**Life and Literature Conference**

**November 14 – 15, 2011**

**Report to the Richard Lounsbery Foundation**

**DRAFT**

**Executive Summary**

To address the future, current and pending issues of the digitization and accessibility of biodiversity literature, the Biodiversity Heritage Library proposed a large, public conference titled “Life and Literature” and received funding from the Richard Lounsbery Foundation. The conference was held at the Field Museum, November 14-15, 2011. 124 participants from approximately 30 countries attended including librarians, biologists, computer scientists, publishers, students, educators, and others. The Life and Literature Conference was an enormous success and appreciated by the attendees. The Conference gave a very effective structure for many BHL users to reflect on the uses and problems of biodiversity literature and to offer concrete suggestions to the BHL on how to improve the situation. BHL Member Institutions will refine and rank suggestions arising from the conference as a key step in preparing a revised work plan for the next three to five years.

**Background: the Biodiversity Heritage Library**

The Biodiversity Heritage Library (BHL), the digitization component of the Encyclopedia of Life, is a consortium of 12 major natural history museum libraries, botanical libraries, and research institutions (see list at end) organized to digitize, serve, and preserve the legacy literature of biodiversity. The BHL is a multi-institutional project because no single natural history museum or botanical garden library holds the complete corpus of legacy literature, even within the individual sub-domains of taxonomy. However, taken together, the consortium of collections represents a uniquely comprehensive assemblage of literature.

Prior to digitization, the resources housed within each BHL institution have existed in isolation, available only to those with physical access to the collections. Much of this published literature is rare or has limited global distribution. These collections are of exceptional value because the domain of systematic biology depends – more than any other science – upon historic literature. . The “cited half-life” of natural history literature is longer than that of any other scientific domain. The so-called “decay-rate” of this literature is much slower than in other fields such as biotechnology.

The relative isolation of these collections presented an antiquated obstacle to biodiversity investigation. This body of biodiversity knowledge, in its current form, is unavailable to a broad range of applications including: research, education, taxonomic study, biodiversity conservation, protected area management, disease control, and maintenance of diverse ecosystems services. This problem is particularly acute for the developing countries that are home to the majority of the world’s biodiversity. Mass digitization projects at large research libraries lack the discipline-specific focus of the Biodiversity Heritage Library Project and have failed to capture significant elements of legacy taxonomic literature.

The primary funding of the BHL has been from the John D. and Catherine T. MacArthur Foundation through a grant to the Encyclopedia of Life project, with smaller amounts from the Gordon and Betty More Foundation and the Richard Lounsbery Foundation. BHL member institutions have contributed significant staff resources to advance the project and have made plans for the continued availability of the already digitized corpus of texts after the current funding expires in 2012. However, without additional funding it will be impossible to digitize the remaining vital legacy biodiversity literature and to continue providing networked services to take advantage of the dynamic technical landscape.

The BHL has demonstrated to date:

* 1. The infrastructure for scanning, mounting, and serving digitized biodiversity literature.
  2. A cooperative working environment involving decision-making, discussion, and work assignment distributed among staff at 12 natural history and botanical libraries.
  3. The provision of web services for reuse of this content by other projects.
  4. A favorable relationship with non-profit scientific publishers and non-profit journal aggregators allowing the BHL to digitize and serve journal backfiles that are post-1923.
  5. A global cooperation to incentivize other national-level projects to host and preserve the content.
  6. The ability to deliver more content through scanning if funding is made available.

However, although the BHL currently makes 37,000,000 text pages of biodiversity literature available, this is only 6.4% of the estimated total domain of core biodiversity literature or 32% of the pre-1923 literature.

To address the current and pending issues of biodiversity literature the BHL proposed a large, public conference titled “Life and Literature” and received funding from the Richard Lounsbery Foundation. The conference was held at the Field Museum, November 14-15, 2011. 124 participants from approximately 30 countries attended including librarians, biologists, computer scientists, publishers, students, educators, and others. <ANY OTHER RELEVANT ATTENDEE INFO?>

**The Life and Literature Conference.**

The purposes of the conference were:

1. Engagement of current and future constituencies concerned with biodiversity literature.
2. Integration and interoperability of the BHL with major biodiversity and digital library programs.
3. Further innovative integration of digitized literature with biodiversity databases.
4. Publishing models and legacy scientific literature.
5. Setting the agenda for the BHL for the next 4 – 5 years.

There were two plenary speakers:

**Dr. Richard Pyle:**

Richard Pyle's main field of expertise involves the taxonomy and biogeography of coral-reef fishes. His other areas of interest include the use of advanced diving technology to document biodiversity inhabiting deeper regions of tropical coral reefs, and also the development of computer database systems (and associated data standards) for managing biodiversity information.He earned his PhD under the guidance of John E. Randall, and has worked in the Department of Natural Sciences at Bishop Museum since 1986. He has authored over a hundred scientific, technical, and popular articles, has been featured in dozens of documentary films, and has received a number of significant awards for his achievements. Pyle is also a commissioner of the International Commission on Zoological Nomenclature.

**George Dyson:**

George Dyson is a historian of technology whose interests have included the development (and redevelopment) of the Aleut kayak (*Baidarka*, 1986), the evolution of digital computing and telecommunications (*Darwin Among the Machines*, 1997), and the exploration of space (*Project Orion*, 2002). *Turing’s Cathedral*, “a creation myth for the digital universe,” will be published by Pantheon (USA) and Penguin (UK) in January 2012.

There were four panels:

**Research, Informatics, and the Published Record.**  
The panelists were Dr. Sandra Knapp (Natural History Museum), Dr. Stanley Blum (California Academy of Science), Dr. Donat Agosti (American Museum of Natural History), and Dr. Elycia Wallis (Museum Victoria).  
  
**Publishers, Aggregators, and Authors - New Models and Access.**  
The panelists were Susan Skomal (BioOne), Jan Reichelt (Mendeley), Lyubo Penev (Pensoft), ?? Catriona MacCallum (PLoS), Abel Packer (SciELO)  
  
**Learning and Education.**  
The panelists were Dr. Marie Studer (Encyclopedia of Life), Dr. Douglas Wilkin (CK-12 Foundation), Dr. Devin Reese (NSF Resource Center), Dr. Ken Walker (Museum Victoria), and Natalie Zamora (INBIO).  
  
**Building Collaborative Networks for Science and the Humanities through Scientific Literature.**  
The panelists were Dr. Janet Browne (Harvard), Dr. Alain Touwaide (Smithsonian), Dr. Peggy Macnamara (Field Museum), and Dr. Chris Wildrick (Syracuse University).

These were followed by breakout sessions that focused on four key questions for the future of the BHL:

1. What new types of content and connections between content should BHL focus on bringing into its collection in the next 5 years?
2. What new technologies should BHL adopt and integrate, or plan for, in the next 5 years?
3. What other institutions, domains, or user groups should BHL engage with in the next 5 years?
4. What does education mean for BHL? What is the audience? Targeted? Broad?

The four breakout sessions reconvened and presented their main findings for what BHL should maintain, expand, innovate, and avoid in future development.

In the appendix to this document are the notes, findings, and recommendations from the four breakout groups from the Conference. What follows is a synthesized summary, which were often expressed in differing terms and with differing frames of reference. The synthesis can be cross checked by comparing with the original notes.

The Life and Literature conference established beyond doubt that the BHL is a pilot project no longer. It is part of the workflow for research and relied on by many.

BHL is relied on by people in diverse fields and in various ways:

1. Practicing taxonomists doing primary zoological, botanical, and, to a lesser extent, mycological research.
2. Historians in such fields as history of science, agricultural history, period re-enactment, and cultural history.
3. Artists, illustrators, conceptual artists.
4. Hobbyists.
5. Teachers? <more examples>
6. Librarians, public, special and academic
7. Bioinformatics Professionals
8. Students in the fields of zoology, botany, history of science, mycology, and more

BHL is used by other research information systems and their users:

1. Biodiversity informatics systems such as the Encyclopedia of Life, BioStor, the European Distributed Institute of Taxonomy (EDIT), <more please> AlgaeBase, AnimalBase, WoRMS
2. Bibliographic aggregator/facilitor systems such as CrossRef, OCLC, Europeana, <more please>
3. Museum Collection Information systems <verify and examples.> California Native Plant Society's Inventory of Rare and Endangered Plants ([www.rareplants.cnps.org](http://www.rareplants.cnps.org/)) Royal Museum of Central Africa (via Europeana)
4. Online library catalog systems such as the library systems at the Smithsonian Institution, Harvard Museum of Comparative Zoology, Missouri Botanical Garden, New York Botanical Garden, California Academy of Science <more please>Field Museum Library, featured link to Biodiversity Heritage Library placed on Fenton's Library website for July 2011 - please visit [http://fentonlibrary.wetpaint.com](http://fentonlibrary.wetpaint.com/),

<more, please> Synynyms—Ryan Schenk’s species name visualization tool. http://synynyms.no.de/

**BHL Users and Stakeholders Want Many Things**

Participants expect a significant expansion in scope and scale of the Biodiversity Heritage Library (BHL). In fact, some participants assigned almost any conceivable biodiversity-information need as a problem the BHL should attempt to solve. This tendency should not be viewed a problem but rather as an opportunity to more clearly define the boundaries of the BHL.

Discussions in the breakout sessions identified a number of areas where BHL requires significant enhancement:

* Contexualizing data for educators
* Improving the accuracy of Optical Character Recognition
* Better citation resolution

The BHL should avoid the following pitfalls:

* Imposing layers of interpretation
* Charging
* Closed access
* Being wedded to old technology
* Being bought out
* Limiting access
* Growing too big too fast
* Loss of focus (“spread too thin”)
* Bad metadata

Innovations that BHL should provide:

* Functional workflow to correct poor OCR
* Identification and annotation of illustrations
* Better ways of sorting search results, e.g. by taxon name occurrences, by date, by original description, by co-occurrence of other terms.
* Full text searching of the OCR by journal volume.
* User created collections at every level of access
* Click & go: Connect grey literature to published literature
* Click & go: Connect published sources to published literature
* Explore ways to serve in-copyright content
* Compare literature to current environmentalist/conservation biologists
* Discovery layer services
* Common name authority file and searching – allowing people to add and correct
* Article-handling, article metadata.
* Multi-lingual access
* Annotation
* BHL should expand**:**content based on taxonomic communities (ViBrant, Scratchpads)
* Accept boutique scanned collections
* Connect w/ Mendeley
* Connect w/ ZooRecord
* Define CiteBank & make a trusted repository
* article level access
* BHL Resources & funding!
* BHL citations in Wikipedia
* Connect better to Google
* Make scan-on-demand easier

**What is feasible in the next 5 years?**

**What can be done in the next year?**

* BHL Expeditions tool made available end of 2012!
* Collect and coordinate feedback and focus groups on interfaces and use cases
  + BHL US performed user testing on UI in October 2011 (see appendix for results)
    - Major outcomes: make it easier to find illustrations and improve pagination
    - BHL US has a feedback link on their website, through which they have and will continue to receive feedback on UI improvements
  + BHLAu beginning to collect feedback and coordinate focus groups on UI
    - Navigation (with multiple searches) is important
    - There is a feedback form on bhl.ala.org.au - this is best way to get feedback
  + News feed on home page
    - BHL China is about to put 2 million photos on, should take advantage of this
    - BHLEU has 14 languages, need to advertise that
  + Link to all other BHL sites and harmonize portals
* Include maps with illustrations
  + so far only on books, need to think of other types media - pure illustrations, sounds/audio, manuscripts, video, etc.
* Have autofill on forms
* See appendix A for immediate post-conference work on a BHL Useability Study

**What can we do in the 1- 2 years?**

* Coordinate globally to develop K-12 educational packages
  + Make site more aesthetically pleasing to draw in users
* ID intermediaries to talk to and get message out
  + teachers, lib, game developers, museum educators
  + volunteer advocates
  + DVDs/CDs and preset materials for students to use to present to students
  + give materials
* Larger presence on social networks
  + right now on US and UK site, very active twitter feed/Facebook/blog/Flickr
* Coordinate Focus Groups!
  + ideas for use, who will use site?
  + not just for science, but for art and other audiences

**What can we do in the next 3-5 year?**

* GIS component into search or be able to use it
* Develop Tools and API development so game designers and media producers can use BHL info in what they create
* Integrate Computable data (wing length, life span, etc) way to pull so people can use and manipulate
* Partnerships with publishers to get more modern materials

**Conclusion**

BHL Member Institutions will refine and rank suggestions arising from the conference as a key step in preparing a revised work plan for the next three to five years. Subsequent to the Life and Literature conference, a matrix of technical development was drafted for review by BHL stakeholders s a first step in the process,.

The Life and Literature Conference was highly successful.. <Can someone assemble some testimonials here?> The conference provided an effective structure for attendees to reflect on the uses and problems of biodiversity literature and to offer concrete suggestions to the BHL on how to improve the situation.

**Appendex A**

# ****BHL USABILITY TEST REPORT 2011****

This Usability Test Report intends to summarize what the [2011 Usability Test Notes](https://bhl.wikispaces.com/2011+Usability+Test+Notes) indicated and convert them into [Portal User Interface and Functionality Requirements](https://bhl.wikispaces.com/BHL+Portal+Requirements) for the BHL-US/UK and BHL-Australia Technical Teams during the [BHL-AU Staff visit to MOBOT 2011](https://bhl.wikispaces.com/BHL-AU+Visit+to+MOBOT+2011" \l "BHL-AU%20Staff%20visit%20to%20Missouri%20Botanical%20Garden.-Goals). It is expected that particular tasks and timelines would come up from this input after review and prioritization.

## ****NAMES****

### ****Summary Issues****

Some users said that finding Names was not apparent enough on the Australian site.  
On both side, many users were confused by the term “Names” and did not understand that it meant “Scientific Names” and not “Personal” “Corporate” or other.

### ****Recommendations****

**BHL-Australia**

* Use a more universal symbol that conveys the concept of Scientific Names instead of uBio.
* Change the rollover text to say “Scientific Names on this Page” which the general public can understand.

**BHL-US/UK**

* Change the text on the Book Viewer page from “Names on this page” to say “Scientific Names on this Page” which the general public can understand.

## ****OCR****

### ****Summary Issues****

Participants did not know what “OCR” was or what they could do with it.  
People did not know what “View Text” meant.  
There is an expectation that users should be able to do Full-text searching on both sites, based on customary behavior in other sites and software applications.  
Testers observed Users trying to search on a page through the Ctrl+F key combination.  
It was suggested that the most appropriate solution to enable text and crowdsourced correction of it is to adopt a simple approach: install MediaWiki and assign every page its own place.

### ****Recommendations****

**BHL-Australia**

* Change “View Text on this Page” label to “View OCR from this Page”
* Change the existing message “This text is generated from uncorrected OCR” to “This text is generated from uncorrected OCR and as such, may contain, inconsistencies with the actual content of the original page.”

**BHL-US/UK**

* Change “View Text” label to “View OCR from this Page”
* Change the existing message “This text is generated from uncorrected OCR” to “This text is generated from uncorrected OCR and as such, may contain, inconsistencies with the actual content of the original page.”

**BHL-US/UK & BHL-Australia**

* The suggestion of allowing crowdsourcing through a simple approach like MediaWiki should be further investigated.

## ****SEARCH****

### ****Summary Issues****

In BHL-US/UK, some users interpreted that by typing a text in the box to search the metadata, they would be also looking within the text.  
For the BHL-Australia site, it was suggested that it would be interesting to search a text within the contents of a book or journal.  
  
**Recommendations**

**BHL-Australia**

* Change the text “Search” that appears within the Search box to “Search another book or journal” to avoid confusion when searching within the book is desired.

**BHL-US/UK**

* Put the text “Search another book or journal” as a within the Search box to avoid confusion when searching within the book is desired.

**BHL-US/UK & BHL-Australia**

* It is strongly suggested that searching text in the page should be enabled (Ctrl+F) and searching text in the whole book should be enabled by including a check box to mark when “Find in the entire Book”. For example, something like this:
* Include functionality for “Keyword in context”, where a keyword is highlighted within the sentence or paragraph where it was found.
* It is advised that allowing to search for text within all the books be enabled; for example, by including a check box to mark when “Search within Books/Journals” in the current Search option:
* Until Full-text searching is enabled, when users try to find (hit Ctrl+F), there should be an automated message pop up that states “This book or journal is not Full-text searchable”.
* Enable keywords search within a title.

## ****FINDING ILLUSTRATIONS AND MULTIPLE-BOUND OBJECTS****

### ****Summary Issues****

Finding illustrations brought a lot of frustration for users. Users are frustrated by lack of page metadata in BHL-Australia.  
In the BHL-US/UK, page metadata should be accurate (i.e. not useful when images say “text”).  
Users said they liked the thumbnails but they also found the page navigation useful as well as the slider from BHL-Australia; they wanted to have available together.  
There should be a marriage of the image page preview and the page metadata.  
Thumbnail view in AUS is not obvious enough.  
Several users looked in the table of contents and indexes for illustrations  
Searching across Multiple bound objects had the same issues as illustrations: page metadata could be greatly improved to assist with this and Thumbnail View is useful when pages aren’t articulated.

### ****Recommendations****

**BHL-Australia**

* BHL-Australia could provide a way for users to scroll through the page metadata as a whole instead of just page by page (for example, provide visual cues in the slider where the pages of certain type are)
* BHL-Australia should consider adding the page metadata to the hover menu of each page (particularly useful for the Thumbnail view)

**BHL-US/UK**

* The page that is shown in the Book Viewer should be highlighted in the page metadata list, synchronize the scrolling of the Book Viewer and the page metadata list.
* Consider including the Thumbnail view and the slider into the BHL Portal as a way to navigate.
* Do not show the Internet Archive window when downloading (PDF, OCR, etc.) files and once the Thumbnail Viewer is implemented, evaluate taking out the Internet Archive link so that the user does not leave the BHL World.
* To avoid confusion with illustrations, consider changing the “Download images” and rephrase it "Download JP2"

**BHL-US/UK & BHL-Australia**

* Investigate the idea of tagging at least the table of contents and the indexes among all the pages.
* Consider not displaying the default value “text” that comes from Internet Archive because it is confusing for the user to see it assigned to images.
* Find a way to improve the Thumbnail loading speed.
* Investigate how to automate the pushing of images to Flickr, let the users tag them and provide a way for BHL to use this tagging and incorporate it in the Portal.

## ****EOL****

### ****Summary Issues****

The BHL-Australia link to EoL is a little bit too subtle and the BHL-US logo is a little bit faint and hard to see.

### ****Recommendations****

**BHL-Australia**

* Consider changing the “[EOL]” string for “[View details in EOL]”.

**BHL-US/UK**

* Consider making the icon linking to EOL more evident one by adding a black background or using the version with a clear background (gray).

**BHL-US/UK & BHL-Australia**

* Consider having a link called “More information…” that then goes to an intermediate page to make the Name integration easier across BHL (EOL,ALA, Wikispecies, Catalogue of Life, GNA, etc.). This way, users will be more aware when they will be leaving the BHL World.

## ****GENERAL****

### ****Summary Issues****

“Titles” is more explicit that “Book/Journals” in the Search Results page.  
The general “look and feel” of the BHL-Australia site is preferred.  
The functionality of the BHL-US/UK site is desired in the BHL-Australia site.

### ****Recommendations****

**BHL-Australia**

* Consider giving the Book Viewer the look and field of a popup window on top of the Bibliography (for example, allow for an easy way to Escape (Esc or X)
* Allow to Find (Ctrl+F) in the Book Viewer

**BHL-US/UK**

* Consider using “Titles” instead of “Books/Journals” in the Search Results page; consider the use of default text in the search box (“Search in the Collection”).
* Consider the BHL-Australia site “look and feel”; take into consideration the use of real estate that the Book Viewer employs (for example, consider the use of a floating bar) and the utilization of graphical elements (like icons, the slider, etc).

**BHL-US/UK & BHL-Australia**

Improve OCR quality, allow for crowdsourced correction and achieve a better rate of scientific names found.