

# 도서관 링크데이터(Library Linked Data) 현황분석과 통합 LLD 설계 및 확산방향에 대한 고찰

## Analyzing Current State of Library Linked Data, Designing an Integrated LLD, and Thoughts on Ways to Expand LLD

오 삼 균(Sam Gyun Oh)\*

김 성 훈(Seonghun Kim)\*\*

장 원 홍(Wonhong Jang)\*\*\*

### < 목 차 >

- |                                      |                        |
|--------------------------------------|------------------------|
| I. 서 론                               | IV. 국내 LLD 적용을 위한 제안사항 |
| II. 링크데이터(Linked Data)의 개념 및<br>선행연구 | V. 한국 통합 LLD 모형        |
| III. LLD 구현기술 및 구축 장애 요소             | VI. 결 론                |

### 초 록

본 연구의 목적은 근래 W3C의 특별 작업반 형태로 운영되어 온 도서관 링크데이터(LLD : Library Link Data) 관련 국제 활동상황을 분석하고, LLD가 도서관 서비스에 미칠 혜택을 사례연구를 통해 우선 숙고한 후, 보다 의미 있는 LLD를 위한 글로벌 식별체계, 이용 가능한 어휘와 데이터셋의 연결 문제 및 RDF, SKOS, Microformats, RDFa 등의 LLD 적용 방안을 논하는 것이다. 아울러 LLD 수용 및 LLD 기반 신 정보서비스 구축의 장애 요인들을 분석하고, 그 분석의 바탕 위에 LLD 기반 신 디지털정보시스템 설계 및 LLD 확산을 위한 권고사항을 제시한다.

키워드: 한국 도서관 링크데이터, 한국 링크데이터

### ABSTRACT

This paper analyzes the activities and recommendations of W3C LLD Incubator group and investigates the benefits of Library Link Data through case studies. It also discusses the necessary global identification systems for developing useful LLD as well as the problems associated with linking available vocabularies with existing data sets. Ways to apply RDF, SKOS, Microformats, and RDFa to LLD are presented, the many challenges related to establishing LLD-based information services and the various barriers that prevent libraries from accepting LLD are also identified. Finally this paper presents a new LLD-based digital information system and provides some recommendations on how to promote and expand LLD in Korea.

Keywords: Korea Library Linked Data, Korea Linked Data

\* 성균관대학교 문헌정보학과 교수(samoh@skku.edu) (제1저자)

\*\* 성균관대학교 문헌정보학과 박사과정(godwmaw0278@gmail.com) (공동저자)

\*\*\* LG 상남도서관(jangwonhong@gmail.com) (공동저자)

• 접수일: 2011년 11월 25일 • 최초심사일: 2011년 12월 9일 • 최종심사일: 2011년 12월 28일

## I. 서론

미국 의회도서관(Library of Congress, 이하 LC)이 목록레코드의 도서관 상호 교환을 목적으로 MARC(Machine Readable Catalog)를 제정한 이래 도서관 메타데이터는 코드화에 기반을 둔 컴퓨터의 체계적 식별의 결과로 축적, 유통되며 고유하게 발전해왔다. 이로써 표준목록 데이터의 운용 및 도서관 데이터의 상호 교환과 사용은 괄목할만하게 활발해졌다. 그러나 도서관 표준목록 데이터를 시맨틱 웹의 관점에서 평가하면 이하의 몇 가지 중요한 문제점에 주목하게 된다.

첫째, 도서관 사이의 데이터 교환을 지향하는 태생적인 특징은 자체 맥락 이외의 상호운용성 면에서 도서관 표준목록 데이터를 매우 취약하게 만드는 근본 원인이다. MARC나 OAI-PMH의 주안점은 도서관 상호교환 체계이므로 도서관을 벗어난 정보 용·통합에서 상당한 난관에 봉착하는 것이다. 외부로 방출되지 않고 단일 기관 혹은 도서관들 사이로 그 사용 범위가 제한된 상태를 답습하는 데이터는 결국 광의의 사회적 자원으로 발전하지 못하고 고립된 자원으로 존재할 위험이 크다.

둘째, 특정기관을 직접 방문하기 보다는 온라인 접근을 통해서 정보자원을 이용하는, 즉 정보환경의 중심이 네트워크로 대거 이동한 추세 하에서 외부 데이터와의 연결성(connection)을 확보하지 못한 도서관 데이터는 실질적인 중요성에도 불구하고 이용 가치가 하감할 수밖에 없다.

셋째, 동일 작품이 도서, 영화, 뮤지컬 등의 다양한 형태로 나타나는 경우가 계속 증가하는 가운데 각각의 콘텐츠 유형을 구분하고 그들 사이의 관계를 형성하는 작업이 더욱 필요해졌으나 도서관 데이터는 식별성(identification)과 연결성(connectivity)의 확보에서도 취약점을 드러내고 있다.

### 1. 연구의 배경 및 필요성

1997년에 FRBR(Functional Requirements for Bibliographic Records) 모형을 제시했던 IFLA는 2008년에 들어서자 FRBR의 기능적 요건을 반영하되 RDA(Resource Description and Access)로 방향을 전환하여 AACR2 목록의 비일관성과 애매함, 중복내용을 제거하고, 규칙 내에서의 지속된 문제를 해소하고자 했다.<sup>1)</sup> 이러한 움직임은 시맨틱웹이 지향하는 URI를 통한 객체의 식별과 자원간의 관계를 명확히 정의하여 정보 검색을 향상시키고, 자원 간의 관계에 따라 검색결과를 의미 단위로 그룹화하려는 노력과 맥을 같이 한다.

웹상에서 유통 가능한 정보는 방대하지만 그 대부분의 형태는 HTML이며, 나아가 정작 중요한

1) 이미화, "MARC 데이터의 RDA 저작 및 표현형 요소 분석을 통한 한국목록규칙 및 KORMARC의 고려사항," 한국도서관·정보학회지, 제41권, 제2호(2010. 6), pp.251-272.

데이터는 개방되지 않는 형태로 DBMS에 저장되는 경우가 많다. 이처럼 웹문서가 HTML 형태를 취하고 도서관 목록데이터가 DBMS에 폐쇄적인 형태로 저장되기 때문에 외부 데이터와 단절된 현실은 정보제공자와 정보이용자 양자에게 다 만족스럽지 못한 결과를 초래한다.

Library Linked Data(이하 LLD로 표기)는 도서관 데이터의 식별과 재사용, 접근성, 그리고 세부적인 데이터의 연결성의 문제를 해결하려는 시도이다. 웹상에서 활성화한 링크데이터의 원칙을 도서관 목록데이터에 적용시켜서 독방(silo)에 갇힌 목록정보를 외부로 방출시키고 시너지 효과가 기대되는 외부 데이터와 연결시키는 노력이 현실적인 대안이기 때문이다. 따라서 국외에서는 이미 미국의회도서관이 LCSH를 SKOS로 방출해 타 기관들이 자유롭게 사용할 수 있는 길을 열었고, 스웨덴 국립도서관 또한 도서관 소장 목록정보를 링크데이터로 제공하고 있다. W3C 산하에 도서관 목록데이터의 링크데이터 변환과 그 효과적인 확산을 목적으로 하는 다양한 도서관 관계자들의 모임인 LLD 보육작업반이 가동한 것도 동일한 맥락의 움직임이다.<sup>2)</sup>

그러나 시맨틱웹 기술을 도서관에 적용하는 가장 바람직한 접근법으로 간주되고 있는 링크데이터의 국내 도입 속도는 매우 더딜 뿐만 아니라 관련연구 역시 상대적으로 미미한 형편이다. 다각도의 체계적인 연구와 실험, 적용이 국내에서 시급히 시행되고 이를 위한 기초 작업을 선행할 필요가 있는 이유라고 하겠다.

링크데이터를 이용할 경우, 예를 들어 검색어가 작가라면 RDF 도서관 Mashup, Project Gutenberg, 그리고 FOAF(Friend Of A Friend) 어휘를 활용해서 작가와 관련된 풍부한 정보를 제공하는 것이 가능해진다. 검색어가 어떤 지역명이라면 그 정확한 현재 위치와 역사, 관련 인물을 더욱 상세히 제시할 기회도 얻게 된다. 기존의 도서관 목록을 링크데이터로 방출하여 다양한 연결고리 정보를 제공하기 때문에 이용자가 여러 시스템을 사용하며 않아도 한 번에 필요한 정보를 취합할 수 있는 것이다.<sup>3)</sup>

## 2. 연구의 목적

본 연구에서는 LLD 관련 전 세계의 활동상황을 분석하고, 최근 W3C 산하 LLD 보육작업반에서 발표한 연구 보고서를 토대로 LLD가 도서관 서비스에 가져올 혜택과 수용에 있어서의 장애요인을 논의하고자 한다. 나아가 국내의 상황을 고려하여 LLD 확산을 위한 권고사항을 제시하고, LLD 기반 새로운 정보시스템 설계 모형을 제안하고자 한다.

2) W3C Library Linked Data Incubator Group, <<http://www.w3.org/2005/Incubator/ld/>>.

3) 조명대, “도서관에서의 Linked Data 활용방안에 관한 연구,” 한국문헌정보학회지, 제44권, 제1호(2010. 2), pp.181-198.

## II. 링크데이터(Linked Data)의 개념 및 선행연구

### 1. 링크데이터의 개념

Tim Berners-Lee가 제안한 시멘틱웹은 정보자원을 인간과 기계가 모두 이해할 수 있는 형태로 정의하고, 그 정보자원 간의 연결을 통해 관련 부가 정보를 용이하게 검색할 수 있도록 하는 데 그 중점을 둔다. 링크데이터는 자원과 자원을 유의미하게 연결해주는 시멘틱웹 기술을 한 단계 더 정교하게 적용하여 데이터 간의 유의미한 연결을 통해 보다 광범위한 데이터의 연결을 가능하게 한다.<sup>4)</sup>

링크데이터의 핵심은 문서(Document) 중심이 아닌, 데이터 중심의 웹(web of data) 구축이다. 다시 말하면 기존의 레거시 데이터를 링크데이터로 변환하여 인터넷상에 발행해서 공동으로 활용하자는 것이다. 따라서 링크데이터는 전혀 새로운 그 무엇을 만드는 것이 아니라 웹의 틀에 약간의 구조 변형을 일으킨다. 기존의 웹이 HTML을 기반으로 Link를 통해 문서를 연결하는 것이었다면, 링크데이터는 가공되지 않은 데이터를 공개해 세밀하게 데이터를 연결함으로써 더욱 효과적인 정보 이용과 생성을 꾀해서 웹을 풍성하게 만든다.<sup>5)</sup> 이의 실행을 위해 Tim Berners-Lee는 링크데이터가 갖추어야 할 네 가지 원칙 <표 1>을 제시하며, 이 원칙을 준수해서 링크데이터를 발행할 것을 권한 바 있다.

<표 1> 링크데이터 4원칙

1. 모든 개체에 URI 식별자 부여.
2. HTTP URI를 사용하여 개체의 이름을 찾아볼 수 있도록 지원.
3. URI를 참조할 때 개체에 대한 유용한 정보를 제공해야 함.
4. 다른 URI와의 연결을 통해 더 많은 정보를 발견할 수 있도록 지원.

링크데이터는 RDF(Resource Description Framework) 데이터 모델을 기반으로 하며, RDF/XML을 비롯하여 N-Triple 등의 구문구조로 표현될 수 있다. RDF로 구축된 다양한 링크데이터에 RDF 질의 언어인 SPARQL를 활용하여 매우 정교한 질의를 수행하게 된다.

4) Tim Berners-Lee, *Linked Data*, 2006, <<http://www.w3.org/DesignIssues/LinkedData.html>> [cited 2011. 7. 29].

5) Tim Berners-Lee, *Presentation on Linked Data In TED Conference*, 2009,

<[http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html)> [cited 2011. 8. 20].

## 2. 링크데이터 구문구조

시맨틱웹 기술을 데이터에 적용한 모범사례인 링크데이터의 구문 표현에는 RDF(resource description framework)가 사용된다. RDF란 자원에 대한 정보를 의미상으로 기술하기 위한 데이터 모델로서, 데이터를 주어, 술어, 목적어 구조로 표현함으로써 데이터 간의 관계를 기계가 이해할 수 있게 한다. RDF 표준구문을 따라서 데이터를 표현해야 의미의 혼동이 없는 데이터의 사용, 용이한 관련 자원검색과 데이터의 상호운용성 확보가 가능해진다.

RDF는 데이터 형식이 아닌 데이터 모델이기에 주어, 술어, 목적어 관계를 웹상에 표현하기 위해서는 적절한 구문 구조가 요구된다. 이 구문 구조로는 W3C가 표준화한 RDF 직렬화 포맷인 RDF/XML, 그리고 프로세스의 향상을 위해 그것을 간결화한 RDF 직렬화 포맷이 있다. 이하는 RDF 간결 구문 구조와 RDF/XML에 대한 기초적인 설명이다.

### N-Triple

RDF의 가장 단순한 형태이며 추약되지 않은 URI를 사용하여 리소스를 표현한다. 주어, 술어, 목적어 각 트리플은 <,>안에 표현되고 트리플 셋은 마침표(.)를 통해 구분되기 때문에 한 행에 모든 내용을 다 표현할 필요는 없다. N-Triple의 용례는 다음과 같다.<sup>6)</sup>

```
<http://biglynx.co.uk/people/dave-smith>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person> .

<http://biglynx.co.uk/people/dave-smith> <http://xmlns.com/foaf/0.1/name>
"Dave Smith" .
```

### Notation 3 RDF(N3)

Tim Berners-Lee에 의해 개발된 포맷으로서 URI의 추약형인 QName을 사용한 간결성이 그 특징이다. 서두에 QName을 정의하고, QName을 사용하여 주어, 술어, 목적어를 서술한 뒤, 마침표(.)로 트리플의 끝을 표현한다. 특히 테이블에서 데이터를 반입할 때 하나의 주어에 여러 술어와 목적어가 오는 경우가 많은데, 세미콜론(;)을 사용하여 하나의 주어에 여러 술어와 목적어를 표현할 수 있다. N3의 용례는 다음과 같다.

6) RDF 구문구조의 용례는 <<http://linkeddatabook.com/editions/1.0/>> 참고.

```

@prefix blxp: <http://biglynx.co.uk/people/>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

blxp:dave-smith rdf:type foaf:Person ;
foaf:name "Dave Smith" .
    
```

RDF/XML

상기한 간결 구문들을 XML 구문의 원리를 따라 RDF/XML로 기술하면 아래와 같다.

```

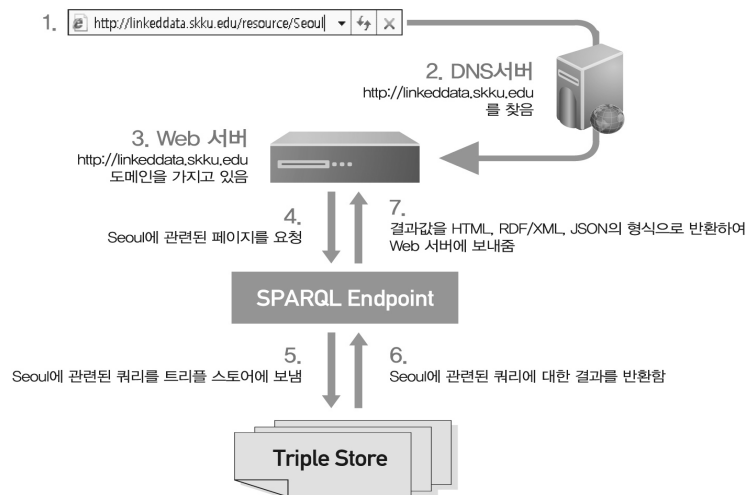
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:foaf="http://xmlns.com/foaf/0.1/">

<rdf:Description rdf:about="http://biglynx.co.uk/people/dave-smith">
<rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
<foaf:name>Dave Smith</foaf:name>
</rdf:Description>

</rdf:RDF>
    
```

3. 링크데이터의 구동 메커니즘

실제적인 링크데이터의 구동 원리를 도식화하여 살펴보면 <그림 1>과 같다.



<그림 1> 링크데이터의 구동 절차 및 단계

링크데이터를 위한 서버는 보통 RDF 저장소인 Triple 저장소와 이를 기반으로 SPARQL 쿼리를 처리할 수 있는 SPARQL 엔드포인트로 구성된다. <그림 1>과 같이 이용자가 특정 URI에 대한 정보를 웹서버로 요청하면 이 요청은 SPARQL 엔드포인트로 전달되고, Triple 저장소에서의 쿼리를 거쳐 반환된 질의 결과 값이 이용자에게 전달된다.

#### 4. 링크데이터 선행연구

국내의 경우, 시맨틱웹의 기반을 구성하는 URI나 온톨로지 등을 활용한 시스템 구축이나 연구는 상대적으로 활발한 편이나 링크데이터에 관한 연구 및 적용 사례는 매우 희소하다.

지리정보를 활용한 사례로는 서로 다른 위치와 openAPI 인터페이스를 통해 제공되던 지도와 같은 공간 정보 데이터를 링크데이터를 기반으로 통합함으로써 상이한 데이터 모델에 따른 매쉬업 문제를 해결하고 공개 데이터 이용의 활성화 가능성을 보여준 연구가 있다.<sup>7)</sup> 도로명 주소 데이터, 네이버 및 야후와 같은 포털 서비스에서 제공하는 공간정보데이터의 분석을 바탕으로 더블린 코어 등과 같은 데이터모델을 참조하여 공간 속성 데이터 모델을 정의한 후, 이것을 가지고 RDF 데이터 생성을 위한 프로그램을 설계한 이 연구는 실제 안드로이드 환경에서의 응용까지 시도한 의미 있는 측면이 있고, 링크데이터 생성 및 활용을 전반적으로 살펴보는 데 유용하다.

국가 공공정보에 대한 연구 사례로는 공공정보를 공개하여 웹에 배포할 시 기존의 openAPI나 벌크데이터 다운로드 방식 보다는 링크데이터를 기반으로 한 SPARQL 엔드포인트를 구축함으로써 정보를 보다 빠르고 효율적으로 서비스할 수 있음을 증명한 연구<sup>8)</sup>가 있다. 이 연구에서는 개별 기관에 의해 서로 다른 방식으로 진행 되는 비효율적인 국내 공공정보 관리 상황을 문제점으로 도출하고 링크데이터를 활용한 시맨틱 클라우드 구축을 제안한다. 실제로 시맨틱 데이터 클라우드 프로토타입 개발을 통해 기존의 웹 크롤링, 벌크데이터 다운로드 및 openAPI를 통해 공공정보를 배포하는 것보다 링크데이터를 구축하여 SPARQL 엔드포인트를 통해 공공정보를 배포하는 것이 훨씬 더 효율적인 방안을 증명하고 있다.

가장 최근에 진행된 연구로는 링크데이터에 대한 이해가 없는 일반인도 링크데이터를 손쉽게 생성하는 것이 가능하도록 링크데이터 생성 시스템을 개발한 사례가 있다.<sup>9)</sup> 이 연구에서는 링크데이터 RDF 데이터의 검색, 파싱 및 링크생성 등을 처리할 수 있는 모듈로 구성된 시스템을 개발하

---

7) 박재준, 김정옥, 유기윤, "Linked Data를 이용한 POI 매쉬업 서비스," 2010 한국지형공간정보학회 추계 학술대회 (2010. 9), pp.123-124.  
8) 김청림 등, "시맨틱 데이터 클라우드 : 시맨틱 기술 기반 공공정보 배포 프레임워크," 한국정보과학회 2010 한국 컴퓨터 학술발표논문집, 제37권, 제2호(2010. 11), pp.20-24.  
9) 정효숙, 김희진, 박성빈, "일반인을 위한 링크드 데이터 생성 시스템 개발 및 활용," 컴퓨터교육학회 논문지, 제14권, 제2호(2011. 3), pp.47-59.

여 교수학습 영역의 링크데이터를 생성한 후, 이렇게 생성된 링크데이터의 실제적인 활용 시나리오를 제시한다. 링크데이터 생성을 위한 기존의 복잡한 시스템과는 달리 기 구축된 RDF 문서에서 의미적 연관성을 갖는 RDF 링크 정보를 간단하게 생성하도록 해서 사전지식 없이도 데이터를 쉽게 연결하는 방안을 제시했다는 데 의미가 있다.

해외에서 링크데이터를 활용한 사례들은 매우 다양하다. 웹서비스의 경우, 영화리뷰 사이트 Revyu는 링크데이터인 DBpedia의 영화 관련정보를 매칭하여 보다 풍부한 영화 정보를 이용자들에게 제공한다.<sup>10)</sup> Talis는 영국의 대학에서 사용되는 핵심적인 강의 자료들을 학생이나 강사가 손쉽게 활용할 수 있도록 하는 링크데이터 기반 구축 강의 자료 서비스이다.<sup>11)</sup> 가장 대중적이라고 할 수 있는 영국의 방송국 BBC는 서로 다른 시스템을 통해 관리해온 수많은 라디오 방송과 TV 채널의 콘텐츠들에 대해 링크데이터를 적용함으로써 콘텐츠 간의 관계 정보 및 외부 링크데이터 소스들과의 연결정보를 생성해서 관리 비용을 크게 줄일 뿐만 아니라 부가적으로 유용한 정보를 이용자에게 제공한다.<sup>12)</sup> 이 외에도 DBpedia 정보를 아이폰의 지도 정보와 연계해서 이용자에게 보여주는 DBpedia 모바일서비스를 비롯하여,<sup>13)</sup> 국가가 보유한 공공정보를 일반인들이 자유롭게 활용하여 다양한 어플리케이션들을 개발할 수 있게 링크데이터 형태로 구축해서 웹에 공개하는 data.gov 서비스나 data.gov.uk와 같은 서비스 등도 찾아볼 수 있다.

## 5. LLD의 혜택

W3C LLD 보육작업반 연구보고서<sup>14)</sup>는 링크데이터의 강점을 공유성(shareable), 확장성(extensible), 재사용성(re-usable), 국제화(internationalization)라는 네 가지 관점에서 살피고, 링크데이터를 도서관에 도입할 때 기대할 수 있는 이점을 4가지 측면에서 논하고 있다. 여러 도서관의 협업을 통해 링크데이터 형태로 표현된 정보자원은 RDF 및 URI에 기반을 두므로 외부 타 기관과의 공유와 확장이 쉽고, 재사용성을 확보한다. 국제적인 연계성이 가능한 식별자를 링크데이터에 할당하면 국내,외 링크데이터 간의 연계가 용이해지는 장점도 있다.

10) Tom Heath and Enrico Motta, Revyu.com - a Reviewing and Rating Site for the Web of Data, 2007, <<http://www.cs.vu.nl/~pmika/swc-2007/Revyu.pdf>> [cited 2011. 8. 21].

11) Talis Platform Home page, <[www.talis.com/platform/](http://www.talis.com/platform/)>.

12) Georgi Kobilarov et al., Media Meets Semantic Web - How the BBC Uses DBpedia and Linked Data to Make Connections, 2009, <[http://wtlab.um.ac.ir/parameters/wtlab/filemanager/LD\\_resources/other/eswc2009-bbc-dbpedia.pdf](http://wtlab.um.ac.ir/parameters/wtlab/filemanager/LD_resources/other/eswc2009-bbc-dbpedia.pdf)> [cited 2011. 9. 14].

13) DBpedia Mobile - A Location-Enabled Linked Data Browser, 2008.

14) W3C Library Linked Data Incubator Report,

<<http://www.w3.org/2005/Incubator/lld/wiki/DraftReportWithTransclusion>> [cited 2011. 9. 14].



〈표 2〉 링크데이터를 통한 혜택

구분	혜택
도서관 이용자	<ul style="list-style-type: none"> <li>• UX 측면에서의 사용성 개선</li> <li>• 기 구축된 링크데이터와의 연결을 통한 풍부한 부가정보 탐색 가능</li> </ul>
도서관 조직	<ul style="list-style-type: none"> <li>• 국제 표준 식별자 사용을 통한 정보의 손쉬운 발행</li> <li>• 메타데이터 소유권 명확화</li> </ul>
도서관 담당자	<ul style="list-style-type: none"> <li>• 목록 등 작업 시 중복 리소스 투입 방지</li> <li>• 정보의 최신성 등 유지관리 노력 감소</li> </ul>
도서관 개발	<ul style="list-style-type: none"> <li>• 특정 벤더에 종속된 솔루션 사용으로부터 독립</li> <li>• 오픈소스 활용을 통한 개발상의 편익과 커뮤니티의 지원</li> </ul>

#### 가. 이용자 측면의 혜택

도서관 이용자들은 LLD 기반의 시스템을 사용하면서 기존의 도서관 시스템이 제공하지 못했던 ‘직관적인 사용성’을 경험하고 ‘풍부한 관계정보’를 제공받게 된다. 이용자들은 DBpedia, Geonames, Musicbrainz 등과 같이 LLD를 통해 연결된 다양한 데이터 자원으로부터 기존의 시스템으로는 불가능했던 정도의 풍부한 부가정보를 제공받을 뿐만 아니라 정교하게 발전하고 있는 LLD 기술과 광대한 링크데이터 자원의 결합을 통해 빠른 시간 내에 원하는 정보를 직관적으로 탐색할 수 있게 될 것이다.

#### 나. 도서관 조직 측면의 혜택

현재 도서관 분야에서 사용하는 데이터의 포맷들은 다른 분야의 데이터들과 상호운용이 어렵고 도서관 시스템을 개발하는 일부 업체의 기술이나 솔루션으로부터 독립성을 확보하지 못하는 경향을 보인다. 링크데이터 도입은 도서관 내부의 데이터 관리절차를 향상시키고 디지털 정보자원과 그에 대한 서술정보 사이의 링크를 보다 잘 유지하도록 할 뿐만 아니라, 기관들 사이의 데이터 교환 절차도 크게 효율화하는 결과를 내게 된다. 국제적으로 통용되는 식별자를 사용하는 특성으로 말미암아 기관이 보유한 특징적인 정보자원을 보다 가시적으로 쉽게 노출하는 것이 가능하며, 메타데이터의 소유권도 기존에 비해 보다 명확하게 표현할 수 있다.

#### 다. 도서관 현업 종사자 측면의 혜택

도서관에서 링크데이터를 사용하면 정보자원을 기술할 때 명확하고 재사용성이 높은 동시에 국제적으로도 원활하게 공유되는 형태의 데이터를 생성하므로 현재의 목록작업 절차와 비교할 때 중복적으로 낭비하는 노력을 최소화한다. 즉, 현재 웹의 기반을 이루고 있는 식별자를 사용하여 정보자원을 기술하기 때문에 목록 작업자에 의한 재사용과 최신성의 유지를 기대할 수 있다.

라. 도서관 개발자 측면의 혜택

링크데이터 기반의 시스템을 구축함으로써 도서관 분야에 최적화하여 개발되었으나 복잡하고 비효율적인 MARC나 EAD 등과 같은 포맷을 다루는 목적의 전용 소프트웨어나 도구는 더 이상 개발하거나 사용할 필요가 없어진다. 즉, 특정 영역에 종속된 소프트웨어의 사용으로부터 자유로워질 뿐만 아니라 링크데이터를 다루는 무수한 종류의 오픈소스 도구를 활용하거나 수많은 개발자들이 참여하는 개발자 커뮤니티를 통한 기술적 지원도 얻을 수 있게 된다.

### Ⅲ. LLD 구현기술 및 구축 장애요소

#### 1. LLD 관련 기술

참여 주체의 수가 급속도로 증가하고 있는 상황이 증명하듯이 링크데이터와 관련한 기술도 그 짧은 역사에 비해 매우 괄목할만한 발전을 거듭하는 중이다. 그 결과 현재 개발단계에 있는 대부분의 다양한 기술 혹은 도구들 역시 가까운 시일 내에 실제 적용되는 수준에 이를 것이다. LLD 관련 기술은 링크데이터의 생성과 관련 기술, 링크데이터 클라이언트 관련 기술, 링크데이터 링크 생성 및 통합관련 기술 등 크게 세 항목으로 나누어 살펴 볼 수 있다.

##### 가. LLD 생성 및 추출관련 기술

링크데이터는 웹의 근간을 이루는 표준 HTTP 프로토콜, URI 체계 및 RDF 표준 등을 기반으로 하므로 이들 표준과 연관된 기술과 도구를 통해 간단하게 생성할 수 있고, 기존 도서관 시스템의 변경이나 대규모 개발 비용이 없어도 도입할 수 있는 것이 큰 장점이다.

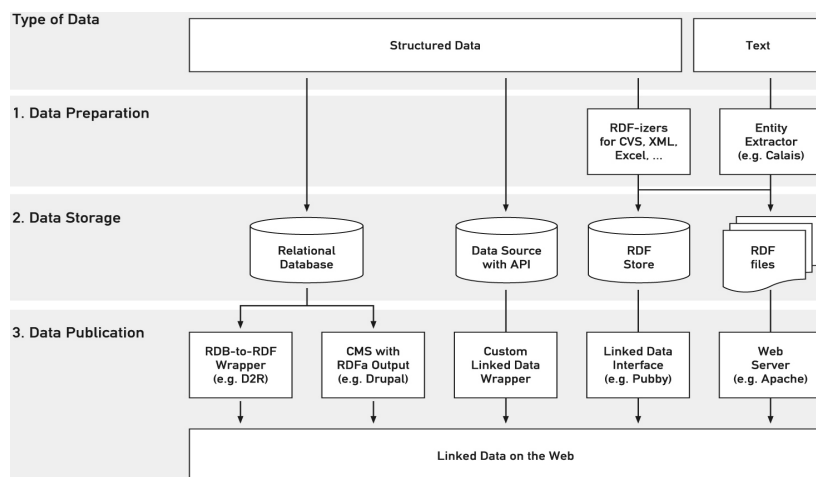
도서관에서 LLD를 생성하는 가장 기본적인 방법은 RDF 생성을 지원하는 GUI 기반의 XML 또는 온톨로지 편집기를 활용하여 LLD 파일을 생성한 후 기존의 웹서버에 업로드하는 것이다. 오래된 웹서버를 사용하는 기관의 경우, RDF/XML 파일에 대한 올바른 MIME 타입을 되돌려 주지 못하는 문제가 발생할 수 있으나 클라이언트로부터 RDF/XML 문서 요청이 있을 때 올바른 MIME 타입을 되돌려 주게 하는 소규모의 작업으로 문제를 해결하면 된다. 이 생성 방식은 기술적으로 가장 간단한데, LLD의 규모가 작거나 접속 트래픽을 고려해서 덤프 파일을 생성해 제공해야 할 경우와 LLD파일을 주기적으로 백업해야 할 경우에는 적합하지만, LLD가 대용량이고 업데이트 주기가 짧은 경우에는 적합하지 않을 수 있다.

기술적으로 단순한 방법을 통한 LLD 제공을 고려하고 있다면 RDFa 포맷을 취하는 것도 가능하다. RDFa는 HTML 페이지 내에 RDF를 기술할 수 있도록 해주며, 정적인 HTML 페이지에

RDFa를 삽입하거나 클라이언트 요청 시 RDFa가 HTML 페이지에 자동으로 삽입되도록 하여 검색엔진에 의한 서비스 노출 효과를 극대화할 수 있다. 실제로 현재 광범위하게 사용 중인 Drupal의 경우, 이용자가 데이터를 입력할 때 이를 바탕으로 한 RDFa의 자동 생성을 지원한다. 서버나 도서관 시스템에서 애플리케이션의 하부를 구성하는 데이터베이스 및 저장소 등의 단계에서 LLD를 생성하게 할 수도 있다. LLD를 요구하는 모든 클라이언트들의 요청을 처리하도록 RDF를 생성하거나 URI Dereference 및 Redirection을 지원하도록 서버 측 스크립트를 작성할 수도 있으며, D2R서버, OpenLink Virtuoso, Triplify 등과 같은 도구에서 제공하는 맵핑 스크립트 또는 테이블 작성을 통해 관계형 데이터베이스로부터 생성하려는 LLD 데이터를 자동으로 추출할 수도 있다.

만약 도서관에서 openAPI 서비스를 제공하는 경우라면 래퍼(Wrapper)를 통한 LLD의 추출 및 생성도 고려할만한 방식이다. 즉 클라이언트에 의한 모든 LLD 요청을 openAPI 요청으로 변환하여 서버에 전달하고 서버의 openAPI 응답을 다시 RDF로 변환하여 클라이언트에 전달해 주는 링크데이터 래퍼를 통해 LLD를 자동으로 생성하는 것이다. 이러한 방식은 openAPI가 광범위하게 활용되고 있는 현재의 웹 환경에서 도서관 이외의 기관도 적극적으로 활용할 수 있다. 도서관이 현재 RDF 데이터 저장을 위한 별도의 Triple 저장소 등을 운영하고 있다면 매우 간단한 설정 변경만으로도 기존의 RDF 데이터를 LLD로 추출해 서비스 하는 일도 가능하다.

〈그림 2〉는 데이터의 종류에 따라 효율적인 링크데이터 발행 방법을 도식화한 것이다. 링크데이터로 발행되어야 하는 데이터 볼륨의 양과 데이터의 변경 빈도수에 따라 데이터의 역동성을 고려해 단순한 문서와 구조화된 문서를 구별하고, 데이터의 사전 준비단계를 거쳐 데이터 저장소의 특징을 파악하여 웹서버로 발행하거나, API의 특성에 따른 래퍼를 사용하거나, D2R 서버와 같은 RDB2RDF 래퍼를 사용하여 링크데이터를 발행하는 것이다.



〈그림 2〉 링크데이터 발행의 종류<sup>15)</sup>

#### 나. LLD 클라이언트 관련 기술

링크데이터 클라이언트는 실제 링크데이터를 탐색하여 필요한 정보를 얻고자 하는 이용자를 지원하기 위한 링크데이터 브라우저, 링크데이터 편집자를 위한 에디터, 효율적인 검색 서비스 제공을 위한 링크데이터 검색엔진 및 인덱서 등으로 구분할 수 있다.

Tabulator, Marbles RDF 브라우저, razorbase, ODE 등이 대표적인 브라우저인데, RDF 탐색을 지원하는 브라우저라면 링크데이터 또한 자유롭게 탐색할 수 있다. 최근에 개발되고 있는 브라우저는 링크데이터 편집기능을 동시에 제공하는 경우가 많으며, 대부분의 웹브라우저들이 지원하는 확장기능 형태로도 개발되어 일반 웹브라우저에서의 링크데이터 탐색을 지원하기도 한다. 링크데이터 생성 및 편집에 활용할 수 있는 수많은 에디터가 존재하며, Protege, Ontostudio, Knoodl 등과 같이 RDF/OWL 온톨로지 생성 시 광범위하게 사용되는 온톨로지 에디터를 활용하면 링크데이터의 용이한 생성이 가능하다.

#### 다. LLD 링크 생성 및 통합관련 기술

LLD의 성패는 링크데이터 간의 링크정보를 생성하고 동일한 개념의 링크데이터를 효과적으로 통합하는 노력을 통해서 이용자에게 얼마나 풍부하고 정확한 정보를 제공해 주는가에 달려 있다고 하겠다. 따라서 링크데이터 간의 링크정보 생성 및 통합과 관련된 기술은 가장 핵심적인 링크데이터 기술이다.

각각의 링크데이터가 취급하는 도메인 영역의 전문가 또는 편집자가 링크데이터 브라우저를 사용하면서 개념이 일치하거나 연관 관계가 있는 링크데이터를 데이터 소스에서 검색하고 연결 관계를 추가하면 서로 다른 링크데이터 소스를 연결하는 링크 정보의 정확성을 최적화할 수 있을 것이다. 하지만 이러한 수작업을 통한 링크정보 생성 및 동일개체의 통합은 비교적 소규모 데이터 소스에 대해서는 적합한 방법일 수 있지만 막대한 양의 시간적, 물질적 자원이 들어가는 단점이 있다.

대규모 링크데이터 소스 사이의 링크생성을 지원하는 도구도 다양하다. Knofuss, Interlinking Distributed Social Graphs, GNAT와 같이 문자열 비교, 유사도 측정 및 다양한 시맨틱 알고리즘 활용을 통해 링크정보 자동생성을 지원하는 도구가 있는 반면, RDF-AI, SILK, GUMF, ExpLOD와 같은 대부분의 도구는 링크정보 생성의 완전 자동화 단계에는 이르지 못해 매칭규칙 정의 및 부가정보제공과 같은 작업의 부가적 수행을 요한다.<sup>16)</sup>

그러나 이처럼 활용 가능한 도구가 많지만 대부분 특정 도메인에 속한 링크데이터 소스에 대한

15) Tom Heath and Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space*, 1st ed. (Washington : Morgan & Claypool, 2011).

16) Stephan Woßlger et al., *Interlinking Data - Approaches and Tools*, 2011.  
<[http://www.sti-innsbruck.at/fileadmin/documents/technical\\_report/Interlinking\\_Data-Approaches\\_and\\_Tools-Technical\\_Report\\_March\\_2011.pdf](http://www.sti-innsbruck.at/fileadmin/documents/technical_report/Interlinking_Data-Approaches_and_Tools-Technical_Report_March_2011.pdf)> [cited 2011. 8. 10].

처리에만 최적화되어 있고, 링크데이터가 속한 도메인과 무관하게 정확한 링크정보 생성을 보장하는 강력한 도구는 아직 나오지 않은 상황이다. 따라서 보유하고 있는 링크데이터 데이터 처리에 가장 적합한 도구를 면밀히 검토한 후에 상황에 맞게 적용할 필요가 있다.

## 2. LLD 구축 장애요소

### 가. 내외부적 환경변화에 대한 도서관의 보수적 대응

도서관이 MARC를 도입하고 컴퓨터를 사용하여 목록의 자동화를 피하기 시작했을 당시, 도서관은 새로운 기술을 빠르게 흡수하고 활용하는 기관으로 인식되었다. 그러나 이후 기술은 급속한 발전을 거듭한 반면, 도서관은 장점이 확실하게 검증된 기술에 국한해서 수용적인 자세를 보여 온 것이 사실이다. 링크데이터라는 신기술이 도서관 외의 맥락에서 쌓아가는 다수의 적용 사례가 도서관의 링크데이터 적용으로 이어지지 못하는 주요한 이유도 거기에서 찾을 수 있다. 도서관 분야에서의 성공 사례를 지켜보고서야 움직이려는 보수적 풍토가 도서관에서의 링크데이터 적용을 저해하는 요인으로 작용하는 것이다.

나아가 도서관관련 표준의 경우, 국립기관이나 국제기관이 주도해서 시범모델 또는 프로토타입 없이 문서로 먼저 작성하는 관행도 LLD 구축에는 장애요소가 된다. 웹기반 표준이 실례 중심의 기술과 세부문제 차후 해결이라는 매우 실행주의적인 절차를 밟아 성숙하는 것과는 대조적이다. 결국 도서관관련 표준의 제정에는 상대적으로 더욱 긴 시간이 소요될 뿐만 아니라 급변하는 기술 환경에 대한 도서관의 대응 속도 역시 느려지는 것이다.

### 나. 도서관 데이터의 상호운용성 부족

현재의 도서관 데이터는 도서관 사이에서만 상호운용이 가능한 형태를 취한다. 소 단위별 URI를 사용하지 않고 레코드 단위로만 표현되는 데이터이기 때문이다. 일부 도서관이 데이터의 RDF 전환을 진행하고 있지만, 전통적인 도서관 데이터의 변환은 단순히 도서관 시스템의 데이터 속성과 RDF 속성을 매핑하는 것 이상의 작업을 요구한다.

### 다. ROI 값 계산의 어려움

LLD 구축을 위해 투자하는 비용과 그 이득(ROI)을 계산하는, 즉 공공서비스의 비용 효과를 추산하는 작업은 용이하지 않다. ROI 값의 산출이 쉽지 않다는 사실은 대규모의 변화를 요구하는 링크데이터 사업의 정당성의 확립과 주장이 난제라는 뜻이다. 현재의 데이터 생성에 소요되는 ROI와 LLD 전환 ROI를 비교하는 어떤 객관적인 측정도구가 존재하지 않는 상황은 LLD 구축을 위한 비용지출의 정당성을 확립하는 작업에 걸림돌로 작용하고 있다.

## IV. 국내 LLD 적용을 위한 제안사항

### 1. 전거 파일이나 통제어 리스트로부터 시작할 것

링크데이터의 초기단계에서는 필요성이 크면서 적은 노력으로도 가능한 것부터 시작하는 것이 바람직하다. 전거 파일이나 통제어 리스트를 SKOS를 이용해서 정리한다면 도서관 데이터를 링크데이터화 하는 작업에 효과적인 교두보가 된다. LCSH를 SKOS 기반으로 변환한 미국의회도서관이 그 좋은 예다. 한국의 전거파일이나 통제어 리스트를 LCSH를 참고해서 연계한다면 보다 빠르고 질 좋은 전거파일과 통제어 리스트를 구성할 수 있으리라 기대된다. 또 LIBRIS를 구축한 스웨덴 국립 도서관의 경우에도 서지 데이터를 링크데이터화 하기에 앞서 이미 잘 알려진 URI를 적극 활용하는 정책을 도입해 저자명, 조직명에 foaf:name을 주고, 주제명 표목과 시소러스 구축에는 skos:concept을 이용했고,<sup>17)</sup> 시스템 전체를 중단 시키기거나 기존 업무에 차질을 주지 않는 가운데 도서관 데이터를 링크데이터화 하여 효과를 신속히 볼 수 있었다. 국내에서도 유사하게 접근한다면 우리의 전거파일이나 통제어 리스트가 국제적으로 보다 원활하게 융합되는 기반이 마련될 것이다.

### 2. 도서관 데이터의 RDF 전환 시도

새로운 기술을 적용할 때 그 결과의 도출이 더딘 상황은 가능하면 조성하지 않는 것이 좋다. 따라서 도서관 시스템 내부에만 갇혀있는 특정 주제에 대한 데이터를 전환해 링크데이터 환경으로 가지고 나와 그 유익을 검증하고 여러 기관의 협업 분위기를 조성해나가는 것이 바람직하다. 링크데이터의 폭발적인 연결 능력을 보기 위해서는 여러 기관의 협력이 필요하기 때문에, 이처럼 도서관내의 특정 영역부터 시작하거나 혹은 전문 도서관 데이터를 링크데이터로 전환할 것을 제안한다. 대학도서관의 경우, 도서관 자체에서 생산하는 정보자원의 거의 대부분을 MARC, MODS 등의 표준에 따라 기술하므로 협의체 구성을 통해 기관별로 특정 주제와 도메인을 할당하고 각 기관이 보유한 데이터를 RDF로 전환한 후 공유한다면 상당한 시너지 효과가 기대된다. 특히 Cloud Computing이 도래한 현 시점에서는 대학도서관 데이터를 RDF로 변환시켜서 거대한 대학통합 링크데이터 Cloud를 생성하는 프로젝트 역시 시도할만한 가치가 있다.

17) Martin Malmstern, *Exposing Library Data as Linked Data*, 2009.  
<<http://www.ifla2009satelliteflorence.it/meeting3/program/assets/MartinMalmstern.pdf>> [cited 2011. 11. 17].

### 3. 메타데이터 생성 및 유통구조의 모범실무 구축

현재 도서와 관련된 메타데이터는 크게 출판계와 도서관계라는 두 주체가 각자의 목적을 따라 생성하고 활용한다. 출판계의 ONIX, BISAC/BIC, ISNI, ISTD와 도서관계의 MARC, LCSH/DDC, 전거통제, FRBR 등은 구조와 어의 면에서 완벽한 상호운용이 불가능하다. 그러나 노력의 중복을 방지할 필요와 상호 데이터 교환으로 발생하는 유익을 고려할 때 이 두 주체는 공동대응을 통한 상승효과를 추구해야 한다.<sup>18)</sup> 링크데이터는 상호호환을 통한 상승효과와 상생효과를 생각할 때 최적의 도구이다. 링크데이터로 출판계와 도서관계의 메타데이터를 변환해서 연결한다면 서로 보다 풍성한 메타데이터를 확보하는 상승효과를 기대할 수 있다. 또한 이렇게 연결된 데이터에 서평, 구입 후기 등 자료 구매 욕구를 일으킬 가능성이 있는 데이터를 더 링크한다면 둘 사이의 상생효과도 커질 것이다.

### 4. 도서관 간의 협업 분위기 조성

도서관이 링크데이터로의 변환을 통해 보다 나은 서비스를 제공하려면 협업은 필수적이다. 각 도서관들마다 개별적으로 데이터 연결을 위한 링크데이터로의 변환을 시도하고 협업하지 않는다면 결국 상당한 자원의 낭비만 초래한다. 그러나 도서관들 사이의 협약을 세우고 데이터 링크를 공유한다면 빠른 서비스 구축에 도움이 될 것이다. 링크데이터를 통해 실제적인 실익을 얻기 위해서는 가능한 한 많은 기관이 LLD에 관심을 가지고 각 기관이 소유한 데이터를 링크데이터로 방출할 수 있는 환경을 조성하는 것이 중요하다. 링크데이터에 관한 논의가 초기 단계에 머물러 있는 국내 상황에서 자원의 중복 투입으로 인한 피해를 막으려면 링크데이터 구축과 관련한 전반적인 노하우를 기관들 사이에 원활하게 공유할 필요가 있다. 이 사안은 각 도서관중별로 참여할 수 있는 별도의 협의체를 구성해 논의를 진행할만하다. 공론의 장을 통해서 각 기관별로 강점을 지닌 정보자원을 링크데이터로 전환하도록 하고, 강한 기술력을 지닌 기관이 링크데이터 발행 전반에 관한 기술적 가이드라인을 작성해서 타 기관을 지원한다면 보다 조속한 시일 내에 링크데이터가 구축되어 국내외 기관들과 상호 연결되는 기반이 마련될 것이다.

### 5. 기관에 적합한 URI 정책 수립

앞서 3장에서 언급했던 바와 같이 링크데이터의 구현에서 가장 중요한 것은 모든 기관이 보유한

18) 조재인, “공급사슬상의 도서메타데이터 생성 유통에 관한 고찰,” 한국문헌정보학회지, 제44권, 제3호(2010, 8), pp.61-80.

모든 정보자원에 고유한 식별자를 할당하는 일이다. 물론 기존의 데이터베이스에서도 각각의 레코드를 구분하는 고유의 식별자를 정의했지만 시스템 내에서만 유효하거나 타 시스템 또는 기관과의 연계를 위해서는 전혀 도움이 되지 않는 식별자에 불과했다. 링크데이터 구축 후에 그 링크데이터가 국내외의 다른 기관과 연결도록 하기 위해서는 개별 정보자원에 대해 URI 식별자를 부여할 필요가 있다. URI는 자유롭게 정의하여 할당할 수 있으나, 현재 기관이 사용하고 있는 도메인 이름을 바탕으로 체계적인 URI 명명정책을 세워서 관리 대상인 모든 자원에 URI를 할당하는 것이 보다 바람직하며 필요에 따라 충분히 확장될 수 있는 체계를 갖추어야 할 것이다. Tom Heath와 Christian Bizer(2011)<sup>19)</sup>가 제시하고 있는 URI 정책은 매우 구체적이고 명확하므로 참고할 만하다. URI 정의에서 각별히 주의해야 할 부분은 URI 자체만으로도 대상 정보자원이 무엇인가를 유추할 수 있을 만큼 직관적이고도 간단한 구조로 설계되어야 한다는 점이다.

## 6. 외부자원과의 연결 노력

도서관 데이터를 링크데이터로 발행한다고 해서 즉각적이고 폭발적인 정보 연결이 일어나는 것은 아니다. 데이터가 다양한 외부자원(서점, 블로그, 소셜미디어 데이터)과 연결될 때 기대하지 않던 정보의 재발견이 용이해지고, 끝없는 정보의 향해가 이뤄지며, 데이터의 메쉬업을 통해 새로운 어플리케이션의 생성에도 이바지하게 될 것이다. 스웨덴 LIBRIS의 경우에도 도서관 데이터를 링크데이터로 발행한 이후 데이터의 발견성과 재 사용률을 높일 목적으로 권위 있는 연결 대상이라고 할만한 DBPedia와 LCSH에 연결한 것을 주목해야 한다.<sup>20)</sup>

## 7. 도서관 이용자의 링크작업 동참 확보

웹2.0이 누린 폭발적인 호황은 기존 이용자의 활발한 지식생산이 배경으로 작용한 바가 크다. 만일 네이버 지식인을 NHN인력만으로 만들어 나갔거나 위키백과의 정보 생산에 위키의 직원들만 가담했다면 웹 2.0의 정신과도 맞지 않았을 뿐만 아니라 오늘날의 엄청난 지식 데이터베이스 또한 생성되지 못했을 것이다. 이는 일반 사용자들의 적극적인 참여가 뒷받침될 때 기존에 생성된 링크데이터가 보다 신속하고 풍성하게 연결되어 재사용과 새로운 창작물의 출현에도 크게 기여할 것을 짐작하게 되는 이유다. 정효숙<sup>21)</sup>은 교육영역에서 일반인이 쉽게 참여하여 링크데이터를 연결할 뿐만 아니라 발행할 수도 있는 시스템을 구축하고, 그 효과로서 데이터 재사용의 용이성과 관련된

19) Tom Heath and Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space - chapter 4*, 1st ed.(Washington : Morgan & Claypool, 2011).

20) Martin Malmstern, *op cit*.

21) 정효숙, 김희진, 박성빈, *전계논문*, pp.47-59.



학습 자료 간의 유의미한 연결로 학습이해를 증진하고, 출처가 다른 교육데이터의 관리를 유용하게 하며, 따라서 더욱 우수한 교육 콘텐츠의 제작이 가능해진다고 주장하였다. 도서관이 기존의 단편적인 서지정보를 벗어나 링크데이터를 통해 관련 데이터를 유의미하게 연결하고 나아가 RDF 데이터를 이용한 다양한 검색을 도모하는 독자적인 플랫폼을 구축한다면 이용자들이 자료와 관련된 보다 유의미한 데이터를 다양하게 링크된 형태로 쉽게 검색할 수 있게 될 것이다.



〈그림 3〉 이용자 참여형 도서관 Linked Data 플랫폼

## 8. 선행연구 및 해외사례의 적극적인 활용

LLD의 개념이 소개된 이후 수년에 걸쳐 다양한 응용사례와 서비스모델을 개발해온 해외의 움직임과 비교할 때 국내에서의 LLD 관련 논의는 비교적 초기단계다. 일례로 현재 활용 가능한 링크데이터 셋의 수를 살펴보면, OKF(Open Knowledge Foundation)가 운영하는 데이터허브(thedatahub.org)에 등록된 데이터 셋의 수는 2011년 10월을 기준으로 2,392개 정도이며, LLD 데이터셋의 수도 52개에 달한다. 이에 반해 국내의 활용 가능한 링크데이터 셋은 거의 전무한 실정이다.

수많은 데이터 셋들 간의 연결을 통해 시너지 효과가 나는 링크데이터의 특성상 현재 가장 중요한 일은 개인, 기관 등 다양한 주체들의 노력을 통해 다양한 종류의 데이터 셋이 생성될 수 있도록 하는 것이다. 링크데이터의 핵심 철학이 데이터 개방과 공유인 만큼 해외에서 생성한 데이터 셋의 대부분이 접근과 취득에 용이하고 링크데이터를 시스템적으로 구현할 수 있는 프레임워크나 도구도 오픈소스가 주류라는 것은 매우 다행스런 사실이다. 따라서 링크데이터의 구현을 고려하는 기관이라면 웹상에서 찾아지는 수많은 연구사례와 적합한 구현사례를 참고할 때 큰 도움을 받을 것이다. 또 링크데이터가 수많은 데이터 셋 간의 연결을 통해 폭발적인 힘을 발휘하는 특성이 있으므로 새로운 어휘를 만들기 전에 상호운용 가능한 기 구축된 해외의 어휘들을 잘 분석해서 사용하는 것이 바람직하고, 그 과정에서 개발자들의 시간과 에너지도 절약할 수 있다. 이미 구축된 서지온톨로지(<http://bibliontology.com>)의 어휘와 국제표준인 더블린코어(Dublin Core) 요소를 사용해서 서지 데이터를 구축한 스웨덴 LIBRIS의 사례가 그 좋은 근거를 제공한다.

## 9. 신기술의 적극적인 활용

도서관은 데이터 유지와 관리의 측면에서 타 분야의 기관들에 비해 매우 조심스럽고 보수적인 자세로 대응하는 경향이 있다. 이는 데이터 저장기술이 하루가 다르게 발전하고 수많은 데이터 모델이 제시되면서 급변하는 상황에서도 도서관은 수십 년간 MARC와 같은 단일한 메타데이터를 사용해온 사실에서도 확인된다. 오랜 기간 동일한 포맷으로 유지하며 관리해온 데이터를 익숙하지 않은 링크데이터로 변환하는 일에 도서관계가 느끼는 부담은 이해할만한 것이다. 그러나 링크데이터 구축의 강점들 중의 하나는 현재 도서관이 운영하고 있는 기존의 시스템을 변경하거나 많은 예산이 들어가는 신규 개발을 거치지 않고도 링크데이터를 구현하는 다양한 방법과 도구가 찾아진다는 것이다. 일례로 D2R 서버와 같은 오픈소스 도구를 사용하면 도서관이 운영 중인 데이터베이스를 링크데이터로 표현하는 맵핑 규칙을 작성하는 것만으로도 링크데이터를 생성해서 웹에 방출할 수 있다. 그러나 향후 다른 기관들과의 연계나 협업 등을 고려할 때 기관 내 로컬서버에서의 작업이 바람직하지 않을 경우라면 클라우드 기술의 활용이 적절한 대안이다. 클라우드 플랫폼의 활용을 통해 하드웨어, 서버 소프트웨어 설치 및 설정에 대한 부담을 크게 줄이고, 기관 간에도 물리적인 제약 없이 동일한 플랫폼 환경에서 협업을 진행할 수 있는 부가적인 장점도 얻게 된다. 이 경우에는 클라우드 플랫폼을 적극적으로 활용하고자 하는 해외 도서관 및 관련 기관들의 움직임도 주목하고 나가야 할 것이다.

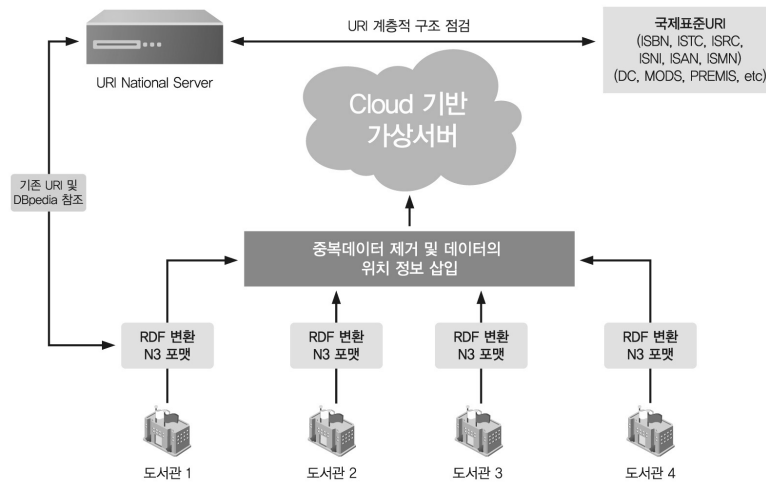
## 10. 링크데이터 모범실무 패턴의 설계

LLD의 요구에 적합한 설계 패턴을 세울 필요가 있다. 이런 패턴은 새로운 기술을 이해하는 데 기존의 설계유형에 의존하는 사람과 개발자의 필요를 채워줄 것이고, LLD의 일관성을 증진시켜 줄 것이다. 또 RDF 어휘를 관리하고 보존할 모범실무정책, URI의 생성과 유지, 적절한 네임스페이스 할당 정책 등을 동시에 관리할 조직도 요구되며, 일반적으로 다음과 같은 업무가 수행되어야 한다.

- 모범실무 가이드라인을 기초로 URI를 설계하는 패턴의 사용
- URI의 영속성 유지
- 온톨로지 설계방법, 구조화된 어휘 생성을 위한 방법론 제공
- 각 URI와 네임스페이스를 관리하기 위한 버전 통제
- 도서관시스템에서 보다 원활한 요소, 어휘, 메타데이터를 활용을 위한 시맨틱웹의 기술과 보조를 맞춘 URL의 사용
- 소규모의 조직에 있어서 네임스페이스 확장가능성 유지
- 레이블과 주석의 다른 언어로의 전환 안내

## V. 한국 통합 LLD 모형

본 연구에서는 앞서 살펴본 LLD 구축관련 기술과 제안사항을 기초로 한국도서관에 적용할 수 있는 한국 통합 LLD 모형, <그림 4>를 제안한다.



<그림 4> 한국 통합 LLD 모형도

- 1) 한국 통합 URI 서버를 구축한다. 각 도서관은 URI를 자동으로 할당할 때 이 URI 서버를 참조하며, 이 URI 서버는 dbpedia 등 주요 링크데이터 데이터소스의 내용을 복사해서 가지고 있거나 실시간 조회기능을 제공하고, 국제적으로 통용되는 URI 정보를 포함한다.
  - a. 특히 현재 다양한 분야에서 활용되고 있는, ISO 산하 기술위원회(TC 46 - 문헌정보)의 하부위원회(SC9)에서 제정한 국제표준식별자(ISO 식별자)를 DOI를 통해 동적식별자(Actionable Identifier)로 변환해서 링크데이터의 URI로 사용할 필요가 있다. ISO 식별자는 FRBR 모델과도 원활하게 매핑되므로 링크데이터를 FRBR 기반으로 그룹화 하는 작업에 적절하다.
  - b. 네임스페이스를 선언한 다양한 메타데이터 스키마 요소에 부과된 다수의 URI를 링크데이터 관계 기술에 적극 활용하여 모든 링크데이터의 국제적 상호호환성을 증진하는 효과를 도모한다.
- 2) 각 도서관은 보유하고 있는 목록데이터를 RDF N3 포맷으로 발행한다. 목록데이터를 RDF N3포맷으로 일괄 변환할 수도 있으나, 기존의 RDBMS 테이블과 RDF 간의 맵핑 테이블 작성을 통해 기존 도서관 시스템을 크게 변경하지 않고 RDF를 발행할 수도 있다. 이 단계의

원활한 진행은 도서관들의 적절한 합의와 협업을 요하며, N3 포맷을 활용하는 이유는 그것이 데이터의 보다 신속한 RDF 전환에 용이하기 때문이다.

- 3) 목록데이터를 RDF Triple로 전환할 때, 통합 URI 서버를 참조해서 동일자원에 서로 다른 URI를 부여하는 오류를 피하고 동일 URI를 적극 활용하여 원활한 링크가 이루어지도록 한다.
- 4) 각 도서관에서 생성된 RDF 중 중복된 데이터를 제거하고, 각 RDF의 생성 주체 및 소장기관의 위치정보를 기록한다. 이 과정은 모든 자원을 서로 연결시킬 뿐만 아니라 자원의 위치파악을 용이하게 한다.
- 5) 마지막으로, 생성되고 관리된 RDF 데이터를 클라우드 기반의 가상서버에 링크데이터로 발행한다. 결과적으로 통합 LLD 구축에 참여하는 모든 기관이 클라우드 기반의 서버에 등록된 모든 자원에 대한 접근권을 소유하며, 모든 도서관이 기 발행된 목록데이터에 대한 검색권을 소유하게 되므로 시간과 에너지를 절약하고 이용자의 편의를 증진하게 될 것이다.

## VI. 결 론

본 연구에서는 현 도서관의 문제점을 극복하고 서비스를 향상시키는 대안으로써 링크데이터에 주목하고 W3C LLD 보육작업반의 연구를 비롯한 선행연구를 분석한 후, 국내 도서관에 적합한 제안사항들과 한국 통합 LLD 모형을 제시하였다. 도서관 분야에서의 국내 사례가 거의 전무하고 ROI 도출 자체가 난제이며 목록데이터에 저작권 문제가 동반되는 등, 다수의 선결과제나 장애요소가 있으나 장기적인 측면에서 링크데이터 구축으로 얻어지는 실익이 매우 크다는 것은 링크데이터 관련 연구사례와 참여기관의 수가 해를 거듭하며 증가하는 추세라는 사실로도 입증된다. 공유성(shareability), 확장성(extensibility), 재사용성(re-usability), 국제성(internationalization)이라는 링크데이터의 장점을 통해 도서관은 이용자들에게 풍부한 정보를 제공하고 데이터를 효율적으로 관리할 수 있다. 나아가 데이터 중복생성에 드는 시간과 에너지를 절약하며, 다양한 오픈 소스를 활용하여 서비스 중단 없는 지속적인 개발도 도모할 수 있다.

한국 통합 LLD 모형의 구축에 있어서는 계속 늘어나는 링크데이터 관련 해외 선행 사례를 적극적으로 발굴하고 적용하여 더 발전된 서비스를 개발하고, 클라우드 컴퓨팅과 같은 인프라와 개발환경을 도입하여 초기 구축비용과 유지비용을 줄이며, 도서관 메타데이터를 클라우드 기반의 서버에 중복 없이 모으고 통합된 도서관 링크데이터와 외부의 유익한 데이터를 연결할 것을 이 연구는 제안하였다. 그 결과로 형성된 링크데이터를 기반으로 가능해진 정보검색의 폭과 깊이에 의거해서 이용자의 참여를 통한 보다 섬세하고 정확한 서비스도 현실화시킬 수 있을 것으로 보인다.

도서관계는 이미 FRBR 및 RDA를 통해서 기존 도서관 목록의 시맨틱 수준을 끌어올릴 준비를

갖추었다. 주제 및 인명에 관한 전거를 가지고 FOAF와 SKOS 등의 어휘를 활용해서 비교적 적은 노력으로 도서관의 데이터를 웹상에 링크데이터로 발행할 경우, 큰 효과도 기대된다. 다시 말해서 도서관계는 이미 상당부분 링크데이터로 나갈만한 준비를 해온 것이다. 새로운 기술의 적용을 위한 선결 사항들이 있다 할지라도 링크데이터는 이용자 감소 및 예산 감축과 같은 도서관의 현실적인 문제를 해결할 기회를 제공할 가능성도 크다. 도서관계가 실질적이고도 명확하며 고유한 이익을 이용자들에게 돌려주는 서비스 주체로서 한 걸음 더 도약하는 발판을 링크데이터의 적극적인 도입에서 찾기를 기대한다.

〈참고문헌은 각주로 대신함〉

