The Biodiversity Heritage Library Growing Globally: Sowing the Seeds in Europe and the US

Boris Jacob[1], Constance Rinaldo[2],
and Bianca Crowley[3]

[1]Royal Museum for Central Africa, Leuvensesteenweg 13
3080 Tevuren, Belgium
{boris.jacob@africamuseum.be}
[2] Ernst Mayr Library of the Museum of Comparative Zoology, Harvard University
26 Oxford St.
Cambridge, MA 02138 USA
{crinaldo@oeb.harvard.edu}
[3]Biodiversity Heritage Library
National Museum of Natural History
PO Box 37012
Smithsonian Institution
Washington, D.C. 20013-7012 USA
{crowleyb@si.edu}

Abstract. The Biodiversity Heritage Library[1] (BHL) was organized to digitize, serve, and preserve the legacy literature of biodiversity with a policy of open access. The BHL started as a small consortium of natural history libraries, botanical libraries and research libraries in the US and UK in 2005. Currently BHL delivers more than 33 million digitized pages. BHL capitalizes on features such as taxonomic intelligence, user feedback, and an article repository to deliver and refine services. This demonstration will show the global linkages, collaborative tools and user experience that forms the global BHL.

Keywords: Digital library, Open access, BHL, BHL-Europe, Biodiversity, Taxonomy, Systematics

1 Introduction to the Biodiversity Heritage Library

What is now a global initiative of connected projects and activities to digitize and make available the legacy literature of biodiversity, started as a small consortium of natural history libraries, botanical libraries and research libraries in the US and UK in 2005 [1].

The idea was then, and still is now, to support biodiversity related sciences (like taxonomy) by providing free and open access to publications that have stood as the pillars of these sciences, such as Linnaeus's *Systema naturae*[2] Darwin's *Origin of*

[1] http://www.biodiversitylibrary.org/
[2] http://biodiversitylibrary.org/page/6981741

Species or Humboldt's *Ideen zu einer Geographie der Pflanzen*, and hundreds of thousands of other works describing animal and plant species.

The idea of a digital library for the world's biodiversity literature has generated excitement in the library community as well as the biodiversity science community. The EC funded project Biodiversity Heritage Library for Europe[3] (BHL-Europe) started in 2009 aiming at leveraging European natural history literature and making it available via the European virtual library Europeana[4]. Other institutions, countries and continents soon followed and now the idea of a global BHL is spreading in Australia, China, Egypt and Brazil.

Biodiversity researchers, and taxonomists in particular, require access to specialized literature, spanning all publication years and geographic locations. The literature, together with the plant and animal specimen collections held in natural history museums and botanical gardens, are the foundation of biodiversity and taxonomic research. What once took a scientist a lifetime of work and extensive travel to gather, is now a few clicks away [2, 3].

Not only is access to the literature important but making use of the data elements embedded within the text is critical to connecting the dots throughout the cycle of research. Applications such as "taxonomic intelligence" expose the occurrences of scientific names within the optical character recognized (OCR'd) text and this rich dataset is made available for re-use via standard protocols, such as OAI-PMH, and application programming interfaces (APIs).

Dissemination of research is now part of a rich interdisciplinary, flexible environment where the legacy literature is digitally archived and made freely accessible, and web services liberate content into new applications and programs.

2 Global Framework: Demonstrating Linkages and Partnerships

The global BHL framework consists of self-managed regional and local nodes that collaborate, serve content and communicate on a global level. BHL has partnered with the Internet Archive[5] (IA) for digitization, server and storage services. BHL-Europe consists of more than 20 institutions in Europe, many of which have their own digitization programs. Some BHL-Europe content can be found through Europeana but not all content is available yet. Currently content from BHL is accessible by searching Europeana and linking to BHL for full-text viewing. Loading content from BHL-Europe into BHL is in the test phase. BHL-China[6] has a portal and has loaded Chinese content into the IA for harvest by BHL. The Egyptian installation of BHL through the Bibliotheca Alexandrina is under development. BHL SciELO[7], the Brazilian regional installation, was launched in December 2010 and will contain more than 700 multidisciplinary journals by December 2011. These journals will be

[3] http://www.bhl-europe.eu/
[4] http://www.europeana.eu/
[5] http://www.archive.org/
[6] http://www.bhl-china.org/
[7] http://biodiversidade.scielo.br

accessible via the article repository for BHL: CiteBank[8]. CiteBank is an open access platform to aggregate citations for biodiversity publications and deliver access to biodiversity related articles.

As the scanning and digitization section of the Encyclopedia of Life[9] (EOL), the BHL is engaged in linking into EOL species pages. Conversely EOL links out from species pages to literature in BHL. These linkages are accomplished using the taxonomic tools provided by uBio[10] and its webservices such as taxonFinder. The Atlas of Living Australia[11] (ALA) mines BHL content for the earliest mentions of Australian species and uses the ALA's literature tab to link to these references in BHL.

3 From Users to Librarians: Use Case Demonstration

Users of the BHL portals have free and open access to over 46,000 titles and growing. As with traditional library catalogs, users can search by standard entry points such as title and author, but what separates BHL from other digital library projects is its exposure of the scientific names embedded within the pages of its collection. Users conducting research on a given organism, for example corn, can enter *Zea mays*[12] into the search box and return an automatically generated index of pages where the scientific name is cited. Users are now able to gain more immediate access to the relevant content they need at the page level, impossible with a brick-and-mortar library collection. Browsing from this page-up, rather than title-down approach, users have the opportunity to flip through adjacent pages of the book or volume and explore any additional content associated with the title.

In digitizing historic works, it is often the case that the holdings of a single library cannot fulfill scanning for all volumes of a given serial title, for example. When users are met with gaps in the BHL collection, they are encouraged to provide a request for the missing content thus enacting a series of collaborative activities among BHL staff to satisfy the user's request. Coordinating the selection of materials for scanning, especially in response to patron-driven requests for content, requires the use of flexible, low-barrier tools for communication and collaboration. Librarians participating in the BHL use various tools to de-duplicate content, select items for scanning prior to digitization and manage the collection post digitization. While the scanning of materials happens piece by piece, the de-duplication happens title by title. Monographs and serials are handled differently through the Monographic Deduper and BHL Scan List, respectively. Working as simple lists, the Monographic Deduper serves as master list of unique monographic titles selected and sent for scanning while the BHL Scan List is a combined list of serial titles used by individual libraries to place bids on titles they intend to scan. Simple tools are better for collaborating within a growing global consortium. The low-barrier tools used for selection and de-

[8] http://citebank.org/
[9] http://www.eol.org/
[10] http://www.ubio.org/
[11] http://www.ala.org.au/
[12] http://biodiversitylibrary.org/name/zea+mays

duplication have the advantage of allowing the BHL to quickly amass a collection of digitized literature and leave room for additional partners to fully participate in content provision even years after the project's inception.

4 Conclusion and Outlook

Digitized literature in and of itself is a powerful resource. In that sense BHL aims to collect, store and make available much more biodiversity related legacy literature in the future. To better organize the scanning on a global scale BHL-Europe will provide the Global References Index to Biodiversity[13] (GRIB), a union catalogue of the existing published literature in natural history institutions, which includes linking to those publications already scanned and available online.

Accessing publications via the different local BHL web portals or via Europeana is but one way. For scientists it is more important to search and retrieve information via OAI-PMH and OpenURL from within special scientific workflow tools like the EDIT Platform for Cybertaxonomy[14].

However, when considered within the broader "knowledge ecology" of biodiversity research [4], the BHL is but one aspect of a larger interdisciplinary framework dedicated to understanding and preserving the diversity of biological organisms. Other organizations and projects like the Global Biodiversity Information Facility[15] (GBIF) and projects like OpenUp![16] aim to make available the vast amount of natural history collections data online.

As a project centered around the principles of open access and open data, the BHL proactively facilitates the linking of information throughout the cycle of scientific discovery and publication. The long term vision is to connect data from specimens collected in the field with field notes taken by collectors, to publications that result from the research cycle. In this way the information can facilitate biodiversity research more broadly than was possible with the information coded in printed books and physical collections.

References

1. Gwinn, N.E. & C. Rinaldo. The Biodiversity Heritage Library: sharing biodiversity literature with the world. IFLA Journal 35(1): 25-34 (2009).
2. Godfray, H.C.J., B.R. Clark, I. J. Kitching, S.J. Mayo & M.J. Scoble. The Web and the Structure of Taxonomy, Systematic Biology, 56(6): 943-955 (2007).
3. Minelli, A. The Status of Taxonomic Literature, Trends in Ecology and Evolution 18(2):75-78 (2003).
4. Garnett, T. Collaboration Trials and Triumphs: Biodiversity Heritage Library. Presentation at Yours Mine and Ours: Leadership through Collaboration, OCLC, September 20-21, 2010.

[13] http://grib.gbv.de/

[14] European Distributed Institute of Taxonomy (http://wp5.e-taxonomy.eu/)

[15] http://www.gbif.org/

[16] Opening up the Natural History Heritage for Europeana (http://open-up.eu/)