

공급사슬상의 도서메타데이터 생성·유통에 관한 고찰

A Study on Book Metadata Creation and Distribution on Supply Chain

조재인(Jane Cho)*

목 차

- | | |
|-------------------------------------|--------------------------|
| 1. 서론 | 4.1 공급사슬상의 도서메타데이터 흐름 분석 |
| 2. 공급사슬상의 관계자들과 도서메타데이터 | 4.2 효율적 생성·유통 방안 제시 |
| 3. 출판계와 도서관계 메타데이터 유사 표준과 상호 운용성 분석 | 5. 결론 |
| 4. 도서메타데이터의 효율적인 생성·유통 방안 고찰 | 5.1 요약 |
| | 5.2 제언 |

초 록

최근 출판계는 이용자의 구매 의사 결정에 있어 메타데이터 활용이 매우 중요한 요소임을 인식하게 됨에 따라, 효율적인 데이터의 작성과 품질 유지, 그리고 공급사슬상의 표준과 교환 시스템에 대해 관심을 갖게 되었다. 도서관계 또한 목록 작성의 경제적 효용성을 추구하면서 좀 더 최초 정보원에 근접한 데이터 소스를 통해 작업을 간소화할 수 있는 모델을 추구하게 되었다. 본 연구는 발생 기원은 동일하지만 각각 출판계와 도서관계에서 각기 다른 형식과 표준으로 활용되고 있는 도서메타데이터의 흐름을 살펴보고 공통의 쟁점과 상호 운용의 가능성을 조망하였다. 최근 출판계와 도서관계를 중심으로 논의되고 있는 각종 이슈들을 점검해 보고 도서메타데이터의 효율적인 생성·유통 방안에 대해 고찰하였다.

ABSTRACT

Recently, the publishing community now recognizes the importance of metadata in customers' buying decisions. As a result, they are more interested in effective metadata creation and quality maintenance, as well as standardization of exchanging system in the supply chain. As the library community also investigates the economic effectiveness of creating metadata, they try to find the best model for simplifying metadata creation by using sources close to the original. This study analyzes metadata work flow which had same source but be used in different fields by their own type and standard. It also discusses the same issues about each section and possibility about interoperation. Finally this paper tries to find an effective creation and distribution model of book metadata which can be used in domestic publishing and the library community.

키워드: 출판 공급사슬, 도서메타데이터
ONIX, MARC, CIP

* 인천대학교 문헌정보학과 조교수(chojane123@naver.com)
논문접수일자: 2010년 7월 19일 최초심사일자: 2010년 7월 26일 게재확정일자: 2010년 8월 17일
한국문헌정보학회지, 44(3): 61-80, 2010. [DOI:10.4275/KSLIS.2010.44.3.061]

1. 서론

출판계는 유통과 판매를 위해 메타데이터를 생산하고 도매상, 서점으로 이어지는 비즈니스 프로세스에 의해 공급사슬상(Supply Chain)¹⁾의 파트너들과 공유하게 된다. 한편, 도서관계는 발견과 식별, 소장을 위해 메타데이터를 생성하고 국가대표도서관, 서지유틸리티기관, 개별도서관으로 이어지는 일련의 흐름 속에서 상호 공유하거나 교환하게 된다. 도서메타데이터는 그 발생 기원은 동일하지만, 출판계와 도서관계의 공급사슬상에서 고유의 목적에 의해 각기 다른 표준과 형식으로 생산되어 진화되어 간다.

온라인 환경에서 출판계는 그동안 비즈니스 목적으로 내부적으로만 활용하였던 메타데이터를 이용자들에게 쉽게 공개할 수 있게 되었다. 더구나 이용자의 구매 의사 결정에 있어 메타데이터 활용이 매우 중요한 요소임을 인식하게 됨에 따라, 효율적인 데이터 작성과 품질 유지, 그리고 공급사슬상의 표준과 교환 시스템에 대해 관심을 갖게 되었다. 한편 도서관계에서도 검색 엔진의 등장으로 도서관 목록의 중요성이 감소되기 시작하면서, 카피목록, 자동처리를 통해 신속성을 제고하고 비용을 삭감하기 위한 총체적 프로세스 정비가 요구되게 되었다. 그에 따라 공동목록이나 아웃소싱을 통해 비용을 절감할 뿐 아니라, 출판계의 정보원(Upstream Metadata)을 통하여 작업을 간소화할 수 있는 모델도 고민하게 되었다. 그러한 맥락에서 LC(Library of Congress)는 출판계 메타데이터 반입을 통한 기술목록 작성 계획을 발표하였으

며, OCLC(Online Computer Library Center)도 'A Symposium for Publishers and Librarians'²⁾를 계기로 도서메타데이터 생성과 유통의 새로운 패러다임을 모색하게 되었다.

본 연구에서는 최근 논의되고 있는 도서메타데이터에 관한 이슈들을 종합하여, 출판계와 도서관계의 공통 요구와 중복 노력, 그리고 상호 운용의 가능성 등을 검토해 보고자 한다. 더불어 이를 기반으로 도서메타데이터 생명 주기에 있어 보다 진화된 메카니즘의 도입 가능성을 모색해 본다. 본 연구는 첫째, 공급사슬상의 관계자들 입장에서 메타데이터의 유통 경향과 관련 쟁점을 분석하였다. 둘째, 출판계와 도서관계 양측에 적용되고 있는 유사 데이터 표준과 상호 운용성을 살펴보았다. 셋째, 현행 도서메타데이터 흐름상의 특징을 파악해 보며, 양측이 중복 노력을 최소화하고 상호 긍정적으로 메타데이터를 운용할 수 있는 방안을 고찰해 보았다.

2. 공급사슬상의 관계자들과 도서메타데이터

NISO(National Information Standards Organization)와 OCLC는 도서메타데이터 관계자들을 출판사, 도서유통업체, 서점, 메타데이터벤더, 포털, 국가대표도서관 및 로컬도서관으로 구분하고 있다. 본 장에서는 NISO와 OCLC의 조사(NISO and OCLC 2009), 'A Symposium for Publishers and Librarians' 관련 자료, 한국출판

1) 공급업체에서 고객에 이르는 모든 거래 파트너와의 정보, 자금, 지식의 흐름을 지칭하는 용어.

2) A Symposium for Publishers and Librarians. <<http://www.oclc.org/publisher-symposium/>>.

유통진흥원(<http://www.booktrade.or.kr/>) 등을 참조하여 개별 관계자들 입장에서 도서메타데이터의 유통 경향과 관련 쟁점들을 종합해 보도록 한다.

2.1 출판사

미국의 대형 출판사는 출판물의 제작과정에서 발생하는 메타데이터를 ONIX(ONline Information eXchange) 형식으로 관리하고 있다. 그러나 아직까지 많은 중소규모 출판사들은 엑셀을 비롯해 독자적인 포맷을 채택하고 있다. 더구나 도서 홍보나 신간 안내를 위해 제작하는 프린트 카탈로그가 메타데이터의 역할을 대신하기도 한다. 이러한 상황은 혼자서 기획, 편집, 영업까지 도맡아 하고 있는 개미출판사가 전체의 1/4 가량을 차지하고 있는 우리나라의 경우에 더 일반적이라고 할 수 있겠다. 기존에는 출판사들이 고유의 관리 시스템을 통하여 데이터를 내부적으로만 관리하였기 때문에 비즈니스 파트너 간의 공유와 교환에 대한 요구가 발생하지 않았다. 그러나 공급 사슬의 유통 전산화와 데이터 관리의 효율화에 대한 수요가 발생하면서, 표준 데이터 형식이 필요하게 되었다. 또한 온라인 환경에서 출판사들이 중간 유통 과정을 생략하고 이용자와 직접 접촉하게 되면서, 메타데이터의 생성과 공급사슬망내에서의 품질 유지에도 관심을 갖게 되었다. 이러한 배경으로 최근 BISG(The Book Industry Study Group)은 ONIX 파일을 평가하고 데이터의 품질과 적시성의 기준을 준수하는 출판사들을 식별하는 프로세스도 운영하게 되었다.

2.2 도서 유통 업체

도서 유통 업체인 Baker & Taylor, Ingram은 매해 다루게 되는 도서의 숫자가 10%씩의 증가율을 보이고 있다고 하였다. 실제 20만권 정도가 증가하지만 디지털 포맷을 비롯한 다양한 구현형의 등장으로, 한 개의 콘텐츠가 보통 2-3개의 매체로 간행되고 있기 때문이라고 한다. 보통 유통 업체들은 CIP(Cataloging in Publication)보다 출판사의 메타데이터를 우선적으로 채택하게 되지만, 수서와 동시에 MARC 데이터 제공을 원하는 도서관들을 위해 미의회도서관의 PCC(Program for Cooperative Cataloging)에도 참여하고 있다. 한편, 복센, 송인서적, 한국출판협동조합과 같은 한국의 대표적인 유통 업체들은 한국출판유통진흥원과 공동으로 자체 입고 도서에 대한 ONIX를 작성하고 있다. 출판과정의 부산물로 생성되는 해외 사례와 달리 유통 단계에서 최초의 ONIX가 작성되고 있는 셈이다.

2.3 서점

한편 대표적인 오프라인 서점인 반스앤노블은 양질의 기술 메타데이터가 매출과 직결된다는 생각을 가지고 있다. 우리나라의 최대 서점인 교보 문고 역시 질 좋은 메타데이터가 마케팅의 경쟁력이라고 판단하여 데이터의 무상 공유에는 다소 제한적인 입장을 보이고 있다. 기술 데이터의 품질 제고와 정확성, 신속성 제고를 위하여 반스앤노블은 출판사와 유통업체에게 44개의 필수 데이터 요소를 요구하고 있다. BISG가 정의하고 있는 요소들이 대부분 포함

되지만, 기술 요소 이외에도 재고 여부, 배포처 리스트, 가격과 같이 물류 변화에 따른 가변적 요소도 매우 중요하게 간주되고 있으며, 그런 측면에서 간혹 ONIX보다 EDI(Electronic Data Interchange)가 더욱 탁월한 구조로 평가되기도 한다. 각기 다른 소스들로부터 데이터를 수집하게 되는 대형 온라인 서점에서는 범용적으로 수용되는 표준이나 최우량사례(Best Practice)가 데이터 유지 기준을 결정하는 중요한 요소가 된다.

2.4 메타데이터 벤더

NISO의 보고서에서는 데이터를 수집하여 보강하고 재배포하는 Bowker, Nielsen Book, 서지 유틸리티 기관인 BDS(Bibliographic Data Services), 그리고 OCLC를 메타데이터 벤더로 정의하고 있다(NISO and OCLC 2009). 우리나라의 경우에는 ONIX를 생성하고 각계에 필요한 형식으로 배포하는 한국출판유통진흥원, 서지유틸리티서비스를 제공하는 한국교육학술정보원 등이 여기에 해당된다고 볼 수 있겠다. 벤더들은 다양한 형식의 데이터를 수집한 후, 특정 표준에 따라 보강하여 MARC 또는 ONIX in XML 형식으로 유통업체, 판매상, 도서관 등에 재배포한다. 데이터를 수집하여 가공하고 다시 배포하는 역할을 하는 이들에게 있어 상호운용성은 매우 중요한 요소인데, ONIX가 50%, 엑셀을 비롯한 기타 디지털 형식이 45%, 나머지는 5%는 인쇄물 형태로 수집되고 있다. Bowker와 Nielsen Book은 서명과 저자명의 전거 통제

를 통하여 데이터를 보강시킬 뿐 아니라, 표지 이미지, 목차, 저자 양력, 북 수상, 독자 수준, 이용자 리뷰와 평가, 추천 정보 등을 여기에 추가시키고 있다. 또한 ISBN을 등록하고 출판사 프리픽스를 부여하며, CISAC(International Confederation of Societies of Authors and Composers), IFFRO(The International Federation of Reproduction Rights Organizations)와 함께 ISTC(International Standard Text Code)를 창설하여, 저작권 관리 업무 효율화를 위한 기반을 마련하기도 하였다. 한편, UK의 CIP 아웃소싱 업무를 맡고 있는 BDS는 75,000건의 출판권 레코드를 만들어낼 뿐 아니라, ONIX를 MARC21로 변환시켜 보급하고 있다고 하며, 매해 6-12만 개의 서지레코드를 수집하고 있는 OCLC는 NextGen Pilot³⁾을 통해 보강된 ONIX를 생성시켜, 역으로 출판계에 재보급할 계획을 가지고 있다고 한다. 그 만큼 메타데이터 벤더들은 공급 사슬내에서 데이터의 품질 제고와 배포에 있어 중요한 중간자 역할을 수행하고 있다. 한편, 우리나라에서는 한국출판유통진흥원이 ONIX를 생성하고 각계에 필요한 형식으로 배포하는 역할을 수행한다. 한국출판유통진흥원은 앞서 언급한 바와 같이 유통 업체인 북센의 창고에 입고된 신간도서에 대한 ONIX를 작성하여 서점, 유통 업체 등에 보급하며, 데이터를 MARC으로도 변환하여 국립중앙도서관, 한국교육학술정보원 및 일선 도서관에 다양한 방식으로 보급하고 있다.

3) Next Generation Cataloging. <<http://www.oclc.org/partnerships/material/nexgen/nextgencataloging.htm>>.

2.5 포털

한편, 저작권 침해 문제로 출판계와 작가 조합 등에 의해 제소된 바 있는 구글은 최근 긴 협상 끝에 저작권 비용에 관련된 Book Rights Registry의 운영을 맡게 되었다. 이 등록소를 통해 권리 보유자들의 데이터베이스를 관리하고, 그들을 확인하여 소재를 파악하며, 지급을 조율하게 된다. 따라서 저자와 저작을 식별하고 저작 간의 관계를 파악하는 문제, 또한 시리즈와 멀티 볼륨 저작을 식별하고 관리하는 문제 등이 이슈가 되고 있다. 구글은 이 문제의 해결을 위해 최근 메타데이터의 관리를 강조하게 되었으며, 더불어 유관 저작 식별 알고리즘을 개발하는 등 도서와 저작의 식별에도 많은 관심을 기울이고 있다. 한편, 우리나라의 네이버는 책서비스를 통하여 출판사들의 신간을 홍보하고, 구매자들과 연결될 수 있는 직접적인 경로를 제공하고 있다. 또한 국내 주요 온라인 서점과 관종별 도서관 종합목록을 통합한 명실공히 도서 통합 포털을 지향하고 있어, 유관 저작의 식별과 품질 유지의 문제가 주목된다.

2.6 국가대표도서관

데이터 교환을 위한 표준과 서지 제어에 관한 각종 정책을 개발하는 미의회도서관은 R2 컨설팅을 통해 북미 지역의 MARC 시장을 분석하고 업계의 불필요한 중복에 대하여 지적한 바 있다 (R2 Consulting LLC 2009). 목록의 경제성에 관한 부분은 최근 도서관계에 화두가 되고 있는데, 이러한 맥락에서 LC는 PCC 프로젝트를 통해 협동 목록과 전거의 중요성을 유도하고 있으

며, 도서 산업과 저작권, 도서관과 정보 분야에서 사용되는 다양한 표준들이 상호 운용되어야 한다는 점을 매우 강조하고 있다. 국가대표도서관은 출판계와의 데이터 공유를 위해서도 많은 연구를 수행하고 있는데, 전자 저널의 XML 포맷에서 핵심 요소를 추출하여 ONIX 레코드로 만드는 프로젝트, 그리고 XML이나 텍스트 베이스 PDF에서 메타데이터를 추출하여 METS(Metadata Encoding and Transmission Standard)와 함께 MODS(Metadata Object Description Schema) 포맷으로 활용하는 연구가 주목되고 있다. 한편, 국가대표도서관은 출판사로부터 받은 정보를 기반으로 CIP를 작성하여 도서관매업자에게 신간에 대한 출판전 도서 목록을 제공하고 있다. 출판전 도서 목록은 출판이 완료된 이후에 제목, 크기, 페이지 등의 수정 작업을 거쳐 완벽한 목록으로 보강되고, 여기에 주제 분석, 전거 통제 등의 지적인 작업이 더해져 도서관계에 여러 가지 창구로 재 보급된다.

2.7 로컬 도서관

NISO와 OCLC의 보고서에 의하면(2009), 일선 도서관에서는 30% 미만에 대해서만 원목 작업을 수행하고 있다고 한다. 그러나 아직까지 다운로드 받은 많은 레코드들이 로컬 도서관의 목록 작성 관행에 의해 수정되고 있다고 보고되고 있다. LC가 보급한 레코드 350,000건 중 80%가 로컬 도서관에 의해 업그레이드되었고, BL이 만든 26만건 중에서도 55%가 업그레이드되었다고 하니, 국가 대표도서관이 완벽한 수준의 서지데이터를 작성하여 배포하여도 로컬 도서관 레벨에서는 많은 부분의 수정과 갱신이 이루

어지고 있는 셈이다. 한편, 최근에는 전자저널과 전자책의 영향으로 많은 이용자들이 목록이 아니라, 구글과 같이 원문의 내용 검색을 통하여

원하는 페이지로 직접 접근하게 되었다. 따라서 목록 작성에 투입되는 인력과 예산 운영의 효율성 문제가 제기되고 있다.

〈표 1〉 공급사슬상 관계자들과 도서메타데이터

관계자	메타데이터 현황	예시	데이터 형식
출판사	<ul style="list-style-type: none"> - 미국의 대형 출판사는 ONIX를 활용하고 있지만 대부분은 엑셀 파일을 비롯해 독자적인 포맷을 선택. 국내 출판사는 거의 대부분 ONIX를 작성하지 않음 - 직접 이미지, 목차, 서평 등의 부가 정보를 구축하여 홈페이지를 통해 홍보하는 경우가 많아지면서 메타데이터의 생성과 유지에 대한 필요성을 인식하게 됨 	- 출판사	ONIX (미국 50%), 엑셀, 인쇄목록, 자체포맷
유통업체	<ul style="list-style-type: none"> - CIP 데이터보다는 출판사의 데이터가 우선적으로 채택 - 도서관이 수서와 동시에 MARC을 원하고 있어, 많은 벤더들이 미의회도서관의 PCC에 참여 - 국내에서는 유통단계에서 최초 ONIX가 작성 	- Baker & Taylor, Ingram - 북센, 한국출판협동조합	ONIX, 자체포맷, MARC
서점	<ul style="list-style-type: none"> - 기술메타데이터가 매출과 직결된다고 인식 - 반스앤노블은 BISG가 정의하고 있는 기술 요소 이외에도 재고 여부, 배포처 리스트, 가격과 같은 가변적인 데이터 중시. 판매를 위한 적시 갱신 강조 	- 아마존 반스앤노블 - 교보문고 YES24	ONIX, EDI 자체포맷 MARC
메타데이터 벤더	<ul style="list-style-type: none"> - 전자 통제를 수행하여 데이터를 보장하며 챗터, 커버 이미지, 목차, 베스트셀러 사이테이션, 저자 양력, 북 수상, 독자 수준, 이용자 리뷰와 평가, 추천 정보 등을 추가 - 공급 사슬내에서 도서메타데이터의 품질 제고와 배포에 있어 중요한 중간자 역할을 수행 	- Bowker Nielsen Book, BDS, OCLC - 출판유통진흥원, KERIS	ONIX MARC 자체포맷
포털	<ul style="list-style-type: none"> - 구글은 수 백만권의 책을 디지털화하여 공개하고 있으며 저작권 비용에 관련된 Book Rights Registry 운영. 저자와 저작을 식별하고 저작간의 관계 파악에 주력. ONIX와 MARC을 모두 수용 - 네이버는 책서비스를 통해 출판사가 신간을 홍보하고 직접 판매할 수 있는 기회 지원. 국내 최대 도서 포털을 위해 각종 서점과 도서관의 데이터를 수집·통합 	- 구글 북, 네이버책	ONIX MARC
국가대표 도서관	<ul style="list-style-type: none"> - 출판사로부터 받은 정보를 기반으로 CIP를 작성하고 도서관에 제공 - 북 벤더에게도 출판 3-6개월 전에 신간에 대한 출판전 도서 목록을 제공. 출간이 완료된 이후에 제목, 크기, 페이지 등의 수정 작업을 거쳐 완벽한 목록으로 보강되고, 주제 분석, 전자 통제 등의 지적인 작업이 더해져 도서관계에 여러 가지 창구로 재 보급 	- LC - 국립중앙도서관	MARC(CIP)
로컬 도서관	<ul style="list-style-type: none"> - 30% 미만에 대해서만 원목 작업 수행 - 다운로드 받은 서지에 대해서도 로컬 관행에 의한 추가 작업 다수 수행 - 도서관 목록의 유효성과 경제성에 대한 논의 가속 	- 개별 도서관	MARC

2.8 도서메타데이터 유통 경향과 관련 쟁점

전술한 도서메타데이터 유통 경향은 아래와 같이 요약해 볼 수 있겠다. 첫째, 공급 사슬의 유통 전산화와 데이터 관리의 효율화에 대한 수요가 급증하면서, 출판사들은 표준 데이터 생성과 품질 유지에 많은 관심을 갖게 되었다. 둘째, 온라인 시장 점유율 증가는 서점을 포함한 출판계 내부에 있어, 기술 메타데이터에 대한 중요성을 더욱 강조하게 되었다. 셋째, 최근 하나의 저작이 다양한 매체로 간행되면서 도서유통업계 역시 더 많은 양의 메타데이터를 처리하게 되었으며, 유관 저작의 효과적 식별과 저작권 관리 효율화의 문제에도 직면하게 되었다. 네 번째, 한편, 도서관계는 아직도 기술 목록에 많은 시간과 노력을 투입하고 있어, 경제적 효율성을 제고하기 위한 대안의 모색이 절실히 요구되고 있다.

도서메타데이터는 출판사, 도서유통업체, 서점, 도서관 등 공급사슬상에서 각기 필요한 형식과 내용으로 활용되고 있어, 생성과 품질 유지에 중복 투입되는 시간과 비용에 대한 문제가 쟁점이 되고 있다. 이와 관련하여 Hachette Book Group의 관계자는 출판사가 도서에 대한 가장 정확한 정보를 가지고 있다고 강조하고 도서유통업체, 서점, 도서관 등은 최초 정보원(Upstream Metadata)인 출판사의 데이터를 효율적으로 활용하는 방안을 모색할 필요가 있다고 논평하였다. 같은 맥락에서 NetLibrary의 Suzanne Kemperman(OCLC 2009a)은 아마존이든 도서관이든 상관없이 이용자들은 동일한 검색 경험을 원하므로, 발견이라는 이슈는 양 커뮤니티에 있어 중요한 공통 컨셉이라

고 말하였다. 따라서 출판계와 도서관계는 서지 데이터를 연계하여 공동 활용해야 할 것이라고 지적하였다. 도서관계도 이제 원목 작업은 도서관계 전체에서 그만두어야 할 시기가 되었다고 논평하고, 최초 정보원(Upstream Metadata) 데이터를 활용하여 목록의 경제성을 추구해야 한다고 논의하였다(OCLC 2009a).

한편, 최근 하나의 콘텐츠가 전자책, 영화 등 다양한 방식으로 사용되면서, 콘텐츠에 대한 저작권 관리 문제가 대두되었다. 따라서 동일 저작의 다양한 구현형을 식별하고, 저자를 정확하게 식별하는 문제에 대한 중요성도 부각되었다. 그 밖에 시리즈와 그 구성물을 식별하는 문제, ISBN 부여 단위를 결정하는 문제 등이 도서 산업과 도서관 분야에서 쟁점이 되고 있다. 출판계와 도서관계는 각각 메타데이터 형식으로 MARC과 ONIX를 채택할 뿐 아니라, 저작과 저작의 식별을 위해서도 각기 다른 표준을 채택하고 있어, 최근 양측의 전문성을 바탕으로 한 공동 대응 노력이 필요하다고 논의되어진다.

3. 출판계와 도서관계 메타데이터 유사 표준과 상호 운용성 분석

출판계와 도서관계는 비즈니스 파트너와 유관 기관간에 데이터가 원활히 흘러갈 수 있도록 각기 표준을 정의하고 있는데, 언급했다시피, 양측은 유사하면서도 상이한 형식을 채택하고 있다. 본 장에서는 몇 가지의 체계에 대해 그 유사성과 차이점, 그리고 상호 운용성을 살펴해보도록 한다.

3.1 ONIX / MARC

ONIX는 XML 메시지 구조를 사용하는 코드 값으로 구성되어 있는 스키마로 도서산업계에서 데이터를 교환하기 위한 표준으로 사용된다. 2001년 1월에 도서 산업의 상품정보를 전자적 형태로 표현하고 전자상거래를 지원하기 위하여 버전 1이 발표되었다. 온라인 거래가 가능한 모든 콘텐츠에 적용 가능한 ONIX에는 목록 정보, 저자정보, 도서정보, 출판사정보, 유통정보 등을 포함하고 있다. ONIX는 출판사로부터 도매상과 메타데이터 벤더에게 다양한 경로로 전달되며, 출판 후 갱신된 리뷰, 가격, 상태 정보는 ONIX 피드로 파트너들에게 재배포 된다. 한편, 최근 발표된 ONIX 3.0은 디지털 콘텐츠를 더욱 효과적으로 다루고 동일 저작의 다양한 구현형이 구조적으로 연계되도록 지원하며, 연속간행물을 좀 더 효율적으로 다룰 수 있도록 개선되었다고 한다.⁴⁾

한편, 출판계에 대응되는 메타데이터 형식상의 표준인 MARC은 출판정보를 기반으로 CIP 데이터로 먼저 탄생된다. 출판사가 교정쇄 또는 전자 버전으로 표제지, 판권기, 색인 등을 국가대표도서관에 보내면, 국가대표도서관은 CIP를 제작하여 출판될 저작에 포함시켜 되돌려 보낸다. 미국에서는 10개 정도의 큰 출판사가 55,000건 정도를 ONIX 형태로 보내고 있지만, CIP 생성에 ONIX가 직접 사용되지 않으며, 국가대표도서관이 작성한 통제 heading과 분류 번호도 출판사의 ONIX 스트림에 의해 다시 환류되지 않는다.

여하튼 두 표준은 기본적으로 구조상 그리고

어의상의 큰 차이를 가지고 있다. MARC은 유통과 판매에 관련된 상세 사항을 취급하지 않으며, ONIX는 수용자에 의해 데이터가 어떻게 사용되고 갱신되며 관리되는지에 대한 정보를 다루지 않는다. 더구나 MARC에서와 같은 접근점에 대한 개념이 부재하다. 이러한 차이로 인해 사실상 두 표준의 완벽한 매핑은 불가능하며, 매핑 가운데 데이터 손실 발생도 불가피하다 (Godby 2010). 그러나 ONIX와 MARC은 LC, OCLC에 의해 서지 정보의 획득을 위한 매핑이 시도되고 있으며, 국내에서도 한국출판유통진흥원이 MARC 생성을 위해 ONIX를 활용하고 있다(한국출판유통진흥원 2006).

3.2 BISAC / LCSH, DDC

BISAC(Book Industry Standards and Communications)은 출판계의 공급사슬상에서 범용적으로 활용되는 주제 분류 표목으로 검색 시스템을 통해 주제 분야를 검색하거나 도서를 배열하기 위하여 활용된다. 9개의 알파벳 문자에 의해 표현되며, 컴퓨터, 소설, 역사 등 52개의 주요 섹션으로 구성 된다.⁵⁾ 가령 아프리카 역사 일반은 HIS001000으로 표현되며, HISTORY/AFRICA/GENERAL로 해석될 수 있다. LCSH가 300,000개의 용어로 구성되어 있는 반면, BISAC은 3,000개에 불과한데, BISAC이 좀 더 대중적이며 이용자 친화적이라는 평가를 받고 있다. 따라서 공공도서관에서도 DDC를 대신하여 BISAC을 채택하고 있는 곳이 늘고 있다고 한다. Maricopa도서관은 포커스 그룹 인터뷰를 통해 DDC 분류가

4) EDItEUR. <<http://www.editeur.org/>>.

5) BISAC. <<http://www.bisg.org/>>.

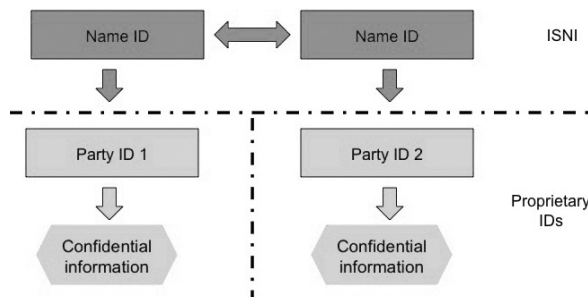
이용자들에게 친숙하지 않아, 서점과 같이 더 쉬운 분류 체계를 도입할 필요가 있음을 입증하였고, 그에 따라 BISAC을 채택하게 되었다고 말하고 있다(Norman Oder 2007). 최근 OCLC에서는 BISAC과 DDC의 매핑을 추진한 바 있을 정도로 도서관계의 BISAC 적용 가능성에 대해서도 논의가 이루어지고 있다(Mitchell 2010).

3.3 ISNI / NACO, VIAF

The International Standard Name Identifier (ISNI)⁶⁾는 ISO Standard(ISO 27729)로 미디어 콘텐츠 산업의 출판, 관리, 콘텐츠 배포와 공급사슬망에서 이름 식별을 위하여 사용된다. ISNI는 혼동될 수 있는 이름의 모호성을 분명하게 하기 위한 도구를 제공하며, 이 정보가 유관 시스템들 사이에 연계되기 위하여 활용된다. 개인에 대한 망라적 정보를 직접적으로 제공하기 위하여 개발된 시스템이 아니기 때문에, 다양한 공급사슬상의 파트너들이 개인 정보를 유출하지 않으면서도 정보를 교환할 수 있는 구조를 가진다. 개인을 구분하기 위한 최소한의 메타데이터

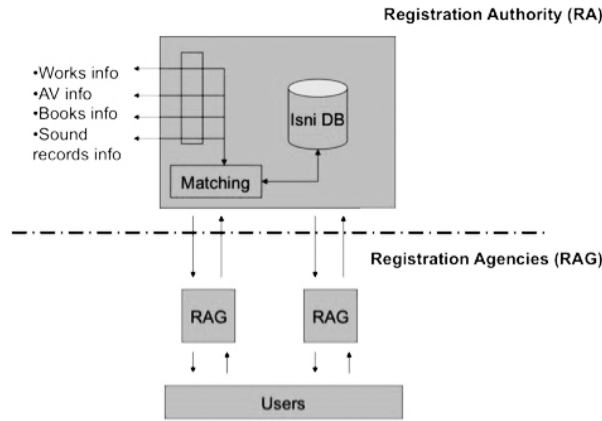
로 구성되며, 기타 정보는 제한 접속이 보장된 데이터베이스내에서 관리된다. ISNI 운영 체계는 참조 데이터베이스를 생성 관리하는 Registration Authority(RA)와 이용자들에게 ISNI 서비스를 제공하는 Registration Agencies(RAG)로 구분된다. RAG는 CISAC(International Confederation of Societies of Authors and Composers), IFRRO(International Federation of Reproduction Rights Organisations), IPDA(International Performers' Database Association), Bowker 이외에도 OCLC(Online Computer Library Center) 등이 맡고 있다.

한편, 도서관계는 전거 통제와 이름 식별 체계 유지를 위하여 NACO(The name authority program component of the PCC)와 VIAF 등을 운영하고 있다. VIAF(The Virtual International Authority File)는 미의회도서관, 독일국립도서관, 프랑스국립도서관의 협동 프로젝트로 하나의 단일 이름 전거에 대하여 세 개 기관의 전거 파일을 가상으로 조합하게 된다. VIAF는 천만개 이상의 인명과 그 이형으로 구성된 전거 레코드를 가지고 있으며, NACO의 회원이 로컬



〈그림 1〉 ISNI 운영 구조
(출처: www.isni.org)

6) ISNI. <www.isni.org>.



〈그림 2〉 RA, RAG의 역할
(출처 www.isni.or)

도서관의 인명 레코드를 추가하게 되면, 자동적으로 VIAF가 갱신되는 구조로 운영된다.

도서관계에서는 1984년에 Standard Authority Data Number(ISADN)를 개발하였으나, 전거레코드의 기능 요건으로 FRAR(Functional Requirements for Authority Records)를 정의할 뿐, 더 이상 인명식별체계는 업데이트 하지 않고 있다. 대신 최근 MARBI(MARC Advisory Committee 2010)는 출판계에서 사용되고 있는 ISNI를 아래와 같이 서지와 전거 레코드에 추가시켜 인명식별방식을 보완하고자 하여 주목된다.

100 1# \$aRendell, Ruth,\$d1930-\$0ISNI
8462 8328 5653 6435

3.4 ISTC / FRBR

한편, The International Standard Text Code (ISO 21047)⁷⁾는 출판사, 도서관매업자와 저작

권 관리 시스템에서 서로 다른 구현형을 가지고 있는 동일 원 저작을 식별하기 위하여 고안된 개념이다. 표준은 2009년 5월에 공식적으로 출판되었으며 International ISTC Agency에서 구현이 시작되었다. ISTC의 기본적인 선택스는 16개의 숫자와 문자로 구성되고 있으며, Registration element, Year element, Work element, Check digit이 포함된다. ISTC registration agency는 저작자와 저작권 대표자의 요구에 의해 고유번호를 부여하게 되는데, 저작권이 만료된 저작물도 공정 이용을 위해 국가도서관이 고유번호 부여를 요구하기도 한다.

한편, 도서관계의 FRBR은 서지적 세계를 개체와 관계에 의해 재해석하는 새로운 개념 모델로 저작, 표현형, 구현형, 아이템의 계층적 구조를 취하고 있다. FRBR은 ICP, RDA 등 차세대 목록의 기본 개념이 되고 있을 정도로 도서관계에서 매우 중요한 개념으로 이해되고 있다. 하나의 저작이 다양한 표현형과 구현형으로 표현되

7) ISTC. 〈<http://www.istc-international.org/>〉.

거나 제작될 수 있다는 개념을 기반으로 하고 있어, 다양한 유관 저작물을 동일 저작 레벨로 그룹핑하는 것이 가장 중요한 과제가 된다. OCLC는 서지 데이터베이스에서 동일 저작에서 파생된 유관 저작을 클러스터하는 워크셋 알고리즘을 개발하였으며, FRBR 워크셋 정보를 웹상에서 활용할 수 있는 xISBN 서비스를 운영하고 있다. 한편, ISTC를 FRBR의 저작 레벨 또는 표현형 레벨에 매핑하는 문제에 대한 논의는 아직 혼란스러운데, ISTC가 텍스트 기반 저작의 식별체계인 만큼, MARBI(MARC Advisory Committee 2010)는 최근 소설류에 한해 MARC 21 포맷에 아래와 같이 식별번호를 포함시키고자 한다.

024 7# \$aISTC 0A3 2009 012445C9 B\$2istc
[ISTC for the work "Winter in Madrid"]

3.5 유사 표준의 상호 운용성

위에서 살펴본 바와 같이 양측은 유사한 표준을 채택하여 도서메타데이터를 관리하고 있다. ① 자료의 발견과 식별을 위한 메타데이터 형식으로 도서관계는 MARC을, 발견과 유통·판매를 위하여 출판계는 ONIX를 사용하고 있으며, ② 자료의 내용적 분류를 위하여 도서관계는 LCSH나 DDC를, 출판계는 BISAC이나 BIC을 사용하고 있다. 한편, ③ 도서관계는 저작의 식별을 위해 전거 통제를 실시하고 있으나, 출판계는 ISNI를 활용하고 있으며, ④ 동일 저작에서 파생된 다양한 구현형을 통제하기 위하여 출판계는 ISTC(International Standard Text Code)를, 도서관계는 FRBR이라는 개념을 활용하고 있다.

전술한 표준들은 도서의 판매와 유통, 도서관 소장과 이용자 접근이라는 상이한 목적을 위해 개발되었다. 그러므로 전자는 가격과 상태에 관한 유통과 마케팅 정보, 그리고 저작권 관리의 효율성 제고에 초점을 맞추며, 후자는 발견과 접근 그리고 유관 저작의 식별을 강조하게 된다. 따라서 양측 체계는 구조상, 어의상의 큰 차이를 가지며, 이로 인해 완벽한 호환은 불가능하다고 평가된다. 그러나 부족한 부분을 보완하기 위하여 상대측 표준이 제한적으로 활용되고 있으며, 축적된 전문성이나 노하우를 채택하여 좀 더 효율적으로 발전시킬 수 있는 방안에 대한 논의도 이루어지고 있다는 점이 주목된다.

4. 도서메타데이터의 효율적인 생성·유통 방안 고찰

앞에서는 도서메타데이터의 유통 경향을 분석하고 관련 쟁점을 도출하였다. 더불어 양측이 채택하고 있는 유사 표준에 대하여 살펴보았다. 도서메타데이터는 출판사, 도서유통업체, 서점, 도서관 등 공급사슬상에서 각기 필요한 형식과 내용으로 활용되고 있어, 생성과 품질 유지에 중복 투입되는 시간과 비용에 대한 문제가 쟁점이 되고 있다. 또한 원소스 멀티유스 경향으로 하나의 콘텐츠가 다양한 방식으로 간행되면서, 동일 저작의 다양한 구현형을 식별하거나 저작을 정확하게 식별하는 문제도 쟁점이 되고 있다. 각기 다른 표준을 운용해 온 양측은 최근 상호 운용하여 경제성을 도모하는 한편, 직면한 복잡한 문제에 공동 대처할 필요를 느끼기 시작하였다.

본 장에서는 도서메타데이터를 둘러싼 여러 가지 쟁점 가운데, 생성과 품질 유지의 효율성을 제고시키는 측면에 초점을 맞춰 양측에 적용 가능한 새로운 메카니즘을 고민해 보고자 한다.

4.1 공급사슬상의 도서메타데이터 흐름 분석

언급했다시피, 생성과 품질 유지의 효율성에 대한 문제가 논의의 쟁점이 되고 있어, 먼저, 도서메타데이터의 흐름을 공급사슬망 차원에서 조망하고, 그 특징과 문제점을 구체화할 필요가 있겠다. 본 장에서는 OCLC(2009b)의 조사를 바탕으로 ONIX 및 CIP 데이터 샘플, 출판사, 유통업체, 서점 등 각 공급사슬 관계자들 사이트를 참조하여 <그림 3>과 같이 메타데이터의 흐름을 재구성해 보았다. ① 출판전 단계, ② 출판 후 유통 단계, ③ 판매 단계, ④ 도서관 소장 및 서비스 단계로 구분하여, 주요 관계자, 데이터 형식과 주요 요소, 그리고 흐름상의 특징을 아래와 같이 정리해 본다.

① 출판전단계

- 주요 관계자: 출판사, 국가대표도서관
- 데이터 형식: ONIX, CIP, 기타포맷
- 주요 생성 요소: 출판/인쇄정보, 기본서지정보, 주제분석 및 전거통제정보(CIP)
- ㉠ 출판 전에는 출판/인쇄에 관한 정보 뿐 아니라, 기본 서지정보가 생성된다. 그러나 출판 완료 후, 서명, 부서명, 사이즈 등이 변화되므로 이 단계의 서지정보는 매우 유동적이다. 일부 출판사들은 ONIX를 이용해 메타데이터를 관리하지만, 대부분의 중소출판사들은 독자적인 형식으로 데

이터를 관리하고 있다. ㉡ 한편, 출판사가 국가대표도서관에 제출한 데이터는 주제명 표목과 분류 기호 등이 추가되어 CIP 데이터로 탄생되기도 한다. 그러나 CIP 신청률이 2008년 기준으로 전체 발행물의 7-13%로(김선에 2009) 매우 저조한 국내에서는 출판된 메타데이터의 존재를 일반화하기 어렵다.

② 출판 후 유통단계

- 주요 관계자: 유통업체, 메타데이터 벤더
- 데이터 형식: ONIX, MARC, 기타포맷
- 주요 생성 요소: 상태, 유통 정보, 독자수준, 복수상, 이미지, 목차 등 보완된 기술 정보
- ㉢ 책이 출판되면 페이지, 크기와 같은 형태사항과 서명, 부서명 같은 서지 데이터가 고정된다. ㉣ 도매상과 유통업체는 도서 입수 후, ONIX를 검증하고 가격과 상태 같은 유통 정보를 업데이트한다. 또한 도서관 납품을 위하여 MARC 데이터를 구축하기도 한다. ㉤ 메타데이터 벤더는 전거 통제, 주제 분석, 챗터 정보, 이미지, 목차, 저자 양력, 복수상 정보 등을 추가해 데이터의 품질을 보강하여, 다양한 형식으로 서점, 도서관 등에 배포한다. 국내에서는 한국출판유통진흥원이 유통 단계에서 최초 ONIX 메타데이터를 생성하여 신간 출판물이라는 이름으로 관련 업체 및 도서관에 배포하고 있다.

③ 판매 단계

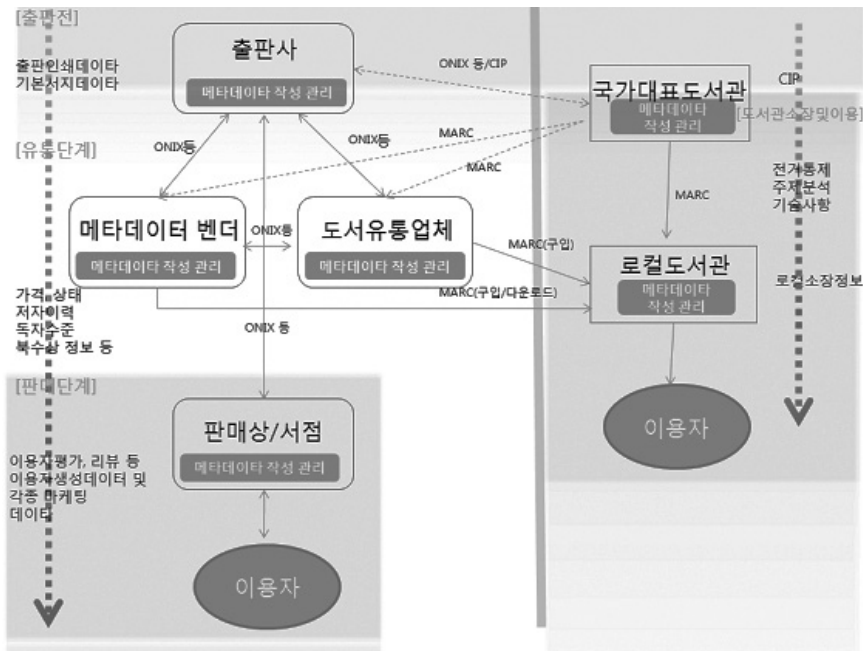
- 주요 관계자: 서점, 이용자

- 데이터 형식: ONIX, MARC, EDI, 기타 포맷
- 주요 생성 요소: 배포처, 가격 정보, 리뷰, 평점 등 이용자 생성 정보
- ㉠ 서점은 도매상, 출판사, 벤더 등으로부터 입수한 메타데이터에 새로운 데이터를 추가시킨다. 리뷰나 평점 같은 이용자 생성 메타데이터를 추가하여 이용자가 구매 결정시 활용할 수 있도록 보강하기도 한다. 한편, ㉡ 재고 여부, 배포처 리스트, 가격과 같은 가변적 데이터 요소가 이 단계에서 갱신된다.
- 데이터 형식: MARC
- 주요 생성 요소: 기술정보, 접근점 정보, 로컬 소장 정보
- ㉢ 도서관은 CIP 데이터 또는 도매상이나 벤더가 제공하는 신간출판정보를 수서 결정 과정에서 활용한다. ㉣ 수서가 완료되면 데이터는 정리파트로 인계되며, 소장정보를 포함한 로컬 목록 작업이 완성된다. ㉤ 도매상이 납품하거나 각종 서지유틸리티에서 제공받은 MARC이 활용되지만, ㉥ 로컬의 고유 정보 입력과 수정, 보완 작업에 많은 시간과 노력이 투입되기도 한다.

④ 도서관 소장 및 서비스 단계

- 주요 관계자: 서지유틸리티기관, 도서관, 이용자

〈그림 3〉과 같이 공급사슬망 차원에서 조망한 도서메타데이터는 출판전 단계에서 이용자에게로 이어지는 생명 주기에 의해 진화되어 가



〈그림 3〉 공급사슬상의 메타데이터 흐름

는 특징을 보인다. 출판전 단계에서는 기본서지 정보와 출판/인쇄에 관련된 각종 정보가 생성되며, 유통단계에서는 다양한 기술정보가 보장될 수 있다. 이용자와 직접 만나게 되는 판매단계에서는 구매자들의 선택을 돕거나 마케팅 의사 결정에 활용 가능한 다양한 정보가 생성될 수 있다. 한편, 국가대표도서관의 CIP, 유통업체의 MARC 납품 등을 통해 출판계와 도서관계에 상호 유입되면서 불필요한 재구축도 감소시킬 수 있어, 이러한 흐름상의 특징을 체계화하면, 도서메타데이터를 효율적으로 생성하여 유통시킬 수 있다.

그러나 이러한 특징에도 불구하고 공급사슬 상의 각 관계자들은 도서메타데이터를 자신들의 목적에 따라 각기 다른 형식과 내용으로 구축하면서 단편적으로 활용하고 있다. 다시 말해 생명 주기 동안 보장된 데이터 요소들을 각 관계자들 사이에 효과적으로 공유하지 않고 중복 구축하거나 재 작성하면서 불필요한 노력을 발생시킨다. 유통업체와 벤더, 온라인 서점 등 출판계 관계자들은 각기 마케팅과 관리 정보의 최신성 유지를 위하여 많은 인력을 투입하고 있으며, 도서관계 역시 각종 서지유틸리티로부터 무료로 활용할 수 있는 데이터를 도매상이나 벤더들로부터 유상으로 구매하거나 별도의 목록 인력을 투입하여 새롭게 작성하면서 불필요한 예산과 노력을 투입하고 있다. 또한 양측이 많은 노력을 투입하여 보장한 데이터 요소들도 상호 간 충분히 공유되지 않고 있다. 도서관계가 주제 분석, 전거 통제 등 지적인 노력으로 보장한 데이터 요소들이 출판계로 적극 유입되어 활용되지 않으며, 출판/인쇄에 관한 상세정보, 이용자 리뷰나 평가, 저자의 이력, 복수상, 이용자 수

준 등 출판계의 공급사슬상에서 생성되거나 보장된 의미 있는 데이터 요소들도 원활히 도서관계에 유입되어 활용되지 못한다.

4.2 효율적 생성·유통 방안 제시

진술한 바와 같이, 도서메타데이터는 동일한 정보원을 가지지만 양측의 목적에 따라 각기 다른 형식으로 활용되고 있으며, 각계의 노력으로 풍부해진 데이터 요소들이 상호 공유되지 못하고 있다. 도서관계와 출판계는 상호 운용 가능성과 필요성을 모두 인지하고 있으며, MARC과 ONIX의 부족한 부분을 적시에 교류하길 원한다. 더 근본적으로는 출판전 단계에서부터 고품질의 서지데이터가 생성되어 출판계와 도서관계에 공유되고, 생명주기에 따라 진화되길 기대한다. 그러한 맥락에서 본 장은 이를 위한 몇 가지 방안에 대해 고민해 보고자 한다.

4.2.1 ONIX와 KORMARC의 상호 보장 메커니즘

출판계에서 유통되는 메타데이터와 도서관계의 목록데이터는 서로 부족한 부분을 보장할 수 있어야 할 것이다. 출판계에서 유통되는 메타데이터에는 출판/인쇄에 관한 상세정보 이외에도 생명 주기에 따라 도서관계에서 확보하기 어려운 다양한 정보가 생성될 수 있다. 반면, 도서관계의 목록 데이터에도 주제 분석과 전거 통제를 통해 출판계에서 확보하기 어려운 다양한 정보가 생성될 수 있다. 또한 소급 서지데이터를 기반으로 한 다양한 마이닝 기술을 통해 서지 요소를 더욱 풍부하게 보장시킬 수 있다. 가령 FRBR 워크셋을 통해 신간도서의 원저작과 관련된 서지레코드를 소급 DB에서 추출한다면,

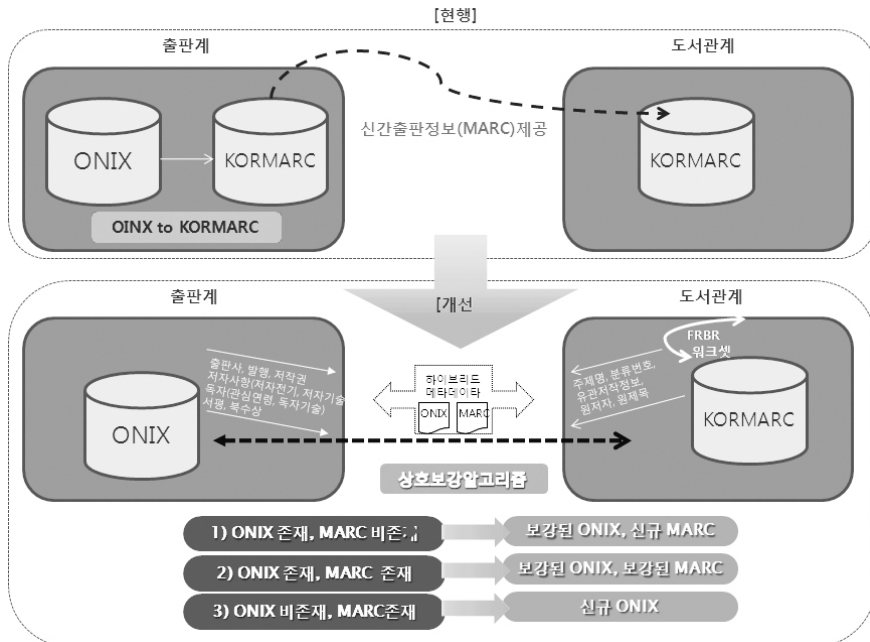
원저작의 서명, 저자, 주제에 관련된 상세 정보, 그리고 동일 원저작에서 파생된 다양한 유관 저작 정보를 기계적으로 확보하여 양측이 모두 유용하게 활용할 수 있을 것이다.

우리나라의 경우, 유통단계에서 작성된 ONIX의 서지 요소가 KORMARC으로 변환되어 서지유틸리티 기관을 통해 도서관계에 보급되고 있다. 여기에 사용되는 변환 알고리즘을 수정하여, 오닉스와 KORMARC이 상호 보장되는 새로운 메카니즘을 도입할 수 있을 것이다. <그림 4>는 본 논문에서 제안하는 ONIX와 KORMARC의 상호 보장 메카니즘 개념도이다. 아래에서 조금 더 구체적으로 설명해 본다.

첫째, 그림 상단에서 보이는 바와 같이, 현재는 유통 단계에서 작성된 ONIX의 서지 데이터 요소가 ONIX to KORMARC 알고리즘을 통해

MARC으로 변환되어 서지유틸리티 기관을 비롯한 도서관계에 신간출판정보로 보급되고 있다.

둘째, ONIX와 KORMARC의 단순 매핑으로는 도서관계에서 지금까지 다루지 않았지만 최근 그 중요성이 강조된 다양한 요소들이 유실될 수 있다. 더구나 기술의 수준 차이 등 구조적 문제로 인하여 매핑이 쉽지 않은 경우도 존재한다. 가령 ONIX의 PR8에 정의된 “공헌자 이름/형태”에는 직위 및 소속기관, 저자의 이력, 저자의 배경 등 다양한 기술사항이 하위요소로 정의되어 있다. 그러나 KORMARC과의 기술 수준 차이는 변환 과정에서 이러한 요소를 원활히 수용하기 어렵게 한다. PR14에서 다루고 있는 “독자”도 마찬가지이다. 역시 하위요소로 관심연령대, 독자기술(예, 컴퓨터 초보자), 독자범위 및 독자범위의 정확도 등이 정의되어 있으나 같은



<그림 4> ONIX와 KORMARC 상호 보장 메카니즘 개념도

이유로 변환이 용이하지 않다. ONIX의 변환 알고리즘은 이러한 요소를 수용하기 위하여 보완될 필요가 있으며, 추가 요소의 수용을 위해 경우에 따라 KORMARC의 구조적 수정도 필요해 보인다.

셋째, 상호 보강을 위해 KORMARC과 ONIX를 매핑하면, 다음과 같은 세 가지 경우가 나타날 수 있다. ONIX와 일치하는 MARC이 존재하지 않는 경우, ONIX와 일치하는 MARC이 존재하는 경우, 그리고 KORMARC은 존재하나 ONIX가 아직 생성되지 않은 경우이다. ① 먼저, ONIX와 일치하는 MARC이 존재하지 않는 경우는, 신간에 대한 ONIX가 생성되었지만, 아직 KORMARC은 생성되지 않은 상태이다. 이러한 경우, ONIX의 서지 요소(ISBN, 총서, 서명저자사항, 저자, 판사항, 발행정보, 독자사항 등)로 KORMARC을 자동 생성시킬 수 있으며, 이 데이터는 도서관계에서 신간출판정보로 활용될 수 있다. ② 한편, ONIX와 KORMARC이 모두 존재하는 경우는 상호 보강 알고리즘에 의해 ONIX 고유의 요소(저자이력, 저자전기, 관심연령, 서평 등)로 KORMARC을 보강시키며, 동시에 KORMARC 고유의 요소(주제명, 분류번호, 유관저작정보 등)로 ONIX를 보강시킬 수 있다. 상호 보강 알고리즘은 매핑 대상이 되는 상대 데이터에 부족한 요소를 발견하여 보완하거나, 우수한 쪽의 데이터 요소로 대체시킬 수 있다. ③ 마지막으로 ONIX보다 KORMARC이 먼저 생성되어버린 경우도 있다. 우리나라에서는 유통 단계에서 최초의 ONIX가 구축되고 있으므로, 간혹 출판사의 직관을 통해 신간을 구입한 도서관이 ONIX보다 먼저 KORMARC을 작성하게 되는 경우가 있을 수 있다. 이러한

경우, 먼저 구축된 KORMARC의 서지 요소(ISBN, 서명저자사항, 주제분석, 전거통제)로 ONIX를 자동 생성시킬 수 있으며, 이 데이터는 완벽한 ONIX를 구축하는데 기초 데이터로 활용될 수 있을 것이다.

넷째, 최근에는 하나의 콘텐츠가 도서, 만화, 영화 등 다양한 매체로 재생산되고 있다. 또한 새롭게 출판된 도서일지라도 기존 저작의 개작, 증보, 번역물 중 하나일 수 있다. OCLC는 이러한 아이디어를 바탕으로 FRBR 알고리즘을 NextGen Pilot에 활용하고 있다. 같은 방식으로 국내 최대 서지데이터베이스인 UNICAT에서 FRBR 클러스터를 생성하면, 대상이 되는 도서의 다양한 유관 저작에 공통적으로 포함되어 있는 서지 정보(주제명, 분류번호, 원제목, 원저자 등)를 기계적으로 추출하여, 신간 서지 데이터의 품질을 제고시킬 수 있을 것이다. 따라서 ONIX와 일치하는 KORMARC이 존재하지 않는 ①번과 같은 경우에도, FRBR 클러스터 정보로 ONIX의 서지 요소를 보강시킬 수 있게 된다.

4.2.2 CIP 활성화를 통한 고품질 서지데이터 생성·공유

한편, 전술한 ONIX와 MARC의 상호 보강 메커니즘이 CIP가 활성화되어 있지 않은 국내 환경을 위한 보완적 대안이라면, CIP 활성화는 출판계와 도서관계가 출판전 단계에서부터 고품질의 서지데이터를 생성하여 공유할 수 있는 좀 더 근원적 방안이라고 말할 수 있겠다. CIP는 출판사, 유통 업체, 서점 등 공급사슬상의 관계자들이 중복하여 서지데이터를 작성하는 노력을 최소화시킬 수 있으며, 동시에 카피 목록 작성을

통해 도서관 업무의 경제성을 제고시킬 수 있다. 따라서 ONIX와 MARC의 상호 보강 메커니즘과 더불어 출판사들의 CIP 신청을 독려할 수 있는 방안에 대한 고민도 필요하겠다. CIP를 제공한 출판사가 적시에 출판 도서를 홍보하고 판매 경로를 확보할 수 있다면 자연스럽게 참여율을 제고할 수 있을 것이다(김선에 2009). 국내 최대 도서 포털이면서, 판매 시스템을 연계하고 있는 “네이버 책”의 출판사 신간 등록 서비스와 제휴한다면, CIP 참여율 제고에 많은 도움을 받을 수 있을 것으로 기대된다.

4.2.3 생명 주기에 따라 진화되는 메커니즘
앞서 언급한 바와 같이, 도서메타데이터는 공급사슬상에서 시간이 지남에 따라 도서의 상태, 재고 여부, 가격 등의 비즈니스 관련 정보 뿐 아니라, 이용자 접점에서 평점, 리뷰, 태그 등 새로운 정보가 추가될 수 있다. 이러한 정보는 출판계 측면에서는 비즈니스 의사 결정 데이터로 활용될 수 있으며, 도서관계에서는 이용자의 판단과 선택을 돕는 정보로 활용될 수 있을 것이다. 데이터가 일관적으로 유지되어서 생명 주기에 따라 진화되어가기 위해서는 먼저 출판계 내부의 표준 채택이 선행되어야 하며, 동시에 데이터 흐름에 따라 추가되고 갱신된 부분이 공급사슬에 역으로 환류되는 메커니즘의 고안도 필요할 것이다. ONIX 피드를 통해 진화된 데이터가 정보원을 향해 자동 피드되고 종합목록과 같은 서지시스템으로도 전달되어 연계적으로 갱신될 뿐 아니라, 그것을 다운로드 받은 로컬도서관에도 피드되는 방안을 고민해 볼 필요가 있을 것이다.

5. 결론

5.1 요약

본 연구에서는 첫째, 공급사슬상의 관계자들 입장에서 메타데이터의 유통 경향과 관련 쟁점을 분석하였다. 하나의 저작이 다양한 매체와 형식으로 간행되면서 출판계는 자세한 기술 정보 제시를 요구받게 되었고, 온라인 시장 점유율 증가로 인해 메타데이터는 출판 마케팅의 주요 요소가 되었다. 도서관계는 목록 작성의 경제적 효용성을 추구하면서, 최초 정보원에 근접한 데이터 소스를 통해 목록 작성 업무를 간소화할 수 있는 모델을 모색하게 되었다. 한편, 원소스 멀티유스 경향으로 하나의 콘텐츠가 다양한 방식으로 간행되면서, 동일 저작의 다양한 구현형을 식별하거나 저자를 정확하게 식별하는 문제가 양측에게 모두 쟁점이 되고 있다. 이러한 배경으로 최근 상호 운용하여 경제성을 도모하는 한편, 직면한 복잡한 문제에 대한 공동 대처 필요가 제기되었다.

둘째, 출판계와 도서관계에서 적용되고 있는 유사 데이터 표준과 상호 운용성을 살펴보았다. 출판계는 판매와 유통을 위해 ONIX, BISAC/BIC, ISNI, ISTC 등을 사용하고 있으며, 도서관계는 도서관 소장과 이용자 접근을 위해 MARC, LCSH/DDC, 전거통제, FRBR 등을 사용하고 있다. 양측 체계는 구조상, 어의상 큰 차이를 가지고 있어 완벽한 호환은 불가능하지만, 부족한 부분을 상호 보완한다면 좀 더 효율적으로 운용될 수 있다고 평가된다. 한편, 이용자의 선택과 구매 결정을 지원하기 위한 양질의 기술 메타데이터 생성, 그리고 유관 저작과 저자의 식별 문제는

양측 표준이 직면한 공통 과제로 그간의 전문성을 바탕으로 한 공동 대응이 절실히 요구된다.

셋째, 출판계와 도서관계에서 유통되고 있는 도서메타데이터 흐름상의 특징과 문제점을 파악해 보았다. 현행 도서메타데이터는 출판전 단계에서부터 이용자에게로 이어지기까지 출판사, 유통업체, 서점, 도서관 등에서 각기 필요한 형식과 내용으로 활용되고 있다. 생명 주기 동안 상호간 복잡하게 교류되면서 진화되기도 하지만, 개별 관계자들에 의해 중복 구축되거나 재작성되면서 불필요한 노력이 발생되기도 한다. 또한 생명 주기 동안 보강된 데이터 요소들이 각 관계자들 사이에 효과적으로 공유되지 못하고 있다.

넷째, 위의 분석을 바탕으로 도서메타데이터 생명 주기에 있어 보다 진화된 메커니즘의 도입 가능성을 모색해 보았다. 아래 제언에서 기술하고 있는 바와 같이, 본 연구는 양측이 중복 노력을 최소화하고 상호 긍정적으로 메타데이터를 운용할 수 있는 3가지 방안을 제시하였다.

5.2 제언

도서관계와 출판계가 새로운 메커니즘을 통해, MARC과 ONIX의 부족한 부분을 적시에

교류할 수 있다면, 도서관계는 신속하고 정확한 기술정보를 확보하고 이를 통해 목록의 경제성을 제고할 수 있다. 또한 출판계도 주제 분석과 전거 통제 등으로 풍부해진 메타데이터를 적시에 확보하고 판매자 시스템으로 양질의 데이터가 신속하게 전달될 수 있을 것이다. 본 연구에서는 도서메타데이터의 새로운 생성·유통 체계를 위한 몇 가지 방안을 제안하였다. 첫 번째, ONIX와 MARC 데이터가 상호 매핑되어 양측의 부족한 부분을 보강할 수 있는 알고리즘이 도입되어야 할 것이다. MARC에 부족한 출판/인쇄에 관한 상세정보와, 저자 전기, 서평 등의 마케팅 정보, ONIX에 부족한 주제 분석, 전거 통제, 연관 저록 정보 등을 상호 보강하여 양측이 적시에 활용할 수 있는 새로운 체계가 고안되어야 할 것이다. 두 번째, CIP 제도를 활성화시켜 출판전 도서 정보가 도서관계에 원활히 유입되고, 도서관계의 지적 노력을 통해 보강된 서지 정보가 출판계에서도 원활히 활용되어야 할 것이다. 세 번째, 도서메타데이터 생명 주기에 따라 추가되고 갱신되는 다양한 정보를 ONIX 피드를 통해 공급 시술에 역으로 환류시키고, 이미 다운로드 받아 활용하고 있는 도서관계에도 자동 피드시킬 수 있는 메커니즘이 고안되어야 할 것이다.

참 고 문 헌

- [1] 김선애. 2009. 우리나라 CIP 프로그램에 관한 고찰. 『제46회 전국도서관대회 세미나 3: 공공도서관 목록, 빠르고 정확하게』. [online]. [cited 2009. 12. 1]. <<http://conference.kla.kr/>>.
- [2] 장지숙. 2009. 공공도서관 목록 품질향상을 위한 방안: CIP를 중심으로. 『제46회 전국도서관대회

- 세미나 3: 공공도서관목록, 빠르고 정확하게. [online]. [cited 2009. 12. 1].
 <<http://conference.kla.kr>>.
- [3] 한국출판유통진흥원. 2006. KORMARC ↔ ONIX 변환표. [online]. [cited 2009. 10. 1].
 <http://www.booktrade.or.kr/new_notice/att_file/KORMARC_ONIX_CHANGE_NEW.pdf>.
- [4] 國立國會図書館, 2007, デュース十進分類法を採用しない図書館, 議論的的に, カレントアウェアネス-E No.111. [online]. [cited 2009. 10. 1].
 <<http://www.current.ndl.go.jp/e677>>.
- [5] A Symposium for Publishers and Librarians. [online]. [cited].
 <<http://www.oclc.org/publisher-symposium/>>.
- [6] BISAC. [online]. [cited]. <<http://www.bisg.org/>>.
- [7] EDItEUR. [online]. [cited]. <<http://www.editeur.org/>>.
- [8] Godby, & Carol Jean. 2010. "Mapping ONIX to MARC." [online]. [cited 2010. 4. 10].
 <<http://www.oclc.org/research/publications/library/2010/2010-14.pdf>>.
- [9] ISNI. [online]. [cited]. <www.isni.org/>.
- [10] ISTC. [online]. [cited]. <<http://www.istc-international.org/>>.
- [11] Mitchell, Joan S. 2010. "BISAC-DDC Mappings. ALA Midwinter Meeting Boston January 16, 2010." [online]. [cited 2010. 3. 1]. <www.oclc.org/dewey/news/conferences/bisac.ppt>.
- [12] MARC Advisory Committee. 2010. "MARC DISCUSSION PAPER NO. 2010-DP03." [online]. [cited 2010. 3. 10]. <<http://www.loc.gov/marc/marbi/2010/2010-dp03.html>>.
- [13] NISO, & OCLC. 2009. "Streamlining Book Metadata Workflow." [online]. [cited 2009. 12. 1].
 <http://www.niso.org/publications/white_papers/StreamlineBookMetadataWorkflowWhitePaper.pdf>.
- [14] Oder, & Norman. 2007. "Behind the Maricopa County Library District's Dewey-less Plan, Library Journal, 5/31/2007." [online]. [cited 2009. 10. 1].
 <<http://www.libraryjournal.com/article/CA6448055.html>>.
- [15] OCLC. 2009a. "Report on OCLC's Symposium for Publishers and Libraries." [online]. [cited 2010. 3. 3]. <<http://www.oclc.org/publisher-symposium/summary/default.htm>>.
- [16] OCLC. 2009b. "From ONIX to MARC and Back Again: New Frontiers in Metadata Creation at OCLC, ALA Midwinter January 25, 2009." [online]. [cited 2010. 3. 13].
 <http://www.oclc.org/us/en/multimedia/2009/files/ONIX_to_MARC.ppt>.
- [17] R2 Consulting LLC. 2009. "Study of the North American MARC Records Marketplace." [online]. [cited 2009. 12. 30].
 <http://www.loc.gov/bibliographic-future/news/MARC_Record_Marketplace_2009-10.pdf>.

- [18] Working Group on the Future of Bibliographic Control. 2008. "On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control." [online]. [cited 2009. 11. 1].
<<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>>.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Sun-Ae Kim. 2009. "Urinara CIP Programe Gwanhan Gochal." *46th KLA General Conference Semina 3: Gonggongdoseogwanmokrok, Bbareugo Jeonghwakhage*. [online]. [cited 2009. 12. 1]. <<http://conference.kla.kr/>>.
- [2] Ji-Suk Jang. 2009. "Gonggongdoseogwan mokrok Pumjilhyangsangeul Wihan Bangan: CIPreul Jungsimeuro." *46th KLA General Conference Semina 3: Gonggongdoseogwanmokrok, Bbareugo Jeonghwakhage*. [online]. [cited 2009. 12. 1]. <<http://conference.kla.kr>>.
- [3] Book Trade Promotion Center. 2006. KORMARC ↔ ONIX Byeonhwanpyo. [online]. [cited 2009. 10. 1].
<http://www.booktrade.or.kr/new_notice/att_file/KORMARC_ONIX_CHANGE_NEW.pdf>.