ECP-2008-DILI-518001

BHL-Europe

Minutes on the workgroupsession on the 1st of December at the Meeting in London , NHM, on the first version of the approved best practice guidelines and standards

Deliverable number	*Minutes 1st of December, NHM*
Dissemination level	######
Delivery date	*6 December 2010*
Status	*Ready to read!*
Authors	######

e*Content*plus

This project is funded under the *e*Content*plus* programme[1],
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
Melita Birthälmer	MfN
Michaela Hierschläger	LANDOE
Tom Gilissen	NAT
Jane Smith	NHM
Constance Rinaldo (conference call)	MCZ
Laszlo Peregovits	HNHM
Antonio Valdecasas	CSIC
Marian Ramos	CSIC
Nancy Chillingworth	NHM
Bernd Sproger	AIT
Gerda Koch	AIT
Henning Knudsen	UCPH
Lesa Ng	RBGE

0.2 Revision History

Revision Date	Author	Version	Change Reference & Summary

0.3 Reviewers

This document requires the following reviews and approvals.

Name	Position	Date	Version

0.4 Distribution

This document has been distributed to:

Group	Date of issue	Version

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 (HIERSCHLÄGER LANDOE) 6

3 BACKGROUND (BIRTHÄLMER MFN) 6

4 PRE-SCANNING SETUP (SMITH NHM JUST MERGES TOGETHER – INPUT MUST COME FROM YOU!!) 6

4.1 MAPPING OF ALREADY EXISTING METADATA TO GRIB (MALICKY LANDOE) 6
4.2 COLLECTION ANALYSIS (VALDECASAS, RAMOS CSIC) 6
4.2.1 IPR issues and how to tackle them (Smith, NHM) 7
4.3 CHECK GRIB/SCAN LIST (GILISSEN NAT) 7
4.4 CHOOSING BOOK/JOURNAL (GILISSEN, NAT) 7
4.5 CALCULATE THE TOTAL COSTS OF THE WHOLE PROCESS (RINALDO MCZ, MALICKY LANDOE, GILISSEN NAT) 7
4.6 METADATA (NHMW, AIT, MALICKY LANDOE AND PEREGOVITS HNHM) 7
4.6.1 Recording Metadata in database for the own library usage (NHMW, AIT, Malicky LANDOE and Peregovits HNHM) 8
4.6.2 Metadata Requirements – Metadata Bucket (NHMW, AIT, Malicky LANDOE and Peregovits HNHM) 8
4.6.2.1 Granularity (NHMW, AIT, Malicky LANDOE and Peregovits HNHM) 8
4.6.2.2 Intellectual Property Rights/Metadata (Smith NHM) 8
4.6.2.3 Volume level (NHMW, AIT, Malicky LANDOE and Peregovits HNHM) 8
4.6.2.4 Article level (NHMW, AIT, Malicky LANDOE and Peregovits HNHM) 8
4.6.2.5 Page level (NHMW, AIT, Malicky LANDOE and Peregovits HNHM) 9
4.6.2.6 Creator (NHMW, AIT, Malicky LANDOE and Peregovits HNHM) 9
4.7 RECORDING AUTHORS, BIOGRAPHICAL INFORMATION (GUSENLEITNER LANDOE) 9
4.8 PUT METADATA ON HOMEPAGE (GUSENLEITNER LANDOE) 9
4.9 FUNDING OF THE SCANNING (MALICKY LANDOE) 9

5 DIGITISATION WORKFLOW (GILISSEN NAT) 9

5.1 PRELIMINARY ACTIONS (GILISSEN NAT) 9
5.1.1 Preparing the documents (Gilissen NAT) 9
5.1.2 Preparing the scanning order (Gilissen NAT) 10
5.1.3 Checklist (Gilissen NAT) 10
5.2 IMAGING REQUIREMENTS (GILISSEN NAT) 10
5.2.1 Basic principles (Gilissen NAT) 10
5.2.2 Resolution recommended, Bit depth (Gilissen NAT) 10
5.2.3 Format recommended (Gilissen NAT) 10
5.2.4 Post-processing (Gilissen NAT) 10
5.3 SCANNING (GILISSEN NAT) 10
5.4 OCR (GILISSEN NAT) 10
5.5 ARTICLE LEVEL ACCESS (GUSENLEITNER LANDOE, PEREGOVITS HNHM) 10
5.6 QUALITY CONTROL (MALICKY LANDOE) 10

6 PREPARATION OF DATA FOR BHL-EUROPE (MALICKY LANDOE) 10

6.1 SET UP TECHNICAL INTERFACE/PROVIDER SOFTWARE INTERFACE (MALICKY LANDOE) 11
6.2 DATA HARMONIZATION (MALICKY LANDOE) 11
6.2.1 File submission Guidelines (Malicky LANDOE) 11
6.2.1.1 Files to be supplied (Malicky LANDOE) 11
6.2.1.2 Delivery method (Malicky LANDOE) 11
6.2.1.3 Example (Malicky LANDOE) 11
6.3 PREINGEST (MALICKY LANDOE) 11
6.4 INGEST (MALICKY LANDOE) 11
6.5 POST OR RE-INGEST (MALICKY LANDOE) 11

7 COLLECTION MANAGEMENT & CURATION (SMITH NHM) 11

8 GENERAL INFORMATION 12

8.1 KEEP THE INFORMATION UP TO DATE (SMITH NHM, RINALDO MCZ) 12
8.2 WORKFLOW EXAMPLES (GILISSEN NAT; WELTER-SCHULTES UGOE; SMITH NHM;…) 12
8.3 WORKFLOW AND TOOLS (ZIELKE UBER) 12
8.4 GENERAL ARCHIVATION – PERMANENT STORAGE (SOMEONE FROM WP3) 12

Appendix (Hierschläger LANDOE) 14

Abbreviations (Hierschläger LANDOE) 14

Workgroupe session 1st of December, NHM London
possible the ideal !!!13:30-17:00

Try to get Connie and Michael via conference call

We state in each chapter the minimum if possible the ideal standard!

2 Purpose (Hierschläger LANDOE)

3 Background (Birthälmer MfN)

4 Pre-scanning setup (Smith NHM just merges together – Input must come from YOU!!)

4.1 Mapping of already existing Metadata to GRIB (Malicky LANDOE)

4.2 Collection analysis (Valdecasas, Ramos CSIC)

Minutes: Information from Antonio and Marian on what they intend to do: Power Point presentation
3 points of view : technical,..
How people shall analyse choose what to digitise; how CSIC is doing it
Laszlo: should not leave out other topics (let's say grey literature)
Set priorities on which themes should be digitised
Section with topics – first digitise these – are the most important themes and then you can go on to these themes
We need a general part for the collection analysis
Demands
Criteria how to choose analysis the right book:

Digitise whole volumes
Responsible for bringing the content of the chapter in the right order: **CISC, Henning Knudsen UCPH, Connie Rinaldo (MCZ) contribution**
General recommendations : with criteria:

Technical Connie Rinaldo (MCZ)
Use cases: CSIC, UCPH (Henning Knudesen), Connie Rinaldo (MCZ)

Global level, national level, institutional level

4.2.1 IPR issues and how to tackle them (Smith,Nancy, NHM)

Minutes:
There is already a BPG on IPR from BHL

4.3 Check GRIB/Scan List (Gilissen NAT)

4.4 Choosing Book/Journal (Gilissen, NAT)

4.5 Calculate the total costs of the whole process (Rinaldo MCZ, Malicky LANDOE, Gilissen NAT)

4.6 Metadata (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

Minutes:
Definitions for the terms, primarly contact Gerda!!!!!!
Clear reasoning why we need metadata
Is there a difference in the Metadata for born digitals?
Laszlo: There are lovely tools to manage pdf metadata which spare a lot of time.
Metadata
Definition – data about data. What does it mean in the real world. Have most of you met before. Yes, but most user are even do not know about it (e.g., digital photos, wordprocessor files)
Why metadata, do we really need them.
What kind of metadata are there (Technical, administrative, IPR, content)?
Where metadata are stored (external – in various files, internal – built into the files themselves)
Metadata standards (content: DC, qDC, IPTC, bibliographic; technical (XML, XMP))
Metadata export and import
Metadata flow in multiple directions
Is there a difference in metadata creation for the born digital materials?
External tools for metadata management; pdf tools slicer, bookmarking tools

Granularity (BHL context)

minimal level – bibliographic description + IPR
optimal – minimal + keywords, abstract
ideal – optimal + taxon and geographic metadata

Volume level
Book and journal volume/issue level (title pages + article pages + other pages)
Article level
Article + related pages (e.g., separate illustrations)
Page Level

4.6.1 Recording Metadata in database for the own library usage (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

4.6.2 Metadata Requirements – Metadata Bucket (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

4.6.2.1 Granularity (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

Minutes:
Adress??

4.6.2.2 Intellectual Property Rights/Metadata (Smith NHM)

Minutes:
If Jane has questions, she is going to ask Tom Gilissen or Melita Birthälmer
Management should take care of these things.

4.6.2.3 Volume level (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

Minutes:
Title, item page – BHL is using these terms
Give a definition about the used terms
Conference call-with Laszlo Michael und AIT Bernd Gerda about which terms should be used. –Title item page, volume article page ?

4.6.2.4 Article level (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

4.6.2.5 Page level (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

4.6.2.6 Creator (NHMW, AIT, Malicky LANDOE and Peregovits HNHM)

4.7 Recording authors, biographical information (Gusenleitner LANDOE)

4.8 Put Metadata on homepage (Gusenleitner LANDOE)

Minutes:
write example on it
Laszlo Peregovits writes example on how you provide Metadata on homepage
Tom Gilissen provides example for this chapter

4.9 Funding of the scanning (Malicky LANDOE)

Minutes:
Suggestion from Laszlo how he gets funding from scanning!

5 Digitisation workflow (Gilissen NAT)

Minutes:
Dennis Zielke (UBER should give Tom Gilissen the questionnaire he is doing.)

5.1 Preliminary actions (Gilissen NAT)

5.1.1 Preparing the documents (Gilissen NAT)

5.1.2 Preparing the scanning order (Gilissen NAT)

5.1.3 Checklist (Gilissen NAT)

5.2 Imaging requirements (Gilissen NAT)

5.2.1 Basic principles (Gilissen NAT)

5.2.2 Resolution recommended, Bit depth (Gilissen NAT)

5.2.3 Format recommended (Gilissen NAT)

5.2.4 Post-processing (Gilissen NAT)

5.3 Scanning (Gilissen NAT)

5.4 OCR (Gilissen NAT)

Minutes:
AIT, does metadata and OCR

5.5 Article level access (Gusenleitner LANDOE, Peregovits HNHM)

Minutes:
Laszlo gives a tool how to dismantle down to article level

5.6 Quality control (Malicky LANDOE)

Minutes:
Jane sends something again, has already sent something to me? Do not know?

6 Preparation of data for BHL-Europe (Malicky LANDOE)

6.1 Set up technical interface/provider software interface (Malicky LANDOE)

6.2 Data harmonization (Malicky LANDOE)

6.2.1 File submission Guidelines (Malicky LANDOE)

6.2.1.1 Files to be supplied (Malicky LANDOE)

6.2.1.2 Delivery method (Malicky LANDOE)

6.2.1.3 Example (Malicky LANDOE)

6.3 Preingest (Malicky LANDOE)

6.4 Ingest (Malicky LANDOE)

6.5 Post or Re-Ingest (Malicky LANDOE)

7 Collection management & Curation (Smith NHM)

Minutes:
all things are uploaded to the servers of NHM, due diligence, take down policy, the minimal requirement and the metadata shall be provided,

Update shall be bidirectional, have to provide institutional the content, to ministry or whatever, dataflow should be bidirectional
OAI harvester
Connie whether BHLE is open access or not
You get the BHL metadata fields
Enrich you own metadata so you get it biodirectional
Jane needs help from AIT Bernd Sproger, Gerda Koch, Nancy, BHL classic
Jane merges together, gets help from Bernd andConnie

8 General Information

Minutes:
just heading for the subchapters, nothing is written here

8.1 Keep the information up to date (Smith NHM, Rinaldo MCZ)

Minutes:
What happens when our funding ends,…

8.2 Workflow examples (Gilissen NAT; Welter-Schultes UGOE; Smith NHM;…)

8.3 Workflow and Tools (Zielke UBER)

Minutes:
If you use tools, send an E-mail to Michaela Hierschläger and she sends it to Dennis; for what are you using this, some information for the survey Dennis is doing
Goobi survey by Dennis Zielke
Goobi can be used for many digitisation processes
Pictures of internet archive Connie sends something

8.4 General archivation – permanent storage (ATOS)

Minutes:
What is important
ATOS – needs some input
What do you want to have in there?
Are there specific questions from you that you would like to have answered
Depends on your own situation
TopTips

Jane reads this guide through in regards to language proofreading

Appendix (Hierschläger LANDOE)

Abbreviations (Hierschläger LANDOE)

https://bhl.wikispaces.com/BHL+Europe+glossar
imp+athena_digitisation_standards-landscape.pdf

BPG: Best Practice Guide (shortcut for "approved best practice guidelines and standards")
DFG: Deutsche Forschungsgemeinschaft (German Research Foundation)
EZOOLO: Early Zoological Literature Online
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
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)
KVK: http://www.ubka.uni-karlsruhe.de/kvk.html (Karlsruhe Vitual Catalogue, offers a uniform user interface for simultaneous research in many union catalogues)
OPAC: Online Public Access Catalogue
RARA: rare works stacks of the SUB Göttingen
SBPK Berlin: Staatsbibliothek zu Berlin - Preussischer Kulturbesitz
SUB Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen
EOL: Encyclopedia 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/
OCLC: Not for profit computer service and research organization whose systems help libraries locate, acquire, catalog, and lend library materials. http://www.oclc.org
BHL Open Access: Images on the portal are available for viewing and reuse with no charges attached (this is just a suggestion!). 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