

## **OpenLibrary.org: Leveraging Digital Technologies to Provide Open, Universal Access to Books**

### **INTRODUCTION**

We propose to build OpenLibrary.org as a step forward in facilitating resource sharing among small and owners of special or hidden collections. The project will build on the strengths of the Internet Archive's book digitization activities and development of universal online access tools, Plymouth State University's (PSU) development of tools to bring small libraries online, the Boston Library Consortium's (BLC) leadership role in the Rethinking Resource Sharing Initiative. OpenLibrary.org will focus on the needs of institutions that struggle with inter-library resource sharing and creating easy community access to their collections.

### **1. Needs Assessment**

Despite the fact that libraries are making their catalogs available online in ever greater numbers, and that books are being digitized through large initiatives such as the Million Books Project, the Open Content Alliance and other thematic projects, books are hard to find. An average Internet user searching the web might never learn of books or other materials that might meet their needs. Many library resources – particularly small and special collections - are part of what is called the 'deep' or 'invisible' web and can't be found in search engines. OCLC data shows that 84% of the general public begins their search for information in search engines, and Pew Internet Project data shows that 80% of Internet users trust what they find there. Thus both libraries and Internet users are at risk: users are at risk of not finding the books they seek, and libraries are at risk of becoming less relevant to their patrons because they are unable reach them via traditional channels.

This is in part because online book sharing and digitization initiatives are being implemented independently from one another. Existing online resources would appear to benefit from a supplemental mechanism to allow readers and librarians to easily find a copy of a book to borrow, find out whether it has been digitized, and to get access to other metadata, edition and interpretive information about the book. This project seeks to address these problems by developing a website with services that serve both readers and libraries in reaching distributed resources.

With its partners, the Internet Archive will build an open, structured website that includes tools to provide public and institutional access to library resources around the world. OpenLibrary.org will offer a system of integrated tools that can be used individually or together to meet a library or patron's needs, that is free to the user and the library. The overall goal of the project is to shorten the distance between initial query and document.

Benefits to the library community will be:

- Open and free tools for exposing and sharing collections
- A mechanism to augment local catalogs with worldwide book resources
- Online representation of local and unique resources
- Better coordination of digitization efforts
- A model for resource sharing that can be replicated by library systems, including school systems and consortia

Benefits to the library patron or reader will include:

- A user-friendly tool to find and access books wherever they may reside
- Remixability of books and information in unique and personalized ways
- Information about where to borrow or purchase books
- Interpretation, analysis and other information to enhance the reading experience
- A forum for discussion of books

The OpenLibrary.org website will provide one page per published book with links to libraries with physical copies, to digitized versions and to community-generated commentary and book sharing. All book pages will have metadata in accepted formats that are harvestable within the OAI model. The initial corpus of pages will be based on three sources of data - the Library of Congress (LoC) bibliographic catalog, publisher information about printed books and the metadata and scanned text of books from the Open Content Alliance (OCA.) Throughout the project period we hope to add records and scanned texts from testbed libraries that participate in the website development..

The Open Content Alliance and other partner libraries will provide access to digitized books via hotlink or upload. As use of the OpenLibrary site grows, we envision access to other scanned books from sources worldwide. To facilitate better coordination and use of digitization resources, we will develop a prototype feature to scan books on demand. Future features of the site may include patron-initiated inter-library loans, even if mediated by local libraries.

These services could be of special use to small libraries. Over half of the nation's libraries serve populations of less than 10,000, and nearly 29% serve populations of fewer than 2,500. These libraries especially need cost effective tools to help them improve service and share resources.

OpenLibrary.org will allow libraries to more fully serve their local communities while maintaining their unique identities. This will be made possible through Scriblio, an open source application being developed by PSU for hosting and integration into OpenLibrary.org.

The Scriblio component of the site will help bring small libraries, including their catalogs and other services, online. It enables libraries to easily build websites, up- and download catalog records, and to present their full range of services to local patrons. In this way libraries can maintain their distinct identities while participating in the larger community of libraries sharing resources through the OpenLibrary site. This can be of special value to school libraries with little or no online presence and small or rural libraries.

Since the website will be built with open source software and will be open itself, it can be used as a model for other online communities, including historical societies, museums, school systems and any other organizations with the desire to more effectively share their physical and digital resources.

Other projects that attempt to address the needs of these audiences do not offer comprehensive, free, open and user-friendly access to books worldwide. ILS systems are designed for internal library use, and online systems such as Open WorldCat contain book data available only under license –with small libraries and special collections significantly under-represented in the catalogs. Newer ‘born-digital’ collections such as Google books and WikiBooks do not offer library tools or mechanisms for resource sharing. Finally, none of these resources represent the library beyond the book, which OpenLibrary.org will be able to do.

### **Institutional Collaboration**

Partners directly involved in the OpenLibrary project are the Internet Archive, Plymouth State University and the Boston Library Consortium.

The Internet Archive is building the OpenLibrary.org website and its underlying storage and hosting architecture plus finding and access tools for books. The site will have a wikipedia-like structure, enabling members of the public, institutions and libraries to add information about individual books. The backbone of the resource will be an open, universal union catalog with one page per published book, starting with publishers' ONIX feeds, the Library of Congress catalog and scanned books contributed by OCA. Internet Archive will add records from other libraries as we test the website functions. The Archive has ten years' experience building petabyte-scale storage and hosting infrastructure, and serves millions of unique users in searching its existing web, book, audio, film and software collections. The Archive serves as a technology partner for many large libraries and archives, including NARA, Library of Congress, Bibliothèque Nationale de France and the British Library. The Archive has received awards from numerous foundations, including the Andrew W. Mellon and Alfred P. Sloan Foundations, for its innovations in open source software development.

Plymouth State University will develop a multi-user version of the Scriblio application to be hosted on the OpenLibrary.org web site. Scriblio is an integrated web content management system with faceted search and browse features. It started as WPopac, an open source OPAC that the Andrew W. Mellon Foundation recognized with a Mellon Award for Technology Collaboration in December 2006, and has since expanded to include web content management features in line with its objective: library software for library users. Scriblio enables small libraries to bring their catalogs and other library services online in a free and easy to use way. PSU and the Internet Archive will work together to integrate up- and down-load features for bibliographic information, simultaneously growing the OpenLibrary's union catalog and enabling the addition of records to a library's local catalog.

The Boston Library Consortium will be active in the development and dissemination phases of this project. Member libraries will serve as testbeds for the website and its functions as they are developed. They may also contribute special collections to the OpenLibrary union catalog as they test the prototype record upload facility. The Rethinking Resource Sharing Initiative (RRSI), which is jointly chaired by the Boston Library Consortium, the California Digital Library and the Pennsylvania Academic Library Consortium, has identified the need to clearly define the opportunities to improve inter-library resource sharing and to explore novel solutions. As a partner in this project, the BLC will bring the OpenLibrary project to the RRSI group for input and participation.

The Open Content Alliance, which is administered by IA, will make available all of its digitized books to the OpenLibrary, and may potentially serve as scanning partners when books are selected for digitization.

The international library community will be engaged in this dialogue through the International Internet Preservation Consortium (IIPC), of which the Internet Archive is a founding member.

Additional informal partners will be sought and welcomed to participate in the development of OpenLibrary.org and its integration with existing systems. The University of Rochester, OCLC and

the University of Illinois at Urbana/Champaign are just a few of the forward thinkers whose collaboration we would welcome.

## **2. National Impact and Intended Results**

OpenLibrary.org will have national impact because it will enable any library, anywhere to easily join the online book community. Small libraries around the country and larger institutions with special or hidden collections will be able to raise their visibility easily and with minimal effort and expense. Moreover, with each contributing institution the value of the union catalog will grow, offering an ever-richer resource for libraries and patrons everywhere.

The success of projects such as PENNtags, through which the University of Pennsylvania has opened up its library catalog to the public, and RAPID, a system to facilitate patron-initiated requests for journal articles, suggests that there is an appetite for open resources and a public community that is eager to participate in building and supporting it.

OpenLibrary.org will help small libraries to have a meaningful online presence. These may be small rural libraries or even school libraries. The project can help foster adoption of library standards in the online community by creating a neutral, open system with which commercial systems will be able to integrate and interoperate, should they so choose. All components of the OpenLibrary and Scriblio system will be created using open source software and will, themselves be open. They will be available for adoption and sharing by any institution or consortium wishing to do so.

A major benefit of this shared resource will be coordinated, rather than independent, book digitization efforts. If a book has already been scanned, it can be found and shared with rights management being exercised by the owner of the scanned copy.

OpenLibrary.org is being designed to serve multiple constituencies – the small library wishing to better serve its local community while maintaining its distinct identity, larger institutions needing a mechanism to expose their hidden collections, library patrons and general internet users seeking books beyond the limits of their local library systems and individuals wanting a forum to discuss, organize and annotate book resources in personalized ways.

The proposed wiki-like editing aspect of the website is also an indicator of its ability to have national impact on members of the public. The trend toward user-generated-content is reflected in many online resources today and we feel that there is a high likelihood it will work for the book-reading community as well. While wiki-style systems sometimes bring concerns about degradation of records information, we believe the information sharing benefit to the community outweighs this risk. To mitigate this risk we will keep original metadata records intact and will evaluate the quality control of the site as usage increases.

Another indicator of broad, multi-institution impact is that interested libraries from around the world have contacted Plymouth State University about making Scriblio available as a free, GPL-licensed download. Because Scriblio was designed to work in conjunction with any collection of bibliographic materials -- including those in commercial ILSs -- the potential audience of users includes most any library worldwide.

### **3. Project Design and Evaluation Plan**

This project is the first in many steps toward a fully functioning OpenLibrary. It is one of many possible approaches to resource sharing, but we believe it is the first to approach the issue from the perspective of the Internet user.

The first stages of project planning will occur through the collaboration of the three partners in a series of face-to-face meetings. The basic timeline and responsibilities for the project have been articulated, but these meetings will enable us to share progress, make midcourse adjustments, and to incorporate feedback from our respective communities as we proceed through development.

The Internet Archive will go live with the first basic version of OpenLibrary.org by November 2007, with book pages for a subset of the records in the Library of Congress Catalog. The first version of the site will have some wiki editing capabilities as well. As libraries and members of the public begin to use the site, we will incorporate their feedback into the development process. Input from the BLC and Scriblio test sites will be critical to refining and prioritizing the functions to be deployed on the website.

In parallel with the ‘test’ site launch, Plymouth State will begin developing the hosted version of Scriblio. Input from single installation test sites will be sought and used to guide further development. This will ensure that we remain in constant touch with the needs of the small library community.

Members of the Boston Library Consortium will assist the development team in understanding some of the larger issues of inter-library resource sharing, including needs and barriers. They will also be invited as early users and testers of the web site, including the functions of hosted Scriblio.

Specific project goals for the two-year period are

1. Fully launch the OpenLibrary.org website with associated physical infrastructure in place
2. Implement preservation policies to ensure the integrity of original records and scanned books
3. Integrate at least 150,000 digitized books into the site
4. Launch the multi-user hosted version of Scriblio on OpenLibrary.org; bring 50 small libraries online over two years
5. Deploy upload and bulk download function for catalog records
6. Develop a prototype “Scan on demand” function and other retrieval options
7. Phase development of search capability within individual library catalogs and across all OpenLibrary collections (the union catalog)
8. Publicize the system and engage public and library usage through conferences, blogging, links from related library and search sites and communication through library associations and consortia.
9. Engage worldwide library community in the resource sharing system.

Specific project activities to support these goals are discussed below.

#### **Work done to date**

The Internet Archive has successfully built and managed petabyte-scale storage and hosting architecture since 1996. The extensive experience in supporting millions of Internet users,

developing search capability for large archived collections, and serving digitized information through user friendly access tools (e.g., the Wayback Machine,) provide the foundation for the development of OpenLibrary.org. The hosting technology and architecture for OpenLibrary.org will be built on the open PetaBox architecture. Selection of open source software for the wiki site and design of the underlying web site and user interface will be ready by the start of the grant period.

Scriblio, initially developed as WPopac, has been developed in a single library per installation version. Scriblio works well to represent a single library, but its architecture, like that of WordPress that it is based on, does not allow a single installation to represent multiple sites or libraries. WordPress has addressed this with a branched version called WordPress MU (multi user) that is available to the community under the GNU Public License and is used to run the WordPress.com blog hosting service, which serves over 250,000 blogs. This application will be the basis for a hosted version on OpenLibrary.org. Scriblio is in use now by Plymouth State University's Lamson Library (<http://www.plymouth.edu/library/opac/>), PSU's Brown Company Collection (<http://beyondbrownpaper.plymouth.edu/>), and Tamworth NH's Cook Memorial Library (<http://tamworthlibrary.org/>).

This project will also build on the work of the Open Content Alliance by linking to the full text of books already digitized by OCA and the addition of scanned books by the Internet Archive. Scanned books are being added to the Archive's public collection at the rate of 12,000 books per month. Scanning centers operated by the Internet Archive in seven locations and by OCA members will be the initial mechanisms by which a "Scan this Book" function can be tested for implementation on OpenLibrary.org. The cost sharing mechanism for scanning will be developed during our ongoing dialogue with the library community. The potential benefit for reducing the cost of inter-library loan and circulation in general will be that a book can be scanned for the cost of a single ILL transaction, and all future patron-initiated requests would be fulfilled at no cost to the hosting library.

We are committed to the use of standards and best practices for open resources for this project. This will facilitate broad participation and easy interoperability with library systems. Some of the standards we will use:

- Data import using accepted metadata format, including MARC and Dublin Core
- Data output in original format as well as potentially in microformat-encoded XHTML via RSS, Atom, and OpenSearch.
- OAI data harvesting

### **Proposed Project Activities**

Grant funds will be used to support three main branches of activity – a) development of the website functions, b) feature development and catalog integration, and c) outreach to the library community and the public for participation in building and sustaining the OpenLibrary system.

#### **a) Site development**

Year 1:

- Build database(s) and related software for underlying structure of the web site, test and revise until stable using BLC and small library partners as testbeds.
- Develop a wiki language that will be familiar to users but also support the structured data inherent in building a union catalog

- Develop methodology and tools to ingest records from sources and convert to a usable online format while still preserving original records
- Design website user interface, including wiki templates for data entry by end users
- Gather catalog data from libraries, publishers, and possibly aggregators – this will be an ongoing process throughout the life of OpenLibrary.org, but will be particularly intensive within the first 3-6 months of the grant period.

#### Year 2

- Develop functions to support the choice points – e.g., Find at a library, download, borrow via ILL, purchase and ‘Scan this book’
- Develop prototype ‘Scan this book’ system to match requests with physical copies, first priority at locations with scanning stations
- Explore patron-initiated ILL transactions, even if mediated by libraries (e.g., affiliation with a library branch required for transaction to be completed.)
- Develop private-label domain hosting in Scriblio MU so that, for example, tamworthlibrary.scriblio.net becomes tamworthlibrary.org.
- Explore possibility of integrating OpenLibrary with an open source ILS program to provide low/no cost solutions for libraries (and then integrate it if possible) – 6 months?

### **b) Feature development and catalog integration**

#### Year 1

- Implement Scriblio in WordPressMU so that libraries can go online simply by creating a username and password on the hosted service.
- Develop self-service batch loading tools that automatically match works in one library's collection with works shared by other libraries.
- Implement the multi-user application on OpenLibrary.org and develop seamless interoperability between Scriblio library sites and the OpenLibrary.org site functions, including finding tools that reach into individual library catalogs and easy up- and download of catalog records

#### Year 2:

- Scalability testing and development.
- Develop OAI interface to Scriblio.

#### Throughout project:

- Development of faceted search and browse features.
- Development of ease of use features based on user feedback and usability testing.
- Development of ease of management features to reduce training requirements for those managing the site.

### **c) Outreach and needs assessment**

#### Year 1:

- Present OpenLibrary.org at 4-6 library and open-source conferences
- Conduct at least one roundtable meeting to engage libraries from across the country.
- Meet with members of the IIPC to invite international participation
- Meet with 50 libraries individually to identify their needs and those of their constituents.
- Identify 20 libraries as development partners; bring them online in beta form.

## Year 2:

- Conduct a conference with development partners for feedback and roadmap extension
- Present OpenLibrary.org at 4-6 conferences
- Conduct at least one roundtable meeting to engage libraries from across the country
- Engage the RRSI in outreach and dissemination discussions
- Go live with the 20 beta partner libraries.
- Meet with an additional 50 libraries to identify their needs and those of their constituents.
- Implement an additional 30 libraries in beta testing.
- Active communication with leaders in the library and Internet communities to build consensus and target development efforts.

**Evaluation Plan**

## Measurable project outcomes

1. Launch the OpenLibrary.org website and construct at least 7 million book pages - *Evaluation will be based on tracking catalog records obtained from publishers, libraries and the LC catalog, as well as progress in building the pages.*
2. Launch the multi-user hosted version of Scriblio on OpenLibrary.org; bring 50 small libraries online over two years. - *Evaluation mechanisms will include feedback from libraries in beta testing mode and tracking progress toward engaging additional libraries.*
3. Link at least 150,000 digitized books to the site - *Evaluation of finding aids will be done at testbed sites during the development phase, and success measured in the number of downloaded books once the site goes live. We will provide a feedback channel on the website for users to report problems with finding or accessing digitized copies of books.*
4. Deploy upload and bulk download function for catalog records – *Evaluation will be done through the testbed libraries of the BLC and small rural libraries*
5. Develop a prototype “Scan on Demand” function – *Evaluation mechanism will be the ability of testbed libraries to initiate on-demand book scanning with successful fulfillment, in prototype mode only.*
6. Publicize the system and engage public and library usage through conferences, blogging, links from related library and search sites and communication through library associations and consortia - *Evaluation mechanism will be usage statistics, including numbers of libraries contributing to and downloading from OpenLibrary.org, numbers of new libraries brought online, numbers of book downloads and web traffic to the site overall.*
7. Engage worldwide library community in the resource sharing system - *Evaluation of progress will be based on communications with international libraries through the IIPC and tracking of the number of international libraries either linking to or contributing records to the OpenLibrary catalog.*

**4. Project Resources****Management Plan**

Alexis Rossi of the Internet Archive will be the project manager for OpenLibrary.org. She currently works in the Collections group of the Archive, managing acquisition and integration of new collections into the Archive’s public collection. Alexis will coordinate communication among partners through face-to-face meetings, a project listserv and telephone and email communication. The partners will make decisions about feature development and prioritization jointly based on feedback from testbed libraries and members of the public throughout site development.



Two people will head PSU's efforts: Lichen Rancourt and Casey Bisson, both working now on the Scriblio project within PSU's Lamson Library. Lichen, who led the implementation with Cook Memorial Library, will lead outreach, training, and documentation efforts; Casey, who developed WPopac, will lead software development. Both will collaborate with Internet Archive on development of shared tools and resources.

Barbara Preece, in her role as Executive Director of the Boston Library Consortium, will coordinate the roles of BLC members as testbeds for the OpenLibrary.org site. These academic libraries may contribute special collections to the union catalog as their mechanism for evaluating the ease of use and functionality of the OpenLibrary site. The BLC will participate heavily in the dissemination phase of the project. Barbara will work with Alexis and Lichen to coordinate meetings and presentations about OpenLibrary.org within the library community.

See the Budget Narrative for discussion of the specific use of grant funds to support the project.

## **5. Dissemination**

The value of OpenLibrary.org and hosted Scriblio as resource sharing tools is entirely dependent upon the participation of libraries and individuals in building site content. Outreach will therefore be an integral aspect of the project from the outset.

Information about the project will be made available online by the Internet Archive, the BLC and PSU. Progress on developing the site and tools to assist small libraries will be made transparent through extensive use of blogging in the library and open source software communities. PSU will also conduct a formal public relations campaign and will coordinate joint announcements of the project. PSU's efforts will be focused heavily on exposing the system to small and rural libraries around the country. One hundred library visits are planned over the two-year period in order to engage this segment of the library community.

The Boston Library Consortium's visibility within the academic library community – and especially its leadership role in the Rethinking Resource Sharing Initiative – will be an important element in promoting this project. BLC will host at least two broad-based roundtable meetings during the project to engage many sectors of the library community and to raise awareness of OpenLibrary.org.

The project team has budgeted funds to make presentations at several conferences throughout the two-year project period. We will also contribute articles to online and print journals that serve the library community. The Internet Archive will participate in an Openness Summit planned by the BLC in the first year of the project, and will at that time share progress on the OpenLibrary project. This summit will be an important opportunity to engage more libraries and members of the open source community in discussion of features and functions to be added to the site.

Brewster Kahle, founder and Digital Librarian at the Internet Archive, is a well known and popular speaker in the Internet and library communities. Casey Bisson and Lichen Rancourt are well known to small libraries as champions of their needs. All three, with others from the three partner organizations, will speak about the project at conferences across the country and internationally.

The power of the online community in dissemination is not to be underestimated. Already, the Scriblio software has shown success in placing library content on the web and generating 'viral' awareness and involvement. PSU's old OPAC enjoyed about 750,000 page loads annually, but the Scriblio-based version has seen almost 20 million page loads within its first year. A considerable portion of that traffic is from search engines, and the results can be seen in web searches for "Joe Monninger," or "A Baby Sister For Frances," and others, which show Lamson Library near the top of the results.

## **6. Sustainability**

The long term sustainability of this project depends on keeping the centralized technical support costs low and engaging the library community and the public at large for contributions and curation services. As noted above, the growth in system openness, user-generated content and wikipedia-like systems gives us some assurance that, once engaged, the library community will support continued growth of OpenLibrary.org.

Once OpenLibrary.org has been launched and 50 libraries have been brought online through the hosted version of Scriblio, we envision continuous impact beyond the life of this project. The model is completely in line with needs identified by the library community to facilitate resource sharing and Internet trends in user-generated content and the desire for better finding tools. The wiki structure of the site is intended to make it easy for the library community itself to support growth of the resource. One-click upload of catalog records, bulk download, metadata editing and other features plus the ease with which patrons can find and request books will help this to become an organically growing, self sustaining initiative.

The Internet Archive is committed to supporting the infrastructure for the website in perpetuity. Since OpenLibrary.org will provide an access point for distributed resources, the need for growing physical architecture is expected to be modest. We anticipate a natural evolution in library and patron behavior to participate in this federated, linked system. Libraries will be able to focus their resources on local needs while tapping into the OpenLibrary union catalog as needed.

Preservation of digitized books and library catalog records will be a distributed responsibility, with each library responsible for ensuring the integrity of their collections. PSU will continue to provide support for the hosted Scriblio application as part of their broad program of services to small libraries. The Internet Archive has preservation policies in place to ensure the integrity of data stored on its systems and will continue to adhere to these policies.

Financial sustainability of the Internet Archive has been demonstrated by ten years of successful operation as a nonprofit with a broad funding base. Plymouth State University is a financially sound institution that has demonstrated its ability to generate funding for the Scriblio application, and we expect that ability to continue. The Boston Library Consortium has a longstanding mission to promote and serve the needs of its member libraries and will support continued experimentation with OpenLibrary.org as a new model for meeting those needs.