Final Report: User Needs Analysis

January 2018

Pamela A. McClanahan, Resident
Smithsonian Institution Libraries
Biodiversity Heritage Library
National Digital Stewardship Residency

Table of Contents

Acknowledgements

Executive Summary

Introduction

Background
- Review of BHL Structure and Systems
  - BHL's Issue Tracking System
- Review of Relevant Literature and Case Studies
- Review of BHL 2010 Survey
- Review of Responses to Informal Questionnaire
- Review of Input from Informal Interviews
  - Smithsonian Institution Staff
  - Biodiversity Organizations

Methodology
- User Groups
  - Consortium Users
  - System Users
  - Individual Users
- Surveys
  - Developing Surveys
    - Internal Review Board
    - Survey Logistics
  - Performing Survey Analysis

Analysis - User Surveys
- Individual User Survey
  - Definition
  - Purpose
  - Goals
  - Distribution
  - Description
  - Overall Sample Size
  - Analysis
    - I1 - Informed Consent
    - I2 - How often do you use BHL?
    - I3 - What best describes your role when using BHL?

- I4 - (Follow up to Q3) If scientist/taxonomist chosen, please briefly describe your area of specialization.
- I5 - How do you utilize the content you access in BHL? [Check all that apply.]
- I6 - What if anything do you find challenging or frustrating about using the BHL website?
- I7 - When using the BHL site, how valuable do you find the following:
- I8 - Is there something more BHL could be doing to help you identify and/or extract data to answer your research questions?
- I9 - (Follow up to Q8) If yes, please describe.
- I10 - The following are some suggestions we've received from BHL users in the past. Which if any of these new services would you like to see implemented by BHL? [Check all that apply]
- I11 - Have you come across gaps in BHL content coverage?
- I12 - (Follow-up to Q11) If yes, please describe (for example, but not limited to - gaps in taxons, time periods, ecologies, geographies, institutions, authors, and volume gaps in serial/journal titles, etc.)
- I13 - Do you use archival/manuscript content in BHL? (e.g. correspondence, field notes, diaries)
- I14 - (Follow-up to Q13) If yes, what information or data do you search for in biodiversity archival items?
- I15 - Do you use illustrations/images in BHL?
- I16 - (Follow-up to Q15) If yes, How do you use those illustrations/images?
- I17 - (Follow-up to Q15) If yes, Are there other features related to image use that you would like to see added?
- I18 - (Follow-up to Q15) If yes, How do you find those images?
- I19 - (Follow-up to Q15) If yes, Are there other ways you would like to be able to search for images?
- I20 - If you have come across features on other digital library sites that you would like to see on BHL, please describe the feature (links appreciated).
- I21 - What other data sources and systems do you use? [Check all that apply]
- I22 - Do you have any additional feedback you would like to share about your experience using the Biodiversity Heritage Library or comments about new features?

Consortium User Survey

- Definition
- Purpose
- Goals
- Distribution
- Description
- Overall Sample Size
- Analysis
  - C1 - Informed Consent

- C2 - Below is a list of the proposed enhancements/tools that garnered the most consensus at the recent BHL Members Meeting and through other discussions with Partners. Please rate your level of interest in these services being implemented in BHL.
- C3 - (Follow up to Q2) Please provide any additional details on what you would like to see come out of implementing and providing the services listed above, or if there is another tool or feature that you feel should be a priority for BHL going forward.
- C4 - What do you see as the primary challenge to implementing these tools and services?
- C5 - (Follow up to Q4) How do you see BHL overcoming this challenge?
- C6 - Please rate your satisfaction with BHL's content coverage in the following areas:
- C7 - Do you have any additional comments related to particular strengths or gaps in BHL's content coverage?
- C8 - If you have come across features on other digital library sites that you would like to see on BHL, please describe the feature (links appreciated).
- C9 - What has been your experience adding content to BHL via Macaw and Admin Dash? Please share any likes/dislikes, difficulties you may have encountered, or areas of improvement you would recommend.
- C10 - Do you have any additional comments or suggestions to share about technical development priorities for BHL?

System User Survey

- Definition
- Purpose
- Goals
- Distribution
- Description
- Overall Sample Size
- Analysis
  - S1 - Informed Consent
  - S2 - Do you use the BHL API?
  - S3 - The following are some suggestions we've received from BHL users in the past for enhancements and tools. Which if any of these new services would you like to see implemented by BHL? [Check all that apply]
  - S4 - How would you like to see BHL data services be more interoperable with other data services and systems?
  - S5 - What do you see as the primary challenge to increasing interoperability with BHL?
  - S6 - How do you see us working together to overcome this challenge?
  - S7 - Please rate your satisfaction with BHL collection coverage in the following areas:
  - S8 - (Follow-up to Q7) Please provide any details on the above areas related to their usefulness and any particular strengths or gaps in BHL's coverage.

S9 - Does the current offering of data through the BHL API meet your needs? Why or Why not?
S10 - Does the available documentation on the BHL API meet your needs? Why or why not?
S11 - If you have come across features on other data service sites that you would like to see in BHL, please describe the feature (links appreciated).
S12 - Do you have any additional feedback to share about your interaction with BHL or any further recommendations on how BHL could make the content you or your users want to work with more easily accessible?

Recommendations
Further User Research
Usability Studies
Focus Groups
Repeated Surveys
User Priorities
Search
Named Entities
Linking

Conclusion

Appendix A - Supplemental Material
Trainings
Presentations and Conferences
Blog Posts

Appendix B - Informal Questionnaire
Q1 - Are you a member of EC, Secretariat, or Tech Team?
Q2 - Where do you see the priorities for the next version of BHL? What features or functionality would be most appreciated by BHL Partners?
Q3 - What does BHL do really well currently and should continue to do as it moves to the next version?
Q4 - What do you see as the biggest challenges to implementing BHL v.2? How can we address those challenges?
Q5 - Who would you describe as BHL's most important external partners and how do you see BHL pursuing further interoperability with these partners in the next version?
Q6 - What do you feel are the most important areas for improvement needed for BHL from the user perspective?
Q7 - Do you feel there is any area where BHL needs a greater understanding of its users' needs? Is there anything you want to know more about from the BHL users?
Q8 - Is there anything else you feel would be helpful for me to know about BHL, its partners, or its users as I gather feedback?

Appendix C - Informal Interviews
- Informal Interview Table
- Informal Interview Summaries
  - Smithsonian Staff
  - Biodiversity Organizations

Appendix D - Individual Survey
- I1 - Informed Consent
- I6 - What if anything do you find challenging or frustrating about using the BHL website?
- I9 - (Follow up to Q8) If yes, please describe.
- I22 - Do you have any additional feedback you would like to share about your experience using the Biodiversity Heritage Library or comments about new features?

Appendix E - Consortium Survey
- C1 - Informed Consent
- C3 - (Follow up to Q2) Please provide any additional details on what you would like to see come out of implementing and providing the services listed above, or if there is another tool or feature that you feel should be a priority for BHL going forward.
- C4 - What do you see as the primary challenge to implementing these tools and services?
- C5 - (Follow up to Q4) How do you see BHL overcoming this challenge?
- C7 - Do you have any additional comments related to particular strengths or gaps in BHL's content coverage?
- C8 - If you have come across features on other digital library sites that you would like to see on BHL, please describe the feature (links appreciated).
- C9 - What has been your experience adding content to BHL via Macaw and Admin Dash? Please share any likes/dislikes, difficulties you may have encountered, or areas of improvement you would recommend.
- C10 - Do you have any additional comments or suggestions to share about technical development priorities for BHL?

Appendix F - System User Survey
- S1 - Informed Consent
- S4 - How would you like to see BHL data services be more interoperable with other data services and systems?
- S5 - What do you see as the primary challenge to increasing interoperability with BHL?
- S6 - How do you see us working together to overcome this challenge?
- S8 - (Follow-up to Q7) Please provide any details on the above areas related to their usefulness and any particular strengths or gaps in BHL's coverage.
- S9 - Does the current offering of data through the BHL API meet your needs? Why or Why not?
- S10 - Does the available documentation on the BHL API meet your needs? Why or why not?

S11 - If you have come across features on other data service sites that you would like to see in BHL, please describe the feature (links appreciated).

S12 - Do you have any additional feedback to share about your interaction with BHL or any further recommendations on how BHL could make the content you or your users want to work with more easily accessible?

Appendix G - Resources

BHL

Biodiversity

Digital Libraries

User Studies

User Study Examples

Acknowledgements

Thank you to everyone who made this project possible, especially to the Library of Congress and Institute of Museum and Library Services (IMLS) for establishing the National Digital Stewardship Residency (NDSR) program, the Biodiversity Heritage Library (BHL) for hosting the residents, IMLS for generously funding the BHL-NDSR residencies, and all of the residents, mentors, and host site staff for their collaboration, input, and guidance.

A big thank you is also extended to all of the BHL Members, Affiliates, and Partners, biodiversity organizations, and BHL users of all kinds who participated in the surveys and informal interviews. This feedback is incredibly valuable for the work of BHL now and going forward and the biodiversity and digital library communities overall.

Executive Summary

This report shares the project overview, findings, and final recommendations of National Digital Stewardship Resident, Pamela McClanahan's User Needs Analysis project for the Biodiversity Heritage Library (BHL). BHL is a consortium of natural history and botanical libraries working to make natural history literature openly accessible.

This project was based at Smithsonian Libraries working with the BHL Secretariat and ran from January 9, 2017 to January 8, 2018 as a part of the BHL National Digital Stewardship Residency (NDSR) program, *Foundations to Actions: Extending Innovations to Digital Libraries in Partnership with NDSR Learners*. Carolyn Sheffield, BHL Program Manager, served as the mentor for this resident.

The User Needs Analysis Project focused on collecting user feedback from each of the BHL user groups to provide data for user-informed decision making going forward as BHL looks to move to its next version. It involved defining the BHL user groups, interviewing users, and deciding on user surveys targeted at each user group as the main method of user feedback.

Included in this report are an introduction to the project, an overview of the background research, a description of the methodology involved in completing the user needs project, an analysis of each of the surveys, and final recommendations based on the top user priorities revealed in the analysis and suggestions for further user research. There are appendices with additional information including all of the trainings, presentations, and blog posts completed throughout the year, supplemental information for the survey analysis sections, and a list of resources consulted and for BHL's use going forward.

Through the background research learning about BHL's structure, systems, and previous user studies along with conducting an informal questionnaire with BHL leadership and interviewing Smithsonian staff and representatives from the biodiversity community, the resident determined the user groups and designed a survey targeted at each group. The three user groups for this project are Individual Users, Consortium Users, and System Users. The resident designed the survey questions based on the preliminary research and best practices and all questions were approved by the Smithsonian Internal Review Board to ensure there were no risks to participants. The surveys were conducted through SurveyMonkey and were each live for approximately three weeks. No personally identifiable information was collected. The resident analyzed the surveys through the SurveyMonkey tools and using Excel.

Through information gathered from the research and analysis, the resident recommends that additional user research be conducted including usability studies, focus groups, and repeated surveys. The top user priorities identified in analysis are search, named entities, and linking.

Introduction

The BHL-NDSR cohort consisted of five residents each working on an individual BHL related project at different host sites that are a part of the BHL consortium. The User Needs Analysis Resident, Pamela McClanahan, worked with the BHL-NDSR mentors and residents to identify key stakeholders and user groups and defined methodologies for gathering their input. Results will be used to define recommendations and requirements for expanding BHL digital library functionality and ensure requirements for new services and features for the BHL platform are informed by members of the larger biodiversity community.

The resident completed preliminary background research to guide the design of the user surveys. Through a review of BHL's structure and systems including BHL's internal issue tracking system, BHL's previous user study work, and conducting an informal questionnaire of BHL leadership and informal interviews with Smithsonian staff and the biodiversity community, the resident gathered the background knowledge needed to define the user groups and create a user survey for each group. Surveys were distributed to each of the groups and the resident completed an analysis of the responses. Through the background research and survey analysis, the resident determined that there is some additional user needs research that would be beneficial for BHL going forward and was able to identify the top user needs as BHL plans its technical development.

While each resident had their own focus, the projects were closely related, and the residents collaborated amongst each other. The residents communicated daily via the business chat tool, Slack, and used the Google suite of products to share documents and have regular video meetings. The residents and mentors had a biweekly conference call. Together, the residents created a blog to share updates on their projects, trends in the digital library field, professional development, and conference presentations.[1] The residents were able to share their work, while promoting BHL and NDSR, at several conferences throughout the year including DPLAfest, NDSR Symposium, and ALA. More details on the blog, professional development, and conferences can be found in Appendix A.

[1] NDSR at BHL: Foundations to Actions. https://ndsrbhl.wordpress.com.

Background

Having just recently celebrated their 10th anniversary, BHL is at point where they know they want to make some technical development and other changes to their site. As BHL looks to move to their next version, the NDSR projects are providing background and research for BHL to make those updates. The User Needs Analysis project is collecting data on user priorities for going forward.

To learn more about the target user groups, determine the best method to gather user feedback, and subsequently design the surveys, the resident conducted preliminary background research. This included an examination of previous work BHL has done on investigating user needs, a review of the current system, a review of the feedback received via BHL's internal issue tracking system, and conducting an informal questionnaire and informal interviews.

Review of BHL Structure and Systems

A thorough understanding of the BHL site and accompanying public and private wiki was necessary to understand the library and its' different user groups. Learning the organizational structure was very valuable as well to know how the Members of the consortium interact with BHL and how technical development currently occurs.

Carolyn Sheffield, Program Manager and NDSR Mentor, was very helpful in advising on BHL, its structure and Members. Other BHL Secretariat staff were also helpful in understanding Gemini, the users, and collections including Grace Costantino, BHL Outreach and Communication Manager and Bianca Crowley, BHL Digital Collection Manager.

Additionally, the resident had the opportunity to participate in some of the BHL committee meetings throughout the year such as the Collections Committee. The resident attended the weekly Technical Advisory Group conference call. There were other opportunities such as participating in a meeting about the workflow for addressing user feedback to the Global Names Architecture (the organization that manages BHL's scientific name finding algorithm) that the resident could be a part of as well.

This review of BHL including these meetings provided the context for technical development progress for BHL as well as some of the needs and preferences of the consortium Members and staff.

BHL's Issue Tracking System

The resident also reviewed BHL's internal issue tracking system, Countersoft's Gemini product for issue-tracking, to learn the types of issues submitted by users. Staff have been tagging some of the issues in Gemini as "wishlist," "2.0," or "NDSR" for features and enhancements that

would be reviewed during the NDSR project and as BHL looks to move to its next version. The resident incorporated the tools, features, and themes mentioned in Gemini within the interviews and surveys to gather more information around these issues.

All of these meetings and review of BHL and Gemini provided background on the current status of BHL and some known user needs and priorities that could be delved into further.

Review of Relevant Literature and Case Studies

The resident also conducted a review of user study publications, case studies, and other examples to have an understanding of best practices in the field of digital libraries and user surveys. Reviewing other user studies and surveys gave background and context to what might work best for a BHL survey. A listing of resources can be found in Appendix G.

Review of BHL 2010 Survey

BHL has conducted user surveys in the past with 2010 being their most recent user survey.[2] Reviewing the kinds of questions asked and the summary of the results provided insight into previous issues for the general users, needs identified by users, and how users interact with BHL. There were some questions that would be helpful to ask again. Other questions proved to be a starting point that would be helpful to ask about different or new features.

Review of Responses to Informal Questionnaire

The resident developed an informal questionnaire based on the review of BHL's structure and systems and distributed it to the BHL Secretariat, Executive Council, and Technical Advisory Group to learn more about the perspective of the BHL leadership on technical development and the vision going forward for BHL.

The responses to the informal questionnaire revealed some of BHL's strengths and weaknesses along with areas to learn more from BHL users, which aided in designing the user survey questions. These responses were shared with the other residents to help inform their research and projects as well. See Appendix B for a list of questions and responses to this questionnaire.

Review of Input from Informal Interviews

The resident conducted informal interviews with Smithsonian Staff, as well as staff from biodiversity organizations to learn more about Smithsonian operations, digital libraries, and the needs of the biodiversity community. A full list of interviews is provided in Appendix C.

---

[2] Welter-Schultes, Francisco, Biodiversity Heritage Library, "BHL Survey 2010, Evaluation of the Results," http://www.planetposter.de/bhl/wien2010.htm.

Smithsonian Institution Staff

The resident met with several staff onsite at Smithsonian to aid in the User Needs Project from technical staff to digital library experts. Staff provided background on BHL and their interactions with BHL, as well as some tips for online surveys and presenting data.

Biodiversity Organizations

Additionally, the resident met with staff from biodiversity organizations who have partnered with BHL and have similar user groups. Some of these meetings took place at the Smithsonian and others were via Skype for international groups. These meetings provided the context and details to understand the needs of the biodiversity community and how BHL can better partner with these groups going forward.

Methodology

The resident identified the target audiences and defined three user groups, decided on a survey as the format for collecting user feedback from each of these groups, designed the surveys including the Smithsonian IRB approval process, and then distributed and analyzed the surveys.

User Groups

Through the background research, the resident identified three BHL user groups and defined them as follows for the purposes of this project:

Consortium Users

A contributor to BHL including Members, Affiliates, Partners, and staff. BHL has over 60+ worldwide partners including universities, government agencies, and other natural history and botanical libraries.

System Users

Organizations or individuals who interact with BHL for the purpose of enriching another system via APIs or manually, such as Encyclopedia of Life and the Integrated Taxonomic Information System.

Individual Users

Anyone visiting the BHL website to search for information to answer their research needs such as, scientists, collection managers, librarians, etc.

While all users of BHL, each of these groups interact with the system in different ways. For example, Consortium Users contribute content to BHL and have their own patrons at their institutions that use BHL for their research. The System Users, on the other hand, may work with the BHL API to provide content to their own networks or represent power users who work with the BHL data in a more technical way. The Individual Users encompass anyone who is coming directly to BHL for their research needs. While individual Users may or may not have in-depth knowledge of BHL or their internal or external partners, they are seen as the primary users of the features of the BHL site. Since these groups interact with BHL in different ways, the resident decided to conduct three separate user surveys each targeted at one of these user groups.

Surveys

The resident designed a survey to be distributed to each of the three user groups focused on the needs of each of the user groups and their preferences for technical development in BHL going forward.

Developing Surveys

Using the research and information gathered as detailed above the resident developed questions for surveys for each of the three users groups. The Individual User Survey was completed first, as it was the largest distribution and had the most questions. It by far received the most responses as well. For the Individual Survey, in addition to the above research, the Smithsonian resident worked with the other residents to develop questions related to their projects. This included a collections coverage question, an archival content usage question, several image use and discoverability questions, and a question about examples of features from other digital libraries. The Consortium User Survey and the System User Survey were about half the length of the Individual User Survey and each had a much smaller, targeted distribution. There were some similar questions to the Individual User Survey, but they were mostly specific to those user groups and following up on broad discussions the resident and BHL had with these groups previously.

Internal Review Board

Another aspect of the process for developing user surveys at the Smithsonian Institution was submitting the project to the Internal Review Board (IRB) for approval. This involved the resident and mentor taking the Smithsonian's Human Subjects in Research course and receiving a certificate of completion. The certificate of completion along with the survey questions, distribution information, and promotion language, along with their OSP HS 01 form needed to be submitted to the IRB. The project received approval and was considered exempt. Any changes need to be submitted to the IRB for approval though. The resident submitted an addendum for the two additional surveys and they were approved as a part of the original project exemption as well. The anticipated end date for this project with the IRB is July 9, 2018 and any additional aspects to the project or extensions will need to be submitted to the IRB for approval.

Survey Logistics

Each of the surveys was developed for distribution through SurveyMonkey, which is the platform that has previously been approved for use at the Smithsonian and the resident had prior experience working with this platform.[3] SurveyMonkey is user-friendly and comes with many options for survey creation and provides internal analysis tools, as well as multiple ways to export the data for analysis with other programs, such as Microsoft Excel.

---

[3] SurveyMonkey, https://www.surveymonkey.com.

Each of the surveys were live for approximately three weeks and included a combination of multiple choice, likert scale, and free text questions. Multiple choice answer options were randomized to decrease bias in responses. When appropriate, multiple choice questions included an “Other” answer choice option for survey respondents to fill in a free text response if they desired. The first question in each survey is a statement of informed consent as required by Smithsonian IRB procedures. If the respondent agreed, they were taken to the next question; if they disagreed, they were taken to a disqualify page. That was the only required question on the survey, as a participant had to agree to it to be able to take the survey. Otherwise, survey participants could answer as many or as few questions as they wished.

Performing Survey Analysis

While any feedback is valuable when it comes to users needs, usually the larger the sample size, the more representative the results are for the entire user population. The overall sample size for each of these surveys was calculated using the SurveyMonkey Sample Size Calculator. This looks at the total population size, the confidence level, and the margin of error. The numbers are provided for each survey in the analysis section.[4]

The resident utilized the SurveyMonkey internal analysis tools to some extent, but mostly used Excel for analysis and graph creation. The results were exported for use in Excel in both a summary format, which provides an excel sheet for each question, as well as with all individual responses to sort and manipulate as needed.

For multiple choice questions, the resident reviewed the results in SurveyMonkey and created a table and graph of the results in Excel. These tables and graphs are used to visually represent the distribution of responses per each option. Data is presented as tallied responses per option, percent of responses for that question, and for likert scale questions, as weighted averages. The responses to the “Other” answer choices were also reviewed and summarized. The analysis section of this report includes a short narrative explaining the results and includes a table and graph for most questions.

For free text questions, the resident reviewed the exported individual responses to the question in Excel and assigned categories based on themes that emerged from reading through the responses. A given response could belong to multiple categories. The responses for each category were tallied and a table and graph of those tallied responses, along with a narrative explaining the results, are included in the analysis of these questions. Each survey has an appendix that provides a description of categories determined for each free text question.

---

[4] “Surveymonkey Sample Size Calculator,” accessed May 4, 2017, https://www.surveymonkey.com/mp/sample-size-calculator/.

Analysis - User Surveys

The resident completed an analysis of each question in the three user surveys: Individual, Consortium, and System. Each survey section contains the definition of the user group, the purpose and goals of the survey, the survey distribution and description, the sample size calculation, and an analysis of each question from the survey.

Individual User Survey

The Individual User Survey section contains the definition of individual users, the purpose and goals of the survey, the survey distribution and description, the sample size calculation, and an analysis of each question from the survey.

Definition

For the purposes of this study, individual users refer to anyone visiting the BHL website to search for information such as scientists, collection managers, librarians, and the general public.

Purpose

The purpose of the Individual User Survey is to gather feedback to inform recommendations and requirements for the next version of BHL including which improvements and features to focus on in the design process for a new platform.

Goals

The goals of this survey are to confirm users and context of use, learn user priorities for important features, and identify any issues users encounter on the site.

Distribution

This survey was promoted on the BHL website through a message and link on the homepage banner. The survey was also distributed via email to the BHL listserv and other biodiversity listservs including iDigBio, Herbaria, SPNHC, GBIF, and Taxacom. Additionally, the survey was promoted through posts on BHL social media including Facebook and Twitter.

Description

This survey was live for approximately three weeks from July 13, 2017 to August 7, 2017. The survey consists of 22 questions including the informed consent, 10 multiple choice, and 11 free text.

Overall Sample Size

During the approximately three weeks the survey was live, there were 606 total responses. Based on BHL's average monthly website usage of 100,000[5], this gives a 95% confidence level and a 4% margin of error overall for the survey.

Analysis

This Analysis section presents each question from the Individual User Survey, the number of participants who answered and the number skipped, notation if it is a multiple choice or free text question, a brief narrative analysis, a table of the results, and when appropriate a graph of the results.

I1 - Informed Consent

[606 Respondents, all agreed to the Informed Consent Statement]

See Appendix D for Informed Consent Statement

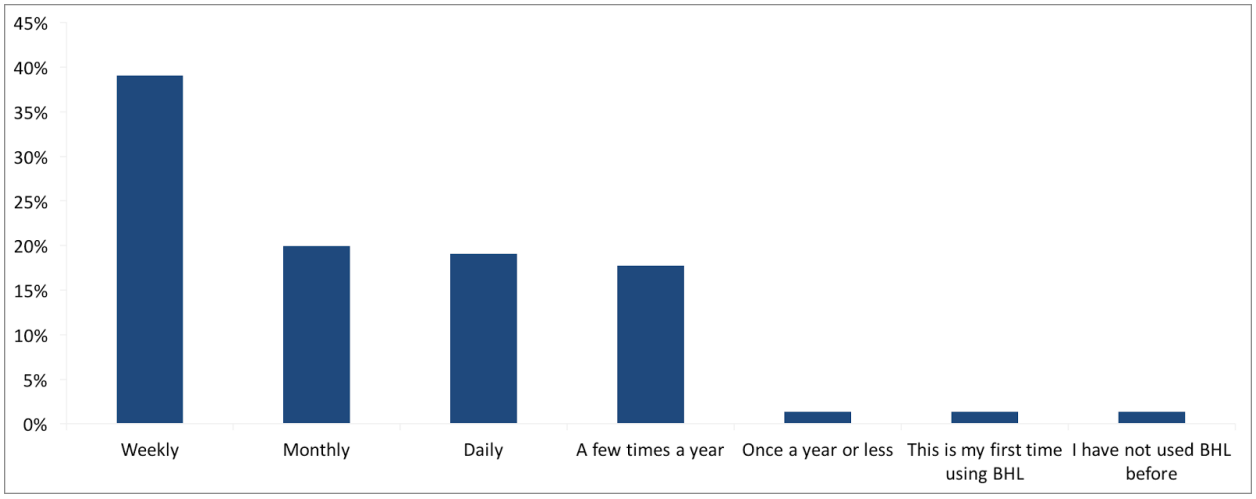
I2 - How often do you use BHL?

[586 answered, 20 skipped] [Multiple Choice]

Nearly 80% of survey respondents use BHL at least once a month and nearly 60% use BHL at least once a week. The top choice was "Weekly" at 39%.

Answer Choices	Responses	
Weekly	39.08%	229
Monthly	19.97%	117
Daily	19.11%	112
A few times a year	17.75%	104
Once a year or less	1.37%	8
This is my first time using BHL	1.37%	8
I have not used BHL before	1.37%	8

[5] Biodiversity Heritage Library, "BHL Reports and Statistics," https://bhl.wikispaces.com/BHL+Reports+and+Statistics.



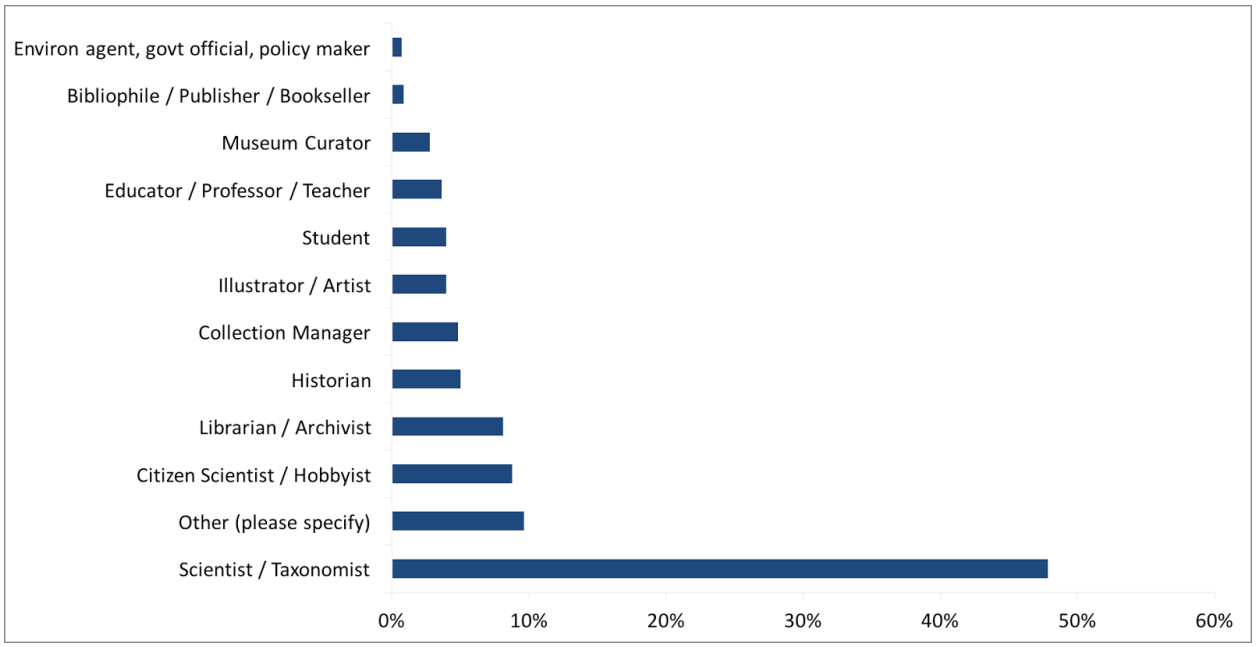
I3 - What best describes your role when using BHL?

[581 answered, 25 skipped] [Multiple Choice]

Almost half of survey participants described their role as “Scientist / Taxonomist.” The next two highest were “Citizen Scientist / Hobbyist” at about 9% and “Librarian / Archivist” at about 8%. There were also several respondents who chose “Other.” Those responses included many of the same categories as above, but slightly different titles although still in the realm of scientist, librarian, curator, and hobbyist. Some responses were retired professionals, interns, or volunteers, or others felt they fit in many categories.

Answer Choices	Responses	
Scientist / Taxonomist	47.85%	278
Other (please specify)	9.64%	56
Citizen Scientist / Hobbyist	8.78%	51
Librarian / Archivist	8.09%	47
Historian	4.99%	29
Collection Manager	4.82%	28
Illustrator / Artist	3.96%	23
Student	3.96%	23
Educator / Professor / Teacher	3.61%	21
Museum Curator	2.75%	16

Bibliophile / Publisher / Bookseller	0.86%	5
Environmental agent, government official, policy maker	0.69%	4



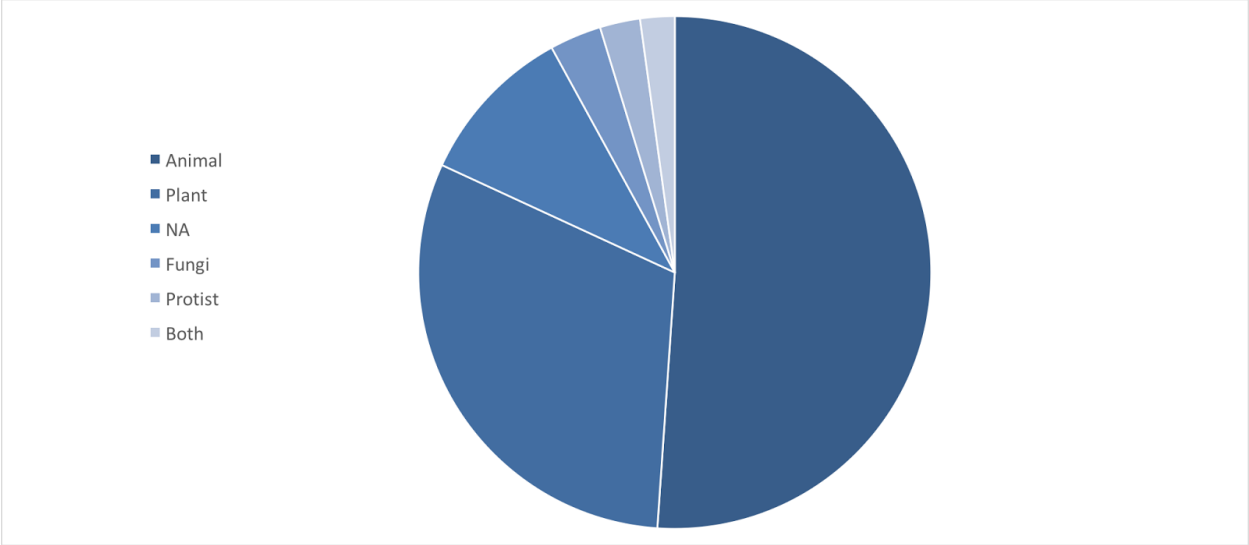
I4 - (Follow up to Q3) If scientist/taxonomist chosen, please briefly describe your area of specialization.

[276 answer; 330 skipped (278 answered scientist/taxonomist to previous question of the 278 this question was visible to, 276 answered; 2 skipped)] [Free Text]

Categories were determined from the free text responses and tallied. Although this was free text, most survey respondents provided the species they specialize in, and this is a tally to estimate the distribution of how many scientist/taxonomists specialized in “Animal” vs. “Plant.” About half were “Animal” and about one third were “Plant.”

Category	Tally	Percent
Animal	141	51.09%
Plant	85	30.80%
NA	28	10.14%
Fungi	9	3.26%
Protist	7	2.54%

Both	6	2.17%



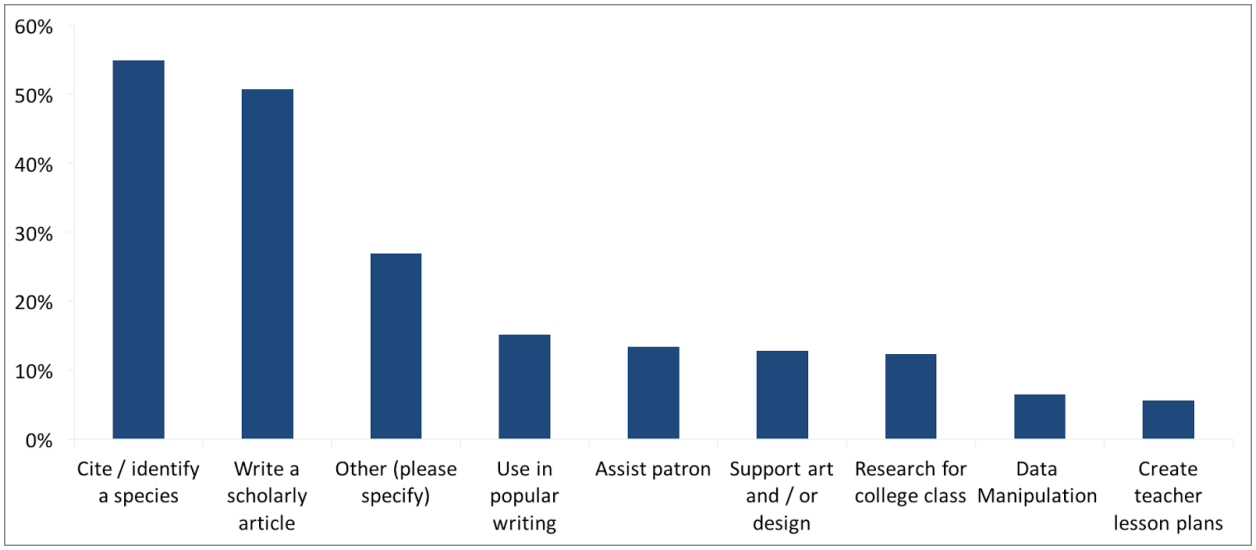
I5 - How do you utilize the content you access in BHL? [Check all that apply.]

[568 answered; 38 skipped] [Multiple Choice]

BHL is most often used for scholarly or professional scientific work with the top categories by far being “Cite / identify a species” at 55% and “Write a scholarly article” at 51%. There were also 153 “Other” responses. The majority of the “Other” responses aligned closely with those top two categories and mentioned professional scientific and taxonomic work such as compiling bibliographies on species, utilizing the original description of a species, writing other scholarly works, and overall research. There were also some responses that were related to crowd-sourced efforts such as contributing to wikipedia articles, as well as several that cited use for personal interest and entertainment.

Answer Choices	Responses	
Cite / identify a species	54.93%	312
Write a scholarly article	50.70%	288
Other (please specify)	26.94%	153
Use in popular writing	15.14%	86
Assist a patron of your library / museum	13.38%	76

Support art and / or design projects	12.85%	73
Complete research for undergraduate or graduate class	12.32%	70
Data Manipulation (e.g. extract data using API)	6.51%	37
Compile information for teacher lesson plans	5.63%	32



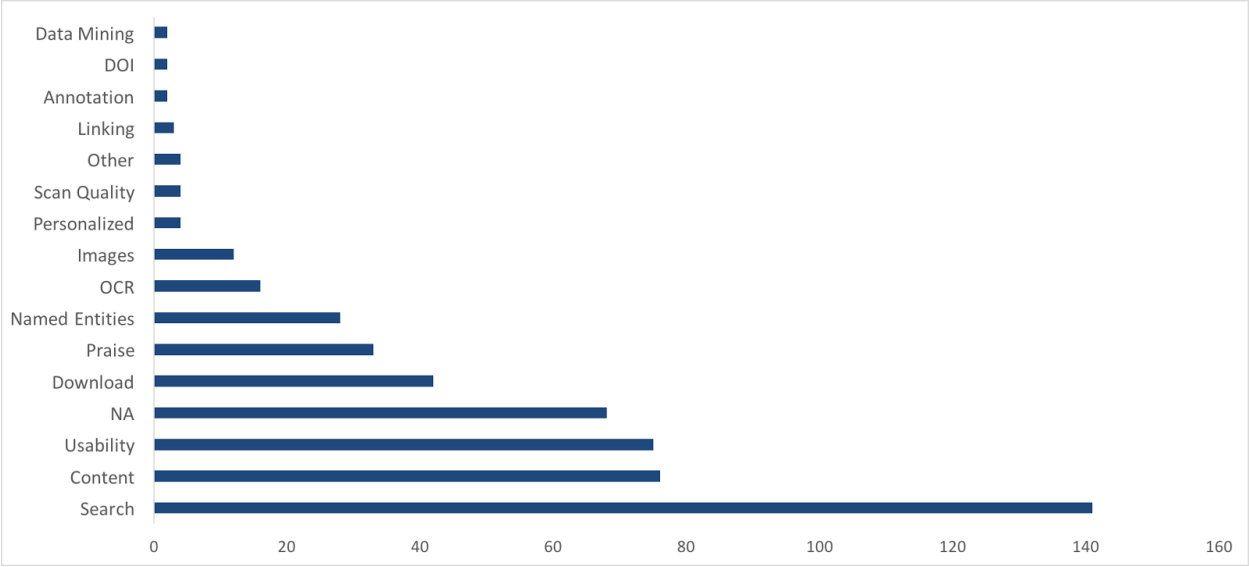
I6 - What if anything do you find challenging or frustrating about using the BHL website?

[429 answered; 177 skipped] [Free Text]

Categories were determined from the free text responses and tallied. Responses in the “Search” category were the most frequent. This included faceted searching, full-text search, fuzzy matches, related to illustrations, browsing, better metadata, search in other languages, saved searches, geographic names, enhanced scientific name search, and overall improved search and ability to retrieve relevant information. It is important to note that in Questions I9 and I10, we delve further into some of these different aspects of search and separate out input specific to illustrations, scientific names, and geographic names. More details on “Search” comments and a brief description of each category can be found in Appendix D.

Category	Tally
Search	141
Content	76
Usability	75
NA	68

Download	42
Praise	33
Named Entities	28
OCR	16
Images	12
Personalized	4
Scan Quality	4
Other	4
Linking	3
Annotation	2
DOI	2
Data Mining	2



I7 - When using the BHL site, how valuable do you find the following:

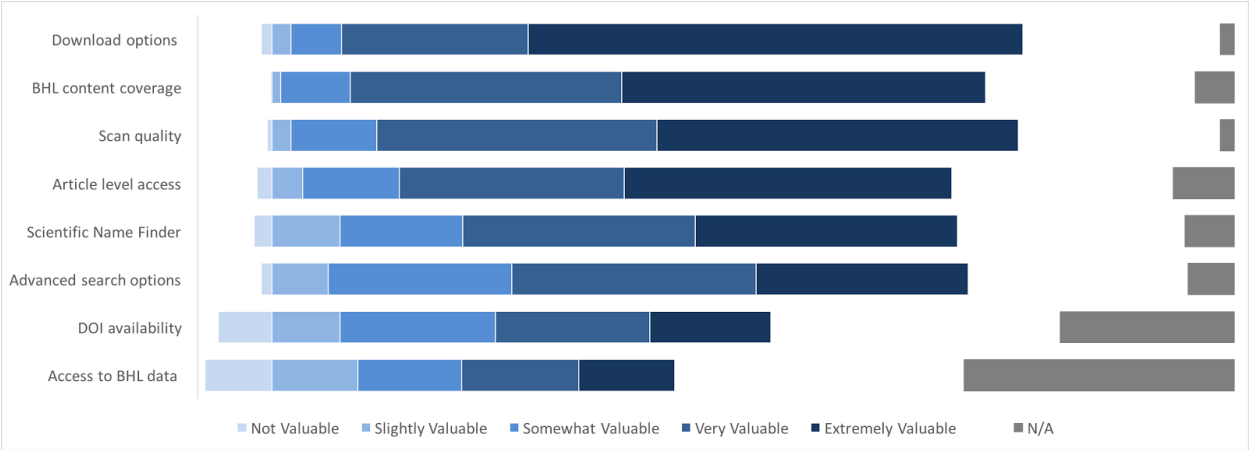
[530 answered; 76 skipped] [Multiple Choice]

Participants were presented with a list of current features and asked to rate each on a scale of not valuable to very valuable. When comparing weighted average, survey respondents found “Download Options” to be the most valuable followed by “BHL Content Coverage.”

The graph below depicts number of responses and “Download Options” and “BHL Content Coverage” are still the highest rated. To the right in gray are not applicable responses. The two choices at the bottom Access to BHL data and DOI availability have a larger number of not applicable responses on the right in gray and a higher number of not valuable responses at the far left in light blue.

There was an “Other” option on this question not included in the graph, and there were some respondents that said they were not familiar with DOIs or APIs, which may explain why there were more not applicable choices for those options. Many respondents were confirming the given choices often noting “BHL content coverage” and others were praising BHL. Many noted the stable page level url’s were a very valuable feature.

Answer Choices	Total	Weighted Average
Access to BHL data for data mining (e.g. APIs)	500	3.19
DOI availability	491	3.4
Advanced search options	509	3.84
Scientific Name Finder	509	3.93
Article level access	511	4.18
Scan quality	517	4.3
BHL content coverage	510	4.38
Download options (e.g. PDF generation)	524	4.49
Other (please specify any other aspects of the website you find very valuable or extremely valuable)	79	-



Graph created in Excel by Richard Naples, SIL and edited by resident.

I8 - Is there something more BHL could be doing to help you identify and/or extract data to answer your research questions?

[506 answered; 100 skipped] [Multiple Choice]

Over half of respondents responded “No,” there is nothing more BHL needs to do with the data for their research questions.

Answer Choices	Responses	
Yes	42.89%	217
No	57.11%	289

I9 - (Follow up to Q8) If yes, please describe.

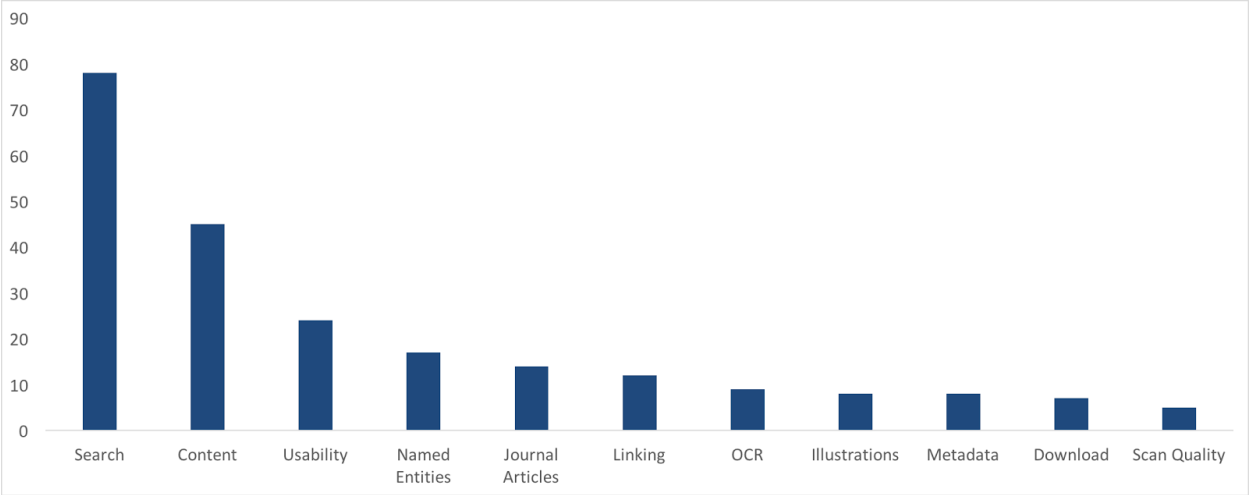
[216 answered; 390 skipped (of 217 “Yes” responses, 216 answered; 1 skipped)] [Free Text]

Categories were determined from the free text responses and tallied. The topic most mentioned by survey participants was “Search.” Responses specifically mentioned faceted searching, full-text search, fuzzy matches, related to illustrations, browsing, search by specimen number, and search by scientific name synonyms. More details on “Search” comments and a brief description of each category can be found in Appendix D.

Category	Tally
Search	78
Content	45
Usability	24
Named Entities	17
Journal Articles	14
Linking	12
OCR	9
Illustrations	8
Metadata	8

Download	7
Scan Quality	5
Annotations	4
Data	4
Documentation	4
Original Description	4
Citation	3
Languages	3
Other	3
API	2
General Info	2
Personalized features	2
Research Tools	2
DOI	1
Feedback	1
Images	1
Related Titles	1
Transcription	1

graph of Top Categories:



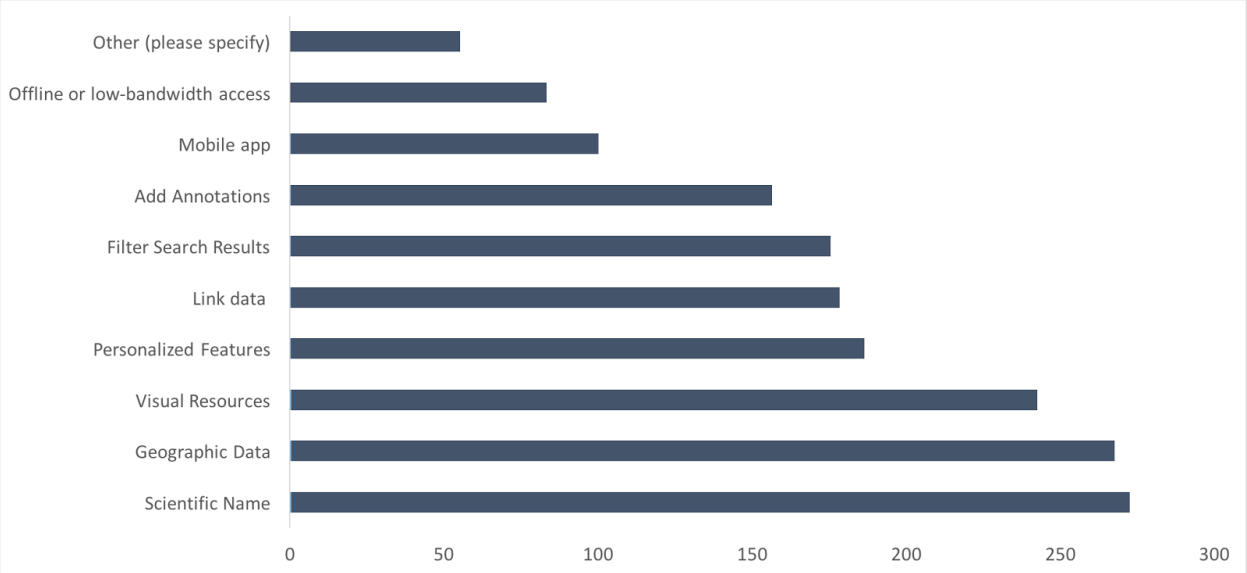
I10 - The following are some suggestions we've received from BHL users in the past. Which if any of these new services would you like to see implemented by BHL? [Check all that apply]

[487 answered; 119 skipped] [Multiple Choice]

The three choices that were checked the most were "View scientific name synonyms in addition to name in literature," "Geographic data searchable like scientific names," and "Easier way to download visual resources" all with half of respondents or more answering this question choosing them. There was an "Other" option on this question. The most mentioned response in "Other" was improving overall search.

Answer Choices	Responses	
View scientific name synonyms in addition to name in literature	55.85%	272
Geographic data searchable like scientific names	54.83%	267
Easier way to locate and download visual resources (e.g. illustrations, graphs)	49.69%	242
Personalized features, such as saved searches, saved items, and customized RSS feeds	38.19%	186
Link data with outside sources (e.g. GBIF)	36.55%	178
Ability to filter search results (e.g. by in-copyright; by archival materials; by visual resources)	35.93%	175
Add annotations and turn on/off annotations from other users (e.g. comments on a particular work, author, or species)	32.03%	156

Ability to work from mobile devices	20.53%	100
Offline or low-bandwidth access to BHL content	17.04%	83
Other (please specify)	11.29%	55



I11 - Have you come across gaps in BHL content coverage?

[492 answered; 114 skipped] [Multiple choice]

Respondents were exactly even on “Yes” vs. “No.”

Answer Choices	Responses	
Yes	50.00%	246
No	50.00%	246

I12 - (Follow-up to Q11) If yes, please describe (for example, but not limited to - gaps in taxons, time periods, ecologies, geographies, institutions, authors, and volume gaps in serial/journal titles, etc.)

[235 answered; 371 skipped (of 246 answered “Yes,” 226 answered; 20 skipped)] [Free Text]

Question I12 was developed in collaboration with fellow resident, Alicia Esquivel, in relation to her project. Results have been shared with Alicia for further analysis. Overall, many responses referred to either in-copyright content or gaps in journals. Respondents also included locations

such as Africa, Russia, Asia, Argentina, Latin American countries, Germany, Australia, New Zealand, and South American countries. Additionally, comments included specific works, specific species, more recent literature, and early medical literature. See Alicia Esquivel's final report for more detail on this topic.

I13 - Do you use archival/manuscript content in BHL? (e.g. correspondence, field notes, diaries)

[500 answered; 106 skipped] [Multiple Choice]

The majority of survey respondents (78.20%) answered "No" to this question and 21.80% answered "Yes."

Answer Choices	Responses	
Yes	21.80%	109
No	78.20%	391

Of those survey participants who answered "Yes", 38% had self-identified as "Scientist / Taxonomist," and the next highest category had self-identified as "Historian" at 13% in question I3. When looking at the top two services these respondents would like to see implemented "Geographic data" and "Visual resources" at 58% and 59% respectively were the highest. Of those that said "Yes" to this question, 80% of them said "Yes" they use images as well. Also, 59% of them use both "GBIF" and "EOL" as other sources.

I14 - (Follow-up to Q13) If yes, what information or data do you search for in biodiversity archival items?

[110 answered; 496 skipped (of 109 answered "Yes," 85 answered; 24 skipped)] [Free Text]

Question I14 was developed in collaboration with fellow resident, Katie Mika, in relation to her project. Results have been shared with Katie for further analysis. Overall, responses included much of the same types of data BHL users look for in other items such as general taxonomic and species information and the original description. They also mentioned occurrence data. Respondents said they looked for field notes, correspondence, diaries, and illustrations / sketches. Some noted that they didn't know about this type of content, but were interested to learn more. Others indicated they were simply not using archival material at this time. See Katie Mika's final report for more detail on this topic.

I15 - Do you use illustrations/images in BHL?

[511 answered; 95 skipped] [Multiple Choice]

The majority of survey respondents answered “Yes” to this question - 71.62% “Yes” and 28.38% “No.”

Answer Choices	Responses	
Yes	71.62%	366
No	28.38%	145

Of those survey respondents who answered “Yes” to this question, 50% of them identified as “Scientist / Taxonomist” and the next highest category was “Citizen Scientist” at 9%.

I16 - (Follow-up to Q15) If yes, How do you use those illustrations/images?

[324 answered; 282 skipped (of 366 answered “Yes,” 320 answered; 46 skipped)] [Free Text]

Question I16 was developed in collaboration with fellow resident, Ari Rehbein, in relation to her project. Results have been shared with Ari for further analysis. Overall, responses included many noting use for identifying species, species description, and for reference, citation, and comparison, as well as linking to other databases. Additionally many survey respondents mentioned using images in lecture slides, publications, presentations, blogs, and social media. Other uses mentioned included crowdsourced projects like wikimedia, teaching / instruction, library outreach / reference, resource for art / drawing projects, and for personal viewing and use. See Ari Rehbein’s final report for more detail on this topic.

I17 - (Follow-up to Q15) If yes, Are there other features related to image use that you would like to see added?

[173 answered; 433 skipped (of 366 answered “Yes,” 171 answered; 195 skipped)] [Free Text]

Question I17 was developed in collaboration with fellow resident, Ari Rehbein, in relation to her project. Results have been shared with Ari for further analysis. Many responses said no, nothing more. Overall, comments included easier way to get to plate from search, easier way to download just the image, easier way to get source citation, and better way to zoom and manipulate images. Some other responses were more metadata for images, higher resolution image options, a way to link the plate to the species description and to other related images, and a way to know if an item has images in it. See Ari Rehbein’s final report for more detail on this topic.

I18 - (Follow-up to Q15) If yes, How do you find those images?

[270 answered; 336 skipped (of 366 answered “Yes,” 270 answered; 96 skipped)] [Free Text]

Question I18 was developed in collaboration with fellow resident, Ari Rehbein, in relation to her project. Results have been shared with Ari for further analysis. Overall, survey respondents

found images through the BHL search feature, browsing, through BHL Flickr feed and other social media channels, through other sites such as Tropicos and Google, with help from a librarian, or by chance. Most comments noted they most commonly found images through what was contained within the book or article they were already viewing. Some survey participants took this question to mean the quality of images and most said they were a good quality. See Ari Rehbein's final report for more detail on this topic.

I19 - (Follow-up to Q15) If yes, Are there other ways you would like to be able to search for images?

[175 answered; 431 skipped (of 366 answered "Yes," 175 answered; 191 skipped)] [Free Text]

Question I19 was developed in collaboration with fellow resident, Ari Rehbein, in relation to her project. Results have been shared with Ari for further analysis. Overall, survey respondents commented they would like to be able to search by artist, species, scientific name, synonym, common name, location, personal name, time period, keyword, attributes such as colored, photo, map, and portrait, and categories. Respondents also asked for icons in the search results, filtering out pages without images, a feature like Google's related image function, and offering a database of just images. See Ari Rehbein's final report for more detail on this topic.

I20 - If you have come across features on other digital library sites that you would like to see on BHL, please describe the feature (links appreciated).

[126 answered; 480 skipped] [Free Text]

Question I20 was developed in collaboration with fellow resident, Marissa Kings, in relation to her project. Results have been shared with Marissa for further analysis. Overall, survey respondents mentioned liking the book viewer, download options, search, annotation feature, and content from other sites and would like to see enhancements in those areas in BHL. Some specific sites mentioned in the comments were Internet Archive, HathiTrust, Gallica, Darwin Online, Trove, World Spider Catalog, Wellcome Images, and Digital Library of the Royal Botanic Garden. See Marissa King's final report for more detail on this topic.

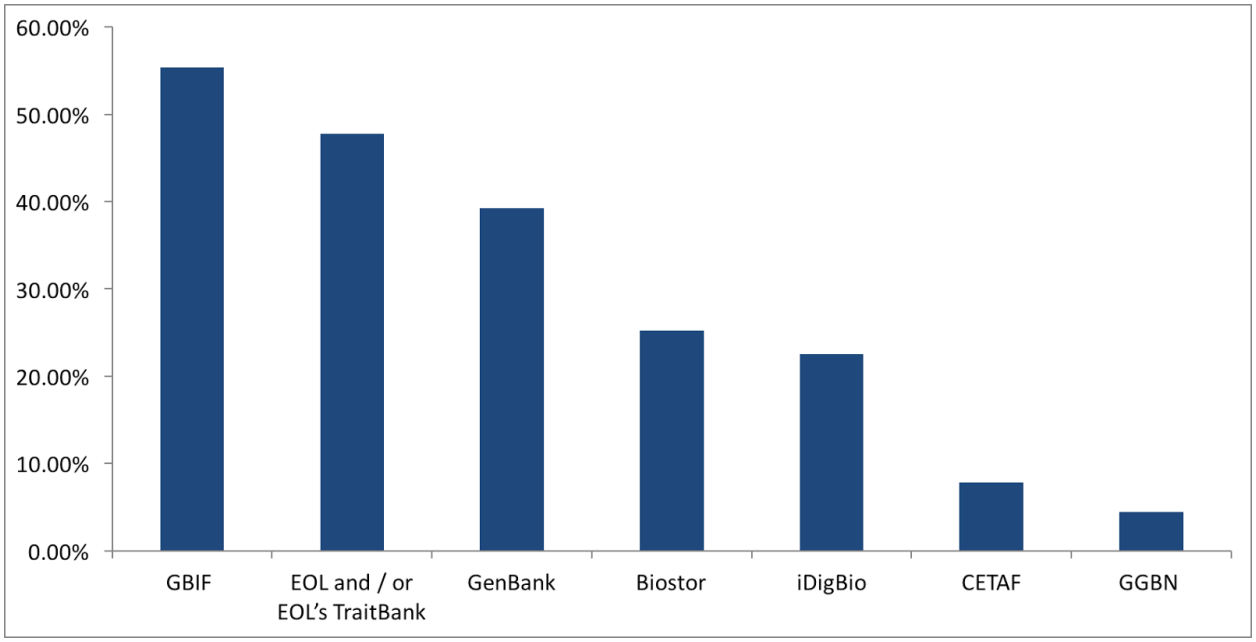
I21 - What other data sources and systems do you use? [Check all that apply]

[408 answered; 198 skipped] [Multiple Choice]

The top chosen data sources and systems by survey respondents were "GBIF" at 55.39% and "EOL" at 47.79%. There were also a high number of "Other" responses to this question - a total of 210 resulting in an additional 185 data systems and services used by survey participants. An analysis of the top "Other" responses are included in the second table and graph below.

Answer Choices	Responses

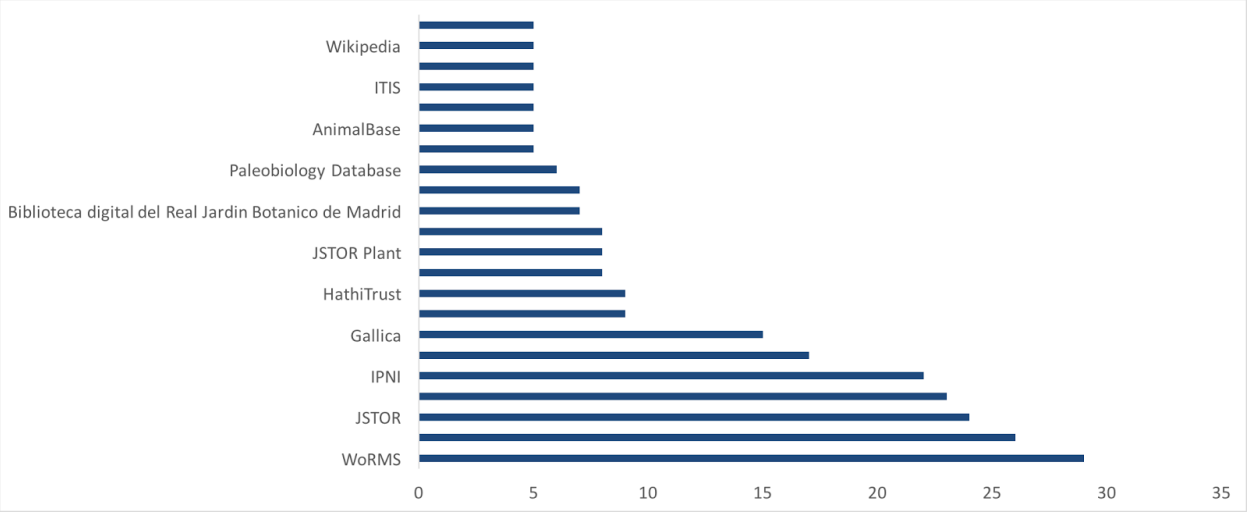
GBIF	55.39%	226
EOL and / or EOL's TraitBank	47.79%	195
GenBank	39.22%	160
Biostor	25.25%	103
iDigBio	22.55%	92
CETAF	7.84%	32
GGBN	4.41%	18
Other (please specify)	51.47%	210



Q21 - "Other" Analysis:

Category	Tally
WoRMS	29
Internet Archive	26
JSTOR	24
Tropicos	23

IPNI	22
Google Books	17
Gallica	15
Botanicus	9
HathiTrust	9
Google Scholar	8
JSTOR Plant	8
The Plantlist	8
Biblioteca digital del Real Jardin Botanico de Madrid	7
BOLD System	7
Paleobiology Database	6
Algaebase	5
AnimalBase	5
Atlas of Living Australia	5
ITIS	5
Kew Catalog	5
Wikipedia	5
Zobodat	5



I22 - Do you have any additional feedback you would like to share about your experience using the Biodiversity Heritage Library or comments about new features?

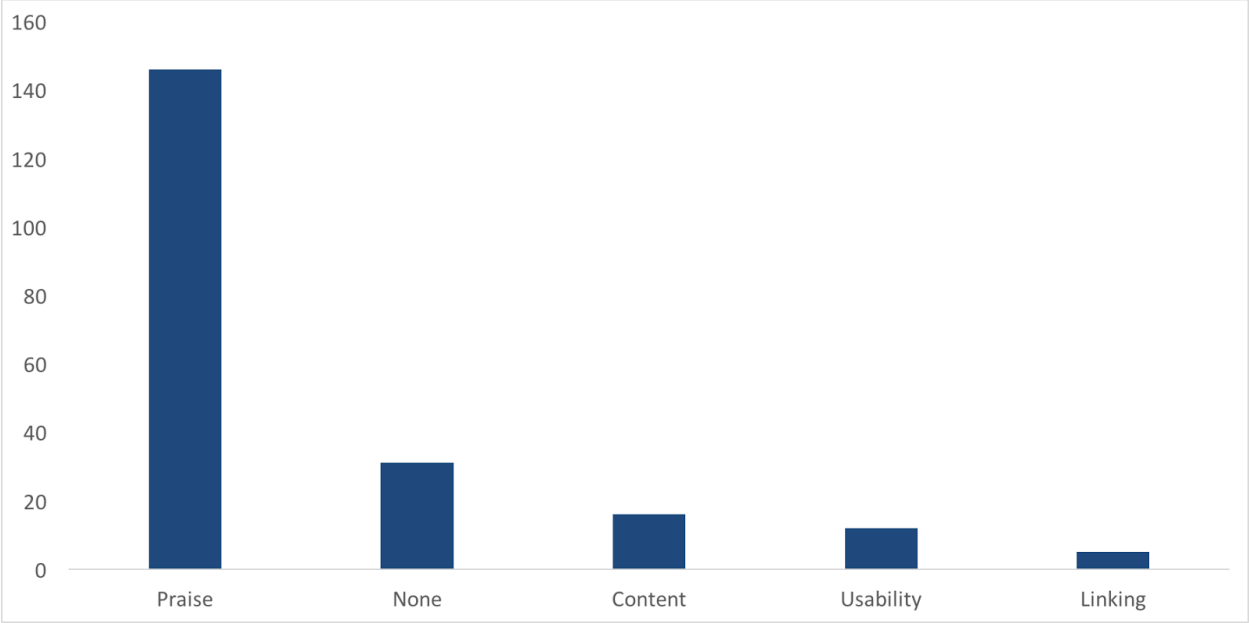
[212 answered; 394 skipped] [Multiple Choice]

Categories were determined from the free text responses and tallied. Nearly 70% of respondents who answered this question commented with a “Praise” for BHL. The second most received response type were participants who said no further feedback. The next highest category was “Content,” and respondents asked for either more content, noted missing volumes or other gaps in coverage, or reiterated this is one of BHL’s most important features. More details on “Praise” comments and a brief description of each category can be found in Appendix D.

Category	Tally	Percentage
Praise	146	69.19%
None	31	14.69%
Content	16	7.58%
Usability	12	5.69%
Linking	5	2.37%
Remove	4	1.90%
Feedback	4	1.90%
Documentation-Outreach	3	1.42%
Illustrations	3	1.42%

Named Entities	3	1.42%
Download	2	0.95%
Features	2	0.95%
Images	2	0.95%
Metadata	2	0.95%
OCR	2	0.95%
Personalized features	2	0.95%
Search	2	0.95%
Annotations	1	0.47%
Articles	1	0.47%
Citation	1	0.47%
Data	1	0.47%

*Percentage is based on number of respondents to question (not number of assigned categories since a response could have more than one category assigned to it)



*graph uses tally numbers

Consortium User Survey

The Consortium User Survey section contains the definition of consortium users, the purpose and goals of the survey, the survey distribution and description, the sample size calculation, and an analysis of each question from the survey.

Definition

BHL Consortium Users encompass individuals contributing content to, or other otherwise participating in, BHL as part of the BHL Consortium including Members, Affiliates, Partners, and staff.

Purpose

The purpose of the Consortium User Survey is to collect input from BHL Member, Affiliate, and Partner resentatives and their staff based on their perspectives as participants in the consortium and how they would like to see the digital library evolve. As part of that, this survey provided an opportunity to further delve into preferences and recommendations put forth by BHL Members at the 2017 BHL Annual Meeting. Special emphasis was placed on areas for new technical development and feedback on the current systems and tools used by BHL Partners for contributing content to BHL.

Goals

The goals of this survey are to focus on technical development and gathering further detail from broad discussions at the BHL Annual Meeting, as well as soliciting input on the systems for uploading content into BHL used by the Partners.

Distribution

This survey was distributed via email to the BHL Partners, Members, and Staff Listservs that include at least one contact representative of each BHL Partner (Member and Affiliate) organization (176 email addresses).

Description

This survey was live for approximately three weeks from September 12, 2017 to October 1, 2017. The survey consists of 10 questions including the informed consent, 2 multiple choice, and 7 free text.

Overall Sample Size

During the three weeks the survey was live, there were 70 total responses. With a distribution of 176 email addresses, this gives a 95% confidence level and a 9% margin of error overall for the

survey. Feedback collected via this survey is especially notable and valuable as it reflects the perspectives of those who have made a formal commitment to participating in BHL and have a vested interest in the future of BHL.

Analysis

The Analysis section presents each question from the Consortium User Survey, the number of participants who answered and the number skipped, notation if it is a multiple choice or free text question, a brief narrative analysis, a table of the results, and when appropriate a graph of the results.

C1 - Informed Consent

[70 respondents, all agreed to the Informed Consent Statement]

See Appendix E for Informed Consent Statement

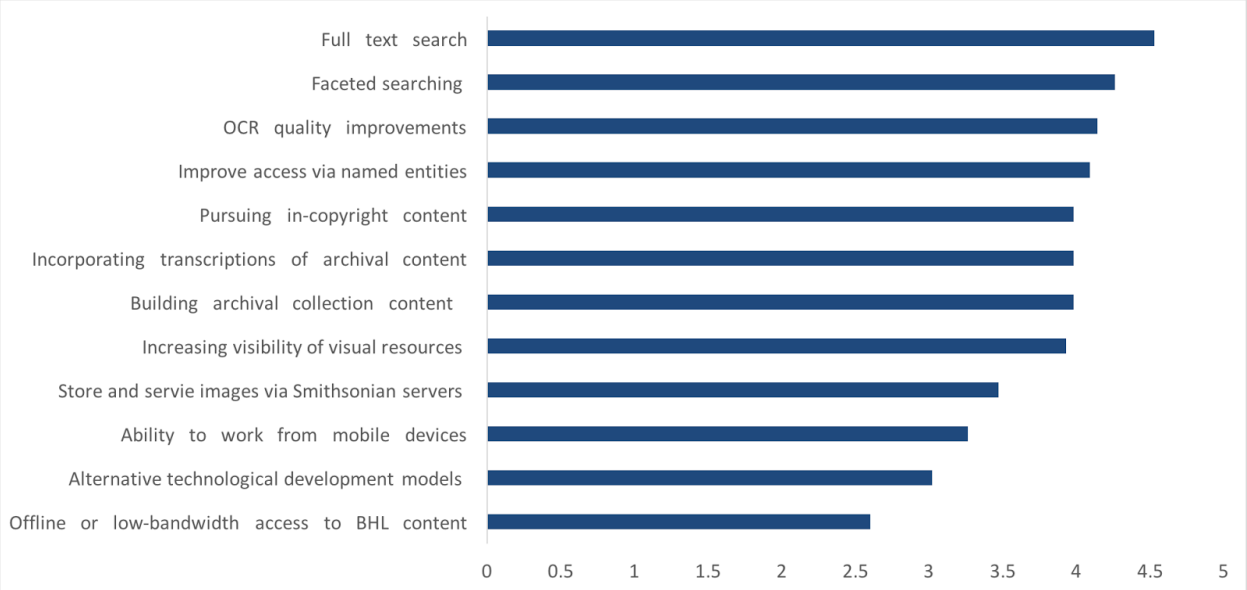
C2 - Below is a list of the proposed enhancements/tools that garnered the most consensus at the recent BHL Members Meeting and through other discussions with Partners. Please rate your level of interest in these services being implemented in BHL.

[58 answered; 12 skipped] [Multiple choice]

Participants were asked to rate areas of BHL technical development on a scale of not interested to extremely interested. When comparing weighted average, survey respondents were most interested in “Full text search” and “Faceted searching.” “OCR quality improvements” and “Improve access via named entities” were the next highest rated.

Answer Choices	Total	Weighted Average
Full text search	57	4.53
OCR quality improvements	58	4.14
Faceted searching (e.g. search by manuscripts vs published, book/journal/archival material, copyright status, illustrations)	58	4.26
Improve access via scientific names and other named entities search (taxon names, author names, geographic names, publisher names)	57	4.09
Ability to store and serve images from Smithsonian servers (i.e., as a failover for Internet Archive delivery and for countries where IA may be blocked)	57	3.47

Ability to work from mobile devices	58	3.26
Offline or low-bandwidth access to BHL content	58	2.6
Alternative technological development models (i.e., where third parties would develop apps and services from BHL open data)	58	3.02
Building archival collection content (e.g. correspondence, field notes, diaries)	58	3.98
Incorporating transcriptions of archival content	58	3.98
Increasing visibility of visual resources (e.g. illustrations, photographs, maps, graphs)	57	3.93
Pursuing in-copyright content	58	3.98



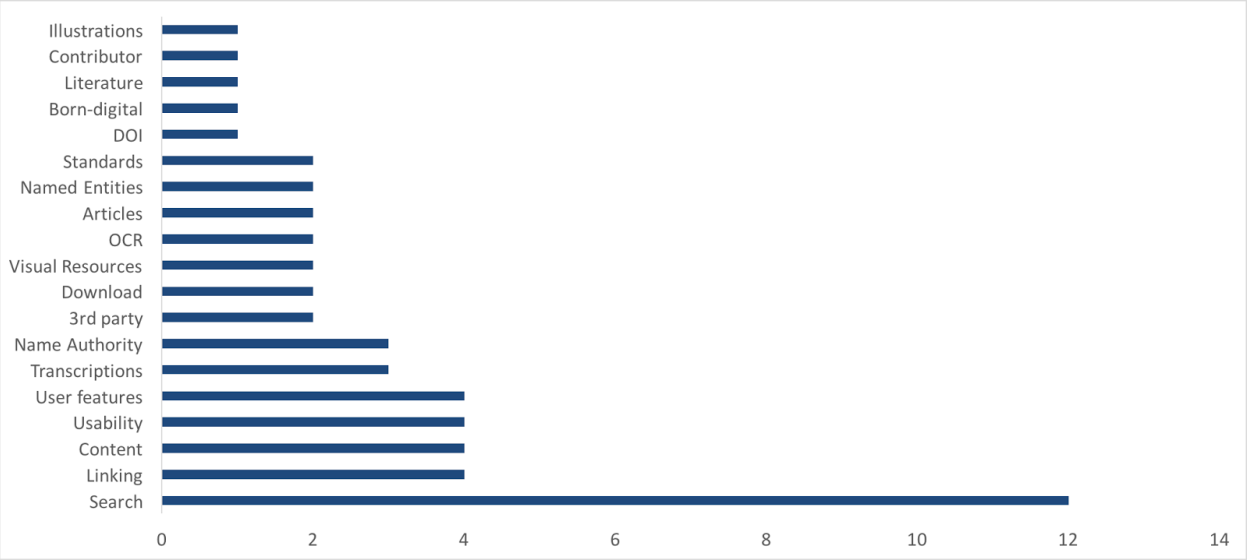
C3 - (Follow up to Q2) Please provide any additional details on what you would like to see come out of implementing and providing the services listed above, or if there is another tool or feature that you feel should be a priority for BHL going forward.

[29 answered; 41 skipped] [Free text]

Categories were determined from the free text responses and tallied. As in C2, the most mentioned topic was “Search.” Survey participants again noted that full-text search should be a number one priority. Faceted searching was again mentioned. Other areas of search that were discussed were semantic search, sorting search results by date, specialized subject headings list, improving search-ability and search rank, increasing discoverability of collections, adding

copy-specific metadata that could be searched on, and de-emphasizing browse and re-emphasizing and restructuring advanced search. A brief description of each category can be found in Appendix E.

Category	Tally
Search	12
Linking	4
Content	4
Usability	4
User features	4
Transcriptions	3
Name Authority	3
3rd party	2
Download	2
Visual Resources	2
OCR corrections	2
Articles	2
Named Entities	2
Standards	2
DOI	1
Born-digital	1
Literature	1
Contributor	1
Illustrations	1



C4 - What do you see as the primary challenge to implementing these tools and services?

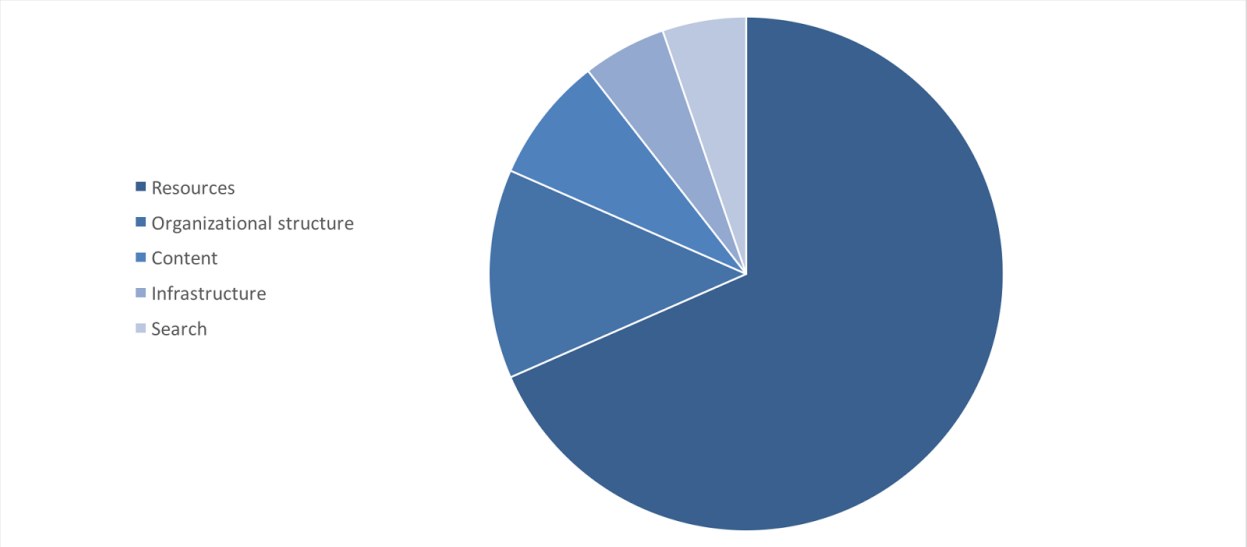
[38 answered; 32 skipped] [Free text]

Categories were determined from the free text responses and tallied. Overwhelmingly, “Resources” was noted as the primary challenge to implementing the tools and services. There were many comments related to this, but they fell into five main themes - money, time/effort, technology, staff, and expertise. A brief description of these themes and each category can be found in Appendix E.

Category	Tally
Resources	26
Organizational structure	5
Content	3
Infrastructure	2
Search	2
Articles	1
Upload	1
Format	1
Metadata	1

IA	1
Transcriptions	1
DOI	1
OCR	1
Named Entities	1
Archival content	1
Usability	1
3rd party	1
Training	1
Update Data model	1
Creativity	1

graph displays top 5 categories:



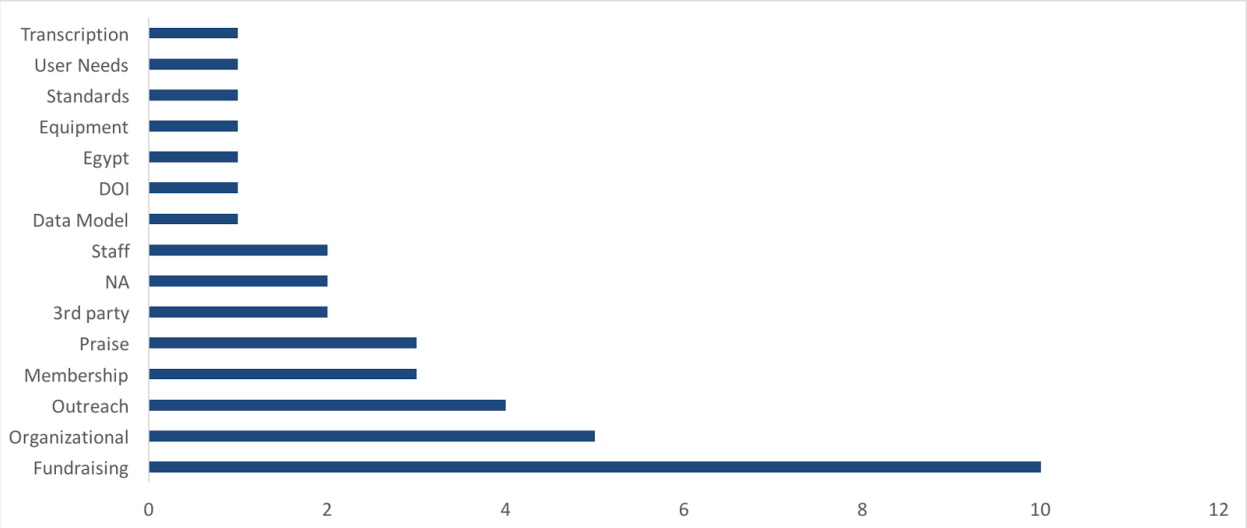
C5 - (Follow up to Q4) How do you see BHL overcoming this challenge?

[32 answered; 38 skipped] [Free text]

Categories were determined from the free text responses and tallied. With “Resources” being the highest tally in the previous question, “Fundraising” is the highest tally for overcoming that challenge. Suggestions included looking for new and different revenue sources, applying to

more grants, trying different membership models, establishing an endowment, and looking for funding specifically for development separate from maintenance and administrative costs. A brief description of each category can be found in Appendix E.

Category	Tally
Fundraising	10
Organizational	5
Outreach	4
Membership	3
Praise	3
3rd party	2
NA	2
Staff	2
Update Data Model	1
DOI	1
Egypt	1
Equipment	1
Standards	1
User Needs	1
Transcription	1

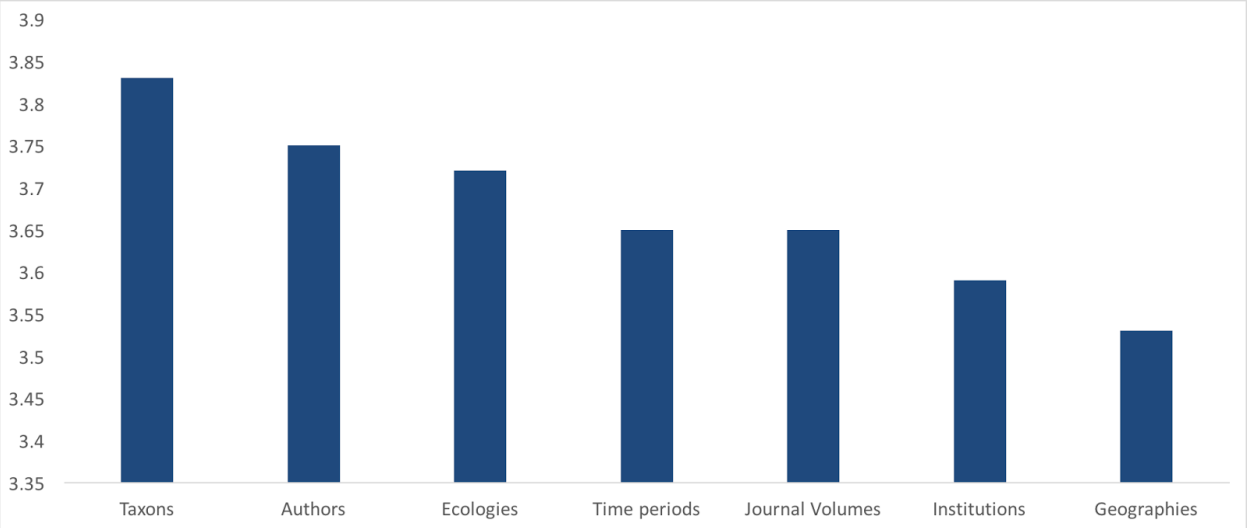


C6 - Please rate your satisfaction with BHL's content coverage in the following areas:

[55 answered; 15 skipped] [Multiple Choice]

Participants were asked to rate areas of BHL collection coverage on a scale of not satisfied to extremely satisfied with an option to choose NA. When comparing weighted average, survey respondents were most satisfied with coverage of "taxons" and across "authors."

Answer Choices	Total	Weighted Average
Taxons	53	3.83
Time periods	54	3.65
Ecologies	53	3.72
Geographies	54	3.53
Institutions	54	3.59
Authors	55	3.75
Journal Volumes	54	3.65



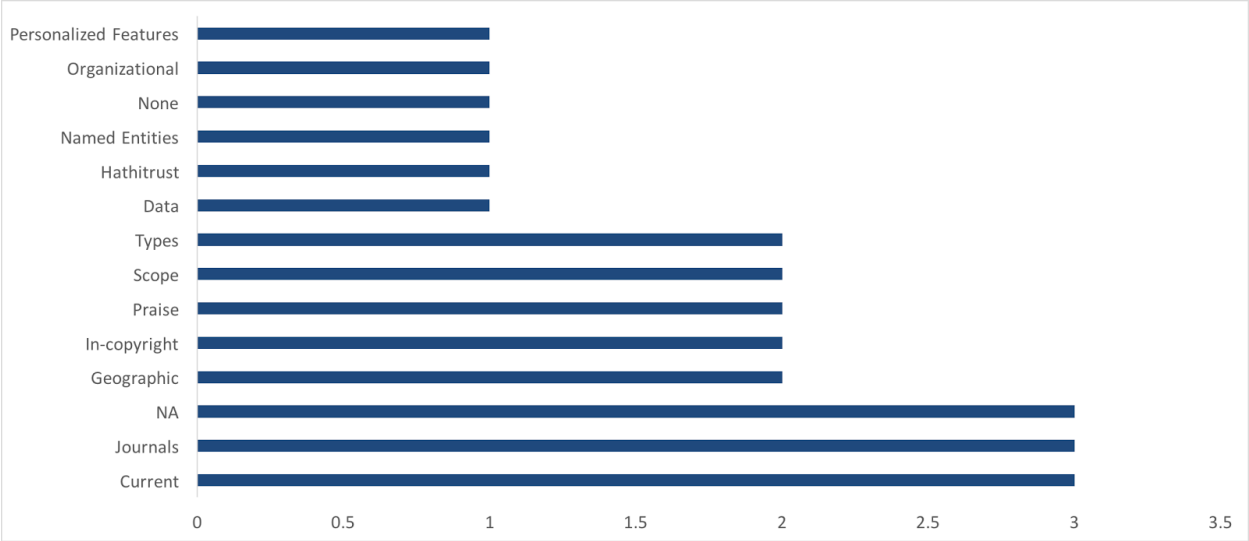
C7 - Do you have any additional comments related to particular strengths or gaps in BHL's content coverage?

[18 answered; 52 skipped] [Free text]

Categories were determined from the free text responses and tallied. There were several respondents that said they were not familiar with BHL collection coverage and those were categorized as NA and not included in the graph. The two categories that tied for most tallies were “Current” and “Journals” and respondents were asking for more current content and to fill in journal runs. A brief description of each category can be found in Appendix E.

Category	Tally
Current	3
Journals gaps	3
NA	3
Geographic	2
In-copyright	2
Praise	2
Scope	2
Types of materials	2
Data	1

Hathitrust	1
Named Entities	1
None	1
Organizational	1
Personalized Features	1



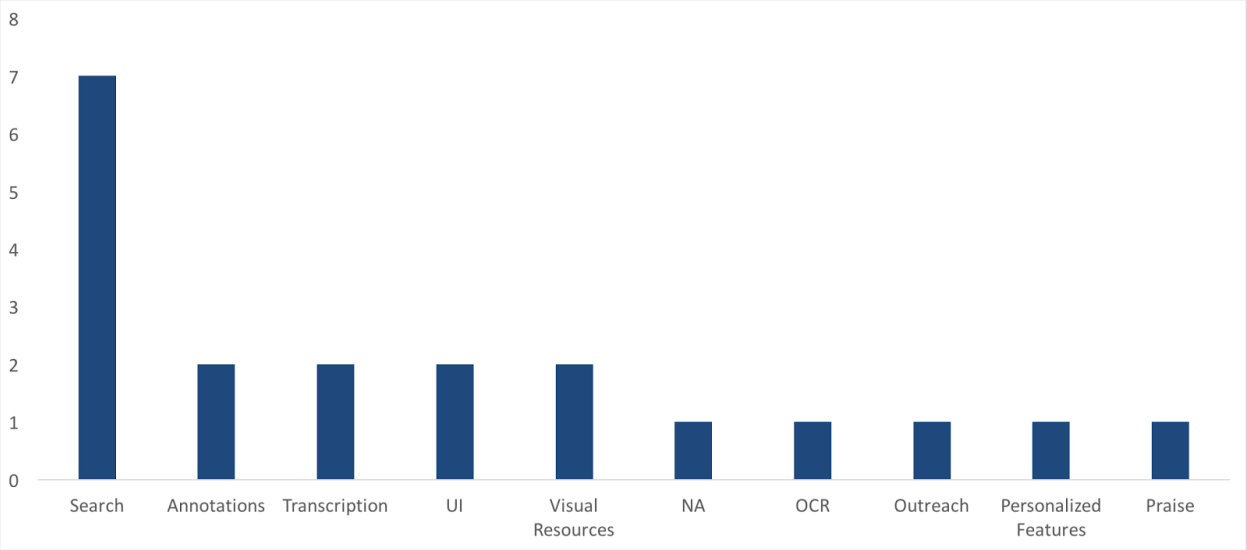
C8 - If you have come across features on other digital library sites that you would like to see on BHL, please describe the feature (links appreciated).

[17 answered; 53 skipped] [Free text]

Categories were determined from the free text responses and tallied. “Search” was the most mentioned feature respondents liked on other sites including full-text search, improved browsing, and keyword tagging of images. A brief description of each category along with examples if provided by survey participants can be found in Appendix E.

Category	Tally
Search	7
Annotations	2
Transcription	2
UI	2

Visual Resources	2
NA	1
OCR	1
Outreach	1
Personalized Features	1
Praise	1



C9 - What has been your experience adding content to BHL via Macaw and Admin Dash? Please share any likes/dislikes, difficulties you may have encountered, or areas of improvement you would recommend.

[31 answered; 39 skipped] [Free text]

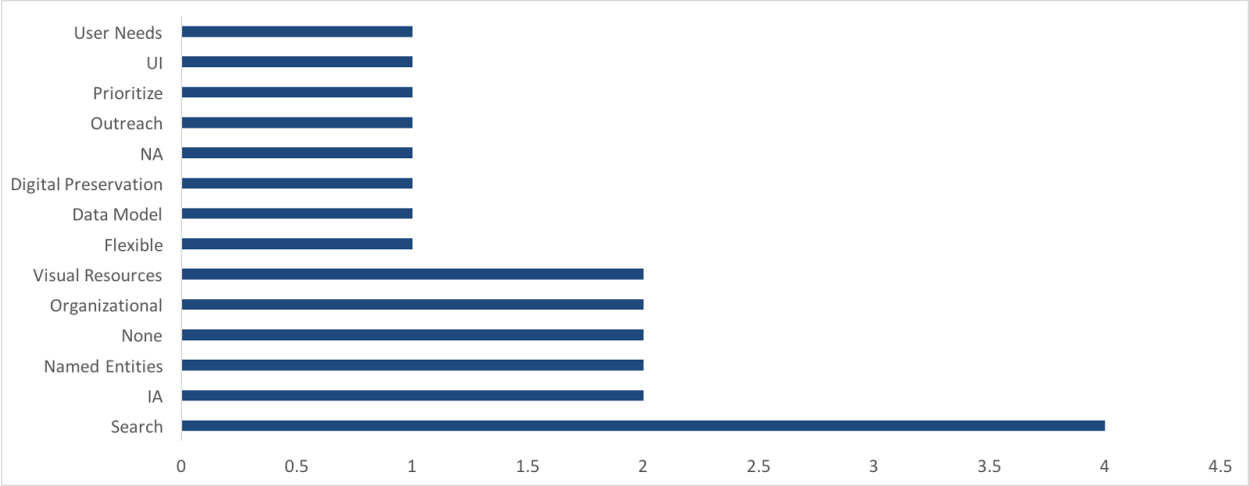
Categories were determined from the free text responses and included details on areas of improvement, pros and cons for Macaw, Admin Dash, and in general. Categories were not tallied due to the multi-part structure of the question. Overall, many survey participants found both systems to be user-friendly and had good experiences working with both systems. Others reported some difficulty learning at first or finding the systems a bit “clunky.” The most common comments on specific areas of improvement included pagination, batch edit, upload time, article markup, and inconsistency in performance. There were many praises for staff being very helpful and resolving any problems quickly. A brief description of the areas of improvement and pros and cons can be found in Appendix E.

C10 - Do you have any additional comments or suggestions to share about technical development priorities for BHL?

[15 answered; 55 skipped] [Free text]

Categories were determined from the free text responses and tallied. The topic most mentioned was again “Search.” Responses included similar comments to previous questions in the survey, such as overall improved search functionality, implementing full-text search, a browse by subject, and faceted search. A brief description of each category can be found in Appendix E.

Category	Tally
Search	4
IA	2
Named Entities	2
None	2
Organizational	2
Visual Resources	2
Flexible	1
Update Data Model	1
Digital Preservation	1
NA	1
Outreach	1
Prioritize	1
UI	1
User Needs	1



System User Survey

The System User Survey section covers the definition of system users, the purpose and goals of the survey, the survey distribution and description, the sample size calculation, and an analysis of each question from the survey.

Definition

For the purposes of this study,he system users are defined as organizations or individuals who interact, on a regular basis or at scale, with BHL usually for the purpose of enriching another system via application programming interfaces (APIs) or manually.

Purpose

The purpose of the System User Survey is to learn how BHL could move toward greater interoperability with other members of the larger biodiversity community (BHL's external partners).

Goals

The goals of this survey are to focus on technical interoperability requirements; and to learn how these partners/users want to be able to work with BHL at the system level either for their own work or for the users of their sites/services and what challenges exist.

Distribution

This survey was distributed via a targeted email to contacts at Encyclopedia of Life, Global Biodiversity Information Facility, Integrated Taxonomic Information System, BIOSTOR, Global

Names, Global Genome Biodiversity Network, Genbank, and the BHL API key holders (245 email addresses).

Description

This survey was live for approximately three weeks from September 12, 2017 to October 1, 2017. The survey consists of 12 questions including the informed consent, 3 multiple choice, and 8 free text.

Overall Sample Size

During the three weeks the survey was live, there were 41 total responses. With a distribution of 245 email addresses, this gives a 95% confidence level and a 14% margin of error overall for the survey. While this survey is a smaller sample size and the results more anecdotal, the information gathered remains highly valuable for BHL as this user group includes both partner organizations and active members of the larger biodiversity community.

Analysis

The Analysis section presents each question from the System User Survey, the number of participants who answered and the number skipped, notation if it is a multiple choice or free text question, a brief narrative analysis, a table of the results, and when appropriate a graph of the results.

S1 - Informed Consent

[41 respondents, all agreed to the Informed Consent Statement]

See Appendix F for the Informed Consent statement

S2 - Do you use the BHL API?

[40 answered; 1 skipped] [Multiple Choice]

Over half of survey respondents answered “Yes” to using the BHL API.

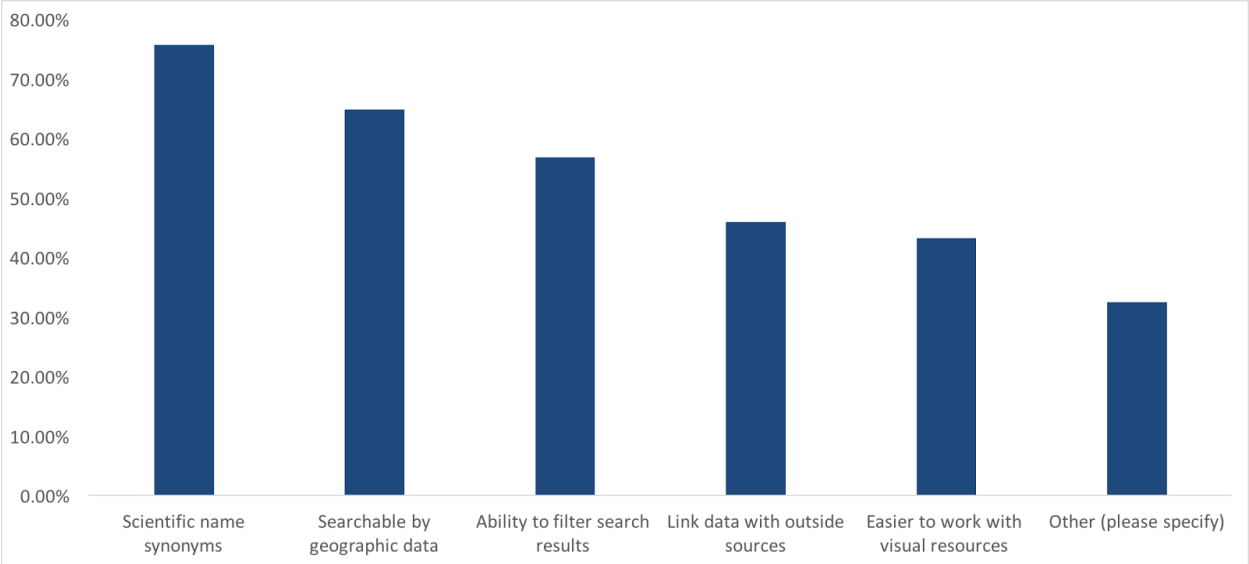
Answer Choices	Responses	
Yes	62.50%	25
No	37.50%	15

S3 - The following are some suggestions we've received from BHL users in the past for enhancements and tools. Which if any of these new services would you like to see implemented by BHL? [Check all that apply]

[37 answered; 4 skipped] [Multiple Choice]

More than half of survey respondents chose "View scientific name synonyms in addition to name in literature," "Geographic data searchable like scientific names," and "Ability to filter search results." The top choice at over 75% was "View scientific name synonyms in addition to name in literature" requesting enhancement to BHL's most unique feature - the ability to search on scientific names. There were 12 "Other" responses that focused mainly on improving overall search functionality, details on improving scientific search, as well as requests for improvements to OCR.

Answer Choices	Responses	
View scientific name synonyms in addition to name in literature	75.68%	28
Geographic data searchable like scientific names	64.86%	24
Ability to filter search results (e.g., by in-copyright; by archival materials; by visual resources)	56.76%	21
Easier way to locate and download visual resources (e.g., illustrations, graphs, maps, photographs)	43.24%	16
Link data with outside sources	45.95%	17
Other (please specify)	32.43%	12



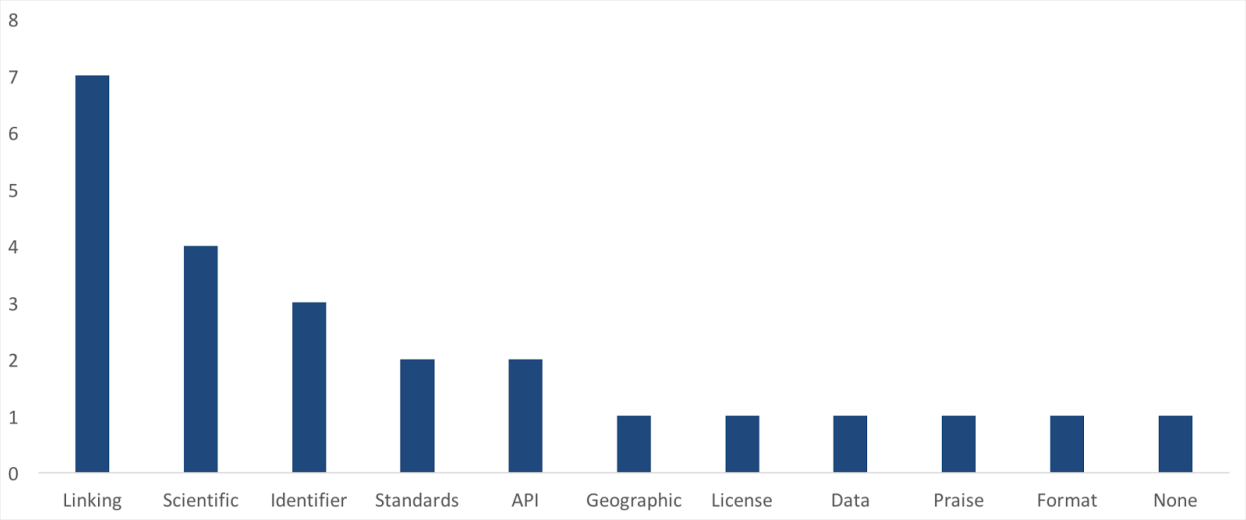
S4 - How would you like to see BHL data services be more interoperable with other data services and systems?

[14 answered; 27 skipped] [Free Text]

Categories were determined from the free text responses and tallied. The topic most mentioned by survey participants was “Linking.” This included linking to other data services, specimens and sequences, aligning with taxonomy (CoL, GBIF, EOL, etc.), and linking to unique identifiers and other data. A brief description of each category can be found in Appendix F.

Category	Tally
Linking	7
Scientific	4
Identifier	3
Standards	2
API	2
Geographic	1
License	1
Data	1
Praise	1

Format	1
None	1



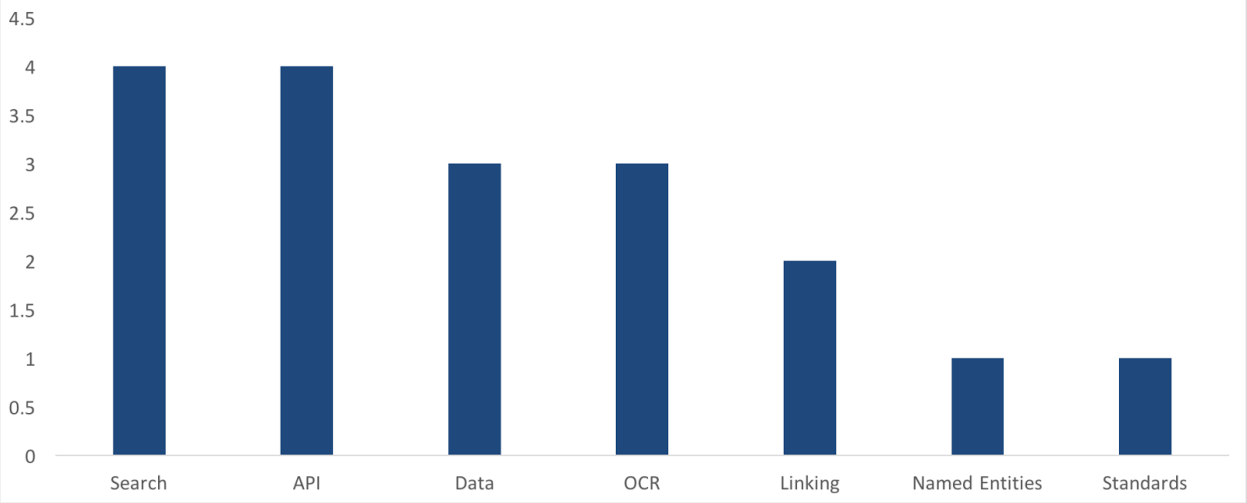
S5 - What do you see as the primary challenge to increasing interoperability with BHL?

[12 answered; 29 skipped] [Free Text]

Categories were determined from the free text responses and tallied. The two categories tied for the top tally were “Search” and “API.” “Search” responses included overall improved search capabilities, fuller results for search terms, better search interface for bibliographic linkages, and search functionality like that of Google Scholar working with BHL. “API” responses included making API a priority and building Google Map like widgets / autocompletes into the API, including linked data definitions in the API, adding a filter mechanism in the API to reduce number of results, and creating standards from the community for the API. A brief description of each category can be found in Appendix F.

Category	Tally
Search	4
API	4
Data	3
OCR	3
Linking	2
Named Entities	1

Standards	1

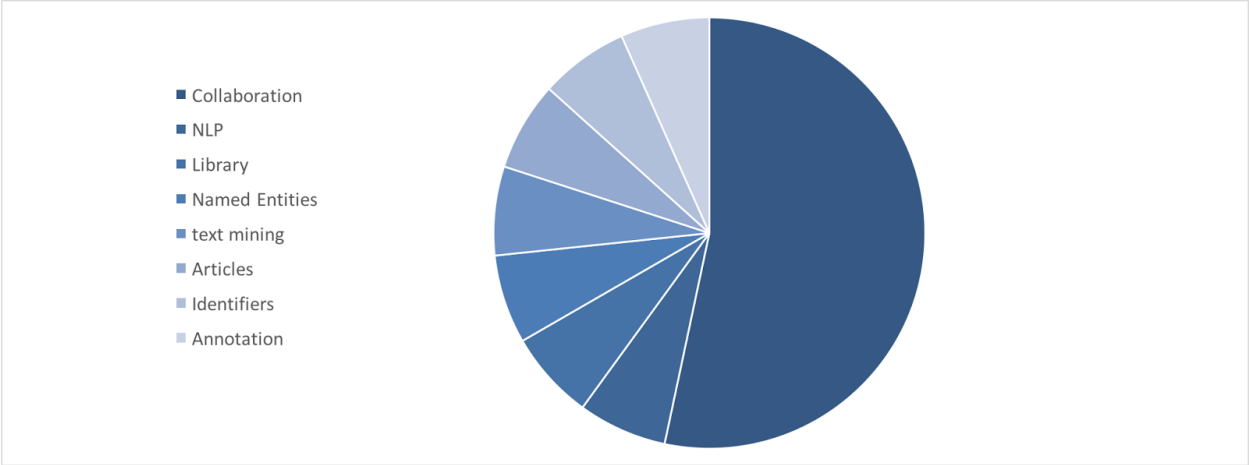


S6 - How do you see us working together to overcome this challenge?

[13 answered; 28 skipped] [Free Text]

Categories were determined from the free text responses and tallied. Overwhelmingly the top tally was “Collaboration.” Survey respondents suggested collaborating with over biodiversity organizations, working with others to agree on unique identifiers, being involved in hackathon projects, working with the BHL community to test prototypes of new versions / features, working with other databases / data services, trying new technologies like NLP and training others, and working more closely with library science experts. A listing of each response can be found in Appendix F.

Category	Tally
Collaboration	8
NLP	1
Library	1
Named Entities	1
text mining	1
Articles	1
Identifiers	1
Annotation	1



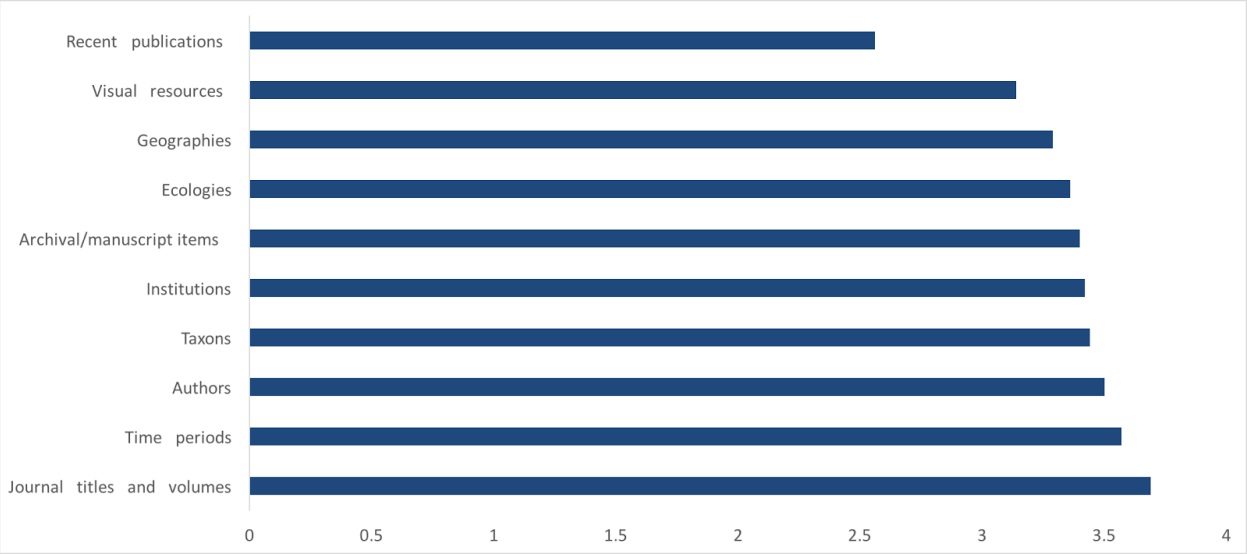
S7 - Please rate your satisfaction with BHL collection coverage in the following areas:

[20 answered; 21 skipped] [Multiple Choice]

Participants were asked to rate areas of BHL collection coverage on a scale of not satisfied to extremely satisfied with an option to choose NA. When comparing weighted average, survey respondents were most satisfied with coverage of “journal titles and volumes” and across “time periods.”

Answer Choices	Total	Weighted Average
Taxons	20	3.44
Time periods	19	3.57
Ecologies	18	3.36
Geographies	19	3.29
Institutions	19	3.42
Authors	19	3.5
Journal titles and volumes	19	3.69
Archival/manuscript items (e.g. correspondence, field notes, diaries)	19	3.4
Visual resources (e.g. illustrations, photographs, maps, graphs)	19	3.14

Recent publications (including in-copyright and born digital materials)	19	2.56



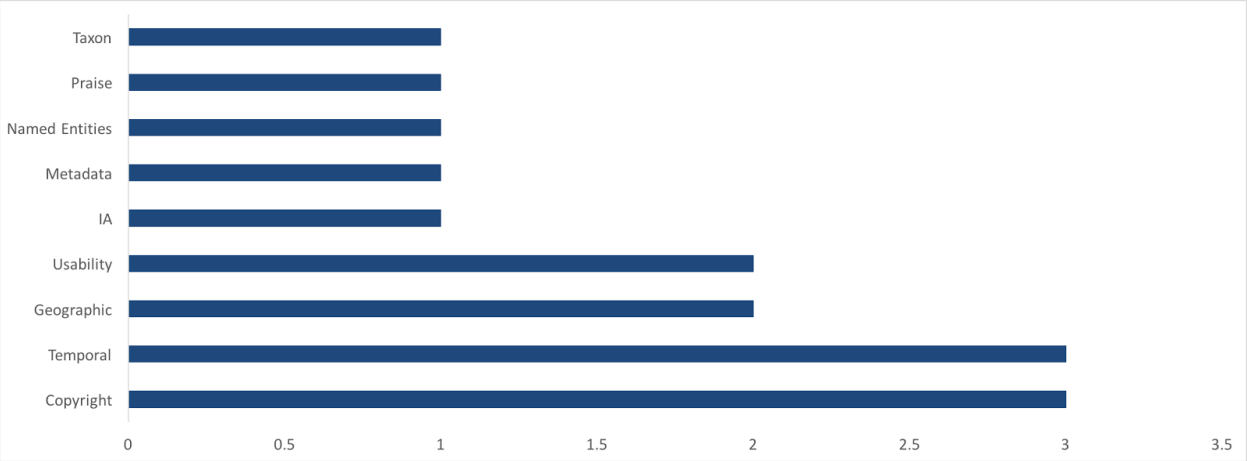
S8 - (Follow-up to Q7) Please provide any details on the above areas related to their usefulness and any particular strengths or gaps in BHL's coverage.

[10 answered; 31 skipped] [Free Text]

Categories were determined from the free text responses and tallied. Four respondents that said they were not familiar with BHL collection coverage and those were categorized as NA and not included in the graph. The two categories with the highest tallys (3 each) were "Copyright" and "Temporal." For "Copyright," three respondents said there are gaps in collection coverage of in-copyright content. Survey participants also had several time periods for suggesting to fill in content. A brief description of each category can be found in Appendix F.

Category	Tally
NA	4
Copyright	3
Temporal	3
Geographic	2
Usability	2
IA	1

Metadata	1
Named Entities	1
Praise	1
Taxon	1



S9 - Does the current offering of data through the BHL API meet your needs? Why or Why not?

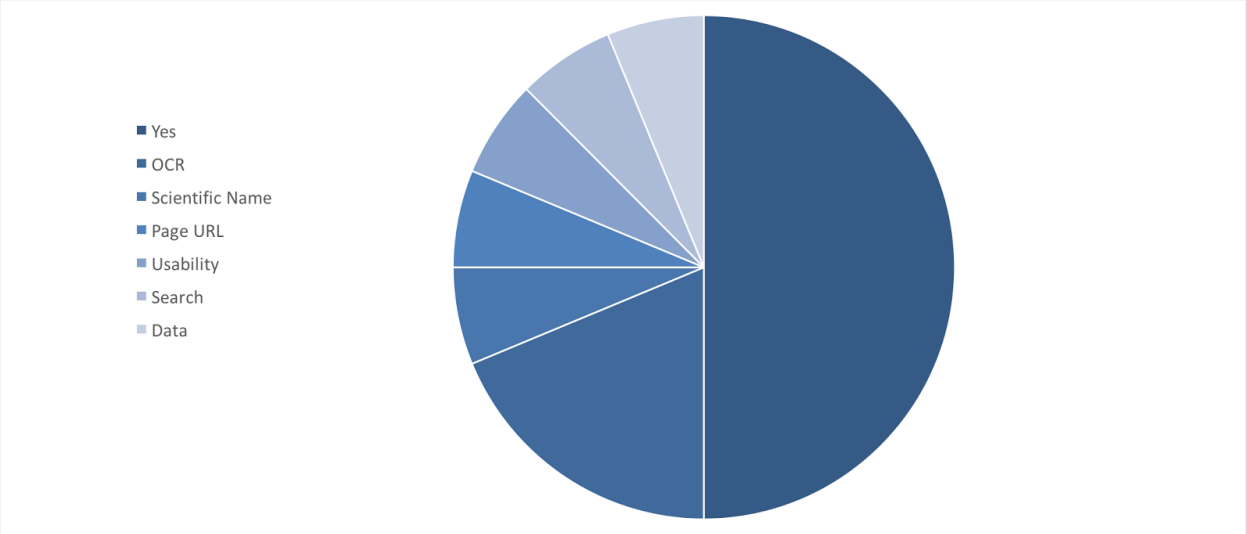
[18 answered; 23 skipped] [Free Text]

Categories were determined from the free text responses and tallied. There were several respondents that said they were not familiar with BHL API and those were categorized as NA and not included in the graph. Eight out of the eighteen respondents said that “Yes” the API meets their needs.

Some suggestions for improvement were easier way to work with the scientific names, better search and filtering, more user-friendly approach to OCR, and cleaner way to get at the tracking data. A listing of each response can be found in Appendix F.

Category	Tally
Yes	8
NA	4
OCR	3

Scientific Name	1
Page URL	1
Usability	1
Search	1
Data	1



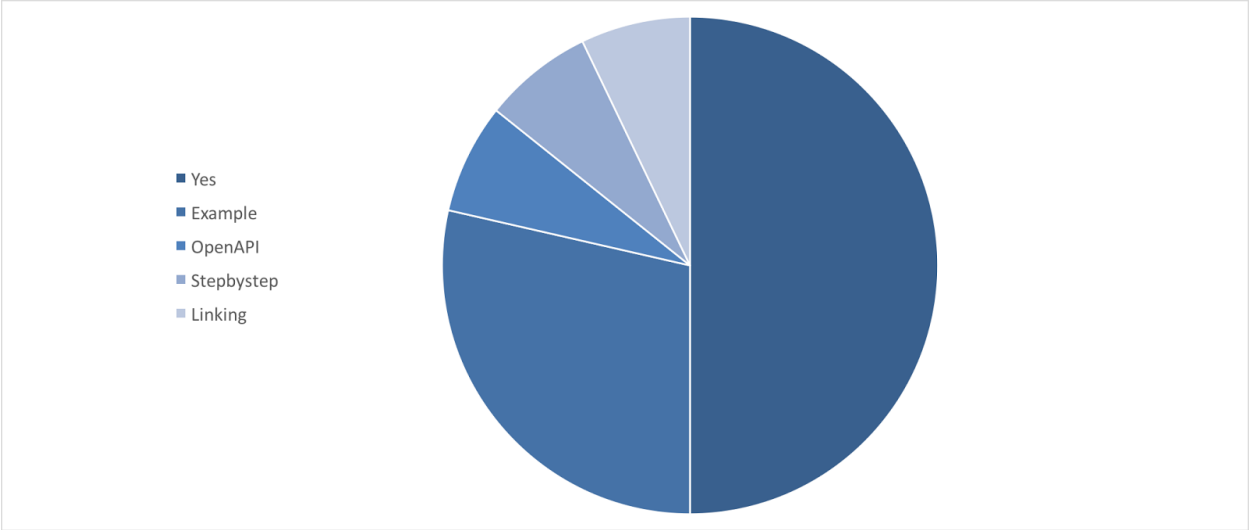
S10 - Does the available documentation on the BHL API meet your needs? Why or why not?

[16 answered; 25 skipped] [Free Text]

Categories were determined from the free text responses and tallied. Five respondents said they did not use the BHL API and those were categorized as NA and not included in the graph. Seven respondents said “Yes” the API documentation met their needs. For the respondents who indicated some improvement was needed for the documentation, four said that adding examples to the documentation would be helpful. A listing of each response can be found in Appendix F.

Category	Tally
Yes	7
NA	5
Example	4

OpenAPI	1
Stepbystep	1
Linking	1

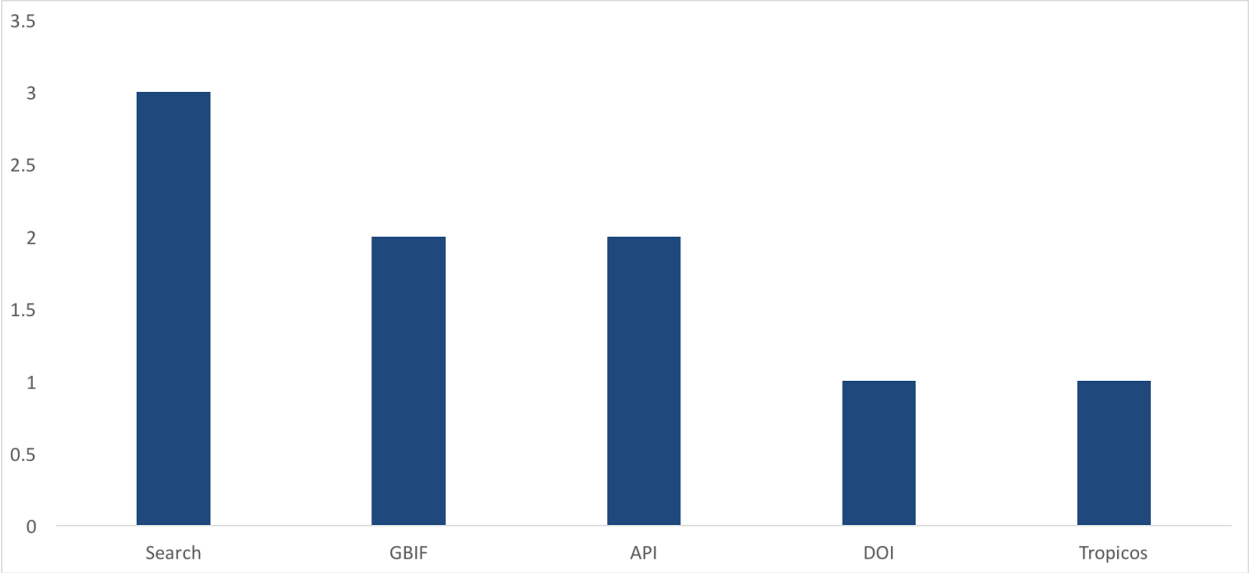


S11 - If you have come across features on other data service sites that you would like to see in BHL, please describe the feature (links appreciated).

[7 answered; 34 skipped] [Free Text]

Categories were determined from the free text responses and tallied. The one “NA” response was removed from the graph. Responses below show both the sites referenced and the features discussed.The most frequently occurring category was “Search.” Respondents commented that search was better on other sites and specifically wanted to be able to sort by date or newest addition at species level. A listing of each response can be found in Appendix F.

Category	Tally
Search	3
GBIF	2
API	2
DOI	1
NA	1
Tropicos	1



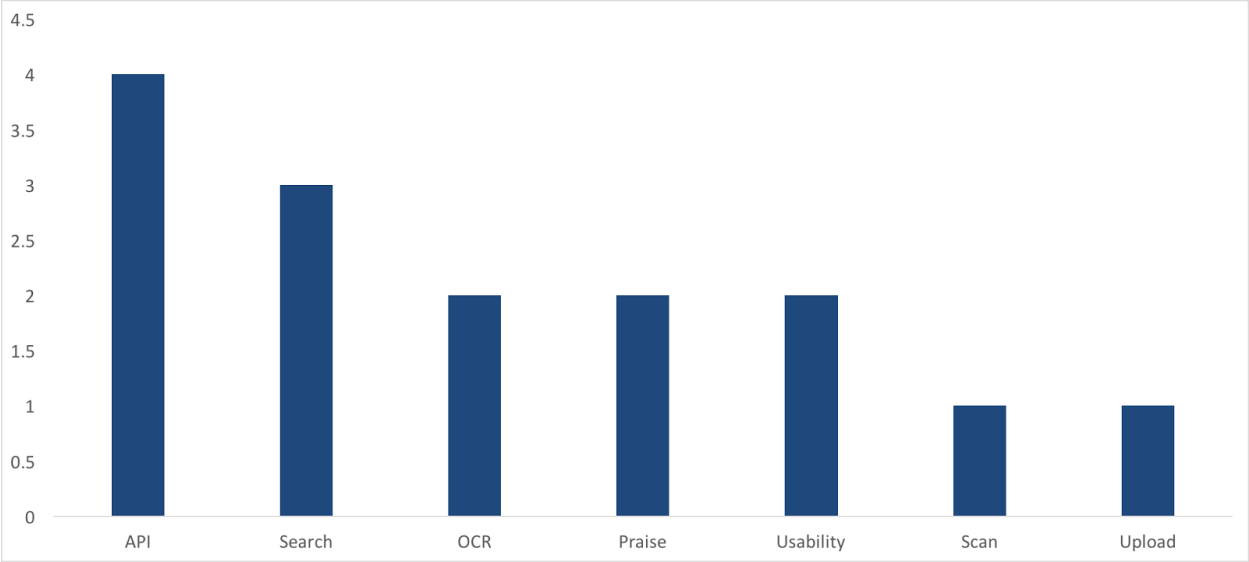
S12 - Do you have any additional feedback to share about your interaction with BHL or any further recommendations on how BHL could make the content you or your users want to work with more easily accessible?

[11 answered; 30 skipped] [Free Text]

Categories were determined from the free text responses and tallied. Four participants provided feedback that was categorized as being related to the BHL“API.” This included a suggestion to provide service over a secure server, more community engagement of developers and informaticians working on BHL, providing more API formats, and one user said the API worked well for their app. Three participants submitted responses related to “Search” including the need for better search-ability, implementing full-text search, and overall improving search capabilities. A brief description of each category can be found in Appendix F.

Category	Tally
API	4
Search	3
OCR	2
Praise	2
Usability	2
Scan	1

Upload	1


4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
API
Search
OCR
Praise
Usability
Scan
Upload

Recommendations

Overall, the user surveys confirmed the primary BHL audience is the scientist / taxonomist community who use BHL frequently as a part of their workflow for professional and scholarly work. Users find the content, download options, and scan quality to be the most highly valuable features.

The resident has compiled recommendations including further user research and user priorities from the surveys. See the resources section in Appendix G for additional reference materials related to these areas.

Further User Research

Building on the surveys, further user research could be beneficial for BHL including usability studies, focus groups, and repeated surveys.

Usability Studies

As identified by the users in the survey along with the Gemini requests and mentions in the interviews, usability does seem to be an issue especially navigating the site, understanding the current features offered by BHL, and the user interface in the book viewer. As BHL is moving forward with its next version, the resident recommends doing usability testing along the way to help ensure the user experience on the site is optimal.

Some specific areas to go deeper for usability are first time users understanding how to use the site and complete specific tasks, navigating the site, especially between the main site and the wiki, and locating the help documentation. Also, user interface with the search options as well as viewing the content and the download, zoom, and other options there in the viewer.

Formative usability testing is done before there is a design prototype and gives feedback for developing the site. These user surveys have already provided a baseline of feedback for developing the site. The resident recommends summative usability testing, which is done once a prototype is ready for review.[6] Internal testing should happen at each stage along the way as designated by the BHL Tech Team. Usability testing should happen with BHL users once a prototype is ready whether that is in stages for a prototype of a new feature such as full-text search, or if that is for many features being launched as a part of a BHL version 1.5 or 2.0.

The resident recommends both moderated in-person and moderated remote usability tests since BHL is a global consortium and recommends learnability usability testing to see how users do accomplishing specific tasks on the site throughout the design process as BHL moves to its

[6] Baxter, Kathy, Catherine Courage, and Kelly Caine, *A Practical Guide to User Research Methods: Understanding Your Users* (New York: Morgan Kaufman, 2015), 432.

next version and implements new features.[7] Multiple rounds of testing are recommended with approximately five participants in each round with representatives from each of the different BHL user groups.[8]

In designing the usability tests, utilizing Google analytics as well as the other analytic tools BHL already uses and the backend user logs for BHL will give baseline information on site statistics for understanding areas that need improvement and developing questions or tasks around these areas. Another site that could provide valuable feedback in usability testing for BHL is www.hotjar.com, which utilizes heatmaps to see user clicks and behavior on a site.

When beginning to design the usability tests and other user studies for BHL, two resources to turn to would be the government user experience site, www.usability.gov, and the industry leader in user testing, the Nielsen Norman Group, https://www.nngroup.com. Additionally, *Research Methods in Library and Information Science* has information on designing user studies for all of these methods specific to the library field.

Focus Groups

Some of the top priority areas are consistent across the survey data, but for some of the other areas of interest, it may be helpful to have some users participate in focus groups to have a deeper understanding of preferences and prioritization from the users. This would be especially helpful when it comes to learning what customization and personalized features should mean for BHL. Focus groups could be used to delve further into the idea of personalized accounts and all of the desired features within that area, as focus groups are most helpful "when you want to answer questions of definitions, value, or policy."[9]

Some specific areas that users have already noted are the ability to save searches, filter RSS feeds, and bookmark favorite items. An annotation feature with user comments that could be turned on and off by users has also shown up in the user research and would be an area of customization to further delve into with a group of users. A focus group may also draw out more information about the interest in data and text mining and learning how users want to manipulate data and / or link data to other sites, which are other topics mentioned across the surveys.

Focus groups are generally most effective with one moderator, one assistant, and about five to ten participants. The moderator will ask questions and engage all participants and the assistant will help with logistics, take notes, and monitor the time. Analysis will work best if the session can be recorded and then later transcribed.[10] For BHL, this will mostly likely need to happen for

---

[7] Sauro, Jeff. "5 Types of Usability Tests," created August 11, 2015, accessed January 7, 2017, https://measuringu.com/five-types-usability/.
[8] Baxter, Courage, and Caine, *Understanding Your Users*, 436.
[9] Stacks, Don W., *Primer of Public Relations Research* (New York: The Guilford Press, 2002), 91.
[10] Stacks, *Public Relations Research*, 91-101.

remote participation with users located throughout the country and world. BHL has several known power users who would most likely provide excellent feedback in this type of format. There are some third party apps that can be used to record Skype calls for example.[11] BHL can also use their communication channels to solicit additional participants if needed or for follow-up focus groups. Keeping in mind this along with all user testing will need to follow Smithsonian Internal Review Board and Privacy Review Board guidelines and approval protocol.

Repeated Surveys

In the past, surveys have been done as needed by BHL, but a consistent check-in with the users would be beneficial to compare data. User surveys "help you to: learn about your current user population and their characteristics; find out the users' likes/dislikes about the current product and track user sentiment over time, and learn how users currently use the system."[12] Surveys help to reach a larger audience than other user testing techniques and can be supplemented with the other user research methods.[13]

A short survey made available on the BHL website on a regular basis, such as annually, would show trends and exactly who is regularly using BHL and what their needs are over time. Three to five questions confirming the audience and use context are a good base for another BHL user survey. A good example to reference is the survey completed by Museum Victoria.[14] As BHL looks to move to its next version, additional surveys or questions related to particular topics on the annual survey may be helpful as implementation is underway. This will provide ongoing feedback and insight from the user community.

Questions should be mostly multiple choice and only one to two open-ended. If using an open-ended question be as specific as possible. In conducting the user surveys analyzed in this report, it was found that many of the same topics and comments came across even when asked different open-ended questions. The amount of time required to analyze open-ended responses is also prohibitive to including a large number of them in surveys going forward.

User Priorities

Across the surveys, the top areas of interest included Search, Named Entities, and Linking.

Search

Search consistently came across as the top priority for users. Users often know the content they want is in BHL, but they struggle with locating it. Many of the issues surrounding search should

---

[11] "How can I record my Skype calls?", accessed January 7, 2017, https://support.skype.com/en/faq/fa12395/how-can-i-record-my-skype-calls.
[12] Baxter, Courage, and Caine, *Understanding Your Users*, 266.
[13] Baxter, Courage, and Caine, *Understanding Your Users*, 267.
[14] Museum Victoria User Survey

be resolved with the implementation of the full text search with faceted search options that BHL already has in progress.

Other search-related areas users identified are correct and complete metadata, ability to fuzzy match, ways to refine search results, better and cleaner browse options for scientific names and authors, ways to view only the most recently added items for a given search, a more intuitive way to display articles within journals, and easier way to find articles. Users also wanted better search options for illustrations, a full list of species to search through, and being able to browse and / or filter by topics such as zoology vs. botany.

Named Entities

Another main priority across the board for users is Named Entities with author names, scientific names, and geographic names being cited most frequently as of interest. This is a known issue for BHL, but name authority for authors was listed as a main frustration in each survey, in gemini, and in interviews.

Also, scientific names is a main priority for BHL. Users find this to be a unique and much appreciated feature, but it also could be improved and being able to search on synonyms and common name are priorities for the users.

Geographic data has also come up along the way as a strong interest for users and would certainly be an area for BHL to pursue going forward. Users would like to be able to search on geographic data like the scientific name feature.

Linking

In each of the surveys and in other feedback, users have mentioned that they would like to see more linking of data. They especially would like to see more or improved linking to other data services like GBIF and EOL.

Users mentioned more linking of taxonomy among biodiversity organizations and within BHL and out to other data services. An issue within the biodiversity community itself is the lack of a common taxonomic backbone to work from, and BHL will need to stay apprised of developments in this area to accomplish linking among different taxonomic hierarchies. BHL is already partnering with ITIS, GBIF, EOL, and others on this issue and should continue to follow them for updates.

Additionally, linking with standardized unique identifiers used by other biodiversity groups and in biodiversity databases and aggregators has been mentioned by many users. Some identifiers mentioned in the surveys were taxon IDs, URIs from OBO Foundry for biological and habitat terms, Geonames, and wikidata.

Conclusion

BHL is an amazing resource to the biodiversity community offering free and open access to natural history legacy literature with unique features, namely the much user-appreciated scientific name finding tool. The projects the BHL-NDSR residents have completed over the past year will be providing research to help BHL move to its next version and offer even more features and services to the natural history and taxonomy users, as well as users from many other disciplines.

The work of the Users Needs Analysis resident has given BHL a tremendous amount of data about BHL users and feedback directly from these users. This will be helpful for BHL as they move forward with additional user studies and as they implement their priority features and services while being informed by their users' needs. Once again, thank you to everyone who made this project possible including BHL, IMLS, NDSR residents and mentors, and all the users who participated in the study.

Appendix A - Supplemental Material

This section includes a listing of the trainings, presentations, and blog posts completed during the course of the residency.

Trainings

A list of the trainings completed by the resident.

Training Name	Date
BHL Bootcamp	Feb 1, 2, 3
iDigBio Darwin Core webinar	Feb 7
DataOne webinar Reproducible Science	Feb 14
ASIST webinar Top Trends and Tips in LIS Research	Feb 16
AAPB NDSR webinar Intro to Data Manipulation with Python CSV	Feb 23
Webinar Social Media Photos as a Source of Primary Biodiversity Data	Mar 1
Smithsonian Training: Human Subjects in Research (Certificate of completion received)	Mar 3
iDigBio webinar Darwin Core Hour: Even Simple is Hard	Mar 7
webinar Basic Principles and Practices of Born-Digital Preservation	Mar 7
iDigBio webinar Georeferencing data	Mar 16
BHL Annual Meeting update and brownbag (Martin)	Mar 27
Introduction to Library Resources for the Sciences (NMNH Main Library)	Mar 28
Library Juice Course: Getting to Know your Users through Interviews and Focus Groups	Apr 3-28
iDigBio webinar Digital Storytelling	Apr 12
iDigBio webinar Darwin Core Hour: GBIF and iDigBio data aggregators	Aug 15

MD SLA - Show Me the Numbers: Storytelling with Data and Visualizations - conference	Oc 12
Smithsonian Environmental Research Center Tour and Science librarianship discussion	Oct 19
NDSA DigiPres conference	Oct 25-26
Folger Shakespeare Library Tour and discussion on digital libraries and user needs with their Digital Media & Publications team	Dec 13

Presentations and Conferences

A list of the presentations completed during the residency with links to slides or recordings as appropriate.

- Presenter, “User Needs Analysis,” BHL NDSR Webinar Series, November 2017, iDigBio AdobeConnect platform, (Recording: https://idigbio.adobeconnect.com/_a1130716096/pd0piuxxvs2z/).
- Co-Presenter, “An Evolving Portal: Planning for the Future of the Biodiversity Heritage Library,” DLF Forum, October 2017, Westin Convention Center, Pittsburgh, Pennsylvania, (Slides: https://osf.io/4axh7/).
- Co-Author, “BHL’s Feedback Tools & User Surveys: Investigating User Needs for Data in Digital Libraries,” TDWG (Biodiversity Informatics) Annual conference, October 2017, Ottawa, Canada, (Slides: https://www.slideshare.net/carolynsheffield/capitale-tuesday-1615sheffieldfinal) .
- Co-Author, “Improving Access to the Biodiversity Heritage Library: Halfway Remarks by National Digital Stewardship Residents,” ALA Annual, June 2017, McCormick Place Convention Center, Chicago, Illinois.
- Co-Presenter, “Building a Communication Network for Collaborative Projects,” NDSR Symposium, April 2017, Library of Congress, Washington, DC.
- Co-Presenter, “Foundations to Actions: Extending Innovations in Digital Libraries in Partnership with NDSR Learners,” DPLAfest, April 2017, Harold Washington Library Center, Chicago, Illinois, (Slides: http://schd.ws/hosted_files/dplafest2017/eb/DPLAFest%20presentation%20FINAL%20with%20notes.pptx).

Blog Posts

A list of the blog posts completed by the resident with links as appropriate.

Resident Blog - https://ndsrbhl.wordpress.com

Posts by resident:

User Survey Update 11/6/2017
https://ndsrbhl.wordpress.com/2017/11/06/user-survey-update/
Getting to Know the BHL Users 5/17/2017
https://ndsrbhl.wordpress.com/2017/05/17/getting-to-know-the-bhl-users/

Posts by cohort:

BHL Bootcamp 3/3/2017
https://ndsrbhl.wordpress.com/2017/03/03/bhl-bootcamp/
DPLAFest and NDSR Symposium 5/4/2017
https://ndsrbhl.wordpress.com/2017/05/04/dplafest-and-ndsr-symposium/
BHL NDSR at ALA 7/7/2017
https://ndsrbhl.wordpress.com/2017/07/07/bhl-ndsr-at-ala/
Preliminary Recommendations from NDSR Residents 9/22/2017
https://ndsrbhl.wordpress.com/2017/09/22/preliminary-recommendations-from-ndsr-residents/
Presentation Roundup 11/30/2017
https://ndsrbhl.wordpress.com/2017/11/30/presentation-roundup/

BHL Blog - http://blog.biodiversitylibrary.org

Posts by cohort:

NDSR Residents Update 9/19/2017
http://blog.biodiversitylibrary.org/2017/09/ndsr-residents-update.html
NDSR Residents Mid Year Update 7/18/2017
http://blog.biodiversitylibrary.org/2017/07/ndsr-residents-mid-year-update.html
An Update from NDSR Residents 5/16/2017
http://blog.biodiversitylibrary.org/2017/05/an-update-from-ndsr-residents.html
Keeping Up with NDSR 3/21/2017
http://blog.biodiversitylibrary.org/2017/03/keeping-up-with-ndsr.html

SI Internal Blog (only accessible to Smithsonian staff)

Introduction post by resident: Hello from Pam, NDSR BHL Resident 1/18/2017
https://collab.si.edu/sites/OUSS-SIL/Libraries/service/highlights/Lists/Posts/Post.aspx?List=1f925d7a%2D9e51%2D4ce5%2Dbd8d%2D3cd42bbe07c4&ID=1265&Web=8fd5b306%2Dd38b%2D4d2d%2D8563%2D05d4945b8326

Appendix B - Informal Questionnaire

The resident developed and distributed an informal questionnaire to the BHL leadership including the Executive Council, Secretariat, and Technical Advisory Group to learn about their priorities and expectations for the next version of BHL.

The questionnaire was created in Google forms and consisted of a total of eight questions (one multiple choice and seven open-ended). It was distributed via email and responses were collected over an approximately one week timeframe. Nine respondents participated in the questionnaire. Each question and all responses received are listed below.

Q1 - Are you a member of EC, Secretariat, or Tech Team?

[9 answered]

EC:	11.1%
Secretariat:	44.4%
Tech Team:	44.4%

Q2 - Where do you see the priorities for the next version of BHL? What features or functionality would be most appreciated by BHL Partners?

[8 answered; 1 skipped]

- Full text search, enhanced ability to work with names, e.g. synonyms nomenclature acts, etc.
- Full text search; Ability to facet and filter results by subjects, keywords, relationships, and maybe even contributors, etc.; easier way to download high resolution images from BHL; integration of crowdsourced data into BHL; easier way to browse and search for just images; hosting and serving BHL images from SIL rather than IA; way for users to annotate content in BHL; improved mobile responsiveness; better integration with our outreach and content marketing; way to mark and search for first description of a species; easier ways to access usage and content stats for partners and users (I'm thinking some sort of public dashboard, but something that doesn't require heavy manual maintenance); way to allow users to enhance content directly within BHL (i.e. tag images, markup taxonomic names, correct OCR); way to link literature and objects in other databases more easily.
- Sometimes 'partner' refers to BHL members and affiliates and sometimes it refers to like-minded organizations such as EOL and IA. I don't know what 'partner' means in this

question. I have a hard time talking about 'the next version' of BHL because to me that language implies a complete redesign and rewrite and I don't see the need for that. Instead, I see the need for continuous, incremental improvement and enhancement. My list of what those improvements should be is long.. Better handling of archival material, faceted search, full text search, interoperability with transcription platforms such as FromthePage, DigiVol and the Smithsonian Transcription platform. Keeping the data scrubbed is of the utmost importance and is almost impossible in the absence of a batch edit function. Why is routine data cleanup being done by BHL's only programmer whose skills are desperately needed for true programming? The ability to download a PDF file with a searchable text layer is long overdue. I'd like to see BHL behave like commercial e-content databases, e.g. offer single-click download of articles.

- Improved search and filtering functionality; provide functionality to correct scientific names on pages; ability to add article metadata at the time of scanning/uploading; ability to correct OCR; implement name authority control; provide better option to download single page high resolution images
- The priorities for the next version of BHL should be to make the literature and archives more accessible in such a way that they more effectively support research that can lead to better understanding and conservation of biodiversity. Many BHL Partners have collected transcriptions of some of their contributed collections. I am sure they would love to see those incorporated into and searchable within the BHL interface. Many BHL Partners would also love to be able to highlight their particular collections. To some extent they can already do this. But what if there were a way to quickly search for gems, i.e., beautiful illustrations in the books they've contributed that they can then point a user or maybe even a donor to and let them know how they too can easily find similar materials from their collections in BHL? BHL Partners are also very committed to the open access component of BHL for facilitating research. I think they would like to see more powerful linking capabilities overall. Links to born digital / moving wall titles for which the content is freely (no paywall) available online but for which we don't yet have permission to display in BHL. Or links to related collections at their institutions (specimens, works of art). Or more links to external biodiversity resources such as GBIF and EOL.
- Image discovery, improved taxonomic authority control (linking of name changes), collection development, improved metadata, expansion of copyrighted material included/deep links to current content. Also adding appropriate members (helps with funding) across the world: filling in gaps in geographic areas for membership (Russia, Asia, South America...and more)
- I'd like see the innovative stuff we did as part of the grant funded projects become core-BHL and part of the daily workflow. Needed Features - image searching, and browsing. Transcriptions should be incorporated into the portal and transcription work by the public should continue whether inside or outside of the portal (doesn't matter as long as we have an automated way to move it into the portal). OCR correction should be an option for the public - gaming is one way to do that but there are other methods. Annotating BHL pages and sections - users want to be able to discuss, comment on and

in some cases argue against the text and images. We need functionality that allows users to do that directly in the portal.
- ➢ Eliminate reliance on Internet Archive (and all that it implies), integration of Macaw into BHL directly, more crowdsourcing (and accepting that it won't ever be perfect data)

Q3 - What does BHL do really well currently and should continue to do as it moves to the next version?

[8 answered; 1 skipped]

- ➢ The ability for users to generate custom PDFs was introduced about 8 years ago, and the usage of that feature has yet to plateau. Probably the most successful feature of the site.
- ➢ The book viewer interface is simple and intuitive; taxonomic name finding is a valuable tool, but OCR needs to be enhanced to allow more name finding; custom PDF download is a oft-praised and valued feature; feedback system that allows users to ask questions and get answers from BHL staff; IA ingest allows us to gain valuable content that we wouldn't otherwise have if we focused just on partners.
- ➢ The content is extraordinary and we should continue to add needed content while avoiding duplication.
- ➢ BHL is and does best as a digital library but even its catalog does not serve some basic functions that a typical ILS provides, like name authority control. We should continue to build a collection of biodiversity relevant materials and provide options for accessing and downloading book like objects. Providing free and open access to content has always been our biggest draw and greatest success, especially to our core user groups.
- ➢ One of BHL's greatest strengths has been a clear scope. Supporting biodiversity research is our north star. I hope that we will continue to maintain focus even as we continue to grow and expand. In terms of features: Scientific names are also an important access point. Feedback submission should remain available through the website, though could perhaps be more prominent. Our bookviewer seems to already work well. We need advanced search options but those could possibly be revised. Links to EOL are a nice touch. Perhaps there could be a way to incorporate more links between systems without getting too messy?
- ➢ Provide full text, quality literature with complete versions of journal runs and content (foldouts).
- ➢ Manage user feedback, coordinate tasks, social media outreach
- ➢ Everything it currently does?

Q4 - What do you see as the biggest challenges to implementing BHL v.2? How can we address those challenges?

[9 answered]

- Money
- Limited developer resources. If bringing on additional developers is out of the question, there may need to be a freeze put on development and maintenance (except for emergencies) on the existing site while v.2 is built.
- I believe the biggest challenge is going to be staff and resources. This includes staff and resources for building the platform, conducting extensive tests with users and technology to assess functionality and usability, and providing adequate training and tutorials for the new tools and features. Honestly, I think we need to get some grants to support this work. I do not see our current staff as being able to keep the day-to-day project running while also building this new platform.
- The biggest challenge is the lack of manpower and the lack of effective decision making. It's frequently hard to determine who has the authority to make a decision. Uncertain funding is also a problem.
- Resources! We need web design expertise to help with UI changes. My understanding is Mike, while great at the backend functionality, is not as skilled for front end design. Also need the ability to perform well-founded user testing and better assess usability of the BHL web site as it stands now and into the future. Addressing the resource challenge is always tricky since it requires funding: I believe BHL Egypt is providing technical development resources at no cost? To what extent can they perform this work? Are there student resources that we can tap into for cheap labor? Are there pro bono tech development programs available that could give BHL a couple days of their time? Would a few BHL Member orgs have the ability to request 2-3 days of technical development time to host a kind of cross-institutional agile development session to implement BHL improvements?
- The biggest challenges will be focus and resources. There are endless possibilities and opportunities for providing access to our content in new ways. To help maintain focus, it will be helpful for us to develop clear criteria for prioritizing changes and choosing between things when necessary. To implement BHL v.2, we will need to increase resources, particularly in terms of development staff. BHL Egypt staff may be part of that solution but we will likely need others. We should consider and plan for needing additional developer time, UI expertise or other expertise for which we will need to raise funds. If we would like to set a specific launch date, we will also need dedicated oversight of selected approach, clearly defined deliverables, and well-managed project schedule.
- Resources--human and dollars. Fundraising, grants, more dedicated staff; adding full members

- ➢ Funding to implement the next version. I'm not confident we're going to get a significant amt of money to be able to do a complete overhaul of both the backend and front end which is what needs to happen ideally. So realistically I think we need to plan for changes in stages over time so that as funding comes in we can address each area in order of priority.
- ➢ Recreating all existing functionality on a new platform in a timely manner.

Q5 - Who would you describe as BHL's most important external partners and how do you see BHL pursuing further interoperability with these partners in the next version?

[8 answered; 1 skipped]

- ➢ GBIF, museum collections systems e.g. emu
- ➢ Unless BHL is going to take over the business of automatically finding names in text/ocr, then BHL needs to keep a close relationship with Dima Mozzherin and the Global Names project.
- ➢ Global Names is obviously a very important external partner, because one of the key value-add services we offer is taxonomic name finding. Our service is built on GNRD, so we need to maintain a healthy and transparent relationship with them (and ITIS by extension) to keep this service strong and effective. More regular meetings and discussions with this organization would be good. I also think the funding agencies that we've worked with are very important (CLIR, IMLS, MacArthur, Sloan, NEH, etc.). Funding remains our biggest challenge (in my opinion), so regularly meeting with people from these foundations is a good way to learn about major funding opportunities to support future activities.
- ➢ IA, DPLA and transcription platforms. Possibly Wikimedia Foundation. I'd also like to see better interoperability with OCLC, specifically a KBART feed that allows discoverability of BHL content in OCLC.
- ➢ I think BioStor, Global Names Architecture and Internet Archive are our most important partners because we already rely on their services to provide for users of our website. Hopefully we can find ways to work more closely with GNA to help improve their scientific name indexes which will in turn improve BHL's scientific name search. It would be great to be able to plug into BioStor at the time of scanning/uploading to more proactively control the BHL article metadata we harvest from BioStor. It would be great if there was an easier way to fix page image corrections and metadata from IA to BHL -- having some kind of synchronization between our 2 sites would be great if at all possible!
- ➢ It's difficult to select the most important as different partners have different roles, each of which is important in it's own way. GBIF occurrence records, as an example, could be enriched if geographic data in the pages of BHL corpus could be extracted and associated with a published observation of a species observation or collection of a

specimen at a particular time in a particular place. This kind of semantic data (place, maybe weather) could also have usefulness for other external groups such as climate scientists and others. I think the best way to be interoperable with our partners is to here from them on both what they want from BHL and how they are currently managing their systems, and how they plan to manage them in the future.

- GBIF, ICTSI, GNRD, TDWG, EOL, Europeana, DPLA, Atlas of Living Australia/New Zealand
- Not sure

Q6 - What do you feel are the most important areas for improvement needed for BHL from the user perspective?

[7 answered; 2 skipped]

- Search, though we are already working on that.
- Hands down, OCR improvement. Almost all of the negative comments I get from users can be traced back to the quality of our OCR and the issues it causes. I also often hear from users that they want an easier way to download single, high resolution pages. I also think many of our training materials need to be improved, and it needs to be easier for users to find these materials. I think many users would also appreciate more transparency regarding what we're planning to digitize, although this is a much trickier problem to solve.
- Faceted search, full text search, single-click download of articles. Standardization of author names.
- Improvements to scientific name functionality, OCR and ability to annotate content
- It depends on the users. For some it's the desire to filter results, even in the feed of new items. For others, it's the inefficiency of having to search through results for 3 or more different variants of the same author name.
- full text search, image recognition, better taxonomic name linkages (authority control?) and common name (authority control), better connections with GBIF and Museum specimens
- Improved response time on suggested improvements or additions, even if we are never going to go add them.

Q7 - Do you feel there is any area where BHL needs a greater understanding of its users' needs? Is there anything you want to know more about from the BHL users?

[8 answered; 1 skipped]

- I would like to know what users opinions are on the need for user accounts and crowdsourcing. I get the sense there are some users that would really appreciate those features... but does that interest extend beyond those dedicated few?
- I think we need a better understanding of the needs and applications of users in the technical community. We have many data-oriented tools, and as big data becomes increasingly important in scientific research, the ability to access BHL data via APIs and other exports can help our content become increasingly relevant and useful for answering big research questions (and puts us in a wonderful position for new funding opportunities!). However, I feel that we have a very poor understanding of our technical community and their needs. I'm also not sure how to connect with and outreach to this community. Where do they communicate? How do they find news and information? What do they need us to provide from a data standpoint? I would love to understand this community's needs and expectations better, and improve our outreach to this community.
- Is there a requirement to download smaller PDFs quickly? I have never seen a commercial e-content database that delivers PDFs of a size comparable to those delivered by BHL. Huge files are important for entrepreneurs who put sci-art on t-shirts but how important are they to scientists and students? Rod Page envisions a BHL where article PDFs of a modest size are prepared ahead of time and downloaded instantly. What do users think? How much interest is there in customized RSS feeds and the ability to quickly identify new content in a specific field?
- I'm really not sure. I think it would be great to have a more regular way to get feedback from users about BHL website functionality. Issuing regular user surveys seems like a lot but what are some other strategies for collecting this kind of feedback? Perhaps we could get a small but committed group of BHL users that would be willing to share this kind of feedback with us and put them all on a listserv or something? It would be great to hear from Rod Page, Siobhan Leachman, and other regular power users about what they'd like to see in a more iterative, informal and regular fashion.
- I am curious about the actual distribution of user types (taxonomists vs conservation scientists vs collections managers vs artists). I don't know if there is an easy way to get this information. I think it would also be useful to know more about (1) how BHL users are using the BHL APIs (2) how BHL users would like to use APIs with our content.
- Image use, taxonomic linkages, specimen links
- Yes what are their annotation needs? We only had annotation software in the portal for a short time so would be nice to do some more testing with users to gather some use cases
- Not sure

Q8 - Is there anything else you feel would be helpful for me to know about BHL, its partners, or its users as I gather feedback?

[4 answered; 5 skipped]

- My impressions... there are two primary audiences for BHL (librarians and taxonomists), and their needs do not always overlap. Also, at times I feel that BHL staff pushes for new features in BHL that benefit primarily BHL staff, and not BHL users (I am not 100% that this is true, but I get that impression).
- BHL Partners and users are a very diverse group of institutions and individuals. They come to BHL with a variety of interests and different kinds of and levels of expertise. The work of building an understanding of all of the myriad, important perspectives is unquestionably a great challenge and of great value. Thank you so much for your work on this.
- It would be helpful to be sure the CETAF group is included in user groups that are asked for feedback along with TDWG members, Atlas of Living Australia, GBIF and others.
- Not sure

Appendix C - Informal Interviews

This section includes a table listing the informal interviews conducted by the resident, as well as a brief summary of each interview.

Informal Interview Table

This table provides a listing of the informal interviews conducted by the resident including the meeting date, name of interviewee, their title and organization, and the purpose of the meeting.

Date	Name	Title	Organization	Purpose of Meeting
2/8/17	Beth Stern	Enterprise Architect	Smithsonian Institution	Overview of Smithsonian Technical Review Board process and format (SI)
2/13/17	Bob Corrigan	Director of Operations	Encyclopedia of Life	Overview of EOL and its technical development initiatives and how EOL interacts with BHL (SI)
2/15/17	Gerald “Stinger” Guala and ITIS staff	Director of ITIS	Integrated Taxonomic Information System	Attend ITIS staff meeting to learn about the work of ITIS and how these power users interact with BHL in their daily jobs (SI)
2/17/17	Dmitry Schigel	Scientific Officer	Global Biodiversity Information Facility	Overview of GBIF, how GBIF partners with BHL, and the work that’s been done on geographic name finding
3/21/17	Rod Page	Founder	Biostor	Learning how this power user interacts with BHL and how he uses it for Biostor
4/17/17	Keri Thompson	Head, Digital Library	Smithsonian Libraries	Learning background on BHL and digital libraries, as well as some BHL member and user issues
8/9/18	Richard Naples	Technical Information Specialist	Smithsonian Libraries	Overview of online surveys and presenting data from surveys with SurveyMonkey and Excel
8/30/17	Jacqueline Chapman	Digital Collections Librarian	Smithsonian Libraries	Learning some common BHL user comments and issues and digitization workflow process

Informal Interview Summaries

This section includes a brief summary of the informal Interviews with the Smithsonian staff and the representatives from biodiversity organizations.

Smithsonian Staff

Beth Stern, Enterprise Architect, SI

Originally, the resident was going to assist in the creation of a technical requirements document that met the Smithsonian's Technical Review Board (TRB) policies and procedures. The resident met with Beth Stern, an Enterprise Architect with SI, to learn about the TRB and the documentation required for submission to the TRB. As the year progressed, it was determined that BHL would not be at the point where it would be ready to start working on TRB documentation for its next version during the course of this project.

Keri Thompson, Head, Digital Library Department, SIL
https://library.si.edu/staff/keri-thompson
As Keri Thompson works with the Smithsonian Institution Libraries (SIL) digital libraries, she provided some insight into overall trends in digital libraries and user research. In the past, Keri has worked more directly with BHL's Technical Advisory Group and provided some background on specific issues with BHL as well and some thoughts on the future of research for BHL users.

Richard Naples, Technical Information Specialist, SIL
https://library.si.edu/staff/richard-naples
Richard Naples serves as the resident expert for online surveys for SIL and shared best practices in conducting and analyzing surveys, as well as presenting findings. Also, Richard was able to assist with some Excel analysis questions and chart creation. Richard additionally met with the resident and mentor to further discuss presenting survey data and manipulating data in Excel.

Jacqueline Chapman, Digital Collections Librarian, SIL
https://library.si.edu/staff/jacqueline-chapman
Jackie Chapman handles the digitization workflow for Smithsonian Libraries and interacts with BHL records and users often in this role. Previously, Jackie assisted with the triage of Gemini (BHL's internal tracking system) requests becoming very familiar with user issues. With this background, Jackie was able to provide some suggestions for BHL going forward dealing with the UI and search, as well as insight into how users interact with the BHL site and content.

Biodiversity Organizations

Encyclopedia of Life
http://eol.org/
The resident met with Bob Corrigan, Director of Operations for Encyclopedia of Life (EOL) to learn about the partnership between EOL and BHL. Corrigan provided an overview of EOL's technical development goals and ways that EOL and BHL could work together. EOL moved to a new platform this year. BHL has the authoritative, trusted information, but needs to be more accessible. EOL expressed a desire to help identify and extract elements of the relevant literature such as the original species description from BHL and emphasized the importance working from a common taxonomic backbone to aid in interoperability and compatibility of data across systems. In addition to this meeting, the Smithsonian resident met with Katja Schulz of EOL to discuss more about extracting the relevant data from BHL with their new text tool. Also, several of the residents met with Bob Corrigan and Jen Hammock, also of EOL, to discuss collections analysis and the potential for EOL to help show the dark spots in the tree of life in BHL collections.

Global Biodiversity Information Facility
https://www.gbif.org
The resident met with Dmitry Schigel, Scientific Officer for the Global Biodiversity Information Facility (GBIF) via Skype. This served as a good introduction to the work of GBIF and the biodiversity field. Schigel also gave some background on a project that BHL and GBIF have been considering in the past about a geographic name finding tool. Schigel shared some work GBIF has done on user surveys as well.

Biostor
http://biostor.org
Founder of Biostor, Rod Page, is a BHL power user and aids BHL in the discovery of articles in the serial publications in their system. The Smithsonian resident along with some of the other residents met with Page via Skype to learn more about how he interacts with BHL and what he would like to be able to do with the system. He shared a little bit about the data mining and visualization work that he does, but he reiterated getting to the most relevant literature is the most important aspect.

Integrated Taxonomic Information System
https://www.itis.gov
The resident along with other members of the BHL Secretariat were invited to a staff meeting of the Integrated Taxonomic Information System (ITIS). This is another group of BHL power users, taxonomists who use BHL in their daily workflow to cite and identify species. Each member of the staff introduced themselves, their work, and how they interact with BHL. It also provided some time for the BHL Secretariat and ITIS staff to share information and ask questions of each

other. This meeting was very informative for the resident in seeing examples of BHL use cases, as well learning the strengths and weaknesses of the site from the user perspective.

Appendix D - Individual Survey

This appendix includes descriptions of the categories determined from the free text response questions in the Individual User Survey.

I1 - Informed Consent

Purpose: The purpose of this survey is to gather user feedback and suggestions for improvements and enhancements to the BHL website. Data from this survey will be used to identify and prioritize improvements and features in the next version of the website.

Duration: This survey is expected to take about 15 minutes.

Risks: There are no known risks for participating in this study.

Confidentiality: We will use the data you provide, along with the data provided by other study participants, to inform the design of improvements and enhancements to the BHL website. Results and conclusions may be included in publications. To ensure confidentiality, we will not associate your IP address with your data. Data collected from this survey will be stored in a secure location.

Benefits: Your input will help inform the design and development of new features and functionality for the BHL website. While there are no direct benefits for individual survey respondents, survey results may also be included in publications on designing digital libraries and biodiversity information needs.

Freedom to Withdraw: You may withdraw from any or all of this activity at any time without penalty.

I hereby certify that I am at least 18 years of age and I agree with these terms.

- Agree
- Disagree

I6 - What if anything do you find challenging or frustrating about using the BHL website?

Description of Categories:

Search

- Saving search results for additional searches

- Needing to search in other languages
- Metadata
    - More metadata and more metadata requirements to find items with similar titles/date ranges
    - Incomplete, incorrect, or inconsistent metadata / cataloging
    - Missing metadata in book viewer - just says text or plate without page number, difficult to look through
- Fuzzy match
- Taxonomic synonymy
- Full list of species to search
- Search by geographic name
- Titles - abbreviations of names when titles too long (need more of and/or need to make it clear)
- Advanced search inadequate (lack of date range, way to refine search, clear display of results)
- Full text search / search within a document
    - Like using internet archive for the full text search
- Topic / subject search difficult
- Sorting search results
    - E.g. getting the seed catalogs out of search results
    - Clicking too many times to get to what you want
- Having to scroll through so many pages during a search and no number of results
- Browse
    - Species search time-consuming and difficult to view
    - Name browse messy
    - Authors always start at “A,” would like to be able to choose letter
    - Search by most recently added to see what’s new since last search (added by BHL date)
- Searching just illustrations
- Serial publications in same run not listed together
- Adding a cover image in the search results
- Finding relevant section within a large book
- Lack of pagination in old books
- Being able to search zoology vs. botany
- Complete search elsewhere to get info to be able to find work in BHL
- *Finding articles within journals - so many pages
    - Search results show journals and articles, if it could display in more intuitive way

Download

- Smaller file size, low-res images
- More streamlined way to select pages (esp for larger works)
- Understanding how to select individual pages
- Would like to download without having to go through email
- Scan quality

- Printing without downloading
- Would like separate files for each article in a journal or each chapter in a book
    - Difficulty navigating to article in a journal
- Difficulty with JPEG format
- Hi-res image downloads

Images

- More images
- Finding books with illustrations
- Some pages labeled text when they should be plates
- Search just illustrations
- Lack of Hi-res images
- Page dimensions

Content

- More content
- Missing Content
    - In copyright content
    - *Moving wall content - volume/issue missing
    - Rare literature
    - 1700-1850
    - pre-Darwin publications on birds
    - 40s, 50s, 60s - eastern countries
    - Recent research
    - Foreign language works
    - 1870's sparse
    - Italy, Argentina, Germany, Latin American countries
    - Applied science and useful arts from 19th C
- Not finding what user wants
- Duplicates of certain entries without clear explanation, confusion with different institutions scanning same thing, mulitple copies
- Overly restrictive copyright restrictions ("switch to CC-BY like your partners")
- More content in tables that can be put in CSV
- Find content in Internet Archive that's not in BHL

Usability

- Navigating website
- Finding and using download options and bibliographic information
- Not understanding where to start / how to use as a first time user
- Finding help section
- Difficulty understanding help section
- Understanding everything BHL offers
- Suggestion to identify top FAQ and pro-tips and share often with BHL community to learn how to use BHL better (faster, easier, unique features)
- More streamlined (Too many steps do things)
- User with older software can't view certain things on BHL (adobe)

- Not understanding connection to Internet Archive
- Understanding citations
- Accessibility when Internet Archive is down
- Printing without downloading / printing from the website
- Font on website too small
- Prefer stricter and more sleek design
- Responsive design for different screen sizes
    - Mobile phone viewing
- Slow
    - Speed loading pages / file size
- Navigating search results
    - No way to filter, too many pages to look through, confusing and time-consuming bibliography to view on species search, no number of search results
    - Understanding journal volumes and finding them and viewing them, display search results for journals and articles in more intuitive way
    - Scrolling through results (arrows stop working)
    - Having to click too many times
    - Keeping seed catalogs separate from scientific literature
- Difficulty with UI
    - Hints when you mouse over things
- UI - image resolution
- *UI - Book viewer
    - looking through a book
    - Flexibility to accommodate different formats (not all left to right orientation)
    - Scroll instead of flip in two page view
    - Easier to view in Internet Archive (to see full OCR and search in text while viewing)
    - Slow browsing looking at pdf
    - Missing page number when viewing text / plates
    - Ability to preview thumbnails
    - Slow and cumbersome
    - Positioning pages
    - Being able to zoom
    - Copying and pasting from pages
    - Thumbnails - more options
    - Show full page button / full page viewing
    - Difficulty navigating to article in journal
    - Being able to collapse left hand and top areas to create larger viewing space

Named entities

- *Author name

    - Name authority (difficulty to search and find what user looking for)
- *Scientific name
    - Searching on synonyms
    - Not always accurate or picked up
    - Difficult to view search (long, time-consuming bibliography to go through)
    - Search taxons within publication
    - Link to TL-2 (Taxonomic Literature II for botanists)
    - Vernacular/common names
- Geographic search

OCR

- Only being able to view single page, not whole book
- Does not maintain page structure
- Sometimes OCR unavailable
- Errors in OCR
    - scientific names not getting picked up
    - From older literature
    - Foreign languages

Personalized

- User accounts
- Save/bookmark favorite items
- Save searches (results by page, save page)

Scan Quality

Annotation

- Add annotations about scientific names or correcting scientific names

Missing a DOI

Data/Text Mining

- Downloads do not work with programs like R to search text

Linking

- Linking to EOL
- Article search results from EOL difficult to sort through, no article title
- Unexact link from bhl to tropicos

Other

More information about member organization that contributed work

Superseded by World Spider Catalog

API - externally held content not API accessible

Transcription/Crowdsourced

- Correct transcriptions while working with article and OCR

I9 - (Follow up to Q8) If yes, please describe.

Description of Categories:

- Annotations

    - Bring back way to add comments
    - A way for users to flag or correct missed or incorrect scientific names
- API
    - Making users more aware of it and how to use it
- Articles
    - Article indexing / improve article level access
- Citation
    - Easily get citation in bibliographic manager
    - Improve Zotero functionality
- Content
    - More content, more coverage
    - Old journal of Latin America
    - Missing volumes of 19th C works
    - Moving wall volumes
    - More taxonomic journals
    - Focus more on expanding coverage of biodiversity literature (not seed catalogs and field notes)
    - Include journals of applied science and the useful arts
    - Increase access to out of print foreign journals (e.g. publications from the Estazione) and more old German taxonomic pubs
    - Increase non-english journals
    - Work to get more permissions for copyrighted material
    - More paleontological literature
    - SE Asian plants
- Data
    - Better way to load and search within in text file
    - Improve quality of XML
    - Improve ability to read stored data
    - Persistent GUIDs for data
    - List of journals covered
    - List of species names sorted and mapped to higher taxa
    - Provide modular code widgets
- Documentation
    - Improve documentation (and locating documentation)
    - Provide more and better tutorials, outreach, online seminars, background information on scope and breadth of collection and reasons for gaps, long term goals and strategic plan
    - Especially improving outreach/documentation/interface for API
    - More instructions and recommendations on most efficient ways to search for different kind of information
- DOI
    - Add more
- Download

    - Increase the quality of the pdfs
    - More options
    - Explaining options / making UI better
    - Provide djvu for download
- Feedback
    - A rapid feedback process where scan requests are acknowledged, processed, and notified
- General Info
    - Featuring content, learning about content in other ways
- Illustrations
    - Search specifically for illustrations
    - Provide google image search - upload an images and find similar ones in the scanned books
    - Higher quality
- Images
    - Higher quality images
- Journal Articles
    - Easier to search
    - Easier to identify authors
- Languages
    - In OCR options
- Linking
    - Links to places to find very old scientific publications that aren't in BHL
    - Connect to other databases like WoRMS, EOL, and Wikipedia
    - Bring type specimens together like in Jstor
    - Link species to recent research publications
    - Link to more sources like Gallica
- Metadata
    - Pagination for all text and plates
    - Missing tags
- Named Entities
    - Authority Control - authors
    - Improve scientific names
        - Scientific names for archival content
        - A way for users to flag or correct missed or incorrect scientific names
        - Search by common name as well
    - Add ability to search by geographic names
- OCR
    - Better ocr
    - More languages
    - Better at other scripts (Fraktur/Gothic; cyrillic, etc)
- Original Description
    - Some way to machine sort through content to find original description

    - Identify pages with protologue for given species
- Personalized features
    - Alerts for material added on specific topic, taxa or geography, author, etc.
- Related Titles
- Research Tools
- Scan Quality
    - Better quality
    - Especially illustrations
- Search
    - Filter
        - by year range, date of publication
        - by categories such as maps, catalog seeds, etc
        - By author
        - By common name
        - Recent additions
    - Full text Search
    - Fuzzy matches/beings with/contains
    - For illustrations
    - Better way to load and search within in text file
    - Bring search interface and resulting sorting options in line with other libraries
    - Improve advanced search capabilities
    - Make browsing by subject easier
    - Make it as easy to find as Jstor
    - Add geographic searching
    - See all works by an author
    - Browse by species
    - Search by specimen number
    - Suggest similar searches and other works
    - Search for synonyms of scientific names
- Transcription
    - Allow transcription to enable scientific names in archival content
    - Allow users to correct OCR
- Usability
    - Confusion with article download
    - Improvements to UI to due to many requests for features that already exist such as documentation on different tools, especially the API, a help section, a place to make a scan request
    - Frustrating or confusing UI
    - Provide information on abbreviations of publications
    - Enable zoom with the mouse
    - Ability to copy text from pages
    - Embedded mulitlanguage translator
- Other

    - Scanning parts wrappers
    - Compile index and contents pages
    - Merge with BHL Europe
- None
    - User said there was nothing more BHL could do
- Praise
    - Positive feedback, appreciation for BHL
- Remove
    - Irrelevant or unusable comment, usually due to not being descriptive enough (for example, “separate files”)

I22 - Do you have any additional feedback you would like to share about your experience using the Biodiversity Heritage Library or comments about new features?

Description of Categories:

- Praise
    - Overall praise
    - Love access to content / content best feature
        - Opens up news areas of research
        - Provides access to content previously inaccessible to most
        - Making sources available that aren’t available anywhere in user's’ country
        - Convenience for day to day work
    - Love being able to download pds of articles and select pages
    - Extremely helpful to be able to access original descriptions quickly and easily through BHL
    - Good API for data mining
    - Enjoy images and updates shared on facebook and instagram
    - Enjoy logging Flickr image tags
    - Great way for researchers to work across geographical and other divides
    - Images from BHL so helpful for lecture presentations
    - Like the wide spectrum of fields
    - Great for filling interlibrary loan requests
    - Makes assisting researchers so much better
    - Enjoy twitter feed
    - “When I report errors they are fixed immediately!”
    - Grateful for the bibliographic services
- Remove
- None

- Annotations
    - “Increases search engine ranking for queries of taxon names”
- Articles
    - Metadata quality for articles
- Citation
    - Make bibliographic of complete works available in case only in complete copies in BHL
- Content
    - More content
        - More in-copyright content
        - Uneven - missing russian journals
        - More archival correspondence and records
        - More Japanese and other asiatic sources
        - More works with illustrations
        - 1920’s-1940’s, Japan and GermanyP
    - Missing journal wrappers which are needed for date of publication
    - Content of BHL is most important feature
    - Grow to size of Hathitrust but remain free and open
- Data
    - “Please make your text parsing algorithms more accessible for personal research use.”
- Documentation-Outreach
    - More instructions
    - Promote Flickr especially the ability to see taxon synonyms (see comment below)
    - “Please do try and tweet about exploratory accounts of various scientists' diaries/journals in your collection. Such journals are great to read during field work.”
- Download
    - Better options
    - Greater availability of unprocessed images for download
- Features
    - “Need major step forwards”
    - Don’t lose what you currently offer by improving or adding new features
- Feedback
    - Would like to see user request / ticket system made public
    - More transparency about what is being digitized and when, providing a new content newsletter
    - On survey results and long-term plans for BHL
- Illustrations
    - More illustrations
- Images
    - Sharper image resolution in scans of text and illustrations

- Linking
    - Links to other repositories that have works not in BHL
    - More links to other sites and repositories
        - Animalbase
        - Zodobat
        - WoRMS
- Metadata
    - Quality of metadata (author names, misspellings)
- Named Entities
    - Author names, different spellings
    - Taxon names tagged on Flickr (see comment below)
- OCR
    - Better options
- Personalized
    - Create personalized publication and illustration libraries from BHL content (see comment below)
    - Bookmarks
    - Book “shelves” for different topics
    - Options to follow other people doing similar research
- Search
    - Fuzzy search to account for OCR quality
    - Add sort to see alpha order, type of publication, etc, also how many you would like to see and most recent or particular day/date added
- Usability
    - Simplify UI
    - Not knowing how to submit scan request
    - 24 hour access is the key need
    - Not set up for other types of users (ie. historians)
    - Visible feedback link on main page

Appendix E - Consortium Survey

This appendix includes descriptions of the categories determined from the free text response questions in the Consortium User Survey. Some include a list of all responses due to the number of responses and variation within the responses not warranting a list of category descriptions. Each question is noted if it is a description of categories or a list of all responses.

C1 - Informed Consent

Purpose: The purpose of this survey is to gather user feedback and suggestions for improvements and enhancements to the BHL website. Data from this survey will be used to identify and prioritize improvements and features in the next version of the website.

Duration: This survey is expected to take about 15 minutes.

Risks: There are no known risks for participating in this study.

Confidentiality: We will use the data you provide, along with the data provided by other study participants, to inform the design of improvements and enhancements to the BHL website. Results and conclusions may be included in publications. To ensure confidentiality, we will not collect your IP address with your data. Data collected from this survey will be stored in a secure location.

Benefits: Your input will help inform the design and development of new features and functionality for the BHL website. While there are no direct benefits for individual survey respondents, survey results may improve the overall usability and usefulness of BHL for both contributing libraries and their users. Survey results may also be included in publications on designing digital libraries and biodiversity information needs.

Freedom to Withdraw: You may withdraw from any or all of this survey at any time without penalty.

I hereby certify that I am at least 18 years of age and I agree with these terms.

- Agree
- Disagree

C3 - (Follow up to Q2) Please provide any additional details on what you would like to see come out of implementing and providing the services listed above, or if there is another tool or feature that you feel should be a priority for BHL going forward.

Description of Categories:

3rd party
- Very big task, may not realize true value for several years
- Not enough tech support for all of goals so reaching out to 3rd parties could be way to resolve this

Articles
- Splitting journal issues into articles
- Increase volume of articles indexation

Born-digital
- Born-digital is the future and should be added to the list to make sure it is a priority for incorporating born-digital content

Content
- Seek best copy (to display wrappers, annotations, inscriptions, etc) and complete sets for both monographs and journals rather than partial/incomplete
- Seek and include bibliographical variants (during hand-pressed period variation between copies is not uncommon)
- Building new content should still be highest priority (in copyright, out of copyright, orphan works, and content at high risk for loss)
- More text content that can be run through OCR
- Help small institutions digitize their content

Contributor
- Clear, resolvable links back to contributing library's catalog

DOI
- Register DOIs for content

Download
- Availability of high-quality downloads
- More streamlined way to get high resolution images

Illustrations
- Critical and have scientific value in their own right

Linking
- Connect to GBIF and other data research repositories
- Linked data and semantic search
- Links to curated additional content (such as finding aids or volume lists for serial monographs)

Literature
- Keep focus on biodiversity literature as it is BHL's mission and strength and have all interface and infrastructure changes support this

Name authority
- Clean-up of author names should be a priority

Named entities
- Disambiguating authors
- Searching on location information

OCR
- Engage community in correcting OCR

Search
- Include bibliographic elements such as wrappers, annotations, inscriptions
- Full-text search should be a number one priority
- Facets (ability to search within a search)
- Increase discoverability of collections which is currently an issue for BHL
- Search and browse for visual resources to see range of content
- Add copy-specific metadata that could be searched on
- Semantic search
- Sort items by date
- Specialized subject headings list
- Improve search-ability and search rank
- De-emphasize Browse, re-emphasize and restructure Advanced Search

Standards
- Implement KBART standard to aid in outreach to library community
- Ensuring BHL content can be found by open access browser extensions like Unpaywall

Transcription
- Display transcription in the existing OCR field
- Build community around transcription
- Uploading of transcriptions should be a priority

Usability
- More user-friendly interface
- See tracking of downloads
- More streamlined and user-friendly download options
- Make scientific name search more obvious on the site
- BHL and IA records sync (edit BHL and IA is also updated)

User features
- Way for users to enhance content (correct OCR, add tags, etc)
- Annotations
- Transcriptions, boards - so BHL users feel more invested

Visual Resources
- Add visual resources that are scientifically important
- Ability to search and browse for visual resources (see range of content)

C4 - What do you see as the primary challenge to implementing these tools and services?

Description of Categories:

3rd party
- Cooperate with 3rd parties and major web stakeholders

Archival content
- Archival content should be expanded however metadata fields and interface viewer are book-centric

Articles
- Making articles in BHL content accessible

Content
- Making sure best copy is made available in BHL (and better copies not being rejected as duplicates)
- Quality of older materials
- Too much literature still missing (esp original descriptions), that should be priority before adding archival content
- Too much focus on visual resource discoverability vs. digitizing missing literature content

Creativity
- Think out of the box for long-term developments

Data model
- Marc based data model

DOI
- work/time/technical skill/cost involved in preparing metadata to register DOI

Format
- Ability to make articles accessible in XML format

IA
- Must eliminate risk involved with dependence on IA

Infrastructure
- Hard to get support for infrastructure changes and requires a lot of effort from staff already spread thin
- Current infrastructure not flexible for different formats

Metadata
- Inadequate catalog records to get needed information
- Resource demands for preparing metadata for registering DOIs

Named Entities
- Improving scientific names

OCR
- Improvements to OCR quality needed and not sure BHL has capacity to make those improvements

Organizational Structure

- Committee-driven decision making problematic
- Competing priorities and resources
- Executive committee micromanagement
- Prioritization of resources
- Lack of technical and organizational leadership

Resources

- Money
    - Hard to get funding for infrastructure changes
    - Chasing grant money that changes BHL priorities
    - Lack of technical and organizational leadership due to financial resource issue
- Time/effort
- Technology
- Staff
    - Limited developer resources and ack of structured process for building out requirements before development begins
    - More developers to split work between front and back end to speed up process
    - Lack of strong technical lead
    - Lack of front-line staff input on technical priorities
    - Project management
- Expertise
    - Lack of bibliographical expertise to seek, recognize, include variant copies, separate states/issues/printings.
    - Technical skill
    - Need UI/UX experts

Search

- Full-text search

Training

- Staff need more training

Transcriptions

- Many different transcription tools, how will BHL work with different formats

Upload

- The uploading process

Usability

- Expand interface options (bookviewer) to work better with archival content

C5 - (Follow up to Q4) How do you see BHL overcoming this challenge?

Description of Categories:

3rd party
- “If we are able to open up development to third parties we would likely have a faster development cycle”

Data Model
- “Overhaul” of data model

DOI
- Assign more DOIs

Egypt
- Good option to expand resources by having Egypt work on technical development

Equipment
- Higher resolution cameras for scanning

Fundraising
- Apply to more grants
- New and different revenue sources
- Establish an endowment
- Out of the box funding (“Donations via Go Fund Me, etc.; Endorsements/Advertising; Memberships (like PBS etc.) with swag for certain levels)”
- Funding for development “separate of maintenance and administrative costs”

Membership
- Try variety of membership categories with adjusted fees
- Increase membership numbers

NA

Organizational
- “Doing one thing well at a time instead of trying to address multiple issues at once”
- Have a few people make decision “Too many people make for a lot of time wasted in details”
- “Follow a strict software development planning process”
- Share more details with members about development goals, progress over time and expected vs. real timeline
- “Clear objectives and timeline and a professional approach (rather than amateur academic attempts) followed thru”

Outreach
- Include BHL in college course requirements (library/archival, biodiversity university courses)
- Improve way of retrieving information from researchers
- Work on marketing and elevator speeches for garnering continued support
- Advocate with web stakeholders (not necessarily natural science field)

Praise
- Great track record in getting funding
- Tech team works well together, glad to see tech team addressing Gemini issues
- Seen BHL operations improve over time

Staff
- Priority to staff training

Standards
- Set transcription standards
- Set cataloging and imaging standards for contributing libraries

Transcription
- Choose a standard and have others adhere (as with scans of digitized material)

User Needs
- Surveys like this one are a good first step, need to rank and prioritize tools and surveys and focus on tops ones first

C7 - Do you have any additional comments related to particular strengths or gaps in BHL's content coverage?

Description of Categories:

Current
- More current content

Data
- Discovery/extraction of traits

Geographic
- More content from Asian countries
- Adding geographic and searchable tags to metadata

Hathitrust
- Get items from Hathitrust in BHL

In-copyright
- More staff to pursue in-copyright content

Journals
- Fill in journal runs

Named Entities
- Authority control

NA

None

Organizational
- Does BHL know it’s strengths and weaknesses? Are those shared with members?

Personalized Features
- Annotations

Praise
- Strength - content that is widely used by research community

Scope

- Defining BHL's scope clearly
- Making collection description with strengths and gaps available

Types

- Collect other types of materials - maps, archival, and images

C8 - If you have come across features on other digital library sites that you would like to see on BHL, please describe the feature (links appreciated).

Description of Categories:

Annotations

- Let users annotate on the site even if need to be moderated at first to create community
- Annotations like Europeana IIIF image delivery

OCR

- OCR inline with text and users able to make corrections - National Library of Australia Trove newspaper correction site (http://trove.nla.gov.au/newspaper/)

Outreach

- Online exhibition - https://dp.la/exhibitions http://spices.biodiversityexhibition.com/

Personalized Features

- User accounts - create customized layout, save queries, etc - "MyBHL"
- Create communities

Praise

Search

- Full-text search - California Digital Newspaper project (https://cdnc.ucr.edu/cgi-bin/cdnc)
- Search images by keyword with tagging (https://images.is.ed.ac.uk/luna/servlet/view/all?showAll=what&sort=work_title%2Cwork_creator_name%2Cwork_shelfmark)
- Improved browsing of BHL subcollections such as (http://library.columbia.edu/find/digital-collections.html)
- Don't stop at full-text search, go semantic and enable in UI, create communities
- Subject class search for biodiversity (https://www.wdl.org/es/topic/)
- Thumbnails and faceted search in results - Gallica (http://gallica.bnf.fr/services/engine/search/sru?operation=searchRetrieve&version=1.2&query=%28gallica%20all%20%22chat%22%29&suggest=0)
- Visualizations of search and browse - UK Medical Heritage Library (https://ukmhl.historicaltexts.jisc.ac.uk/home)

Visual Resources

- Map display - David Rumsey Historical Maps collection (https://www.davidrumsey.com)

- Search Images by Keyword with tagging (https://images.is.ed.ac.uk/luna/servlet/view/all?showAll=what&sort=work_title%2Cwork_creator_name%2Cwork_shelfmark)

Transcription

- Adding transcriptions to BHL
- Page by page display of original and transcription so that you can scroll through both at same time like - Smithsonian Transcription Center

User Interface (UI)

- Improved book viewer like - Wellcome library book viewer
- Turning pages

C9 - What has been your experience adding content to BHL via Macaw and Admin Dash? Please share any likes/dislikes, difficulties you may have encountered, or areas of improvement you would recommend.

Description of Categories:

Details on Areas of Improvement:

Articles

- Ability to add articles with Macaw
- Article/chapter level markup for Macaw would be *extremely* good (I can't emphasize this enough).

Batch

- Ability to do batch edits in admin dash

Inconsistency

- Macaw can be a rather capricious program - when it works, it works very well but it can hit a hurdle and just fail to work. This has been our biggest problem with it.
- Uploading pages to Macaw is sometimes spotty, and a couple pages will usually encounter an error and not be uploaded.

Pagination

- Improve page characterisation process to make it faster.
- On the Admin Dash when paginating or defining articles, please allow the user to start by entering the item identifier or the title identifier.
- I would really like to see the pagination interface in the admin dashboard be improved to match that of Macaw's, where you can view the pages via thumbnails to quickly select images.

- The admin dash for pagination is so clunky it gets really annoying and makes the process take longer than it needs to.

Upload time

- larger files tend to take much longer to upload
- Slow to upload tif files

Pros/Cons:

Macaw Pro

- ergonomy. The guide is clear. Great team, very quick answers.
- Adding content to Macaw has been pretty easy
- Easy to use
- Excellent tool
- Very friendly
- Very good experience
- The ability to add metadata and paginate via spreadsheets rather than through the UI is fantastic. Allows me to take advantage of some of the power of Excel to do repetitive tasks.

Macaw Con

- Marc format conversion (form Unimarc to USMarc) is challenging for us. Macaw could have a function of automatic save. Improve page characterisation process to make it faster.
- larger files tend to take much longer to upload
- I'd like to be able to upload articles via Macaw
- It would be nice to be able to search Macaw by other than just standard number.
- Having learnt how to scan and digitize documents and other archived material it is still difficult for me to upload the scanned material into BHL.
- Article/chapter level markup for Macaw would be *extremely* good (I can't emphasize this enough).
- The magnifying glass over the preview pane at right would be easier to use if it appeared where one clicks on the image, instead always at the center and then having to be moved.
- It's great in Macaw. But that darn button still won't close at file upload without refreshing the page. There is a gemini ticket for it. I wish there were a couple more page types available. There are gemini requests for that too.
- Macaw can be a rather capricious program - when it works, it works very well but it can hit a hurdle and just fail to work. This has been our biggest problem with it.
- Macaw has a learning curve but works once one knows the process.
- Slow to upload tif files
- Some bugs when uploading PDFs although the staff have been helpful in trying to solve these errors

- Uploading pages to Macaw is sometimes spotty, and a couple pages will usually encounter an error and not be uploaded. It would be great if there was a "retry" option for pages that encounter an error directly on the upload page. Currently the only way to load these error pages is to write down the un-uploaded file numbers, accept the upload as is with missing pages, and then go back in and insert the missing pages from the upload page again. Otherwise, adding page metadata in Macaw is fairly smooth and straightforward.
- Would like to see the ability to create segments via spreadsheets at the time of uploading.

Admin Dash Pro

- easy to work with
- Extremely useful

Admin Dash Con

- Difficulties to search and find a particular document (the title research may be confusing so the item codes are needed). Interesting to have the possibility to access a proper dashboard for each institution.
- could do with another 'save' button near the top of the record. Drop down lists for copyright fields.
- a bit clunky for reordering volumes.
- i'd like to see improved site layout design for Admin Dash - batch editing would be great (but potentially dangerous I suppose); is there ANY way we can have these tools help us manage page inserts somehow? It would be SUPER to have a way to create MARC records via these tools to help with uploading content.
- On the Admin Dash when paginating or defining articles, please allow the user to start by entering the item identifier or the title identifier. Currently, the user has to enter the title identifier and then drill down to the item identifier. This is a real time waster.
- I would really like to see the pagination interface in the admin dashboard be improved to match that of Macaw's, where you can view the pages via thumbnails to quickly select images.
- I'd like the ability to make easier batch edits in the admin dash. For instance, currently I cannot add a bunch of separate items under the same title to the same collection.
- The admin dash for pagination is so clunky it gets really annoying and makes the process take longer than it needs to.
- The Admin dash is good but there are a few functions that are really clunky and manual.
- Admin Dash is complicated for one who doesn't use it on a regular basis.

General Pro

- BHL has been very responsive about making requested additions/improvements.
- Both essential tools
- Both very good
- Both easy to work with

- Both friendly interfaces
- Very few issues and any problem is solved quickly with no inconvenience

General Con

- Go beyond the admin/user modes. Create Access Control Lists that allow some users to view and access more functionality and do some admin / citizen science work in the Portal or in the satellite input programs…
- I would like to be able to have items contributed by more than one contributor automatically feed into a collection (at the moment this is only possible for a single contributor).
- I am constantly tripped up by differences between pagination in Macaw and pagination via the dashboard. I get specification of roman numeral page numbers wrong on a regular basis.
- It is a bit clunky and needs to be streamlined along with more cohesion between Macaw and Admin Dash.
- There are too many systems to learn and understand.

C10 - Do you have any additional comments or suggestions to share about technical development priorities for BHL?

Description of Categories:

Data Model

- Reconcile with EDM

Digital Preservation

- Establish digital preservation standards and include DP system as part of new BHL

Flexible

- More flexible portal

Internet Archive

- The only other things is, apart from our technical developments for BHL it would be good to know if there is any way to leverage similar improvements with IA which is upstream from all our efforts. This is really pie-in-the-sky but any kind of mirror hosting of BHL content found in IA is an important long-term concern for the permanence of the project.
- Move away from IA as storage platform

Named Entities

- Name authority control
- Authority control high priority

NA

None

Organizational

- Rush products out, too much time evaluating, approving, controlling... Just get prototypes up and running, quick, fast, dirty... the new development methodology you are about to implement will require it, don't prune its benefits by "customizing it to our own special BHL ways"... it's a software development endeavor, that's it! Get your requirements straight (as I believe you are doing here) staff it and fund it properly and move on with it... We've been waiting for a while now! If we are not going to experiment with skunk projects anymore, then here's where we will need to take the leap then.
- Tech Team needs a leader, someone to triage and set priorities. I am afraid to submit tech problems in case the team enacts them without thinking them through or talking to others who may have knowledge that I do not. I am also afraid of being dismissed due to group think and pre-made decisions.

Outreach

- Share detailed plan with members - “It would be helpful to be aware of key priorities already committed and a layout of new and different possibilities and how difficult or easy they may be to accomplish.”

Prioritize

- Improve existing features (like search and name authority control) before adding new features

Search

- Better searching of all kinds
- Improved search functionality
- Full-text search
- Browse by subject
- Faceted search

User Interface

- “One minor point - the name of the contributing library can overlay the text depending on window size, length of the name of the contributor, and which view is chosen. Would be nice if that could shift somewhere else or be a bit more mobile. Still think it is important but a pain when QAing. “

User Needs

- Ask for feedback from members and users more often and make updates to site more often

Visual Resources

- Better integration of images into portal
- Priority to incorporate visual resources

Appendix F - System User Survey

This appendix includes descriptions of the categories determined from the free text response questions in the System User Survey. Some include a list of all responses due to the number of responses and variation within the responses not warranting a list of category descriptions. Each question is noted if it is a description of categories or a list of all responses.

S1 - Informed Consent

Purpose: The purpose of this survey is to gather user feedback and suggestions for improvements and enhancements to the BHL website. Data from this survey will be used to identify and prioritize improvements and features in the next version of the website.

Duration: This survey is expected to take about 15 minutes.

Risks: There are no known risks for participating in this study.

Confidentiality: We will use the data you provide, along with the data provided by other study participants, to inform the design of improvements and enhancements to the BHL website. Results and conclusions may be included in publications. To ensure confidentiality, we will not collect your IP address with your data. Data collected from this survey will be stored in a secure location.

Benefits: Your input will help inform the design and development of new features and functionality for the BHL website. While there are no direct benefits for individual survey respondents, survey results may inform improvements to the website (biodiversitylibrary.org) that may benefit BHL users in general. Survey results may also be included in publications on designing digital libraries and biodiversity information needs.

Freedom to Withdraw: You may withdraw from any or all of this activity at any time without penalty.

- Agree
- Disagree

S4 - How would you like to see BHL data services be more interoperable with other data services and systems?

Description of categories:

Linking

- Align taxonomy with CoL, GBIF, EOL, etc.
- Allow for data linking using Open Annotation standards and integration with other data aggregation services
- Link scientific names to BHL resources and other data sources
- Linked data
- Links to specimens and sequences
- Link references and authors to unique identifiers

Identifier

- Include identifiers for key text annotations such as GBIF taxon IDs, URIs from OBO Foundry for biological and habitat terms, Geonames, wikidata, PlantList IDs
- Cross link references and authors to unique identifiers

Scientific

- Discovering scientific names in BHL is a largely manual process and would need to be improved to be more interoperable
- Linking scientific name / aligning taxonomy with others sources (eg CoL, GBIF, EOL)
- Adding scientific name synonyms

Standards

- API implementation standards (data value normalization, data structure, metadata of APIs)
- Common standards for field (TDWG DwC and Dublin Core)
- Controlled vocabularies for values

API

- API implementation standards (data value normalization, data structure, metadata of APIs)
- Swagger/OpenAPI support

Geographic

- Geographic data

License

- a lot of BHL content can't be used on Wikipedia until non-commercial licenses (such as CC-BY-NC), so filtering for guaranteed public-domain content (say, publications before 1850, or those explicitly licensed CC-BY or CC-BY-SA)

Data

- Harvesting data from text

Praise

- I think the BHL is doing a LOT of things right already: widely reusable content (including fully reusable content in its vast public domain holdings), permanent URLs to books, articles and pages, and easy to use APIs.

Format

- Allowing results in JSON format

S5 - What do you see as the primary challenge to increasing interoperability with BHL?

Description of categories:

Search
- Better search interface for bibliographic linkages
- Improved search capabilities
- Fuller results for search terms
- For those that don't use an API, "automagic" search functionality such as Google Scholar cooperating with BHL

API
- Keep prioritizing the API, regardless of how much use it currently has it is key to long term success. Start building pluggable widgets that leverage the API and that can be injected straight into other applications (autocompletes, etc.). Think Google maps like widgets.
- Format of data received in API (Linked data definitions)
- Standards from the community for API
- Adding filter mechanism in API syntax (to reduce number of results)

Data
- BHL data is structured in a librarycentric way which can add challenges when wanting to view literature from other groupings (ex: a little cumbersome to get a list of article titles for an author about a specific taxon at a location)
- Data mining from text
- Data structure

OCR
- Improve OCR

Linking
- Improving linking on names
- Provide linked data definitions of fields returned by API

Named Entities
- De-duplication of authors and references
- Detecting valid scientific names in text

Standards
- Get consensus within API providers community for the standards

S6 - How do you see us working together to overcome this challenge?

List of all responses:

- I'd love to see BHL at more digital biodiversity workshops and hackathons. https://epandda.org/ is very interesting in collaborating with BHL to do specimen - publication matching
- Involving community (e.g. TDWG, GBIF, iDigBio)
- Maybe an NLP project focused on training on BHL text?
- maybe we need a new vision on how to integrate BHL and ITIS
- ongoing BHL effort to annotate text (automated or crowdsourced), early release of alpha products so communities of interest can poke them and provide feedback
- Probably requires work with library scientists
- Start with a nearby community and promote for a wider community little by little
- The International Plant Names Index springs to mind as a well-curated and extensive resource for plant name and author resolution / verification
- There is a lot of potential to integrate BHL into TaxonWorks, with the goal to produce exemplar "widgets" that are generically usable (i.e. not just in TaxonWorks). Let's do a sponsored hackathon?!
- Via the application of Text Mining techniques such as Named Entity Recognition to enrich and index BHL's contents
- We could try out APIs with defined examples from our work to see if APIs and filters do what we would like them to do
- Why not use Zooniverse to slice scanned journals up into articles.
- work with other organizations to agree on unique identifiers for authors, references and scientific names.

S8 - (Follow-up to Q7) Please provide any details on the above areas related to their usefulness and any particular strengths or gaps in BHL’s coverage.

Description of categories:

NA

- Several responses that said they weren’t familiar with BHL collection coverage

Copyright

- Problem area in-copyright content from 1920’s - 1950’s

Temporal
- 1920’s-1950’s in copyright
- 17th - early 19th C (only in German or European repositories)
- Higher priority to fill geographic/taxon pre-1920 first

Geographic
- 17th - early 19th C (only in German or European repositories)
- Fill geographic coverage pre-1920 first

Usability
- Understanding BHL’s relationship with IA

IA
- Understanding BHL’s relationship with IA

Metadata
- Difficult to get metadata from publication from scanned copy

Named Entities
- Variants on author name

Praise
- “Love the ability to get at old journals”

Taxon
- Fill taxon coverage pre-1920 first

S9 - Does the current offering of data through the BHL API meet your needs? Why or Why not?

List of all responses:

- Could be improved: What I need: More different possibilites for API's, especially working on a level of Genus (ex. search "Fagus**") and resulting species List, today I've to use large amount of single requests (ex. Fagus sylvatica, Fagus antarctica ...)
- For the quick experiments I ran, yes. I haven't spent enough time truly testing the system though.
- Generally yes, but not for my particular purpose. (Answered in Question 8).
- I am unsure how to answer bc I am unsure whether using the API is distinguished from using the website's interface. For example, I almost never download PDFs from BHL, but the provided direct addresses of specific pages within documents is essential to my work and something I use constantly
- I found it difficult to use. What I wanted out of it was the OCR text or failing that the JPG page-images (which I would then OCR myself) but I didn't manage to get it to work - I tried to feed it the info on the pages I wanted but it didn't seem to work.
- I have no detailed experience with the API, sorry.

- I haven't used the API much
- It would be useful to have cleaner means of gaining tracking data.
- No, I'd like to see a standard Boolean interface in order to do OR'ed searches for names. I'd also like to see faceted searching where I could provide additional filter to restrict the results output (by geography, language, publication date, etc.)
- not a user of the api
- Not used API
- Yes
- Yes
- yes
- Yes for the most part. Would love to see more OCR text more easily requested
- Yes, no need for more information.
- yes.
- Yes. It is only used occasionally, but if memory serves, it seems to be work well.

S10 - Does the available documentation on the BHL API meet your needs? Why or why not?

List of all responses:

- an interactive example form such as the one of http://eol.org/api/docs/pages could be informative
- Documentation is enough to get started for a developer. Some sort of non API key sandbox or more contextual examples with sample data return could help less experienced users understand endpoints and the associated parameters.
- Documentation seems adequate. Would be nice to have machine readable access to services (OpenAPI).
- Don't know
- Generally good, but good documentation is vital for data re-users. So it can be better. A user-friendly step-by-step guide for novice developers will be particularly welcome.
- I had no look into the documention yet.
- I would like stronger links to GBIF so I can dload articles related to specimen info published in GBIF. One missing link is the itineraries of the collectors - this could help link places to dates and vice versa - perhaps Zooniverse could help build these.
- It could be significantly improved. Try Raml. Provide a test framework that doesn't hit your production sites for validation. Move issues/tracking/documentation to Github so that developers can track more closely specific issues/evolution of the API.
- n/a
- not a user of the api
- Not used API

- The existing API's are described in good quality; it is important to have at least one (better two) practical example for every API-description, and than it's easy to understand.
- yes
- Yes
- Yes
- Yes it does. Have not encountered any difficulty implementing the API.

S11 - If you have come across features on other data service sites that you would like to see in BHL, please describe the feature (links appreciated).

List of all responses:

- 1) Online demo/testing page for APIs. 2) APIs which do not require API keys and manual registration.
- DOIs would also be nice to see in results from API calls (if not already implemented).
- Downloadable search results similar to GBIF https://www.gbif.org/occurrence/search
- I find the Search does not meet my needs, I end up having to browse towards my goal, through screen after screen. Don't understand why if I put a journal name in, is not found, but browsing eventually reveals that the journal has been scanned and is in BHL.
- I use two data services in my daily work, which I find very useful and well documented: 1) the GBIF taxonomy backbone: https://www.gbif.org/developer/species 2) the Tropicos API: http://services.tropicos.org/ And for our
- none come to mind
- today it's not possibel to see / to track changes, means new added works in BHL on a species-Level! for example today I make a query on Scientific name "Zea mays" and check the masses of results; I repeat the query two weeks later and the amount of results changed from 9200 to 9205 - how can I track/find the 5 new ones? To have a tool would be very, very useful.

S12 - Do you have any additional feedback to share about your interaction with BHL or any further recommendations on how BHL could make the content you or your users want to work with more easily accessible?

Description of categories:

API

- Provide access over secure service (https)
- More community engagement of developers and informaticians working on BHL, mailing list for users of the BHL API?
- Praise for API working well for one user's app
- Provide more API formats

OCR

- More accurate and improved OCR

Praise

- Stable URL is "perfect (short and simple)" for linking to pages
- "I just want to thank BHL, I would spend ages in the library answering questions if it weren't for the website. "

Scan

- Understanding how scans get into BHL and who can submit scans

Search

- Better search-ability
- Use a dedicated full text search (not just SQL database queries) such as SOLR and ElasticSearch
- Improved search capabilities

Upload

- "We found that the upload interface and metadata entry interface was tricky to use. We also encountered some difficulties updating already uploaded volumes."

Usability

- More friendly interface
- Something like FSI-viewer (better book viewer)
- Trouble with upload interface

Appendix G - Resources

This section encompasses some of the most useful resources consulted by the resident this year, those sources referenced in this report, and other resources that may be helpful to BHL going forward. It is divided into BHL resources, Biodiversity resources, Digital Library resources, and User Study resources.

BHL

Biodiversity Heritage Library. BHL Annual Member Meeting 2017. “Strategic Planning Session.” Accessed April 17, 2017. https://docs.google.com/spreadsheets/d/1pZUOF2so2p8VNyykvsIWCs5FW09Ll9nI74owjoHRi_M/edit#gid=0.

Biodiversity Heritage Library. BHL Blog, Notes & News from the BHL Staff. “BHL User Series.” http://blog.biodiversitylibrary.org/search/label/bhl%20users.

Biodiversity Heritage Library. “BHL Outreach & Communication Plan.” https://bhl.wikispaces.com/Communication+Plan.

Biodiversity Heritage Library. “BHL 2010 US Survey.” https://bhl.wikispaces.com/BHL-US+Survey.

Biodiversity Heritage Library. “BHL 2011 Usability Test.” https://bhl.wikispaces.com/2011+Usability+Test.

Biodiversity Heritage Library. “BHL Reports and Statistics.” https://bhl.wikispaces.com/BHL+Reports+and+Statistics.

Costantino, Grace. “Final Web Analytics Brief: Biodiversity Heritage Library.” Graduate Certificate, University of Denver University College, 2017.

Welter-Schultes, Francisco. Biodiversity Heritage Library. “BHL Survey 2010, Evaluation of the Results.” http://www.planetposter.de/bhl/wien2010.htm.

Biodiversity

Alba, Christina. iDigBio. “Natural History Collections as Primary Data in Ecological Research.” Accessed October 28, 2017. https://www.idigbio.org/content/natural-history-collections-primary-data-ecological-research.

Astrin Jonas J. and Hannah C. Schubert. “Community Perception of Natural History Collections - An Online Survey.” *Bonn Zoological Bulletin* 66, no. 1 (April 2017): 61-72. Accessed October 28, 2017. http://zoologicalbulletin.de/BzB_Volumes/Volume_66_1/061_072_BzB66_1_Astrin_JJ_and_Schubert_HC.pdf.

Biostor. http://biostor.org.

Culley, Theresa M. “The Frontier of Data Discoverability: Why We Need to Share Our Data.” *Applications in Plant Sciences* 5, no. 10 (October 2017): 1700111. Accessed October 28, 2017. http://www.bioone.org/doi/abs/10.3732/apps.1700111#.WesaAsa4X_M.twitter.

Encyclopedia of Life. http://eol.org.

Field Notes on Science & Nature. Edited by Michael R. Canfield. Cambridge, MA: Harvard University Press, 2011.

Global Biodiversity Information Facility. https://www.gbif.org.

Holland, Peter. *The Animal Kingdom: A Very Short Introduction.* New York: Oxford Press, 2011.

Integrated Taxonomic Information System. https://www.itis.gov.

Soltis, Pamela S.“Digitization of herbaria enables novel research.” *American Journal of Botany* 104, no. 9 (September 2017): 1-4. Accessed October 28, 2017. http://m.amjbot.org/content/early/2017/09/28/ajb.1700281.full.pdf.

Tuominen, Jouni, Nina Laurenne, and Eero Hyvonen. “Biological Names and Taxonomies on the Semantic Web -- Managing the Change in Scientific Conception.” *ESWC 2011*, Part II, LNCS 6644 (2011): 255-269.

Watson, Mark F., et al. *Descriptive Taxonomy: The Foundation of Biodiversity Research.* Cambridge, UK: Cambridge University Press, 2015.

Wright, John. *The Naming of the Shrew: A Curious History of Latin Names.* London: Bloomsbury Publishing, 2014.

Digital Libraries

Blumer, Eliane, Jasmin Hugi, and Rene Schneider. “The Usability Issues of Faceted Navigation in Digital Libraries.” JLIS.it 5 no. 2 (July 2014): 85-100. Accessed September 28, 2017. https://www.jlis.it/article/viewFile/10072/9421.

The DPLA Archival Description Working Group. “Aggregating and Representing Collections in the Digital Public Library of America.” Created November 2016. Accessed April 3, 2017. https://docs.google.com/document/d/16r_px4GajLIOZMlyXyN0pJFfWpstJIKxHF3brgM4YOo/edit.

Europeana. “Europeana Data Model Documentation.” Created November 18, 2014. Accessed April 14, 2017. https://pro.europeana.eu/page/edm-documentation.

Europeana. “White Paper on Best Practices for Multilingual Access to Digital Libraries.” Created June 23, 2016. Accessed April 14, 2017. https://pro.europeana.eu/files/Europeana_Professional/Publications/BestPracticesForMultilingualAccess_whitepaper.pdf.

International Image Interoperability Framework. http://iiif.io.

Marz, Nathan and James Warren. *Big Data: Principles and Best Practices of Scalable Real-time Data Systems*. Shelter Island: Manning Publications Co., 2015.

Museums Computer Group. “Welcome to the new National Galleries of Scotland website.” Created August 20, 2017. Accessed August 24, 2017. http://museumscomputergroup.org.uk/welcome-to-the-new-national-galleries-of-scotland-website/.

Scott, Tom. Wellcome Collection. “A digital strategy for Wellcome Collection.” Created October 18, 2017. Accessed October 28, 2017. https://stacks.wellcomecollection.org/a-digital-strategy-for-wellcome-collection-1b43e536533.

User Studies

Baxter, Kathy, Catherine Courage, and Kelly Caine. *A Practical Guide to User Research Methods: Understanding Your Users*. New York: Morgan Kaufman, 2015.

Connaway, Lynn Silipigni and Marie L. Radford. *Research Methods in Library and Information Science*. Denver, Colorado: Libraries Unlimited, 2017.

Digital Library Federation Assessment Interest Group, User Studies Working Group. DLF. “Surveying the Landscape: Use and Usability Assessment of Digital Libraries.” Created December 2015. Accessed February 25, 2017. https://osf.io/uc8b3/.

Hotjar. www.hotjar.com.

Nielsen Norman Group. https://www.nngroup.com.

Pew Research Center. “U.S. Survey Research. Questionnaire Design” Accessed May 8, 2017. http://www.pewresearch.org/methodology/u-s-survey-research/questionnaire-design/.

Rogers, Yvonne, Helen Sharp, and Jenny Preece. Interaction Design: Beyond Human-Computer Interaction. United Kingdom: John Wiley & Sons Ltd, 2011.

Sauro, Jeff. Measuring U. “5 Types of Usability Tests.” Created August 11, 2015. Accessed January 7, 2017. https://measuringu.com/five-types-usability/.

Skype. “How can I record my Skype calls?” Accessed January 7, 2017. https://support.skype.com/en/faq/fa12395/how-can-i-record-my-skype-calls.

Stacks, Don W. *Primer of Public Relations Research*. New York: The Guilford Press, 2002.

Storytelling with Data. http://www.storytellingwithdata.com.

Surveymonkey. https://www.surveymonkey.com.

Surveymonkey. “Sample Size Calculator.” Accessed May 4, 2017. https://www.surveymonkey.com/mp/sample-size-calculator/.

Usability.gov. https://www.usability.gov.

User Studies for Digital Library Development. Edited by Milena Dobreva, Andy O’Dwyer, and Pierluigi Feliciati. London: Facet Publishing, 2012.

User Study Examples

Europeana. “Europeana Cloud: Unlocking Europe’s Research via The Cloud.” Last Revised November 2015. Accessed April 18, 2017. https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.3%20D1.6%20User%20Requirements%20Analysis%20and%20Case%20Studies%20Report%20Content%20Strategy%20Report.pdf.

Hope, Courtney. Museums Victoria. “MV Website User Research Survey.” Accessed May 3, 2017.

Schigel, Dmitry. Global Biodiversity Information Facility. “GBIF User Survey Samples.” Accessed February 17, 2017.

Sheffield, Carolyn. Smithsonian Institution. “Field Book Project Cataloger Survey.” Submitted to Smithsonian IRB May 26, 2011.