



Using Web 2.0, Open Source Technology and Social Networking Services to Facilitate Collaboration and Access to Genealogy and Local History Information

Melvin P. Thatcher
FamilySearch
Utah, USA

Meeting: 204. Genealogy and Local History

WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL
23-27 August 2009, Milan, Italy
<http://www.ifla.org/annual-conference/ifla75/index.htm>

Abstract:

By understanding and utilizing the Web as a platform, organizations can leverage the Internet as a structure that will enable record custodians (archives, libraries, and museums) to preserve, publish and provide access to data anytime, anywhere, at less cost. Many organizations in the genealogy and local history environment, whether large or small, face three basic challenges as digitized records become increasingly important. These three areas include:

- *Capacity: How does an organization create or increase its capacity to digitize and publish records?*
- *Technology: How can an organization take advantage of today's technology with limited technical expertise and budget?*
- *Opportunities: How can an organization become part of a larger community and leverage its combined knowledge and assets?*

While Web 2.0, Open Source and Social Networking Services have been established in the Internet environment for a period of time, it is important to consider these developments from a holistic approach when trying to understand how to leverage the power of the Internet. This paper will present issues that should be considered and how they apply in the genealogy and local history space.

Introduction

The theme of this session, "Opening up our cultural heritage through digitisation and collaboration," states concisely the path FamilySearch International (referred to hereafter as "FamilySearch," d.b.a. Genealogical Society of Utah) has chosen to follow. Therefore, we are pleased for this opportunity to share some thoughts on this topic.

Our concern is how to make more digitized and electronic record data available to more people faster at lower cost to institutions and users. Encouraging and supporting the creation of open information systems and global, participatory communities of users, software developers, and information service providers are critical components of this endeavor. We are confident that the contributions of participants in vibrant global information communities to open information systems will result in making more record data available to more people faster and at lower costs for everyone.

The utilization of new technologies and web services are essential to opening up and facilitating access to our family and local history heritage. This paper will provide a brief overview of how libraries, archives, and museums can and do leverage Web 2.0 technologies and concepts, Open Source software and its community, and social network services to make more of their records accessible sooner, improve access and user services through collaboration, and lower the cost of opening up cultural heritage collections. Examples of relevant developments will be provided from the community and FamilySearch.

Definitions

Before providing and commenting on definitions of the key concepts of Web 2.0, open source, creative commons license, and social networking services, I wish to draw attention to some common themes in these concepts, namely community, dynamic, standards, open, free, freedom, collective, collaboration, developing, and sharing. They are all about making more information accessible through individuals and organizations working together in a free, open, and sharing environment.

Web 2.0. Web 2.0 began life as a technology and was transformed into a way of thinking about and doing business. It “is characterised as facilitating communication, information sharing, interoperability, and collaboration on the World Wide Web. It has led to the development and evolution of web-based communities, hosted services, and web applications.” The participatory aspect of Web 2.0 encourages users to add value as they use an application. Additional characteristics of Web 2.0 include “dynamic content, metadata, web standards, and scalability” as well as “openness, freedom, and [the creation] of collective intelligence by way of user participation.”¹ Implementation of Web 2.0 technologies and business concepts enables libraries to build and leverage patron and peer communities for the common good. The net result can be a better user experience through increased and improved access to popular and little or never used materials in a library’s collections.

Open Source. The purpose of the Open Source Initiative is to make the evolution of software development easy. An open source software license allows for the free distribution and redistribution of the software program and source code in original or compiled form, or source code patches when needed for modifying the program if the source is restricted. It permits the modification and distribution of the program and source code or creation and distribution of derivative works under the original license with no additional licenses required. The license must not discriminate against individuals or groups or fields of endeavor. It must not tie distribution of the program to a specific product, restrict it from distribution with another product, place restrictions on other software that is distributed with

¹ http://en.wikipedia.org/wiki/Web_2.0.

it. Finally provisions of the license must not be tied to a particular technology or style of interface.²

Collaboration communities grow around open source software to evolve the software in order to meet the specific needs of users. Precious budget can be saved and technical expertise accessed by identifying free open source software that addresses your needs and/or by participating in the community as a user to specify additional functionality that is needed. Users are also encouraged to contribute to the community through developing and sharing open source software programs, components, modules, and plugins.

Creative Commons license. The concept of a creative commons for textual materials and data grew out of the open source movement. Creative Commons is a US non-profit organization founded in 2001 that works to counter increasingly restrictive copyright trends. The purpose of this organization is to expand the quantity of creative works that are “available to the public for free and legal sharing, use, repurposing, and remixing.” This is done through copyright licenses that retain some rights for the creator of content while granting specific use rights of the user.³ For example, access to information could be furthered in a Creative Commons license by requiring the output resulting from use of the licensed program to be open to public access.

Social Network Service. “A social network service focuses on building online communities of people who share interests and/or activities, or who are interested in exploring the interests and activities of others. Most social network services are web-based and provide a variety of ways for users to interact, . . .”⁴ Facebook and Twitter are well-known social network services that are being increasingly utilized by North American libraries as interactive ways for building communities with patrons and peers.

Creating and increasing capacity to digitize and publish

Virtually all libraries, archives, and museums, as well as family and local history societies, need more capacity to open their collections faster and improve user services. Cooperating and collaborating with interested institutions with these common aims, mobilizing a volunteer workforce, and using the internet as a platform for collaborative work are important ways of creating and increasing the capacity to digitize, publish, and provide ready access to collections.

Cooperation and collaboration. Resources can be augmented and capacity increased through cooperation and collaboration with other institutions and non-profit and commercial affiliates. Some examples of the types of arrangements that make getting more records data to more people faster and at lower cost possible follow.

The Family History Archive project is an example of institutions working together to make out-of-copyright family and local histories, periodicals and magazines, gazetteers and other materials accessible online. So far more than 35,000 titles scanned by or with the help (loan equipment, provide volunteers if needed) of FamilySearch at seven public and

² <http://www.opensource.org/docs/definition.php>.

³ <http://www.creativecommons.org>

⁴ http://en-wikipedia.org/wiki/Social_network_services.

university libraries are hosted on the website of the Harold B. Lee Library at Brigham Young University (USA) and accessed through links from the participants' OPACs.⁵

The principles of mutual respect and mutual benefit underpin successful, cooperative relationships. In the instance of commercial affiliates that have an interest in the records, this means that agreements may need to have three components: build, operate, and transfer. This will allow the commercial affiliate to obtain a return on its investment before ownership of the product of cooperation ultimately reverts to the records custodian. The agreement that was announced in 2007 between the National Archives and Records Administration (USA) with Footnote.com to digitize and provide access to historical records in the National Archives is a prime example of this type of agreement. Footnote is responsible for creating the digital images. The National Archives can provide immediate access to the images in the reading rooms in its archival system. Footnote.com is permitted to charge for access to the images on its website for a period of five years. Thereafter, the images can be published on the National Archives website for free public access.⁶

Findmypast.com (UK), the Origins Network (UK), and FamilySearch are collaborating to make the 1841 to 1901 England/Wales censuses available online. The indexes will be available free online in Record Search at www.pilot.familysearch.org. The indexes will be linked to Findmypast.com where the images can be viewed for a small fee that permits Find My Past to recoup its costs over a fixed period. FamilySearch members, however, are granted free access to the images.

Volunteers. Harnessing the help of volunteers is another important way to create and increase capacity. Many institutions have a small number of loyal volunteers who come onto their premises to perform any number of tasks. Volunteering can be made more convenient through digitization and utilization of the internet. This will allow tasks such as cataloging and indexing records to be done offsite and the geographical base for recruitment to be expanded, even globalized, if desired.

FamilySearch is a pioneer in this field. It is in the process of moving its volunteer record indexing program online at www.indexing.familysearch.org where volunteers from anywhere on the planet can sign up, download indexing software and images, work online or offline anytime, and submit the results of their work. This service is also available for free to other non-profit organizations who want to mobilize and coordinate the work of volunteers to index genealogy and local history sources of interest to local users. The Ohio Genealogical Society (USA) project to index early Ohio tax records is an early example.⁷

Internet. Now let us position the Internet as a platform for increasing the capacity to deliver and access information and to facilitate collaboration through a brief review of the evolution of electronic access to information.

Prior to the advent of the Internet, introduction of the personal computer facilitated sharing information in files stored on floppy disks. The information consumer was required to use various kinds of proprietary software to access the information based on the programs

⁵ <http://www.lib.byu.edu/fhc/index.php>

⁶ <http://www.archives.gov/press/press-releases/2007/nr07-41.html>

⁷ <http://www.ogs.org/about/taxproject.php>. For a list of current FamilySearch Indexing partner projects go to: http://www.familysearchindexing.org/projects/current_partner_projects.jsf.

that were used to generate the files. Sharing information in static files was made easier through the development of computerized internal networks. Later version control enabled collaboration on dynamic files. The creation of electronic databases increased our ability to store and access information on the network. These conveniences were extended globally within an organization through a virtual private network (VPN).

The rise and development of the Internet and WWW have provided a global network for sharing information and collaborating in trusted relationships without the cost of a VPN when the necessary security and communication protocols and collaboration tools are present. Browser based applications permit access to information anywhere, anytime and facilitate collaboration without the need for other software. Resource requirements have decreased and the capacity for sharing and accessing information and for collaboration have increased because of the Internet, WWW, Web 2.0 technology and business concepts, and social networking services.⁸

Taking advantage of technology today

In order to provide digital access to printed and manuscript materials in their collections, archives, libraries, and museums need a free, or at least an affordable, software toolkit to digitize, describe, and publish and access to records in the reading room and online. Many also need help with storing and hosting digital content. Let us take a look at these technology needs and how they might be addressed.

Imaging software. Imaging software must be capable of high production speeds for large scale digitization projects. It must support capture of big format documents that are of interest to genealogists. A module for creating metadata during the imaging process would facilitate earlier publication of the images. FamilySearch is in the process of developing imaging software with these and many other capabilities for enhanced image quality. It intends to share the software free with interested records custodians. A Twain interface will be provided for using the software with a desktop scanner.⁹

Record description. Robust metadata created after images are captured will enhance the user's research experience. Free, open source tools for describing materials are now becoming available and providing great savings for libraries and archives. For example, many may be familiar with the Evergreen automated library system which is an enterprise-class, open source software.¹⁰ This system is MARC 21 compliant.¹¹ Participating members of the Public Libraries Association in British Columbia, Canada, are expecting a savings on the order of \$10 million dollars over a few years.¹² In the world of archives, the International Council on Archives (ICA) is leading the way with the ICA-AtoM software project to create a free, open source archival description application. It complies with all of ICA's description

⁸ I wish to thank my colleague Dean Taylor for sharing these insights.

⁹ For the really adventurous, here is a communal do-it-yourself website that promotes building your own camera set up and creating imaging software: <http://www.diybookscanner.org/>.

¹⁰ <http://www.open-ils.org/>

¹¹ <http://www.loc.gov/marc/marcsysvend.html>

¹² Peter Van Garderen, "The ICA-AtoM Project and Technology," (2009), p. 4.5, sownladed June 3, 2009 from http://ica-atom.org/VanGarderen_TheICA-AtoMProjectAndTechnology_AAB_RioDeJaniero_16-17March2009.pdf

standards. Dublin Core and Canadian Rules for Archival Description templates are provided. GNU/GPL licensing allows modification of the software to meet other description standards.¹³ Beta release 1.0.7 is available now at <http://ica-atom.org/>.

Publication and access. Free open source software applications for publishing and providing access to digital content are also available.

Dspace¹⁴ and Greenstone¹⁵ are examples of open repository software designed specifically for building a digital library. They have standards based fields for creating metadata, provide simple and advanced search capability, and handle multi-media materials.

Drupal, Joomla!, and Plone are some of the content management systems that can be used to create websites as well as to publish and access digital content. These popular free open source software have caught the attention of libraries and library communities. For example, the PLINKIT Collaborative, formed by a number of state libraries in the USA, has used Plone to create its Public Libraries Interface Kit (PLINKIT) to enable libraries to create their own websites. The toolkit provides templates for the home page and topical pages, a search link to the library's catalog, links to other websites, RSS feeds, etc. The toolkit can be adapted to other kinds of libraries.¹⁶ Drupal was used by the Idaho Commission for Libraries (USA) to develop its E-Branch in a Box toolkit.¹⁷ The Whatcom County Library System (USA) has recently launched a Joomla! powered website that provides access to digital content.¹⁸ Drupal powers the digital library of McMaster University (Canada)¹⁹ and the digital collections of the Art Institute of Chicago.²⁰ The website and content management capabilities of these free tools can potentially reduce the cost of publishing and providing access to digital content.

From this overview, the possibility of assembling a kit of completely free tools for digitizing, describing, publishing, and providing access to family and local history materials is clearly evident. Now the problem is how to cope with the cost of storing and hosting the digital content that can be created with these free tools. Help is at hand!

Storage and hosting. Digital library services are now available to store, host, and facilitate access to digitized records and other forms of electronic media. For example, Internet Archive, founded in 1996, is a non-profit organization based in the USA. Its purpose is to provide "permanent access for researchers, historians, and scholars to historical collections that exist in digital format." Since 1999 it has also been accepting other media such as texts, audio, moving, software as well as archived web pages." It offers book scanning services to libraries, archives, and museums and hosts and provides free access to digital records and information on its website. Books and records in the public domain are eligible for inclusion in the archive. The collection scope accommodates most subjects. Authors are encouraged to donate their digital content under a Creative Commons license or to provide a clear statement of use rights. Internet Archives performs an OCR scan of texts to facilitate user

¹³ Van Garderen, *passim*.

¹⁴ <http://www.dspace.org>

¹⁵ <http://www.greenstone.org/>

¹⁶ <http://www.plinkit.org/>

¹⁷ <http://libraries.idaho.gov/landing/e-branch-a-box>

¹⁸ <http://www.wcls.org/>

¹⁹ <http://digitalcollections.mcmaster.ca/>. To view an example go to

<http://digitalcollections.mcmaster.ca/browseby?op0=OR&filter0%5B%5D=manuscript>

²⁰ <http://www.artic.edu/aic/collections/>

access to the content. The page that provides access to the records of a particular institution is branded with the institution's name. Of course, the institution is also noted in the description of the digital item.²¹

FamilySearch is well-known for storing and preserving information on microfilm. It is equally committed to providing the same service for records custodians who cooperate in the digitization of historical sources of genealogical information. As to the preservation aspect, note that FamilySearch is following a redundancy and migration strategy, while searching for a more permanent solution for this need. FamilySearch is also willing to host family and local history digital content for access from the website of the owner of the records as a part of cooperation in opening the records for online use by researchers.

Building and Leveraging Communities

Free social networking services and tools are made to order for building and leveraging communities of individuals and organizations with shared interests. These developments on the Internet and the Web not only give a lot of exposure for information that participants want to share at no cost, they are also being used by libraries, archives, and museums to harvest useful information about/from users and institutions. Collective community intelligence is used to identify needs and interests, improve user services and experiences, find potential collaborators, and facilitate collaboration.

Using the OPAC. The Social Online Public Access Catalog (SOPAC) created by John Blyberg is a remarkable example of building and leveraging patron and library communities. SOPAC 2.0 takes full advantage of Drupal's functionality to give the patron the ability to rate, tag, review, and comment on titles and to form book discussion groups. The ratings are tied to the bibliographic records to enable the return of search results ranked by patron ratings. Tags are incorporated into the catalog where they are searchable as key words, thus improving the search experience of future users. Social data can be shared between libraries to enrich their catalogs and improve their ability to serve patrons.²²

Using photo sharing. The Digital Past initiative, which was created in 1998 and is managed by the Northern Suburban Library System (NSLS) in Illinois (USA), provides an opportunity for individuals and libraries, historical societies, museums, and other cultural heritage organizations to make their digitized genealogy and local history holdings available for free on the web. Individuals are invited to submit images through the Digital Past group page on Flickr or with the help of a participating institution. NSLS provides backend support and facilities and training for more than 40 institutional participants who are responsible for digitizing, describing and submitting their own content. The initiative employs

²¹ <http://www.archive.org/index.php>. See the Allen County Library(USA) page at http://www.archive.org/details/allen_county. The World Digital Library which was recently launched by the Library of Congress also performs many of these same services; however, its interests are more selective; see <http://www.wdl.org/en/>

²² <http://www.thesocialopac.net>. For more background see <http://www.blyberg.net/2007/01/21/aadlorg-goes-social/> and listen to an interview with John Blyberg at http://lisnews.org/audio_interview_john_blyberg_creator_sopac_social_opac. Create and account and take a tour at the Darien Library (USA) website at <http://www.darienlibrary.org/catalog>.

CONTENTdm as its publication engine. The metadata templates are based on the Dublin Core standard. And most importantly, the Digital Past initiative is OAI (Open Archives Initiative) compliant,²³ meaning that it observes interoperability standards that promote the dissemination of digital content and the harvesting and sharing of metadata.²⁴

Using the Wiki. The purposes of the FamilySearch Wiki at <http://wiki.familysearch.org> are to facilitate the creation and sharing of information about all things related to family history while building a community of contributors and users. Individuals and staff of institutions are welcome to contribute and edit articles for the benefit of all users of this open website. Institutions are encouraged to make use of this wiki in serving family and local history patrons.

Using forums. FamilySearch is beta testing a genealogy forum website. It can be accessed through the FamilySearch Wiki. When it is publicly launched, the URL will be www.forums.familysearch.org. Among other things, the website will allow registered participants to form public and invitation-only social groups around topics of shared interests.²⁵ The forums will facilitate collaboration among family historians while helping FamilySearch obtain feedback on user needs and concerns which can be used to guide its records access activities and to improve patron services.

Conclusion

More records data can be made available to more people faster and at lower costs through cooperation and collaboration between interested parties and by leveraging Web 2.0 technologies and business concepts, open source software communities, and social networking services. When a library, archive, or museum does not have enough resources to make its records accessible, it may turn to other institutions and non-profit and/or commercial affiliates to pool resources and share the workload needed to open its records and provide online access. Creation of a community of volunteers to index records and perform other relevant tasks will increase an institution's capacity to provide good access to its records. Judicious selection of free open source software will give an institution the tools that it needs to digitize, describe, and provide access to its records. Participating in the open source development community will provide an institution access to technical expertise that it cannot afford. Taking advantage of storage and hosting services will ensure long-term preservation of digital content and make it accessible to users at reduced cost to the institution. Using the collaborative tools and concepts of Web 2.0 and social networking services to obtain feedback from and build communities of users and peers will help an institution collect more family and local history materials, provide better access to more of its collection, and attract more volunteers. By adding a donate button to your website you can allow satisfied users to make voluntary monetary contributions towards access to more records and further improvements!

²³ <http://www.digitalpast.org/>

²⁴ <http://www.openarchives.org/>

²⁵ <http://www.familysearchsupport.org/forums>