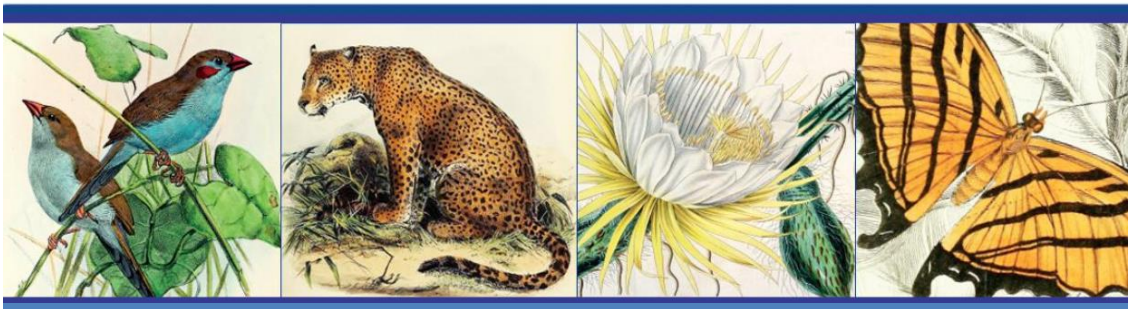


"TREASURES UNLOCKED" FINAL REPORT

ENABLING IMAGE DISCOVERY WITHIN THE BIODIVERSITY HERITAGE LIBRARY

Ariadne Rehbein
January 2017

Table of Contents

EXECUTIVE SUMMARY **2**

INTRODUCTION **3**

User-centered development best practices: Preceding design through user research 3

Project management best practices: Technical, Stakeholder, and Comparative systems requirements gathering 4

Crowdsourcing best practices: Lessons learned and ways to move forward 4

Pursuing new technologies: Computer vision in the sciences, art, and digital libraries 5

STUDIES AND RESULTS **5**

BHL Portal illustration users' needs research 5

Illustrations in taxonomic publications 5

Previous studies regarding BHL illustrations 6

NDSR Residency studies 6

BHL Scientists User Stories Blog post analysis 7

Interviews 10

Survey 12

Additional audiences for BHL images: Historians of science and artists/ illustrators 15

Crowdsourcing Study 16

Summary and best practices 16

Interviews 17

Disucssion of crowdsourced metadata from BHL's Flickr stream and Science Gossip 21

Metadata production through computer vision 26

Summary and best practices 26

Experimenting with Clarifai 27

RECOMMENDATIONS **30**

Recommendations for improved BHL Portal functionality 30

Recommendations related to figures and descriptions 31

Crowdsourcing recommendations 32

Computer vision recommendations 34

Future Work and Next Steps 35

CONCLUSION **39**

APPENDICES **41**

WORKS CITED **51**

Executive Summary

The "Treasures Unlocked" National Digital Stewardship Residency (NDSR) project is a way for the Biodiversity Heritage Library (BHL) to take the next steps in illustration discovery following the "Art of Life" National Endowment for the Humanities-funded project. "Art of Life" identified illustrations through automated means and welcomed crowdsourcing participants to contribute metadata through BHL's Flickr stream[1] and Science Gossip[2], a Zooniverse platform-based project focused on illustrated scientific periodicals from the Victorian era.

Research into the goals of taxonomists, historians of science, and artists/ illustrators surrounding natural history illustrations was conducted, resulting in the following recommendations for the BHL Portal and BHL leadership:

- An International Image Interoperability Framework (IIIF) server and viewer to enable fast and rich delivery of images.
- User accounts to save illustrations and offer page-level citations to facilitate creation of citations for illustrations.
- The ability to download up to every illustration in a book, with a high resolution option.
- A "View illustrations only" button within the BHL book navigation panel.
- Asking users to leave notes or links regarding the location of illustration plates.
- Collaborating with creators of biodiversity data to integrate metadata regarding illustrations tied to original descriptions of species.

The successes and challenges of BHL's current illustration crowdsourcing projects, according to crowdsourcing best practices and interviews with participants, indicate that BHL should continue this work with increased project management and a clear statement of purpose for the data produced. Expanded illustration crowdsourcing project ideas are presented, such as corrections of automated descriptions through a game and Wikipedia editing.

Experiments with computer vision applications and industry research suggest that BHL should pursue collaborations with external subject matter and technology experts. This would enable BHL to develop more nuanced search terms for illustrations, combine the results with human review, and develop other possible search services.

Future approaches towards interface design should have a basis in behavioral user research, BHL's existing crowdsourced metadata, users' requested search fields, aspirations from product stakeholders, and image delivery and searching standards.

---

[1] www.flickr.com/photos/biodivlibrary/albums

[2] www.sciencegossip.org

Introduction

Interface design, cultivating engagement and metadata production through crowdsourcing, automated means of metadata production, and integrating crowdsourced metadata into the technical architecture of BHL are exciting efforts. They present challenges as inherently interlinked undertakings. These are areas in which BHL has some experience, and can therefore learn from prior successes, roadblocks, as well as suggestions for, if not application of methodologies, technologies, and philosophies within the course of the project year.

As explained by the BHL NDSR grant Project Director, Constance Rinaldo, the project was not intended to take BHL to "the point that we'll be able to do illustration searching, but rather [provide] ways to move forward" towards this larger goal based upon research into or application of evolving best practices in digital libraries and the larger biodiversity community.

This project allowed me to identify or demonstrate best practices in digital library interface and searching standards, cultivating engagement through crowdsourcing, future means of metadata production, and user-centered research. My final report describes best practices resulting from this research, recommends actions or decisions that BHL would need to make to work in alignment with them, and future research or work to be accomplished towards the goal of illustration discovery through the BHL Portal.

User-centered development best practices: Preceding design through user research

Regarding recommendations for enhancements to the BHL portal, the BHL Lead Technical Developer, Mike Lichtenberg, suggested that I consider providing use cases or user stories (or both) as a way of capturing system requirements. These could be accompanied by mockups or screenshots. Mike would then create functional specifications with this as a basis. According to usability.gov, functional specifications provide details of how a product should behave and specify what is needed for development.

A use case is a sequence of simple steps taken to accomplish a particular task, beginning with a user's goal and ending when that goal is fulfilled. Use cases designate how a website should respond to an action. User stories are informal explanations of what the software should accomplish, but do not state all of the details necessary to implement a feature. They are a central part of agile technology development projects. While user stories can be written throughout the

release cycle, typically, everyone on the team participates in writing them to create a product backlog at the kickoff of a product. In the agile approach, this is preceded by research into the intended users and their goals. In some cases, prior to the technology building cycles, high fidelity user interface designs are provided to developers. These are based upon user stories and honed through user testing and application of design best practices. User experience professionals may also work slightly ahead of developers in providing designs for certain aspects of an interface. The process applied by different organizations of course varies, and many are learning about this approach and how it may be adopted in a useful way to meet their goals.

I focused largely on taxonomists' goals, attitudes, and problems surrounding illustrations as a foundation of all work to come. Examples of the ideation process and enhancements to the BHL Portal based primarily on taxonomists' needs are included as part of this report. With this foundation in user research, I recommend that BHL continues to pursue user-centered methodologies, including design and testing through behavioral research, in its product development.

Project management best practices: Technical, Stakeholder, and Comparative systems requirements gathering

Various requirements gathering approaches were explored towards the larger goal of enabling illustration access through the BHL Portal: discussing the primary users of the BHL Portal and broader audiences for illustrations with BHL leadership, stakeholder interviews, a metadata evaluation, and investigations into searching standards and image access standards among digital libraries.

Crowdsourcing best practices: Lessons learned and ways to move forward

The role of crowdsourcing volunteers in improving access to illustrations through the BHL Portal is an important consideration, encompassing both the potential for engagement around BHL content and the opportunity for producing new metadata. I took the approach of exploring current work and the meaning it holds for participants. Application of best practices in evaluating current efforts and conversations with participants resulted in lessons and various ideas for crowdsourcing topics and methods that could be of benefit to BHL, beyond metadata production for access to illustrations through the BHL Portal. This report includes guidelines and various ways that BHL can make illustration-based crowdsourcing a core activity.

Pursuing new technologies: Computer vision in the sciences, art, and digital libraries

A small number of illustrations within the BHL corpus have tags connected to them: less than 1% out of an estimated 4-5 million. Computer vision is a booming area, with existing applications in the biodiversity and the cultural heritage sector. An investigation into current uses among digital libraries, internal experimentation, and ideas for bringing together computer vision and human tagging were explored.

Studies and Results

BHL Portal illustration users' needs research

Findings regarding taxonomists', historians', and scientific illustrators/ artists' goals, attitudes, and problems surrounding scientific illustrations are presented. This research included a review of illustrations in the taxonomic research and publishing process, knowledge gathered by BHL, and studies conducted as part of the Residency. Taxonomists were selected as the primary user group to study since the focus of the BHL Portal is to support scientific research, in alignment with the BHL mission to improve research methodology through open access to biodiversity literature.

Illustrations in taxonomic publications

As expressed by Willi Eglof et al in their article, "Copyright and the Use of Images as Biodiversity Data," "Since the first accepted contributions to taxonomy (Clerck 1757, Linnaeus 1753, Linnaeus 1758), taxonomic publications have contained treatments identifying a taxon using a scientific name in a hierarchical classification, list characteristics that define the taxon and distinguish it from all others, report where the taxon has been found, and cite earlier publications with content on that taxon… Published revision of taxonomic groups, such as genuses or subfamilies require that all relevant previous treatments be cited." Illustrations within works targeted at the professional taxonomic community are intended to serve as aids in identifying biodiversity. In establishing a newly named species in a taxonomic publication, it is ideal to include an image highlighting the structural features of the type specimen in high detail. For example, Recommendation 16F of the International Code of Zoological Nomenclature (ICZN) suggests that

holotypes or syntypes should be illustrated in a work in which the new nominal taxon is established. While a line drawing created by a scientific illustrator may be the prototypical concept of an illustration, the Code describes an illustration as "a work of art or a photograph depicting a feature or features of an organism, e.g. a picture of a herbarium specimen or a scanning electron micrograph." Maps and other diagrams are often also included to provide further context.

Similarly, manual classification of a selection of BHL illustrations as part of the "Art of Life" project allowed the team to estimate that 52% of illustrations within the BHL corpus are drawings, paintings, or diagrams, 24% are photos, 19% are charts or tables, 4% are maps, and <1% are bookplates. By running algorithms across BHL Pages, the Art of Life project team found that approximately 10% of BHL corpus contained illustrations. In addition to illustrations, Harris, Bridgewater, and Moutsambote describe referring to identification keys and physical specimens as potential components of the identification process.

Previous studies regarding BHL illustrations

Knowledge about illustration-related functionality within the BHL Portal has been gathered in several ways. Surveys spearheaded in 2009 and 2010 by Francisco Welter-Schultes and Bianca Lipscomb regarding BHL Europe recorded the rate of finding illustrations as half as frequent as accessing text. The researchers surmised that "the lower rating might have to do with plate numbers not being given in the page level metadata." A usability test conducted by BHL staff in 2011 provided similar conclusions: finding illustrations was a frustrating experience. Users were hindered by lack of or inaccurate page level metadata in the book navigation panel. Kelli Trei (Biosciences Librarian and Assistant Professor at the University of Illinois Urbana Champaign) conducted an analysis of BHL User Story blog posts in 2016 as part of her study, "The Impact of the Biodiversity Heritage Library on Scientific Research" Trei noted that 16% of scientists' blog posts mentioned downloading illustrations.

NDSR Residency studies

To gather a more current and comprehensive understanding of taxonomist user needs, I analyzed BHL User Stories blog posts, held interviews with Research staff at the Missouri Botanical Garden, and conducted a survey of BHL users (as part the larger NDSR Individual Users Survey.) It was valuable to capture language surrounding illustrations through the blog posts, deepen insights through interviews, and confirm findings and gain further ideas through survey results.

These studies reached a broad base of participants: taxonomists spanning different research areas who actively use BHL and the illustrations made available through this resource as well as scientists within a specific focus of taxonomic research (botany) who could speak to illustration usage more generally.

These methods were informed by Christian Rohrer's article, "When to Use Which User-Experience Research Methods" provided on the Nielsen Norman Group website[3]. The studies were attitudinal rather than behavioral (discussing behaviors rather than observing them) and were intended to discover important issues to address among this group. With limited access and functionality regarding illustrations in the BHL Portal, attitudinal research was considered the most appropriate. The studies inspired new ideas and recommendations for serving the illustration needs of this group (the strategize phase of product development), though the BHL User Stories blog post analysis and NDSR survey also spoke explicitly to desired improvements to BHL (the assessment phase of product development.)

Best practices were also applied through the elicitation of stakeholder requirements for illustration related design and functionality in the BHL Portal. According to Angela Wick, in the *Lynda.com* video, "Requirements Elicitation for Business Analysts," the leaders and managers of the area impacted by the project "provide a unique perspective to the problem and opportunity addressed by the solution." With her depth of knowledge regarding BHL illustrations, BHL Outreach and Communication Manager, Grace Costantino, was consulted regarding aspirations for image access through the BHL Portal. Dedicated BHL volunteers, Michelle Marshall and Siobhan Leachman were also consulted as subject matter experts in BHL illustrations and crowdsourcing.

BHL Scientists User Stories Blog post analysis

Methodology

This study built upon the classification of BHL User Stories blog posts developed by Kelli Trei as part of her study, "The Impact of the Biodiversity Heritage Library on Scientific Research." Trei's main questions were: "How are people, scientists in particular, using the BHL?" and "What impact, if any, does layperson use of the BHL have on science?" At the time of her research, there were 73 blog posts (February 28, 2011 to June 9, 2016.) There were 14 posts written after 6/9/2016.

[3] Nielsen Norman Group is a leader in the user experience field, providing consulting and resources.

Due to a glitch on the BHL blog, Kelli's analysis did not include the 2/15/2011 post. The analysis done for this report included 88 posts, ending with Dr. Alejandro Bortolus' post on 8/3/2017.

The additional blog posts were classified by user type and tagged as "illustrations mentioned" if they included one or more of the following terms: illustrated, illustration, images, image, illustrations, plate, #sciart, art. Then the paragraphs in which illustrations were mentioned by scientists were gathered and classified . There was only one case in which one scientist's blog post was classified in two ways. The dataset may be reviewed in the Google Folder [4] that accompanies this report.

Results

The following bar chart (Fig. 1) captures user types represented in BHL Blog User Stories and the number of posts mentioning illustrations.

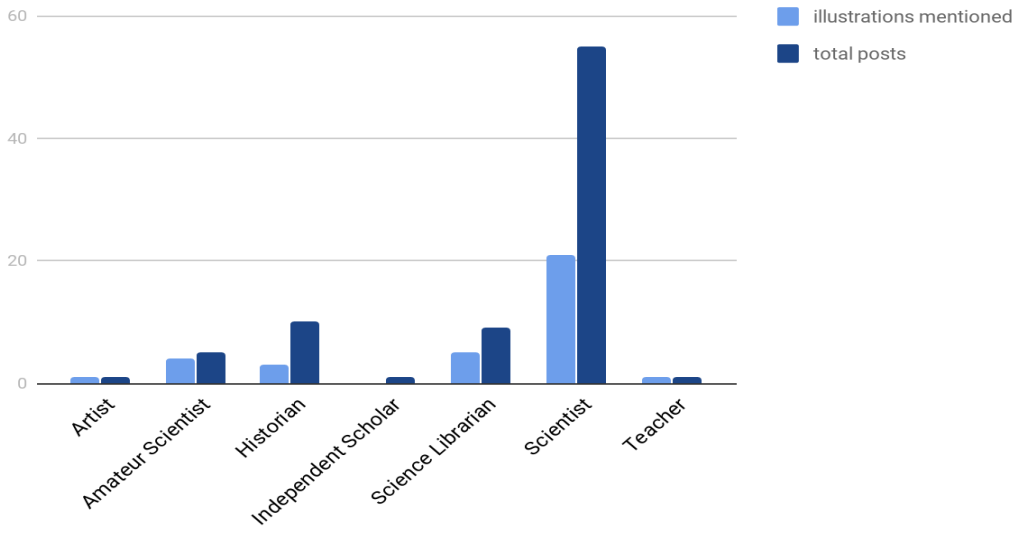


Fig. 1 – User types from BHL Blog User Stories

[4] https://drive.google.com/drive/folders/1niiEgd7JPkHvzISZztdwoFsx8bQIZSEy?usp=sharing

Scientists' blog posts revealed a range of responses (Fig. 2). It is compelling to see the emotional value provided by illustrations. Users enjoy the beauty of historic illustrations, with two users feeling more connected to the history of biology. For example, Christopher Buddle is personally inspired by BHL illustrations and he sees their prominent role, and BHL's most valuable contribution overall, as igniting passion and engagement regarding biodiversity across all sectors of society. Use of illustrations for fundraising, as discussed by Karen James, ties into this concept.

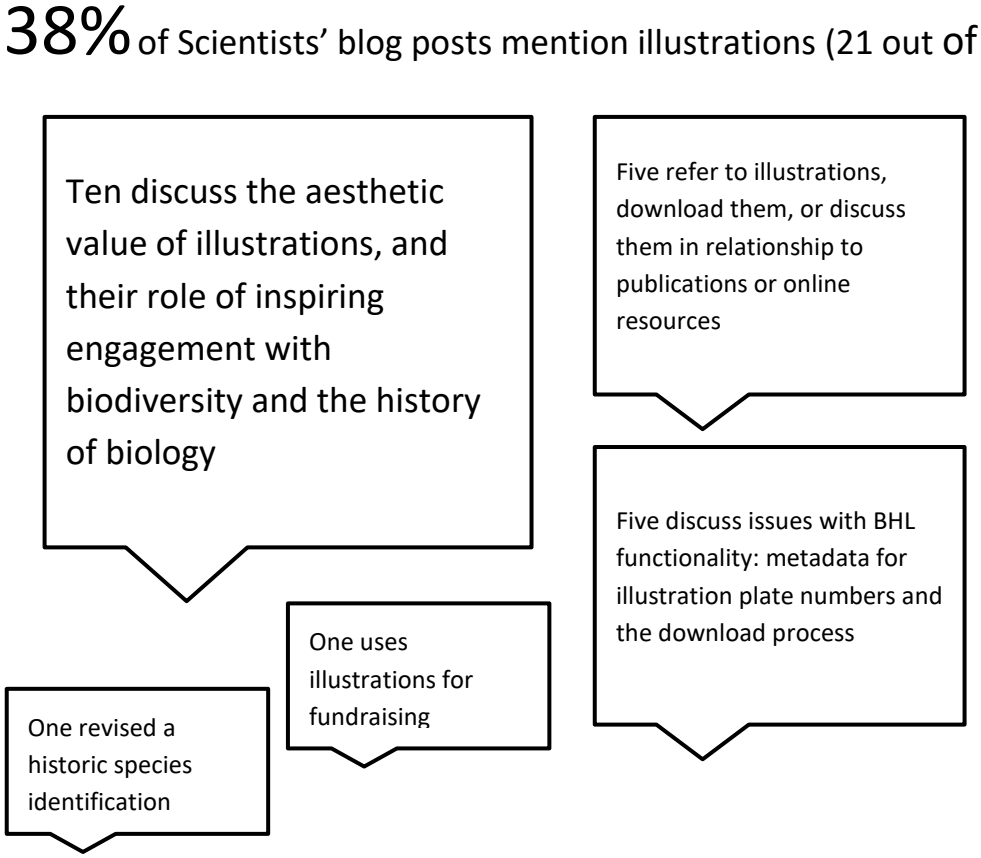


Fig. 2 – Summary of scientists' blog post analysis

The difficulty of using the BHL Portal to access and download illustrations was confirmed from previous user studies. Moreover, the same advice for metadata improvements was provided, in addition to improved download functionality. This quote from Barna Páll-Gergely is one of several related to improved page-level metadata and download options: "In many cases the plates of figures and the blank pages in-between appear marked as 'text' on the left-hand navigation menu. It would be great to enumerate them. The other thing is that selecting pages of a single publication

and finding the plates at the end of each volume takes a long time. Sometimes it would be more user-friendly to provide a way to automatically group these together for download."

Five users discuss referring to illustrations, downloading them, or in relationship to publications or online resources. Further information regarding motivations to refer to and download illustrations was gathered through interviews.

Interviews

Solicitation for Participation

Interviews were conducted among Research staff members at the Missouri Botanical Garden to learn how illustrations fit into this group's work. A call for interviewees was sent to the Research staff email listserv as well as through the employee-wide e-mail newsletter.

The main interview questions were informed by user experience research practice: gathering contextual information and goals related to illustrations. Interviewees were asked about the goals of their position and what they achieve through use of illustrations. Interviews were exploratory in nature; further detail was prompted throughout the conversation.

Interviews were conducted over the course of March 2017, and included:

Gordon McPherson, Curator, Africa and Madagascar

Amy Pool, Curatorial Assistant, Research

Barry Hammel, Curator, Research

Jim Miller, Senior Vice President, Science and Conservation

Jan Salick, Senior Curator, William L. Brown Center

Yevonn Wilson- Ramsey, Senior Botanical Illustrator

The interview data was analyzed by gathering related statements and classifying them. Consent forms were not developed for these interviews, therefore directly attributable quotes are not provided in the results.

Results

The analysis indicated that primary uses for illustrations include:

1. Illustrations of type specimens are ideal for identifying specimens (though resources other than scientific drawings may also be used, such as photographs and specimens.)
2. Citation of or inclusion of illustrations in publications or online resources, such Google map projects or exhibitions and presentations to teach about botany
3. Aid understanding of historic conceptualizations (species names and place names)
4. Aid understanding of unclear species description in historic literature (a rare occurrence)
5. Various resources are used to access species images: Plantillustrations.org (historic illustrations, used for exhibitions) and JSTOR Plants (images of specimens), and Google Images (photographs of live plant, scientific name used for searching)
6. Assignment of lectotype in lieu of physical specimen (a rare but important occurrence)

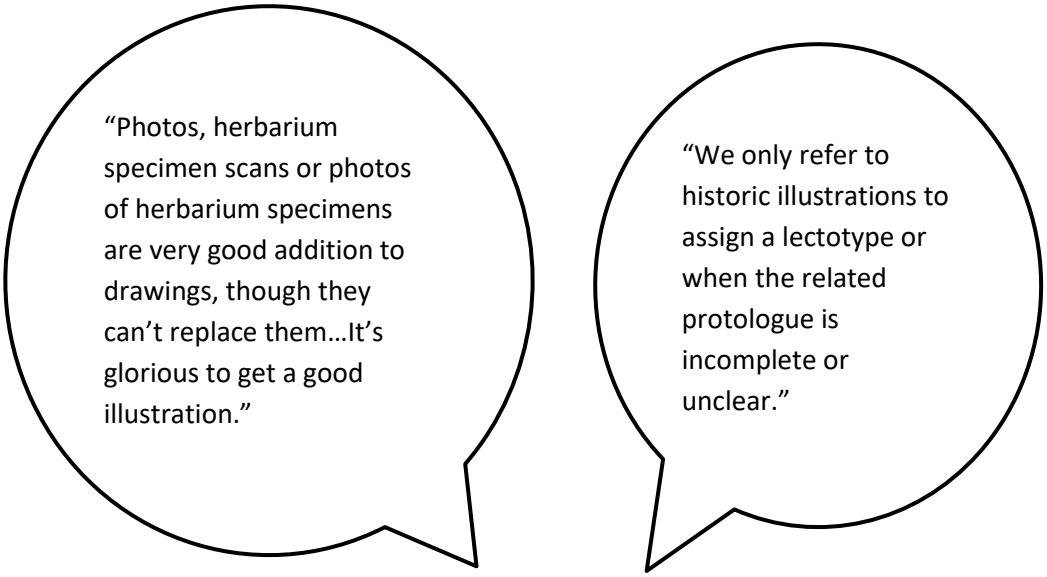


Fig. 3 – Quotes from interview participants

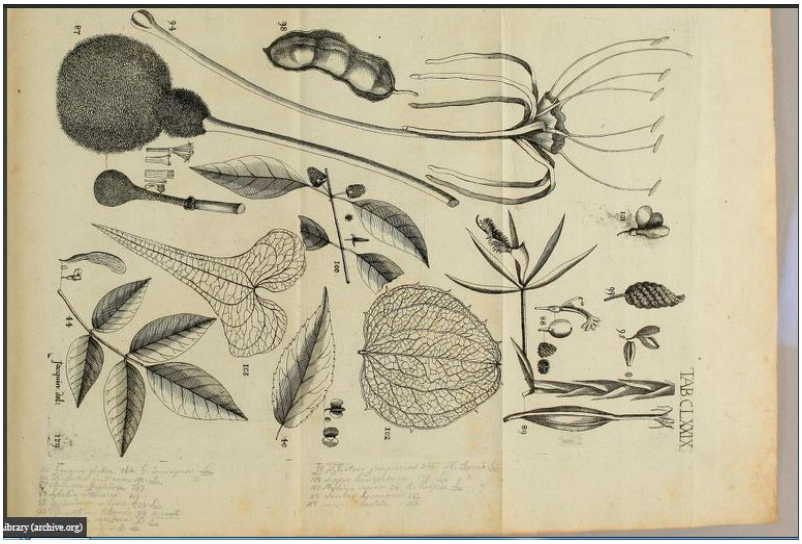


Fig. 4 – Lectotype of *Gouania glabra* Jacq. Selectarum Stirpium Americanarum Historia ... 264, pl. 179, f. 40. 1763

Pool, A. 2014. Taxonomic revision of Gouania (Rhamnaceae) for North America. Ann. Missouri Bot. Gard. 99(3): 520. www.botanicus.org/page/2552302.

Survey

Methodology

Illustration-related questions from the Individual Users Survey conducted by NDSR Resident Pamela McClanahan were analyzed. Pamela and I worked together to create the following questions:

1. Do you use illustrations/ images in BHL? (classified by user type)

Among those who responded “yes”:

2. How do you use those illustrations/images?
3. Are there other features related to image use that you would like to see added?
4. How do you find those images?
5. Are there other ways you would like to search for images?

Questions 2, 3, 4, and 5 were free text responses. These questions were only available to respondents who responded "yes" to Question 1. Similar responses within each question were gathered and then classified for scientists. Since there were relatively few historian and artist/illustrator responses, these were simply counted.

Appendix 1 includes graphics representing results for each of the survey questions. The Google Folder associated with this project includes both the raw survey data and analysis document for scientists.

Results

Fifty-nine percent of Scientist / Taxonomists, or 165 of 278 responded that they use illustrations/ images in BHL.

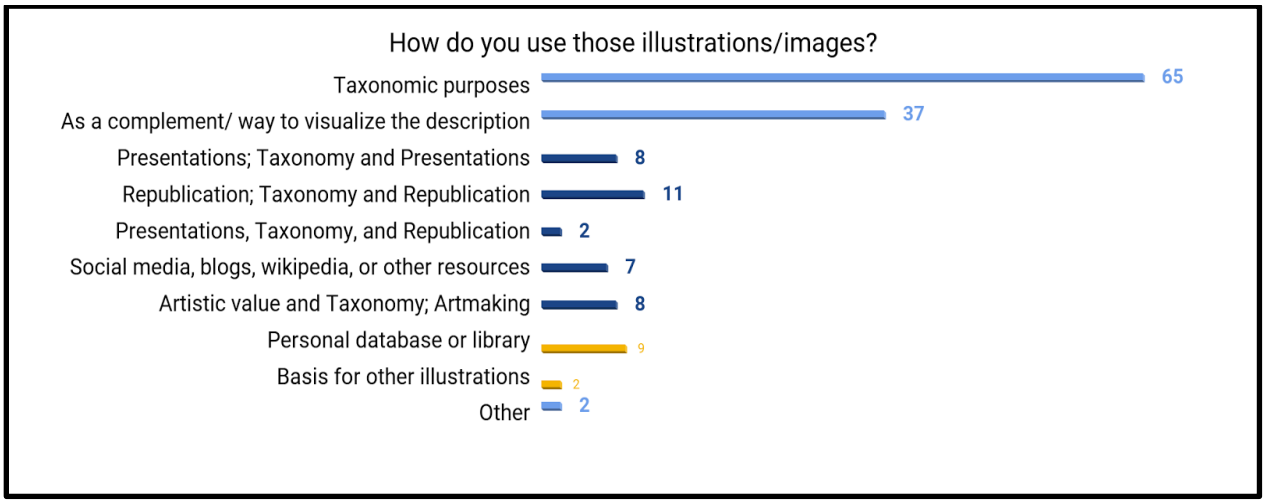


Fig. 5 – Results of Question 2

Uses:

- The top uses for illustrations expressed by taxonomists were for specimen or species identification. Visualization of species descriptions was specifically mentioned by a high number of participants, and so has been broken out into a separate category.
- Unlike in the BHL User Story Blog post analysis, an appreciation for the aesthetic value of illustrations was minimal.

- The survey confirmed the republication of illustrations in articles and books as well as sharing them through social media, online resources, and presentations.
- Newly expressed uses of illustrations were artmaking and saving in personal libraries.

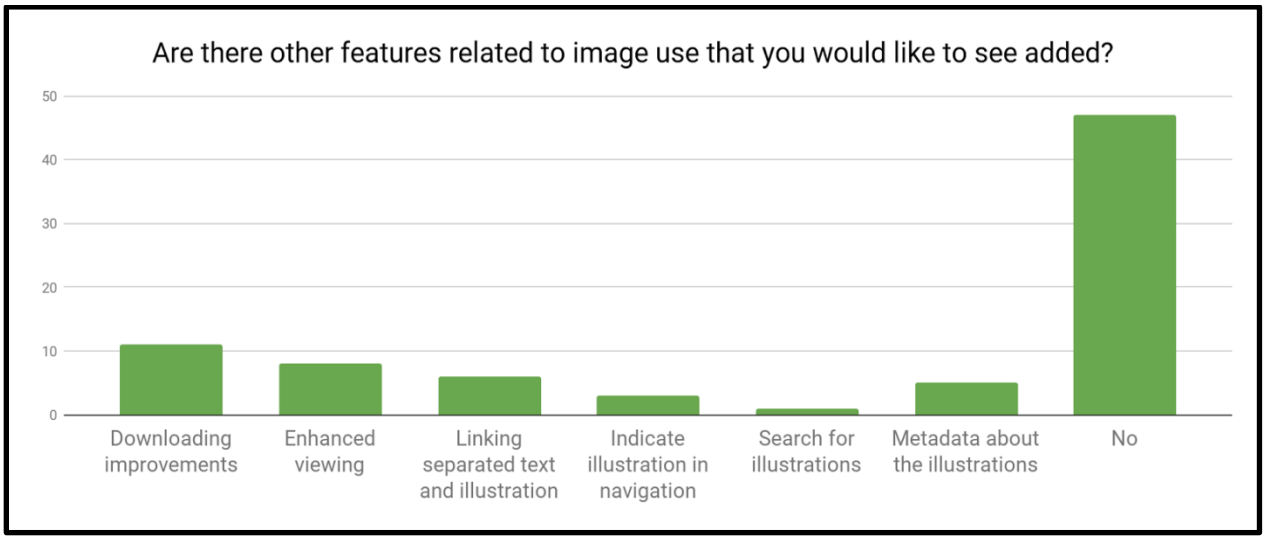


Fig. 6 – Results of Question 3

User access challenges and requests for improvements:

- Problems were confirmed from prior research, such as the need to improve downloading options, for example higher quality TIFFs or JPEGs and an easier way to download individual or all images from books. Also, there is difficulty in locating illustrations tied to a species description of interest. A new suggestion to overcome this problem is linking illustrations to species descriptions that are on non-contiguous pages.
- The option to show only image pages has been suggested in internal interviews with BHL stakeholders and was also suggested in response to Question 4. Previous suggestions were repeated, such as indicating plate information in the book navigation panel.
- A few respondents requested new features, such as enhanced viewing options and downloading an index of images from each book (Question 4.)

Current discovery processes:

- Beyond manually browsing through books, the survey confirmed that taxonomists often gain access to illustrations through their citation in another work or online resource such as Tropicos. (Question 5)

Illustration search terms:

- Many participants responded that there are not other ways that they would like to search for illustrations, perhaps because scientific name is already available for searching text resources on BHL. The small number of positive responses included:
    a. By scientific name (16%, or 14 of 85 respondents to this question)
    b. By scientific name in addition to other categories, such as taxonomic category, artist, author. (<1%)
    c. Visually related images (<1%)
    d. General request for image searching (<1%)

Additional audiences for BHL images: Historians of science and artists/ illustrators

A review of the NDSR Individual Users survey results and informal conversations with an "Art of Life" Advisory Board member and a Missouri Botanical Garden scientific illustrator uncovered uses for illustrations among historians of science, artists, and professional scientific illustrators. Within the NDSR Individual Users survey, 65% of Historians, or 19 of 29 responded that they use illustrations/ images in BHL, and 78% percent of Artists/ Illustrators, or 18 of 23.

Like taxonomists, six historians spoke of including illustrations in presentations and publications, and two spoke of sharing them online (through a blog and an app). A few examples of research interests regarding illustrations were landscape gardens, researching artists themselves, identifying species, and understanding the observations of 18th/19th century scientists when terms did not exist in that time. Several artists used illustrations in their creation of collages. Among those who used BHL illustrations in their scientific illustration work, searching by scientific name was requested. Botanical scientific illustrators primarily referred to physical herbarium specimens, as well as a variety of photographs of live species from online sources in their work.

Search fields requested by historians of science include: location, species, scientific name, illustrator, and time period. Artists requested the same search fields, but also included author, publisher, and media type.

Crowdsourcing Study

Summary and best practices

Integrating crowdsourced illustration metadata into the backend architecture and frontend user interface of the BHL Portal should prompt BHL to consider the sustainability of existing projects and to review the experiences of contributors if not improve them according to new technological capabilities, system design, and project management best practices. A fuller understanding of this work can inform organizational decision making and next steps. Moreover, framing crowdsourcing projects as learning experiments is a common philosophy in this field. As stated by Chris Lintott (astronomer and founder of both Galaxy Zoo and Zooniverse) at CrowdCon 2015: "…the idea of internally framing our projects as experiments rather than as complete systems. So instead of just saying, 'We are going to produce this data,' or 'We are going to produce this information to add to our collection,' or 'We are going to enable this audience, we say, 'We will also learn about how best to build these projects.'"

BHL Flickr tagging started as an effort to automatically ingest BHL images into the Encyclopedia of Life through the use of machine tags for species (ex. taxonomy:binomial=Picea excelsa.) Science Gossip was established as part of the "Art of Life" project. This a Zooniverse platform-based project, collaboratively built with the Constructing Scientific Communities grant project, upon conferral of an award to the "Art of Life" team for their project proposal. This site is focused on exploring and adding useful metadata about the work of Victorian scientific illustrators. Participants on both BHL's Flickr stream and Science Gossip have pursued (and discussed their shared) interests in locating, researching, and sharing information about women scientists/artists on Wikipedia. An important component of BHL Flickr tagging for top participants is adding current taxonomic names, synonyms, common names, and even Pre-Linnaean names to illustrations.

According to Donelle McKinley, a leading researcher in cultural heritage crowdsourcing, projects may be evaluated according to their level of participation and the quality of their contributions. Understanding and connecting to participant motivation, effective project management and system

design, and evaluation and refinements over time are essential to project success. These served as the basic evaluation criteria in reviewing these platforms and informed my studies.

Interviews

Solicitation for Participation

Dedicated BHL volunteers, Siobhan Leachman and Michelle Marshall were interviewed through Skype using the same interview protocol explained below. They served as a continuing resource through blog comments and emails and also provided me with a Google Doc explaining their viewpoints on BHL illustration crowdsourcing.

Flickr

The most prolific Flickr taggers as of February 2017 were contacted for interviews, starting with the top 18 who contributed over 450 tags. Beyond Siobhan Leachman and Michelle Marshall who have contributed in the tens of thousands, most of these taggers had contributed several thousand tags. Four of the group appeared to have deleted their Flickr account names, as they did not appear in the Flickr account search results. Another top tagger, Chris Freeland, was not contacted; having served previously as the Technical Director of BHL he wouldn't provide the perspective of a typical participant. A list of top participants may be reviewed in the Google Folder associated with this project.

As invitations were sent through the BHL Flickr account, Grace Costantino provided feedback on the invitation text. This text was also used to invite Science Gossip participants directly. Among the group of most prolific taggers, the participants who responded but did not participate hadn't tagged BHL illustrations on Flickr in a very long time and felt that they couldn't provide enough information about their participation to justify an interview. The Flickr account comment history revealed that many of the top taggers last participated a year and a half ago or more. Invitations were only sent to the most recent taggers, resulting in the following participants:

John Tann - Skype call (7/25)

Hollis Marriot - phone call (8/9)

Shelley James - written response (8/17)

Ashley Bradford - phone call (8/21)

Science Gossip

I introduced the "Treasures Unlocked" project to Science Gossip community through the Science Gossip blog and the Science Gossip Talk Research forum on May 2, 2017. A call for participants was later posted on the Research board on June 5, 2017. The moderators provided six additional names of potential participants, including formerly active Science Gossip users and a moderator. Seven of the most recent Science Gossip participants were contacted directly with interview invitations. The following participants were interviewed:

Geoff Belknap (Science Gossip lead researcher) Skype calls 3/9, 3/12, and 5/18 (the latter for discussion of interview questions)

Jules Wilkinson (Science Gossip moderator) – written interview (6/22)

Zuza Macháčková (Science Gossip moderator) - written interview (7/1)

Elisabeth Baeten - Skype call (7/19)

Interview Protocol

Participants were asked to address the themes below. Additional questions were included as prompts if needed (See Appendix 2.) The interviews were conversational and exploratory in nature, allowing expansion upon new topics of interest. Interviews were analyzed according to common themes.

Themes

1. Understand the reasons for your involvement in BHL illustration crowdsourcing in relationship to any involvement in crowdsourcing/ citizen science overall

2. Evaluate successes and challenges of Science Gossip/ BHL Flickr according to best practices

3. Illustration use and sharing

4. Future BHL illustration crowdsourcing

Results

Interviewees spoke of several motivations for participating in crowdsourcing on BHL's Flickr account or Science Gossip, including the dedication to sharing information, the desire to make a difference, personal interests in the content, the ability to make new discoveries and pursue them further, and a temporary "obsession" with certain crowdsourcing projects. A range of interests were found among just this small group of interviewees: For example, Flickr taggers were interested in Australian species, New Zealand species, and American West plants species, bugs, and bees. Science Gossip interviewees noted Arctic/Antarctic exploration, taxonomic tagging of ocean creatures, wild animals, or plants, being drawn to beautiful historic scientific illustrations, and the study of historic scientific illustrators. The interest in women in science emerged on both platforms. Matching personal interests with the content of the project was cited as the major factor in choosing to participate in future BHL illustration crowdsourcing projects, but supporting the other motivations would be highly important for future project success. These findings match the common altruistic and self-directed motivations found among cultural heritage crowdsourcing volunteers and citizen scientists respectively identified by Donelle McKinley and Gitte Kragh.

Additionally, two participants felt that BHL should continue to use existing outside platforms rather than develop a crowdsourcing platform or initiative directly within the BHL Portal. For one, the benefit of an outside platform was the potential for a worldwide audience. She perceived the primary audience of BHL to be from English-speaking countries. The other felt that the creation and maintenance of bespoke technology for crowdsourcing by BHL staff would be too onerous.

Regarding the discovery of BHL's illustration crowdsourcing projects, the Science Gossip interviewees heard of the project through Zooniverse advertising. They were long-time participants in Zooniverse projects, starting in 2007, and 2012 respectively. Additionally, they were not familiar with BHL until participating in Science Gossip. Three of the four Flickr participants had existing knowledge and involvement with BHL, as well as professional roles related to biodiversity. One of the Flickr interviewees had a major interest in sharing and adding descriptive tags to their own and others' biodiversity related photographs on Flickr, but was not very familiar with BHL.

Limitations in system design, project management, and a high taxonomic skill level are all challenges for participation on BHL's Flickr account. Almost all interviewees used self-created

spreadsheets to copy and paste tags. Participants spoke of having multiple windows open, including to BHL to search for descriptions of the illustration, a tagging spreadsheet, and current scientific name, common name, and synonym references such as the Catalog of Life or the Plant List. Requiring machine tags was seen as particularly frustrating for one of the participants and two users were unclear about their purpose. For example, one had questions about whether it was valuable to add editor or artist tags, since she only saw this encouraged for particular materials. Moreover, she was unclear about where to go for questions. This technical overhead combined with the high level of expertise in taxonomy required to fully tag illustrations with synonyms, common names, and current taxonomic names led two participants to say, "no one else would do this."

By contrast, Science Gossip provides greater support for participation through more flexible system design and project management capabilities. Science Gossip Moderator, Jules, shared that "the workflow works well and the interface is clean and easy to use." One of the benefits is a form-based interface for inputting various values as appropriate. Another benefit is the easily accessible discussion board where resources can be shared. As questions have arisen, the moderators, Geoff Belknap, and participants have been able to work together to construct FAQs. Moderators have also added resources such as list of common contributors, including links to databases of scientific illustrators to help decipher names. Participants have the option of performing technically and intellectually challenging work (the same species research conducted on BHL's Flickr account) or researching illustrators. They furthermore have the option of asking for the research assistance of others and sharing their findings, particularly through the use of hashtags. The relatively simpler tasks of describing illustrations through free text keywords and transcribing figure inscriptions allows for broader participation as well. While these tasks are possible on BHL's Flickr account, taxonomic machine tagging has been promoted as the main focus in posts such as "We need your Help to Tag over a Million Biodiversity Images in Flickr" by Trish Rose-Sandler on the BHL Blog.

Interviewees on both platforms expressed frustration with the lack of project updates and results from their work. This is a challenge to participants' motivation to make an impact. As long time Zooniverse participants, the Science Gossip interviewees expected a higher level of engagement from the research team such as updates on their research findings as the project progressed, new goals for the group, or progress towards the eventual creation of a resource. This was tied to the loss of community and participation in their view. Due to their experience with other Zooniverse projects, they were also able to provide suggestions for reinvigorating the group.

BHL Flickr interviewees mostly worked independently from other participants on this platform (only Siobhan Leachman and Michelle Marshall mentioned staying in contact through Twitter.) Among these interviewees, Siobhan and Michelle in particular were frustrated that tags and illustrations were not being ingested and shared by EOL or BHL. They therefore found alternative purposes and ways of sharing their work, such as tagging broad taxonomic categories commonly featured through the Historical Science Art social media accounts, adding Wikidata tags to support ingestion of illustrations into Wikimedia Commons, and writing Wikipedia articles and accompanying them with tagged illustrations.

Discussion of crowdsourced metadata from BHL's Flickr stream and Science Gossip

BHL's Flickr metadata is backed up once a month to BHL storage, through an API call managed by the BHL Lead Technical Developer. Science Gossip metadata is not backed up frequently, though older metadata is available through BHL's Github page.[5] Up-to-the moment Science Gossip metadata is available through "The Illustrated Milkman" website[6]. There is also a dashboard tracking participation over time. [7]

In its raw form, the BHL Flickr metadata is a tab separated value spreadsheet, listing each tag and its associated BHL page ID, Flickr page ID, machine tag status, and the tag's creator (Flickr user ID and name.) The following (Fig. 7) is an example record for one illustration:

[5] https://github.com/gbhl/bhl-us-data-sets/tree/master/Zooniverse

[6] http://explore.sciencegossip.org/groups

[7] https://api.zooniverse.org/projects/illustratedlife/status

BHL page ID	Flickr page ID	Machine tag (1/0 =yes or no)	Tag	Flickr user ID	Flickr User
188765	8358234364	1	taxonomy:binomial =Cathartes aura	124202008@N07	lyoung22205
188765	8358234364	1	taxonomy:binomial l=Vultur aura	124202008@N07	lyoung22205

Fig. 7 – Example of a BHL Flickr metadata record

Science Gossip metadata is available as both an HTML page and a comma separated value spreadsheet. In both of these forms, the metadata is nested in JSON (JavaScript Object Notation, a file format used to transmit data objects.) The following (Fig. 8) is an example record for a page with one illustration:

Science Gossip ID	Number of taggers	BHL ID	vol	page	Date	Image scale	Type	Coordinates of identified type	**keywords**
ASC00009wy	6	1634794	v. 7 (1871)	Page 4	1871	0.38043478 26086957	Drawing	[0, 0, 165, 239, 201, 235]	**keywords** ["white ant", "female", "winged ant", "winged", "insect", "diagram", "Insect", "White Ant", "Termite", "winged female", "termite", "India"]
ASC00009wy	6	1634794	v. 7 (1871)	Page 4	1871	0.38043478 26086957	Species	[293, 374, 0, 0, 0, 0]	**common** ["Termite", "winged termite", "White Ant, Termite"] **scientific** [""]

Fig. 8 – Example of a Science Gossip metadata record

There is a great variance in the structure and depth of tags in the Science Gossip and Flickr metadata sets. For example, on BHL Flickr, 31,771 images are tagged with 134,176 tags. However, the number of tags range between 1 and 71 per image. On Science Gossip, 9,529 pages have been linked to over 32,000 concepts, such as a page, drawing, photo, inscription, or species, demonstrating the depth of information created through this platform. Each of these concepts has a wealth of free text terms associated with them. The gaps in crowdsourced descriptions should also be considered in the determining the feasibility of search and browse functionality and the ability to present standardized information regarding each item.

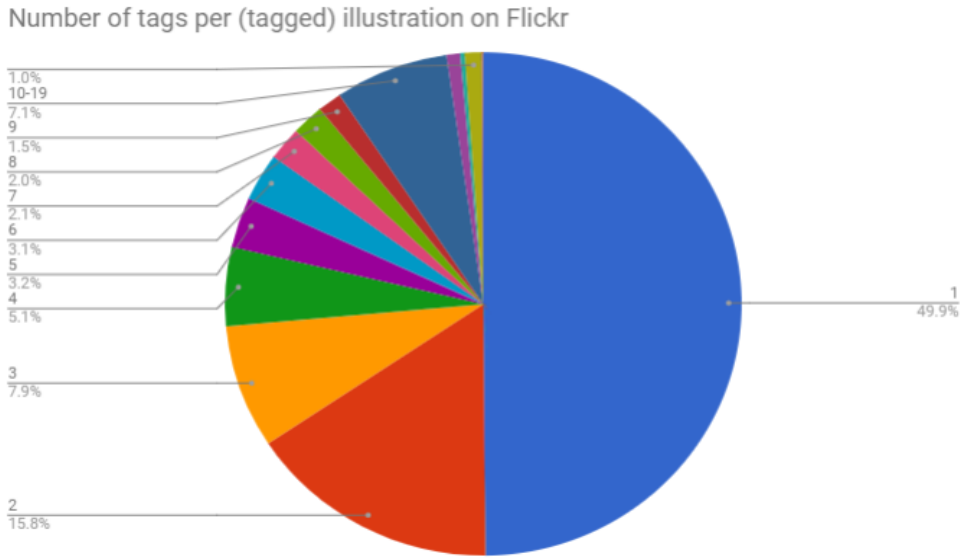


Fig. 9 – Number of tags per (tagged) illustration on Flickr

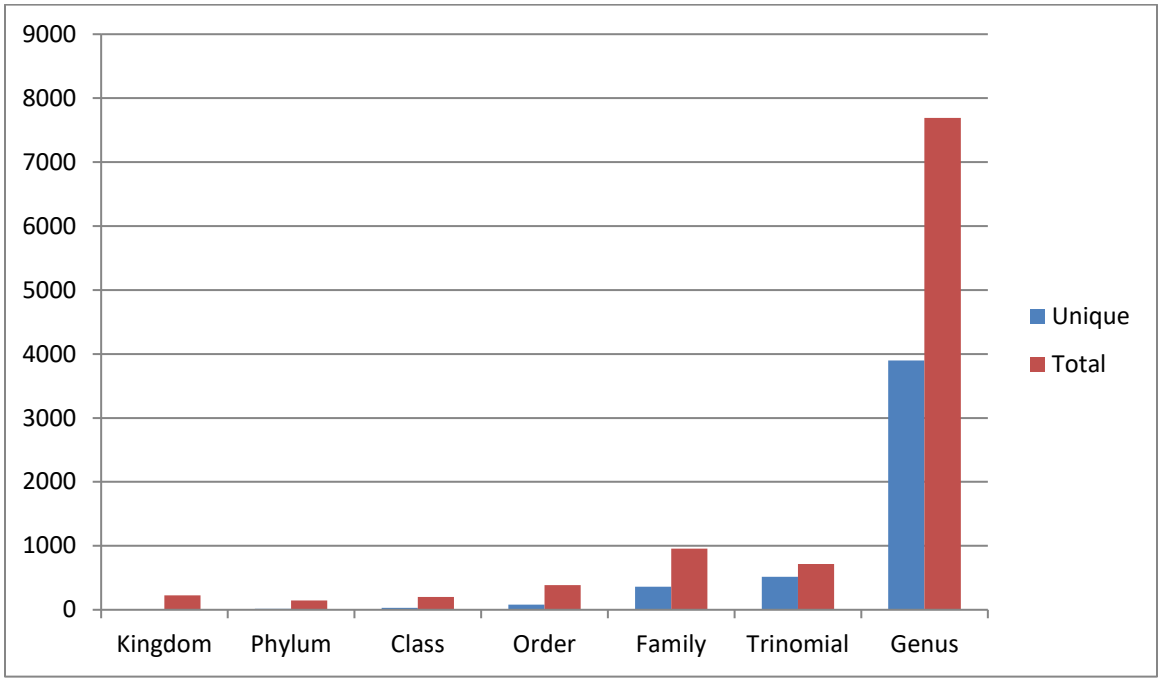


Fig. 10A. Number of tags by level of classification on Flickr

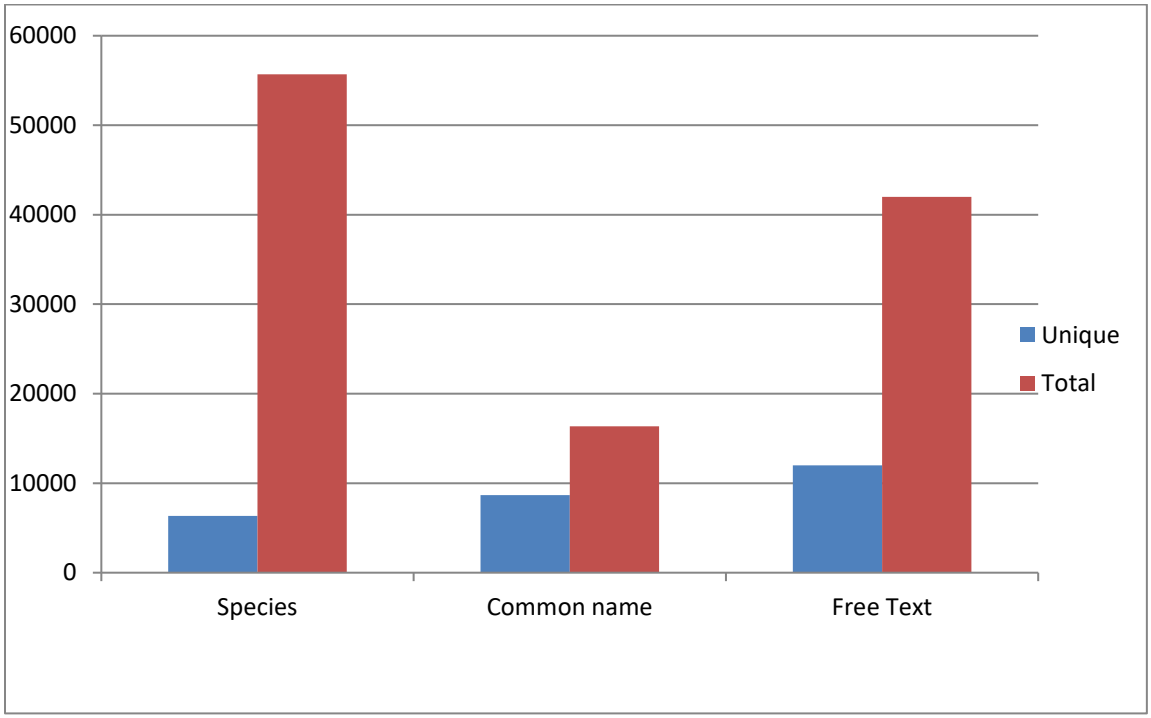


Fig. 10B – BHL Flickr machine tag count for species and common name and free text count

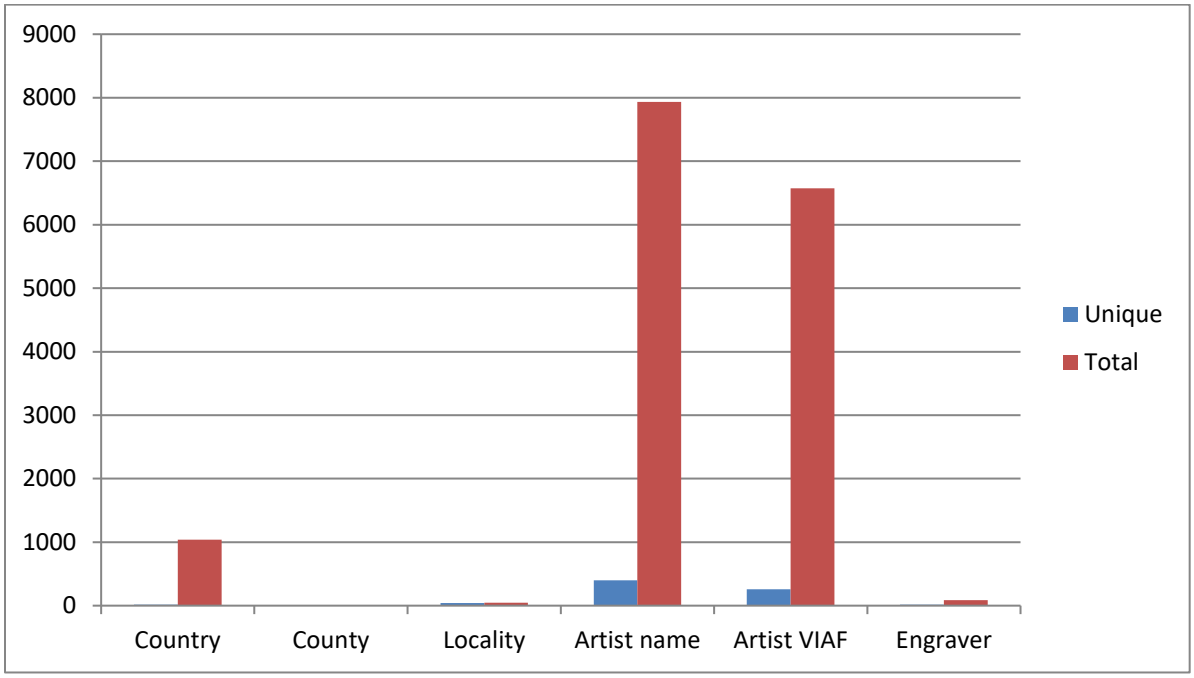


Fig. 10C – BHL Flickr machine tag count for country, county, locality, artist name, artist viaf, and engraver

Artist tagging on Science Gossip is more precise than on BHL Flickr. This is due to the nature of the tagging activities: illustrator/artist names are transcribed when and how they appear to participants on Science Gossip (and may be researched concurrently or subsequently.) For example, Walter Hood Fitch may be listed as W Fitch, W. Fitch, Fitch, del et lith, Pitch, among others. One of the benefits of the tagging process on Science Gossip is that the same keywords may be added by different individuals to each concept, offering validation of useful terms. While a variety of free text terms have been applied to illustrations on BHL Flickr, they are rarely applied multiple times per page.

Several metadata samples may be reviewed in the Google Folder associated with this project.

Metadata production through computer vision

Summary and best practices

Machine learning is an emerging best practice for digital biodiversity, art, and library image collections. Advanced algorithms may be created and used to process images and return information. Some exciting potentials for BHL include creating new descriptive keywords for searching, categorizing images in various ways, and the ability to search by similar images.

Examples within the realm of biodiversity include recognizing observation photographs (and improving the algorithms to do so) through the Lepidoptera app and iNaturalist; specimen recognition, such as a new automatic identification system designed by Hernández-Serna and Luz Fernanda Jiménez-Segura to recognize fish, plant, and butterfly species from Europe and South America with over 90% accuracy; and a successful test case of classifying specimen images at the species level using the 26 most common tree species in Germany. Within the realm of art, researchers at Rutgers University have trained algorithms on database of 80,000 paintings to recognize artists with 60 percent accuracy and recognize style with 45 percent accuracy.

At the current time, digital libraries are in an experimental phase with computer vision. Some projects are led by external professionals with the underlying goals of new ways of exploring image collections and public engagement. Institutions like the Metropolitan Museum of Art, the British Library Labs, and Europeana have placed their datasets online for anyone to use, invite collaborations, hold challenges, and even crowdfund such work. Mario Klingemann's "The Order of Things" project to add meaningful tags to the British Library's Flickr Commons images and create thematic collections as well as new works of art was a celebrated collaboration. It earned him the British Library Labs Creative/ Artistic Award in 2015.

Digital libraries are also experimenting with commercial services and the skills of machine learning experts but have yet to integrate this work into fully-fledged production environments. For example, Cogapp provided a presentation at the 2017 International Image Interoperability Framework conference at the Vatican Library about their experiments across several services. Major commercial computer vision APIs such as Microsoft, Google Cloud Vision, and Clarifai are available. Gaurav Oberoi's comparative research noted similar features such as text recognition, color analysis, grouping/ searching images by visual similarity, and generating descriptive keywords. Combining commercial services and machine learning expertise, the British Library has also invited computer science students to build a prototype for its image collections using "two public image recognition APIs (Alchemy and Imagga) and a bespoke algorithm." The "British Library Machine Learning Experiment" is available at blbigdata.herokuapp.com. Europeana has also built an Image Similarity search prototype as part of the Advanced Service Search and Enhancing Technological Solutions (ASSETS) project.

Experimenting with Clarifai

Term extraction through commercial computer vision API services offers the potential of creating descriptive metadata that is currently created through crowdsourcing activities. Discussing the wealth of illustration content and the nature of the BHL NDSR program with Clarifai, they offered

up to 10 million free uploads as part of their Collaboration program: 5 million using the general model and 5 million using up to 10 custom concepts. [8]

Clarifai offers both advantages and drawbacks in generating and evaluating results. Clarifai offers the ability to upload content, create, and train custom sets through a viewer "app", in addition to interacting with the API. The viewer previews the same content, results, and functionality that may be used in a production environment. Unfortunately, uploading and interacting with the viewer presented significant challenges; glitches and slow speed halted progress entirely. Troubleshooting by uploading smaller batches of content did not resolve this issue. Ultimately, these services are ideally intended to be interacted with as APIs with the results fed into separate user interfaces for either production or evaluation purposes (as demonstrated through Cogapp's experiment, in which the results were indexed through Elasticsearch and visualized through SearchKit.)

An evaluation of the general model was the primary experiment conducted, first, using several book albums from Flickr. For each image, 20 descriptive terms are returned with an estimate of accuracy (rarely below 90%) Examples are included below:

Fig. 11 – Word cloud representing frequent terms for illustrations in *112 superb varieties for market gardeners season of 1926*

[8] The contact person is Liz O'Sullivan, Director of Customer Experience, reached at liz@clarifai.com.

Fig. 12 – Word cloud representing frequent terms for illustrations in *Album de Aves Amazonicas*

Secondly, results for individual images within the following popular categories used on BHL social media were evaluated from a set of random images from Flickr: Algae, Ferns, Fungi & Mosses; Birds; Fish; Flowering Plants; Insects & Spiders; Mammals; Mollusks and Invertebrates; Reptiles & Amphibians; Fossils. General terms relevant to biodiversity appeared at least once or twice within the set of 20 terms returned, such as mushroom/ fungus, bird/ owl, fish, flora, insect, mammal, shell, turtle, and skull/dinosaur. Appendix 3 includes example results from this process.

While these results are very promising for producing these categories as descriptive metadata, the slow rate of uploading (32 images at a time) and limited number of custom concepts (10 whether on the free, standard, or collaboration plan) demonstrate that internal staff work on as large and nuanced a corpus as BHL's illustrations is not the most efficient or ideal approach. If using the general model, BHL might develop a script to eliminate unhelpful terms (such as retro, print, illustration, or antique) and keep terms relevant to biodiversity after results are received. Pending the skillset and resources available, BHL may wish to pursue additional categorization experiments by interacting with the API directly and tracking results, or sharing them through a search prototype. Working independently may be cost-effective, but not efficient: BHL may elect to use all 5 million custom uploads and 5 million general uploads available. Regular pricing for Clarifai includes 10,000 free inputs (number of items uploaded) per month, plus $0.80 per extra 1,000 inputs per month. There are also 5,000 free operations per month (any action such as tagging or searching tags), with $1.20 per additional 1,000 operations. With 4 million illustrations, this would cost approximately $8,000 with just one operation per input.

Recommendations

The recommendations that follow discuss functionality of the BHL Portal and improvements regarding figures and descriptions based upon identified user needs; staffing, strategies, and technology for supporting illustration crowdsourcing based upon interviews and best practices; and leveraging of computer vision through collaborative approaches based upon experiments and industry research. Further interface design and behavioral user research are also recommended, with concomitant staffing and management approaches.

Recommendations for improved BHL Portal functionality

User stories were created as a best practice in the technology development process. According to Germaine Satia, in her article, "How to Write a Painless User Story," "User stories are meant to be living artifacts that can be updated and modified as a project evolves." More stories can be created pending further user research, such as testing of prototypes, or broken down into smaller achievable components. Satia states that "User stories explain the problem or need that the user will solve through a particular piece of product functionality and its value, but they don't explain how it should be addressed from a technical or design perspective." User stories permit the project team to focus on needs, explore ideas to meet them, and facilitate conversations about their technical feasibility, the design around them, and their priority. The following recommendations were developed through brainstorming and reference to best practices in digital libraries.

User stories

- As a taxonomist, I want to zoom in on a selected illustration page so that I can compare it to a specimen.
- As an artist, I want to view/zoom in on illustrations in high resolution for inspiration.
- As a historian of science or a taxonomist, I want to download illustrations for my papers and presentations.
- As a taxonomist, I want to download illustrations so that I can save them in my personal library.
- As an artist, I want to download illustrations in high resolution for collages.
- As a taxonomist, I want to download high resolution illustrations so that I can print them as artwork or for personal enjoyment.

Recommendations

I. Set up a IIIF server and viewer to enable fast, rich, zoom and pan delivery of images, in addition to manipulation of size, scale, region of interest, rotation, quality and format. The International Image Interoperability Framework (IIIF) is a set of shared application programming interface (API) specifications for interoperable functionality in digital image repositories. Tristan Roddis, Director of Web Development at CogApp, explains the technical skills and resources required to implement the image API and presentation API in his article, "Everything You Always Wanted to Know About IIIF* (*But Were Afraid to Ask)" as skills in systems administration and JSON, and a compatible server.

II. Integrate user accounts to save illustrations. This is a feature offered by Google and DPLA, among other digital libraries, and would overcome the need to download and store illustrations on a personal computer for future reference. IIIF is compatible with various authentication/ access control systems that BHL may choose from as part of this service.

III. While BHL offers connections to the citation manager, Mendeley, items can only be cited at the book level. For citing illustrations for publications or presentations, it would be highly useful to offer this service at the page level. IIIF APIs also offer stable URIs that may be shared.

IV. Offer additional download features, including the ability to download all illustrations in a book, with the option of doing this in high resolution. Eliminating the multiple steps and URL editing required to download individual high resolution images is also recommended.

Recommendations related to figures and descriptions

User story

- As a taxonomist, I want access to illustrations separated from the species description I am looking at so that I can better understand the description. (Note that plates separated from their reference in the text or not bound within the text, compounded by incorrect or missing plate metadata in book navigation are problematic.)

Recommendations

I. Create a "View illustrations only" button within the book navigation panel.

II. Ask users to leave notes or links to one another regarding the location of plates through the future annotation feature in BHL version 2. This information may feed into a future project to create direct hyperlinks between figures and their reference in the text.

III. Collaborate with other creators of biodiversity data such as checklist producers. For example, a priority for the World Register of Marine Species (WORMS) is linking original descriptions to taxa. Some editors also provide page numbers and references to illustrations where the taxon in question was first described and illustrated. Integrating this data into BHL through technical means or tying into this existing workflow through the upcoming annotation feature of BHL version 2 would benefit others referring to BHL directly and strengthen BHL's ties to biodiversity informatics community.

Crowdsourcing recommendations

As stated by Trevor Owens, in "Making Crowdsourcing Compatible with the Missions and Values of Cultural Heritage Organizations," Crowdsourcing is "not limited to being an instrument which enables us to better deliver content to end users...It is about providing meaningful ways for the public to enhance collections while more deeply engaging and exploring them." BHL has the opportunity to make crowdsourcing of illustrations a core activity, and to consider potential platforms, topics, and management practices. This will require at minimum that a BHL staff person adopt part time crowdsourcing project management duties. He or she might work with taxonomists or historians of science with particular research questions to design projects with an endpoint, a statement as to how the metadata might improve the BHL collection, increase the visibility of BHL content, or provide useful data for research. The Zooniverse Project Builder is an excellent candidate for this use. This also presents the opportunity for BHL staff across institutions to get involved in sustaining projects, with the benefit of serving as hands on training in crowdsourcing.

While there aren't drawbacks to continuing to encourage volunteers to tag BHL's Flickr collections or within BHL's upcoming annotation tool in the same mode, a more sustainable approach requires dedicated project management, a more specific and time-limited purpose, and cultivating a sense of community. Using platforms purpose built for crowdsourcing descriptive metadata and finding ways to tap into existing communities like those found on Zooniverse would be helpful towards meeting these goals. There may also be opportunities to collaborate with the creators of existing species observation apps to share BHL illustrations of interest through these platforms, or to invite citizen scientists to participate in limited term taxonomic tagging of BHL illustrations, akin to observation "bio-blitzes." Adding descriptive text to images in the vein of the American Museum of Natural History's Project Describe, based upon the Scribe platform, would provide a low barrier to entry for participants and a wealth of searchable terms for general audiences of the BHL Portal.

Related to this, BHL is strongly encouraged to work together with the Constructing Scientific Communities group and Zooniverse to develop a plan and timeline for reinvigorating Science Gossip. Particularly as the lead Science Gossip research team member has taken on new professional obligations requiring a great deal of attention, another ConSciCom member or someone else in their professional network might consider this role. Additional suggestions provided by the Science Gossip moderators include:

- A mass e-mail to Zooniverse volunteers setting a challenge to complete a publication (particularly a highly illustrated one)
- A mass e-mail with a challenge to look out for certain types of images that are related to a researcher's interest, such as the work of a particular engraver
- Continuing to share research publications related to (or ideally based upon) the discoveries in Science Gossip
- Progress updates on completing publications
- Sharing interesting images or facts from Science Gossip via social media

Coordinated efforts and engagement with the Wikipedia community would build upon the activities and knowledge of existing BHL illustration crowdsourcing participants and further expand the reach of and knowledge production surrounding BHL content through illustrations. At the current time, participants work on an individual basis to create articles integrating illustrations from Flickr, or research on historic illustrators. While there has been interest in coordinated efforts on Science Gossip, progress has been hindered by a lack of a distinct plan. For example, as described by dedicated BHL volunteer, Michelle Marshall, at minimum, BHL could invite Wiki contributors to tag BHL Flickr illustrations with Wiki friendly machine tags so that they could be ingested into Wiki Commons. BHL staff members may also batch upload content using the GLAMwiki Toolset.

This work would ideally be established and coordinated by a Wikipedian in Residence. The ultimate goal of hosting a Wikipedian in Residence is to build relationships between the host organization and the Wikipedia community. The impact of the Resident at the Naturalis Biodiversity Center in Leiden may serve as an example. Genus and species descriptions were added to thousands of images on Wikimedia Commons, benefitting a wide community. While there is the possibility of publishing such images and metadata in the BHL portal (a potential stated in Europeana's Content Strategy regarding Metadata and Content from WikiMedia Foundation) the primary benefit is wider engagement. BHL might also follow the example of Europeana in developing online exhibitions at the conclusion of limited-time initiatives.

As the popularity of and press surrounding Science Gossip, BHL's Flickr collections, and open access initiatives like Wikipedians in Residence attest, elevated brand recognition and reputation are additional benefits to crowdsourcing programs. This recognition and support among the public may even build into gaining additional sources of funding for BHL from independent individuals.

Computer vision recommendations

The most effective route for developing additional metadata for searching for illustrations in the BHL Portal is conversation and collaboration with external subject matter and technology experts who can leverage computer vision. This would prove to be more efficient for extracting terms of interest and categorizing images at a more nuanced level. This would also help support the creation of a mutual vision for the potentials and expected outcomes of such work. Further ideas suggested by Rod Page (evolutionary biologist at the University of Glasgow, Scotland and creator of Biostor) include: "how many specimens are included, what the specimen orientation is, what part of the organism is being displayed, the country depicted in a map." With this expertise, the existing crowdsourced metadata may be leveraged in training, an extremely powerful step. In an ideal scenario, human effort and automated description will be combined, allowing immediate improvement/ training of the algorithm. All of these more advanced stages will require the expertise of others beyond the current capacity within BHL to simply extract terms or categories and use them for searching directly.

To explore a related option, I consulted with a local media-focused technology company, Rampant Interactive. Their proposed solution for illustrations lacking metadata is the development of a game or app. The first stage would be using a commercial API such as Amazon Web Services' Rekognition for general classification of all images, at the approximate cost listed for Clarifai regular pricing. Users of the game would verify or correct general tags, with the option to add inscription information or other notes. Stats and leaderboards would be included in the game, to encourage use. Completing more difficult tasks would result in higher points. Their typical pricing for apps or games is $10,000-30,000. This metadata could be integrated into a centralized database supporting searching through the BHL Portal. While Rampant Interactive currently has minimal expertise in training machine learning algorithms, we noted the continuing improvement in quality, pricing, and availability of commercial computer vision services.

Future Work and Next Steps

The goal of access to illustrations through the BHL Portal through integration of existing crowdsourced metadata and potential future metadata will ideally be achieved through multiple cycles of user research, designing, and testing/ prototyping, and eventually building technology. The user experience/ agile development cycle is meant to provide immediate validation for the usability of a technology product and provide value to users as quickly as possible. This would require BHL to develop an integrated approach to user story development, technical architecture and front-end website development, and testing with users. The Europeana ASSETS (Advanced Service Search and Enhancing Technological Solutions) project provides an example of tying requirements gathering, design, and testing closely to users. In this case, prototypes mirroring the technical architecture of Europeana were developed for eventual integration into the core Europeana platform.

Staff with skills in scripting to efficiently parse the existing metadata into the upcoming database/ data model, in interface design and user testing, and in web development will be required for future development. A deeper look at the crowdsourced metadata to understand the relationships between terms is recommended as an interim step prior to building a database. Additionally, further internal experimentation with computer vision APIs to assign various categories may be possible pending time and skills in this area, though collaborative projects are more strongly recommended. As explained by project mentor, Trish Rose-Sandler, BHL's technical development during the course of this project has not included planning for integration of crowdsourced (or machine-generated) illustration metadata into the technical architecture of BHL. Rather, this will be part of BHL's upcoming work on BHL version 2 to develop a more flexible data model that accommodates digital objects rather than just digitized books.

Illustration access through the BHL Portal should at minimum meet standard practices for presenting image content. As stated by Menard and Smithglass in their article, "Digital image access: an exploration of the best practices of online resources," the major image search engines all offer similar functionalities for image retrieval, i.e. keyword-based query resulting in a grid of image thumbnails. Various query refinements are available, including image size, aspect ratio, color, and different types of content." They found that digital library collections either presented natural language keyword search or advanced search with pre-defined categories using drop-down menus.

The following design aspirations provided by Grace Costantino are also high priority recommendations for illustration access:

- Designate the ability to search for illustrations vs. books on homepage
- Provide illustration/ page level search results
- Utilize the bounding boxes that have been created by crowdsourcing participants around illustrations on pages (Bounding boxes may be used in showing search results, or shown when browsing through illustration pages within a book)

A rough mockup has been developed according to the above standard practices for presenting image content, reviews of DPLA, Europeana, and Trove, searching best practices, a review of crowdsourced illustration metadata and search fields of importance to taxonomists, historians of science, and artists. Nick Babich's articles, "Best Practices for Search" and "Best Practices for Search Results" are excellent resources for search interaction and results, and informed these suggestions. As with all design ideas, this mockup may be used to prompt a conversation regarding its technical feasibility and invite team members to discuss additional ideas that are more cost effective or innovative.

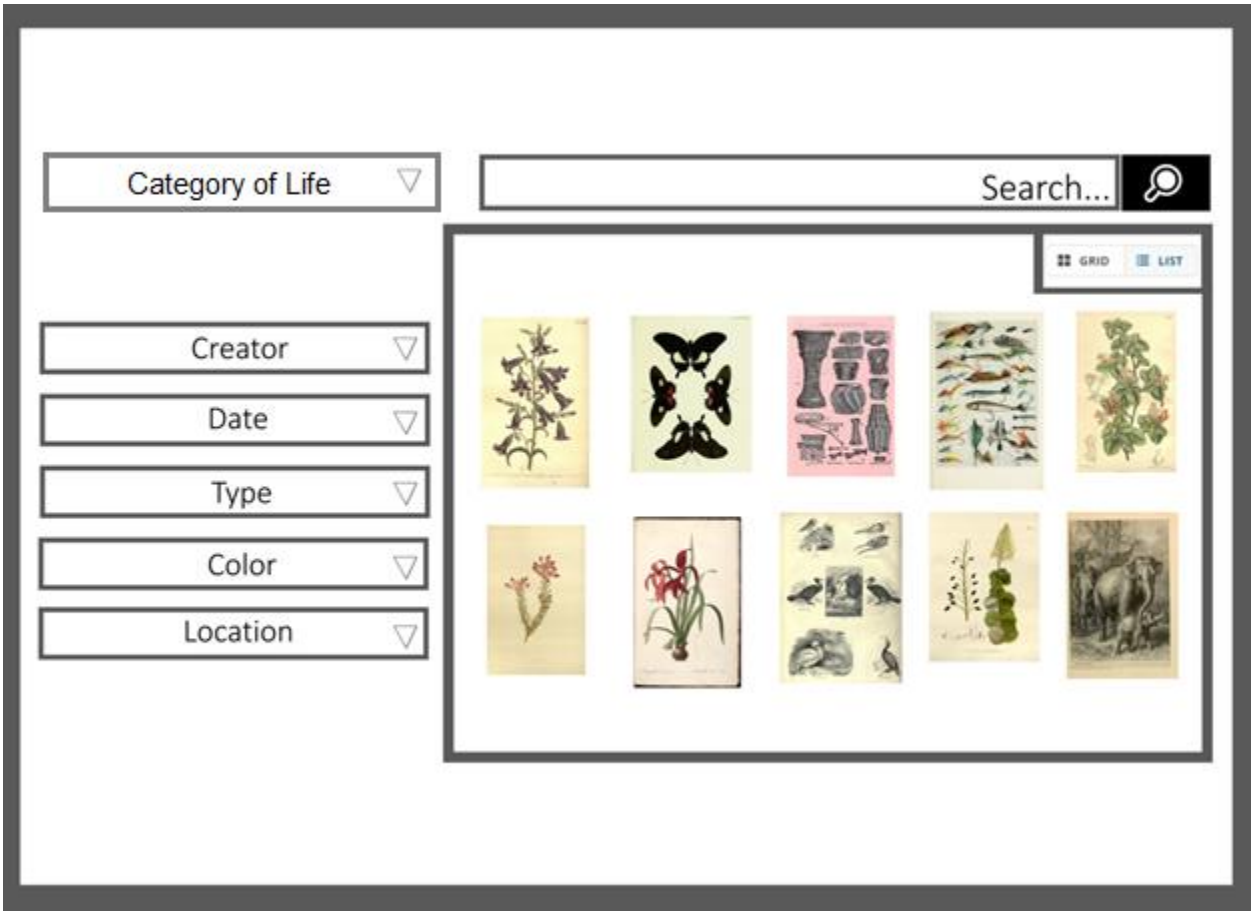


Fig. 13 – Basic search and browse mockup

Features recommended in the mockup:

1. All crowdsourced metadata terms will be available for searching through the search bar, including free text terms from BHL's Flickr account, keywords from Science Gossip, inscriptions, scientific, and common names and the faceted terms on the left, other than color and date. A grid or list view is available for results.

2. Including scientific names, related synonyms, and common names in the search process and search results for illustrations would expand access among both taxonomists and the general public and build upon the goals of taxonomic tagging on BHL's Flickr account. BHL currently runs all text through Global Names Recognition and Discovery, a scientific name discovery service based on algorithms. While this same strategy could also be used for crowdsourced metadata, BHL is

urged to consider validating and expanding names through the Catalog of Life -- the most "comprehensive and authoritative global index of species currently available." It consists of a single integrated species checklist and taxonomic hierarchy, and provides scientific, common names and synonyms. In the meantime, it is hoped that BHL will use the well-researched current names, synonyms, and common names applied to BHL Flickr images to expand its search functionality.

3. Search auto-suggestions are a standard practice online and have been implemented by Europeana. This encompasses spelling auto-corrections, recognition of root words, and predictive text. All crowdsourced metadata would need to be compared to controlled vocabularies such as Thesaurus of Geographic Names for locations and the Virtual International Authority File for artists names. The metadata would be pre-processed prior to being made available to search by matching terms in the metadata to identifiers in these vocabularies.

4. Location information may be pulled from the crowdsourced metadata and as well as geographic subject terms for the book containing the illustration. Similarly, creator information may be pulled primarily from artist name machine tags and artist VIAF machine tags, or contributor tags. In cases where no creator has been assigned to an illustration through crowdsourcing, BHL may consider providing this information from book metadata where possible. It would be valuable to provide links to Wikipedia articles created through BHL's illustration crowdsourcing efforts next to creator names in the faceted search.

5. Faceting by date and image type are standard expectations across digital libraries like DPLA, Europeana, and Trove, and can be easily implemented through the metadata available for all crowdsourced illustrations. Date is a particularly important access point for users distinguishing between historic and more modern illustrations for specimen identification, or those seeking out historic materials for their particular research purposes.

6. Color browsing is a standard image searching practice, as exemplified by Google Images. Among digital libraries, this is available on Europeana and as an app on DPLA. This may be especially appealing for artmaking.

7. It is also important to include the option of wildcards for free text searching of scientific names. This may be modelled after the search mechanism of the Plant List. According to one interviewee, taxonomists will be persistent in searching inscriptions or assigned names to find illustrations of interest.

8. Beyond this, the option to browse for illustrations by readily available popular taxonomic categories, rather than searching directly, would be useful for general audiences. The categories of life used for BHL social media may be used for this purpose (shown in the mockup as a dropdown menu.) They are Algae, Ferns, Fungi & Mosses; Birds; Fish; Flowering Plants; Insects & Spiders; Mammals; Mollusks and Invertebrates; Reptiles & Amphibians; Fossils.

Future decisions regarding illustration searching and browsing designs for the BHL Portal may also be informed by:

- A review of image platforms serving artistic, taxonomic, citizen science, and historic communities, as well as general audiences to determine common (or divergent) expectations for design, interaction/task flows, and ways of exploring related content or interacting with content. Instagram, the UK Medical Heritage Library project's "wall of images", the "Lost Visions: Retrieving the Visual Element of Printed Books from the Nineteenth Century" Illustration archive, Old Maps Online, the Getty Museum Collections, the Smithsonian Institution Entomology Department Illustration Archive and iNaturalist's photo browser are suggested as starting points. BHL is also encouraged to stay abreast of developments in the Biodiversity Literature Repository image searching interface. The broad appeal of illustrations will need to be balanced with the research needs of taxonomists in determining design priorities.
- Additional user research, particularly among educators, citizen scientists, historians, and artists. Ideally, a sponsor will be designated from among these groups who can develop support among his or her network to review and test out future wireframes/mockups and eventual prototypes. For participants, this will be a tangible method of contributing to illustration access. For BHL, this will provide clear rationale for design ideas before they go live. Moreover, united support and a vision for this work will increase the likelihood of its achievement, potentially leading to additional funding or collaboration with additional technologists to achieve it.

Conclusion

This project has provided foundational research into the goals of taxonomists, historians of science, and artists/ illustrators as users of natural history illustrations. According to Jakob Nielsen in his article, "Agile Development Projects and Usability," this is an important predecessor to product development; it is a way of understanding these audiences and BHL's unique opportunity to serve them. This research has been combined with digital library best practices and brainstorming to recommend enhancements to the BHL Portal.

A review of the existing crowdsourced metadata and requested search terms, design aspirations from product stakeholders, and common image delivery and searching standards were the primary factors considered in suggesting future interface modifications for the BHL Portal.

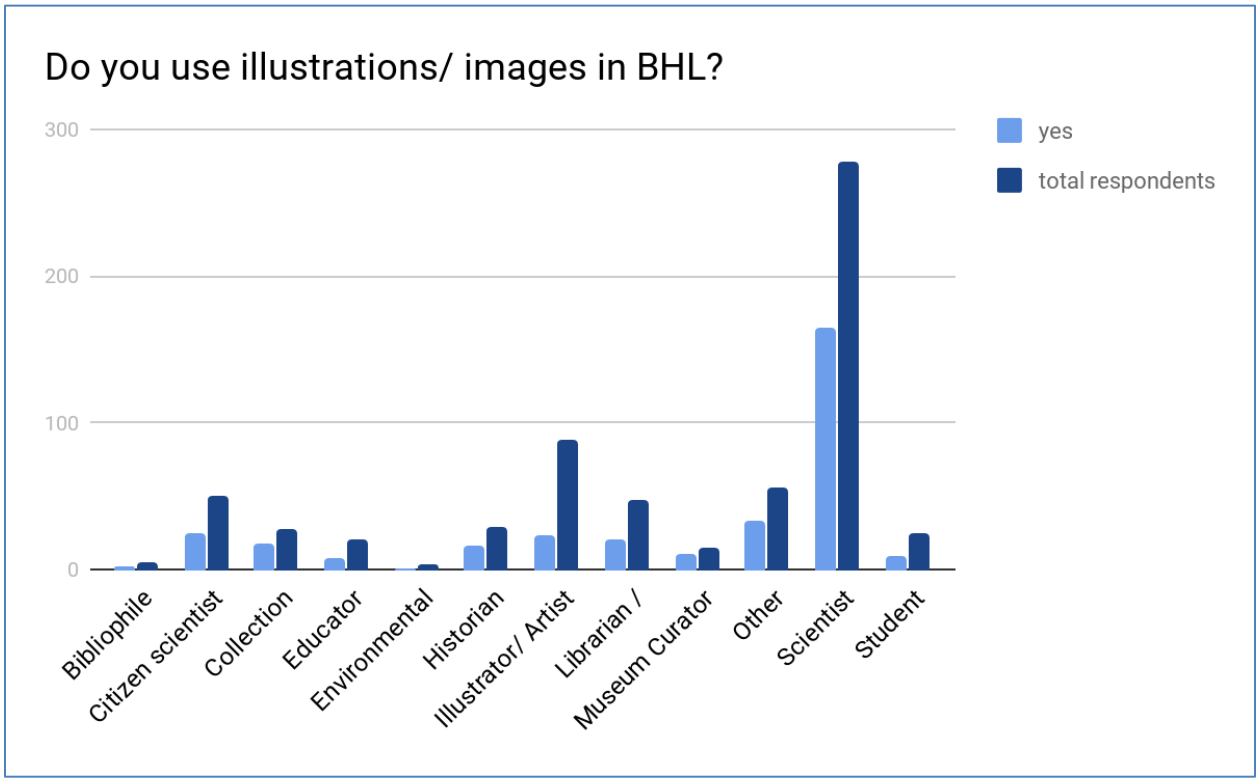
Experimentation with a commercial computer vision service revealed the ability to identify basic categories of life. A route for combining such results with crowdsourced corrections and additional notes was explored. Ultimately, it is suggested that BHL pursue collaborations with external subject matter and technology experts who can leverage computer vision in more robust ways to develop search terms for the BHL Portal. This will prove more efficient and will likely build excitement and support for this work.

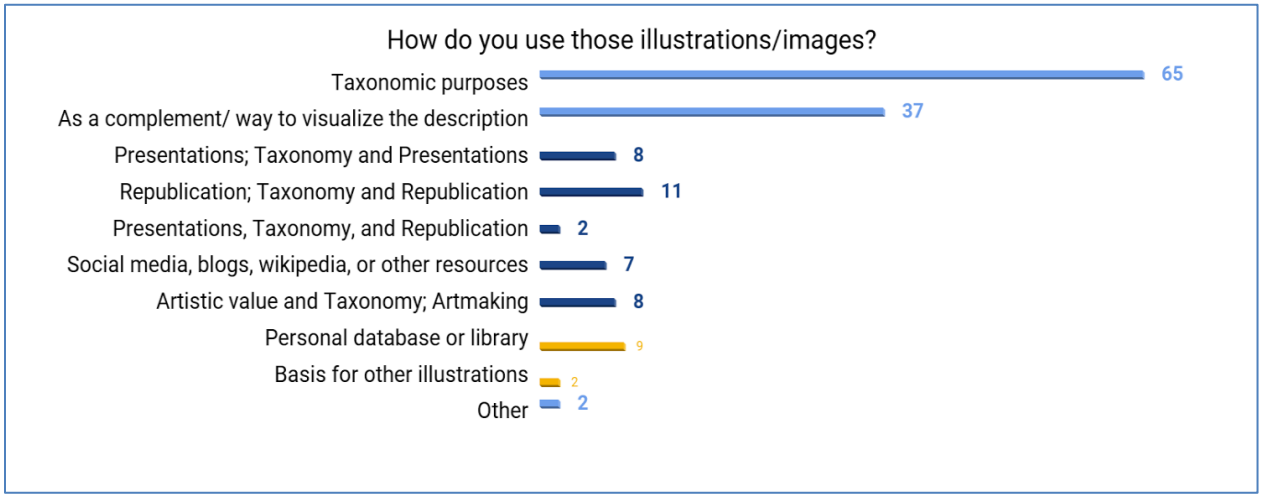
After a review of the successes and challenges of existing illustration crowdsourcing, BHL has been encouraged to continue to pursue illustration-based crowdsourcing with stronger project management and a clearer statement of purpose. Many types of crowdsourcing are possible, such as taxonomic research and tagging, free text tagging, corrections of automated descriptions through a game, and Wikipedia editing. Applying project management and system design best practices, considering suggested external platforms, and working with researchers in determining project topics may help scope future work.

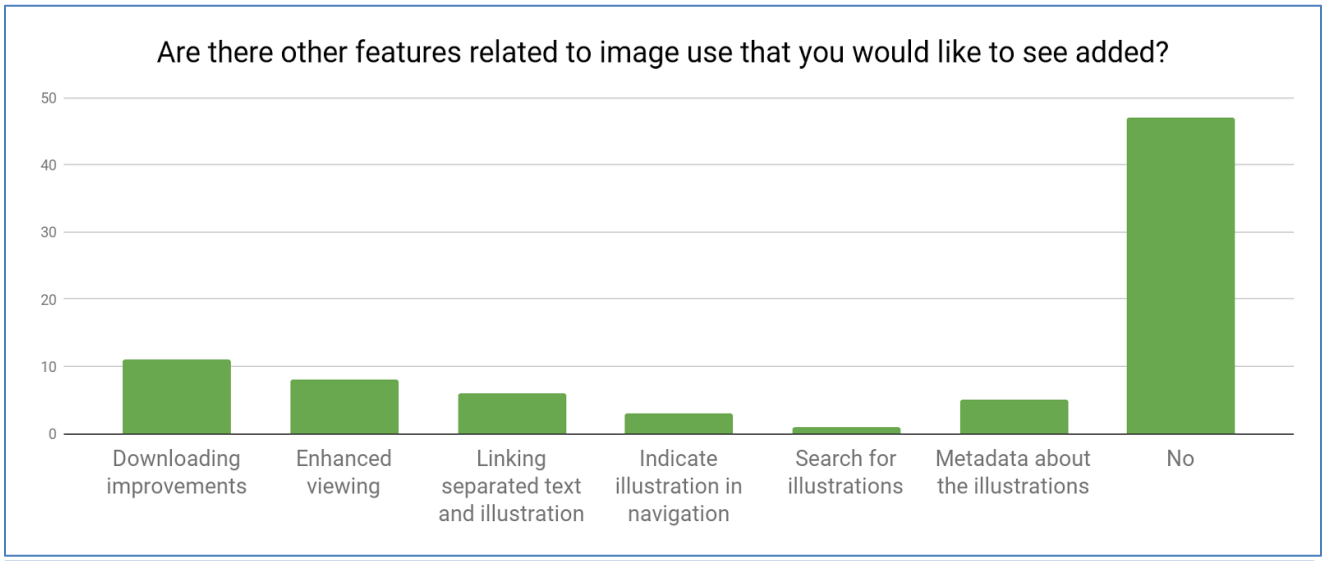
Lastly, approaches towards additional interface design and behavioral user research have been recommended.

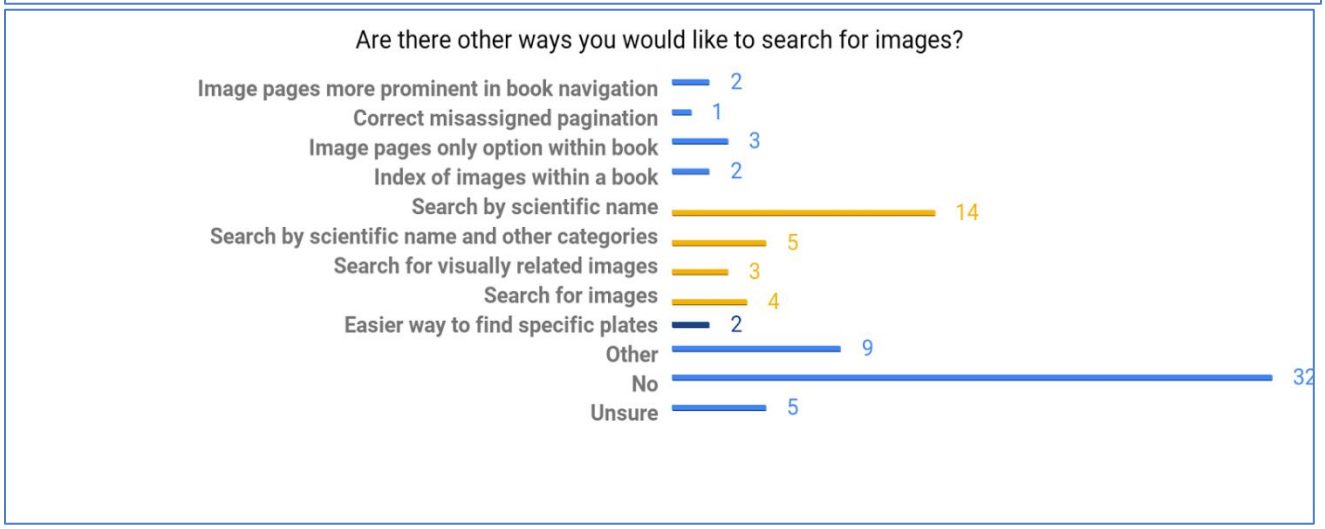
Appendices

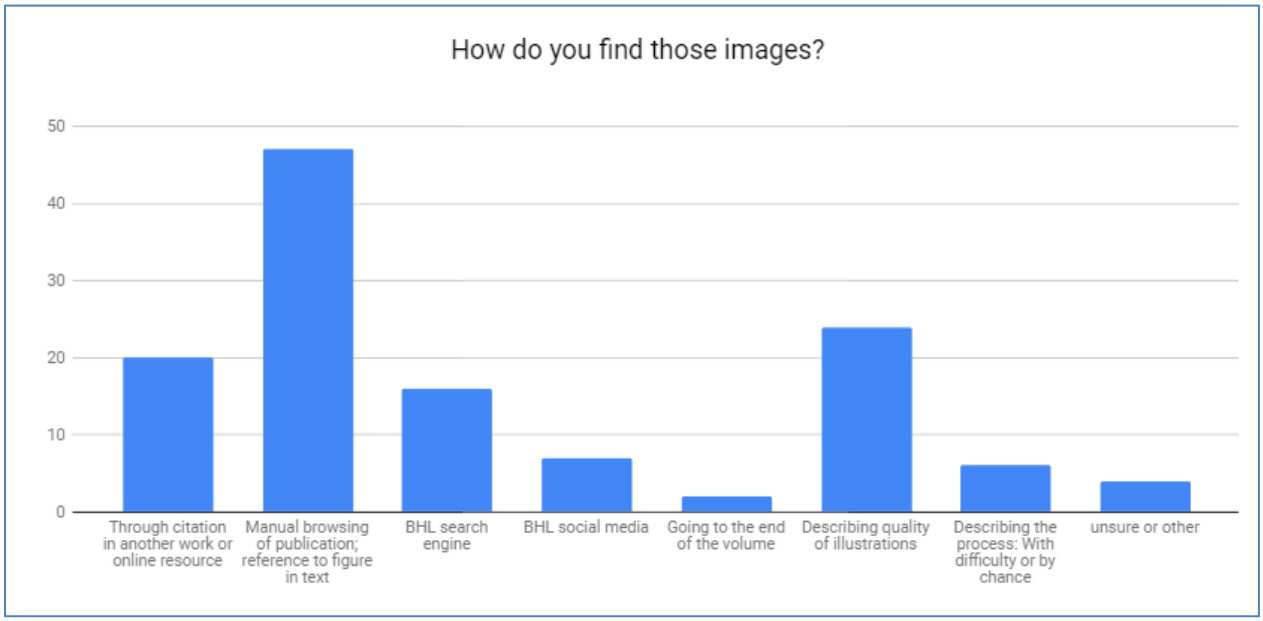
Appendix 1 - NDSR individual users survey, illustration related responses for scientists






Are there other features related to image use that you would like to see added?
50
40
30
20
10
0
Downloading improvements
Enhanced viewing
Linking separated text and illustration
Indicate illustration in navigation
Search for illustrations
Metadata about the illustrations
No


Are there other ways you would like to search for images?
Image pages more prominent in book navigation 2
Correct misassigned pagination 1
Image pages only option within book 3
Index of images within a book 2
Search by scientific name 14
Search by scientific name and other categories 5
Search for visually related images 3
Search for images 4
Easier way to find specific plates 2
Other 9
No 32
Unsure 5


How do you find those images?
50
40
30
20
10
0
Through citation in another work or online resource
Manual browsing of publication; reference to figure in text
BHL search engine
BHL social media
Going to the end of the volume
Describing quality of illustrations
Describing the process: With difficulty or by chance
unsure or other

Appendix 2 - Crowdsourcing interview questions

How did you find out about BHL Flickr or SG? How do you find out about other citizen science or crowdsourcing opportunities?

While BHL directly serves people in the biodiversity sciences, and citizen scientists, bibliophiles, artists, historians, educators gain from its content and programs. Do you consider yourself a member of any of these groups? Is your work on Science Gossip or BHL Flickr related to your personal interests or professional role?

What motivates you to contribute to Science Gossip/ BHL Flickr?

Do you contribute to any other crowdsourcing initiatives? Which, and are they related to your work on SG or BHL Flickr?

How do you feel about BHL using your metadata tagging work to enhance access to illustrations through search and browse functionality in the BHL digital library? Is this a motivator for your work?

Do you seek out natural history illustrations outside of Science Gossip/ BHL Flickr? Why? Can you tell me about the resources you use and about your process?

Do you use or share the same illustrations, metadata from SG or BHL Flickr outside of your work on these platforms? How interested are you in doing this, or in allowing others to find new uses?

I'd like to gauge how well BHL Flickr/ SG fits supports your participation, according to some elements that researchers such as Mia Ridge, Donelle McKinley, and others have found important.

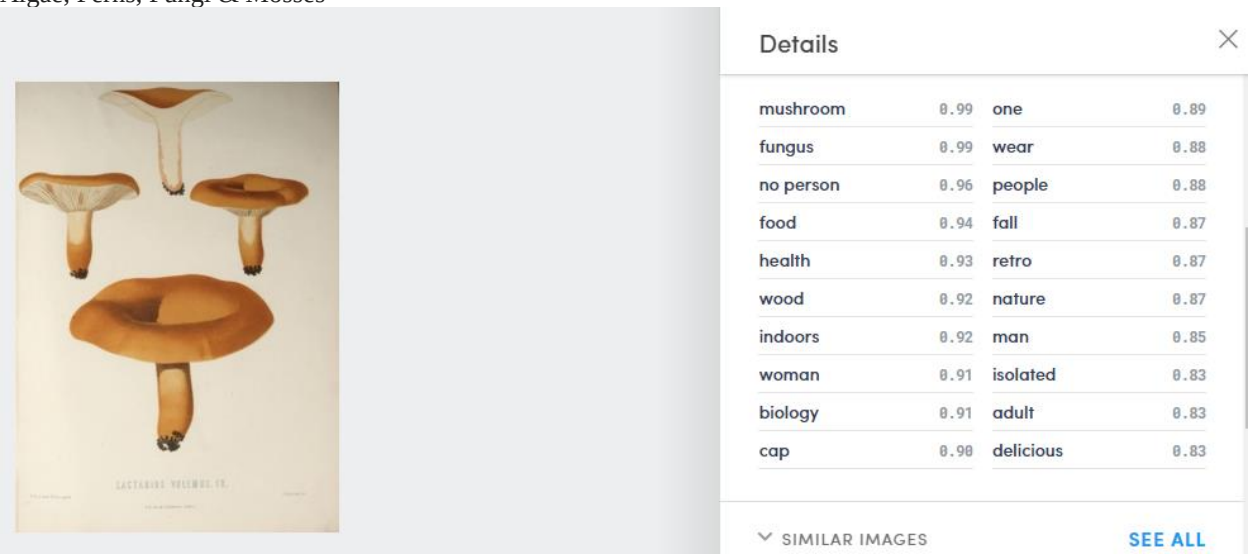
- How clear are the task instructions?
- How well designed is the process of working on SG or BHL Flickr?
- Do you consider learning and skill development a part of your work? How?
- How informed do you feel about the goals of SG/ BHL Flickr and progress you or the community makes?
- How strong of a sense of community do you find on SG or BHL Flickr?

Are there any other overall strengths or weaknesses you'd like to discuss?

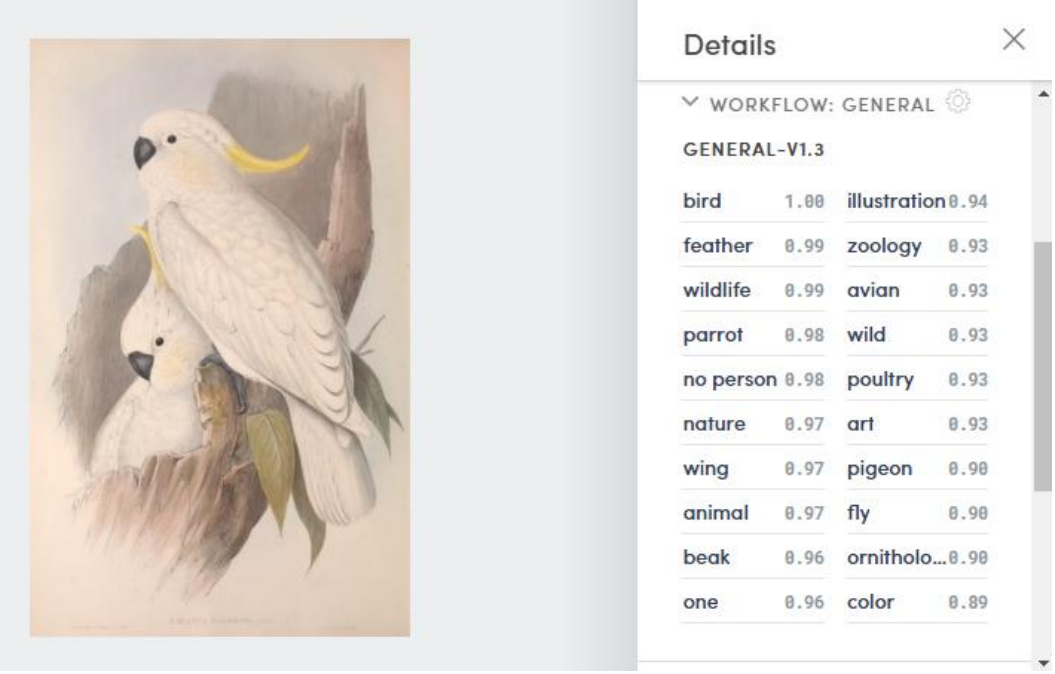
Are you interested in future crowdsourcing initiatives using BHL illustrations? Any recommendations for content?

Appendix 3 - Clarifai general model results

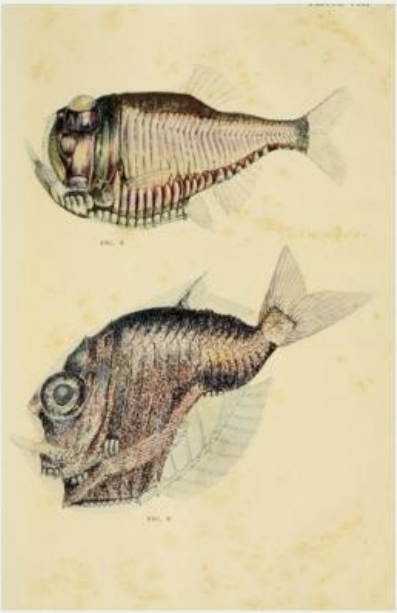
Algae, Ferns, Fungi & Mosses



Birds



Fish



Details

WORKFLOW: GENERAL

GENERAL-V1.3

fish	1.00	art	0.92
illustration	0.98	biology	0.92
one	0.97	seafood	0.91
print	0.97	water	0.91
no person	0.95	engraving	0.90
nature	0.95	desktop	0.89
sea	0.94	animal	0.88
freshwater	0.93	underwater	0.88
painting	0.93	color	0.87
fin	0.93	tail	0.87

Flowering Plants

Details

WORKFLOW: GENERAL

GENERAL-V1.3

flora	0.99	color	0.93
leaf	0.99	art	0.92
nature	0.99	closeup	0.91
flower	0.98	growth	0.91
no person	0.98	beautiful	0.90
garden	0.97	illustration	0.90
summer	0.96	freshness	0.89
floral	0.96	botanical	0.89
desktop	0.96	blooming	0.89
decoration	0.93	bright	0.87

Insects & Spiders

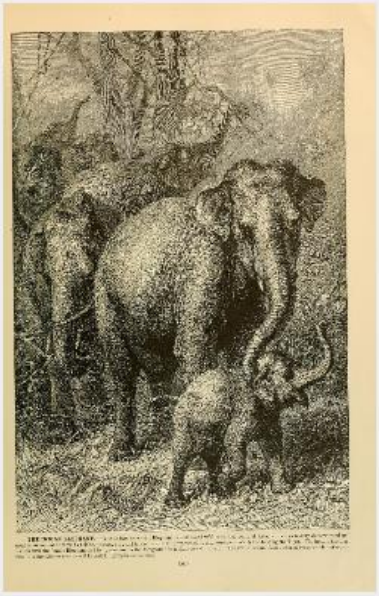
Details

WORKFLOW: GENERAL

GENERAL-V1.3

dragonfly	1.00	mosquito	0.98
insect	1.00	delicate	0.95
wing	1.00	fragility	0.94
fly	1.00	thorax	0.94
nature	0.99	wild	0.92
damselfly	0.99	flora	0.91
Odonata	0.99	garden	0.91
invertebrate	0.99	net	0.91
animal	0.98	tail	0.91
wildlife	0.98	closeup	0.90

Mammals



Details

GENERAL-V1.3

illustration	1.00	elephant	0.96
print	1.00	paper	0.95
vintage	0.99	etching	0.94
art	0.99	mammal	0.94
old	0.99	one	0.93
engraving	0.99	portrait	0.93
antique	0.98	man	0.93
people	0.98	ancient	0.93
visuals	0.98	animal	0.92
retro	0.97	painting	0.92

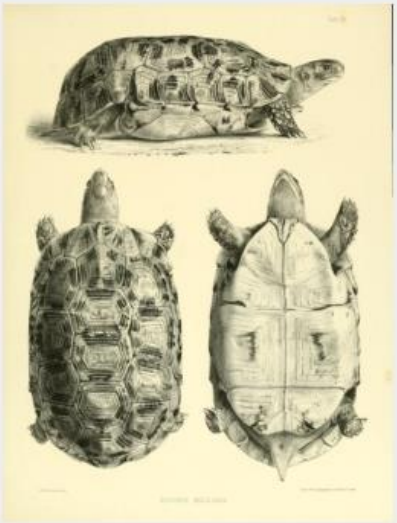
Mollusks and Invertebrates

WORKFLOW: GENERAL

GENERAL-V1.3

illustration	0.99	animal	0.93
art	0.97	disjunct	0.92
decoration	0.97	desktop	0.92
seashell	0.96	shape	0.92
biology	0.96	ancient	0.91
nature	0.95	symbol	0.90
design	0.94	souvenir	0.90
shellfish	0.94	conch	0.89
no person	0.94	chalk out	0.88
shell	0.94	vintage	0.88

Reptiles & Amphibians



WORKFLOW: GENERAL

GENERAL-V1.3

turtle	1.00	old	0.98
tortoise	1.00	biology	0.97
armor	1.00	one	0.97
illustration	1.00	slow	0.97
reptile	0.99	people	0.97
shield	0.99	science	0.96
shell	0.99	art	0.94
print	0.99	wear	0.93
shell (food)	0.98	protection	0.92
engraving	0.98	etching	0.92

Fossils

WORKFLOW: GENERAL

GENERAL-V1.3

illustration	0.99	biology	0.92
print	0.98	ancient	0.90
art	0.98	vintage	0.89
animal	0.97	painting	0.89
two	0.96	dinosaur	0.89
nature	0.96	mammal	0.88
no person	0.96	retro	0.88
museum	0.95	skull	0.88
old	0.94	frame	0.87
group	0.94	science	0.87

Works Cited

"Amazon Rekognition Overview." *Amazon Web Services*. aws.amazon.com/recognition. Accessed Nov. 2017.

"ASSETS: Advanced Service Search and Enhancing Technological Solutions." *Europeana Pro*, 9 Sept. 2014, pro.europeana.eu/project/assets. Accessed Aug. 2017.

Babich, Nick. "Best Practices for Search." *UX Booth*, 2 May 2017, www.uxbooth.com/articles/best-practices-for-search. Accessed July 2017.

Babich, Nick. "Best Practices for Search Results." *UX Planet*, 2 Mar. 2017, uxplanet.org/best-practices-for-search-results-1bbed9d7a311. Accessed July 2017.

Baker, James. "The British Library Machine Learning Experiment." *British Library Digital scholarship blog*, 29 Apr. 2015, blogs.bl.uk/digital-scholarship/2015/04/the-british-library-machine-learning-experiment.html. Accessed Aug. 2017.

Baeten, Elisabeth. Personal interview. 19 July 2017.

Belknap, Geoff. Personal interview. 9 Mar. 2017.

Belknap, Geoff. Personal interview. 14 Mar. 2017.

Belknap, Geoff. Personal interview. 5 May 2017.

BHL Staff. "2011 Usability Test Wiki page." *BHL Staff Wiki*. Accessed Mar. 2017. [no longer accessible as of Mar. 2018.]

"The Code Online" *International Commission on Zoological Nomenclature*, 1999, www.iczn.org/iczn/index.jsp. Accessed June 2017.

Cook, Michael. Personal interview. 21 June 2017.

Costantino, Grace. "Download How To." *Biodiversity Heritage Library Wiki*. 18, Oct. 2017, biodivlib.wikispaces.com/Download+How+To#highres. Accessed Aug. 2017.

Costantino, Grace. Personal interview. 21 Mar. 2017.

Costantino, Grace. Personal interview. 11 June 2017.

Costantino, Grace. Personal interview. 10 Aug. 2017.

Costantino, Grace et al. "The Case of the Mistaken Manakin." *BHL Blog/ News & Notes from BHL Staff*, BHL, 13 Oct. 2016, blog.biodiversitylibrary.org/2016/10/the-case-of-mistaken-manakin.html

Costantino, Grace and Barna Páll-Gergely. "Taxonomic research on Southeast Asian land snails." *BHL Blog/ News & Notes from BHL Staff*, BHL, 15 Oct. 2013, blog.biodiversitylibrary.org/2013/10/taxonomic-research-on-southeast-asian.html. Accessed Mar. 2017.

Costantino, Grace and Christopher Buddle. "BHL: Continuing to Inspire a Love of Natural History." *BHL Blog/ News & Notes from BHL Staff*, BHL, 14 Jan. 2016, blog.biodiversitylibrary.org/2016/01/bhl-continuing-to-inspire-love-of.html. Accessed Mar. 2017.

Costantino, Grace and Karen James. "The Plants of Acadia National Park." *BHL Blog/ News & Notes from BHL Staff*, BHL, 15 Nov. 2013, blog.biodiversitylibrary.org/2013/11/the-plants-of-acadia-national-park.html. Accessed Mar. 2017.

"Drawings." *The J. Paul Getty Museum*, www.getty.edu/art/drawings. Accessed Dec. 2017.

Egloff, Willi, et al. “Copyright and the Use of Images as Biodiversity Data.” *Research Ideas and Outcomes* vol. 3 no. e12502, Mar. 2017, doi:10.1101/087015.

"Entomology Collections Illustration Archive." *Smithsonian Museum of Natural History Department of Entomology*, collections.nmnh.si.edu/search/ento/?preconf=illustrationshttp://collections.mnh.si.edu/search/ento/?preconf=illustrations. Accessed Apr. 2017.

Europeana Foundation. "Europeana Content Strategy." Mar. 2017, pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Content%20Strategy.pdf.

"Expanded access to images in the Biodiversity Literature Repository." *Plazi*. 11 Aug. 2017, plazi.org/news/beitrag/expanded-access-to-images-in-the-biodiversity-literature-repository/b86d1880198c5c7315c2d38ae3a82820. Accessed Nov. 2017.

"GLAM/Naturalis." *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. 31 Aug. 2017, en.wikipedia.org/wiki/Wikipedia:GLAM/Naturalis. Accessed Aug. 2017.

"GLAMwiki Toolset." *Wikimedia Commons, the free media repository*. Wikimedia Foundation, Inc. 5 Jun. 2017, commons.wikimedia.org/wiki/Commons:GLAMwiki_Toolset. Accessed Aug. 2017.

Göldi, Emil August. *Album de Aves Amazonicas*. Impressão do Instituto Polygraphico, 1900-1906, doi.org/10.5962/bhl.title.13945.

Gordea, Sergiu. "An Image Similarity Search for the European Digital Library and Beyond." *Theory and Practice of Digital Libraries*, 22-26, Sept. 2013, Valletta, Malta, edited by

Bolikowski, Łukasz et al., Springer, Cham, 2014, pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Creative/WP2%20%20Infrastructure%20for%20Content%20Reuse/An%20Image%20Similarity%20Search%20for%20the%20European%20Digital%20Library%20and%20Beyond.pdf.

Hammel, Barry. Personal interview. 8 Mar. 2017.

Harris S., et al. "Training in Tropical Plant Identification" *Descriptive Taxonomy: The Foundation of Biodiversity Research, edited by Watson M.F. et al.* The Systematics Association, Special Volume 84, 2015, pp. 160-170.

"Help - The Plant List." *The Plant List: A working list of all plant species*, 2013, www.theplantlist.org/1.1/help. Accessed Oct. 2017.

Hernández-Serna, Andrés, and Luz Fernanda Jiménez-Segura. "Automatic Identification of Species with Neural Networks." Ed. Mark Costello. *PeerJ* vol.2, no. e563, 4 Nov. 2014, doi.org/10.7717/peerj.563

Hindle, Adrian. "Automated image analysis with IIIF." *Cogapp*. 20 Jun. 2017. blog.cogapp.com/automated-image-analysis-with-iiif-6594ff5b2b32

The Illustration Archive. Lost Visions project, Cardiff U., 2015, illustrationarchive.cf.ac.uk.

Illustrated Milkman. Zooniverse, explore.sciencegossip.org/groups

"iNaturalist Computer Vision Explorations." *iNaturalist*, 23, Sept. 2017, inaturalist.org/pages/computer_vision_demo

International Image Interoperability Framework. International Image Interoperability Framework, iiif.io. Accessed May 2017.

James, Shelley. Personal interview. 17 Aug. 2017.

Kapsalis, Effie. "The Impact of Open Access on Galleries, Libraries, Museums, & Archives."

Smithsonian Emerging Leaders Development Program, 27, Apr. 2016, siarchives.si.edu/sites/default/files/pdfs/2016_03_10_OpenCollections_Public.pdf

Kragh, Gitte. "The motivations of volunteers in citizen science." *They walk among us: The rise of citizen science,* special issue of *Environmental Scientist*, vol. 25, no. 2, Aug. 2016, the-ies.org/analysis/motivations-volunteers-citizen.

Kroh, Andreas (World Register of Marine Species Steering Committee Member.) Personal interview. 17 Oct. 2017.

Leachman Siobhan. Personal interview. 5 May 2017.

Leachman Siobhan. Comment on "How do you solve a problem like illustrations?" *BHL NDSR Blog*, 29 Mar. 2017, 6:38 p.m., ndsrbhl.wordpress.com/2017/03/27/how-do-you-solve-a-problem-like-illustrations.

Lintott, Christopher. *Engaging the Public: Best Practices for Crowdsourcing Across the Disciplines*, 6-8 May 2015, College Park, Maryland, Crowd Consortium.

Macháčková, Zuza. Personal interview. 1 July 2017.

"The Machine Vision Algorithm Beating Art Historians at Their Own Game." *MIT Technology Review*, 11 May 2015, www.technologyreview.com/s/537366/the-machine-vision-algorithm-beating-art-historians-at-their-own-game. Accessed May 2017.

Marriot, Hollis. Personal interview. 19 July 2017.

Marshall, Michelle and Siohban Leachman. "Thoughts on Illustrations in BHL." 11 Aug. 2017. *Google Doc* file. docs.google.com/document/d/1Oa_XLa1gVUKGP59z3M8qUAz337o2Do55qJ4xef1hq2I/edit. Accessed 11 Aug. 2017.

Marshall, Michelle. Comment on "How do you solve a problem like illustrations?" *BHL NDSR Blog*, 28 Mar. 2017, 1:00 a.m., ndsrbhl.wordpress.com/2017/03/27/how-do-you-solve-a-problem-like-illustrations.

Marshall, Michelle. Personal interview. 11 July 2017.

Marshall, Michelle. Personal interview. 11 Aug. 2017.

Marshall, Michelle. Personal interview. 16 Sept. 2017.

McKinley, Donelle. Motivations to crowdsource and contribute. *Non-profit Crowd,* nonprofitcrowd.org/crowdsourcing-motivations. Accessed 3 Apr. 2017.

McPherson, Gordon. Personal interview. 2 Mar. 2017.

Menard, Elaine and Margaret Smithglass. "Digital image access: an exploration of the best practices of online resources", *Library Hi Tech*, vol. 32, no. 1, 2014, pp.98-119, doi.org/10.1108/LHT-05-2013-0064.

Miller, James. Personal interview. 20 Mar. 2017.

Nielsen, Jakob. "Agile Development Projects and Usability." *Nielsen Norman Group*, 17 Nov. 2008, www.nngroup.com/articles/agile-development-and-usability. Accessed Sept. 2017.

[Neil]. "Wiki Loves Monuments 2012: Words From The Winner." *Europeana Blog*, 12 Dec. 2012, blog.europeana.eu/2012/12/wiki-loves-monuments-2012-words-from-the-winner. Accessed Oct. 2017.

Oberoi, Gaurav. "Comparing the Top Five Computer Vision APIs." *Goberoi*. 11 Jul. 2016, goberoi.com/comparing-the-top-five-computer-vision-apis-98e3e3d7c647. Accessed July 2017. Accessed May 2017.

Old Maps Online. Klokan Technologies GmbH, Switzerland and The Great Britain Historical GIS Project at the University of Portsmouth, UK, Jan 2013, www.oldmapsonline.org. Accessed Oct. 2017.

“The Order of Things.” *British Library Labs.* labs.bl.uk/The+Order+of+Things. Accessed June 2017.

Owens, Trevor. “Making Crowdsourcing Compatible with the Missions.” *Crowdsourcing our Cultural Heritage*, Edited by Mia Ridge, Ashgate Publishing, Ltd., 2014, 269-280.

Page, Roderic D.M., "Exploring images in the Biodiversity Literature Repository." *iPhylo*. 10 Nov. 2017, iphylo.blogspot.com/2017/11/exploring-images-in-biodiversity.html. Accessed Nov. 2017.

Page, Roderic. Personal interview. 21 Mar. 2017.

Pool, Amy. Personal interview. 6 Mar. 2017.

Pool, Amy. Personal interview. Oct. 2017.

Project Describe. American Museum of Natural History, projectdescribe.amnh.org. Accessed Aug. 2017.

"Rampant Interactive Homepage." *Rampant Interactive*. therampant.com. Accessed July 2017.

Reitmeyer, Mai. Personal interview. 14 June 2017.

Rinaldo, Constance. Personal interview. 11 Sept. 2017.

Roddis, Tristan. "Everything You Always Wanted to Know About IIIF* (*But Were Afraid to Ask)." *Cogapp*, 10 Nov. 2017, blog.cogapp.com/everything-you-always-wanted-to-know-about-iiif-but-were-afraid-to-ask-6f7acc7d56c3. Accessed Nov. 2017.

Rohrer, Christian. "When to Use Which User-Experience Research Methods." *Nielsen Norman Group*, 12 Oct. 2014, www.nngroup.com/articles/which-ux-research-methods. Accessed June 2017.

Rose-Sandler, Trish. "We need your Help to Tag over a Million Biodiversity Images in Flickr." *News and Notes from the BHL Staff*, 29 Jan. 2015, http://blog.biodiversitylibrary.org/2015/01/hundreds-of-thousands-of-bhl-images.html. Accessed Feb. 2017.

Rose-Sandler, Trish and Chris Freeland. "Art of Life." *BHL Wiki*, 6 Jan. 2016, biodivlib.wikispaces.com/Art+of+Life. Accessed Jan. 2017.

Salick, Jan. Personal interview. 1 Mar. 2017.

Satia, Germaine. "How to Write A Painless User Story" *Studio by UXPin*. www.uxpin.com/studio/ux-design/how-to-write-a-painless-user-story-for-agile-ux-teams. Accessed Aug. 2017.

Smocovitis, Betty. Personal interview. 20 Mar. 2017.

Tangerini, Alice. Personal interview. 8 May 2017.

Tann, John. Personal interview. 19 July 2017.

Trei, Kelli. "The Impact of the Biodiversity Heritage Library (BHL) on Scientific Research." DLF (Digital Library Federation) Forum 2016, 7-9 Nov. 2016, Milwaukee, WI.

"UK Medical Heritage Library." *Jisc*, www.jisc.ac.uk/rd/projects/uk-medical-heritage-library. Accessed Sept. 2017.

Unger, Jakob et al. “Computer vision applied to herbarium specimens of German trees: testing the future utility of the millions of herbarium specimen images for automated identification.” *BMC Evolutionary Biology,* vol. 16, no. 248, 16 Nov. 2016. doi.org/10.1186/s12862-016-0827-5.

"Use of British Library Digital Content and Data." *British Library Labs.* labs.bl.uk/Other+Uses+of+Collections. Accessed June 2017.

"Use Cases." *usability.gov,* www.usability.gov/how-to-and-tools/methods/use-cases.html. Accessed 5 June 2017.

Welter-Schultes, Francisco and Bianca Crowley (Lipscomb). "BHL Survey 2010, evaluation of the results." Planet Poster Editions, 2010, www.planetposter.de/bhl/wien2010.htm. Accessed Mar. 2017.

Wick, Angela. “Requirements Elicitation for Business Analysts.” *Lynda.com*, 19 Jan. 2016, www.lynda.com/Communication-tutorials/Using-interviews-elicit-requirements/410330/461470-4.html. Accessed July 2017.

Wilkinson, Jules. Personal interview. 22 June 2017.

Wilson-Ramsey, Yevonn. Personal interview. 9 Mar. 2017.

Zooniverse Project Builder. Zooniverse, 2015, www.zooniverse.org/lab. Accessed Sept. 2017.

Zooniverse: Science Gossip. Zooniverse. api.zooniverse.org/projects/illustratedlife/status. Accessed Mar. 2017.

112 Superb Varieties for Market Gardeners Season of 1926. Francis C. Stokes & Co., 1926. doi.org/10.5962/bhl.title.68866.