



# Biodiversity Heritage Library

Smithsonian Institution Libraries

[www.biodiversitylibrary.org](http://www.biodiversitylibrary.org)

The Ithaka S+R Case Studies in Sustainability were conducted in partnership with the Association of Research Libraries and were made possible by the generous support of the Institute of Museum and Library Services. A summary report, “Searching for Sustainability: Strategies from Eight Digitized Special Collections,” is available at <http://www.sr.ithaka.org/research-publications/searching-sustainability>. This specific case study was researched and written by Deanna Marcum, Managing Director, Ithaka S+R.

## ABSTRACT

The Biodiversity Heritage Library (BHL), created in 2006, is the result of a collaboration of ten natural history museum and botanical garden libraries seeking to digitize core taxonomic literature and to make it free and openly available throughout the world. Today, the BHL includes fifteen member institutions whose efforts have shaped a collection of over 60,000 titles. It is supported through a combination of membership dues, in-kind support from member institutions, contributions from the user community, and direct support from the Smithsonian Institution Libraries, and it reaches tens of thousands of users each year. While managing the complex partnership has not been easy, BHL offers an instructive model for multi-institutional, international collaboration.

## INTRODUCTION

The Biodiversity Heritage Library was created as a way to gather and share several major collections of taxonomic literature and associated materials as collected by participating institutions. As of June 2013, the Biodiversity Heritage Library included 60,202 titles, representing 114,336 volumes and 40.8 million pages. The collection is growing at the rate of approximately 1,500 volumes per month. According to the program director, the Library now contains approximately 30 percent of the extant biodiversity literature and approximately 15 percent of all biodiversity literature (including that under copyright). The BHL has reached agreements with 125 publishers to provide access to 273 titles that include materials under copyright.



The Biodiversity Heritage Library developed out of the recognition that scientists needed research materials that were difficult to locate and retrieve. By its nature, taxonomic research relies on the long tail of species descriptions to describe and organize life on earth. Museums had already started to digitize their specimen collections; libraries would make an enormous contribution to scientific scholarship by digitizing the scientific literature based on such specimens. By 2004, the Smithsonian Libraries had already made significant progress toward establishing a digital library of taxonomic literature, and the idea of developing an even larger digital collection, which was clearly needed by the scientific community,

was appealing. The digital library work done by other organizations that were to become partners in BHL, particularly the Missouri Botanical Garden with its Botanicus and Rare Books from the MBG Library: Illustrated Garden projects,<sup>1</sup> made the startup of BHL simpler. All of the eventual BHL partners furthermore found it a challenge to meet the needs of their users with their limited budgets, so the possibility of developing an important digital resource in collaboration with other institutions presented a significant opportunity. The primary focus at the outset of the project was to identify grant funding to begin digitizing and making accessible the collections of those partner institutions that had not already begun to do so.

The Biodiversity Heritage Library has developed beyond project status to become a service that researchers in systematic biology have integrated into their daily work. Its highly specialized nature means that it serves a well-defined audience, and the user feedback that is an essential feature of the website ensures that the resource meets the needs of the audience it is meant to serve. Beautiful illustrations of flora and fauna, made freely available via Flickr, will greatly benefit other disciplines and a general audience, but the constant focus of the BHL partners on scientists' needs is what defines the Library's success.

## BACKGROUND

The Biodiversity Heritage Library began as a grant-funded project. Initial meetings of scientists funded by The Andrew W. Mellon Foundation in 2003 identified the need to make taxonomic literature more widely available. Describing what is known as the "taxonomic impediment," the scientists pointed out that the literature, specimen, and archival collections upon which their research depends are housed in museums and libraries in large cities, and those vital collections are out of easy reach of many researchers, particularly when they are out in the field. Tackling the published literature housed in library collections, while no small task, was identified as an achievable first step toward relieving the burdens of this impediment.

The immediate outcome of these discussions was the funding, by the Smithsonian's Atherton Seidell Endowment, of the digitization of the *Biologia Centrali-Americana* (BCA). The BCA, published 1879–1915, is an important 63-volume compendium of taxonomic information on Central America and Mexico that exists in complete sets primarily in large research libraries. The BCA digital site was launched in July 2004.



*Mémoires sur les lépidoptères.*

St. Pétersbourg: Imprimerie de

M.M. Stassuléwitch, 1884-1901

Courtesy of the Biodiversity

Heritage Library.

Nearly concurrently, Tom Garnett of Smithsonian Libraries attended the February 2004 Telluride Institute symposium that proposed the creation of the Encyclopedia of Life. These meetings led to an international meeting, "Library and Laboratory: The Marriage of Research, Data, and Taxonomic Literature," held in February 2005 at the Natural History Museum in London and attended by over seventy individuals from libraries, the field of bioinformatics, and the biological research community. The outcomes of this meeting led directly to the formation of the BHL during an organizational meeting held at the Smithsonian Libraries in June 2006. Shortly thereafter, the BHL was included in the successful Encyclopedia of Life funding proposal to the John D. Rockefeller and Catherine T. MacArthur and Alfred P. Sloan Foundations.

The BHL used the funds it received through the MacArthur Foundation to provide funding for rapid mass scanning to begin in May 2007. Brewster Kahle's Internet Archive was selected as the primary scanning partner because of the Archive's commitment to open access a requirement of the MacArthur award, its proven expertise in mass digitization, the reasonable rates that were a consequence of the Archive's not-for-profit status, and its ability to ramp up quickly. BHL funding went to existing Internet Archive scanning facilities in Boston (Boston Public Library) and New York (New

York Public Library), and also to provide supplemental funding for the creation of the FedScan facility at the Library of Congress. Additionally, single scanning workstations were set up at the Smithsonian Libraries and the Natural History Museum Library (London).

The initial governance structure of the BHL included a council made up of representatives from the ten original member organizations. The members elected a chair, vice chair, and secretary from the membership. In February 2007, Tom Garnett, Associate Director of Smithsonian Libraries and a key figure in the creation of the BHL, was selected as the first BHL program director. The Smithsonian Libraries agreed to assume responsibility for the administration of the initiative (forming the BHL Secretariat), while the Missouri Botanical Garden took the lead on technical operations under the leadership of Chris Freeland, who was named the BHL technical director.

By 2012–13, at the time the project was coming to the end of its grant funding, there were fifteen member organizations, thirteen of which had been dues-paying members from the outset. The membership is now made up of the major natural history organizations in the world, including both academic institutions and public and private research and heritage institutions of all sizes.<sup>2</sup>



From *Illustrated natural history of the animal kingdom*. Courtesy of the Biodiversity Heritage Library.

Each of the participating institutions developed its own digitization plan, focusing on the materials from its collections that would be most valuable to the BHL project. This planning is facilitated by BHL collections coordinator, Bianca Crowley, located at the BHL Secretariat (Smithsonian Libraries).

The grant funding and contributed labor by the members of the BHL provided stability until 2012, when a series of changes occurred that forced a new way of thinking about the organization. The MacArthur funding ran out in July 2012. A no-cost extension of the MacArthur grant allowed remaining funds to be carried through July 2013, time that allowed for BHL’s development of a plan for transition to a sustainable model. Tom Garnett retired from Smithsonian Libraries in March 2012 (he was replaced as BHL program director by Martin Kalfatovic), and shortly thereafter Chris Freeland, BHL’s technical director, left the Missouri Botanical Garden to become Associate Director of Libraries at Washington University. Freeland was succeeded by William Ulate, also of the Missouri Botanical Garden. During late 2011, knowing that personnel and funding sources were in flux, BHL instituted a revised governance

structure that created a Steering Committee consisting of those BHL members paying annual membership dues of \$10,000. With this new structure in place at the start of 2012,<sup>3</sup> the group turned its attention to the long-term sustainability that was vital for an organization already proving to be invaluable to the scientific community.

## SUSTAINABILITY STRATEGY

The BHL is a good example of a collaborative project focused on a highly selective audience. There is a finite number of institutions holding the specialized collections that support research in biodiversity. Partnership is necessarily limited to organizations that house such collections; however, the BHL continues to identify and recruit new members who can contribute to its mission.

### Economic Model

One of the requirements of the MacArthur grant—the need to provide content for the Encyclopedia of Life—plus the strong request of the taxonomic community to “get as much content out there as possible” drove the BHL to concentrate more heavily in its first years on solving immediate technical and content problems than

on developing a business model for the organization. Nancy Gwinn, Director of the Smithsonian Libraries, was concerned about how the successful project, increasingly becoming part of the infrastructure of taxonomic research, would be sustained, and she made long-term sustainability and funding a priority at the start of her term as BHL chair.

BHL's partnership model is the cornerstone of its sustainability strategy. BHL partners think about funding the Library's operations from two different perspectives: funding for central activities, including the operations of the Secretariat and the technical team, and funding for operations of the partner institutions or "franchises," which contribute content and content-related work via their existing library staff.

The BHL's current funding comes from its member institutions, individual donations, and new grants. Revenue from each source is described below.

- **Membership dues.** Payment of \$10,000 per year in membership dues entitles institutions to participate in BHL at the Steering Committee level. With thirteen institutions currently participating at this level, the project garners \$130,000 per year from this source. Even though the partner institutions vary greatly in size, the group determined that there would be a single amount for annual dues: \$10,000 per institution was deemed enough to make an impact on the project's bottom line without being overly onerous for most of the partners.

The BHL continues to recruit members; Cornell University Library and the Library of Congress became Steering Committee members in late 2012. The BHL executive group believes that fifteen to twenty members would be an effective number to manage at the Steering Committee level. Other membership categories that would serve both the needs of BHL and of smaller institutions, and/or those unable to afford the \$10,000/year that would make them eligible to participate in the Steering Committee, were defined in May 2013.

- **In-kind contributions from member organizations.** BHL member organizations contribute staff labor amounting to 14.2 FTEs and other costs per year, or just over \$2 million. This includes 5.3 FTEs contributed by the Smithsonian Libraries, which translates to a financial investment of \$400,000 per year.

The Smithsonian has funded BHL through the Atherton Seidell Endowment (~\$300,000/year); Smithsonian Libraries also receives a small federal allocation for BHL as part of the overall federal support for biodiversity received by the Institution. Harvard's Museum of Comparative Zoology has also supported its participation with approximately \$300,000/year. The types and sizes of partner institutions mean that institutional support varies widely.

The Smithsonian Libraries has designated the BHL as a key priority. It has added the duties of program director to those of the associate director for digital services and converted this into a BHL position: BHL program manager. Smithsonian Libraries Director Gwinn has also named the BHL as the Libraries' priority for fundraising over the next eighteen months, as part of the entire Smithsonian Institution's national campaign.

- **Direct donations from BHL's user community.** BHL instituted an online donation option in late 2012. A small initial success with an end-of-year (2012) online donation campaign encouraged more proactive campaigns in FY 2013, and BHL set a goal for the year of ~\$10,000. To date, user community donations have totaled just under \$10,000, and the Library continues to receive small donations from donors around the world.

- **Grant support to member institutions.** Supporting foundations and agencies include the Moore Foundation, Lounsbery Foundation, and JRS Foundation, and the NSF, NEH, and IMLS.

To supplement the original MacArthur funding, BHL partners have found various ways to advance the BHL program. To support advanced technical development, the Missouri Botanical Garden actively and successfully sought grants from the Moore Foundation, the NEH, and the NSF that funded technical staff for BHL. Harvard's Museum of Comparative Zoology identified internal endowment funds to support the work of BHL staff at that location. Likewise, other partners have used internal subventions or grants from outside sources (including IMLS and the Mellon Foundation) to leverage library activities to support BHL work. The Smithsonian Libraries was fortunate in being able to apply funds generated by the Atherton Seidell Endowment, originally established to provide for reproduction and distribution of previously published scientific work via microfilming, to support BHL work at the Smithsonian.

### Public Benefit

#### Open, harvestable data

When the BHL was first organized, all participating institutions agreed that this resource must be freely and openly available to the world. After all, its whole purpose was to make the literature associated with taxonomic specimens available to a much broader community of scholars. At the moment, BHL is scanning only the taxonomic literature. Scanning of specimens is a separate digital project, outside the scope of the project. In the future, BHL has aspirations to scan more archival and manuscript material and to consolidate various archival field notes projects into BHL. Funders were originally skeptical, noting that Google has already digitized some of this content. What would be different about BHL? The greatest difference is that Google data are not open for data mining or mass downloads, both of which are freely and easily possible through BHL.

From the beginning, one of BHL's primary goals has been to expose data for reuse. The users of BHL requested machine and human harvestable data. As a result, BHL has been built as a massive dataset that contains bibliographic data, page images, and text that has been processed through optical character recognition software and that indexes scientific names. BHL provides its data through a number of mechanisms: data exports (MODS, EndNote, Bibtext, and text files), APIs, OpenURL, and OAI-PMH.<sup>4</sup>

BHL defines its success through the use made of the content by both humans and machines. Much of the use of the BHL comes from other databases and systems—services that are pulling in information from queries about species. Using the BHL's API that provides access to the species names found through BHL's taxonomic name-finding algorithm, these services look for species name occurrences and then pull BHL references into their own systems. For example, Tropicos users do not see BHL content directly, but through links in the Tropicos database. In statistics gathered from January 1, 2010 through March 29, 2011, BHL reported 1,068,586 visits from 233 countries/territories; 48.12 percent of these were from search engines, 33.47 percent from referring sites, 18.38 percent from direct traffic, and .03 percent from other sources.<sup>5</sup>

### Serving user interests

The primary users of this resource are researchers in systematic biology. A secondary audience is those teaching life sciences at the undergraduate and graduate level. The website encourages patron interaction, and all of the scanning operations are patron driven.

BHL has become aware through user statistics that there is a great deal of interest in the image collections included in the Library. In 2012, BHL received an NEH grant to develop support for data mining of BHL for

taxonomic illustrations. This tool will greatly increase the 80,000-plus illustrations identified as part of the scanning workflow and posted on Flickr. Based on feedback from the Flickr community and other users, BHL believes that these illustrations will be a rich resource for humanities researchers, as well as for the art and K-12 communities.

## KEY FACTORS OF SUSTAINABILITY

**1. A committed partnership.** For all of the smaller partner institutions and for the Smithsonian Libraries, the BHL is a highest-priority project to which they donate both money and time, for it is the digital platform that allows them to provide information resources to their user communities. Cornell, Harvard, and the Library of Congress, all comprehensive research libraries with a multitude of priorities, view the BHL as an important project for meeting the needs of a subset of their users.

The priority given to this project by the partners is evidenced by the staff time they contribute. Each institution participates in the project by scanning material from its own library collections, managing its own digitization workflows, and volunteering its own staff resources.

In addition, the BHL executive committee meets weekly by teleconference. The institutional members meet by telephone quarterly. There is an annual membership meeting where all participants are invited for in-person discussions. The location of the annual meeting rotates among the partner institutions.

**2. Global participation.** The BHL began as a United States/United Kingdom project, but the partner institutions are keenly aware of the global nature of the data it holds. Materials are in more than two hundred languages, and BHL has been working with other countries to ensure that the Library is converted to different national instantiations. With three years of funding from the European Union, the BHL Europe project was the first to launch, in 2009. The Chinese Academy of Science has developed a BHL China component, as have counterpart organizations in Brazil, Egypt, and Africa. BHL Australia, which was formed in 2010 with funding from the Atlas of Living Australia, has contributed both content and technical expertise in the design of the recently relaunched BHL portal. BHL Africa, under the leadership of the South African National Biodiversity Institute (SANBI) and the National Museums of Kenya, officially launched in April 2013. BHL global nodes in Southeast Asia, Japan, India, Russia, and South Korea are forward goals.

**3. User research resulting in interface changes.** The Biodiversity Heritage Library partners recognize that the resource must be first and foremost useful to the scholarly community. They are continually adding features to the website that make the content more useful to taxonomic scientists. The most recent improvements include:

- Updated design. The website's design has been upgraded (March 2013) to reflect the admired aesthetics of the BHL Australia portal and enhanced functionality of the US/UK portal.
- Article and chapter access. To date, over 92,000 articles and chapters have been indexed and are searchable within BHL. Additional articles and chapters will become available as the collections continue to be indexed.
- Open data enhancements: BHL's APIs, OpenURL interface, and data exports have been modified to include available article and chapter information.

- **Book viewer updates:** The BHL book viewer allows users to view multiple columns of pages on screen at once, to more easily navigate to a specific page within a book. Users can also view OCR text alongside page images, and for books that have been indexed, they can navigate directly to the articles or chapters within using a new Table of Contents feature that uses the article and chapter data mentioned above.
- **Improvements in PDF creation.** The custom PDF creation process allows users to select pages for PDF copying while in the book-viewer mode and more easily review the selected pages before creating the PDF.

## ONGOING SUSTAINABILITY ISSUES

- **Specialized content and finite primary audience.** The Biodiversity Heritage Library recognizes that it is a highly specialized resource; while there are some secondary uses of the collection, it is primarily for biologists who specialize in taxonomy. The major physical collections of taxonomic literature reside in a limited number of institutions in large urban areas, primarily in the developed world. There are, of course, many smaller natural history museums and smaller botanical gardens that may have some resources to contribute, but they are not likely to share the same interests and concerns as the current members of BHL. They are also unlikely to be willing or able to pay current membership fees. BHL is still in the process of figuring out how to address the free-rider problem in what is, essentially, an open resource, and this situation is a challenge to BHL sustainability. At the same time, providing some level of partnership appropriate to smaller institutions and inviting them to be “on the team” is important to the BHL, given its collaborative and participatory origins. Among the options being explored by BHL’s executive group is the possibility of allowing such institutions to participate by contributing staff time or outreach initiatives in lieu of cash payments for membership.
- **Fragile technical infrastructure.** BHL considers its major problem the fragility of the technical infrastructure. Technology has been the responsibility of the Missouri Botanical Garden. Providing an appropriate level of support for the technical infrastructure will need to be addressed by the wider BHL partnership. To that end, the program director and technical director have created a BHL Technical Advisory Group (TAG) comprised of key technical members from partner institutions. The TAG will work with the technical director and program director in planning for the current and future needs of the BHL platform. While there are some elements of the technical needs of BHL that can be more widely distributed, there are practical considerations and efficiencies in centralized core IT activities for this project.
- **Intellectual property rights.** BHL is committed to openness, but the current copyright environment is a major barrier to success. Using the legal counsel capacity at the Smithsonian, BHL actively negotiates with publishers for the right to include on the website materials that are under copyright. This is a slow and expensive—and ultimately unsustainable—process. Large institutions such as the Smithsonian and Harvard have legal counsel on staff, but this one digital project is unlikely to have priority. Currently, the Smithsonian is assuming responsibility for negotiating rights, but this may not be practical as a long-term solution.
- **Varied digitization standards.** Digitization standards have been an issue for the group. The standards established by the Internet Archive are effective for providing access to the materials, but digital copies made under the Archive’s auspices do not serve as true preservation copies. The Internet Archive workflow is unable to handle certain types of materials (due to size, fragility, rarity, etc.). These materials are digitized, whenever possible, through in-house partner scanning solutions. These materials scanned outside the Internet Archive are then ingested into BHL via either the Macaw software (developed at Smithsonian

Libraries based on an earlier software tool developed at the Missouri Botanical Garden) or through other digital library workflow tools (Cornell's workflow being the example here). These non-Internet Archive scanning workflows are usually in conformity with those of other large-scale digital library projects (e.g., in maintaining TIFF output, etc.).

- **Collaboration challenges.** The BHL has learned a great deal about the promises and the challenges of collaborative efforts. As current Program Director Martin Kalfatovic describes it, this initiative started as a science project, turned into a library project, and has now morphed into a social experiment. The organizational partnerships are the most challenging part of building and delivering this specialized library. Managing institutional attitudes and navigating through “me first” tendencies proved to be especially challenging when the grant funds came to an end and an agreement was struck, and a structure developed, to self-fund the project while continuing to pursue collaborative grants.

## LESSONS FOR OTHERS

1. **The importance of building trust among partners.** Time and experience with the collaborative effort have given the partners a level of trust that was not necessarily there at the beginning of the project. At first, partners were concerned about what the involvement of the largest among them, Smithsonian Libraries, would mean. Now, the partners are comfortable with the collaborative model and have a better sense of the project's ongoing costs. All of them realize that the success of this effort rests with the Smithsonian taking a strong funding and organizational role.
2. **Consistent support and institutional buy-in provides stability.** “Federalizing” the program through appropriations at the Smithsonian Institution has given stability to the project, although the Smithsonian Libraries must make a request for federal funds every year. Through this annual appropriation process, the project is reasonably assured of \$125,000 each year, but because BHL is a priority for the Smithsonian, the amount is sometimes greater (in 2012, the amount from the Smithsonian Institution was \$200,000). Nancy Gwinn believes that it will be necessary to devote another staff position at the Smithsonian Libraries to the BHL program in the near term. True sustainability, she is convinced, will come when this project becomes a regular part of the Smithsonian budget, and this seems likely because the goals of the BHL are intertwined with the mission of the Smithsonian Institution.

## APPENDIX

### Interviewees

Nancy Gwinn, Director, Smithsonian Institution Libraries

Martin Kalfatovic, Associate Director, Smithsonian Libraries and Program Director, Biodiversity Heritage Library

### Resources Consulted

BHL: Biodiversity Heritage Library, <http://biodiversitylibrary.org>.

Quarterly Reports, BHL, <http://biodivlib.wikispaces.com/Reports>.

## NOTES

- 1 Botanicus, <http://www.botanicus.org/>; Rare Books from the MBG Library, <http://www.illustratedgarden.org/>.
- 2 The member organizations are the Academy of Natural Sciences Library and Archives; American Museum of Natural History Library; California Academy of Sciences Library; Cornell University Library; the Field Museum Library; Harvard University Botany Libraries; Ernst Mayr Library of the Museum of Comparative Zoology; Library of Congress; Marine Biological Laboratory and Woods Hole Oceanographic Institution Library; Missouri Botanical Garden Library; Natural History Museum, London, Library and Archives; New York Botanical Garden Library; Royal Botanic Garden, Kew, Library and Archives; the Smithsonian Libraries; and United States Geological Survey Libraries.
- 3 At the March 2013 BHL Steering Committee meeting, the group's name was simplified to "BHL Members."
- 4 Trish Rose-Sandler, Keri Thompson, Constance Rinaldo, Martin R. Kalfatovic, and William Ulate, "Collaboration and Communication Tools Used by the Biodiversity Heritage Library: Refining Strategies for Success," JCDL '12: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries (New York: ACM, 2013), 361–62.
- 5 See "Homepage," BHL: Biodiversity Heritage Library, at <http://www.biodiversitylibrary.org>.

This project was made possible in part by the Institute of Museum and Library Services, LG-06-10-0189-10.



Copyright 2013 Association of Research Libraries. This work is licensed under a Creative Commons Attribution-Non Commercial-ShareAlike 3.0 United States License. To view a copy of the license, please see <http://creativecommons.org/licenses/by-nc-sa/3.0/us/>.