PROJECT DOCUMENTATION

INTERNAL INTERIM PROGRESS REPORT

Project: BHL-Europe

Release: Final version

Date: 17 October 2011

Author: Jana Hoffmann

Table of Contents

Contributors

Person	Partner
Jana Hoffmann	MfN
Henning Scholz	MfN
Melita Birthälmer	MfN
Lola Obajuluwa	NHM
Chris Sleep	NHM
Jiri Kvacek	NMP
Tine Mallaerts	NMP

Circulated to the consortium for review: 21 September 2011

Internal Interim Progress Report

1   Purpose

To report on a quarterly base the status of work for each member of the consortium. This report will not duplicate the content of the official progress and annual reports requested by the EC every 6 and 12 months, respectively.

2   Follow-ups from previous reports

For further information on past activities we refer to [Progress Reports](#)[1] D1.1, D 1.2, D 1.4 and D1.6.

3   Activities

General Remark

The following activities reported here are taken from the status report of the following partners:

MfN, NHM, AIT, ATOS, FUB-BGBM, UGOE, NHMW, LANDOE, HNHM, NAT, NBGB, RBINS, BnF, MNHN, CSIC, RBGE, UBER, EDLF, BHLUS (MOBOT, SIL), RMCA, Sp2000, UH-Viikki, NMP.

Not submitted reports:
MIZPAS, UCPH, MSN

No work:
Wiley

This time, collection of status report took much longer than expected, which results in a delay of the Interim report #5, and only 19 partners submitted a report in time. The timely production of the reports needs to improve in the next reporting period. Furthermore, the quality of some of the status reports has to improve in terms of submission of narrative reports rather than only keyword or bullet points. It would also be helpful if partners would assign their tasks to the task (numbers) described in the Description of Work (DoW).

WP1

1.1.1 - Resource planning, monitoring and controlling
The project coordination, project partners and BHL-Europe associates followed the project progress on the BHL Wiki. They used the information provided for planning tasks and meetings as well as for the preparation of their reports.

Farewell to BHL-Europe colleagues
Nancy Chillingworth – The BHL-Europe Work Package 4 was officially closed end of April thus Nancy's contract ended for BHL-Europe. However, she is still located at the NHM and will be available for IPR related questions.

---

[1] https://bhl.wikispaces.com/BHLE_WP1_reports_Progressreports

Adrian Smales – Adrian's contract ended end of April and Work Package 3 lead was handed over to Graham Higley, Lola Obalujuwa and Chris Sleep.

Lesa Ng - BHL-E Project Worker at RBGE completed her nine months contract at the end of April 2011.

1.1.2 – Liaison between the Commission, consortium partners, EUROPEANA, BHL, and external experts; effective communication with the consortium members, Work Package leaders, the Commission, and interested parties; coordination of meetings and progress reviews.

Project Management Group conference calls (PMG calls)
PMG calls were held on a regular basis in this reporting period, in particular in order to plan the 2nd Annual Review Meeting. The minutes from all PMG calls are provided on the BHL Wiki[2].

2nd BHL-Europe Annual Review Meeting in Tervuren/ Meise, Belgium
The planning of the 2nd BHL-Europe Annual Review Meeting already started in January 2011.
It was agreed that the meeting will be organised by NBGB in collaboration with RMCA and held at two different places, in Meise and Tervuren, Belgium. Next to the organisation of logistics the preparation of the tight meeting's agenda including organisation of the review meeting and multiple parallel workgroups was a time consuming and challenging task. It required a high discipline and perfect level of alignment in order to use the time slots efficiently. In preparation of the actual review meeting the WP leaders had to align their presentations in terms of content and layout in order to give a consistent presentation of the project's outcomes during the Review Meeting. The workgroup session before and after the actual Review Meeting intended to cover all aspects from content management over portal development to sustainability.
The meeting started 7 June in Meise (NBGB) with a PMG meeting for preparation of the official Review Meeting. In the afternoon, two parallel sessions took place: (1) collection policy and (2) evaluation and testing. The meeting then continued 8-10 June in Tervuren, Belgium, at RMCA. The official review meeting on 8 June was attended by members of 24 BHL-Europe partners. It was a very successful review and BHL-Europe received a green flag. This means that BHL-Europe can continue the work of the project as planned. The reviewers were satisfied particularly seeing that we take all their advice serious and considered many of their comments. However, one deliverable needs to be amended and re-submitted to provide a more detailed content ingest plan to Europeana and the BHL-Europe portal in the third project year. The payment of the 3rd pre-instalment depends on the approval of this updated deliverable (D 2.7). In the morning of 9 June two parallel workshops were held, one dedicated to dissemination activities with a particular focus on the virtual exhibition, and a second workshop for our content provider to demonstrate the metadata mapping tool and the Pre-Ingest tool. In the afternoon there were three parallel workshops: a content provider workshop on the GRIB, a business development workshop to investigate gaps in the existing D1.5 document for further development into D1.9, and the first part of the portal development workshop for the development team (portal review). The outcomes of the workshops were presented to the entire consortium in the early evening. The next day, 10 June, the portal development workshop continued. A parallel workshop was held to plan the next steps in finalising the Best Practice Guide for M36 of the project. The meeting was closed by a TMB session to discuss the last open questions and existing issues towards the Public Core Release in M27 of the project (August 2011) with the developers and decision makers in the project. All details including agenda, notes and participants are available on the BHL Wiki[3].

---

[2] https://bhl.wikispaces.com/BHL-E_PMG_calls
[3] https://bhl.wikispaces.com/BHLE_AnnualReview2

Preparation of the Public Core release meeting in The Hague, Netherlands
The next BHL-Europe meeting will take place from 30 August to 2 September 2011 in The Hague, The Netherland, and is hosted by Europeana at the [Koninklijke Bibliotheek](#)[4] (KB). A meeting's Wikipage was set up with a preliminary agenda, directions and a tentative participants list. The meeting will host the PMG meeting, a business plan workshop with Europeana, the Portal Demo, Review and Testing meeting, a TMB meeting and a CWG meeting. All relevant information can be found on the [BHL Wiki](#)[5].

1.1.3 – Production and consolidation of periodic external reports; including costs statements; internal quality assurance

Financial Statements
The financial statements of all project partners need to be submitted after the end of the second project year together with the request for the 3rd pre-instalment. The project office prepared and updated the [guidelines](#)[6] and planned the process to allow a timely delivery of all documents. However, the administrative process is very complicate and slow. Thus, the financial statements are not yet finalised. The project coordination is still waiting for the final documents to be approved by the partner's local administration department and the project office.
This year, we are in a much better position to really check and validate the financial statements compared to last year due to the improved reporting process in BHL-Europe (activity reports). The project assistant checked all financial statements against the monthly activity reports to identify inconsistencies in personnel commitment, for example. This supports all partners in validating their internal project documentation.
After the submission of the financial statements, we will continue with the budget planning for the final year of the project to make sure we spend the available budget efficiently until the end of the project in April 2012.

Production of reports and internal quality assurance
In this reporting period the activity reports were collected from the consortium partners and send out for approval to the Work Package leaders.

Quality assurance (Risk and Issue management)
Risk and issues of the project were collected with the status report from each partner and completed by the respective project manager or coordinator. The risk and issue log of the project was discussed and evaluated during the PMG calls and appropriate counter measurements were decided.

Periodical internal and external reports: D1.6 Progress Report and D1.7 Annual Report
The progress report D1.6 and Annual Report D1.7 was finalised in this reporting period based upon the two status reports send by each partner. The progress report was published for review by the entire consortium and an external reviewer. We included valuable feedback from consortium partners. A draft version of the report was sent to the EC. The finalisation of the financial statements and the resource table is still in process.

1.2.2 - Project management
Some specific project management work was required in the last reporting period. End of April the contract of our WP3 leader Adrian Smales at NHM ended and in month 24 of the project WP4 was officially closed as planned according to the DoW. In order to plan all related actions in more detail an additional meeting was held in London 7 April with the NHM team and the project coordinator. It had to be assured that before closing WP4 all relevant tasks were carried out and

---

[4] http://www.kb.nl/
[5] https://bhl.wikispaces.com/BHLE_Public_Core_Release_Meeting
[6] https://bhl.wikispaces.com/BHLE-FS

Work Package results were documented properly. Furthermore, it was discussed how to proceed with IPR question coming up during the final year of the project. Nancy Chillingworth, the WP4 leader, will continue to work at NHM and agreed to act as a consultant for IPR questions during the last year of the project. Nancy will also continue to work for the Legal Working Group of the CCPA, which facilitates the follow up of legal aspects related to Europeana. This will be of particularly interest when the Europeana Data Exchange Agreement needs to be finalised. Furthermore, handover activities of WP3 management tasks from the current Work Package leader, Adrian Smales, to the NHM project management team and future roles and responsibilities were discussed. Graham Higley will act as the official WP3 leader for the transition period until NHM has hired a new head of IT. Chris Sleep will be responsible for managing all technical aspects related to the development of the BHL-Europe system and Lola Obajuluwa for the overall WP3 project management. This assures a smooth continuation of all of the WP3 management tasks in this important period of the project.

1.2.3 - Production of deliverables
In total nine deliverables were due in month 24 of the BHL-Europe project. It was a major management effort for the project office to organise the work between all involved WPs and team members. The final compilation and review of all documents by the project office was successfully finished in time to deliver all documents to the external reviewers and the EC as well as for the preparation of the Annual Review Meeting.

1.4.3 - Business development
In project month 36 BHL-Europe has to present a business plan for long-term sustainability (D1.9). The more general plan (D1.5) was further elaborated during the Business Development Workshop in Tervuren. A survey was planned to collect information from our content providers on their future perspective in providing content to BHL-Europe after the end of the project.

1.5.1 - Networking activities (Europeana and Natural Europe)
The second Europeana Group of Projects meeting was held on 20 May 2011 in The Hague, Netherlands, and was attended by 13 projects. A key point on the agenda was the discussion of the "Project share space" to be built over the summer 2011. This web space will bring together content from all of the Europeana Projects to facilitate knowledge exchange and communication. BHL-Europe is participating in the pilot project to set up this exchange service. The BHL-Europe project coordinator will coordinate these activities. An afternoon session was dedicated to a "lessons learnt" session and small workshops to discuss issues and various options. The dissemination session was particularly important for BHL-Europe as it showed that Europeana is the public end-user space for all content. Other project's sites or portals are either for their professional users, or, if end-user directed, for a scientific audience. Thus, it is not a primary task of our dissemination activities to reach out to the public. More collaboration with Europeana is certainly beneficial to cover this important user group, the public users, more efficiently.
The [Natural Europe](#) [7] project was launched in October 2010 in the framework of the ICT-PSP programme. Natural Europe aims at placing museum visitors in the centre of an exhibition that is both physical & virtual using technology, services and learning content to provide personalized; appealing educational experiences, changing the way museum education is perceived and conducted, worldwide (Natural Europe vision taken from the 1st Natural Europe Newsletter). The project coordinator is currently investigating on what level a collaboration of BHL-Europe with Natural Europe is possible. Knowledge transfer and data exchange are two possible fields of collaboration. A face to face meeting is required to explore all options in more detail.

API Email Europeana
The newly launched Europeana Search API and Search Widget API were circulated in the BHL-Europe consortium. They enable integration of European cultural and scientific heritage in content

---

[7] http://www.natural-europe.eu/

provider's online environments in a format that is convenient and meaningful to end-users. These Europeana API services can help to successfully engage online audiences by delivering richer and relevant content to them in new, intuitive, visually appealing ways. The improved user experience translates into an increase in high-quality traffic, active online communities and stronger brand association. At the moment these services are only available to Europeana network partners having an own portal. More details can be found here: API website[8].

Lizzy Komen (EDLF) presented the integration of the Europeana API by RMCA[9] as a best practice example of Europeana API integration during the BHL-Europe Annual Review Meeting.

WP2

2.1.3 - IT Development

D2.5 Final and enhanced Web-database for content management and collection analysis
Support has been given to Boris Jacob for D2.5 who finalised the document at the beginning of this reporting period. The document has been reviewed a first time before sending it to the consortium and to the external reviewer and a second time before finally handing it over to the project coordinator. A major task of work package leader was to assure the timely production of this deliverable. D2.5 serves currently as a manual for the GRIB and describes the functionalities of the system.

GRIB
The Digitisation Management Widget of the GRIB has been finalised during this reporting period. The GRIB was planned to be up and running for the BHL-Europe Annual Review meeting, but due to technical difficulties at the VZG, the GRIB couldn't be put into production for the BHL-Europe partners in time. The technical issue could be identified but not solved by the end of the reporting period. To date the VZG is solving the issue and plans to have the GRIB finally up and running reliably in September.
NBGB installed the Z39.50 protocol in order to facilitate harvesting by the GRIB.
During April RBGE created further bibliographic records of mycological and phytopathological books for inclusion into the Global References Index for Biodiversity.

Scanlist
The Scanlist is running in parallel to the GRIB to ensure a smooth change from the Scanlist to the GRIB until the GRIB is 100% stable. During this transition period the Scanlist is still maintained.

2.2. - Analysis of domain content

D2.7 Content analysis and management status report 2
At the beginning of this reporting period the writing, layout, and internal and external review of this deliverable took place and occupied most of the time.

2.2.3 - Collection Policy
The discussion on a BHL-Europe collection policy has been initiated at the beginning of this working period. A collection policy should help to clarify the question how we are going to organise the content contribution and prioritisation of content for BHL-Europe. The goal is to achieve a consensus with the content providers.
For this purpose this task has been given to Michaela Hierschläger from LANDOE in order to gather ideas from our content providers on our BHL Wiki[10]. Furthermore a collection policy

---

[8] http://version1.europeana.eu/web/api
[9] http://bit.ly/AfricamuseumEuropeanaAPI
[10] https://bhl.wikispaces.com/BHLE_WP2_CoPo

workshop was held during the Annual Review Meeting to reach a consensus on the BHL-Europe collection policy. This is an ongoing task for the next reporting period. At the end of this reporting period a first version of the collection policy has been sent out for review to the attendees of the collection policy workshop. To date the collection policy is in the process of finalisation.

2.3 - Management of the content acquisition process

This Annual Review Meeting has been used to provide detailed information related to content and train content provider. Therefore one day was planned for workshops related to content provider's tasks. Three different Content Provider Workshops were planned to take place during this day as followed:

- Pre-ingest
- Mapping
- GRIB

The *Pre-ingest workshop* aimed at giving an introduction to the Pre-ingest tool developed by AIT within WP3. Therefore, AIT has been asked well before the meeting to provide a step-by-step tutorial in order to instruct all content providers how to use the pre-ingest tool for the ingestion of content after content has been uploaded to the NHM server and to discuss examples regarding the pre-ingest file submission guidelines. Unfortunately, the pre-ingest tool is currently not available on the NHM server and thus our content providers could not be instructed how to use the tool to trigger the ingest process. The *Mapping workshop* aimed at introducing the developed mapping tool - developed by NHMW – to the content provider in order to give them the opportunity to do the mapping of their metadata to the OLEF schema (= BHL-Europe schema) autonomously. NHMW provided a well structured introduction to the mapping tool including live examples. The *GRIB workshop* was a continuation of previous workshops and aimed at giving an update on the GRIB development and introduction of its functionalities and use. This workshop was held by Boris Jacob in cooperation with Andreas Krausz from the Head office of the common library network GBV (VZG). The content providers were informed about progress in the GRIB development and all functionalities available at that time have been presented to the content providers. Further, content providers have been asked what their expectations are after the project life time. One major need indentified, is to keep up the communication in the community of natural history librarians. It was suggested to have a technical system, like a help desk with forum functionalities, which supports an easy and fast communication between librarians.

The management of the current content providers is an ongoing task within WP2. During this reporting period a lot of time has been invested into communications with our current content providers and the WP3 team (AIT, NHMW) in order to coordinate the upload of content and feedback for content providers regarding their content. Each content provider is different and has own needs and therefore communication with each of our content provider was and still is necessary.

More up-to-date information about the content management can be found on the BHL Wiki. [11]

There you find a documentation of the current status in order to inform our partners as well as the technical team, to make the content management process more visible and to facilitate communications and future planning of their tasks.

2.3.2 – Assist partners in implementation and evaluation of scanning operations; discuss and distribution of data standards and specifications; attracting new content providers

Analysis and preparation of (meta)data for Pre-Ingest

This chapter gives an overview over work done in this reporting period regarding analysis and preparation of scanned material and the corresponding metadata. It is separated in paragraphs by partners to facilitate reading.

---

[11] https://bhl.wikispaces.com/Content+_Management

BnF

BnF worked on the selection of documents for digitization for BHL-Europe and checked the quality of already digitized pages. In more detail, they took a look at the *Annales de la Société Entomologique de France* ranging from year 1938 to 2002. In total 29.523 pages have been quality checked. Furthermore, the data transfer to BHL-Europe has been technically supervised. The transfer of BnF content to BHL-Europe had to be initiated twice as errors occurred during the transfer. According to the check up of the already transferred content by BnF, one directory is still missing after the second transfer. However, because of the immense data volume BnF cannot find out which directory is missing using their FTP account and is waiting now on the ingestion of their content to the BHL-Europe portal in order to check the availability of content online. Subsequently, they will be able to provide the missing directory to BHL-Europe.

BnF is currently preparing their second delivery of content to BHL-Europe in September 2011.

CSIC

CSIC worked on metadata for already digitized content, uploaded several officially published and documents considered as "grey literature":

- "Unidad de Biodiversidad", a book on biodiversity that has been produced by the FECYT (Spanish Foundation of Science and Technology) for teachers at the high school level.
- Celso Arevalo, 1929, "La vida en las aguas dulces", the first limnological monograph in Spanish.
- Three volumes of the "International Symposium and First World Congress on the Preservation and Conservation of Natural History Collections".
- "Iberus", vol. 22.2 (as it was not available in BHL-US).
- Particular mention deserves the completion of ingestion of the Spanish Malacological Journal "Iberus", which was only partially available in BHL-US. Now, taking BHL-US and BHL-Europe together, the journal is complete as following parts of Iberus have been provided by CSIC: vol. 24.2, vol. 25.1, vol. 25.2, vol. 28.1, and vol.28.2.
- "Memoria presentada para optar a la plaza de conservador del Museo Nacional de Ciencias Naturales", the first natural history monograph on museums curators in Spanish (regarded as "grey literature", A.G.Valdecasas, 1984).
- Furthermore, CSIC got in touch with following copyright owners:
- Maria Soledad Vicente Rosillo, co-author of the five volumes of "Catálogo de los fondos especiales de la Biblioteca del Museo Nacional de Ciencias Naturale".
- Dr. Borja Sanchiz has been contacted in order to digitize "Primer Congreso Español de Biogeografía", 1978, Universidad Complutense de Madrid, a Ph.D. thesis on Biodiversity for BHL-Europe.

To date, no final decision was received in any of those cases.

MNHN

After numerous checks, reviews and feedbacks a template (Excel sheet) for the content indexing has been validated by the MNHM team and by their digital library provider (the vendor of portal). Then, MNHN sent the template to their digitisation and indexing service provider. First, a small batch of 50 books has been ordered for tests and adjustments. The main difficulty is that each collection of books or journals has its own specifications (for example two rules of numbering for some serials, missing TOC, etc.). Thus MNHN has to point out the particularities of the collection to the service provider each time, before the service provider is able to start the work. The tests ended in April and the production has been started in May.

In May, MNHN updated and validated their OAI API for Gallica. This API can also be used by BHL-Europe to harvest the metadata of the ~200 books that are already available at the NHM server. 200,000 pages in tiff and the 200,000 pages in jpg, recently scanned, are ready to be uploaded to BHL-Europe. However for the moment, they only can provide metadata in UNIMARC on journal title level. Besides, MNHN has started to receive the template for the content indexing. This is an ongoing process and samples of the Excel file will be sent to the BHL-Europe ingest

team. MNHN is waiting on the ingestion of their content to the BHL-Europe portal in order to get a feedback how their content is displayed in the portal before uploading further content.

In April, the analysis and description of a first batch of selected new content finished (Flores, Index seminum, History of the Botanical Garden and of the MNHN). 28,500 new pages are ready to be scanned in future, which however still depends on funding.

MNHN has been contacted by some other institutions about its participation in the BHL-Europe project and its other digitisation projects, such as the libraries of University of Renne, the École des Ponts Paris Tech and the Interacademic library of Health in Paris. Communication between the WP2 lead and the University of Renne are currently in progress.

NATURALIS

Naturalis started a new large scale digitisation project. At the start of this project a lot of attention was paid on selecting the right materials for digitisation, checking the availability in BHL (Europe), and setting up the work processes so that the results can be uploaded to BHL-Europe.

NBGB

NBGB tested the FTP upload to the BHL-Europe server. Further work was image quality control and checking metadata which are currently in process.

RBINS

RBINS discussed with their library personnel which digital publication they can provide to BHL-Europe and defined priorities. Furthermore, they uploaded ~ 2000 pages as test content to BHL-Europe. The RBINS metadata is not in a standard format and thus the preparation of the mapping takes more time and is still in process.

RBGE

In April 2011 *Notes from the Royal Botanic Garden Edinburgh v.1-46 1900-1990* and *Flora of Bhutan Parts 1(1)-3(2) 1983-2001* were uploaded to the BHL-Europe server. This content has been harvested by Europeana, and quality checked by RBGE.

Following discussions and a meeting with staff at the National Library of Scotland (NLS), arrangements were set in place for staff operating the Internet Archive station at NLS, to create digital versions of selected books from the Library. The following titles have been scanned and made available:

- Sutherland, James. Hortus medicus Edinburgensis… (1683)
- Alston, Charles. Index plantarum… in horto medico Edinburgensis (1740)
- Hope, John. [Dr. Hope's elementa botanici] (178?); Alston, Charles. A third dissertation ion quick-lime and lime-water. (1757)
- Fletcher, H.R. & Brown, W.H. The Royal Botanic Garden Edinburgh 1670-1970. (1970)
- Alston, Charles. A dissertation on botany. (1754); Braithwaite, M.E. and Long, D.G. The botanist in Berwickshire. (1990)
- The Botanical Society of Edinburgh 1836-1936. (1936). Permission was granted by the Berwickshire Naturalists' Club and the Botanical Society of Edinburgh, to include the last two titles. These books will be provided to BHL-Europe.

Furthermore, some work on rights clearance was done, because of permission from external organisations to arrange for digitization of three books using Internet Archive station at National Library of Scotland. This has not been charged to the project.

LANDOE

LANDOE increased the amount of content delivered to Europeana. To date they supply slightly more than 3500 books to the Europeana portal. The metadata for EUROPEANA is still provided through an OAI-PMH interface.

LANDOE started the content upload to the BHL-Europe server and nearly finished this task within the reporting period. This initial upload took more than 3 weeks, all digital books that are available in Europeana have now also been provided to BHL-Europe for ingest to the portal. Missing

metadata will be delivered until July/August 2011. In the meantime they have digitised and metadata enhanced 600 additional books. This material – including more books in the current digitisation process - will be provided for the BHL-Europe Portal and Europeana end of 2011.

HNHM

Following bulk digitisation of HNHM journals, HNHM is working further on the compilation and key-in of the bibliographic journal article level metadata sets (xml format). The splitting of scanned journal volume to article level was continued. The following journals' metadata and quality assurance and database conversation of the following journals' data were completed:

- Anthropologica Hungarica,
- Crania
- Parasitologica Hungarica.

The compilation of Acta zoological hungarica metadata was also started. HNHM is working on a continuous update with recent volumes of journals in order to submit more new content as requested by users.

Because of the change in their workflow all formerly uploaded content had to be resubmitted. In order to facilitate this process - as HNHM is providing a huge amount of content on article level - scripts were created to extract data from their DB (MySQL) to MARC21 format (XML). A further shell script was developed for creating the necessary library structures and facilitating the data upload to BHL-Europe.

Live demonstration and negotiation started with the regional natural history museums and the Hungarian Biological Society in order to incorporate their biodiversity literature into the BHL-Europe corpus. As soon as official agreements are available, HNHM will provide further content to BHL-Europe.

UGOE

In collaboration with Naturalis Leiden (Tom Gilissen) a list of important historic zoological books to be scanned by Naturalis digital library was extracted from AnimalBase resources. These were books that had been scheduled for digitisation in Göttingen, Germany, but which could not be scanned either due to technical reasons or due to lack of funding. The list mainly derived from detailed analyses of Sherborn's Index Animalium published between 1902 and 1931, where all literature was listed that contained new taxonomic names in zoology. It was checked if the works were digitised somewhere else, and if not, they were included in the list. The first part of this list contained 140 books published before 1780 and was sent to Naturalis in June.

RMCA

RMCA is continuously working on selecting literature to be scanned, coordinating the scanning process and updating their local Scanlist[12], which is a common list for the three Belgian BHL-Europe partners: RMCA, NBGB, and RBINS.

From April to June more than 15,000 additional pages of biodiversity literature from the zoological department were digitized in the framework of other projects and in-house activities. All files were controlled and renamed on page level. Furthermore, metadata processing is one of the important tasks in preparation of the content for the ingestion. This includes the input of new data, the enrichment of catalogue data with additional information and the extraction of metadata from the LIBIS library system.

Public domain literature is continuously uploaded to the NHM server. The first 52 items (~24000 pages) have been harvested by Europeana so far. Routine copyright clearance procedures are one of RMCA's tasks, too. In the last three months RMCA was able to obtain 3 signed BHL-Europe agreements.

---

[12] http://193.190.223.46/wiki_ext/index.php/BHL

<u>UH-Viikki</u>
During this reporting period the first part of digitized content were checked and processed for inclusion into Europeana. Currently 30 items from UH-Viikki are available in Europeana. In May the second part of publications were handed to our digitization provider Vaasan Micro Copy and were received in June.
UH-Viikki received the signed BHL-Europe Licence Agreements from two Finnish biological societies, Biological Society of Finland Vanamo and Societas pro Fauna et Flora Fennica. The agreements allow UH-Viikki to digitize publication series of these societies until 1940. The licence used in these agreements is Creative Commons Attribution-NonCommercial-ShareAlike 3.0.

Assist partners in implementation and evaluation of scanning operations
UBER continued the work on a Goobi test server for our content providers.

Best Practice guidelines and standards
One task within WP2 is assisting partners in implementation and evaluation of scanning operations and to deliver a first draft of the Best Practice Guidelines and Standards. This delivery aims to give a first standardized guideline for Europe. The approved best practice guidelines and standards shall be understandable for every person working with it. They shall guide new and current contributors in a simple and clear way through a digitisation workflow from the either print or digital version of an item to the digitised representation of that item in the BHL-Europe portal. The guide will describe the minimum required standard for BHL-Europe and where possible specifying the ideal standard. The first version of this guide was due at the end of the second project year, 30th April 2011. During this reporting period the first version has been finalized and handed over to the project coordinator. However, the work on the Best Practice guidelines will continue. The preparation for the final version started during this reporting period in cooperation with our partner LANDOE. A third workshop has been organized and held during the Annual Review Meeting in June. All contributors for the first version have been invited to take part in this workshop and give feedback on the first draft and how to move forward based on the final document. Contributors have been instructed how to review the chapters for the final version in order to facilitate reading by users. Further missing information have been identified and contributors have been assigned to the respective chapters.
More detailed information about the workshop and outcome can be found on the [BHL Wiki](https://bhl.wikispaces.com/BHLE_WP2_BPG)[13].
After the workshop several partners were asked to contribute to the final Best Practice guidelines and standards. NBGB already described their internal digitisation workflow as an example for the document.

Expansion of the content provider network
Attracting new partners until the end of the project is another major task within WP2.
During this reporting period several potential content providers have been contacted for acquiring new content to BHL-Europe. Communications with prospective content providers requires a lot of time and preparations in order to answer various occurred questions, for instance questions regarding IPR, metadata, OCR or the MoU. This is an ongoing task.
Furthermore, the "BHL-Europe in a nutshell" document has been translated into German by our partner LANDOE in addition to the already existing versions in Spanish, English, French and Russian. Further information about content acquisition, e.g. the "BHL-Europe in a Nutshell" document can be found on the [BHL Wiki](https://bhl.wikispaces.com/Content+_Management)[14].
The BHL-Europe project office at MfN is involved in planning the full implementation of the newly purchased scanner. As the scanner will not only be used for books but for many other applications as well, priorities need to be defined.
In May 2011, a digitisation project of the university library Frankfurt/Main was approved to be funded by the German Research Foundation (DFG). BHL-Europe was significantly involved in the

---

[13] https://bhl.wikispaces.com/BHLE_WP2_BPG
[14] https://bhl.wikispaces.com/Content+_Management

project design and the initial planning process through the PCO. It is a great success for BHL-Europe that funding was received. MfN is currently discussing the final project plan and supports the digitisation project through BHL-Europe WP2.

Coordinate with Europeana

Coordination with Europeana took place with Lizzy Komen to ensure that new content will be ingested to Europeana in time.

There is ongoing involvement of BHL-Europe in the work of the Europeana Council of Content Providers and Aggregators (CCPA). A meeting of the core WG was held alongside the Open Culture[15] conference in Birmingham on 7 June 2011. This meeting focussed on agreements of the work plan for all WG and starting with the work. Within this reporting period the work plan for the Technical WG, where Henning Scholz as the CCPA secretary is the liaison officer, is finalised and the first white paper already drafted. Altogether 10 white papers will be prepared by the core group members of the Technical WG and presented during the CCPA plenary alongside DISH[16] on 6 December 2011.

WP3

3.1.1 - Management of the technical development team.

In this reporting period the technical development team have mainly focused on further Development and the Implementation phase of the technical and functional architecture of the final BHL-Europe System. The tech group has continued to hold intensive technical meetings through various forums (conference calls, Skype and TMB meetings) to discuss ongoing Development work within WP3/WP2. The main technology partners (AIT, Atos) have continued to work intensively on the BHL-Europe System architecture and on the implementation plan for BHL-Europe system. In preparation for the 2nd Annual Review Meeting in Tervuren AIT continued with the Portal development work and implementation of the Priority 1 BHL-Europe Core Features ahead of the next Key Milestone M28 – Public Core Release.

BHL-Europe Annual Review Meeting, Tervuren

Lee Namba (Atos) and Lola Obajuluwa (NHM) organised two workshop sessions on the Portal Development: - The first session consisted of the Evaluation & Review of Existing Portal Features that had been implemented so far. During this session AIT (Bernd Sproger) provided us with a step by step walkthrough of the existing portal features and their functionalities in the AIT work environment. Bernd demonstrated the BHL-Europe portal core features which have been further implemented and refined e.g. search queries can be saved, search results can be viewed in different layouts (list, grid), different XSL transformations can be applied to the results view, highlighting of search terms in the results views. This was the first opportunity to gather initial feedback from both users and developers that attended the workshop. At the end of the first session the demonstrated portal functions, especially Search and Retrieval, were evaluated positively and the implemented features match with the description in the CoR (Catalogue of Requirements) to a large extend. Three versions of the Advanced Search have been implemented:-

- A Google like search, where the user can search in the categories "with all words", "with exact phrase", "with any word" and "without words"
- A Europeana like search, where the user can dynamically add search fields and concatenate them with logical operators.
- A code word search, where the user can enter code words in order to make the search query.

In the first 2 searches there are also elements for selecting the metadata fields (ESE) to search in. The searches are configurable via a user configuration menu.

---

[15] http://www.openculture2011.org.uk/
[16] http://www.dish2011.nl/

The 2nd session was used to review the Priority 1 features yet to be implemented for M28 and review/prioritise the remaining work for the portal implementation.

On the last day of the Annual Review Meeting, the TechGroup held a Closed TMB session. During this session various topics were covered and decisions made on ongoing work within the Development process. Areas of focus include an update from Atos (Lee, Hengdi) on the Shared Development Environment which is now up and running. This Software environment strategy and planning will allow for an organised and harmonised shared development environment between our multiple development partners. Atos also provided a demo of the GitHub Issue Tracker which is being used for our Source Code and will allow for a distributed development between the various team members. Furthermore, now that we are at the development phase of the project in terms of the portal all Issues/Bugs have been migrated to the Issue Tracker. The Issue tracker will be used as a tool for tracking Issues, such as features, enhancements, bugs, etc. Other topics covered include Kick-off discussions to determine the scope of work required for our next deliverable D3.8 (M30) as well as planning for the Public Core Release meeting which is scheduled to take place at the end of August in The Hague.

3.1.5 - Development and adaption of specific tools; Implementation and adaption of TaxonFinder and name recognition tools; improvement and implementation of OCR techniques.

Development of the Pre-Ingest Tool

The deployment of the development version of the Pre-Ingest tool on NHM's servers has been successfully completed. The current deployed version is not yet configured and ready for Content Providers to access directly, however the deployment of the shared codebase via Jenkins will allow future developments to be automatically deployed and visible to the users, which will enable access for review and testing as soon as that can be supported.

WP3 recognise the impact that this availability for WP2 and Content Providers has on the dependant WP2 processes, and is working to resolve this issue as a priority. Content Provider logon access to the tool has been requested and this is due to be implemented shortly.

The Pre-Ingest source code has been re-factored to ease readability and understandability. In particular, the messaging configuration has been renamed to use terms which are widely known from package delivery services, since pre-ingest is used to package submissions and deliver them to other components for further usage. Consequently, refactoring was also necessary to prepare pre-ingest for additional unit testing using the spring frameworks test facilities and mock up objects. The pre-ingest backend has also been investigated to enable the multi-user capabilities of the web application by using session scoped proxied beans. This forms the basis for concurrent interaction with the tool and concurrent processing of data. The continuous integration platform, Jenkins has been used and tested for deployment, but since it was difficult to achieve a quick deployment strategy mainly due to complexities with user rights based on NFS shares, pre-ingest is deployed directly to bhl-mandible through a tunnelled connection. Furthermore, important optimisations of memory consumption have been introduced concerning the directory listings – this also speeds up browsing performance.

A tracking object and micro service has been implemented so that Pre-Ingest messaging becomes transparent for users. Also a UI component has been worked on so that the logs of pre-ingest are visible on the UI. A micro service has been created which is based upon the Aperture library and creates RDF/XML representation of a submission's file/folder structure. This was an unresolved issue which is needed to later recreate SIPs from AIPs, but also to determine which submission is linked to what files and folders. Tesseract 3.00 is being used to implement the OCR micro service. So each page will be OCR'ed during Pre-Ingest. A micro service has been created to include taxon finder (perl server) and taxon-finder web services (ruby). Each pages taxa are now recognized once the page's OCR'ed text is sent through the taxon finder micro service.

In cooperation with NHM London AIT also received VPN access to specific servers and ports and will continue to create a stable deployment mechanism for future development.

Collaboration on OCR improvements (IMPACT)
As part of the ongoing collaboration between BHL-Europe and IMPACT, in April 2011 we produced a test set for ground truthing of page images through the EU-funded IMPACT project. This test set selection was based on the IMPACT guidelines for building the test set. BHL-US content was used for this test set as this content has the variability of material needed for the process, e.g. languages, font types. BHL-US also has experience with OCR based on previous projects and initiatives to work on the challenges with high quality OCR. Henning Scholz and Chris Freeland were the main actors on our side to prepare and finalise the test set in collaboration with IMPACT. Francisco Welter-Schultes is now the liaison to follow up the ground truth process over the next weeks. For this process we received software developed within the IMPACT project for testing purposes. Now the ground truth pre-production has started involving the following steps:
- Conversion of supplied JP2 images to TIFF required for OCR and long-term storage
- OCR of the derived TIFF files to produce PAGE.XML, the format used for ground truth and evaluation in IMPACT
- Assigning unique IDs to both images and initial OCR result
- Ingest of images and metadata into our central repository
- Selection of a small set of representative images for the tendering process
- Sending out a tender to a number of qualified service providers

We hope to discuss some first outcomes of the process alongside the Public Core Release Meeting of BHL-Europe at the end of August in The Hague.

Schema Mapping Tool
Wolfgang Koller (NHMW) ran a workshop session during the Annual Review Meeting and presented a live demonstration of the schema mapping tool. Further improvements and some adaptations of the schema mapping tool have also been implemented in order to improve the handling for the content providers. Metadata provided by content providers were mapped to the OLEF schema in order to be able to ingest them into the BHLE-Portal. Over the next few months NHMW will map more content providers to OLEF as well as assist with the portal development.

Integration of CoL
In June Frank Bisby (Sp2000) visited Heimo Rainer at the Vienna Natural History Museum to discuss the requests made by Heimo for Species 2000 to provide the Catalogue of Life to both the BHL-Europe and OpenUp! projects.
The principal discussion was on the structure and usage that can be made of the Catalogue of Life synonymy, to provide synonymic indexing and a taxon user-view for both the BHL-Europe and OpenUp! Portals. Furthermore different ways of delivering the Catalogue of Life – from imported editions to web-services were discussed as well as the new web-services being implemented for the Catalogue in the 2nd Batch of services in the 4D4Life project, due for testing at the end of June 2011, and for public release in October 2011. These will be made available for test during the summer. The nature of the license or permission to use the CoL that will be provided by Species 2000 to BHL-E and OpenUp! was another topic. The permission will be to use the CoL as an indexing and annotation tool, but restricted to this use and not extending to publishing the CoL or derivative works.

3.2.1 - 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).
The ingestion process for metadata into Europeana has been described taking into account the newly released version 3.4 of the Europeana Semantic Elements (May 2011). Once the content of the partners is in the OAI repository the partners receive a link to their content in the Europeana content checker tool where they can preview their records. Partners could give feedback on their content presentation in the content checker tool. Their feedback was incorporated before the data was finally provided to Europeana.

Europeana office also received the link to the new content in the content checker and provided feedback. The May 2011 Europeana publication included the following BHL-Europe sets:

- BHLUS (87572)
- CSIC (9)
- LANDOE (3511)
- NBN (3516)
- RBGE (183)
- RMCA (52)
- UBBI (1634)
- UBER (60)
- UCPH (28)
- UHVIIKKI (30)

The BHL-Europe Europeana ingest planning was revised and a new planning till project end was forwarded to Europeana office.

3.2.2 - Develop a distributed data model that will allow countries to retain control of their data, while enabling the material to be available through EUROPEANA, the BHL Portal (with Taxonomic Intelligence tools), and any national portals.

Translation modules for the portal have been added and a workflow for the translation process has been designed.

The style and the pagination of the result list have been adapted. Additionally, there are now configuration options for the view of the result list and of a record. Search queries are now saveable and can be managed in a user interface.

Nodes of the BHL-Europe portal have been installed for test purposes using a virtual machine. This virtual machine was handed over to NHM and was mounted at the NHM facilities in London as a test engine.

3.3 - Addressing distributed access and storage – long-term sustainability

US Data Transfer to BHL-Europe Servers

Over the past couple of months Chris Sleep (NHM) has been liaising with our US partners with regards to the upload of content. Once the decision was made to ship the data from the US on a set of discs holding the remainder of the initial content, this significantly sped up the process of data transfer and allowed the progress to the next phase of data synchronisation over the WAN (Wide Area Network). The process of copying data from the last batch is now complete and spot checks were run before the drives were shipped onward to Bibliotheca Alexandrina in Egypt.

GUID Infrastructure Setup

Andreas Kohlbecker (BGBM) coordinated the development and setup of the GUID-Minting and Handle.net resolver infrastructure. Furthermore he was responsible for the installation and setting up of the Handle.net server on the BHL-Europe server, as well for the registration of the handle prefix.

3.4.1 - Work with the EUROPEANA to create multiple language access to the BHL Portal.

Port functionality from Prototype to Access.

Data harmonization of RBGE, RMCA, CSIC, UH-Viikki has been continuously improved and been transformed into ESE as part of the negotiation of the Europeana harvesting procedure. The BHL-Europe portal prototype basic search functionalities have been translated into seven European languages.

3.4.3 - Apply the model to the key European languages (English, French, German, Italian, Portuguese, Spanish, Danish, Dutch, Czech, Hungarian, Polish, Swedish), enabling access to the BHL Portal throughout Europe.

Further Development of the Portal System

A preliminary Drupal 7 theme has been created based upon the official BHL-Europe Drupal 6 theme. Drupal 7 has been setup to be using different localization and internationalisation modules thus facilitating the implementation of different languages. Therefore all 7 necessary languages prepared for the Annual Review have been incorporated.

The current available version of Islandora was not compatible with Drupal7 and therefore Atos (Lee Namba & Hengdi Zhang) have rebuilt some of the essential modules (the Islandora digital repository module and book module) to make the basic functions available on our Drupal7 portal and replace fedora item bootstrap class with fedora API module for Drupal7.

Atos have also developed web services for downloading books in different formats e.g. PDF pages and books with a range, OCR packages of books, images and books. These web services enable the download of different formats of ingested pages and books from the Fedora repository. Furthermore, Pentaho Drupal Integration has been tested and implemented using HTTP Post with METS to create new Drupal nodes in the portal. Data harmonization for RBGE, RMCA, CSIC, UH-Viikki has been developed, specifically pre-processing steps to harmonize the structure of submitted files. This pre-processing is done on a per-collection basis using various bash-scripts and XSL transformations. Also, the IA BookReader for preview purposes of created archival information packages has been created. An image conversion backend has been created using Perl and PHP, the PHP backend is currently in use. New tests needed to be conducted using Tesseract 3 since old tests and development plans were based on Tesseract 2.

Pentaho Drupal has been used to create ESE mappings for metadata. It has also been used to create transformations specifically needed for OAI-PMH. During development it has proven itself to be a tool which should be used for transforming file/folder structures so that submissions adhere to the pre-ingest file submission guidelines.

Implement BHL-Europe Core Features and preparation of the Portal Release (M28)

One of the outcomes of the Portal Development workshop held in Tervuren was for the Catalogue of Requirements to be further revised into a single specifications document capturing all the relevant features for the portal development and also serves as a reference for testing. This document (Catalogue of User Requirements BHL-Europe Portal Feature List & Priorities) also captures all the relevant information from a user's perspective of the key functionalities of the BHL-Europe portal as well as prioritising the features according to user's feedback and contractual obligations. The document has since been revised by Jana Hoffmann and has been made public to all the consortium members.

Furthermore, the List of features have been migrated to the Issue tracker and prioritised. Up until the end of July the developers will be busy dealing with the Portal Issues (features, enhancement, bugs etc) and ensuring that Core Features are tested and fully functional ahead of the Public Core Release in M28.

D3.7 Key Components documented for output of D3.5

D3.7 was submitted to the EC in May 2011. The main objective of the document was to cover the technical architecture status in terms of where our technology stands to date with regards to the Implementation Phase. Overall the document described the technical architecture, workflows and interdependencies for the Key Components previously identified in deliverable D3.5 including: - BHL-Europe Portal, Pre-Ingest tool and Archival Storage.

WP4

Work package four was closed in April and all relevant deliverables, D4.2 and D4.3, were submitted in time.

WP5

5.1 - CWG meetings

During the Annual Review Meeting, Tervuren and Meise (Belgium) a CWG meeting was held. Several employees of the RMCA interested in the project joined this meeting of the communication working group. Jiri Kvacek and Jiri Frank presented their work on the virtual exhibition BLE[17] on spices. All the different features were discussed and suggestions were made. It was agreed that BHL-Europe will continue working on a virtual exhibition on expeditions until the November release.

Furthermore, Patricia Mergen presented RMCA's virtual exhibition on expeditions[18], where they combined information from different projects (BHL-Europe, OpenUp!, Europeana).

Furthermore the timeline for dissemination activities related to the portal launch were discussed and agreed. For further information please refer to the minutes on the BHL Wiki.[19]

5.2.1 – BHL-Europe Website - http://www.bhl-europe.eu/

During May and June we repeatedly edited the project web site and also BHL Wiki. The number of languages was increased to 8, three other languages – Hungarian, Finish and Polish are in preparation. MIZPAS provided a translation of the project web site into Polish. RBINS is currently working on a Dutch translation.

5.3.1 – Work with promotion material (bike jerseys)

A few BHL-Europe staff members were equipped with bike jerseys to promote BHL-Europe to the general public during sports events. The first event we used the jerseys was a cycling time trial in Halle[20] on 29 May 2011. The second event was the Berlin Triathlon[21] on 19 June 2011. Dennis Zielke and Henning Scholz were the two starters in both events for BHL-Europe. The success of both starts justifies the extension of these activities. Two more jerseys were ordered for the two NHMW colleagues in Austria for disseminating BHL-Europe to other people at other sport events.

5.3.1 – BLE - Virtual exhibition

In May and June most affords of NMP staff was focussed on finishing a draft of the virtual exhibition BLE[22]. The purpose of BLE is to promote the content of BHL-Europe and the BHL-Europe portal to a wider public. In cooperation with IT4Care we finished the topic "Spices" which includes 52 books from 19 countries, in 17 languages ranging from 15th to 20th century. They are divided into 10 subjects and can be sorted by 50 spices plants (groups of plants in some cases). There are 52 spices cards underlined with more than 150 pages of interesting information. In total there are 200 illustrations of spices on display. The BLE was successfully presented during the Annual Review Meeting in Tervuren and Meise, Belgium.

5.3.2 - BHL-Europe Newsletter

Newsletter #7 was produced in April 2011. Henning revised and approved this newsletter issue and RBGE did the English editing. 100 copies were brought and distributed during the Annual Review Meeting. The annual report is going to be issue #8 of the BHL-Europe Newsletter to be finalised in July 2011. The work on the annual report, however, was done in April 2011 as this document is an official deliverable for M24 of the project (D1.7). 50 copies were printed and send to partners and to the EU commission. The work on the graphical design of year three issues and Newsletter #9 was initiated.

---

[17] http://www.biodiversityexhibition.com/
[18] http://cybertaxonomy.africamuseum.be//drupal7/
[19] https://bhl.wikispaces.com/BHLE_AnnualReview2_minutes
[20] http://www.velo-ev.de/index.php?page=135
[21] http://www.berlintriathlon.de/
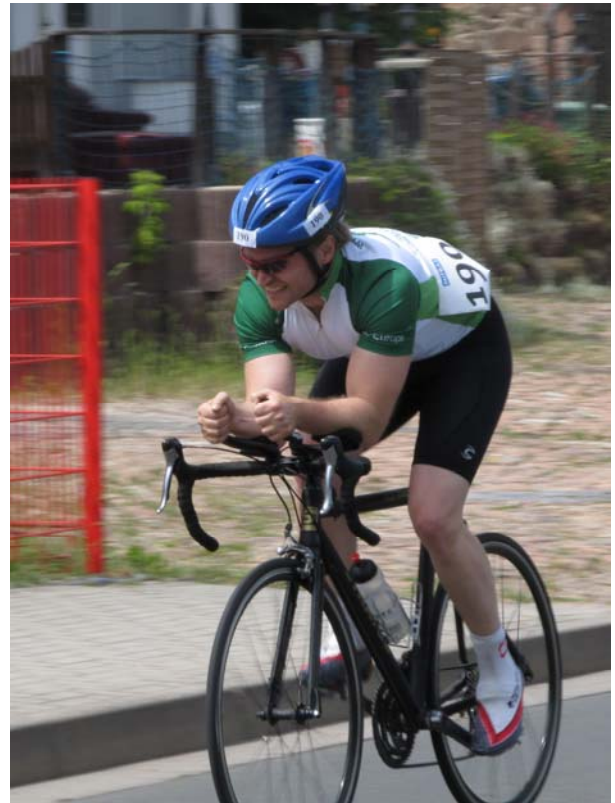[22] http://www.biodiversityexhibition.com/

Figure 1: Henning (left) and Dennis (right) during the individual time trial in Halle.

Figure 2: Henning (front) and Dennis (back) during the cycling part of the Berlin Triathlon.

Figure 3: BLE entry page.

5.3.3 – Ensure proper press release; maintain the press review; distribution of articles for various types of print and online media.

International Innovation Report & other publications
In this reporting period, Henning finalised the BHL-Europe contribution for the International Innovation Report in collaboration with ResearchMedia. This report was published in May 2011 and has a 3-page publication on BHL-Europe. In the following, ResearchMedia also produced a brochure with additional material. Brochure and article is available for dissemination activities of BHL-Europe. We have the full copyright for distribution of the material in printed form or through the project website.
The list of publications on BHL-Europe written by BHL-Europe partners can be found on the BHL-Europe website[23] or the BHL Wiki[24].

Facebook, twitter and blog
At the end of June 2011 we have 103 'members' of the BHL-Europe Facebook site[25]. On Twitter BHL-Europe[26] has 80 followers at the end of June 2011, which is almost twice as much as the previous report. There are some articles in preparation for the blog, but they are not online yet.

5.4.2 – Maintain and update database of conferences
An up to date list of events where BHL-Europe is presented by project partners is available on the BHL Wiki. [27]

5.4.5 - Dissemination activities of partners
Several partners participated in various institutional or project meetings and conferences to represent BHL-Europe.

CSIC
CSIC staff translated the BHL-Europe Portal Technical Demos notes into Spanish. Furthermore, CSIC worked on an Exhibition on "Nombrando Especies" (Naming Species) which will include a

---

[23] http://www.bhl-europe.eu/de/publikationen/veroeffentlichungen
[24] https://bhl.wikispaces.com/BHL-Europe-publications
[25] http://www.facebook.com/pages/BHL-Europe/151086001600041
[26] http://twitter.com/BHLEurope
[27] https://bhl.wikispaces.com/dissemination+of+project+by+partners

stand on BHL-Europe. The exhibition is divided in three parts. The first part introduces "What is in a scientific name". The second part gives a review of the "Top Ten Species" in 2008 to 2011. The third part is dedicated to BHL-Europe and how digital libraries support the work in the field of biodiversity. This part also includes a video (1:30 minutes) and a booth with info and photos. The exhibition will open in the first week of August 2011.

EDLF
EDLF communicate BHL-Europe within Europeana network and Europeana and promoted BHL-Europe via the Europeana social media streams like Facebook, Twitter, and RSS. ELDF was engage in the set up of the BHL-Europe virtual exhibition plan by providing information on existing Europeana and The European Library virtual exhibitions. Furthermore, EDLF kept the information page for BHL-Europe on the Europeana Group page up to date and provided guidance for the transfer to the new Europeana knowledge sharing platform[28].

BGBM
W. Berendsohn gave details on the aims and the infrastructural setup of BHL-Europe during a talk titled "Sustainable infrastructures for e-taxonomy in Europe" for the Royal Danish Academy of Sciences and Letters during the Symposium on "Biodiversity in the Silicon Age", Copenhagen 17-18 May 2011.

MNHN
A paper has been published on the Gallica blog in May, after OAI-PMH updating at MNHN. This paper refers to the partnership of Gallica with BHL-Europe.

NBGB
Short notes describing the project were published in the internal newsletter for the staff of the Botanical Garden and in the in the Botanical Garden electronic newsletter. The BHL-Europe project was presented at the Librarians Day of the Federal Belgian Libraries Forum in Brussels, 17 May 2011, and at the European Botanical and Horticultural Libraries group Annual Meeting in Oslo, 19 May 2011.

UBER
During the German Librarians Day, 100. Bibliothekartag 2011 in Berlin 6-8 June 2011, Dennis Zielke presented the BHL-Europe project at the booth of the German Initiative for Networked Information (DINI e.V.) within the exhibition area.
During the Workshop on virtual research environments (VRE) at the Humboldt-University 21 June 2011 Dennis Zielke presented the BHL-Europe project to a staff member of the Konstanz University Library. The University of Konstanz is working on a new DFG-project called movebank (documenting moving traces of animals worldwide).

UGOE
A report on the collaboration with Firenze was written and submitted to the BHL-Europe newsletter, as well as an article which analysed the content of BHL in terms of distribution of digital content in the timeline of publication of the works.

RBINS
RBINS provided the Dutch and French translation of the BHL-Europe portal Demo site. Together with Nicole Hanquart from NBGB, Tine Mallaerts prepared a Dutch and French presentation[29]

---

[28] https://version1.europeana.eu/web/guest/details-bhl/
[29] http://www.slideshare.net/Bibforum/bhl-europe-biodiversity-heritage-library

about the project for the "Dag van de bibliothecaris" in Brussels. On this day about 60 librarians from different Belgian federal institutes came together and discussed different subjects.

RMCA
For the dissemination workshop RMCA prepared an integration of the Europeana API into the museums website to allow searching for RMCA content within Europeana (see: http://bit.ly/AfricamuseumEuropeanaAPI). RMCA also created a proof of concept for a BHL-Europe virtual exhibition around the theme of Lemaire's expedition to Katanga[30] showing connections between the published part of the expedition literature with other RMCA content and content in Europeana. In the frame of this activity, RMCA staff attended the E-corpus training, where an introduction to the creation of virtual libraries and satellite-websites within the e-corpus platform was presented.

5.5. – Plan and evaluate user activities
7 June Andreas Kohlbecker and NMP organised the Evaluation and Testing Workshop during the BHL-Europe Annual Review in Tervuren and Meise, Belgium. This workshop was to decide on the various testing approaches (including the technical testing) and identify responsibilities. The minutes can be found on the BHL Wiki[31].
Andreas also requested quotations from usability testing agencies and evaluated them.
Jiri Frank (QA Manager) and Jana Hoffmann (MfN) have established the testing group and created a testing workflow to do the initial testing and provide feedback for the developers.
The main focus for the developers and the QA team is to ensure that the Priority 1 Core Feature functionalities are tested and fully functional ahead of the Public Core Release meeting planned for the end of August.

BHL-US – Extract from the Program Director's Report

On behalf of Tom Garnett, Director of the BHL, Mr. Graham Higley, head of library and information services at the Natural History Museum in London, accepted the 2010 John Thackray medal[32], which was awarded to the BHL in recognition of the past three years of work building its digital library.
Chris Freeland and William Ulate attended the Society for the Preservation of Natural History Collections annual conference[33] in San Francisco, where BHL was offered a table in the vendor hall. Chris Freeland also presented a poster on the Engelmann's project and attended the stand with William.
Bianca Crowley and William Ulate attended the BHL-Europe Annual Review and shared with BHL-Europe colleagues on their common experiences, requirements and lessons learned. William Ulate gave a talk during the BHL-Europe Annual Review to update participants on the advances of the Global BHL Project.
Chris Freeland attended the Technical Workshop of the Digital Public Library of America[34] at the Associate Librarian for Library Services Office at the Library of Congress, which convened a small working group meeting to begin to make recommendations for the overall technical architecture of a Digital Public Library of America (DPLA) and to converge on a set of key technical principles upon which the DPLA should be built. Particularly, the objective was "*to surface and identify concrete insights, including areas for future research; development and policy proposals; and other tangible outcomes*".
Chris Freeland led an initial discussion on the overall goals and architecture of the DPLA. His presentation "DPLA Technologies: Foundations for Growth & Sustainability" at

---

[30] http://cybertaxonomy.africamuseum.be/drupal7/
[31] https://bhl.wikispaces.com/file/view/Evaluation+and+Testing+Workshop.pdf
[32] http://www.shnh.org.uk/awards/the-john-thackray-medal/2010.html
[33] http://www.spnhc.org/news/view/26
[34] http://bit.ly/oIvb7J

http://www.slideshare.net/chrisfreeland/dpla-technologies-foundations-for-growth-sustainability
was used as the base for discussion during the first day of the meeting.
Chris Freeland worked together with participants at the end of the DPLA meeting to summarize the key takeaways from the workshop: http://blog.chrisfreeland.com/2011/06/dpla-summary-from-jun-mtg.html
Charles Darwin's Library[35], a BHL special collection that draws on original copies and surrogates from other libraries and provides full transcriptions of his annotations and marks, was released online. This collection is composed by a digital edition and virtual reconstruction of Charles Darwin's 330 most heavily annotated books he owned by the moment of his death.

4   Deliverables

These deliverables were submitted in M24 (April):

D1.5	First overview of the business plan for long-term sustainability	M24	M25, May 11
D1.6	Progress report 4 including pre-financing request	M24	M25, May 11
D1.7	Annual report	M24	M25, May 11
D2.5	Final and enhanced Web-database for content management and collection analysis	M24	M25, May 11
D2.6	Delivery of the first version of the approved best practice guidelines and standards	M24	M25, May 11
D2.7	Content analysis and management status report 2 (metadata, page numbers, content providers)	M24	M25, May 11
D3.7	Key components documented for output of D3.5 e.g. BHL-Europe Portal, OCR demonstrators, distributed storage model, etc.	M24	M25, May 11
D4.2	Complete signed agreements with first group of rights holders to enable material to be used in the BHL Europeana, and established process for addition of further material.	M24	M25, May 11
D4.3	Deliver IPR framework to support long-term access and sustainability of the digitise material.	M24	M25, May 11

5   Schedule status/problems/issues/deviations

Risk & issues
Although we received a green flag during the review meeting, we can not confirm we are fully on track. We have identified a number of issues and delays in the project that have implications for the success of the project. The most important issues are explained in some detail in the following.
The timely production of high quality status report by all partners is still an issue for the project. The same is true for the monthly activity reports and the financial statements. Experience has shown that it is very difficult to get the full set of documents in time. Status reports need to be narrative on task level to facilitate the compilation of the full report. Activity reports need to be on

---

[35] http://www.biodiversitylibrary.org/collection/darwinlibrary

task level so we are able to match the hours with the narrative report. Eventually, the financial statements need to reflect all work so the narrative progress report is consistent with the financial report to justify all costs. If we are not able to produce all reports in high quality in shorter time next year, we risk not being finished by the end of the project and then lack the resources to finalise the paperwork. This will at least delay the payment of the final rate to the project.

The GRIB is still not working, as demonstrated and seen during the Annual Review Meeting in Tervuren. Serious technical issues with the GRIB infrastructure at the hosting site affect the finalisation of the GRIB prototype. It is planned to solve the problems over the summer. With the delay of the GRIB we risk that we are not able to integrate GRIB functionalities into the BHL-Europe portal. We also risk to not being able to replace the BHL scan list with the GRIB.

By the time of writing, the Pre-Ingest tool is not accessible, useable and testable for the content providers and no tutorial is available. This has a series of implications for several areas in the project.

First, we can not run all available data through the process and produce the various files we need for the various OAIS components (SIP, AIP, DIP). This delays the testing of the entire ingest process and the test of the technical infrastructure. It also delays the test of the data performance in the portal, for example. This again influence the entire data upload to NHM servers. Content providers need to understand how the data are displayed before uploading the full corpus. Quality checks need to come at an early stage of the process to avoid duplication of work. Eventually we risk not being able to ingest all content we want and need into our system by the end of the project.

Second, our content providers need time, training and clear instructions to fully understand the process to work with the tool independently. We need to understand the work required by all content providers to make sure sufficient resources (and expertise) are available on the content provider side. This is particularly important for the non-consortium content providers, as they don't have any budget from us to do the work. Currently we can not expect our content providers to do the work with the Pre-Ingest tool.

Third, a delayed Pre-Ingest tool delays the content acquisition process. New and prospective content providers need detailed information on what we expect them to do and how to prepare the content best for us. This requires a successful test of the Pre-Ingest process to identify problems and errors that may require a revision of the current file submission guidelines.

Forth, the Pre-Ingest is an important part of the Best Practice Guide (D2.9). Without being tested and used by our content providers, this section in the Best Practice Guide can not be finalised. A delayed D2.9 will also delay the work on deliverable D5.11, which is based on D2.9.

The current progress towards the Public Core Release end of August is slow. We still don't have all code fragments in one place (NHM server) to fully work in a distributed way. This also has an impact on the testing procedure as testing needs a fully working portal version with all parts in place. We soon need to decide if the August milestone is still achievable or not. A delay of the Public Core Release has some important implications. Most importantly, we are loosing time to incorporate and test all features we have identified for the final BHL-Europe portal. We also can not launch our virtual exhibitions without a working and stable portal at the base. This will delay the start of our final dissemination activities of the project. Without a public portal, a user evaluation on a larger scale is also not possible. This will delay D5.9, our 2nd user evaluation report. Also workshop and training sessions for users can not be offered without the portal. A public portal with content properly ingested into the system is also required to actively promote our work to prospective content providers and thus facilitate the content acquisition process. Otherwise, we are loosing time to negotiate with new content providers. The meeting in The Hague at the end of August is in any case required to review the whole status and re-allocate tasks in a way to bring us back on track.

6   Work planned

The work planned for the next phase is explained in more detail in Annex 1 (DoW).

7   Products to be completed in the next phase

33	D1.8 Progress Report 5
34	D3.8 Sustainability policy for continuation of services e.g. hosting, future development, helpdesk provision for service/ content provider etc.
35	D5.9 (postponed to January 2011) Second user evaluation report