Consumers as Creators: Understanding the Annotation Needs of the Scientific Community through the Domain of Botany

The Missouri Botanical Garden (MOBOT), with partners at Saint Louis University (SLU), propose a $50,000 Planning Grant to analyze Web annotation needs of the scientific community and develop a prototype of how those needs may be met within a digital library platform. *Consumers as Creators* will assess the practicality of using existing annotation tools to satisfy this community's needs, including technical, economic, and operational considerations and will identify a set of best practices to integrate an annotation tool within a virtual library. Results will help to illuminate and inform about the annotation needs of botanists as well as those within the broader scientific research community

Statement of Need

The New Media Consortium's *Horizon report: 2015 Library Edition*[1], identifies the Semantic Web and Linked Data as key technologies that will significantly impact academic and research libraries in the next two to three years. Libraries increasingly understand it is insufficient to simply provide online collections access; users want integrated Semantic Web tools among library site services. Annotating (i.e., making comments on a resource) is an important part of the vision for the Semantic Web. While annotation tools can be re-purposed by libraries, most fall within a proprietary environment for particular groups, are not well-suited to general audiences' needs, and do not allow easy sharing and discovery of annotations across the Web. This conflicts with Semantic Web principles and limits access and value.

Despite the important role annotation plays in the Semantic Web, cultural heritage institutions have been slow to adopt it. Digital libraries have rarely incorporated annotation type functionality into their interfaces with a few exceptions, namely, National Library of Australia's Trove repository[2] and Europeana's Sounds project[3]. In 2016, findings from the *I Annotate* conference held in Berlin concluded that the uptake of web annotation could be sufficiently moved forward by tackling these key issues: 1) interoperability, 2) domain use cases, and 3) user centered design[4]. Significant strides in addressing interoperability were made in 2017 with the release of the Web Annotation Data Model (WADM) from the World Wide Web Consortium (W3C) Web Annotation Working Group [5]. With it the opportunity now exists to create a single, common, Resource Description Framework (RDF) based specification for annotating digital resources.

Consumers as Creators aims to make inroads into the second issue, use cases, by honing in on a specific scientific domain and identifying best practices therein. This

[1] https://www.nmc.org/nmc-horizon-news/nmc-releases-the-nmc-horizon-report-2015-library-edition/
[2] https://trove.nla.gov.au/
[3] http://www.eusounds.eu/
[4] https://pro.europeana.eu/page/issue-5-annotations#introduction
[5] https://www.w3.org/TR/annotation-model/

project will advance the annotation needs of the scientific community in its broadest sense, with the goal of developing methods that are expandable to other communities. Our prototype will focus on a particular group of scientists, namely systematic botanists, whose advanced creation of vocabularies and definitions of data model elements will allow it to leverage adoption of annotations. Lessons learned can be applied to demonstrate their potential impact in scientific research, humanities and other scholarly arenas.

The staff of the Center for Biodiversity Informatics (CBI) at MOBOT is well-acquainted with this researcher community, with a decade-long history of developing content repositories to access biodiversity literature (Biodiversity Heritage Library[6]), plant specimen data (Tropicos[7]), and living collections information (LCMS[8]). Through close collaboration with botanists and the librarians who serve them, the CBI has identified valuable use cases for developing in-depth user assessments of annotation needs. With this grant, the CBI will perform a landscape review of existing tools and test the applicability of one of those tools within a platform called Botanicus (http://www.botanicus.org/). Botanicus is a freely accessible portal to historic botanical literature that was developed by staff from CBI in collaboration with librarians from the Peter H. Raven Library at MOBOT.

Botanists' research process and their annotation needs

Systematic botanists, also known as taxonomists, are scientists who describe and identify species. Their research process results in two primary outputs - specimens and publications. Specimens are collected in the field, pressed and dried, mounted on archival sheets and stored in a museum's herbaria for inspection. These then serve as "type specimens" or exemplars to anchor or centralize the defining features of that particular taxon within a description. During the inspection process the specimens are also compared to potentially related specimens found in other herbaria which are acquired through cooperative lending programs. Once inspection is complete, the detailed physical description of the organism must then be published in a scientific journal (whether printed or online) in order for the taxonomist to officially name a species.[9]

The need for annotation comes into play during several stages of the botanists' research process. First as part of the process of inspecting specimens on loan through other herbaria, botanists will often add their more current research data to the loaned specimen sheets. As explained by staff at the University of Florida Herbarium,

> *The value of herbarium specimens are improved by careful annotation. Annotations bring the scientific names of specimens up-to-date to conform with*

---

[6] https://www.biodiversitylibrary.org/

[7] http://www.tropicos.org/

[8] http://livingcollections.org/mobot/

[9] As required by the International Code of Nomenclature for algae, fungi, and plants(ICN) https://iapt-taxon.org/nomen/main.php

current species concepts. This helps herbaria organize their collections and they are an integral part of curatorial management. Annotations also document the use of specimens in research studies. This cross-referencing is integral to the scientific method and is important for future researchers.[10]

Herbaria have even developed guidelines for how these annotations should be recorded which dictate how much information to include, their format, and the type of paper it should be recorded on - preferably archival.

Botanists' need for annotations during the publication stage can happen either as part of the peer review process and/or during the post-publication stage. Peer reviewers, editors and copy editors all use annotations to suggest changes and communicate feedback to authors for improving a text before it goes to publication. Scientists have annotated published books and journals for centuries. Hand-written annotations have served many purposes since the earliest printing of books, including to indicate ownership, respond to the text, or record mottoes and proverbs. A few scientists' preserved, personal libraries provide a wealth of information about the influence of their contemporaries on themselves, and their hand-written marginal notes offer personal reflections on the theories these books contain.

Annotation on the Web

Web annotations constitute a re-creation and extension of these age-old functionalities as a new, interactive mode built on and linked through Web technology. Online annotations were possible as far back as 1993 with an early version of the Mosaic web browser, but during that era, Web 1.0 users were mostly viewed as mere content consumers. Few platforms supported content creation until the move into a Web 2.0 environment. Consumers became creators, and the more recent spread of social media throughout the Web has led to a new understanding of its value as a place to connect, build and share data. As we reach the era of Web 3.0, the original vision of the Semantic Web is being realized although little Web content currently exploits all possibilities.

Herbaria have begun transitioning their specimen collections from analog to online environments. Mostly larger institutions have taken the lead on this since the move requires significant equipment and staffing to image the specimen sheets and transcribe the data contained on them. Botanists have come to expect that annotation will become a core functionality within these user interfaces. For example AnnoSys[11], a web-based annotation system developed at the Botanical Garden and Botanical Museum Berlin, is now being incorporated into at least a dozen specimen data portals.

The move from print to online publications has made it easier not only to create but now share annotations. Publishers are already beginning to build annotation functionality into

---

[10] https://www.floridamuseum.ufl.edu/herbarium/anno/
[11] https://annosys.bgbm.org/

their sites and have adopted multi-platform tools such as hypothes.is[12] which can integrate with html, PDF, and EPUB formats.   Librarians are strategically poised to serve an important role in this research ecosystem as well.   As providers of digitized historic literature and ebooks, librarians can expand beyond their traditional role as content providers by having annotation capabilities integrated directly into their platforms where users can contribute more current expert knowledge and generate broader access points for library content.

Annotation use cases for Botanists

Motivations for creating annotations are somewhat generalizable across domains. For example, they could include wanting to complete or correct an idea; relating objects within or external to a repository together; or simply tagging. But particular domains may have more specific variations on these motivations. For example, a botanist might want to:
- Provide details about taxonomic changes (see Appendix 1)
- Link field notes to crowdsourced transcriptions (see Appendix 2)
- Provide species and common name information for scientific illustrations (see Appendix 3)

The examples above come from real-world annotation actions initiated by users of the Biodiversity Heritage Library (BHL). BHL is a consortium of natural history and botanical libraries that cooperate to digitize the legacy literature of biodiversity held in their collections and to make that literature available for open access. The BHL portal, developed and maintained by the CBI, provides access to over 50 million pages of text from the 15th century to today. Annotation functionality was made available as a trial within the portal from December of 2015 through June 2016 as part of the IMLS-funded *Mining Biodiversity* project[13]. For that project, the CBI chose a social commenting tool called Disqus[14]. BHL received 188 individual annotations during that time[15]. While brief, the trial did demonstrate a desire for botanists and citizen scientists to want to actively engage in the annotation process within a digital library interface.

After the trial the tool was discontinued within BHL for a variety of reasons. Disqus was a proprietary tool that would not have served well as a long term scalable solution. Customizations to the tool were limited and annotations were stored on Disqus and not BHL servers. This goes against principles of the scholarly community to be open and interoperable. For libraries interested in adopting annotation tools they will need to seek open source solutions so that both the tool and data gathered can be preserved and shared more easily.

Shortly following the trial, the CBI was approached by staff from the Walter J. Ong, S.J. Center for Digital Humanities (CDH) at Saint Louis University (SLU) about an annotation

---

[12] https://web.hypothes.is/

[13] http://www.nactem.ac.uk/DID-MIBIO/

[14] https://disqus.com/

[15] https://disqus.com/by/BioDivLibrary/

tool they had developed for humanities scholars called RERUM[16]. Within the humanities, scholarly annotation and assertions are the foundation building blocks of new knowledge. Digital Humanities has magnified the efforts of libraries and museums who have been digitizing and sharing their resources by bringing together the academic conversation and the resources it comprises. The emergence of standards for annotation and content delivery, such as Web Annotation and the International Image Interoperability Framework (IIIF), creates a space for simple repositories and applications to host and share new resources and relationships. RERUM has accelerated this even further by not only hosting IIIF and other digital surrogates and containers alongside the annotations, but also offering the service free and open to the public. The diversity of objects supported by RERUM accelerates the promise of annotation in the humanities (and research in general) by supplying both the content, targets, and containers in an open and discoverable way.

Through a generic API, RERUM has proven adaptable to several public projects including Broken Books[17], a collaboration between SLU Libraries, Digital Scriptorium, and Biblioteca Apostolica Vaticana. Broken Books reassembles and describes dismembered manuscripts through flexible aggregation and annotation. RERUM is also used by the French Renaissance Paleography project[18], a collaboration between the Newberry and University of Toronto Libraries. It implements a customized front-end for a transcription tool called T-PEN, thereby enabling pedagogy and paleographic self-study of early French documents. As IIIF has gained traction in the scientific imaging and cultural heritage communities, CDH has sought opportunities to apply the advantages of RERUM beyond the Digital Humanities. *Consumers as Creators* would provide an interesting test case for the use of RERUM outside of the audience for which it was built and identify its adaptability for botanists and the sciences more broadly.

Relevant annotation tools, projects and standards

This project will focus on open source annotation tools that follow established standards such as W3C's Web Annotation Data Model (WADM)[19] and/or the International Image Interoperability Framework[20] (IIIF). At least four tools meet these criteria: RERUM, Hypothes.is[21], digilib[22], and Annotorious[23]. RERUM will be used as a prototype for testing with data found in the Botanicus platform. Its ability to create and store Web annotations as well as store IIIF documents gives it an advantage over other annotation tools. RERUM developers from SLU will participate in this project by providing technical support on its setup. CBI will also seek input from outside experts and organizations such as the Annotating All Knowledge (AAK) Coalition - a group of key scholarly publishers, libraries, and technologists.

---

[16] http://rerum.io

[17] http://brokenbooks.org

[18] newberry.org/french-renaissance-paleography

[19] https://www.w3.org/TR/annotation-model/

[20] http://iiif.io/

[21] https://web.hypothes.is/

[22] http://digilib.sourceforge.net/

[23] https://annotorious.github.io/

Project Design

Project Goals/Outcomes/Assumptions

The intended outcome of the proposed project is to illuminate literature annotation needs of scientific and other research communities by honing in on the annotation needs of a well-defined user group in systematic botany. Assessment of the practicality of an existing tool to satisfy the annotation needs of botanical users, including technical, economic, and operational considerations, will inform developers on best practices to integrate an annotation tool within a virtual library. Ultimately, a list of planning activities and partner commitments needed for a robust project proposal will result.

Audiences

This project will be useful to the following audiences:
- Librarians looking to improve their virtual library by enabling their users to add value to their content.
- Botanists who want to enhance the corpus of their digital library collection by augmenting knowledge through the annotations provided.
- Developers who want to choose a tool to enable annotations in their online solutions, particularly specialized online library systems.

Deliverables

a. **Needs Analysis Report** with prioritized list of annotation needs for users of a botanical virtual library.
b. **Feasibility Study** with a thorough evaluation of four open source existing annotation tools based on their potential to address the needs identified in the Analysis Report
c. **Proof of concept prototype** installed within a virtual library to demonstrate the functional capacity of one of the evaluated tools
d. **Outcomes Assessment** with next step recommendations to propose a full-scale project adopting an annotation tool as part of a virtual library.

Activities

The following activities will be conducted within one year, beginning May, 2018:

1. In order to understand the needs that users of a specialized digital library could have in relation to annotate their content, we will employ the case research approach, interviewing at least ten users of a botanical virtual library from at least five different institutions to discover concepts and patterns in case data that relate to their annotation needs. Multiple case design is considered more appropriate for theory testing, for establishing generalizability of inferences, and for developing richer and more nuanced interpretations of a phenomenon[24]. So,

as available, current users of annotation tools would be thoroughly questioned about their procedures and workflow when annotating. All answers will be analyzed to get annotations needs described and classified in terms of user type, purpose and function (months 1-3).

2. Four existing annotation tools will be thoroughly evaluated against the previous needs analysis to develop a feasibility study for how they could satisfy botanists' needs. While the project team has done some preliminary assessments of these tools, the feasibility study will conduct a more in-depth look into technical considerations including hardware and software requirements, as well as functionality available for creating user roles. Finally, when possible, staff will estimate time needed for installing and setting up the tools. This learning curve assessment will be useful to libraries with limited technical support (months 3-7).

3. RERUM will be integrated within a digital library platform (Botanicus) as a proof-of-concept on how an existing annotation tool could support the different types of annotations needs that the botanical users may have. This prototype will run against a digital library to test the integration and effectiveness of the requirements compliance. By performing the actual installation of a prototype of one of the tools within a digital library platform, we will be able to corroborate our estimations and determine how to cope with any new issues and risks that have not been foreseen. Several annotations of each type of need identified in activity 1 will be input as a test of the prototype efficiency (months 5-11).

4. Outcomes from this project will be assessed to identify requisites, best practices, needed tasks and further developments required, as well as the appropriate partners needed for a full-scale Project Plan. Particular consideration should be given to the activities needed for the proper expansion and scaling of the prototype (month 10-12).

Project Resources

Total Planning Grant cost is $50,000 Two CBI researchers, Trish Rose-Sandler and William Ulate, will gather and analyze necessary information about scientific annotation needs, handle project management, and implement and test a prototype. Salaries & wages for the year are $34,954, plus $8,040 (23% of salary) included as fringe benefits. In addition, costs include travel to 1 conference $2,461 plus indirect costs of $4,545 (10%). MOBOT will provide hardware, software and server support staff in-kind and SLU will provide technical consultancy support in-kind.

Communication Plan

The project team will use a variety of publishing channels, including traditional print media, conference presentations and emerging social media, to promote its content, services, and activities. These will be used to disseminate project results to the

[24] Bhattacherjee, Anol. Social Science Research: Principles, Methods, and Practices. University of South Florida. 2012. Available at: http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks#page102. (accessed Jan. 16, 2018)

biodiversity and broader scientific communities as well as librarians and technologists. Presentations at conferences may include: TDWG which holds a yearly biodiversity standards group conference; the Biodiversity Heritage Library's yearly member's meeting; the Digital Library Federation (DLF); and the I Annotate conference. Project progress will be shared through various listservs including: Taxicom, DLF, and TDWG as well as social media accounts including Twitter, Facebook and blogs (MOBOT http://discoverandshare.org/ and SLU http://blog.ongcdh.org/),

Diversity

Diversity will be addressed during the project, and in the creation of any follow-on proposals, starting with the project team. The team reflects diversity in both gender and ethnicity. Project leads at MOBOT include both a female librarian from the U.S. and a male computer scientist immigrant from Costa Rica. Staff at SLU includes a male designer born in Ireland. This will allow the team to bring a variety of perspectives to bear on this project. We will also actively try to identify and recruit botanists who are diverse in terms gender, work at small and large institutions and represent varying ethnicities. Perhaps more importantly, we will want to interview botanists of varying ages as we suspect younger botanists may be more open and engaged with the idea of online annotation than older, more seasoned botanists. We hope this will consciously address and limit any discriminatory effect from a generational digital divide.

National Impact

Curation is the act of selecting and interpreting content - a role previously limited to content providers such as publishers and librarians. As Web users have an increasing desire to move from consumers to creators, they want to actively engage in the curation role and their annotations can offer additional access points beyond traditional bibliographic information that libraries provide. User-added annotations can lead to a richer dialog and broader context around curated collections than is otherwise had by restricting the curation role to traditional curators, thereby increasing the collections' impact, value and reach.

Annotations can better facilitate discovery and extraction of knowledge from scientific literature. Until recently, the idea of an integrated standardized annotation as "a unit of conversation built into the very fabric of the Web" was a far reaching objective. Today it is becoming a high priority need that libraries must address as part of their services in order to impact the future practice of research. The AAK Coalition posits this in turn will "transform scholarship" by enabling "personal note taking, peer review, copy editing, post publication discussion, journal clubs, classroom uses, automated classification, deep linking, and much more".

Consumers as Creators will build upon current strategic initiatives in the implementation of the Semantic Web to demonstrate the importance of annotation functionality that comments on, characterizes, extends or links entities from different realms in science. We will demonstrate through use cases from a specific scientific community and

through a prototype within a digital library portal, how annotations can produce far-reaching impacts across virtual libraries of any type of cultural heritage institution.