Collaboration and Communication Tools used by the Biodiversity Heritage Library: Refining Strategies for Success

Trish Rose-Sandler
Center for Biodiversity Informatics, Missouri Botanical Garden
4344 Shaw Blvd.
St. Louis, MO 63110
1 314-577-9473 x6396
Trish.rose-sandler@mobot.org

Keri Thompson
Smithsonian Institution Libraries
NMNH, 10th and Constitution Ave NW
Washington, DC 20560
1 202-633-1716
thompsonk@si.edu

William Ulate
Center for Biodiversity Informatics, Missouri Botanical Garden
4344 Shaw Blvd.
St. Louis, MO 63110
1 314-577-9473 x6398
William.Ulate@mobot.org

Constance Rinaldo
Ernst Mayr Library of the Museum of Comparative Zoology, Harvard University
26 Oxford St.
Cambridge, MA 02138
617-495-4576
crinaldo@oeb.harvard.edu

Martin Kalfatovic
Smithsonian Institution Libraries
NMNH, 10th and Constitution Ave NW
Washington, DC 20560
1 202-633-1705
kalfatovicm@si.edu

ABSTRACT

Through the application of multiple strategies and tools, the Biodiversity Heritage Library has created an effective and collaborative multi-institutional virtual organization. The purpose of this paper is to explore the communication and collaboration strategies used by the BHL to create, maintain, and provide open access to its corpus of biodiversity literature. BHL, in its seventh year, is a mature service and no longer a pilot project. Largely driven from the ground up, and without any institutional mandate, the BHL has successfully and organically fostered an organizational model that has encouraged innovation, user engagement, and global expansion.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]:User Issues

Additional Categories:

K.4 [**Computers and Society**], K.6.1[**Project and People Management**]

General Terms

n/a

Keywords

communication, collaboration, digital library, biodiversity, libraries, biodiversity heritage library, social media, subject repository

1. INTRODUCTION

The purpose of this paper is to explore the communication and collaboration strategies used by the Biodiversity Heritage Library (BHL) to create, maintain, and provide open access to its corpus of biodiversity literature. We will review how, through the application of multiple strategies and tools, BHL has created an effective and collaborative multi-institutional virtual organization. We will also discuss the tools and strategies employed by BHL to communicate with users in order to leverage their experiences and feedback to improve and maintain the content and functionality of the digital library.

The Biodiversity Heritage Library is a global initiative of collaborating projects and activities that seeks to digitize, preserve and make available to the world the legacy literature of biodiversity. What started as a consortium of 10 natural history, research and botanical libraries in the United States and the United Kingdom in 2005 has now grown to 14 members as of January 2012. Additionally, BHL now includes collaborating projects in Europe, Australia, China, Brazil, and Egypt with over 40 individual institutions as members. The BHL portal[1] is the primary access point through which the BHL resources are served and the content now encompasses more than 37 million pages from around 53,000 titles.

The idea was then, and still remains, to support biodiversity related sciences by providing free and open access to biodiversity science publications such as Linnaeus's *Systema naturae*[2] and Darwin's *Origin of Species*[3] along with hundreds of thousands of other works describing the biodiversity of species. BHL has received strong support both in the library and the biodiversity

[1] http://www.biodiversitylibrary.org/
[2] http://biodiversitylibrary.org/page/6981741
[3] http://biodiversitylibrary.org/item/18405

science communities. Now entering its seventh year, the BHL is no longer a pilot project but a mature service which researchers in systematic biology have come to rely on for their daily work. Reaching this important milestone provides an opportunity to reflect on how collaboration and communication both amongst its members as well as with its users has played a key role in the growth and continued enhancement of this digital repository.

2. COLLABORATION AND COMMUNICATION WITHIN THE ORGANIZATION

2.1 BHL administrative structure

Largely driven from the ground up, and without any institutional mandate, the BHL has successfully and organically fostered an organizational model that has encouraged innovation, user engagement and global expansion. While not a legal entity, it is a loose consortium of members, all committed, through a Memoranda of Agreements, to support the goals of the BHL. The heads of these libraries form the institutional council, which provides strategic direction, reviews current issues, monitors the budget, and has an annual face-to-face meeting as well as frequent conference calls. An elected executive committee consisting of a Chair, Vice-Chair, and Secretary/Treasurer meets weekly via conference call along with the BHL Program Manager and Technical Director.[2] As BHL has matured, the governance structure has shifted. We now have an executive steering committee of dues-paying members, along with the elected executive committee, the institutional council, and an additional global governance structure.

The global structure allows partners from all over the world to determine how to achieve global BHL goals within their own cultural, technical and political contexts. With the addition of global members, yearly face-to-face international meetings were added, which focused on topics specific to global coordination such as content mirroring, portal development, management, and oversight. This global BHL business model has the potential to create a sea change in natural history institutions. By establishing these global nodes, biodiversity literature networks within larger geographic regions will be more firmly established. It should be noted that while some grant funding is available for BHL conferences and meetings, member institutions are expected, when possible, to provide some degree of financial support to their staff for attendance.

2.2 Strategies for communication among BHL staff

Though the BHL has a formal governance structure, the coordination of the routine activities of the organization have from the beginning occurred in a more informal, non-hierarchical way. The organizational culture has always been highly collaborative due in part to the common missions of member institutions, shared core user groups, and shared values[1][2]. To communicate across a large multi-institutional virtual organization, multiple methods and tools are used from traditional face-to-face meetings and conference calls to common, generally web-based, tools such as wiki, Skype, and Doodle -- an absolute necessity for scheduling conference calls within an organization that includes more than 50 institutions and spans over 9 time zones!

Initially, BHL communications developed around the immediate need to rapidly develop systems, select standards, and coordinate content creation in order to meet grant-driven deadlines. Multi-day face-to-face meetings of key staff from the original 10 institutions met to work through the core issues. Subgroups, made up of area experts, coalesced around particular topics such as collections development, metadata, and technical development. An effort was made by group members to ensure that multiple institutions were represented in each subgroup, and decisions made by them were readily adopted by the members at large.

Discussions continued after the in-person meetings via conference or video calls, using the internal wiki to record meeting minutes, ideas, and decisions. From these early meetings grew a tradition of yearly meetings for all levels of BHL staff. BHL members also took every opportunity to have a BHL-focused meeting, even if it were limited to two staff members in a casual setting, e.g., in between sessions at a library conference. The informal brainstorming of solutions or discussions by those staff would then be passed up to the larger group, or to focused smaller groups.

As the organization grew and matured, additional periodic communication methods were adopted. Weekly or monthly conference calls were scheduled for the formal leadership as well as for technical staff and collections focused staff. Agendas, minutes and action items were all recorded on the wiki. The wiki has been a key tool in preserving and providing access to policies, best practices and general information about the BHL. Particularly as additional members have joined, the ability to refer them to existing documentation has made their transition easier. To ensure continued efficient and effective operation of the BHL, staff from new institutional members are actively included in the existing communication channels and discussions as they join.

2.3 Strategies for collaboration among BHL staff

As a multi-institutional, multi-member project, BHL has the advantage of being able to draw upon individuals with a wide range of knowledge and skills. Collaboration across the members has been a key success factor. Member institutions share tools, best practices, workflows, and sometimes resources such as scanning facilities. Shared tools are developed by institutions who have the programming and hosting resources. Though core systems are located in one institution (the Missouri Botanical Garden) other members have taken on the responsibility of creating additional tools[3] needed for workflow management in order to distribute the technical workload and, in the initial rush to meet grant-related deadlines, enable members to begin creating content rapidly. The Marine Biological Laboratories/Woods Hole Oceanographic Institute Library rapidly developed an online tool to aid in deduplication of monographs before scanning. Natural History Museum-London developed a quasi-union catalog of serials that also aided deduplication of effort by allowing members to claim or “bid” on serial runs they intended to scan. The serials tool has since been expanded to include additional member institutions and is now managed by BHL-Europe.

Best practices, such as procedures for doing quality assurance or standards for how to record enumeration and chronology of journal volumes, are generally completed through a shared workload model. Often one institution or staff person will take the lead on a particular issue based on their area of expertise or availability. Discussions in person or on conference calls

involving interested staff are organized, and deadlines are set for making decisions, usually of no more than two to three months. Once the best practice is agreed upon by members, it is recorded on the wiki.

Though in the beginning workflow tasks were coordinated via phone or email, as the BHL grew larger and the volume of scanning greater, this became laborious. The commercial-off-the-shelf (COTS) issue-tracking system, Countersoft's Gemini, has been employed as a key tool in BHL's collaboration strategies[4]. While the adoption of this type of tool is not new in digital libraries, the manner in which it is used by BHL is somewhat novel. Rather than managing software development tasks, it is used instead as a communication and collaboration hub: tracking user requested digitization, assigning "gap fills"[4] as well as coordinating correction and enhancement of metadata. Each institution is encouraged to devote some staff (or volunteer and intern) time to enhancing records they've contributed, usually through "paginating" (adding page-level metadata), title-level record cleanup and merging, or through volume enumeration. Support of BHL through dedicating staff time to improve content or through supplying other resources, such as content mirroring, is one way that institutional members who can't afford to directly contribute funds to central services can help ensure the continued viability of BHL.

3. COLLABORATION AND COMMUNICATION WITH USERS

3.1 Communication strategies with users

Our user communication and collaboration strategies are a key part of the larger strategies we employ as an organization to maintain and improve digital library collections and services. Our experience has taught us that creating content and basic platforms are relatively easy compared to the prospect of preserving, expanding, and improving content and services into the future. Communicating actively with users and then acting on the feedback they give us are two key elements in the ongoing success of BHL, particularly in how we grow and maintain collections.

BHL has a core fulltime staff of twelve[5], based throughout the member institutions, that are funded to work only on BHL. All member institutions, however, contribute the time and effort of at least a portion of their regular staff in support of BHL. The contributed staff time from all member institutions has been estimated at approximately eighteen FTEs total. Two of the core staff share responsibility for managing centralized social media accounts and communications on platforms such as Facebook, Twitter, and Flickr, as well as on our blog and public wiki. In addition to the centralized communication channels, some staff from throughout the consortium also actively seek feedback from users and disseminate information about the BHL using individual accounts on these platforms. Though using social media to create a dialog with users is not a novel concept, BHL relies on active communication with our users as a key strategy in the development and ongoing maintenance of our digital library. Feedback is used to refine collections, find and correct metadata errors, improve the interface and system functionality, and prioritize development of features and enhancements.

In order to leverage user feedback we have integrated social media communications (along with feedback received via email and traditional user surveys) into our central communication management system, Gemini. All scanning and functionality requests, anecdotes about BHL's value to their work, or found errors submitted by users are entered into this system where they can be assigned to staff throughout the consortium for resolution and monitored for actions taken or further communication. Reports can be generated from the system to ensure that we are acting on user-supplied information and requests in a timely manner, as well as to analyze types and frequency of requests and error reports. Anecdotes from users managed through the system are also shared with the Executive, Steering, and Institutional Council and publicly through quarterly reports.

BHL also tries to maintain a dialogue and create trust[5] with our users through posting user and staff profiles on our blog. These monthly features let users tell us how they conduct their research, how they use our content, what they need from us, and how they and their work benefits from our existence. It is a valuable source of information for making decisions about content, functionality, and development of BHL. We also post staff profiles to highlight roles and workflows within the organization and help users understand the processes that go into creating and maintaining the BHL.

As we learn more about our users we target specific social media platforms to reach distinct user communities. Many of our most vocal research-focused users, particularly data informaticians and taxonomists, are active on Twitter, so we carefully monitor Twitter and respond quickly to comments and questions. Increasingly we are gaining other types of users, such as educators and artists, who have found our content through Flickr and prefer to communicate via email. We are putting additional resources towards increasing the amount of content we have on Flickr for these users, as well as creating an email newsletter appropriate for general-interest audiences.

3.2 Collaboration strategies with users

To be a part of the global "biodiversity commons" means going beyond merely providing a portal for 37 million plus and counting pages of historic texts to users. One key goal is to expose data for reuse. Many BHL users requested machine and human harvestable data. This allows data which was created for a specific purpose and audience (e.g. historic texts, nomenclatural services, encyclopedias) to interact with other data and serve new, previously unimagined, roles.

BHL is a massive dataset: not only of bibliographic data but also page images (often extraordinarily beautiful illustrations), text that has been run through optical character recognition (OCR) software, and of particular note, scientific names that are useful in multidisciplinary research (e.g. systematics, natural language processing, humanities). Scientific names found in the historic literature demonstrate the power of open data because they are an important access point for researchers in biodiversity. When doing

[4] a "gap fill" is when one or more volumes are seen to be missing from a run of a serial title digitized by one institution, the missing volumes are then located in another institution's collection and a request is made to that institution to contribute scans of those missing volumes.

[5] Fulltime staff include the Project Director, Project Manager, Technical Director, Collections Coordinator, Global Coordinator, Lead Programmer, and Data Analyst. These staff are supplemented by five staff at Smithsonian Libraries and the Museum of Comparative Zoology Library who are funded by their institutions solely for BHL related work.

research on a taxon, it often is necessary to know when the first mention of that species occurred in the published literature. The first publication establishes a species name, although names can change as research intensifies. Each page in BHL has been OCR'd and that text is run against a webservice available via UBio's Taxonfinder[6] which has identified 80 million scientific names in our corpus. Once identified, names can be used to generate a bibliography of all pages within BHL's books and journals that have made reference to a term such as "Zea mays" (scientific name for corn). The OCR'd pages, names files, and bibliographic references are saved and made available to anyone for downloading and reuse.

BHL has multiple paths for providing its data. These include data exports (MODS, EndNote, Bibtex, text files); APIs (a web service invoked via HTTP queries with responses in JSON, XML, or XML wrapped in a SOAP envelope); OpenURL (citation request made to a resolver that returns results in form of JSON, XML, and HTML); and OAI-PMH (records are provided in MODS and Dublin Core formats).

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 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).

Numerous projects have taken advantage of this "open data" and some have even developed applications that recontexualize BHL data such as the Encyclopedia of Life, BioStor, and Synynyms.

3.2.1 Encyclopedia of Life

BHL was initially funded to serve as the primary literature component of the Encyclopedia of Life[7] (EOL). This funding resulted in the large scanned corpus that exists today and enabled bi-directional linking - from species page in EOL to literature in the BHL and from pages in the BHL portal that mention a species directly back to the EOL page that expands on the knowledge of that species. These linkages are accomplished using the taxonomic tools provided by uBio and its webservices such as taxonFinder.

3.2.2 BioStor

Rod Page, a science researcher based in Glasgow, exploits data exports and APIs from BHL by grabbing metadata, page images, OCR'd text, and names files in order to build applications that give context to the data. For example, Page's BioStor[8] application provides data at a greater level of granularity than in the BHL portal, particularly at the article level. The BHL portal is discoverable primarily at the taxonomic name, volume or journal level. A citation search within BioStor will expose a specific article if it exists in a volume from the BHL data.

3.2.3 Synynyms

Ryan Schenk is a web developer at the Marine Biological Laboratory in Woods Hole, MA whose application called Synynyms[9] provides visualizations of taxonomic name changes. Organisms are known at any point in time by a particular scientific name but these names may be disputed and can change with new research. Schenk's application extracts species name synonyms from EOL, searches BHL for publications that mention these names, and then uses the data to illustrate the popularity of an organism's different names in the published literature over time.

4. CONCLUSION

The BHL has employed multiple collaboration and communication strategies over its seven year of existence. Its loosely coupled operational structure allows for nimbleness of decision making (i.e. recommendations do not necessarily need to go through multiple hierarchical levels of decision making in order to be turned into action items). As a result, BHL has been able to respond to its users and the needs of the biodiversity community much more quickly.

Adoption of collaboration tools among BHL partners, both freely available and commercial, and employment of those tools in sometimes non-traditional ways, has enabled a virtual organization to perform as effectively as a non-virtual organization. Varying time zones became less important as BHL found ways to communicate with its members in both synchronous and asynchronous ways. In recognition of its outstanding collaborative partnerships, the BHL was awarded by the Association for Library Collections & Technical Services, a division of the American Library Association, the Outstanding Collaboration Citation Award in 2010[10].

Experimenting with multiple social media platforms simultaneously has allowed BHL to reach diverse audiences, receive user input through multiple channels, and engage with users in a way that simulates in-person interactions. User input is given equal weight as input from BHL staff and oftentimes users' suggestions result in newly available content or user interface changes within days or weeks rather than months or years. In this way, users feel they are active contributors to the BHL rather than just passive users of its content and become much more invested in its success and longevity.

Not satisfied with just building a silo of biodiversity content, BHL wanted to push that content out into the biodiversity commons to allow it to be recontextualized and built upon by others working in similar domains. This openness of data has been well received by the community and users have done some commendable work giving it new life beyond its original form of publication. BHL is now looking to bring some of that enhanced content, for example articles derived from BHL journals, back into the portal to further enrich its services. In this way the crowdsourced data completes an object's "digital content life cycle" whose primary processes include: creation, management, discovery, use and reuse. [6]

We acknowledge there are, what may seem, some weaknesses in our organizational model, such as, redundancy and duplication of effort, because our priorities were often driven by the varying funding streams for BHL. But in considering them within the model, by making intentional choices to accept these overlaps to achieve cost efficiency and not derail timelines, we have turned these issues into strengths that provide resilience to the model. In

[6] http://www.ubio.org
[7] http://eol.org
[8] http://biostor.org
[9] http://synynyms.no.de/

[10] http://www.ala.org/ala/mgrps/divs/alcts/resources/ano/v21/n3/feat/system.cfm

the future, the BHL will continue to investigate new collaboration and communication strategies, abandon those it feels no longer serve its mission and adopt and even develop new tools as needed for the community we serve.

5. ACKNOWLEDGMENTS

The authors would like to thank all of our BHL colleagues for sharing their expertise so freely in all venues.

6. REFERENCES

[1] Waibel, Gunter. 2010. Collaboration Contexts: Framing Local, Group and Global Solutions. Report produced by OCLC Research.http://www.oclc.org/research/publications/library/2010/2010-09.pdf

[2] Gwinn, N.E. & C. Rinaldo. The Biodiversity Heritage Library: sharing biodiversity literature with the world. *IFLA Journal* 35(1): 25-34 (2009).

[3] Pilsk, Suzanne C., Person, Matthew A., DeVeer, Joseph M., Furfey, John F. and Kalfatovic, Martin R. 2010. The Biodiversity Heritage Library: Advancing Metadata Practices in a Collaborative Digital Library. *Journal of Library Metadata*, 10: 136-155. DOI:10.1080/19386389.2010.506400

[4] Costantino, G. , B. Crowley, R. Morin, and E. Thomas. 2011. Heeding the call: user feedback management and the digital library. *66th IFLA Council and General Conference*, San Juan, Puerto Rico.

[5] Schrier, Robert A. 2011. Digital Librarianship & Social Media: the Digital Library as Conversation Facilitator *D-Lib Magazine*, July/August 2011 Volume 17, Number 7/8. DOI:10.1045/july2011-schrier

[6] Ooman, Johan and Lora Aroyo. 2011. Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. http://www.cs.vu.nl/~marieke/OomenAroyoCT2011.pdf