

Collaborating on open science: The journey of the Biodiversity Heritage Library

Jane E. Smith^{a,*} and Constance A. Rinaldo^b

^a *Natural History Museum, London, UK*

^b *Ernst Mayr Library Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA*

Abstract. The Biodiversity Heritage Library, BHL,¹ is an established and successful digital library, formed by a global consortium of natural history libraries, with engaged and enthusiastic users. The extensive partnerships, curated content, innovative tools and services, the ease of mining the data all combine to establish an open science resource that advances scientific progress through linking, use and reuse. The aim of BHL as stated on the web page is: “Inspiring discovery through free access to biodiversity knowledge. The Biodiversity Heritage Library works collaboratively to make biodiversity literature openly available to the world as part of a global biodiversity community. BHL also serves as the foundational literature component of the Encyclopedia of Life (EOL)”. BHL and EOL are linked via taxonomic names and bibliographies. BHL is linked in a similar way to the Global Biodiversity Information Facility (GBIF) and thus has broad exposure to scientists across the globe as well as a global public.

Keywords: Data reuse, digital library, open access, outreach, global partnership, Taxonomic Intelligence

1. Introduction

This paper describes how the Biodiversity Heritage Library (BHL) has become an established and successful digital library with engaged and enthusiastic users [4,5]. The extensive partnerships, curated content, innovative tools and services, the ease of mining the data all combine to establish an open science resource that advances scientific progress through linking, use and reuse. The aim of BHL as stated on the web page is: “Inspiring discovery through free access to biodiversity knowledge. The Biodiversity Heritage Library works collaboratively to make biodiversity literature openly available to the world as part of a global biodiversity community. BHL also serves as the foundational literature component of the Encyclopedia of Life (EOL)”. BHL and EOL are linked via taxonomic names and bibliographies. BHL is linked in a similar way to the Global Biodiversity Information Facility (GBIF) and thus has broad exposure to scientists across the globe as well as a global public. Both EOL and GBIF present bibliographies that lead back to the BHL portal. Additionally, content is represented in Europeana² and the Digital Public Library of America³ (DPLA). DPLA has harvested metadata for approximately 85% of the items in BHL, further broadening the audience for BHL collections.

*Corresponding author: Jane E. Smith, Natural History Museum, Cromwell Road, London, SW7 5BD, UK. E-mail: jane.smith@nhm.ac.uk.

¹<http://www.biodiversitylibrary.org/>.

²<http://www.europeana.eu/portal/>.

³<http://dp.la/>.

2. BHL content

The Biodiversity heritage Library (BHL) currently provides scientists, scholars, citizen scientists and the public free and open access to a critical mass of over 46.6 million pages of digitised text and grey literature on biodiversity. This content represents over 97,000 titles and 163,000 volumes, approximately 17% of the biodiversity literature.

BHL partners have worked collaboratively since 2005 to make biodiversity literature openly available to the world as part of a global biodiversity community. The partners and contributors include libraries, natural history, botanical and research institutions that collectively hold a substantial part of the world's published literature and original material including published scientific papers and books, grey literature such as field notebooks and extraordinary illustrations all related to biodiversity. The digitized pages and associated metadata have been produced through scanning centres operated by technology partner, Internet Archive and through local institutional digitization efforts. The core audience for BHL is scientists, particularly taxonomists who need access to literature spanning all publication years from pre 1700 to currently published material that is often related to active specimen collections. The published literature on biological diversity is rare or has limited global distribution and much of it is available in only a few libraries. Figure 1 shows the subject areas relevant to BHL.

From a research perspective, these collections are of exceptional value because the domain of systematic biology depends upon the historic literature as much as the recently published. Yet this wealth of knowledge was, until BHL's development, available only to those researchers who could gain direct access to significant Library collections. Until recently, to access this content scientists would have to spend considerable time travelling to libraries and museums with rare and unique collections and this slowed scientific study and collaboration. Literature about the biota existing in countries outside the

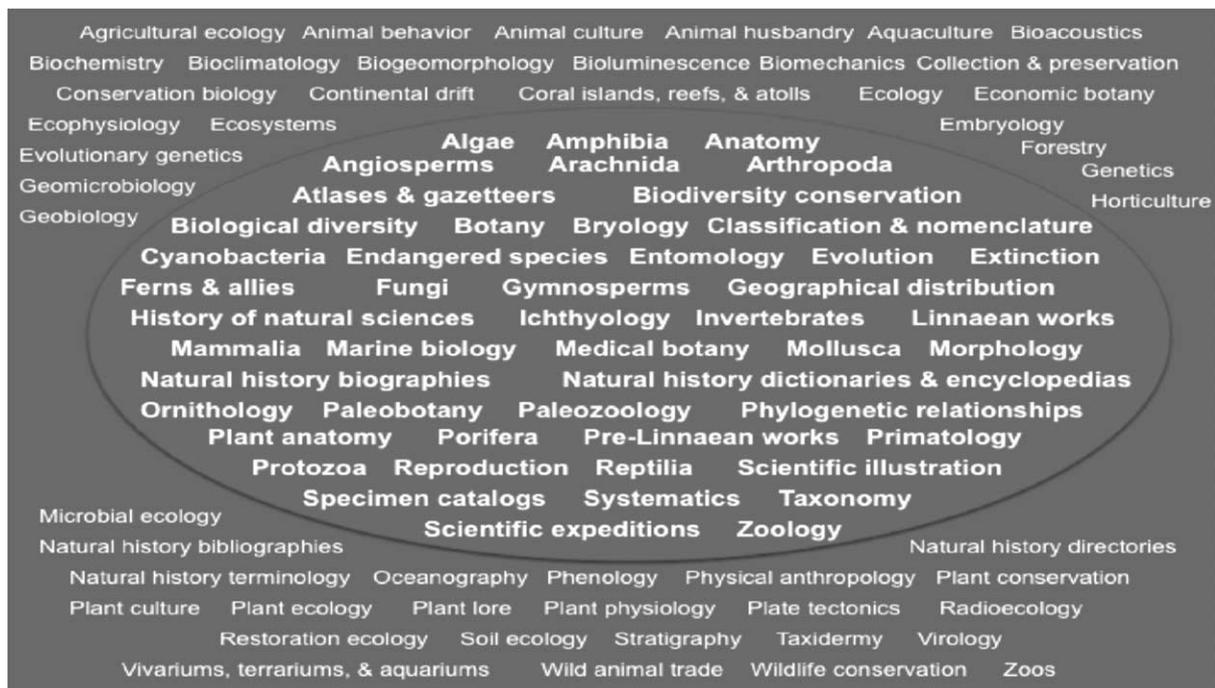


Fig. 1. Primary collecting areas for BHL. Core collecting subjects are seen within the inner circle.

western hemisphere was often not available within their borders. Biologists considered that access to the published literature was one of the chief impediments [3] to the efficiency of research in the field. Among other outcomes, free global access to digitized versions of the literature has repatriated information about the earth's species to all parts of the world [4].

BHL partner libraries addressed the challenge of this “taxonomic impediment” [3] by building a digital library of biodiversity literature, designed for taxonomists and systematists and utilising their language of taxonomic names for accessibility. There are more than 157 million scientific names that have been identified in BHL. All bibliographic and taxonomic data are free to download, re-use and remix thus providing open data services to anyone, anywhere.

3. Copyright

While the bulk of the BHL repository includes public domain literature, by negotiating with publishers and other rights holders, BHL also provides relevant literature that is current and in-Copyright. Many Natural History and related Learned Societies have contributed content directly or provided permission for their titles to be digitized by BHL partners for inclusion in the repository. Doing so has exposed the content of these often specialised publications to a much wider audience. To date, nearly 400 publishers have given permission for the inclusion of their titles.

Copyright impacts our ability to contribute and ingest content and is complicated by the variation in legislative frameworks within countries and regions across the globe. Therefore, the BHL partners have been careful to establish a Copyright and licence framework that reflects those regional differences while providing a common approach to which all contributors can agree. It is a framework that can be updated as the Copyright landscape changes. The framework supports the open access philosophy underpinning BHL, but also aims to respect the rights of contributors including an explicit take down policy. For example, an early piece of work undertaken by the BHL-Europe regional members, funded by the EU, was the development of a Copyright framework and guidance for EU partners intending to contribute content. BHL partners follow due diligence practices before digitizing titles and written permissions are actively sought from rights holders for material still in Copyright. Out of Copyright material within BHL, in the Public Domain can be used, with appropriate attribution, or reused for multiple purposes. Reuse of the material in-Copyright is subject to Creative Commons Attribution Non-Commercial Share Alike Licence.

4. Supporting research

“I’ve been working on a large synonymy project, and could not have done it without the BHL. In fact, I’ve been working on this since 1993, and made more progress in the past 3 years than in all the time before.” R. Marcelo.

By continuing to add relevant content, curating that content and enhancing access through the application of innovative tools, BHL supports scientists and researchers in multiple disciplines in their day-to-day work and collaboration with others. The availability and reusability of the scientific data accessible via BHL ranges from taxonomic study and training, biodiversity research, biodiversity conservation and maintenance of diverse ecosystems, climate change, animal and plant disease control through to audiences beyond core science constituencies, including historical and cultural research, exploration, and global commerce.

In addition to the core Science audiences, it became rapidly apparent that other users were drawn to the biodiversity treasure trove in the BHL. The heritage natural history volumes are rich in extraordinary illustrations that are coveted for study and re-use by art historians, citizen scientists, commercial artists and designers and the general public. Innovative projects and services result from interest in illustrations, field notebooks and journals formerly found only in museum and library archives. For example, the Art of Life Project has developed an algorithm to extract illustrations within volumes and provide enough metadata to make them discoverable [6]. These illustrations have been used to make greeting cards and wedding invitations, as well as for scientific study. Another example of how BHL data can be enhanced is the development of a game to aid in the crowdsourcing of transcription for handwritten field notes and complex seed and nursery catalogues.

The BHL technical team is working with collaborators to bring new perspectives to the interpretation of the material held in BHL. A current example, using crowdsourcing, is the joint project Science Gossip, a collaboration among BHL partner Missouri Botanical Gardens, *Zooniverse* and the UK's Arts and Humanities Research Council (AHRC) funded Constructing Scientific Communities: Citizen Science in the 19th and 21st Centuries. The project is an investigation into the making and communication of science in the Victorian period and today, and in the process provides metadata to enable discovery of illustrations from BHL.⁴

In addition, tools have been developed so that citizen scientists and the general public can consult virtual exhibitions of curated content on broad interest topics such as exploration, spices and women in science.⁵ Other projects support data mining and the improvement of Optical Character Recognition (OCR).⁶

5. User engagement and feedback

User engagement and feedback is critical to the advancement and sustainability of BHL. As well as supporting communication and collaboration in a virtual organisation, BHL has a highly developed social media strategy, which encourages user access and engagement. BHL is well connected and respected within the blogosphere, Twitter, Pinterest, Flickr and Facebook communities. Campaigns such as "Monsters are Real"⁷ attracted attention from the general public and the news media. Social media is also how BHL shares the outcomes from the feedback received through issue-tracking software on the website with a direct avenue for user comments. All feedback receives an answer. Content additions are welcome through the BHL portal's feedback tool and in-scope requests are added to the scanning queue. If there are scope concerns about a request, the BHL Collections Committee will review the request and recommend an action. User surveys and conference panels provide opportunities for targeted feedback and guidance in the development of services and content [2].

6. Sustainability

Initially BHL was funded by the John D. and Catherine T. MacArthur Foundation and Alfred P. Sloan Foundation through EOL, and the Gordon and Betty Moore Foundation for Fedora integration by the

⁴<http://www.sciencegossip.org/#/classify>.

⁵<http://latinonaturalhistory.biodiversityexhibition.com/>.

⁶<http://miningbiodiversity.org/>.

⁷<http://blog.biodiversitylibrary.org/2014/10/monsters-are-real.html>.

Missouri Botanical Gardens. The sustainability of BHL depends on a mixed funding model, including direct support by BHL partners, single and jointly awarded grants, and more recently a dues structure for members and fees for services to non-members to support basic administrative costs, content preparation and deposit assistance when needed. BHL operates as a virtual organisation and its strength is the participation and long-term commitment of the members through alignment of strategic goals of the constituent institutions and the contribution of collection and technical expertise from across the globe. Organisations that participate in BHL may provide substantial day to day staff support to improve and maintain the digital collections, support travel to conferences and BHL meetings across the globe, and support digitization of collections directly as well as through the dues mechanism.

7. Conclusion

“BHL is radically changing the status quo and democratizing access to knowledge about biodiversity,” lauds Dr. Sullivan. “Now anybody in the world has instant access to the original species description in a couple of clicks!” [1].

Working in partnership, with a common approach, has enabled the participating organisations to bring together and link their collections in ways that provide a more complete research resource for the scholarly community. By providing the linked content and tools to interact with the data, the audience for BHL has expanded beyond academic scientists to the academic humanities, citizen scientists and artist communities. Additionally, the prominence and focus of BHL has attracted publishers and other rights holders, allowing BHL to negotiate more easily to include material still in Copyright. Collaboration on standards, best practice and infrastructure solutions has enabled higher quality images, metadata and support tools to be produced, long term digital storage solutions to be achieved and the sharing and cost reduction of scanning operations and best practices. The development of a digital library may seem an ordinary task in 2015 but the BHL is unusual and exciting because by envisioning the initial goal as providing open data services and promoting open science as a direct response to user demands by focusing on the language of biological taxonomy, BHL has become a valuable global tool beyond its primary audience.

Acknowledgements

We would like to thank our colleagues, past and present, for all of their work, and discussions that have contributed to the Biodiversity Heritage Library and our presentation, particularly, the BHL Secretariat: Martin Kalfatovic, Carolyn Sheffield, Grace Costantino and Bianca Crowley.

References

- [1] G. Costantino, We need books to save Biodiversity, Vol. 12, 2014, BHL blog post, <http://blog.biodiversitylibrary.org/2014/12/we-need-books-to-save-biodiversity.html>.
- [2] G. Costantino, B. Crowley, R. Morin and E. Thomas, Heeding the call. *User* **40**(4) (2011), 146–157. doi:10.1515/mdr.2011.019.
- [3] Global Taxonomy Initiative, Convention of Biological Diversity <https://www.cbd.int/gti/problem.shtml> (described in 1992).

- [4] N.E. Gwinn and C.A. Rinaldo, The Biodiversity Library: Sharing biodiversity with the world, *IFLA Journal* **35** (2009), 25–34.
- [5] C.A. Rinaldo and J.E. Smith, Moving through time and culture with the Biodiversity Heritage Library, in: *Migrating Heritage: Experiences of Cultural Networks and Cultural Dialogue in Europe*, P. Innocenti, ed., Ashgate Publishing Group, Surrey, 2014.
- [6] T. Rose-Sandler, N.E. Gwinn and C.A. Rinaldo, The art of life: Merging the worlds of art and science. Libraries, citizens, societies: Confluence for knowledge. In session 149 – Art libraries with science and technology libraries, in: *IFLA WLIC*, 16–22 August 2014, Lyon, France, 2014.