

# Best Practice and Standardization Initiatives for Managing Electronic Resources

by Rafal Kasprowski

## Information Standards

The processes of managing and accessing electronic resources involve a number of participants – libraries, subscription agents, content providers, hosting services – and tend to be complex, time-intensive and susceptible to human error. Licensing, exchange of acquisition data, research and usage data collection are some of the processes that have been performed manually or have no or only poorly implemented best practices or standards. As a consequence, their potential usefulness still cannot be fully exploited (for research, usage data collection), remains generally too time-consuming for many libraries (licensing) or leads to miscommunication (exchange of acquisition data).

With the advent of the online environment, however, opportunities have arisen for unprecedented data integration and linking, which if used to advantage can lead to a high level of automation and an increase in productivity and discoverability. New best practices and standards related to the management of electronic resources as well as new approaches to earlier standardization efforts are currently underway to address these opportunities.

Not all standardization efforts lead to full-fledged standards. Depending on the scope of the problem or degree of interest by the stakeholders some begin as best practices and develop into standards, while others remain best practices. Some of the problems mentioned above have already been partially resolved by recent standardization initiatives; others have only been addressed recently. This report, though not all inclusive, does discuss some of the more important current standardization efforts related to the management and accessibility of electronic resources.

Rafal Kasprowski is with Rice University, where he can be reached by mail at Fondren Library, MS 44, 6100 Main Street, Houston, TX 77005; or by email at [rk11@rice.edu](mailto:rk11@rice.edu)

### Link Resolvers and Knowledge Bases

**Open URL.** OpenURL is a framework standard ([www.niso.org/standards/z39-88-2004/](http://www.niso.org/standards/z39-88-2004/)) developed under the auspices of the American National Standards Institute (ANSI) and the National Information Standards Organization (NISO) to transport packages of information over a network. OpenURL has a broad range of potential applications. This report will limit the discussion of the standard to its use in transferring bibliographic metadata within a URL to facilitate access to library holdings. As access to electronic content is central to electronic resource management, OpenURL plays a significant role in a number of standardization efforts in this area.

Before the advent of OpenURL, if a search yielded no full text for a citation, as in an abstracting and indexing database or a journal issue not held by the library, users would continue to manually search other library resources, such as full-text databases or journals. In platforms where an OpenURL link is set up, an additional click triggers a search that uses the metadata from that same citation to check all of the library's holdings and retrieve the full text if a match is found.

OpenURL linking involves building a URL string with the bibliographic metadata, such as ISSN, volume number or starting page, from a citation following an OpenURL-compliant syntax. The resource where the search and the OpenURL linking process begin is called the *source*. Figure 1 illustrates an example of a source link – or outbound link – for articles (the brackets are provided as an aide to indicate place holders for specific citation data):

FIGURE 1. OpenURL outbound or source link

[http://anylibrary.anyresolver.com/?genre=article&sid=\[source ID\]&issn=\[ISSN\]&title=\[journal name\]&article=\[article title\]&volume=\[volume\]&issue=\[issue\]&spage=\[start page\] &date=\[yyyy\]](http://anylibrary.anyresolver.com/?genre=article&sid=[source ID]&issn=[ISSN]&title=[journal name]&article=[article title]&volume=[volume]&issue=[issue]&spage=[start page] &date=[yyyy])

KASPROWSKI, continued

The link loaded with the specific metadata is sent to the knowledge base that contains the library's online subscriptions, including exact holdings. The information in the link is then checked against the holdings in the knowledge base. When a match is found, the citation data is plugged into an OpenURL string pointing to the site of the full-text provider, called the *target*. Figure 2 is a sample structure of a target link – or inbound link:

FIGURE 2. OpenURL inbound or target link

```
http://www.anytarget.com/openurl.aspx?genre=article&issn=[ISSN]&volume=[volume]&spage=[start page]&date=[yyyymmdd]
```

The target provider may not always use all the different metadata passed on from the source provider, but both follow the OpenURL standard syntax and specifications (e.g., “spage” to mean “start page”). Both of the above linking examples follow the 0.1 standard version. A version 1.0 also exists and follows a similar, but more complex, syntax.

The tool that does all the linking in the background is the link resolver. The link resolver pulls the citation data from the source, sends it to the knowledge base using the OpenURL syntax, checks the data against the library's holdings and builds the OpenURL link to the target. The link resolver is integrated with the knowledge base, the front end of which is the library's online journal interface, also known as an A-to-Z or e-journal list.

For OpenURL linking to work, the source provider and link resolver have to be OpenURL-compliant. The source has to be able to pass OpenURL strings to the resolver via links embedded next to each citation; the target has to make its content available via the URL address generated by the link resolver. The exact target URL structures are maintained in the knowledge base.

**KBART.** To ensure that access to electronic content reflects actual subscription holdings, librarians, content providers and resolver suppliers currently depend on each other for accurate title and holdings information. Content providers send title data to link-resolver suppliers, create correct source links and share proper target data. Link-resolver suppliers maintain correct title and target link data in the knowledge bases using the

information they receive from content providers. Libraries rely on correct information in the knowledge bases, with title data provided by resolver suppliers and holdings data that must be updated by content providers.

Despite this interdependence, no formal policies exist among the three groups regarding the information exchanged, its level of accuracy, formatting, choice of data elements, update schedule or OpenURL compliance. The lack of a unified approach regarding these data exchanges has been a cause of common access problems.

For example, update schedules may differ between the content providers and the resolver supplier or the content providers themselves may not keep their systems up-to-date. The source provider may indicate a recent record, but the target provider may not have uploaded the full text for this record yet, which causes resolvers to create dead-end links. Although most content providers are aware of the OpenURL standard, they may decide not to adopt this linking technology because they do not realize the significant impact it has on the usage of their content. There is also a lack of uniformity in data formats used and in the construction of inbound linking syntaxes among content providers. The source provider may only plug in the year for the date element in its outbound OpenURL, but the target provider may require a more precise date format for linking to its content. Whereas a standard for the linking structure exists in the form of the OpenURL syntax, no standard exists yet for managing the content transported in OpenURL links.

KBART (Knowledge Bases and Related Tools) is a working group formed by NISO and the UKSG (United Kingdom Serials Group) to develop best practices for the exchange of data among suppliers of knowledge bases and link resolvers, content providers and libraries in the knowledge base supply chain ([www.niso.org/workrooms/kbart](http://www.niso.org/workrooms/kbart)). Its first objective is to determine why the problems occur, how they affect the supply chain and what the stakeholders can do to solve them. As of June 2008, the group has developed an initial list of terms that should be clarified and defined. It also built a flowchart to identify areas of miscommunication associated with the processes in the supply chain. The initiative may consider working on a standard based on the best practices developed by its

initial efforts and positive feedback from the stakeholders. Its initial objective, though, is to bring about a unified understanding of the issues and a strong community of practice.

### The Work, Its Manifestations and Access Points

**DOIs and CrossRef.** An identifier closely associated with access to electronic resources is the Digital Object Identifier (DOI). It identifies and is used to link to full-text electronic items, whether they are books, book chapters or journal articles. The DOI System ([www.doi.org](http://www.doi.org)) is currently being standardized through the International Standards Organization (ISO). The final standard is expected to be published by 2009.

DOIs are composed of a prefix and a suffix. The prefix always starts with a “10” and includes a unique number designating a publisher (e.g., “1000”). The suffix is a numeric or alphanumeric combination of any length, such as “182,” or “JoesReport1108.” It can also be another standard number like an ISBN, in which case it refers to a book. DOIs become actionable online through the addition of a base string: <http://dx.doi.org>. An example of a complete actionable DOI is <http://dx.doi.org/10.1000/182>.

DOIs are unique numbers and, like OpenURL links, can also generate persistent links. An ever growing number of publishers use DOIs to identify their content, journal articles for the most part, and store the DOIs in a common repository through a member association, CrossRef (<http://crossref.org>). Link resolvers use DOI links in conjunction with OpenURL links to provide access to full texts. While DOI linking requires the use of a single uniform number, successful linking using the OpenURL system requires the publisher of the source citation and the publisher of the full-text target to follow similar indexing and link resolving practices for most if not all the tags in a link. For example, if a source just provides the year as the publication date to the link resolver, then the OpenURL link may not work with targets that only accept a year-month-day format for resolving publication dates. Because neither the indexing nor the link resolving practices are standardized across the industry, DOI links – when available – tend to be more reliable than OpenURL links.

**ISBN-13.** The International Standard Book Number (ISBN) is a unique identifier for each edition or media version of a book or audio book

recording and provides a means to distinguish between hardcover and paperback editions of a title ([www.isbn.org](http://www.isbn.org)). Products that are not books are currently assigned a UPC (Universal Product Code) or an EAN (the International Article Number, formerly the European Article Number). Through 2006, the ISBN was a 10-digit number (or ISBN-10). The International ISBN Agency determined, however, that it would soon run out of 10-digit numbers to assign to books. In order to ensure that the ISBN remains a unique identifier, it was expanded to 13 digits (or ISBN-13). Existing ISBN-10 numbers now receive the 978 prefix; once the supply of 978 numbers has been depleted, the 979 prefix will be used for ISBN-13s.

The EAN used for the bar code on books (called the Bookland ISBN) is actually identical to the ISBN-13. The ultimate objective of the ISBN-13 standardization effort is to replace the use of the UPC on books with the EAN, which would make the ISBN-13 the standard machine-readable code. In an effort to further standardize product identifiers, the Book Industry Study Group (BISG), has proposed a 14-digit product identifier, the Global Trade Identification Number (GTIN), to be used in the future for all product identifiers and has suggested that organizations expand their database fields to 14 digits right away in preparation for this change.

**ISSN-L.** The ISSN-L, or the linking ISSN, is the new ISO standard for the ISSN. It is meant to allow linking between different media versions of a serial (for example, journal, newspaper and other continuing resources). For example, if the ISSN is 0264-2875 for the print version of a journal and 1750-0095 for its online version, then ISSN-L 0264-2875 could be its linking ISSN. (The table listing the designated linking ISSNs for the corresponding ISSNs is available at [www.issn.org](http://www.issn.org).) While different ISSNs for the same continuing resource are used to distinguish among the printed, online and CD-ROM products the resource may be used for, the ISSN-L is used to group all the media versions under a single identifier to improve content management. The ISSN-L can also facilitate discovery of content across all of its media versions in services like OpenURL. Even if a print ISSN is used by the source provider and an online ISSN by the target provider, the ISSN-L will void this discrepancy and thus help the OpenURL link resolve consistently. It should be noted that the ISSN-L can change when the title of

a resource changes. A resource will then have one ISSN-L for the old title and another ISSN-L for the new title.

**ISTC.** The International Standard Text Code (ISTC) number is meant to encompass all printed and digital manifestations and editions of a particular work. Whereas the ISBN of a work may change depending on the edition of the work or whether it is a print or audio version of it, the ISTC will remain constant for that work. In other words, the ISTC identifies the work separately from the different manifestations or expressions of the work. As of August 2008, an ISO draft standard has been approved for the ISTC and is pending formal publication. An ISTC is a 16-digit number, for example 0A9-2009-12C4E315-F, consisting of four elements: the registration agency element; the year element; the work element and a check digit. In the future, the correspondence between works (ISTC) and their derivative products (ISBNs) will be captured in bibliographic databases utilized by search engines, libraries, retailers and other platforms providing discoverability. For a single ISTC, there could be several editions of the work with ISBNs and DOI links for books and book chapters. Providers are already beginning to make their content accessible using this linking method.

The push for uniformity in the electronic resources environment reflects the need to pull all the manifestations of a work together to make a desired content easily accessible regardless of the editions or media it may exist in. The ISSN-L promises a better success rate in resolving OpenURL links compared to its print or electronic counterparts. The ISTC hierarchy coherently organizes books and book chapters and provides a rationale for the ISBN to be used in DOI links to enable access to these full-text items.

### Integration of Usage and Cost-Related Data

**COUNTER and SUSHI.** Over the past decade, a continuous effort to facilitate the collection of usage data for electronic resources has first led to the development of the Counting Online Usage of NeTworked Electronic Resources, or COUNTER, protocol ([www.projectcounter.org/](http://www.projectcounter.org/)), and more recently to the ANSI/NISO Z39.93 Standardized Usage Statistics Harvesting Initiative, or SUSHI, standard ([www.niso.org/workrooms/sushi](http://www.niso.org/workrooms/sushi)).

The objective of the COUNTER project is to standardize usage data reports generated by content providers in order to help institutions analyze and develop their collections. COUNTER reports include Journal Report 1: Number of Successful Full-Text Article Requests by Month and Journal; Journal Report 2: Turnaways by Month and Journal; Database Report 1: Total Searches and Sessions by Month and Database; and others.

The objective of the SUSHI standard has been to pull data from COUNTER reports and send it to an ERMS using an automated transfer protocol. This process has the advantage of integrating usage data from various sources into a single repository for easier data management. By the same token, SUSHI also promotes the use and interpretation of COUNTER reports by eliminating the oftentimes excessively time-consuming process of usage data collection [1]. Several vendors (including Innovative Interfaces, Serials Solutions and Ex Libris), a usage data collection service (MPS Technologies), content providers (such as EBSCO, Gale, HighWire Press, IOP, Metapress and ProQuest) and subscription agents (such as EBSCO and Swets) have started to implement or are in the process of implementing SUSHI, currently in its final 1.5 version (ANSI/NISO Z39.93-2007). Previous versions were the early proof of concept SUSHI draft (0.1) and the near final SUSHI version (1.0).

Following the adoption of SUSHI as a standard, a committee was formed in mid-2008 to promote its development and implementation. Several developers (affectionately known as the “Sushi-Shokunin” or sushi experts) have been recruited to monitor the SUSHI developers list, help organizations adopt SUSHI and assist early adopters in moving to the final version. A common problem with current implementations is COUNTER data lumped together with the transfer protocol in the SUSHI message. The content of the message and the way the message is delivered are meant to be treated separately, since the SUSHI protocol can be written to retrieve usage data based on other reporting standards. It is projected that these issues will be resolved by the end of 2009. The most recent Release 3 of the COUNTER code of practice, published in March 2008, already makes it mandatory for content providers to support SUSHI in order to be COUNTER-compliant.

**CORE.** As part of the second phase of its charge, the Electronic Resource Management Initiative (ERMI), a collaboration of the Digital Library Federation (DLF), investigated which financial data elements from integrated library system (ILS) acquisitions modules would be suited for cost management from within electronic resource management systems (ERMS). Following interviews with librarians, ERMS suppliers and ILS vendors, an ERMI subcommittee produced a white paper in January 2008 ([www.diglib.org/standards/ERMI\\_Interop\\_Report\\_20080108.pdf](http://www.diglib.org/standards/ERMI_Interop_Report_20080108.pdf)).

Based on ERMI's findings, a group of ILS and ERMS representatives conceived the Cost of Resource Exchange (CORE) standard specifying a protocol for transferring financial data from the ILS ("the source") to the ERMS ("the requestor") in order to facilitate the exchange of cost, fund, vendor and invoice information between both systems. For libraries that have both an ILS and ERMS, sharing information across these systems would eliminate duplicate data entry and promote data integration. Cost-related reports could be run directly in the ERMS and specifically cost-per-use reports when combined with usage data imported via the SUSHI protocol. Ultimately that same standard could also be applied between any two other business systems capable of using this data exchange format ([www.niso.org/workrooms/core](http://www.niso.org/workrooms/core)).

The CORE standard proposal was approved in May 2008 by NISO and has since progressed to the formation of a new working group. Work to be done includes the creation of a data element dictionary, with all data elements clearly defined (e.g., cost = invoice amount + amounts of all supplemental invoices for subscription period). Another important step is completing basic use cases, where the study of key data transfer scenarios will help define important data elements. The data could be requested from within the ERMS, for example, or be pushed to the ERMS automatically whenever an ILS acquisition record is updated. The data could also be delivered in batch loads or in single transactions. Trialing these and possibly other use cases could drive new data elements such as an ERMS resource ID stored in the ILS.

The CORE Working Group plans to focus on core data elements common to most ILS and resource management systems and, as was the

case with the SUSHI protocol, separate the definition of the data elements from the development of the data transfer protocol since the two are not necessarily interdependent. Possible compatibility with existing standards also needs to be investigated, such as the ONIX (Online Information Exchange) for serials standard and related formats: SOH (serial online holdings) format; SPS (serial products and subscriptions) format; SRN (serial release notification). Equally critical to the success of the standard will be library participation in promoting the adoption of the future standard by the vendor and publisher communities.

As with other standardization efforts that have reached this stage, the next step on the CORE Working Group's agenda will be writing a draft standard. After input from interested parties, a Draft Standard for Trial Use (DSFTU) will be written and tested to address any remaining issues. The final draft will be put to a vote and become an official standard if the vote is favorable.

### Coding License Terms and Defining Consensus

**License Expression and ONIX-PL.** The complex process of producing, selling, repackaging and providing electronic content involving its creators, publishers and aggregators has contributed to the proliferation of licensing as a measure to protect the rights of content producers and distributors where copyright law with its fair use provisions does not suffice. Reviewing and negotiating licensing terms is a time-consuming process, however, involving significant administrative cost for both content providers and librarians.

Based on the license element descriptions in the 2004 report on electronic resource management ([www.diglib.org/pubs/dlf102/](http://www.diglib.org/pubs/dlf102/)) produced during the first phase of the ERMI initiative, NISO, DLF, EDItEUR and Publishers Licensing Society (PLS) agreed in 2005 to form the License Expression Working Group (LEWG). LEWG was to investigate how the ERMI data elements could be used, and possibly expanded, to express licenses in a machine-readable format, load them into ERMS, link them to digital resources and communicate key usage terms to library staff and users. Early on in this investigation, ERMI and LEWG drew a distinction between license expression and rights expression, as rights expression languages were considered too restrictive in the context of the inherent

ambiguities of negotiated agreements and copyright law. The concept of interpretation would have to be part of the license expression code.

An XML-based format for license terms that is part of the EDItEUR ONIX (Online Information Exchange) family called ONIX-PL (ONIX for Publications Licenses) was developed, and a provisional draft was published in March 2007. Work on the ONIX-PL License Editor (OPLE) to help librarians code licenses without having to know XML also began with the development of the ONIX-PL format [1].

It became evident in this standardization effort as well that the data to be used for the ERMS had to be chosen in a separate step from the means for delivering this data to the ERMS. In mid-2008, NISO replaced LEWG with two new teams to assume this work. The ONIX-PL Working Group will continue to develop the ONIX license messaging specification, while members of NISO's Business Information Topic Committee will conduct a survey to see how vendors and libraries are applying ERMI's license element descriptions.

**SERU.** The SERU (Shared E-Resources Understanding) initiative represents an alternative approach to a machine-readable license scheme. Instead of defining licensing data elements, SERU is attempting to define a basic set of terms of use that publishers and librarians can agree on and use in situations where a formal license could be considered unnecessary. This understanding would be embodied in a standard SERU document reducing the need for license review and negotiation.

SERU, currently in draft version 0.9, defines common understandings such as subscription and subscribers, appropriate and inappropriate uses, confidentiality and privacy, online performance and service, archival access and perpetual rights. It has been suggested that SERU is one example of a baseline license template from which variants are possible. In fact, once agreement to use SERU is reached, its terms can be loaded into the ERMS.

First conceived in 2005, SERU became a NISO recommended practice (RP-7-2008) in February 2008, with around 20 publishers, over 40 academic libraries and several consortiums registered as participants. To promote the use of this best practice effort, interested parties are asked to register at the NISO/SERU website ([www.niso.org/workrooms/seru](http://www.niso.org/workrooms/seru)) and

use SERU ([www.niso.org/serudraft0\\_9.pdf](http://www.niso.org/serudraft0_9.pdf)) when both parties agree to; otherwise, a traditional license can be used. Price-related terms, such as content, access period and cost, can be written into the purchase order when SERU is used. The next steps in promoting SERU include recruiting highly regarded publishers, targeting platform providers to bring smaller publishers into the fold, and encouraging librarians to sign up and use SERU whenever possible.

### Data Exchange Using Institutional Identifiers

**I<sup>2</sup>.** The NISO Working Group on Institutional Identifiers, the I<sup>2</sup> (pronounced "I 2"), is one of two initiatives developing institutional identifiers ([www.niso.org/workrooms/i2](http://www.niso.org/workrooms/i2)). The work of I<sup>2</sup> builds on the recommendation of the Journal Supply Chain Efficiency Improvement Pilot ([www.journalsupplychain.com](http://www.journalsupplychain.com)) and the initial efforts of Ringgold E-Marketing Services to implement an institutional identifier as a means to improve the exchange of data between all the participants in the electronic content supply chain, which includes libraries, publishers, and subscription agents.

The I<sup>2</sup> number will describe a "licensing unit," a concept used throughout the e-content supply chain and applicable to all the relevant transactions. At the moment the stakeholders involved share some identifiers, but they tend to differ for each transaction. Existing institutional identifiers keep track of locations or business entities, such as SAN (Standard Address Numbers), GLN (Global Locator Numbers) or ISIL (International Standard Identifier for Libraries). However, an institution can be renamed and its address can move due to takeovers, either change quickly leading to miscommunication across the supply chain. A licensing unit identifier would continue to refer to the same entity despite changes in its address and other information. Licensing units are also becoming increasingly discrete, often representing not just entire organizations, but campuses, specialized library branches (main library, law library, medical library, business library), even departments, so a single address may actually represent several different licensing units. While there may be a need to look at the relationship between the current identifiers and I<sup>2</sup>, the objective of this working group is to create a standard number with standard metadata that can be used for all

KASPROWSKI, continued

transactions in the supply chain. The project is currently at the data-gathering stage; a final standard may be approved as soon as mid-2010.

Currently Ringgold is maintaining the I<sup>2</sup> numbers, but once the full standard is implemented, questions may arise as to who will continue to keep the I<sup>2</sup> records up to date and whether any new administrative entities will try to recover the maintenance costs by making this activity a paid service.

**WorldCat Registry.** The other initiative using an institutional identifier is the WorldCat Registry, an online directory developed by OCLC in the past two years to enable libraries and library consortia to maintain a single source for information pertaining to their institutional identity ([www.worldcat.org/webservices/registry/xsl/about](http://www.worldcat.org/webservices/registry/xsl/about)). Participating institutions are designated by a non-editable WorldCat Registry ID.

On a secure web platform, institutions create and maintain a single profile that includes administrative information such as address, IP locations, consortial memberships, main and branch institutions and administrative contacts. Institutions can also promote their web-based services by providing information on their online catalog, virtual reference and OpenURL servers. The institutional information can be shared with third parties, including technology vendors, electronic content or service providers and fellow consortial institutions, via a read-only link and eliminates the need for updating changes to institutional information multiple times with content providers and vendors.

OCLC distributes the registry data across many popular open-source web services, such as WorldCat.org. With links such as the online catalog's

web address or the OpenURL resolver shared in this manner, a broader Internet audience discovers and uses the institution's content.

### Relationships Among Standards

Standardization projects that are better defined from the start are certainly expected to follow a smoother development process, but an equally if not more important factor for a successful standard implementation is broad participation by librarians, vendors, content providers and subscription agents. Both factors become apparent when revisiting older standards that have not garnered the broad recognition necessary to provide substantial improvement to the processes they were meant to streamline.

As the number of standardization initiatives grows, so does the probability of relationships among the final standards. A clear example of this is the complementary OpenURL and DOI linking syntaxes or even the relation between the ISTC and DOI via the ISBN. Institutional identifiers appear to integrate multiple processes. Since the WorldCat Registry ID, for example, defines not only a library's administrative information, but also its OpenURL resolver and the underlying knowledge base, it can be seen as facilitating both the acquisitions process and full-text access.

### Acknowledgments

Beth R. Bernhardt (University of North Carolina at Greensboro); Adam Chandler (Cornell University); Helen Henderson (Ringgold Ltd); Ed Riding (SirsiDynix) ■

### Resources Cited in the Article

[1] For more details on this standardization effort, read the author's report entitled "Standards in Electronic Resource Management," published in the August/September 2007 issue of the *Bulletin* at [www.asis.org/Bulletin/Aug-07/Kasprowski.pdf](http://www.asis.org/Bulletin/Aug-07/Kasprowski.pdf).