

An Interview with Terry Reese

Abstract

In July 2005, Bonnie Parks spoke with Terry Reese, digital production unit head at Oregon State University's Valley Library. In this interview, Reese shares his views on topics ranging from electronic journal management, the role of the catalog, to the impact Google scholar might have on the way libraries provide access to information.

Reese is the author of several freely-downloadable software applications including the EBSCO Records Wizard (ERW) and the often cited MARCEdit. He has written hundreds of utilities and tools, many of which are used in libraries around the globe. Still considered a relative newcomer to the profession, Reese was named one of Library Journal's 2005 "movers and shakers."¹ In his spare time Reese enjoys spending time with his family, riding his bicycle and writing code.

Professional Questions

BP: What sparked your interest in working in the library field?

TR: To be honest, I didn't start out wanting to be a librarian (though I think many people that end up in the field can probably say the same thing). I kind of fell into it and found that I enjoyed the work. I always thought it was odd because I've never been a real voracious reader. I have some favorites, but I never have been the type of person that can simply read for the fun of it – I always read with a goal in mind. But that's neither here nor there. As I

was saying, I kind of fell into it. When I first started college at the University of Oregon (UO) in 1995, I took a position at the university's Map and Aerial Photography Library. It didn't take me long to realize that I loved working with maps and aerial photography, so I stayed in the map library throughout my time at the UO. However, I probably wouldn't have chosen librarianship as a profession had it not been for two reasons: Peter Stark, the head librarian at the time, and the bringing of Geographic Information Systems (GIS) data and development into the library. First, I credit much of my personal success in libraries to Peter. His enthusiasm was infectious and I really appreciated his patience with me when I first started cataloging cartographic materials. I honestly can say that while working with Peter, there was a time where I stopped seeing the work that I did as a job and more as a way of life. I found my work and hobbies converging – making my job just fun. Second was the inclusion of GIS into map libraries. I've always been a bit of a techie (I hate that word). I've been programming in one language or another since I was about thirteen or fourteen years old, and by the time I got to college, I really was ready to give it up. I no longer enjoyed it, and for me, writing code has always been something that I do for fun. When the library started dabbling in the world of GIS, though, I finally was able to find a place where I could start programming again and have fun doing it. After doing a little research before graduation, I found that libraries were moving in a direction that I thought I could help facilitate (and have a little fun in the process). So, when a position opened at Oregon State University (OSU) after graduation, I applied to test the waters, and I haven't looked back.

BP: What do you enjoy most about working in an academic environment?

TR: I think what I enjoy most is the variety of the work and the fact that I end up learning something new all the time. It also helps that I enjoy the people with whom I work.

BP: For years, the Integrated Library System (ILS) has been the standard tool used by most libraries to track and provide access to selected resources. Now that libraries are moving away from print and acquiring more and more electronic resources, what do you see as some of the challenges of organizing and maintaining access to these publications?

TR: Libraries traditionally have viewed the ILS as the central knowledge repository, really its most valuable asset outside of the actual materials that the ILS describes. However, I think that we see that we now live in a more distributed world that we did in the past – wait, no, we’ve always lived in this distributed world – its just that only recently, people have finally started to notice. While the ILS never will go away, I think that libraries need to realize that the ILS is now just one small piece of an always-growing digital whole. We have electronic journals and books, but why stop there? Add to them websites, knowledge bases, software applications, et cetera, and what we have is a much more eclectic blend of resources to which our patrons now expect to have access. To some degree, libraries have started to offer solutions to these problems. Many libraries now have OpenURL resolvers to aid users in finding full text from abstracts, and scores of libraries are implementing federated search technologies to allow patrons to search multiple databases from a single search. Also, the library community’s interest in and continued development of services built using the Open Archives Protocol for Metadata Harvesting (OAI-PMH)² are all good first steps. However, with each of these technologies, new problems arise that did not exist in the recent past. When all information was stored in the ILS, there was only one database, one vendor and one very controlled metadata schema that needed accommodation. Now, libraries must account for multiple metadata schemas, controlled vocabularies, vendors, database types and query languages as well as new issues like access restrictions and reliability of the data (will it be

there in one week, three months, five years?). This is in addition to the more mundane problems of de-duping resources from multiple databases and providing relevancy in search results from across multiple targets.

BP: Many libraries are now moving away from A to Z lists to manage access to their electronic resources. They are opting for more sophisticated solutions involving access via the OPAC or using a federated search tool. Do you think A to Z lists are becoming redundant or do they still play a viable role in enabling access to electronic resources?

TR: I actually think that an A-Z list is still relevant today – if only to allow researchers the ability to quickly see if a library owns a particular journal. While this information also is likely to be in the OPAC, I personally prefer the simplicity of the A-Z list when I'm just looking to get to a particular title that is available online. Depending upon how you've created your A-Z list, it also may provide a way to find other like journals quickly, without having to go through the trouble of working with the ILS.

BP: How much do you think technological advances, changing attitudes, user needs, even luck have contributed to libraries' positive approach to electronic resource management?

TR: Well, given the choices above, I think that changing attitudes have probably played the biggest role in how we do things today compared a with few years ago. I know that some will disagree, but I don't think that technology advances really have played too large a role in this. I just think that we have done a better job using the technology that we have available. The only caveat to this is that technology is now much less expensive than in the past, so now more organizations have access to technology than did previously. But back to changing

attitudes. I think the change has been on two fronts. First, in the library community, the ILS, MARC, AACR2 – these have been the profession's sacred cows, so to speak. I think in the past, these concepts were much more immutable than they are today, and I think this is largely due to the current and next generation of professionals coming into the library. Second, I think our patrons want electronic materials and have grown much more comfortable working with these resources. Each new generation that passes through the university seems to be more and more comfortable with electronic resources to the point that they consider having solely a print resource insufficient. And finally, many faculty and researchers, also our patrons, have embraced electronic resources – again, something that is a relatively new phenomenon.

BP: At OSU you created the EBSCO Records Wizard (ERW),³ a program that modifies vendor-supplied MARC records for electronic journal titles in such a way that technical services staff can load thousands of MARC records into the catalog with a very small investment of staff time. This system is particularly effective for providing bibliographic records for titles in aggregated databases, and it also supports record maintenance. Will you tell us more about this system?

TR: I developed ERW primarily because we needed to be able to get a large amount of ephemeral bibliographic data into our ILS system and aid in the overall maintenance of the ILS as titles changed overtime. The project came about when the library started working with EBSCO Information Services. Through our subscription to EBSCO, the library has access to some 19,000 journal titles (though not all unique). But unlike titles from other vendor packages to whom we subscribe, the journal titles purchased through EBSCO change constantly. It was because of their ephemeral nature that OSU Libraries decided that

manually cataloging these serial titles wasn't particularly feasible. However, there was a strong desire to have the journal information loaded into the catalog. The library looked at a couple of vendor solutions (Serials Solutions, EBSCO) and while each would have worked for our situation, the cost was more than the library wanted to invest. So I worked on building a little utility that could take the MARC records provided by EBSCO, and modify and de-dupe the provided records so that in the end, we had one record with each of the various EBSCO access points represented. This included modifications to the records to accommodate our local electronic resources cataloging guidelines. The process allowed me to quickly crunch the 19,000 records over multiple MARC files into a single, 12,000 record file of unique titles that were ready to load. The only remaining issue to address was that of maintenance. Since the records didn't come with an OCLC number, I decided to construct a fake OCLC number for use just in our catalog. The number started at 9999999999999 and then counted down to one. This method had two benefits: First, I knew that all vendor records were loaded in the upper block of our catalog and second, when overlaying these records each month, I didn't have to worry about what title was being overlaid, only the number of titles being overlaid. So long as an equal number or greater were loaded into the ILS each month, all EBSCO titles would be overlaid and corrected. If fewer titles were loaded than the previous month, I simply needed to delete the handful of records that were not overlaid. (This only happened once during the three years that we used the process.) This simplified maintenance, so much so that as we worked with other vendors that provided MARC records, we incorporated the vendors into our process. At its height, I was processing approximately 30,000 journal titles each month (not all unique) through ERW.

BP: Your system is very flexible and responsive to the needs of technical services staff. How important do you think internal communication is when conceptualizing, designing and

testing a system such as ERW? How do you keep the balance so that the librarian utilises the technology rather than the technology controls the librarian?

TR: I definitely would say that communication is important, partially in terms of building a system or tool that can meet your community's needs but also so that your community is aware of what is going on. The desire to have these records included in the catalog was shared throughout the library. I think that had I developed the tool solely within technical services and not allowed reference services to comment, there likely would have been problems, and vice versa. It's a bit of a balancing act: you don't want to clog up the process with too much feedback, but there needs to be just enough to make sure that concerns have been met. Finding the balance, though, is sometimes a hit and miss process. With this tool, I actually wanted feedback relating to how the information should look in the catalog. I also wanted to know what additional data or linkages our librarians or patrons might need to be successful. Since the tool allowed for easy manipulation of the MARC data during the building process, I wanted to provide staff the opportunity to give me their best-case outcomes; that is, how would they want the records to look if everything could be accomplished?

Once I was able to make the tool work, I started working with staff both at OSU and outside OSU to develop an interface that would be simple and straightforward so that anyone could pick it up and use the utility if they wanted. I believe that if someone develops a tool to simplify a process, the developer should make the tool as simple as possible so people can actually use it.

BP: It is a common debate as to whether it is better for libraries to develop management systems at a local level or purchase a commercial product. At Oregon State University, administrators initially chose to use technology to develop their own system to load vendor supplied MARC records for electronic journal titles into their OPAC. You were heavily involved in this work. What do you see as the advantages in developing these tools at the local level?

TR: Well, obviously, tools developed at the local level can be more flexible to your local needs. At OSU, it really came down to looking at what the vendors could provide and what we could do in house. When we looked at the differences, we could get a better product by developing in house at a lower cost. With that said, I should point out that since the library has moved to utilizing Innovative Interfaces' Electronic Resource Management software (ERM), we load vendor records on a much less frequent and more selective basis. Using ERM, we have been able to streamline the workflow of updating and maintaining holdings data within the catalog for electronic materials. In fact, ERM also will generate records for titles for materials not in the catalog. However, this has come at a cost. ERM records are only so flexible. The records are very stripped down, comprised merely of titles, links and holdings data. Because these records are so minimal, we've had to purchase our holdings data from a third-party vendor so that we have an updated knowledge base to load into the ERM software each month. So while we picked up some functionality, we lost other.

BP: The management of electronic resources certainly is a popular project topic at present. Libraries, consortia and commercial ventures are all exploring ways they can use and develop technology to better manage these resources. What is the next technological development you would like to see emerge?

TR: Actually, there already is something that would help make journal management much easier, or at least, facilitate further development, and that is good holdings data. I'd like to see publishers not only make their holdings data available but also produce it in both a standard format and as accurately as possible. I think too often this information is provided simply as an afterthought. Vendors like Serials Solutions, EBSCO and others do a lot of work collecting and normalizing this data, but there remains room for improvements. These enhancements would allow developers to focus on a standard set of data elements and content in those data elements. In thinking about issues that I deal with during our monthly data loads into our ERM software, nearly all of the problems I encounter relate to errors in the holdings data or unstructured date information in the holdings data. Taking it a step further, I'd like to see aggregators and journal producers provide a standard method for allowing their data to be harvestable (maybe OAI) or searchable (SRU⁴ or OpenSearch⁵) to help facilitate distributed search and retrieval. I'd also like to see ILS vendors move more towards an open, modular model. This way, libraries would be able to purchase the best modules from vendors without having to worry whether they will interact with their existing ILS software. A more open system also would allow the library open source community the ability to participate in these developments. Presently, if an institution wants to integrate its electronic resource management with its ILS, there exist very few options.

BP: How important do you think it is that libraries take an active role in finding technology-based solutions to library problems? By employing people like yourself, who have technical skills, creativity and an interest in problem solving? By giving such people the support and tools they need to discover or create new solutions?

TR: Personally, I wish every library had an R&D department. I sometimes worry that libraries take a much too passive role when it comes to solving their own technical problems. I think that this is changing, particularly with respect to the strong role that groups like the Digital Library Federation are playing in pushing the development of digital library solutions. However, I still think that for many things, libraries tend to prefer vendor solutions, even when open source alternatives are available.

BP: Speaking of creating solutions, can we talk a little bit about your MARCEdit⁶ tool? MARCEdit enables librarians to edit MARC records effectively in a windows environment. Like most of your creations you have made the software freely available and MARCEdit is probably your most notable piece of software. What was the impetus that led to its development? How does the tool work?

TR: Well, it started as a hobby application that I began writing during my final year at the UO to get more familiar with MARC. Peter had been mentoring me, teaching and reviewing my work doing cartographic cataloging and I was finding it rather slow going. I'd been reading through the cartographic cataloging resources that we had in the office, but I'm very much an active learner (I don't learn well from books), so I decided to write a cataloging program to learn MARC and familiarize myself with the rules, and it served its purpose for me very well.

When I got to OSU, a large Marcive clean-up project was in the planning stages. The proposed clean up was going to affect approximately 50,000 records and the timeline to completion looked to be fairly long. Well, I dusted off the programming that I'd done earlier and put together a MARC engine that could then be called from a scripting language. At the

time, only the MARC.pm⁷ utility existed, but in testing, the sheer number of records and changes needed seemed to overwhelm the tool. I suppose had it worked, I probably would have used that instead. Anyway, after the project, Kyle Banerjee, who at the time worked at OSU, started needling me to put an interface on the engine and make it available for other folks to utilize. It wasn't something that I was all that interested in doing initially, partially because this was one of those programs I developed in my spare time, and because I wasn't sure if I wanted to support it. But eventually I gave in and made it available. To my surprise, folks were very interested in it, and I've been developing it ever since.

Essentially, the program started out as a windows-based replacement for the U.S. Library of Congress's DOS-based MARCBreaker and MARCMaker,⁸ but it has since evolved into a more complete MARC-editing suite of tools. I wrote the program in C#, so in theory it is operating system agnostic (uses .NET on Windows and Mono on 'nix and Mac platforms) and provides users with the tools that they need to move between MARC and XML metadata schemas. Since the program originally was a MARC-editing tool, it has a MARC WordPad for doing global edits on fields, subfields, text, et cetera, as well as a Z39.50 client, an OAI harvester, a MARC validator for simple validation and tools for moving between different character sets (for example, MARC-8 to UTF-8 and back again). As the program grew, I added an XML API to facilitate the movement of data between different metadata formats. This allows users to create Encoded Archival Description (EAD) records from MARC records and vice versa or generate Dublin Core records from Federal Geographic Data Committee (FGDC) data. With each version, I expand the XML functionality as well as enhance the MARC editor to make the editing of large record sets simpler and easier.

BP: I understand that MARCEdit now is utilized around the globe, and in a variety of different kinds of libraries. What kind of accommodations did you have to make in order to make the program accessible to a diverse international audience?

TR: Well, at first they were small. Last year, for example, I worked with some folks in India to make sure that the MarcEditor could support 8-bit Unicode Transformation Format (UTF-8) characters so that librarians there could create MARC records using one of India's several native scripts. This year, much of the development that I've been doing has been to accommodate my international audience. I think that the two most important developments are the creation of a language engine that will allow international users to use MARCEdit in their own native language (either by using the provided language files or building their own) and building Unicode support throughout the application so that global edits, regular expression matching, et cetera, would all support Unicode-based operations. I'm hoping that both of these changes will make it easier for folks to use MARCEdit to do their work and encourage users to create more metadata in their native scripts for their users.

BP: Google Scholar is the phrase on all librarians' lips at the moment. You played a key role in OSU's arranging for its OpenURL resolver to be added to Google Scholar. What kind of work was involved in this process? What advice would you give to another library thinking of doing the same thing? What do you think this development means for the library and its users?

TR: Google actually has made the process fairly straightforward. In fact, they have a website specifically for librarians who want to add their electronic holdings and use their OpenURL resolver with their service.⁹ In terms of what we needed to do to set it up – well, there is an

XML config file that needs to be edited and placed on your server. Then you need to provide a method so that your holdings information can be harvested by Google on a regular basis using the XML markup language that they have specified. At OSU, we had one additional step: we had to get III to make some changes to allow WebBridge (III's OpenURL resolver) to work with Google's service.

In terms of impact for users and the library, well it's too early right now for me to say. As of this moment, it really is beta, and you can tell. A quick survey of five searches yields nine links to OSU resources on the first results page. However, only three actually work with our OpenURL resolver, and because of missing or incorrect data the others fail to display any resources. So at this point, using Google Scholar to search for OSU resources would probably be more frustrating than beneficial as items would display as being held in Google but not located when the search is resolved by our OpenURL resolver. In the long run, however, I believe that Google will get it working, and depending on how libraries promote GoogleScholar, it may prove useful. If nothing else, it's one more point of access. However, what I find more interesting (and this relates to an earlier question), is how this is another example of the library community looking to shift responsibility for some of its services to an outside vendor, and in this case, a vendor that provides no assurances that the service will be maintained. I think libraries will (and some have) look to using Google to manage access to their electronic materials, and I'm not sure that's a good thing. I'm also not comfortable with the direction the profession seems to be moving towards canonizing Google as the de facto reference source. Were I Google, I'd love it – you can't buy this type of credibility – or maybe you can. Take, for instance, Google's current digitization projects at select libraries around the country.

BP: For one so early in his career, you already have built up an impressive list of projects in which you have been involved and software you have developed, only some of which we have been able to mention in this interview. That large output begs the question, what projects are you currently working on? Will you tell us about them?

TR: The project I am devoting most of my time right now to is an open source¹⁰ federated search tool that hopefully will begin to look at moving away from distributed searching to much more harvesting and custom indexing of metadata to provide consistent search and retrieval across resources.

BP: Most librarians like to stay aware of new library products, tools and solutions. You obviously have a high level of awareness of technological developments. Often your own solutions are designed to work with existing tools and commercial products. Can you tell us about any particular product that you have found inspirational or with which you have particularly enjoyed working?

TR: In terms of the applications that I find myself using most often and enjoy working with the most, they would be IndexData's Yet Another Z39.50 library (YAZ) and SAXON, an XML/XSL processor. Over the last two years, nearly every project that I've found myself working on has somehow involved working with these two products. In terms of commercial vendors, I've found working with DiMeMa's CONTENTdm application (partly because they are open to suggestions and change) and Google's Application Program Interface (API), both traditional and Maps API, a lot of fun. I use the traditional Google API all the time in applications that require search support or spell checking. The Maps API is something that I'm just playing with right now.

BP: What is your vision of the role of the twenty-first-century academic library?

TR: My vision of a twenty-first-century academic library is one that is busy! Actually, I think that academic libraries will need to continue to be both data producers and facilitators.

Producers in the sense that libraries will need to make digital their own materials, particularly those items in Special Collections and Archives, as well as starting to look at digitizing important areas of their print collections. Academic libraries will need to be facilitators in the sense that libraries will need to be able to link their users to these new digital resources. As the publishing model continues to shift towards all-digital publication and government print publications continue to become a thing of the past, libraries will need to continue to be able to build the bridges necessary for their patrons to access this world of information. I think libraries will also find new roles to serve, particularly in terms of becoming more active members of the educational process.

Notes

1. "The Whiz Kid." *Library Journal* Suppl. 130, no. 5 (2005): 26.
2. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides an application-independent interoperability framework based on metadata harvesting. There are two classes of participants in the OAI-PMH framework. Data providers administer systems that support the OAI-PMH as a means of exposing metadata; and service providers use metadata harvested via the OAI-PMH as a basis for building value-added services. "The Open Archives Initiative Protocol for Metadata Harvesting." Updated Oct. 12, 2004 <http://www.openarchives.org/OAI/openarchivesprotocol.html> (accessed July 30, 2005).
3. Terry Reese. "OSU's EBSCO Records Wizard." Updated May 10, 2002 <http://oregonstate.edu/~reese/ebSCO/html/> (accessed July 30, 2005).
4. For an explanation of Search and Retrieve URL Service (SRU) see Eric Lease Morgan, e-mail to Web4Lib mailing list, July 27, 2005, <http://lists.webjunction.org/wjlists/web4lib/2005-July/037994.html> (accessed Aug. 2, 2005).
5. OpenSearch is a collection of technologies, all built on top of popular open standards, to allow content providers to publish their search results in a format suitable for syndication. "OpenSearch." <http://opensearch.a9.com/> (accessed July 30, 2005).
6. Terry Reese. "MARCEdit, Your Complete Free MARC Editing Utility." Updated July 25, 2005 <http://oregonstate.edu/~reese/marcedit/html/> (accessed July 30, 2005).
7. "MARC/Perl: Machine Readable Cataloging + Perl." <http://marcpm.sourceforge.net/> (accessed July 30, 2005).
8. Library of Congress. "MARCMaker and MARCBreaker User's Manual." Updated May 1, 2001 <http://www.loc.gov/marc/makrbrkr.html> (accessed July 30, 2005).
9. Google. "Google Scholar Support for Libraries." <http://www.opensource.org/> (accessed July 31, 2005).
10. Open Source Initiative. "Open Source." <http://www.opensource.org/> (accessed July 30, 2005).