name	Obligation	Occurrence	Description
FileName	Required	Once	File name of associated image file page.
Sequence	Required	Once	Sort order for given page, relative to other pages in item.
Page number	Optional	Repeatable	Physical printed page number of scanned page.
Page Type	Optional	Repeatable	A description of the contents of a page. One of: <ul style="list-style-type: none"> • Appendix • Blank • Cover • Illustration • Index • Map • Table of Contents • Text • Title Page
Year	Optional	Once	Year page was published.
Volume	Optional	Once	Volume in which page was published.
IssuePrefix	Optional	Once	Description of subdivision below Volume, if present in item. One of: <ul style="list-style-type: none"> • Issue • Part • Number • Heft
Issue	Optional	Once	Numeral in which page was published.
Embedded rights	Required only if they differ from	Once	Only required where relating to a specific page.

	title's rights.		
Date Created	Required	Once	Date page scanned.
Date Last Modified	Required	Once	Date page modified.

4.6 Creator

4.6.1 Definition

A "Creator" is defined as a person or corporate body responsible for the creation of the Title.

4.6.2 Description

Where possible an authority list such as NACO (Name Authority Cooperative Program ¹ or Personen-Namen-Datei (PND)² should be used when entering this metadata.

Name	Obligation	Occurrence	Description
Name	Required	Once	Name of person or organization responsible for creation of title. Last Name, First Name. An alternative is to enter (if relevant) "[Anonymous]" or [S. n.]
Role	Required	Repeatable	One of: Author Publisher Illustrator Added Author
DOB	Optional	Once	Date of Birth
DOD	Optional	Once	Date of Death
Biography	Optional	Once	Free text biography of creator.

4.7 Rights

4.7.1 Definition

Rights which persist within the object being digitised and licence under which digitisation is permitted. Where separate and more restrictive rights apply to component parts, these should be embedded within the appropriate component section(s) above.

¹ <http://www.loc.gov/catdir/pcc/naco/naco.html>

² http://z3950gw.dbf.ddb.de/z3950/zfo_get_file.cgi?fileName=DDB/searchForm.html

4.7.2 Description

Name	Obligation	Occurrence	Description
Rights status	Required	Once	One of “in copyright”, “out of copyright”, “public domain”
Rights owner	Required	Repeatable	
Organisation issuing licence	Required	Once	Default: “Rights owner”
Licence terms	Optional	Once	e.g. CC3.0 or CC0 CCPD etc
More than one rights holder	Required	Once	Default: 1

5 File Submission Guidelines

5.1 Files to be supplied

There are two options available:

Option 1 – METS¹ format file containing

- 1) MARC details,
- 2) Technical information
- 3) Administrative information
- 4) Page structure information (See worked example below for relationships to be captured).

Option 2

- 1) A MARC or MARCXML format document from your library catalogue for the digitized title, named with your local title identifier.
- 2) One folder for every book (item) scanned, named using the local library identifier.
- 3) Within each folder, master images files named using the local library identifier + 4-digit sequence number in form “identifier_XXXX.ext” where ext = file extension (see worked example below for relationships to be captured).
- 4) If unable to provide a MARCXML record, BHL staff will attempt to obtain an appropriate record following consultation in advance.

¹ www.loc.gov/standards/mets

5.2 Delivery method

5.2.1 Digital image files

The preferred delivery method for files is via external hard drives. Multiple DVDs, if properly labelled, are an acceptable alternative. For external hard drives, it is the responsibility of the contributor to ship the drive to BHL, but BHL will pay to return the drive. Though BHL will take care of the drives, it does not take responsibility for the handling in the mail.

5.2.2 Metadata

Any of the methods for digital image files will be accepted. In addition, availability via OAI PMH¹ is accepted.

5.3 Worked Example

The Library of the Natural History Survey (LNHS) would like to submit their scans of Linnaeus' *Species Plantarum*. This 2 volume title has a local title identifier of b12069590 in the LNHS system. Volume 1 has a local library identifier of 31753000802824 and Volume 2 has a local library identifier of 31753000802832. The following files & folders would be submitted to BHL:

An example of the metadata can be seen at:

<http://www.botanicus.org/MARCXML/b12069590.xml>

/B12069590.xml

 /31753000802824/

 /31753000802824_0001.tif

 /31753000802824_0002.tif

 ...

 /31753000802824_0566.tif

 /31753000802824_0567.tif

 /31753000802832/

 /31753000802832_0001.tif

 /31753000802832_0002.tif

 ...

 /31753000802832_0671.tif

 /31753000802832_0672.tif

¹ www.openarchives.org/OAI/openarchivesprotocol.html