Purposeful gaming and BHL: engaging the public in improving and enhancing access to digital texts

1.   Statement of Need

The Missouri Botanical Garden (MOBOT) and partners at Harvard University, Cornell University, and The New York Botanical Garden propose a project to significantly improve access to digital texts through the applicability of purposeful gaming for the completion of data enhancement tasks needed for content found within the Biodiversity Heritage Library (BHL). This project tackles a major challenge for digital libraries: full-text searching of texts is significantly hampered by poor output from Optical Character Recognition (OCR) software. Historic literature has proven to be particularly problematic because of its tendency to have varying fonts, typesetting, and layouts that make it difficult to accurately render. The European Union's IMPACT project[1], a 2008-2012 effort to improve access to texts, states that poor OCR "does in many cases not produce satisfying results for historical documents. Recognition rates are poor or even useless. No commercial or other OCR engine is able to cope satisfactorily with the wide range of printed materials published between the start of the Gutenberg age in the 15th century and the start of the industrial production of books in the middle of the 19th century." This state of affairs illustrates the pressing need to identify additional solutions to OCR for improving access to digital texts.

The BHL is an international consortium of the world's leading natural history libraries, including the Missouri Botanical Garden's Peter H. Raven Library, that have collaborated to digitize the public domain literature documenting the world's biological diversity. This has resulted in the single largest, open-licensed source of biodiversity literature made available both through the Internet Archive and through a customized portal at http://www.biodiversitylibrary.org/. BHL is a perfect testbed for investigating alternate solutions to the generation of digital outputs both because it is a significantly large corpus (40 million pages of scanned texts accompanied by 40 million OCR outputs) and because the majority of its content is historic literature (the majority of BHL content was published between 1450s-1900s). OCR is also largely ineffective on hand-written texts such as field notebooks–a growing content type in the BHL. Transcription tools can be implemented to manually generate the digital output from the public, and notable examples of crowd-sourced transcription projects within the U.S. include: the New York Public Library's *What's on the Menu* and the National Archives' *Transcription Pilot Project*. Yet, even manual transcription of texts is subject to error because it relies on a single user interpretation of a page.

Purposeful Gaming and BHL will demonstrate whether or not digital games are an excellent tool for analyzing and improving digital outputs from OCR and transcription activities because large numbers of users can be harnessed quickly and efficiently to focus on the review and correction of particularly problematic words by being presented the task as a game. Digital gaming as entertainment has been around for several decades but only recently has it been adopted by other fields. Organizations are finding gaming useful for any number of tasks, such as folding protein strings, encouraging brand loyalty, and training employees. This transition from gaming as mere play into gaming with a purpose may explain its appeal and wider uptake. By contrast, there has been a more limited uptake of purposeful gaming within libraries. DigitalKoot[2], a game

---

[1] EU's IMPACT Project http://www.impact-project.eu/about-the-project/concept/
[2] DigitalKoot http://www.digitalkoot.fi/

commissioned by the National Library of Finland for improving text correction, is one of the few shining examples. When surveying U.S. libraries, the picture does not improve. Beyond a few notable projects, such as *Metadata Games*[3] deployed at Rauner Library at Dartmouth College and *Gaming Against Plagiarism*[4] developed at the Marston Science Library at the University of Florida, U.S. libraries do not seem to have widely exploited this tool.

Additionally, of the projects listed above, only DigitalKoot tackles the improvement of OCR output from digital texts. *Purposeful Gaming and BHL* will build upon the goal of text correction from DigitalKoot but diverge in important ways. First, *Purposeful Gaming and BHL* will test the use of gaming for text correction against three unique textual content types: books/journals, field notebooks, and horticultural catalogs. These three types serve as exemplars of some the most difficult challenges for OCR interpretation. Challenges for books/journals and field notebooks were previously outlined. Horticultural catalogs tend to exhibit very idiosyncratic layouts, contain multiple columns, and are often a mix of text and illustrations, all of which introduce problems for OCR software and make them a perfect candidate for improvement through a gaming tool. Secondly, *Purposeful Gaming and BHL* will develop a framework for how to manage digital outputs generated by multiple tools (i.e., the results from the OCR, transcription, and gaming tools) in order to quantitatively measure and assess their effectiveness so that other libraries can do a cost/benefit analysis when deciding whether or not to implement them.

In order to build the testbed needed for the *Purposeful Gaming and BHL* project, staff will utilize a mix of existing digitized content and content that will need to be digitized as part of this project. BHL has already digitized and ingested into its portal a significant number of books and journals. Field notebooks are currently being digitized by Harvard, one of the project's partners, as part of the IMLS-funded *Connecting Content* project–an effort to develop an innovative framework for integrating primary source documents with museum specimens and digitized published literature. A selection of these notebooks will be used for *Purposeful Gaming and BHL*. Horticultural catalogs are currently lacking in the BHL corpus and will need to be digitized as part of this project by MOBOT and project partners at Cornell and The New York Botanical Garden (NYBG), all of whom have sizable collections in these areas.

Purposeful Gaming and BHL will leverage tools developed as part of the NEH-funded *Art of Life* project. For that project, MOBOT is developing algorithms for automated identification of illustrations found within the text pages of the BHL corpus and pushing those illustrations to crowd-sourcing environments (e.g. Flickr and Wikimedia Commons) for their description. All pages digitized as a result of *Purposeful Gaming and BHL* will be run through the illustration identification algorithms, and those pages will be sent to crowd-sourcing environments for further description. All three textual content types to be used in this project are rich in botanical and zoological imagery, particularly horticultural catalogs which often contain lavish full- color illustrations.

2. Impact

The most direct benefit from *Purposeful Gaming and BHL* will be to improve access to content in the largest open-access repository in biodiversity-the Biodiversity Heritage Library. Because it has rich

---

[3] Metadata Games http://www.tiltfactor.org/metadata-games

[4] Gaming Against Plagiarism http://blogs.uflib.ufl.edu/gap/the-team/

textual and illustrative content that is freely available for download and reuse without restrictions, the BHL has appeal and use by a wide range of audiences beyond its primary community including artists, historians, illustrators, graphic designers, publishers, archivists, educators, students, and citizen scientists. Researchers from the biological sciences have expressed an ongoing need to tap this wealth of information using automated methods from computer science in order to index the whole corpus. This would allow for full-text searching, data mining, and mark-up of content in order to analyze the knowledge contained within. At the BHL annual staff meeting in 2012, implementation of full text search was identified as a "high impact" activity that BHL should pursue in the next one to two years because of the anticipated benefits to its users. All of these activities require accurately rendered digital outputs in order to be successful.

Field notebooks are a gold mine of unpublished observations, journal notes, sketches, weather reports, specimen lists and travel narratives–they are primary source data at its most raw and unevaluated. The field notebooks to be used for *Purposeful Gaming and BHL* are those of William Brewster, a well-known, highly published ornithologist working during the late 19th and early 20th centuries. The ability to do full-text searching of these notebooks will enhance contemporary studies and make the entire research cycle of this scientist's work available for analysis by historians of science, scientists, social scientists, and humanists.

The addition of horticultural catalogs to the BHL corpus will provide a new content type that is sorely lacking. These catalogs help to trace the nation's formative and continued development by illustrating its growth from an agrarian society, in which plants and seeds were bought and sold primarily for utilitarian purposes, to an industrial one, in which these materials informed interior design sensibilities and gardening for leisure and aesthetic purposes. These catalogs and their accompanying illustrations provide some of the earliest information available on the history and development of botany, horticulture, and commercial agriculture in the country, and constitute the material heritage of the nursery and seed industries in America. Audiences to benefit from digitization and full text access to them would include gardeners, landscape designers, horticultural historians, plant breeders, and even art historians.

More importantly, benefits from the results of *Purposeful Gaming and BHL* would extend to the broader digital library community. Any institution managing large text collections will benefit from novel and more cost-effective approaches to generating searchable texts, such as engaging the public in text correction through interaction with a game. *TrendsWatch 2012: Museums and the pulse of the future* urges exploration of "the intersection of crowdsourcing with gaming, making the input and anticipation fun—not like work—while capturing hundreds or thousands of hours of volunteer time"[5]. While aimed at museums, the plea could easily be directed toward any cultural heritage organization. The anticipated benefits of gaming include improved access to content by providing richer and more accurate data; an extension of limited staff resources; and exposure of library content to communities who may not know about the collections otherwise.

Finally, project results could be particularly beneficial for special collections and archival collections that are digitizing their manuscript materials but often have little more than collection-level descriptions of their content to search on. Digital outputs from the scanned pages become the key to accessing these collections. Transcription tools, while useful, have a limited appeal to those audiences who are willing to spend considerable time viewing and transcribing detailed pages of hand-written text. They are also

---

[5] http://www.aam-us.org/docs/center-for-the-future-of-museums/2012_trends_watch_final.pdf

prone to error because their interpretation is limited to a single transcriber. Gaming would serve as a complementary tool to transcription by providing a more efficient and effective way to enlist larger numbers of people in text correction through a fun and interactive gaming environment and ultimately ensure accurate text for mining. Combining the aspects of play, rewards, and purpose, provides a potentially powerful tool.

Assessment method

We will use several methods to assess the quality and gradual improvement of digital outputs created through the project. The quality of a single OCR output will be automatically scored based on the number of non-alpha characters it contains. Two or more OCR or transcription outputs from the same page will be automatically scored based on the number of differences between them. This will help determine which outputs are good candidates for the gaming tool. Outputs from the game will then be compared against a set of "ground truth" files which are manually keyed pages (i.e., 100% accurate). By assigning scores at each stage of digital output, it will be possible to accurately measure the level of improvement. Success of the gaming tool will be affected by a number of factors such as the quality of design, promotion of the game, and the number of users who engage with it. Ultimately, digital libraries will need to weigh the game's level of improvement against costs and investment of staff time in its deployment and management.

3. Project Design

The development teams at MOBOT, Cornell, Harvard, and NYBG, working with a professional game software developer, will collaborate to design, develop, and deploy the applications and infrastructure required to achieve project objectives within a two-year timeframe. The development process will consist of frequent in-person meetings at MOBOT for cognizant MOBOT staff and bi-weekly meetings with all project partners using video and/or telephone conferencing. Project management will be done through MOBOT's web-based project management system, Redmine. These utilities will help the development teams share ideas and communicate progress on project deliverables in an open and collaborative environment. The MOBOT project manager will have responsibility for overall coordination of the project and its deliverables within the prescribed timeframe. *Purposeful Gaming and BHL* has eight primary objectives, described below: 1) digitizing horticultural catalogs; 2) transcribing field notebooks and horticultural catalogs; 3) building a technical framework for management of outputs; 4) comparing digital outputs for accuracy; 5) developing and deploying a game; 6) evaluating accuracy scores from the game against ground truth pages; 7) generating an error matrix for clean-up; and 8) producing a report and disseminating findings.

The document called Workflow Diagram for Processes and Decision Points in the Appendix helps to illustrate the sequence of activities and decision points discussed in the objectives.

Objective 1: Digitizing horticultural catalogs and generating two digital outputs (Months 1-16)

Horticultural catalogs will be digitized by staff or contractors at Cornell, NYBG, and MOBOT. OCR pages will be generated from the resulting digital scans as part of the upload process to Internet Archive and the BHL portal. The quality of the OCR will be automatically assessed based on the number of non-alpha characters it contains. If it contains less than 20% of non-alpha characters, a second OCR

will be generated by MOBOT. If it contains more than 20% of non-alpha characters, the content will be sent to the transcription process as described in Objective 2.

Objective 2: Transcribing digital scans of field notebooks and horticultural catalogs (Months 1-16)

As mentioned previously, field notebooks are being digitized by Harvard as part of the IMLS *Connecting Content* grant and are in the process of being loaded and made accessible in the BHL portal. The resulting digital scans from the pages of notebooks cannot be effectively run against OCR software because the original text is hand-written and not typeset. Therefore, we will implement a transcription tool that will allow users to manually transcribe the text while viewing side-by-side digital images of the hand-written text. We will require at least two transcription outputs from different users in order to compare the texts. We will utilize one of several open source transcription tools that are available, rather than developing our own. Potential candidates include: Scripto (http://scripto.org/); FromThePage (http://beta.fromthepage.com/); and T-PEN (http://t-pen.org/TPEN/).

Once it is determined which open-source transcription tool will provide the functionality needed for this project, the software will be installed either as a stand-alone application or as part of the BHL portal user interface. Once installed, the tool will be promoted through traditional (e.g., newspaper articles) and social media outlets (blogging, Twitter, Facebook, etc.). Users will include primarily the public and some interns from Harvard. We will also conduct an experiment with using a small number of workers from Amazon's Mechanical Turk services (https://aws.amazon.com/mturk/) to see if that is a more cost effective way to complete transcriptions. Reports from successful transcription projects have indicated the need for oversight and feedback to participants in order for the transcription project to be productive.

Objective 3: Building the technical framework for managing and comparing different versions of page text (Months 1-8)

Whether gathered via automated means (i.e., OCR) or manual means (i.e., transcription or gaming), outputs from the digital texts will need to be stored and managed with a system. The project will implement a technical framework to support the workflow and to measure outputs at each stage as described in Objective 4. The current BHL portal contains an OCR text storage module which will be adapted to accommodate the additional functionality needed for this project. Part of that functionality will include development of an algorithm to classify OCR pages according to the number of non-alpha characters they contain.

Objective 4: Comparison of pages generated through OCR and transcription activities in order to assign an accuracy score and help determine which words are best candidates for gaming tool (Months 13-16)

Having two or more digital outputs for a text page can help assess whether there is agreement between the words on the page and provide some measure of accuracy. Accuracy rates will be determined by the percentage of words that are in disagreement compared to overall number of words on the page. Those outputs with a low accuracy score will be set aside for manual review and considered for potential manual correction. Those outputs with a medium to high accuracy score will be candidates for the gaming tool. Their problematic words will be extracted and serve as content for input into the game.

Objective 5: Developing and deploying an online game that will allow crowdsourcing of text correction (Months 5-21)

Development and deployment of an online game will be an essential step in the workflow of correcting digital outputs. Gaming will provide a fun and interactive way to engage the public in text correction activities where OCR and transcription have failed. While we have conducted an extensive landscape review of purposeful games and have been in discussions with a local game design company based in St Louis, we have not had an opportunity to engage in discussions across multiple gaming companies to find the best candidate. We believe this is a necessary step in the success of this project. Therefore, we will write and disseminate a Request for Proposals (RFP) to multiple gaming companies and sign a contract with a company which can best meet our needs within the budget we have identified.   Details about what qualities we are looking for can be found in the Resumes of Key Project Staff and Consultants document under Gaming Company Qualifications. Minimum requirements are that the company or organization can provide a solution that is not tied to a particular software or hardware environment, is scalable, can meet project requirements with the specified budget of $110,000, and preferably, would have an understanding of, if not experience in working with, libraries.

Objective 6: Evaluating accuracy scores from the game against "ground truth" pages (Months 22-23)

In order to be able to assess how well the gaming tool improved the digital outputs, we will need to compare a sample of the corrected text output from the game with a set of "ground truth" pages. These are pages that have been manually transcribed. We will utilize existing ground truth pages that were created for the BHL-Europe and IMPACT collaboration[6], but this is only a small set (12 books and 100 random pages). Additional ground truth pages will be created as a result of manual review of digital outputs having low accuracy scores. We will then compare outputs from the game to these pages and assign accuracy scores. The results from the evaluation will be used to fine-tune the text correction process.

Objective 7: Generating an error matrix in order to automate correction across entire BHL corpus (Months 22-23)

In order to apply results from the game to the entire BHL corpus, it will be necessary to generate an error matrix, also known as a "confusion matrix" in the field of artificial intelligence. This would serve as a probabilistic model of the errors introduced by both OCR and the transcription processes. The matrix generates a set of strings that are in error or incorrect with their corresponding correct strings, (e.g. the letter "u" being replaced by "ii", the degree symbol ° being replaced by symbols such as "^"). A robust error matrix will automate further cleanup of the entire corpus enabling results from the game to become much more scalable.

Objective 8: Writing final report and disseminating findings (Months 1-24)

We will produce a final report that outlines the impact of a gaming tool on improving digital outputs for text in measurable and quantitative ways. This report will include detailed information on the technical framework needed for managing and comparing digital outputs. It will also include instructions on creating an error matrix so that results from a game can scale to a much larger digital

---

[6]Results of collaboration of BHL-Europe and IMPACT http://www.bhle.eu/sk/node/1685

corpus. Findings will be disseminated throughout the duration of the project via BHL's social media channels and as part of conference presentations (see Communication Plans for details). Staff will also publish final results in a primary digital library publication, such as *D-LIB* magazine.

4. Project Resources: Personnel, Time, Budget

Budget
This project will require a total of $903,226 to complete. The total amount of requested funding from IMLS is $449,641 divided among MOBOT, the three partner institutions, and a game design company. The total amount of in-kind/cost-sharing support provided by all participating institutions is $453,585.

Project Staff
This project is a top priority for the Missouri Botanical Garden to advance the mission of the Peter H. Raven Library and of the Biodiversity Heritage Library as a whole. It will tackle a key challenge in biodiversity informatics. It builds upon the institution's investments over many years and will create content and tools that will help move many other efforts forward. As such, we plan to dedicate significant amounts of MOBOT staff time to all phases of the project and this time is viewed as fundamental to (not outside of) those individual's regular duties. Those individuals include:

- *Charles Miller*, Vice President of Information Technology and Chief Information Officer, Missouri Botanical Garden
  Miller is the lead technologist at MOBOT, helping guide design and delivery of all systems supporting MOBOT's mission, including operational applications like ticketing systems and retail point of sale applications, as well as BHL and other informatics tools developed for management of MOBOT's rich scientific datasets. Miller will provide project oversight and evaluation of the use and delivery of the proposed software resources.
- *William Ulate Rodriguez,* Senior Project Coordinator, Center for Biodiversity Informatics, Missouri Botanical Garden and Technical Director for BHL
  Ulate is an integral part of BHL, helping coordinate project management activities across its partner institutions and funded initiatives and overseeing the technical development of the BHL portal. Ulate will strengthen the project through his domain knowledge of scientific data and his strong background in custom software development projects. He will serve as grant administrator and play a shared role with the project manager in identifying functionality requirements for system architecture, writing the RFP for the gaming company, and working with the chosen company on design, testing, and deployment.
- *Trish Rose-Sandler,* Data Analyst, Center for Biodiversity Informatics, Missouri Botanical Garden
  Rose-Sandler works with stakeholders within the Center for Biodiversity Informatics and its associated projects to provide content and data management assistance to the Biodiversity Heritage Library, Tropicos, and other informatics systems at MOBOT. She is currently the PI for the NEH project *Art of Life.* For this project, Rose-Sandler will serve as project manager, assuming responsibility for overall coordination, deliverables, dissemination of project information and findings, and meeting IMLS reporting requirements. Other tasks include: identifying functionality requirements for system architecture; writing the RFP for

the gaming company and then working with the chosen company on design, testing, and deployment; and development of the error matrix.

- *Douglas Holland,* Director, Missouri Botanical Garden Peter H. Raven Library
  Holland, through his role as Library Director, has unparalleled expertise in the domain of natural history literature and will provide administrative oversight and consultation on matters relating to the horticultural catalogs. He will be responsible for identifying which catalogs to digitize from the MOBOT Library and clarifying their public domain status. He will assist with communication and dissemination activities and oversee the work of Mike Blomberg.
- *Mike Blomberg,* Imaging Lab Coordinator, Missouri Botanical Garden Peter H. Raven Library
  Blomberg has been instrumental in the setup and implementation of scanning facilities within the Library and has been actively involved in metadata description for materials in BHL since its inception in 2007. Blomberg will support this project by scanning horticultural catalogs and generating OCR outputs for MOBOT, Cornell, and NYBG. He will be involved in transcription activities when needed.
- *Biodiversity Informatics Architect (this position is in the process of being re-filled),* Missouri Botanical Garden
  The Informatics Architect will work with the programming contractor on development, testing, and deployment of the various systems and tools for the project. He/she will also support grant-funded project deliverables through institutional means following the end of grant funding, as well as migrating materials and resources into new systems as those are instantiated within MOBOT's technical infrastructure.
- *Mike Lichtenberg,* Programming Contractor (Missouri Botanical Garden)
  Lichtenberg is an expert contract programmer with more than ten years of experience programming large, feature-rich web sites. Lichtenberg has been the lead developer of BHL for more than six years, helping guide its expansion from a prototype to the fully realized production system it is today. Lichtenberg will build the technical framework for managing and comparing different versions of page text and be involved in the implementation of the transcription and gaming tools and their interactions with existing BHL applications.

Each partner institution has appointed key staff to oversee their institutional contributions. These include:

Ernst Mayr Library of the Museum of Comparative Zoology at Harvard University
- James Hanken, Director of the Museum of Comparative Zoology
- Constance Rinaldo, Librarian of the Ernst Mayr Library
- Joe deVeer, Project Manager
- Robert Young, Special Collections Librarian

Mann Library, Cornell University
- Martin Schlabach, Librarian
- Kevin Nixon, Professor of Botany

The LuEsther T. Mertz Library, The New York Botanical Garden
- Susan Fraser, Director
- John Mignault, Systems Librarian

- Kevin Nolan, Digital Projects Manager
- Lisa Studier, Metadata Cataloger
- Yumi Choi, Catalog Librarian
- Andrew Tschinkel, Scanning Technician

Duration
This project will be conducted from December 1, 2013-November 30, 2015. There will be at least one face-to-face meeting with MOBOT staff and partners during the duration of this project.

MOBOT has an extensive financial and accounting system in place to manage federal grants and receives federal awards from multiple agencies each year. MOBOT had a significant (and planned) income surplus in 2011 due to the influx of funds in response to our current capital campaign. Virtually all of these additional funds are restricted to capital campaign projects and endowment and not applicable to the annual operating budget.

5. Communications Plan

This project will utilize a variety of communication channels, including traditional print media, conference presentations, and social media, to disseminate its findings to the BHL community of enthusiasts as well as the broader digital library community.

BHL has an active social media presence and uses the following technologies in support of promoting program deliverables and achievements:
- *BHL Blog*: http://biodiversitylibrary.blogspot.com

The BHL Blog is used to report preliminary results from new initiatives and gives historical background on a given book, author, illustrator, organism or theme. More than 2,600 visitors a month consult the BHL Blog.
- *Twitter*: http://twitter.com/BioDivLibrary

BHL's Twitter feed is used to send out links to interesting content within BHL, including a BHL Page of the Day. Regular status updates on *Purposeful Gaming and BHL* and overall progress will be sent out to the more than 2,354 people who follow BHL on Twitter.
- *Facebook*: http://www.facebook.com/pages/Biodiversity-Heritage-Library/63547246565

BHL has more than 4,365 people who "like" its page on Facebook, which features quizzes on species identification and other challenges to enthusiasts interested in BHL content.

MOBOT and its partners will present project findings to relevant professional groups of humanities scholars, biodiversity scientists, and library and information science curators. We will prepare presentations for key conferences that occur during project funding, such as IMLS' Webwise, the Digital Library Federation Forum, the Taxonomic Database Working Group (TDWG) meeting, as well as conferences hosted by the American Library Association and Society of American Archivists. An important forum for sharing project progress and results will be the annual BHL staff meetings.

6. Sustainability

MOBOT has a demonstrated history of supporting grant-funded project deliverables through institutional means following the end of grant funding, as well as migrating materials and resources

into new systems as those are instantiated within MOBOT's technical infrastructure. MOBOT has served as the host for the technical development team for BHL since it was founded in 2007 and has developed a robust preservation infrastructure for long-term maintenance of the interfaces, applications, and digital products resulting from BHL programming activities

Digital products that will result specifically from this grant include:

- Metadata and their related content files for horticultural catalogs
- Digital outputs from transcription software
- Digital outputs from OCR software
- Correct digital outputs from the game
- Ground truth files

These products will be created and maintained within the same stringent standards of the BHL workflow. For scans, BHL follows the Digital Library Federation's (DLF) "Benchmark for Faithful Digital Reproductions of Monographs and Serials" (http://www.diglib.org/standards/bmarkfin.htm). Metadata is encoded in appropriate standards for descriptive (MARCXML, MODS), technical (PREMIS), and structural data (METS).

Long-term preservation of all data is ensured through storage in a preservation system built from the Fedora Commons framework as well as mirroring of BHL content at more than one location. BHL uses file synchronization utilities to copy all of the digital files for a scanned book from Internet Archive's servers onto a High Performance Computing Cluster located in Woods Hole, Massachusetts. These servers are managed by the Marine Biological Laboratory, one of BHL's partner institutions.

It is also important to mention when discussing sustainability of project results that all of the content within BHL is published freely online and can be viewed and downloaded by anyone without restriction or fee, thereby furthering BHL's strong commitment to public domain literature having broad, unfettered use. BHL already contributes its metadata records to OCLC in order to increase findability. Through completion of project objectives, will have a set of corrected text pages that can also be reused within other portals and enable full-text searching of the content.

Perhaps more important than the digital products that will result from *Purposeful Gaming and BHL* is the expansion of the BHL architecture to help manage and compare different versions of page text. MOBOT staff will gain experience in building a system that can manage multiple outputs for a single digital scan, coming from different sources and requiring version tracking. Algorithms will be developed to analyze agreements across versions in order to apply accuracy scores. The scores can then be used to determine which output provides the best candidate for full-text search. Both the architecture and output from varying text correction tools will serve as a benchmark for libraries to follow when determining which tools and workflow techniques to adopt locally in order to improve access to their own digital texts and improve the search experience for their users.

As well, if gaming is demonstrated as an excellent tool for analyzing and improving digital outputs from OCR and transcription activities, we expect that its use will expand among BHL partners and others in the digital library community. This will help meet the urgent and growing worldwide need to significantly improve access to digital texts.