

ABSTRACT: Foundations to Actions: Extending Innovations in Digital Libraries in Partnership with NDSR Learners

The Biodiversity Heritage Library (BHL), led by the Ernst Mayr Library of the Museum of Comparative Zoology, Harvard University (MCZ), proposes to host a National Digital Stewardship Residency (NDSR) cohort for the 2016-2017 year. The formal partners are the Field Museum of Natural History (FMNH), Chicago Botanic Garden (CBG), Missouri Botanical Garden (MOBOT), Natural History Museum of Los Angeles County (NHM) and the Smithsonian Institution (SI). The BHL NDSR cohort would support five geographically distributed residents, graduates of LIS or related master's programs, in a collaborative project to improve tools, curation and content stewardship of a key digital life sciences library: BHL. Outcomes will include a best practices guidelines document for digital libraries incorporating transcriptions, image searching, collection analysis techniques and better connections to museums, archives and other relevant databases. Better functionality for the BHL, based on user needs, will feed into other data systems and provide enhancements for BHL biodiversity and library partners such as such as the Encyclopedia of Life (EOL), Global Biodiversity Information Facility (GBIF), Digital Public Library of America (DPLA), and Europeana. This project will be completed in 2 years from June 1, 2016 to May 31, 2018, with the residents at the host institutions for 12 months Jan. 1, 2017 - December 31, 2017.

Although the focus of BHL is on Life Sciences, there are connections to history, art and culture. For instance, all of the grant participants are associated with cultural heritage museums that incorporate more than biodiversity into their exhibits, research and education activities. This series of resident projects is an opportunity to develop a bridge to connect museum and library/archives collections.

Each host institution will provide mentorship to a resident for a specific project designed to improve the functionality of BHL and identify how tools and processes may be transferred to or from other digital library and museum environments. The Chicago partners' resident will address content and gap analysis, reviewing the domain of literature underpinning the field of biodiversity, the amount of that literature in the public domain, and explore methodologies to scope and visualize digital collections. The MCZ resident will build on work done with MOBOT to develop methodologies and propose tools for verification and integration of crowdsourced data corrections. The MOBOT resident will explore user interface modifications to the BHL portal to enable image searching, browsing and display. The NHM resident will consult with BHL partners such as DPLA and Europeana to determine how BHL data works in these large scale digital libraries and categorize high value tools and services. Finally, the SI resident will work with the BHL Secretariat to to determine additional digital library needs and services that will provide increased value to a broader set of BHL users.

TITLE Foundations to Actions: Extending Innovations in Digital Libraries in Partnership with NDSR Learners

STATEMENT OF NEED

“There is a broad need to improve all librarians’ and archivists’ computational literacy. At the same time, there is a particular need to develop a cadre of library and archives software developers, engineers and digital project managers who can contribute to the core development, management, and direction of software and services.”¹

This proposal addresses the IMLS priority of a national digital platform, and the concepts of user engagement and collaboration beyond libraries, by using the Biodiversity Heritage Library (BHL) as a test bed for improving accessibility and connecting data among a variety of data providers. Biodiversity informatics organizations and their users such as Global Biodiversity Information Facility (GBIF) and the Encyclopedia of Life (EOL) contain data that can aid in identifying tools to create an effective BHL for the next generation based on user priorities. BHL thrives on user engagement, as evidenced by effective use of social media tools for collecting, managing and responding to user feedback promptly.^{2 3} For example, crowdsourced tagging for BHL Flickr images has allowed BHL to share images with EOL. The BHL NDSR cohort will work together with mentors to plan the next generation of BHL and characterize the tools necessary to improve data management and reuse. Each resident would be placed at one of BHL’s partner institutions (or geographic regions) that have appropriate mentors with skills and demonstrated capacity; matches will be made according to local BHL expertise and the resident’s interests.

The BHL residency supports many goals and priorities of the NDSR program: ⁴

- strong mentorship component
- service learning
- supporting the future of digital design and management by training professionals.

“IMLS and others will need to work effectively and simultaneously with local institutions, with state and national service providers, and with global initiatives”; “...end users... can also be contributors.”⁵

The connection to national programs and best practices is important for building stronger and more well-rounded professionals with leadership capabilities. The opportunities for residents in the program to partner with professionals in biodiversity fields as well as libraries, museums and archives are a key feature of this residency.

¹ IMLS Focus Report, April 28, 2015 p.10

<https://www.imls.gov/publications/imls-focus-summary-report-national-digital-platform>

² BHL Summer 2015 Quarterly Report <https://goo.gl/Z5jm3t>

³ Costantino, Grace., Bianca Crowley, Rebecca Morin, Erin Thomas. 2011. Heeding the call: user feedback management and the digital library. Microform & Digitization Review

⁴ National Digital Stewardship Residency <http://www.digitalpreservation.gov/ndsr/>

⁵ IMLS Focus Report, April 28, 2015 pp. 8, 7

The BHL⁶ has successfully provided open, global access to biodiversity literature dating from the 15th century through recent publications (via an active copyright permissions program).⁷ However, like many research institutions, the BHL is facing the need to assess and transform capacity to manage changing user expectations for “big data” and interoperability with data stored elsewhere. Examples of user-driven requests and essential enhancements to the current suite of tools and services include implementing full-text search of Optical Character Recognition (OCR) files, enabling the import of crowdsourced data (including manuscript transcriptions, OCR corrections and descriptive metadata about illustrations), linking museum specimen data to publication data, addressing collection development issues to assess the scope of biodiversity literature and identifying significant gaps in the BHL collection⁸ ⁹. The overall goal is to improve the usability, discoverability and functionality of BHL’s content, such as images, based on best practices of other digital libraries, for example the Digital Public Library of America (DPLA), Europeana, and HathiTrust. Most of these issues are not limited to biodiversity systems and solutions and recommendations could improve the utility of other digital libraries, especially regarding connections to systems that manage data that is complementary or supplementary to a specific digital library. Like libraries and other organizations, BHL needs professionals with the digital curation, analytical, user experience and data management skills that residents will gain through this practical program.

The BHL NDSR program model resembles the distributed, virtual model of the WGBH NDSR program, but adjusts the model by pre-identifying hosts and projects instead of staging a competition to attract hosts. Additionally, with at least one of the five residents, we will explore the effectiveness of working at two different sites in the same city. The virtual, distributed model and the opportunity to participate in multiple libraries will provide an exceptional and practical learning experience for each resident. Practical post-graduate experiences are key to successful advancement in digital stewardship careers.¹⁰

IMPACT

Residents will gain skills in digital stewardship by working with an established digital library. Specific stewardship aspects include: collection management and services; understanding of the links among archives, museums, libraries and information systems both within a physical location and across a global collaborative; and how these initiatives can be connected to the greater world of information systems and users. These skills will prepare residents for leadership roles as digital stewards in archives management, data

⁶ <http://biodiversitylibrary.org>

⁷ Gwinn, Nancy E. and Constance A. Rinaldo. 2009. The Biodiversity Heritage Library: Sharing biodiversity with the world. *IFLA Journal* 35(1): 25-34

⁸ Rinaldo, Constance and Mark Phillips. 2013. DLF Forum Working Session Collection Assessment in a Collaborative Environment: BHL and DPLA . November 6.

⁹ <http://biodivlib.wikispaces.com/Collection+Development+Policy>

¹⁰ Bastian, JA, MV Cloonan and R. Harvey. 2011. From Teacher to Learner to User: developing a digital stewardship pedagogy. *Library Trends* 39 (4): 607-622.

curation, information systems management, informatics and community management as well as working in a globally distributed, collaborative virtual environment. The **BHL** will gain defined objectives to enhance innovative tools and services, and newly articulated best practices; **users** will have better connected information systems and improved discovery; **residents** will have new skills to bring to their future employers, including virtual collaboration.

The BHL cohort, working closely with their mentors, will generate outcomes that will be reproducible and transferable to other disciplines and digital domains. This will be accomplished by documenting user needs, assessing services and tools, prioritizing enhancements, modeling domain independent digital library metrics, and establishing informatics collaborations. This cohort of residents will have the opportunity to network among themselves, within the host organizations (libraries embedded in scientific institutions), with professional societies, and also with the global BHL community. The impact of the program goes beyond the effect of the individual projects and residents. The impact will be felt additionally, in how the host's resources and the cohort's engagement make the projects more effective. This can happen because the residents are able to test ideas with BHL partners, convene collaborators from partner institutions, learn from the methodologies of the institutional projects and one another, and interact with global partners such as GBIF, DPLA and BHL members and affiliates.

Outcomes will include a best practices and guidelines document for digital libraries incorporating metadata enhancement, improved OCR displays, collection analysis techniques and better connections to museums, archives and other relevant databases such as DPLA, Europeana, GBIF and EOL.

PROJECT DESIGN

The Biodiversity Heritage Library (BHL), led by the Ernst Mayr Library of the Museum of Comparative Zoology, Harvard University (MCZ), proposes to host a National Digital Stewardship Residency (NDSR) cohort for the 2016-2017 year. The BHL NDSR cohort would support five geographically distributed residents, graduates of LIS or related master's programs, in a collaborative project to improve tools, curation and content stewardship of a key digital life sciences library.

Because BHL is a content hub for DPLA and associated with with GBIF, EOL and Europeana, a broad set of content providers are available to test and disseminate best practices.

Better functionality for the BHL, based on user needs, will feed into other data systems and provide enhancements for BHL biodiversity and library partners such as such as the Encyclopedia of Life (EOL), Global Biodiversity Information Facility (GBIF), Digital Public Library of America (DPLA), and Europeana. All of the participants are associated with cultural heritage museums that incorporate more than biodiversity into their exhibits, research and education activities. For example, MCZ is associated with the Harvard

Museums of Science and Culture (HMSC).¹¹ Three BHL partner institutions have significant transcription programs or are actively working on transcription for material that can not be read by OCR software.¹² This series of resident projects is an opportunity to develop a bridge to connect museum and library/archives collections. Although the focus of BHL is on Life Sciences, there are connections to history, art and culture.¹³

The following five project outlines are proposed:

Content Analysis. This project will be an analysis of the quantity of literature underpinning the field of biodiversity, the amount of that literature in the public domain, the representation of each discipline (delineated by taxon group) within BHL, an exploration of methodologies to scope the collections, and areas where BHL may target development to better serve the research population. Host: *BHL Chicago partners* (The Field Museum of Natural History and the Chicago Botanic Garden) with mentors Library Directors Christine Giannoni and Leora Siegel.

Import of Crowdsourced Data Corrections and Enhancements. Building on the IMLS grant received by the Missouri Botanical Garden, *Purposeful Gaming and BHL*, in which BHL worked with Mary Flannagan and Tiltfactor¹⁴, the mentors for this project will work with the resident to develop methodologies and propose tools for integration of crowdsourced data corrections and enhancements back into the BHL portal. Best practices will be documented for verification, trust, and multi-tier review. Host: *Harvard University: MCZ*, mentors Library Managers Constance Rinaldo and Joseph deVeer.

Enabling image discovery within the Biodiversity Heritage Library Building on the successful NEH *Art of Life*¹⁵ grant, crowdsourced metadata around BHL images hosted on Flickr and Zooniverse would be integrated back into the BHL through user interface modifications to the BHL portal to enable image searching, browsing and display. The resident will work with the BHL technical team to propose best practices for integration of this data into BHL as well as sustainable methodologies for augmenting image tagging for BHL content. Host: *Missouri Botanical Garden (MOBOT)*, mentors Library Director Douglas Holland and Data Projects Coordinator, Trish Rose-Sandler.

Digital Library Best Practices Analysis. The mentor will work with the resident to consult with BHL partners such as DPLA and Europeana to determine how BHL data works in these large scale national and pan-national digital libraries. The resident will propose analyses of other large-scale digital libraries (HathiTrust, the Internet Archive, National Digital Library of New Zealand, Trove, for example) to categorize high value tools and services that can be built into the next version of BHL or developed with existing APIs from

¹¹ <http://hmsc.harvard.edu/fourmuseums>

¹² For example, Smithsonian Institution Transcription Center <https://transcription.si.edu/>

¹³ Art of Life and Purposeful Gaming Project descriptions. <https://biodivlib.wikispaces.com/Current+Projects>

¹⁴ Greenemeier, L. 2015. Smorball: Citizen Science. *Scientific American* December 30
<http://www.scientificamerican.com/citizen-science/smorball/>

¹⁵ <http://biodivlib.wikispaces.com/Art+of+Life>

partners. Host: Los Angeles County Museum of Natural History, mentor Chief Librarian and Curator Richard Hulser.

User Needs and Usability Analysis. The mentor will work with the resident to identify members of the larger taxonomic community (e.g. TDWG, GBIF and biodiversity informatics specialists) to determine additional digital library needs and services that will provide increased value to BHL content. Building on a nearly ten year relationship with this community, BHL staff will introduce the resident to relevant initiatives hosted at the Smithsonian, such as the Consortium for the Barcode of Life (CBoL) and EOL. The close proximity of many biodiversity organizations makes the Smithsonian an ideal choice for this project. Host: BHL Secretariat/Smithsonian Libraries, mentor Martin Kalfatovic, BHL Program Director.

Outcomes:

- a) Content analysis methodology that can be applied to other digital libraries;
- b) Guidance to identify targets for digitization;
- c) Transferable methodology for integration of crowdsourced enhancements using OCR and metadata corrections supplied by past grant projects.¹⁶
- d) Best practices documentation for verification, trust, and multi-tier review of crowdsourced transcriptions and metadata; Best practices document for each project.
- e) Description and prioritization of 1) added-value digital library tools and services to be built; and 2) existing tools in other services that can be leveraged to facilitate the next generation digital library as tested in BHL
- f) User needs assessment and prioritization of enhancements;
- g) Guidance in data implementation and linking schema.

Deliverables:

1. A series of public reports on each of the 5 focus areas for use by the digital library and biodiversity informatics community.
2. Report on transferability to other digital library or biodiversity data services (GBIF, EOL, DPLA, Museums) of each discrete project. Mechanical Turk was tested as a transcription tool in the IMLS *Purposeful Gaming and BHL*¹⁷ grant thus a summary and analysis of the utility of Mechanical Turk in can be provided in this setting.
3. An overall review of the BHL NDSR distributed program with an analysis of the shared resident case.
4. Each resident will work with their mentor to plan a webinar on a topic related to their project with a local expert as presenter.
5. Each resident will present a final report with details of their work, lessons learned and results to the mentors and other residents.
6. Each resident will present an aspect of their work at a professional conference.

¹⁶ Art of Life and Purposeful Gaming Project descriptions. <https://biodivlib.wikispaces.com/Current+Projects>

¹⁷ <http://biodivlib.wikispaces.com/Purposeful+Gaming>

The first six months is the planning period for the mentors and the selection of residents.

Phase 1 Planning: mentor digital project management training, protocol for mentorship, review of existing training and project documentation, review of past NDSR models, planning for the residencies, identification and confirmation of an educator in the field of library science or biodiversity as an advisor, develop a baseline survey of skills and post-residency survey of skills, establish a framework for virtual meetings and trainings, develop the project website; review and finalize the projects to ensure each is well-defined and the host institution is prepared for the resident, development of a contingency plan for drop-outs or project failures. (2-3 months).

Phase 2 Resident selection: review and revise resident requirements and tailor to finalized projects; advertise with library, archives, records management, informatics, museum studies and computer science programs nationally and locally; mentors as a group review and select applicants in a virtual environment (3-4 months)

Phase 3: Residencies: preparation a) Immersive training for the Biodiversity Heritage Library including collection management (selection, deduplication, digitization, curation, user feedback), workflows, internal and external communication, software, metadata, scope of biodiversity b) Digital library overview including open data, partnerships and linkages, current best practices guidelines¹⁸, apps and interoperability. Post training, the residents will have one week to disperse and settle in at their host institution.

Phase 4 Residencies: Once settled, a pre-residency skill inventory will be completed for each resident. Then each resident will spend the first 2 months immersed in all aspects of BHL at the local level, encompassing content identification, digitization, using BHL for discovery and administrative purposes, building collections within the interface, enhancing metadata, testing tools and reviewing documentation. This experience will inform each resident's focus on their area of concentration: collection analysis/gaps, user studies, guidelines for incorporating crowdsourced metadata and OCR corrections and articulating best practices. Residents would spend the rest of their program working with mentors designing, conducting, and supporting projects that target these areas to demonstrate to BHL what can be achieved with current resources and what requires an immediate increase in capacity to best satisfy changing user needs. At the end of the residency, a post-residency skill inventory will be completed. The mentors will collaborate with the residents to review and evaluate the projects and process, document lessons learned, complete final reports and develop recommendations for BHL.

Residents will attend weekly meetings with mentors, bi-weekly virtual meetings with the cohort, BHL monthly staff and twice monthly collections calls. The bi-weekly cohort calls will help the residents coordinate activities, chart progress and discuss results of user

¹⁸ NISO Framework Working Group/IMLS. 2007. A Framework of Guidance for Building Good Digital Collections: A NISO Recommended Practice. 3rd edition December 2007

research and pilot projects. The initial training session will be held at the Smithsonian Institution led by the BHL Secretariat. Residents also will be encouraged to attend the annual meeting of the BHL and present their work. Mentors will coordinate presentations at professional meetings (Biodiversity Information Standards known as TDWG, CNI, IMLS Focus) for presentation experience for residents.

Project updates will be presented via BHL & institutional blogs, the project website, BHL and institutional social media and presentations. Each host institution will have travel stipends available for residents and mentors. Face to face meetings were considered important to resident cohorts by previous programs,¹⁹ thus support for face to face meetings has been built in to the budget along with virtual opportunities. BHL has extensive experience communicating among multiple time zones in a telephone conferencing or videoconferencing environment.²⁰

Phase 5: Evaluation by mentors of residencies (host and residents) and projects, testing transferability of project results with GBIF and DPLA, analyse pre and post skill tests, review and revise “best practices” document for each project; compile and post “lessons learned” document, write final report, ensure that all documentation pertaining to the grant is accessible via the project website and that availability of these reports is broadly communicated.

Conversations about lessons learned with leaders from the Harvard (Andrea Goethals), Metro NY (Margo Padilla) and WGBH (Karen Cariani) NDSR cohorts lead us to propose 12 month terms for each resident and provide stipends that would allow the residents to purchase health insurance to address challenges identified for past cohorts.

PROJECT RESOURCES: PERSONNEL, TIMELINE, BUDGET

There will be at least one mentor, who is also a BHL onsite representative, at each host institution. Existing BHL committees, consisting of staff from multiple partners, will consult in all aspects of the project, for example Collections (content analysis), Technical (tools, best practices, data linking), and Staff (all partners) committees. Prof. James Hanken will serve as Principal Investigator, in his role as MCZ Director and Harvard Faculty.

Personnel

Constance Rinaldo (Harvard MCZ Ernst Mayr Library: Project Director, Mentor) Rinaldo has 25 years of experience in collaborative projects, digitization and grants management. She will coordinate resident stipends through Harvard University, organize meetings, ensure timely project execution and lead planning and reporting efforts. The Ernst Mayr Library is a founding member of BHL.

¹⁹ Margo Padilla and Karen Cariani, personal communication

²⁰ Collaboration and communication tools used by the Biodiversity Heritage ILbrary

Trish-Rose Sandler, Constance Rinaldo, Keri Thompson, William Ulate, Martin Kalfatovic

June 2012 JCDL '12: Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries

Joe deVeer (Project Manager & Mentor), is Project Manager and Museum Liaison at the Ernst Mayr Library of the Museum of Comparative Zoology, where he has worked since 2006. deVeer manages library and museum digitization projects, curates the resulting digital assets archived in the Harvard University Library Digital Repository Service, and oversees contribution of Ernst Mayr Library collections to BHL.

Martin R. Kalfatovic (Mentor) is Associate Director, Digital Programs and Initiatives at the Smithsonian Libraries and Program Director, Biodiversity Heritage Library. He has twenty years of progressively complex work in digital libraries and information technology with a focus on the areas of information management, data access, and digitization in a major museum complex.

Christine Giannoni (Mentor), is Museum Librarian at The Field Museum, where she has worked since 2002. Giannoni oversees all of the library's operational areas, including the general research collections, the Mary W. Runnells Rare Book Room, Photo Archives & Museum Archives, and has taken a leading role in the library's participation in the BHL since 2009.

Leora Siegel (Mentor), is the Director of the Lenhardt Library of the Chicago Botanic Garden where she has worked in various capacities since 2000. She has served as project director for several grants awarded for digitization, conservation, and cataloging volumes from the rare book collection. In 2014, the Lenhardt Library joined BHL as an affiliate member.

Richard Hulser (Mentor), is Chief Librarian and Curator at the Natural History Museum of Los Angeles County, where he has worked since 2010. Hulser oversees operations of the research library, institutional archives and related records management initiatives. He is the representative for the NHMLAC participation as a BHL affiliate since 2013.

Douglas Holland (Mentor) is the Director of the Missouri Botanical Garden Library. He has more than 15 years of experience in libraries and archives including managing digital projects. MOBOT is a founding member of BHL.

Trish Rose-Sandler (Technology Mentor), is Data Projects Coordinator in the Center for Biodiversity Informatics at the Missouri Botanical Garden, where she has worked since 2010. Besides managing several grant-funded digital library projects, she has served as a data analyst and a member of the technical development team for BHL since 2010.

5. ADDITIONAL STAFF Carolyn Sheffield (BHL Program Manager) in conjunction with Bianca Crowley (Collections Manager) will coordinate the initial BHL Training.

Timeline

This project will cover 2 years from June 1, 2016 to May 31, 2018

6/1/2016-8/31/2016: Phase 1

9/1/2016-12/31/2-16: Phase 2

01/01/2017-12/31/2017: Phase 3 & 4

01/01/2018-05/31/2018: Phase 5

Budget

Harvard is requesting **\$370,756** to support the BHL NDSR proposal. The majority of the budget is allocated to resident support. Over the 2 years of the project, resident support will be **\$250,000** for 5 residents plus additional funds for travel related to professional development. Total non-student costs are **\$120,756** for partial mentor support, travel support for mentors, honoraria for trainers and educator consultant, and webinar software; this amount also includes applicable indirect costs. Cost share will be provided by participating institutions as in-kind support based on mentor time. Each host institution will provide a desk, space and necessary in-house training and equipment.

DIVERSITY PLAN

Harvard University is committed to selecting faculty and staff without discrimination against individuals on the basis of race, color, sex, gender identity, sexual orientation, religion, creed, national origin, ancestry, age, veteran status, disability unrelated to job requirements, genetic information, military service, or other protected status.²¹ Applicants will be recruited according to this statement although each host institution is responsible for ensuring that they adhere to the standards set by their institution.

COMMUNICATION PLAN.

BHL has a detailed [Outreach and Communication](#)²² plan. As per this plan, the BHL has an active presence (presentations, papers, demonstrations, discussions) in multiple digital library, biodiversity informatics and museum environments. Residents will have an opportunity to attend professional meetings and participate in interest groups associated with professional associations²³ such as the Digital Library Federation (DLF), American Library Association (ALA), Coalition for Networked Information (CNI), Museums and the Web, LITA Forum and DPLAFest as well as science venues such as the Society of the Preservation of Natural History Collections (SPNHC), American Association for the Advancement of Science (AAAS) and TDWG. Residents also will contribute to the creation of content highlighting the work of the projects to be distributed on the BHL blog and related institutional blogs and websites. An online bibliography of publications and presentations about the projects will be maintained. These activities will foster dialogue about digital libraries and biodiversity topics within BHL and beyond, including with scientific and professional communities.

SUSTAINABILITY PLAN

“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

²¹ REAFFIRMATION OF THE UNIVERSITY’S POLICY CONCERNING AFFIRMATIVE ACTION AND EQUAL EMPLOYMENT OPPORTUNITY BY THE PRESIDENT AND FELLOWS OF HARVARD COLLEGE <http://goo.gl/U03jRy>

²² https://docs.google.com/document/d/1daoP4CKwACFPkIrMyE_BX-r1VAN2bBj8i2WP5QhEzFY/edit

²³ DLF Assessment Interest Group <https://wiki.diglib.org/Assessment> for example

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.”²⁴

The residents that participate in this learning opportunity will receive hands-on practice, be cross trained in digital library tools and innovations and become part of the expert practitioner communities cited as a “fundamental sustainability issue” by Bethany Nowviskie, Council on Library and Information Resources .²⁵ BHL is self-sustaining using a membership and donation model. BHL is well-known in the digital library and international scientific, taxonomic and natural history communities such as museums all over the world, EOL and GBIF. These organizations have a vested interest in the sustainability of BHL. The BHL Secretariat is headquartered at the Smithsonian Institution and has significant financial and staffing support there as well as from the other institutions participating in this grant proposal and beyond that, more than a dozen BHL members and affiliates.²⁶ CY 2014 “in kind” contributions from Members and Affiliates totaled \$1.6m and 14.5 FTE.

The BHL had identified the following as key areas contributing to sustainability: a committed partnership, global participation, and user research resulting in interface changes. Among the outcomes of this NDSR residency will be the experience of observing challenges to sustainability such as: specialized content and finite primary audience; technical infrastructure; intellectual property rights in the context of collection definition; varied digitization standards; and the unique opportunities for global collaboration.

All project documents, reports and resources will be publically available and remain so on the project website. The completion of this program will enable residents to move into careers with competencies and experiences to become digital library and curation leaders, thus the benefits will have ripple effects well into the future. The institutions participating in this proposal will gladly offer their expertise to future grantees. In conjunction, BHL will pursue funding to enhance the necessary innovative services or tools identified during this 2 year project. The host institutions will continue to recruit new professionals, refine training materials, and share expertise beyond the end of the program. Improved access to biodiversity and environmental data will elicit support from new, as well as current sectors.

²⁴ Deanna Marcum. "Searching for Sustainability: Strategies from Eight Digitized Special Collections: Biodiversity Heritage Library". Ithaka S+R Case Study. 2014. <http://www.sr.ithaka.org/research-publications/searching-sustainability>.

²⁵ IMLS Focus Report, April 28, 2015 p.10

²⁶ <http://biodivlib.wikispaces.com/BHL+Consortium#BHLparticipatingInstitutions> Appendix

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded research, data, software, and other digital products. The assets you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products is not always straightforward. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and best practices that could become quickly outdated. Instead, we ask that you answer a series of questions that address specific aspects of creating and managing digital assets. Your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

If you propose to create any type of digital product as part of your project, complete this form. We define digital products very broadly. If you are developing anything through the use of information technology (e.g., digital collections, web resources, metadata, software, or data), you should complete this form.

Please indicate which of the following digital products you will create or collect during your project
(Check all that apply):

	Every proposal creating a digital product should complete ...	Part I
	If your project will create or collect ...	Then you should complete ...
<input type="checkbox"/>	Digital content	Part II
<input type="checkbox"/>	Software (systems, tools, apps, etc.)	Part III
<input type="checkbox"/>	Dataset	Part IV

PART I.

A. Intellectual Property Rights and Permissions

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1 What will be the intellectual property status of the content, software, or datasets you intend to create? Who will hold the copyright? Will you assign a Creative Commons license (<http://us.creativecommons.org>) to the content? If so, which license will it be? If it is software, what open source license will you use (e.g., BSD, GNU, MIT)? Explain and justify your licensing selections.

A.2 What ownership rights will your organization assert over the new digital content, software, or datasets and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users about relevant terms or conditions.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

Part II: Projects Creating or Collecting Digital Content

A. Creating New Digital Content

A.1 Describe the digital content you will create and/or collect, the quantities of each type, and format you will use.

A.2 List the equipment, software, and supplies that you will use to create the content or the name of the service provider who will perform the work.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to create, along with the relevant information on the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

B. Digital Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance (e.g., storage systems, shared repositories, technical documentation, migration planning, commitment of organizational funding for these purposes). Please note: You may charge the Federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the Federal award. (See 2 CFR 200.461).

C. Metadata

C.1 Describe how you will produce metadata (e.g., technical, descriptive, administrative, or preservation). Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, or PREMIS) and metadata content (e.g., thesauri).

C.2 Explain your strategy for preserving and maintaining metadata created and/or collected during and after the award period of performance.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of digital content created during your project (e.g., an API (Application Programming Interface), contributions to the Digital Public Library of America (DPLA) or other digital platform, or other support to allow batch queries and retrieval of metadata).

D. Access and Use

D.1 Describe how you will make the digital content available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

D.2 Provide the name and URL(s) (Uniform Resource Locator) for any examples of previous digital collections or content your organization has created.

Part III. Projects Creating Software (systems, tools, apps, etc.)

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) this software will serve.

A.2 List other existing software that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software (systems, tools, apps, etc.) and explain why you chose them.

B.2 Describe how the intended software will extend or interoperate with other existing software.

B.3 Describe any underlying additional software or system dependencies necessary to run the new software you will create.

B.4 Describe the processes you will use for development documentation and for maintaining and updating technical documentation for users of the software.

B.5 Provide the name and URL(s) for examples of any previous software tools or systems your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under an open-source license to maximize access and promote reuse. What ownership rights will your organization assert over the software created, and what conditions will you impose on the access and use of this product? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain any prohibitive terms or conditions of use or access, explain why these terms or conditions are justifiable, and explain how you will notify potential users of the software or system.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

C.3 Identify where you will be publicly depositing source code for the software developed:

Name of publicly accessible source code repository:

URL:

Part IV. Projects Creating a Dataset

1. Summarize the intended purpose of this data, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

2. Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

3. Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

4. If you will collect additional documentation such as consent agreements along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

5. What will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

6. What documentation (e.g., data documentation, codebooks, etc.) will you capture or create along with the dataset(s)? Where will the documentation be stored, and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

7. What is the plan for archiving, managing, and disseminating data after the completion of the award-funded project?

8. Identify where you will be publicly depositing dataset(s):

Name of repository:
URL:

9. When and how frequently will you review this data management plan? How will the implementation be monitored?

Original Preliminary Proposal

Planning BHL Version 2.0:

The Biodiversity Heritage Library (BHL), led by the Ernst Mayr Library of the Museum of Comparative Zoology, Harvard University (MCZ), proposes to host a National Digital Stewardship Residency (NDSR) cohort for the 2016-2017 year. The BHL NDSR cohort would support five geographically distributed residents, graduates of LIS or related master's programs, in a collaborative project to improve curation and content stewardship of a key digital life science library.

Need/ Impact The BHL has successfully provided open, global access to biodiversity literature dating from the 15th century through recent publications (via an active copyright permissions program). However, like many research institutions, the BHL is facing the need to assess and transform capacity to manage changing user expectations for “big data.” Examples include implementing full-text search of OCR files, enabling the import of crowdsourced data (including manuscript transcriptions, OCR corrections, enhanced illustration metadata), linking museum specimen data to publication data, addressing collections development issues to assess the scope of biodiversity literature, and improving the discoverability, usability, and functionality of BHL's content based on other best of practice digital libraries such as the DPLA, Europeana, and the HathiTrust.

As noted in the *IMLS Focus Report*¹: “IMLS and others will need to work effectively and simultaneously with local institutions, with state and national service providers, and with global initiatives”; “...end users... can also be contributors.” This proposal addresses the IMLS priority of a national digital platform, and the concepts of user engagement and collaboration beyond libraries, by working with biodiversity informatics organizations such as Global Biodiversity Information Facility (GBIF) and the Encyclopedia of Life (EOL) to identify tools to create an effective BHL for the next generation based on user priorities. BHL thrives on user engagement, as evidenced by effective use of social media tools and prompt responses to comments. For example, crowdsourced tagging for BHL Flickr images has allowed BHL to share images with EOL.

The BHL NDSR cohort will work together planning “BHL Version 2.0.” Each resident would be placed at one of BHL's partner institutions (or geographic regions) that have appropriate mentors with skills and demonstrated capacity; matches will be made according to local BHL expertise and the resident's interests. Conversations with leaders from the Harvard, Metro NY and WGBH NDSR cohorts lead us to propose 12 month terms for each resident and employment benefits, as the shorter term and lack of benefits in those programs were challenges for past cohorts. Our program model resembles the virtual model of WGBH, but improves on this by pre-identifying hosts and projects (instead of staging a competition to attract hosts); with two residents, we are exploring the effectiveness of working at two different sites in the same city.

Workplan: The first six months is the planning period for the mentors covering: 1) digital project management; 2) review of projects to ensure each is well-defined and the host organization is prepared for the resident; 3) contingency planning for drop-outs or project failures; and 4) advertising for and selecting residents. Once selected, each resident would spend the first month immersed in all aspects of BHL at the local level, encompassing content identification, digitization, using BHL for discovery and administrative purposes, building collections within the interface, enhancing metadata, testing tools and reviewing documentation. The first four months of the project also require the cohort to collaborate to identify user groups, conduct user and usability surveys, gather and conduct focus groups, and thereby narrow and focus the specific functionalities BHL needs to implement. Residents would spend the rest of their program designing, conducting, and supporting projects that target the highest priority functions to demonstrate to BHL what can be achieved with current resources and what requires an immediate increase in capacity to best satisfy changing user needs. Once the residency is completed, the mentors will collaborate to review and evaluate the projects and process, document lessons learned, and complete final reports.

The following five project outlines are under consideration:

Content Analysis. This project will be an analysis of the quantity of literature underpinning the field of biodiversity, the amount of that literature in the public domain, the representation of each discipline (delineated by taxon group) within BHL, and areas where BHL may target development to better serve the research population. The project will also explore methodologies to scope existing BHL collections based on full-text analysis. Proposed host: **BHL Chicago partners (The Field Museum of Natural History and the Chicago Botanic Garden)** mentors Library Directors Christine Gianonni and Leora Siegel.

Import of Crowdsourced Data Corrections and Enhancements. Building on the IMLS grant received by the Missouri Botanical Garden for purposeful gaming, in which we worked with Mary Flanagan from Tiltfactor, this project will develop methodologies and tool proposal for integration of crowdsourced data corrections and enhancements. Best practices will be documented verification, trust, and multi-tier review. Proposed host: **Harvard University (Museum of Comparative**

¹ IMLS Focus Report, April 28, 2015; <https://www.ims.gov/publications/ims-focus-summary-report-national-digital-platform>

Zoology, and Harvard Botany Libraries) mentors Library Managers Constance Rinaldo, Judith Warnement and Joseph deVeer.

Import of Crowdsourced Image Metadata. Building on the successful NEH “Art of Life” grant, crowdsourced metadata around BHL images hosted on Flickr would be integrated into the BHL user interface to enable image searching. The resident will work with the BHL technical team and members of the visual resources community to propose best practices for integration of this data into BHL as well as sustainable methodologies for augmenting image tagging for BHL content. Proposed host: **Missouri Botanical Garden**, mentor Peter H. Raven Library Director Douglas Holland.

Digital Library Best Practices Analysis. The resident will work with BHL partners at DPLA and Europeana to determine how BHL data works in these large scale national and pan-national digital libraries. The resident will also analyze other large-scale digital libraries (HathiTrust, the Internet Archive, National Digital Library of New Zealand, Trove, for example) to determine high value tools and services that can be built into BHL Version 2.0 or developed with existing APIs from partners such as DPLA or Europeana. Proposed host: **Los Angeles County Museum of Natural History**, mentor Head Librarian Richard Hulser.

User Needs and Usability Analysis. The resident will work with members of the larger taxonomic community (e.g. TDWG, GBIF and biodiversity informatics specialists) to determine additional digital library needs and services that will provide increased value to BHL content. Building on a nearly ten year relationship with this community, BHL staff will introduce the resident to relevant initiatives hosted at the Smithsonian, such as the Consortium for the Barcode of Life (CBOL) and EOL. The close proximity of many biodiversity organizations makes the Smithsonian an ideal choice for this project. Proposed host: **BHL Secretariat/Smithsonian Libraries**, mentor Martin Kalfatovic, BHL Program Director.

Residents will attend monthly BHL staff calls and conduct weekly cohort calls to coordinate activities, chart progress and discuss results of user research and later the pilot project. The residents will travel at least three times to a central location for a two-day intensive collaboration on the then-current segment of the project to ensure each resident’s work is aligning with the cohort and the overall goals of BHL. An initial meeting will be held at the Smithsonian. Residents will be invited to attend the annual meeting of the BHL.

It is expected that each resident will attend an appropriate major taxonomic or informatics meeting during the course of the residency. Examples of these meetings include TDWG, GBIF, and SPNHC. Residents also will contribute project updates to BHL and institutional blog posts.

The BHL cohort will generate outcomes that will be reproducible and transferable to other disciplines and digital domains. This will be accomplished by documenting user needs assessment and tools, modeling domain independent digital library metrics, and establishing informatics collaborations. This cohort of residents will have the opportunity to network among themselves, within the host organizations (libraries embedded in scientific institutions) and also with the global BHL community.

Outcomes:

a) Content analysis methodology that can be applied to other digital libraries; **b)** Guidance to identify targets for digitization; **c)** Transferable methodology for integration of crowdsourced enhancements; **d)** Best practices documentation for verification, trust, and multi-tier review; **e)** Identification of 1) high value tools and services to be built; and 2) existing APIs that can be leveraged to facilitate BHL Version 2.0; **f)** User needs assessment and prioritization of enhancements; **g)** Identification of added-value digital library services; **h)** Guidance in data implementation and linking schema.

Project Management & Staffing: The project lead and manager of the residency will be Constance Rinaldo (MCZ) who has 25 years of experience in collaborative projects and grants management. Rinaldo will coordinate subawards through Harvard University, organize meetings, ensure timely project execution, and lead planning and reporting efforts. There will be at least one mentor, who is also a BHL onsite representative, at each host institution. Existing BHL committees, consisting of staff from multiple partners, will consult in all aspects of the project, for example **Collections** (content analysis) and **Technical** (tools, best practices, data linking) committees. Prof. James Hanken will serve as Project Director, in his role as MCZ Director.

Budget. We estimate the total budget requested from IMLS to be \$620,719 for the 2 year period. IMLS funds will directly cover support for 5 NDSR residents (stipends and fringe benefits), plus overhead, travel for mentors and residents to meetings, laptops for each resident, and immersion training in the production and maintenance of the BHL. Cost share ~\$54,000 will be provided by participating institutions as in-kind support, based on mentors’ time.

Timeline. This is planned as a two year project, running from July 1, 2016 – June 30, 2018.