

# 도서관 데이터 공유개방을 위한 링크드 데이터(Linked Data)의 활용

:: 현미환\_한국과학기술정보연구원

최근 글로벌 트렌드로서 오픈데이터에 대한 각국 정부의 관심과 노력이 확대되면서 공공데이터 분야를 중심으로 링크드 오픈데이터(Linked Open Data)에 대한 관심 높아지고 있다. 특히 미국과 영국 등 해외 주요국을 중심으로 공공데이터를 링크드 데이터로 구축하는 작업을 진행하고 있으며, 국내에서도 국가 차원에서 링크드 데이터 시범사업으로 공공DB피디아를 구축하면서 공공데이터 활용을 위한 링크드 데이터 구축이 활발하게 논의되고 있다. 이렇게 오픈데이터에 대한 인식이 확산되고 있는 환경에서 질적, 양적으로 수준 높은 데이터를 보유한 도서관에서 데이터의 공유, 개방에 적극 동참한다면 자관이 보유한 데이터의 활용성을 높이는 것은 물론 도서관 데이터 간의 상호연계와 의미적 연결을 통해 데이터를 기반으로 한 지식생태계 조성이 가능해 질 수 있을 것이다. 본 고에서는 도서관에서 링크드 데이터를 통해 얻을 수 있는 혜택이 무엇인지를 생각해 보고, 이를 통해 도서관의 링크드 데이터 활용가능성에 대해 논의해 보고자 한다. 이를 위해 링크드 데이터의 정의와 특징, 도서관 링크드 데이터 구축 동향, KISTI의 링크드 오픈데이터 구축 사례를 중심으로 살펴보고자 한다.

## 1. 링크드 데이터의 정의와 특징

링크드 데이터(Linked Open Data : LOD)는 시맨틱 웹을 구현하기 위한 방법으로 역참조할 수 있는 URI(dereferenceable URIs)를 활용하여 웹 상에 데이터를 공개, 공유, 연결하는 방법으로 웹에 존재하는 다양한 정보 자원의 공개와 공유를 위해 정보 자원을 연결하는 기술이다. 현재 문서 중심의 웹(Web of Document)이 데이터가 중심이 되는 새로운 계층(Web of Data)으로 진화하면서 보다 정확한 의미 연계를 통해 정보를 생산하고, 유통하는 표준을 적용한 데이터 집합이 링크드 데이터이다.

링크드 데이터는 2006년 팀 버너스 리에 의해 소개된 이후, 많은 데이터들이 RDF 형태로 공개되고 상호 연계되어 활용되기 시작했다. 2007년부터 2010년까지 링크드 데이터는 매년 2배 이상의 종류와 크기로 성장하여 링크드 데이터 클라우드를 형성하였다. 링크드 데이터 클라우드는 RDF 트리플(Triples) 및 링크(Links)를 이용한 데이터셋 구성 프로젝트인 LOC(Linking Open-data Community)를 통해 생성된 링크드 데이터셋의 시각화 표현이다. 2007년에 20억 개였던 RDF 트리플은 2011년

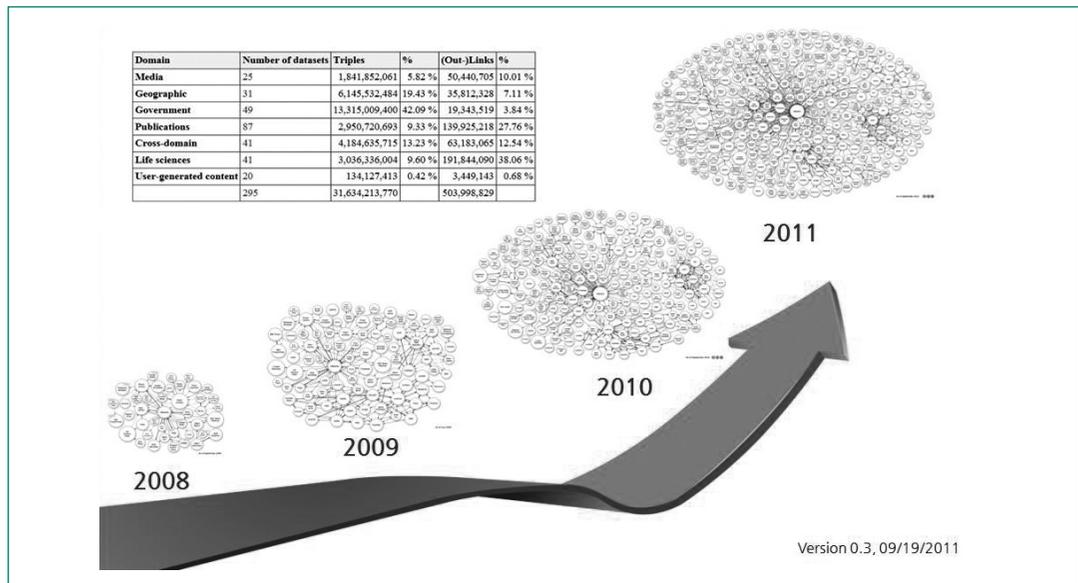
〈링크드 데이터 생성원칙(Linked Data 4 Principles)〉

- 개체 식별을 위해 URI를 사용한다.
- HTTP 프로토콜을 통해 접근할 수 있는 URI를 사용한다.
- 개체가 URI를 통해 참조될 때 유용한 정보를 제공하기 위해 RDF, XML과 같은 표준 형식을 사용한다.
- 다른 개체와의 링크 정보를 포함시켜 관련 정보의 검색 수준을 향상시킨다.

310억 개로 증가하였고, 200백만 개이던 RDF 링크는 5억 개로 증가하였다. 2011년 링크드 데이터 클라우드를 기준으로 도메인별 데이터셋의 개수와 트리플, 링크 현황은 <그림 1>과 같다. 링크드 데이터 클라우드는 총 8개의 도메인으로 구분되어 있으며, 미국이나 영국 등의 국가에서 정부 주도의 공공데이터를 링크드 데이터로 공개하면서 전체 중 40% 이상을 공공데이터가 차지하고 있다.

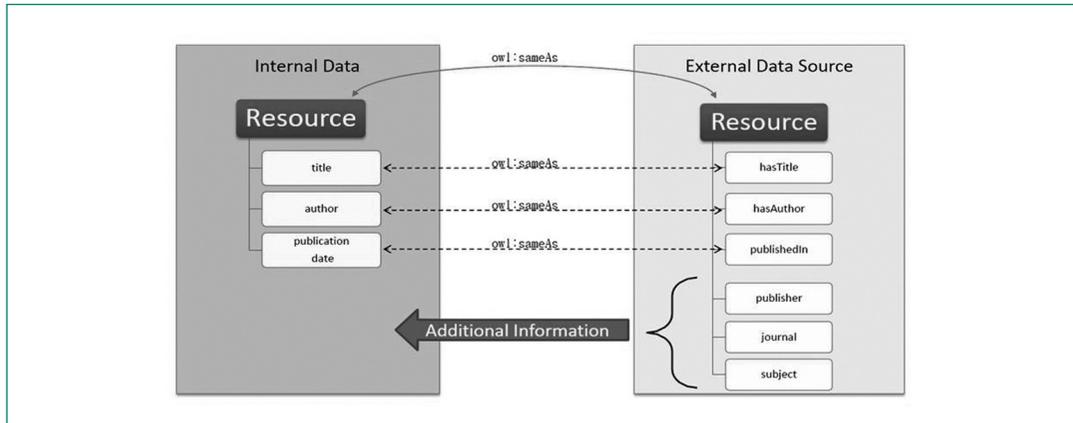
링크드 데이터를 이용하면 데이터의 재사용이 가능한데, <그림 2>에서 보듯이 각각의 데이터에 링크드 데이터 식별자인 URI 링크가 부여되어 외부에서 지속적으로 재사용할 수 있는 장점이 있다. 데이터 재사용을 통한 데이터의 중복 문제를 해결할 수 있기 때문에 도서관에서는 데이터 구축이나 관리를 위한 비용을 절감할 수 있는 효과가 있다. 또한 링크드 데이터를 이용하면 하나의 데이터를 통해 관

그림 1 ▶ 링크드 데이터 클라우드



(출처 : <http://lod-cloud.net/state/>)

그림 2 ▶ 링크드 데이터의 활용 : 데이터 구축 중복 방지



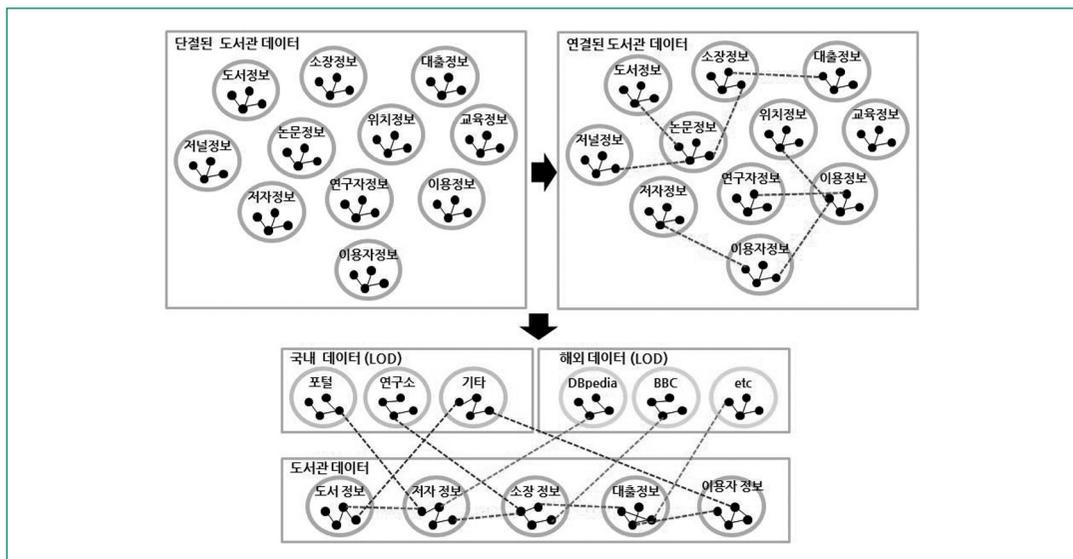
련된 또 다른 데이터로 연결이 가능하며 이를 통해 정보의 확장도 가능하다.

〈그림 3〉에서 보듯이 단절되어 있던 도서관 데이터가 링크드 데이터를 통해 지속적으로 개방, 공유되고 연결되면 도서관 분야 외의 데이터베이스와도 링크되면서 하나의 글로벌 데이터베이스가 가능해 질 수 있다.

## 2. 도서관의 링크드 데이터 구축 동향

도서관 링크드 데이터는 전통적으로 도서관이 관리하고 있는 데이터를 변환하는 것을 기본으로 하며, 이 범주에 속하는 데이터는 서명, 저자, 날짜 등을 포함하는 서지 데이터(bibliographic data)와 분

그림 3 ▶ 링크드 데이터 적용 전/후 비교



류, 주제명, 저자명, 시소러스, 텍소노미 등을 포함하는 전거 데이터(authority data)가 있다. 현재 기존의 도서관에서 관리하는 데이터를 링크드 데이터로 변환하여 발행하는 해외 사례와 데이터 범주는 다음 <표 1>과 같다.

보통 도서관에서 전거데이터는 주제명, 저자명(개인, 단체명)을 지칭하지만 실제로 각 도서관에서 별도로 관리하고 있는 다양한 통제 어휘들이 존재하기 때문에 이를 링크드 데이터로 변환하여 발행하는 것도 중요한데, 특히, 전 세계 72,000 여개 도서관과 연계하여 목록, 전자자원 통합검색 등을 제공하는 OCLC(Online Computer Library Center)에서는 다양한 3가지 유형의 전거 데이터를 링크드 데이터로 구축하여 제공한다. OCLC에서 링크드 데이터로 구축한 국제적인 전거 통제를 목적으로 개발된

VIAF(Virtual International Authority File)의 경우 전거 데이터의 공유 증진을 통해 도서관의 목록 구축 비용 절감을 기대할 수 있다.

미국의회도서관에서는 2011년 ‘네트워크’라는 개념을 중심에 두고 다양한 형태의 콘텐츠를 수용할 수 있는 새로운 서지환경을 구축하고자 BIBFRAME(Bibliographic Framework)을 발표하였는데, BIBFRAME의 목적은 미래의 정보자원과 MARC21로 인코딩 된 전통적인 서지데이터를 모델링 할 수 있는 패턴을 제공하는 것이다. BIBFRAME은 링크드 데이터를 기반으로 한 도서관 간에 데이터 교환이 가능한 서지모델을 지향하고 있다.

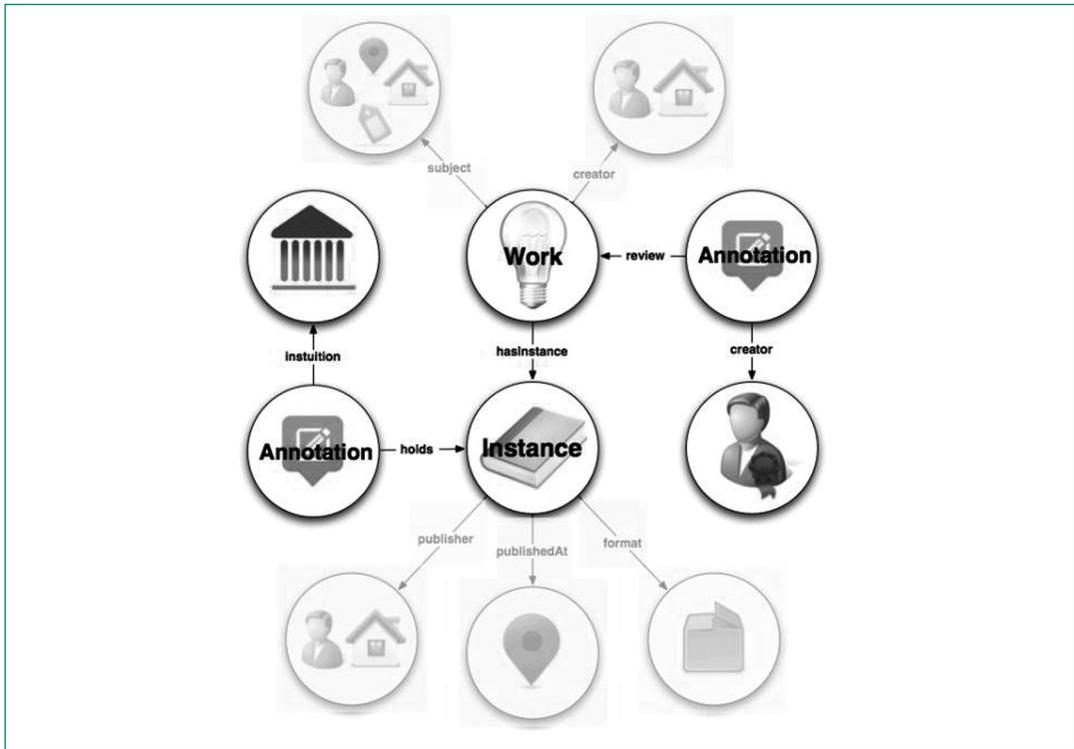
표 1 ▶ 해외 도서관의 링크드 데이터 구축 현황

기관명	데이터 범주	접근 URL
영국국립도서관	서지	http://bnb.data.bl.uk/
스웨덴 종합목록	서지/전거	http://data.libris.kb.se/
독일국립도서관	서지/전거	http://www.dnb.de/EN/lds
네이처	서지	http://data.nature.com/
OCLC	전거	VIAF, DDC, FAST
미국의회도서관	전거	http://id.loc.gov/
Lobid	서지/단체/조직	http://lobid.org/
유로피아나	전거	http://data.europeana.eu/

표 2 ▶ OCLC 링크드 데이터 자원

대상	특징
VIAF (Virtual International Authority File)	<ul style="list-style-type: none"> <li>• URL : http://viaf.org</li> <li>• 제공형태 : HTML, RDF/XML</li> </ul>
DDC (Dewey Decimal Classification)	<ul style="list-style-type: none"> <li>• URL : http://dewey.info</li> <li>• 제공형태 : HTML/RDFa, RDF/XML, Turtle, JSON-SPARQL</li> </ul>
FAST (Facet Application of Subject)	<ul style="list-style-type: none"> <li>• URL : id.worldcat.org/fast</li> <li>• 제공형태 : HTML, RDF/XML – Download</li> </ul>

그림 4 ▶ BIBFRAME 링크드 데이터 모델



### 3. KISTI 링크드 데이터 구축 사례

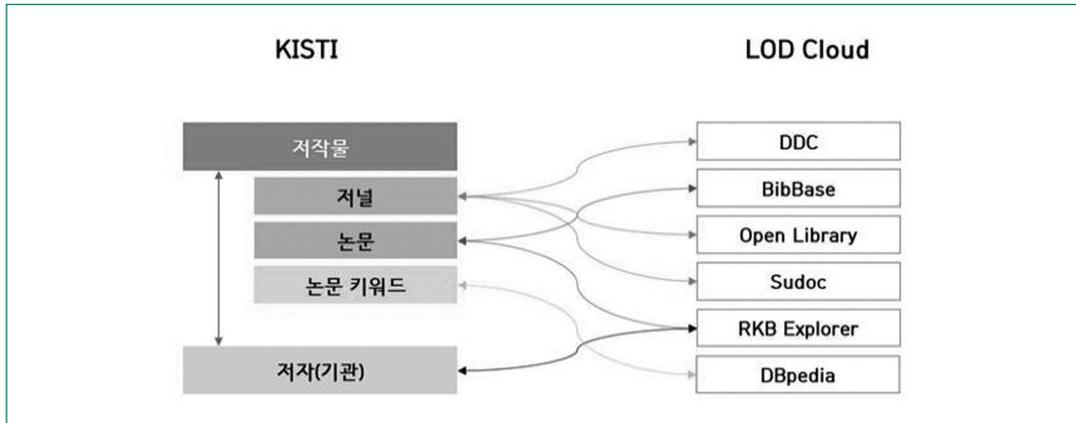
KISTI는 2007년부터 NDSL 웹사이트를 통해 제공하고 있는 논문, 특허, 보고서, 표준, 동향 등의 과학기술정보를 OpenAPI와 OAI-PMH 형태로 개방하여 공유하고 있다. 2013년부터는 공유 데이터의 활용성을 높이기 위해 기존 OpenAPI를 중심으로 한 실시간 정보 공유방식에서 벗어나 데이터 수준으로 정보를 개방하고, 개방된 데이터 간의 의미적 연결을 통해 데이터 기반의 지식 생태계를 조성하고자 국내 학술정보를 대상으로 링크드 데이터 서비스를 베타서비스 형태로 제공하고 있다.

NDSL 링크드 데이터는 최근 3년간 발행된 국

내학술지 770종에 수록된 논문과 저널, 저자, 기관 정보를 대상으로 서지 및 전거데이터를 링크드 데이터로 구축하였다. 또한, 구축된 데이터의 활용성을 높이고자 링크드 데이터 클라우드의 학술 데이터셋을 대상으로 논문, 저널, 저자, 기관, 참고문헌 등에 대한 인터리킹을 시도하였다.

NDSL 링크드 데이터 스키마 설계를 위해 외부 데이터와 연결이 용이하도록 과학기술정보의 특성을 잘 반영하고 있도록 3가지의 어휘집(vocabulary)을 사용하였는데, 서지 정보를 설계를 위해 DC(Dublin Core)와 PRISM(Publishing Requirements for Industry Standard Metadata)을 활용하였고, 저자 정보 설계에는 사람 혹은 기관을 정의할 때 사용되는

그림 5 ▶ 국내논문과 외부 데이터셋간의 인터링킹



어휘집인 FOAF(Friend of A Friend)를 활용하였다. 참고문헌 정보는 Bibtex를 활용하여 스키마를 정의하였다. <그림 6>은 NDSL 학술정보 링크드 데이터 스키마 설계에서 각 개체간의 관계를 나타낸 개념도이다.

NDSL 링크드 데이터 서비스는 다음 페이지 <그림 7>에서 보듯이 기존의 RDB 형태의 데이터를 RDF 형태의 트리플 데이터로 변환하고, 이것을 트리플 저장소에 저장한다. 저장된 트리플 데이터는

링크드 데이터 질의 언어인 SPARQL Endpoint를 통해 검색이 가능하다.

베타서비스로 제공되는 NDSL 링크드 데이터 웹사이트에서는, 기본적인 RDF 검색, 다운로드 기능을 제공하고 있으며, 자원간의 관계를 그래프 형태로 보여주는 오픈소스 프로그램인 RefFinder, 특정 URI의 연결정보를 탐색할 수 있는 LODLive 서비스를 응용서비스로 제공하고 있다.

그림 6 ▶ NDSL 학술정보 링크드 데이터 모델링

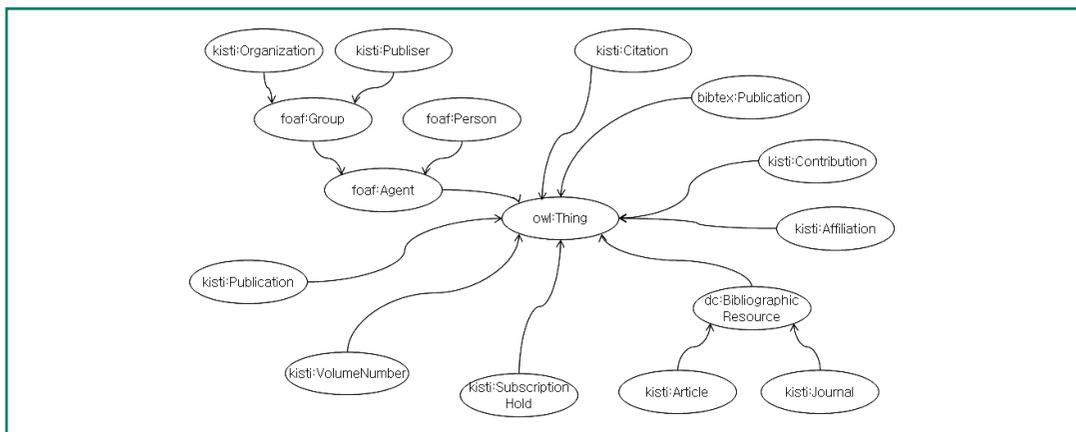


그림 7 ▶ NDSL 링크드 데이터 구축 체계

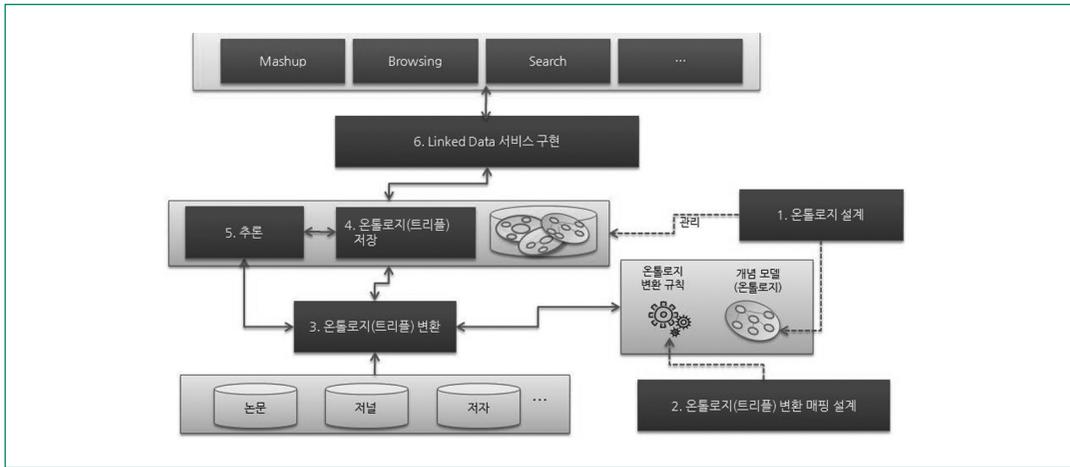


그림 8 ▶ NDSL LOD 베타서비스(<http://lod.ndsl.kr>)

**Linked Open Data  
for NDSL**

About    Search    SPARQL Endpoint    Application

Linked Open Data  
Beta Service

**NDSL LOD란?**

**NDSL LOD**

개방형 고부가 콘텐츠 기반의 과학기술정보 공유 모형 연구의 일환으로 시범적으로 구축, 발행된 트리플을 포함하고 있습니다.

< 자세히 보기 >

**SPARQL Endpoint**

SPARQL Endpoint를 이용해 보세요.

▶ NDSL LOD에 대한 SPARQL 질의를 지원합니다.

< 자세히 보기 >

**Application**

NDSL LOD에서 제공하는 데이터를 시각적으로 확인해 보실 수 있습니다.

< 자세히 보기 >

## 4. 결론

링크드 데이터는 웹에서 데이터를 의미있게 연결하고 공개하여 데이터의 활용성을 높이고자 한다는 측면에서 정보의 공동 활용을 촉진시킬 수 있는 중요한 기술이 될 수 있다.

또한 불필요한 데이터의 중복을 방지하여 전체적인 관리 비용을 절감할 수 있고, 핵심 데이터 외에는 외부 데이터를 연계하므로 네트워크 효과에

의한 데이터의 부가가치 생성이 가능하다. 그리고 RDF, URI 등의 표준을 활용하여 데이터를 표현하기 때문에 데이터의 표준화가 가능하고, 데이터간의 상호운용성이 높아져 데이터 통합이 수월해지는 장점이 있다. 본 고를 통해 그 동안 도서관이 소홀히 해왔던 데이터의 공유와 개방이 지식생태계 활성화 측면에서 중요한 의의가 있음을 인식하고 고민해 보는 계기가 되기를 희망한다.

### ● 참고 문헌

- [1] 한국과학기술정보연구원(2014). 도서관 분야의 링크드 데이터 구축 동향. 지식리포트 40호, 서울 : 한국과학기술정보연구원.
- [2] 한국과학기술정보연구원(2014). 링크드 데이터 기반 학술정보 콘텐츠 구축 및 서비스 제공 방안. 서울 : 한국과학기술정보연구원.
- [3] W3C. (n.d.). W3C SEMANTIC WEB ACTIVITY. Retrieved from <http://www.w3.org/2001/sw/>
- [4] 박진호 (2013a, 7월). 도서관은 웹에 존재하는가? : 링크드데이터, 글로벌 데이터베이스, 국가전자도서관 세미나. 경기도 : 국가기록원
- [5] Berners-Lee, T. (1998). Semantic web road map, September 1998. W3C Draft <http://www.w3.org/DesignIssues/Semantic.html>.
- [6] Berners-Lee, T. (2007). Giant global graph. online posting, Networks," Proceedings of the National Academy of Sciences of the United States of America, 98, 404-9.