

ECP-2008-DILI-518001

BHL-Europe

Content analysis & management status report 3 (metadata, page numbers, content providers)

Deliverable number	<i>D2.8</i>
Dissemination level	<i>Restricted</i>
Delivery date	<i>18 May 2012</i>
Status	<i>Final</i>
Authors	<i>Melita BIRTHÄLMER, Henning SCHOLZ</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
Henning Scholz	MfN
Boris Jacob	RMCA
Andreas Krausz	VZG
Graham Higley	NHM
Graham Hardy	RBGE

0.2 Revision History

Revision Date	Author	Version	Change Reference & Summary
02 April 2012	Melita Birthälmer	0.1	First draft
19 April 2012	Boris Jacob	0.1	Adding additional information on chapter 4.1
24 April 2012	Melita Birthälmer	0.2	Draft incorporating all available information
15 May 2012	Henning Scholz	0.3	Draft incorporating input by NHM & RBGE
18 May 2012	Henning Scholz	1.0	Final version

0.3 Reviewers

This document requires the following reviews and approvals.

Name	Date	Version
BHL-Europe PMG & Consortium	10 May 2012	0.2

0.4 Distribution

This document has been distributed to:

Group	Date of issue	Version
BHL-Europe Consortium	2 May 2012	0.2
BHL-Europe Consortium	21 May 2012	1.0

1 Table of Contents

0	DOCUMENT HISTORY	2
0.1	CONTRIBUTORS.....	2
0.2	REVISION HISTORY	2
0.3	REVIEWERS.....	2
0.4	DISTRIBUTION.....	2
1	TABLE OF CONTENTS	3
2	PURPOSE	4
3	BACKGROUND.....	4
4	CONTENT MANAGEMENT	5
4.1	GLOBAL REFERENCES INDEX TO BIODIVERSITY (GRIB)	5
4.1.1	<i>Delivery and update of the GRIB data</i>	5
4.1.2	<i>Data heterogeneity</i>	6
4.1.3	<i>Data status of the GRIB</i>	7
4.1.4	<i>GRIB outlook</i>	10
4.2	CONTENT PROVIDERS.....	11
4.2.1	<i>Primary content providers</i>	11
4.2.2	<i>Additional acquired content providers</i>	12
4.2.3	<i>Content acquisition</i>	14
4.2.3.1	<i>Recommended methodology</i>	14
4.2.4	<i>Content Acquisition - Lessons learned</i>	16
4.2.5	<i>Outlook</i>	17
5	CONTENT ANALYSIS	18
5.1	CONTENT UPLOAD	18
5.2	UNDERLYING CONTENT OF BHL-EUROPE.....	20
5.2.1	<i>Interconnected repositories</i>	22
5.3	CONTENT AVAILABLE THROUGH EUROPEANA	23
5.4	EUROPEAN CULTURAL HERITAGE DISTRIBUTION.....	24
5.5	CONTENT AVAILABILITY IN EUROPE	25
5.6	LESSONS LEARNED.....	27
	APPENDIX.....	28
	A: CALCULATION TABLE FOR ESTIMATION OF NUMBER OF PAGES WITHIN EUROPEANA	28
	B: BHL-EUROPE IN A NUTSHELL.....	29

2 Purpose

This document aims to give an overview of the content specific progress within BHL-Europe after the third and final project year. Overviews have already been given for each project year in the “Content analysis & management status report 1” D2.4 for the first year and the subsequent delivery D2.7 for the second year. This report gives updated information with reference to the D2.4 and D2.7.

3 Background

BHL-Europe supports European institutions in their digitisation efforts, manages the acquisition of biodiversity content in Europe and hosts digitised biodiversity content. Although this document is a deliverable within WP2, it also contains information on WP3 outcomes as they highly impact WP2 outcomes, particularly in the area of content analysis and content acquisition.

Chapter 4 gives an overview of the content management within BHL-Europe. First of all, the status of the Global References to Biodiversity Index (GRIB) is given in Chapter 4.1. Detailed information about BHL-Europe content providers including a section lessons learned and an outlook for possible activities with regards to content acquisition is given.

Chapter 5 addresses content analysis of the BHL-Europe content. The process of uploading content to the BHL-Europe server is described in Chapter 5.1. Information about the underlying content of BHL-Europe is presented in Chapter 5.2. BHL-Europe acts additionally as a content aggregator for Europeana and therefore Chapter 5.3 shows the availability of BHL-Europe content within Europeana. Furthermore, the distribution of European cultural heritage is shown and an estimation of published content within the domain of biodiversity in the European Union is given in Chapter 5.4 and Chapter 5.5. The content analysis closes with a lessons learned section.

4 Content Management

Content Management is an important task within BHL-Europe and comprises the management of the scanning process, the coordination of and communication with the content providers and data providers, and the content acquisition of new content for BHL-Europe. Within BHL-Europe we differentiate between data and content provider. A data provider is a partner of the BHL-Europe project providing only metadata (e.g. library catalogue) to BHL-Europe for the Global References Index to Biodiversity. A content provider is in contrast a partner of the BHL-Europe project providing content and metadata to the BHL-Europe repository and system.

In order to facilitate communication with our content and data providers a mailing list (bhl-e.cp@lists.hu-berlin.de) has been set up and is in use on a regular basis. This mailing list is used for updating BHL-Europe content and data providers and announcing important information to our content providers.

4.1 *Global References Index to Biodiversity (GRIB)*

The Global References Index to Biodiversity (GRIB)² is the joint literature index from the projects Biodiversity Heritage Library for Europe (BHL-Europe) and European Distributed Institute of Taxonomy (EDIT), developed together with the head office of the Common Library Network GBV (VZG). It is based on the OCLC PICA-CBS (Central Library System) hosted by the VZG and accessible via standard interfaces (WWW, Z39.50, SRU, internal XML and OAI). The GRIB incorporates bibliographic data from library catalogues provided by natural history institutions that are part of these projects. While importing library catalogues to the GRIB, data sets are matched and merged, if they have been identified as duplicates (de-duplication). The GRIB lists the libraries where publications can be found and links records to publications already digitised. The GRIB also offers the possibility to indicate publications to be digitised that are not yet accessible online, allowing for an inter-institutional scanning planning among BHL-Europe partners. Via different search mechanisms the GRIB can help content providers in their digitisation planning. For scientists the GRIB is an advantage because they receive search results that not only show publications that are already digitised, but also those who exist in libraries of natural history institutions.

4.1.1 **Delivery and update of the GRIB data**

The mechanism to update bibliographic data into the references index consists of three major steps:

1. Delivery of the data by the data provider
2. Conversion of the delivered data to Pica+ format
3. De-duplication and merging of the records

² Detailed information about functionalities and technical specifications of the GRIB can be found in the deliverable “Final and enhanced Web-database for content management and collection analysis” D2.5.

While method and frequency of delivering the data in step 1 varies on the partners' side, there is a fixed process on the side of the VZG in step 2 and 3.

A data provider can provide bibliographic data to the VZG by manually exporting the data or by providing a Z39.50 or an OAI-PMH interface for automatically harvesting by the VZG. Once the VZG receives the data in the format of the providing institution, it has to be mapped manually to the GRIB database format Pica+, which is the basis of the "Match & Merge" process and thus de-duplicates records. During the de-duplication process different results can occur:

- A match is detected and the new imported data is stored together with the already existing data.
- A new imported item is detected as a possible match and is kept distinct but marked for further manual checking.
- No match is detected and all imported records are kept independently in the GRIB.

The amount of time spent mapping the original data to the Pica+ format, and identifying duplicates, is dependent on the quality of data supplied by each data provider.

In general, one can say, that the better the data format and the richer the provided data, the better are the results. Good formats in this sense are standard bibliographic data formats like different flavours of MARC, Pica, MAB or even DC.

However, in the majority of cases received data for the import to the GRIB is not in a standard bibliographic data format. The duration for the de-duplication process is reduced if adequate fields for identifying a record as unique are provided, such as "ISBN" or "OCLC number". If unique fields like that are not included in the data, the de-duplication algorithm uses the "Title" of the record which might not identify a new imported record as a duplicate because of variations in spelling, for instance. Additionally, items with the same or very similar title might erroneously be marked as duplicates. Thus it must be marked for further manual checking. The quite low quality of available data slows down the process of de-duplications and thus the final import and availability of the data within the GRIB.

4.1.2 Data heterogeneity

There are some general remarks that need to be mentioned in order to improve the outcome of future projects dealing with metadata in this domain. The following given factor had a high impact on the outcomes of the GRIB. During the project time it has been recognised that the availability of library catalogues in a standard metadata format is not naturally, it is more over very rare. Metadata of library catalogues have been mostly provided in a proprietary format (e.g. .xls files) to us which makes it very difficult and time consuming to understand the delivered data and map it to a standard format such as Pica+. The mapping to a standard format is necessary to import the data and run automatic processes such as the de-duplication algorithm for merging duplicates within the provided data.

In order to understand the fields of a proprietary metadata format and thus being able to map and merge the data, it is necessary that data providers provide additionally a description of the format and fields they use. In a nutshell, aggregating data from a large amount of libraries with a maximum heterogeneity of content is a complex task and needs by far more efforts than available and manageable within this project only as a subtask of a work package.

4.1.3 Data status of the GRIB

As of 18 May 2012 the GRIB holds 785,608 de-duplicated records³ from the provided library catalogues of NAT, NHM, FU-BGBM, MfN, NBGB, RBGE, RBINS, UH-Viikki. This is the status after the re-import of the data from January/February 2012, which was done to enhance the data quality.

Table 1 gives an overview of the library that has been created from the imported 747,035 bibliographic data records provided by the consortium. 100,616 records have been identified as duplicates and merged. 40,239 records have been marked as possible duplicates by the system. Those possible duplicates cannot be merged automatically, but would need manual checking first.

The table shows the available data at the institutions according to the collected library questionnaire from May/June 2010. Not all catalogues have been delivered or harvested for import into the GRIB because testing with the ones submitted hasn't been finished so far, due to reasons mentioned in the chapter above and non GRIB related technical hardware issues at the VZG. Apart from the ones incorporated into the GRIB the catalogues of HNHM and RMCA are exported but need further description, UBER, UGOE and AnimalBase need to be harvested from their source systems by the VZG and reviewed for import into the GRIB. The remaining catalogues are still due to be imported after the official end of the BHL-Europe project. Additional information is given in the chapter *GRIB outlook*.

³ The number of de-duplicated records in the GRIB is retrieved by searching for the unique record ID's Pica Production Number (PPN). The related search term for the simple search interface at <http://grib.gbv.de> is "ppn = 0?"

Nr.	Name	Abbreviation [GRIB internal library number]	Available data	Delivered data	Imported data into the GRIB	Number of records
1	Stichting Nationaal Natuurhistorisch Museum Naturalis	NAT [bib 4792]	1 library catalogue ⁽¹⁾	1 library catalogue	1 library catalogue	76,173
2	Natural History Museum	NHM [bib 4793]	1 library catalogue	1 library catalogue	1 library catalogue	100,992
3	Freie Universität Berlin	FU-BGBM [bib 0188]	1 library catalogue	1 library catalogue	1 library catalogue	63,787
4	Museum für Naturkunde Berlin	MfN [bib 4795]	1 library catalogue	1 library catalogue	1 library catalogue	201,980
5	Royal Belgian Institute of Natural Sciences	RBINS [bib 4796]	1 library catalogue	1 library catalogue	1 library catalogue	196,206
6	National Botanic Garden of Belgium	NBGB [bib 4797]	1 library catalogue	1 library catalogue	1 library catalogue	107,897
7	Royal Botanic Garden Edinburgh	RBGE [bib 4794]	1 library catalogue	1 library catalogue	1 library catalogue	No figure ⁽⁴⁾
8	Helsingin yliopisto	UH-Viikki [bib 4799]	1 library catalogue	1 library catalogue	1 library catalogue	No figure ⁽⁴⁾
9	Humboldt- Universität zu Berlin	UBER	1 repository database	1 repository database	1 repository database	No figure ⁽⁴⁾
10	Nationaal Herbarium Nederland	NHN	1 library catalogue ⁽¹⁾	1 library catalogue	1 library catalogue	No figure ⁽⁴⁾
11	Royal Museum for Central Africa	RMCA	10 library catalogues	10 library catalogues	10 library catalogues	No figure ⁽⁴⁾
12	Biodiversity Heritage Library	BHL-US	1 repository database	1 repository database	Import not finished ⁽⁵⁾	

Nr.	Name	Abbreviation	Available data	Delivered data	Imported data into the GRIB	Number of records
13	Hungarian Natural History Museum	HNHM	1 library catalogue 1 article database	1 library catalogue 1 article database	missing field description ⁽²⁾	
14	Consejo Superior de Investigaciones Cientificas; Museo nacional de Ciencias Naturales	CSIC	1 library catalogue 1 archival catalogue			
15	Museum and Institute of Zoology, Polish Academy of Sciences	MIZPAS	1 library catalogue			
16	Museum National d'Histoire Naturelle	MNHN	1 library catalogue 1 repository database			
17	Università degli Studi di Firenze	MSN	1 library catalogue 1 repository database			
18	Narodni muzeum	NMP	1 library catalogue			
19	Georg-August-Universität Göttingen	UGOE	1 library catalogue 1 repository database ⁽³⁾			
20	Smithsonian Institution	SIL	1 library catalogue			

Total amount of imported records

747,035

Table 1: Available library catalogues provided by content providers & status of import to the GRIB (18.05.2012)

⁽¹⁾ ...NAT provides a second library catalogue from the Nationaal Herbarium Nederland. This catalogue is listed under Nr. 10.

⁽²⁾ ... Both catalogue has been delivered by HNHM, however could not be imported to the GRIB as a description of the used fields within the exported catalogue was not available. In order to correctly map the fields to the PICA+ format, we need to be able to understand which information is provided in which field. Without this additional information we are not able to import catalogues to the GRIB.

⁽³⁾ ... the AnimalBase repository catalogue is available from UGOE.

⁽⁴⁾ ...All catalogues are already imported but the figures are not displayed separately to give a resolution per library. An index update is necessary to display the data correctly, which is planned for end of May. Therefore, the number of total records in the GRIB is higher than the number of individual records.

⁽⁵⁾ ...The BHL-US data are available but not imported in the database yet.

4.1.4 GRIB outlook

MfN and VZG have an agreement for the VZG to host and maintain the GRIB infrastructure, as well as updating the index with information from BHL-Europe, EDIT and BHL partners until the end of February 2020. The fact that expectations for technology and import of data in regard to the GRIB were not met due to reasons of data quality and non GRIB related technical difficulties at the VZG, does not influence the future of this agreement. After the BHL-Europe project ends the VZG aims to fix the technical problems and make the GRIB accessible 24 hours, 7 days a week. The library catalogue data submitted by partners is to remain within the GRIB and updates from already provided catalogues, as well as new data provided will be incorporated into the GRIB in cooperation with the BHL-Europe office at the MfN. The copyright of the original data is retained by the institution supplying the data. A physical export of the GRIB is only possible in agreement with the MfN.

As a first step to plan the future of the GRIB, a working group will be established to discuss and agree on data submission guidelines to facilitate the mapping process for the VZG and thus facilitate the ingest and update of catalogue in the GRIB. In this process we also need to solve some issues related to the handling of serial data (volume, issue, article) in the GRIB. After this process is finished successfully (including a test of ingest following the guidelines), the remaining libraries will be approached to deliver their catalogues in a format that is approved to be ingested effectively. In parallel, the technical infrastructure at VZG will be updated and the widget functionality will be improved to work more stable and effectively. This will then allow to use the GRIB for digitisation management as planned originally.

4.2 Content providers

The following chapter gives an overview of primary content providers since the beginning of the project. Content providers were acquired during the first, second and third project years. The process of content acquisition within BHL-Europe and lessons learned are also documented within this chapter.

4.2.1 Primary content providers

BHL-Europe content providers aim to provide high quality digital content representing the biodiversity domain. The content providers within the consortium have been selected on the basis of their ability to contribute key biodiversity and taxonomic literature on as wide a range of organisms as possible. Thus national botanic gardens and natural history museums within European Union partner states are content providers because of the richness of their library collections. From the beginning of the project BHL-Europe has had 17 primary content providers involved as consortium members which are listed in Table 2.

CP No.	Name	Acronym	Country
1	Natural History Museum	NHM	United Kingdom
2	Národní museum	NMP	Czech Republic
3	Georg-August-Universität Göttingen Stiftung Öffentlichen Rechts	UGOE	Germany
4	Land Oberösterreich (Oberösterreichische Landesmuseen)	LANDOE	Austria
5	Hungarian Natural History Museum	HNHM	Hungary
6	University of Copenhagen (The Natural History Museum of Denmark)	UCPH	Denmark
7	Stichting Nationaal Natuurhistorisch Museum Naturalis	NAT	Netherlands
8	National Botanic Garden of Belgium	NBGB	Belgium
9	Royal Museum for Central Africa	RMCA	Belgium
10	Royal Belgian Institute of Natural Sciences	RBINS	Belgium
11	Bibliothèque nationale de France	BnF	France
12	Muséum national d'Histoire naturelle	MNHN	France
13	Consejo Superior de Investigaciones Científicas	CSIC	Spain
14	Royal Botanic Garden Edinburgh	RBGE	United Kingdom
15	Smithsonian Institution	SIL	United States of America
16	Missouri Botanical Garden	MOBOT	United States of America
17	Helsingin yliopisto, University of Helsinki, Viikki Science Library	UH-Viikki	Finland

Table 2: List of primary content providers at the beginning of the project

4.2.2 Additional acquired content providers

First year of the project

BHL-Europe aimed to acquire further content providers to increase the amount of digital content significantly during the project lifetime and thus make as much content as possible available to numerous groups of users such as scientists, decision makers and the European population. New content providers have joined the project and will contribute content to BHL-Europe. During the first year of the project two institutions entered the project as new content providers (Table 3).

CP No.	Name	Acronym	Country
18	Humboldt-Universität zu Berlin	UBER	Germany
19	Universitätsbibliothek Bielefeld	UB-Bielefeld	Germany

Table 3: Additional content provider after first year of project lifetime

The *Humboldt Universität zu Berlin (UBER)* was already part of the consortium with the role of a disseminator and digital library expert and has added the role of a content provider.

The *Universitätsbibliothek Bielefeld (UB-Bielefeld)* has been acquired as a new content provider for BHL-Europe.

Second year of the project

During the second year of the project 6 institutions agreed to act as new content providers for BHL-Europe (Table 4).

CP No.	Name	Acronym	Country
20	Università degli Studi di Firenze	MSN	Italy
21	Museum and Institute of Zoology, Polish Academy of Sciences	MIZPAS	Poland
22	Museum für Naturkunde Berlin	MfN	Germany
23	Institute of Paleobiology Polish Academy of Sciences	PAS	Poland
24	Gesellschaft für Biologische Systematik	GfBS	Germany
25	Landesbibliothekszenrum Rheinland-Pfalz - dilibri	dilibri	Germany

Table 4: Additional content providers after second year of project lifetime

The *Università degli Studi di Firenze (MSN)* and the *Museum and Institute of Zoology, Polish Academy of Sciences (MIZPAS)* were part of the consortium since the beginning of the project. Both already had the role of a domain and language expert and during the second year of the project added the role of content provider.

MSN worked together with our consortium partner and content provider UGOE in a collaborative project and prepared biodiversity literature for digitisation within the UGOE

library. Literature - most of Italian origin - was transported from Florence to Göttingen in winter 2010 and is provided to BHL-Europe through UGOE.

MIZPAS successfully submitted a digital repository project in 2010 with the start of project in 2011 and started digitizing their resources in winter 2011.

The *Museum für Naturkunde Berlin* (MfN) was successful in raising funds for a book scanner beginning of 2011. During 2011 the scanner has been purchased and a scan operator has been employed. The scanning of the first books started beginning of 2012.

The *Institute of Paleobiology of the Polish Academy of Science*⁴ has been acquired as a new content provider to BHL-Europe and provides digital content from their already online available archive (<http://www.app.pan.pl/archives.html>) to BHL-Europe including future born digital publications.

The *Gesellschaft für Biologische Systematik*⁵ (GfBS) is a German society for biological systematics and contributes all their publications older than two years to BHL-Europe.

*Dilibri*⁶ is the digitised collection of regional studies about the federal state Rhineland-Palatinate and provides biodiversity related content to BHL-Europe as a new content provider.

Third year of the project

During the third year of the project three institutions entered the project as new content providers for BHL-Europe (Table 5).

Nr. of CP	Name	Abbreviation	Country
26	University library Johann Christian Senckenberg	UB-Frankfurt	Germany
27	Freie Universität Berlin	FUB-BGBM	Germany
28	Université de Rennes 1	Rennes 1	France

Table 5: Additional content providers after third year of project lifetime

The *University library Johann Christian Senckenberg* (UB-Frankfurt) and the *Freie Universität Berlin* (FUB-BGBM) were successful in submitting a digitisation project for German Botanical Journals in the period of 1753-1914. The project started in summer 2011. Content from FUB-BGBM will be provided also by the University library Johann Christian Senckenberg.

The *Université de Rennes 1* has been acquired as a new content provider. The library of the Université de Rennes 1 digitises books and manuscripts from its heritage collections from the 18th to early 20th century in life science and related areas such as botany, zoology and medicine.

⁴ www.paleo.pan.pl

⁵ www.gfbs-home.de

⁶ www.dilibri.de

4.2.3 Content acquisition

Attracting new content providers was and needs to be an ongoing process also after the project lifetime in order to keep the BHL-Europe portal attractive for our users.

A “BHL-Europe in a nutshell” document has been prepared within WP2 during the project lifetime as a dissemination document for the content acquisition. The document aims to be a very brief but informative document including the most important facts about BHL-Europe, benefits and IPR issues. The document has been translated by our consortium and is available in following languages:

- English
- Spanish
- French
- Russian
- German

The English document is available in the Appendix within this document. All available languages have been made available to the dissemination team for publication on the official BHL-Europe website⁷ and have been uploaded to the BHL-wiki⁸. The document has already been used by our consortium during the last two project years.

4.2.3.1 Recommended methodology

During the second year of the project consortium members took part in a new and additional method for content acquisition. More precisely, the identified method took advantage of our consortiums’ connections in order to attract new content providers within their own countries using the “BHL-Europe in a nutshell” document. This new method is designed to be sustainable to ensure the continued growth of BHL-Europe not only during the project time but also after the end of the financial aid from the European Union.

The new method was tested by our consortium partners RBGE during the project lifetime. The process of using the “BHL-Europe in a nutshell” document, feedback and lessons learned are described below in order to impart knowledge beyond the project lifetime for the proposed incremental growth of BHL-Europe in the business plan deliverable 1.9.

Process of using BHL-Europe in a nutshell for content acquisition

RBGE identified two ways of using the “BHL-Europe in a nutshell” document for the content acquisition:

- Send emails to known and unknown individuals within the biodiversity community of the own country – with and without prior telephone contact.
- Use in poster sessions at conferences.

Both ways are described in more detail hereafter.

⁷ www.bhle.eu

⁸ [https://bhl.wikispaces.com/Content+ Management](https://bhl.wikispaces.com/Content+Management)

1. *Send emails to known and unknown individuals within the biodiversity community of the own country*

- Send a cover letter along with “BHL-Europe in a nutshell” to institutional staff with contacts within the biodiversity community to receive feedback and comments for prospective content providers that can be approached.
- Prepare a review of your own imprints included in BHL-Europe to give an example of current content.
- Prepare - if necessary - an introductory information sheet about your institution in regards to BHL-Europe
- Send emails to secretaries of selected natural history organisations. Where no secretary can be identified send information to the general email contact for the society. Include following documents:
 - i. Cover letter
 - ii. BHL-Europe in a nutshell
 - iii. Review of own imprints
 - iv. Introductory information sheet about your institution - if necessary
- Try to answer questions that arise from the content providers to give the prospective content provider an understanding of how the processes work for you within BHL-Europe, providing a real and practical example.
- Seek assistance from BHL-Europe Executive Director- if necessary - for questions that arise.
- Assist the prospective content provider with preparing to sign the MoU or connect the prospective content provider and BHL-Europe Executive Director for continuing negotiations.

2. *Use in poster sessions at conferences*

- Print “BHL-Europe in a nutshell” for display at conferences, e.g. poster session.

Although the main focus of attending conferences might not be acquisition of new content providers, it is an opportunity to publicise BHL-Europe, Europeana and your involvement with BHL-Europe. Consequently, this exposure might help indirectly with acquiring new content for the project and Europeana.

Feedback and lessons learned

Acquiring new content providers for BHL-Europe is not an easy task. RBGE sent out information about BHL-Europe to 32 prospective content providers within Scotland via email and received 3 responses, 2 asking for further information and one noting that the project sounded very interesting and he would forward the message to the society committee and put it on the agenda for the next meeting. Feedback from further telephone discussions centred around issues relating to IPR and the open access nature of the project. The feedback from advertising BHL-Europe in poster sessions was limited with only one request for further information. Building on experience, it shows that content acquisition is not a short-term

process and much effort is needed to communicate with prospective content providers. The whole process from the beginning until signing the MoU might take several months.

The experience so far shows that there is a general support for the central aims of BHL-Europe - namely free access to biodiversity literature. However, there are two key problems that need to be overcome:

1. The matter of funding sources for the digitisation of potential new content.
2. Intellectual property rights (IPR) issues with potential content.

Taking each of these in turn, content is potentially available but funding for digitisation is not. In the current constrained financial climate it is increasingly more difficult to attract funds. The first problem might be avoided by first contacting only those institutions with digital content available on their own website. However, this makes it more difficult to attract those institutions for BHL-Europe as they already have their own online presence.

Questions relating to the issue of intellectual property rights, covered areas such as how do you identify, record and contact third party rights holders, and how much time does implementing and following through these processes take.

4.2.4 Content Acquisition - Lessons learned

Content acquisition was an important task until the end of the project, taken with a lot of efforts. Potential content providers request a high amount of additional information on the whole process and can efficiently only be acquired successfully with a working environment. Building on experience, this task needs to be taken up in future projects with additional effort by the whole project team in order to increase the visibility of the project - such as BHL-Europe - to as broad a range of potential content providers. One of the most frequently asked question by potential and also already acquired content providers was “When and where can we see any results of the already provided content”. Without providing evidence that proposed values can be delivered it is very difficult to convince potential providers to join the project. Therefore, delayed deadlines with the visible project outputs led to not being able to acquire new content providers during the last project year. Moreover, instead of being able to concentrate on new prospective content providers the main task during the last project year was keeping the already acquired ones and of course also the primary content providers still interested in providing us content, even if we couldn’t show any visible results to them. Therefore, the process of acquiring new content providers requires first of all visible results, simple and fast communication through several work packages in order to be able to give recommendations and feedback to prospective content providers in a timely manner.

In general, communications with content providers is labour-intensive and the time interval beginning with the first contact varies from a minimum of half year to even several years. This might be caused by diverse IPR questions and issues, non-availability of metadata in a standard format, or the matter of funding sources for digitisation projects.

4.2.5 Outlook

There is much more potential content available in the domain of biodiversity that could not be included in BHL-Europe due to the limited timeframe of the project. To keep BHL-Europe alive and attractive for the users it might be beneficial to approach additional institutions that already have digitised content.

Therefore a list of potential content providers - with already available digitised content - including all necessary contact details has been prepared and handed over to the Executive Director. This list can be used for further content acquisition for an ongoing BHL-Europe or future projects.

Table 6 shows an extract of institutions that have been identified as potential content providers.

Name	Country
Aboca Museum - Bibliotheca Antiqua	Italy
Biblioteca Digital de Botânica University of Coimbra	Portugal
BioLib - Kurt Stübers Online-Bibliothek	Germany
Cyberliber: an Electronic Library for Mycology	UK
Digital Library of the Real Jardín Botánico of Madrid	Spain
Digitale Bibliothek Braunschweig- section botany zoology	Germany
Herzogin Anna Amalia Bibliothek - section biology	Germany
Jean-Baptiste Lamarck (1744-1829): works and heritage	UK
The Linnaean Correspondence	France
Archbishop chateau and gardens in Kromeríž	Czech Republic
Oxford Digital Library: Key 19th century entomological literature	UK
SEALS - Swiss Electronic Academic Library Service	Switzerland
University of Strasbourg	France
Tela Botanica	France

Table 6: Extract of institutions as potential content providers for BHL-Europe

However it is recommended to approach these institutions only after the BHL-Europe portal is ready for dissemination and BHL-Europe content is visible to prospective content providers.

5 Content Analysis

Content within BHL-Europe 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 license. The digital content must either be in the public domain, or else the content contributors must have permission from intellectual property owners to provide access under Creative Commons. In this part of the report we aim to give a detailed analysis of the freely available content provided by BHL-Europe content providers. The process of uploading content is described as well as the underlying content of BHL-Europe. BHL-Europe content that is available through Europeana is also described. Additionally the relationship between European cultural heritage and content within BHL-US, as well as estimates of published biodiversity content within the European Union are presented. The content analysis chapter closes with a lessons learned section.

5.1 Content upload

Managing data upload to the BHL-Europe server began in autumn 2010 in cooperation with WP3. Before content providers were able to upload content to the BHL-Europe server hosted at the Natural History Museum (NHM) in London requirements had to be reviewed. Functional and technical requirements were gathered and then the feasibility of meeting these requirements was discussed with NHM staff, in line with NHM infrastructure and security constraints. Simultaneously all content providers were asked to provide primary and backup contacts for the content upload within their institution. In order to coordinate the required information a wiki page⁹ was established. According to the information provided by our content providers and the functional and technical specifications, NHM installed the FTPS server and created for each content provider an account and sent out logon details in November 2010. It turned out that some content providers had difficulties using a FTPS connection within their own institution. Therefore all content providers were asked during the Content Provider and Technical meeting in London to check their FTPS connection and report any problems. The above mentioned wiki page contains a list of all BHL-Europe content providers, general information about the FTPS account and a 'How to use' section. For those content providers who had problems with the primary FTPS solution a second solution has been provided.

Content providers have been asked individually to upload content to the NHM server. First they uploaded test content to the NHM server including the scanned images and the corresponding metadata according to the old file submission guidelines¹⁰ until January 2012 and according to the updated and new file submission guidelines¹¹ from January 2012 on. The change in the guidelines was an essential step after it has been decided within the Project Management Group to move forward with a new Pre-Ingest tool. BHL-Europe decided to

⁹ https://bhl.wikispaces.com/BHL-Europe_Data_Upload

¹⁰ https://bhl.wikispaces.com/BHL-E_WP3_PREINGEST

¹¹ https://bhl.wikispaces.com/file/view/BHL-Europe_File+Submission+Guidelines.pdf

complement the already available rules in order to allow an easier automatic ingest of the provided content and to clarify open questions from the content providers.

According to the designed workflow a content provider should receive feedback for its uploaded test content before uploading additional content to the BHL-Europe server. This should on the one hand serve as security for the future work content providers are doing and on the other hand avoid additional work for our ingest team with the bulk content. Therefore, after the first test upload, the ingest team is asked to check the provided test content and ingest it using the Pre-Ingest tool until it is finally visible in the portal. Unfortunately the feedback step couldn't be reached for many of our content providers because of the technical problems with the developed tools. Therefore most of our content providers have been waiting for feedback almost until one month before the end of the project or have yet to receive feedback. This situation makes it very difficult for our content providers to provide all their promised content according to the new guidelines in time.

Figure 1 shows the exemplary content upload workflow that should be followed in practice.

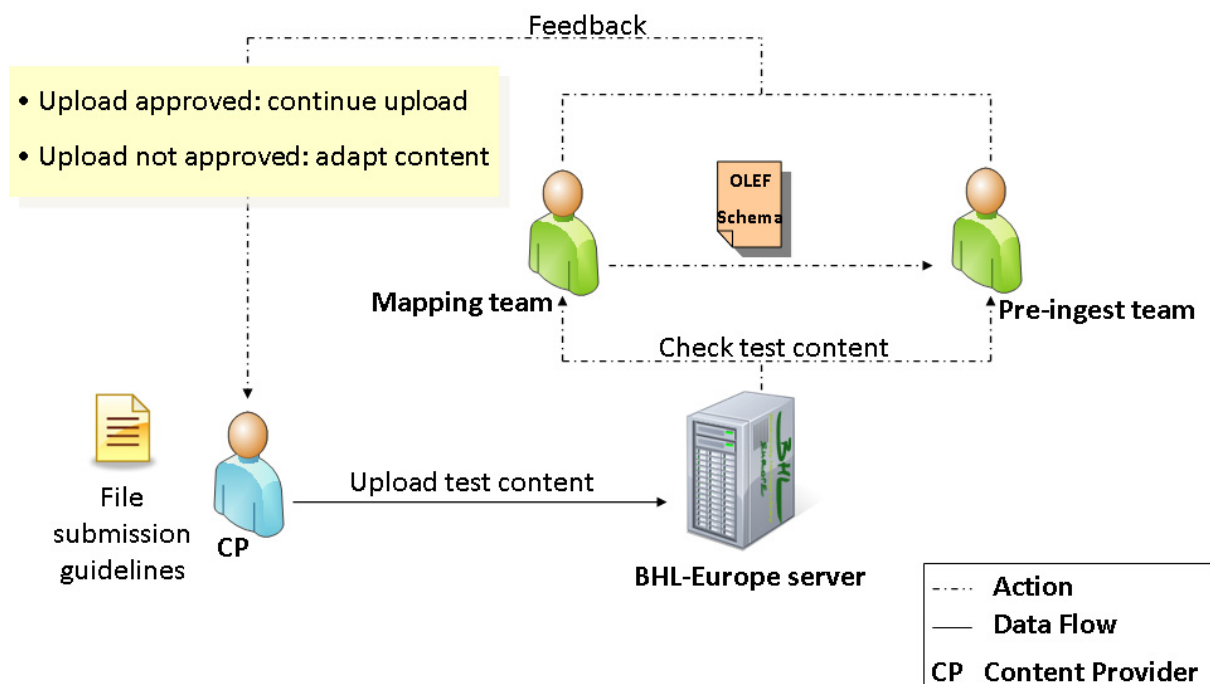


Figure 1: content upload workflow

5.2 Underlying content of BHL-Europe

The following section shows the underlying content of BHL-Europe according to the information given by our content providers within the signed Memorandum of Understanding or the Description of Work. This quantity will also be available at the BHL-Europe Portal.

Quantity according to MoU [pages]			
European content providers	April 2010	April 2011	April 2012
NHM ⁽²⁾	2,382,713	2,782,713	3,182,713
NMP	300	2,000	3,000
UGOE	100,000	100,000	100,000
LANDOE	500,000	600,000	650,000
HNHM	23,000	44,000	50,000
UCPH		50,000	100,000
NAT	86,500	88,500	90,500
NBGB	5,000	35,000	50,000
RMCA	15,000	50,000	50,000
RBINS		30,000	70,000
BnF	150,000	150,000	200,000
MNHN	112,000	172,000	232,000
CSIC ⁽¹⁾	21,000	15,000	14,000
RBGE	25,211	25,211	25,211
UH-Viikki	5,985	12,985	19,985
UBER	12,200	12,200	12,200
UB-Bielefeld	21,800	21,800	21,800
PAS			10,800
Dilibri			4,400
GfBS			3,157
MSN ⁽³⁾			55,216
MIZPAS			2,000
Rennes 1			5,000
UB-Frankfurt ⁽⁴⁾			330,000
MfN			2,000
Total amount (European content providers)	3,460,709	4,191,409	5,283,982

BHL-US content providers			
BHL-US (without NHM)	30,200,335	36,000,000	39,000,000
Total amount (European & US content provider)	33,661,044	40,191,409	44,283,982

Table 7: Underlying content of BHL-Europe

⁽¹⁾ ... According to DoW

⁽²⁾ ... NHM provides content to BHL-Europe & BHL-US. To facilitate process BHL-Europe harvests NHM content from BHL-US.

⁽³⁾ ... MSN provides content to BHL-Europe over UGOE. This content is not included in the stated page numbers for UGOE.

⁽⁴⁾ ... UB-Frankfurt will provide also content from FU-BGBM as the content from both institutions is an output of the same project

More detailed information about the specific partners is given in the following paragraph.

FU-BGBM (Nr. 27, Table 5) is not included in Table 7 because UB-Frankfurt will provide content of FU-BGBM to BHL-Europe. Both institutions are collaborating in the same digitisation project. Thus the outcome will be provided by only one institution; however MoUs have been signed by both institutions.

NHM is a content provider for BHL-Europe as well for BHL-US. To facilitate the process of providing content for NHM and avoiding duplication of work, BHL-Europe harvests NHM content from BHL-US. NHM contributed content has been included in the statistics as it is already available online on BHL-US.

MSN digitised zoological works in collaboration with our partner UGOE. 430 volumes had been delivered to Göttingen, but many journal volumes were not digitised for conservation reasons, and a few titles have been recently digitised by BHL-US partners. In total, 274 volumes (55,216 pages) have been digitised and will be provided by UGOE to BHL-Europe. The 55,216 pages are not included in the stated page numbers for UGOE in Table 7.

Up-to-date information on the current status of content upload of each content provider can be found on the wiki¹². The already uploaded content to the BHL-Europe server is visible using the following URL: <http://bhl-celsus.nhm.ac.uk/uploads/>. For safety reasons the respective logon details for the FTPS connection is requested to look at the uploaded content.

¹² [https://bhl.wikispaces.com/Content+ Management](https://bhl.wikispaces.com/Content+Management)

5.2.1 Interconnected repositories

Interconnection has been shown by WP3 to be technically possible for the following partners by mapping metadata from various BHL-Europe content providers and ingesting the content to the BHL-Europe Portal (Table 8).

Nr. of CP	Content Provider	Status
1	NHM	Interconnection successful
2	NMP	Interconnection successful
3	LANDOE	Interconnection successful
4	HNHM	Testing in process
5	UCPH	Testing in process
6	NAT	Interconnection successful
7	NBGB	Interconnection successful
8	RMCA	Interconnection successful
9	RBINS	Testing in process
10	BnF	Not interconnected
11	MNHN	Not interconnected
12	CSIC	Interconnection successful
13	RBGE	Interconnection successful
14	UH-Viikki	Testing in process
15	UBER	Testing in process
16	UB-Bielefeld	Not interconnected
17	PAS	Interconnection successful
18	Dilibri	Interconnection successful
19	GfBS	Not interconnected
20	MSN	Not interconnected
21	MIZPAS	Testing in process
22	Rennes 1	Interconnection successful
23	UB-Frankfurt	Testing in process
24	FU-BGBM	Testing in process
25	MfN	Interconnection successful
26	UGOE	Not interconnected
27	SIL	Interconnection successful
28	MOBOT	Interconnection successful

Table 8: Status of interconnected repositories.

As of 18 May 2012, we have 14 repositories interconnected and the mechanisms in place to ingest all content remotely. The data and content harmonisation process for all of them is finished and the content is following the File Submission Guidelines. We need to provide some feedback back to the content providers to allow them uploading new content in the same structure and format.

The test ingest for eight other repositories is in process, after data and content were harmonised. Some minor tweaks still have to come before these data can be ingested remotely.

The repositories of six of our content providers was not able to connect to the BHL-Europe system in order to ingest the content. The reasons for that are explained below. The content from UGOE and MSN is coming with a harddrive that has not yet arrived in London to be uploaded to the servers. However, as UGOE is working with a standardised approach (Goobi), we expect the content to be in good format for ingestion. The content from MNHN and UB-Bielefeld needs some more work to bring the content into a format from where we can actually ingest it. This content is still not following the File Submission Guidelines. This is even more serious with the content from BnF and GfBS. First of all we need to do a lot of reverse engineering to understand how to bring the content into the right format before we can actually harmonise the data. We also need to liaise with the content providers to get new metadata or support on how to decode the metadata correctly. The delay with this process is due to the fact that we were late in providing the necessary guidance for content providers on how to prepare the data properly, which is a direct result of the late delivery of the Pre-Ingest tool. As the processes are now in place and functional and the people that have gained the experience in managing the process continue to work for BHL-Europe, we are optimistic to get as much content ingested as possible even after the end of the BHL-Europe project.

It is obvious that the numbers above are lower than in previous reports. These numbers now reflect the interconnection to the Live BHL-Europe System and not to any kind of prototype as it was before.

5.3 Content available through Europeana

BHL-Europe acts as a content aggregator for Europeana and therefore provides the available biodiversity material not only through the BHL-Europe portal, but also through Europeana.

With reference to the new published Data Exchange Agreement from Europeana, BHL-Europe decided to draft an additional Appendix for the BHL-Europe MoU in order to be able to provide the content to Europeana also from 1st of July 2012 on. Content providers have been asked to sign the additional Appendix in order to be able to make the provided content available through Europeana. Content providers who will not sign the additional Appendix will be excluded from the metadata transfer to Europeana. To date of writing it is foreseeable that BnF will not sign the MoU until the official end of the project. This is caused by legal issues particular to the regulatory framework in France. BnF is currently in discussion with the France Ministry of Culture to overcome these issues and is optimistic to sign the Data Exchange Agreement before the Europeana deadline of 1st of July 2012.

Table 9 shows the quantity of content available using the Europeana indicator (items) as well as the BHL-Europe indicator (pages).

According to Table 7, BHL-Europe will provide 44,283,982 pages to Europeana by the end of the project. By the end of the project BHL-Europe has provided 81% of the expected content to Europeana. Currently, all content available in Europeana is still harvested through our prototype system. We now have a new OAI-PMH provider that is exposing ESE already¹³. We are also in close contact with the Europeana ingest team to make sure our content meets the requirement of Europeana. Thus, BHL-Europe has all processes in place to ensure a continuous delivery of content to Europeana. However, we still need to ingest the majority of our content into the new system to then expose the final URL to Europeana for harvesting. As we will continue to ingest content after the end of the project, all content on our servers will eventually be available in Europeana through our new infrastructure.

Content Provider	Europeana indicator [items]	BHL-Europe indicator [pages]
LANDOE	3,511	540,062
NAT	3,516	96,604 ⁽¹⁾
UBER	61	12,200
UBBI	1,634	20,449
UCPH	103	5,577
UH-Viikki	30	11,455
RBGE	183	25,776
RMCA	52	26,668
CSIC	9	2,896
NHM	4,404	2,502,593
BHL-US	87,058	32,633,312
Total amount	100,561	35,865,514

Table 9: BHL-Europe content in Europeana (14.05.2012)

⁽¹⁾ ... Estimated numbers of volumes and corresponding pages given within the MoU have been used as data set for calculation of the mean value of pages per volume for each content provider if number of pages could not be counted. Calculation table is included in the Appendix.

5.4 European cultural heritage distribution

European cultural heritage is distributed all over the world and is not only available within Europe. Biodiversity heritage literature is part of our European cultural heritage and is also scattered worldwide. For example, the Marine Biological Laboratory (MBL), founded 1888, is the oldest private biological laboratory in the United States. 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 at MBL working for European institutions. Their careers often include responses to mid-20th century forces for which working at MBL was often part of a path leading to freedom from the upheaval caused by

¹³ <http://bhl-test.nhm.ac.uk/oai-pmh/?verb=ListRecords&metadataPrefix=ese>

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 extended European cultural heritage outside of Europe. Hence, European researchers have travelled to the MBLWHOI (Marine Biological Laboratory Woods Hole Oceanographic Institution) Library for many years to read and study literature of European origin and also to enrich the libraries during their time in residence. 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-US, 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 easily.

An estimation of the total BHL-US corpus in June 2010 showed that ~ 53 % of all pages available to that date were of European origin (published in Europe).

Continent	Amount of content [%]
Europe	53.2
North America	44
South America	0.2
Australia	0.8
Africa	0.2
Asia	1.3
Oceania	0.3

Table 10: Origin of BHL-US corpus (June 2010)

This high level of European content within BHL-US shows the importance of this content for BHL-Europe: in making BHL-US data and content interoperable with European content providers and managing the content acquisition process over all partners in cooperation with BHL-US, our European content providers avoid duplicating scanning efforts. Thus, collaborating with BHL-US also has the important economic consequence of saving tax money for Europeans. This aspect of the partnership should not be underestimated. At the time of writing there are 34 million pages available from BHL-US. Among these pages it is estimated that 18.1 million pages are of European origin. BHL-Europe participates also in global events of global BHL which is a federation of BHL nodes (Australia, China, Brazil, US, Europe) in the whole world. This worldwide collaboration allows on the one hand to repatriate European cultural heritage and on the other hand to make European cultural heritage freely available all over the world.

5.5 Content availability in Europe

This section aims to give an estimate of available natural history content within BHL-Europe. The numbers presented are an extrapolation based on published content in Austria. Our consortium partner LANDOE has calculated the amount of natural history content already published and available for digitisation planning in Austria.

Table 11 shows the calculated numbers of available biodiversity content in Austria. In Austria published natural history content ranges from 2.2 to 2.7 million pages. These estimated numbers and the population figures of Austria, Germany and Europe are used in the following calculation to estimate the published natural history content within Europe.

Type of material	Available content [pages]
Digitised Serials	500,000
Serials published, not digitised yet	700,000
Monographs published, not digitised yet ⁽¹⁾	1,000,000 - 1,500,000
Total amount	2,200,000 - 2,700,000

Table 11: Estimation of available biodiversity content in Austria

⁽¹⁾ ... based on numbers from NHMW. Lower estimate of published biodiversity monographs is 3,000; upper estimate of monographs is 5,000. Assumption: mean value of pages for a monograph is 300 pages.

Nowadays, Austria has ~ 8 million inhabitants and published 2.2 - 2.7 million pages of natural history content. Applying this ratio to Germany, which has ~80 million inhabitants, the amount of published natural history content should be ten times higher and thus ranging from 22 - 27 million pages published in Germany. The European Union has a population of ~500 million which is ~5.5 times the population of Austria and Germany together. Therefore it is estimated that natural history content published within the European Union ranges from 133,100,000 - 166,335,000 pages.

Region	Population [million]	Amount of pages
Austria	~8	2,200,000 - 2,700,000
Germany	~80	22,000,000 - 27,000,000
Austria & Germany	~90	24,200,00 - 29,700,000
European Union	~500	133,100,00 - 166,335,000

Table 12: Estimation of available natural history content available within the European Union

Further estimates made by our consortium partner LANDOE show that within the domain of earth sciences there is additionally a minimum of one million pages and within the domain of agriculture a minimum of 1.9 million pages available for digitisation for serials within the current borders of Austria.

These results indicate that a huge amount of biodiversity content is available and only a small share of this content can be made available within the BHL-Europe project time. Most of this content exists only as print versions and still needs to be digitised.

5.6 Lessons learned

In order to facilitate the work for future projects, lessons learned from the BHL-Europe project are described with regards to content management work package.

For future projects it is recommended to choose the performance measurement more independently for each work package. Key performance indicators should not be highly dependent on the outcomes of several work packages, as it was within BHL-Europe. In this context it would be wise to avoid too many dependencies between work packages, which might result into slowing down the possible work within other work packages or even bring the work of a team to a halt. A good understanding of what is important within the project and also understanding crucial milestones is necessary from the beginning across all involved partners.

Milestones should be set from the beginning in compliance with milestones of all work packages already with the description of work. For projects with content provision, it is advised to plan the final deadline for content provision at least 2 months before the official end of the project already in the description of work in order to allow enough time to process the content within the project time frame.

Appendix

A: Calculation table for estimation of number of pages within Europeana

European Content Providers	Quantity according to MoU [pages]			Quantity [volumes]			Mean value [pages per volume]	Items in Europeana	Pages [BHL-Europe indicator]
	April 2010	April 2011	April 2012	April 2010	April 2011	April 2012			
NAT	86,500	88,500	90,500	3,153	3,220	3,290	27	3,516	96,604

B: BHL-Europe in a nutshell

BHL-Europe in a nutshell

BHL-Europe vision: Build a Digital Open Access Library for Biodiversity Literature.

- **What is BHL-Europe?**

The Biodiversity Heritage Library for Europe (BHL-Europe) is an ambitious and innovative European Commission funded program, started on 1 May 2009. BHL-Europe aims towards a digital global library of life by bringing together existing digital collections of biodiversity literature from all over Europe into one freely accessible online portal.

- **Why do we need BHL-Europe?**

The libraries of the European natural history museums and botanical gardens collectively hold the majority of the world's published knowledge on the discovery and subsequent description of biological diversity. As yet this wealth of knowledge is only currently available to those few people who can gain direct access to these collections. The body of biodiversity knowledge is thus effectively withheld from use for a multiplicity of potential users.

Much of the early published literature is rare or has limited global distribution and is available in only a very few libraries. From a research perspective, these collections are of exceptional value because the domain of systematic biology depends – more than any other natural science – upon historic literature. Once the collections of biodiversity literature are freely available on the Internet, this will be of great value to scientists, and also to a much wider public.

- **What is the focus of BHL-Europe?**

BHL-Europe focuses on the interoperability of existing European digital libraries and repositories with the goal of providing open access to the general public and scientists via one unique Web portal and Europeana. There is no single natural history museum or botanical garden library which holds the complete corpus of legacy literature. Therefore, BHL-Europe needs to be a federation of numerous institutions in which your institution can also be a part.



▪ **What are the benefits of becoming a partner of BHL-Europe?**

- BHL-Europe increases the visibility of your content, thus increases the visibility of your institution.
- BHL-Europe establishes a sustainable preservation and archive system to store, curate, manage and migrate your data. Thus BHL-Europe will have strategies and processes in place for long-term preservation of the data produced by biodiversity digitisation programmes.
- BHL-Europe provides multilingual access to your content, thus saving you the costs for creating a multilingual portal.
- BHL-Europe enriches your metadata so that every content provider can retrieve and reuse the enhanced data.
- BHL-Europe provides you access to an important up-to-date and cost-effective pool of information and network of collaboration to help you use best practice approaches.
- BHL-Europe provides tools and technologies to cost-effectively present the digital content and manage digitisation projects in the biodiversity domain.
- BHL-Europe helps you to create good quality OCR text from the scanned page images for further data analysis and metadata enrichment.
- BHL-Europe provides you access to Taxonomic Intelligence tools to facilitate the search for taxon specific information.
- BHL-Europe dissemination activities will reach a large group of target users to increase the use of your data and content. Thus, BHL-Europe gives a cost-effective way to multiply the impact of your editorial activities and build a presence with users, the professional community, and other organisations including national governments.
- BHL-Europe makes all the content available through Europeana – a search platform to a collection of European digital libraries. This will increase the visibility of BHL-Europe content, and it will also increase the visibility of your data. As Europeana enriches your metadata, this new metadata can be retrieved through BHL-Europe too. BHL-Europe benefits from the network and knowledge building capacity of Europeana, thus increasing the access to state-of-the-art research and technologies. A sustainable Europeana will also support the sustainability of BHL-Europe.



Intellectual Property Rights

The main aim of the BHL-Europe project is to make biodiversity knowledge accessible on an open access basis to a wide spectrum of end-users. The value the project delivers is not merely in making works available online, but in the open access terms under which the works are available. Therefore understanding Intellectual Property considerations is fundamental in achieving this objective.

The purpose of this summary is to outline Intellectual Property Rights (IPR) factors that need to be considered in relation to the scanning and online display of library biodiversity collections of BHL-Europe project partners. Scanning copyright works and making them publicly available without due consideration of rights ownership and clearance presents serious risks.

Adherence to BHL-Europe's Intellectual Property principles and requirements is essential to the success and sustainability of the project. All BHL-Europe partners must therefore understand the following:

- BHL-Europe partners are responsible for providing content and data to the project that does not infringe the intellectual property rights of third parties.
- The activities of BHL-Europe cross many legal jurisdictions in terms of the partners involved, origin of material to be scanned and applicable IPR legislation. It is therefore necessary for project partners to know the relevant legislation of their own country.
- The project shares its data (and some content) with other projects, including BHL, EOL and Europeana. It is essential that project partners ensure that the works and data they make available to BHL-Europe are also licensed for sharing with each other and these other projects.
- BHL-Europe partners must ensure that digital content provided to the project are licensed under Creative Commons open access terms as stipulated in the project's Description of Work to the European Commission.
- Neither the project BHL-Europe nor the data/content providers will seek to assert any intellectual property rights over digital copies of public domain original work. Original public domain material can be reused or exploited by anyone who wishes to use it including educational, non-commercial, and commercial users.



- The use of technical protection measures such as visible digital watermarking or copy prevention are not compatible with the Open Access principles of BHL-Europe. Such enforcement technologies cannot respect or acknowledge the subtle and subjective concepts of fair dealing or fair use or other exceptions to the monopoly rights copyright confers.
- Neither data nor content provided to BHL-Europe may be restricted by subscription fees or other access controls that require payment.

New BHL-Europe partners will be required to sign a memorandum of understanding which further details the terms upon which content and data are made available to the project. BHL-Europe will provide project partners with an IPR best practice guide that includes further information about IPR risk management, rights clearance, due diligence and Creative Commons licensing.

