PaLA | College & Research Division (CRD)

Pennsylvania Libraries: *Research & Practice*

Open Source Opens Doors

Repurposing Library Software to Facilitate Faculty Research and Collaboration

Sandra L. Stump & Rosemary L. Deegan

Sandra L. Stump is Associate Library Director/Collection Management Librarian at Albright College Library, sstump@alb.edu
Rosemary L. Deegan is Library Director at Albright College Library, rdeegan@alb.edu

Asked to convert a faculty-created Microsoft Word document of biblical references found within popular films into a searchable database for scholars, the Albright College library staff helped create a multi-access database called *Bible in the Reel World.* The database relied on student workers for inputting data, used MARC standard formatting for future portability, and encouraged interactive feedback, enabling scholars to submit comments and suggest additional films and references. Using the open source integrated library system Koha, MarcEdit software, and free record exporting from IMDb, library staff created a fully searchable database for researchers and scholars to examine the use of scripture in popular film.

Background

In the fall of 2011, Dr. Robert Seesengood, a faculty member in the Albright College Religious Studies Department, approached the Provost of the college to obtain support in constructing a fully searchable database from a large Microsoft Word document created with a colleague from Britain, Dr. Larry Kreitzer from Regent's Park College, Oxford. Titled *Bible in the Reel World: A Reference Guide to Scriptural Quotations and Allusions in Film*, the 285-page document was an alphabetical listing of almost 1,100 popular films analyzed for scriptural references. Each entry included cast names, the names of characters, the film's ratings, a description of the film, and an analysis of each film for scriptural references.

Dr. Kreitzer wanted to share his many years of documenting films because the study of the Bible and film had been growing in recent years. Patrick Gray (2006) noted the growing popularity of Bible and Film courses where "an increasing number of instructors are incorporating movies into existing biblical studies courses on an ad hoc basis." Furthermore, at the 2008 meeting of the Society of Biblical Literature, the conference agenda listed a new unit

formation session for Bible and Film (Society of Biblical Literature, 2008). Renamed Scripture and Film in 2012, this program unit "focuses on the critical analysis and interpretation of Bible/Jesus films and other films incorporating biblical themes or motifs in terms of the films' biblical and extra-biblical content, cultural and historical significance, and ideology" (Society of Biblical Literature, 2013). With this growing interest in the intersection of the Bible and film, any tool that could facilitate research in this area would appeal to scholars and students.

The Provost asked the library to assist with the project, and it was against this backdrop that the Albright librarians became involved with the *Bible in the Reel World* project. Determining the crucial needs for the proposed database involved lengthy discussions between librarians and the faculty. The process to create the database took two years of back-and-forth communication with the faculty and pushed the librarians to use their training in information retrieval to help generate a dialog that focused on end user needs. In addition to the basic need to convert the Word document into a web-accessible database with little to no money, it was decided that the final database should have

- flexibility of searching by using as many entry points as possible;
- easy data entry so student workers could do most of the work with minimal training;
- ability of worldwide scholars to login and submit comments or new film suggestions through the interface; and
- future sustainability and portability.

With the parameters established, the discussions focused on finding an appropriate software platform for the database. As a result of minimal financial support for the project, open source software was investigated. After examining a variety of library and archival database products, staff decided that adapting an integrated library system (ILS) product would provide the desired search functionality and access. Loading an open source ILS software product onto a library server was deemed feasible because the library had a technician on staff with the skill set necessary to install and configure the server and software. Staff reasoned that by converting the film entries to MARC bibliographic records, the ILS would provide the ability to search names, subjects, and full-text content notes fields. Since student workers were going to provide the backbone of the labor, it was hoped that using authority-controlled fields would provide standardization to aid in the uniformity of the entries.[1] Because the library was already considering Koha as a possible replacement for its ILS, the group decided to experiment to see if it could be adapted for the project.

Koha (koha-community.org) is an enterprise-class ILS that has modules available to support all of the major library functionality services, such as circulation, cataloging, acquisitions, and serials. Originally developed by the Horowhenua Library Trust in New Zealand in 1999 and built on standards-compliant technologies, the Koha software is freely available to download and licensed under the GNU General Public License.[2] The Koha community is vibrant and supportive, and the software receives major upgrades about twice a year. Although open source, there are vendors who provide Koha hosting and support, the two largest in the United States being Liblime (koha.org) and ByWater Solutions (bywatersolutions.com).

Figure 1

Librarians at Albright College chose the open source Koha ILS as the platform for the Bible in the Reel World *project*

Process

Once the database platform was selected, the library technical support specialist downloaded the current Koha 3.8 version and installed it on a library server. Since there was no need to provide circulation, acquisitions, and serials information, those modules were turned off during the initial configuration, using only the cataloging and online catalog display modules. Several changes were made to the basic cascading style sheets (CSS) and the main page was simply configured to introduce the product. In most cases, configuration defaults were selected to keep future software updates easy to maintain.

While the technical support specialist built the database framework, students and staff looked for established MARC records for the films to minimize the typing and cataloging knowledge required to build the database. Library staff located a free Internet Movie Database (IMDb) to MARC converter [3] which allowed the user to enter the IMDb number or the movie title and return a MARC record with tags for film genre, cast, the Motion Picture Association of America (MPAA) rating, the poster image, and a brief synopsis. Each record was exported to a MARC file with a .mrc extension. Once records were downloaded, staff used MarcEdit,[4] a free software program designed to manipulate MARC records, to compile the single records into multi-record files capable of being uploaded to Koha.

Once the records were captured from IMDb and imported into Koha, project coordinators decided to accept the IMDb data without modification. Through trial and error, the librarians established a set of inputting procedures that adjusted basic cataloging rules both to facilitate data entry by anyone with minimal training and to enable the searching and online display features envisioned by the faculty.

With the records now in Koha, the next phase of the project involved transferring the film descriptions and the scriptural analysis text from the original Word document (see Figure 2) into the formatted records. Bowing to basic cataloging principles, the library staff used four major MARC tags to format that information. These were 520 (summary field), 630 (subject tag for uniform or standard title), 650 (subject tag for topical entries), and 700 (personal name). Although the basic tags and recommended indicators were used, no attempt was made to define the information by subfield codes, e.g., the subject form division codes in the 650s.

To stay true to the original document, it was decided to replace the IMDb-supplied film summary with the one crafted by the faculty member. Students and staff typists cut and pasted this synopsis into the original 520 tag. Since a 520 tag can be duplicated, the description of how and where the scriptural references were used in the film was copied into a second 520 tag. This double tag format kept the stylistic separation that was apparent in the original Word document and facilitated clarity in the online catalog display.

> **Absence of Malice (1981) Director: Sydney Pollack Starring Paul Newman as Michael Gallagher, Sally Field as Megan Carter, Bob Balaban as Elliot Rosen, Melinda Dillon as Teresa Perrone, and Wilford Brimley as James A. Wells. [116 minutes] Rated PG (UK TV).**
>
> The son of a long-dead Mafia boss is falsely implicated in a murder investigation and decides to take on the prosecutor who is engineering it. A newspaper reporter gets involved in the case and is duped into printing false stories about the case before realising the damage she has done, including pushing a friend of the accused man to suicide.
>
> At the funeral service of Teresa Perrone the minister presiding reads from John 12:23-25. At the climactic confrontation between all of the principal players in the investigation before the Assistant Attorney General James J. Wells, one of the district attorneys admits that there was a leak to the press in the proceedings. Wells replies, 'A leak! You call what is going on around here a leak? Boy, the last time there was a leak like this, Noah built himself a boat!' This is a reference to the story of Noah and the ark contained in Genesis 6:14-9:8.

Figure 2

This sample film entry from the Word document, Bible in the Reel World, *includes the film description, the film summary, and scriptural references*

Once the 520 summaries were entered, the typists examined each entry not only for the scriptural references but also for any other keywords, phrases, or names that scholars might choose to search. Once identified, they were entered as authority-controlled, index-searchable entries. The Koha software was set to create authority records automatically, so entering these additional terms into authority-controlled tags allowed hyperlinks to be created in the online catalog to assist scholars in finding information more readily.

Thus, all references to scripture and important subjects, characters, and places, e.g., Noah or Garden of Gethsemane, were captured and entered as subject terms. All records were identified as Old Testament and/or New Testament, and these terms were added in the 630 subject tag to group the information together. Any additional scriptural references were entered into additional 630 tags using the common order of books of the Bible. For example, the tags for the movie *Absence of Malice* are listed as

630 04 Old Testament

630 04 Genesis 6:14-9:8

630 04 New Testament

630 04 John 12:23-25

The entry order of books was deemed important since the display in the online catalog matched the order of entry in the MARC record. The authority-controlled 630 tag enabled all films with a specific reference (e.g., John 12:23-25) to be retrieved by clicking on the link in the online catalog.

All of the topical references, such as subjects, places, and character names, were entered into 650 topical subject tags. As described above, this method provided some standardization and more importantly, allowed Koha to create authority records and hyperlinks to these key terms as well. Finally, staff compared the 700 (personal name) MARC tags imported from IMDb with the names listed in the original Word document and added additional 700 tags as needed. The final result is seen in Figure 3.

Since the work on this project was interspersed with daily library work and had no real budget, the project took over two years to get to final testing. During that time, Koha had a major upgrade to version 3.10, which was easily incorporated. Much staff time was spent making Koha look and feel like the faculty wanted, so the CSS files that control the layout were adjusted and labels were added to the notes fields. Testing confirmed that Koha was satisfactory for the faculty needs, and final editing and future control of the project will be transferred to the

Figure 3

An online catalog view of the Absence of Malice *entry—Note the subject and personal name authorities as well as the bolded labels that identify each segment of information; Koha allows authorized patrons to log in and provide comments, a requirement of the project*

faculty shortly. This use of Koha demonstrated the possibilities of this technology for use with any large collection (e.g., equipment or tools) needing circulation and tracking that could be converted to some form of the MARC record.

One unique feature built into Koha was its Google-like keyword searching with automatic truncation and retrieval of alternate spellings in case of typographical errors. This could result in high or slightly misleading retrieval. At this point, the project coordinators configured the software to allow this flexibility to accommodate those more comfortable with Google-style searching. Scholars needing more precision and reliability could use the advanced search options, field designations, or Boolean logic to find their materials.

Lessons Learned

Believing that open source materials showed great promise for expanding library operations, and following the Albright College motto of "a different way of thinking," library staff enjoyed working on this project. Confident that this project was a unique use of an open source library product, they felt this database demonstrated how such a product could be repurposed and adapted for uses beyond the traditional ILS. One caveat: Open source software, although free, requires staff with technological expertise and a willingness to try new things.

The Albright College Library staff was able to complete the project on a shoe-string budget because the major skill sets necessary to successfully implement the Koha installation and populate the database were already in-house. Although the Albright Library had a computer specialist who could install, configure, and troubleshoot a Linux server and understood advanced programming like Perl/C++ and basic web creator skills in HTML and CSS, these skills could be obtained for a price from a third party-managed site, such as one of the Koha support vendors. In addition, there were library staff members who not only understood basic cataloging practices but had also experienced the original automation of the library's collection and understood how to configure Koha correctly to obtain the desired search results.

The resulting database available at ibf.albright.edu encompassed more features and possibilities than the faculty envisioned when they first proposed the project. However, the advantages were not just for the faculty. The project team members realized two major benefits for the library. Primarily, work on this project and the increased familiarity with Koha encouraged the Albright Library to select it as the new ILS software. After investigating a variety of options for a new ILS, the library decided to move from its outmoded proprietary system to an open source product. Working with Koha allowed staff to understand the additional configuration options available. The library migrated its data and had its Koha-based *cataLION* library system (catalion.albright.edu) operational by the end of January 2013. Secondarily, a closer working relationship developed between the library and faculty. Confident that the project enhanced the library's reputation for flexibility, creativity, and collegiality with the college administration, library staff gained the confidence to try more innovative and collaborative solutions to solve future problems.

Notes

[1] For more information about authority control and its importance to standardize catalog search results, see libraries.vermont.gov/sites/libraries/files/tsu/lesson1authority.htm

[2] For Koha information, see koha-community.org/download-koha; for GNU General Public License details, see www.gnu.org/licenses/gpl.html

[3] The MARC converter was supplied by Charles Ledvina at libcat.org. According to the site, "Libcat.org hosts two web applications which create MARC records from data freely available on the world wide web. These converters transform XML product descriptions into bibliographic records which are readable and loadable by most integrated library systems (ILS)." Libcat.org provides an *Amazon to MARC convertor* as well as the *Internet Movie Database to MARC* convertor at amazon.libcat.org/cgi-bin/imdb2marc.pl that was used for the project.

[4] MarcEdit was developed by Terry Reese at Oregon State University (people.oregonstate.edu/~reeset/marcedit/html/about.html). The software can be downloaded at people.oregonstate.edu/~reeset/marcedit/html/downloads.html

References

Gray, P. (2006). The Bible and film: Current pedagogical trends. *SBL Forum.* Retrieved from http://sbl-site.org/Article.aspx?ArticleID=479

Society of Biblical Literature. (2008). Bible in film: New unit formation session. In *Annual Meeting Session Guide* (p.83). Retrieved from http://www.sbl-site.org/assets/pdfs/2008_SessionGuide_Mon.pdf

Society of Biblical Literature. (2013). Program units. Retrieved from http://www.sbl-site.org/meetings/Congresses_ProgramUnits.aspx?MeetingId=23