

ECP-2008-DILI-518001

BHL-Europe

**Delivery of the final revised best practice
guidelines and standards -
Hints-working document**

Deliverable number	<i>D2.9</i>
Dissemination level	<i>public</i>
Delivery date	<i>30 April 2012</i>
Status	<i>Draft</i>
Authors	<i>Birhölmer Melita Hierschläger Michaela</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.

0 Document History

0.1 Contributors

Person	Partner
Michael Malicky	LANDOE
Fritz Gusenleitner	LANDOE
Tom Gilissen	NAT
Dennis Zielke	UBER
Jane Smith	NHM
Constance Rinaldo	MCZ
Laszlo Peregovits	HNHM
Antonio Valdecasas	CSIC
Marian Ramos	CSIC
Manuel Sanchez Ruiz	CSIC
Francisco Welter-Schultes	UGOE
Lee Namba	ATOS
Wolfgang Koller	NHMW
Graham Hardy	RBGE

0.2 Revision History

Revision Date	Author	Version	Change Reference & Summary
15.02.2011	M. Hierschläger	1	First compliment of chapters sent to Melita BIRTHÄLMER for first corrections
22.02.2011	Melita BIRTHÄLMER	2	Proofread of M. BIRTHÄLMER back to M. Hierschläger
25.02.2011	Jane Smith	3	Linguistic proofread by Jane Smith
01.03.2011	M. BIRTHÄLMER	4	BPG sent to Consortium plus selected people from BHL-US
21.03.2011	M. Hierschläger	5	Feedback incorporation
07.04.2011	M. Hierschläger	6	BPG revision
11.04.2011	M. BIRTHÄLMER	7	BPG revision

12.04.2011	M. Hierschläger	8	BPG review incorporation
------------	-----------------	---	--------------------------

0.3 Reviewers

This document requires the following reviews and approvals.

Name	Position	Date	Version
Youssef Bassily	Mikhail Head of Software and System Development Section, ICT Department, Bibliotheca Alexandrina	11.04.2011	4

0.4 Distribution

This document has been distributed to:

Group	Date of issue	Version
BHL-Europe consortium plus selected people from BHL-US	01.03.2011	4

1 Table of Contents

0	DOCUMENT HISTORY	2
0.1	CONTRIBUTORS.....	2
0.2	REVISION HISTORY	2
0.3	REVIEWERS.....	3
0.4	DISTRIBUTION.....	3
1	TABLE OF CONTENTS	4
2	PURPOSE	6
3	BACKGROUND	6
4	PRE-DIGITISATION SETUP	6
4.1	IMPORTING OF ALREADY EXISTING METADATA TO THE GLOBAL REFERENCES INDEX TO BIODIVERSITY 6	
4.2	WHAT TO DIGITISE?	6
4.2.1	<i>Collection Policy</i>	6
4.2.2	<i>Collection Analysis</i>	7
4.2.3	<i>IPR issues and how to tackle them</i>	7
4.3	CHOOSING BOOK/JOURNAL	7
4.4	CALCULATE THE TOTAL COSTS OF THE WHOLE PROCESS.....	8
4.5	CHOOSE THE FINAL BOOK/JOURNAL	8
4.6	METADATA.....	8
4.6.1	<i>Recording Metadata in database</i>	9
4.6.2	<i>Metadata Requirements – Metadata Bucket</i>	9
4.6.2.1	Granularity	9
4.6.2.2	Intellectual Property Rights/Metadata.....	9
4.6.2.3	Volume level	9
4.6.2.4	Article level.....	9
4.6.2.5	Page level.....	9
4.6.2.6	Creator.....	9
4.7	PUT METADATA ON HOMEPAGE.....	10
4.8	FUNDING OF THE DIGITISATION PROCESS	10
5	DIGITISATION WORKFLOW	10
5.1	PRELIMINARY STEPS	10
5.1.1	<i>Analysing and selecting the material</i>	10
5.1.2	<i>Preparing the scanning order</i>	11
5.1.3	<i>Checklist</i>	11
5.2	IMAGING REQUIREMENTS.....	11
5.2.1	<i>Basic principles</i>	11
5.2.2	<i>Resolution recommended, Bit depth</i>	11
5.2.3	<i>Post-processing</i>	11
5.3	SCANNING.....	11
5.4	OCR.....	11
5.5	ARTICLE LEVEL ACCESS.....	11
5.6	QUALITY CONTROL	11
5.7	FURTHER REFERENCES	12
6	PREPARATION OF DATA FOR BHL-EUROPE	12
6.1	SET UP TECHNICAL INTERFACE/PROVIDER SOFTWARE INTERFACE.....	12

6.2	DATA HARMONIZATION	12
6.2.1	<i>File submission Guidelines</i>	12
6.2.1.1	Files to be supplied.....	12
6.2.1.2	Example.....	12
6.2.1.3	Delivery method.....	12
6.2.1.3.1	Implementing an OAI-PMH web service.....	12
6.3	SCHEMA MAPPING TOOL	12
6.4	PRE-INGEST	13
6.5	INGEST.....	13
6.6	POST OR RE-INGEST.....	13
6.7	INGEST TO EUROPEANA	13
7	GENERAL INFORMATION.....	13
7.1	WORKFLOW EXAMPLES	13
7.1.1	<i>Example from the Georg-August Universität Göttingen Stiftung Öffentlichen Rechts, Göttingen (Germany)</i>	14
7.1.2	<i>Example from the Natural History Museum Library, London, UK</i>	14
7.2	WORKFLOW AND TOOLS.....	14
7.2.1	<i>Individual steps in the digitisation process</i>	14
7.2.2	<i>Workflow management software Goobi</i>	14
7.2.3	<i>Workflow management software DAF</i>	14
7.3	GENERAL ARCHIVATION – PERMANENT STORAGE.....	14
	ABBREVIATIONS	16
	FIGURES.....	20
	TABLES.....	21
	REFERENCES.....	21
	APPENDIX.....	26
	A: EXAMPLE OF A MODEL BILATERAL CONTRACT	26
	B: PRE-INGEST FILE SUBMISSION GUIDELINES	27

In the following the “Delivery of the first version of the approved best practice guidelines and standards” is referred to as “BPG” (Best Practice Guide) or “Approved Best Practice Guidelines and Standards”.

2 Purpose

The BPG aims to be an “easy to follow” guide step by step from either a print or digital version of an original publication to the final digitised form of that publication in the BHL-Europe portal.

Commentary. Is there any brief guideline about e.g. minimum requirements for upload to BHL-Europe which can be mentioned as a requirements list? (second paragraph)

3 Background

4 Pre-Digitisation setup

The Approved Best Practice Guidelines and Standards start here with important steps which should be done prior to the real digitisation workflow, which is described in chapter 5 “Digitisation Workflow”.

4.1 *Importing of already existing Metadata to the Global References Index to Biodiversity*

Short explanation of the GRIB – have a look at published Deliverables can be copy and pasted and adjusted from **D2.5** (D2.7)

Importing existing metadata to the GRIB is a very important thing to do as it avoids literature being digitised twice. It ensures the proper management of literature digitisation.

An explanation how this is done can currently not be given because the GRIB was still in the test phase at the point of writing.

4.2 *What digitise?*

4.2.1 Collection Policy

Insert outcome of collection policy – WS Tervuren!!!!

Maybe use this for Collection Policy?

Example :

I Congreso Iberoamericano del Medio Ambiente (1975). Three volumes with biological, physical and chemical contributions. Criteria for inclusion? Whole work? Only biodiversity related papers? – answers should follow in Collection Policy

Discuss also scan volumes where not solely biodiversity is published

4.2.2 Collection Analysis

Shorten for midi
 further content input
 Practical approach
 Literature from own library
 Literature from external libraries

Useful references

Carolina M.-Albadalejo: Entomological Bibliography of Spanish authors, 1758 - 2000

This chapter provides an answer to the main question: “What shall be digitised?” in the wide scope of biodiversity; and its surrounding themes.

4.2.3 IPR issues and how to tackle them

Shorten and simplify for midi – for spee include a practical example perhaps of LandOE
 The chapter should be that way that it can be solely taken and be sent to future CP.
 Simplify this chapter for midi

Intellectual Property Rights (IPR) questions may quickly arise after content for digitisation has been chosen.

This chapter aims to identify and to clarify the IPR factors that need to be considered when selecting and digitising titles for BHL-Europe.

Perhaps make an extra example of Land OÖ way of how they deal with IPR: is a bit provided by the example below:

What are bilateral contracts and what are they for? Answers are provided in the following paragraph:

Why are bilateral contracts entered?

Periodicals finance themselves with sales; the exchange of works facilitates the development of one’s own library. Hence the editors categorise the issue of digitisation and the making available of PDFs as a sensitive matter. A written agreement, including stipulations about all conditions and the at-will terminability which is signed by the editor and the website provider is of utmost importance in order to clarify why, in which way, and based on which agreement, PDFs may be presented on the Internet. This makes the editors feel secure and encourages them to grant the rights to make their periodicals publicly available. An example of a model contract can be found in the Appendix on page 26.

Of help can also be chapter 4.7 Put Metadata on homepage in the strategic right gaining procedure.

4.3 Choosing Book/Journal

Is the old chapter of D2.6 but inclusive check GRIB – can perhaps be taken out of D2.5

How to check the GRIB - write an explanation on how to do it – D2.5

Before you choose the final item – also check the GRIB if your chosen item is still available and not already scanned or being scanned.

Hence if you have considered the GRIB, IPR issues you can start with cost calculation.

4.4 Calculate the total costs of the whole process

For midi shorten for spee leave it like it is; For midi explain the main points which should be considered in cost calculation

Inclusion of further example

This chapter demonstrates the calculation of costs for the whole digitisation process by referring to the experiences of two BHL-Europe partners.

In this regard may also be chapter 4.9 funding of the scanning of interest. This chapter is at the end of pre digitisation setup because the point at which one start considering the funding may be at various stages of the working order and each institution has to decide on its own when to do it. However two ways are presented at which stage the considerations on funding can take place.

4.5 Choose the final Book/Journal

New chapter – but not much to write here – just strategic explanation of working steps

After cost calculation you can reconsider whether you want to digitise your chosen item or not. The working steps that follow now are very work intensive and at this point of the procedure you should mark your chosen item in the GRIB as being scanned to avoid duplication. “Description how signing literature to the GRIB should be inserted here.” – Can be taken out of **D2.5**

When you have done the previous points you should start to think about Metadata on the technical side but also how they can help you in terms of IPR which is explained in chapter 4.7 Put metadata on homepage. This can be a strategic move to gain the favour of rightsholders.

4.6 Metadata

How to record metadata

For midi short description how to make metadata

OLEF description

Consistent use of OLEF terms in the whole BPG

Term definition clearance – what is a volume, journal, article, page in the OLEF schema...

Shall be a real guideline step after step how to make the Metadata and how to further proceed with the metadata – I guess the further proceeding is chapter 6 pre-Ingest – shall be sustainable so that a lot can be done in-house – so as not to give money to external enterprises.

Schema/standard definition in terms of metadata

1. **Scheme/schema:** is a formal description of a standard

Standard: agreement of things verified by an authority

MDA – model driven architecture - <http://www.mda.org.uk/spectrum.htm> link does not work

The metadata chapter provides some general aspects as well as an example and requirements for recording metadata.

4.6.1 Recording Metadata in database

4.6.2 Metadata Requirements – Metadata Bucket

Where are metadata stored?

4.6.2.1 Granularity

“Nevertheless, a relatively low cost way of solving part of this problem is provided by BHL-Europe automatically enriching the metadata with taxonomic terms.” – [keep an eye on this if this is really going to happen](#)

4.6.2.2 Intellectual Property Rights/Metadata

4.6.2.3 Volume level

Has to be renamed after OLEF

4.6.2.4 Article level

Has to be renamed after OLEF

4.6.2.5 Page level

Has to be renamed after OLEF

Recording on this level is a huge effort and one should weigh the cost-benefit ratio. The page level metadata allows access to the targeted page for a scientific research directly, but in most situations this can be done also very quickly on article level. On this level the following information is required:

- 1) Page number
- Etc.

4.6.2.6 Creator

Has to be renamed after OLEF

4.7 Put Metadata on homepage

In this chapter it is explained why it will be helpful if metadata are provided on the institutional homepage with regard to receive funding and in the respect of building up a well structured offer of digitised literature.

4.8 Funding of the digitisation process

2 Methods:

1. Method: hunt for money before the start of digitisation (before you start to work down the pre-digitisation setup working steps)
2. Method: follow all the pre-digitisation setup working steps and then go “hunting” for money

Further examples??!!!

Several possibilities and examples with better or less chances are given by our partner LANDOE for the funding of scanning hereafter:

- 1) Overheads from other projects – As our Technical/Scientific staff do the administrative work for the projects themselves, we get the overheads as additional free budget. E.g.: A project with worth 120.000 Euros with 20 percent overhead included raises our funds by 20.000 Euro that is approx. 150.000 pages with OCR.
- 2) Internal budget – if your department head is asking for some ideas at the end of a fiscal year, raise your hands!
- 3) National funds from public bodies. In Austria Museums can ask for additional budgets (10.000-20.000 Euros) at the Ministry of Culture for digitisation processes.
- 4) Very rare international projects with funds for such work.

5 Digitisation workflow

In this chapter we will try and describe the basic steps of a digitisation project for books and / or journals. How you will actually organise each of the steps will depend on your own

5.1 Preliminary steps

5.1.1 Analysing and selecting the material

After parts of your collection have been identified as possible candidates for digitisation, and collection analysis and deduplication matters have been investigated thoroughly, the identified books and journals need to be carefully assessed.

5.1.2 Preparing the scanning order

Do not reference to the Catalogue of Content holder requirements

5.1.3 Checklist

A bit confusing is this checklist – make it more understandable

5.2 Imaging requirements

5.2.1 Basic principles

5.2.2 Resolution recommended, Bit depth

5.2.3 Post-processing

According to the “Catalogue of content holder requirements”
https://bhl.wikispaces.com/file/view/BHL-E_2pt1_20090805.pdf² “

5.3 Scanning

5.4 OCR

Latin and Arabic ocr features already included – perhaps include ocr features for other languages like Cyrillic or Greek.

D3.7

5.5 Article level access

Figures shall be in English and not in german – problem has anyone an English version of the dobe acrobat – he/she shall remake the figure

This chapter is based on the experienced BHL-Europe partner LANDOE and aims to provide help, new ideas and work patterns about developing access at article level.

5.6 Quality control

“What about general quality control for the processing and the OCR quality? This section only focuses on quality control for article level access” – Suggestion by the official reviewer.

This chapter draws on the experience of LANDOE to describe a three step quality control workflow.

First quality step: enhance the Article level data of the Author(s) with biographical data, so one can more easily distinguish and match author names (full names, address).

² Non-BHL partners will gain access to these sites when they become a content provider of the project.

Second quality step: Names of article level PDFs have a schema that has to match the metadata; otherwise the article level PDF will not be shown in the web:

Series code, number of volume, start page and end page are cross checked here.

Third quality step: The quality control person has to read through the web table of contents and check that the volume or part is complete.

5.7 Further References

6 Preparation of data for BHL-Europe

For see more elaborate and exact guideline how this works pre ingest etc.

Really describe all necessary working steps – do not state “ups” at the moment not available

For midi just what is really necessary

D3.7 can be used in this regard

6.1 Set up technical interface/provider software interface

6.2 Data harmonization

6.2.1 File submission Guidelines

6.2.1.1 Files to be supplied

6.2.1.2 Example

6.2.1.3 Delivery method

6.2.1.3.1 Implementing an OAI-PMH web service

in Europeana semantic elements schema

in BHL-Europe schema OLEF

6.3 Schema Mapping Tool

The schema mapping tool will be open source and documentation will be available before the end of the project

Is there a schema-Handbook?

6.4 Pre-Ingest

[Handbook](#)
[Is D3.7 of help?](#)

A content provider is asked to log into the Pre-Ingest module to trigger the conversion of all metadata into the BHL-Europe schema based on the previously defined configuration file (see chapter 6.3 for more details) by selecting the folder(s) for procession via the Pre-Ingest interface. The Pre-Ingest module will also provide up to date information on the status of the processing and the results.

More detailed information on the Pre-Ingest module can only be given in the next BHL-Europe Deliverable 2.9 (final revised BPG due 30.04.2012), as to date of writing the module was in progress.

6.5 Ingest

6.6 Post or Re-Ingest

6.7 Ingest to Europeana

7 General Information

[What is with Collection Management and Curation – what was meant initially with this?](#)
[Perhaps this is now collection Policy?](#)

[What is with the chapter “Keep the information up to date”](#)

[For midi leave out the whole chapter 7 – not really necessary for the working steps](#)
[Simply for midi leave out everything that comes after 6.7 – because then the practical workflow is over](#)

7.1 Workflow examples

[Further examples](#)

This chapter will show workflow examples from BHL-Europe partners.

7.1.1 Example from the Georg-August Universität Göttingen Stiftung Öffentlichen Rechts, Göttingen (Germany)

7.1.2 Example from the Natural History Museum Library, London, UK

7.2 Workflow and Tools

7.2.1 Individual steps in the digitisation process

Include Arabic ocr tools

This chapter provides an overview of software tools and solutions that may be used as automated workflow systems to control and document each step in a digitisation process.

7.2.2 Workflow management software Goobi

This chapter gives a general overview of the scanning workflow management software Goobi³.

7.2.3 Workflow management software DAF

This chapter gives a rough overview of the workflow management software DAF.

7.3 General archivation – permanent storage

For midi version leave out:

Abbreviations – the midi version should be written that no abbreviations section is needed and if though only really a short one

Figures

Tables

References (if yes then only short – no link collection)

Appendix

³ <http://www.digiverso.com/en/products/goobi>



Abbreviations

ABBYY: is a commercial company, which provides software for OCR

ABCD: Access to Biological data Collections <http://rs.tdwg.org/dwc/index.htm> – A data and metadata specification for the exchange of biological collections and observations data.

AIP (Archival Information Package): An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

ANSI: American National Standards Institute

API: Application Programming Interface

ATOS: Atos Origin Integration France (Paris)

BA: Bibliotheca Alexandrina

BDS: BHL Deduplication Subset

BHL: Biodiversity Heritage Library

BHL-E: Biodiversity Heritage Library Europe

BHL-Europe: Biodiversity Heritage Library Europe

BHL Open Access: Images on the portal are available for viewing and reuse with no charges attached. See the Berlin declaration: http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf Link to BHL Copyright and Licensing statement: <http://biodivlib.wikispaces.com/Licensing+and+Copyright>

Bitmap: or raster graphic is a data structure representing a generally rectangular grid of pixels, or points of colour, viewable via a monitor, paper, or other display medium. Raster images are stored in image files with varying formats.

Link: http://en.wikipedia.org/wiki/Bit_map

BPG: Best Practice Guide (shortcut for “Delivery of the first version of the approved best practice guidelines and standards”), also referred to as “Approved best practice guidelines and standards”.

BSD: Berkeley Software Distribution is a version of the operating system Unix.

Colour depth: or bit depth is the number of bits used to represent the colour of a single pixel in a bitmapped image or video frame buffer. This concept is also known as bits per pixel (bpp), particularly when specified along with the number of bits used. Higher colour depth gives a broader range of distinct colours.

CSDGM: Federal Geographic Data Committee Content standard for Digital Geospatial Metadata <http://www.nbi.gov/> - National Biological Infrastructure Initiative metadata standard including geographical metadata.

CSIC: Consejo Superior de Investigaciones Científicas (Madrid), Spain

CVS: Concurrent Versions System or Concurrent Versioning System is a client-server free software revision control system in the field of software development. Version control system software keeps track of all work and all changes in a set of files and allows several developers (potentially widely separated in space and/or time) to collaborate.

DAF: Digital Assets Factory

DC: Dublin Core is a standardised metadata element set which provides small and fundamental groups of txt elements.

DFG: Deutsche Forschungsgemeinschaft (German Science Foundation), Germany

DIP (Dissemination Information Package): The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS.

DLF: Digital Library Federation, USA

DPI (Dots Per Inch): is a measure of spatial printing or video dot density, in particular the number of individual dots that can be placed in a line within the span of 1 inch (2.54 cm). The DPI value tends to correlate with image resolution, but is related only indirectly.

DwC: Darwin Core <http://rs.tdwg.org/dwc/index.htm> – An extension of the DC for biodiversity information.

EAD: Encoded Archival Description <http://www.loc.gov/ead/> – Mostly used by archives.

EDIT: European Distributed Institute of Taxonomy

EOL: Encyclopaedia of Life, a project to create an online reference source and database for every one of the 1.8 million species that are named and known on this planet. <http://www.eol.org/>

EUROPEANA: Europeana is a search platform to a collection of European digital libraries with digitised paintings, books, films and archives. The project was initiated by the European Commission.

EXIF: Exchangeable image file format is a specification for the image file format used by digital cameras and scanners. The specification uses the existing JPEG, TIFF Rev. 6.0, and RIFF WAV file formats, with the addition of specific metadata tags. It is not supported in JPEG 2000, PNG and GIF.

EZOLO: Early Zoological Literature Online

FTE: Full-time equivalent

Gallica: Bibliothèque numérique, Gallica digital library – Over a million books and documents... <http://gallica.bnf.fr/>, France

GBV: Gemeinsamer Bibliotheksverbund (Common Library Network, a public non-profit institution by seven northern German federal states Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen and the Stiftung Preußischer Kulturbesitz) <http://www.gbv.de> →VZG

GPL: General Public Licence

GRIB: Global References Index to Biodiversity

GVK: Gemeinsamer Verbundkatalog (GBV Union Catalogue at <http://gso.gbv.de>, a multimaterial bibliographic database, comprises the library holdings of 7 northern German federal states, →GBV)

HNHM: Hungarian Natural History Museum (Budapest), Hungary

IA: Internet Archive (Universal access to all knowledge) <http://www.archive.org>

ILS: Integrated Library System

Image resolution: describes the detail an image holds. The term applies to digital images, film images, and other types of images. Higher resolution means more image detail.

http://en.wikipedia.org/wiki/Image_resolution

IPR: Intellectual Property Rights

IPTC: Information Interchange Model is a file structure and set of metadata attributes that can be applied to text, images and other media types. It was developed in the early 1990s by the International Press Telecommunications council (IPTC) to expedite the international exchange of news among newspapers and news agencies.

ISO: International Organization for Standardization; the primary international standards development organization.

JSF: JavaServer Faces is a Java-based application Framework to building web-based user interfaces

JPEG: Joint Photographic Experts Group, is an image format that used method of compression for digital photography

JPEG2000: is an image format for raster graphics with image compression.

KVK: <http://www.ubka.uni-karlsruhe.de/kvk.html> Karlsruher Virtueller Katalog (Karlsruhe Virtual Catalogue) (KVK) offers a uniform user interface for simultaneous research in union catalogues located around the world. Germany

LANDOE: Land Oberösterreich (Oberösterreichische Landesmuseen, Linz, Österreich), (Upper Austrian State Museums, Biology Centre), Austria

LDAP: Lightweight Directory Access Protocol is an application protocol for reading and editing directories over an IP network.

MAB2: Machine exchange format for metadata used by libraries, especially in the German library system.

MARC: MACHine-Readable Cataloging, standards for the representation and communication of bibliographic and related information in machine-readable form

MARC21: is a standard bibliographic metadata format, see also MARC; a formatting, record structure, and encoding standard for electronic bibliographic cataloguing records developed by the Library of Congress. The "21" refers to the version of MARC issued in 1998 that integrated the U.S. and Canadian versions of MARC.

MBLWHOI Library: Marin Biological Laboratory Woods Hole Oceanographic Institution Library, USA

MCZ: Museum of Comparative Zoology Harvard University, USA

MfN: Museum für Naturkunde (Berlin), Germany

Metadata: structured information that describes, explains, locates, and otherwise makes it easier to retrieve and use an information resource.

METS: Metadata Encoding and Transmission Standard; is an XML-based container for metadata; a metadata scheme for complex digital library objects

MODS: Metadata Object Description Schema; is an XML-based bibliographic description schema and a derivative of MARC 21; a metadata scheme for rich description of electronic resources.

MySQL: is the most popular relational database management system.

NAT: Stichting Nationaal Natuurhistorisch Museum Naturalis (Leiden), Netherlands

NHM: Natural History Museum (London), UK

NHML: Natural History Museum Library London, UK

NHMW: Naturhistorisches Museum Wien (Museum of Natural History Vienna), Austria

NISO: National Information Standards Organization (United States of America); a standards development organization, accredited by the American National Standards Institute that develops library and information related standards.

OAI: Open Archives Initiative, develops and promotes Standards for Web Content Interoperability.

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting

OAIS: Open Archival Information System

OCLC: Not for profit computer service and research organization whose systems help libraries locate, acquire, catalogue, and lend library materials. <http://www.oclc.org>

OCR: Optical Character Recognition; the electronic identification and digital encoding of scanned, printed or handwritten characters by means of an optical scanner and specialized software.

OPAC: Online Public Access Catalogue, is an online catalogue of a library collection that is available to the public

PDI: Preservation Description Information

PHAIDRA: is the acronym for Permanent Hosting, Archiving and Indexing of Digital Resources and Assets <https://phaidra.univie.ac.at/>

PHP: Hypertext Pre-processor - a scripting language

PICA: The PICA format is the internal data format for cataloguing in “Gemeinsamer Bibliotheksverbund” (GBV)

PNG: Portable Network Graphics, is a raster graphics image format

PPI (Pixels Per Inch): or pixel density is a measurement of the resolution of devices in various contexts; typically computer displays, image scanners, and digital camera image sensors.

http://en.wikipedia.org/wiki/Pixels_per_inch

http://en.wikipedia.org/wiki/Dots_per_inch

PPN: Pica Production number

PREMIS: Preservation Metadata. Implementation Strategies, is an international working group concerned with developing metadata for use in digital preservation.

QA: Quality Assurance

qDC: Qualified Dublin Core, more detailed Dublin Core

RARA: rare works stacks of the SUB Göttingen, Germany

RBGE: Royal Botanic Garden Edinburgh, UK

RDF: Resource Description Framework, is a standard model for data interchange on the Web

REST: Representational State Transfer

RSLP: Collection-level description <http://www.ukoln.ac.uk/metadata/cld/> – General, covers collections of all subjects, domains and types.

SBPK Berlin: Staatsbibliothek zu Berlin - Preussischer Kulturbesitz

SDK: Software Development Kit is a set of development tools that allow an application to be created.

Semantics: The names and meanings of metadata elements.

SIP (Submission Information Package): An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.

SMB: Server Message Block is a network protocol that you can use to provide shared access to files, printers and serial ports.

SOAP: Simple Object Access Protocol, is a simple XML-based protocol to let applications exchange information over HTTP, <http://www.w3schools.com/soap/default.asp>

SPARQL: Is the recursive acronym for SPARQL Protocol and RDF Query Language.

SPECTRUM: <http://www.mda.org.uk/spectrum.htm> -- Describes museum objects.

SQL: Structured Query Language, is the database language of data manipulation (e.g. update, delete) and database creation (e.g. create)

SUB Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen, Germany

TAR: Tape ARchive is a method to unite several single files in one file. No compression takes place.

TEI: Text Encoding Initiative <http://www.tei-c.org/> – Digital texts.

TGZ: A TGZ file is a TAR (Tape ARchive) compressed with GZIP.

TIFF: Tagged image file format

Tip ins: Loose leaf page (or pages) that have been added later; usually containing corrections.

UBER: Humboldt-Universität zu Berlin, Germany

UGOE: Georg-August-Universität Göttingen, Germany

URL: Uniform Resource Locator; a unique address for identifying and locating a resource on the Internet.

URN: Uniform Resource Name, identifies a resource or unit of information independent of its location.

VZG: Verbundzentrale des GBV – Head office of the Common Library Network, Germany

XML: Extensible Markup Language; a set of rules for encoding documents in machine-readable form.

XMP: Extensible Metadata Platform is a standard, created by Adobe Systems Inc. for processing and storing standardised and proprietary information relating to the contents of a file.

Z39.50: is a client–server protocol for searching and retrieving information from remote computer databases; National Information Standards Organization Z39.50 Information Retrieval Protocol (Z39.50/ISO 23950), a computer protocol that can be implemented on any platform, defines a standard way for two computers to communicate for the purpose of information retrieval; a NISO and ISO standard protocol for cross-system search and retrieval. Officially, international standard, ISO 23950, Information Retrieval (Z39.50): Application Service Definition and Protocol Specification, and ANSI/NISO standard Z39.50.

ZR: Zoological Record, general database for zoological literature

Figures

Fig. 1: Limits of database in relation to search criteria..... **Fehler! Textmarke nicht definiert.**

Fig. 2: Criteria for choosing content to be included in BHL-Europe**Fehler! Textmarke nicht definiert.**

Fig. 3:Risk band and licensing guide **Fehler! Textmarke nicht definiert.**

Fig. 4: short summary of the journal **Fehler! Textmarke nicht definiert.**

Fig. 5: Screenshot database www.zobodat.at **Fehler! Textmarke nicht definiert.**

Fig. 6: Example **Fehler! Textmarke nicht definiert.**

Fig. 7: Bitmap **Fehler! Textmarke nicht definiert.**

Fig. 8: <http://www.landesmuseum.at/biologiezentrum/> online presentation of individual articles belonging to a volume of a journal **Fehler! Textmarke nicht definiert.**

Fig. 9: window divided into two horizontal halves (first step)**Fehler! Textmarke nicht definiert.**

Fig. 10: second step **Fehler! Textmarke nicht definiert.**

Fig. 11: third step **Fehler! Textmarke nicht definiert.**

Fig. 12: fourth step **Fehler! Textmarke nicht definiert.**

Fig. 13: fifth step **Fehler! Textmarke nicht definiert.**

Fig. 14: sixth step **Fehler! Textmarke nicht definiert.**

Fig. 15: seventh step **Fehler! Textmarke nicht definiert.**

Fig. 16: eighth step **Fehler! Textmarke nicht definiert.**

Fig. 17: Processing Gemini Requests (screenshot) **Fehler! Textmarke nicht definiert.**

Fig. 18: DAF Architecture **Fehler! Textmarke nicht definiert.**

Fig. 19: OAIS **Fehler! Textmarke nicht definiert.**

Tables

Tab. 1: Risk bands.....	Fehler! Textmarke nicht definiert.
Tab. 2: Cost spreadsheet example BHL-US.....	Fehler! Textmarke nicht definiert.
Tab. 3: elements identified by Europeana.....	Fehler! Textmarke nicht definiert.
Tab. 4 elements, requirements, rights	Fehler! Textmarke nicht definiert.
Tab. 5: elements, requirements, rights	Fehler! Textmarke nicht definiert.
Tab. 6: General condition factors.....	Fehler! Textmarke nicht definiert.
Tab. 7: Detailed condition factors.....	Fehler! Textmarke nicht definiert.
Tab. 8: Commonly used bit depth rates http://en.wikipedia.org/wiki/Color_depth	Fehler! Textmarke nicht definiert.
Tab. 9: Resolution and Bit depth.....	Fehler! Textmarke nicht definiert.
Tab. 10: Chart how many books in a given shipment should be QA'd	Fehler! Textmarke nicht definiert.

References

Non digital References:

- BROCKE v. J. & A. SIMONS (2008): Towards a Process Model for Digital Content Analysis – The Case of Hilti. Available at: [http://ecom.fov.uni-mb.si/proceedings.nsf/0/4c1351e08e8d0f87c1257482003c3852/\\$FILE/44Brocke.pdf](http://ecom.fov.uni-mb.si/proceedings.nsf/0/4c1351e08e8d0f87c1257482003c3852/$FILE/44Brocke.pdf)
- DEUTSCHE FORSCHUNGSGEMEINSCHAFT (2009): Scientific Library Services and Information Systems (LIS): DFG Practical Guidelines on Digitisation for programmes funding Scientific Library Services and Information Systems http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung_en.pdf
- DRAKE K-M., B. JUSTRELL & A.M. TAMMARO (2003): Good Practice Handbook. Version 1.2. edited by the Minerva Working Group 6: Identification of good practices and competence centres. Minerva Knowledge Base Digitising Content Together. http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbook1_2.pdf
- EUROPEANA TEAM (2010): Europeana Aggregators' Handbook Edition 1. Europeana think culture. http://version1.europeana.eu/c/document_library/get_file?uuid=94bcddbfc3625-4e6d-8135-c7375d6bbc62&groupId=10602
- HANKEN J. (2010): Retooling Special collections Digitization in the age of Mass Scanning (LG-50-08-0058-08. URL: <http://biodivlib.wikispaces.com/file/detail/Retooling+Special+Collections+Digitization+in+the+Age+of+Mass+Scanning.odt>
- INTERNET ARCHIVE OPERATIONS TEAM (2010): Internet Archive Book Digitization Process. Ed. Robert Miller, Director of Books, Internet Archive.
- MCKENNA G., COLLECTION TRUST (UK), C. DELOOF & ROYAL MUSEUMS OF ART AND HISTORY (BE) (2009): Recommendation and best practice report regarding the application of standards, including recommendations for a harvesting format and fact sheets for dissemination. D3.2. Athena Access to cultural heritage networks across

Europe. <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>

MCKENNA G., COLLECTION TRUST (UK), C. DELOOF & ROYAL MUSEUMS OF ART AND HISTORY (Belgium) (2009): Digitisation: Standards Landscape for European museums, archives, libraries. Athena Access to cultural heritage networks across Europe. La Tipografia di Umbero Frisardi, Roma, Italy. <http://www.emuseum.cz/download/athena-soubory/27-imp-athenalibrettinoStandardCopertinaLastINCIANO.pdf>

SCAIFE B. (2009): Catalogue of content holder requirements (quality, quantity, accessibility, standards, specifications of content and metadata), (ECP-2008-DILI-518001) URL: https://bhl.wikispaces.com/file/view/BHL-E_2pt1_20090805.pdf

The Catalogue of content holder requirements is largely based on the Digital Library Federation's (DLF) "Benchmark for Faithful Digital Reproductions of Monographs and Serials": <http://www.diglib.org/standards/bmarkfin.htm> (2002)

THE DIGITAL LIBRARY FEDERATION BENCHMARK WORKING GROUP (2002): Benchmark for Faithful Digital Reproductions of Monographs and Serials. The Digital Library Federation. <http://www.diglib.org/standards/bmarkfin.pdf>

Digital References (Summary of Internet Links):

<https://bhl.wikispaces.com/BHL-Europe>

http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung_en.pdf

http://www.minervaeurope.org/structure/workinggroups/goodpract/document/bestpracticehandbook1_2.pdf

<http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>

<http://www.emuseum.cz/download/athena-soubory/27-imp-athenalibrettinoStandardCopertinaLastINCIANO.pdf>

http://version1.europeana.eu/c/document_library/get_file?uuid=94bcddbdf-3625-4e6d-8135-c7375d6bbc62&groupId=10602

<http://www.archive.org/details/ProcessDocument&reCache=1>

<http://www.diglib.org/standards/bmarkfin.pdf>

<http://ecom.fov.uni->

[mb.si/proceedings.nsf/0/4c1351e08e8d0f87c1257482003c3852/\\$FILE/44Brocke.pdf](http://mb.si/proceedings.nsf/0/4c1351e08e8d0f87c1257482003c3852/$FILE/44Brocke.pdf)

<http://www.bhl-europe.eu/en/outcomes/documents/ipr-working-documents>

<http://www.bhl-europe.eu/en/outcomes/documents/ipr-working-documents>

<http://tyler.hrc.utexas.edu/>

<http://biodivlib.wikispaces.com/Licensing+and+Copyright>

<http://www.bhl-europe.eu/en/outcomes/documents/ipr-working-documents>

<http://grib.gbv.de/>

<http://bhleurope.gbv.de/>

<https://bhl.wikispaces.com/Global+References+Index+to+Biodiversity+%28GRIB%29>

<http://www.e-taxonomy.eu/>

<http://www.gbv.de/vgm/index?lang=en>

<http://www.gbv.de/vgm/info/biblio/01VZG/>

<http://bhl.nhm-wien.ac.at/scanlist/>

<https://bhl.wikispaces.com/Documentation+from+BHLSeriallist>

<http://dublincore.org/>

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

<http://www.loc.gov/ead/>
<http://www.loc.gov/marc/>
<http://www.ukoln.ac.uk/metadata/cld/>
<http://www.tei-c.org/>
<http://rs.tdwg.org/dwc/index.htm>
<http://rs.tdwg.org/abcd/>
<http://marinemetadata.org/references/iso19115/>
<http://www.nbio.gov/>
<http://knb.ecoinformatics.org/software/eml/>
<http://dublincore.org>
<http://dublincore.org/documents/dcmi-terms/>
<http://www.bhl-europe.eu/en/outcomes/documents/ipr-working-documents>
http://version1.europeana.eu/c/document_library/get_file?uuid=a830cb84-9e71-41d6-9ca3-cc36415d16f8&groupId=10602
<http://www.version1.europeana.eu/web/guest/technical-requirements/>
<http://purl.org/dc/elements/1.1/rights>
<http://version1.europeana.eu/web/guest/technical-requirements/>
<http://www.version1.europeana.eu/web/guest/technical-requirements/>
www.zobodat.at
<http://biodivlib.wikispaces.com/file/detail/Retooling+Special+Collections+Digitization+in+the+Age+of+Mass+Scanning.odt>
https://bhl.wikispaces.com/file/view/BHL-E_2pt1_20090805.pdf
https://bhl.wikispaces.com/file/view/BHL-E_2pt1_20090805.pdf
<http://www.library.cornell.edu/preservation/tutorial/metadata/metadata-01.html>
<http://www.ahds.ac.uk/creating/information-papers/checklist/index.htm>
http://en.wikipedia.org/wiki/Color_depth
https://bhl.wikispaces.com/BHLE_WP2
https://bhl.wikispaces.com/file/view/BHL-E_2pt1_20090805.pdf
https://bhl.wikispaces.com/file/view/BHL-E_2pt1_20090805.pdf
<http://www.diglib.org/standards/bmarkfin.htm>
<https://bhl.wikispaces.com/QA+Procedures>
http://en.wikipedia.org/wiki/List_of_optical_character_recognition_software
<http://www.landesmuseum.at/biologiezentrum/>
<http://www.diglib.org/standards/bmarkfin.htm>
https://bhl.wikispaces.com/BHLE_WP2_BPG
<https://bhl.wikispaces.com/BHL+Cookbook>
https://bhl.wikispaces.com/BHLE_WP2
www.openarchives.org
<http://www.biologiezentrum.at/oai/oai2.php>
<http://gso.gbv.de>
<http://www.animalbase.uni-goettingen.de/>
<http://www.fishbase.org/>
<http://www.exlibrisgroup.com/category/Aleph>
http://si-pddr.si.edu/dspace/bitstream/10088/7182/1/Gwinn_Rinaldo_IFLA_2009.pdf
<https://bhl.wikispaces.com/Workflow+documents>
<https://bhl.wikispaces.com/BHL+Cookbook>
<https://bhl.wikispaces.com/QA+Policy>
<https://bhl.wikispaces.com/Cookbook+Quality>

<https://bhl.wikispaces.com/QA+Procedures>
<https://bhl.wikispaces.com/QA+Policy>
<https://bhl.wikispaces.com/QA+sampling+chart>
<http://www.carpet-project.net/en/catalogue/carpet/production-level-goobi-1/>
http://www.arcsoft.com/de-de/software_title.asp?productCode=SNSD
<http://www.gimp.org/docs/>
<http://www.gimp.org/downloads/>
<http://www.gimp.org/about/introduction.html>
<http://www.irfanview.com/eula.htm>
<http://www.gimp.org/about/introduction.html>
<http://www.gnu.de/documents/gpl-3.0.en.html>
<http://www.irfanview.net/>
http://www.irfanview.net/main_formats.htm
<http://www.software.com/irfanview-9337-1>
<http://www.heise.de/software/download/gocr/7147>
<http://jocr.sourceforge.net/download.html>
<http://www.dspace.org/latest-release>
<http://www.dspace.org/training-summary/training-materials.html>
<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
<http://openrepository.com/products/open-repository-dspace>
<http://atmire.com/consultancy.php>
http://www.carpet-project.net/fileadmin/user_upload/documents/2010-11-26-Repositoryienhosting
<https://phaidra.univie.ac.at/>
<https://fedora.phaidra.univie.ac.at/fedora/get/o:29162/bdef:Content/get>
<http://sourceforge.net/projects/metsbuilder/>
<http://sourceforge.net/projects/jhove/>
<http://hul.harvard.edu/jhove/documentation.html>
<http://www.loc.gov/standards/mets/version18/mets.xsd>
<http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd>
<http://www.loc.gov/standards/mods/mods.xsd>
<http://en.wikipedia.org/wiki/JHOVE>
<http://www.carpet-project.net/en/catalogue/carpet/JHOVE/>
http://www.java.com/en/download/inc/windows_upgrade_xpi.jsp
<http://www.carpet-project.net/en/catalogue/carpet/Fedora%20Commons>
<http://www.w3schools.com/soap/default.asp>
<https://wiki.duraspace.org/display/FCR30/Getting+Started+with+Fedora>
<http://www.fedora-commons.org/about/features>
<http://download.oracle.com/javase/index.html>
<http://tomcat.apache.org/>
<http://maven.apache.org/>
<http://www.digiverso.com/en/products/goobi>
<http://goobi-wiki.slub-dresden.de/index.php/Hauptseite>
<http://www.digiverso.com>
<http://www.carpet-project.net/en/knowledge-base/>
<http://www.carpet-project.net/forum/>
<http://www.digiverso.com/en/products/viewer>
<http://www.carpet-project.net/en/>

<http://www.minervaeurope.org/>
<http://www.digitalisierung.ethz.ch/>
<http://www.digiverso.com/en/>
<https://phaidra.univie.ac.at/>
<http://fedora-commons.org/>
<http://www.arcsoft.com/en-us/>
<http://www.irfanview.com/>
<http://www.gimp.org/>
<http://rs.tdwg.org/dwc/index.htm>
http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf
<http://biodivlib.wikispaces.com/Licensing+and+Copyright>
http://en.wikipedia.org/wiki/Bit_map
<http://www.nbio.gov/>
http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf
<http://biodivlib.wikispaces.com/Licensing+and+Copyright>
http://en.wikipedia.org/wiki/Bit_map
<http://www.nbio.gov/>
<http://rs.tdwg.org/dwc/index.htm>
<http://www.loc.gov/ead/>
<http://www.eol.org/>
<http://gallica.bnf.fr/>
<http://www.gbv.de>
<http://gso.gbv.de>
<http://www.archive.org>
http://en.wikipedia.org/wiki/Image_resolution
<http://www.ubka.uni-karlsruhe.de/kvk.html>
<http://www.oclc.org>
http://en.wikipedia.org/wiki/Pixels_per_inch
http://en.wikipedia.org/wiki/Dots_per_inch
<http://www.ukoln.ac.uk/metadata/cld/>
<http://www.w3schools.com/soap/default.asp>
<http://www.mda.org.uk/spectrum.htm>
<http://www.tei-c.org/>
<http://www.w3.org/XML/>
<http://www.loc.gov/standards/mods/>
<http://www.ubio.org/>
<http://www.biodiversitylibrary.org/>
<http://makeit.digitalnz.org/guidelines>
<http://www.library.cornell.edu/preservation/tutorial/>
<http://www.cambridgeincolour.com/tutorials/bit-depth.htm>
<http://wiki.bibalex.org/DAFWiki>

Appendix

A: Example of a model bilateral contract

City and date

Name of institution

Entitled Person

Address

Website provider

Authorised proxy

Address

Agreement

Herewith **### (Name of institution)** approves of the digital versions of the following periodicals “**Name of periodical**”, “**Name of periodical**”, and “**Name of periodical**” being made available for download on the website of the **provider (www.websiteprovider.xx)** .

The following volumes/issues will be provided for free download:

- 1) #####
- 2) #####
- 3) #####

The remaining volumes/articles will be for sale for **##** Cent per page on **www.websiteprovider.xx**. After the financial service provider’s deduction of expenses the entitled person will receive **##** Cent per downloaded page. If a client downloads 500 pages or more they will be given a discount (as shown in the list below). Such a discount respectively reduces the gains. The financial service provider draws up an annual account including the transfer of the earned money and list of the itemised downloads.

The website provider is obligated to subsequently transfer the money to the account **###** of the entitled person.

It shall be hereby stated that either party may rescind this contract without giving any reasons. At the same time it shall be emphasised that free supply of data can be changed to a fee-based service and vice versa as proposed by the entitled person.

For reasons that are inherent in the system the following discounts are defined for all downloads which the entitled person hereby approves:

Pages	Discount
>=500	10%
>=1000	20%

>=3000	30%
>=5000	40%
>=10000	50%

As of now the entitled person provides the web provider with a PDF-document (printable version) of every volume if available to include in the download platform. Furthermore a written note has to be included about whether the download is subject to a charge and if so for which period the service is fee-based. The conditions as described above apply to future volumes as well. The web provider includes a watermark in the header of every PDF-page reading as follows: © **entitled person, downloaded at www.websiteprovider.##**

Signature of the entitled person

Signature of the authorised website provider

B: Pre-Ingest File Submission Guidelines