ECP-2008-DILI-518001

BHL-Europe

Content analysis & management status report 1 (metadata, page numbers, content providers)

Deliverable number	*D2.4*
Dissemination level	*Restricted*
Delivery date	*21 July 2010*
Status	*Final*
Authors	*Henning Scholz, Melita Birthälmer, Boris Jacob*

*e***Content***plus*

0   Document History

0.1   Contributors

Person	Partner
Henning Scholz	MfN
Melita Birthälmer	MfN
Boris Jacob	MfN
Gerda Koch	AIT
Wolfgang Koller	NHMW
Michael Malicky	LANDOE
Carolina Santamaria	CSIC
Matt Person	external

0.2   Revision History

Revision Date	Author	Version	Change Reference & Summary
15 Apr 2010	Henning Scholz	0.1	First draft
15 Apr 2010	Boris Jacob	0.2	Second draft
16 Apr 2010	Henning Scholz	0.3	Third draft for internal and external review
28 Apr 2010	Henning Scholz	1.0	Final version incorporating internal and external reviews
17 Jun 2010	Henning Scholz	1.1	Revision of final version after EC Annual Review
13 Jul 2010	Melita Birthälmer	1.2	Revision after receipt of consortium comments
16 Jul 2010	Boris Jacob	1.3	Revision after receipt of consortium comments
21 Jul 2010	Henning Scholz	2.0	Final version incorporating internal and external reviews

0.3   Reviewers

This document requires the following reviews and approvals.

Name	Position	Date	Version
Manuela Schulz	Academic Librarian	24 April 2010	0.3
Constance Rinaldo	Academic Librarian	17 June 2010	1.1

0.4 Distribution

This document has been distributed to:

Group	Date of issue	Version
Manuela Schulz	16 April 2010	0.3
BHL-Europe consortium	16 April 2010	0.3
BHL-Europe consortium	3 May 2010	1.0
Constance Rinaldo	17 June 2010	1.1
BHL-Europe consortium	17 June 2010	1.1
BHL-Europe consortium	23 July 2010	2.0

1 Table of Contents

2 Purpose

This document is to give an overview of the content specific progress of the project since it started in May 2009. It is an update of the information on the content of BHL-Europe given in the Description of Work (Annex 1 of the BHL-Europe Grant Agreement).

3 Background

BHL-Europe will support the digitisation and manage the acquisition and hosting of the digitised material contained in European institutions. The project will assist in the process by which each institution digitises its biodiversity material so it is done efficiently and effectively. Several processes and tools have been being established by BHL-Europe to analyse the content and support the management of the scanning initiatives of each partner. Processes are questionnaires, phone conferences, e-mail groups and face to face discussions at the five meetings with content providers so far: May 2009 Berlin, August 2009 Leiden, November 2009 Prague, March 2010 and May 2010 Vienna. Tools are system prototypes that can be tested by the partners and advanced with their input. Current prototypes are the BHL Scan list, the GRIB (both described in D2.2 and D2.3), and the Pre Ingest (described in D3.5).

A library questionnaire was sent out to the partners in the beginning of the project to better understand the content, workflow and requirements of every content provider. 13 Partners answered this questionnaire (see Appendix A, row 4) and the feedback received from this influenced a number of processes, in particular, the harmonisation of metadata for the BHL-Europe system and the preparation of D2.1. Based on D2.1, further discussions and work on the metadata schema was leading to the BHL-Europe Master Data Element List (see section 5.1 and Appendix C), taking into account the content specific requirements of the individual content providers.

The Memorandum of Understanding (MoU) was designed with an appendix to further specify the underlying content of BHL-Europe (see section 5.2 and Appendix B). This appendix is tought of as an amendment to section 4.1 of the Description of Work and is influencing the ingest planning to BHL-Europe and Europeana.

A second library questionnaire was sent out in April 2010 (LQ 2010) to collect further requirements from the content providers, and to better understand the local library systems and IT infrastructure. 11 partners answered LQ 2010 (see Appendix A, row 5).

Three system prototypes were developed over the last year. The AIT demonstrator (http://bhl.ait.co.at/) has been available since summer 2009 to collect metadata and manage the test mapping of content provider metadata. The GRIB prototype and the BHL Scan List were established in December 2009 to build a common management system of the content, library catalogues, and digitisation processes of our partners.

All of these are described in detail below.

4   Content management

The Global References Index to Biodiversity (GRIB), which is a database system with content management and deduplication functionalities, is our most important tool to analyse the content and support the management of the scanning initiatives of each partner. At the moment, the first prototype is ready for testing and usage by BHL-Europe consortium members only.

The GRIB infrastructure will harvest the library catalogues of institutions with libraries focused on taxonomic and biodiversity literature. These institutions are mainly natural history museums and botanical gardens, but also zoological, botanical and palaeontological libraries of universities, and national libraries. This common library catalogue of all these institutions will represent the vast majority of the literature relevant for the scientific community, but also for other target users of the system like citizen scientists, students, teachers, policy makers, and general interest audience.

Several tools and procedures will be used for building a priority list of books and journals to be included in BHL-Europe in digital form. We will analyse such major indexes as *Index Kewensis,* Sherbourne's *Index Animalium,* and Neave's *Nomenclator Zoologicus*. We are also working on mining other resources to determine those journals that have been most cited in the literature of species identification and description. Furthermore, we will use the experiences of our partners in building the priority list. Most of the BHL-Europe partner institutions are natural history museums or botanical gardens, thus a large number of experts in the biodiversity domain are available to select relevant content for BHL-Europe. Some of our partners already make use of this expertise in building local priority lists of journals and monographs (e.g. NBGB, RMCA, UH-Viikki). Furthermore, BHL colleagues in the USA are working on additional collection development strategies and criteria to prioritise items for digitisation. All priority lists will be included in the GRIB once this system is fully functional. We expect this to happen towards the end of 2010. In addition to the joint efforts of our consortium to build priority lists of content for BHL-Europe, the GRIB will have the functionality to allow users to nominate content for digitisation.

The work on the GRIB is still in progress. Therefore, we need an interim solution to manage the digitisation process and help with the analysis of available content. This interim solution is the BHL Scan List as a BHL-Europe adaptation of the Biodiversity Heritage Library Serials Union Catalogue developed at NHM London for the BHL project. The current functionality of the BHL Scan List includes:
  ▪ MARC exchange format records (minimum required fields stipulated)
  ▪ Records matched and merged through batch php scripting (see above)
  ▪ Merged set made available over the Web and fully searchable/sortable
  ▪ Authentication (1 per BHL and BHL-Europe partner institution)
  ▪ Ability to bid to scan part or all of a title's holdings at a per title level
  ▪ Ability to manually merge records which the scripting did not detect
  ▪ Merging the entries based on the ISSN using a stored procedure inside MySQL
  ▪ Ability to link to record and download to our local scanning management system

Currently, the BHL Scan List has 84,314 records. As numerous duplicates are still not merged, the number of serials is difficult to estimate. However, 30,742 items in 2,554 serial titles are currently available in digital format via the BHL Portal (12 volumes per serial average). 2,891 titles are currently in progress of being digitised (1,193 complete bids, 1,698

partial bids). Further analysis is necessary to estimate the number of serials that still need to be digitised based on the current corpus of serials in BHL and BHL-Europe partner libraries.

As we are still building the infrastructure for effective content analysis, we need to have an overview of the available corpus of biodiversity literature to be scanned by BHL-Europe national partners and BHL. This also helps us in building a priority list of literature for digitisation programmes. Therefore, we did some calculations and estimations of that corpus. We based our calculations on real numbers of available serials from Austria, good estimates of German language biodiversity literature and estimates from BHL. Based on the number of inhabitants figures we extrapolated the page number estimates of German language biodiversity literature for the EU. Based on these first estimates, we have a range from 88,000,000 pages up to 166,375,000 pages of natural history literature published within the EU until today. Towards the end of the current year we expect more detailed estimates after we harvest actual catalogue records into the GRIB.

4.1 Content providers

The content providers within the consortium will create a critical mass of high quality digital content representing the biodiversity domain. The content providers have been selected on the basis of their ability to contribute key biodiversity and taxonomic literature. Although an accurate definition of the key biodiversity and taxonomic literature is still in discussion, large natural history museums and botanical gardens are considered to have large collections of primary taxonomic literature comprising all kingdoms of life.

The content will not be restricted by proprietary third-party rights or any other constraints which would limit its use in an open access environment using a Creative Commons licenses[2]. The digital content must either be in the public domain, or else the content contributors must have permission from IP owners to provide access under Creative Commons.

The initial focus on public domain material is not a limitation because systematic biology depends more than any other natural science upon historic literature. Another reason to focus on historical literature is that many old and important monographs are themselves inherently very rare, fragile or in need of conservation. This makes "hands on" access very difficult. This project will substantially reduce the need for handling of these rare and valuable materials.

From the beginning of the project BHL-Europe has had 17 content providers involved as consortium members (see also Figure 1):

1) Natural History Museum (NHM), UK

2) Narodni muzeum (NMP), CZ

3) Georg-August-Universitat Gottingen Stiftung Offentlichen Rechts (UGOE), DE

4) Land Oberosterreich (Oberoesterreichische Landesmuseen) (LANDOE), AT

5) Hungarian Natural History Museum (HNHM), HU

6) University of Copenhagen (The Natural History Museum of Denmark) (UCPH), DK

---

[2] See http://creativecommons.org/

7) Stichting Nationaal Natuurhistorisch Museum Naturalis (NAT), NL

8) National Botanic Garden of Belgium (NBGB), BE

9) Royal Museum for Central Africa (RMCA), BE

10) Royal Belgian Institute of Natural Sciences (RBINS), BE

11) Bibliotheque nationale de France (BnF), FR

12) Museum national d'histoire naturelle (MNHN), FR

13) Consejo Superior de Investigaciones Cientificas (CSIC), ES

14) Royal Botanic Garden Edinburgh (RBGE), UK

15) Smithsonian Institution (SIL), US

16) Missouri Botanical Garden (MOBOT), US

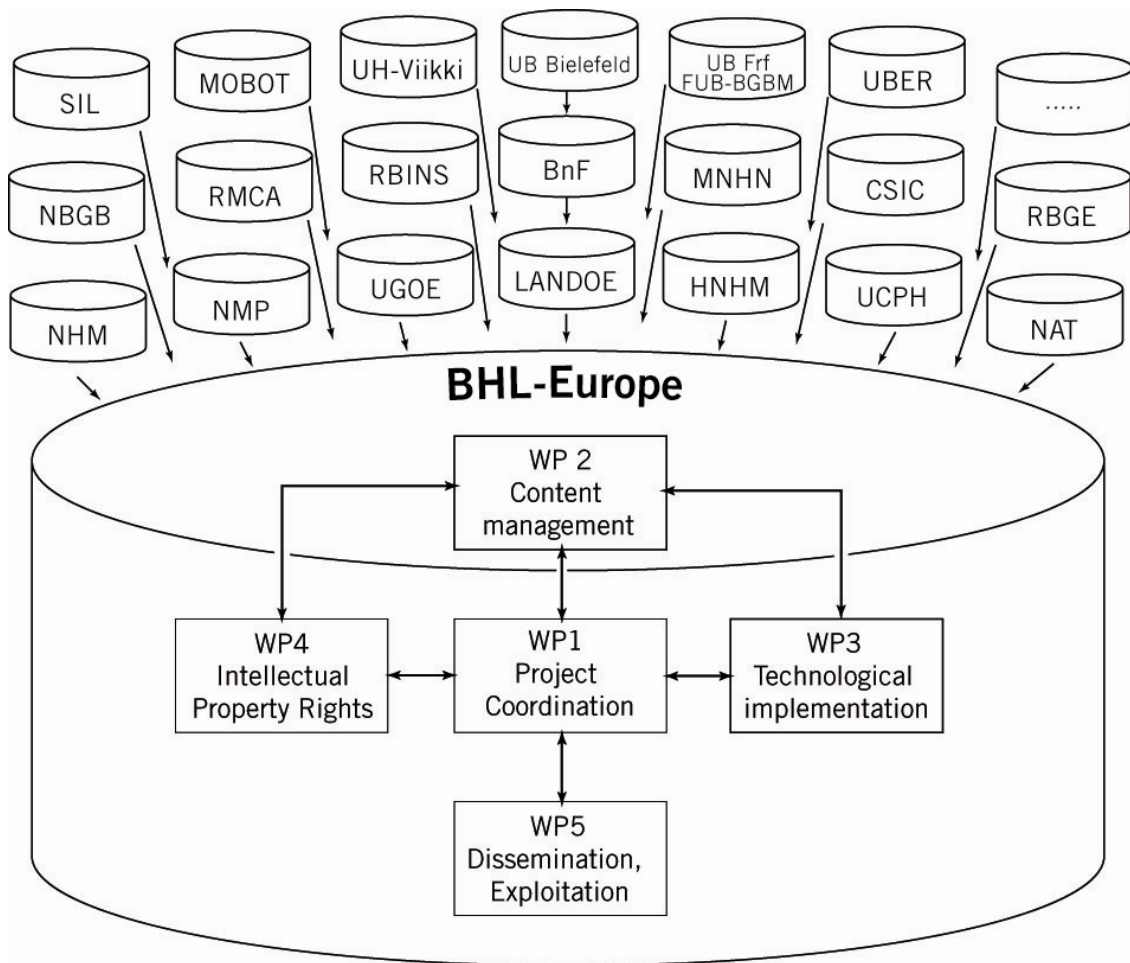17) Helsingin yliopisto, University of Helsinki, Viikki Science Library (UH-Viikki), FI

Figure 1: BHL-Europe internal organisation and the network of content providers today and in the future.

In addition to the above mentioned content providers, we expect new partners to provide further content during the project's lifetime. These new partners may be other important libraries in countries not yet engaged in the project, or learned societies. The following two institutions have signed the MoU and will provide content to BHL-Europe (see Appendix A, row 1):

1) Universitätsbibliothek Bielefeld, DE

2) Humboldt-Universität zu Berlin (UBER), DE

Some BHL-Europe consortium members are not content providers yet, but do have libraries and are thus being treated as potential content providers (see Appendix A, row 2). Further potential content providers may be found within the EU project EDIT (European Distributed Institute of Taxonomy). In addition to 15 institutions that are both BHL-Europe and EDIT consortium members (see Appendix A, colour: green and yellow), there are 13 more EDIT only partners that are mostly specialised natural history/biodiversity institutions.

25 BHL-Europe partners can provide library catalogue data for the GRIB (see Appendix A, row 3). The majority of these institutions have catalogues accessible online (see Appendix A, row 7). The analysis of the library questionnaires furthermore revealed that the partners that do have part of their content published in an online catalogue often have these data in separate databases. Some have different catalogues for monographs and serials, or library holdings and digital holdings. In general each catalogue is stored in one database. The analysis of the questionnaires revealed so far, that 17 partners have altogether 38 catalogues which are relevant to BHL-Europe (see Table 1 and Appendix A, row 6).

Partner Acronym	Number of Library Catalogues
CSIC	3
FUB-BGBM	1
HNHM	2
MfN	1
MIZPAS	1
MNHN	2
MSN	2
NAT	2
NBGB	1
NHM	3
NMP	1
RBGE	1
RBINS	2
RMCA	10
UBER	1
UGOE	2
UH-Vikki	3
Total sum	38

Table 1: Partner and their catalogues

Furthermore, the questionnaires indicated that partners may have a high complexity within their various catalogues. Here are some examples: CSIC has as three different catalogues (bibliographic catalog, archival catalog and authority catalog) in different databases which incorporate already scanned objects. MfN and FUB-BGBM have one catalogue each, stored in a shared DB together with the catalogue of the Biology department of the Freie Universität Berlin (see Figure 2). In RMCA the different branch libraries have altogether nine different catalogues/databases and some of the holdings are already catalogued in the shared LIBIS catalogue.
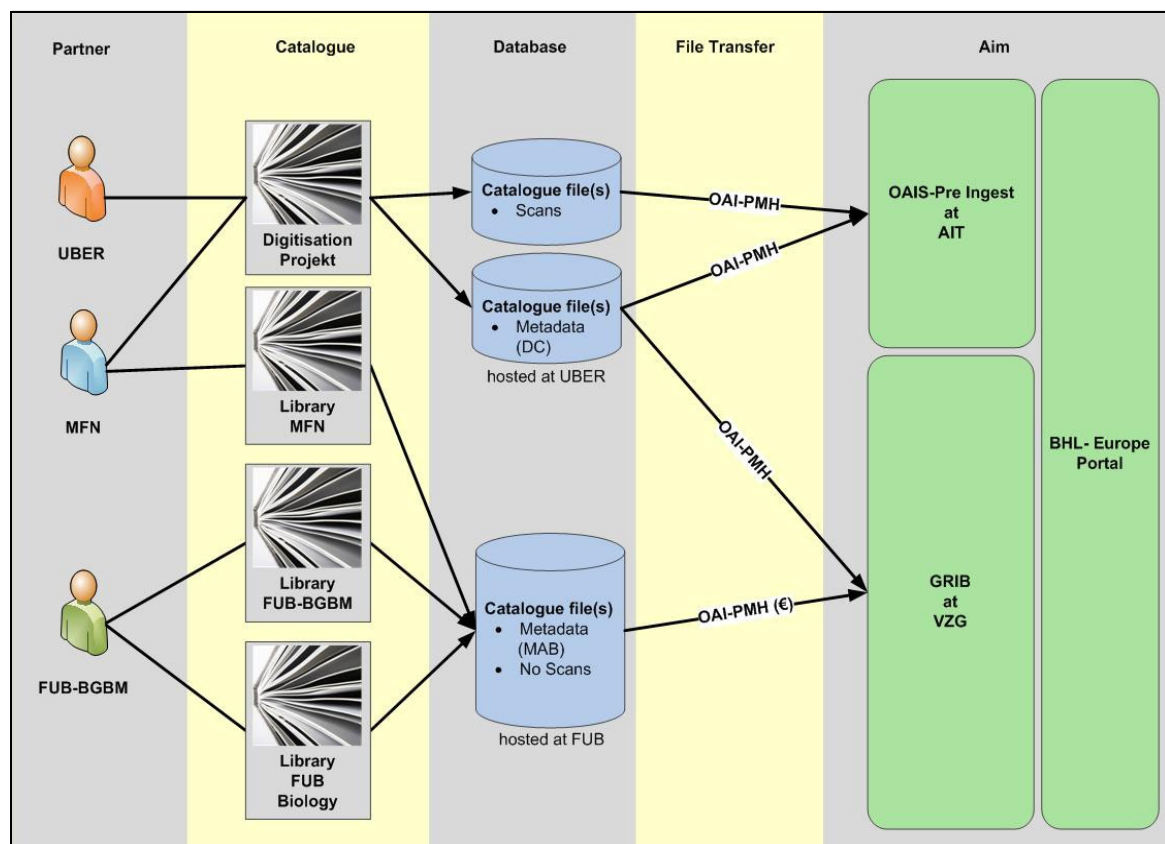
Figure 2: Data management complexity

There are different possibilities of exporting library catalogue metadata, Figure 3 shows the amount of metadata exported by partner libraries using 3 possible export methods." (see Appendix A, rows 8 to 11).

14 partners provide (parts of) their catalogue records to other bibliographic utilities, like OCLC-WorldCat or the Common Library Network GBV. This opens the possibility of harvesting them from those sources directly. (see Appendix A, row 11). A few partners also deliver to Europeana and need to align with Europeana. Individual partners must be able to manage the data matching and updating. It is also important to coordinate the deduplication of this content in Europeana, as it would come from two different sources to Europeana. BnF, for example, is a content provider for Europeana and provides the biodiversity material again

to Europeana via BHL-Europe. Europeana has to find ways to identify duplicate records and filter the search results accordingly.
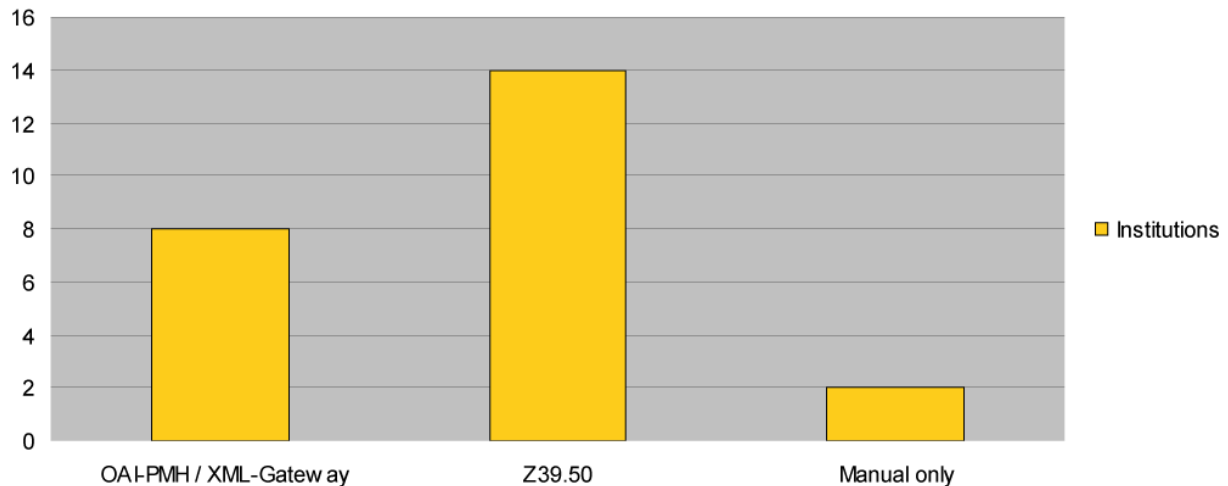

Figure 3: Library Catalogue export possibilities

4.2 Pre-Ingest: progress and status

Pre-Ingest represents a provider of the Open Archival Information System (OAIS). It will help the content providers (libraries, digitisation centres, etc.) to match the data before it will be ingested. As soon as the content providers are finished with scanning, they will upload their data to the Pre-Ingest component via FTP and use the Submission Information Package (SIP) creator module to check for errors within the uploaded Pre-SIP via a web browser. Pre-Ingest will push the SIP into Ingest once it is created. Within the SIP Creator, metadata mapping tools can be used to map the local library data to the BHL-Europe schema. The entire Pre-Ingest process will be managed by experienced BHL-Europe team members to assist content providers to understand the process and to successfully finish the harmonisation process.

A Pre-Ingest testing phase was scheduled during the content provider meeting in March 2010 (Vienna) in order to test the file submission guidelines (see D2.1) and other processes associated with the Pre-Ingest (see also D3.5), as well as the BHL-Europe metadata elements (see Appendix C) in a real life context. BHL-Europe content providers were asked to provide a few samples of scanned books or serials that are available in their local repositories together with the corresponding metadata. For this purpose AIT provides FTP access to the content provider for uploading test metadata and the corresponding test scans. Five content providers already have provided samples of scans and metadata for the Pre-Ingest test to date (Table 2). The communication with our partners will be intensified now in order to get the appropriate samples within the next weeks, with the objective of conducting the first test phase of the BHL-Europe Pre-ingest module this summer.

Partner Archive	Metadata received - amount/ description level	Metadata format	Scan amount received	Scan Format	FTP account
HNHM	6 samples	.xml, .txt, .sql			no
NAT	NFM , issues, articles, pages,	xls	9 folders	tif, pdf	yes
NBGB	books	.mrc	4 books	tif	yes
RMCA	32 books, csv	mrc	32 books	.tif	no
UH-Viikki	MARC21, Volume1 Memoranda Societatis pro Fauna et Flora Fennica	.mrc; .mrk	11 tiff 1 pdf (164 pages)	tiff	yes

Table 2: Submitted data for Pre-Ingest testing

5 Content analysis

An important prerequisite to managing the content effectively is to analyse both the metadata and the digital content. We need to understand the partner requirements, their local infrastructures, workflows and data formats in more detail in order to match the data in GRIB, in the BHL-Europe system and in Europeana. We also need to understand the quality and quantity of the content BHL-Europe is receiving over time to plan the ingest process accordingly. The work done so far is summarised below.

5.1 Metadata analysis

In order to provide a sustainable system within BHL-Europe, the metadata needs to be defined properly. Information which may be requested, but has no definition in our metadata schema, cannot be displayed or used. Therefore, the discussion of metadata standards and schemas has been a continuous task since the start of the project in May 2009. We included a number of BHL-Europe target users and experts as members of the BHL-Europe consortium in these discussions: taxonomists, librarians, ICT specialists, and others. In August 2009, BHL-Europe ratified a first document on metadata fields to be used by BHL-Europe (Deliverable D2.1: Catalogue of content holder requirements (quality, quantity, accessibility, standards and specifications of content and metadata)). Based on the D2.1 fields, several available standard schemes (like ONIX, METS, MARCXML, etc.) were evaluated in depth by all content providers in Email discussions and during meetings. All of those schemas are able to contain the basic information from D2.1.

A first content analysis review carried out in spring/summer 2009 revealed that MARC in all its flavors and versions is widely used by the BHL-Europe content providers (Figure 4 and Appendix A, row 12 to 15). Therefore, MARC can be considered as our baseline for mappings to the BHL-Europe schema. This is also supported by the fact that a number of MARC mappings are available, including MARC 21 to MODS, MODS to MARC 21, DC to MARC 21, MARC 21 to DC, ONIX to MARC 21, UNIMARC to MARC 21[3] Dublin Core (DC) is also widely used, which is also easy to handle for BHL-Europe. As the ESE schema (Europeana Semantic Elements) is qualified DC, we are working with DC mappings anyway to map BHL-Europe data to Europeana.

In a first test of the data harmonization in summer 2009, metadata of 14 BHL-Europe content providers were ingested into the AIT demonstrator (see Appendix A, row 18 and 19) and MODS was used as the common format for the descriptive metadata. The use of MODS was justified by the Description of Work that suggested MODS as the standard to be applied. This test was a technical proof-of-concept to match data from different sources and different formats on one single platform and search via this platform across the various databases and catalogues provided by the partner libraries. The experiences of this process were used for the follow up discussions on metadata schema and Pre-Ingest methodology.

---

[3] http://www.loc.gov/marc/marcdocz.html
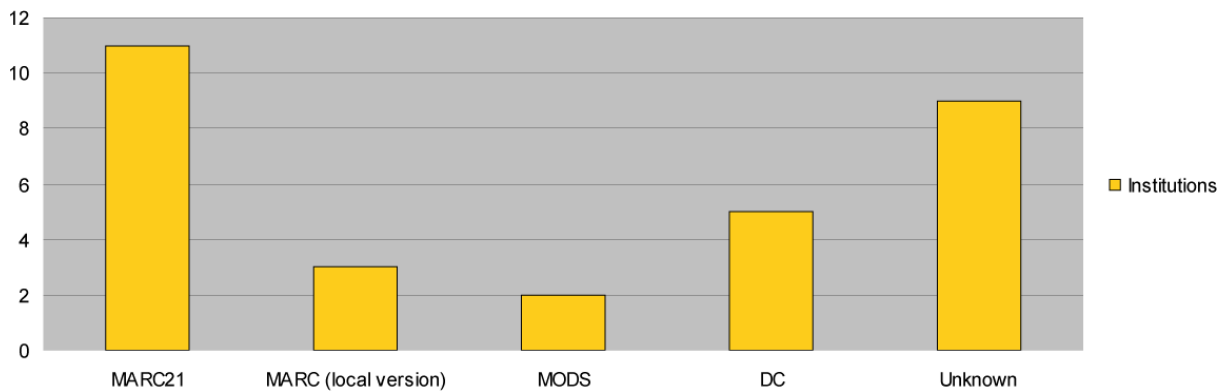
Figure 4: Metadata Formats used by BHL-Europe Content Providers

BHL-Europe aims to not only provide a sophisticated search facility based on the FRBR model (Functional Requirements for Bibliographic Records[4]) but also to allow users to search on article and chapter level. This requires data and descriptions on article and chapter level in the provided metadata to the BHL-Europe system and the information on how title, item and article/chapter are connected with one another. BHL-Europe will therefore have to provide automatic routines (wherever possible) and tools for the content providers to add these descriptions during the Pre-Ingest process of the data to the BHL-Europe system. This also requires a BHL-Europe metadata schema that provides data fields for these additional descriptions.

Deliverable D2.1 outlined the baseline of such a schema which in consequence led to the suggestion and decision to create a BHL-Europe METS profile which will be able to carry information on administrative, descriptive and structural metadata (Figure 5).

Figure 5: BHL-Europe METS profile

---

[4] http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records

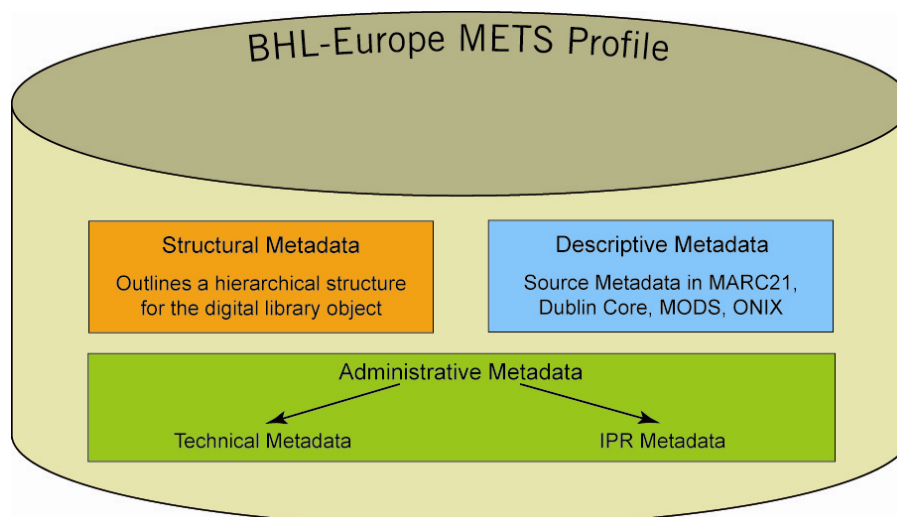METS (Metadata Encoding & Transmission Standard) is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, using the XML schema language of the World Wide Web Consortium. This standard is maintained by the United States Library of Congress which guarantees its long-term sustainability.

In its descriptive section, METS allows the introduction of various metadata standards. As the 2009 findings showed a dominance of MARC and Dublin Core metadata being used by BHL-Europe content providers (see Figure 4), and these standards are well maintained and widely used standards in the library domain as mentioned above, it was suggested that all pre-existing Metadata that comes into BHL-Europe be converted into MARC – and presented in MARCXML. Nevertheless the additional fields BHL-Europe will need for the advanced search facilities but which are not included in this standard have to be filled in during the BHL-Europe data submission process.

The list of BHL-Europe metadata fields as described in D2.1 was therefore further elaborated in the BHL-Europe Master Data Element List (Appendix C). The list already provides a mapping of the various data fields to Dublin Core (ESE), MARC and MODS. It shows metadata fields for the description levels: Title, Item, Page, Rights, Chapter/Article. This schema will be used for our descriptive metadata instead of MODS, as it better supports the requirements of the BHL-Europe content providers.

Both D2.1 and the BHL-Europe Master Data Element List are non-public working documents at the moment, which will further be refined based on the outcome of the Pre-Ingest tests to be carried out in summer 2010. The results of these tests will show possible shortcomings and affect the final design of the BHL-Europe metadata schema. In addition the refinement of the preliminary BHL-E use cases may reveal the need for new metadata fields (or elimination of existing fields) and provide information on what fields should be mandatory, recommended, or optional). The final list of metadata elements, however, will be available to the public in the second year of the project. The current list of metadata fields will be used for the BHL-Europe prototype.

5.2 Underlying content of BHL-Europe

Section 4.1 of the Description of Work gives an overview of the underlying content of BHL-Europe. The table specifies the type, quantity, definition, IPR, current use, metadata, and language of the content per partner. This table, however, does not provide a roadmap of content delivery and also does not specify the level of access to the individual content per partner. Based on initial discussions we concluded that individual content providers may have different policies in sharing their content with BHL-Europe. Therefore we need to agree what content each provider shares with BHL-Europe and what content has to remain in the repository of the provider. This information will aid in decisions about the systems architecture and the service we can provide for our partners. The Memorandum of Understanding as foreseen in the Description of Work, was extended by an Appendix to collect the necessary information and agree on the level of access to the BHL-Europe content. The wording of the MoU was agreed in several rounds of discussions in summer and fall 2009 (Email, wiki, meetings). The collection of data from all content providers also involved discussions and negotiations between the PMG, the TMB and the individual content provider to clarify the consequences of various decisions. This is a time-consuming process and it continues. To date, only one of the 17 BHL-Europe content providers have not signed the

MoU yet (CSIC). Very recently (June 2010), we received the signed MoU from UCPH, BnF and HNHM and their data are included in the following analysis.

Specific information on the digitisation projects of our content providers may be found in the BHL-Europe newsletter, where we provide well illustrated details on individual projects and activities. The quarterly newsletters can be downloaded from our project Web site under http://www.bhl-europe.eu/newsletter.php. Newsletter #1 (July 2009) has articles on digitisation in the Czech Republic and BHL digitisation at the Natural History Museum London. Newsletter #2 (December 2009) has articles on activities in the Viikki Science Library, the Royal Belgian Institute of Natural Sciences and the digitisation at BnF for Gallica and Europeana. Newsletter #3 (March 2009) has an article about the digitisation activities in Goettingen. Further contributions will follow with the next newsletters.

5.2.1 Quality of content

The data format provided by the content providers is shown in Figure 6. The majority of the data will be provided as TIFF and JPG format. Furthermore PDF files are provided in addition by most of the content providers.
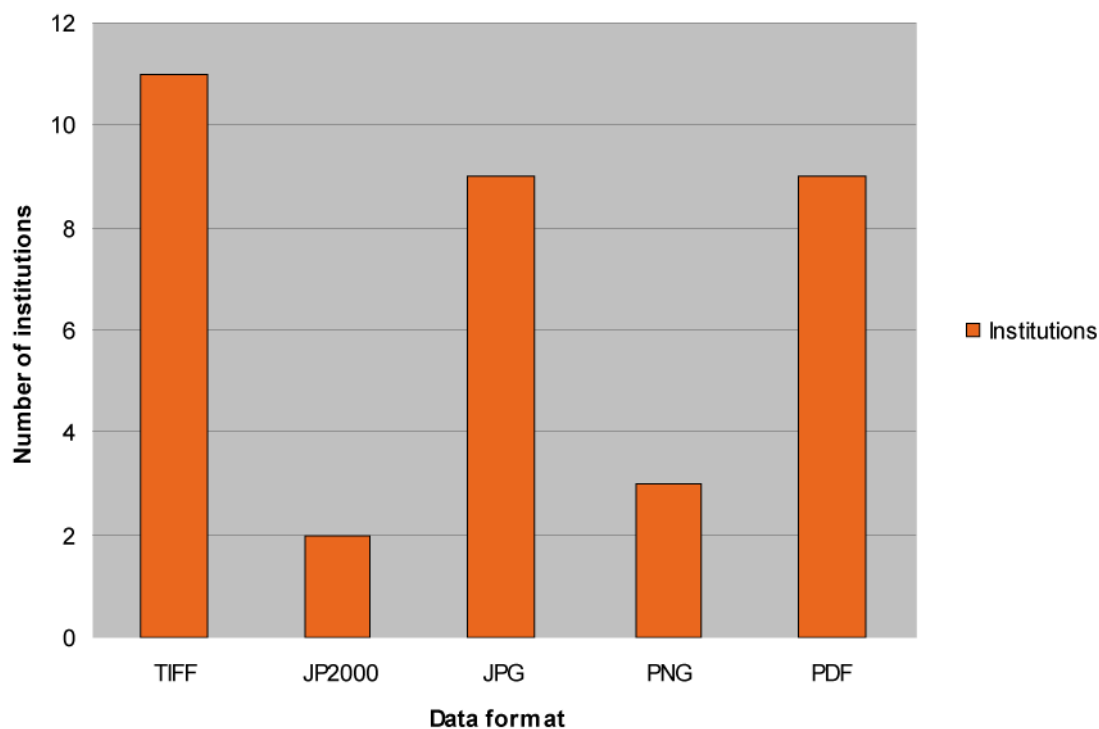
Figure 6: Format of data provided to BHL-Europe

According to the catalogue of content holder requirements (D2.1), the minimum resolution for submitted content is 300 dpi. The evaluation of the MoUs regarding the resolution of the provided content (see Figure 7) shows the acceptance of the imaging requirements recommended in D2.1. Except for NMP (Narodni muzeum), all content providers agreed to provide a minimum resolution of 300 dpi. Some of our partners supply a higher resolution than advised in D2.1.
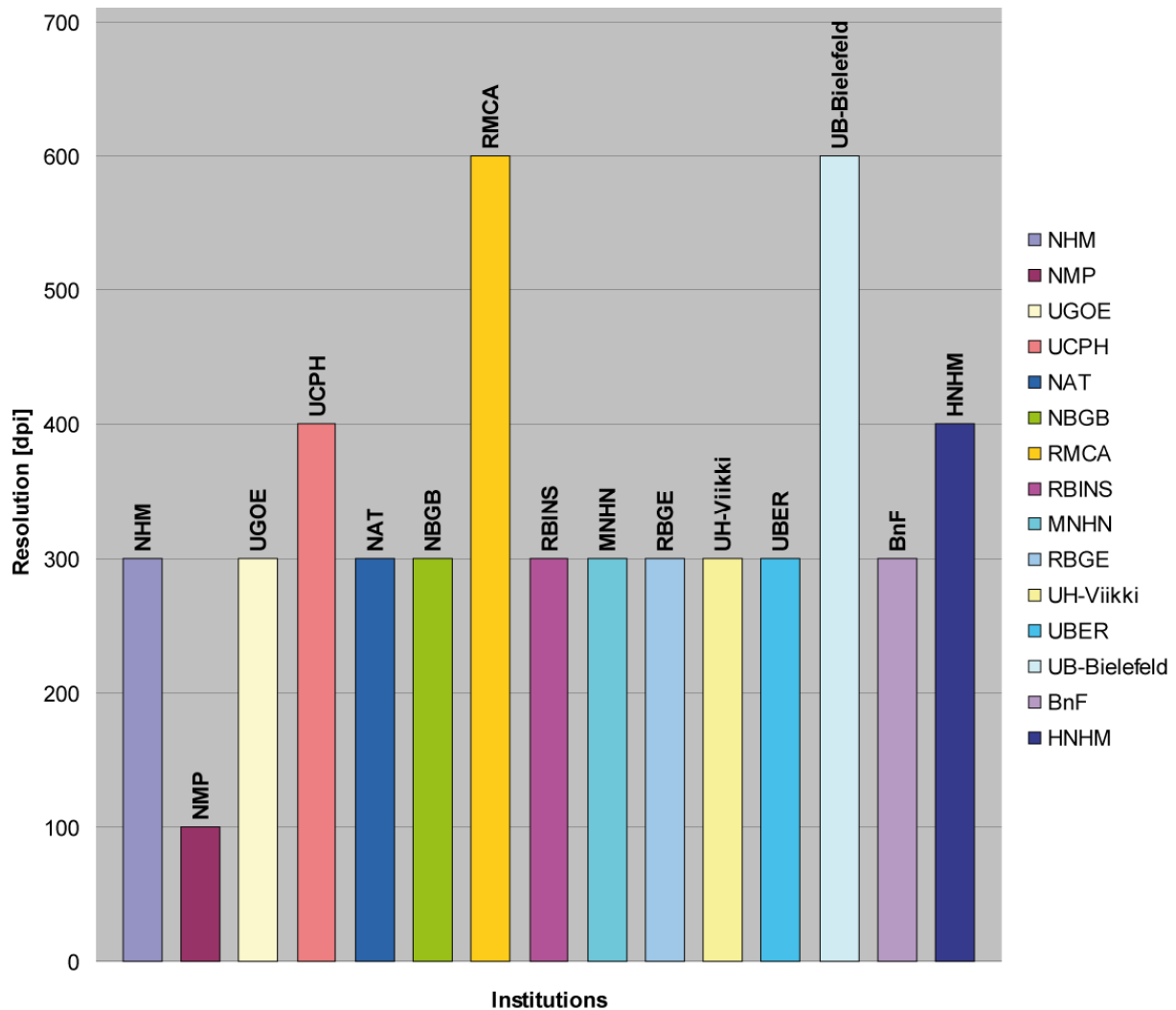
Figure 7: Resolution of provided content

5.2.2 Quantity of content

With the MoU, we also have a detailed overview on the quantity of content available from BHL-Europe over the next years. BHL-Europe counts its volume of content in two units: volume and pages. BHL-Europe provides page level access to all of the content harvested by us. Therefore, every content provider is required to provide individual page images and not only PDF or multipage TIFF files representing a volume. We will have a book viewer on our portal as this is the way BHL currently is exposing the content to the user (see http://www.biodiversitylibrary.org). Europeana, however, is counting individual books as objects to be searched and displayed via http://www.europeana.eu. One object is represented with a thumbnail on the portal, and it makes no sense displaying every single page present in our repository. That would make search and retrieval via Europeana very difficult and is therefore not a user friendly approach. Hence, we need to know the numbers of pages and volumes contributed by our partners.

As we get the volume numbers from our partners and use them for our ingestion planning, we will use volume numbers herein and in future reports. Page numbers are more abstract and library users are probably more used to counting books rather than pages. This makes the comparison of the quantity of content easier. However, as a rule of thumb, one volume has an average of 350 pages.

Currently, there are over 10,000 volumes from all European content providers to contribute to the BHL-Europe project. This number is continuously increasing as scanning projects are still in progress. Taking all the European partners of our network into account, we expect about 13,000 volumes to be available via BHL-Europe and Europeana in about 18 months time. This number does not take into account the contribution from BHL-Europe partners that have not signed the MoU to date, and from BHL. This number also does not take into account the contribution of additional content providers that will contribute to BHL-Europe in the near future. Figure 8 provides an overview of content coming to BHL-Europe over the project duration. In addition to the total corpus of BHL-Europe we mark the content from European institutions and US institutions separately. For detailed information about the number of volumes per partner see also Appendix B.

Figure 8: Quantity of provided content
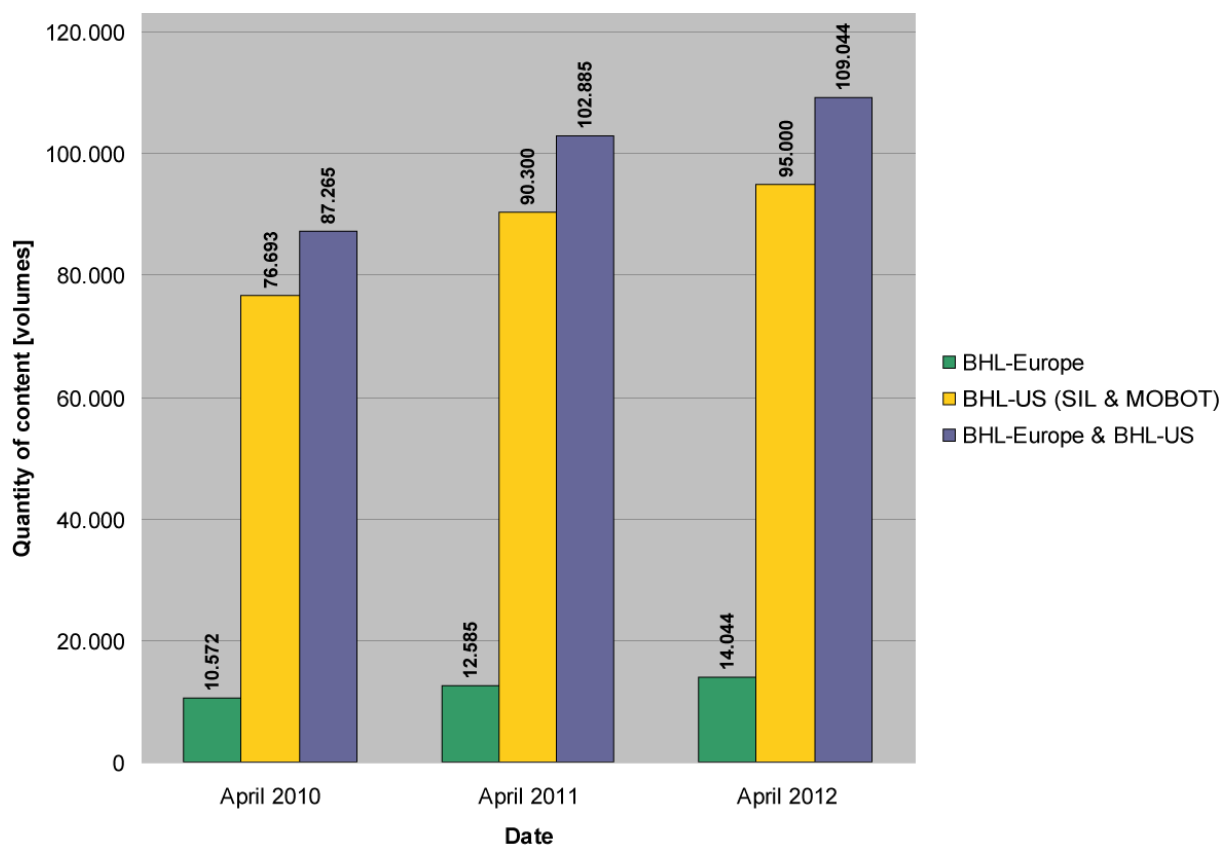
5.3 European cultural heritage and BHL

Today European cultural heritage is distributed all over the world. For example, since Jamestown, the first colonial settlement in the New World, was founded 400 years ago, immigrants from Germany and German speaking countries have had considerable influence on the economic and political development of the USA. The German-American Heritage

Foundation of the USA (http://www.gahfusa.org/) has as its goal to inform and educate the American public about the heritage of German immigrants and their valuable contributions to the development of the USA and to protect the cultural heritage of Germans in the USA.

Biodiversity heritage literature is also part of our European cultural heritage and is scattered worldwide. For example, the Marine Biological Laboratory (MBL), founded 1888, is the oldest private biological laboratory in the United States. Louis Agassiz, born in Switzerland, consulted with Anton Dohrn of Stazione Zoologica in Napoli in the late 1800's on the construction of a library of the biological sciences at the Marine Biological Laboratory. Since 1930 this library has also served the Woods Hole Oceanographic Institution and the entire Woods Hole Scientific Community which is now known as MBLWHOI Library (Marine Biological Laboratory Woods Hole Oceanographic Institution). Since 1920, 56 Nobel Prize winners have been associated with MBL during their careers. Of these awardees, 22 were European citizens who spent part or all of their careers working for European institutions. The histories of their careers often include responses to mid-20th century historical forces for which working at MBL was often part of a path leading to freedom from the upheaval caused by historical events. August Krogh, Otto Meyerhof, Salvador Luria and Albert Claude are only a few names of Nobel Prize awardees of European background that worked at MBL, and hence brought European cultural heritage outside Europe. Hence, European researchers have travelled to the MBLWHOI library for many years to read and study this literature of European origin and also enrich the libraries during their working time. It is known that some of these researchers worked in the MBLWHOI library because publications of European origin were present there, but no longer present in Europe.

By collaborating with BHL, we are able to repatriate this European heritage literature back to Europe via BHL-Europe and Europeana and consequently enable European citizens to access this literature. Within the total BHL corpus of about 30 million pages and 80,000 items (volumes, books) currently available at http://www.biodiversitylibrary.org, the origin of about 24 million pages and 65,000 volumes is clearly defined in the metadata. Among this dataset, 53.1% of all pages (Figure 9) and 45.7 % of all volumes (Figure 10) within BHL are of European origin (published in Europe). This high level of European content within BHL shows the importance of BHL content for BHL-Europe: in making BHL data and content interoperable with European content providers and managing the content acquisition process over all partners, including BHL, these European content providers avoid a duplication of scanning efforts. Thus, collaborating with BHL also has an important economic consequence of saving Europeans tax money and this aspect should not be underestimated.
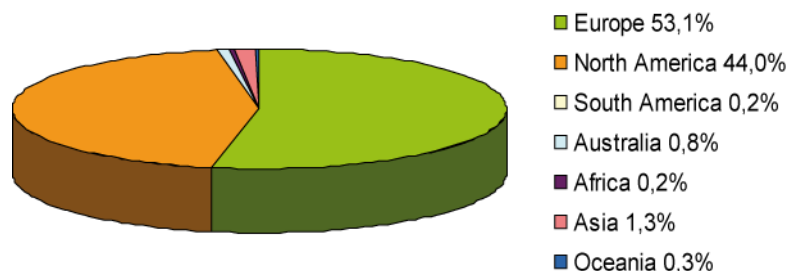

Figure 9: Pages within BHL ranged by continent of origin

Europe 45,7%
North America 51,0%
South America 0,5%
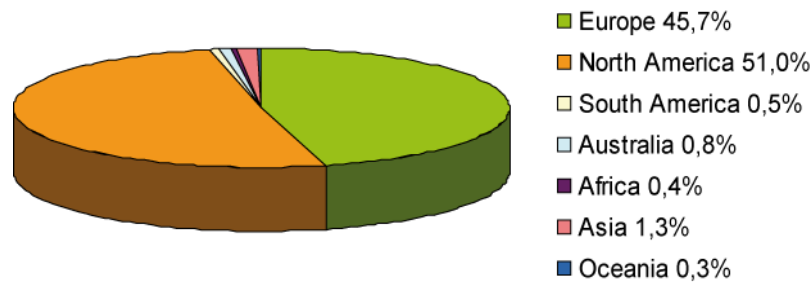Australia 0,8%
Africa 0,4%
Asia 1,3%
Oceania 0,3%

Figure 10: Volumes within BHL ranged by continent of origin

There is a high demand in Europe for biodiversity content. This requirement is shown by the number of European researchers travelling to institutions and libraries around the world, such as the MBLWHOI Library, the Smithsonian Institution Libraries and the Harvard University botanical and zoological libraries, but also by European users already using the BHL Pportal. Figure 11 demonstrates that over 40% of all BHL users are already Europeans, which indicates that BHL-Europe is and will be of great value for the European community. Hence, the aim of BHL-Europe is to provide a service situated in Europe and to extend the availability of biodiversity literature within Europe and make this content available to numerous groups of users.

Europe 47,9%
North America 29,9%
South America 7,6%
Australia 2,4%
Africa 1,4%
Asia 10,6%
Oceania 0,1%
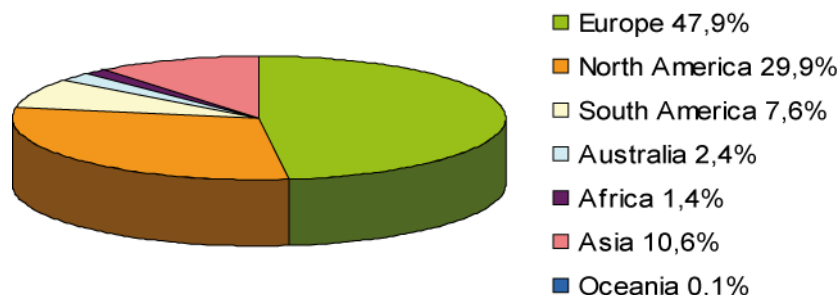
Figure 11: Users of the BHL Portal ranged by location (period of time: 01. April 2009 – 31. March 2010)

5.4 Europeana ingest plan

The Europeana ingest plan reflects the consensus on metadata and content delivery by our partners to BHL-Europe. Thus, it is the result of preliminary work, discussions and meetings within the first year of the project. Without the signed MoU in hand, for example, it is difficult to give enough details to Europeana for their ingest plan. The ingest plan also takes into account some of the requirements of our partners on technical aspects. All the work of the last months that is documented herein was thus leading towards the plan provided below.

In a first step we deliver the content of three BHL-Europe content providers to Europeana to test the ESE mapping and ingest procedure in cooperation with Europeana. We selected three experienced partners for this test that also reflect different local workflows. The content of BHL (NHM, SIL, MOBOT) and NAT is ingested first, followed by LANDOE in June 2010

for the Europeana Rhine release. By 10 June, 82,845 items from BHL institutions and Naturalis (NAT) have been ingested and made accessible to Europeana through BHL-Europe[5].

Partner Archive	April 2010 [volumes]	June 2010 [volumes]	Autumn 2010 [volumes]	January 2011 [volumes]	April 2011 [volumes]	Winter 2011 / Spring 2012 [volumes]
BHL (SIL; NHM; MOBOT)	79.329	79.329	79.329	79.329	94.966	100.337
NAT	3.516	3.516	3.516	3.516	3.516	3.516
LANDOE		2.300	2.300	2.300	2.750	3.000
UGOE			300	300	300	300
RMCA			20	20	100	100
BnF			510	510	510	680
MNHN			287	287	437	587
UH-Viikki			35	35	60	85
NMP				2	6	10
RBGE				46	46	46
NBGB				15	90	90
CSIC				7	120	120
UCPH					100	200
RBINS					100	230
HNHM					160	190
UB-Bielefeld					66	66
UBER					61	61
Sum of volumes	82.845	85.145	86.297	86.367	103.388	109.618

Table 3: Europeana Ingest (June 2010)

In a second step, we will deliver the content of more BHL-Europe partners to Europeana in autumn 2010. These are experienced partners that have a large amount of content, and sufficient personnel resources available by that date to work efficiently on the data harmonisation and ingest. This second group of content providers also includes partners that already deliver content to Europeana. The Göttinger Digitalisierungszentrum and the

---

[5] See "Darwin Among Scientific Treasures On Europeana" at http://version1.europeana.eu/web/guest/news/-/blogs/darwin-among-scientific-treasures-on-europeana and "BHL Europe Collection" in Europeana at http://europeana.eu/portal/brief-doc.html?query=europeana_collectionName:087*&view=table

Bibliothèque nationale de France both aggregate content for Europeana. At this stage we then have the chance to further coordinate with Europeana to deduplicate content in Europeana.

In a third and final step, all content providers that currently are in the beginning of their scanning programmes and still working on the establishment of their local repositories will be connected to Europeana.

Taking all the various aspects described above, Table 3 shows the ingest planning in volumes for Europeana based on the signed Memorandum of Understanding for those partners from whom we already have received the MoU (see Appendix B), based on the Description of Work for those partners we did not receive the MoUs yet (CSIC) and based on the already available data on the AIT server (April 2010: BHL, NAT).

Appendix

A: Content Provider Information

Question key (column numbers):

1. Current Content Provider [b]
2. Potential Content Provider [a]
3. Metadata Provider [a]
4. Answered Lib. Questionnaire 2009 [b]
5. Answered Lib. Questionnaire 2010 [b]
6. Number of Library Catalogues [a]
7. Online Catalogues Available [a]
8. Export via OAI-PMH/XML-Gateway possible or planned [a]
9. Export via Z39.50 possible [a]
10. Only manual export possible [a]
11. Records also hold in another bibliographic utility [a]
12. Libraries using MARC 21 [a]
13. Lib. using MARC local version [a]
14. Libraries using MODS [a]
15. Libraries using DC [a]
16. Provided data for GRIB import [b]
17. Data already imported into GRIB [b]
18. Provided metadata for AIT test [b]
19. Metadata records for BHL-E Test [b]
20. Pre Ingest Samples Receive [b]

Provider	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Wiley	n/a	1	1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
UH-Virkki	1	n/a	1	1	1	3	1	1	1	0	1	1	0	0	1	1	0	1	707	1
UGOE	1	n/a	1	0	2	1	0	0	0	0	1	0	0	1	1	0	1	1000		0
UCPH	1	n/a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
UBER	1	n/a	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	
UB-Bielefeld	1	n/a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Sp2000	1	n/a	1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	
SIL	1	n/a	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	
RMCA	1	n/a	1	1	1	10	1	1	1	0	1	1	0	0	0	1	1	0	1	
RBINS	1	n/a	1	1	1	2	1	0	0	1	1	0	0	0	1	0	0	6	0	
RBGE	1	n/a	1	1	1		1	0	1	0	1	1	0	0	1	1	1	744	0	
NMP	1	n/a	1	1	0	1	1	0	1	0	1	1	0	0	0	0	1	54	0	
NHMW	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
NHM	1	n/a	1	1	1	3	1	1	1	0	1	1	0	0	1	0	0	0	0	
NBGB	1	n/a	1	1	1	1	1	0	1	1	1	1	0	0	1	0	1	22	1	
NAT	1	n/a	1	1	1	2	1	1	0	0	1	1	0	1	1	0	1	114	1	
MSN	0	1	1	0	1	2	0	0	1	0	1	1	0	0	0	0	0	0	0	
MOBOT	1	n/a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10	0	
MNHN	1	n/a	1	0	1	2	1	1	1	0	1	0	0	0	0	0	1	100	0	
MIZPAS	0	1	1	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	
MfN	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	
LANDOE	1	n/a	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	100	0	
HNHM	1	n/a	1	0	1	2	0	1	1	0	0	0	0	0	0	0	0	0	1	
FUB-BGBM	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	
CSIC	1	n/a	1	1	1	3	1	1	1	0	1	1	0	0	0	0	1	1	0	
BnF	1	n/a	1	1	0	0	1	0	1	0	1	0	1	0	1	0	0	700	0	
Results	19	6	25	13	11	38	14	8	14	2	14	11	3	2	5	8	3	14	4558	5

[a] 1=yes, n/a = not applicable, 0=not known
[b] 1=yes, n/a = not applicable, 0=not yet

Partner of BHL-Europe
Partner of BHL-Europe & EDIT
Partner of BHL-Europe & EDIT & BHL

B: Underlying Content - Memorandum of Understanding

Category		NHM	NMP	UGOE	LANDOE	HNHM	UCPH	NAT	NBGB	RMCA	RBINS	BnF	MNHN	RBGE	SIL/MOBOT (BHL-US)	UH-Viikki	UBER	UB-Bielefeld
Quality of data hosted by content providers	TIFF	No	No	Yes		Yes	No	No	Yes	Yes	Yes	Yes	Yes		No	No	No	Yes
	JP2000	Yes	No	No		No	No	No	No	No	No	No	No		Yes	No	No	No
	JPG	Yes	Yes	No		No	No	No	Yes	Yes	No	Yes	Yes		Yes	No	Yes	No
	GIF	No	No	No		No	No	No	No	No	No	No	No		No	No	No	Yes
	PNG	No	No	No		No	No	No	No	No	Yes	No	No		No	No	Yes	No
	PDF	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No		Yes?	Yes	Yes	No
	OCR	Yes	No	No	Yes	Yes	No	Yes	Yes<1	No	Yes	Yes	No	No	Yes	Yes	Yes	No
	Metadata	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Resolution of data [dpi]	TIFF			300		400			300	300	300	300	300	300			600	600
	JP2000	300													300			
	JPG	300	300					300					150				300	
	GIF																	
	PNG																300	
	PDF	300							600	600	300					300		
Quality of data hosted by BHL-Europe	TIFF	No	No	Yes		Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
	JP2000	Yes	No	No		No	No	No	Yes	Yes	No	No	No		Yes	No	No	No
	JPG	Yes	Yes	No		No	Yes	No	Yes	No	No	Yes	Yes		Yes	No	Yes	No
	GIF	No	No	No		No	No	No	No	No	No	No	No		No	No	No	No
	PNG	Yes	No	No		No	No	No	No	No	Yes	Yes	No		No	No	No	No
	PDF	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No		Yes?	Yes	No	No
	OCR	Yes	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No
	Metadata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Resolution of data [dpi]	TIFF	300		300		400			300	300	300	300	300	300	300	300		600
	JP2000	300	100															
	JPG	300					400	300					150		300		300	
	GIF																	
	PNG																	
	PDF	300						300		600	300							
Quantity [volumes]	April 2010	3.995	2		2.300	80		3.153	10	20	0	510	287	53	76.693	35	61	66
	April 2011	4.666	6	300	2.750	160	100	3.220	60	100	0	510	473	53	90.300	60	61	66
	April 2012	5.337	10	300	3.000	190	200	3.290	85	100	0	680	587	53	95.000	85	61	66
Quantity [pages]	April 2010	2.382.713	300	100.000	500.000	23.000		86.500	5.000	15.000	0	150.000	112.000	25.211	30.200.335	5.985	12.200	21.800
	April 2011	2.782.713	2.000	100.000	600.000	44.000	50.000	88.500	35.000	50.000	30.000	150.000	172.000	25.211	36.000.000	12.985	12.200	21.800
	April 2012	3.182.713	3.000	100.000	650.000	50.000	100.000	90.500	50.000	50.000	70.000	200.000	232.000	25.211	39.000.000	19.985	12.200	21.800
IPR	Public domain (in terms of Open Access)	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
	Creative Commons/Science Commons	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No		Yes	Yes	No	Yes	No
	Public acces but restricted to commercial use															Yes		
Metadata	Dublin Core		No	Yes		Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No
	MARC/MARC21	Yes	No	No			Yes	No	Yes	Yes	Yes	Yes	Yes	Yes		Yes	No	No
	METS		No	Yes			No	Yes	No		No		No	No		No	Yes	No
	MODS		No	No			No	Yes	No		No		No	No		No	Yes	No
	Own schema		Yes	No	Yes				No	Yes	Yes			No		No	Yes	Yes
	Endnote		Yes						Yes									

C: BHL-Europe Master Data Element List

BHL-E Metadata Elements

based upon and mapped to the
Library of Congress Metadata for Digital Content -- Master Data Element List

Version 2: March 2010

Legend: NEW BHL metadata fields
Column "O" (Obligation) may have the following values: M= Mandatory, R= recommended, O= Optional

MARC INFO

Bibliographic Elements

	Element Name	DUBLINCORE/ESE	BHL	MARC	MODS element
O	TITLE INFORMATION*	DC TITLE			
M	Title	DC Title	Title	245	<titleInfo><title>
O	Translated Title			242	<titleInfo type="translated"><title>
O	Alternative Title	Dcterms: alternative		246	<titleInfo type="alternative"><title>
O	Abbreviated Title				<titleInfo type="abbreviated"><title>
O	Uniform Title			240	<titleInfo type="uniform"><title>
O	- Non-sort Title element				<titleInfo><nonSort>

Master Data Element list 4/14/2010

	Element Name	DUBLINCORE/ESE	BHL	MARC	MODS element
O	- Subtitle			245 $b	<titleInfo><subTitle>
O	- Title Part Number			245 $n	<titleInfo><partNumber>
O	- Title Part Name			246 $p	<titleInfo><partName>
	NAME INFORMATION*	DC CREATOR			
R / M	Name (NB: Either the creator name OR the publisher name is mandatory– one of these has to be filled in)	DC Creator	Name[11]	100/700	<name><namePart> <name authority="">
	Personal Name				<name type="personal"><namePart>
	Corporate Name			110/710	<name type="corporate"><namePart>
	Event Name			111/711	<name type="meeting"><namePart>
	- Role		Role		<name><role>
	- Part of Name				<name><namePart>
	- Family Name				<name><namePart type="family">
	- Given Name				<name><namePart type="given">
	- Date				<name><namePart type="date">
	PUBLICATION OR ORIGIN	DC PUBLISHER			
R / M	Publisher (NB: Either the creator name OR the publisher name is mandatory– one of these has to be filled in)	DC Publisher	Publisher	260 $ b	<originInfo><publisher>
R	Place			260 $ a	<originInfo><place>
	DESCRIPTION	DC DESCRIPTION			
	Description	DC description			
	Abstract	Dcterms: Abstract			

Master Data Element list 4/14/2010

	Element Name	DUBLINCORE/ESE	BHL	MARC	MODS element
	Contents	Dcterms: TableOfContents			<tableOfContents>
	RESOURCE DATE INFORMATION*	DC DATE			
	Date Created	Dcterms: created *The date of creation of the original object. Not the date of digitization (notes of europeana)			<originInfo><dateCreated> <originInfo><dateCreated encoding="">
M	Date Issued	Dcterms: issued * issued date of the original object (europeanna description)	Date indicated on Title	260 $c	<originInfo><dateIssued> <originInfo><dateIssued encoding="">
	Copyright Date				<originInfo><copyright Date> <originInfo><copyright Date encoding="">
	- Start date				<dateCreated point="start"> etc.
	- End date				<dateCreated point="end"> etc.
	EDITION INFORMATION				
	Edition		Edition	260	<originInfo><edition>
	LANGUAGE INFORMATION	DC LANGUAGE			
R	Language	DC Language of resource	Language	008/35-37 or 041	<language><languageTerm> <language><languageTerm authority="iso639-2b" type="code"> <language><languageTerm authority="code" objectPart="summary">
	TYPE OF RESOURCE*	DC TYPE /ESE TYPE			
M	Type of Resource	ESE type: TEXT	Type	008	<typeOfResource>

Master Data Element list 4/14/2010

	Element Name	DUBLINCORE/ESE	BHL	MARC	MODS element
		IMAGE SOUND VIDEO			
	GENRE*				
R	Genre	Periodicals, Monograph…	Genre http://www.loc.gov/marc/sourcecode/genre/genrelist.html#terms (Use of the MARC genre term list suggested)	008 http://www.loc.gov/marc/sourcecode/genre/genrelist.html#related	<genre>
	GEOGRAPHIC INFORMATION	DC COVERAGE (spatial)			
	Geographic (textual) *possibly interesting for BHL-E	Dc coverage *The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. This may be a named place, a location, a spatial coordinate, a period, date, date range or a named administrative entity.			<originInfo><place><placeTerm type="text"> -or- <subject><geographic>
	Geographic (coded)				<originInfo><place><placeTerm type="code">
	Hierarchical geographic				<subject><hierarchical Geographic>
	Coordinates				<subject><cartographic s> <coordinates>
	SUBJECT INFORMATION*	DC SUBJECT			
R	Classification	DC Subject	Subject	080-084	<classification authority="">
R	Topic	DC Subject	Subject	650	<subject authority=""><topic>
	Geographic			651	<subject authority=""><geograp

Master Data Element list 4/14/2010

	Element Name	DUBLINCORE/ESE	BHL	MARC	MODS element
					hic>
	Temporal			650	<subject authority=""><tempora l>
	Name			610	<subject authority=""><name></namePart>
	Title				
	NOTE INFORMATION				
	Note			500	<note type="" >
	FREQUENCY INFORMATION				
R	Frequency (SERIALS)			310	<originInfo><frequency>
	PHYSICAL DESCRIPTION	DC FORMAT			
R	Form	DC Format	Format		<physicalDescription><form>
	Mime Type				<physicalDescription><internetMediaType>
	Extent				<physicalDescription><extent>
	Digital Origin				<physicalDescription><digitalOrigin>
	RELATED ITEMS*	DC SOURCE *A related resource from which the described resource is derived in whole or in part. DC RELATION *A related resource. The recommended best practice is t identify the resource using a formal identification scheme.			
	Collection Title			772/773	<relatedItem type="host">
	Constituent Part			774	<relatedItem

Master Data Element list 4/14/2010

Element Name	DUBLINCORE/ESE	BHL	MARC	MODS element
				type="constituent">
Series			760	<relatedItem type="series">
Translation of			775	<relatedItem type="otherVersion">
Different Format			776	<relatedItem type="otherFormat">
Preceding Title			780	<relatedItem>
Succeeding Title			785	
Non-specific				<relatedItem>
- Title				<relatedItem><titleInfo><title>
- URL				<relatedItem><titleInfo><identifier>
- Grouping Identifier				<extension><relatedItem type="recordGroup"><identifier type="recordGroupID">
IDENTIFIERS*	DC IDENTIFIER			
M Identifier	DC identifier * identifier of the original object. For the URL of the original object see Europeana isShownBy (europeana note)	Identifier	01X-09X	<identifier type=""> e.g. <identifier type="hdl"> etc.
AUDIENCE INFORMATION				
Audience			008/22 521	<targetAudience>

NO MARC FIELDS (BHL-E FIELDS)

Article/Chapter LEVEL

Master Data Element list 4/14/2010

D2.4

M/R/O	Element Name	Use Note	Repeatable
O	Title		
O	Author		
O	Collection	Name of the serial or collection where the article or chapter are included	
O	Vol. number	Vol. number where the article or chapter are included	
O	Begin Page		
O	End Page		
O	Description or abstract		
O	Subject		
O	Links or relationship???		

Holdings Elements

	Element Name	DUBLINCORE/ESE	BHL	MARC	MODS element
	LOCATION INFORMATION				
R	Institution	DC PROVENANCE / DC provenance	Provenance *Institution that owns the original object	852 $a	<location><physicalLocation> <location><physicalLocation authority="">
	- Shelf locator		Call number	852	<location><shelfLocator>
	Electronic Location			856 $u	<location><url>
	RIGHTS INFORMATION	DC RIGHTS			
	Rights			506/540	<accessCondition>
	RECORD INFORMATION				
R	Record Identifier [12]		Record Identifier *number of the record assigned by the owner institution= control number	001/014	<recordInfo><recordIdentifier>
	Record Creation Date			008/00-05	<recordInfo><recordCreationDate>
	Record Change Date			005	<recordInfo><recordCh

Master Data Element list 4/14/2010

	Element Name	ESE	Element Definition	MARC	MODS element
					...angeDate>
	Record Creator				<recordInfo><recordContentSource>
	Record Origin				<recordInfo><recordOrigin>
			Barcode — The barcode affixed to the physical object scanned. (a library may have more than one copy of an item)	852 $p	
R	Language of Cataloging	ESE language	Record Language	040 $b	<recordInfo><languageOfCataloging>
			Issues (serial)		
			Volume (serial or Multivolume)		
			Number or part (serials)		
			Year (serials)		

NO MARC INFO (INFORMATION FROM DG)

ITEM

	Element Name	BHL	Element Definition	DCORE/ESE	MARC	MODS element
R	Date captured	Date Created	Date item scanned			<originInfo><dateCaptured><originInfo><dateCaptUred enconding>
R		Date item modified	Date item modified			
		Scanning Contributor	Institution(s) assisting with scanning. (roles: taxonomist, scan centre etc.)			
R		Scanning Institu	Institution res ponsible for			

Master Data Element list 4/14/2010

		tion	scanning.			
M		Provider	Name of BHL-E Content provider			
R		Local library Identifier [13]??? Digital Identifier??	Unique identifier for ite m scanned, from scanning library catalogue. Note: Use same as local library identifier if this information is not easily accessible			
R		Sequence	If a multivolume monograph or serial, the sort order for this item relative to the other objects scanned in this title.			
R		Start Date	Beginning year of bound object			
R		Start Volume	Beginning volume number of bound object			

PAGE

	Element Name	BHL	Element Definition	DCORE/ESE	MARC	MODS element
R	Date captured	Date Created	Date page scanned			
R		Date last modified	Date page modified			
R		File name	File name of associated image file page			
R		Sequence [14]	Sort order for given page, relative to order of pages in item			
		Embedded rights	Only required where relating to a specific page			
R		Page number	Physical printed page number. Values [1]			

[1] Appendix, blank, cover, illustration, index …

Master Data Element list 4/14/2010

RIGHTS

Element Name	BHL	Element Definition	DCORE/ESE	MARC	MODS element
	More than one Rights …	Default: 1			
	Organisation Issuing….	Default: "rights owner"			
R	Rights owner				
R	Rights status	One of "in copyright" "out of copyright" "public domain"			

Suggested other additional BHL-E fields

Element Name	BHL	Element Definition	DCORE/ESE	MARC	MODS element
	Foreign Keys	Identifier for titles as recognized by popular indices (Linneus number?)			
	Taxonomic names				
	Geographical codes (see above)				
	other				

Master Data Element list 4/14/2010