

공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축에 관한 연구

A Study on National Linking System Implementation based on Linked Data for Public Data

윤소영 (So-Young Yoon)*

초 록

공공데이터는 공공기관의 정책 수립 등의 공적 기능을 수행하기 위하여 많은 비용을 들여 수집되고 있으나 대부분은 사일로(silos)로 방치되고 있다. 우리나라도 공공데이터를 개방하려는 움직임이 활발해지고 있으나 일부 기관을 중심으로 OpenAPI 형식으로 데이터를 제공하는 등 제한적인 수준에 머물고 있다. 국가차원에서 LOD(Linking Open Data) 시범사업으로 공공DB 피디아를 구축하면서 최근 들어 공공데이터 활용을 위한 링크드 데이터 구축이 활발하게 논의되고 있다. 본 연구에서는 관련연구 및 공공부문을 중심으로 한 링크드 데이터 구축 사례를 바탕으로 공공데이터 활용을 위한 국가 연계체계 사업인 공공DB 피디아 구축과정에서 도출된 문제를 파악하고 기존의 국가DB 연계체계 구축과정을 참고하여 그 해결 방안을 제시하였다.

ABSTRACT

Public information has been collected in various fields with huge costs in order to serve public purposes such as public agencies' policy-making. However, the collected public information has been overlooked as silos. In Korea, many attempts have been made to open the public information to the public only to result in limited extent, where OpenAPI data is being presented by some agencies. Recently, at the national level, the LOD(Linking Open Data) project has built the national DB, initiating the ground on which the linked data will be based for their active availability. This study has outlined overall problems in earlier projects which have built up national linking systems based on linked data for public data use. A possible solution has been proposed with a real experience of having set up an existing national DB of Korean public agencies.

키워드: 공공데이터, 링크드 데이터, 국가 연계체계, 도메인 온톨로지, 시맨틱 웹, LOD, 링크드 데이터 클라우드

public data, linked data, national linking system, domain ontology, semantic web, linked open data, linked data clouds

* 국사편찬위원회 사료연구위원(syoon@mest.go.kr)

■ 논문접수일자: 2013년 3월 15일 ■ 최초심사일자: 2013년 3월 16일 ■ 게재확정일자: 2013년 3월 22일
■ 정보관리학회지, 30(1), 259-284, 2013. [http://dx.doi.org/10.3743/KOSIM.2013.30.1.259]

1. 서론

정보가 경쟁력이 된 지금 시대에는 누가 얼마나 많은 데이터를 가지고 있느냐보다 데이터 사일로에서 얼마나 의미 있는 정보를 잘 끌어내어 공개하고 얼마나 잘 활용시키느냐가 중요한 이슈가 되었다. 링크드 데이터는 데이터를 의미 있게 연결하고 공개하여 데이터의 활용성을 높이고자 한다는 측면에서 Berners-Lee가 초창기 웹을 설계할 때 데이터 간의 연결을 목표로 하였던 웹의 기본 취지와도 잘 맞아떨어진다.

데이터 간의 연결이라는 웹의 초기 목적을 달성하기 위해 시맨틱 웹이 제안되었고 이를 실행하기 위한 실천 방안의 하나로 링크드 데이터가 대안으로 제시되고 있다. 링크드 데이터는 웹 페이지를 구성하는 데이터 간의 연결을 목표로 하여 웹 페이지가 아닌 각각의 자원(resource)을 중심으로 연결이 이루어져 데이터 중심의 웹이 됨을 강조하고 있다. 링크드 데이터로 실현되는 시맨틱 웹은 단지 데이터를 웹으로 제공하는 것이 아니라 데이터 간의 의미있는 링크를 형성하여 인간이나 기계 모두 데이터의 웹을 탐험할 수 있도록 해준다. 링크드 데이터를 통해 유용한 데이터를 얻게 되면, 그 데이터에 관계된 데이터로 계속되는 향해가 가능해진다.

공공데이터 개방을 통한 공공정보의 활용을 위해 이미 유럽연합, 미국, 영국 등 세계 각국에서는 법적, 제도적 기반을 마련하고 강력한 국가주도의 정책을 실행하여 사회·경제성장의 원동력으로 링크드 데이터를 적극 활용하고 있다. 시맨틱 웹과 링크드 데이터를 활용하면 데이터의 일관된 개방 형태와 연결을 지원할 수 있고 이를 기반으로 다양한 응용 기반을 구축하

여 해당 국가의 정보경쟁력 강화를 통해 국가 경쟁력 확보를 할 수 있을 것으로 기대된다. 데이터를 공개하는 기관 차원에서는 데이터 이용을 통한 홍보 및 가치 상승효과를 기대할 수 있고 데이터 이용 활성화 및 이용방식의 다양화를 통해 예산절감 효과도 기대할 수 있다.

초기의 공공데이터 개방을 통한 활용은 국가 및 공공영역에서 공공의 정보를 국민에게 개방함으로써 공공영역의 업무와 경영실태를 투명하게 운영하는지를 감시하는 차원이었다. 민간의 입장에서는 이를 감사 및 공개행정 입장에서 열람하고 활용하는 용도를 넘어 산업적으로 활용하는 방향으로 확대되는 재사용(reuse) 방안이 상대적으로 더욱 활발하게 진행되었다.

우리나라의 경우 민간의 급증하는 활용수요에도 불구하고 공공 영역에 축적되어 있는 데이터의 제공이 아닌 대국민 활용을 위한 서비스 수준에 머무르고 있다. 이는 공공영역에서 생산하는 공공데이터의 양이 충분하지 못해서가 아니라 여러 가지 제약으로 인해 공공기관이 이를 적극적으로 공개하지 못하고 있기 때문이다. 그럼에도 공공데이터 활용과 개방을 위해 몇몇 기관들을 중심으로 시맨틱 웹검색 서비스 구축이나 LOD(Linked Open Data) 프로젝트 등이 진행되고 있다. 국가차원에서도 한국정보화진흥원이 LOD 시범사업을 진행하여 국가DB 피디아(현, 공공DB 피디아)를 구축하고 공공데이터포털을 통해 서비스 하는 등 작업이 한창 진행 중이다. 그러나 개별 기관의 사업은 각 기관의 정책이나 필요에 따라 제한적으로 진행되고 있고 공공DB 피디아도 각 기관의 링크드 데이터 서비스 기반이 마련되지 않은 상황에서 그 가능성을 점검해보기 위한 아직 초기 단계의 프

로토타입 수준에 머무르고 있다.

본 연구에서는 공공데이터 활용을 위해 공공 부문을 중심으로 한 링크드 데이터 구축 사례 및 관련연구를 바탕으로, 공공DB 피디아 LOD 시범사업을 분석하여 문제점을 도출하고 국가 지식포털 등 기존 국가DB 연계체계 구축사례를 참고하여 공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축을 위한 발전방안을 제시하고자 한다.

2. 링크드 데이터 구축

2.1 링크드 데이터

위키피디아에서는 링크드 데이터를 “URI (Unified Resource Identifier)와 RDF(Resource Description Framework)를 이용해 시맨틱 웹 상에 올려있는 데이터, 정보, 지식을 노출하고 공유하며 연결하기 위해 추천되는 최고의 방법을 설명하는 용어”라고 명시하고 있다.

링크드 데이터를 통해 관련된 데이터를 서로 연결함으로써 다음과 같은 장점을 얻을 수 있다. 첫째, 데이터를 URI, RDF, HTTP를 통해 연결하여 사용할 수 있으므로 생산해낸 데이터가 아니라도 링크드 데이터로 연결되어 있는 데이터를 하나의 지식베이스처럼 사용할 수 있다. 둘째, 링크드 데이터를 통해 공개된 데이터를 이용하면 자신이 원하는 데이터가 이미 존재하는지, 어디에 존재하는지 알 수 있으므로 시스템의 사일로 문제에 의해 발생된 불필요한 데이터 중복의 문제를 해결할 수 있다. 셋째, 시맨틱 웹 표준인 RDF 형태의 데이터 발행으로 마치 하

나의 글로벌 데이터베이스처럼 질의하고 이용할 수 있으며, 이를 통해 상호운용성을 높이고 데이터 통합을 용이하게 할 수 있다. 넷째, URI로 구별되는 데이터 리소스의 자유로운 접근 및 이용이 가능하고 SPARQL Endpoint를 통해 SPARQL 질의가 가능하며 이를 응용 프로그램 상에서 이용할 수 있으므로, OpenAPI에 비해 데이터 접근을 더욱 구체화할 수 있어 데이터 지향의 매쉬업이 가능하다. 마지막으로, 데이터의 자유로운 연결과 이용은 새로운 데이터의 생산을 증가시키고 데이터가 서로 연결되는 네트워크화(network of data, cloud of data)가 진행된다.

링크드 데이터는 시맨틱 데이터의 웹으로, 접근가능하고 식별 가능하며 역참조할 수 있는 URI와 RDF로 기술된 트리플 구조를 이용하여 시맨틱 데이터를 기술하고 HTTP URI를 사용하여 서로 관계를 맺어주도록 하고 있다. 이때 공개된 링크드 데이터의 활용성을 보장하기 위해서는 데이터와 함께 그 의미하는 바도 함께 제공되어야 한다. 이를 위해 URI의 namespaces를 사용하여 의미 간 충돌을 해소하는 방법과 온톨로지 정보를 담고있는 XML 형식의 RDF 제공으로 링크드 데이터 구축을 제안하고 있다. 즉 데이터의 활용성 및 재사용성을 위해서는 데이터의 의미를 함께 제공할 수 있도록 온톨로지 구축이 선행되어야 하고 온톨로지에 기반한 링크드 데이터의 발행이 이루어져야 한다.

박옥남(2012)은 한국형 링크드 데이터 클라우드를 위한 단계로 국가기록원 전자 데이터 셋의 링크드 데이터화를 위해서 온톨로지 모델링을 거쳐 링크드 데이터를 구축하는 방법을 채택하였다. 노영희(2012)는 한국교육학술정보원

dCollection의 링크드 데이터 구축을 위해 학위 논문 데이터를 대상으로 기존에 구축된 도메인 온톨로지 및 도메인 사전을 참조하여 시맨틱 태깅을 수행하고 온톨로지 변환규칙을 적용하여 인스턴스화를 수행한 후 트리플로 저장하여 링크드 데이터를 발행하는 실험적인 연구를 수행하였다.

조명대(2010)는 도서관에서의 링크드 데이터 활용방안을 위한 모델로 기존 도서관의 서지 사항 및 전거데이터 그리고 메타데이터를 이용하는 방법과 토폭맵 기반 온톨로지 모델링을 거쳐 구축하는 방법을 제안하였다. 박지영(2012)은 서지 정보의 확장을 위해 링크드 데이터를 이용한 연구에서 서지 온톨로지를 이용하여 링크드 데이터와 결합시켜 서지데이터를 확장시키는 방식으로 연구를 진행하였다.

본 연구에서도 공공기관을 중심으로 한 공공 데이터 개방 및 활용을 위한 국가 연계체계 구축을 위한 방법으로, 데이터 자원(resource)과 그 의미도 함께 제공하여 데이터의 재사용성을 보장하는 온톨로지 기반 링크드 데이터 구축을 제안한다.

2.2 링크드 데이터 기반 기술

팀 버너스리(Bernes-Lee, 2006)가 시맨틱 웹 환경을 조성하기 위한 전략으로 링크드 데이터를 제안한 이래로 LOD 프로젝트(Linked Open Data Project)가 급속히 확산되고 있다. 2009년 미국 정부가 데이터를 국민에게 공개하는 공공 데이터 개방사이트(<http://data.gov>)를 개설하고 뒤이어 2010년 영국도 이와 유사한 공공데이터 개방사이트(<http://data.gov.uk>)를 개설하

면서 링크드 데이터를 이용한 정부의 공공데이터 공개가 확산되었고 기술적인 수요도 급속히 증가하는 추세이다.

BBC의 링크드 데이터 어플리케이션의 성공적인 적용에 힘입어 뉴욕타임즈가 링크드 데이터 프로젝트를, 미의회도서관이 LCSH의 링크드 데이터화를, 그리고 미국의 전자제품 소매회사인 베스트바이가 GoodRelations 온톨로지를 이용해 제품 정보를 링크드 데이터 형식으로 발행하는 등 다양한 분야의 적용 사례가 발표되고 있다. 공개된 데이터를 이용한 다양한 어플리케이션들도 등장하게 되었는데 웹 기반의 어플리케이션뿐만 아니라 DBpedia Mobile 등과 같이 링크드 데이터를 통해 공개된 데이터들을 서로 연계, 활용한 위치기반 스마트폰 어플리케이션들도 등장하였다. 이러한 어플리케이션을 구축하기 위해서 링크드 데이터를 위한 기반 솔루션이 필요한데, 미국, 유럽의 시맨틱 웹 회사들은 2007년 이후부터 링크드 데이터에 관심을 가지고 링크드 데이터의 발행, 저장 및 관리, 제공을 위한 솔루션을 연구하기 시작하여 2009년을 기점으로 링크드 데이터 솔루션을 대거 발표하고 있다. 링크드 데이터 구축을 위해 필요한 기술 및 도구로는 크게 링크드 데이터 발행 플랫폼/프레임워크(Linked Data Publishing Platforms/Frameworks), 링크드 데이터/RDF 편집기 및 유효성검증기(Linked Data/RDF Editors and Validators), 링크드 데이터 이용을 위한 브라우저 및 클라이언트 라이브러리와 검색엔진(Tools for Consuming Linked Data), 그리고 최종 이용자를 위한 링크드 데이터 어플리케이션(Linked Data Applications for End Users) 등이다("Linked Data

Community” [Online]). 가장 대표적인 솔루션으로는 DBpedia의 링크드 데이터 제공 시스템으로 유명한 OpenLink사의 Virtuoso와 영국의 data.gov.uk 서비스와 링크드 데이터 어플리케이션인 BBC Music 서비스 구축에 도입된 영국 시맨틱 솔루션 회사 Talis의 링크드 데이터 솔루션 Talis Platform이 유명하다. 독일의 Ontotext사는 LOD 데이터 셋 중 생명공학분야의 UniProt, PubMed, EntrezGene 등의 데이터 셋 40억개 데이터에 대한 SPARQL Endpoint 및 검색 인터페이스를 제공하는 linked life data (<http://www.linkedlifedata.com>) 서비스를 운영하고 있다. 미국의 Franz사의 대표적인 상용 시맨틱 웹 데이터 저장소(트리플 저장소)인 AllegroGraph, 브라우저인 AGWebView 및 Graph 기반 브라우저를 제공하는 뷰어인 Gruff를 패키지화한 솔루션은 DBpedia 독일 사이트를 구축하는 데 이용되었다. TopQuadrant사의 TopBraid Composer는 온톨로지 모델링 도구로 시맨틱 웹 모델을 개발하고 시맨틱 웹 어플리케이션을 개발하기 위해 사용된다. TopQuadrant Korea에서 개발한 OntoBase 2.0은 RDF 트리플 구조의 그래프 모델을 위한 전용 저장소이다.

KISTI가 개발한 OntoFrame은 온톨로지 기반의 시맨틱 웹 기술 적용을 위한 시맨틱 웹 프레임워크로 개체 식별을 위한 엔진인 OntoURI와 온톨로지 저장소이자 추론엔진인 OntoReasoner로 구성되어 있다.

현재 LOD 클라우드에 등록되어 있는 LOD 커뮤니티 프로젝트와 그 외의 단독 혹은 기관에 의해 수행되고 있는 프로젝트의 수는 70여개에 이를 정도로 활발하다(“CKAN”, [Online]).

3. 공공데이터 개방

정보의 축적과 가공, 전달이 쉬워져 공공정보가 지식정보의 원천으로 재인식되기 시작하면서 공개된 공공정보를 이용하여 개발된 스마트폰 앱이 폭발적인 인기를 누리는 등 공공데이터의 수요와 활용 가능성이 검증되고 있다. 이와 같이 공공데이터는 민간에 제공되어 새로운 상품과 서비스를 개발하는 기초자료로 이용되는 등 다양하게 활용되고 있다. 이를 통해 행정기관은 업무의 부담을 덜고 고품질의 행정서비스를 제공할 수 있게 되며 국민은 다양한 정보와 서비스를 제공받을 수 있는 기회를 얻게 되며, 이는 산업진흥으로 이어져 국가경쟁력 강화로 이어지게 된다(최진원, 2012).

공공기관은 정책 수립과 같은 공적 기능을 수행하기 위하여 교통이나 물가, 범죄율 등의 정보가 필요하므로 많은 비용을 들여 여러 분야의 폭넓은 정보를 수집하고 있으나 목적을 달성하고 나면 사일로로 방치되는 것이 일반적인 모습이다. 그러나 인터넷과 디지털 기술의 발전, 웹 2.0의 OpenAPI와 매쉬업 등의 등장으로 공공정보의 새로운 활용 가능성이 열리게 되었다(이현정, 남영준, 2012). 현재 우리나라는 국가차원의 지원사업을 통해 공공데이터 활용을 위한 방안으로 2016년까지 100종 구축을 목표로 OpenAPI 사업이 진행되고 있다. 그러나 OpenAPI는 정의된 형식으로만 데이터에 접근이 가능하므로 기상예보 등과 같은 정형화된 데이터일 경우에는 효과가 크지만, 대부분의 정형화되어 있지 않은 데이터의 경우에는 원하는 형식의 조건을 구성해 질의하는 것이 불가능하기 때문에 이용할 수 있는 데이터의 범위에 한계가 있고 이를

해결하기 위해서는 지속적인 업데이트가 수행되어야 하는 등의 불편함이 있다. 공공데이터의 링크드 데이터 구축으로 제한된 정보 접근의 한계를 극복하고 데이터 중복도 해소시킬 수 있다. 또한 접근할 수 있는 데이터의 내용이 구체적이므로 활용성 및 사용성 또한 높아질 수 있게 된다.

공공데이터의 효과적이고 성공적인 공개를 위한 전략방안으로 시맨틱 기술과 함께하는 링크드 데이터기술에 대한 초기부터의 집중적인 투자와 개발이 제안되고 있다(이상윤, 윤희주, 2012). 링크드 데이터를 활용하여 새로운 가치를 창출하려는 민간 사업자 및 개인들은 원하는 데이터들을 찾을 수 있는 방법을 필요로 하게 된다. 또한 링크드 데이터를 제공하는 정부 및 기관들도 제공하는 데이터들이 잘 활용되고 링크드 데이터의 장점인 관계성을 확장해가길 원할 것이다. 이런 차원에서 링크드 데이터를 생산, 확장, 소비하는 다양한 사용자들의 요구를 만족시킬 수 있는 플랫폼을 제공하기 위한 노력이 세계적으로 국가차원, 공공부문, 그리고 민간 차원에서 활발히 진행되고 있다. 이를 통해 사용자들이 보다 적극적으로 공공 데이터를 활용하여 부가가치를 창출할 수 있게 돕고, 데이터를 제공하는 조직 및 기관들 역시 링크드 데이터 포맷의 데이터를 통해 자연스럽게 데이터 간의 연결이 확장될 수 있도록 지원하는 데이터 생산, 활용의 선순환 구조를 만들어 내고 있다.

3.1 공공분야 링크드 데이터

공공분야의 대표적인 사례는 미국과 영국으로, 정부 분야의 공공데이터를 링크드 데이터화

하는 작업을 진행하고 있다. 정부의 투명성 증진과 국민의 알권리를 충족시켜 주는 효과에 더해, 정부가 생산, 관리하는 정보를 통해 보다 창의적인 서비스와 어플리케이션 개발을 위해 전세계 정부는 공공 데이터의 개방에 심혈을 기울이고 있다. 또한 데이터의 재사용성 증진과 함께, 어플리케이션에서 보다 유연하게 활용할 수 있도록 하기 위해 링크드 데이터를 적극 도입하고 있다. 미국의 data.gov에 의하면 현재 공공정보 개방에 참여하고 있는 국가는 20여개에 달한다.

미국 정부는 연방정부를 더욱 투명하고, 참여적이며, 협력적으로 만들기 위해 오바마 정부 출범 직후 Open Government Initiative를 시작하였다. 이를 위해 차세대 웹 기술을 이용하여 국민이 정책 결정에 참여할 수 있도록 프로세스를 구축(regulation.gov)하고 공공정보(non-personal information)를 이용할 수 있도록 공공정보 개방사이트(data.gov)를 2009년 5월 시작하였다. 이를 통해 연방 정부의 데이터에 대한 접근을 제고하고 이에 따라 혁신적인 아이디어를 장려함으로써 정부의 벽을 뛰어넘어 데이터의 창조적 이용을 확대하고자 하였다. 팀 버너스리(Tim Berners-Lee)와 함께 시맨틱 웹의 창시자인 짐 헨들러(Jim Hendler) 교수의 Tetherless World 팀이 프로젝트를 수행하였으며, data.gov의 일부를 RDF로 변환하여 링크드 데이터 클라우드(Linked Data Cloud)에 합류하고자 하였다.

영국 정부는 정책의 투명성 제고, 국민의 알권리 향상, 공공정보의 사회경제적 가치 증대, 그리고 국가 인지도를 높이기 위해 링크드 데이터로 공공정보를 개방하고 있다. 2009년 2월 '정

보의 힘(Power of Information, POI)' 보고서에서 공공부문의 정보공유 및 활용에 따른 가치 창출을 위해 공공정보에 접근하고 활용할 수 있도록 해야 하고, 국민의 창의적 혁신을 지원하고, 개인과 단체가 삶의 질을 개선할 수 있도록 더 나은 정보에 기반한 판단을 통해 솔루션을 만들어 낼 수 있도록 도와야 한다고 주장하였다. 이를 기반으로 2010년 1월 21일 공공정보 개방 사이트(www.data.gov.uk) 서비스를 시작하였으며 3,000개 이상의 링크드 데이터 셋을 서비스하고 있다. 링크드 데이터를 이용하여 다양한 웹 기반, 또는 스마트폰 기반의 개방형 어플리케이션들이 등록되고 있으며, 산업계와 연구 기관 간의 협력을 위한 허브로서의 역할을 수행하고 있다. 또한 링크드 데이터의 재활용 및 창의적 이용을 위해 레퍼런스, 교육, 교통, 육지 측량부(ordnance survey)의 4가지 데이터 자원에 대한 SPARQL Endpoint를 제공하고 있다.

핀란드는 기업들이 서로 협력하여 지능적인 서비스를 만들 수 있는 환경을 제공하기 위해 시맨틱 웹 온톨로지 인프라 구조를 국가적인 차원에서 개발하려는 프로젝트 FinnONTO를 진행하였다. 그 중에 ONKI는 온톨로지 기반의 메타데이터를 구축하여 웹상의 콘텐츠와 서비스를 연결하기 위한 프로젝트인데 2008년 9월부터 3개월 동안 36,000명의 방문자와 104,000번의 방문수 등 이용자 참여율이 높은 대표적 시맨틱 웹서비스로 역사, 문화, 자연과학, 의학 정보, 경영 정보, 지리 정보 등의 분야에서 약 40여 개의 도메인을 통해 서비스되고 있으며 무료로 이용할 수 있는 어플리케이션을 제공한다. Health Finland는 시민들의 건강관리를 위한 포털로 온톨로지와 메타데이터를 이용하여 다

이어트, 운동, 금연, 건강 생활 등에 관한 콘텐츠를 제공하고 있다(오원석, 2009, "FinnONTO", [Online]).

3.2 언론, 방송분야 링크드 데이터

언론, 방송 분야의 링크드 데이터 서비스는 DBpedia를 중심으로 활발히 진행되고 있는데, BBC, CNN, NewYork Times 등의 주요 방송 및 언론사뿐 아니라 MusicBrainz 등의 음악 데이터베이스와 연계하는 등 링크드 데이터 서비스 체계가 확산되고 있다.

BBC는 다양한 국가와 지역의 라디오 방송과 TV 방송을 운영하고 있는데, 링크드 데이터를 적용하기 이전에는 각 지사와 방송 채널들이 개별적으로 사이트를 운영하고 정보를 관리하였다. BBC가 생산하여 제공하는 정보는 다양한 채널로 다양한 지역을 통해 제공되는데 이러한 정보들은 해당 채널별, 지역별로 분리되어 운영됨으로써 데이터 중복의 문제와 데이터 관리 및 운영비용의 과다 지출을 초래하였다. 또한 데이터가 서로 공유, 연계되지 못함으로써 관련 있는 데이터 간의 단절 상황이 가속화되었다. 데이터 연결의 중요성을 인식하게 되어 다양한 분야의 라디오 및 TV 채널 각 도메인 간의 정보 융합이 BBC의 가장 큰 현안으로 부각되었다. 이러한 문제를 해결하기 위해 BBC는 데이터 분야와 서비스 분야를 분리하고 링크드 데이터를 통해 데이터를 공유, 연계함으로써 데이터를 재활용하고 동적인 제어를 통해 BBC News, BBC Music, BBC Sports 등에서 링크드 데이터를 콘텐츠 생산에 활용하고 있다(Bartlett, 2013).

BBC가 링크드 데이터를 적용하며 얻게 된

혜택은 크게 2가지로 나눌 수 있는데, 그 하나는 데이터의 재활용성 증진이며 다른 하나는 사용자의 참여이다. 데이터의 재활용성을 통해서 한 번 생산한 정보를 또 다른 도메인, 또 다른 사이트, 또 다른 방송에서 활용하는 것이다. BBC Programmes와 Music, BBC News는 링크드 데이터 어플리케이션을 이용하여 정적인 데이터의 조합인 HTML 형식의 웹사이트를 동적인 웹사이트로 전환하고 있다. 이 어플리케이션은 BBC Programmes Ontology, Music Ontology 등을 이용하여 구조화된 자체 데이터와 대중음악 도메인인 MusicBrainz, DBpedia 등이 제공하는 링크드 데이터 형식의 외부 데이터를 융복합할 수 있도록 한다(BBC programmes, [Online]). BBC는 자신의 데이터에 대한 원본 데이터 공급자(raw data Provider)와 링크드 데이터 공급자로 역할을 할 뿐만 아니라, DBpedia를 통해 원본 데이터 공급자인 위키피디아의 모든 데이터에 대한 링크드 데이터 어플리케이션도 공급하고 있다. 이를 통해 콘텐츠 소비자뿐만 아니라 외부의 서비스 개발자에게 적절한 콘텍스트의 의미 있는 콘텐츠를 소비할 수 있도록 해줌으로써 콘텐츠의 부가가치를 높이고 있다.

미국의 뉴욕타임즈(New York Times)는 지난 150년 동안 기록이나 문헌의 내용을 표현하는

색인어-주제형 표목을 정리한 The New York Times Index를 발행하고 있다. 이 색인을 통해 특정 주제 아래 관련 기사가 오랫동안 쌓이면서 그 가치도 커지게 되었다. 뉴욕타임즈는 지난 2009년 Topic이라는 이름의 링크드 데이터 프로젝트를 통해 이를 공개하였다. Topic에는 약 3만 건의 키워드가 등록되어 있는데 이 프로젝트의 핵심은 기사에 포함되어 있지 않은 다양한 외부정보를 자유롭게 연동해 쓸 수 있도록 Freebase, DBpedia, 그리고 GeoNames와 정보를 연계하는 것이다. 2009년 10월에 인명 주제명 표목 약 5,000건을 공개하고 이어 2010년 1월에 기관명, 기업명, 지명 등 약 5,000건의 주제형 표목을 링크드 데이터로 발행하였다(〈표 1〉 참조). 뉴욕타임즈가 공개한 링크드 데이터는 OWL, SKOS, DC(Dublin Core), 그리고 CC(Creative Commons)의 내용을 담고 있는 RDF/XML, HTML 형태로 제공되고 있으며 출처만 밝히면 어떤 용도로든 조건없이 사용할 수 있다. 또한 개발자 네트워크를 통해 기사검색 API를 공개하고, SPARQL 사용법, API 사용법 등과 같은 정보를 공유하고 있으며 API를 활용한 매쉬업과 외부 응용 어플리케이션 개발을 장려하고 있다("NewYork Times Linked Data" [Online]).

〈표 1〉 뉴욕타임즈 링크드 데이터 공개 현황(2010.1.13 기준)

유형	수작업 연계	자동 연계	합계
인명	4,978	0	4,978
조직	1,489	1,592	3,081
위치	1,910	0	1,910
디스크립터	498	0	498
			10,467

〈<http://data.nytimes.com>〉

MusicBrainz는 음반 데이터베이스로 오디오 CD와 MP3 등의 디지털 음악 트랙을 포함하여 곡에 대한 메타데이터(음악가, 곡명, 앨범명, 트랙번호, 출시일, 오디오 트랙번호, CD Disc ID 등)를 제공하는 등 포괄적인 음악 백과사전을 목표로 하고 있다. 메타데이터 표현방법에 대한 공식적인 가이드라인을 제공하여 누구나 자유롭게 자료를 채우고 수정할 수 있는데, 간단한 데이터는 사용자들이 바로 수정하도록 하고 음반을 새로 추가하는 등의 편집은 사용자들의 투표를 통해 반영하도록 하고 있다. MusicBrainz의 음반 정보는 퍼블릭 도메인으로 공개되며 나머지 정보는 CC-BY-NC-SA(저작자 표시-비영리-동일조건 변경 허락)으로 공개되어 있다. LinkedBrainz는 MusicBrainz Database를 링크드 데이터로 발행하기 위한 프로젝트로, 연계를 통해 음악 정보 공동 활용에 앞장하고 있다. 음악가, 앨범, 트랙 정보 등이 대상이며 이러한 정보들을 링크드 데이터로 발행하여 음악 관련 커뮤니티나 정보 서비스 사이트, 다양한 어플리케이션에서 활용할 수 있도록 하고 있다("MusicBrainz" [Online]). LinkedBrainz는 Music Ontology와 함께 Event Ontology, Timeline Ontology 등 외부 온톨로지를 연계하여 사용하고 있다. Music Ontology는 음악의 작곡부터 작곡된 곡을 연주하고 그 연주를 녹음하여 음반으로 나오기까지의 과정과 사용된 악기들, 연주자들, 지휘자, 작곡자, 편곡자 등 다양한 음악 관련 정보를 표현할 수 있다. Music Ontology는 음악과 관련된 세밀한 정보들까지도 모두 표현을 할 수 있도록 FRBR 모형에 기반하여, 음악가, 앨범에 대한 기본적인 명세(레벨 1), 레벨 1에 더해 소개와 Event, Time 등이 확장된 명

세(레벨 2), 그리고 레벨 2보다 좀 더 확장된 명세(레벨 3)의 서로 다른 3가지 수준으로 데이터를 표현하고 있다("LinkedBrainz" [Online]).

3.3 도서관, 정보서비스 분야

도서관 분야에서의 정보 유통 체계는 내부적으로는 MARC를 중심으로 확고한 반면, 웹을 중심으로 유통 구조를 형성하고 있는 다른 분야와의 협력에는 한계를 보이고 있다. 이러한 제한을 극복하기 위한 여러 노력의 일환으로 최근에는 링크드 데이터를 이용해 정보를 공개하고 연계를 하려는 움직임이 활발하다. 도서관 분야에서의 링크드 데이터 접근의 특징은 개념적 접근과 기술적 접근이 동시에 진행되고 있다는 것이다.

미의회도서관은 LCSH를 RDF로 표현하여 링크드 데이터를 발행하고 있으며 MARC 21 포맷으로 되어 있는 전거 데이터를 SKOS 어휘(vocabulary)를 이용하여 RDF로 변환하고 이를 웹상에 발행하고 있다. 전거 레코드의 각 항목을 skos:Concept의 인스턴스로 표현하고 LCCN(Library Congress Control Number)을 이용하여 URI를 부여하고 있다(<http://id.loc.gov>).

VIAF(Virtual International Authority File)는 전 세계 전거 데이터의 공유를 통해 목록 비용의 절감과 함께 국제적인 전거 통제를 목적으로 OCLC에 의해 운영되고 있다. 전 세계 16개국이 참여하고 있으며, 1,300만개의 Name Records를 각 국가가 사용하는 언어로 표현하여 서비스하고 있다. Personal Names, Geographic, Corporate, Title, Family, Events 등으로 구성된 각 항목마다 URI를 부여하여 링크드 데이터로

이용 가능하도록 하고 있으며 RDF, HTML 형식 등을 지원하고 있다(<http://viaf.org>).

LIBRIS는 2008년 도서관 전체 데이터를 링크드 데이터로 구축한 최초의 종합목록으로 대학도서관, 연구도서관, 박물관, 아카이브, 공공도서관 등 170여개 기관이 이용하는 종합목록 시스템이다. 6백만 개의 서지 데이터 및 2천만 개의 도서관 장서목록, 2십만 개의 전자 데이터로 구성되어 있으며 목록과 함께 데이터 자동 반입/반출 기능 및 openAPI를 제공하고 있으며 링크드 데이터를 이용할 수 있도록 RDF 파일 형식을 지원하고 있다(<http://libris.kb.se>).

유로피아나는 유럽의 문화유산 컬렉션을 위한 서비스로, 도서관, 박물관, 기록관 등의 서로 다른 메타데이터를 활용하고 있는데 이런 데이터들은 상호 간에 연결이 필수적인데 이는 링크드 데이터를 이용하여 가능하다. 이러한 목적을 위해 EDM(Europeana Data Model)을 개발하여 유로피아나의 컨텍스트 사이의 차이를 이어주는 역할을 수행하고, 좀 더 풍부한 자원 검색 환경과 복잡한 데이터를 활용한 보다 개선된 디스플레이 변환이 가능하도록 하고 있다(<http://www.europeana.eu>).

생명과학과 더불어 현재 전 세계 LOD 클라우드에서 가장 활발한 분야는 지리정보 분야로, 문화정보 분야와 더불어 거의 모든 영역에서 지리정보를 응용한 서비스가 가능하기 때문에 융합 정보 생산의 중요한 위치를 담당하고 있다. 모바일 기기의 성장과 네트워크의 보편화로 그 중요성은 더욱 커지고 있다. 지리정보 분야에서 가장 두각을 나타내고 있는 곳은 GeoNames이다. GeoNames는 세계 각국의 지리정보를 생산, 운영하고 있으며 누구나 활용 가능하도록 개방

하고 있다(<http://www.geonames.org>).

DBpedia는 사용자 참여형 백과사전이라 불리는 위키피디아를 웹상에서 보다 정교하고, 구조적인 데이터로 활용할 수 있도록 지원하기 위한 서비스이다. DBpedia는 위키피디아의 데이터를 기반으로, 위키피디아 데이터를 보다 더 잘 찾고 응용 어플리케이션이나 외부 서비스에서 잘 연계할 수 있도록 지원한다. 이러한 목표를 위해 DBpedia는 시맨틱 웹 기술, 즉 구체적으로는 링크드 데이터 기술을 적용하고 있으며 “DBpedia is the Semantic Web mirror of Wikipedia”라고 설명하고 있다. DBpedia의 중요성은 지식베이스라는 측면에 있는데 이는 위키피디아의 중요성에서 시작된다. 지식베이스는 특정 서비스나 어플리케이션에서 원하는 목적을 달성하기 위해 디지털화된 정보를 구성하는 가장 핵심이 되는 자원이다. 기존의 지식베이스는 특정 도메인에 국한되어 구축되었고 이로 인해 도메인별 지식베이스는 해당 도메인만의 지식과 용어를 주로 사용하여 다른 도메인과 융합되고 협업되기 어려운 구조로 구축되었다. 위키피디아는 사용자의 참여를 통해 전문적인 지식이 축적될 수 있는 개방형 구조를 채택하여 도메인 간 용어의 불일치로 나타나는 문제, 특정 도메인 위주로 지식화되는 문제, 그리고 도메인 간 협업을 위한 장벽의 문제를 해소하고 있다. DBpedia는 다양한 도메인을 수용하고 커뮤니티의 합의에 의해서 축적된 위키피디아의 데이터를 제공하며, DBpedia Knowledge Extraction Framework을 통해 위키피디아의 데이터를 수집하고 추출하고 가공하고 있다.

DBpedia의 또 다른 중요성은 시맨틱 웹, 그 중에서도 링크드 데이터의 적용 측면에 있다.

DBpedia는 웹을 중심으로 보다 지능적이고 유연한 정보 유통과 소통을 위해 위키피디아의 인터페이스 혹은 허브 역할을 수행하고 있다. 기존의 사람 중심적인 표현과 활용을 위한 문서 웹(Document Web)에서 데이터 간 연계, 협업을 위한 Data Web, 즉 Web of Data로 패러다임 전환이 가속화되고 있는 시점에서 DBpedia의 역할은 Web of Data를 위한 허브 역할로 그 중요성이 더욱 부각되고 있다. Web of Data를 위한 DBpedia의 역할을 요약해 보면 다음과 같다.

- ① 역참조(Dereferenceable) 식별기호 정의
 - 모든 DBpedia 엔티티에 대하여 Web 중심의 역참조 식별기호 정의
 - Web of Data의 발달을 저해하는 개체 식별기호 누락문제를 극복하기 위한 노력
- ② RDF 링크 발행 및 링크 발행자 지원
 - DBpedia에서 다른 웹데이터 소스로서의 RDF 링크 발행
 - 다른 발행자들의 RDF 데이터에서 DBpedia로 연결을 설정할 수 있도록 지원
- ③ 위키피디아의 활용
 - 웹 중심의 사용자 참여형 백과사전인 위키피디아를 Web of Data를 위한 다양한 도메인 지식베이스로 변환
 - DBpedia Knowledge Extraction Framework을 통해 항상 최신의 위키피디아 정보를 DBpedia에서 Web of Data로 활용할 수 있도록 지원

DBpedia는 자원(resource) 생성 식별기호, 즉 URI를 생성하기 위해 위키피디아의 영어 이름을 사용하며 다른 언어버전의 경우는 해당 언어의 이름을 사용한다. 예를 들어 서울(Seoul)의

URI는 위키피디아에서는 “<http://en.wikipedia.org/wiki/Seoul>”이며 DBpedia에서는 “<http://dbpedia.org/resource/Seoul>”이다. DBpedia는 다양한 속성(property)을 통해 자원을 표현하는데 특정 자원에 대해 다른 언어의 정보가 존재하면 해당 언어로 된 간략 요약(short abstract)과 상세 요약(long abstract)을 동시에 제공하며, 해당 위키피디아 페이지 링크도 포함한다. 자원 개체는 약 20,000개의 클래스로 구성된 온톨로지를 기반으로 한 다양한 분류를 통해 표현된다.

DBpedia는 외부 어플리케이션이나 서비스에서 DBpedia에 접근하기 위한 다양한 방법을 제공하는데, SPARQL 프로토콜을 통해 DBpedia의 SPARQL Endpoint에 SPARQL 질의를 전송하고 결과를 반환받거나 N-Triple형태의 RDF 형식으로 다운받을 수도 있다.

3.4 국내 사례

국내의 링크드 데이터 서비스 구축이나 기술 보급이 상대적으로 느리게 진행되고 있다. 시맨틱 웹 기술을 이용한 검색이나 네트워크 분석 등의 사례나 링크드 데이터 관련 연구는 진행되고 있으나, 실제로 웹이라는 정보 유통 공간을 이용한 링크드 데이터 구축은 아직 초기 단계에 머무르고 있다. LOD 클라우드에 속해 있는 KISTI 사례와 bibleontology 사례가 가장 근접한 링크드 데이터 서비스에 해당한다고 볼 수 있다. KISTI는 사우스햄튼 대학의 RKBExplore (<http://www.rkbexplorer.com>)에 데이터를 제공하고 그 사이트를 통해 링크드 데이터를 서비스하고 있다. 또한 KISTI는 국내 사용자들

대상으로 논문, 특허, 보고서, 동향 정보 등을 서비스하고 있는 NDSL에 포함되어 있는 일부 데이터를 트리플로 변환한 데이터를 서비스하고 있다. 데이터는 논문 등의 발행물과 연구자 정보 등으로 구성되어 있다. bibleontology는 성경 속 인물과 사건에 대한 온톨로지를 설계하고 TopQuadrant의 링크드 데이터 솔루션을 이용하여 발행한 공개형 링크드 데이터 서비스이다 (<http://bibleontology.com>).

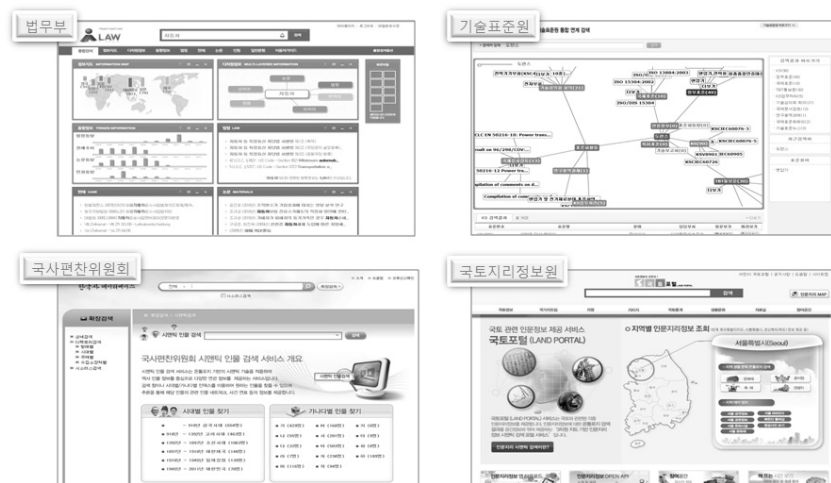
최근에는 공공기관에서 보유한 공공데이터를 민간에 공개하려는 움직임이 활발하게 이루어지고 있는데 서울시의 열린 데이터 광장(<http://data.seoul.go.kr>)이나 한국정보화진흥원의 공공데이터포털(<http://data.go.kr>)을 통해 다양한 데이터들이 API를 통해 공개되고 있다. 서울시는 공공데이터를 활용한 매쉬업 등 창의적인 어플리케이션을 직접 제작할 수 있도록 OpenAPI와 개발자 공간을 운영하고 있다. 서울시 공공데이터는 초기에는 시민생활과 밀접한 교통, 지리 정보부터 점차 서울시 보유 대다수 데이터를 개방하고 단계적으로 2014년까지 150종의 공공DB 개방을 목표로 하고 있다.

한국정보화진흥원은 공공데이터포털을 통해 국가가 보유하고 있는 다양한 공공정보를 국민에 개방하여 이를 편리하고 손쉽게 활용할 수 있도록 지원하고, 많은 자원들이 개방될 수 있도록 계획을 추진하고 있다. 공유서비스로 지정된 공유자원에는 정형데이터 및 비정형데이터를 직접 활용할 수 있도록 개방하는 데이터 공유자원 약 1천6백만건과 정보, 소프트웨어, 하드웨어, 개발환경 등을 직접 활용 또는 2차 가공 활용할 수 있도록 개방하는 OpenAPI 공유자원 241건이 있다. 한국정보화진흥원은 더 나아가 국가지능

형 연계체계 구축을 목표로 2012년 수행한 LOD 시범사업인 공공DB 피디아(<http://lod.data.go.kr>)를 통해 공공데이터를 대상으로 링크드 데이터를 시범적으로 구축하였다.

그 외에 시맨틱 웹 서비스를 제공하고 있는 대표적인 공공부문의 사례로는 국토지리정보원의 3차원 지도 기반 인문지리정보 시맨틱 검색 포털서비스인 국토포털이 있다(<http://www.land.go.kr/portal/main.do>). 인문지리정보의 지식기반(KB) 구축에 머물지 않고 지리적 위치를 기반으로 한 온톨로지를 중심으로 인문지리정보를 연계하고 통합하여 구축하였다. 온톨로지 구조를 통해 검색된 각종 인문지리정보(human-oriented geographic information)를 지오웹 플랫폼 기반의 3차원 공간정보상에 융·복합한 차세대 GIS 서비스체계를 구축하여 제공하고 있다. 스마트폰을 통해 사용자 위치 중심의 다양한 인문지리정보를 현장에서 서비스 받을 수 있도록 하고, 유관기관 및 민간에서 직접 다양한 목적에 맞추어 서비스를 개발하고자 할 경우 표준화된 개방형 연계서비스(OpenAPI)를 통하여 효과적으로 연계·활용할 수 있도록 제공하고 있다. 국토지리정보원은 지속적인 구축사업을 위해 '인문지리정보 통합기반 구축 작업규정'[국토지리정보원예규 제42호, 2012.7.11, 제정] 예규를 제정하여 시행하고 있다. 이 예규에는 인문지리정보 분류체계 표준, 통합DB 속성 정의서, RDF 데이터 형식, 온톨로지 구축 품질관리표(온톨로지 모델링), 온톨로지 구축 품질관리표(온톨로지 인스턴스), 그리고 온톨로지 구축 품질관리총괄표를 제시하는 등 작업을 상세하게 규정하고 있다.

국사편찬위원회는 역사용어 온톨로지 구축



〈그림 1〉 시맨틱 웹 검색서비스 기관 사례

을 통해 시맨틱 웹검색서비스를 단계별로 추진하고 있는데, 그 첫 단계로 검색 수요가 가장 높은 인물을 대상으로 인물 온톨로지를 구축하고 이를 기반으로 시맨틱 웹 검색서비스를 제공하고 있다. 이 서비스는 인물, 조직 및 단체, 사진, 지리 정보 등으로 대상을 확장해나갈 예정이며 데이터 재사용성을 높이기 위해 링크드 데이터로 데이터를 공개하기 위한 준비를 진행하고 있다(<http://db.history.go.kr>).

법무부의 아이로시스템은 핵심어에 대한 정보 지도, 다차원정보, 동향정보, 법령, 판례, 논문, 민원, 일반문헌 정보를 원스톱으로 이용할 수 있도록 제공하는 서비스이다. 다차원정보 서비스는 검색어와 관련이 있다고 선정한 관련어를 종합하여 시각적(온톨로지 맵)으로 제시하는 서비스로, 시스템이 시맨틱 추론을 이용하여 자동으로 용어 관계를 추출하여 제공하는 서비스와 용어 사전을 이용하여 개념적으로 상하관계를 제시하는 서비스로 구성되어 있다(<http://ilaw.go.kr>).

기술표준원 통합연계검색 서비스는 대상 정

보 간 연계와 추론을 통해 새로운 정보를 제시하는 차세대 검색기술을 이용하여 기술표준정보의 제공 범위와 활용성 향상을 목표로 하고 있다. 표준용어사전을 통해 검색어의 의미를 분석하고 검색어와 각 기술표준 정보를 연계함과 동시에 상호 간 관계를 가시화하여 제공하고 있다(<http://www.standard.go.kr>).

4. 링크드 데이터 국가 연계체계 분석 및 발전 방안

공공데이터 활용을 위한 체계를 구축하기 위한 노력이 몇몇 기관을 중심으로 진행되고 있으나 그 수가 매우 적고 독자적인 사업을 통해 정보 공개를 진행하고 있는 경우가 대부분이다. 기관 간 데이터 연계를 통한 데이터의 재사용성 및 활용성을 높이고자 하는 담당자의 의지가 있더라도 이를 실현시키기에는 현실적으로 기관 간 협조를 비롯하여 진행과정에 있어서 많은 제

약이 따른다. 일부 기관을 중심으로 협의를 통해 API를 이용한 데이터 연계가 진행되고 있으나 이는 앞서 언급한 OpenAPI의 한계를 그대로 지니고 있어 정보의 공공 개방 및 활용이라는 목적을 달성하기에는 충분하지 못하다. 이러한 상황은 전면적이고 완전한 정보 공개 방법을 찾아 공공데이터 개방을 진행하고자 하는 기관들이 협력하여 풀어나가야 할 과제이다.

공공데이터 개방과 활용을 위한 국가차원에서 노력은 국가지식포털에서부터 현재 공공데이터포털에 이르기까지 지속적으로 진행되고 있다. 이러한 노력의 일환으로 2012년에 한국정보화진흥원에서 '시맨틱 웹 기반의 국가지능형 연계체계 구축을 위한 사업'을 진행하여 LOD 시범서비스인 국가DB 피디아(현, 공공DB 피디아)를 시작하였다.

4.1 공공DB 피디아 분석

공공데이터의 개방과 활용을 위해 구축된 공공DB 피디아는 통계청, 국토해양부, 서울시, 교육과학기술부(한국교육학술정보원) 등 5개 기관과 네이버(영화정보)가 제공한 24개 데이터셋과 50,184개 자원을 대상으로 링크드 데이터를 서비스하고 있다. 데이터 제공 외에도 지역 기반의 링크드 데이터를 활용하여 개발 가능한 지역생활 정보 모바일 서비스 등 링크드 데이터를 활용한 서비스 사례를 제공하고 있다.

4.1.1 데이터 제공 기관 및 범위

공공DB 피디아 시범사업은 LOD의 가능성을 시험하는 프로토타입을 목표로 진행되었다. 따라서 기관과 공공DB 피디아가 시스템적으로

데이터를 연계하는 체계가 아니라, 기관에서 데이터를 제공하고 공공DB 피디아에서 온톨로지 모델링을 통해 링크드 데이터로 변환, 가공하여 사이트에서 공개하고 SPARQL Endpoint로 데이터를 제공하는 방식으로 구축되었다.

사업 초기에는 15개 영역을 대상으로 연계대상 DB를 선정하였으나 서비스 모델 후보 검토를 통해 대상 범위를 줄이고 데이터 변환 및 제공 동의가 이루어진 기관의 데이터를 중심으로 웹 공개 데이터를 추출하여 링크드 데이터 대상을 선정하였다. 지역 정보에 기반하여 연계될 수 있는 기관의 데이터를 중심으로 구성하였는데, 그 이유는 제한된 구축기간 내에 여러 기관에서 구축한 데이터들을 연결할 때 기준이 되는 명확한 연결고리로 지역이 가장 적절하고 데이터 활용성 및 향후의 확장성을 고려하였을 때에도 효과적이었기 때문이다. 따라서 기관에서 제공할 수 있는 데이터가 지역을 중심으로 연결될 수 없는 경우에는 일단 제외하였고 링크드 데이터로 가공, 변환이 용이한 구조화된 데이터, 즉 통계 데이터 위주로 대상 데이터를 구성하였다. 그 외에 민간영역의 데이터가 일부 참여하고 있는데 시범사업을 수행한 네이버가 저작권 문제가 해결된 영화정보를 제공하고 있다.

공공DB 피디아 사이트에서 서비스하고 있는 데이터의 주제별 구성은 다음과 같이 6개 분류로 나뉘어져 있다.

- 교육 - 유치원, 초중고교 정보, 사설학원수, 공공도서관 정보
- 통계 - 인구동향, 의료기관 정보, 폐수배출업소, 범죄발생건수
- 공공시설 - 공공체육시설, 도시공원, 문화기반시설

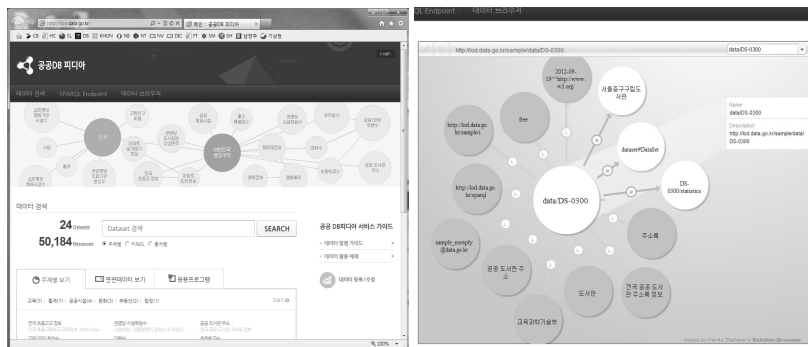
- 문화 - 문화재 정보, 영화정보
- 부동산 - 아파트 단지 정보 및 실거래가 정보
- 행정 - 행정구역 정보

대상이 된 DB는 지역을 중심으로 데이터 간 연계가 구성될 수 있는 지역별 교육정보, 도서관 정보, 부동산 정보, 지역별 통계정보 등이 주를 이루고 있으며 통계데이터와 같이 정형화된 형식의 데이터가 대부분을 차지하고 있다. 따라서 간단한 매핑이나 비교적 간단한 데이터 분석을 통해 링크드 데이터로 변환이 가능하였다. <그

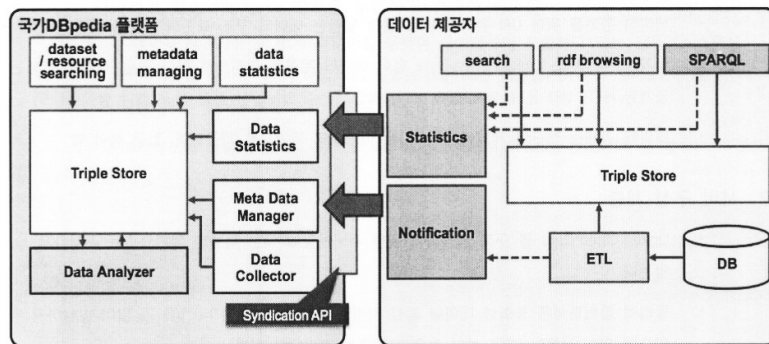
림 2>는 공공데이터포털을 통해 서비스하고 있는 공공DB 피디아의 메인화면과 데이터를 비주얼하게 보여주는 데이터 브라우저 화면이다.

4.1.2 연계방안 설계

링크드 데이터 구축 및 활용을 위한 공공DB 피디아와 데이터 제공자 간의 시스템 아키텍처는 <그림 3>과 같이 요약할 수 있다. 링크드 데이터를 구축하고 공공DB 피디아에 등록하는 역할은 데이터를 제공하는 공공기관이나 민간이 담당을 하고 공공DB 피디아는 등록된 메타 데이터를 대상으로 데이터를 활용하려는 사용



<그림 2> 공공데이터포털의 공공DB 피디아
<<http://lod.data.go.kr>>



<그림 3> 공공DB 피디아 시스템 아키텍처(한국정보화진흥원, 2012)

자에게 검색 및 조회를 할 수 있도록 환경을 제공해주면 되는 구조이다. 이때 데이터 제공자인 기관의 환경은 링크드 데이터 구축 및 제공을 위한 시스템이 갖추어져 있다고 전제하고 제공된 데이터의 무결성 유지를 위한 동기화 API(syndication API)를 개발하였다. 공공DB 피디아 시범사업에서는 기관의 데이터를 모두 제공받아 자체적으로 링크드 데이터로 변환하여 제공하고 있으므로 다양한 기관의 환경이나 운영상에서 발생할 수 있는 변수에 대해서는 고려하지 않고 향후 기관과의 연계체계 구축을 위한 표준적인 연계방안을 마련하는 데 중점을 두고 사업을 진행하였다.

4.1.3 링크드 데이터 URI 생성

RDF 데이터에서 리터럴(literal)을 제외한 모든 자원은 URI를 유일한 식별기호로 사용하므로 유일성을 보장할 수 있는 적절한 이름규칙을 가지도록 하며 데이터를 활용하려는 이용자도 알 수 있도록 직관적인 형태를 유지하도록 한다.

자원(resource) URI는 RDF에서 자원을 데이터로 표현하려는 사물을 의미하며 rdf:type을 이용하여 표현되는 클래스의 인스턴스에 해당한다. 자원 URI에 자원의 이름이나 표지(label)를 쓸 경우에는 유일성이 보장될 수 있도록 지정해야 한다. 기존의 레거시(legacy) 데이터베이스에서 RDF로 변환하는 경우에는 기본 키(primary key)를 URI로 이용하는 것도 가능하다. 공공DB 피디아의 자원 URI는 데이터 제공 기관의 DB마다 DS-0200, DS-0106와 같은 표지를 붙여 유일성을 보장하고 있으나 직관적이지는 못하다. RDF 데이터에서 사용하는 자원에 대한

클래스 이름과 프로퍼티명(predicate)을 RDF 데이터에 대한 메타데이터로 표현하는데 이때 어휘(vocabulary)들은 RDFS/OWL로 기술되며 URI로 식별한다. 공공DB 피디아의 어휘 URI들은 RDF 데이터의 URI와는 구분되는 다른 base URI를, namespace를 이용하여 논리적, 물리적으로 분리하고 있다. 즉 어휘는 '/vocabulary', 자원은 '/resource' 등으로 구분하여 사용하고 있다.

URI 생성시에 가장 문제가 되는 것은 데이터 이용자가 알 수 있는 직관적인 형태의 유일성을 보장하는 이름규칙을 정하는 것이다. 일련번호 코드나 시스템 id 등을 이용하는 방법도 제시되고 있는데 이는 유일성은 보장하나 규칙성이 없고 데이터 이용자가 불편하게 만드는 URI 생성 방법이다.

4.1.4 온톨로지 구축

링크드 데이터의 재사용성을 위하여 자원과 자원의 의미를 함께 제공하게 되는데 일반적으로 온톨로지를 기반으로 링크드 데이터를 생성하거나 변환하게 된다. 공공DB 피디아 시범사업에서는 데이터를 제공하는 기관들이 보유하고 있는 온톨로지가 존재하지 않아서 공공DB 피디아 구축과정에서 데이터 셋을 위한 온톨로지 모델링을 수행하고 그에 따른 변환 작업을 거쳐 링크드 데이터를 발행하였다.

공공DB 피디아의 대상 DB가 비교적 간단하고 지역을 중심으로 연결될 수 있는 통계 데이터가 중심이었으므로 온톨로지 모델링도 지역(region)과 지역통계(regionStats)라는 2개의 클래스를 생성하고 그 하위의 프로퍼티를 정의하는 것으로 간략하게 진행하였다. 공공기관이

보유하고 있는 대부분의 데이터는 업무 수행을 위한 행정 데이터가 대부분이므로 이 모델링의 활용성 및 확대 가능성도 상당히 높을 것으로 예상된다. 그러나 지식자원을 중심으로 기관이 보유한 데이터는 이러한 단순한 구조의 온톨로지 모델링으로는 자원의 의미를 충분히 표현할 수 없다. 따라서 향후 연계체계 구축을 위해서는 주제별 혹은 분야별 데이터와 데이터 간의 관계를 충분히 표현할 수 있는 온톨로지 구축이 선행되어야 할 것이다.

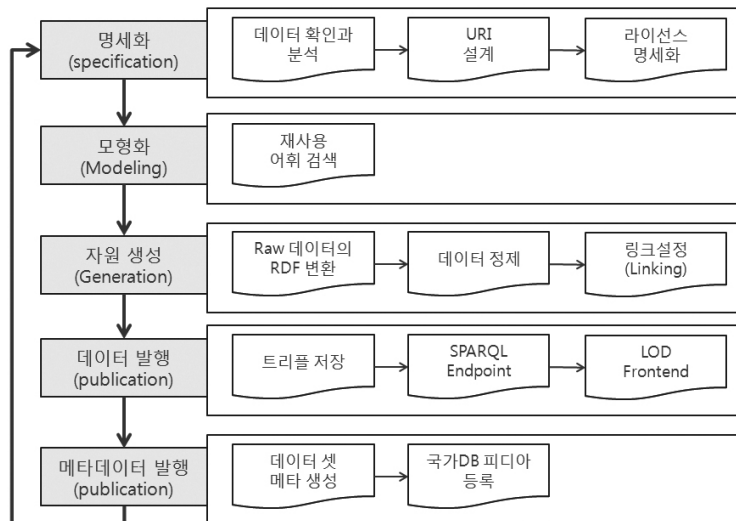
4.1.5 가이드라인 작성

공공DB 피디아 구축사업에서 작성된 가이드라인은 링크드 데이터 등록 기관에서 자원을 등록할 때 고려해야 하는 사항을 프로세스 절차별로 기술하고 있다. 링크드 데이터 자원의 생산 과정은 개발 프로젝트처럼 라이프사이클을 갖고 있으며 특히 링크드 데이터 자원의 지속적인 개선과 확장이 가능한 반복적인 생애주기 모델

을 따르고 있다.

공공데이터의 링크드 데이터 발행은 <그림 4>와 같이 명세화 → 모형화 → LOD 자원 생성 → 데이터 발행 → 메타데이터 발행 등의 과정을 거쳐 이루어진다. 모형화과정에서는 어휘의 재사용성을 통한 온톨로지 개발, 행정표준용어인 아담스(<http://www.adams.go.kr>)와 같은 어휘의 우선사용, 신규 어휘에 대한 별도의 생성 작업 등이 포함된다. LOD 자원 생성 과정에는 연계 오류가 발생된 링크정보의 수정 작업이 포함되며 메타데이터 발행시에는 공공기록물의 관리적 측면도 고려할 것을 제안하고 있다.

또한 데이터 관리절차와 관련하여 보안, 변환, 저장, 프로그램 개발 등과 같은 기능적 요구사항과 개인정보, 데이터 품질 등과 같은 비기능적 요구사항에 대해서도 기술하고 있다. 그 외에 연계체계 구축을 위한 필요한 데이터, 시스템, 서비스 모델 표준화에 대한 내용을 담고 있으면 관련 기술 및 산업 표준에 대한 분석도



<그림 4> 링크드 데이터 발행과정(한국정보화진흥원, 2012)

제공하고 있다. 그러나 이 가이드라인은 기존의 사례를 분석하고 참고하여 작성된 초안으로 참여기관 및 사업자가 활용할 수 있는 지침으로 실효성을 거두기 위해서는 실제 구축사업을 진행하면서 계속 내용을 보완하고 더불어 연계시스템 운영을 위한 매뉴얼 등을 작성하여 제공해야 할 것이다.

4.2 링크드 데이터 국가 연계체계 발전방안

4.2.1 참여기관 협의체 구성

국가 연계체계를 통해 공공데이터의 개방과 재활용을 활성화시켜 사회경제적 가치를 창출하도록 지원하며 공공정보를 요청하고 활용하는 것이 시민의 권리라고 인지하는 기관의 인식 변화가 필요하다(이정아, 2010). 기관의 인식변화가 선행된다 하더라도 공공정보를 생산하고 있는 공공기관이 링크드 데이터를 위한 준비가 되어 있지 않은 상태에서 공공데이터 활용을 위한 링크드 데이터 체계를 구축하는 과정은 쉽지 않을 것이다. 공공DB 피디아가 기관의 지속적인 협력 없이 데이터를 연계, 통합하여 링크드 데이터 서비스를 제공한다고 하더라도 데이터 공개를 통한 데이터 재사용의 효과를 충분히 거두기는 어려울 것이다.

현재 링크드 데이터 구축을 위해 준비되어 있거나 준비하고 있는 기관은 KISTI나 국토지리정보원 등 몇몇 기관을 제외하고는 찾아보기 힘들다. 어느 정도 준비되어 있다고 판단되는 기관들도 링크드 데이터 국가 연계체계 구축을 위한 거시적인 관점에서보다는 기관이나 해당 분야를 중심으로 시맨틱 웹 서비스를 위한 과정에

서 링크드 데이터 연계체계를 따르고 있다고 봐야 할 것이다. 국가차원에서 지원을 통해 각 기관이 링크드 데이터 연계시스템을 구축한다고 해도 국가지식포털과 같은 기존의 국가DB 연계체계 구축 과정에 비추어 보면 기관의 상황에 따라 다양한 변수가 발생할 것으로 예상된다. 어느 기관은 연계체계 구축을 위해 적극적으로 기관의 시스템 아키텍처까지 변경하면서 참여하고 어느 기관은 데이터 공개에는 찬성하나 기관의 기존 시스템을 링크드 데이터 발행이 가능한 체계로 전환하는 것을 부담스러워한다. 또 일부 기관은 데이터를 공개하였을 때 기관이 연계 되는 혜택이 명확하지 않으면 적극적으로 참여하지 않는다.

이러한 사례는 기존의 국가지식포털을 비롯한 국가DB 연계체계 구축 및 운영과정에서 이미 경험한 문제 상황으로, 기관들의 협의체를 구성하여 지속적으로 논의하고 합의를 통해 문제를 해결해 나가는 과정을 거쳐 어느 정도 해결할 수 있다. 전체 기관을 대상으로 협의체를 구성하되 분야별로 그룹을 나누어 논의를 진행하고 그룹별 의견을 모아 전체적인 협의를 진행하는 것이 효율적이다. 지식사업이라 불리던 국가지식정보자원관리사업에서 출발하여 국가DB 사업까지 한국정보화진흥원에 의해 10년 넘게 유지하고 있는 국가DB 연계체계를 활용하는 것도 좋은 방법이 될 것이다. <그림 5>의 구성도에 나타나 있는 종합정보센터 혹은 전문정보센터로 참여하고 있는 기관들의 DB를 대상으로 통합검색서비스를 제공하던 국가지식포털은 현재 공공데이터포털에 통합되어 있다. 국가지식포털은 5개의 종합정보센터를 중간 서비스 제공자로 두고 데이터 제공자인 전문정보센터의 데



〈그림 5〉 국가지식포털(현 공공데이터포털에 통합)의 구성도
 〈http://www.data.go.kr〉

이터를 통합해서 제공하도록 함과 동시에 전문 정보센터에서 직접 데이터를 제공받는 이원화된 체제를 활용하였다.

공공DB 피디아에 참여하는 기관의 범위가 명확히 정해지지 않은 상태이나 국가지식포털의 참여기관과 대동소이할 것으로 판단되므로 이 연계체계를 활용하는 것이 바람직할 것이다. 최근에는 국가의 지원규모가 축소되면서 명맥을 유지하는 정도로만 운영되고 있으나 국가에서 지원받아 공공데이터를 구축한 거의 대부분의 기관이 데이터 제공자로 참여하고 있으므로 이를 링크드 데이터 참여기관으로 재구성하는 것이 노력을 아낄 수 있는 방안이라 할 수 있다. 기관 협의회체를 구성한다고 해서 모든 문제가 해결되는 것은 아니지만 협의회체를 통해 링크드 데이터 국가 연계체계 구축과정에서 발생하는 문제를 공동으로 해결하고 발전을 위해 합의된 방

향으로 나아가도록 노력해야 할 것이다. 기관이 담당하고 있는 역할과 책임의 수행을 위해서 기관 스스로가 노력하는 것은 당연하나 그 협의회체를 운영하고 지원하는 한국정보화진흥원의 동기 부여 역할이 가장 중요하다고 할 수 있을 것이다.

4.2.2 연계시스템 구축

공공DB 피디아는 참여 기관이 링크드 데이터 구축 및 공개를 위한 시스템적인 기반이 갖춰져 있다는 가정 하에 링크드 데이터 연계체계를 구축하는 시스템 아키텍처를 기본으로 하여 동기화 API(Syndication API)만을 제공하는 것으로 시스템 구축 범위를 한정하고 있다. 그러나 데이터 제공자인 기관이 앞의 〈그림 3〉과 같은 시스템을 구축하고 있지 못한 상황이라면 동기화 API는 무용지물이 될 수밖에 없다. 이

러한 상황은 국가 연계체계뿐 아니라 기관의 협력체계 구축에 있어서도 발생할 수 있는 문제로 담당자나 기관의 의지만으로 해결될 수 없는 부분에 해당한다.

기준에 OAI(Open Archives Initiative) 기반으로 구축된 국가지식포털이나 한국역사정보통합시스템(<http://www.koreanhistory.or.kr>)은 서비스 제공자(Service Provider)로서 지금의 공공DB 피디아에 해당한다. 서비스 제공자는 데이터 제공자(Data Provider)로서 실제 데이터를 보유하고 있는 각 기관(전문정보센터)들이 보내주는 데이터를 제공받아 이용자에게 통합검색을 제공하는 역할을 수행한다. 그러나 연계 대상이 되는 기관들이 데이터 제공자로서 역할을 수행할 수 있는 환경을 마련하기에는 여러 가지로 장애가 많아 실제 시스템을 구축할 때에는 서비스 제공자가 주도적으로 각 기관의 메타데이터를 추출하고 수집할 수 있는 API를 개발하여 각 기관의 개별적인 환경에 맞도록 수정하여 배포하였다. 몇몇 기관의 경우에는 기관의 정책에 따라 기관에서 메타데이터를 오프라인으로 제공하기도 하고 때로는 직접 방문하여 데이터를 추출하고 수집해 오는 경우도 발생한다. 그러나 기관 담당자와 사업자가 변경되면서 이러한 부분에 대한 주의를 기울이지 않고 시스템 개발을 진행하는 경우도 발생하므로 이 부분에 대한 협조 요청도 필요하다.

이러한 문제는 공공DB 피디아 구축, 운영에 있어서도 얼마든지 발생할 수 있을 것으로 예상된다. 예를 들어 <그림 3>의 시스템 아키텍처에서 공공DB 피디아와 데이터 제공자 간의 동기화 API를 통한 데이터 최신성 유지 매커니즘은 기관의 상황에 좌우될 가능성이 있다. API를

이용하지 못하는 상황에 맞춰 시나리오를 준비하고 대안을 마련해야 한다. 데이터 제공기관이 링크드 데이터 구축을 위한 준비가 되어있지 않다면 공공DB 피디아에서 시범사업과 같은 방식으로 데이터를 제공받아 온톨로지를 구축하고 링크드 데이터로 변환하여 서비스하는 방법도 병행할 수 있을 것이다.

4.2.3 링크드 데이터 URI 생성

링크드 데이터 발행 과정에서 제일 먼저 거쳐야 하는 URI 생성과 관련해서도 namespace로서 구별하는 것만으로 중복 문제를 해결하면서 재사용성을 보장할 수는 없다. 공공데이터를 개방하고 활용하기 위한 방안으로 링크드 데이터를 구축한다는 것은 각 기관에서 구축한 링크드 데이터가 동일한 사물을 가리키는 경우 namespace로서 구별하여 중복되지 않도록 하는 것이 아니라 동일한 URI를 가져야 한다는 의미를 전제하고 있다. 이러한 문제는 LOD 구축 초기부터 제기되었고 해결이 쉽지 않은 과제인데, 이 문제를 해결하기 위해 언론, 방송 분야에서 위키피디아를 중심으로 URI를 생성하여 채택하고 있다.

공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축에는 대부분 공공데이터를 보유하고 있는 기관들이 참여하게 될 것이다. 따라서 앞서 4.2.1에서 기관협력 방안으로 제시한 기관 협의체가 구성될 것이므로 논의를 통해 데이터별로 주 담당기관을 정하여 그 기관이 주도적으로 URI를 생성하도록 하면 공공데이터 전반에 걸쳐 단일한 URI 생성이 가능해진다. 예를 들어 역사적인 인물은 역사분야에서 주도적으로 URI를 생성하도록 하고 현대 시기 인명은

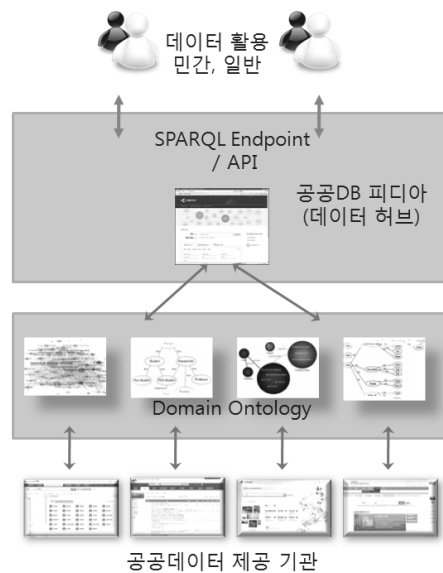
교육학술분야에서 많이 출현하므로 그 분야에
서 담당하는 것으로 하면 된다. 그 두 분야에 해
당되지 않는 경우에는 그 다음으로 인명의 출현
이 많은 기관이 담당하기로 미리 합의하면 된다.
공공DB 피디아 시범사업처럼 데이터를 일괄
제공받는 경우에도 데이터별로 한국정보화진흥
원과 데이터별 담당기관들이 협의하여 URI를
생성하면 규칙성을 유지할 수 있을 것이다. 여
기에 해당하지 않는 예외적인 경우가 발생하는
경우에는 namespace를 이용하여 해결하면 될
것이다.

4.2.4 온톨로지 구축

링크드 데이터 발행을 위한 또 다른 문제는
온톨로지와 관련되어 있다. 첫째, 기관의 온톨
로지가 존재하지 않는 경우, 공공데이터를 제공
하는 대부분의 기관이 링크드 데이터를 구축하
고 있지 않으므로 데이터와 그 의미를 파악할

수 있는 도구인 온톨로지도 없다는 사실은 이미
시범사업을 통해서도 파악이 된 부분이다. 둘째,
기관이 온톨로지를 구축하고 있다하더라도 각
데이터 제공자가 이용하는 개별적인 온톨로지
의 활용 또한 쉽지 않다는 것이다. <그림 6>에서
와 같이 각 분야의 도메인 온톨로지 구축을 유
도하여 국가차원에서 도메인 온톨로지에 기반
한 시맨틱 포털을 구축하고 이를 활용하여 링크
드 데이터를 발행하는 방법이 가장 바람직한 방
향이라 할 수 있다.

링크드 데이터 구축시에 온톨로지 모델링을
통한 도메인 온톨로지 결합으로 시너지 효과를
일으킬 수 있는 분야나 데이터는 상당히 많다.
예를 들어 이용자가 '이순신'을 검색하여 링크드
데이터에 접근한다고 하면 이순신이 만든 '거북
선'과 관련된 역사적인 사실은 역사분야를 통해
확장하고 거북선의 원리 등은 과학기술분야를
통해 접근하며 거북선과 관련된 유물, 유적 등



<그림 6> 도메인 온톨로지 적용 체계

에 대한 데이터는 문화분야를 통해 확장하여 이들 데이터를 활용한 다양한 어플리케이션도 얼마든지 개발할 수 있게 된다. 이렇게 데이터에 접근하고 활용할 수 있는 것이 링크드 데이터의 혜택을 단적으로 보여주는 예가 될 것이다.

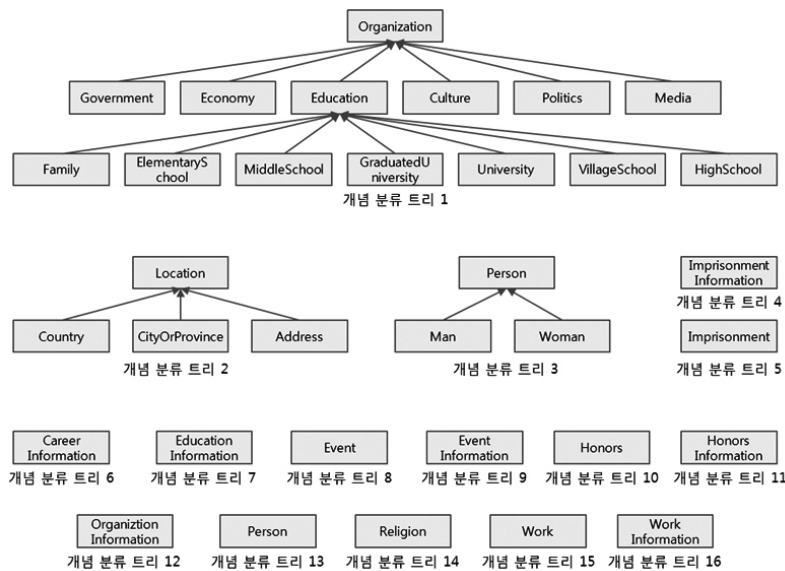
온톨로지를 신규로 구축해야 하는 기관의 경우 모델링에 투자하는 비용대 효과를 고려해 보았을 때, 공공데이터 개방을 계속 미루는 것보다 <그림 7>의 예와 같이 시소러스 수준의 상하위 개념 관계(subClassOf)만으로 구축한 온톨로지 모델을 통해 링크드 데이터를 발행하고 공공데이터를 활용하는 것도 고려해 볼 만하다.

과학기술분야나 법률분야처럼 데이터가 명확한 경우에는 온톨로지 자동 구축 도구를 활용하여 온톨로지를 구축하는 방법도 고려할 만하다. 통계청 데이터, 기술표준원의 표준기술정보 등과 같이 정형화되어 있거나 어느 정도 정형화시킬 수 있는 공공데이터의 경우에는 온톨로지 자

동 구축이 가능하다. 그러나 시간이 지나면 모두 역사가 되고 시간이 흐른 후에 사건의 명칭이 확립되는 등 다양한 변수가 발생하는 역사분야나 문학분야에서의 온톨로지 자동 구축은 불가능하며 실효성도 없다.

4.2.5 가이드라인의 상세화

공공데이터 활용을 위한 링크드 데이터 구축을 위해 꼭 필요한 또 하나는 기관의 참여 유도 및 참여의 방법을 제시하는 공개지침인 가이드라인 제공이다. 공공DB 피디아 시범사업을 통해 작성된 가이드라인은 기본적인 골격을 갖춘 정도로 기관이 지침으로 사용하기 위해서는 보완이 필요하다. 데이터의 광범위하고 효과적인 재사용을 보장하기 위해 데이터에 대한 라이선스를 명확히 하여 데이터 재사용이 법적인 문제를 일으키지 않도록 해야 한다. 예를 들어 라이선스에 데이터 셋 간 상호운용성이 가능하도록



<그림 7> 상하위 개념 관계(subClassOf)

CC(Creative Commons)와 같이 사람들이 이미 알고 있는 기존의 라이선스를 명시하는 것도 한 방법이다(Bauer & Kaltenböck, 2012). 그와 더불어 실제 링크드 데이터 구축에 참여하는 기관 및 사업자를 위해 더 상세한 매뉴얼이 필요하다. 운영매뉴얼에는 각 분야의 기관들이 참여하고 링크드 데이터를 발행하는 데 있어 당면하게 되는 URI 생성, 온톨로지의 구축의 상세 정도, 메타데이터 발행 등에 내용이 들어가야 한다. 가이드라인과 운영매뉴얼은 기존 사례들을 참고하고 기관 협의체에서의 논의를 통해 국가 연계체계 담당기관인 한국정보화진흥원에서 작성하여 배포해야 한다. 또한 지속적인 갱신을 통해 내용을 보완하고 그 이력을 남겨서 연계체계 구축 과정에서 반복될 수 있는 시행착오를 줄이도록 해야 한다. 이런 지침류는 시스템을 운영하고 유지하는데 있어 표준이 되며 신규 기관들의 링크드 데이터 참여 유도를 위한 안내서로 역할을 하게 된다. 또한 담당자가 바뀌었을 때 교육교재로 사용이 되며 시스템의 지속성을 유지하는 데 있어서도 중요한 역할을 하게 된다.

5. 결론

공공데이터 공개 및 활용을 통한 국가 경쟁력 확보라는 목표를 이루기 위해서는 링크드 데이터의 체계적이고 광범위한 구축이 우선적으로 이루어져야 할 것이다. 이를 위해서는 공공기관의 적극적인 참여와 더불어 공공데이터가 민간에게 가능한 편리하게 제공되고 이용될 수 있는 체계가 구축되어야 한다. 현재 공공정보를 공개하려는 노력이 국가 중앙부처나 지자체 등을 중

심으로 가시화되고 있으나 대부분 OpenAPI를 통해 미리 정해진 형식으로 제공되는 등 전면적이고 완전한 개방이 이루어지고 있지는 않다. 공공데이터의 재사용성 및 활용성을 높여 기관이 보유하고 있는 데이터의 가치를 높이고자 하는 기관의 목표는 온톨로지에 기반한 링크드 데이터 구축으로 달성될 수 있을 것이다.

본 논문에서는 공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축을 위해 공공DB 피디아 구축 과정에서 발생한 문제를 파악하고 해결 방안을 제시하고자 하였다. 국가차원에서 공공정보를 공유하고 활용하고자 하는 노력은 오래전부터 시작되었으며 데이터의 재사용을 높이기 위해 링크드 데이터로 공개하는 프로젝트도 그 연장선상에서 이루어졌다고 할 수 있다. 미국을 비롯한 세계 각국에서도 이러한 움직임은 빅데이터 논의와 함께 지속적으로 추진, 확대되고 있다. 정부기관뿐만 아니라 언론·방송분야에서도 위키피디아의 지식베이스인 DBpedia를 중심으로 활발하게 링크드 데이터 프로젝트가 진행되고 있다. 도서관 분야에서도 미의회도서관을 중심으로 FRBR 도입 등과 맞물려 의미에 기반한 서지데이터 연계 및 공개가 활발히 진행되고 있다.

우리나라도 국가차원에서 온톨로지를 활용한 시맨틱 웹 서비스 구축에 대한 논의에서 시작하여 링크드 데이터 구축으로 논의가 구체화되면서 LOD 시범사업으로 공공DB 피디아를 구축하는 단계에 이르렀다. 공공데이터 활용을 위한 국가 연계체계 구축을 위해서 공공DB 피디아 구축사업을 통해 파악된 문제의 대부분은 기존의 국가DB 연계체계 구축과정에서 해결의 실마리를 찾을 수 있다. 첫째, 공공데이터를 보

유하고 있는 기관의 적극적인 참여를 유도하고 연계체계의 지속적인 유지를 위해 기관 협의체 구성이 필요하다. 기관의 역할 수행과 관련하여 링크드 데이터 제공 기관들의 역할을 크게 기대할 수 없는 경우에는 링크드 데이터 구축 사업의 실효성을 위해 국가차원에서 그 역할을 대신할 수 있는 방안까지 염두에 두어야 한다. 둘째, URI 생성과 관련하여 구축한 링크드 데이터가 동일한 사물을 가리키는 경우 namespace로 서로 구별하여 중복 문제를 해소하는 것만으로는 충분하지 않다. 즉 동일한 URI를 가져야 하는데 이 경우 링크드 데이터의 기준이 되는 URI 생성이 필요하다. 기관 협의체 내에서 논의를 거쳐 데이터별 주 담당기관을 선정하여 일관성 있게 URI를 생성하고 관리하도록 한다. 언론 방송분야에서는 위키피디아를 중심으로 한 URI 생성을 채택하고 있다. 셋째, 각 분야별로 합의하여 도메인 온톨로지를 구축하고 이를 결합하여 링크드 데이터를 발행하도록 한다. 온톨로지가 구축되어 있지 않은 경우, 개방을 무한정 미루는 것보다 상하관계 정도의 간단한 모델링이

나 정형화된 데이터의 경우 온톨로지 자동 구축 도구를 이용하는 것도 한 방법이다. 기관에서 링크드 데이터 발행이 어렵다면 공공DB 피디아 시범사업에서처럼 데이터를 제공받아 온톨로지 모델링을 거쳐 링크드 데이터를 발행하는 방법도 병행하면 된다. 마지막으로 가장 중요하면서도 간과하고 지나갈 수 있는 부분인 가이드라인 및 매뉴얼 작성이다. 기관 참여를 적극적으로 유도하기 위해서는 정확한 가이드라인과 상세한 운영매뉴얼이 필수적이다.

공공정보의 링크드 데이터 공개는 민간의 창의성과 결합하여 고부가가치 산업 육성은 물론 스마트 환경기반의 정부행정 혁신을 통해 국민의 삶의 질을 향상시키고 우리나라를 정보 선진국으로 도약시키는 기반이 될 수 있을 것이다. 공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축의 성공은 기관의 적극적인 참여를 통해서만 가능할 것이다. 따라서 기관 간 협의체를 통한 지속적인 논의와 합의, 정확한 방향 제시, 그리고 국가차원의 적극적인 지원이 뒷받침되어야 할 것이다.

참 고 문 헌

- 노영희 (2012). dCollection의 링크드 데이터 구축에 관한 연구. 한국도서관·정보학회지, 43(2), 247-271.
- 박옥남 (2012). 기록물 전자통제 기반 Linked Data 구축에 대한 연구. 한국비블리아학회지, 23(2), 5-25.
- 박지영 (2012). 링크드 데이터 방식을 통한 서지 정보의 확장에 관한 연구. 정보관리학회지, 29(1), 231-251. <http://dx.doi.org/10.3743/KOSIM.2012.29.1.231>
- 오원석 (2009). Linked Data 동향과 전망. 서울: 탐쿼드란트코리아.
- 이상윤, 윤홍주 (2012). 공공데이터를 활용한 국가정보화 전략연구: 시나리오플래닝을 적용하여. 한국전자통신학회논문지, 7(46), 1259-1273.

- 이정아 (2010). 스마트 정부의 공공정보 개방과 이용활성화 전략(CIO Report 28). 서울: 한국정보문화진흥원.
- 이현정, 남영준 (2012). 공공데이터베이스의 Linked Open Data구축을 위한 가이드라인 설계. 제19회 한국정보관리학회 학술대회 논문집, 63-68.
- 조명대 (2010). 도서관에서의 Linked Data 활용방안에 관한 연구. 한국문헌정보학회지, 44(1), 181-198.
<http://dx.doi.org/10.4275/KSLIS.2010.44.1.181>
- 최진원 (2012). 공공정보 이용활성화를 위한 법제도적 과제에 대한 연구. 정보법학, 16(1), 237-266.
- 한국정보화진흥원 (2012). 국가지식 지능형연계체계 가이드라인. 서울: 한국정보화진흥원.
- Bartlett, O. (2013, February 19). Linked data: Connecting together the BBC's online content. Retrieved from <http://www.bbc.co.uk/blogs/internet/posts/Linked-Data-Connecting-together-the-BBCs-Online-Content>
- Bauer, F., & Kaltenböck, M. (2012). Linked open data: The essentials. Retrieved from <http://www.semantic-web.at/LOD-TheEssentials.pdf>
- BBC programmes. <http://www.bbc.co.uk/programmes/developers>
- Berners-Lee, T. (2006). Linked data: design issues. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
- CKAN - Data Hub. Retrieved from <http://datahub.io>
- FinnONTO-National Semantic web ontology project in Finland. Retrieved from <http://www.seco.tkk.fi/projects/finnonto/>
- LinkedBrainz - A project to provide MusicBrainz NGS as linked data. Retrieved from <http://linkedbrainz.c4dmpresents.org>
- MusicBrainz - The open music encyclopedia. Retrieved from <http://musicbrainz.org>
- NewYork Times Linked Data. Retrieved from <http://data.nytimes.com>
- UK Open Government. Retrieved from <http://data.gov.uk>
- US Open Government. Retrieved from <http://www.data.gov>
- W3C - Linked data. Retrieved from <http://www.w3.org/standards/semanticweb/data>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Cho, Myungdae (2010). A study on applications for linked data in libraries. Journal of Korean Society for Library and Information Science, 44(1), 181-198.
<http://dx.doi.org/10.4275/KSLIS.2010.44.1.181>

- Choe, Jin-Won (2012). A study on the legal issues in PSI re-use. *Journal of Korea Information Law*, 16(1), 237-266.
- Lee, Jong-A (2010). Strategy to open and utilize public information for smart government(CIO Report 28). Seoul: National Information Society Agency.
- Lee, Sang-Yun, & Yoon, Hong-Joo (2012). The study on strategy of national information for electronic government of S. Korea with public data analysed by the application of scenario planning. *The Korean Institute of Electronic Communication Sciences*, 7(46), 1259-1273.
- National Information Society Agency (2012). Guideline for national knowledge linking system. Seoul: National Information Society Agency.
- Noh, Young-Hee (2012). A study on configuring dCollection as the linked data. *Journal of the Korean Library and Informtion Science Society*, 43(2), 247-271.
- Oh, Won-Seok (2009). *Toward the data Web*. Seoul: TopQuadrant Korea.
- Park, Ok Nam (2012). The design and development of linked data from authority data in national archives of Korea. *Journal of the Korean Biblia Society for Library and Information Science*, 23(2), 5-25.
- Park, Ziyong (2012). Extending bibliographic information using linked data. *Journal of the Korean Society for Information Management*, 29(1), 231-251.
<http://dx.doi.org/10.3743/KOSIM.2012.29.1.231>
- Yi, Hyun-Jung, & Nam, Young-Joon (2012). A study on designing guidelines for Linked Open Data organization of national databases. *Proceedings of the 19th Conference of Korean Society for Information Management*, 63-68.