The Exchange of Serials Subscription Information

By Ed Jones

A white paper prepared for the National Information Standards Organization, with support from the Digital Library Federation

Published by the National Information Standards Organization
Bethesda, Maryland

NISO Press, Bethesda, Maryland, U.S.A.

Published by
NISO Press

NISO Press
4733 Bethesda Avenue, Suite 300
Bethesda, MD 20814
Telephone: 301-654-2512
Email: nisohq@niso.org

About NISO White Papers

NISO White Papers are contributed or solicited papers that address an issue that has implications for standards development. White Papers can be viewed as a pre-standardization activity. A NISO White Paper might define and explore some of the questions that come into play before formal standardization work is started. Or, a NISO White Paper might identify areas that are opportunities for standards development and suggest possible approaches. All White Papers are posted on the NISO website (http:www.niso.org).

Suggestions for topics to be addressed in the White Paper series or unsolicited manuscripts should be sent to nisohq@niso.org.

NISO White Papers now available:

- Issues in Crosswalking Content Metadata Standards
  by Margaret St. Pierre and William P. LaPlant, Jr.
- The Myth of Free Standards: Giving Away the Farm
  by Andrew N. Bank
- The Case for New Economic Models to Support Standardization
  by Clifford Lynch
- Internet, Interoperability, and Standards - Filling the Gaps
  by Janifer Gatenby

Contents

Acknowledgements

In the course of this study, several dozen individuals were contacted in the communities involved, and a library-specific survey was posted on the NISO Web site to collect input from the library community at large. I would like to express my sincere gratitude to all those contacted, who generously contributed their time and expertise.

Diane Bednarski and Bill Schickling, Gaylord Information Services
Bob Boissy, divine, inc.
Christie Burke, EBSCO
Adam Chandler, Cornell University
Tim Devenport, consultant on ONIX for Serials
Todd Fegan, ProQuest Information and Learning
Dave Fisher and Stacy Nelson, UC San Diego
David Fritsch, TDNet
Brian Green, EDItEUR
Eric Hellman, Openly Informatics
Helen Henderson, OCLC Resolution Services
Todd Holbrook, Simon Fraser University
Tim Ingoldsby, American Institute of Physics
October Ivins, consultant
Mike Kaplan, Ex Libris (Aleph)
Bill Kara et al., Cornell University
Katharina Klemperer, Harrassowitz
Mike and Steve McCracken, Serials Solutions
Kirsty Meddings, ingenta/CatchWord
Jim Mouw, University of Chicago
Norman Paskin, International DOI Foundation
Glenn Patton, OCLC, Inc.
Tom Peters, Committee on Institutional Cooperation
Jan Peterson, Infotrieve
Sara Randall, Endeavor Information Systems
Nathan Robertson, The Johns Hopkins University
Tony Ross, Elsevier Science (USA)
Norman Roth, High Wire Press
Harry Samuels, LinkFinder Plus
Jenny Walker, Ex Libris USA, Inc.

Any errors or misrepresentations of fact in this study are the responsibility of the author.

Ed Jones
San Diego, California

Executive summary

This study was undertaken in order (1) to identify current and potential applications in which serials subscription data are exchanged, (2) to identify the formats currently in use for such exchange, and (3) to ascertain the perceived utility of standards to support such exchange, including standard identifiers for subscribers and services. A large number of interested individuals and organizations were contacted in the effort to accomplish these tasks.

Current and potential applications were identified and described in some detail. (See Appendix 2 for a summary table of parties, functions, and formats.)

It was found that, where formats appropriate to a given exchange (e.g., the MARC 21 format for holdings data or a common spreadsheet format) were already in existence, these tended to be used. Otherwise the formats currently in use tended to be proprietary in nature. Beyond this, the "exchanges" involved tended to consist of the manual retrieval from a remote site of files in these various formats.

There was widespread support among those contacted for the development of a standard format for exchanging serials subscription data. Among those familiar with ONIX for Serials, especially within the publisher community and organizations working closely with that community, there was strong support for incorporating as much as possible the work already completed in creating that industry standard.

It was found that a major stumbling block in the attainment of a standard for exchanging serials subscription data is the absence of standard identifiers for certain crucial elements of any exchange (especially parties to the exchange and complex products such as aggregations and subscription packages).

The study recommends, inter alia, that standards work in this area (1) occur in a context that assumes ONIX for Serials as the standard for most downstream transmission of bibliographic data (and consequently becomes involved in the evolution of that standard); (2) find a solution to the problem of identifiers in these exchanges--possibly the Global Trade Item Number [GTIN] and Global Location Number [GLN])—without which a standard will not function; and (3) encourage the participation of all interested parties, both for the creation of as complete a standard as possible and to ensure that competitive concerns are addressed over appropriate uses of data.

1 Introduction

This study was undertaken in order (1) to identify current and potential applications in which serials subscription data are exchanged, (2) to identify the formats currently in use for such exchange, and (3) to ascertain the perceived utility of standards to support such exchange, including standard identifiers for subscribers and services.

It is assumed that in an ideal world, such an exchange would be in the form of messages sent from one system to another, each message triggering an appropriate response.

Currently the ***exchange*** (in this sense) of serials subscription data occurs on a meaningful scale only between publishers and their customers—whether directly or via subscription agents—where such exchange is essential to conducting business. Here various proprietary and standard formats—particularly the various standards for electronic data interchange (EDI)—have been developed to facilitate the exchange. EDI use has been especially successful where subscription agents have played a role. The volume of transactions handled by these agents has been sufficient to justify the associated investment in systems modification, and the benefits arising from the consequent availability of standard data has encouraged other players, particularly publishers and vendors of integrated library systems (ILS), to modify their systems in turn to generate and receive data in these formats.[1]

The data exchanged in EDI requires a detailed format that can accommodate all the variables potentially applicable to the transactions involved, specifically invoicing, claiming, and dispatch notification. The subscription—either its initiation or its continuance—is the primary focus of the exchange.

It should be noted that the current study is not concerned with these exchanges. It addresses itself rather to exchanges that occur between the library and other parties once a subscription is in place, that arise from the ***fact*** of the subscription but are not directly concerned with it. It is expected that the format that will eventually arise to handle these subscription-based exchanges will entail fewer elements than are necessary in the EDI transactions that enable or maintain the subscription itself. It should also be noted that the subscription-related applications for which EDI messages are currently employed (other than claims and claim responses) are also addressed in the emerging ONIX for Serials metadata standards. (See under **The current standards landscape** later in this study.)

Subscription-based exchanges can be divided into those that enable services related to subscriptions and those that monitor changes in the makeup of subscriptions in order to trigger corresponding changes in related systems and services, both local and remote. These exchanges are exemplified by the following:

1 Exchanges between libraries or their intermediaries and publication access management services such as ***jake*** to communicate information on the journals, subscription packages, and aggregations to which the libraries have rights, and to enable the links from such services to the relevant journals, subscription packages, and aggregations;

2 Exchanges between libraries or their intermediaries and publishers, hosting services, document delivery services, and aggregators to communicate information on the journals to which the libraries have rights in electronic form as well as those for which they locally

[1] For a list of formats used in exchanges between divine Faxon Library Services and various ILS vendors, see http://www.faxon.com/edi/ils.htm

hold backfiles, either electronic or print, and to enable subscriptions, suppress pay-per-view options, and enable links to representations of local holdings;

3 Exchanges between libraries or their intermediaries and extended linking services to communicate information on the "appropriate copies" to which the libraries would like to enable links, along with the associated rights and services.

Further communications could flow from these initial communications, including the periodic return provision of updated data by the recipient, based on the parameters set by the library or intermediary in preceding messages. For example, the communication of library rights to a given subscription package could trigger updates from a hosting service to that library whenever additions or changes were made to the list of journals included in that package or the extent of the corresponding backfiles. Likewise, the communication of library rights to a given aggregation could trigger updates from an aggregator to that library whenever additions or changes were made to the list of journals included in the aggregation, which in turn could trigger local accessioning and/or cataloging operations (for a new or changed title) or de-accessioning operations (for a deleted title). While the expectation of such reciprocal communications is basic to this study, their precise format lies outside its scope.

The standardization of such exchanges entails three standardizable components: (1) the format of the exchange message; (2) identifiers for the objects of the message (serials, aggregations, etc.); and (3) the parties involved (library, link resolution service, etc.).

Definitions

For purposes of this study, discrete classes of service (which may be standalone services or components of larger services) are defined in terms of the particular functionalities that are of interest to the study. These classes, with their definitions, are as follows:

Aggregation service. An agency offering packages of content (aggregations)—usually subject-oriented with associated indexing—that may include the abstracts and full text of the indexed items. Access to the full text of publications in aggregations is typically included in the cost of subscribing to the aggregation and is independent of any subscription to the underlying publication. The publications included in an aggregation may change over the course of a given subscription to that aggregation as publications are added and dropped, and individual publications may be subject to embargoes on access to recent content, in order to encourage separate subscriptions to the publications involved. Examples: Gale, ProQuest, EBSCOhost.

Document supply service. An agency offering access to non-subscribed content on an item-by-item basis. Such access may be to the entire available run of a publication or to a customer-specified subset of issues. Delivery may be by means of fax, Ariel, etc., but typically excludes online delivery. Example: Infotrieve.

Link resolution service. An agency that resolves URLs on OpenURL-enabled Web sites into customer-specific URLs providing for a choice of customer-defined services relating to the work represented by the original URL. Examples: SFX, LinkFinder Plus, 1cate.

Publication access management service. An agency offering customers basic and updated data on the publications to which they have access rights, whether these publications are hosted locally or remotely. The data may be used to update a variety of local systems, including local management tools, local e-publication Web pages, and records in the local catalog. Examples: jake, Serials Solutions, TDNet.

Publication hosting service. An agency that hosts the full text of publications on its Web site. The agency may be a publisher or an agency acting on behalf of one or more publishers. The service may offer access by subscription or by means of various limited-access options (limited by number of accesses or by a time window). Subscriptions are typically to individual publications, but may also be to packages of publications. Such packages may be generally

available or they may ad hoc (e.g., available to members of a given consortium or of an ad hoc group). The publications included in a package may change over the course of a given subscription to that package. Examples: CatchWord, High Wire Press, ingenta.

Subscription agent. An agency acting on behalf of a customer in its day-to-day dealings with publishers, representing the customer to the publisher and the publisher to the customer. Examples: divine Faxon Library Services, EBSCO, SwetsBlackwell.

2 The current state of exchanges

The various EDI standards have greatly facilitated the exchange of subscription data among libraries, subscription agents, and publishers. However, few exchanges (in this sense) of the subscription-based data that is the focus of this study are currently taking place. What is much more common is for agency A to go to the Website of agency B and either download or otherwise "harvest" the relevant data (what might kindly be called "data receipt") or manually update a form on that Website with relevant data of their own (what might be called "data transmittal").

At present, the variety of formats employed by exchange partners is almost as great as the number of partners. The few formal exchanges in place tend to be party-specific, with legal constraints on outside use of the data provided.

Exchanges typically involve a variety of elements, their exact nature depending on the exchange partners. The number of elements will also vary, in both requests and responses, depending on the nature of the exchange.

The following section looks at the current state of these "exchanges" between various classes of service (as defined in the **Introduction**) both in terms of the data involved and of the parties to the exchange. The numbers in square brackets refer to the corresponding sample exchange scenario described in a later section (**Sample message-based exchange scenarios**). Each exchange is here characterized in tabular form as to parties (from/to), function, and the formats involved, and is followed by a narrative description of the exchange. A consolidated table of all these exchanges appears as Appendix 2 to this study.

From aggregation services to publication access management services [1]

Parties	Function	Format
Aggregation services Publication access management services	Download/transmit title lists	MS Excel, CSV, other-character-delimited, PDF, HTML, proprietary (jake, Serials Solutions)
Aggregation service Publication access management services	Download/transmit updates to title lists	MS Excel, CSV, other-character-delimited

Aggregation services typically make available lists of the publications included in their aggregations. From time to time they also make available lists of changes to these lists. They may or may not be the originators of these lists. For example, aggregations produced by the H.W. Wilson Company are distributed by numerous aggregators; the title lists in these cases originate with H.W. Wilson.

These lists are typically made available as downloadable files on the aggregator Website. They may be made available in one or more formats, sometimes depending on the use to which the list is likely to be put. For example, a list that in which the data is expected to be manipulated by the user will be offered in a spreadsheet (MS Excel, CSV, etc.) or special-character-delimited ASCII format (tab, pipe, etc.), whereas lists that are not expected to be manipulated (e.g., publications in a given discipline) may be offered in PDF or HTML. Exceptionally, lists may also be offered in a format prescribed by one or more of the publication access management services.

While these lists may be transmitted to customers or other agencies (e.g., publication access management services) or placed on an FTP site for retrieval, this is not the norm. Typically, a customer or agency goes to the aggregator Website and downloads the list there.

In general, these lists include, for each publication, its title, ISSN (when available), and start and end dates for indexing, abstracting, and availability in various delivery formats (e.g., ASCII text, page image, PDF). Some services also make available proprietary publication identifiers used internally. Start and end dates occur at varying levels of specificity, from year alone to year, month, and day, and may include free text in some cases (e.g., "spring 1997"). Start and end dates may be stored as separate elements or they may be combined in a single element, separated from one another by a hyphen. Start and end numeric designations (e.g., "volume1, issue 1") tend not to be used. Currency may be indicated in a variety of ways: by an empty end date element, by a plus sign, or by a free-text word or phrase (e.g., "current"). Information on embargo periods is likewise indicated in a variety of ways (e.g., number of months, number of days, or a free-text phrase).

Updates are generally offered on a monthly basis or less frequently, and will carry an indication of the type of change, in free text (e.g., "Changed", "Lost rights"). These lists may or may not be product-specific; when they are not, the product may be followed by a list of publications rather than being tied to each publication in the list).

From aggregation services to libraries [4]

Parties	Function	Format
Aggregation services Libraries	Download/transmit title lists	MS Excel, CSV, other-character-delimited, PDF, HTML, proprietary (jake, Serials Solutions), MARC 21 bibliographic records

In addition to the aggregator title lists described in the preceding entry, aggregators may make available MARC 21 bibliographic records representing publications in their aggregations.

From libraries to aggregation services [3]

Parties	Function	Format
Libraries Aggregation services	Transmit holdings lists	character-delimited files

Aggregators also typically enable linking from records within an aggregation to holdings represented in a library's OPAC or other site. Libraries can batch upload delimited files representing these holdings for display to users.

From libraries to document supply services [3]

Parties	Function	Format
Libraries Document supply services	Transmit holdings data	Special-character-delimited files, MARC 21 holdings records

Document supply services make available citations to articles in publications for which the full text is available for delivery, typically by fax or Ariel.

Libraries may be able to transmit their holdings to the service as MARC 21 holdings records, and these will then be used to suppress delivery of items that are identified as being already in the library in print or accessible electronically through an existing subscription (or through ongoing access to an online back file).

In some cases, the mechanism is fairly simple: presence of a given publication in the customer's library is recorded at the title level only (in which case the transmittal consists of a tab-delimited list of ISSNs).

From libraries to link resolution services [5]

Parties	Function	Format
Libraries Link resolution services	Transmit holdings lists	None (select subset representing holdings from a supplied database)

Existing link resolution services provide customers with a database pre-populated with information on Web sites hosting publications and offering services relating to those publications (URLs and templates). The customer activates the links desired for their own subscribed publications (and coverage) and adds links for any publications not already present in the database (e.g., locally hosted publications). In addition to URLs and templates, the data typically stored includes title; abbreviated title; p-ISSN; e-ISSN; CODEN; start and end year, volume, and issue; a "payment pending" indicator (SFX), customer information (e.g., which sites of a consortium such as the California Digital Library), and supplier (e.g., SwetsNet Navigator).

OCLC is currently engaged in developing a centralized rights management and link resolution service that is intended to maintain information on IP ranges, institutions, content, access rights (ILL, course pack, etc.), time period, and referred source ("appropriate copy"). OCLC hopes to acquire the necessary rights data from subscription agents and IP range data from the library. Each record would be accessible by any involved party (the relevant publisher, subscription agent, and library). The target date for going live with this service is the second quarter of 2003.

From libraries to publication access management services [2]

Parties	Function	Format
Libraries Publication access management services	Transmit holdings lists	MS Excel, CSV, other special-character-delimited files

Libraries may communicate holdings information to a publication access management service to produce a single integrated list of all serials held by the library. Such information is typically conveyed as a spreadsheet with a link to the appropriate holdings record in the library OPAC. For Serials Solutions, the spreadsheet has five columns: Title, ISSN, DateStart, DateEnd, and URL (DateStart and DateEnd are optional). Specific holdings are often not communicated due to their volatility—a realtime link to the OPAC is seen as more effective.

From publication access management services to libraries [2]

Parties	Function	Format
Publication access management services Libraries	Transmit title lists	MS Excel, CSV, other special-character-delimited files
Publication access management services Libraries	e-mail change lists	character-delimited files

Serials Solutions data can be distributed as a spreadsheet with the following elements: title, p-ISSN, e-ISSN, full text start, full text end, images start, images end, ASCII start, ASCII end, abstract start, abstract end, citation only start, citation only end, URL, provider, and database. Libraries identify the titles links they want to activate in the list.

Change data can be distributed as a pipe-delimited file (TDNet) with the following elements: title, p-ISSN, e-ISSN, aggregator name (if any), publisher, subject headings, URL, coverage dates, library ILS control number (if needed).

From publication hosting services to libraries [4]

Parties	Function	Format
Publication hosting services Libraries	Transmit title lists	MS Excel, CSV, other special-character-delimited files
Publication hosting services Libraries	e-mail change lists	unformatted

Lists are made available on Web sites in varying degrees of accessibility. For example, one service provides a continually updated list in CSV format with the following elements: title, ISSN, publisher, start date, "True PDF" start issue and date, end issue and date, notes, and journal URL. Other publishers do not offer title information in an easily usable (consolidated) format

Publication hosting services typically require registration at the Web site. Publisher sites may require some specific form of identification (e.g., a subscription identifier that it may be necessary to get from the library's subscription agent) to verify access rights. Other hosting services may verify rights against publisher-supplied data or bounce the registration information to the publisher for verification. One service has a variety of access mechanisms, including username/password, IP range, pay-per-view, and a special-access URL (e.g., for authors, in lieu of preprints).

One service gets around the problem of library identifiers by using the email address used in correspondence as the identifier.

Libraries may subscribe to publication hosting services (and aggregations) directly or as members of a consortium. Likewise, they may subscribe to individual publications or to generally available or ad hoc packages of publications.

Publication hosting services typically notify of title changes and newly available titles by e-mail. These notifications may or may not indicate that a given title is included in a given package or that a given customer is entitled to the title as a member of a consortium.

Agencies may request from hosting services information to enable linking via a link resolution service. For example, one consortium requests both an algorithm for computing article-level URLs (if such a one exists) as well as the following for each title it subscribes to as a consortium: title, p-ISSN, e-ISSN (if assigned), title code (if needed to enable the URL), start date, start volume/issue, format (PDF, HTML, TeX, etc.)

From publication hosting services to publication access management services [1]

Parties	Function	Format
Publication hosting services Publication access management services	Download/transmit title lists	jake, Serials Solution XML DTD, partner-specific format.

Both jake and Serials Solutions have preferred formats in which to receive data (see Appendix 1), though these are more honored in the breach.

Efforts are currently under way by Serials Solutions to collect customer-specific data from publication hosting services managed by subscription agents (Harrassowitz, SwetsBlackwell), but the details of these lists is not available due to confidentiality concerns. In each case, the format is specific to the partners and represents a compromise that, while usable, is not yet stable and falls short of the ideal (Serials Solutions). ***The exchange agreements also involve clauses that prohibit the reuse of the data.***

From publication hosting services to publishers and back [6]

Example: CatchWord

Parties	Function	Format
Publication hosting services Publishers	Verify/enable subscription access	Web interface

Before enabling a subscription, CatchWord sends to the publisher (via a Web interface): CatchWord ID, institution name, institution address, administrator e-mail address, institution IP addresses, and subscription number (library-supplied). The publisher responds with a validate/reject judgment, date ranges, and any additional subscription, which feed into the access control database.

From publishers to publication hosting services [6]

Example: ingenta

Parties	Function	Format
Publishers Publication hosting services	Verify subscription access	CSV

Publishers use an ingenta-supplied template to create CSV files with the following elements: subscriber identifier, subscription type (personal/institutional), subscriber name, address, postal code, country, abbreviated journal name, and subscription start and end dates.

From subscription agents to libraries [4]

The data maintained by subscription agents is often unsatisfactory for accomplishing the tasks that libraries and others would like to achieve. Local decisions may have led, for example, to the use of p-ISSNs only on records for the print versions of publications, making links to electronic versions and to combined print-with-electronic records for the same publication nonexistent. Ultimately, the database needs to be reconstituted to enable transactions of the sort envisaged here (e.g., explicitly coding titles where electronic access is free with a print subscription, as well as titles where electronic access is free regardless of print subscription [e.g., BMJ and, after a specified embargo, various High Wire Press titles]). Additionally, some way needs to be found of identifying the multiple distinct entries that correspond to a single ISSN. One subscription agent's efforts to facilitate a customer's links via SFX ultimately failed due to the difficulty of easily determining the customer's rights to electronic journals based on the data stored in the agent's database.

3 The current standards landscape

As is apparent from the preceding section, most subscription-based exchanges do not take place following standardized formats. Having said that, a number of pertinent international, national, and industry standards are in varying stages of development and implementation in the various communities concerned with such exchanges. The organizations promulgating the relevant standards are given below, along with brief descriptions of the standards concerned and their bearing on the present study.

Standards for messages

EDItEUR (International Group for Electronic Commerce in the Book and Serial Sectors)
http://www.editeur.org

EDItEUR's brief is to promote the use of EDI in the book and serial industries. In 2000, they published ***EDI Implementation Guidelines for Serials. Issue 1, Version 1.3: Despatch, Claims, and Invoices*** governing the content and structure of certain EDI messages. This EDI-related work is mentioned here for reference purposes only. As noted in the **Introduction**, the current study is not concerned with EDI exchanges.

More recently, EDItEUR has moved into the area of XML DTDs/schemas for conveying publisher catalog product data. A DTD for book product information (ONIX 1.0), based on the earlier EPICS (EDItEUR Product Information Communication Standards) was published in 2000, and is intended for use in conveying product information from publishers to other participants in the supply chain.

EDItEUR has proposed three record types for communicating serial information from publishers to other parties (who may then pass it on to further parties). These are, in order of increasing comprehensiveness, the Serial Item record, the Serial Title record, and the Subscription Package record. ***It should be noted that these record types are currently designed for general distribution and do not include an element for matching against a query from a specific subscriber.***

In August of 2000, EDItEUR undertook to extend the ONIX model to serials, and in late 2001 the first drafts of ONIX for Serials were made available. These have since been updated and are, as of April 2002:

Serial Item record.[2] The Serial Item record is intended to support alerting, dispatch, and library check-in applications (in which cases it is treated as a "release notice") and to give a full bibliographic description of such a part for any local database applications (in which cases it is treated as "content description"). The data in a Serial Item Record can be used by subscribers and by intermediaries for a multitude of purposes. For example, as "content description" it can be used to create a local representation of an issue table of contents on a local Web site, on an intranet, or to be dispatched to local interested parties in an e-mail message, either as part of a

[2] ***ONIX for Serials: Serial Item Record.*** Draft 2, compiled by Tim Devenport and David Martin. London: EDItEUR, 2 April 2002.

current awareness service (in which case it is also operating as a "release notice") or on an ***ad hoc*** basis; or it can be used to populate a local database of journal articles, either universal or selective, so that local users can better exploit the resource. As a "release notice", it can be used to update local holdings records—communication of the extent of newly available backfiles is one of the functions of the record—and to enable or suppress related issue-specific services (such as interlibrary lending).

Serial Title record.[3] The Serial Title record is intended to support at least three applications: Bibliographic description at the level of the serial; Product catalog item description (sometimes in conjunction with the Subscription Package record); and Tailored packages with price quotations (in conjunction with the Subscription Package record).

Subscription Package record.[4] The Subscription Package record is intended to support two applications, in conjunction with both the Serial Title record and, as applicable, ONIX book Product records:[5] the Product catalog item description and Tailored package with price quotations. The Subscription Package record describes the subscription package and identifies the constituent titles (e.g., ISSNs and titles). The data in a Subscription Package record could be used by an organization to identify those packages to which customers have rights and to monitor the constituent titles (serial or book) included in the package for purposes of notifying any changes to the subscribing customers. If received by the customer, the data—along with that in any associated Serial Title or Book Product records—could be used to update any affected local database or catalog.

While the Serial Title record is used to signal changes in status (title changes, etc.), the Serial Item record is used to signal changes in access (e.g., availability of backfiles).

Because both the Serial Item and Serial Title records include URLs or similar locators, as well as ISSNs and similar identifiers, they can also be used to populate databases that store such locators and associate them with such identifiers.

If ONIX for Serials is implemented by any part of the publisher community, those (such as hosting services) who receive content from those publishers will need to modify their applications to receive data in this format. Additionally, those who currently "harvest" data from publisher sites will have an incentive to set up formal relationships with these publishers to receive that data as ONIX for Serials records and to develop applications that can automatically manipulate, store, and pass these records (or their content, in whole or part) on to others.

- ***Subscription Package Record*** (2002: draft for consultation)
  [intended for use in conveying Rich Catalogue Information relating to standard and ad hoc packages comprised of individual serial titles as well as other items]
- ***Serial Title Record*** (2002: draft for consultation)
  [intended for use in conveying Rich Catalogue Information relating to individual serial titles]
- ***Serial Item Record*** (2002: draft for consultation)
  [intended for use in Alerting, Despatch, Library Check-In, and Structured Multi-level Bibliographic Description]

---

[3] ***ONIX for Serials: Serial Title Record.*** Draft 2, compiled by Tim Devenport and David Martin. London: EDItEUR, 2 April 2002.
[4] ***ONIX for Serials: Subscription Package Record.*** Draft 1, compiled by Tim Devenport and David Martin. London: EDItEUR, 2 April 2002.
[5] ***ONIX Product Information: <Product> Record***. Guidelines Release 2.0, compiled by David Martin. London: EDItEUR, jointly with Book Industry Communication, London, and Book Industry Study Group, New York, 2001. http://www.editeur.com/onixfiles2.0/ONIXProductRecord2.0.pdf

The ONIX book standard has been widely implemented—including more than sixty major British and American publishers—and major recipients of bibliographic data (e.g., OCLC and the Library of Congress) are able to receive data in the ONIX format. The standard for serials is currently awaiting its first tests.

International Organization for Standardization www.iso.ch

The ISO, through its Technical Committee 46, promulgates international standards in the area of information and documentation. ISO has defined a standard for serials holdings data at a summary level (ISO 10324:1997) but it is not as comprehensive as the corresponding American standard (see **National Information Standards Organization** below).

Library of Congress http://lcweb.loc.gov/marc/

The Library of Congress is the maintenance agency for the various MARC 21 formats and their associated identifier lists. The format of primary interest to this study is the MARC 21 format for holdings data, used by libraries and related agencies to communicate data on their holdings, as well as related bibliographic and control data. The format is designed to accommodate holdings statements conforming to ANSI/NISO Z39.71 (see **National Organization for Standardization** below).

National Information Standards Organization http://www.niso.org

NISO promulgates American national standards in the area of information management. The standard of primary interest to this study, insofar as messages are concerned, is ***ANSI/NISO Z39.71 – 1999 – Holdings Statements for Bibliographic Items***, which is related to the MARC 21 format for holdings data (see **Library of Congress** above).

Standards for identifiers

EAN International http://www.ean-int.org

Originally a bar-coding agency assigning European Article Numbers (EANs), EAN International now oversees these and related activities worldwide. In the context of the current study, EAN International assigns GLNs (Global Location Numbers) and GTINs (Global Trade Item Numbers).

- GLNs are used to identify legal entities (e.g., a registered company), functional entities (e.g., a specific department within a legal entity), and physical entities (e.g., a door of a warehouse).
- GTINs are used to identify any item in trade. They encompass EANs and UPCs barcoded on most commercial items.

GLNs and GTINs are typically expressed as bar codes as well as numeric strings, and they are most thoroughly exploited in environments where these bar codes can be scanned. One problem that results from this close relationship is that GTINs associated with serials are associated (through the price code) with single purchasable issues rather than with the serial as an entity.

ISSN International http://www.issn.org

ISSN International and the associated national ISSN centers maintain the authoritative data relating to assigned ISSN and their rules for assignment. The ISSN arose in the context of the indexing and abstracting community (specifically, the Abstracting Board of the International Council of Scientific Unions [ICSU/AB]), was subsequently brought under the auspices of Unesco, and now functions as an international system for identifying serial publications. ISSNs distinguish between print and electronic versions of the same journal, between the different

successive titles of a single journal, and between different journals with the same title. ISSN are typically assigned at the national level by the relevant national bibliographic agency.

International Organization for Standardization www.iso.ch

The ISO, through its Technical Committee 46, promulgates international standards in the area of information and documentation. A standard currently under development—the ***International Standard Identifier for Libraries and Related Organisations (ISIL)—***would play a role on the international level similar to that of the MARC Code List for Organizations within the US (see **Library of Congress** below).

The ISBN and ISSN are likewise ISO standards [ISO 2108 and 3297 respectively]. They are discussed in more detail in a later section (**Identifiers and other stumbling blocks**).

Library of Congress http://lcweb.loc.gov/marc/

The Library of Congress is the maintenance agency for the various MARC 21 formats and their associated identifier lists. The identifier list of interest to this study is the MARC Code List for Organizations, a list of holding symbols of libraries and related organizations, primarily American.

National Information Standards Organization http://www.niso.org

NISO promulgates American national standards in the area of information management. Two standards are of interest to this study, insofar as identifiers are concerned:

- ***ANSI/NISO Z39.43 – 1993 (R2001) – Standard Address Number (SAN) for the Publishing Industry***. This standard is analogous to the GLN (see **EAN International** above), though shorter in length.
- ***ANSI/NISO Z39.56 – 1996 (R2002) – Serial Item and Contribution Identifier (SICI).*** This standard is intended to unambiguously identify the issues and individual contributions to a serial publication. Its structure is built upon the ISSN, which serves as its highest-level constituent.

4 Sample message-based exchange scenarios

The initial call for this study described a number of scenarios where it was felt a standard format for exchanging subscription data (or more broadly, rights data) would expedite transactions and maintain the currency and accuracy of rights data among the partners. Following are descriptions of these scenarios, amended to accommodate others that could be served by a standard format for exchanging serials subscription/rights data.

It should be borne in mind that most subscription-based exchanges do not currently take place following standardized formats and cannot properly be called "exchanges" in that typically one party visits the Web site of another party and downloads a file after manually entering various selection criteria. The elaborations of the scenarios assume a true "exchange" environment and propose various means for reaching this environment using existing standards, standards under development, and standards yet to be developed.

An underlying assumption of all these scenarios is the protection, as much as possible, of any information that might be commercially useful to one or more of the parties involved. To this extent, the scenarios do not envisage the exchange among third parties of information identifying the subscriptions of a given library without the consent of all parties involved. For example, the integrity of subscriber lists is important to many parties and might be undermined by a too open exchange of data, driving those parties from participation in the exchange.

The scenarios incorporate, wherever feasible, the work that has already taken place in of EDItEUR's ONIX for Serials program, in the belief that ONIX for Serials may provide some components of a solution to these scenarios. The final version of ONIX for Serials will be published in the form of an XML DTD or schema. The version used in preparing this study was that of April 2002.

Scenario 1: Publishers, hosting services, and aggregators communicate to third-party services such as Serials Solutions information about the titles they include in their e-journal collections

This scenario represents two separate flows, the first from publishers and hosting services to third-party services, the second from aggregators to third-party services. It is assumed that prior to this scenario, hosting services and aggregators (the latter via the indexing and abstracting services represented in their aggregations) received ONIX data for the relevant titles from their publishers (as envisaged in ONIX for Serials).

Publisher/hosting-service-to-third-party-service flow

In the case of the publisher/hosting-service-to-third-party-service flow, ONIX for Serials might be a serviceable carrier for the data required by the third-party services. For example, a publication access management service would require from publishers/hosting services, at a minimum, the following elements for each title:

- Publisher/hosting service
- Title
- ISSN
- Start date
- End date
- Base URL

Because the publisher/hosting service would be transmitting a comprehensive title list to the publication access management service, the ONIX for Serials Subscription Package record might suggest itself as the most likely vehicle (with component Serial Title records), though independent Serial Title records might represent an alternate mode of transmittal.

Assuming the first mode, the following is a sample ONIX message using the pertinent elements, assuming no changes to ONIX for Serials:

```
<SubscriptionPackageRecord>
      <RecordReference>9876543</RecordReference>
      <NotificationType>03</NotificationType>
      <SubscriptionPackageIdentifier>
      <SubscriptionPackageIDType>01</SubscriptionPackageIDType>
      <IDTypeName>CatchWord product no</IDTypeName>
      <IDValue>0001</IDValue>
      <SubscriptionPackageDescription>CatchWord product catalog access
elements<SubscriptionPackageDescription>
      <TitlePackage>
            <SerialTitleIdentifier>
                  <SerialTitleIDType>07</SerialTitleIDType>
                  <IDValue>08954852</IDValue>
            </SerialTitleIdentifier>
            <Title>
                  <TitleType>01</TitleType>
                  <TitleText>Academic Questions</TitleText>
            </Title>
            <ProductForm>JD</ProductForm>
            <Website>
                  <WebsiteLink>http://cherubino.catchword.com/vl=
9159233/cl=49/nw=1/rpsv/cw/tranpub/08954852/contp1.htm</WebsiteLink>
            </Website>
            <JournalIssue>
                  <JournalIssueRole>04</JournalIssueRole>
                  <JournalVolumeNumber>12</JournalVolumeNumber>
                  <JournalIssueNumber>1</JournalIssueNumber>
                  <CoverDate datetype="03">19991</CoverDate>
            </JournalIssue>
            <JournalIssue>
                  <JournalIssueRole>05</JournalIssueRole>
                  <JournalVolumeNumber>15</JournalVolumeNumber>
                  <JournalIssueNumber>1</JournalIssueNumber>
                  <CoverDate datetype="03">20021</CoverDate>
            </JournalIssue>
      </TitlePackage>
      <TitlePackage>
            <SerialTitleIdentifier>
                  <SerialTitleIDType>07</SerialTitleIDType>
                  <IDValue>08989621</IDValue>
            </SerialTitleIdentifier>
            <Title>
                  <TitleType>01</TitleType>
                  <TitleText>Accountability in Research: Policies and
Quality Assurance</TitleText>
            </Title>
[… etc.]
</SubscriptionPackageRecord>
```

No major changes would be needed to ONIX for Serials as it exists in the current (April 2002) draft to accommodate these elements, though the various elements of the <JournalIssue> composite would need to be expanded as noted in a later section (see **Identifiers and other stumbling blocks**). Additionally, the <SerialTitleIDType> would need to be revised to explicitly differentiate between electronic and print ISSNs, either one of which may be used in practice to identify an online resource. It may also be necessary to explicitly identify the level to which the <Website> element applies (The example above assumes it applies to the title level).

Aggregator-to-third-party-service flow

No format analogous to ONIX for Serials exists for communicating aggregator content. However, the existence and scope of ONIX for Serials argue strongly for its use in this context as well, in full or in some abbreviated form. ONIX documentation explicitly mentions its use in transmittal of metadata from publishers to abstracting and indexing services, though no examples are given.

Because aggregator content is subject to limitations and conditions different from those to which publisher content is subjected, it would be necessary either to amend ONIX for Serials to accommodate these limitations and conditions (e.g., embargo periods and format limitations) or to apply a different format not yet developed. ***It is assumed that ONIX for Serials could be modified to support these changes.***

Any solution to this scenario will involve substantial investment by indexing and abstracting services and/or aggregators to bring the content of aggregations under greater bibliographic control. Specifically, it will entail a greater use of standard identifiers for the bibliographic items included in indexing and abstracting services and aggregations, and may require the development of standard identifiers for the aggregations themselves (or their accommodation within an existing scheme).

A publication access management service would require from aggregators, at a minimum, the following elements for each title:

- Publisher/hosting service
- Title
- ISSN
- Start date
- End date
- Embargo period
- Full text delivery format
- Base URL

Assuming a robust flow of metadata from publishers to abstracting and indexing services, this scenario would require the same changes to ONIX for Serials as the preceding, with the additional need for elements to accommodate values for the embargo period and additional values for the electronic publication format. (Electronic publication format is defined in the ONIX <EpubFormat> element, but currently only four values are defined.)

Scenario 2: Third-party services (such as Serials Solutions) communicate to libraries information about the titles to which they subscribe

This scenario implies two separate flows, the first from libraries to third-party services to identify the titles, aggregations, and serial packages to which they have rights (either directly or through membership in a consortium), the second from third-party services to libraries to give them details about those titles, aggregations, and serial packages.

One caveat regarding this scenario is that it assumes a message-based model that does not yet exist. Third-party services currently distribute their databases to libraries, and the libraries manually identify in those databases the titles, aggregations, and serial packages to which they have rights.

Library-to-third-party-service flow

No standard currently exists for the library-to-third-party-service flow. It assumes in its turn a number of predecessor flows, each using ONIX for Serials as in Scenario 1:

(a) from publisher or aggregator to library/subscription agent for those titles, aggregations, and subscription packages directly subscribed;

(b) from subscription agent to library for those titles, aggregations, and subscription packages subscribed via the agent;

(c) from publisher or aggregator to library/consortium for those titles, aggregation and subscription packages subscribed via the consortium;

(d) from consortium to library for those titles, aggregation and subscription packages subscribed via the consortium.

Assuming these predecessor flows, the library would have a comprehensive list of content to which they have rights that could be used to generate queries to the third-party service. These queries would need to unambiguously identify

- Library
- Subscribed item (title, aggregation, serial package)
- Supplier
- Term of the subscription
- Type of response expected

Software at the library end would need to be able to receive ONIX for Serials messages and to generate from those messages queries that could be sent to third-party services.

Third-party-service-to-library flow

Presumably this flow would use ONIX for Serials as in Scenario 1, if it is assumed that aggregations would be treated as Subscription packages in the ONIX for Serials record hierarchy. Once queries were received from libraries indicating the titles, aggregations, and serial packages to which they had rights, third-party services could send periodic and ad hoc update responses using ONIX Serial Title and Subscription Package records that would alert libraries to changes in the subscribed content (e.g., titles dropped from an aggregation).

No changes to ONIX for Serials would be required beyond those specified in Scenario 1.

Scenario 3: Libraries communicate to publishers, hosting services, aggregators, and document supply services information about the titles they hold

This scenario assumes that “holdings” may represent a variety of formats, and that the information communicated will be dependent on the parties involved and the intended consequence of the communication. For example, a communication intended to provide information about print holdings of a given title to be available to users attempting to access the electronic version will be different from one intended to suppress document delivery of content available locally in ***any*** format.

This exchange could in theory take place in two sequences:

(a) a service notifies a library of the content available through the service, and the library responds with an indication of its corresponding holdings;

(b) a library notifies a service of its holdings.

In both cases, the presumed outcome would be the appropriate action by the service involved (e.g., adding holdings information or an OPAC link for the titles involved, suppressing document delivery for the content involved)

The first (predecessor) flow in (a) would use ONIX for Serials as in Scenario 1. The second flow in (a) as well as the flow in (b) might involve different levels of detail, corresponding to the variety of ways in which holdings information is currently represented in such services:

- Holdings start date
- Holdings end date
- URL of holdings record in a library OPAC
- Detailed indication of holdings (e.g., XML version of a MARC 21 holdings record)
- Binary indication of holdings (yes/no)

These would also require elements to identify:

- Library
- Subscribed item (title, aggregation, serial package)
- Supplier

Note: an ambiguous case exists in the provision of deposit accounts where for a given fee a library is granted a fixed number of electronic accesses to a given range of titles. In these cases, presumably a library's policy on representing these titles internally would govern its external representations as well.

Scenario 4: Publishers, hosting services, and aggregators communicate to libraries about the titles they subscribe to

This is the predecessor flow mentioned in Scenario 2. It appears that ONIX for Serials could satisfy this scenario, in that the elements required are present in that format. It is assumed that, if such communications were received from a publisher by a subscription agent, consortium, or other intermediary, then that intermediary would pass the communication on to the library or libraries involved, modified or enhanced, when appropriate, with its own data.

Scenario 5: Libraries communicate to providers of extended linking services about their preferred sources for various resources

This scenario assumes that the communications in Scenario 4 have taken place. In this scenario, ONIX for Serials would again provide a basic structure, though it would need to be amended to accommodate information on formulas for constructing URLs and associated holding ranges (assuming different sources for different ranges).

For the system envisaged in these scenarios to work, standards must exist for the records flowing in each direction, and standard identifiers must exist that will facilitate the matching of records representing the same entity in the two flows.

A significant buy-in from the major parties involved will also be needed in order for the system to function, as the transactions envisaged are complex and invariably mutually dependent, moreso as one moves farther downstream, where most of the demand lies.

Additionally, as this involves subscription data that may be considered sensitive by one or more of the parties involved, a security key may be necessary to assure confidentiality or invoke related legal conditions to the exchange.

5 Identifiers and other stumbling blocks

No standard for the exchange of serials subscription data can succeed without the use of other standards that unambiguously and expeditiously identify both the resources involved and the parties involved in the exchange. While such standards currently exist to varying degrees, each has problems in application that will invariably place limits on the universality, and consequently the relative success, of any exchange mechanism.

This section looks in turn at existing candidate standards for identifying resources and at those for identifying parties to an exchange.

Standards for identifying resources

Two classes of resources need to be accommodated in a standard for exchanging serials subscription data: the online resources that are the primary target of the standard and the print resources that serve as either a secondary target (for example, in the case of activating links from electronic resources to print resources represented in a library's catalog) or as a trigger (for example, in the case of electronic resources where rights derive from rights in a corresponding print resource).

We will look at standards for print resources first, since this is where the most mature standards exist.

Standards for identifying print resources

Two international standards have long been in use for the identification of print resources, and have gained widespread acceptance: the International Standard Book Number (ISBN [ISO 2108]) and the International Standard Serial Number (ISSN [ISO 3297]).

ISBNs are assigned by registered publishers to their published output. They serve primarily as inventory control numbers in the book trade and secondarily as somewhat imperfect bibliographic control numbers in the library community. It is the former use that has assured the ISBN's success. However, the ISBN suffers both from its inconsistent application—ISBNs have been known to be re-used by publishers over time—and from the fact that, after nearly half a century of use, it is still far from universal, even in countries with a mature book trade. While many small publishers are not registered with their national ISBN agencies, some very large publishers fall through the net as well, the most notable being the US Government Printing Office. Beyond this, the ISBN system is approaching capacity, and the standard is currently being revised with a view to expanding that capacity (e.g., by prefixing an additional EAN flag and incorporating it and the existing "Bookland" EAN into the ISBN proper, or by converting the ISBN into either a "dumb" number or, less probably, a hexadecimal number).[6] Ephemeral materials collected in various aggregations are examples of materials that currently fall outside the ISBN system but would fall within the scope of any standard for exchanging serial subscription data.

ISSNs are assigned by national and regional ISSN centers—typically located within national bibliographic agencies (NBAs)—as well as by an international center, to serial publications

[6] "W2. Expansion of the ISBN System" ***ISBN Newsletter*** 16 (2002) http://www.isbn-international.org/html/publicat/letter16/expansion.htm

deposited with the NBAs. While ISSNs, like ISBNs, do not cover all of their target population, they do cover those most likely to find their way into library collections. However, because ISSNs are not assigned by publishers, they are frequently assigned "after the fact"—after a new serial is published or its title changes—and even then may not be used by the publisher. Likewise, because it is linked to a bibliographic feature of the serial—its title—its ***de jure*** extent (as prescribed in its bibliographic record at the NBA) may not correspond to its ***de facto*** extent (as defined by its appearance on the printed issues of a serial). From a publisher's point of view, a title change—which triggers an ISSN change—does not typically signal the creation of a new product. Likewise ISSNs used to distinguish a main serial from a serial supplement may not be meaningful in terms of the corresponding purchasable product.

As an example of the current limitations of the ISSN system, one publication access management service relies on a combination of title and ISSN to identify the same serial in different aggregations and hosting services, but even so, they find for example, some 16 unlinked representations of JAMA (Journal of the American Medical Association) in their database, requiring manual intervention to correlate them.

Additionally, as in the ISBN system, ephemeral materials collected in various aggregations currently fall outside the ISSN system but would need to fall within the scope of any standard for exchanging serial subscription data.

Standards for identifying electronic resources

Both ISBNs and ISSNs are being assigned to electronic resources, though fixed rules for such assignment do not exist. For example, some ISBN agencies assign numbers to the same publication appearing in different electronic formats (PDF, HTML, etc.) while others do not.[7] Likewise, it is possible for a publication to be a serial in print (where its discrete numbered parts qualify it as a serial) but not in electronic form (where it may exist as a database of dynamically changing records). The former would receive an ISSN but the latter would not.[8]

Beyond this, as with print resources, a substantial portion of electronic resources fall outside the ISBN/ISSN system altogether, primarily through a failure to seek assignment of the requisite identifier (though failure to discover and recognize the "unidentified" resource can also play a role). In many cases, these resources are ephemeral constituents of aggregations and carry an "identifier" only within the context of that aggregation (which itself will not carry a universal identifier). The same resource in different collections will carry different collection-specific identifiers, with a resulting failure to be identifiable outside the collection.

In terms of resources eligible for ISSN assignment, this situation can be somewhat remedied by greater vigilance on the part of libraries, alerting the various ISSN centers when a serial is not yet in the system. However, this introduces delay into the process, and a dependence on the efficiency and resources of the responsible ISSN center.

In the realm of resources eligible for ISBN assignment, the situation is even more problematical. Because ISBNs are assigned by publishers, the principal reason that a publication would lack an ISBN would be not because the publication is not yet in the system but because the publisher itself is not yet in the system. Beyond this, some publishers—including some major ones such as the US Government Printing Office—may be partially in the system and partially out of it. And ISBNs are less likely than ISSNs to be separately assigned to electronic versions.

Additionally, there is sometimes an implementation problem in secondary services of failing to test the validity of standard numbers, which may be incorrectly transcribed in the record creation process. Standard numbers such as ISBNs, ISSNs, SANs, and the various EAN.UCC numbers include a check digit as their final character. This check digit is used to verify the validity of the entire number. Unfortunately, processes creating records that include standard numbers often do

[7] "W1. Development and Progress in Numbering Electronic Products (ep), eBooks, etc." ***ISBN Newsletter*** 16 (2002) http://www.isbn-international.org/html/publicat/letter16/development.htm
[8] "ISSN for Electronic Serials" http://lcweb.loc.gov/issn/e-serials.html

not include validity tests on the included numbers, and some incorrectly lowercase the roman numeral "X" in the check digit, resulting in validation failure.

DOI (Digital Object Identifier)

For electronic content, the DOI is rapidly becoming the identifier of choice, at least among major publishers of scholarly content. In this context, the DOI serves as the basis of the CrossRef reference linking initiative, involving some 124 publishers and a database of more than five million articles. The DOI is a "dumb" number consisting of a publisher prefix and a publisher-assigned object identifier that may be any combination of characters. The DOI does not include a check-digit, but use of the DOI does not typically involve manual transcription, so transcription errors are unlikely.

In theory, DOIs are assignable to electronic content at any level of specificity—journal, issue, article—and serve as persistent identifiers in the online environment. In practice, however, DOIs are assigned to articles but not to the journals that include them. Hence, DOIs may have a role in identifying electronic books (which may be constituents of aggregations[9]) but not electronic serials. Beyond this, the decision as to what should receive a DOI is up to the individual publisher rather than any set of universally applicable rules.

Because the DOI infrastructure is supported through user assessments (preferably via "registration agencies"), publishers may, for their own reasons, opt not to participate. However, the structure is designed to encourage participation, with registration agencies (e.g., CrossRef) existing to fulfill the needs of various communities.[10]

Standards for identifying parties to an exchange

Standards for identifying libraries and related organizations

Various national and proprietary schemes have evolved for identifying repositories in union catalogs. Examples of these national and proprietary schemes are, respectively, the lists maintained by the Library of Congress (LC)[11] and OCLC.[12] The limitation of these schemes is that they are for the most part restricted to library "holding" locations rather than the range of functional units that may be involved in the exchange of serial subscription data, both in libraries and in other organizations, and they tend to be national or organizational in scope. The development of an international standard to encompass these schemes has at present faltered due to failure to include schemes such as OCLC's that are themselves international (if not universal) in scope.[13]

Standards for identifying organizations in the book trade

A commercial scheme that attempts "to identify organizations and businesses interacting with the publishing industry (including book and serial manufacturers, libraries, publishers, etc.)" is the Standard Address Number (SAN) for the Publishing Industry.[14] The SAN is used primarily in EDI transactions. While the SAN has a broader functional application than library-identification

[9] In this context, it should be noted that aggregation services do not typically seek DOI assignment for their content.
[10] International DOI Foundation, FAQ http://www.doi.org/faq.htm
[11] MARC Code List for Organizations http://lcweb.loc.gov/marc/organizations/
[12] OCLC Libraries http://www.oclc.org/contacts/libraries/
[13] For the current version of the draft International Standard Identifier for Libraries and Related Organisations, DIS 15511, see http://www.niso.org/international/SC4/wg8n20.pdf
[14] Standard Address Number (SAN) for the Publishing Industry http://www.niso.org/standards/resources/Z39-43.pdf

schemes such as those of LC and OCLC, it suffers from being geographically concentrated in the US.

Standards for doing both: identifying resources and parties to an exchange

EAN.UCC: A single identifier system for organizations, services, and publications

Analogous to the SAN on the international level is the EAN.UCC scheme of EAN International and the Uniform Code Council, encountered in everyday life in the barcodes found on most commercial products. This scheme, initially restricted to trade items, has now expanded to encompass locations (the SAN analog) and services.

It should be noted that the various EAN.UCC numbering schemes are exclusive of one another: A number in one scheme may recur in another scheme, in which it identifies a different entity. The structure of the number depends on whether the EAN13 or UCC12 standards are applied. In general, the UCC12 standard is applied in North America and the EAN13 elsewhere. Both are unique and unambiguous when right justified in a 13-digit field.

The two sub-schemes considered here are the Global Location Number (GLN) and the Global Trade Item Number (GTIN). The schemes suffer from the shortcomings of all such self-supporting schemes: The associated fees will tend to limit participation to those for whom the benefit demonstrably exceeds the cost.

Global Location Numbers (GLNs)

Like a Standard Address Number (SAN), a GLN identifies commercial entities at various levels of specificity (down to the level of a warehouse door, if necessary). Like all EANs, the GLN consists of 13 characters, the last of which is a check digit.

Global Trade Item Numbers (GTINs)

Books and serial publications have long been assigned their own three-digit EAN prefixes: 978 and 977 respectively. Current discussions regarding the revision of the ISBN may lead to additional prefixes being assigned. The “Bookland” EAN consists of the 978 prefix followed by the first nine digits of the ISBN and a check digit. Similarly, the “Serial-land” EAN consists of the 977 prefix followed by the first seven digits of the ISSN, a two-digit price code (typically two zeros), and a check digit. (The price code may be changed for a special issue carrying an exceptional price, but this would not be the EAN typically carried by a serial.)

It should be borne in mind that the focus of GTINs, as their name implies, is trade, and there is an underlying assumption of a one-to-one correspondence between GTIN and price (hence the two-digit price code in the serial GTIN). In the case of serials, this is a single-issue (newsstand) price. This may or may not affect their usability in the context of the exchange of serial subscription data.

For GTINs with prefixes of “977” (serials) and “978” (books), there is no charge for GTINs, which can be extrapolated from the corresponding ISSN or ISBN. However, there would be a charge if GTINs using other prefixes were assigned to entities (subscription packages, aggregations) currently outside the ISSN/ISBN system. This would argue for bringing subscription packages and aggregations within that system, though it is unclear how this could be done without a major revision of the criteria for assigning ISSNs and/or ISBNs.

Discussion

There are various ways to deal with the problem of identifying subscription items and parties. One is to enable multiple schemes for identifying these entities, though such a solution must then

deal with the limited overlap among the various schemes and the need for translation tables to establish equivalences between identifiers from different schemes. It must also deal with items that are subscribable but are living "off the grid" in terms of the standard identifier schemes. This category encompasses both ephemeral materials, the publishers of which have never applied for a standard number, and aggregations and subscription packages, which are "out of scope" of the existing schemes.

A possible means of tackling this problem of "off the grid" items might lie in the creation of pseudo-identifiers for items lacking true identifiers. These pseudo-identifiers might be assigned, by the first interested party, to the catalog record for the item in a recognized master-record database such as that of the CONSER Program. When a true identifier was assigned to an item by the appropriate authority in the master-record database, this would supplant any already-assigned pseudo-identifier. Two drawbacks of such a solution are (1) that this would entail a maintenance structure at least as costly as the schemes it would be augmenting and (2) that there is no universal master-record database.

Another possibility might be some sort of "fast tracking" within existing schemes, where a given agency would expedite assignment when a request came from a party needing the identifier for a transaction. But even this would be feasible only within a scheme like ISSN, where there is no fee associated with assignment, and even there different national centers would have different resources to devote to ISSN assignment, much less "fast tracking".

Neither of these solutions deals with the problem of identifying aggregations and subscription packages. These are neither books nor serials, and in some cases they may not exist in more than one copy (e.g., a subscription package negotiated for members of a particular consortium). For these non-book, non-serial entities, there currently is no identifier scheme. To bring these within the ISBN/ISSN systems would require a revision of the assignment criteria currently in place for these systems.

Because these items are "subscribable", the ISSN system would be the most likely home for them. The major drawback here involves unnamed subscription packages, because the ISSN system quite sensibly requires that a "serial" have a "title". While aggregations have titles, as do some subscription packages (e.g., those offered by Project MUSE), many subscription packages would be more problematical (e.g., those negotiated between a publisher and a particular group of libraries). It is unclear how these might be accommodated within the ISSN system.

A third solution, though one which again cannot get around the question of fees, is an all-encompassing identifier scheme. There is only one of these—the EAN.GCC system—and it is all encompassing both because it applies to all the entities needing identifiers in this paper and because it has the flexibility to be structured when structure is useful (when incorporating ISBNs and ISSNs) and unstructured when structure serves no purpose (when identifying units of an organization and when identifying multi-component collections such as aggregations and subscription packages).

The EAN.GCC system could accommodate all the resources accommodated by the ISBN/ISSN systems (including the expansion to aggregations and subscription packages mentioned above), as well as those that cannot be shoe-horned in, such as untitled subscription packages. In the case of such out-of-scope resources, arbitrary GTINs might be assigned, either from a publisher cache or from some reserved group similar to the serial and book EANs. One possibility might be the reserving of a certain range of GTINs from any future expansion of ISBNs within that scheme. For example, if the ISBN system were expanded to include the 979 prefix, then a sufficiently large group within the 979 range might be reserved for assignment to "untitled" subscription packages.

Other stumbling blocks

Beyond identifiers, some other problems lurk in the path of developing a standard for exchanging serial subscription data.

Concerns over costs of modifying software

For the purposes envisaged in this study, subscription data would be transmitted from an agency holding such data (e.g., a library, subscription agency, or publisher) to an agency where this data would be used to select other data (e.g., a journal data management or extended linking service). The receiving agency would need to modify its software both to receive and interpret the standardized data, and to format and transmit an appropriate response (e.g., a list of matching journals and attributes). In many cases, this may entail adjustments to internal data storage. Formats that were designed to generate human-readable and -interpretable lists and formats (e.g., PDFs) may not be sufficiently granular to generate machine-readable and -interpretable lists. As a minor example, one agency involved in this study produces lists that include an asterisk after a journal title when that title is not available in a given format.[15]

Concerns over confidentiality of data

In the two cases where subscription agents have agreed to exchange data with Serials Solutions, the exchange has required a contract in place that would prohibit re-use of the data by other parties. Likewise, publishers have expressed concern that such exchanges could enable competitors surreptitiously to gain access to their circulation lists. One states that "cooperation with third party services … would only be done under rigid written agreements and careful data security."[16]

This suggests that work on an acceptable standard form for such agreements would form a useful parallel activity to any work to develop a standard to exchanging serials subscription data.

Representation of dates: applicable standards

While an international standard exists for the representation of dates [ISO 8601], it is incomplete for the range of values found on bibliographic items. The standard defines three classes of dates: (1) a standard date in the form CCYYMMDD; (2) the date in the form of a day of the year (CCYYDDD); and (3) the date in the form of a day of a given week (CCYY"W"WWD, e.g., 2002W234 for fourth day in the twenty-third week of 2002). In all cases, a given level of specificity can be eliminated when it is not applicable. While the standard does not address seasons or quarters, representations for these are provided in the NISO standard for the SICI [ANSI/NISO Z39.56; no ISO equivalent] in the ranges 21-24 for seasons (CCYYSS) and 31-34 for quarters (CCYYQQ). The NISO standard also employs the forward slash to indicate a span (e.g., 2001/2002, 20020301/20020307).

ONIX provides for the specification of the date through a combination of a "datetype" attribute and a numeric string representing the date in terms of that attribute. ONIX accommodates all the date types accommodated by the ISO and NISO standards, except the "given day of a given week" of ISO 8601, which likewise is not accommodated in ANSI/NISO Z39.56.

One essential access characteristic of online resources that is not accommodated in either set of standards, nor in the NISO standard for holdings statements [ANSI/NISO Z39.71], is the "moving window" of access, where the window is defined by reference to a rule (e.g., "latest five years", "all issues except current calendar year"). The "moving window" is, however, accommodated to a very great extent in the "specific retention policy" element of the MARC 21 holdings format.[17] However, even MARC 21 does not provide coding that can represent embargo periods for resources that become available only after the passage of a given period of time or a given date.

Needless to say, the data encountered in the existing lists of organizations involved in this study ran the gamut from highly structured representations of dates to natural-language representations, and the level of specificity varied from the year to the cover date (or a formalized representation).

---

[15] H.W. Wilson Company.
[16] Tim Ingoldsby, American Institute of Physics, e-mail message to author, 13 May 2002.
[17] "MARC 21 Concise Holdings: Control Fields" http://www.loc.gov/marc/holdings/echdcntr.html

Representation of numeric and/or alphabetic designations

Numeric and/or alphabetic designations (e.g., volume and issue numbers) do not occur in the metadata maintained by many of the agencies (e.g., aggregators and publication access management services) likely to be engaged in the exchange of serials subscription data. Where publication access management services maintain preferred formats for data exchange (e.g., jake, Serials Solutions), elements are not even provided for accommodating this data. However, these elements are customarily used by publishers to designate issues, and by libraries to maintain inventory control over the stock of serial issues, suggesting that their exclusion from a standard for exchange would be ill-advised.

These designations are carried in ONIX for Serials records, but at two levels only—volume and issue—creating potential difficulties where greater hierarchy is involved.

While it might seem expedient to eliminate these designations from exchanges, there are cases where these designations may be the only useful means of delimiting the content to which a subscription applies (for example, when a subscription applies to a volume that does not correspond to a calendar year, and individual issues are not differentiated by date).

6 Conclusion and recommendations

There was widespread support among those contacted in this study for a standard format for exchanging serials subscription data. Among those familiar with ONIX for Serials, especially within the publisher community and organizations working closely with that community, there was concern that work on such a standard avoid as much as possible "re-inventing the wheel" by taking into account ONIX for Serials, ensuring that any new standard is compatible with it, and endeavoring to involve those who have worked substantially on ONIX for Serials in any new standards-development effort. One contact expressed skepticism that a useful standard would result from this effort.

The study suggests an exchange of serials subscription data using a query/response model (see Appendix 4), with queries comprising a core group of elements (identifying the subscribed item, the parties to the exchange, the nature of the query, and an optional security key) that would be used to initiate the exchange—in either direction—and responses comprising this core group and a supplementary group of elements that would be specific to the nature of the query. For example, a library (SubscriberIDType/IDValue) might send a query to an identified aggregation service (SupplierIDType/IDValue) regarding the constituents (QueryType) of a subscribed aggregation (SubscriptionItemIDType/IDValue). The aggregation might respond by (1) returning the query with an appended file detailing the constituents of the aggregation; (2) returning a series of discretely identified copies of the query, each with the specified details of a single constituent item; (3) updating its system to dispatch a response to the query whenever a change occurs in the details of the constituents of the aggregation; (4) returning a response that it was unable to satisfy the query, for a specified reason (e.g., library is not among its listed subscribers). This is one model among many.

The ongoing development of ONIX for Serials at this time offers a serendipitous opportunity for the co-evolution of that set of standards with any standard to emerge from the current study. The ONIX for Serials model offers a useful framework for standards work on encoding serials subscription data. Likewise, the needs of the users of serials subscription data, as it has emerged in this study and will emerge in any subsequent standards development work, will provide useful input to the refinement of the ONIX for Serials model. Ideally, in the future data encoded and carried in ONIX for Serials records would be re-usable, with little or no modification, in any records exchanged in a standard format for serials subscription data.

Any system for the efficient exchange of serials subscription data will be critically dependent on the widespread adoption of standard identifiers for both the subscribed items and the entities engaged in the exchange. The current systems for identifying subscribed resources—ISBNs and ISSNs—will be useful means of identifying subscribed items if:

(1) It becomes truly universal, with more vigilant efforts to assign these numbers to collectible electronic items;[18]

(2) All entities that exchange these items make use of the relevant identifiers (and incorporate validation mechanisms);

[18] This will be easier for ISSN, which are assigned free of charge, than for ISBNs, which are assigned based on the purchase of a publisher identifier (a disincentive to participation).

A different problem exists with subscribable items that are not electronic equivalents of traditional books and serials. Aggregations and publisher subscription packages (as well as ***ad hoc*** packages negotiated between specific parties) will need identifiers of their own, possibly outside the ISBN and ISSN systems. It can be argued that these electronic collections are analogous to boxed sets and so are candidates for collective ISBNs. Or it may be argued that, while they are tradable items, they are not "books" and so fall outside the ISBN system (though ISBNs are now applied to a broad range of "non-book" items). At any rate, every effort should be made to capture them within the GTIN system in some form, whether identified by ISBNs, ISSNs, or by non-ISBN/ISSN GTINs (either within or outside the ISBN/ISSN GTIN prefixes), as this would provide the most consistent identifier system.

The parties currently engaged in the exchange of serials subscription data do not use identifiers outside of EDI messages. International EDI standards prefer GLNs while accepting SANs. One drawback of both GLNs and SANs is that there are fees associated with their assignment that may discourage participation to some extent. SANs have been assigned to the major players in North America and the UK, and can reportedly be extended into GLNs.[19] Both SANs and GLNs have advantages as identifiers in that they can be assigned to any unit of an organization, functional or otherwise, and like ISBNs and ISSNs they include a check digit for validity checks. For this reason, they are to be preferred to identifiers such as the MARC and OCLC identifier lists (and the ISIL framework) that primarily identify content repositories and are not currently designed for a more flexible role.

Finally, concerns about privacy and security are strong, especially among agencies for which subscription data holds commercial and competitive value. These concerns are exacerbated by the fact that many players now wear many hats: the same organization my encompass a subscription agency, an aggregator, a traditional publisher, and a publication hosting service, increasing the potential for conflicts of interest in exchanging subscription data. Any standard for exchanging serials subscription data will need to address these concerns satisfactorily or it will fail due to the non-participation of critical parties. One way of addressing these concerns might be through the development of a standard contract regarding the use of subscription data, adherence to which would be required of all parties to subscription data exchange. A central registry might make available a list of contracting parties and support contract enforcement.

Recommendations

1. That NISO become an active participant in the development and refinement of the various ONIX record types—ONIX for Serials, ONIX Product—to ensure that information necessary for the successful exchange of serials subscription data further downstream will be accommodated in the ONIX record structure. While the ONIX record types address several of the applications examined in the current study, the individual types still need to be fleshed out in many areas, some crucial (e.g., record reference number, type, and source). In addition, the capacity for representing chronological and enumerative data in ONIX for Serials is not as comprehensive as that in the MARC 21 holdings format.
2. That NISO actively monitor the "piloting" of ONIX for Serials. The organizations involved in this piloting stretch across the matrix of serials activity (other than libraries) that are the focus of this study, and should provide an effective test of its capabilities in its present form. It will be useful to see the extent to which they make full use of the capabilities of ONIX for Serials, in that many elements are listed as "optional" that might be considered critical if the data were to be passed on to an exchange format for serials subscription data. NISO should encourage the participation of libraries (via subscription agents and library systems vendors) in the piloting stage of ONIX for Serials.

[19] See the description at http://www.booksellers.org.uk/newlibrary/document_view.cfm?document_id=285

3. That NISO standards development work in this area focus on a record type for "subscription-based" messages that assumes that ONIX for Serials will provide a successful mechanism for distributing initial "publication catalog" data downstream that can subsequently be re-used in subscription-based messages. This assumes NISO participation in ongoing ONIX development. One possible approach is presented in Appendix 4.
4. That the concerns—especially the competitive concerns—of upstream users be addressed in the development of the standard through the participation of appropriate parties representing those concerns.
5. That these record types be capable of accommodating the data elements routinely included in exchanges among parties at the current time (see **The current state of exchanges, page 8)**, including MARC 21 holdings and bibliographic records as supplementary data.
6. That these record types accommodate rights data likely to affect library use of remote content (rights to include in course packs, interlibrary lending, etc.).
7. That the various user communities represented by the parties to a given exchange-type participate in the specification of the content of the messages that would be used in exchanges of that type. This might or might not be part of an eventual standard proper.
8. That NISO investigate GTINs as identifiers for aggregations and subscription packages, either within the ISBN system (EAN prefix 978) or outside that system.
9. That NISO investigate GLNs as identifiers for parties to exchanges, including, if necessary, a mechanism to moderate the effects of any fee schedule on parties less able to afford such fees.

Appendix 1

Preferred formats: jake, Serials Solutions

jake

resource	name of your e-resource, be it database, interface, etc.
title	name of the source provided in your e-resource
issn	extremely useful if source is a serial, format 1234-5678
citation_start	preferably numeric date, format YYYY-MM-DD
citation_end	preferably numeric date, format YYYY-MM-DD
fulltext_start	preferably numeric date, format YYYY-MM-DD
fulltext_end	preferably numeric date, format YYYY-MM-DD
db_identifier	an identifier for this title specific to your resource
urlbase	a base url specific to this title

Serials Solution (XML DTD)

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT Databases (Database+)>
<!--
        Element - Databases
        A set of databases and the journals and availability offered within them
        Attributes
                created: a timestamp from the generation of this holdings list
                Provider: the party offering these databases
-->
<!ATTLIST Databases
        created CDATA #REQUIRED
        Provider CDATA #REQUIRED
>
<!ELEMENT Database (Journals*)>
<!--
        Element - Database
        A database offered (including holdings information)
        Attributes
                Name: the name of the database
                HomeURL: the databases home page
                JournalURL: the databases main entry point for journal access
                PubURL: a publication URL for linking directly to the journal (by
appending the PubCode)
-->
<!ATTLIST Database
        Name CDATA #REQUIRED
        HomeURL CDATA #IMPLIED
        JournalURL CDATA #IMPLIED
        PubURL CDATA #IMPLIED
>
<!ELEMENT Journals (Journal)>
<!ELEMENT Journal (Holding+)>
<!--
        Element - Journal
        A Journal being offered and its availability
```

```
        Attributes
                Title: The journal's main title
                ISSN: The journal's print ISSN
                eISSN: An ISSN for an electronic format of the journal
                Publisher: The journal's publisher
                Frequency: Frequency of publication
-->
<!ATTLIST Journal
        Title CDATA #REQUIRED
        ISSN CDATA #IMPLIED
        eISSN CDATA #IMPLIED
        Publisher CDATA #IMPLIED
        Frequency CDATA #IMPLIED
>
<!ELEMENT Holding EMPTY>
<!--
        ELEMENT - Holding
        A range of availability for the parent journal
        Attributes
                Availability: completeness of availability for this coverage range
                DateStart: When this coverage begins
                DateEnd: When this coverage ends.  If omitted, present is assumed
                JournalURL: URL used to access a title list for these issues
                PubCode: Databases key for this journal (appended to the PubURL to make a
link directly to the journal)
                Format: The format of the journal's content as accessed by the patron
-->
<!ATTLIST Holding
        Availability (complete | selected_articles) "complete"
        DateStart CDATA #REQUIRED
        DateEnd CDATA "present"
        JournalURL CDATA #IMPLIED
        PubCode CDATA #IMPLIED
        Format (full-text | text-graphics | image) #IMPLIED
>
```

Appendix 2

Table of parties, functions, and formats

Parties (from/to)	Function	Format
Aggregation services Publication access management services	Download/transmit title lists	MS Excel, CSV, other-character-delimited, PDF, HTML, proprietary (jake, Serials Solutions)
Aggregation service Publication access management services	Download/transmit updates to title lists	MS Excel, CSV, other-character-delimited
Aggregation services Libraries	Download/transmit title lists	MS Excel, CSV, other-character-delimited, PDF, HTML, proprietary (jake, Serials Solutions), MARC 21 bibliographic records
Libraries Aggregation services	Transmit holdings lists	character-delimited files
Libraries Document supply services	Transmit holdings data	Special-character-delimited files, MARC 21 holdings records
Libraries Link resolution services	Transmit holdings lists	None (select subset representing holdings from a supplied database)
Libraries Publication access management services	Transmit holdings lists	MS Excel, CSV, other special-character-delimited files
Publication access management services Libraries	Transmit title lists	MS Excel, CSV, other special-character-delimited files

Parties	Function	Format
Publication access management services Libraries	e-mail change lists	character-delimited files
Publication hosting services Libraries	Transmit title lists	MS Excel, CSV, other special-character-delimited files
Publication hosting services Libraries	e-mail change lists	unformatted
Publication hosting services Publication access management services	Download/transmit title lists	jake, Serials Solution XML DTD, partner-specific format.
Publication hosting services Publishers	Verify/enable subscription access	Web interface
Publishers Publication hosting services	Verify subscription access	CSV

Appendix 3

Summary of NISO Librarians' Serial Survey

by Marilyn Geller

Demographics

NISO created a Librarians' Serials Questionnaire in conjunction with the NISO/DLF Serials Study and announced its availability on a number of library and serials related electronic mailing lists. The questionnaire was on the NISO website for approximately eight weeks in April and May of 2002. A total of 55 valid responses were captured. Responses came predominantly from the United States (46) with other countries including Australia, Belgium, Canada, Greece, Netherlands and Spain representing the remaining responses. Academic librarians represented the largest group with medical and health sciences librarians, government librarians, and special librarians also contributing. Respondents represented a variety of positions within their libraries including serials, administration, collection development and electronic resources among others.

Current system: rights to electronic resources

Only slightly more than one third of the respondents said their libraries communicate information on rights to electronic resources to other organizations. Of this group, almost all took this information either from their integrated library system or a local spreadsheet or database application, but this generally involved some manual processing to extract the data. The majority of respondents also said that the data were unstructured and would require human intervention to be processed at the other end. Typically, the elements included in these communications are title, publisher, ISSN and URL. Less frequent but still common are various combinations of starting and ending chronology and enumeration, licensed rights and IP range. Other elements that librarians listed included vendor and subscription number. Most commonly, this information is communicated to subscription agents and less frequently to electronic resource aggregators, publishers and indexing and abstracting services. Librarians also reported passing this information to electronic resource tracking services, link resolution services and document delivery services.

Current system: library holdings

Approximately two thirds of the respondents said they currently communicate print or electronic serials holding information to other organizations. Like the communication of electronic rights, most respondents said they take this information from either their integrated library system or a local spreadsheet or database application, and this extraction process involves some manual intervention. Unlike the data representing rights to electronic resources, the data representing library holdings are almost evenly divided between structure and unstructured form, and almost half of the structured data conform to the MARC 21 holdings format. In addition to title, publisher and ISSN, the most common elements of this communication are starting and ending chronology and enumeration and holdings gaps. Licensed rights and URLs are less common. Document

delivery services, electronic resource aggregators and subscription agents are the most frequently mentioned recipients of this type of communication with indexing and abstracting services, publishers, link resolution services and electronic resource tracking services also receiving some of these communication. OCLC and consortial union lists were also mentioned although many respondents said that this is more common for print serials holding information than for electronic serials holding information.

Current system: publisher/distributor/aggregator collections

When asked how libraries receive information about electronic resources collections and inclusive coverage, almost three quarters of the respondents listed publishers and subscription agents while slightly less than two thirds also listed electronic resource aggregators and less than a quarter also listed electronic resource tracking services. Librarians also mentioned consortia, websites and electronic mailing lists. This information is generally stored in integrated library systems and local spreadsheet or databases applications. Less frequently, it is stored in electronic resource management systems, on library websites, or in paper files. This generally involved some manual processing, and the data are most commonly in an unstructured format requiring some human intervention. The elements included in these communications mirrored the elements libraries communicate to other organizations about their holdings with one significant exception. The most common elements of this communication are title, publisher, ISSN, and starting and ending chronology and enumeration. In this category, however, URLs are extremely common. Licensed rights and holdings gaps are less common.

Current system: use of identifiers

Almost two thirds of the respondents use their OCLC symbol to identify their institution in communicating subscription data to other organizations and most do not use any identifiers to define these other organizations. When asked what proportion of their electronic journal resources are identified by a standard number in their catalogs, four respondents said all electronic resources carry a standard number such as ISSN or ISBN, seventeen said that more than 90% carried standard numbers, twelve estimated between 75 and 90%, eight answered between 50 and 75%, ten indicated the number was less than 50%, and four respondents did not answer.

Other conclusions

One general conclusion that can be drawn from this survey is that standards in the area of serials holdings have begun to make inroads in supporting industry needs for the communication of information. Where the majority of respondents have stated that data related to rights and to electronic collections is generally unstructured, a much larger percentage of respondents said that data related to holdings was structured.

In addition, there is a clear opportunity for the development of standards in areas beyond serials holdings structures. Most respondents said they use either an integrated library system or a local spreadsheet or database application to store and retrieve communicated data, and these systems and applications are fertile ground for structured data based on developing standards.

Appendix 4

A sample structure for message-based exchanges

Most subscription-based exchanges currently take place in a manual or semi-manual mode between two parties, with most data structured in an ad-hoc fashion to the specifications of one or the other of the parties in the exchange.

One alternative to this is a message-based mode of exchange, in which exchanges would have the capability of being multi-party exchanges, and the data exchanged would be structured according to a standard specification that would be adhered to by all the parties involved.

Structure of messages

Messages would consist of basic queries and more elaborate responses. The minimal elements necessary to an effective exchange of such subscription-based data would be those necessary to uniquely identify:

(1) the owner of the subscription,

(2) the target of the subscription (e.g., a given serial), and

(3) the ultimate provider of the subscription.

Standard identifiers would be needed for all of these elements, which might or might not be satisfied from one or more existing identifier schemes (in which case they would be represented by element pairs—one for the scheme and one for the value).

Because security is a principal concern, especially for providers of subscriptions, an optional fourth element in the form of a security key might be required in some circumstances to establish the legitimacy of a given message. Security concerns might also argue for not publicizing the identifiers used for the owners of subscriptions (e.g., libraries).

In this sample structure, a basic message would serve as a query from one party to another. The fifth element of such a query would identify the nature of the query, preferably in coded form, which in turn would determine the nature of the expected response.

While queries might apply to a collective subscription (e.g., an aggregation or subscription package), responses to them would typically be broken out at the level of the constituent subscribed item (serial or monograph) for effective transmission of detailed data and effective interaction with the bibliographic systems of the parties (where data are typically recorded at this level of detail).

Responses to queries would consist of the original query with a set of optional query-specific elements appended to it. For example, a response from a provider of subscription access (e.g., publisher, hosting service, aggregator) might involve the scope of the subscription and any associated restrictions in scope or access (assuming these data were maintained by the provider in a form that could be automatically accessed in response to a legitimate query). Where multiple responses were triggered by a single query (e.g., a request for the serials included in a given

aggregation), an element might be appended indicating the total number of messages in the response group and the position of the current message in that group.

Responses might occur on a one-time basis, following receipt of the query, or an ongoing basis, as the data requested by the query is updated over time on the responding site; in these cases, the responses would cease only when a second query was received terminating the exchange. An example of the former would be a request to activate full-text access to a subscribed serial from a publication hosting service. An example of the latter would be a request to provide details on constituent parts of a subscribed aggregation (which might be subject to change over time) from an aggregation service.

Message elements

The following elements are for illustrative purposes only and may or may not be part of any standard for exchanging subscription data. They are designed to accommodate re-used data from ONIX for Serials messages that may have been part of earlier exchanges with other parties.

Identifiers are represented as element pairs, the first giving the type of identifier (in coded form) and the second the corresponding value.

Basic query/response elements

The following elements would be part of all queries and all responses to those queries.

SubscriberIDType

If a uniform identifier for subscribers were not part of a standard, it would be necessary to identify here the scheme from which the identifier were drawn (e.g., GLN, SAN, proprietary (sender), proprietary (recipient), e-mail address)

IDValue

The appropriate identifier from the scheme idenitifed in SubscriberIDType

SubscriptionItemIDType

If a uniform identifier for subscription items were not part of a standard, it would be necessary to identify here the scheme from which the identifier were drawn (e.g., GTIN, ISSN (print), ISSN (electronic), ISBN, DOI, LCCN, proprietary (sender), proprietary (recipient)). If no identification scheme were applied to the item in the target database, then the unique name of the item in that database would be supplied as the SubscriptionItemID, and the value “proprietary (recipient)” would be entered here. This element would be repeatable for the same item (e.g., when including both print and electronic ISSN will increase the likelihood of success). This value and the one following might be taken from an identifier pair in an ONIX record from an earlier exchange (e.g., SubscriptionPackageIDType/IDValue, SerialTitleIDType/IDValue).

IDValue

The appropriate identifier from the scheme idenitifed in SubscriptionItemIDType. This element would be repeatable for the same item.

SupplierIDType

If a uniform identifier for suppliers were not part of a standard, it would be necessary to identify here the scheme from which the identifier were drawn (e.g., GLN, SAN, proprietary (sender), proprietary (recipient), e-mail address)

IDValue

The appropriate identifier from the scheme idenitifed in SupplierIDType

QueryType

A coded value would represent the nature of the query. This would govern the optional elements included in the reply to the query and the data source from which values for those elements were taken.

Examples of query types (and respondent) would be:

(1) activate full-text access (library/consortium/subscription agent to publication hosting service)

(2) update publication list (library/consortium/publication access management service to aggregation service)

(3) provide bibliographic records (library to aggregation service)

(4) provide holdings records (aggregation service/document supply service to library)

(5) update links to “appropriate copy” and selected services (library to link resolution services)

(6) update publication list (library to publication access management service)

(7) verify subscription rights (publication hosting service to publisher)

(8) provide linking formula (link resolution service to publication hosting service)

SecurityKey

This element would contain a security key to authenticate the message as a legitimate query.

Optional response elements

Based on the data being exchanged, optional elements in a response might include, but not be limited to, the following:

ComponentItemIDType

This element would be used when the item identified in SubscriptionItemID represents a higher-level item such as an aggregation or subscription package. If a uniform identifier for subscription items were not part of a standard, it would be necessary to identify here the scheme from which the identifier were drawn (e.g., GTIN, ISSN (print), ISSN (electronic), ISBN, DOI, LCCN, proprietary (sender), proprietary (recipient)). If no identification scheme were applied to the item in the target database, then the unique name of the item in that database would be supplied as the SubscriptionItemID, and the value “proprietary (recipient)” would be entered here. This element would be repeatable for the same item (e.g., when including both print and electronic ISSN will increase the likelihood of success). This value and the one following might be taken from an identifier pair in an ONIX record from an earlier exchange (e.g., SerialTitleIDType/IDValue,).

IDValue

The appropriate identifier from the scheme idenitifed in SubscriptionItemIDType. This element would be repeatable for the same item.

AccessBegin

Date and time access to a subscription (or one of its components) begins

AccessEnd

Date and time access to a subscription (or one of its components) ends

AccessScope

A coded value would indicate the scope of AccessNumber, AccessFundAcct, AccessFundCurrency, and AccessFundAcct (item identified SubscribedItemID or item identified in ComponentItemID)

AccessNumber

Number of accesses remaining as of the date and time of the query

AccessFundAcct

AccessFundCurrency

AccessFundAmt

These elements would relate to cases where a given package includes an associated deposit account that can be drawn down by accessing serials offered by the provider outside the formal subscription. The three elements ould identify, respectively, the identifier of the deposit account, the currency involved, and the amount remaining in the account as of the date and time of the query

HoldingsBegin

For electronic items, the date of the earliest issue to which the subscription applies. Omission implies no beginning limit. For print items, the date of the earliest issue held.

HoldingsEnd

For electronic items, the date of the latest issue to which the subscription applies. Omission implies no ending limit. For print items, the date of the latest issue held. An appropriate code would indicate that holdings were current.

HoldingsLinkIDType

This would identify the holdings record format represented by the identifier in HoldingsLinkID (e.g., MARC 21).

IDValue

This would contain the identifier of a related holdings record of the type identified in HoldingsLinkIDType. This identifier might be universally unique or unique in the context of the subscriber (i.e., unique in combination with the value in SubscriberID).

HoldingsURL

This would contain the URL of a holdings record in a library's OPAC.

Embargo

An embargo period would be indicated by a coded value (for example: 60d = 60 days; 3m = 3 months; 1y = 1 year; cy = current calendar year)

Permissions

This element would include coded values for permitted uses of the subscribed content (e.g., inclusion in course packs, interlibrary lending)

IPRange

A repeatable element containing discrete IP addresses or ranges from which access to the subscribed content is permitted