기록물 전거통제 기반 Linked Data 구축에 대한 연구*

The Design and Development of Linked Data from Authority Data in National Archives of Korea

박 옥 남(Ok Nam Park)**

초 록

본 연구에서는 한국형 Linked Data Cloud를 위한 단계로 국가기록원 전거데이터 셋을 Linked Data화하는 것을 목적으로 한다. 이를 위해 국가기록원 전거 데이터의 구조 및 검색시스템을 분석한 후, 전거데이터 셋을 RDF/OWL로 모델링하는 과정을 거쳐 Linked Data로 구축하였다. 또한 기록물의 Linked Data화를 위해 Dublin Core, SKOS(Simple Knowledge Organization System)를 활용하였고 온톨로지 설계도구로 TopBraid Composer TM을 사용하였다. 구축된 RDF/OWL 데이터는 시각화화면을 통해 전거데이터의 자유로운 확장, 전거데이터 용어간의 탐색, 용어의 상세 전거데이터의 접근이 가능하도록 하였으며, 기록물과 연계를 통해 기존 기록물 검색에서 한계로 지적되었던 기록물과 전거데이터간의 연계의 부족을 보완하였다.

ABSTRACT

The purpose of this study is to develop linked data of authority data in national archives of Korea as a cornerstone for linked data cloud of Korea. The study analyzed data structure of authority data as well as a retrieval system. It finally developed linked data based on RDF/OWL, Dublin Core, and SKOS. The study also employed TopBraid ComposerTM as a tool for ontology construction. The visualization of the tool provides users with flexible search and browsing between data as well as access of detail authority data. It complements the search of a current system in terms of flexible linking between records and authority data. The study also suggests future work to publish linked data of archival data set itself and make rich relationships among data in museums, libraries, and other archives.

키워드: 기록물, 전거데이터, 링크드 데이터, RDF/OWL, SKOS
Archive, Authority Data, Linked Data, RDF/OWL, SKOS

---

* 이 논문은 2012년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.
** 한남대학교 문헌정보학과 조교수(ponda@hnu.ac.kr)
논문접수일자 : 2012년 5월 21일 논문심사일자 : 2012년 6월 5일 게재확정일자 : 2012년 6월 14일

1. 서 론

웹이 등장함에 따라 정보환경도 웹을 중심으로 급속히 재편되고 있다. 웹은 더 이상 인간이 정보를 소비하는 환경이 아니라 정보를 생산하고 유통하는 플랫폼으로 발전하고 있다. 우리는 웹 중심의 정보사회에서 정보의 홍수를 경험하고 있으며 기존의 데이터 처리방식을 넘어 새로운 패러다임으로의 전환이 요구된다. 최근에는 '빅데이터(Big Data)'라는 용어가 이러한 변화의 흐름을 보여주고 있으며, 이 가운데 중요한 정보를 어떻게 효율적으로 습득할 것인지가 강조되고 있다(이만재 2011).

공유와 개방의 웹 정보환경은 여전히 데이터 상호교환의 관점에서 본다면 사일로(silos)인 상황이다. 대표적인 사례가 도서관 분야로 도서관은 변화하는 환경에 부응하기 위하여 OPAC, MARC, FRBR, Dublin Core와 같은 표준을 제정하고 도서관 시스템에 적용하기 위하여 노력하였다. 이러한 노력으로 도서관 내의 정보공유에는 어느 정도 성과를 거두었으나 도서관 외부의 정보환경와의 소통에는 어려움을 겪고 있다. 폐쇄적인 정보시스템에 포함된 데이터들은 웹 인터페이스를 통해서만 접근이 가능하며 도서관 외부의 이용자가 이해하고 사용하기 힘든 도서관만의 표준이라고 지적되어 왔다(김성혁 2011; 조명대 2010).

이러한 사일로를 극복하면서 개방과 공유의 시대적 흐름에 부응하기 위하여 데이터의 연계라는 Linked Data의 움직임이 일어나고 있다. Linked Data는 시맨틱 웹이 표방하는 데이터의 웹(Web of Data)을 만들기 위한 구체적인 방법으로 버너스 리가 웹에 데이터를 출판하고 연결하는 데이터의 구조로 제안한 것이다. 이러한 Linked Data를 개발하고 공유하려는 움직임에 맞물려 현재 민간, 도서관, 공공정부에서 각각의 Linked Data Cloud 제작에 노력하고 있으며 더 나아가 한국형 Linked Data Cloud 제작을 위해서는 우선 핵심 데이터 셋을 선정하고 이를 Linked Data화하여 활용・확산하는 노력이 요구된다.

이에 본 연구에서는 한국형 Linked Data Cloud를 위한 단계로 국가기록원 전거데이터 셋을 Linked Data화하는 것을 목적으로 한다. 국가기록원 전거데이터 셋과 같은 공공정보에 대한 Linked Data화는 중요하다. 공공정보의 개방과 공유에 있어 시대적 요구가 증가하고, 이에 따라 공공지식정보를 민간에 공유하거나 공공DB를 적극 활용하기 위한 법률 제정안을 발의하는 등 공공정보의 개방에 대한 시대적 요구에 부합하기 위한 노력이 다양한 방식으로 진행되고 있으며(김태동 2010), 박물관 - 도서관 - 아카이브의 연계에 대한 요구(배정현 2007; 서혜란 2005) 역시 증가하고 있기 때문이다. 이러한 시대적 요구에도 불구하고 기록검색 서비스는 여전히 사일로에 머물고 있다. 기록을 검색하기 위한 다양한 검색방법이 제공되고 있음에도 불구하고 여전히 기록검색은 일반 이용자가 이해하기에는 어려우며, 기록원 밖의 정보환경과는 고립되어 있어, 공공정보의 공개 및 유통이라는 현재의 흐름에 부응하지 못하고 있다. 본 연구는 시맨틱 웹 분야의 모범실무 중의 하나인 Linked Data를 활용하여 실용화하면서 국가기록원의 데이터를 지능화시키는 것을 목적으로 한다.

이를 위해 국가기록원 전거 데이터의 구조

및 검색시스템을 분석한 후, 전거 데이터 셋의 온톨로지화, 온톨로지를 Linked Data로 전환하는 것을 목표로 한다. 특히 본 연구에서는 기록물 데이터 셋 중 전거데이터에 집중한다. 이는 전거 데이터가 기록물 검색의 집중(collocation)을 가능하게 하는 것은 물론, 기록물의 맥락 및 출처정보를 제공하는 중요한 정보원이기 때문이다. 그러므로 전거데이터의 Linked Data가 되어야 진정한 기록물 데이터의 공유 및 상호운용성 확보가 가능할 것이다.

연구의 결과는 향후 도서관, 대학기록관, 박물관 데이터와의 국가적, 세계적 정보융합을 선도할 수 있으며 Linked Open Data에 발행함으로써 국가기록원 데이터 부가가치를 상승시키고, SKOS나 FOAF와 같은 표준어휘를 활용하는 다른 데이터와의 연계로 풍부한 웹 생태계 구성을 조력할 수 있다.

2. 국가기록원 검색

국가기록원은 2007년 나라기록포털이 구축된 이래 중앙・지방 및 공공기관에서 생산된 기록물을 수집하고 조직하여 기록정보 공개 및 열람서비스를 제공하고 있다. 2012년 1월 기준으로 문서 300만권, 도면 22만건, 카드 20만건, 시청각기록물 235만점, 행정박물류 5만 7천점, 행정간행물 22만종, 44만 6천권에 달하는 공공기록은 물론 역대 대통령 기록물, 민간기록물, 해외기록물, 국가 지정 기록물, 구술 기록물 등 다양한 형태의 특수기록물을 보유하고 있다.

나라기록포털을 통해 국가기록물이 제공하는 기록물에 대한 검색서비스가 이루어지고 있는데, 검색방식은 키워드 검색 및 전거 데이터 기반 검색으로 구분된다. 첫 번째, 키워드 검색은 단순 검색 및 고급검색을 통해 이용자가 입력하는 키워드와 기록물 목록데이터가 일치하는 정보를 검색하는 방식이다. 둘째, 생산기관 연혁데이터 베이스를 통한 기관별 검색이 있으며 마지막으로 기록물의 군-시리즈-철-건에 따른 기술 계층별 검색과 주제별검색으로 구분된다(류미숙 2009).

이 중 전거데이터를 사용하고 있는 주제별 검색 및 생산기관별 검색을 살펴보았다. '대통령선거유세(대전1)' 기록을 검색한다고 가정하고 검색을 실시해보았다. 먼저 생산기관별 검색을 살펴보면, 행정기관의 계층정보와 대/중/소 카테고리 정보를 따라 브라우징 검색을 제공하고 있으며, 철・건 단위로 검색된 기록물을 〈그림 1〉과 같이 제시하고 있다. 이는 생산기관의 전체적인 계층적인 구조를 브라우징으로 파악할 수 있다는 장점이 있으나, '국가행정조직 -〉 중항행정기관 -〉 처 -〉 총무처'의 계층구조를 통해서 접근 가능하므로 생산기관의 소속기관에 대한 지식이 없을 경우 검색 시스템을 사용하기 어렵다는 단점이 있다.

주제별 검색을 통해서는 국정분야별, 주제유형별 검색을 제공한다. 국정분야별 검색으로 정부기능연계모델(Business Reference Mode, 이하 BRM)에 따른 28개 분류체계별 디렉토리를 따라가며 검색을 제공하며 주제유형별 검색을 위해 주제유형을 10가지로 구분하고 하위 주제어를 개발하여 각 주제어에 대하여 기록물의 근거, 배경, 내용, 참고자료, 집필자, 관련주제, 관련 기술서를 사용하여 설명하고 있다(〈그림 2〉 참조). 그러나 국정분야별, 주제유형별 검색

〈그림 1〉 생산기관별 검색

〈그림 2〉 주제별 검색 화면

은 생산기관별 검색과 연계되지 않아, 생산기관 검색을 통해 국정분야 및 주제분야를 파악할 수 없으며, 주제어 역시 미완성된 상태로, '선거유세'에 관한 국정분야 및 주제유형을 찾을 수 없다.

국가기록원의 검색은 다음과 같은 문제점을 지적할 수 있다(류미숙 2009; 설문원 2010).

첫째, 검색이 용이하지 않다. 계층별 검색, 생산기관별 검색 같은 수직적 구조는 이용자의 생산기관 또는 출처에 대한 지식을 전제로 하

고 있어 이에 대한 지식이 충분하지 않은 일반 이용자는 기록물 검색도구를 이용한 검색이 용이하지 않다.

둘째, 계층별 검색, 생산기관별 검색과 주제별 검색이 연계가 이루어지지 않아 유연한 정보검색을 제공하지 못하고 있다.

셋째, 생산연혁데이터베이스와 국가기록원 검색시스템이 연계가 이루어지지 않아, 검색된 기록의 출처정보를 통해 기관의 연혁정보를 획득할 수 없거나, 기록의 출처정보를 통해 이전 기관의 기록물을 획득할 수 없다. 또한 기관 정보에서 기관이 생산한 목록레코드로의 이동이 불가능하다.

이러한 문제점을 해결하는 것은 기존의 계층별 검색, 생산기관별 검색과 같은 수직적 구조 외에 주제별 검색이나 생산기관 연혁정보와의 유연한 연계를 통해 다양한 접근수단을 제공하는데 있다. 즉, Linked Data의 개발을 통해 데이터 간의 상호연계를 통한 유연한 연동이 요구된다.

3. Linked Data

3.1 Linked Data 소개

정보에 의미를 부여하고 기계 스스로 정보에 접근하여 분석할 수 있도록 하는 기술은 시맨틱 웹이라는 이름으로 시작되었다. 그러나 시맨틱 웹이 의도하는 바와 달리 다수의 커뮤니티가 채택하지 않아 시맨틱 웹이 추구하는 의미(Semantics)전달이라는 의도가 명시적으로 드러나지 않았다(이만재 2011). 이에 버너스 리는 'Raw Data Now'라는 연설에서 기존의 인터넷을 문서에 연결하고 더 나아가 데이터의 연결을 가능하게 하자는 Linked Data의 보급을 강조하였다(Berners-Lee 2006).

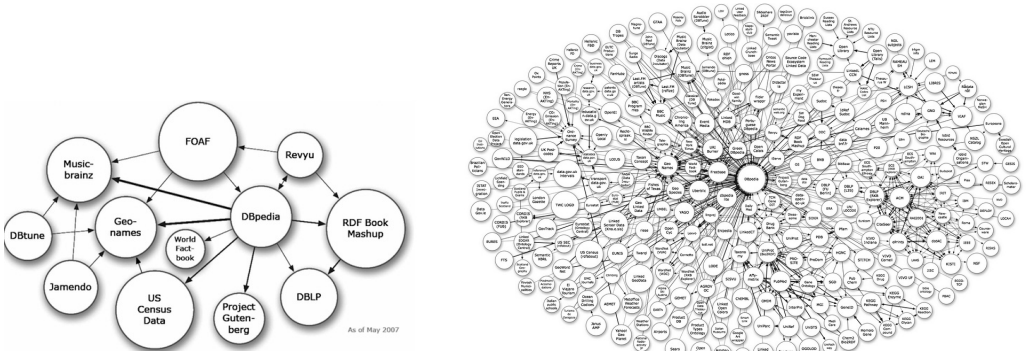
Linked Data는 "URIs와 RDF를 사용하여 웹 상에 데이터, 정보, 지식을 공개하고, 공유하고, 하나로 연결하기 위한 모범 사례를 설명하기 위하여 사용되는 용어로, 시맨틱 웹이 표방하는 데이터 중심의 웹(Web of Data)을 만들기 위한 구체적인 방법(김성혁 2011, 24)"이다. 기존의 웹은 하이퍼링크를 사용하여 노드와 노드, 즉 정보자원을 서로 연결하는 방식으로 이러한 웹에서 이용자는 하이퍼링크를 통해 정보자원을 이동하는 것은 물론 가독함으로써 정보자원 간 상호연계를 파악할 수 있으나 컴퓨터는 정보자원간의 연계를 인식할 수 없다는 한계를 가진다. 즉, 컴퓨터는 인간과 같이 읽고 해독하여 정보의 내용을 파악하거나 데이터를 연결할 수 없는 구조를 가졌다. 이러한 웹의 한계를 극복하기 위하여 버너스 리는 웹에 데이터를 출판하고 연계하고자 하는 모범 사례로 Linked Data를 제안하였다. Linked Data는 하이퍼링크를 사용하여 정보자원을 연결하던 문서 중심의 웹(Web of Documents)에서 문서 내의 데이터에 의미를 부여하고 데이터를 상호 연결하는 데이터 중심의 웹(Web of Data)으로의 패러다임 전환을 의미한다. 문서 중심의 웹에서 데이터 중심의 웹으로의 전환을 위해서는 공통된 언어 및 공통된 구조가 필요한데 여기서 공통된 언어와 구조는 메타데이터 모델(metadata model) 또는 공통 어휘(common vocabulary)를 의미하며 데이터를 공통된 어휘와 메타데이터 모델로 표현하기 위한 프레임워크로 RDF

(Resource Description Framework)와 OWL(Web Ontology Language) 등의 시맨틱 웹 언어를 제시하고 있다(김성혁 2011). 즉, Linked Data는 인간은 물론 기계가 읽고 처리할 수 있는 형태로 데이터를 출판하고 공개하여 데이터의 매쉬업(Mash-Up)을 통해 웹 상에서 데이터의 자유로운 탐색이 가능하게 하는 것을 지향한다.

버너스 리가 2006년 연설을 통해 Linked Data의 중요성을 역설한 뒤, W3C eGovernment Internet Group의 SWEO(Semantic Web Education & Outreach Interest Group)에 의해 2007년 Linking Open Data Project(이하 LOD 프로젝트)가 본격화되었다. 이는 데이터 셋을 공개하고 기계 가독형 시맨틱 웹 언어를 통해 출판(publishing)하여 다양한 데이터 간의 연계를 규정함으로써 데이터를 통한 웹의 확장을 목적으로 한다.

LOD 프로젝트는 위키피디아의 RDF 버전인 DBPedia를 시작으로 출발한 이래 현재까지 급속한 성장을 보이고 있다. 〈그림 3〉과 같이 2007년의 12개의 데이터 셋에 그치던 것이 2011년 10월에는 295개의 데이터 셋이 LOD 프로젝트에 발행되었다. LOD 다이어그램에서 원형의 노드(node)는 Linked Data형태로 출판된 데이터 세트를 의미한다. 화살표는 데이터 세트 간의 연계로, 화살표의 두께가 두꺼울수록 데이터 세트간의 연계가 많이 이루어지고 있음을 의미한다. 또한 각 노드의 크기는 데이터 세트의 크기를 의미한다. 2007년 5월과 2011년 10월을 기준으로 비교하여 볼 때 LOD의 가장 중심에는 변함없이 DBpedia가 위치함을 알 수 있는데, 이는 다양한 Linked Data 프로젝트에서 DBpedia와의 연계를 시도하고 있음을 의미한다. 국내에서 LOD에 발행된 Linked Data 데이터 세트는 BibleOntology가 유일하며 이 역시 DBPedia와의 연계를 시도하고 있다.

현재 Linked Data Cloud는 DBPedia는 물론, 뉴스 및 음악, 소셜웹, 지리정보, 인구정보, 도서관, 생명과학 등 다양한 분야로 확산되고 있으며 최근에는 미국과 영국의 정부 데이터의 Linked Data화를 위한 시도가 이루어지고 있어, 분야를 망라하여 데이터의 공개 및 Linked Data에 대한 관심이 높아지고 있음을 알 수 있다(김성혁 2011).



2007년 5월 Linked Data Cloud　　2011년 10월 Linked Data Cloud

〈그림 3〉 Linked Data Cloud현황(http://richard.cyganiak.de/2007/10/lod/)

3.2 Linked Data 현황

3.2.1 도서관 Linked Data 현황

Linked Data를 도서관에 적용시키는 예는 김성혁(2011)의 연구에서 자세하게 조사되었으며 크게 전거통제와 서지데이터의 Linked Data발행으로 구분할 수 있다.

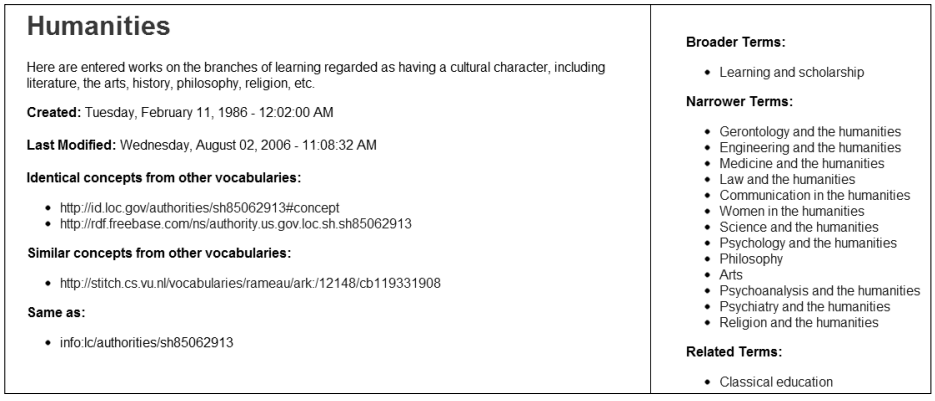
첫 번째는 전거데이터의 Linked Data구축이다. 주제 전거통제나 분류를 위해 사용되는 시소러스, 주제명 표목표와 같은 지식조직시스템(Knoweldge Organization Systems)을 SKOS(Simple Knowledge Organization Systems)을 이용하여 Linked Data로 발행하고 FOAF(Friend-Of-A-Friend)를 기반으로 인명 및 기관전거데이터를 Linked Data로 발행하는 것이 대표적이다.

전거데이터에 대한 Linked Data의 구현은 중요한데, 이는 전거데이터가 정보의 맥락에 대한 정보를 제공하고 있어 정보의 검색은 물론 정보의 식별에 있어서 유용하기 때문이다(류미숙 2009). 즉, 전거데이터의 Linked Data가 되어야 진정한 서지데이터의 공유 및 상호운용성 확보가 가능하다.

주요 사례로 미국 국회도서관의 주제명 표목표(LCSH)를 들 수 있다. 각 주제명 표목표를 SKOS 모델을 사용하여 변환하였으며 LCCN이라는 주제명 표목표 식별자를 기반으로 URI를 부여하고 이를 Linked Data로 출판하였다. 이는 사용자 관점에서 데이터에 대한 유연한 접근은 물론, 다양한 포맷으로 주제명 표목표의 다운로드가 가능하게 하여 필요에 따라 미의회도서관의 주제명 표목표 Linked Data를 재사용하도록 하기 위함이다. Linked Data는 미의회도서관 관점에서 기존의 도서관에서만 사용되던 범위를 확장하여 미의회도서관 데이터의 재사용성을 향상시키며 이를 통해 미의회도서관 데이터 노출을 극대화시킬 것으로 기대된다(오원석 2009).

다음은 "Humanities"라는 미의회도서관 LCSH 주제명에 대한 연관어 및 정보에 대한 화면(〈그림 4〉 참조)이다. 미의회도서관 LCSH 웹사이트에서는 주제어의 RDF파일을 살펴볼 수 있다.

두 번째는 서지 데이터의 Linked Data 발행

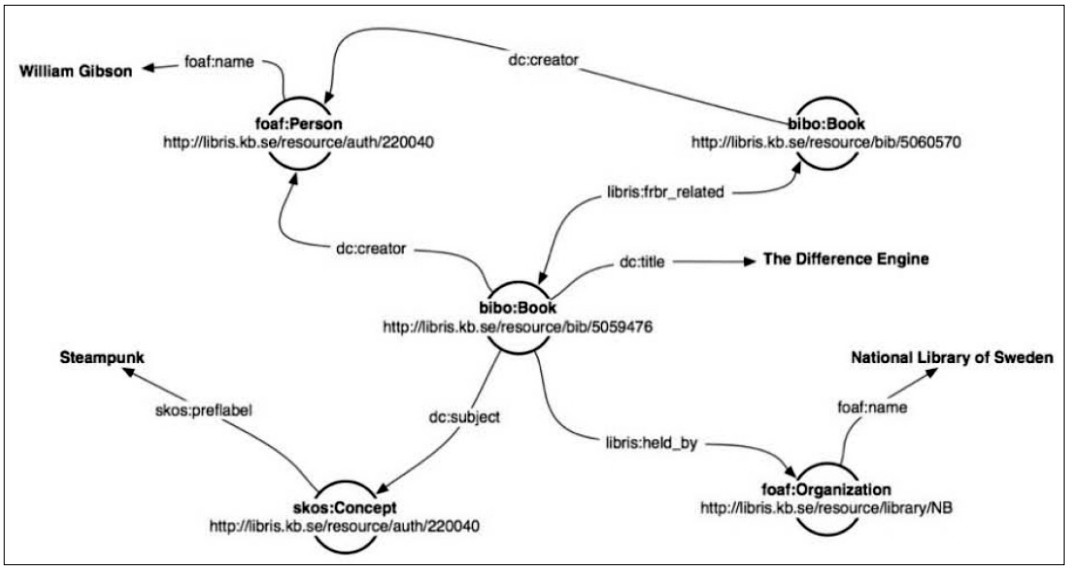


〈그림 4〉 LCSH 주제명

이다. 서지 데이터의 Linked Data화를 위해서는 웹 온톨로지와 같은 기계가독형으로 서지데이터를 변환할 수 있는 프레임워크가 필요하다. 서지데이터의 Linked Data를 위해 주로 사용되는 온톨로지는 RDF/OWL로 Dublin Core나 Bibliontology 어휘를 기반으로 주로 구축되어 왔다.

대표적인 예가 스웨덴 국립도서관 종합목록(LIBRIS), 헝가리 국립도서관, 유럽 원자핵연구기관(CERN)도서관이다. LIBRIS는 전체 서지데이터를 Linked Data로 구현한 최초의 시도이다. LIBRIS는 단지 서지데이터를 기계가독형 데이터로 구현하는 것에 제한되지 않고, DBPedia, MusicBrains, VIAF, LCSH 등 다른 데이터 셋과 연계함으로써 서지 데이터의 접근 및 이용을 가능하게 하였다(김성혁 2011). LIBRIS는 서지데이터의 Linked Data 발행을 위해 Bibliontology, Dublin Core, FOAF, SKOS를 사용하였다. 또한 헝가리 국립도서관도 OPAC과 디지털 도서관 데이터를 Linked Data로 발행하였고, 유럽 원자핵연구기관(CERN)도서관 역시 서지데이터를 오픈데이터로서 공개하였다. 이러한 시도는 도서관 서지데이터의 자유로운 접근과 이용을 통해 도서관 및 다른 데이터간의 매쉬업(Mash-Up)을 확보하고자 하는 것이다(오원석 2009)(〈그림 5〉 참조).

도서관 분야의 Linked Data 사례는 끊임없이 확산되고 있다. 독일국립도서관은 2012년 5월 12일 전거데이터를 Linked Data로 제공한다고 발표하였으며, 영국, 체코, 핀란드와 같은 여러 국가의 국립도서관에서 전거데이터 및 서지데이터의 Linked Data화를 추진하고 있다. 또한 OCLC의 VIAF(Virtual International Authority File)는 인명에 대한 전거 데이터 항목에 URI를 부여하여 전거데이터의 Linked Data로 활용이 가능하도록 공개하고 있다. 국내에서도



〈그림 5〉 LIBRIS Linked Data의 예(Anders 2009)

국립중앙도서관이 2011년 인명전거 및 주제전거데이터의 Linked Data 구축사업을 통해 한국형 Linked Data Cloud의 기반을 마련하였으며 이를 위해 Dublin Core, FOAF, SKOS를 사용하였다.

3.2.2 정부세트 Linked Data현황

공공분야에서도 Linked Data를 구현하는 움직임이 있으며 미국과 영국이 그 대표적인 사례이다.

미국은 오바마 대통령 정권 이후 정부 투명성 및 공개성 확보 차원에서 정부 데이터 접근을 자유롭게 개선하려는 시도가 이루어지고 있다. 이 연구는 시맨틱 웹의 창시자인 버너스 리와 짐핸들러 교수가 주도하여 진행하고 있는 것으로 Data.gov 사이트의 데이터를 공개하고, 이 중 일부를 RDF로 변환하여 Linked Data Cloud와 연계하는 것을 목적으로 한다. 이는 정부 데이터의 풍요한 가치를 인정하고 정부 데이터가 공공행정 영역에서만 사용되던 기존의 방식을 벗어나 일반 이용자의 사용을 통해 공공정부의 부가가치를 증진시키기 위함이다.

미국 정부데이터의 공개는 짧은 시간 내에 놀라운 가시적인 변화를 보여주고 있다. Data.gov는 2009년 시작당시 47개의 데이터 세트에 불과하였으나 2010년 825만개의 데이터 세트로 급속한 성장을 보였으며, 액세스 역시 9,760만 건에 이르렀다. 또한 정부기관이나 지자체가 가공되지 않은 데이터(Raw Data)를 공개하는데 동참하기 시작하였고, 다양한 영역에서 공개된 공공데이터를 활용하려는 움직임이 나타나고 있다(오원석 2009).

영국 Data.gov.uk 역시 국민의 알권리 및 정부의 투명성을 위하여 공공 정보 공개사이트 구축하고 있으며 이를 통해 정부 데이터 공개 및 데이터의 연계를 시도하고 있다. 일반 이용자의 정부 데이터에 대한 자유로운 접근이 가능해짐은 물론 자신만의 방식으로 데이터의 재사용이 가능해질 것으로 기대한다. 이 연구는 고든 브라운 총리가 직접 버너스 리를 책임자로 임명하는 등 정부차원에서 적극적인 지원이 이루어지고 있다.

4. 선행연구

김성혁(2011)은 도서관에서의 Linked Data 활용의 중요성을 논하면서 Linked Data의 동향, 도서관 적용가능성, 기대효과를 제시한바 있다. 김성혁의 연구에서 기존의 Linked Data가 도서관에 적용된 예를 서지데이터와 전거데이터를 중심으로 고찰하였으며 이를 통해 국내의 적용가능성을 제시하였다.

조명대(2010)는 문서중심의 웹에서 데이터 중심의 웹으로의 전환이 필요함을 역설하고 있다. 또한 도서관 역시 웹의 패러다임의 전환에 편승해야 하며 이를 위해 Linked Data를 활용해야 한다고 강조하고 있다. 또한 미의회도서관 및 LIBRIS와 같은 도서관 분야의 Linked Data의 적용을 조사하고, 한국영화 주제 분야를 중심으로 Linked Data의 구현가능성을 모색하였다.

이연호 외(2011)는 동영상 콘텐츠에 등장하는 객체에 직접 부가정보를 기술하는 키워드 기반 어노테이션 연구를 수행하였다. 다수의 어노테이션을 수행하는 사용자들이 협업적으로 동영상 콘텐츠에 등장하는 객체에 대한 어노테이션을

수행할 수 있도록 하는 모듈을 Linked Data를 기반으로 제시하였다. 이를 통해 어노테이션 데이터를 온톨로지 형태로 표현하여 다수의 사용자가 필요시에 어노테이션 데이터를 객체 URI를 이용하여 Linked Data로부터 콘텐츠를 보다 쉽게 공유하고 확장할 수 있도록 하였다.

정효숙, 김희진, 박성빈(2011)은 일반 사용자측면에서 Linked Data를 생성하는 것을 가능하게 하기 위하여 서로 다른 RDF 문서에 존재하는 데이터를 연결하여 Linked Data를 생성할 수 있는 시스템을 연구한바 있다. owl에서 〈owl:sameAs〉은 두 개체를 서로 연결시키는 속성으로써, 두 개의 URI가 실제로 같은 것을 참조하는 것을 의미한다(dereferencing). 본 연구에서는 〈owl:sameAs〉의 속성을 시스템에서 용이하게 사용할 수 있도록 하여 다른 RDF 문서에 존재하는 데이터를 손쉽게 연결하도록 하는 기능을 제공하였다. 이를 통해 교수자가 컨텐츠를 제작할 시에 본인이 직접 제작한 것은 아니지만 학습자의 학습 수준 및 목적에 맞는 다양한 데이터를 찾아 Linked Data로 연결함으로써 학습자의 학습을 조력할 수 있음을 강조하였다.

Fripp(2010)은 전통적인 패싯 분류 스킴과 Linked Data 환경에서의 시맨틱 웹 문서 어노테이터 간의 관계를 분석하기 위하여 패싯분류와 Linked Data 아키텍처의 원칙의 유사성을 논의하였다. DBPedia 온톨로지 콘텐츠를 패싯으로 분류하여, 장소(Place), 인물(Person), 업적(Work), 기관(Organization), 건물(Building), 종(Species)으로 클래스를 추출하고 토스카(Tosca)에 대한 위키피디아 아티클을 바탕으로 패싯을 추출하여 Linked Data를 구성하고 Thinkpedia라는 도구를 사용하여 시각화하였다.

Stuart(2010)는 공공데이터의 가치를 높이 평가하고 이에 대한 처리가 중요하다고 강조하였다. 또한 미국, 영국 등 선진국에서는 공공 데이터의 활용이 정부의 대민 서비스와 직결되어 있음을 인지하고 공공 데이터의 활용을 정부 정책으로 진행되고 있음을 소개하였다. 이렇게 공공데이터의 처리를 Linked Data화 하기 위해서는 무엇보다도 데이터 공개에 대한 인식이 확산되어야 하며, RDF로의 전환 및 RDF파일의 일반인의 접근을 허용하는 것이 이루어져야 한다고 강조하였다.

국내의 사례분석을 통해 Linked Data의 영역이 학습, 동영상, 도서관 등 다양한 분야에 활용될 수 있음을 파악할 수 있다. 또한 Linked Data가 추구하는 데이터의 재사용 및 공유가 가능하기 위해서는 우선적으로 온톨로지 기반의 Linked Data의 구현 및 개방이 이루어져야 함을 파악할 수 있다.

5. 연구방법 및 내용

본 연구에서는 기존의 국가기록원의 전거데이터 셋을 RDF/OWL로 모델링하는 과정을 거쳐 Linked Data를 구축하는 방법을 채택하였다. 또한 안태성(2010)이 제시한 Linked Data 방법론에 근거하여 본 연구에 맞게 수정하였다.

어휘(Vocabulary)는 합의된 표준(de facto)이므로 의미를 공유할 수 있으며, 새로운 어휘의 제안 및 재사용이 가능하다는 장점이 있다. 이에 본 연구는 국가기록원 기록물의 Linked Data화를 위해 RDF/OWL기반으로 Dublin Core, SKOS(Simple Knowledge Organization

System)를 활용하였으며, SKOS와 Dublin Core의 요소로 표현되지 않는 어휘에 한해서 확장하기로 하였다. 또한 온톨로지 설계도구로 W3C 시맨틱 웹 표준을 준수하는 GUI 환경의 온톨로지 모델링 도구인 TopBraid Composer TM을 사용하였다.

자세한 방법론은 다음과 같다.

5.1 RDF/OWL 활용

전거데이터를 Linked Data화하기 위하여 RDF/OWL을 기반으로 하였다. 이는 Linked Data가 지향하는 기계 가독형 웹을 위한 Linked Data의 기본 원칙을 준수하기 위함이다. RDF는 Linked Data로 구현하기 위한 시맨틱 웹 표준 데이터 모델로 트리플 구조를 가진 표준이며 XML기반이다. RDF는 자원(Resource)간의 관계를 그래프로 표현하기 위한 프레임워크이다.

여기서 Resource는 URI를 갖는 모든 것을 의미하며, 선언문을 통해 기술하고자 하는 자원을 의미한다. Description은 자원의 속성이나 관계를 설명하기 위한 것이며, Framework은 자원과 자원에 대한 속성 및 관계를 기술하기 위한 모델 및 문법을 지칭한다. 여기서 자원과 관계는 URI를 갖는다. RDF는 주어(subject), 술서(predicate), 목적어(object)의 모델로 정보를 설명한다.

다음 예를 통해 살펴보면,

> http://www.w3.org/DesignIssues/LinkedData.html의 생성자(Creator)는 http://dbpedia.org/resource/Berners_Lee이다.

와 같은 선언문(statement)이 있다면, 주어는 URI인 http://www.w3.org/DesignIssues/LinkedData.html, 술어는 URI인 http://purl.org/dc/elements/1.1/creator, 목적어는 http://dbpedia.org/resource/Berners_Lee가 된다. 주어는 항상 URI를 가져야 하며, 목적어는 URI를 가지거나 문자열(Literal)일 수 있다.

OWL(Web Ontology Language)은 웹의 리소스를 기술하는 온톨로지 언어로 목적은 RDF 스키마와 동일하게 클래스, 속성, 관계를 정의하기 위함이다. 그러나 RDF 스키마가 지니는 클래스의 비교차성, 논리적 조합 등의 표현제약, 속성의 상세유형 표현제약과 같은 한계를 극복하기 위하여 보다 풍부한 표현을 가능하게 한다. 예를 들어, OWL은 속성은 데이터유형(Datatype Property), 객체속성(Object Property), 주석속성(Annotation Property)으로 구분한다. 객체속성은 자원과 자원을 연결하기 위함이며, 데이터유형 속성은 자원과 값, 주석속성은 자원의 값이 주석의 형태를 지니는 경우를 표현하기 위함이다.

5.2 주제전거 SKOS 어휘 활용

국가기록원 주제기반 어휘를 Linked data하기 위하여 SKOS(Simple Knowledge Organization System)를 기반으로 한다. SKOS는 지식 조직 체계를 분산된 데이터를 연결된 방식으로 표현하기 위한 도구를 제공하기 위해 만들어졌다. SKOS는 W3C에서 시맨틱 웹 기반의 지식 구조화 프레임워크 표준으로 개발하여 2005년 발표한 것으로, RDF기반의 프레임워크로서 분산 정보 처리 환경에서 유연하고 확장이 가능

〈표 1〉 전거데이터 Linked Data 전환 고려사항

세부단계	설명	고려사항
기록물 데이터 분석	기록물 데이터를 RDF모델링하기 위하여 주어, 술어, 목적어 관계 분석	출처원칙을 적용하기 위한 RDF모델링
		기록물과 전거데이터와의 연계표현
주제어분석	국정분야 및 주제유형 체계 분석 뒤 Linked Data로 전환고려	국정분야 및 주제유형 검색에서 SKOS를 적용할 범위 선정
주제어 온톨로지 모델링	국정분야 및 주제유형의 온톨로지 그래프화	클래스, 속성, Cardinality 고려
주제어 Linked Data 전환	국정분야 및 주제유형 분석 뒤 Linked Data로 전환	국정분야 및 주제유형 모델링에서 SKOS외의 어휘를 사용하여 확장할 요소선택
		LCSH SKOS활용 벤치마킹
생산기관 전거데이터 분석	ISAAR(CPF)에 대한 분석으로 기관 전거데이터를 Linked Data로 모델링하기 위한 분석	생산기관계층에 대한 모델링(생산기관 분야 고려)
전거데이터 온톨로지 모델링	ISAAR(CPF)에 근거하여 기술된 요소 표현여부(식별영역, 기술영역, 관계영역, 통제영역)	클래스, 속성, Cardinality 고려
		Dublin Core, SKOS 적용할 범위선정
생산기관 Linked Data 전환	RDF/OWL 형식으로 구성된 그래프 구성 및 전환	LIBRIS 및 국립중앙도서관 인명 및 기관 전거 활용 벤치마킹

하며, 공유와 재사용할 수 있는 다양한 지식 구조와 방법을 제공한다(한성국, 이현실 2006, 290). SKOS는 통제어, 용어, 시소러스, 택소노미, 주제표목 등을 RDF 표현체계를 통해 모델화 하는데 적합한 기능을 갖고 있다. SKOS를 이용하면 기존의 지식 체계를 시맨틱 웹으로 표현하는 것이 가능해진다.

SKOS는 시소러스 용어의 상위어(BT), 하위어(NT), 관계어(RT) 등의 계층 관계를 정의할 수 있는 어휘 외에도 정의, 범위, 범례를 작성할 수 있는 주석어휘, 다른 주제어휘와 연계할 수 있는 어휘, 어휘의 컬렉션을 지정할 수 있는 어휘 등 다양한 어휘를 사용하고 있다. 이에 SKOS를 기반으로 하되, 국가기록원의 국정분야별, 주제유형별 어휘를 표현할 수 없는 경우는 확장하기로 한다.

다음은 기록물 전거데이터를 Linked Data화 할 경우 고려사항이다(〈표 1〉 참조).

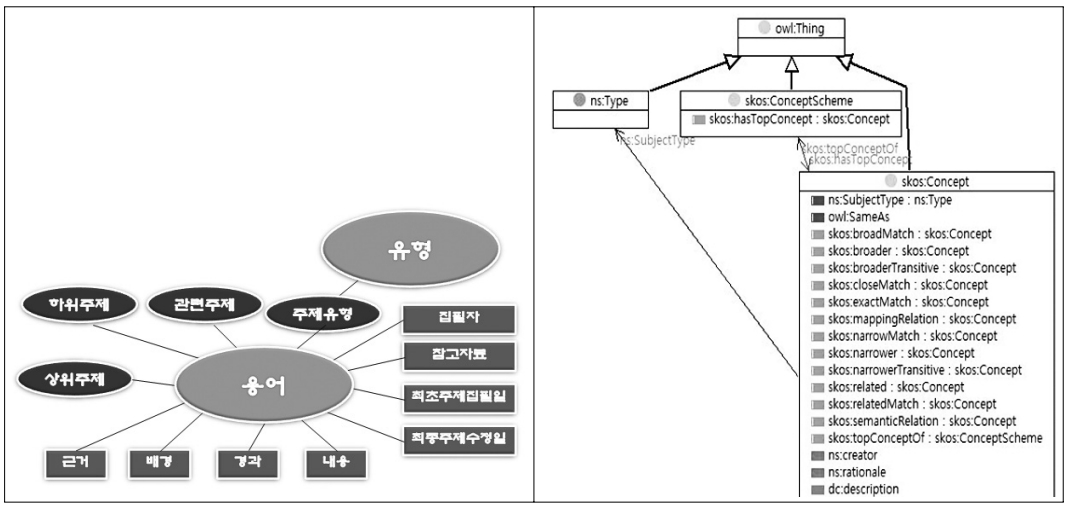
6. 연구결과

6.1 주제전거 Linked Data 전환

주제전거를 OWL을 사용하여 Linked Data로 전환하기에 앞서, 주제전거의 모델링이 이루어졌다. 이 단계는 시맨틱 웹 용어에서 온톨로지를 구현하기 위해 클래스(Class)와 속성(Property)을 구분하는 것이다. 클래스는 온톨로지에서 "kinds of things"을 의미하는 것으로 예를 들어 웹페이지, 사람, 문서유형, 추상적인 개념과 같은 유형(Type)이나 카테고리(Category)를 의미한다. 속성은 특정 클래스를 기술하기 위하여 사용하는 것이다(RDF Primer 2004).

〈그림 6〉과 같이 주제어와 유형을 클래스로, 나머지 주제전거의 요소를 속성으로 모델링하여 주제어의 국정분야 및 주제유형이 클래스를 통해 연결되도록 하였다. 또한 주제유형이라는 객체속성을 사용하여 주제어와 유형 클래스를 연결하였다.

각 속성은 OWL의 객체속성(Object Property), 데이터유형 속성(DataType Property), 주석 속성(Annotation Property)을 사용하여 구분하였으며, SKOS, Dublin Core를 최대한 재사용하였으나 속성의 의미상 재사용할 수 없는 것에 한해 속성어휘를 확장하였다. 예를 들어, 주제어 집필 근거, 집필자, 주제유형은 확장하였다(〈표 2〉 참조).



〈그림 6〉 주제전거 모델링

〈표 2〉 주제전거 사용어휘

구조	OWL 속성	사용어휘
근거	데이터유형 속성	ns:rationale
배경	주석 속성	ns:background
경과	주석 속성	skos:historyNote
내용	주석 속성	dc:description
참고자료	데이터유형 속성	dc:source
집필자	데이터유형 속성	ns:creator
최초 주제 집필일	데이터유형 속성	dcterms:created
최종 주제 수정일	데이터유형 속성	dcterms:modified
상위주제	객체 속성	skos:broader
하위주제	객체 속성	skos:narrower
관련주제	객체 속성	skos:related
주제유형	객체 속성	ns:subjectType

이렇게 모델링된 주제전거데이터는 RDF/OWL을 기반으로 구현되었다. 다음 RDF/OWL 코딩파일을 살펴보면 주제어(skos:Concept)는 객체속성(hasType)을 통해 클래스 유형(Type)과 연결된 것을 파악할 수 있으며(〈그림 7〉 참조), 또한 유형의 인스턴스로 국가기록원 10개의 주제유형을 추가하여 주제어와 유형이 연계되도록 하였다. 각 주제어(인스턴스)가 모델링 어휘를 기반으로 다른 어휘와의 관계가 표현된 것을 알 수 있다(〈그림 8〉 참조).

이렇게 구현된 스키마 레벨에 주제어 인스턴스를 추가한 OWL파일은 시각화(visualization)를 통해 이용자가 검색할 수 있도록 하였다. 즉, 〈그림 9-1〉을 살펴보면, 국민연금의 상위어(사회보험제도), 하위어(유족연금; 국민연금기금; 장애연금; 국민연금급여; 국민연금관리; 노령연금; 국민연금개혁; 국민연금법제정), 관계어(군인연금)와의 관계를 한 화면에서 파악할 수 있도록 하였으며, owl:sameAs를 통해 다른 리소스의 URI와 연계하여 Linked Data에서 주장하는 데이터 간의 연계가 가능하도록 하였다. 본 예는 국가기록원 주제명 전거와 Dbpedia, 국립중앙도서관 주제명과 연결한 것이다.

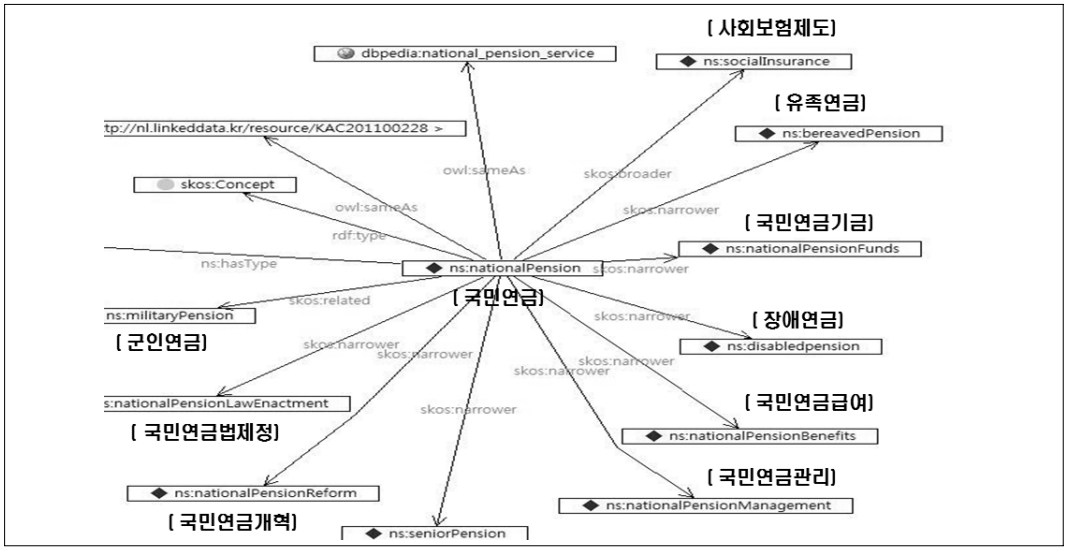
용어를 통해 다른 용어의 관계로의 확장 역시 가능하다. 〈그림 9-2〉와 같이 국민연금의 상위어인 사회보험제도의 상위어, 하위어, 관계어를 한 화면에서 살펴볼 수 있어서, 용어간의 관계를 통해 주제명 사이의 자유로운 브라우징이 가능하다. 또한 시각화화면을 통해 주제명의 전거데이터를 살펴볼 수 있다. 〈그림 9-3〉과 같

```
<owl:Class rdf:ID="Type">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
</owl:Class>
<owl:ObjectProperty rdf:ID="hasType">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Object property 1</rdfs:label>
  <rdfs:domain rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <rdfs:range rdf:resource="#Type"/>
</owl:ObjectProperty>
 <ns:Type rdf:ID="business">
<ns:Type rdf:ID="conference">
<ns:Type rdf:ID="others">
<ns:Type rdf:ID="event">
 <ns:Type rdf:ID="people">
<ns:Type rdf:ID="accident">
```

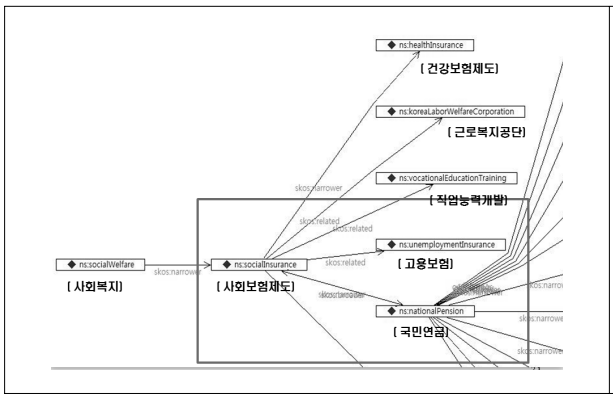
〈그림 7〉 주제전거 클래스 및 속성 코딩

```
<skos:Concept rdf:about="http://archives.linkeddata.go.kr/subjects#nationalPension">
   <skos:prefLabel rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
     >National Pension</skos:prefLabel>
   <ns:hasType rdf:resource="http://archives.linkeddata.go.kr/subjects#policy"/>
   <skos:broader rdf:resource="http://archives.linkeddata.go.kr/subjects#socialInsurance"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#nationalPensionFunds"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#disabledpension"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#nationalPensionBenefits"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#nationalPensionManagement"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#seniorPension"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#bereavedPension"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#nationalPensionReform"/>
   <skos:narrower rdf:resource="http://archives.linkeddata.go.kr/subjects#nationalPensionLawEnactment"/>
   <skos:related rdf:resource="http://archives.linkeddata.go.kr/subjects#militaryPension"/>
   <owl:SameAs rdf:resource="http://dbpedia.org/resource/National_Pension_Service"/>
```

〈그림 8〉 주제전거 인스턴스 코딩



〈그림 9-1〉 주제전거데이터 시각화



〈그림 9-2〉 주제전거데이터 시각화 확장

〈그림 9-3〉 주제전거데이터 시각화 전거데이터 상세

이 국민연금의 상위어인 사회보험제도의 전거데이터를 통해 사회보험제도의 근거, 생산자, 연혁 등을 살펴 볼 수 있다. 이는 용어간의 전거데이터를 탐색하는 것은 물론 전거데이터를 살펴봄으로써 기록관리 전거데이터에서 중시하는 맥락정보를 제공하는 것을 가능하게 한다.

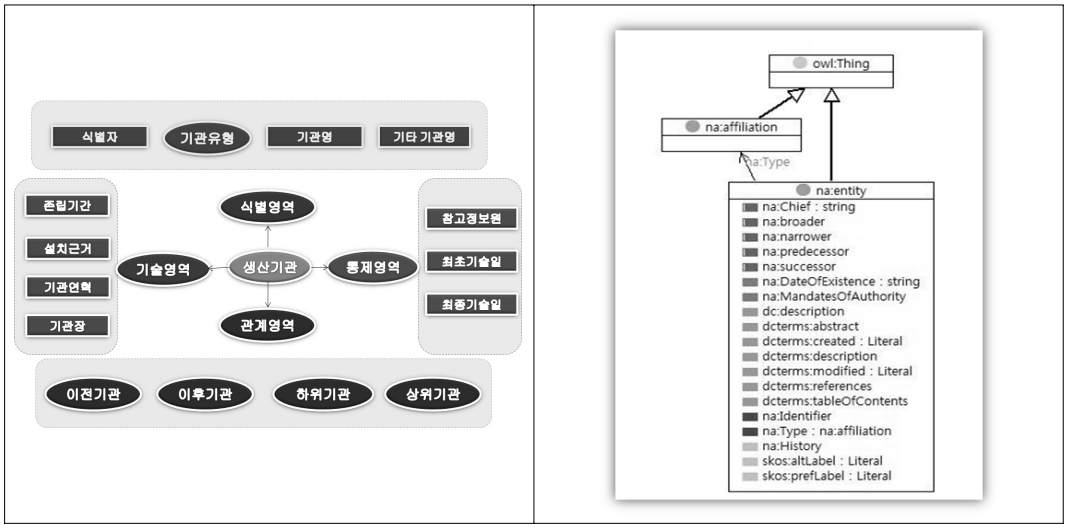
6.2 생산기관 전거 Linked Data 전환

생산기관 전거를 OWL을 사용하여 Linked Data로 전환하기에 앞서, 주제전거와 마찬가지로 모델링이 이루어졌다. 생산기관 전거데이터는 식별영역, 기술영역, 관계영역, 통제영역으로 구분하고 있으며, 각 영역은 여러 요소로 구

분되고 있다. 이를 표현하기 위하여 본 연구에서는 네 가지 영역을 속성(Property)으로, 영역의 요소를 하위 속성(SubProperty)으로 구분하여, 속성과 하위속성의 관계를 통해 각 영역의 요소 출처를 밝힐 수 있도록 하였다. 또한 생산기관 전거는 생산기관의 기관유형을 주의 깊게 살펴볼 필요가 있다. 기관유형은 생산기관의 식별영역에 속하는 동시에 기관유형에도 속하는 요소로 각 기관은 전거데이터 지침에 의거하여 국가행정조직, 자치행정조직, 교육행정조직, 입법조직 등 소속기관으로 구분된다. 또한 각 소속기관은 계층구조에 따라 상세화된다. 예를 들어, 보건복지부는 '국가행정조직 - 중앙행정기관 - 부' 계층구조의 예가 된다. 본 연구에서는 이를 표현하기 위하여 기관 및 소속기관(affiliation)을 클래스로, 나머지 생산전거의 요소를 속성으로 모델링하여, 생산기관의 전거데이터의 식별영역 중 기관유형 속성을 통해 소속기관(affiliation)으로 연계되도록 하였다(〈그림 10〉 참조).

각 속성은 OWL의 객체속성(Object Property), 데이터유형 속성(DataType Property), 주석속성(Annotation Property)을 사용하여 구분하였으며, SKOS, Dublin Core를 최대한 재사용하였다. 그러나 기관전거의 경우 SKOS, Dublin Core를 재사용할 수 없는 경우가 다수 있었는데, 이는 생산기관 전거의 경우 기록물의 출처를 제공하는 중요요소이므로 기관의 맥락정보를 제공하는 요소가 많기 때문이다(〈표 3〉 참조).

이렇게 모델링된 생산전거데이터는 RDF/OWL을 기반으로 구현되었다. 다음 RDF/OWL로 코딩파일을 살펴보면 생산기관(entity)은 객체속성(Type)을 통해 클래스 소속기관(affiliation)과 연결된 것을 파악할 수 있으며(〈그림 11〉 참조), 각 생산기관(인스턴스)이 모델링된 어휘를 기반으로 다른 생산기관과 관계가 표현된 것을 알 수 있다(〈그림 12〉 참조).



〈그림 10〉 생산전거 모델링

〈표 3〉 생산기관전거 사용어휘

구조	OWL 속성	사용어휘	속성계층
식별영역	**객체 속성**	**na:identification**	**속성**
기관영역	데이터유형 속성	na:type	하위속성
기관코드	객체 속성	na:identifier	하위속성
기관명	데이터유형 속성	skos:prefLabel	하위속성
기타 기관명	데이터유형 속성	skos:altLabel	하위속성
기술영역	**객체 속성**	**na:description**	**속성**
존립기간	데이터유형 속성	na:dateOfExistence	하위속성
설치근거	데이터유형 속성	na:mandatesOfAuthority	하위속성
기관연혁	주석 속성	na:history	하위속성
기관장	데이터유형 속성	na:chief	하위속성
관계영역	**객체 속성**	**na:relationships**	**속성**
이전기관	객체 속성	na:predecessor	하위속성
이후기관	객체 속성	na:successor	하위속성
하위기관	객체 속성	skos:narrower	하위속성
상위기관	객체 속성	skos:broader	하위속성
통제영역	**객체 속성**	**na:control**	**속성**
참고정보원	데이터유형 속성	dct:references	하위속성
최초기술일	데이터유형 속성	dct:created	하위속성
최종기술일	데이터유형 속성	dct:modified	하위속성

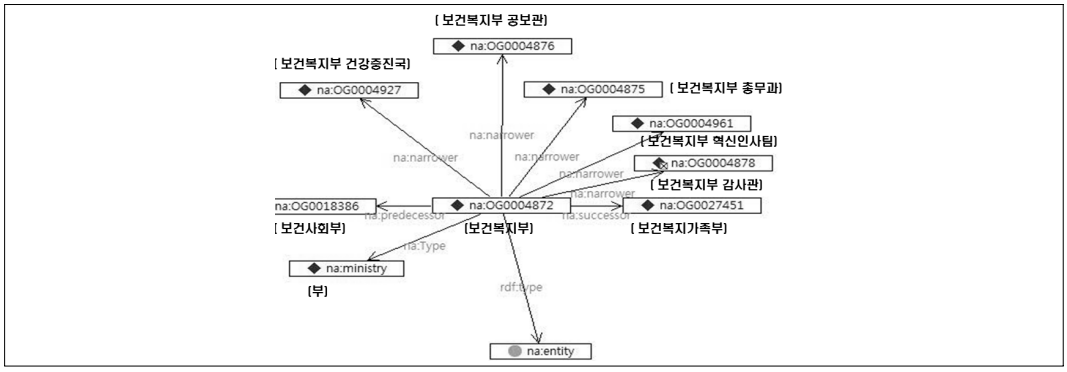
```
<owl:Class rdf:about="http://archives.linkeddata.go.kr/organization#entity">
<owl:Class rdf:about="http://archives.linkeddata.go.kr/organization#affiliation">
<rdf:Property rdf:about="http://archives.linkeddata.go.kr/organization#Type">
  <rdfs:range rdf:resource="http://archives.linkeddata.go.kr/organization#affiliation"/>
  <rdfs:domain rdf:resource="http://archives.linkeddata.go.kr/organization#entity"/>
  <rdfs:subPropertyOf rdf:resource="http://archives.linkeddata.go.kr/organization#identification"/>
</rdf:Property>
<owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#identification">
<owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#control">
<owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#relationships">
<owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#description">
<owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#broader">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >broader</rdfs:label>
  <rdfs:subPropertyOf rdf:resource="http://archives.linkeddata.go.kr/organization#relationships"/>
  <rdfs:domain rdf:resource="http://archives.linkeddata.go.kr/organization#entity"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#narrower">
  <rdfs:subPropertyOf rdf:resource="http://archives.linkeddata.go.kr/organization#relationships"/>
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >하위기관</rdfs:label>
  <rdfs:domain rdf:resource="http://archives.linkeddata.go.kr/organization#entity"/>
  <owl:inverseOf rdf:resource="http://archives.linkeddata.go.kr/organization#broader"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#successor">
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="http://archives.linkeddata.go.kr/organization#predecessor"/>
  </owl:inverseOf>
  <rdfs:domain rdf:resource="http://archives.linkeddata.go.kr/organization#entity"/>
```

〈그림 11〉 생산기관전거 클래스 및 속성 코딩

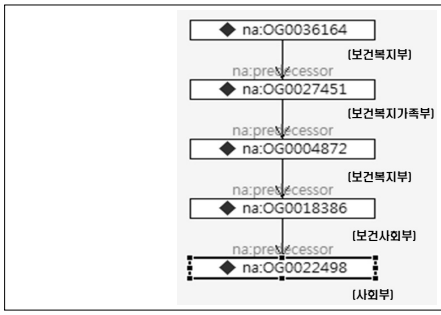
```
<na:entity rdf:about="organization#OG0004872">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >보건복지부</rdfs:label>
  <na:Type rdf:resource="organization#ministry"/>
  <na:narrower rdf:resource="organization#OG0004927"/>
  <na:narrower rdf:resource="organization#OG0004876"/>
  <na:narrower rdf:resource="organization#OG0004875"/>
  <na:narrower rdf:resource="organization#OG0004878"/>
  <na:narrower rdf:resource="organization#OG0004961"/>
  <na:predecessor rdf:resource="organization#OG0018386"/>
  <na:successor rdf:resource="organization#OG0027451"/>
```

〈그림 12〉 생산기관전거 클래스 및 인스턴스 코딩

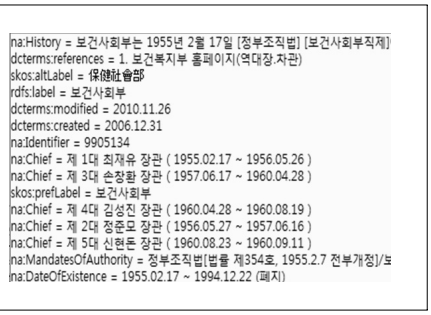
이렇게 구현된 OWL파일은 시각화(visualization)를 통해 이용자가 검색할 수 있도록 한다. 즉, 〈그림 13-1〉을 살펴보면, 보건복지부의 이전기관(보건사회부), 이후기관(보건복지가족부), 하위기관(보건복지부 건강증진국; 보건복지부 공보관; 보건복지부 총무과; 보건복지부 혁신인사팀; 보건복지부 감사관) 등을 파악할 수 있으며, 보건복지부가 중앙행정부의 부에 속해있음을 파악할 수 있다. 또한 〈그림 13-2〉와 같이 보건복지부의 전체 기관 연혁을 기관명의 확장을 통해 살펴볼 수 있는 것은 물론, 〈그림 13-3〉과 같이 시각화화면을 통해 특정 기관의



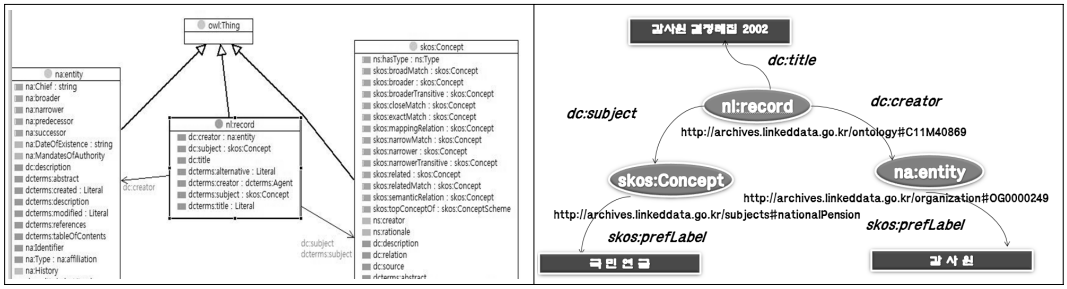
〈그림 13-1〉 생산전거데이터 시각화



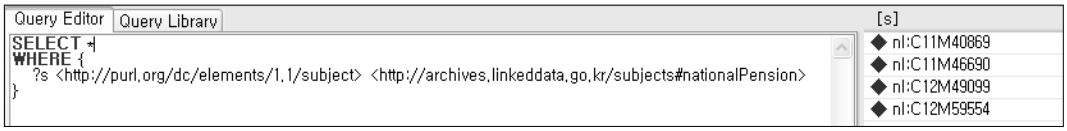
〈그림 13-2〉 생산전거데이터 시각화 확장



〈그림 13-3〉 생산전거데이터 시각화 전거데이터 상세



〈그림 14〉 기록물과 전거데이터의 연결



〈그림 15〉 SPARQL 질의

전거데이터를 살펴볼 수 있다. 이는 생산기관 간의 관계를 탐색하는 것은 물론 전거데이터를 제공함으로써 생산기관 전거데이터의 접근 및 정보제공을 용이하게 한다.

6.3 기록물 데이터와 전거연계

구축된 생산기관 전거데이터와 주제전거데이터를 기록물 데이터로 연결하기 위하여 기록물 데이터를 클래스로 선정하고 생산기관 및 주제전거데이터 온톨로지를 취합하였다. 또한 구축된 온톨로지를 기록물 데이터와 연계하기 위하여 생산자(dc:creator) 및 주제(dc:subject) 속성을 사용하여 연결하였다(〈그림 14〉 참조).

또한 SPARQL 질의를 통하여 주제어, 생산기관에 대한 검색, 또는 기록물과 주제어, 생산기관을 연계한 질의가 가능하도록 하였다. 다음의 예는 SPARQL 질의를 사용하여 국민연금이라는 주제의 기록물을 검색한 화면이다(〈그림 15〉 참조).

7. 결 론

공유, 개방, 참여로 대변되는 웹 정보환경에 대응하기 위하여 기존의 데이터 처리방식을 넘어 지능형 웹으로 전환이 요구된다. 지능형 웹은 웹 상의 모든 데이터가 연결되어 하나의 지식구조를 형성하는 개방형 구조로의 진화를 의미하는 것은 물론 인간을 비롯하여 기계도 이해할 수 있는 웹을 지칭한다. 버너스 리는 지능형 웹의 실현을 위해 시맨틱 웹의 구체적인 모범사례인 Linked Data를 제시하였으며 Linked Data의 가시화를 위해 각 국가의 핵심 데이터 세트를 선정하여 Linked Data로 출판하려는 노력이 요구된다.

이에 본 연구에서는 한국형 Linked Data Cloud를 위한 단계로 국가기록원 전거데이터 셋을 Linked Data화하는 것을 목적으로 한다. 국가기록원은 우리나라 기록 관리의 중심축이며 종이문서는 물론, 행정박물, 시청각 기록물에 이르기까지 다양한 형태의 기록물을 취급하고 있

다. 이러한 기록물의 활용을 위해 국가기록원 데이터의 개방 및 연계가 절실하다. 이를 위한 첫 번째 단계로 국가기록원은 국가기록원 전거데이터의 Linked Data화가 우선되어야 한다.

이를 위해 본 연구에서는 국가기록원 전거 데이터의 구조 및 검색시스템을 분석한 후, 전거데이터 셋의 온톨로지화, 온톨로지를 Linked Data로 전환하는 것을 목표로 한다. 기록물 기술표준인 기존의 국가기록원의 전거데이터 셋을 RDF/OWL로 모델링하는 과정을 거쳐 Linked Data를 구축하는 방법을 채택하였다. 또한 기록물의 Linked Data화를 위해 RDF/OWL기반으로 Dublin Core, SKOS(Simple Knowledge Organization System)를 활용하였고 온톨로지 설계도구로 W3C 시맨틱 웹 표준을 준수하는 GUI 환경의 온톨로지 모델링 도구인 TopBraid Composer TM을 사용하였다.

구축된 RDF/OWL 데이터는 시각화 및 SPARQL을 통해 전거데이터의 자유로운 확장, 전거데이터 용어간의 탐색, 용어의 상세 전거데이터의 접근이 가능하도록 하였으며, 기록물과의 연계를 통해 기존 기록물 검색에서 한계로 지적되었던 기록물과 전거데이터간의 연계의 부족을 보완하였다.

본 연구에서는 기록물 데이터의 Linked Data화는 이루어지지 않았으므로 기록물 데이터와 전거데이터와의 보다 유연한 연계 및 확장을 위해서 기록계층 및 ISAD에 근거한 기술요소 표현여부 등에 대한 분석을 통해 기록물 데이터의 Linked Data가 이루어져야 할 것이다. 또한 Linked Data가 제시하는 '연계(relationship)'를 구현하기 위하여 도서관, 박물관 등 다른 데이터 셋으로의 연계가 향후 계속 연구되어야 할 것이다.

이렇게 구축된 Linked Data는 웹에 출판(publishing)하는 것은 기록물 데이터를 다른 데이터 셋과 연계하여 자유롭게 활용 가능하게 함으로써 국내 국가기록원의 재사용 가능성을 높이는데 기여할 수 있을 것이다. 이는 기존의 이용자들뿐만 아니라 보다 넓은 이용자층에게 기록물 소장 자료의 존재를 노출시킬 수 있는 기회가 될 것이며, 정보 가치를 상승시키고 활용도를 향상시키는데 기여할 것이다. 또한 본 연구에서 제공하는 모델링은 다른 기관의 모델링에 사용될 수 있으므로 이를 통해 기록물 및 도서관 데이터의 Linked Data 모델링에 기여할 수 있을 것이며, 더 나아가 국가기록원 Linked Data의 권위를 신장시키는데 중요한 역할을 할 것으로 기대된다.

참 고 문 헌

김성혁. 2011. 도서관 자료공개와 활용을 위한 링크드 데이터 기술. 『국회도서관보』, 48(3): 24-33.

김태동. 2010. Gov2.0시대의 공공정보 공유·연계 강화를 위한 전략. 『국회도서관보』, 47(7): 22-28.

류미숙. 2009.『행정기관의 기록물 전거레코드 구축 및 활용에 관한 연구』. 석사학위논문. 부산대학교 대학원, 기록관리학협동과정.

배정현. 2007.『디지털 문화 유산 콘텐츠 공동활용을 위한 협력 방안 연구』. 석사학위논문. 연세대학교 대학원, 문헌정보학과.

서혜란. 2005. 도서관과 기록관의 협력방안 모색.『한구비블리아학회 발표논집』, 13: 111-128.

설문원. 2010. 기록 검색도구의 발전과 전망.『한국기록학연구』, 23: 3-43.

신동수, 이상철, 김신열. 2008. FOAF 온톨로지 기반 프로젝트 팀 구성 시스템.『한국정보과학회』, 35(2): 228-233.

안태성. 2010. Linked Open Data(LOD) 소개. [cited 2012.5.12]. 〈http://blog.saltlux.com/?p=2438〉.

오원석. 2009.『Linked Data 동향과 전망』. 서울: (주)탑쿼드란트코리아.

이만재. 2011. 빅데이터와 공공데이터의 활용.『Internet and Information Security』, 2(2): 47-64

이연호, 오경진, 신위살, 조근식. 2011. 링크드 데이터를 이용한 협업적 비디오 어노테이션 및 브라우징 시스템.『지능정보연구』, 17(3): 203-219.

정효숙, 김희진, 박성민. 2011. 일반인을 위한 링크드 데이터 생성 시스템 개발 및 활용.『컴퓨터교육학회논문지』, 14(2): 47-59.

한성국, 이현실. 2006. 시소러스를 활용한 온톨로지 구축방안 연구: 시소러스의 SKOS 변환을 중심으로.『한국비블리아학회지』, 17(1): 285-303.

Berners-Lee, T. 2006. Linked Data - Design Issues. Retrieved. [cited 2012.5.12]. 〈http://www.w3.org/DesignIssues/LinkedData.html〉.

Bizer, C., T. Heath, and T. Berners-Lee. 2009. "Linked Data－The Story So Far." *International Journal on Semantic Web and Information Systems*. [cited 2012.2.17]. 〈http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf〉.

Fripp, D. 2010. "Using linked data to classify web documents." *Aslib Proceddings*, 62(6): 585-595.

Söderbäck, A. Why Libraries should embrace Linked Data. [cited 2012.2.17]. 〈http://code4lib.org/files/LIBRIS_code4lib.pdf〉.

Stuart, D. 2010. "Linked Data and Government Data." *Online*, 34(3): 36.

Linked Data. [cited 2012.4.15]. 〈http://linkeddata.org〉.

Linked Data Open Project. [cited 2012.5.5]. 〈http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData〉.

RDF Primer. [cited 2012.5.10]. 〈http://www.w3.org/TR/rdf-primer/〉.

SKOS Primer. [cited 2012.5.4]. 〈http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/〉.