

ECP-2008-DILI-518001

BHL-Europe

Progress Report

1 May 2009 – 31 October 2009

Deliverable number/name	<i>D1.1</i>
Dissemination level	<i>Confidential</i>
Delivery date	<i>7 November 2009</i>
Status	<i>Final</i>
Author	<i>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
Kai Stalman	MfN
Adrian Smales	NHM
Jiri Kvacek	NMP
Alexander Herzog	AIT
Gerda Koch	AIT
Walter Koch	AIT
Francisco Welter-Schultes	UGOE
Wolfgang Koller	NHMW
Michael Malicky	LANDOE
László Peregovits	HNHM
Tine Mallaerts	RBINS
Doriane Ibarra	MNHN
Kimmo Koskinen	UH-Viikki

0.2 Revision History

Revision Date	Author	Version	Change Reference & Summary
27 Oct. 2009	Henning Scholz	0.1	Initial draft version for external review
28 Oct. 2009	Martin Gordon	0.2	Review of 0.1
28 Oct. 2009	Henning Scholz	0.2a	Draft for external review
29 Oct. 2009	Martin Gordon	0.3	Review of 0.2a
29 Oct. 2009	Henning Scholz	0.4	Draft for internal review
2 Nov. 2009	Henning Scholz	0.5	Draft for external review
2 Nov. 2009	Martin Gordon	0.6	Review of 0.5
5 Nov. 2009	Henning Scholz	0.7	Draft after completion of internal review
7 Nov. 2009	Henning Scholz	1.0	Final version

0.3 Reviewers

This document requires the following reviews and approvals.

Name	Position	Date	Version
Martin Gordon	Project Manager, RBB	28 October 2009	0.1
Martin Gordon	Project Manager, RBB	29 October 2009	0.2a
Martin Gordon	Project Manager, RBB	2 November 2009	0.5
Christoph Häuser	Head of Project Coordination, International Cooperation and Science Policy, MfN	3 November 2009	0.4
BHL-Europe consortium		4 November 2009	0.4

0.4 Distribution

This document has been distributed to:

Group	Date of issue	Version
External reviewer	28 October 2009	0.1
External reviewer	28 October 2009	0.2a
Reviewers and BHL-Europe consortium	29 October 2009	0.4
External reviewer	2 November 2009	0.5
BHL-Europe consortium	10 November 2009	1.0

1 Table of Contents

0	DOCUMENT HISTORY	2
0.1	CONTRIBUTORS.....	2
0.2	REVISION HISTORY	2
0.3	REVIEWERS AND APPROVALS	3
0.4	DISTRIBUTION.....	3
1	TABLE OF CONTENTS	4
2	SUMMARY.....	5
3	STATUS	6
3.1	RESOURCES EMPLOYED	21
3.2	WORK PACKAGE OVERVIEW	22
3.3	DELIVERABLES STATUS	34
3.4	PERFORMANCE INDICATORS	35
4	AWARENESS AND DISSEMINATION	36
4.1	OVERVIEW OF AWARENESS AND DISSEMINATION ACTIVITIES.....	36
4.2	EVENTS AND MEETINGS	36
5	CONCLUSIONS.....	38

2 Summary

The objectives of the reporting period (May to October 2009) of the project were:

- (1) Establish a Technology Management Board and a Communications Working Group
- (2) Review existing content analysis tools and bibliographic services to identify algorithms for deduplication of data and processes for the establishment of a virtual taxonomic library
- (3) Evaluation and review approaches for the establishment and management of multilingual biodiversity digital libraries
- (4) Review state-of-the-art technologies used by EUROPEANA and BHL for processing digital content
- (5) Initial design of a multilingual access point for the search and retrieval of biodiversity content through at least two portals (EUROPEANA and BHL)
- (6) Define and agree on the requirements of the content provider for page images and metadata, in particular (standards and specifications)
- (7) Define and agree the technical and functional architecture of the BHL-Europe system
- (8) Develop operational strategies and processes for long-term preservation and sustainability of the data produced by national biodiversity digitisation programmes
- (9) Develop the BHL-Europe dissemination plan and promotion material
- (10) Raise awareness and ensure that the project outputs are known and used by the target users and that the proposed approach directly addresses user needs
- (11) Negotiate with Rights Holders, publishers to enable access to in-copyright content
- (12) Extension of the network of content providers

Both the Technology Management Board and the Communications Working Group were established during the Kick-off Meeting of BHL-Europe in May 2009. This meeting was also used to start the review process of existing technologies for the establishment of a multilingual biodiversity digital library. This technology review process led towards D3.2 and D3.3 (which will be available at the end of this reporting period). The content provider requirements are analysed and summarised in D2.1. A metadata prototype portal was designed to provide a multilingual access point for search of content and as a test of the harmonisation of metadata of our partners to ensure interoperability.

BHL-Europe is in discussion with the virtual taxonomic library developed in the EDIT programme (ViTaL) and with a virtual library of biology literature (Vifabio) developed in Germany to integrate existing bibliographic databases. A first evaluation of deduplication algorithms was also performed and a testbed was designed to compare deduplication procedures.

In order to guarantee sustainable data access, a large data centre was designed that is currently undergoing validation by our partner NHM. BHL-Europe was very successful in presenting the project at various national and international conferences and congresses to collect feedback for our project development process. We are also in discussion with a number of potential content providers (rights holders as well as digital libraries) in order to aggregate

their content to BHL-Europe. A user requirement survey was initiated at the end of the reporting period, to define the functionalities and services required by our target users.

3 Status

Library questionnaire: The first activity of the project was the distribution of library questionnaires to content providers. This was in preparation for the first round of technology review during the Kick-off meeting of BHL-Europe from 10 to 15 May 2009. The completed questionnaires were evaluated during the first months of the project to be the basis for the configuration of the pre-ingest procedures. It was also the basis for the wording of the Memorandum of Understanding. Thirteen of the BHL-Europe content providers completed the questionnaire. The questions to be answered were as follows:

- (1) What is the number of bibliographic records in your library catalog?
- (2) How many physical volumes (serials and monographs) does your library hold?
- (3) Are they barcoded with links to your catalog? What percentage?
- (4) How many monographic titles does your library hold? How many serial titles?
- (5) Do you have detailed serial volume/item-level information in your catalog? How complete is it?
- (6) What classification/call number scheme does your library use?
- (7) What type of subject headings, e.g. LCSH, does your library use in its catalog?
- (8) Are your bibliographic records also held in a bibliographic utility, e.g. OCLC, RLIN? What is your institution's code?
- (9) Do you have an online library catalogue, if so what is the URL?
- (10) What library system do you use (supplier, product and version; e.g. Voyager, ExLibris, locally developed)?
- (11) Does your system support Z39.50?
- (12) Do you see any problems in providing BHL the original page scans for storage in BHL data centres and further processing in the BHL workflow (image enhancement, OCR) or are you in favor of a decentralised/distributed system?
- (13) Are exchanging your data with others, is your system interoperable with other services like Europeana? If yes: What data are you sharing and who is the other service/project you are working with?

Grant agreement: The grant agreement was received on 30 April 2009 from the EC by the PCO to be distributed to the partners. A-forms were sent out and collected during the Kick-off Meeting and the following days. The signed A-forms were sent back to the EC by 10 June 2009. The grant agreement was signed by the EC and sent to the PCO in early June 2009. After completion of the grant negotiation, the first pre-installment was received and distributed to the partners in July 2009.

Kick-off Meeting: The Kick-off Meeting of BHL-Europe including two initial workshops took place in Berlin (MfN) from 10 to 15 May 2009. It started with an initial PMG meeting followed by a welcome event in the main exhibition hall of MfN. On 11 May the framework for the project was discussed, and the composition of the Technology Management Board

(TMB) was decided (see below). 12 May was mainly reserved for presentations by external projects that are important for BHL-Europe. The BHL-Europe Project Officer from the European Commission (Mr. Marcel Watelet) was in Berlin on 13 May to introduce important aspects of project governance, administration and dissemination. On 12 and 13 May all content providers had the opportunity to present their projects in order to start the process of identifying the content holder requirements. The first workshop of the technology experts was held on 14 and 15 May to identify roles and tasks for WP2 and WP3. The Communications Working Group (CWG) was formed and had its first workshop on 14 May to discuss designs and dissemination activities.

Technology Management Board: The Technology Management Board (TMB) was established with members from a number of institutions and companies within BHL-Europe, currently comprising:

Kai Stalman	MfN, Leader WP2
Adrian Smales	NHM, Leader WP3, Chair TMB
Bernard Scaife	NHM
Walter Koch	AIT
Christopher Freeland	MOBOT, BHL
Julie Verleyen	EDL Foundation, Europeana
Anton Guentsch	BGBM, EDIT WP5

The board will find a suitable replacement for Bernard Scaife, as he left the NHM and the BHL-Europe project as of 14 August 2009. The individual who replaces Scaife at NHM is foreseen as his replacement in the TMB following the successful recruitment process at NHM.

The initial meeting of the TMB was conducted on 14 May 2009. The main aims and goals of the project were discussed as well as detailing the deliverables for the next 12 months; owners and contributors were assigned for each of the deliverables within WP2 and WP3. The minutes for the meeting can be viewed at: https://bhl.wikispaces.com/file/view/BHL-E_TechWorkshop_all.pdf

Meeting in Paris with Atos/AIT (8 June) and BnF/MNHN (9 June): Several questions concerning tasks and roles in WP3 and the TMB remained unclear after the Kick-off Meeting. Therefore, the meeting in Paris was necessary to identify the tasks of the key personnel involved. The meeting was attended by Henning Scholz (MfN), Adrian Smales (NHM), Walter Koch (AIT), Roger Essoh (Atos) and Lee Namba (Atos). The minutes and action items of that meeting are available on the BHLwiki:

https://bhl.wikispaces.com/file/view/minutes_bhl-e_atos_paris_090608.pdf.

The key decision of that meeting was to establish a pre-ingest process with tools developed by AIT based on the existing tools of the DISMARC project. Five digital libraries were identified (BHL, BnF, UGOE, LANDOE, NAT) to submit test data (metadata) for the adaptation of the tools. It was also decided to create an Architectural Design Document (ADD) authored by Lee Namba (Atos). To follow up with the developments leading to the Duracloud System, it was decided that Adrian Smales (NHM, WP3 leader), Roger Essoh (Atos) and Walter Koch (AIT) should attend the SUN PASIG conference in Malta in order to discuss with Sandy Payette and others.

The next day, Henning Scholz (MfN) and Adrian Smales (NHM) met with colleagues from BnF (Hervé Colinmaire, Eva Fuentes, Frédéric Martin, Elisabeth Freyre, Danielle Maricar, Thomas Ledoux, Emmanuelle Bermes) and MNHN (Francoise Romagne, Denis Silvestre) in series of three meetings. The first meeting was to discuss recent development in BnF leading to the development of the Gallica portal, mainly from the librarian point of view. The second meeting was to discuss tasks and responsibilities of all the staff members of BnF and MNHN involved in BHL-Europe. The third meeting was to obtain detailed insights into the development of the preservation and archive system of the BnF (SPAR project). The experiences of BnF and lessons learned are important for BHL-Europe mainly for the use of Fedora in the large scale implementation of preservation systems. These experiences are considered for the development of the BHL-Europe architecture.

BHL-Europe general and workgroup meeting: From 10-12 August we had our 2nd large project meeting in the Nationaal Natuurhistorisch Museum Naturalis, Leiden, The Netherlands. 33 members from 21 partners of the BHL-Europe consortium came together to review the first months of the project and establish a workplan for the following three months. In the technology workshop we reviewed existing tools and approaches and evaluate them for use in BHL-Europe. The cooperation with Europeana in particular but also with other projects like EDIT on a technical based was discussed. The Communication Working Group discussed the future directions for the BHL-Europe Newsletter and discussed the dissemination plan in detail.

BHL-Europe communication: During the meeting in Leiden we agreed on our communication rules in the project. We keep initial discussions on outcomes and draft deliverables within the WP or workgroups. Google Groups may be helpful to organise communication in such workgroups more effectively. Actually, a BHL-Europe tech group was established by end of October to cover all IT experts. The BHLwiki is used for votes or discussions of the full consortium. Each WP has its own section on the BHLwiki that is maintained by the WP leader. The Project Server facilitates the communication of the Workplan and related project management aspects. Email communication is not used for extensive discussions of many people involved. All Email communication should have the Project Office in Cc (bhl-europe@mfn-berlin.de) to facilitate filing of Email threads. We have established various internal mailing lists to address the appropriate group of the consortium. The PMG has weekly conference calls to discuss urgent issues. The minutes of these calls, if not confidential, will be uploaded to the BHLwiki. The PCO will write quarterly status reports for the consortium in addition to the regular progress reports.

Test of mapping, ingest, and post-processing of data and content: At <http://bhl.ait.co.at> a DISMARC port was created for BHL-Europe with a first draft of an appropriate metadata set with the same base functionality. Metadata from various BHL-Europe content provider institutions was mapped against the MODS schema and inserted into the demo server. To date the test server contains 41,832 records from 14 institutions. The mapping of this data was carried out using tools already developed during the DISMARC-project or using ETL-tools (extract-transform-load) available in the open source domain (Pentaho). The same technologies will be used to map this data into an ESE (Europeana Semantic Elements) compliant schema.

To try out available Ingest procedures, the input of scanned books accompanied by bibliographic metadata was tested, using the infrastructure of Internet Archive (IA:

<http://www.archive.org>). This was done to gain better understanding of the relationship between Ingest and Access components. BHL-Europe partner CSIC has provided scanned pages (TIFF), the metadata record has been retrieved using the Z39.50 facility of the Exlibris library system available at CSIC premises. The book was uploaded in the IA biodiversity collection to generate the derivatives (JPG, thumbnails, OCR) for access (<http://www.archive.org/details/contritomorpha00marcus>; see also Fig. 1). The uploaded and generated files provide input to the BHL Portal which was done afterwards. A second book was uploaded to IA to test the processing of Fraktur type face, which was popular in German-speaking areas until 1941 (<http://www.archive.org/details/bilderatlasmw4>). An instruction document on how-to upload content to the Internet Archive was provided.

Based on the experiences and documentation of AIT, the procedure of uploading files to IA to generate derivatives for the BHL Portal is now established further for use by partners. Partners doing digitisation can use this AIT-designed technology to enhance and process page image files at IA, and can pull the new files back into BHL for display through BHL Portal and Europeana. This helps to use existing technologies for post-processing of content to deliver the page images (i.e. the actual digital content) in the format agreed by BHL-Europe project partners.

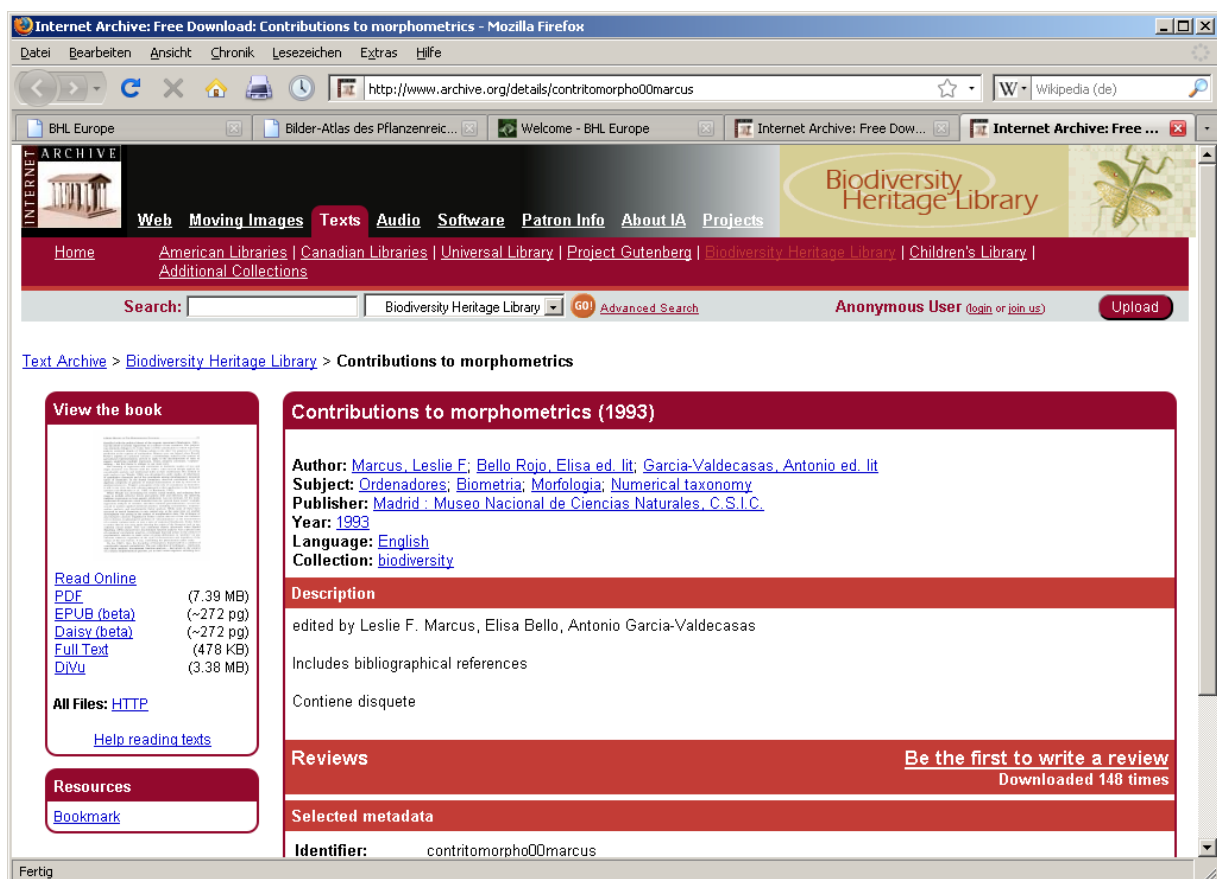


Fig. 1. Screenshot of the Internet Archive portal to illustrate the upload of books.

Metadata generation from library catalogues: The technologies provided above are suitable for harmonising of metadata and mapping to the data standards used in BHL-Europe and Europeana. The mapping of internal library catalogue data to XML/MARC21 standards or other accepted metadata standards is not covered by the tools mentioned above. This makes the use of library catalogues difficult or even impossible to build a bibliographic database system and metadata repository using the input of partners without a working metadata database. Therefore, a Schema-Mapping Tool is created by NHMW to currently map the NHMW library-catalogue data, which are in CDS-ISIS Format to an XML / MARC21 standard, in order to provide a more common data base (Fig. 2). The Tool uses existing libraries like MARC4J and the Java XML-API in order to process the files. At the moment the tool is already capable of:

- Creating the mapping
- Process the mapping (read content, map content, write content)
- Read schema-definitions from XML, XSD, FDTs and the default MARC21 specification
- Process data-files in MARC21 and MARCXML format

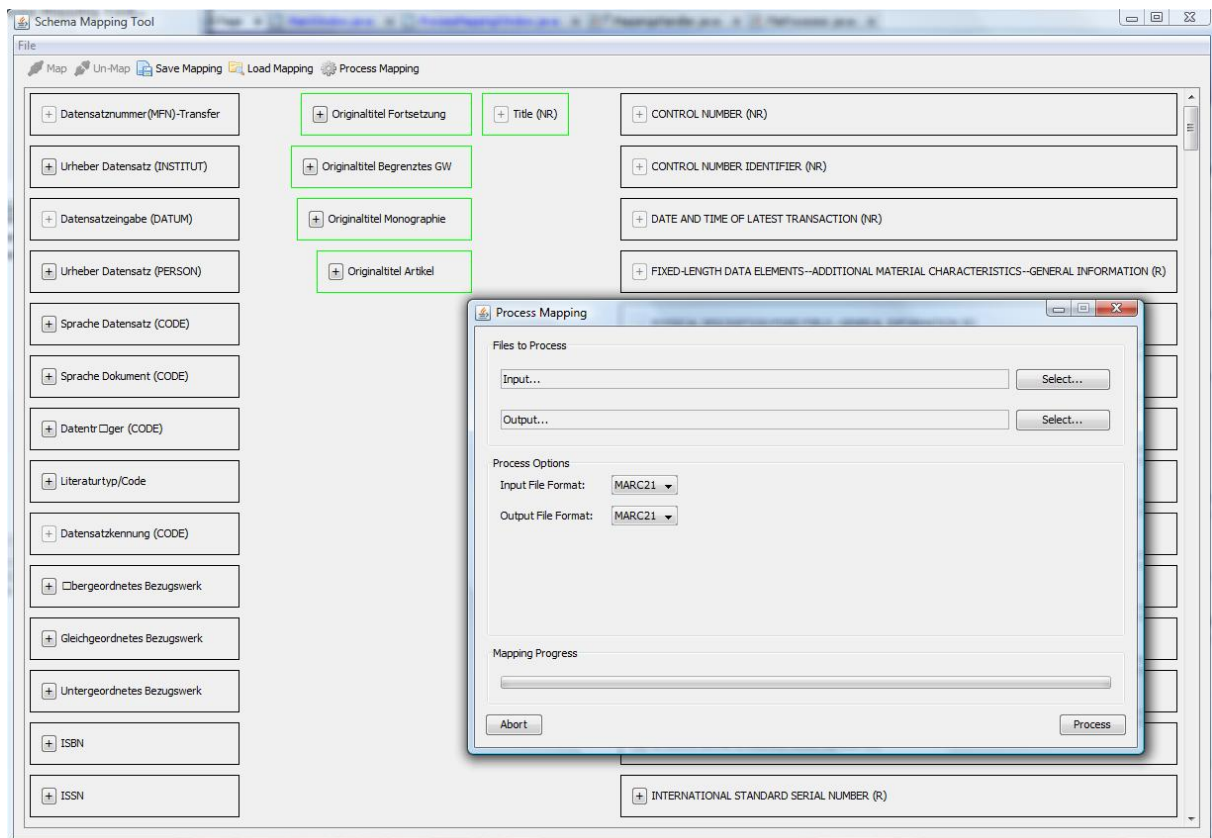


Fig. 2. Screenshot of the Schema-Mapping Tool created by NHMW.

The Schema-Mapping Tool is a Java-Based Application (Java-Applet). The intention of the tool is to provide an application which allows both running the mapping on and offline. The tool is designed to be also used by librarians (the tool offers a simple click and connect interface).

In addition, as it is Java based, parts of the Schema-Mapping Tool (in particular the mapping-process and handling) may be used in the Java-Application Server which will be used for the BHL-Europe Portal (the mapping components of the tool are designed to work even outside an JApplet environment).

The tool itself is designed in a modular manner, so that various data-formats will and can be supported. Interface specifications for classes that can handle different data-formats are already in place. As soon as the data formats of the individual BHL-Europe data providers are available, they can be integrated into the tool using simple classes for reading / writing from / to those formats.

A Google code repository is in place for the schema-mapping tool, which can be found here: <http://code.google.com/p/schema-mapping-tool/>

Deduplication of content: Although not in the scope of OAIIS (see below), it is a very strong requirement to deduplicate catalogue entries and content databases. This is to structure the digital content already existing in the repositories of the BHL-Europe content provider. This is also to avoid duplication of scanning to save resources and to use resources more effectively (by scanning items not already digitised). A first step towards the deduplication of the database is a specific metadata schema. For that purpose, the BHL Deduplication Subset (BDS) was created in the last months based on Dublin Core and enriched by a number of unique fields for the purpose of deduplication.

Available deduplication algorithms were tested, compared, and modified to meet the needs of BHL-Europe. Currently, the BDS is applied to the BHL metadata and the algorithms are tuned to identify duplicates. The current test phase will lead to the development of the deduplication tools needed for the project. A prototype is expected for M9 of the project.

BHL-Europe Architecture: An OAIIS compliant system was set up at <http://bhl.ait.co.at/dev> which uses DISMARC source for PreIngest, Ingest and Access components (Fig. 3). As the DISMARC DataManagement equivalent is not suitable for BHL needs, the component will be replaced by a new component as first implementation.

As overall architectural basis, the OAIIS Standard was chosen to be used for the BHL-Europe system. As the OAIIS components do not describe data augmentation with vocabulary services, data harmonization and the general digitising processes, two additional components are described which wrap the OAIIS System with PreIngest and a Portal.

For each of these components, the best suitable available system was chosen for further integration. For PreIngest and the Portal, the DISMARC system provides multilingual search functionality and data harmonisation tools, whereas ArchivalStorage will be based on Fedora. There are preservation concepts available in Fedora but the search functionality is very limited. All these components are described in depth in D3.2.

As mentioned, the most suitable, proven technology will be used as basis for each component. As these technologies often do not have an open standardized interface, the technology will be wrapped and standardized interfaces will be created in order to supply modular component deployment. Each component is meant to be replaced by better technology as soon as it is available which renders the overall BHL-Europe system stable and sustainable.

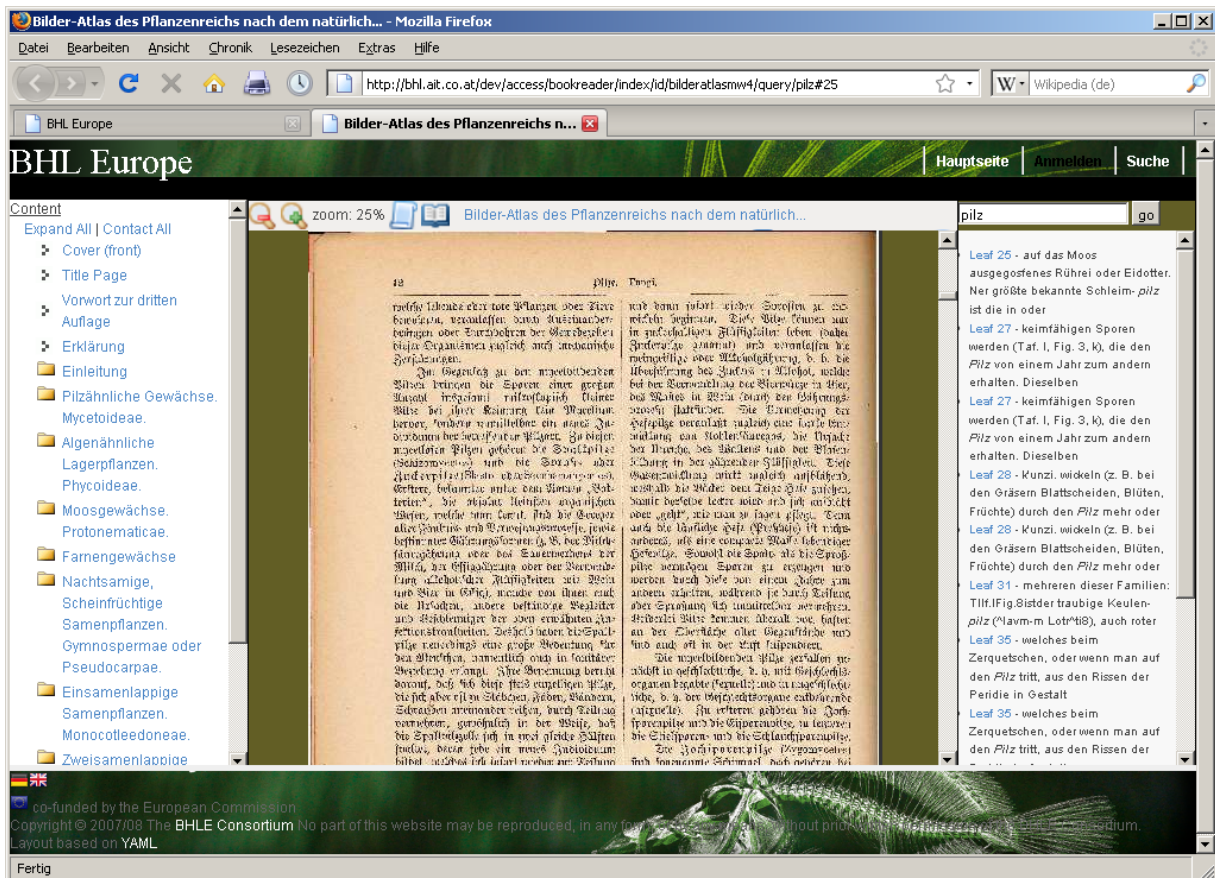


Fig. 3. Screenshot of the BHL-Europe development portal. This also illustrates the integration of the book viewer used by Internet Archive. This book viewer will be adapted according to the requirements and feedback of the BHL-Europe consortium partners.

The overall BHL-Europe system architecture was further elaborated based on the discussions in Berlin, Paris, Malta, St. Louis, San Francisco, and Copenhagen. As tool for repository management FedoraCommons was selected. Emphasis is put on three components: pre-ingest (metadata harmonisation and creation, semantic enrichment using ontology WebServices), storage (integrating cloud technology) and access (providing multilinguality and ontology support for query expansion).

The development work of WP3 is supported by the open source product agile42 which provides web based access to different functionalities used in the SCRUM approach. The necessary IT infrastructure was set up and tested by AIT. The document structure of the ADD (architectural design document) was delivered. The structure of the document follows ESA standards:

http://www.esa.int/TEC/Software_engineering_and_standardisation/TECBUCUXBQE_0.html.

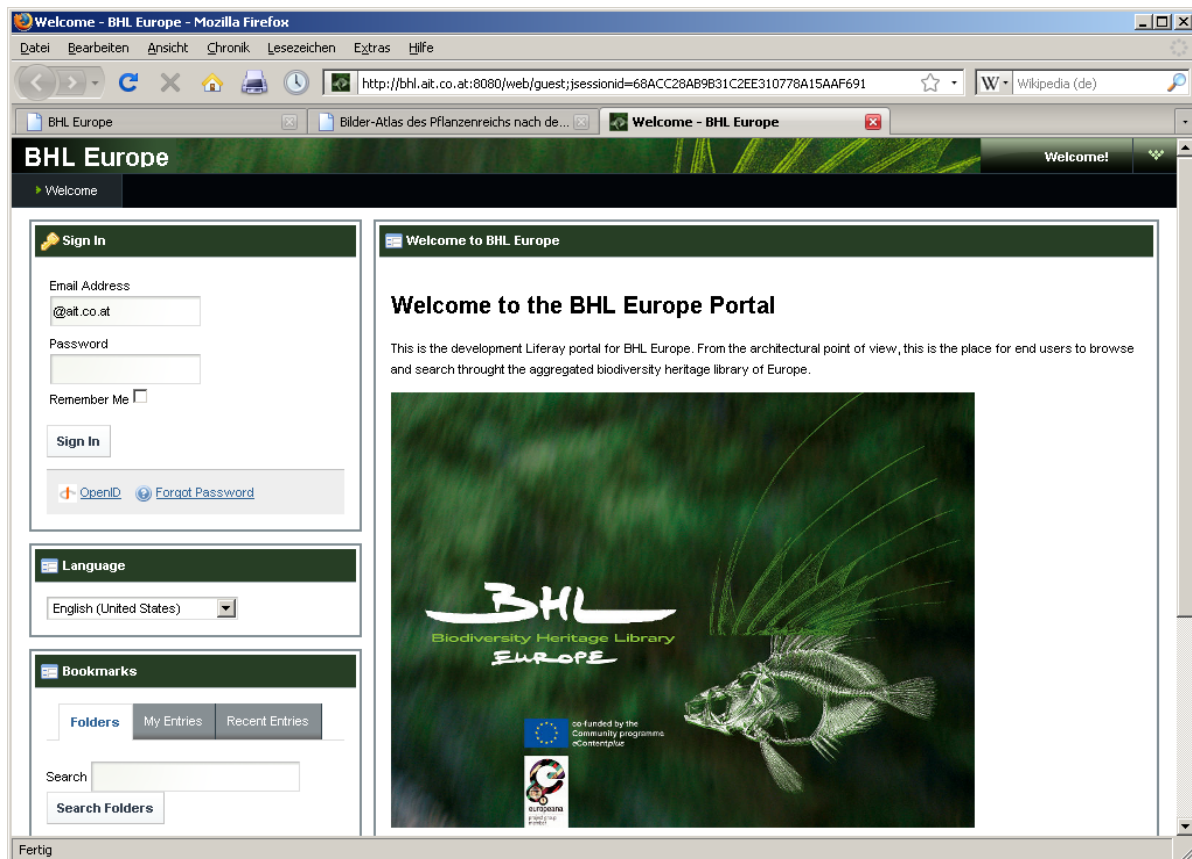


Fig. 4. Screenshot of the draft BHL-Europe Portal component implemented as Liferay Java Portlet Container.

During the eSciDoc-meeting in Karlsruhe the performance issue of Fedora was discussed and the use of the eSciDoc system (eventually for publishing) could be one option for the post processing of BHL-Europe data. The representative of Max Planck Society (MPG) was informed about BHL-Europe, since MPG also supports related projects in the field of natural sciences.

The Portal component for BHL-Europe will be implemented as Liferay Java Portlet Container which provides a solid Framework for user management and multilingual portal interfaces. A first draft of the template was set up at <http://bhl.ait.co.at:8080> (Fig. 4).

The current set of Web sites and test portals are set up to get in depth understanding of the component's requirements and architectural impact.

Multilinguality: AIT has extensive experience in creating multilingual portals - the DISMARC system is fully multilingual (30 languages implemented). In addition, BHL-Europe is investigating progress in other projects to use state-of-the-art technologies in this field. BHL-Europe WP3 leader attended the TrebleCLEF Summer School on Multilingual Information Access in Pisa (15-19 June 2009). The focus of the school was on "How to build effective multilingual information retrieval systems and how to evaluate them". Furthermore, we are looking at the progress of the Europeana Connect WP2 on "Multilingual Access to Content". The leader of that WP is based at the Berlin School of Library and Information Sciences of the Humboldt-University Berlin. Members of the BHL-Europe team are also working for the same institution to also facilitate the alignment of BHL-Europe and Europeana Connect.

DuraCloud: BHL/BHL-Europe was selected as a candidate for the pilot phase of the implementation of the DuraCloud system. The DuraCloud meeting took place at the New York public library on 14 July 2009 and attended by Adrian Smales (NHM, WP3 leader). Within the workshop details were discussed on (amongst other things) how BHL and BHL-Europe would replicate information between their repositories. BHL will move data into a cloud storage environment that incorporates DuraCloud (<http://duraspace.org/duracloud.html>). Once BHL is integrated into the DuraCloud (Uploaded/ingested data) it will be possible to synchronise BHL and BHL-Europe data. The work required to enable the synchronisation by BHL-Europe is to create a connector that will allow the BHL-Europe repository to “see” the same cloud space as BHL. The Akubra connector (<http://www.fedora-commons.org/confluence/display/AKUBRA/Akubra+Project>) will provide the bridge between the preservation repository (Fedora Commons in Europe) and the BHL storage area (Cloud S3 environment in America).

BHL-Europe data centre: One component of BHL-Europe is the building of a large data centre to ensure long term preservation and curation of the digital content provided by the partners. The Wroughton Data Centre (AKA Darwin Repository) will provide the long term sustainability for the BHL-Europe project and provide a European/Global facility for data access and preservation. It was designed during the reporting period, but the decision regarding the building of the centre is still pending (expected to be available before the end of 2009). Although this task was not explicitly foreseen for this period, the basic conditions for setting this up at this early stage of the project were considered as very advantageous to successfully establish the centre.

The Wroughton Data Centre will provide 4 streams of revenue that can be exploited:

1. Storage for printed/physical media and objects
2. Scanning and OCR
3. Data Centre Services (storage, access, preservation etc)
4. Business Services (Disaster recovery, Business Continuity and Vendor Hospitality)

Each of the 4 areas will be investigated to determine the local (i.e. the UK) market demand and requirement for each. At this stage it is anticipated that the demand will be sufficient to provide the incentive to move forward and build the facility.

Currently, the initial investigations have provoked significant interest from various tier 1 vendors, including Cisco, IBM, EMC, Dell and others. The estimated projection for completion of the facility will be 24 months from the commencement of the build phase. The concept visualisation can be found here: <http://www.web3d.co.uk/wroughtondatacentre/>.

The Wroughton Data Centre will be complemented by additional local or regional data centres and access nodes.

Optical Character Recognition: Full text search and the BHL taxonomic intelligence depends very much upon high-quality OCR. However, 35% of the scientific names found by OCR are still inaccurate. Therefore, BHL-Europe is actively seeking ways of improving the OCR quality.

BHL-Europe is in contact with the EU-funded IMPACT project. It aims to significantly improve access to historical text and to take away the barriers that stand in the way of the mass digitisation of the European cultural heritage. One important objective is to innovate

OCR technologies. A meeting was held on 24 June 2009 in the Austrian National Library (attended by Henning Scholz, Heimo Reimo for BHL-Europe), to investigate the cooperation potential and the participation of BHL-Europe in the test period of the new IMPACT tools. In return, BHL-Europe may help IMPACT with quantitative measures of OCR quality.

BHL-Europe is also further investigating the work of other projects on OCR. On 24 September 2009, BHL-Europe WP3 leader attended a workshop entitled “Optical Character Recognition (OCR) for the mass digitisation of textual materials: Improving Access to Text” at the University of Bath (<http://www.ukoln.ac.uk/events/ocr-2009/>). The workshop provided an opportunity for participants to learn about the current state-of-the-art in the digitisation of historical texts, to look at improvements in digitisation techniques currently being explored in research projects such as the EU-funded IMPACT project, and to explore how Optical Character Recognition (OCR) is used in practical digitisation contexts and workflows.

In cooperation with NHM, a JISC project is carried out with the overall aim to establish and extend information extraction techniques from scanned taxonomic literature in the Biodiversity Heritage Library. The Automatic Biodiversity Literature Enhancement project (ABLE) aims to extract mark-up and metadata from scanned literature in the biodiversity domain, including proper nouns (taxon, people and place names) and dates. ABLE also intends to enhance the searchability of those terms using associative techniques from Natural Language Processing combined with likely OCR errors. If fully successful the software developed in the ABLE project will be applied to the BHL library soon.

Another approach currently investigated by our colleagues in SIL and MOBOT in cooperation with Internet Archive is a book wiki (FLOSS Manuals). This wiki includes tools for collaborative OCR re-keying and correction. Using this tool helps BHL-Europe to engage the members of the scientific community to help with OCR improvements. This collaborative effort has proven to work successful for specific communities, see for example the Australian Newspapers Digitisation Program (<http://www.nla.gov.au/ndp/>).

Content holder requirements: Initiated by Bernard Scaife (NHM), discussions were carried out over the BHLwiki to work on the content holder requirements (D2.1). These discussions were based on the BHL specifications. A draft document was released on 24 June 2009 to summarise first round of discussion. Eleven content-providing partners in BHL-Europe took part and provided their views. In a second round of discussion, more contributions were collected to finish the document on 5 August 2009. Based on this documents, BHL-Europe content providers are currently adapting their internal processes for digitisation and post-processing of content. This facilitates the actual ingest of content into BHL-Europe in the near future.

Metadata requirements for IPR has been also documented as part of D2.1. The WP4 leader has worked closely with WP2 colleagues for a pragmatic solution to content with embedded rights and layers of rights – given that the majority of the content digitised by our partners is out of copyright it is thought this will affect minimal content.

Consortium Agreement: The Consortium Agreement was prepared in the reporting period. Although not required in projects of the *eContentplus* programme, it is strongly recommended. A first draft was available in August based on the DESCAs templates. The document was customised, but the liability section of the document is still in discussion. There are also ongoing discussions regarding the extent to which IPR issues are being addressed in the Consortium Agreement.

Memorandum of Understanding: A major step to actually start the ingest of digital content into BHL-Europe is the MoU. This document specifies the type of content every partner will hand over to BHL-Europe as an amendment of the content table in section 4.1 of the Description of Work (Annex 1 of BHL-Europe grant agreement). It also defines the models of access and gives an overview of the technical specifications of the content. The document was drafted by the WP2 leader and discussed in the PMG of the project. The wording then was agreed during a meeting of several project partners in Tervuren on 22 October (MfN, NHM, NHMW, RMCA, NBGB, RBINS). The final review of this document is now finished and signature by each content provider is expected soon.

Project Server: NHM established the Project Server for BHL-Europe. MS Project Professional 2007 is used as the project management software. The Project Server helps to structure reports and statements, to plan tasks in terms of work time and budget, to monitor all work carried out and to control the budget. The Project Server is working in a PRINCE2 framework. In July a one-week meeting was scheduled in London with all WP leaders of BHL-Europe to learn the functionalities of the Server and to customise the Server for our needs. However, the customisation also includes the input of all resources with details of their work contribution and hourly rates. Data collection is still in progress as some partners have difficulties accessing the required information from their administrative departments.

EDIT and BHL-Europe integration: The FUB-BGBM is coordinating the development and implementation of the EDIT Platform for Cybertaxonomy which is a service-oriented infrastructure supporting the taxonomic workflow from fieldwork to the creation of printed manuscripts for publication of scientific results. The platform consists of a suite of loosely coupled taxonomic software applications (<http://dev.e-taxonomy.eu/platform/>) communicating via data standards as well as instances of the EDIT common data model (CDM, <http://dev.e-taxonomy.eu/trac/wiki/CommunityServer>).

Within BHL-Europe, the BGBM (in cooperation with the NHM London and the EDIT Virtual Taxonomic Library - ViTaL) will integrate BHL-Europe services with the EDIT Platform for Cybertaxonomy in order to facilitate access to BHL-Europe resources and feedback to BHL-Europe.

The EDIT CDM-Stores will not only consume BHL-Europe services, but will contribute by exposing catalogue data as well as digitised literature. For this, additional services have to be implemented into the EDIT CDM service-layer allowing BHL-Europe to harvest reference-repositories belonging to CDM community stores. CDM community stores may also expose literature digitised by EDIT communities to BHL-Europe harvesters. All services should conform to agreed BHL-Europe bibliographic data standards and rights management.

The workplan for this work was developed at FUB-BGBM to actually start with the end of the consensus building phase of BHL-Europe in November 2009.

On a different level, BHL-Europe is cooperating with EDIT to facilitate the access to the 124 titles of publications (research journals, etc.) of the 28 EDIT partner institutions. Currently, the entire publication process is in review to face the challenges of the digital age and to enable open access to the research results of the EDIT institutions. On the other hand, the libraries of the EDIT partners currently are discussing the de-accessioning of original material in European libraries to save space and costs. BHL-Europe is helping EDIT scoping the problem and work on solutions. A well maintained catalogue of library holdings with

integrated deduplication tools that BHL-Europe needs for the content management can be used as an overview of the number of physical copies that should be kept in Europe.

Content specific progress of consortium partners: Although the digitisation of biodiversity heritage literature is not part of the BHL-Europe project, activities towards the increase of content are important. BHL-Europe needs to provide access to 25 million pages of heritage literature from at least 30 content providers. In this section, examples of the activities of partners within their national framework are reported.

MfN, FUB-BGBM: Both institutions are actively negotiating for biodiversity content mobilisation in the German library system. A first proposal is prepared to be submitted to the German Research Foundation to scan all German-language botanical and palaeobotanical serials. An important focus of this project is the detailed databasing of article-level metadata.

UGOE: The main focus of UGOE in the past 6 months has been in analysing BHL resources in terms of user requirements and to report shortcomings of BHL web presentations of digitised literature. This contributed to help improving the BHL online presentations, notably in terms of speed and metadata accuracy. UGOE also worked on the establishment of guidelines for functionalities of the future BHL/BHL-Europe Portal.

Digitisation of literature in the EZOOLO/AnimalBase project continued and approximately 350 monographs were digitised and went online in the past reporting period.

LANDOE: The major focus of LANDOE in the first reporting period (6 months) was to expand the portfolio of scanned books and collecting and digitising metadata of natural history literature. The metadata of several medium sized series (up to 50 books) have been completed and it was started with one large series (more than 200 books - Plant systematics and evolution). Up to date, 100,000 additional pages have been digitised, of which 40,000 are already OCR-scanned. The archive now holds 450,000 OCR scanned pages. 390,000 pages are freely available to the general public. LANDOE now has 80 series in different stages of digitisation on the Web pages available.

HNHM: More than 50 colleagues from Slovakia, Hungary, Romania attended the 27th Annual Meeting of the Natural History Curators of the Carpathian Basin (Zlatá Idka, Slovakia, 8–11/09/2009). Discussions were initiated with a short introductory lecture on metadata using BHL-Europe as an example. A follow-up discussion with representatives of the 5 largest regional natural history museums was also organised. All have their own journals and periodicals, and cooperation with these institutions would help BHL-Europe to increase the amount of content available through the BHL Portal and Europeana. All institutions are very interested in contributing their publications. Further negotiation and technical implementation will be expected during the next reporting period. Based on experiences, the network can also be extended to include Slovakian, Romanian museums as well.

HNHM performed a detailed survey of content available digitally at HNHM. The scattered and incomplete description of the journals published by HNHM at article level was converted to a unified database. This is finished for four journals, two of them are already released to the web (<http://annales.nhmus.hu>, <http://miscellanea.nhmus.hu>). Further quality checks of the remaining journals are in progress before they will also be released to the web too.

UH-Viikki: UH-Viikki has concentrated its activities on task 2.3.3 of the BHL-Europe project (coordinate with EUROPEANA, BHL and national scanning projects to ensure that material scanned by BHL-Europe is available through these portals). The team at UH-Viikki has set up a local steering committee consisting of the following representatives: two Finnish learned societies in biology, namely Societas pro Fauna et Flora Fennica and Vanamo, scientists from the Finnish Museum of Natural History working with taxonomy or with the digitisation of biological specimen collections. The committee also includes a professor of botany at the University of Helsinki, Viikki Campus. The steering group has held two meetings during this period. It has produced a formal letter to the National Library which operates a well equipped mass digitisation centre. The letter explains the BHL-Europe project and makes an initiative of cooperation in digitisation of Finnish biological heritage collections. The response from the National Library is expected soon.

Relating to WP5 dealing with dissemination the team in Helsinki has set up a public Finnish language wiki site for the project containing background information and access to a RefShare database aimed at collecting comments and suggestions from researchers for setting priorities for the selection of publications and series that should be digitised. The metadata in this database was transferred from the HELKA library catalogue which is a joint database of the University of Helsinki libraries and the National Library of Finland. The same database was used in harvesting test data for the BHL-Europe portal prototype at AIT. The Finnish wiki is located at <http://wiki.helsinki.fi/display/BHL/Etusivu>.

Activities related to dissemination included participation in a panel discussion on Europeana and the Finnish national digital library project on 20.5.2009 at the Ministry of Education where BHL-Europe was the subject of our presentation. Another meeting organized by the Ministry of Education on 3.9.2009 focused on presenting all Finnish projects that contribute to Europeana as aggregators. As a result the public web site for the National digital library project now contains information on BHL-Europe (<http://www.kdk2011.fi/fi/europeana-jamuuut-hankkeet/bhl-for-europe>). Information about BHL-Europe and instructions on how to use the RefShare database were published on the homepage of the museum. BHL-Europe was presented for the faculty and cooperating research institutes in the three subject oriented advisory committees of our library.

The University of Helsinki has recently combined four existing campus repositories to establish a unified DSpace repository for the university called HELDA – University of Helsinki Digital Archive. The BHL-Europe team has created a special collection named Fauna & Flora in this repository to enable future harvesting of metadata to BHL-Europe. The collection is located at <http://helda.helsinki.fi/handle/10138/4>. HELDA also contains the digitised collections of the Finnish Society of Forest Science and of old forestry literature.

Relating to IPR issues in WP4, UH-Viikki has organized the sending of a letter to all members of Societas pro Fauna et Flora Fennica explaining the BHL-Europe project and asking collectively all authors and IPR holders of articles published before 1940 in the society's series for a permission to digitize and deposit these articles on portals like BHL-Europe and Europeana. The rights holders objecting to this are asked to notify the society.

It might also be of interest to note that Viikki Science Library has been invited to the steering group of FinnONTO 2.0, a large national semantic web project running during 2010-2011. Another participant in this project is the Finnish Museum of Natural History which is starting to build an ontology for 50,000 species found in Finland.

Wiley-Blackwell: As a commercial publisher, Wiley-Blackwell is included in the BHL-Europe consortium. During a meeting in Oxford and further Email communication, it was investigated to what extent BHL-Europe may access the data produced by the digitisation programme of Wiley. To date, about 500,000 pages of various journals have been digitised. Accessing metadata of these publications is not problematic, but access to the actual digital content needs further investigations and negotiations. We propose to use an agreement with Wiley as a model agreement for other publishers as well.

The meeting in Oxford was attended by Henning Scholz (MfN), Graham Higley (NHM), Mike Davis (Wiley), and Liz Ferguson (Wiley).

Dissemination: The working team in the National Museum, Prague prepared the deliverables for M3. NMP worked on corporate design of the project including its logo. The Web site of the project was designed including a demonstrator. The need for terms and conditions of use for the project Web site were identified in cooperation with WP4 leader, who also wrote these terms. The need to employ legal professionals to ensure accurate translation of legal text has been also identified. The translation of the current content of the Web site in German, French, Spanish, and Czech is finished and needs to be uploaded. Further content enhancements will be discussed during the next project meeting to be held at the home institution of the core WP5-team at NMP. The first issue of the newsletter has been published already, the second issue is underway and will be online mid-November. In cooperation with Tine Mallaerts (Royal Belgian Institute of Natural Sciences), NMP prepared a mailing list and a database of relevant conferences. The first dissemination plan was delivered in M3, the final one will be available at the end of M6. NMP also designed a poster template to be used at various conferences to present BHL-Europe results using the identical design concept. This template was populated by the PCO of BHL-Europe for the use at six conferences in fall and winter 2009. During the Kick-off meeting the composition of the Communications Working Group was delivered. The following members were selected:

Nancy Chillingworth	NHM, Leader WP4
Thomas Garnett	SIL, Observer
Lizzy Komen	EDLF, Europeana
Jiri Kvacek	NMP, Leader WP5, Chair CWG
Tine Mallaerts	RBINS
Henning Scholz	MfN, Leader WP1 and PCO
Antonio G. Valdecasas	CSIC

Project publications to date:

Hardy, G. 2009. Press release.- Scottish newspapers.

Hardy, G. 2009. Press release.- New Scientist.

Koch, G. 2009. Österreich und die Digitale Bibliothek Europas. - BIT Online, autumn 2009

Koch, W. & Scholz, H. 2009. DISMARC and BHL-Europe: multilingual access to two aggregation platforms for Europeana.- AT4DL.

Kvacek, J. & Scholz, H. 2009. Biodiversity Heritage Library for Europe - Towards a Global Library of Life.- EDIT Newsletter No. 15.

Scholz, H. 2009. Press release.



Scholz, H. 2009. Biodiversity Heritage Library for Europe - Auf dem Weg zu einer globalen Bibliothek des Lebens.- GfbS Newsletter.

Scholz, H. 2009. Auf dem Weg zu einer globalen Bibliothek des Lebens.- Gmit.

Valdecasas, A.G., Ramos, M. & Moron, I. 2009. Hacia la Biblioteca Digital Universal de Historia natural.- Periódico del MNCN, N 10: 12.

Valdecasas, A.G. 2009. Hacia la Biblioteca Digital Universal de Historia natural.- Graellsia.

3.1 Resources employed

Resources employed for the reporting period (person-months)						
Beneficiary short name	WP 01	WP 02	WP 03	WP 04	WP 05	TOTAL
MfN	5.5	3.5	0.2	0.1	0.35	9.65
NHM	0.11	2.45	4.62	1.82	0.15	9.15
NMP	0.12	0	0	0.04	7.42	7.58
EDLF	0.16	0.3	0.6	0	0.15	1.21
AIT	0.33	0	9.66	0	0.21	10.20
ATOS	0.14	0	0.73	0	0	0.87
FUB-BGBM	0.27	0.30	0.27	0	0	0.84
UGOE	0.01	0.27	0	0.01	0.02	0.31
NHMW	0	0	1.18	0	0	1.18
LANDOE	0.14	0.48	0.02	0.14	0.01	0.79
HNHM	0.05	0.25	0	0	0.1	0.4
MIZPAS	0.14	0.14	0	0	0.14	0.43
UCPH	0.12	0.34	0.18	0	0	0.64
NAT	0.29	1.24	0	0	0	1.53
NBGB	0.36	1.14	0	0.08	0	1.58
RMCA	0	0	0	0	0	0
RBINS	0.26	1.47	0.04	0.11	1.96	3.84
BnF	0.06	0.73	0.31	0	0.11	1.21
MNHN	0.18	1.18	0	0.05	0.09	1.50
CSIC	0.05	0.46	0.08	0.07	0.23	0.89
MSN	0.24	0.15	0.02	0	0.48	0.88
RBGE	0.21	0.24	0	0.01	0.01	0.47
Sp2000	0	0	0	0	0	0
Wiley	0	0	0	0	0	0
SIL	0	0	0	0	0	0
MOBOT	0	0	0	0	0	0
UH-Viikki	0.15	2.37	0.17	0.07	0.18	2.94
UBER	0.02	0.02	0	0	0.02	0.06
TOTAL	9.27	16.74	18.01	2.50	11.63	58.15

3.2 Work Package Overview

Work package description

Work package number :	1	Start date:	M1	End date:	M36
Work package title:	Project Coordination and management				

Objectives for the period

Administrative objective: Ensure adherence of the consortium to the rules, regulations, and financial guidelines of the *eContentplus* programme; establish the project in the European biodiversity community.

Technical objective: Implement the project as set out in the work plan; ensure exchange of information and communication between partners; ensure progress of the project; guarantee timely deliverables.

QA objective: Ensure verifiable progress of the project and be committed to high quality output that has tangible impact on *eContentplus* programme objectives.

Description of work carried out and achievements

T1.1.1: Resource planning (financial, personnel, material), monitoring and controlling.

Achievements: The Project Server in interaction with MS Project Professional helps BHL-Europe to plan and monitor resources, budget, and material.

T1.1.2.1: Liaison between the Commission, consortium members, EUROPEANA, BHL, and external experts; effective communication with the consortium members, Work Package leaders, the Commission, and interested external parties.

Achievements: Email, phone, Skype, BHLwiki, and regular personal meetings ensure effective communication between all parties involved in BHL-Europe and related to the work carried out in the project. Every Tuesday, the Project Management Group of BHL-Europe has a conference call to update on progress and to discuss recent developments and issues.

T1.1.2.2: Coordination of meetings and progress reviews.

Achievements: The Kick-off meeting was held between 10 May and 15 May 2009 in the Museum für Naturkunde Berlin. A second project meeting was held from 10 to 12 August in Naturalis Leiden. The third big project meeting is currently planned for 15-18 November in Prague. In between, smaller workshops were organised to work on specific aspects (mainly technical) of the project. Sometimes these workshops were held as side meetings of conferences. These workshops were held in London (3 June, 20-24 July), Paris (8-9 June), Malta (24-26 June), St. Louis (17-21 August), Graz (3-4 September), The Hague (13-16 September), Berlin (29 September), and Copenhagen (13 October).

T1.1.3: Production and consolidation of periodic external reports; set-up of the Web-based project management portal.

Achievements: This document is the first periodic report of the project. The Web-based project management portal was set up in July based on a MS Project Server. It

will be fully customised and running by 15 November 2009. Further details are provided above in section 3 of this report, see also below. In addition, the BHLwiki (<https://bhl.wikispaces.com/BHL-Europe>) is used for the exchange of documents, management of the preparation of deliverables, and discussion of questions of a more general importance.

T1.1.4: Representing the project.

Achievements: The project was presented at various conferences and meetings (see below). The Project Coordinator represented the project on various occasions (e.g. Europeana Aggregator Round Table).

T1.1.2.1: Work-package and task coordination.

Achievements: The Project Server interaction with MS Project Professional helps BHL-Europe to manage the work packages and tasks of the project. WP leader are setting the WP work plan for the next working period based on the project plan (see DoW). This is then approved by the PCO and aligned by WP1 with all other WP. Resources are aligned with the tasks and will be informed about the tasks and the estimated work time by the WP leader. Resources report back to the server about achievements on this task.

T1.2.2: Project plan maintenance; monitoring of project progress and milestones; identification and trouble shooting of technical and organisational problems.

Achievements: The Project Server in interaction with MS Project Professional helps BHL-Europe to plan the project, maintain the workplan, monitor the project progress, and help in risk and issue management. The server is not fully working because the customization is not finished yet. It is expected to be finished by 15 November 2009.

T1.2.3: Timely production of deliverables.

Achievements: The Project Office of BHL-Europe is the final instance for ensuring timely production of deliverables. In cooperation with the WP leaders, the members of the consortium will be asked to contribute in fulfillment of their commitments according to the Description of Work.

T1.2.4.1: Quality control against the technical and contractual aspects.

Achievements: QA procedures are in place (see below) to ensure quality control against the technical and contractual aspects.

T1.2.4.2: Coordination with EDL Foundation Office.

Achievements: One member of the EDLF Office is member of the BHL-Europe Project Management Group to ensure proper alignment of all activities of BHL-Europe with Europeana.

T1.3: Quality assurance: Definition and communication of quality assurance procedures via project management portal, coordination of quality assurance process (internal as well as external review procedures for the various project results and deliverables).

Achievements: The QA procedures were communicated via the BHLwiki and agreed during the PSG meeting in August. External reviewers are identified and will be consulted to proofread the deliverables of the project.

T1.4.2: Identify related networks or organisations interested in the implementation of project results.

Achievements: Networks that have been identified in the reporting period as important for the future development of BHL-Europe and implementation of project results are LifeWatch (Data Provider Platform), PESI (brings citations to BHL-Europe), D4Science (BHL as content partner), and AquaRing (semantic enrichment)

T1.5.1: Coordinate with relevant partner networks

Achievements: The most important partner networks we are coordinating with for our current work are Europeana (ingest for Rhine release), EDIT (ViTaL implementation), BHL (alignment of standards, content ingestion procedures), Key2Nature (adoption of Use Cases), and IMPACT (testbed for OCR improvement).

T1.5.2: Contribute to the clustering activities of the EC.

Achievements: No actual cluster meetings of the EC were held. However, BHL-Europe was present and contributed to all relevant Europeana meetings and conferences (Plenary, Aggregator Round Table) and also to events of the EC (Digital Libraries and Digital Preservation: ICT Call 6 Information Day)

Deviation from work plan & remedial action

The Project Server was ready for use as our project management portal in July, but was only available for the use of the WP leader by that time, and was not available to not other members of the BHL-Europe consortium. More work on the customisation of the server was carried out in the last three months. This will be finished in the next two weeks. The next project meeting in Prague (15-18 November) will be used to give detailed instructions for all team members during a Project Server workshop.

No work was foreseen for Task 1.4 (Business Plan) in our work plan, but was in fact carried out as part of our networking activities.

Work package description

Work package number :	2	Start date:	M1	End date:	M36
Work package title:	Analysis of domain content and management of the content acquisition process				

Objectives for the period

IT objective: Establish bibliographic database systems, metadata repositories and Web-based content management systems.

Management objective: Ensure that all content providers agree on the technical architecture of the project; ensure linkage to EUROPEANA; ensure extension of the content providing network.

Description of work carried out and achievements

T2.1.1: Establish a bibliographic database system and metadata repository for monographs and serials based on the Virtual Taxonomic Library developed in EDIT, and also the experiences of the BHL.

Achievements: BHL currently is in the final phase of releasing an open access repository for biodiversity publications (articles) including a user-vetted bibliographic database (CiteBank). BHL-Europe and EDIT agreed to work on an integrated use of ViTaL components for BHL-Europe. At the moment, we are reorganising the ViTaL components to make them BHL-Europe compliant and work on the actual implementation approach.

T2.1.2: Develop and enhance deduplication tools for monographs and serials based on BHL technologies.

Achievements: Deduplication algorithms for monographs and serials were tested on various datasets. For serials, the actual BHL serials bidlist and an extract of the ZDB (Zeitschriftendatenbank, the central Germany serials database) were used as testsets. For monographs, small BHL sets were used. The fine tuning of the algorithms is underway using the entire BHL dataset to prototype a web tool for the deduplication of monographs.

T2.2.1: Establish a list of monographs and serials that are relevant for the biodiversity community.

Achievements: UH-Viikki established a RefShare platform for collecting comments on books that are proposed for digitisation. This currently works in Finland only. It needs to be discussed, if and how this will be extended for the BHL-Europe community or integrated in the ViTaL service, that will be available for BHL-Europe as well.

T2.3.1: Identify content holder requirements; develop the Memorandum of Understanding.

Achievements: D2.1 identifies the content holder requirements. The MoU defines the models of access and datatypes provided by each content provider in detail. The signment on a bilateral basis is expected soon.

T2.3.2.1: Discussion and distribution of data standards and specifications

Achievements: The data standards and specifications were extensively discussed during the first three months of the project and agreed upon in July. This is

documented by D2.1.

T2.3.2.2: Work with individual donors and governments to facilitate the funding of the scanning

Achievements: Several of BHL-Europe partners are continuously approaching the national governments for funding for scanning projects on a national basis.

T2.3.2.3: Attracting new content providers.

Achievements: The library of Bielefeld University (Germany) is interested to provide content (page images) for BHL-Europe. A metadata sample is available already to test the data harmonisation. The Institute of Speleology "Emil Racovita" (Cluj, Romania) is interested to provide metadata for the BHL-Europe bibliographic database. A contact is made to the Danish National Library in adding more content from Denmark too. Various other partners made contact with potential content providers that needs to be investigated further.

T2.3.3: Coordinate with EUROPEANA, BHL and national scanning projects to ensure that material scanned by BHL-Europe is available through these portals.

Achievements: BHL-Europe has established a timeline and procedure for the content ingestion into Europeana. Actually, BHL-Europe content will be available through Europeana with the Rhine release in July 2010.

T2.3.4: Addressing IPR issues in cooperation with WP4.

Achievements: IPR issues were important for the development of the MoU. WP4 was consulted to align the wording of this document with the work plan of WP4.

Deviation from work plan & remedial action

The MoU was not signed by the end of this reporting period as indicated in the Description of Work. However, the wording is agreed and the MoU will be signed on a bilateral basis between the content provider and BHL-Europe to facilitate the process. As the wording has been already agreed, signature is now a formal process that will not delay the content ingestion. If content providers are late with signature, BHL-Europe can in any case still start with other content providers or first establish ingest procedures for any content providers late in signing the MoU.

Although foreseen to start after M6 of the project, the attraction of new content providers, was done as part of our networking activities.

The task 2.2.3 foreseen to start in this reporting period is delayed because no working Web-database is available. This database will appear as a prototype in M9 together with the deduplication tool and the bibliographic database. However, the specific function identifying the distribution of literature is also dependent on the developments of the ViTaL platform that is delayed as well. Solutions to this problem will be sought in the coming months.

Work package description

Work package number :	3	Start date:	M1	End date:	M36
Work package title:	Technological implementation				

Objectives for the period

Management objective: Management and coordination of technological development.

IT objective: Design of the technical and functional architecture of BHL-Europe; Definition of standards and specifications; Development and adaptation of tools; Building the BHL-Europe prototype portal

Description of work carried out and achievements

T3.1.1: Management of the technical development team.

Achievements: To facilitate the decision making process and the implementation of the architecture, two working groups were created. The first is the Technology Management Board to help on a strategic level, the second is the Technical Implementation Group to actually implement and discuss technological solutions.

T3.1.2: Adaptation of EUROPEANA and BHL data model, workflow, harvesting procedure, standards, specifications.

Achievements: The BHL standards were adapted to meet the requirements of the BHL-Europe content providers and to be compliant with the Europeana Semantic Elements.

T3.1.3: Definition of own standards for images, metadata, harvesting according to partner requirements (consensus-building).

Achievements: We adopted the BHL data model and standards and discussed them in the BHL-Europe consortium. This was adapted then to meet the requirements of our partners.

T3.1.5: Development and adaptation of specific tools.

Achievements: The harmonisation of metadata of partners and library catalogues require specific mapping tools to facilitate the migration of data without losing the original data. For this the mapping tool of the Dismarc reference system will be adopted. Furthermore, another tool is required to export data from library system to actually make metadata available for ingestion into BHL-Europe. This tool was designed by NHMW and is currently undergoing development and testing.

T3.2.1: Ensure that the standards for data management and image formats are consistent with international requirements and specifically the Europeana and BHL, adoption of Europeana Semantic Elements (ESE).

Achievements: We adopted the BHL data model and standards and discussed them in the BHL-Europe consortium. These standards are also compatible with the latest version of the Europeana Semantic Elements.

T3.3: Addressing distributed access and storage – long-term sustainability: Develop a distributed access and storage system to enable national and international storage of the scanned materials; to develop the storage system in such a way that long-term sustainability of the data is secured.

Achievements: We decided to build an OAIS compliant repository during our

consensus building phase. The actual development of the system, however, will follow after the approval of the implementation plan (D3.4) in month 9 of the project. A large data centre was designed for the sustainable storage and curation of the digital content and metadata of BHL-Europe.

Deviation from work plan & remedial action

T3.1.5 was started in the last months of the reporting period to kick-off the mapping process. The tools now are working with standard metadata formats. In the next period the actual adaptation of these tools to the agreed standards will follow to push the metadata harmonisation process forward.

T3.2.1 was started earlier than planned as part of our consensus building process. From now onwards, the Europeana Content Checker is available to actually map BHL/BHL-Europe metadata against ESE and ensure our content is displayed correctly in the Europeana framework.

T3.3 is foreseen to start after M12 of the project. However, the design of the architecture of that system and the design of the physical storage solution is a very time consuming process. The architecture is currently discussed as part of the consensus building process. Architectural design documents were prepared for D3.2. This process will be pushed forward in the next weeks (D3.4) to be prepared for the actual building of the distributed access and storage system. For the data centre, the basic conditions for setting this up at this early stage of the project were considered as very advantageous to successfully establish the centre.

Work package description

Work package number :	4	Start date:	M1	End date:	M36
Work package title:	Intellectual Property Rights				

Objectives for the period

Management and coordination of the intellectual property rights (IPR) framework for BHL-Europe. Ensure that BHL-Europe, EUROPEANA, and the BHL are using common approaches and common agreements, such that data can be exchanged between these partners without further Rights activity.

Description of work carried out and achievements

T4.1: IPR Framework: Establish IPR working documents – including best practice guide, due diligence guide, pro-forma agreements, and process for formally agreeing rights management with rights holders; align the approach with EUROPEANA and BHL and maximise level of interoperability including Rights metadata exchange. Achievements & Activities: WP4 leader, Nancy Chillingworth, attended Kick-off meeting to communicate to partners the role of IPR in the successful delivery of the project. She established what IP is out of scope of this WP and the importance of IPR to sustainability, particularly in regard to open source software, establishing Brand in the absence of rights control over Public Domain or Open Access content.

An IP working group was established in Leiden during the M3 project meeting. This group met to discuss issues and principles about Rights in digital object. An agreement was made that whether or not rights exist they will not be asserted for works where the original is out of copyright. It was agreed that legal advice should be sought to be certain of legality of actions in this WP. This project meeting was also used to establish common IPR terminology – OA / PD, Identify key cross-work package dependencies, the prioritisation of content ingestion around rights restrictions, acceptance of content from non-partners, and to set up an IPR survey.

Following the Leiden meeting, the IP questionnaire was sent to partners to establish rights restrictions on content for project and existing practices and documentation across partner digitisation projects. Results have been limited but broadly encouraging.

WP4 leader has undertaken responsibility to understand the legal issues affecting licensing of content across national borders: Investigated areas include legal jurisdiction, liability of DPs, liability of Aggregators, position of BHL-Europe as a nonlegal entity, licensing implications for BHL-Europe, OCR translations and moral rights/right to translation. This work has informed the content of the Consortium Agreement and MoU for BHL-Europe partners, as well as contribution at MoU meeting with WP2 leader and partner representatives.

Rights clearance – preparation for this task has involved:

- Avoidance of uncoordinated/duplication of permission requests
- Building rights clearance into workflow
- Batch approach key publishers/societies

- Flexibility in licensing solutions
- Out of copyright works
- Orphan works and due diligence for rights holder searches

WP4 leader is keeping abreast of EC consultation and meeting with government representatives regarding UK legislative proposal for heritage institutions to licence Orphan Works. Colleagues within the ARROW project for Orphan Works have also been heavily involved in discussions for a solution to this problem. The speed of this solution depends on legal change at a European level. BHL-Europe is ensuring it facilitates the use of best practice in this area until such a change comes about.

Alignment with Europeana and BHL: WP4 leader has conducted a review of existing BHL copyright documentation, including Web site terms, due diligence guidance, and pro forma permission request letter to establish suitability for this project. This led to the identification of inaccuracies in BHL current licensing model – BHL is attributed as rights holder for digitised content. Requests for commercial uses are directed to BHL rather than Rights Owner. Attribution is stated as necessary to BHL rather than rights owner.

WP4 leader is also a member of the Europeana IPR Cluster Group and attended the Europeana Workgroup Kick-off in April 2009, the IPR Cluster Group meeting in May 2009 and the Europeana Plenary in September 2009, to address:

- Rights in digital surrogate
- De-duplication of deliverables across Europeana content providers
- Legal jurisdiction and discrepancies between national laws
- Licensing models for content and data
- Public Domain Charter – contribution to the breakout discussion about this controversial initiative.
- Europeana data provider agreement – clean hands model. WP4 leader has had close collaboration with Patrick Pfeiffer of Europeana in the development of this document.

Deviation from work plan & remedial action

Common approaches and common agreements may need to be interpreted as ‘comparable’ rather than ‘identical’, for the following reasons: national IP legislation and ideologies differ greatly to the extent that fundamental copyright principles differ. Secondly, BHL-Europe is a cog in a larger machine and our IP solution regarding licensing content may not be the solution driving others.

Work package description

Work package number :	5	Start date:	M1	End date:	M36
Work package title:	Dissemination, Exploitation and Evaluation				

Objectives for the period

Dissemination objective: Develop dissemination strategy; raise awareness, understanding and action of the project among the community and stakeholders in EU member states; ensure effective dissemination of project goals and results to the target users; ensure good communication within the European scientific community

Presentation and demonstration objective: Ensure dissemination of project results at conferences, public events and among the networks of the consortium members

Description of work carried out and achievements

Task 5.1: Establish BHL-Europe Communications Working Group and dissemination plan
Achievements: The CWG was established during the Kick-off meeting in Berlin (May 2009) and had the first meeting on 14 May 2009. The CWG is the first instance for discussion of any issues related to the dissemination activities of the project and for revision of deliverables of WP5 and publications of BHL-Europe. Associated to this task was the preparation of the dissemination plan.

T5.2.1: Plan, design, and publish BHL-Europe Web site for internal and external communication with publication of all relevant results and links to partner networks.
Achievements: The BHL-Europe Web site went online end of July 2009: <http://www.bhl-europe.eu>. It was designed under the leadership of our project designer Pavel Šťastný. The CWG was mainly involved in the review process of the Web site. The Web site informs about the project, provides basic information about its major goals, and information about all 28 partner institutions. Among others it contains a demonstrator, newsletter and conference list.

T5.2.2: Develop a mailing list to facilitate communication of new developments (internal and external); update the list regularly to include all relevant parties.
Achievements: For communication with users we prepared a mailing list. This list will be used for various dissemination activities, particularly for sending the newsletter. Currently, nearly 100 people outside the BHL-Europe consortium receive the newsletter and the number of members is growing.

T5.2.3: Develop self-running demonstrators to illustrate BHL-Europe use and functionality.
Achievements: A first version of the self-running demonstrator is a Power-point presentation. It will be updated depending on project progress. A new design will be applied to make it more attractive and suitable for the general public

T5.3.1: Develop, prepare, and update target group specific promotion kits (fact sheets, flyers, posters, presentations).
Achievements: First flyers were already available in print in August 2009. First presentations were given in June 2009 already, first posters using the official template in October 2009. For all presentations and posters the same corporate design concept applies. It was developed under the leadership of our project designer Pavel Šťastný.

T5.3.2: Prepare a quarterly newsletter to be published online.

Achievements: The first newsletter of BHL-Europe was published in August 2009. It is placed on the Web site where users can easily download and subscribe. The newsletter contains news about the project, present and future meetings, and relevant articles.

T5.3.3: Distribution of articles for various types of print and online media (journals, newsletters).

Achievements: Several publications about the project (for conference volumes, newsletters, and museum journals) and one press release were published. A selection is available on the project Web site. The complete list is provided above.

T5.4.1: Prepare and update multimedia presentation (automatic demonstrator) as information about the project for the general public.

Achievements: This presentation is included on the BHL-Europe Web-site.

T5.4.2: Establish the database of conferences and public events relevant for BHL-Europe; identify consortium members to present the project at selected events.

Achievements: This database was developed by RBINS. The update and maintenance involves a questionnaire sent to all partners every three months to collect up-to-date information on conferences.

T5.4.3: Alignment of public awareness activities of related projects (e.g. EDIT).

Achievements: BHL-Europe has investigated how to align these activities with Europeana for the Rhine release and the Year of Biodiversity in 2010. A meeting is scheduled for April 2010 to discuss details.

T5.4.5: Represent BHL-Europe at conferences and special events (responsibilities depend on conference location and theme).

Achievements: Project members gave oral or poster presentations in various conferences. The complete list is provided below.

T5.5.2: Develop online questionnaires to identify user requirements, preferences, experiences, benefits, and un-met needs.

Achievements: A first user requirement survey was developed in September to develop a first set of information about the functionality and services needed for the BHL-Europe prototype. It will be evaluated during our next project meeting in Prague (November 2009) and results will be presented in the second progress report. The design of the survey is described in D5.5.

Deviation from work plan & remedial action

Deliverables in the first three months were submitted late to the Project Coordinator. This was due to the late recruitment of staff members at NMP caused by the delay of money transfer. However, this has affected neither the delivery of the following documents nor the work plan of the project, so no remedial action was carried out.

No public events or talks (Task 5.4.4) were organised in the first reporting period, but numerous presentations were given during conferences in order to reach as many target users as possible.

The preparation of the online questionnaires to identify user requirements was planned for M9 of the project. However, a first survey was prepared before that time in order to facilitate the building of the prototype based on actual user requirements, and to further engage the consortium partners. This will also help us preparing the next survey in focusing of specific items and functionalities for the target users of BHL-Europe. The review of the survey questionnaire during the revision of the dissemination plan has shown that several questions

are not appropriate for all target users and that the design of the survey should be adapted for future surveys. Therefore, this test survey was of added value to use the full potential of this powerful evaluation tool.

3.3 Deliverables Status

Deliverables List

Deliverable No¹	Deliverable title	Delivery due date²	Actual date of delivery³
D2.1	Catalogue of content holder requirements (quality, quantity, accessibility, standards and specifications of content and metadata)	Month 3	Month 4
D3.1	Deliver composition of Technology Management Board and initial meeting	Month 3	Month 4
D5.1	Web site, including multimedia presentation	Month 3	Month 4
D5.2	BHL-Europe newsletter and mailing list	Month 3	Month 4
D5.3	Database of relevant conferences/events and ownership for BHL-Europe presentations	Month 3	Month 4
D5.4	Deliver composition of Communications Working Group and 1 st dissemination plan	Month 3	Month 4
D1.1	Progress Report 1	Month 6	Month 7
D3.2	Document agreed standards, best practice and system components	Month 6	Month 7
D3.3	Plan for managing interoperability issues, data harmonisation and the integration of the content into BHL-Europe, EUROPEANA and the BHL	Month 6	Month 7
D5.5	BHL-Europe dissemination plan	Month 6	Month 7
D5.6	BHL-Europe promotion kit	Month 6	Month 7

¹ Deliverable numbers in order of delivery dates: D1 – Dn. Deliverable numbers must indicate which workpackage they relate to, e.g. D2.1 for the first deliverable from workpackage 2).

² Month in which the deliverables will be available. Month 0 marking the start of the project, and all delivery dates being relative to this start date.

3.4 Performance Indicators

The below table will not be applicable until the first year of the project is reached but we are already approaching these numbers. The current numbers of pages of biodiversity literature available in the repositories of our current content providers already exceed the 17m pages specified below (currently about 18.5m pages).

Analysing the current usage of the BHL Portal, the expected pageviews of 1m users in one year is a very conservative measure. In the first 3 months of 2009, 650,000 pages were viewed. These numbers are still growing and the Portal is actually well established and used.

The biodiversity literature we have will be ingested by Europeana from February 2010 onwards. However, the actual display of the content will come with the Rhine release of Europeana in July 2010. Thus, the Europeana users will not see BHL-Europe content before July 2010. Therefore, no user statistics will be available in April 2010 (end of year 1 of BHL-Europe). This is due to the workplan of Europeana. This will be explained in more detail in our annual report.

Indicators	Expected vs Actual					
	Year 1 (exp.)	Year 1 (act.)	Year 2 (exp.)	Year 2 (act.)	Year 3 (exp.)	Year 3 (act.)
Number of accessible pages of biodiversity literature (in million)	17		21		25	
Percentage of literature available through Europeana	20%		50%		100%	
Number of interconnected repositories	7		20		30	
Number of content providers	20		25		30	
Number of portal languages	1 ¹		7 ²		12 ³	
Page views through BHL Portal (in million)	1.0		2.0		5.0	
Page views through Europeana Portal (in million)	0.25		1.0		3.0	
Case studies of successful usage of the material by non-scientists	5		10		15	
Agreements with Rights Holders / Publishers	2		4		6	

¹ English; ² + French, German, Italian, Portuguese, Spanish; ³ + Danish, Dutch, Czech, Hungarian, Polish, Swedish

4 Awareness and Dissemination

4.1 Overview of awareness and dissemination activities

The activities and dissemination mechanisms envisaged to raise user awareness are explained in more detail in the BHL-Europe dissemination plan (D.5.4, D5.5). The BHL-Europe project Web site has been online since July 2009 and already has a first set of information available in English language. More languages will come online (German, French, Spanish, Czech) and also more content will become available as BHL-Europe produces more results. The BHL-Europe Newsletter will be published quarterly and contains information about the project, the project related activities of the consortium partners, meetings and event, and updates of projects of the BHL-Europe network, in particular BHL and Europeana. The promotion kit is already in use and will be developed during the next months.

In the first months of the project, BHL-Europe was active in approaching various target users through official conferences and publications. Depending on the audience, the presentations or posters were adapted accordingly. We also sent the relevant expert from the consortium depending upon the the topic of the conference, in order to adequately represent the project. The following of our target users are covered by the events listed below: Libraries, Digitisation centres, Digital library / Open Access networks, Students of different levels (primary to academic), Scientists (e.g. History, Cultural heritage), Scientists (e.g. Biology). With our press release in May we targeted the general public. More press releases and other publications relevant for the public will be developed as the project progresses.

BHL-Europe carried out a user requirement survey at the end of the reporting period. This was an initial step with particular focus on the users employed in the institutions of the consortium partners in the project. The aim of this survey was to define the basic functionalities and services of the BHL-Europe portal with particular focus on the user needs. Future surveys will target a broader audience in order to cover all important target users. These surveys will be accompanied by target user analyses. We will establish target user groups working with the BHL-Europe prototype portal. Their work will be logged to provide direct information about usability and problems. These users will be interviewed directly to get a direct feedback.

BHL-Europe is currently undertaking dissemination activities in order to reach the broader public. In that context we are cooperating with Europeana in a joint PR campaign next year. The Rhine release with the first major contribution of biodiversity heritage literature is simultaneous with the International Year of Biodiversity. During the public biodiversity events planned for 2010, we can present the use of Europeana and BHL to, for example, the public interested in regional wildlife. A meeting is already scheduled for April 2010 in Edinburgh to coordinate these activities.

The BHL colleagues in the USA currently are very active in social networking. They are also linking animal and plant descriptions in Wikipedia to the BHL Portal, which significantly increase the use of the Portal. It can be assumed that most of these users coming to BHL via Wikipedia are non-scientists. BHL-Europe needs to align the activities in this field with the BHL-colleagues to further increase the use of the Portal.

4.2 Events and meetings

List of Events & Meetings

Below is a list of meetings and workshops of members of the BHL-Europe team in order to discuss and work on various aspects of the project to build consensus on architecture and specifications.

Name	Location	Start Time	End Time
BHL-Europe Kick-off meeting	Germany	10/05/2009	15/05/2009
Side Meeting during the e-Biosphere Conference	UK	03/06/2009	03/06/2009
Work-Meeting in Paris (Atos, BnF/MNHN)	France	08/06/2009	09/06/2009
PASIG Summer Meeting	Malta	24/06/2009	26/06/2009
BHL-Europe general & workgroup meeting	The Netherlands	10/08/2009	12/08/2009
BHL@STL	USA	17/08/2009	21/08/2009
WP2-AIT Meeting	Austria	03/09/2009	04/09/2009
Side Meeting during Europeana Plenary	The Netherlands	14/09/2009	16/09/2009
BHL-Europe architecture meeting	Germany	29/09/2009	29/09/2009
BHL-Europe technical management meeting	Denmark	13/10/2009	13/10/2009
BHL-Europe MoU meeting	Belgium	22/10/2009	22/10/2009

Below is a list of conferences and events attended by members of the BHL-Europe team in order to disseminate the project to the various audiences or to collect important information for the progress of the project.

Partner	Title	Start date	End date	Country	Level of dissemination
UH-Viikki	Synergy afternoon of the Finnish National Digital Library	20/05/2009	20/05/2009	Finland	panel discussion
UH-Viikki	Three Liaison Committees of Viikki Science Library	29/05/2009	29/05/2009	Finland	presentation
NHM	International Conference on Biodiversity Informatics	1/06/2009	3/06/2009	UK	flyer
UH-Viikki	Three Liaison Committees of Viikki Science Library	2/06/2009	2/06/2009	Finland	presentation
UH-Viikki	Three Liaison Committees of Viikki Science Library	8/06/2009	8/06/2009	Finland	presentation
MSN	EIPub2009	10/06/2009	10/06/2009	Italy	presentation
AIT	eSciDoc Days	15/06/2009	16/06/2009	Germany	N/A
MfN	EDIT meeting "Scientific publishing in natural history institutions"	22/06/2009	22/06/2009	Slovak Republic	presentation
RBINS	Willi Hennig XXVIII Annual meeting	22/06/2009	26/06/2009	Singapore	personal talk

CSIC	Primer Encuentro Ibérico de Biología Subterránea	10/07/2009	12/07/2009	Spain	flyer
CSIC	I Iberian symposium on geometric morphometrics	23/07/2009	25/07/2009	Spain	presentation, discussion, flyer
MfN	Systematics 2009	10/08/2009	14/08/2009	The Netherlands	presentation
AIT	Workshop on Advanced Technologies for Digital Libraries 2009	8/09/2009	8/09/2008	Italy	presentation
HNHM	27 th Annual Meeting of the Natural History Curators of the Carpathian Basin	8/09/2009	11/09/2009	Slovakia	presentation and discussion
CSIC	XXVI Jornadas de la Asociación Española de Entomología	12/09/2009	15/09/2009	Spain	poster and flyer
CSIC	International Symposium on Islands and Evolution	14/09/2009	17/09/2009	Spain	presentation, flyer
AIT	30. Österreichischer Bibliothekartag	17/09/2009	17/09/2009	Austria	presentation
RBINS	BHL-Europe presentation at RBINS	24/09/2009	24/09/2009	Belgium	presentation
NMP	Athena project workshop	30/09/2009	30/09/2009	Czech Republic	presentation
UBER	DINI Jahrestagung	30/09/2009	1/10/2009	Germany	poster
UH-Viikki	Two membership meetings at UH-Viikki	1/10/2009	31/10/2009	Finland	presentation
MfN	79. Jahrestagung der Paläontologischen Gesellschaft	5/10/2009	7/10/2009	Germany	poster
UBER	Open Access Days	7/10/2009	8/10/2009	Germany	poster
NHM	SUN PASIG	7/10/2009	9/10/2009	USA	presentation
MfN	Digital Libraries and Digital Preservation: ICT Call 6 Information Day	9/10/2009	9/10/2009	Luxembourg	poster
MfN AIT	Improving Access to European Cultural Heritage – goals, ways, and means , Round Table event for aggregators, Europeana	13/10/2009	16/10/2009	Sweden	flyer, personal talk
MfN	2nd LIBER-EBLIDA Workshop on Digitisation of Library Material in Europe	19/10/2009	21/10/2009	The Netherlands	presentation
NHM	TDWG - Taxonomic Database Working Group yearly conference	9/11/2009	13/11/2009	France	poster
AIT	14th international congress "Cultural heritage and new Technologies"	17/11/2009	17/11/2009	Austria	presentation
NHMW	LAPI III, annual meeting of the Latin American Plants Initiative	30/11/2009	4/12/2009	Colombia	flyer, poster
CSIC	4 Congreso Ibérico de ornitología (SPEA)	5/12/2009	8/12/2009	Portugal	flyer
MfN	EDIT General Meeting 2009	15/12/2009	17/12/2009	Portugal	poster or talk

5 Conclusions

For the first six months of the project, BHL-Europe is well on track. Communications among partners are effective, the work is being carried out in a timely fashion and there are currently no barriers foreseen that can be expected to significantly delay the major outcomes of the project. We do look forward to the next period that will deliver tangible initial project results.