WHITE PAPER

An Introduction to the Rapidly Changing World of ERM Standards

Release 1.0

An Introduction to the Rapidly Changing World of ERM Standards

Ted Koppel
Verde Product Manager
April 2006

I. Introduction

Standards—particularly those approved by national or international standards bodies—have been widely adopted in the library automation industry. From the early days of library systems (arbitrarily assigning the starting date of the library automation industry to the development of the MARC Communications format by the Library of Congress in the mid-1960s), the use of industry-developed and industry-accepted standards has made interoperability and the sharing of bibliographic data possible. As an outcome of a widely accepted MARC platform, other standards and protocols evolved. Z39.50 catalog searching, the Bath Profile, the US National Profile, and several others have their roots in early work done with the MARC record.

Electronic resource management (ERM) is now in the spotlight as a crucial management tool in the library world. Delivery of information through electronic resources has been part of library service for several decades, beginning with SDC Orbit databases, NASA RECON, Lockheed Dialog, and BRS, all of whom delivered abstracting and indexing services. The "modern" era of ERM probably began some time around 1999 or 2000, when the confluence of several different trends occurred. First, technology was such that large collections of full-text material could be stored and retrieved rapidly. High-density magnetic and optical storage hardware costs dropped. Second, the cost of delivery plummeted. Broadband, high-speed access lines, and the ubiquity of the Internet meant than any library, and, indeed, any person, could download large chunks of data in mere seconds. As end users came to depend on the Internet for all of their information needs, an expectation of immediate information delivery resulted. User needs and user behavior changed. Finally, and no less important than the first two factors, economics in the information industry changed. The cost of library resources (paper and electronic) skyrocketed and caused libraries to closely examine their needs. Recession and economic malaise kept money tight and put many libraries into contraction mode. Rather than searching for new services to provide, libraries had to make decisions about which services to cut.

We see the result of these three factors today. In many university libraries, electronic resources constitute 50% or more of the library's acquisitions budget. Some special libraries, particularly in the medical and pharmaceutical industries, spend 90% or more of their materials budget on

electronic resources. This shift from paper-based resources to electronic resources has meant that the old, paper-centric management tools have become inadequate. Our industry reacted by developing and integrating new tools, developed specifically for the complex world of electronic resource management.

This article discusses the standards and protocols that make up the foundation of current ERM applications. It describes existing standards and how they are adapting and changing in an ERM world. Finally, it introduces new, developing standards.

II. The Early Days of ERM Standards

Electronic resource use began to rise in the late 1990s because of the confluence of factors described earlier. For fiscal and other reasons, libraries had to track their contracts, licenses, and similar ER-related data elements. A few libraries developed their own ERM systems—MIT's VERA, Harvard University's Harvard ERM, and Boston College's ERMdb, for example. Most libraries used spreadsheets or Microsoft® Access (or other database) software to organize and maintain their records. Some libraries developed multiple hybrids—one system to handle contracts, another to handle statistics, and a third to handle workflow and acquisitions steps. Of course, some libraries based their record keeping on large collections of photocopies kept in ever-growing file folders. Each of these different record-keeping systems was idiosyncratic; each followed its own rules, maintained locally relevant data elements, and interoperated with practically nothing else.

Cornell University was one of the first institutions to recognize the trend towards electronic resources and the need for management tools. Staff from Harvard, Yale, MIT, UCLA, Johns Hopkins, Cornell, the University of Washington, and others organized the Digital Library Federation Electronic Resource Management Initiative (DLF ERMI) and began to design a specification for an electronic resource management system. This small group worked with a larger "reactor panel" to ensure that there ld be widespread understanding and acceptance of the work being done. In late 2004, this group published, on the Web, a document called Electronic Resource Management: The Report of the DLF Initiative, which was released in paper in mid-2005. This report, along with its six substantial appendices, has come to be known as the "DLF spec." In particular, Appendices A, D, and E (covering ERM function, ERM data elements, and ERM data structure) are the basis of the Ex Libris Verde® ERM system and most of the other competitive systems on the market today.

Parallel with the DLF ERMI project was the creation of COUNTER (Counting Online Usage of NeTworked Electronic Resources). The COUNTER group defined sets of data elements, several usage reports, and the delivery format for those usage reports. COUNTER's goal was to ensure that statistics that

were released from any publisher would be delivered in the same standard XML format and therefore would be easier for any system working with them to ingest and manipulate. COUNTER put weight to its efforts through the use of its *Code of Practice,* which outlines the minimum levels of data presentation that qualify as COUNTER compliant.

Another standards effort taking place during the early 2000s was ONIX for Serials. ONIX (Online Information Exchange) was an initiative of the publishing industry, largely in the UK, to enable different players in the supply chain to share information about books being published. ONIX for Books is a highly defined descriptive XML structure that carries not only bibliographic elements but also data about a book's color and weight, advertising plans and participation, and the number of cartons of that book that would be shipped on a freight pallet. The serials publishing industry saw the potential for a well-defined group of data elements, and ONIX for Serials was begun. Participating in this project was the U.S. National Information Standards Organization (NISO), and EDItEUR, which stands for the International Group for Electronic Commerce in the Book and Serials Sectors. Together, NISO and EDItEUR created the Joint Working Party for the Exchange of Serials Subscription Information. The Joint Working Party has three subgroups, each working with a different area of serials information exchange:

- Serials Online Holdings (SOH), an XML format for communication serials holdings in electronic form
- Serials Release Notification (SRN), an XML format for the announcement of a new serials issue (or article). The SRN message might be sent by the publisher or a subscription agent to announce the release of an issue and thereby forestall a cacophony of claims should that issue be late in shipping.
- Serials Products and Subscriptions (SPS), an XML format for communicating serials catalog data (from vendor to customer, for example) or details of existing subscriptions

The Joint Working Party and its three subgroups are, at the time of this writing, finishing their work and publishing the three XML formats for trial and use by the library and publishing communities.

In addition to the two standards efforts mentioned earlier, several other developments were taking place. Link resolution (using Z39.88, the OpenURL standard) began with the Ex Libris SFX® link server and has grown to be a necessity for libraries that want to deliver full text to users efficiently. As OpenURL link resolution caught on, metasearching also moved from being a theoretical experimental curiosity to a mature application on its own. Finally, as automation vendors began to deliver OpenURL resolvers and metasearch engines, derivative products, such as automated A-Z lists for the public, became part of the electronic resource service expectation.

In 2006, three years after these seminal developments in the library automation industry, the following trends appear to be continuing:

- Electronic resource adoption continues to grow at double-digit rates.
- Sophisticated management tools such as Verde are entering the ERM market.
- Libraries (customers) increasingly realize that paper-resource management tools do not work for electronic resources.
- Current e-resource management applications *still* require tedious effort to collect statistics, enter license data, and perform many other tasks.

III. Emerging Standards in the ERM Industry

All electronic resources management systems have certain characteristics in common. Among other features, ERMs all promise to organize, retain, store, and report on licensing information and the permissions granted to the library by the publisher or publisher's agent. They all promise to collect, manipulate, and calculate user statistics from all the different vendors and publishers that are supplying electronic resources. Current software requires libraries to collect these data elements by themselves and enter the data into the ERM, tasks that are tedious and time consuming. An ERM program such as Verde, which brings central management to electronic resources, relies on the library's entry of complete and accurate data to be truly useful as a management tool.

SUSHI

SUSHI stands for Standardized Usage Statistics Harvesting Initiative and is a protocol that will allow an ERM system to ask a publisher, through a Web service request, to create and download an XML file (in COUNTER format) of usage statistics for a particular customer during a particular month. SUSHI, when adopted by the publisher community, has the potential of practically eliminating the need for libraries to chase down and retrieve statistical data from hundreds of different electronic resource suppliers.

Original SUSHI participants were Ex Libris, Innovative, Swets, and EBSCO. Ex Libris and Innovative delegates wrote proof-of-concept client applications and Swets and EBSCO wrote proof-of-concept server Software. Early success was reached in November 2005 when both client applications were able to retrieve message payloads (that is, statistical files) from each of the test servers. Since November, Thomson Scientific has joined the group and is working on software that would let their product act as both a server and a client. In addition, a number of other companies in the library automation industry have signed on as observers and are waiting until the SUSHI standard matures.

SUSHI is a Web service, meaning that data is passed between parties using "normal" protocols and ports (80 and 443). With a Web service (as opposed to another telecommunications approach), problems with security and firewalls are minimized. Furthermore, the use of Web services means that

the “payload” being shared between parties is XML, which is a well known and easily parsed data structure.

As of this writing, SUSHI testing among parties continues. The Verde product from Ex Libris will include a SUSHI protocol data collector in version 2.5, to be released late in 2006. The SUSHI steering group is encouraging publishers and vendors to become SUSHI servers (and COUNTER compliant) by making available a server development kit. Since vendor participation is crucial to the success of this protocol and to widespread automated statistical data harvesting, anything that can be done to reduce technological hurdles to implementation is beneficial. A SUSHI meeting is planned in late 2006, primarily aimed at publishers and vendors, to promote the adoption and use of these standards across the industry.

SUSHI planners are looking at several protocol enhancements as the protocol begins to be used and accepted as a means of gathering statistics. Some of the areas of future development include the issuing of a “claim ticket” when a server is too busy to provide the report at the time of a request, and the expansion of the menu of reports available (currently SUSHI handles COUNTER Journal Report-1 [JR1]; COUNTER has defined several other reports which will also be retrievable).

License Expression

A major role for any ERM software is to store and make accessible information that relates to license terms—that is, the rights and responsibilities that a library has with relation to the resource being used.

Current practice is for libraries and vendors (or agents) to negotiate a license, which is memorialized on paper and stored in some piece of furniture. Important access and digital usage permissions may be displayed on portals or Web pages as interpretations derived from the license document. If a library happens to be using an ERM system, entries in the licensing section are usually transcribed from a marked up version of the paper license. This method of license data entry is not only slow and inefficient but is also prone to error (both typographical and substantial, based on content).

The publishing world and the library world (and its ERM systems) would both benefit if license terms were distributed in a machine-readable, structured format that could be easily parsed and loaded. Replacing the onerous manual entry process, an XML loader would immediately fill the ERM’s licensing module with whatever data had been sent by the publisher. Then, if updates were required based on negotiations between the parties, the draft license would be archived and replaced by data from more recent downloads. In principle, the machine updating of complicated fields would be a huge benefit to publishers and vendors, to the libraries that are covered by the licenses and to other consumers of the resource’s data.

For this distribution and loading mechanism to work, there must first be agreement on what terms are to be included in the structure and how they are to be represented. The original DLF ERMI group defined about 160 licensing terms as part of the original specifications. EDItEUR also prepared a similar, but not identical, approach to licensing in its draft document *ONIX for Licensing Terms*.

Rather than work at cross-purposes, these two groups decided to combine efforts and explore whether a single standard for the exchange of license information between libraries and publishers was possible. This combined effort, along with the License Expression Working Group (LEWG), which was created to work on this project, began its work in January 2006. Support for the group comes from the DLF, EDItEUR, and the Publishers Licensing Society (UK). Membership as of mid-February was 59 institutions, including ILS and ERM vendors, publishers, universities, digital rights management organizations (such as the Copyright Clearance Center in the U.S.), and several national libraries.

LEWG is basing its work on a draft *ONIX for Licensing* document released last August, as well as several sample publisher licenses made available by large publishers with an interest in the outcome of this group. As of mid-March 2006, the group is doing its work primarily through Internet discussions and e-mail, and a number of lively colloquies have taken place. Fundamental questions are being discussed—for example, What is a license? Are financial arrangements part of a license or part of a contract?

Work of the LEWG is sensitive, because the results of this effort may have weighty consequences for stakeholders in this group. Publishers and agents are interested in asserting their content ownership and unambiguously describing the rights that they are granting to licensees. Libraries, on the other hand, want to respect the rights of publishers while at the same time providing the greatest access and service to their user populations.

As this group continues its work, it will split into two subcommittees. One, the technical group, will discuss the specific elements and terms to be included in the standard. The other, to be called the "review group," will analyze and respond to the technical group's ideas. In time, the groups will reach agreement on a mechanism for the automated delivery of license terms, and the resulting work product will become the draft for a national or international standard.

Future Opportunities for Standardization

The ERM industry is in its infancy. Systems like Ex Libris' Verde (and others sold by our competitors) have begun to meet the needs of librarians in 2006. However, as libraries integrate ERM functionality into their processes, higher expectations of functionality, and increased standardization requirements will

begin to emerge. The author sees the following areas as emerging opportunities for standardization in the next several years:

a. **IP address communication and a standardized protocol for communicating IP addresses:** Despite the tendency of IP addresses to change as networks expand and network topology at user sites in improved, IP address authentication is still the most used method of user authentication for electronic resources. [In fact, the NISO Metasearch Initiative (Standards Committee BA – Access Management) in 2005 noted in its recommendations that, at the present, IP authentication and password authentication were the two best practices for authentication for electronic resources use.] Communication of IP address changes from libraries to vendors could be made far easier through the use of a communications protocol that would pass IP address information updates and acknowledgments electronically. Some early discussions on this matter took place in late 2005, but no further progress has been made.

b. **Communication of incidents and breaches from ERM to Vendor**: Most ERM systems track performance and service interruptions on the part of the vendor. A system of automatically reporting service glitches, through a communications protocol, would capitalize on the ERM's tracking logs and directly communicate with the vendor's system. Among the benefits of this approach: immediate notification of problems to the vendor, and a large reduction in telephone call handling of problems, because notification would have been automatic.

c. Similar to (b) – A vendor-initiated protocol to advise customers of service interruptions.

d. ERM-to-ERM Communications protocol: Each ERM system vendor wishes, of course, to keep and satisfy their own customers. Nevertheless, there are times when the sharing of information between ERM systems will be a requirement. Intra-consortium sharing of data is an example of this need for sharing. Currently, almost all ERM systems use the DLF ERMI specification as the core set of data elements in their ERM. Sharing those elements is a logical next step.

e. **An international unique identifier assigned to collections (including packages and interfaces):** As of this time, MARC records and unique identifiers exist for e-journals and e-books, but not for the collections (often packages and interfaces) in which they are delivered. As a result, there is no consistent way of referring to collections. Titles are sometimes used for this purpose, and artificial, temporary identifiers are sometimes assigned, but no system that has international acceptance or recognition has been devised. The ERM industry would rapidly take advantage of such an identifier, when created.

IV. ERM Standards and Ex Libris

Verde has been designed as a staff tool to assist in the management of electronic resources. As an ERM system, it sits in the center of a number of interactions—some internal to the library (such as interoperability with an OPAC, a link resolver, and a metasearch engine) and some calling for interoperability with external sources of data. Interoperation across an industry is far easier and more efficient if shared, negotiated, and accepted standards exist. Lacking standards, time and effort are wasted on idiosyncratic solutions to individual problems.

Therefore, by working with vendors and publishers to deliver ERM data efficiently through standards-based mechanisms, Ex Libris aims to enhance the value of Verde as a core component in delivering library management solutions.

Author contact:

Ted Koppel - ted@exlibrisgroup.com