Plans to improve the quality of book metadata through mutual enhancement of ONIX and KORMARC

Jane Cho
Department of Library & Information Science, University of Incheon, Incheon, Korea

Received 15 July 2011
Revised 14 September 2011
Accepted 23 September 2011

Abstract

Purpose – With the recent expansion of the online environment, the importance of descriptive metadata assisting in a user's purchase and selection has come to the fore. Therefore the publishing circle and libraries began discussions about mutual management in order to minimize common requirements and reiteration of efforts incurred in the process of book metadata creation and distribution. In this context, the purpose of this paper is to analyze the ONIX to KORMARC conversion algorithm currently used in Korea and to propose an advanced mechanism capable of mutually assisting and enhancing metadata of both worlds.

Design/methodology/approach – The ONIX to KORMARC conversion algorithm used in Korea was analyzed, and ONIX elements which were not used in conversion but considered necessary for KORMARC to accommodate in the future, were extracted; and then a new mechanism which consists of three scenarios was proposed, with which both ONIX and KORMARC could assist in creation and reducing deficiencies of the other.

Findings – This study extracted the ONIX descriptive elements that were considered necessary for KORMARC to accommodate in the future and proposals were made for these elements to be mapped in KORMARC. In addition a more advanced mechanism was conceived with which ONIX and KORMARC could help to eliminate any deficiences in the other.

Originality/value – The mutual enhancement mechanism proposed in this study will contribute to minimizing reiteration of efforts exerted in production and distribution by providing high-quality book metadata at the right time, to both the publishing circle and libraries.

Keywords Books, Metadata, Publishers, Libraries, Korea, ONIX, KORMARC, Book metadata, Catalogues

Paper type Research paper

1. Introduction

The publishing industry and libraries are implementing metadata of different styles and contents under the different objectives of sale/distribution and discovery/identification. However, with the recent expansion of online environments, the importance of descriptive metadata assisting in a user's purchase and selection process has come to the fore in both worlds. In addition, as a single work is published in various media and styles, the publishing circle and libraries have come to face the common issue of "identification of the relevant works". Against this background, the publishing circle and libraries began discussions in order to minimize common requirements and reiteration of efforts incurring in the process of book metadata creation and distribution and to achieve mutual management of book metadata.

The Electronic Library
Vol. 31 No. 3, 2013
pp. 345-358

0264-0473
DOI 10.1108/EL-07-2011-0111

Meanwhile, the Korea Book Trade Promotion Center (www.booktrade.or.kr) with support from the Ministry of Culture, Sports and Tourism is creating ONIX (ONline Information eXchange) for new books and distributing them to not only bookstores and wholesalers, but also to libraries after KORMARC (Korean MAchine Readable Cataloging) conversion. However, the metadata systems of both communities are significantly different in terms of structure and semantic, and are therefore not completely compatible with each other. Moreover, the key elements of ONIX, such as author biographical notes, book reviews and etc., not adopted as descriptive elements in traditional catalogs, can become lost in the process of conversion into KORMARC. But these lost elements can be used in libraries afterwards for identifying or assessing relevant works or assisting in a users' selection. For such reason it is necessary to supplement the conversion process to accommodate these elements through KORMARC as well. Meanwhile ONIX can also be improved in its deficiencies, such as subject analysis and authority control, through KORMARC. Due to these reasons, it is necessary to develop the current ONIX to KORMARC conversion algorithm into a more advanced mechanism capable of mutually enhancing insufficiencies of one another.

The mutual enhancement mechanism can not only create a new KORMARC through ONIX, but also enhance insufficiencies of the existing KORMARC using ONIX data elements. Conversely, KORMARC can be used not only in the creation of a new ONIX, but also in enhancing the deficiencies of ONIX. Therefore, this mechanism will contribute to minimizing the reiteration of efforts incurring in the process of creation and distribution by providing more powerful book metadata to the publishing industry and libraries at the appropriate times.

In this context, this study is aimed to analyze the ONIX/ KORMARC conversion algorithm used for distribution of ONIX to libraries by Korea's Book Trade Promotion Center and, based on the analysis, to propose an advanced mechanism capable of mutually assisting in and enhancing metadata of both worlds.

2. Overview of book metadata

2.1 Book metadata on supply chain

The metadata managed in a business process from a publishing company to a wholesaler and a bookstore for distribution and sale of books is called book metadata. In a broad sense, MARC data created by libraries for discovery, identification and holding is a type of book metadata. National Information Standards Organization (NISO) and Online Computer Library Center (OCLC) classify the parties related to book metadata into publisher, distributor, bookstore, metadata vendor, portal site, national library and local library (NISO and OCLC, 2009).

As the demand for computer-based distribution and efficient data management in the supply chain has sharply increased as of late, the publishing circle became highly interested in the standard exchange format and quality maintenance. In addition, due to an increase of the online market share, there is now an awareness that the quality of descriptive metadata is directly connected to sales. Accordingly, the publishing group began searching for ways to create and share metadata more efficiently. Meanwhile, as a single work is published in various media and styles, descriptive elements with which relationships among diverse works and works themselves can be accurately identified were given significance in libraries. However, as the increased dependence

on search engines led to issues on the economic aspect of cataloging in libraries, libraries faced the necessity to search for alternatives through which the cost can be lowered, while quality of catalogs is improved (Cho, 2010).

Under these circumstances, libraries and publishers started discussions recently for the feasibility of mutual management of various standards in the book industry and library-related fields. OCLC organized the Symposium for Publishers and Librarians (www.oclc.org/publisher-symposium) in order to discuss the possibility to apply a more advanced mechanism in terms of the life cycle of book metadata. Through the symposium, diverse opinions were collected from the publishing circle and libraries in relation to mutual metadata management. The opinions indicated that libraries intended to use publisher data in order to increase the economic aspects of cataloging and for prompt and accurate data implementation, while the publishers desired joint handling with libraries of issues relating to securing high-quality descriptive metadata and identifying relevant works. Libraries and publishers not only employed MARC and ONIX respectively as a metadata format, but also employed different standards for identification of authors and works. Therefore, it was predicted that mutual management was not to be accomplished easily. However, they formed a consensus on endeavoring together to build a new mechanism based on the expertise of both worlds.

2.2 Metadata standard in the publishing circle and libraries

The publishing group and libraries define their own standards in order to facilitate data flow into business partners and the related organizations. As mentioned above, they employed similar, yet different formats. The publishers use ONIX for metadata management, while libraries employ MARC. In terms of classification system, the publishers uses The Book Industry Standards Advisory Committee (BISA), while libraries use Dewy Decimal Classification (DDC). In addition, while the publishing circle uses International Standard Name Identifier (ISNI) and International Standard Text Code (ISTC) for identification of authors and works, libraries use the authority control methods of Name Authority Component of the PCC (NACO) and Virtual International Authority File (VIAF). In this study, we look at the differences in metadata formats and classification systems employed by the publishing industry and libraries.

ONIX/MARC. ONIX is a schema consisting with code values that use XML message architecture and this is used as a standard in the book industry for data sharing. It is applicable to all contents of online transaction including bibliographic info., author info., book info., publisher info. and distribution info. However, unlike MARC, ONIX does not provide information on how data are used, updated and managed by users. In addition, ONIX lacks the concept of access points (Godby, 2010). On the other hand, MARC, is divided into the case in which CIP data are created first based on publishing information and the case of original cataloging by holding library. Although providing diverse bibliographic information about discovery, holding and access point, MARC does not offer details relating to distribution and sale.

BISAC/DDC. Book Industry Standards and Communications (BISAC) is a subject heading widely used in the supply chain of the publishing world and is used in arranging books or searching subject areas through search system. BISAC is expressed with nine alphabets and is divided into 52 key sections including computer, fiction and history. DDC used in libraries, on the other hand, consists of a hierarchy

that begins with ten basic subjects leading into outlines and main contents. Since its development was premised on classification of academic fields, DDC is considered to be inferior to BISAC in terms of user friendliness (Mitchell, 2010).

3. Status of book metadata in the Korean publishing circle and libraries

Korea's publishing circle and libraries have been managing the same book metadata by repetitively implementing them in different styles and contents. Therefore, a culture of mutual development and cost reduction through sharing and exchange is yet to be established in this field. In the publishing world, there is not even any interest in data sharing since it has only been managing data internally using its own management system. Distributors and bookstores, which create metadata for sale, distribution and inventory control, only implement data on the same book in different styles and contents, but do not share the data each other. On the other hand, libraries use bibliographic utility or Cataloging In Publication (CIP) data in order to increase the economic usefulness of cataloging. However, since CIP is being drawn up for only 10 percent of all publications and, moreover, shared cataloging is not activated, therefore each library is repetitively investing high levels of human resources and cost in cataloging.

To resolve these problems, the Korean Ministry of Culture, Sports and Tourism recently implemented a publication and distribution information system (Korea Book Trade Promotion Center, 2003). Through this system, which was developed with private capital from the publication-related organizations and government subsidies through the Ministry of Culture, Sports and Tourism, the ONIX-based book metadata for new publications are being distributed to the related parties on the supply chain. The Book Trade Promotion Center is creating the first ONIX with a large number of personnel based on the books published and stocked in BOOXEN (www.booxen.com), a large distributor. Then, the data implemented are supplied to book distributors, wholesalers and publishers. At the same time, the Book Trade Promotion Center is converting the bibliographic elements of ONIX into KORMARC and distributing them to libraries through the union catalog operated by Korea Education and Information Center (www.keris.or.kr). Through these activities, the Book Trade Promotion Center is spreading awareness of mutual data management, which has not been established at all in this field, and is encouraging joint use of the data.

However, it is necessary to think twice about the fact that ONIX is being produced for the first time in the stage of distribution, not publication, and that it is being managed through a business process that excludes libraries. Originally, ONIX is created as a byproduct of publication and must evolve through mutual exchange among partners in the process of business that starts from publishers. However, in Korea, ONIX is being created with a large number of personnel based on books published and stocked by large-scale distributors. This inevitably incurs high implementation costs (Kim, 2007) and, furthermore, it is difficult to include in ONIX the various descriptive elements that can be produced throughout the life cycle of publication ranging from production to sale (Kim, 2009). In addition, the fact that it fails to produce mutual synergy due to the exclusion of libraries is also a problem. In this system, libraries play an extremely passive role, which is to receive information on new publications from the publishing world. However, there should be consideration about the role of libraries to be redefined so that they can mutually interact with the

publishing circle more actively. Libraries are required to search for more developed roles, such as to supply useful information created through librarian's intellectual endeavor to the publishers in addition to receiving descriptive information created over the life cycle of publication at the right time.

4. Analysis of the next-gen project of OCLC

In this study, it is intended to analyze the ONIX/KORMARC conversion algorithm used by the Korean Book Trade Promotion Center for distribution of ONIX to libraries and, based on this, to propose the ways to mutually enhance metadata of both the publishers and libraries. Prior to this, the concept and data handling process of Next-GEN project (www.oclc.org/partnerships/material/nexgen/nextgencataloging.htm) implemented by OCLC for the purpose of mutual metadata management by the publishing industry and libraries will be discussed.

OCLC proposed the conceptual model outlined below in order for mutual sharing of various descriptive data from the publishing circle and authority data from libraries and to map the different data styles and terminologies used by them (OCLC, 2009). Roughly speaking, the process begins with capturing ONIX metadata from WorldCat (www.worldcat.org), matching the metadata against MARC and enhancing them with various algorithms. Then, the results are redistributed to both sides. In this process, enhanced software is used, with which ONIX and MARC enrich one other. In addition, rules for data mining, algorithm using Functional Requirements for Bibliographic Records (FRBR) work-set and rules for mapping of the DDC and BISAC terminologies are included in this model. The enhanced ONIX is added with LCCN, DDC, subtitle and LCSH, etc. Also, clustering information of related versions derived from the same one work is included through FRBR work-set based on the WorldCat database. Conversely, this model can be used to enhance MARC by extracting various descriptive elements from ONIX or to automatically create MARC data for new books. The afore-stated mechanism is described below.

(1) The first process is to match ONIX to MARC record of WorldCat. The result of matching may indicate existence of a record that does not match WorldCat. ONIX on a new book that has just recently been published falls in this case. For this:

- Create a new MARC record by extracting bibliographic information from ONIX.
- Execute FRBR work-set clustering on the newly created MARC record. This is because, even a newly published book may be subject to an iteration, such as a new edition of a previous work, paperback, audio book, e-book and large print. The work level information of FRBR (subject analysis, original work, author information and etc.) can be inherited from these related works. These elements enhance bibliographic data of a newly created MARC record and establish a chain of connection with related works derived from the same work.
- Data elements can equally enhance ONIX and the enhanced ONIX can be used in the publishing circle, too.

(2) The result of ONIX and MARC record matching may also indicate that ONIX corresponds to the MARC record of WorldCat. This happens when some of the

libraries participating in WorldCat already have this book in their collection and therefore created a catalog record. For this:

- Extract data elements that are imperfect or do not exist in MARC from ONIX record and enhance MARC record with the elements extracted.
- Conversely, extract data elements that are imperfect or do not exist in ONIX from MARC record and enhance ONIX with the elements extracted.
- At the same time, extract FRBR work-set information created in MARC and connect it to the matched ONIX.

5. KORMARC/ ONIX mutual enhancement mechanism

In the previous section, the Next-GEN project algorithm of OCLC was briefly discussed. In this section, the mutual enhancement mechanism of KORMARC and ONIX is developed in the following order by applying the idea from OCLC. First, we will look at the conversion algorithm currently used by the Korean Book Trade Promotion Center, which is converting ONIX into KORMARC and distributing them to libraries. Then, issues will be derived from it. Second, among ONIX elements, those found necessary to be accommodated by KORMARC will be derived and appropriate mapping positions of KORMARC fields will be proposed. Third, based on the findings, the KORMARC/ ONIX mutual enhancement mechanism suitable to the conditions in Korea will be conceived.

5.1 Analysis of the current ONIX to KORMARC conversion algorithm

The Korean Book Trade Promotion Center, which create the first ONIX, converts bibliographic elements of ONIX into KORMARC and opens them to be used in necessary places. Table I summarizes the ONIX to KORMARC conversion algorithm (Korea Book Trade Promotion Center, 2006) used by the Center.

With an exception of the following fields – PR4 (E-publication detail), PR16 (Link to media file), PR21 (Territorial right and other sales restriction), PR23 (Related product), PR24 (Supplier, availability and price) and PR25 (Sales promotion information), the Korean Book Trade Promotion Center's conversion algorithm crosswalks ONIX with the basic elements of KORMARC. ONIX and KORMARC are difficult to be completely compatible due to their structural differences including description level. The current algorithm processes it by simplifying or eliminating this problem. In particular, the descriptive elements of biographical notes, readers' descriptions and book reviews dealt with in subfield of PR 14 and 15 fields of ONIX have not been handled by libraries through cataloging. Therefore, these elements are excluded from the mapping rules above. However, due to the reasons outlined below, the importance of these elements is being emphasized as of late:

- assisting users to easily select the desired works among relevant works published in diverse media and styles;
- effectively understanding authors' authority and readers' age groups or education levels; and
- helping in making decisions for purchase and selection through diverse information of book assessment.

ONIX tag	MARC tag
PR.1 RecordReference	001 Control Number
PR.2 ISBN	020 ISBN
PR.3 ProductForm	001 Control Number (add initial to control number)
PR.5 Title Of Series	440 Series Statement
PR.6 Item Number Within Set	440 Series Statement
PR.7 Distinctive Title Translation OfTitle	245 Title Statement
	507 Original Version Note
PR8. Person Name Contributor Role	245 Title Statement
PR.9 Conference Name	111 Conference Name
PR.10 Edition Statement	250 Edition Statement
PR.11 Language Of Text	041 Language Code
PR.12 Pages Arabic Pages Roman Number Of Pages	300 Physical Description
PR.13 Subject Code Subject Scheme Identifier	050 Library of Congress Call Number
Subject Heading Text	082 DDC
	056 KDC
	653 Subject Heading
PR.13 Person As Subject	700 Added Entry – Personal Name
PR.14 AudienceCode	008 23
	506 Restrictions on Access Note
PR15 Main Description	520 Summary
PR.17 Prizes Description	586 Awards Note
PR.18 Title, Title Text Title Type	505 Formatted Content Note
PR.19 City Of Publication	008 6-10
PR.20 Publication Date Copyright Year	260 Publication, Distribution
PR.20 Announcement Date	005 Date and Time of Latest Transaction
PR.24 Price Amount	950 Price

Table I.
Summary of ONIX to KORMARC conversion algorithm

Therefore, it is necessary to supplement the current conversion algorithm and to look for ways with which the algorithm can be accommodated even if it requires partial modification of KORMARC structure.

5.2 Proposing ONIX elements to be accommodated by KORMARC

Table II proposes the positions of fields that can be used in extracting and converting ONIX elements found necessary to be accommodated by KORMARC. The positions have been indicated for a case in which the existing KORMARC field can be used and a case in which additional definition of the KORMARC field is necessary. KORMARC was developed on the basis of MARC21 structure. Therefore there is not a big difference, but in case of subfields, some differences exist. For example, in MARC21 100$g is defined as 'Miscellaneous information', but in KORMARC 100$g is defined as "Korean or Chinese reign rank". In addition, many subfields which have been assigned in MARC21, such as 100$u (Affiliation) 100$j (Attribution qualifier) are not assigned in KORMARC. Therefore to avoid violating the MARC21 structure, if some tag has to be added to KORMARC, preferentially use appropriate MARC 21 fields. Otherwise assign new codes which have not used in MARC 21. Existing/New field in Table II shows each case which using existing KORMARC fields or suggesting new fields. The key process is summarized as follows (elements, which already included in the current ONIX to KORMARC algorithm have been omitted.)

Table II.
ONIX elements to be additionally accommodated by KORMARC

Section	ONIX Tag	Element name	KORMARC Tag	Element name	Existing/new
Authorship	PR.8.22	ProfessionalPosition	100$c	Titles and words associated with a name	Existing
	PR.8.23	Affiliation	100$u	Affiliation	New (marc21 has)
	PR.8.27	BiographicalNote	100$1	Biographical note	New
	PR.8.31	ContributorDescription	100$ 2	Contributor description	New
Audience	PR.14.6	InterestAge(s)	521$a	Target Audience	Existing
	PR.14.12	AudienceDescription	521$c	Interest age(s)	New
	PR.14.8	AudienceRangePrecision (list 31)	521$d	Audience range precision	New
Description and other supporting text	PR.15.2	MainDescription	500$a	General Note	Existing
	PR.15.4	OtherTextFormat (list 34)	556$q	Other text format	New
	PR.15.5	OtherText	556$b	Other text	New
	PR.15.6	OtherTextLinkType (list 35)	556$2	Other text link Type	New
	PR.15.7	OtherTextLink	556$u	Other text link	New
	PR.15.8	AuthorOfOtherText	556$c	Author of Other text	New
	PR.15.9	CorporateSourceOfOtherText	556$d	Corporate source of other text	New
	PR.15.10	TitleOfSourceOfOtherText	556$t	Title of source of other text	New
	PR.15.11	PublicationDateOfOtherText	556$e	Publication date of other text	New
	PR.15.12	TextValidFromDate	556$f	Text valid from date	New
	PR.15.14	ReviewQuote(complete)	556$g	Review quote	New
Image/audio/video file link composite	PR.16.1	FrontCoverImageFileFormatCode (list 36)	856$q	Electronic format type	Existing
	PR.16.2	FrontCoverImgaeFileLinkTypeCode (list 37)	856$2	Access method	Existing
	PR.16.3	FrontCoverImageFileLink	856$u	Uniform Resource Identifier	Existing

(*continued*)

Section	ONIX Tag	ONIX Element name	KORMARC Tag	KORMARC Element name	Existing/new
Publisher	PR.19.17	Co-publisherName	260$h	Co-publisher name	New
	PR.19.18	SponsorName	260$i	Sponsor name	New
	PR.19.19	PublisherOfOriginal-languageVersion	260$j	Publisher of original-language version	New
Copyright statement composite	PR.20.4	CopyrightYear	260$k	Copyright year	New
	PR.20.8	PersonName	260$m	Person name	New
	PR.20.9	CorporateName	260$n	Corporate name	New
	PR.20.11	YearFirstPublished	260$o	Year first published	New
Related products	PR.23.1	Replaced-byISBN	785$z	Succeeding Entry ISBN	Existing
	PR.23.3	Alternnative-formatISBN	775$z	Other Edition Entry ISBN	Existing
	PR.23.8	ISBNOf Related Item (list 51)			
		< RelationCode >	762$z	Sub series Entry ISBN	New
		1. Include			
			760$z	Main series Entry ISBN	New
		3. replaces	780$z	Proceeding Entry ISBN	Existing
		5. replaced by	785$z	Succeeding Entry ISBN	existing
		6. alternative format	775$z	Other Edition Entry ISBN	Existing
		7. has ancillary product	773$z	Host Item Entry ISBN	Existing
		8. is ancillary to	772$z	Supplement Parent Entry ISBN	Existing
		12. Publisher's suggested alternative	787$ z	Other relationship entry ISBN	Existing
		13 Epublication based on (print product)	776$ z	Additional Physical Form ISBN	Existing
		14. FRBR cluster entry	787$ z	Other relationship entry ISBN	Existing

Table II.

First, professional position, affiliation and biographical note have been additionally defined as sub-fields to be accommodated into tag 100. Second, the sub-fields have been added to tag 521, which is used as a user target, so that interest age(s), audience description and audience range precision could be indicated. Third, since an appropriate tag to arrange the elements of book review, etc., which takes a large part of ONIX, does not exist, tag 566 has been newly defined in the 500s. In ONIX, sub-fields are defined to link book reviews consisting of long and difficult sentences that are updated on a real-time basis. Therefore, by defining the related sub-fields in tag 556, it was enabled to include short book reviews in KORMARC or, if necessary, to link them to external resources. For this, it is required to define sub-fields with which to handle text link. This was configured by referring to the method of tag 856. Fourth, image and catalog information have been arranged for linking through field 856 with resources located outside. Fifth, a new sub-field was proposed so that co-publisher name, sponsor name and publisher of original-language version, which are not handled in tag 260 of KORMARC, can be accommodated into details of publication. And information for copyright holder identification was added. Sixth, ONIX elements about the related products were mapped in the linking entry field of KORMARC. In ONIX, diverse relation types about "include", "is part of", etc. are defined and it has been arranged to have the elements accommodated using a similar field in KORMARC. Seventh, in ONIX, a variety of code lists on text link type, link type and relation type not defined in KORMARC are used. It was proposed to convert the code name into text and to enter it into the corresponding KORMARC field to be made reference of.

5.3 Conceiving ONIX/ KORMARC mutual enhancement scenario

In the previous section, the current KORMARC to ONIX conversion algorithm was analyzed and ONIX elements considered necessary to be accommodated by KORMARC in the future were extracted and therefore the appropriate mapping positions were proposed. In this section, the method to capture ONIX elements from bibliographical database, to mutually enhance them and then to re-share them between libraries and the publishing circle will be proposed.

First, ONIX created by the Korea Book Trade Promotion Center is captured and then the ONIX is matched against KORMARC which is constructed in UNICAT, the largest union catalog database of Korea. The following three cases may occur in this process and thus different scenarios are proposed considering each of the three cases. The first case is where ONIX exists, but KORMARC does not. Most new publications fall into this category. The second case is where both KORMARC and ONIX exist. This case occurs when some of the libraries participating in UNICAT have a book in collection and create catalog data on the book. The third case is where KORMARC exists, but ONIX does not. Although it is not considered in the Next-GEN project of OCLC, the instance of direct distribution by publishers or libraries getting hold of them before distributors and creating MARC may take place in Korea where the first ONIX is created in the stage of distribution. In addition, the instance where CIP is created before ONIX is creating may also fall into this category.

A. Result of MARC and ONIX matching indicating that the same MARC as ONIX does not exist in bibliographical DB

(1) Processing:

- Supplement "ONIX to KORMARC conversion rules" by adding the elements proposed in Table II of the previous section. Then, create a new KORMARC according to the supplemented conversion rules.
- Based on the newly created KORMARC, extract FRBR cluster from bibliographical database. A variety of related titles, such as the original work of a translated version, version in other media and the modified and enhanced version, can be extracted.
- Enhance KORMARC by extracting information with which work level data (subject, classification No., original author, title of original edition, etc.) commonly exist in related versions can be linked to all related titles.
- Conversely, enhance ONIX with elements extracted and redistribute it through Book Trade Promotion Center.
- Redistribute it through Book Trade Promotion Center to be utilized by the publishing circle.

(2) Used elements:

- ONIX: (Ex.) ISBN, TitleOfSeries, DistinctiveTitle, Authorship, EditionStatement, LanguageOfText, NumberOfPages, SubjectCode, SubjectSchemeIdentifier, AudienceCode, TableofContent, FrontCoverImage, DescriptionandOtherSupportingText, PrizesDescription, PublisherName, Co-publisherName, CopyrightYear, PriceAmount, RelatedProducts, etc.
- FRBR Cluster Info.: (Ex.) Original Title, Original Author, Classification Number, Subject heading, Relationship Entry, etc.

(3) Outcomes:

- Enhanced ONIX;
- New KORMARC.

B. Result of MARC and ONIX matching indicating that the same data exist

(1) Processing:

- Using the ONIX to KORMARC conversion rules supplemented by Table II, extract elements not existing in MARC from ONIX and enhance the elements extracted.
- Reversely, extract and enhance elements not existing in ONIX from MARC.

(2) Used elements:

- ONIX elements not existing in KORMARC: Ex.) ProfessionalPosition, BiographicalNote, Affiliation, InterestAge(s), AudienceDescription, TableOfContent, FrontCover Image, DescriptionAndOtherSupportingText, PrizesDescription, CopyrightYear, PriceAmount, Co-publisherName, SponsorName, RelatedProducts, etc. and other omitted or weak KORMARC elements.

- KORMARC elements not existing in ONIX: Ex.) Original Title, Original Author, Classification Number, Subject Heading, Relationship Entry, (Authorized) Main Entry etc. and other omitted or weak ONIX elements

(3) Outcomes:

- enhanced ONIX;
- new KORMARC.

C. Result of KORMARC and ONIX crosswalk indicating that ONIX the same as KORMARC does not exist in ONIX DB

(1) Processing:

- Create the first ONIX using KORMARC(CIP) data. Here, the "KORMARC to ONIX conversion rules", which make references of "ONIX to KORMARC", are used. In this ONIX, the basic author and author control details as well as information on subject analysis are created, and then accommodation of the related work information, created by FRBR cluster are added.
- Details of publication, distribution and physical description must be checked in the process of ONIX creation.

(2) Used elements: Bibliographical elements of KORMARC(CIP): ISBN, Title Statement, Subject Heading, Classification Number, (Authorized) Main Entry, Relationship Entry

(3) Outcome: new ONIX.

Figure 1 illustrates example "A" where the same ONIX record is not matched in bibliographical DB. This is the case in which new KORMARC data are created by ONIX data and are enhanced by FRBR work-set mining. From ONIX data, ISBN, Title, Subtitle, Contributor, Contributor description, Language, Audience level, Book award, Other text, Other text link, Other text format, Table of Content, Publication date,

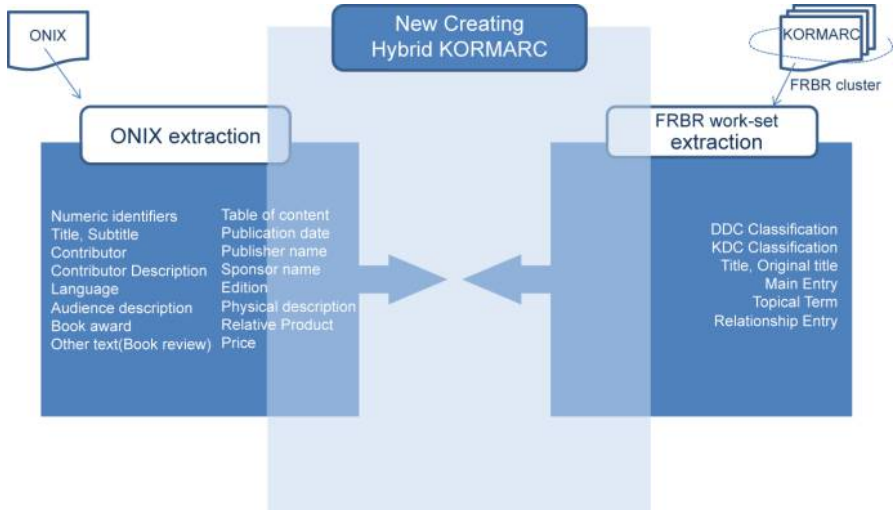

Figure 1.
Example of hybrid MARC creation by ONIX

Publisher name, Sponsor name, Edition, Physical description, Pagination, Subject text, Price, etc. are extracted. With FRBR mining result, DDC, KDC, Original title, Main Entry, Original Author, Linking entry, Subject Heading, and etc. can be extracted and reestablished as hybrid records. Although not listed in the table below, the enhanced data elements can also be enhanced in ONIX at the same level.

6. Conclusion

The Korea Book Trade Promotion Center is converting ONIX into KORMARC for distribution of new publication information to libraries. However, the fundamental structural difference between ONIX and KORMARC inevitably generates large data loss in the process of conversion. In the recent online environment, it became necessary to propose detailed descriptive information with which to identify or assess relevant works or to assist in users' decision making for purchase and selection. Accordingly, although not traditionally employed by libraries, it became necessary to actively accommodate diverse descriptive elements created in the process of publication and distribution.

In this study, the current ONIX to KORMARC algorithm was complemented and among ONIX descriptive elements, those considered necessary to be accommodated by KORMARC were extracted. In addition, proposals were made to enable these elements to be mapped in KORMARC. Moreover, a more advanced mechanism was conceived with which both ONIX and KORMARC can assist in mutual creation and enhancing insufficiencies of one another. We proposed scenarios of the processes where not only a new KORMARC is created through ONIX, but also insufficiencies of the existing KORMARC can be enhanced with ONIX data elements and, reversely, not only a new ONIX is created through KORMARC, but also insufficiencies of the ONIX can be enhanced.

The publishing world recently came to recognize that using metadata is an extremely important factor in a users' decision making for purchase and, accordingly, began emphasizing efficient data preparation and quality maintenance. Libraries, too, started to pursue a model with which operation could be simplified using data sources closer to the first information source at the same time as pursuing economic usefulness of cataloging. The mutual enhancement mechanism proposed in this study will contribute to minimizing reiteration of efforts exerted in the process of creation and distribution by providing high-quality book metadata to libraries and publishers at the right time through mutual management of ONIX and KORMARC.

References

Cho, J. (2010), "A study on the book metadata creation and distribution on supply chain", *Journal of the Korean Library and Information Science Society*, Vol. 44 No. 3, pp. 61-80.

Godby, C.J. (2010), "Mapping ONIX to MARC", available at: www.oclc.org/research/publications/library/2010/2010-14.pdf. (accessed 13 September 2011).

Kim, K. (2007), "Research on the method to modernize publication distribution in Korea", Master's thesis, Jung-ang University.

Kim, S. (2009), "Study about Korean CIP program", 46th national library conference. available at: http://conference.kla.kr/. (accessed 12 September 2011).

Korea Book Trade Promotion Center (2003), "Research on the actual condition and improvement directions for the Korea Book Trade", available at: www.booktrade.or.kr/publication/att_file/Research_report.pdf. (accessed 12 September 2011).

Korea Book Trade Promotion Center (2006), "KORMARC ↔ ONIX conversion table", available at: www.booktrade.or.kr/new_notice/att_file/KORMARC_ONIX_CHANGE_NEW.pdf. (accessed 14 September 2011).

OCLC (2009), "From ONIX to MARC and back again: new frontiers in metadata creation at OCLC", ALA Midwinter, January 25, available at: www.oclc.org/us/en/multimedia/2009/files/ONIX_to_MARC.ppt. (accessed 13 September 2011).

Mitchell, J.S. (2010), *BISAC-DDC Mappings*, available at: www.oclc.org/dewey/news/conferences/bisac.ppt. (accessed 12 September 2011).

NISO and OCLC (2009), "Streamlining book metadata workflow", available at: www.niso.org/publications/white_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf. (accessed 14 September 2011).

About the author

Jane Cho (PhD) is currently an Associate Professor in the Department of Library & Information Science, University of Incheon, Korea. From 2008-2010 she was a full-time Lecturer in the Department of Library & Information Science at Incheon City College. Prior to that she worked as Senior Researcher of the Korea Education & Research Information Service (KERIS) from 1997-2007. She has published several articles on OPACs, resource sharing and library management systems. Jane Cho can be contacted at: chojane123@naver.com