국립중앙도서관 Linked Data 연구보고서
– 주제명, 저자명 전거데이터 중심 –

2011.12

국립중앙도서관
THE NATIONAL LIBRARY OF KOREA

제출문

국립중앙도서관장 귀하

국립중앙도서관 Linked Data 연구보고서를 제출합니다.

2011년 12월

연 구 진

책임 연구자 조명대 서울대학교 기술융합대학원

오원석 (주)탑쿼드란트코리아 이사

박진호 국립중앙도서관 전문연구관

보조 연구원

연구 보조원

책임연구자 조명대, 오원석

목 차

1. 서론 **1**

1.1 Linked Data 3

1.2 도서관 분야의 Linked Data 5

2. Linked Data 현황 **9**

2.1 Linked Data Vocabulary 11

2.1.1 SKOS 25

2.1.2 FOAF 25

2.1.3 SIOC 25

2.1.4 Dublin Core 25

2.1.5 Bibliographic Ontology 25

2.2. Linked Data 현황 25

2.2.1 Linked Open Data Cloud 25

2.2.2 영역별 Linked Data 25

2.3 도서관 분야 Linked Data 25

2.3.1 미국 의회도서관 25

2.3.2 VIAF(Vritual International Authority File) 25

2.3.3 LIBRIS – 스웨덴 국립도서관 25

2.4 공공분야 Linked Data 25

2.4.1 영국 data.gov.uk 25

2.4.2 미국 data.gov 25

2.4.3 핀란드 25

3. 시범 서비스 적용 **41**

3.1 시범 서비스 대상 데이터 43

3.1.1 저자명 전거데이터 25

3.1.2 주제명 전거데이터 25
3.2 온톨로지 모델 설계 44
3.3 온톨로지 명세 67
3.4 데이터 변환 67
3.5 Linked Data 발행 서비스 67

4. 결론 73

부 록 89

부 록 90

표 목 차

〈표 1〉 더블린코어 메타데이터 요소 ····· 13
〈표 2〉 도메인별 LOD 트리플 현황 ····· 19
〈표 3〉 VIAF가 지원하는 동일 개체의 표현 방법 ····· 22
〈표 4〉 data.gov의 카테고리별 다운로드 통계 ····· 35
〈표 5〉 국립중앙도서관 저자명 전거 데이터 구조 예시 ····· 38
〈표 6〉 국립중앙도서관 주제명의 주요 용어 관계지시기호 ····· 46
〈표 7〉 국립중앙도서관 저자명, 주제명 데이터의 온톨로지 명세 ····· 47
〈표 8〉 저자명 용어에 대한 온톨로지 사용어휘 매핑 ····· 48
〈표 9〉 저자명 데이터에 대한 변환된 트리플 예시 ····· 48
〈표 10〉 주제명 용어에 대한 온톨로지 사용어휘 매핑 ····· 49
〈표 11〉 주제명 데이터에 대한 변환된 트리플 예시 1 ····· 50
〈표 12〉 주제명 데이터에 대한 변환된 트리플 예시 2 ····· 51

그 림 목 차

<그림 1> Linked Data의 개념 ····· 14
<그림 2> Cataloging 개념 ····· 16
<그림 3> MARC 예시 ····· 16
<그림 4> 도서관 분야 표준 ····· 23
<그림 5> 고립된 도서관 분야 네트워크 ····· 24
<그림 6> 미국의회도서관의 SKOS 사용 예시 ····· 68
<그림 7> DBPedia의 SKOS 사용 예시 ····· 70
<그림 8> FOAF-a-Matic ····· 72
<그림 9> SIOC 온톨로지 개념 ····· 72
<그림 10> SIOC을 이용하는 시스템 구조 ····· 85
<그림 11> SIOC의 역할 ····· 85
<그림 12> 2010년 9월 LOD Cloud ····· 86
<그림 13> 도메인별 LOD 트리플 분포 ····· 86
<그림 14> 도메인별 LOD 링크 분포 ····· 86
<그림 15> 미국의회도서관의 WWW Linked Data ····· 86
<그림 16> 미국의회도서관 LCSH의 개념(Concept) 관계 ····· 86
<그림 17> VIAF에서 Obama, Barack ····· 86
<그림 18> VIAF의 http://viaf.org/viaf/29541064 ····· 86
<그림 19> VIAF에서 날짜(Date) 정보를 이용한 개체 판별 예시 ····· 86
<그림 20> LIBRIS 서비스 페이지 예시 ····· 86
<그림 21> LIBRIS의 http://libris.kb.se/data/bib/12130810 소스 일부 ····· 86
<그림 22> LIBRIS의 MARC와 RDF 매핑 1 ····· 86
<그림 23> LIBRIS의 MARC와 RDF 매핑 2 ····· 86
<그림 24> 공공정보 개방에 참여하고 있는 국가들 ····· 86
<그림 25> data.gov.uk를 이용한 주택 가격 동향 서비스 ····· 86
<그림 26> data.gov.uk를 이용한 Find GPs ····· 86
<그림 27> data.gov.uk를 이용한 Research Funding Explorer ····· 86
<그림 28> data.gov.uk를 이용한 Research Funding Explorer의 개념 ····· 86
<그림 29> data.gov.uk의 SPARQL Endpoint ····· 86

〈그림 30〉 data.gov의 하루당 방문자 수 (2011년 06월) ·····86
〈그림 31〉 data.gov의 월별 방문자 수 (2011년) ·····86
〈그림 32〉 data.gov의 SPARQL Endpoint ·····86
〈그림 33〉 FinnONTO의 데이터 구성 ·····86
〈그림 34〉 국립중앙도서관 주제명 전거 데이터의 Term 필드 설명 ·····86
〈그림 35〉 국립중앙도서관 주제명 전거 데이터의 Thesuarus 필드 설명 ·····86
〈그림 36〉 국립중앙도서관 Author 클래스의 Diagram ·····86
〈그림 37〉 국립중앙도서관 Concept 클래스의 Diagram ·····86
〈그림 38〉 국립중앙도서관 Librarian 클래스의 Diagram ·····86
〈그림 39〉 국립중앙도서관의 저자명, 주제명 데이터에 대한 온톨로지 Diagram ·····86
〈그림 40〉 저자명 데이터에 대한 변환된 트리플 Diagram ·····86
〈그림 41〉 주제명 데이터에 대한 변환된 트리플 Diagram 1 ·····86
〈그림 42〉 주제명 데이터에 대한 변환된 트리플 Diagram 2 ·····86
〈그림 43〉 Linked Data 발생 서비스의 메인 화면 ·····86
〈그림 44〉 Linked Data 발생 서비스의 Support 화면 ·····86
〈그림 45〉 Linked Data 발생 서비스의 SPARQL Endpoint 화면 ·····86
〈그림 46〉 SPARQL Endpoint에서 질의한 결과 화면 ·····86
〈그림 47〉 국립중앙도서관 온톨로지 데이터에 대한 Visualization Browser 화면 ·····86

1. 서론

1.1 Linked Data

1.2 도서관 분야의 Linked Data

1. 서론

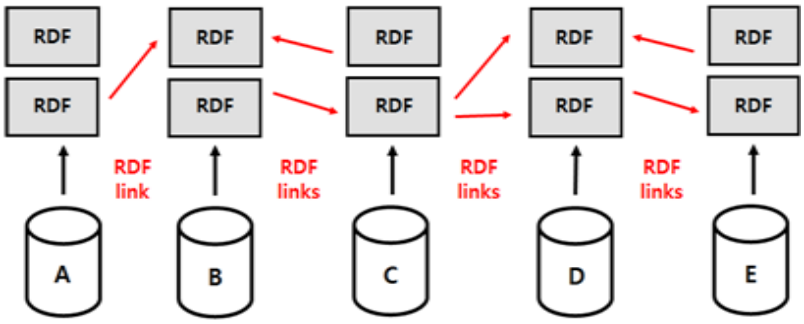
1.1 Linked Data

시맨틱 웹(Semantic Web)은 이용자들이 정보를 좀 더 쉽게 찾고, 공유하고 조합할 수 있게 함으로써 현재의 웹을 한 단계 진화한 형태로 이끌 수 있게 하는 것이 목적이며 이상이다. Linked Data는 시맨틱 웹이 실현되고 있는 모습을 잘 보여주는 개념으로 문서 중심의 현재의 웹과 별도로 데이터가 중심이 되는 웹을 말한다. 현재의 웹은 HTML 문서들이 서로 연결되어 거대한 생태계를 이루고 있다. 마찬가지로 이러한 연결 개념은 문서가 아니라 데이터 간에도 가능하다는 것이 그 핵심이다.

기술적으로 Linked Data는 개방형 데이터(Open Data)들을 네트워크(HTTP 프로토콜)를 통해 개방하고, 유통하여 연계, 협업하고자 하는 것이다. URI기반의 HTTP 역참조(dereference)를 통해 외부 자원에 접근하고, RDF를 통해 기계가독형으로 보유 정보를 웹으로 발행하며, 외부 정보 또는 내부 보유 정보에 질의하기 위한 질의 언어로서 SPARQL을 이용한다. SPARQL을 통해서 다수의 곳에 한 번에 질의하는 것도 가능하다. URI, RDF, SPARQL은 모두 W3C의 시맨틱 웹(Semantic Web) 표준으로 Linked Data는 시맨틱 웹 서비스의 실제적인 모범실무이다.

위키피디아(Wikipedia, http://www.wikipedia.org)에서는 Linked Data를 "URI(Unified Rdsource Identifier)와 RDF(Resource Description Framework)를 이용해 시맨틱 웹 상에 널려있는 데이터, 정보, 지식을 노출하고 공유하며 연결하기 위해 추천되는 최고의 방법을 설명하는 용어"라고 설명하고 있다. 좀 더 구체적으로 설명하면, Linked Data는 사실 정보(fact)를 포함하는 데이터 개체(entity)에 URI를 부여하고 이를 웹 프로토콜인 HTTP를 통해 발행하여 누구나 웹상에서 자유롭게 활용할 수 있게 하는 방법 또는 기술을 말한다. 이는 흔히 Linked Data 플랫폼을 통해 구

현되는데, Linked Data 플랫폼은 기존의 레거시 데이터를 Linked Data로 변환하여 저장, 관리하며, 이를 웹상에 발행하여 공동 활용할 수 있도록 하는 일련의 기술을 지원한다. Linked Data를 통해 문서 중심, 즉 페이지 중심의 기존 웹상의 정보 전달 방식은 페이지가 아닌 웹 페이지를 구성하는 데이터 간의 연결을 통해 보다 풍부한 자원의 공유가 가능한 방식으로 전환될 것이다. 기존의 웹이 하나의 문서를 매개로 연결되어 웹을 구성하였다면, 이제는 각각의 리소스들이 RDF로 표현되고, 연결되어 데이터 중심으로 웹이 연결되는 것이다. 아래는 이러한 데이터 중심의 웹을 위한 Linked Data의 개념을 설명하고 있다.



<그림 1> Linked Data의 개념
(출처 : Christian Bizer: How to Publish Linked Data on the Web)

Linked Data는 시맨틱 웹의 창시자인 팀버너스리(Tim Benners Lee)에 의해 2006년 처음 주장되었으며, 팀버너스리에 의한 Linked Data에 대한 설명은 다음과 같다.

> *시맨틱 웹은 단지 데이터를 웹으로 제공하는 것이 아니라, 데이터 간의 링크를 만듦으로써, 인간이나 기계 모두 데이터의 웹을 탐험할 수 있도록 해준다. Linked Data를 통해 유용한 데이터를 얻게 되면, 그 데이터에 관계된 데이터로 계속되는 항해가 가능하다.*
>
> *- Tim Berners-Lee, Linked Data, 2006,*
> *http://www.w3.org/DesignIssues/LinkedData.html*

2009년 2월 TED 컨퍼런스에서 팀 버너스리는 'Design Issue: Linked Data' 강연을 통해 Linked Data의 네 가지 원칙을 제시하였다.

1) Use URIs as names for things

2) Use HTTP URIs so that people can look up those names.

3) When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)

4) Include links to other URIs. so that they can discover more things.

위의 네 가지 단계가 Linked Data를 사용하는 기본원칙인데, URI를 통해 특정 개념을 표현하고, HTTP 프로토콜을 이용해 해당 개념에 접근(룩업, Look up)할 수 있도록 하고 있다. 또한 RDF를 이용해 접근한 URI가 포함하고 있는 데이터들을 제공 받으며, 이 데이터 안에 포함되어 있는 또 다른 URI로의 접근을 통해 데이터를 연결시킬 수 있게 된다.

Linked Data의 4가지 기본 원칙을 활용해서 셰익스피어(Shakespeare)를 DBpedia에 나타내는 예는 다음과 같다.

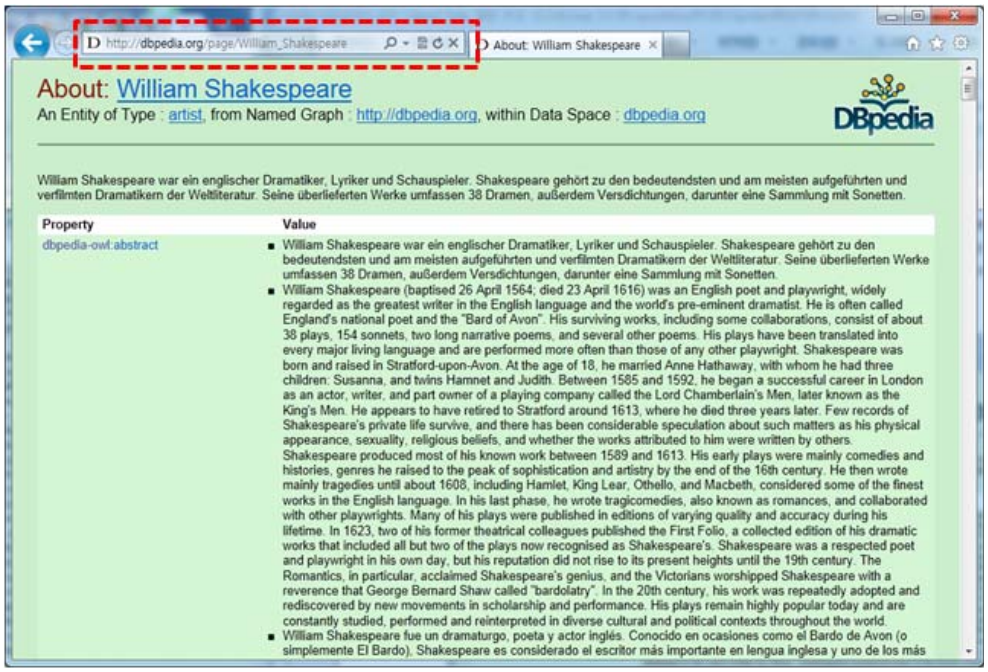
첫 번째 단계는 URI로 개체를 표현하고 식별하는 것이다. DBpedia에서 셰익스피어는 아래와 같은 URI로 표현된다.

· http://dbpedia.org/resource/William_Shakespeare

Content Negotiation을 위해 웹브라우저로 해당 주소에 접근하면 개체를 나타내는 “resource”가 “page”로 변경된다. 이는 웹브라우저를 통해 데이터에 접근하는 경우이면 실제로 개체에 대한 URI는 “resource”로 표현되기 때문이다.

두 번째 단계는 HTTP 프로토콜을 통해 룩업(look-up)이 가능하도록 하는 것이다. 이는 “dbpedia.org”라는 도메인에 HTTP 프로토콜로 접근했을 때 해당 리소스에 접근할 수 있어야 한다는 의미이며, 이를 위해서는 해당 도메인의 물리적 위치에서 웹서버가 동작하고 있어야함을 의미한다. 또한 웹서버는 해당 리소스의 요청에 반응할 수 있어야 하며, 이러한 역

할을 지원하는 것이 Linked Data 플랫폼이다. DBpedia에서 셰익스피어가 해당 URI를 통해 두 번째 단계를 만족하는 모습은 그림과 같다.



<그림 2> Linked Data의 제 2 원칙이 만족된 모습

위의 그림에서 알 수 있듯이, 웹브라우저의 주소 입력란에 해당 URI를 입력한 후 페이지를 요청하면 해당되는 리소소를 화면에 보여준다. 해당 URI가 HTTP를 통해 룩업되고 있는 것이다. 빨간 점선 안에 있는 URI가 해당 부분이다. 해당 URI가 HTTP 프로토콜을 통해 룩업되어야 하는 원칙이 Linked Data의 기본 4 원칙 중에 가장 중요한 원칙이며, 이 원칙을 만족해야만 웹상에서의 정보 개방과 공유, 협업이 가능해진다.

다음으로는 해당 URI에 접근했을 때 풍부한 정보를 제공할 수 있어야 한다는 것이다. 이는 또한 시맨틱웹 표준에 따라 RDF와 SPARQL 로 접근할 수 있어야 하며 셰익스피어의 예제를 통해서는 아래와 같은 그림으로 확인할 수 있다.

<그림 3> Linked Data의 제 3 원칙이 만족된 모습

또한 아래의 그림에서 확인할 수 있듯이 해당 URI를 통해 접근된 데이터는 CSV, JSON 등의 다양한 형식으로 반환이 가능하다.

<그림 4> 다양한 형식으로 제공되는 DBpedia의 Linked Data

해당 데이터에 접근하는 방식은 포맷에 따라 다음과 같다.

- http://dbpedia.org/resource/William_Shakespeare.json
- http://dbpedia.org/resource/William_Shakespeare.ntriples
- http://dbpedia.org/resource/William_Shakespeare.atom

마지막으로 4번째 원칙은 다른 외부 URI로의 접근을 제공하는 것이다. 실제로 이 원칙의 만족이 다양해질수록 시맨틱웹, 특히 Linked Data가 실현하기 위한 웹 자원의 공유 및 협

업을 통한 차세대 웹으로의 접근에 한걸음 더 다가갈 수 있다.

DBpedia의 셰익스피어 예제에서 해당 원칙을 만족하고 있는 모습은 아래와 같다.

<그림 5> Linked Data의 제 4 원칙이 만족된 모습

그림의 빨간 점선 안쪽을 보면 확인할 수 있듯이 뉴욕타임즈와 같은 다른 URI로의 접근을 제공하고 있다.

이와 같은 Linked Data 4가지 원칙을 모두 만족하는 웹상의 자원이 많아질수록 시맨틱 웹 즉, 차세대 웹을 조금 더 앞당길 수 있다. Linked Data를 통해 관련된 데이터를 서로 연결(link)함으로써 얻어지는 장점은 다음과 같다.

- 개방된 데이터를 URI와 RDF, HTTP를 통해 데이터 차원에서 연결하여 사용할 수 있으므로, 내가 만든 데이터가 아니라도 하나의 웹을 하나의 거대한 지식베이스처럼 사용할 수 있다.

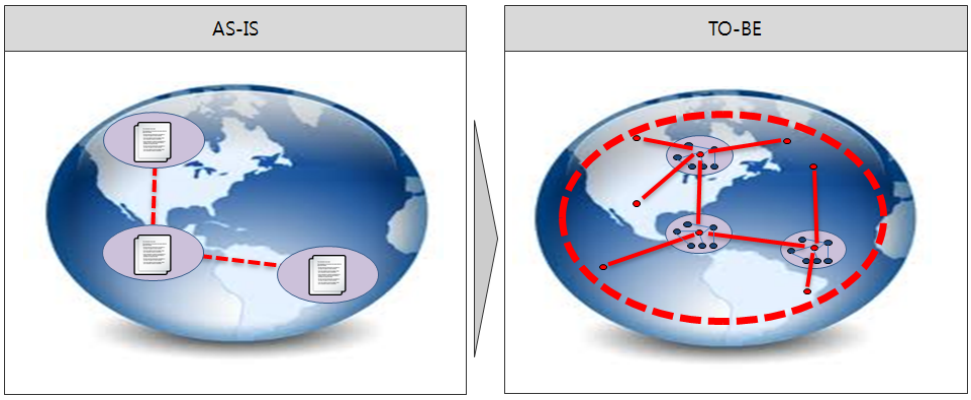
- Linked Data를 통해 공개된 데이터를 이용하면 내가 원하는 데이터가 이미 존재하는지, 어디에 존재하는지 알 수 있으므로 중복된 데이터의 혼란을 감소시킬 수 있다.

- 시맨틱웹 표준인 RDF 형태의 데이터로 발행하므로 마치 하나의 글로벌 데이터베이스처럼 질의하고 이용할 수 있으며, 이를 통해 상호운용성을 높이고, 데이터 통합을 용이하게 할 수 있다.

- URI로 구별되는 데이터 리소스의 자유로운 접근 및 이용이 가능하고, SPARQL Endpoint를 통해 SPARQL 질의가 가능하며, 이를 응용 프로그램에서 이용할 수 있으므로, OpenAPI에 비해 데이터 접근을 더욱 구체화할 수 있어 데이터 지향의 매쉬업을 할 수 있다.

- 초창기의 웹과 같이 데이터의 자유로운 연결과 이용은 새로운 데이터를 생산하고, 양질의 데이터는 트래픽이 증가하게 되는 데이터 네트워크화(Network of Data, Cloud of Data)가 지속될 것이다. 또한, 초기에 LOD Cloud에 진입한 데이터 셋들은 향후에 선점 효과를 누릴 수 있을 것이다.

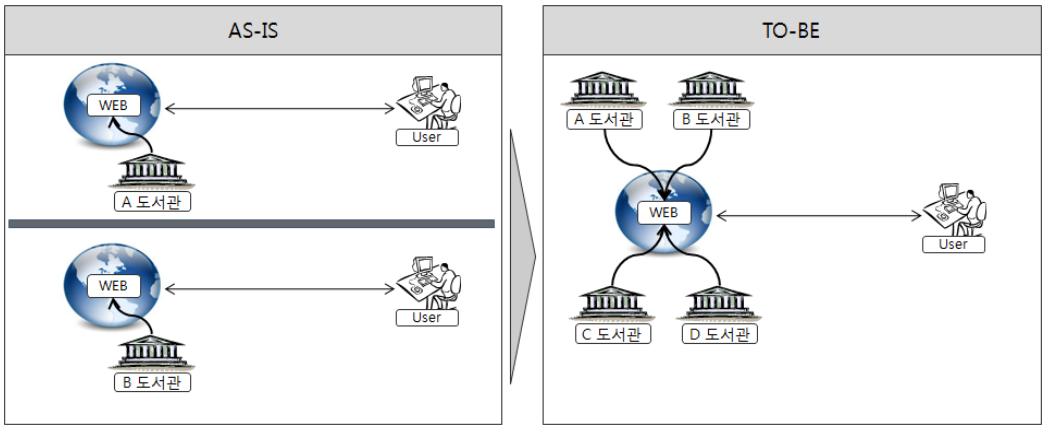
앞에서도 언급했지만, 시맨틱웹 그리고 Linked Data는 웹을 하나의 거대한 지식베이스처럼 활용하고자 하는 목표를 내포하고 있다. 문서위주의 기존 웹에서는 이것이 불가능했지만 데이터 위주의 웹을 통해 데이터를 서로 연결하여 모든 데이터들이 연결되어지면 이를 지식베이스처럼 활용할 수 있게 되는 것이다.



<그림 6> 기존의 웹과 시맨틱 웹의 모습

기존의 웹이 문서 위주의 고립된 데이터를 통한 단순한 연결이었다면, Linked Data는 데이터 위주의 연결된 데이터를 통한 의미적 연결로 웹을 구성하고 활용하고자 하는 노력이며, 기술적 접근이라 할 수 있다.

서비스의 관점에서 살펴보면 기존의 웹을 통한 서비스 방식은 고립된 개별적 서비스만이 가능했다. 하지만 Linked Data를 활용하면 연결된 서비스가 가능해진다. OpenAPI와 매쉬업을 통해 어느 정도 해결할 수도 있지만 OpenAPI 방식은 기능에 의한 매쉬업이지 데이터 자체가 연결되는 모습을 갖을 수는 없다. 단적으로 특정 포털의 OpenAPI 서비스와 또 다른 포털의 OpenAPI 서비스를 연결할 수 없는 한계가 그것이라 할 수 있다. 아래의 그림처럼 Linked Data를 활용하면 각각의 고립된 개별적 서비스를 연결하여 활용할 수 있게 된다.



<그림 7> 고립된 서비스와 연결된 서비스

1.2 도서관 분야의 Linked Data

시맨틱 웹에서는 RDF, RDFS, OWL 등을 통해 데이터를 표현하고 유통하여 기계가 데이터를 처리할 수 있는 환경을 조금 더 제공함으로써 웹상의 정보 생산, 소비 메카니즘을 조금 더 스마트하게 하는 것이다.

이미 DBPedia(http://www.dbpedia.org)를 선두로 하여 공공정보 분야, 도서관 분야, 문

화정보 분야, 언론 분야, 생명과학 분야 등 여러 분야에서 정보의 전달과 융합을 위해 시맨틱 웹을 실현하고 있으며, 점점 더 많은 곳에서 시맨틱 웹이 이용될 것이라고 예상하고 있다. 특히, 많은 분야에서 정보의 공동 활용 및 융합을 위해 시맨틱웹 기술을 활용하고 있으며 특히, 인터넷과 웹을 기반으로 문화정보의 유통과 Cross-Domain간 문화정보의 융합의 방법으로 시맨틱 웹을 적용하는 사례가 늘고 있다.

도서관 분야는 데이터 유통을 위해 많은 표준과 유통 표준을 통해 발전해 왔다. 도서관 분야 발전의 첫 단계는 목록(Cataloging)이다. 목록은 정보를 메타데이터 형식으로 가공하고 식별자를 부여하여 이용 가능하도록 한 것이다.

도서관 분야에서 정보 유통에 또 다른 큰 역할을 했던 것은 MARC(MAchine Readable Cataloging)이다. MARC는 기계가독형 데이터 교환 포맷으로 도서관에서 이용자 서비스를 위해 가장 기초가 되는 작업이며, 도서관 간 상호협력의 시작점이라고 할 수 있다. 미국 의회도서관이 1965년 12월에 도서관 자원위원회(Council onLibrary Resources: CLR)로부터 지원을 받아 기계가독형 목록데이터의 배포에 대한 가능성과 효용성 실험을 위한 프로젝트를 수행하였으며, 이 프로젝트의 이름인 MAchine Readable Cataloging에서 MARC라는 이름이 유래하였다. Furrie(연도입력)는 MARC를 “정보교환 활성화를 위한 업계표준(an industry-wide standard whose primary purpose is to foster communication of information)”이라고 정의하였으며, Cooper(연도입력)는 “기관 간 또는 시스템 간 서지데이터 교환을 위한 표준 포맷(a standard format for interchanging bibliographic data between organizations or systems)”이라고 정의하였다.

리더	00568nam 2200205 k 4500

TAG	IND	내용
001		KMO199509611
005		20050812114024
008		950616s1994 ulk 000 f kor
020		a8901010607g33810:c5500
040		a011001c011001
049	0	IEM1238778v1994 IEM3352837v1994c2
052	01	a813.6b공784ㄱㄷc1994
056		a813.623
082	0	a895.734219
100	1	a공지영
245	10	a(공지영 장편소설)고등어/d공지영 지음
260		a서울:b웅진출판,c1994
300		a287p.;c22cm
653		a고등어
950	0	b5500

<그림 8> MARC 예시

MARC은 초기에는 미국 국회도서관의 정부관련 고문서의 목록을 축적하기 위한 방법이었기 때문에 미국 국회도서관의 요구에 편중되어 개발되었고, 이 때문에 국제적 표준형식으로 적용시키는데 한계가 있었다. 이를 극복하기 위해 LC MARC은 USMARC으로 개정되었고 이후 새로운 다양한 형태의 정보자료나 통제방식을 지원하도록 확장,개발되어 왔다.

USMARC을 기반으로 캐나다는 CANMARC, 호주는 AusMARC, 영국은 UKMARC, 독일은 MAB, 중국은 ChineseMARC, 일본은 JapanMARC, 우리나라에서 KORMARC등 자국의 자료처리에 가장 적합한 MARC을 개발하였다. 또는 실제 도서관마다 USMARC을 사용할 경우 일부내용을 확장하기도 하는데, 이렇게 확장된 포맷(extended format)을 슈퍼셋(supersets)라 한다. OCLC MARC, RLIN MARC, Utlas MARC 같은 것이 USMARC을 확장한 대표적인 예이다. 확장된 포맷은 순수 포맷(pure format)으로 처리할 수 없는 기능을 제공하여 자관의 처리요구를 만족시켜줄 수 있다.

MARC 레코드 구조는 1971년에는 ANSI Z39.20, 1973년에는 ISO 2709로 채택되어, 서지 레코드를 위한 국가 및 국제 표준으로 제안되었다. 그러나 각국이 다른 MARC 형식을 제정함에 따라 국가 간 목록 데이터베이스의 교환이 어려운 문제가 되었고, 이를 해결하기 위해 국제도서관협회(IFLA)에서는 1977년 UNIMARC 첫 판을 발간하기도 하였다. 그리고 1999년 USMARC과 CANMARC이 통합되면서 MARC21이라는 새로운 이름으로 바뀌게 되었다.

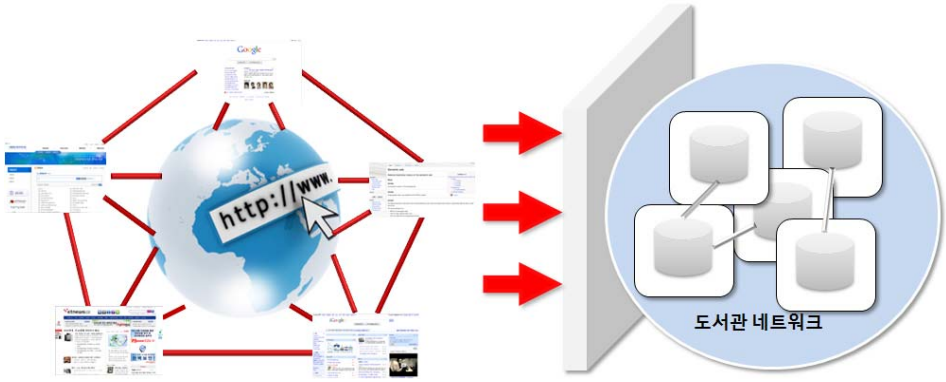
하지만 MARC는 도서관 분야에서만 유통되는 표준이기 때문에 도서관 관련 시스템 사이의 정보 유통에만 활용될 뿐, 웹을 중심으로 하는 정보 유통 환경에서 도서관 이외의 분야와 정보를 융합하거나 유통하는 과정에서는 활용이 어렵다. MARC 이외에도 도서관 분야는 도서관 유통과 도서관 정보 표현을 위한 다양한 표준들이 존재한다. 도서관 분야는 특히 관련 표준이 많으며, 이는 타 분야와의 정보 융합에 복잡한 현상으로 받아들여지곤 한다.

<그림 9> 도서관 분야 표준들

MARC와 같이 도서관 내에서의 유통 구조는 도서관 분야의 노력으로 어느 정도 목표 달성을 이루었다. 그러나 웹을 중심으로 유통 구조를 형성하고 있는 다른 분야에서 도서관 분

야와의 협력을 하지 못하기 때문에 타 분야와의 유통 개선에는 한계를 보이고 있다. 최근에는 Linked Data를 이용해 도서관 분야의 정보 유통 및 표현을 웹 세상으로 끌어내고자 하는 노력이 가속화 되고 있다.

아래의 그림에서처럼 도서관 분야 내에서의 정보 유통은 어느 정도 성과를 거두고 있으나, 외부 네트워크 즉, 타 분야와의 정보 융합이 불편하게 여겨지고 있다. 도서관 분야는 모든 타 분야의 정보들이 문헌이나, 보고서, 데이터 등으로 집결되어 있는 또 다른 전 분야를 포함하고 있는 특성을 갖고 있으며, 타 분야와의 융합이 절실히 요구되는 분야이기도 하다.



<그림 10> 고립된 도서관 분야 네트워크

2. Linked Data 현황

2.1 Linked Data Vocabulary

2.2 Linked Data 현황

2.3 도서관 분야 Linked Data

2.4 공공분야 Linked Data

2. Linked Data 현황

2.1 Linked Data Vocabulary

2.1.1 SKOS

SKOS(Simple Knowledge Organization System)는 시소러스, 택소노미, 분류체계와 주제명 표목표 같은 지식어휘체계를 표현하기 위한 RDF 용어집(vocabulary)이다. SKOS가 RDF(Resource Description Framework)에 기반하고 있기 때문에 SKOS로 표현된 지식어휘체계는 기계가 이해 가능하고(machine-readable), 소프트웨어 어플리케이션 간에 상호호환이 가능(interoperability)하며 웹으로 발행(publishing)이 가능하다.

시맨틱 웹 어플리케이션에서 지식어휘체계로 정의된 용어를 사용하기 위해서는 어플리케이션에서의 참조와 재사용을 증진시키는 방식으로 어휘의 용어들을 정의하고, 문서화하고 발행해야 한다. 다양한 정보원의 데이터를 병합하고 이를 재사용가능하도록 설계된 시맨틱웹의 특성은 SKOS를 기반으로 지식어휘체계를 관리함으로써 다른 시맨틱웹 어휘들과 융복합될 수 있으며 계속적인 확장이 가능하다.

콘텐츠 색인에 있어서 시맨틱 웹 온톨로지 또는 어휘를 사용하는 가장 큰 장점으로는 URI를 통해 통일된 방법으로 명확한 개념을 참조할 수 있다는 점이다. 특히, 이용자가 Linked Data로 발행된 지식어휘체계를 URI로 접근할 때, 개념에 대한 부가 정보를 제공할 수 있고 개념간의 관계를 표현하기 때문에 관계를 따라 브라우징과 탐색이 가능하다.

또한, SKOS 데이터 모델은 사용하는 개별 기관이나 이용자의 목적에 따라 확장이 가능하므로 원래의 SKOS 데이터 모델과 일관성을 유지하면서도 개별 도메인에 특화된 특성을 구현할 수 있는 장점이 있다. SKOS 같은 공통 표현모델을 사용하면 여러 지식어휘체계를 하

나의 어플리케이션에서 연계하여 사용가능하고 시맨틱웹 환경에서 하나의 지식어휘체계를 여러 어플리케이션에서 공유하는 등 연계, 공유가 쉽고 따라서 재사용이 가능하다.

SKOS는 기존에 존재하는 개념적 지식어휘체계, 즉, 시소러스, 분류체계, 주제어 표목표, 택사노미의 형태로 생산된 각종 어휘들을 대체하기 위한 것이 아니라, RDF 기술을 토대로 한 경량의, 직관적인 개념적 모델링 언어로 새로운 KOS를 개발하고 공유하기 위한 언어이다. 이를 통해 저비용의 마이그레이션 방법을 지원함으로써 기존의 지식어휘체계가 시맨틱웹에 포팅(porting)되고, 재사용성과 상호운용성을 증진할 수 있도록 하는 것이 SKOS의 목적이다. 또한 SKOS는 소셜 태깅 어플리케이션과 같은 웹 기반의 협업 도구에서 사용되는 비구조적이고 비형식적인 태깅 시스템과 OWL과 같이 논리적 형식화가 잘된 온톨로지 언어 사이의 연결고리로서의 역할을 담당한다.

SKOS의 개념 모델(concept schema)은 ISO2788 시소러스 표준의 영향을 받아 유사한 구조를 가지고 있다. SKOS 개념 모델은 RDF의 구성자(construct)만으로 정의되어 있는데 그 이유는 RDFS/OWL이 의미적 표현이 더 풍부하나, SKOS만으로도 시소러스를 비롯한 지식어휘체계(KOS)를 표현하는데 충분하기 때문이다.

SKOS는 레이어로 구성된다. SKOS Core는 가장 개발이 진행된 부분이고 시소러스 표준에 직접적으로 대응되는 부분이다. SKOS Mapping은 시소러스 개념들을 하나의 정보원에서 다른 정보원으로 매핑시키기 위해 여러 개의 특정한 속성(property)를 정의하는 SKOS의 확장이다.

가) 개념 정의(Concepts)

SKOS 개념 모델의 가장 기본 요소인 skos:Concept 클래스는 지식어휘체계에 존재하는 객체(object), 의미(meaning), 아이디어 또는 이벤트를 표현하는데 사용된다. skos:Concept을 이용하여 주어진 리소스를 개념으로 명시하기 위해서는

1. 개념에 대한 유일 식별자(URI)를 생성(또는 재사용)하고,

2. rdf:type 프로퍼티를 이용하여 RDF 형식으로 표현한다.

예를 들어, RDF로 표현된 다음 구문은

〈http://www.example.com/animals〉 rdf:type skos:Concept.

"animals"이라는 리소스가 URI로 “http://www.example.com/animals”를 가지며, skos:Concept 클래스의 인스턴스(즉, animals의 타입이 skos:Concept)임을 나타낸다.

위의 구문을 네임스페이스 prefix ex로 축약해서 표현하면 다음과 같다.

ex:animals rdf:type skos:Concept.

나) 레이블 정의(Labels)

skos:Concept로 선언된 개념을 자연어, 즉 인간이 읽을 수 있는 리소스의 이름으로 표현하기 위해 사용하는 것이 rdfs:label이다. SKOS는 rdfs:label을 확장하여 다음과 같은 세가지의 레이블 속성을 제공한다.

1. skos:prefLabel

skos:prefLabel은 주로 두 가지 용도로 사용된다. 첫 번째는 색인 시스템에서 사용되는 색인어(descriptor)를 지정하기 위하여 사용되고, 두 번째는 다국어 레이블을 지정하기 위한 용도로 사용된다.

```
ex:animals rdf:type skos:Concept;
skos:prefLabel "animals".
skos:prefLabel "animals"@en;
skos:prefLabel "animaux"@fr.
```

2. skos:altLabel

skos:altLabel은 skos:prefLabel 외에 보통 약어 표현이나 동의어 표현과 같은 부가적인 표기가 존재할 때 사용된다.

```
ex:fao rdf:type skos:Concept;
skos:prefLabel "Food and Agriculture Organization"@en;
skos:altLabel "FAO"@en.
```

3. skos:hiddenLabel

skos:hiddenLabel은 텍스트 기반의 색인, 검색을 위해 기계가 접근할 수 있는 텍스트 데이터를 표현하면서 데이터가 이용자에게는 보이지 않도록 하고 싶을 때 사용한다. 예를 들어, 철자가 틀린 텍스트 데이터를 색인어에 포함시키고자 할 때 사용한다.

```
ex:animals rdf:type skos:Concept;
skos:prefLabel "animaux"@fr;
skos:altLabel "bêtes"@fr;
skos:hiddenLabel "betes"@fr.
```

다) 개념 간의 의미관계 표현(Semantic Relationships)

지식어휘체계에서 개념 정의를 하는데 의미관계는 중요한 역할을 한다. 개념의 의미는 레이블로 표현되는 자연어로서가 아니라 어휘체계 안에서 다른 개념들과의 관계를 통해 나타내지기 때문이다.

1. 협의어/광의어(Broader/Narrower 관계)

skos:broader와 skos:narrower는 개념들 간의 계층관계, 즉, 어떤 분류체계 내에서의 상/하위 관계, 부분/전체 관계를 표현하는데 사용된다.

```
ex:animals rdf:type skos:Concept;
skos:prefLabel "animals"@en;
skos:narrower ex:mammals.
ex:mammals rdf:type skos:Concept;
skos:prefLabel "mammals"@en;
skos:broader ex:animals.
```

2. 관련어(skos:related)

두 개념 간의 수평적인 연관 관계를 표현하는 데는 skos:related가 사용된다. 이것은 대칭 프로퍼티(symmetric property)로서, 즉 A skos:related B 이면 B skos:related A가 추론된다.

```
ex:birds rdf:type skos:Concept;
skos:prefLabel "birds"@en;
skos:related ex:ornithology.
ex:ornithology rdf:type skos:Concept;
skos:prefLabel "ornithology"@en.
```

라) 지식어휘체계 명시(Concept Schemes)

skos:ConceptScheme은 이 개념 어휘들이 속해 있는 지식어휘체계를 기술하도록 한다. 시소러스나 분류체계 명을 기술하거나 작성자 등을 표현하는데는 더블린 코어(Dublin Core)의 속성을 이용할 수 있다.

```
ex:animalThesaurus rdf:type skos:ConceptScheme;
```

```
dct:title "Simple animal thesaurus";
dct:creator ex:antoineIsaac.
```

마) 지식어휘체계 간의 매핑(Mapping Concept Schemes)

SKOS가 제공하는 여러 프로퍼티를 사용하면 여러 지식어휘체계를 서로 연계한 네트워크를 구성할 수 있다. 서로 다른 지식어휘체계의 개념들이 서로 연결됨으로써 분산적이고, 이형적인 글로벌한 지식어휘체계를 형성하게 된다. 지식어휘체계의 웹은 지식어휘체계 간의 의미적 네비게이션을 가능하게 함으로써 새로운 어플리케이션을 만드는 기반이 된다.

SKOS는 다양한 지식어휘체계 간의 개념을 서로 매핑하는 여러 속성을 제공하는데, skos:exactMatch는 어떤 개념에 정확히 대응되는 관계를 정의할 때, skos:closeMatch는 근접한 유사어를 정의할 때 사용된다. 이 외에도 skos:broadMatch, skos:narrowMatch, 그리고 skos:relatedMatch이 사용 가능하다.

```
ex1:referenceAnimalScheme rdf:type skos:ConceptScheme;
dct:title "Extensive list of animals"@en.
ex2:eggSellerScheme rdf:type skos:ConceptScheme;
dct:title "Obsessed egg-seller's vocabulary"@en.

ex1:platypus skos:broadMatch ex2:eggLayingAnimals.
ex1:platypus skos:relatedMatch ex2:eggs.
ex1:animal skos:exactMatch ex2:animals.
```

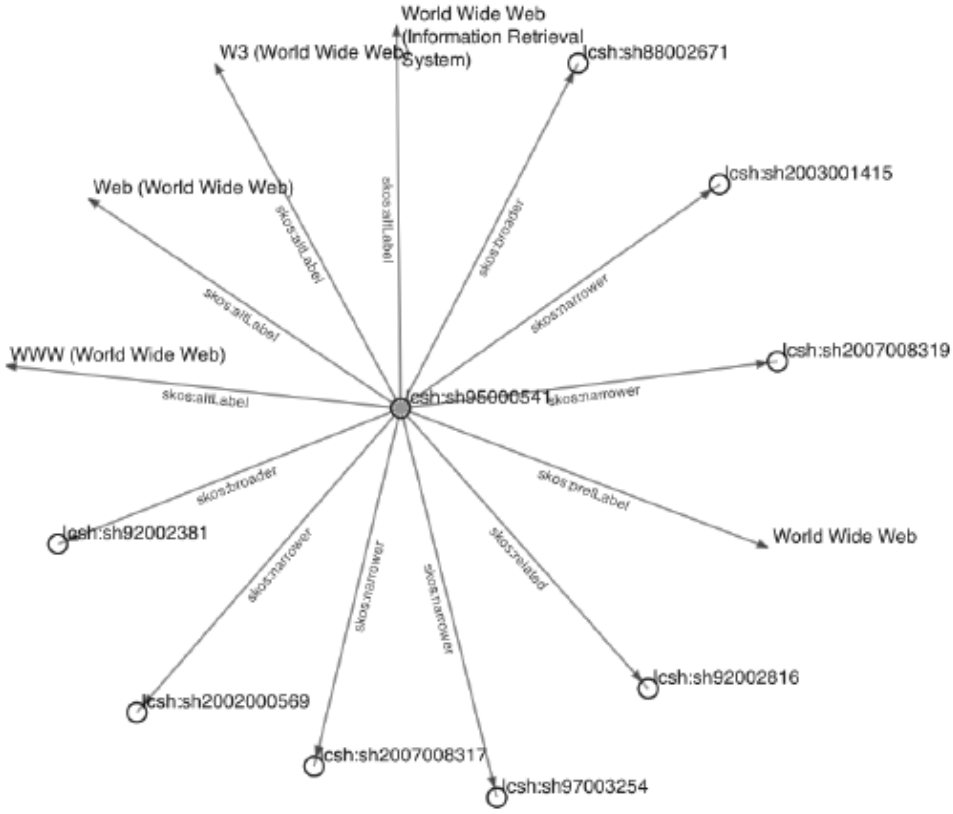
개념 간의 매핑을 통해 서로 다른 두 지식어휘체계를 연계하는 것뿐만 아니라 시맨틱 웹 상에서는 리소스의 URI를 이용하여 분산 방식의 공유와 재사용이 가능하다. 따라서 어떤 하나의 개념이 동시에 여러 지식어휘체계에 소속되는 것이 가능하고 SKOS 관리자는 새로운 개념을 선언할 때 어떤 스키마에 존재하는 개념인지 skos:inScheme 프로퍼티를 이용하여 선언함으로써 이미 존재하는 개념을 차용하여 재사용할 수 있다.

```
ex1:referenceAnimalScheme rdf:type skos:ConceptScheme;
dct:title "Reference list of animals"@en.
ex1:cats rdf:type skos:Concept;
skos:prefLabel "cats"@en;
skos:inScheme ex1:referenceAnimalScheme.
```

SKOS를 이용하는 장점은 개별 시스템 분류체계의 독립성을 보장하면서도 다양한 분류체계 간의 관계를 정의하는 것이 가능하므로 다양한 분류체계 간의 통합이 용이하다는 것과, 새로운 분류체계를 통합할 때 기존의 시스템에 영향을 주지 않으면서도 계속적인 확장이 가능하다는 것을 들 수 있다. SKOS 데이터 모델을 사용하는 개별 기관이나 이용자의 목적에 따라 확장이 가능하기 때문에 원래의 SKOS 데이터 모델과 일관성을 유지하면서도 개별 도메인에 특화된 특성을 구현할 수 있는 장점도 있다. SKOS 같은 공통 표현모델을 사용하면 여러 지식어휘체계(분류체계, 시소러스 등)를 하나의 어플리케이션에서 연계하여 사용 가능하고 웹 환경에서 하나의 지식어휘체계를 여러 어플리케이션에서 공유하는 등 연계, 공유가 쉽고 재사용에 효과적이다. 또한 SKOS로 표현된 분류체계간의 장벽이 없는 용어 간의 관계를 탐색하여 이용자의 검색요구에 적합한 검색어 선택 가이드를 통해 개념 기반 네트워크 탐색을 통한 정보 검색에도 활용될 수 있다. SKOS의 integrity constraints(일관성 제약조건)은 데이터의 적합성을 보장함으로써 데이터의 품질 향상에 기여할 수 있으며, 분류체계에 규칙(rules, ex: SWRL: Semantic Web Rule Language)을 추가함으로써 자동 분류체계 생성도 가능하다.

o SKOS 구현 사례

- LCSH 예시



<그림 12> LCSH의 SKOS Concept Graph

– DBPeida 예시 (http://dbpedia.org/page/WorldWideWeb)

rdfs:label	▪ WorldWideWeb
owl:sameAs	▪ fbase:WorldWideWeb
skos:subject	▪ category:Public_domain_software ▪ category:Free_web_browsers ▪ category:Free_HTML_editors ▪ category:NeXT ▪ category:1990_software ▪ category:Free_software_programmed_in_Objective-C ▪ category:History_of_web_browsers ▪ category:Discontinued_web_browsers

<그림 13> DBPedia의 SKOS 사용 예시

2.1.2 FOAF

FOAF(Friend of a Friend)는 사람과 사람간의 관계 표현을 위한 RDF 어휘집으로써, XML과 RDF에 기반하여 어휘 명세를 정의하고, 이를 기반으로 친구 관계를 명세할 수 있도록 표현한 온톨로지이다. FOAF는 사회적 특성에 따라 몇 단계만 거치면 알 수 있는 사람이 기하급수적으로 늘어나는 현상을 대변하고 있다. 내가 알고 있는 사람이 백 명이고, 그 백 명이 또 다른 백 명씩을 알고 있다면 내 관계에는 1단계만 거쳤을 뿐이지만 만 명이 된다는 특징을 잘 이용하고 있는 것이다. 트위터나 페이스북 등 SNS도 이러한 특징에 기반하고 있는 것이다.

FOAF의 주요 용어(terms)는 크게 3가지 범주로 나눌 수 있다.

첫째는 핵심용어(Core)로 FOAF에서 가장 기본이 되는 클래스와 속성들이다. 이 용어들은 시간이나 기술에 독립적으로 사회적 그룹과 사람들의 특징을 기술하는 것은 물론 역사적 문화적 유산, 디지털도서관의 컨텍스트 등을 설명하기 위한 기본 정보로 활용된다. FOAF에서는 agent의 다른 유형으로 Project, Organization, Group으로 클래스를 정의하고 있다. FOAF 핵심용어와 관련된 작업들은 Dublin Core(http://www.dublincore.org), SKOS(http://www.w3.org/2004/02/skos/), DOAP(http://trac.usefulinc.com/doap), SIOC(http://sioc-project.org/), Org vocabulary(http://www.epimorphics.com/public/vocabulary/org.html), Bio vocabulary(http://vocab.org/bio/0.1/.html)가 있다.

둘째는 소셜 웹(Social Web)과 관련된 추가용어들로 인터넷 계정, 주소록과 같이 웹에서의 활동과 관련된 거서들을 기술하고자 할 때 사용하기 위한 용어들이다. 관련된 작업들은 Portable Contacts(http://portablecontacts.net/), W3C Social Web group(http://www.w3.org/2005/Incubator/socialweb/)이 있다.

마지막은 링크드 데이터(Linked Data) 유틸리티들이다. FOAF는 'RDFWeb' 프로젝트로 시작되었고, 연결된 RDF 문서들의 네트워크를 통해서 간단한 사실 데이터들을 게시하기 위

한 모델로 채택되었다. FOAF는 계속 성장하고 있는"Linked Data"커뮤니티에 중요하게 남아 있다. FOAF는 사람과 관련된 비디오, 책, 스프레드 시트, 3D 모델 등과 같은 documents 정보뿐만 아니라 사람의 머릿속에 존재하는 정보들을 통합하는 웹을 사용하기 위한 시도이다.

FOAF Core	Social Web
· Agent	· nick
· Person	· mbox
· name	· homepage
· title	· weblog
· img	· openid
· depiction (depicts)	· jabberID
· familyName	· mbox_sha1sum
· givenName	· interest
· knows	· topic_interest
· based_near	· topic (page)
· age	· workplaceHomepage
· made (maker)	· workInfoHomepage
· primaryTopic (primaryTopicOf)	· schoolHomepage
· Project	· publications
· Organization	· currentProject
· Group	· pastProject
· member	· account
· Document	· OnlineAccount
· Image	· accountName
	· accountServiceHomepage
	· PersonalProfileDocument
	· tipjar
	· sha1
	· thumbnail
	· logo

<표 > FOAF 용어의 분류

현재의 FOAF 용어 분류(FOAF 명세서 0.98 버전 기준)는 소셜 웹이 등장과 링크드 데이

터 등 새로운 기술의 활용성이 높아지는 것을 반영한 결과라고 할 수 있다. 반면 이전 버전(http://xmlns.com/foaf/spec/20100101.html)에서는 5가지 범주로 용어를 분류하고 있는데 현재의 분류법보다 FOAF를 활용하고 이해하는데 보다 쉬운 체계를 제공하고 있다.

구분	클래스와 속성
FOAF Basics	Agent, Person, name, nick, title, homepage, mbox, mbox_sha1sum, img, depiction (depicts), surname, familyName, givenName, firstName, lastName
Personal Info	weblog, knows, interest, currentProject, pastProject, plan, based_near, age, workplaceHomepage, workInfoHomepage, schoolHomepage, topic_interest, publications, geekcode, myersBriggs, dnaChecksum
Online Accounts / IM	OnlineAccount, OnlineChatAccount, OnlineEcommerceAccount, OnlineGamingAccount, account, accountServiceHomepage, accountName, icqChatID, msnChatID, aimChatID, jabberID, yahooChatID, skypeID
Projects and Groups	Project, Organization, Group, member, membershipClass
Documents and Images	Document, Image, PersonalProfileDocument, topic (page), primaryTopic (primaryTopicOf), tipjar, sha1, made (maker), thumbnail, logo

<표> FOAF 0.97에서의 용어 구분

FOAF 어휘집은 현재 13개의 클래스(Class)와 62개의 속성들(Properties)로 구성되어 있다. 현재 버전을 기준으로 모든 클래스와 속성들이 안정화 상태는 아니며, 명세서 상에는 안정화(stable), 비안정화(unstable), 점검중(testing) 세 가지 상태로 구분하고 있다.

FOAF의 각 클래스별 상세 명세는 다음과 같다.

Class: foaf:Agent	
행위자(agent)로 사람, 그룹, 소프트웨어 혹은 물리적 물체(artifact). 가장 잘 알려준 하위 속성이 사람(Person)이고, 조직(Organization)과 그룹(Group)등도 여기에 속한다. IM 채팅 아이디 등의 속성은 사람에게도 속하지만 소프트웨어 봇(bots)에 속하기도 한다.	
상태	안정
포함하는 속성	weblog, icqChatID, msnChatID, account, age, mbox, yahooChatID, tipjar, jabberID, status, openid, gender, interest, holdsAccount, topic_interest, aimChatID, birthday, made, skypeID, mbox_sha1sum
함께 활용하는 속성	member, maker
하위클래스	Organization, Group, Person

Class: foaf:Group	
Group 클래스는 개별 agent들의 모음이며 그 자체로 agent 역하라 수행이 가능함	
상태	안정
포함하는 속성	member
상위클래스	Agent

Class: foaf:Organization	
Organization 클래스는 회사, 학회 등과 같은 사회적 조직에 대응하는 Agent를 나타냄	
상태	안정
포함하는 속성	member
상위클래스	Agent
서로소 클래스 (disjoint)	Person, Document

Class: foaf:Person	
Person 클래스는 사람을 나타내는데, 생존여부나 실제여부에 대해서 문제시 하지 않음.	
상태	안정
포함하는 속성	myersBriggs, familyName, publications, lastName, family_name, firstName, currentProject, surname, knows, workInfoHomepage, pastProject, geekcode, schoolHomepage, workplaceHomepage, img, plan
함께 활용하는 속성	knows
상위클래스	Agent, Spatial Thing, Person
서로소 클래스 (disjoint)	Organization, Project

Class: foaf:Document	
Document 클래스는 광범위하게 문서라고 간주할 수 있는 것들을 나타냄. 모든 이미지는 문서이기 때문에 Image클래스는 Document의 하위 클래스임. 현재까지는 물리적인 문서와 전자문서 사이의 구분을 명확하게 하고 있지 않다. 또한 저작물의 사본과 추상화된 사본들을 명확하게 구분하지 않음.	
상태	테스트 중
포함하는 속성	topic, sha1, primaryTopic
함께 활용하는 속성	weblog, openid, tipjar, accountServiceHomepage, isPrimaryTopicOf, workplaceHomepage, homepage, interest, workInfoHomepage, page, publications, schoolHomepage
하위클래스	PersonalProfileDocument Image
서로소 클래스 (disjoint)	Organization Project

Class: foaf:Image	
Image 클래스는 Document의 하위클래스로 이미지로 칭할 수 있는 문서들에 대응하는데, JPEG, PNG 등 디지털 이미지들이 대표적임.	
상태	테스트 중
포함하는 속성	thumbnail, depicts
함께 활용하는 속성	thumbnail, depiction, img
상위클래스	Document

Class: foaf:OnlineAccount	
OnlineAccount 클래스는 accountServiceHomepage를 통해서 간접적으로 표시되는 서비스 양식, 조항 등을 나타냄. 소셜네트워크 사이트의 사용자 프로필 페이지와 같이 잘 알려진 URIs를 갖고 있는 문서 혹은 개체를 가리키게 하는 것이 좋은 예임.	
상태	테스트 중
포함하는 속성	accountServiceHomepage, accountName
함께 활용하는 속성	account, holdsAccount
상위클래스	Thing
하위클래스	Online E-commerce Account, Online Gaming Account, Online Chat Account

Class: foaf:PersonalProfileDocument	
PersonalProfileDocument 클래스는 document클래스와 같은 것들과 문서를 만든 사람들의 속성을 기술하기 위해 RDF를 활용함. 문서에서 설명하는 한 사람이 존재하는데 만든 사람이 primaryTopic임.	
상태	테스트 중
상위클래스	Document

Class: foaf:Project	
Project 클래스는 공식, 비공식, 하나의 것 혹은 다수의(모음)의 것일 수 있으며 프로젝트 홈페이지 등을 지칭함. currentProject, pastProject 의 연결을 지정하는 것과 같은 추가작업이 더 필요함.	
상태	테스트 중
서로소 클래스 (disjoint)	Person, Document

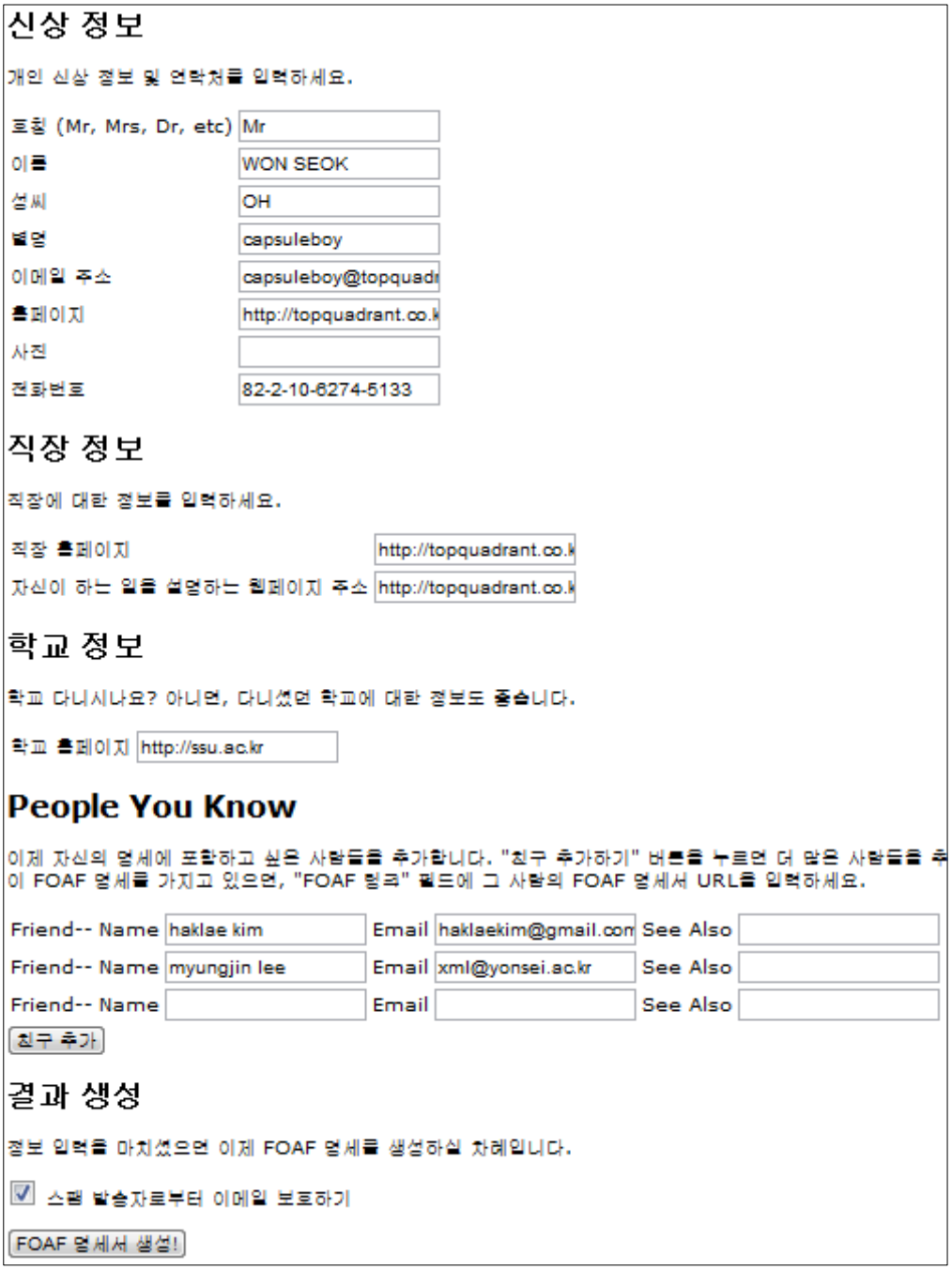
Class: foaf:LabelProperty	
LabelProperty 클래스는 레이블형태로 제공되는 텍스트형 값을 갖는 RDF 속성임. LabelProperty는 효과적인 rdfs:label임.	
상태	불안정(unstable)

Class: foaf:OnlineChatAccount	
OnlineChatAccount 클래스는 채팅, 인스턴트 메시징에 활용되는데, jabberID, aimChatID, skypeID, msnChatID, icqChatID, yahooChatID 등의 속성을 활용할 수 있음	
상태	불안정(unstable)
상위 클래스	Online Account

Class: foaf:OnlineEcommerceAccount	
OnlineAccount의 하위 클래스로 상품, 서비스의 구매, 판매를 표현하는데 활용함(예, Amazon, eBay, PayPal, thinkgeek 등)	
상태	불안정(unstable)
상위 클래스	Online Account

Class: foaf:OnlineGamingAccount	
온라인게임에 활용하는 온라인 계정을 말한다. Xbox live 등은 물론이고 고전적인 텍스트 기반의 시스템 계정도 포함할 수 있다.	
상태	불안정(unstable)
상위 클래스	Online Account

나만의 FOAF를 간단하게 만들 수도 있다. FOAF-a-matic이라는 간단한 자바스크립트 기반의 어플리케이션을 이용하면 간단한 FOAF를 생성해 준다.



<그림 14> FOAF-a-Matic (http://www.ldodds.com/foaf/foaf-a-matic)

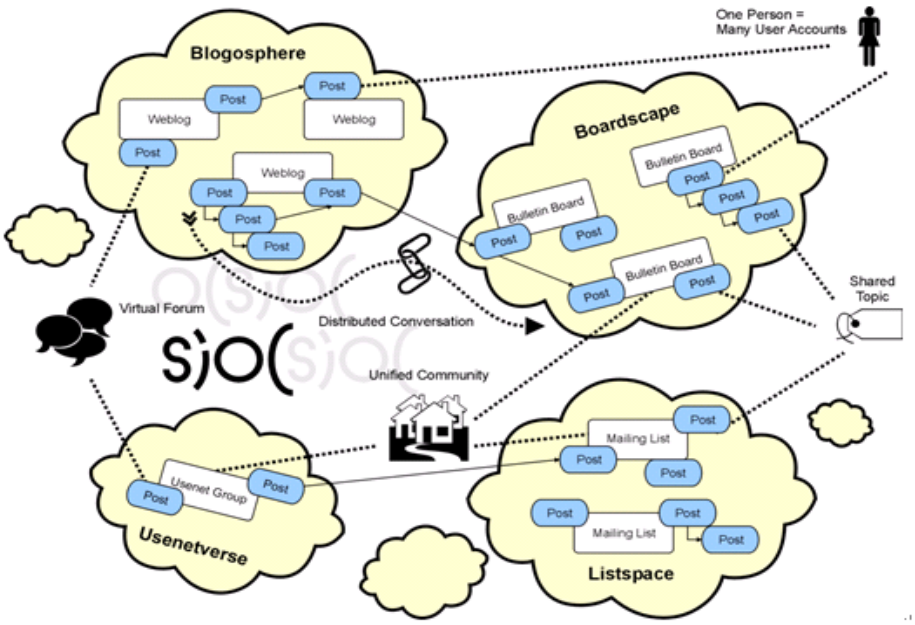
위의 그림에서처럼 간단한 명세를 입력하고 FOAF 명세를 생성하면 아래와 같은 RDF 구문이 생성된다.

```
<rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
        xmlns:foaf="http://xmlns.com/foaf/0.1/"
        xmlns:admin="http://webns.net/mvcb/">
<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="#me"/>
  <foaf:primaryTopic rdf:resource="#me"/>
  <admin:generatorAgent
   rdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
  <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>
<foaf:Person rdf:ID="me">
<foaf:name>WON SEOK OH</foaf:name>
<foaf:title>Mr</foaf:title>
<foaf:givenname>WON SEOK</foaf:givenname>
<foaf:family_name>OH</foaf:family_name>
<foaf:nick>capsuleboy</foaf:nick>
<foaf:mbox_sha1sum>cfdbcd83e0b7a02630bf52aabe24dfd93f9b3736</foaf:mbox_s
ha1sum>
<foaf:homepage rdf:resource="http://topquadrant.co.kr"/>
<foaf:phone rdf:resource="tel:82-2-10-6274-5133"/>
<foaf:workplaceHomepage rdf:resource="http://topquadrant.co.kr"/>
<foaf:workInfoHomepage rdf:resource="http://topquadrant.co.kr"/>
<foaf:schoolHomepage rdf:resource="http://ssu.ac.kr"/>
<foaf:knows>
<foaf:Person>
<foaf:name>haklae kim</foaf:name>
<foaf:mbox_sha1sum>c9dee97e61431d1bb6b5127ace29ede7146c9cb6</foaf:mbox
_sha1sum></foaf:Person></foaf:knows>
<foaf:knows>
<foaf:Person>
<foaf:name>myungjin lee</foaf:name>
<foaf:mbox_sha1sum>2d75f0ee91aa7cdac91b5b0691074600aa5d4036</foaf:mbox_
sha1sum></foaf:Person></foaf:knows></foaf:Person>
</rdf:RDF>
```

FOAF는 RDF 구문을 기반으로 하고 있기에, 그룹(group)에 소속되어 있는 개인(person), 조직(organization)에 의해 생성된 문서(document), 한 개인과 연결된 다른 개인 정보와 같이 의미적 연결 처리를 할 수 있으며, FOAF 문서들의 통합도 가능하다. LiveJournal, MyBlogLog, 그리고 Hi5.com 등의 경우 FOAF export를 제공하고 있으며, FOAF 정보들을 연계 처리하는 응용들로는 Force Directed Graph(자바 스크립트 기반 그래프 라이브러리)를 이용해 FOAF 문서에 있는 정보를 표현해 주는 Visual FOAF Explorer와 FOAF 데이터를 이용해 소셜 네트워크를 분석하여 시각적으로 보여주는 Flink와 같은 응용들이 있다.

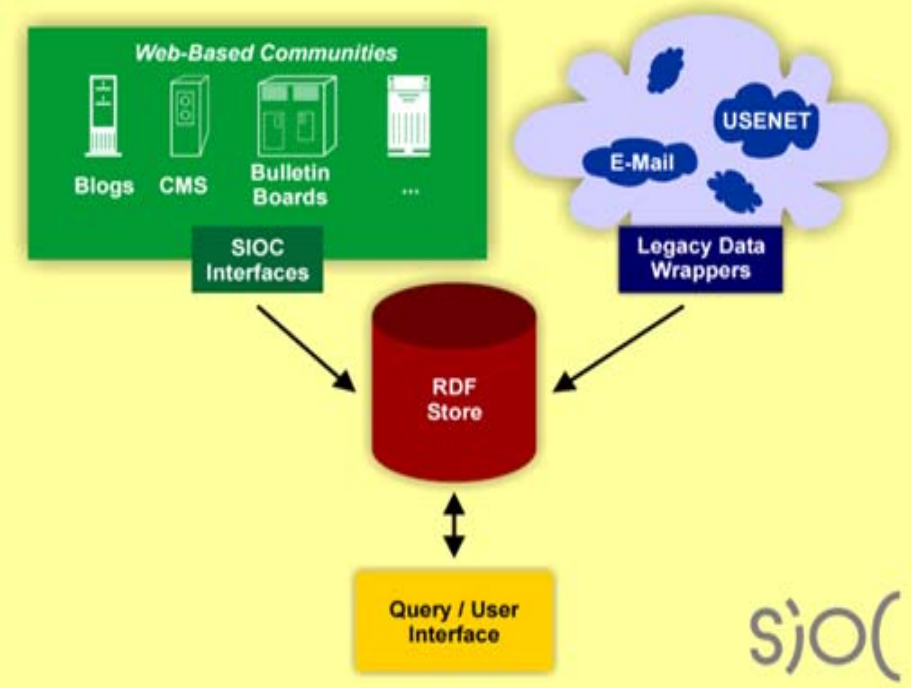
2.1.3 SIOC

SIOC(Semantically-Interlinked Online Communities)은 시맨틱웹 기술을 이용하여 커뮤니티를 연결하고자 하는 목적으로 개발된 어휘집이다. 2004년 DERI를 중심으로 시작되었으며, 시맨틱웹 기술을 이용해 온라인 커뮤니티를 의미적으로 연결하기 위한 시도의 프로젝트이다. SIOC은 커뮤니티 사이의 관계를 표현하기 위해 RDF/RDFS를 이용하는 SIOC 온톨로지를 사용하며 FOAF, RSS 등과 같은 다른 메타데이터 어휘 집합을 이용해 확장 가능한 구조를 갖고 있다. SIOC 온톨로지는 DC(Dublin Core)와 FOAF를 재활용하고 있으며, 시맨틱 소셜네트워크, 시맨틱 소셜미디어 서비스 구축을 위해 포괄적으로 사용될 수 있다. SIOC의 활성화를 위해 다양한 Exporter가 개발되고 있으며, Exporter는 블로그, 포럼 및 소셜 네트워크 서비스와 플러그인 형태로 연결되어서 SIOC 메타데이터를 생성한다. 현재 WordPress, Drupal, Twitter, phpBB, BlogEngine.Net 등이 이러한 기능을 지원한다. 아래의 그림의 SIOC 온톨로지를 설명하고 있다.



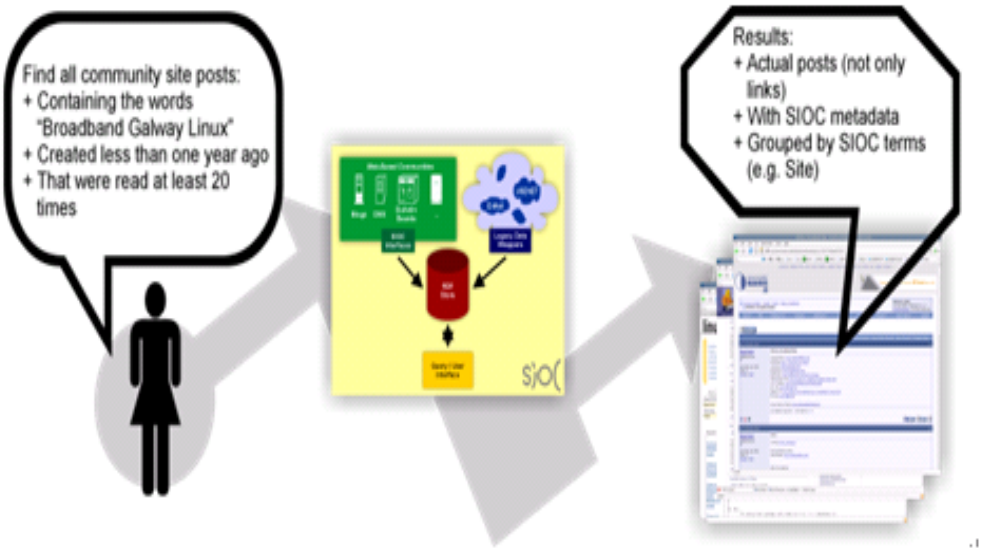
<그림 15> SIOC 온톨로지 개념

아래 그림은 SIOC 아키텍처이다. 기존의 폐쇄된 커뮤니티 사이트에서 SIOC 인터페이스인 Legacy Data Wrapper를 통해 RDF로 저장되고 이를 사용자들이 SIOC와 SPARQL을 통해 폐쇄성을 허물 수 있도록 하고 있다. 또한 SIOC Crawler가 SIOC Repository로 부터 원하는 데이터를 수집할 수 있도록 지원한다.



<그림 16> SIOC을 이용하는 시스템 구조

SIOC은 다양한 커뮤니티 사이트에서 SIOC 데이터를 사용할 수 있도록 하고 있으며, FOAF와 SKOS, Dubin Core와 어휘를 같이 조합할 수 있어 상호 연동이 가능하고 SIOC exporter를 통해 다른 메타웹과 시스템적으로 연동이 가능하여 다양한 응용 비즈니스로의 확장이 가능하다. 이렇게 함으로써 과거의 1차원의 게시판의 글 등을 연결해주는 서비스가 가능하고, 폐쇄되어 있는 공간의 지식을 확장하여 집단지성의 영역으로도 확장될 수 있는 커다란 장점을 갖게 된다.



<그림 17> SIOC의 역할

2.1.4 Dublin Core

더블린 코어(Dublin Core)는 ISO 15836으로 표준화된 메타데이터 요소 집합이다. 다시 말해서, 이 표준은 메타데이터들에서 사용되는 기초적인 관례들을 표준화하여 검색 및 처리가 용이하게 한다. 더블린 코어는 동영상, 소리, 이미지, 텍스트, 웹페이지 등의 디지털 매체들을 기술하는 데 널리 사용되며, 보통 XML과 RDF를 사용하여 구현된다.

DC의 목표는 데이터의 형식과 구조를 단순화하여 원문의 저자나 발행자가 메타데이터를 직접 작성하고, 이를 바탕으로 특정분야에서 요구되는 수준으로 확장하여 사용할 수 있도록 하는 것이다. 이처럼 다양한 영역의 네트워크 자원 기술을 위한 표준이며, 단순성, 확장성, 구문 독립성 등과 같은 특성을 지니고 있다. 비디오, 오디오, 이미지, 텍스트 뿐만 아니라 웹페이지와 같은 복합매체에도 쉽게 적용이 가능하다. 메타데이터의 요소가 너무 단순하여 디지털도서관, 교육매체관리 등과 같은 응용 영역에는 부족하다는 비판이 있지만, 단순한 구조의 시스템 개발, 교환 및 통합용 메타데이터, 응용 프로파일 개발의 기본 메타데이터로 등에 있어서는 유용성이 매우 많다.

더블린 코어 표준에는 두 가지 단계가 들어 있다. 하나는 단순 더블린 코어이며, 다른 하나는 한정적 더블린 코어이다. 또한 요소가 적용되는 범위를 제한하기 위한 한정자(qualifier)도 추가되어 있다.

단순 더블린 코어 메타데이터 요소 집합은 아래와 같다.

요소		설명
1	Title(표제)	자원에 부여되는 이름
2	Creator(제작자)	자원의 내용물을 만드는 데 일차적인 책임을 가지는 개체(엔티티)
3	Type(유형)	자원의 내용물의 성격이나 장르
4	Contributor(기여자)	자원의 내용물에 기여한 책임을 가지는 개체
5	Publisher(발행처)	자원을 이용할 수 있도록 만드는 데 책임을 가지는 개체
6	Date(날짜)	자원의 존재 기간 동안 발생하는 이벤트에 관련된 일자
7	Language(언어)	자원의 지적 내용의 언어
8	Format(형식)	자원의 물리적 구현 형식 또는 디지털 구현 형식
9	Description(설명)	자원의 내용물에 대한 설명
10	Subject(주제)	자원의 내용물에 대한 주제
11	Relation(관계)	관련 자원에 대한 참조
12	Identifier(식별자)	특정 맥락 내에서의 자원에 대한 명백한 참조
13	Rights(권한)	자원이 가지는 권리나 자원에 대한 권리에 관한 정보
14	Source(출처)	현재 자원이 파생된 자원에 대한 참조
15	Coverage(범위)	자원의 내용물의 수량이나 범위

<표 1> 더블린코어 메타데이터 요소

2.1.5 Bibliographic Ontology

Bibliographic Ontology는 시맨틱 웹 상의 RDF로 문헌정보를 기술한다. 인용 온톨로지로, 문헌 분류 온톨로지로, 또는 어떤 종류의 문헌들을 RDF로 기술하기 위한 방법으로 사용되어질 수 있다. 현재 존재하는 다양한 문헌 기술 메타데이터 포맷들의 영향을 받으며 다른 문헌 데이터 소스로 변환하는 바탕으로 이용되어질 수 있다.

Bibliographic Ontology는 http://purl.org/ontology/bibo/ 네임스페이스 URI로 구별되어진다. Bibliographic Ontology는 RDF를 사용함으로 효과적인 확장 가능성 메카니즘을 획득

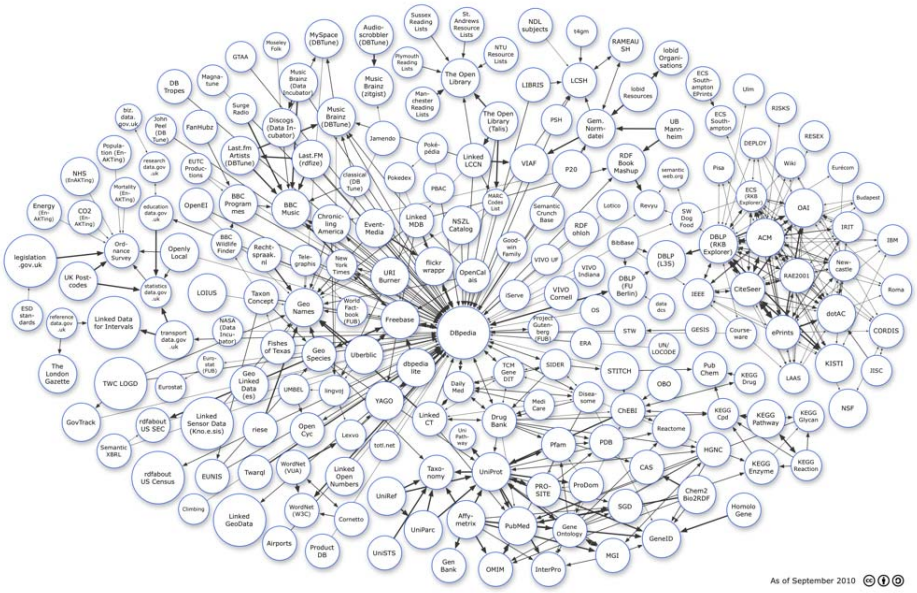
하고, 문헌 온톨로지 기반 표현을 다른 RDF 어휘로 표현된 다른 것과 어울리게 되는 것을 허용한다.

Bibliographic Ontology는 인용과 문헌참조와 관련된 것에 대해 논의하기 원하는 모든 것을 포함할 수 없다. Bibliographic Ontology에 모든 주제를 포함하는 것 대신에 기본 주제를 기술하고 RDF와 같은 큰 프레임워크에 구축한다. 이는 보다 명확한 기술 어휘들에서 처리하는 이점을 제공한다.

2.2. Linked Data 현황

2.2.1 Linked Open Data Cloud

2011년 10월 현재 발표된 LOD(Linking Open Data) Cloud는 2010년 9월에 발표된 아래의 그림이 최신 버전이다.



<그림 18> 2010년 9월 발표된 LOD Cloud

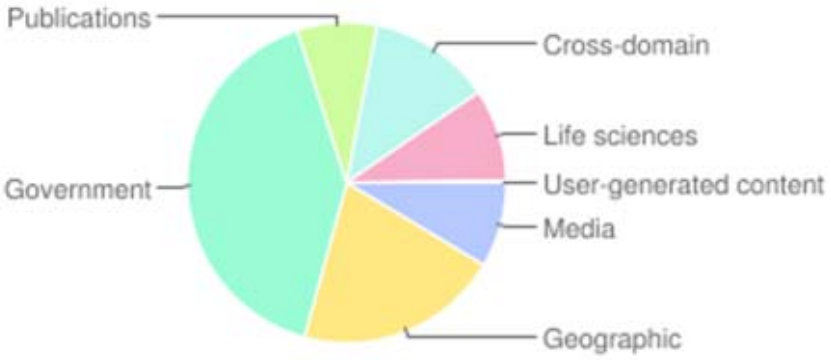
2.2.2 영역별 Linked Data

Linked Data는 다양한 영역에서 활발하게 확산되고 있다. 아래 표에서는 각각의 영역별 트리플 수 및 데이터 셋, 또 다른 외부 데이터 셋과의 연계와 같이 현황을 담고 있다.

Domain	Number of datasets	Triples	%	(Out-)Links	%
Media	26	2,454,298,811	8.59%	50,376,504	12.73%
Geographic	16	5,907,260,125	20.68%	16,539,378	4.18%
Government	26	11,630,305,143	40.72%	17,944,125	4.53%
Publications	68	2,307,901,534	8.08%	77,955,398	19.70%
Cross-domain	22	3,541,130,435	12.40%	29,120,577	7.36%
Life sciences	42	2,664,119,184	9.33%	200,417,873	50.64%
User-generated content	7	57,463,756	0.20%	3,402,228	0.86%
		28,562,478,988		395,756,083	

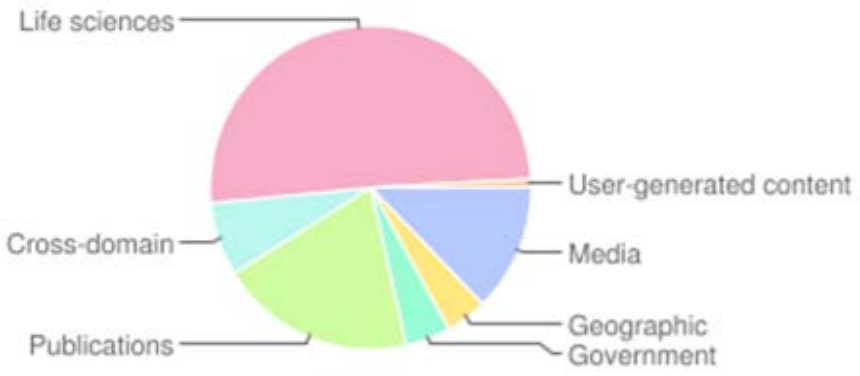
<표 2> 도메인별 LOD 트리플 현황

아래는 도메인별 트리플 분포를 가시적으로 나타낸 것이다.



<그림 19> 도메인별 LOD 트리플 분포
(출처:http://www4.wiwiss.fu-berlin.de/lodcloud/state/)

Linked Data는 제공하는 데이터의 품질도 중요하지만, 데이터들이 웹이라는 유통 환경에서 서로간의 의미 있는 링크를 통해 웹의 정보 유통 환경을 보다 더 유의미하게 하기 위한 것이다. 그래서 도메인별 링크 분포는 무엇보다 중요한 요소로 볼 수 있으며 아래는 도메인별 링크 분포를 가시적으로 나타낸 것이다.



<그림 20> 도메인별 LOD 링크 분포
(출처:http://www4.wiwiss.fu-berlin.de/lodcloud/state/)

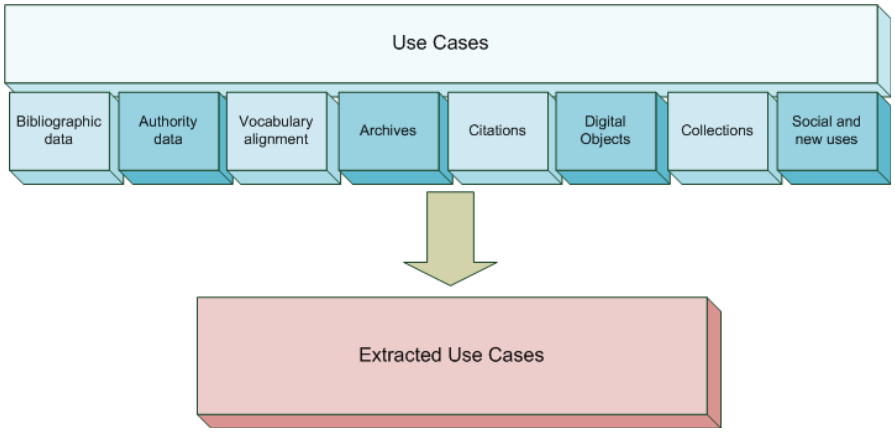
2.3 도서관 분야 Linked Data

2.3.1 Library Linked Data Incubator Group

LLD(Library Linked Data Incubator Group)는 W3C의 Incubator Group으로 도서관 분야의 상호운용성 확보를 위해 Linked Data를 적용하는 방법을 논의하는 장으로 주목받고 있다. 2010년 5월에 창설되었으며, 현재의 최종 보고서는 2011년 10월에 발표된 버전으로 "http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/"에서 확인할 수 있다. LLD의 미션이자 목표는 시맨틱웹 기술, 특히 Linked Data를 이용해 도서관 분야의 데이터를 도서관 내부에서만 유통하지 말고 도서관 외부로 끌어내어 타 분야와의 융합을 용이하게 하고, 도서관 분야 내에서도 웹을 이용해 상호운용성을 높이고자 하는 것이다. 결국은 도서관 분야 데이터의 상호운용성 확보가 주된 목표이다. 도서관 분야와 관련된 많은 Linked Data 사례들은 “Library Linked Data Incubator Group: Use Cases”에서 설명하고 있다.

LLD의 Final Report에서도 강조하고 있는 것은 현재 도서관 분야의 데이터들이 도서관 분야 내에서만 다양한 도서관 표준을 통해 유통되고 있다는 것이다. 타 분야와의 융합과 좀 더 효율적인 도서관 정보를 활용하기 위해서는 웹을 통한 유통이 이루어져야함을 강조하고 있다. 두 번째로는 아직도 많은 데이터들이 사람이 보기 위한 자연어 구조로 되어 있으며, MARC와 같은 구조를 통해 최소의 정보만이 컴퓨터가 이해할 수 있는 구조로 되어 있어, 식별될 수 있는 많은 정보들이 연결된 정보로써 사용될 수 없는 한계에 대해 지적하고 있다. 또한 용어 통제에 대한 문제도 언급하고 있다. 도서관 분야와 타 분야가 서로 같은 개념을 서로 다른 용어로 사용함으로써 발생하는 문제가 그것이다.
최근 들어 도서관 분야에서도 Linked Data를 적용하는 사례들이 늘고 있다. 특히 국가 대표 도서관들이 앞장서고 있는데, 대표적으로는 미국, 스웨덴, 헝가리, 독일, 프랑스, 영국 등이 있으며, UN의 FAO(Food and Agriculture Organization)와 OCLC(Online Computer Library Center)도 Linked Data의 적용에 앞장서고 있다.

도서관 분야를 포함해서, 그 이외의 분야도 그렇지만, 항상 정보라는 것에는 권리라는 것이 수반된다. 저작권이나 개방의 범위 등이 이에 해당되는 문제인데, Linked Data가 무조건 Open Data일 필요는 없지만, Open Data의 특성이 크기 때문에 중요한 이슈로 여겨야함은 당연하다.
W3C의 LLD에서는 도서관 분야의 사례를 8개의 클러스터로 구분하여 설명하고 있다. 8개의 클러스터 구성은 아래 그림과 같다.

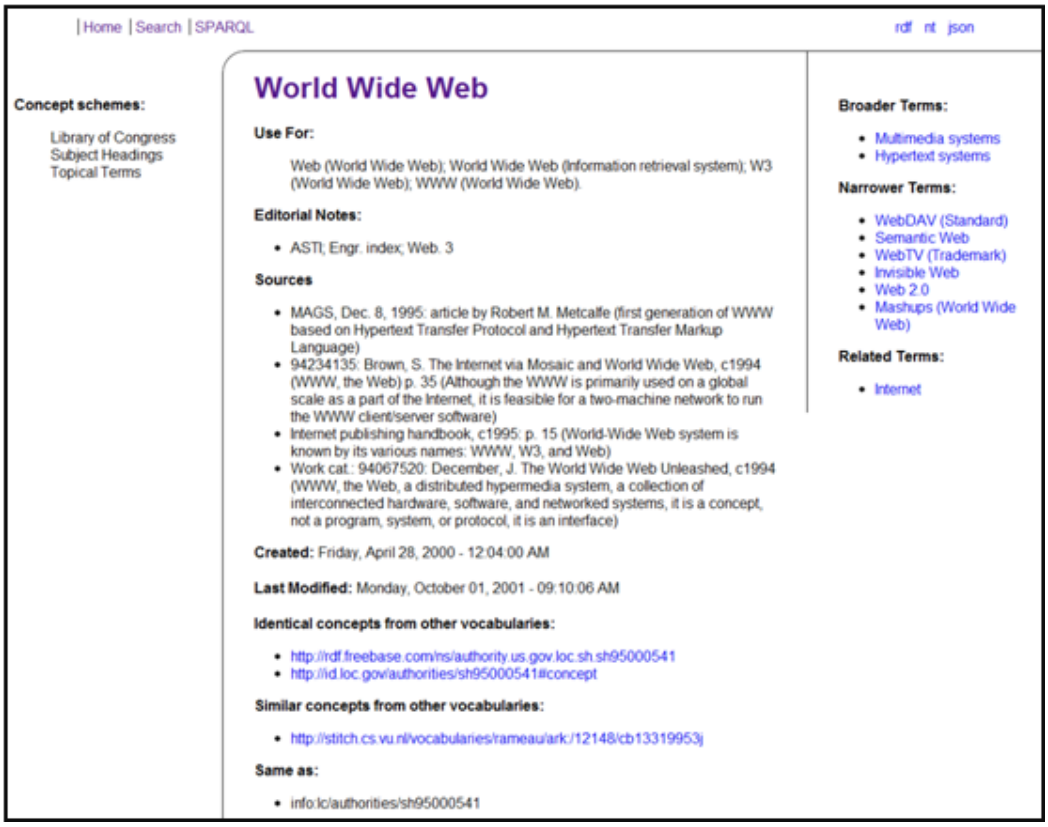


<그림 21> Library Linked Data Use Case Cluster

2.3.2 미국 의회도서관

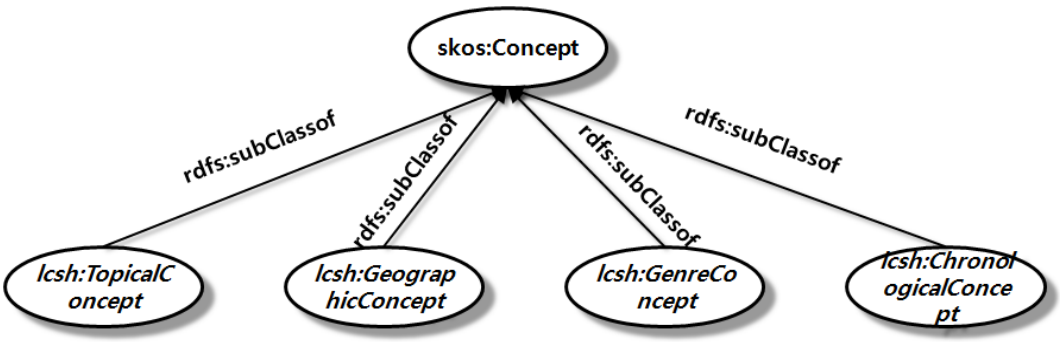
미국 의회 도서관은 자료의 주제목록(Subject Cataloging)을 위한 Subject Headings를 Linked Data, 즉 RDF로 표현하여 발행하고 있으며, MARC21 포맷으로 되어 있는 전거 데이터(Authority Record)를 SKOS Vocabulary를 이용하여 RDF로 변환하고 이를 웹상에 발행하여 서비스하고 있다. 전거 레코드의 각 항목(Entry)을 skos:Concept의 인스턴스로 표현하고 이를 LCCN(Library Congress Control Number)을 이용하여 URI를 부여하고 있으며, 미국 의회 도서관이 LCCN을 이용해 발행하는 URI는 아래와 같은 형태를 갖는다.

- http://lcsubjects.org/subjects/sh95000541



<그림 22> 미국 의회 도서관에서의 World Wide Web을 표현한 Linked Data

미국 의회 도서관에서는 SKOS를 활용하고 있으며, 개념(Concept) 관계는 아래와 같다.



<그림 23> 미국 의회 도서관 LCSH에서 활용하는 개념(Concept) 관계

미국 의회 도서관은 MARC 형식을 SKOS를 이용해 변환하여 Linked Data를 발행하는데 MARC와 SKOS의 매핑 관계는 다음과 같다.

MARC 필드	Feature/Function	RDF Property
010	Control Number	rdf:about
150	Topical Term	skos:prefLabel
151	Geographic term	skos:prefLabel
450	See From Tracing(Topical Term)	skos:altLabel
451	See From Tracing(Geographic Name)	skos:altLabel
550	See Also From Tracing(Topical Term)	skos:broader
550	See Also From Tracing(Topical Term)	skos:related
551	See Also From Tracing(Geographic Name)	skos:broader
551	See Also From Tracing(Geographic Name)	skos:related
667	non public general note	skos:note
670	source data found	dcterms:source
675	Source data not found	skos:editorialNote
678	Biographic or historical data	skos:definition
680	Public general note	skos:scopeNote
681	Subject example tracing note	skos:example
682	Deleted heading information	skos:changeNote
688	Application history note	skos:historyNote
008	Fixed Length Data Elements	dcterms:created
005	Date and time of last transaction	dcterms:modified
053	LC Classification Number	dcterms:lcc

<표 3> LCSH/SKOS 매핑 관계

또, 미국 의회 도서관에서 발행하는 LCSH Linked Data는 외부 데이터와의 연계를 지속적으로 확산하고 있으며, 주요 데이터 셋과의 연계는 아래와 같다.

o Geographical Headings와의 연계
- GeoNames(http://geonames.org)
- CIA World Fact Book(http://www4.wiwiss.fu-berlin.de/factbook/)
o LCSH Concept이 제시된 자료와 연계
- RDF BookMashup (http://www4.wiwiss.fu-berlin.de/bizer/bookmashup)
o 일반적인 정보연계
- DBPedia (http://dbpedia.org)
o LCSH와 REMEAU(프랑스 국립도서관)Subject Headings와 상호연계

2.3.3 VIAF(Vritual International Authority File)

VIAF(Virtual International Authority File)는 전 세계 전거 데이터의 공유 증진을 목적으로 하고 있으며, 이를 통해 목록 비용의 절감과 함께 국제적인 전거 통제를 목적으로 하고 있다. OCLC(Online Computer Library Center : http://www.oclc.org)에 의해 운영되고 있으며, 전 세계 16개국이 주요 참여국으로 참여하고 있다. 1,300만개의 Name Records를 각 국가가 사용하는 언어로 표현하여 서비스하고 있으며, 각 국가별로 VIAF에 참여하고 있는 기관은 다음과 같다.

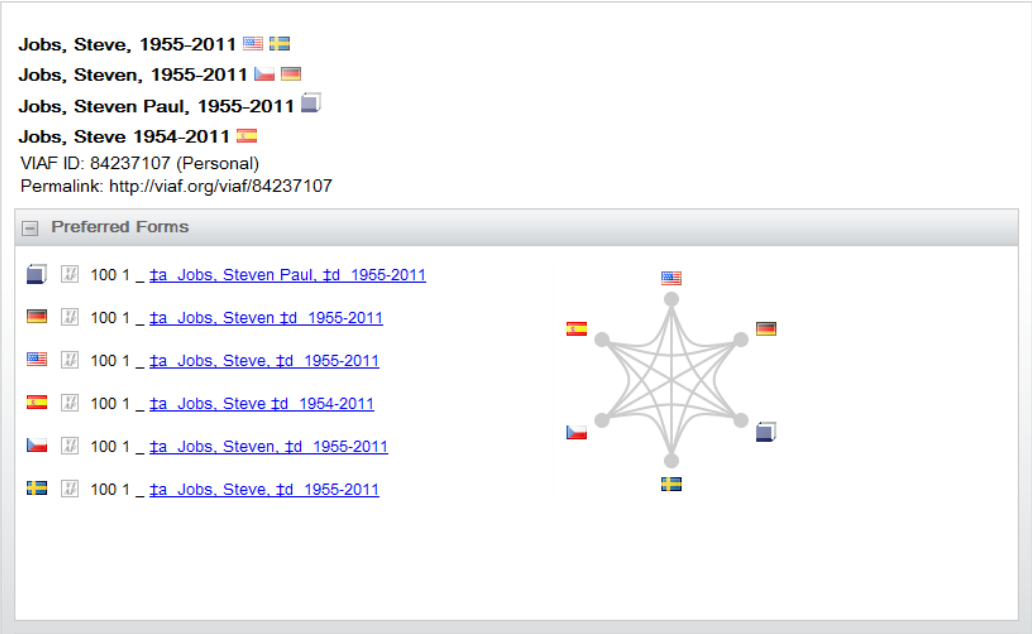
- 오스트레일리아 : National Library of Australia
- 체코 : National Library of the Czech Republic
- 이집트 : Bibliotheca Alexandrina (Egypt)
- 프랑스 : Bibliothèque nationale de France
- 독일 : Deutsche Nationalbibliothek
- Getty Research Institute
- 이스라엘 : National Library of Israel

- 이탈리아 : Istituto Centrale per il Catalogo Unico (Italy)
- 포루투칼 : Biblioteca Nacional de Portugal
- 스페인 : Biblioteca Nacional de España
- 스웨덴 : National Library of Sweden
- 스위스 : Swiss National Library
- 미국 : Library of Congress/NACO
- 바티칸 : Vatican Library
- 폴란드 : NUKAT Center (Poland)
- 캐나다 : Library and Archives Canada
- 헝가리 : National Széchényi Library (Hungary)
- 스위스 : RERO (Switzerland)
- 러시아 : Russian State Library-test
- 벨기에 : Flemish Public Libraries-test

VIAF는 각 항목마다 URI를 부여하여 Linked Data로 이용 가능하도록 하고 있으며, VIAF가 다루고 있는 범위는 다음과 같다.

- Personal Names
- Geographic
- Corporate
- Title
- Family
- Events

VIAF가 서비스하고 있는 Steve Jobs에 대한 서비스 페이지는 다음과 같다.



<그림 24> VIAF에서의 Steve Jobs

VIAF는 하나의 이름이 여러 사람에게서 사용되는 경우, 한명의 이름이 다양한 표기법으로 통용되는 경우의 문제점을 해결하고자 시작하였다. VIAF는 기본적으로 하나의 개체, 즉 한명이 여러 개의 언어와 여러 개의 방법으로 표현되어질 수 있도록 하고 있다. http://viaf.org/viaf/29541064에 대한 표현 예시는 다음과 같다.

<그림 25> VIAF의 http://viaf.org/viaf/29541064 예시

또한, VIAF는 동일 개체에 대해 MARC는 물론 Linked Data를 위한 RDF, HTML 형식 등을 지원하며, VIAF가 지원하는 표현은 다음과 같다.

형식	URI
Default	http://viaf.org/viaf/9855044
Real World Object	http://viaf.org/viaf/9855044.rwo
HTML	http://viaf.org/viaf/9855044.html
XML	http://viaf.org/viaf/9855044.viaf
RDF(FOAF)	http://viaf.org/viaf/9855044.rdf
MARC21	http://viaf.org/viaf/9855044.m21
UNIMARC	http://viaf.org/viaf/9855044.unimarc

<표 4> VIAF가 지원하는 동일 개체의 표현 방법

VIAF는 어떤 개체가 동일 개체인지, 아니면 다른 개체인지 판별하기 위해 여러 가지 메타데이터를 이용하는데, 주로 이용하는 메타데이터와 2009년 07월까지의 데이터 사이즈는 다음과 같다.

- 1,705,555 Title
- 846,722 Double date
- 123,487 Joint author
- 71,851 LCCN
- 24,587 Partial date and partial title
- 11,010 Partial date and publisher
- 9,179 Partial title and publisher
- 6,415 Name as subject
- 3,168 Standard number

이 중에서 날짜(Date) 정보를 통해, 동일 개체인지, 다른 개체인지 판단하는 예시는 다음과 같다.

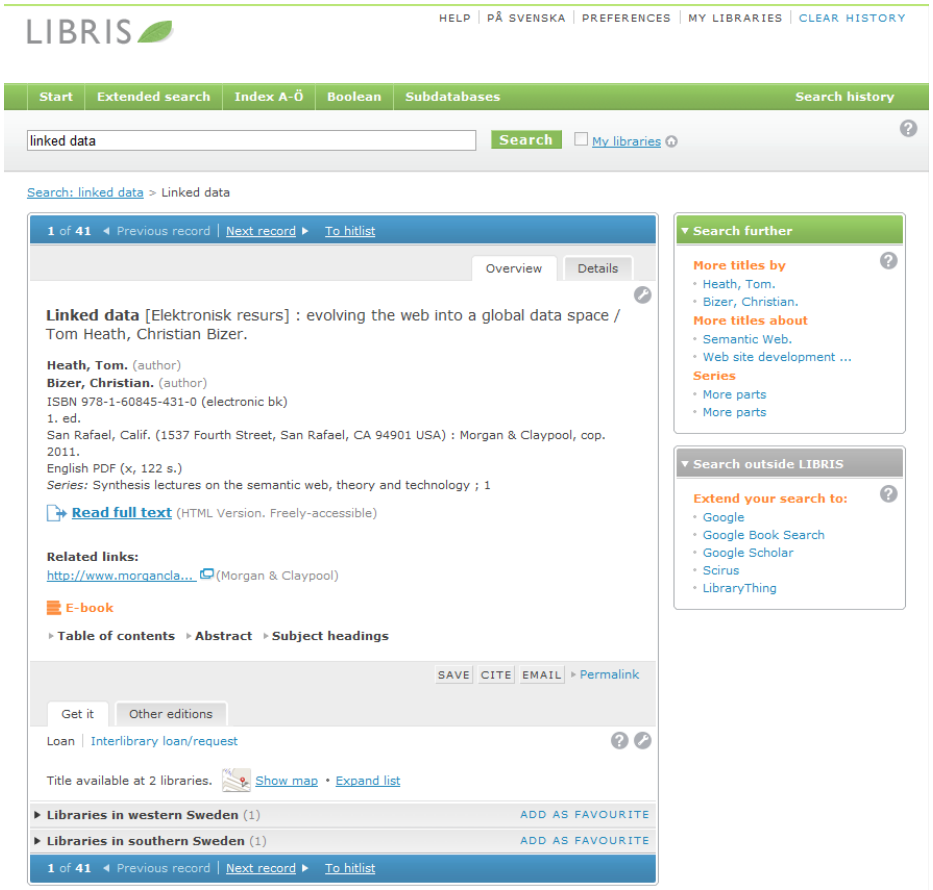
60 headings found for *1946 1947*

	Name	Sample Title
1	Spielberg, Steven, 1947- Spielberg, Steven, 1946-	Amistad, "give us free" : a celebration of the film b essays Jurassic park
2	Mathieu, Mireille 1947- Mathieu, Mireille, 1946-	Mireille Mathieu Mireille Mathieu
3	Chopra, Deepak Chopra, Deepak, 1947- Chopra, Deepak 1946-	Buddha a story of enlightenment La voie du magicien vingt leçons spirituelles pou Alle Kraft steckt in dir
4	Roussos, Demis 1947- Roussos, Demis Roussos, Demis, 1946-	Demis Roussos Demis Roussos Demis Roussos

<그림 26> VIAF에서 날짜(Date) 정보를 이용한 개체 판별 예시

2.3.4 LIBRIS - 스웨덴 국립도서관

LIBRIS는 도서관 전체 데이터를 Linked Data로 구축한 최초의 종합목록(2008년)으로, 대학도서관 및 연구도서관, 박물관, 아카이브, 공공도서관 등 170여개 도서관 기관이 이용하는 종합목록시스템이다. 6백만 개의 서지 데이터 및 2천만 개의 도서관 장서목록, 2십만 개의 전거 데이터로 구성되어 있으며, 목록작업과 함께, 데이터 자동 임포트/익스포트 기능 및 Open API기능을 제공하고 있다. 또한 LIBRIS는 RDF파일 포맷을 제공하여 Linked Data를 구현하고 있는데, "Linked data : evolving the web into a global data space / Tom Heath, Christian Bizer"라는 도서목록에 대한 HTML 서비스 주소는 “http://libris.kb.se/bib/12130810“이며, 서비스 페이지의 일부는 아래와 같다.



<그림 27> LIBRIS 서비스 페이지 예시

위의 HTML 서비스 페이지를 Content Negotiation에 따라 "data" 키워드를 추가하여 “http://libris.kb.se/data/bib/12130810“로 요청하면 해당 개체에 대한 RDF를 반환하여 준다.

- http://libris.kb.se/bib/12130810
- http://libris.kb.se/data/bib/12130810

“http://libris.kb.se/data/bib/12130810“로 요청한 RDF의 일부는 다음과 같다.

```
▼<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bibo="http://purl.org/ontology/bibo/"
 xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:dc="http://purl.org/dc/elements/1.1/"
 xmlns:libris="http://libris.kb.se/vocabulary/experimental#" xmlns:rda="http://RDVocab.info/Elements/"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  ▼<rdf:Description rdf:about="http://libris.kb.se/resource/bib/12130810">
     <rdf:type rdf:resource="http://purl.org/ontology/bibo/Book"/>
     <dc:publisher>Morgan & Claypool</dc:publisher>
     <dc:identifier rdf:resource="URN:ISBN:1608454312"/>
   ▼<dc:description xml:lang="en">
       7. Summary and outlook -- Bibliography -- Authors' biographies.
     </dc:description>
     <dc:description xml:lang="en">Series from website.</dc:description>
     <dc:language rdf:resource="http://purl.org/NET/marccodes/languages/eng#lang"/>
   ▼<dc:description xml:lang="en">
       6. Consuming linked data -- Deployed linked data applications -- Generic applications -- Domain-specific applications --
       Developing a linked data mashup -- Software requirements -- Accessing linked data URIs -- Representing data locally using
       named graphs -- Querying local data with SPARQL -- Architecture of linked data applications -- Accessing the web of data --
       Vocabulary mapping -- Identity resolution -- Provenance tracking -- Data quality assessment -- Caching web data locally --
       Using web data in the application context -- Effort distribution between publishers, consumers and third parties.
     </dc:description>
     <dc:description xml:lang="en">Includes bibliographical references (p. 111-120).</dc:description>
     <dc:identifier>http://linkeddatabook.com/book</dc:identifier>
     <libris:held_by rdf:resource="http://libris.kb.se/resource/library/Zdig"/>
     <libris:held_by rdf:resource="http://libris.kb.se/resource/library/Ldig"/>
     <dc:creator>Christian Bizer</dc:creator>
   ▼<dc:identifier>
       http://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001
     </dc:identifier>
   ▼<dc:description xml:lang="en">
       4. Linked data design considerations -- Using URIs as names for things -- Minting HTTP URIs -- Guidelines for creating cool
       URIs -- Example URIs -- Describing things with RDF -- Literal triples and outgoing links -- Incoming links -- Triples that
       describe related resources -- Triples that describe the description -- Publishing data about data -- Describing a data set
       -- Provenance metadata -- Licenses, waivers and norms for data -- Choosing and using vocabularies to describe data -- SKOs,
       RDFs and OWL -- RDFs basics -- A little OWL -- Reusing existing terms -- Selecting vocabularies -- Defining terms -- Making
       links with RDF -- Making links within a data set -- Making links with external data sources -- Setting RDF links manually --
       Auto-generating RDF links.
     </dc:description>
     <dc:identifier rdf:resource="URN:ISBN:9781608454310"/>
     <bibo:isbn10>1608454312</bibo:isbn10>
   ▼<dc:description xml:lang="en">
       Part of: Synthesis digital library of engineering and computer science.
     </dc:description>
   ▼<dc:description xml:lang="en">
       5. Recipes for publishing linked data -- Linked data publishing patterns -- Patterns in a nutshell -- Additional
       considerations -- The recipes -- Serving linked data as static RDF/XML files -- Serving linked data as RDF embedded in HTML
       files -- Serving RDF and HTML with custom server-side scripts -- Serving linked data from relational databases -- Serving
       linked data from RDF triple stores -- Serving linked data by wrapping existing application or web APIs -- Additional
       approaches to publishing linked data -- Testing and debugging linked data -- Linked data publishing checklist.
     </dc:description>
   ▼<dc:description xml:lang="en">
```


<그림 28> LIBRIS의 http://libris.kb.se/data/bib/12130810 소스 일부

LIBRIS는 MARC21과 RDF를 매핑하고 있으며, 주요 매핑 형식은 다음과 같다.

o 전거파일

- 인명 -> foaf:Person
- 기관 -> foaf:Organization

o MARC와 온톨로지(FOAF, DBPedia) 매핑

o 서지 레코드의 기술을 위해 Doublin Core와 Bibliontology 사용

Field	subfields	RDF Property	Comments
100,400	a,b,c,d	foaf:name	personal name
100,400	a,b,c	foaf:name	personal name in straight form
100,400	d	dbpedia:dateofBirth dbpeida:dateofDeath	the d subfield is split on hyphen
110,410	a,b,c,d	foaf:name	corporate name
150	all except 0,6,8	skos:prefLabel	
550	all except i,w,0,6,8	skos:broader	subfield g='w'
550	all except i,w,0,6,8	skos:narrower	subfield g='h'
750	all except 0,6,8	skos:altLabel	

<표 5> LIBRIS의 MARC와 RDF 매핑

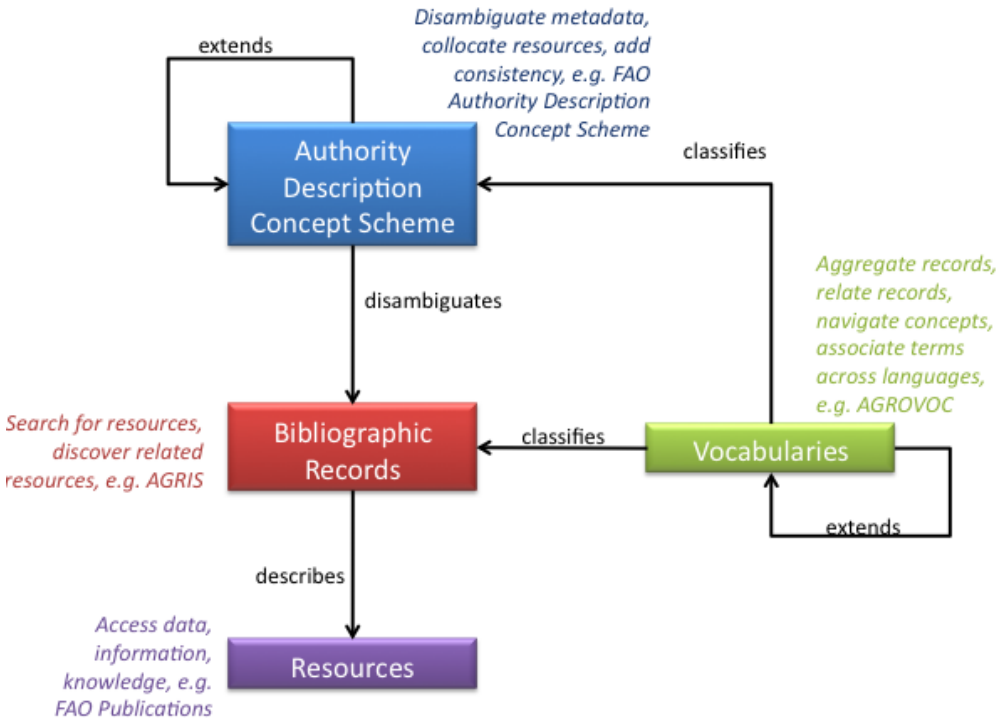
2.3.5 기타

o Bibliographic Network

- IFLA(The International Federation of Library Associations and Institutions)가 주도
- Linked Data를 통해 서지 정보의 정확한 명시와 공유 증진을 통한 상호운용성 확보에 노력

o AGRIS

- FAO(Food and Agriculture Organization of the United Nations)가 주도
- 전세계 150개 이상 기관, 100개 이상 국가의 농업과 관련된 과학, 기술 정보에 대한 논문, 연구보고서 등을 관리하는 AGRIS(International Information System for the Agricultural Sciences and Technology) 운영
- AGRIS Linked Data를 통해 농업 과학, 기술 정보에 대한 웹 기반의 상호운용성 확보에 주력
- AGRIS Linked Data Use Cases



<그림 29> Relationships Between FAO Linked Data USe Cases

2.4 공공분야 Linked Data

정부의 투명성 증진과 국민의 알권리를 충족 시켜 주는 효과에 더해, 정부가 생산, 관리하는 정보를 통해 보다 창의적인 서비스와 어플리케이션 개발을 위해 전 세계 정부는 공공데이터의 개방에 심혈을 기울이고 있다. 또한 데이터의 재사용성 증진과 함께, 어플리케이션에서 보다 유연하게 활용할 수 있도록 하기 위해 Linked Data를 적극 도입하고 있는 추세이다. 미국의 data.gov(http://www.data.gov/opendatasites)에서 조사하여 서비스하고 있는 정보에 의하면 현재 공공정보 개방에 참여하고 있는 국가는 다음과 같다.

<그림 30> 공공정보 개방에 참여하고 있는 국가들

- 호주 : http://data.gov.au/
- 영국 : http://data.gov.uk/
- 티모르 : http://www.transparency.gov.tl/public/index
- 스페인 : http://opendata.euskadi.net/
- 싱가폴 : http://data.gov.sg/
- 노르웨이 : http://data.norge.no
- 뉴질랜드 : http://www.data.govt.nz
- 모로코 : http://data.gov.ma/
- 몰도바 : http://data.gov.md/
- 케냐 : http://opendata.go.ke/
- 이탈리아 : http://www.dati.piemonte.it/
- 아일랜드 : http://www.statcentral.ie/

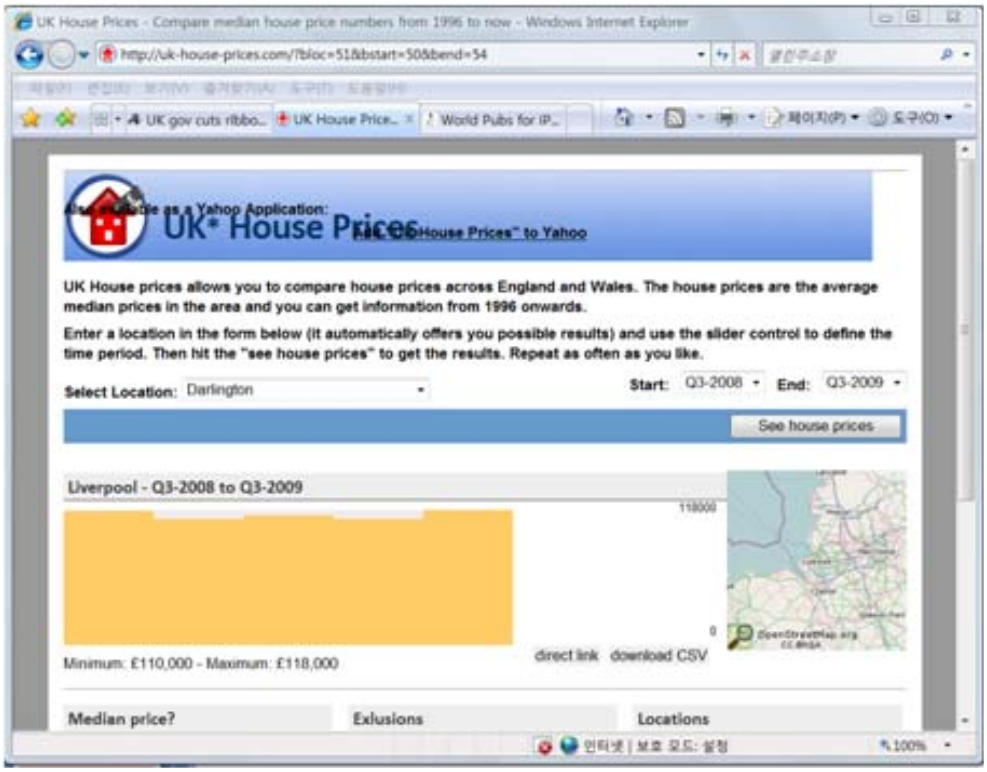
- 홍콩 : http://www.gov.hk/en/theme/psi/welcome/
- 그리스 : http://geodata.gov.gr/geodata/
- 독일 : http://www.portalu.de/ingrid-portal/portal/default-page.psml
- 핀란드 : http://www.suomi.fi/suomifi/tyohuone/yhteiset_palvelut/avoin_data/
- 프랑스 : http://opendata.montpelliernumerique.fr/,
  http://www.data.rennes-metropole.fr/,
  http://opendata.paris.fr/opendata/jsp/site/Portal.jsp
- 에스토니아 : http://pub.stat.ee/px-web.2001/Dialog/statfile1.asp
- 덴마크 : http://digitaliser.dk/resource/432461
- 캐나다 : http://www.data.gc.ca/default.asp?lang=En 외 다수
- 오스트리아 : http://data.wien.gv.at/

2.4.1 영국 data.gov.uk

영국 정부는 정책의 투명성을 높이고, 국민의 알권리를 향상시키며, 공공정보의 재사용성 증진을 통해 경제적 사회적 가치 증대를 유발시키고, 차세대 웹(web of data, 즉, Semantic Web)기술에서 자국의 인지도를 상승시키기 위해 Linked Data 기술을 활용하여 공공정보를 개방하고 있다. 2009년 02월 Power of Information Taskforce Report를 통해 정부가 공공정보의 접근과 활용을 제공하여 국민의 창의적 혁신을 지원하고, 개인과 단체가 삶의 질을 개선하기 위해 더 나은 정보에 의한 판단을 통해 솔루션을 만들어 낼 수 있도록 도와야 함을 주장하였으며, Digital Britain Report를 통해 영국의 인터넷 접근성(connectivity)을 높이기 위한 광대역 네트워크 및 무선 네트워크에 대한 인프라 투자 뿐만 아니라 모든 공공 데이터의 공개와 접근을 위한 작업에 착수해야 함을 주장하였다. 또한, 영국의 수상 고든 브라운은 2009년 07월 “공공정보(government information)를 광범위한 모든 사람들이 접근할 수 있고 이용할 수 있게 하기 위해, 특별 조언자로써 Tim Berners-Lee에게 웹으로 정부 데이터의 공개 접근을 추진하도록 도움을 줄 것을 요청하였다. 2010년 1월 공식적으로 data.gov.uk 서비스가 운영을 시작하였으며, 3,000개 이상의 Linked Data 데이터 셋을 서비스하고 있다. Linked Data를 이용하여 다양한 웹 기반, 또는 스마트폰 기반의 개방형 애

플리케이션들이 등록되고 있으며, 산업계와 연구 기관 간의 협력을 위한 허브로서의 역할을 수행하고 있다. 영국 정부가 개방한 데이터 셋을 Linked Data를 이용해 전 국민이 유용하게 활용할 수 있게 함으로써 다양한 응용 서비스들이 개발되고 활용되고 있는데, 주요 응용 서비스를 살펴보면 다음과 같다.

o UK HOUSE PRICES - 주택 가격 동향 서비스
- 주택가격의 동향을 지역별, 연대별로 시각화하여 서비스
- 부동산 등기소(the Land Registry)의 공공 정보를 이용



<그림 31> data.gov.uk를 이용한 주택 가격 동향 서비스

o Find GPs
- 주변 지역의 가까운 개인병원의 위치 검색 및 추천 서비스
- 데이터 셋 : Health and Social Care Information Center(HSCIC), the Office for national Statistics Access to Services Team

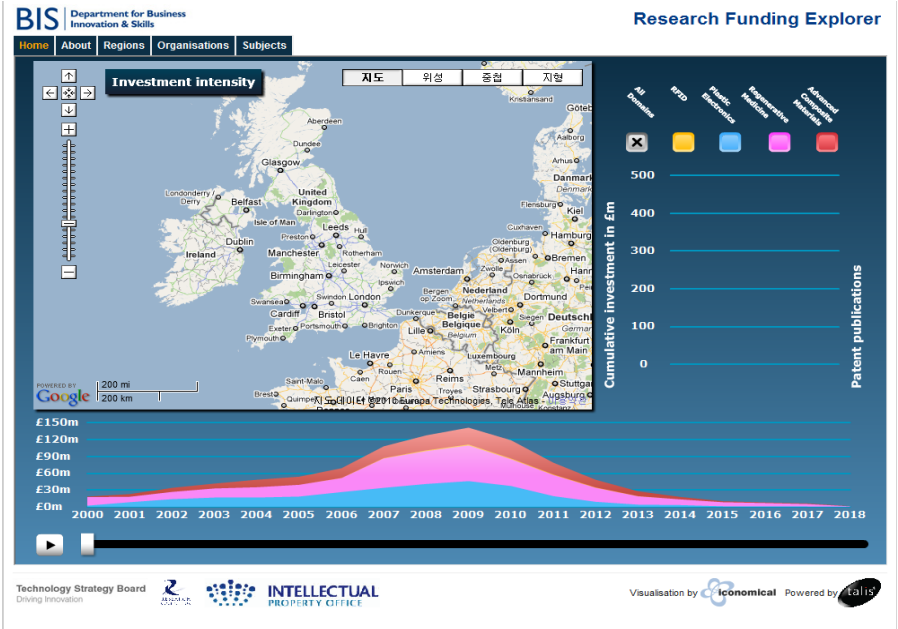
- 현재 위치에서 가장 가까운 개인병원을 검색하고 길찾기 안내
- 아이폰의 GPS를 이용한 위치기반 애플리케이션으로 iTunes에서 0.99$(0.59파운드)로 판매

<그림 32> data.gov.uk를 이용한 Find GPs

o Research Funding Explorer

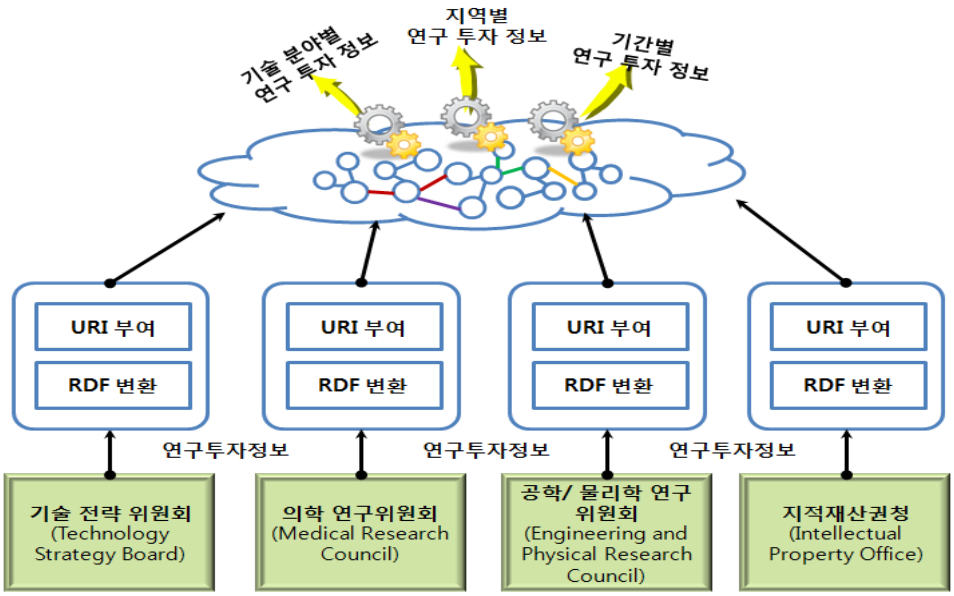
- http://bis.clients.talis.com/
- BIS(the Department for business innovation and skills)부처의 요구로 여러 정부기관에서 분산적으로 보유하고 있는 연구 투자 정보를 통합하여 하나의 뷰로 서비스하기 위하여 Linked Data 활용
  - 기술전략 위원회, 의학 연구위원회, 공학/물리학 연구위원회, 지적재산권청에 분산되어 있는 연구 투자 정보를 Linked Data를 통해 활용
  - RFID, Advanced Composites, Regenerative Medicine, Plastic Electronics 에 대한 투자 정보를 지역, 기간, 주제 분야에 따라 서비스

- 영국 정부의 Linked Data 기반 Research Funding Explorer 구현 사례



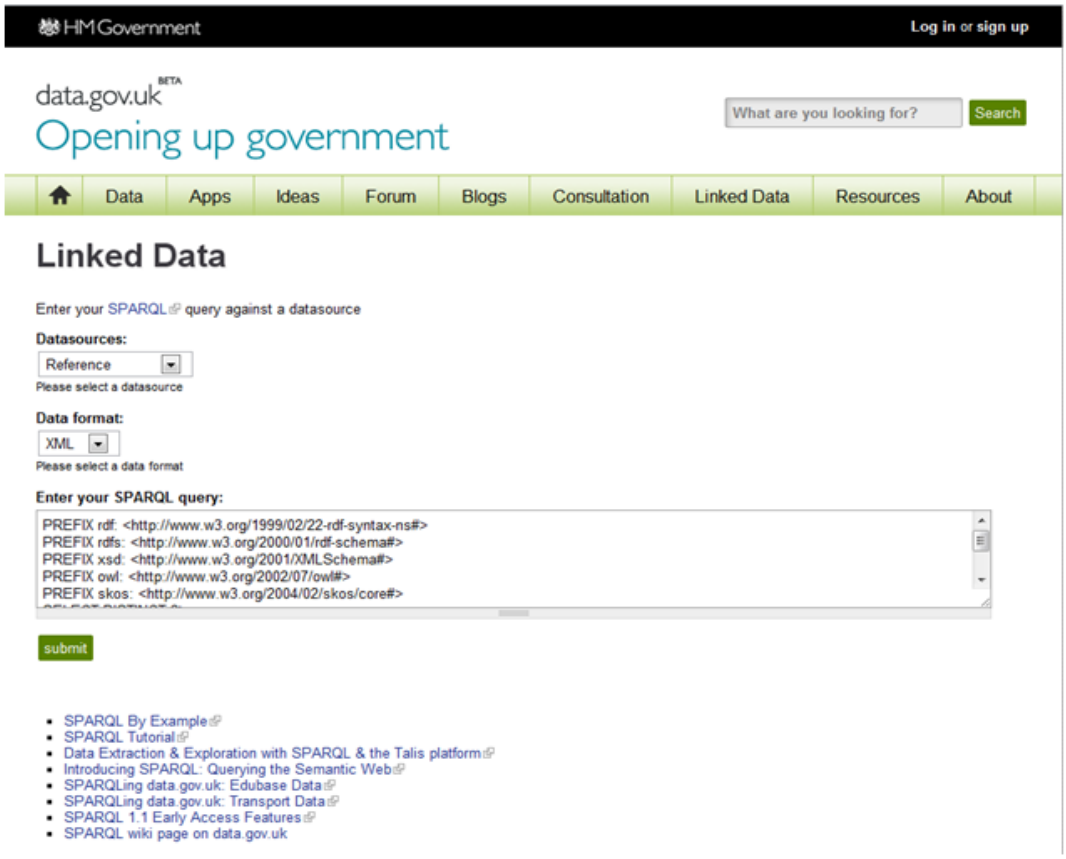
<그림 33> data.gov.uk를 이용한 Research Funding Explorer

- Research Funding Explorer의 개념



<그림 34> data.gov.uk를 이용한 Research Funding Explorer의 개념

영국의 data.gov.uk는 Linked Data의 재활용 및 창의적 이용을 위해 SPARQL Endpoint를 제공하고 있으며, http://data.gov.uk/sparql에서 확인할 수 있다.



<그림 35> data.gov.uk의 SPARQL Endpoint

영국의 data.gov.uk가 2011년 08월 현재 제공하는 SPARQL Endpoint의 Data Source는 다음과 같다.

- Reference
- Education
- Transport
- Ordnance Survey

또한, 영국의 data.gov.uk가 2011년 08월 현재 제공하는 SPARQL Endpoint의 질의 요청에 대한 반환 형식은 다음과 같다.

- XML
- JSON

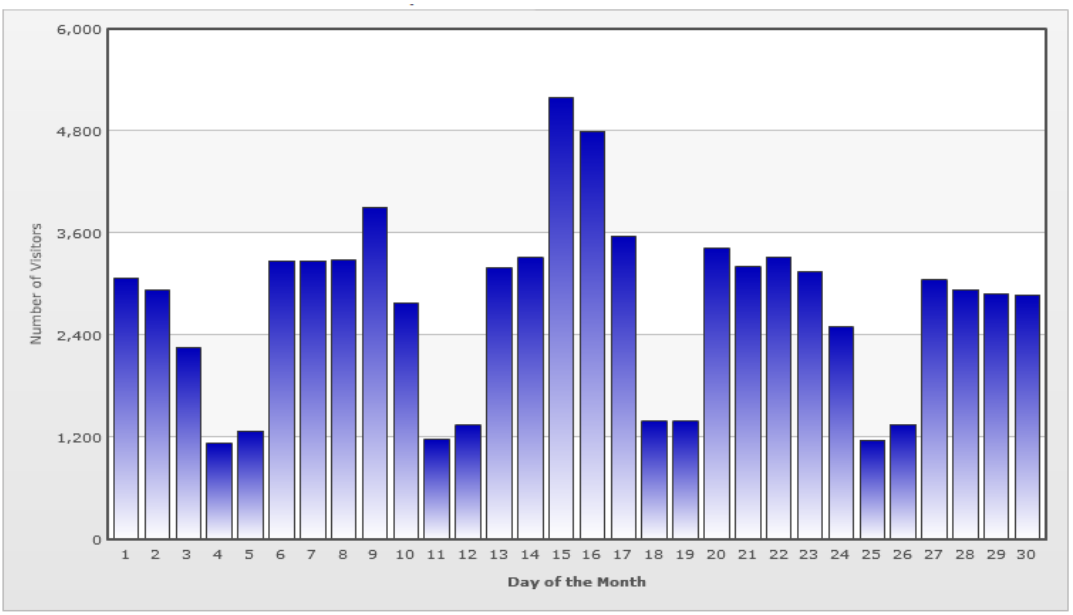
2.4.2 미국 data.gov

미국의 오바마 정부는 출범 즉시 연방정부를 더욱 투명하고, 참여적이며, 협력적으로 만들기 위한 Open Government Initiative 를 선포하였으며, 이를 실현하기 위하여 차세대 웹 기술을 이용한 국민의 정책 결정 참여 프로세스와 웹사이트(regulation.gov)를 구축하는 한편, 공공기관에서 보유하고 있는 공공정보(non-personal information)를 국민이라면 누구나 바로 이용할 수 있도록 하기 위해 2009년 5월 data.gov를 오픈하였다. data.gov 는 600개의 원(raw) 데이터 셋을 CVS/TXT, XML 및 SHP 포맷으로 발행하고, 10만 개 이상의 공간정보 데이터 셋을 보유하고 있으며, data.gov를 통해 미국의 국민과 영리적인 목적의 기관도 포함하는 여러 기관들은 자율적으로 필요한 각 기관의 공공데이터를 재사용할 수 있도록 하고 있다.

data.gov의 데이터 중 116개의 데이터 셋이 Linked Data로 서비스 중이며, 그 범위를 넓혀가고 있다. data.gov의 데이터를 활용한 다양한 종류의 어플리케이션들이 등장하고 있으며, http://explore.data.gov/catalog/apps에서 확인할 수 있다.

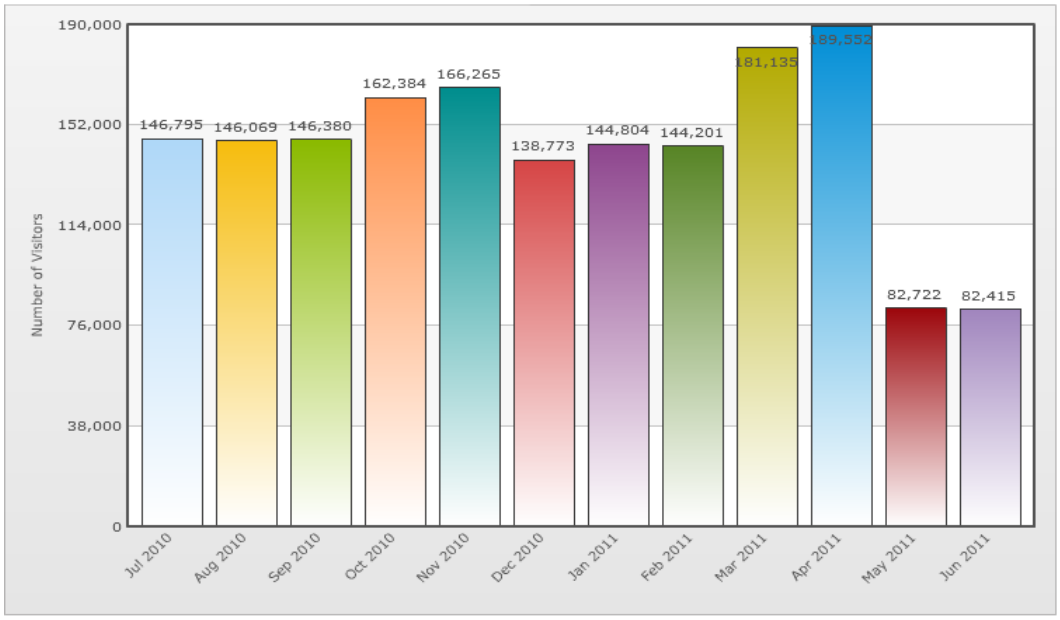
미국의 data.gov는 미국 인구 통계(US Census Data) 정보 헬스케어(http://www.data.gov/health) 정보등을 포함하고 있으며, RDF로 가공된 데이터 셋은 직접 RDF로 제공 받을 수도 있다.

2011년 06월 현재 하루당 방문자 수는 다음과 같다.



<그림 36> data.gov의 하루당 방문자 수 (2011년 06월)

2011년 06월 현재 월별 방문자 수는 다음과 같다.



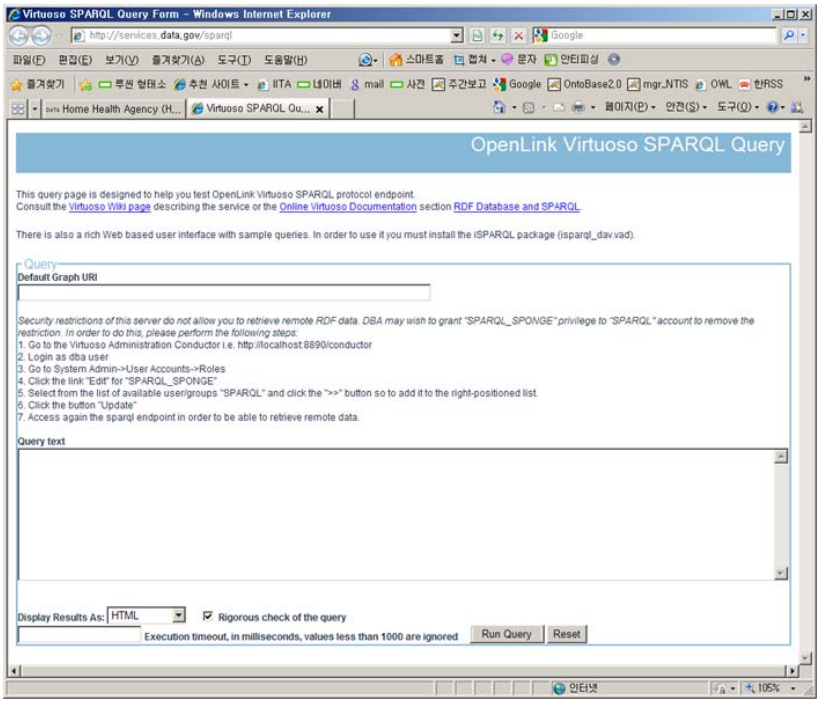
<그림 37> data.gov의 월별 방문자 수 (2011년)

2011년 06월 현재 카테고리별 다운로드 통계는 다음과 같다.

Data Category	Total Downloads
Agriculture	5,318
Arts, Recreation, and Travel	579
Banking, Finance, and Insurance	3,604
Births, Deaths, Marriages, and Divorces	1,884
Business Enterprise	2,082
Construction and Housing	1,154
Economic	0
Education	3,247
Elections	1,191
Energy and Utilities	4,799
Federal Government Finances and Employment	6,502
Foreign Commerce and Aid	1,498
Geography and Environment	71,433
Health and Nutrition	13,487
Income, Expenditures, Poverty, and Wealth	1,634
Information and Communications	5,179
International Statistics	983
Labor Force, Employment, and Earnings	7,993
Law Enforcement, Courts, and Prisons	4,260
Manufactures	53
National Security and Veterans Affairs	10,274
Natural Resources	544
Other	5,749
Population	6,544
Prices	773
Science and Technology	4,234
Social Insurance and Human Services	5,431
State and Local Government Finances and Employment	304
Transportation	3,683
Wholesale and Retail Trade	1,029

<표 6> data.gov의 카테고리별 다운로드 통계

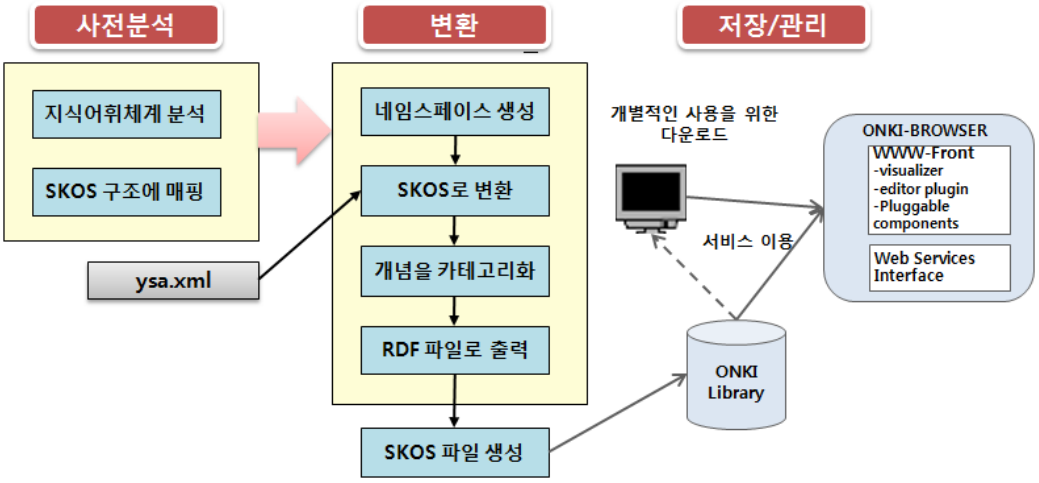
또한 SPARQL Endpoint를 통해 데이터의 재활용성 증진에 앞장서고 있다.



<그림 38> data.gov의 SPARQL Endpoint

2.4.3 핀란드

핀란드는 기업들이 서로 협력적이며, 지능적인 서비스를 만들 수 있는 환경을 제공하기 위한 목적으로 FinnONTO를 개발하였다. 핀란드 국립 도서관에서 개발한 시소러스인 YSA(Finnish General Upper Ontology) 및 Mesh 등을 SKOS를 이용하여 매핑하고 서로 연계하여 공통 온톨로지를 구축한 것이 FinnONTO이다.



<그림 39> FinnONTO의 데이터 구성

FinnONTO는 시맨틱웹 온톨로지 인프라 스트럭처를 국가적인 차원에서 개발하려는 프로젝트이며, 이 중 하나로 ONKI는 온톨로지를 기반으로 웹상의 콘텐츠와 서비스를 메타데이터에 연결하여 명확하고 이해하기 쉽도록 제공하기 위한 온톨로지 프로젝트이다. 2008년 9월부터 3개월 동안 36,000명의 방문자와 104,000번의 방문을 기록하는 등 이용자들의 참여율도 높아 시맨틱웹 기반 정보 서비스의 모범 사례로 꼽히고 있으며, 현재 약 40여 개의 도메인을 가지고 역사, 문화, 자연과학, 의학 정보, 경영 정보, 지리 정보 등의 분야에서 서비스되고 있다. 대부분 무료로 누구나 이용할 수 있는 어플리케이션을 제공하고 있으며, 주요 예로 Health Finland는 시민들의 건강관리를 촉진시키기 위해 만들어진 포털로 다이어트, 운동, 금연, 건강 생활 등에 관한 콘텐츠를 온톨로지와 메타데이터를 이용하여 하나의 포털에서 집합적인 콘텐츠를 제공하고 있다. 모든 카테고리는 라벨과 위계로 위치가 결정된 하위 개념을 가지고 있으며, 검색어가 입력되면 등록된 라벨과 위계에 맞게 콘텐츠를 일치시켜 제시한다. '다이어트'에 대해 검색을 하면 온톨로지 맵을 따라 '다이어트'에 관한 것뿐 아니라 관련 개념인 '몸무게 조절', '체중 감량', '식이요법', '에너지 섭취'등의 하위 개념과 연결하여 추천도 해준다. ONKI가 제공하는 대부분의 온톨로지는 현존하는 백과사전에 기반을 두어 만들어졌지만, 사용자가 직접 참여하는 매우 진화한 형태를 보이고 있으며, ONKI Selector는 사용자가 작성한 주석 같은 메타데이터와 해당 개념을 바로 연결할 수 있도록 'Fetch button'을 가지고 있다. 이는 시맨틱웹이 웹2.0의 집단 지성과 결합되는 양상으

로, 이용자를 사회적으로 연결하는 사회적 연결성(social connectivity)과 정보를 연결하는 정보적 연결성(informational connectivity)이 모두 높은, 미래의 웹이 나아가야 하는 바람직한 방향이라 볼 수 있다.

3. 시범 서비스 적용

3.1 시범 서비스 대상 데이터

3.2 온톨로지 모델 설계

3.3 온톨로지 명세

3.4 데이터 변환

3.4 Linked Data 발행 서비스

3. 시범 서비스 적용

3.1 시범 서비스 대상 데이터

국립중앙도서관의 데이터를 Linked Data로 발행하기 위해서 시범 서비스를 구성하였다. 시범서비스의 대상은 저자명 전거 데이터와 주제명 전거 데이터로 선정하여 구축하였으며, 각각의 데이터를 현재 잘 알려진 어휘를 바탕으로 트리플화 하여 시범 서비스로 적용하였다.

3.1.1 저자명 전거데이터

FOAF로 전환하기 위한 국립중앙도서관의 저자명 전거 데이터는 RDF로 파싱이 가능한 구조의 편집가능 마크 데이터를 대상으로 했다. 전체 데이터 중 한국인 개인인명을 대상으로 하였으며, 분석 및 변환한 마크 데이터의 구조는 아래와 같은데 개인인명의 구조는 실제 필드가 3자리 수이며 4자리 필드의 마지막 숫자는 지시기호로 구성되어 있다.

	00751nzm a2200217nc 4500
①제어번호	001KAC200108735-
②최종수정일	00520111005161725-
	008011122 n aznnnabbn a a a a-
③저자명	1001 - a이광수= - h李光洙, - d1892-1950-

④다른표기(see 보라항목)	4000 -a춘원=-h春園,-d1892-1950- 4000 -a장백산인=-h長白山人,-d1892-1950- 4001 -a향산광랑=-h香山光郎,-d1892-1950- 4001 -a가야마 미쓰로,-d1892-1950- 4001 -aKayama, Mitsuro,-d1892-1950- 4001 -a카야마 미츠로,-d1892-1950- 4001 -a카야마 미쓰로,-d1892-1950- 4001 -aYi, Gwang-su,-d1892-1950- 4001 -aLee, Gwang-su,-d1892-1950- 4001 -aI, Gwang-su,-d1892-1950-
⑤정보원	670 -a세계인명대사전(고려출판사, 1999)-
⑥	678 -a소설가, 6.25때 납북

<표 9> 국립중앙도서관 저자명 전거 데이터 구조 예시

① 제어번호 001KAC200108735-

001이 필드명이며, 저자를 구분하기 위한 식별번호를 담고 있다.

② 최종수정일 00520111005161725

005는 필드명이며, 뒤의 입력 값은 레코드의 최종수정일을 나타낸다.

③ 저자명, ④ 다른표기(see 보라항목)부터 실제 유용한 값이 들어간다.

100, 400, 678등 앞부분 세자리는 각각 의미를 갖는 필드(요소)이며 뒷자리는 지시기호이다. 지시기호가 개인 인명에서 의미하는 바는 아래와 같다.

0 : 성으로 시작하지 않는 이름(forename)

1 : 성으로 시작하는 이름(surname)

2 : 가계명(family name)

4 : 묘호, 시호, 봉호, 추호, 궁호

영문자는 식별기호(속성)로 의미하는 바와 입력규칙은 아래와 같다.

▾a : 개인명(성과 이름) [반복불가]
▾b : 이름(名)에 포함되어 세계(世系)를 칭하는 숫자 [반복불가]
▾c : 이름과 관련 정보 (직위, 칭호 및 기타 명칭. 역조(歷朝), 국명(國名), 한국 및 중국의 세계(世系) [반복]
▾d : 생몰년 [반복불가]
▾e : 역할어 [반복]
▾f : 저작 연도 [반복불가]
▾g : 기타 정보 [반복불가]
▾h : 개인명의 한자표기 [반복불가]
▾j : 속성 한정어 [반복]
▾k : 형식부표목 [반복]
▾l : 저작의 언어 [반복불가]
▾n : 책·권차, 편차 [반복]
▾p : 저작의 편제 [반복]
▾q : 이름의 완전형 [반복불가]
▾t : 저작의 표제 [반복불가]
▾u : 소속 [반복불가]
▾x : 일반세목 [반복]
▾4 : 역할어 부호 [반복]
▾y : 연대세목 [반복]
▾8 : 필드 링크와 일련번호 [반복]
▾z 지리세목 [반복]

단체(기관 등)명의 구조는 아래와 같다.

0XX 0으로 시작하는 것은 제어정보로, 식별정보, 분류기호 등을 포함한다.
1XX 1로 시작하는 것은 표목이다.
2XX 2로 시작하는 것은 “보라(see)”

3XX　3으로 시작하는 것은 “도 보라(see also)”
4XX　4로 시작하는 것은 “보라”의 부출필드
5XX　5로 시작하는 것은 “도 보라”의 부출필드
6XX　6으로 시작하는 데이터처리정보, 주기 등

1XX, 4XX, 5XX필드의 뒷 두 자리는 다음을 의미한다.

X00　개인명
X10　단체명
X11　회의명
X30　통일서명
X50　주제어
X51　지리명
X80　일반세목
X81　지리세목
X82　연대세목

저자명 전거 데이터에서는 저자에 대한 정보로 저자의 이름, 저자의 다른 표기명, 출생연도, 사망연도, 저자에 대한 설명들을 획득할 수 있다. 또한 저자 정보에 대한 최종 수정일, 저자에 대한 출처 정보 등을 부가적으로 획득할 수 있다. 따라서 각각의 저자에 대해 그 저자의 정보를 나타낼 수 있는 어휘를 선정하여 데이터를 매핑하도록 하였고, 저자에 대한 부가적인 정보도 표현할 수 있도록 어휘를 선정하였다.

3.1.2 주제명 전거데이터

주제명 데이터는 변환을 위한 원천 데이터는 MS SQL의 데이터베이스 파일로 입수하였는데, 주요 용어 관계지시기호는 다음 표와 같다.

BT (broader term) : 상위개념어 BTG (broader term/generic) : 상위개념어/屬 BTI (broader term/instance) : 상위개념어/사례 BTP (broader term/partial) : 상위개념어/부분 NT (narrower term) : 하위개념어 NTG (narrower term/generic) : 하위개념어/屬 NTI (narrower term/instance) : 하위개념어/사례 NTP (narrower term/partial) : 하위개념어/부분 RT (related term) : 관련어 SN (scope note) : 범위주기 TT (top term) : 최상위개념 UF (used for 혹은 use for) : 비우선어 USE (use) : 비우선어에서 우선어로의 참조

<표 10> 국립중앙도서관 주제명표의 주요 용어 관계지시기호

이 외에 관련 분류표를 나타내는 관계지시기호(KDC, DDC)나 북한어(NK) 또는 학명(SNN)을 나타내는 관계지시기호 등도 사용하고 있으며, 이외 다양한 외국어 코드를 정의하여 사용하고 있다. 실제 데이터는 term과 thesaurus 데이터로 구성된다. term 데이터는 주제명 자체에 대한 이름, 정렬이름, 입력날짜, 확정날짜, 입력사서, 확정 사서 등 여러 가지 세부 사항에 대한 데이터이고, thesaurus는 주제명과 주제명 간의 관계를 나타내는 데이터로 BT, NT, RT 등의 관계를 표현하는 데이터이다.

필드 이름	데이터 형식	설명
id	숫자	PK
name	텍스트	주제명
sortname	텍스트	불필요
reversesortname	텍스트	불필요
baseflag	텍스트	불필요
childcount	숫자	불필요
inputdate	날짜/시간	입력일자
updatedate	날짜/시간	수정일자
confirmdate	날짜/시간	확정일자
remark	메모	비고 메모
inputuser	텍스트	입력자(사서)
updateuser	텍스트	수정자(사서)
confirmuser	텍스트	확정자(사서)
gorip	텍스트	불필요
status	텍스트	불필요
category	텍스트	불필요
flag	텍스트	불필요

<그림 41> 국립중앙도서관 주제명 전거 데이터의 Term 필드 설명

필드 이름	데이터 형식	설명
termid	숫자	기준어(term테이블의 id)
serno	숫자	
facet	텍스트	
relation	텍스트	관계(BT,NT 등 시소러스 구조어하고 KDC, DDC 분류기호포함)
reltermid	숫자	관계어 ID값 (term테이블의 id)
status	텍스트	
inputuser	텍스트	입력자
confirmuser	텍스트	확정자
inputdate	날짜/시간	입력일
confirmdate	날짜/시간	확정일

<그림 42> 국립중앙도서관 주제명 전거 데이터의 Term 필드 설명

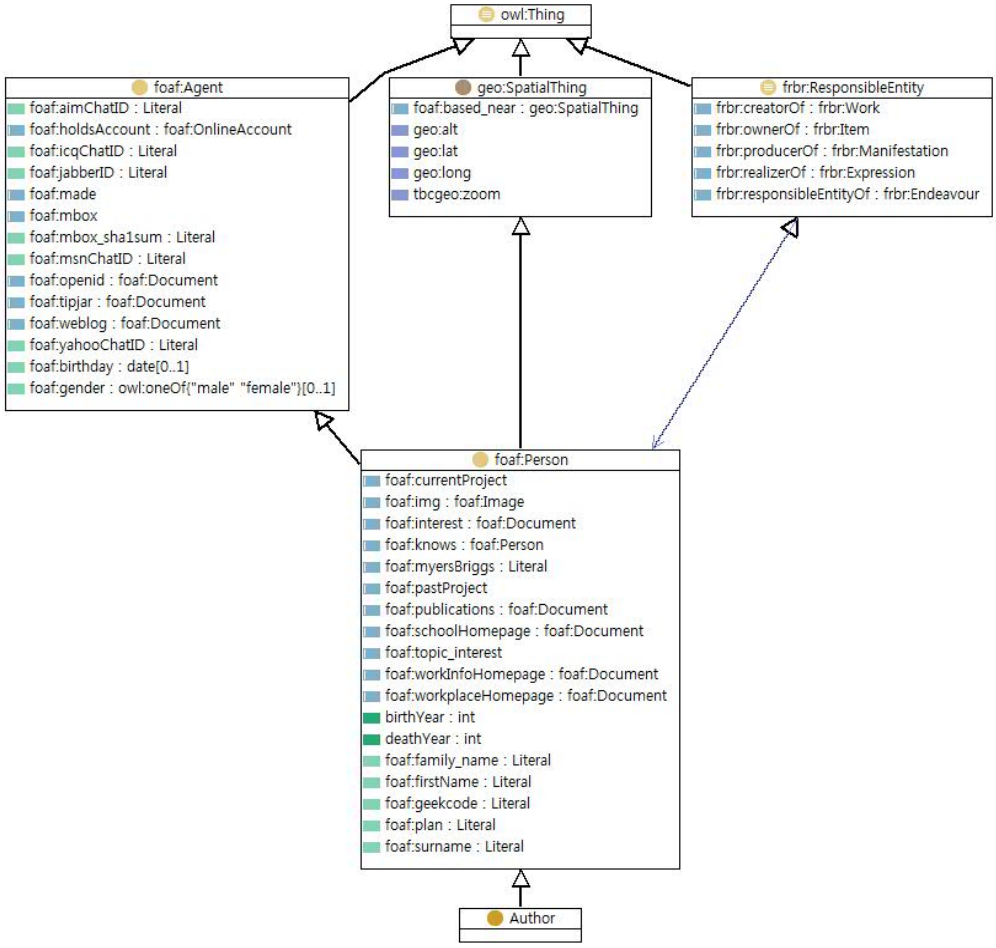
3.2 온톨로지 모델 설계

국립중앙도서관의 저자명, 주제명 데이터를 Linked Data로 생성하기 위하여 FRBR 온톨로지, SKOS, FOAF 어휘를 사용하였다. 국립중앙도서관 데이터 온톨로지는 FRBR 온톨로지를 바탕으로 필요한 항목을 확장하여 사용하였으며 저자명 데이터는 FOAF 어휘를, 주제명

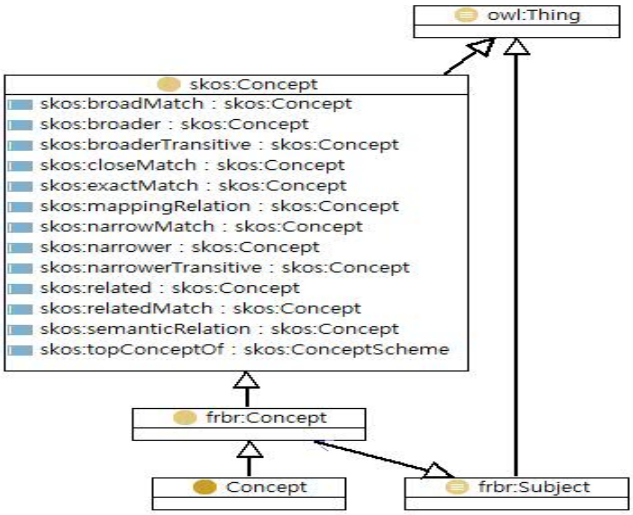
데이터는 SKOS 어휘를 활용하였다. 설계한 온톨로지 모델의 Base URI는 full uri로서 http://nl.linkeddata.kr/ontology/ 을 사용하였으며 prefix는 nl 로 명시하였다.

저자명 데이터는 국립중앙도서관에서 보유하고 있는 인명 데이터로서 FRBR의 Person 클래스의 하위인 Author 로 정의하였다. 그리고 저자명의 구조에서 표현하고 있는 최종수정일, 저자명, 다른표기명, 정보원 등의 데이터는 Dublin Core, FOAF, SKOS의 어휘를 각각 사용하여 DatatypeProperty로 구성하였다.

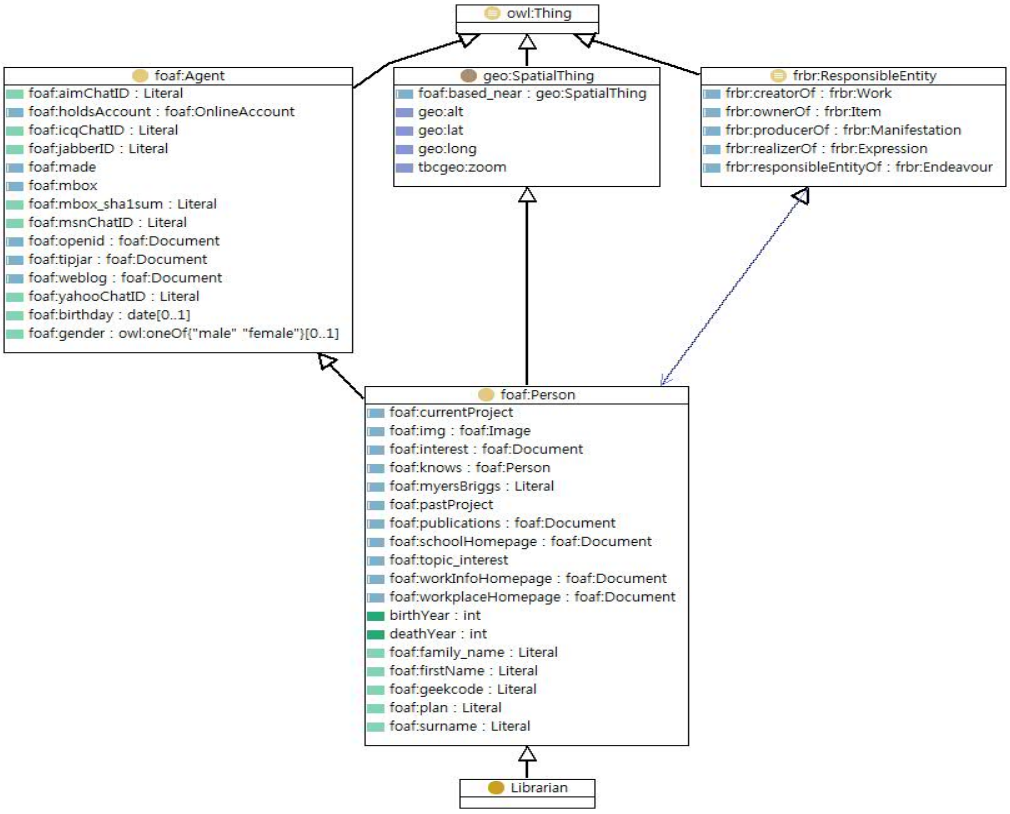
주제명 데이터는 SKOS의 Concept 클래스의 하위인 FRBR의 Concept 클래스의 속성을 받아 그 하위로 Concept 클래스를 구성하였다.



<그림 43> 국립중앙도서관 Author 클래스의 Diagram



<그림 44> 국립중앙도서관 Concept 클래스의 Diagram



<그림 45> 국립중앙도서관 Librarian 클래스의 Diagram

3.3 온톨로지 명세

국립중앙도서관의 저자명, 주제명에 대한 온톨로지는 FRRB, SKOS, FOAF를 import 하여 모델링하였으며 데이터에 따른 필요한 클래스와 프로퍼티를 정의하였다. 저자명 데이터에서는 Author 클래스를 생성하였고 각각의 저자에 대한 생몰년 데이터를 표현하기 위해 birthYear, deathYear 프로퍼티를 생성하였다. 주제명 데이터에서는 Concept 클래스와 Librarian 클래스를 생성하였다. 그리고 공통적으로 데이터에 대한 annotation으로 inputedBy, managedBy, confirmedBy, inputDate, modificationDate, confirmDate 프로퍼티를 생성하여 각각 표현하였다.

온톨로지 모델링 도구는 TopQuadrant 사의 TopBraid Composer를 사용하여 온톨로지 모델을 생성하고 관리하였다.

국립중앙도서관 데이터에 대한 온톨로지 전체 명세는 아래와 같다.

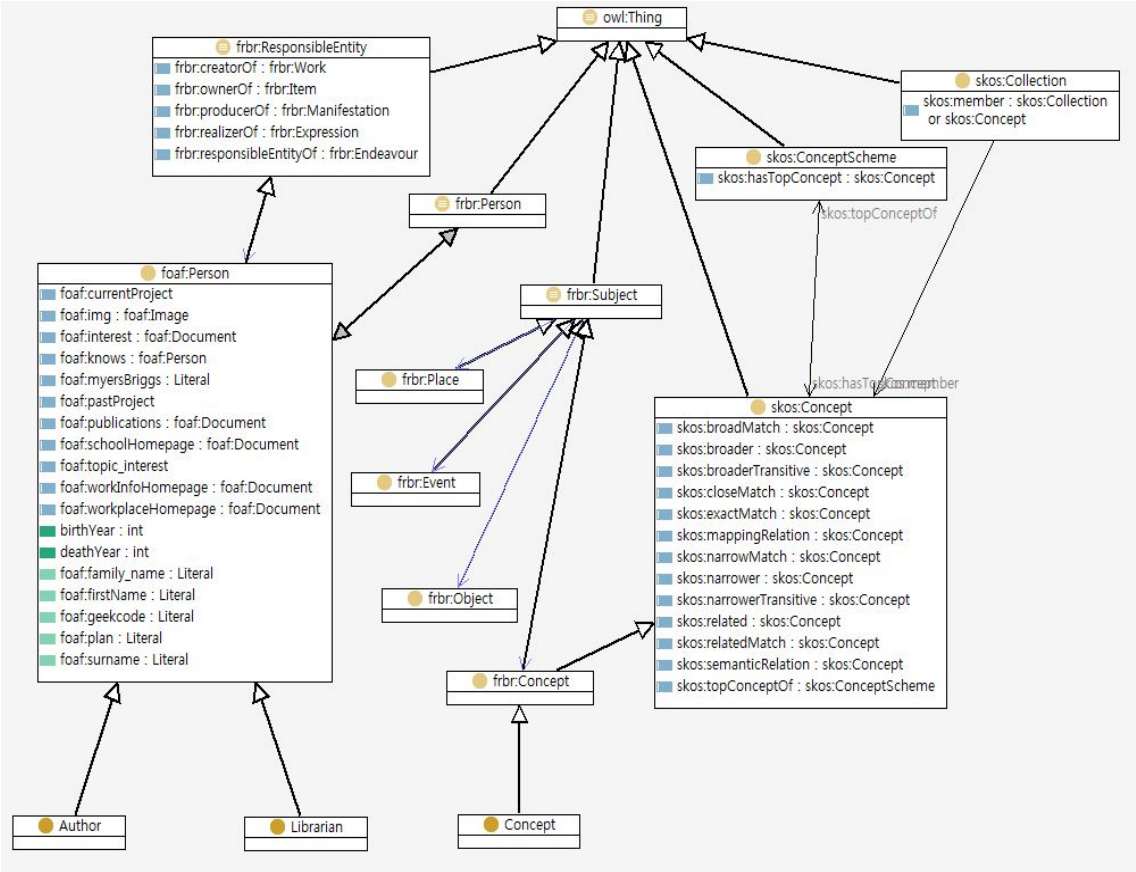
```
<?xml version="1.0"?>
<rdf:RDF
    xmlns="http://nl.linkeddata.kr/ontology/"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:frbr="http://purl.org/vocab/frbr/core#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xml:base="http://nl.linkeddata.kr/ontology/">
  <owl:Ontology rdf:about="">
    <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Created with TopBraid Composer</owl:versionInfo>
```

```
    <owl:imports rdf:resource="http://purl.org/vocab/frbr/core"/>
    <owl:imports rdf:resource="http://purl.org/dc/elements/1.1/"/>
    <owl:imports rdf:resource="http://xmlns.com/foaf/0.1/"/>
    <owl:imports rdf:resource="http://www.w3.org/2004/02/skos/core"/>
  </owl:Ontology>
  <owl:Class rdf:about="http://nl.linkeddata.kr/ontology/Author">
    <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  </owl:Class>
  <owl:Class rdf:about="http://nl.linkeddata.kr/ontology/Concept">
    <rdfs:subClassOf rdf:resource="http://purl.org/vocab/frbr/core#Concept"/>
  </owl:Class>
  <owl:Class rdf:about="http://nl.linkeddata.kr/ontology/Librarian">
    <dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >a librarian</dc:title>
    <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
    <dc:contributor rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    > an information professional trained in library</dc:contributor>
  </owl:Class>
  <owl:ObjectProperty rdf:about="http://nl.linkeddata.kr/ontology/inputedBy">
    <rdfs:subPropertyOf>
      <owl:ObjectProperty rdf:about="http://nl.linkeddata.kr/ontology/managedBy"/>
    </rdfs:subPropertyOf>
    <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="http://nl.linkeddata.kr/ontology/managedBy">
    <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="http://nl.linkeddata.kr/ontology/confirmedBy">
    <rdfs:subPropertyOf rdf:resource="http://nl.linkeddata.kr/ontology/managedBy"/>
    <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:about="http://nl.linkeddata.kr/ontology/modifiedBy">
    <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
    <rdfs:subPropertyOf rdf:resource="http://nl.linkeddata.kr/ontology/managedBy"/>
```

```
  </owl:ObjectProperty>
  <owl:DatatypeProperty rdf:about="http://nl.linkeddata.kr/ontology/birthYear">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="http://nl.linkeddata.kr/ontology/deathYear">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    <rdfs:domain rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="http://nl.linkeddata.kr/ontology/inputDate">
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/date"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#dateTime"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="http://nl.linkeddata.kr/ontology/modificationDate">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#dateTime"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/date"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="http://nl.linkeddata.kr/ontology/confirmDate">
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/date"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#dateTime"/>
  </owl:DatatypeProperty>
</rdf:RDF>

<!-- Created with TopBraid Composer -->
```

<표 11> 국립중앙도서관의 저자명, 주제명 데이터에 대한 온톨로지 명세



<그림 46> 국립중앙도서관의 저자명, 주제명 데이터에 대한 온톨로지 Diagram

3.4 데이터 변환

국립중앙도서관의 저자명, 주제명 전거데이터를 Linked Data로 발행하기 위하여 Triple 데이터로 변환을 수행하였다. 데이터 변환은 크게 저자명 트리플, 주제명 트리플 두 부분으로 나누어 진행하였으며 각각의 전거데이터의 기본 구조를 반영하여 생성하도록 매핑 규칙을 생성하였다.

저자명 전거데이터의 구조에서는 제어번호가 고유한 구분자로 사용되고 있음으로 "http://nl.linkeddata.kr/resource/"를 붙여 하나의 resource로 생성하였다. 저자명은 기본적

으로 foaf:name을 사용하여 생성하였으며 묘호, 시호 등 저자명의 다른 표기가 상당 수 존재하기 때문에 skos:prefLabel로 저자명을 생성하고 묘호, 시호 등 다른 표기는 skos:altLabel로 생성하였다. 그 외에 이 resource에 대한 최종수정일, 정보원, 기타 설명 정보는 Dublin Core의 어휘를 사용하여 표현하였다.

구조	필드 구분		사용 어휘
①제어번호	001		고유 이름으로 사용
②최종수정일	005		dc:date
③저자명	1001		foaf:name skos:prefLabel
④다른표기 (see 보라항목)	4000 4001	a	skos:altLabel
		h	
		d	nl:birthYear nl:deathYear
⑤정보원	670		dc:source
⑥기타	678		dc:description

<표 12> 저자명 용어어 대한 온톨로지 사용어휘 매핑

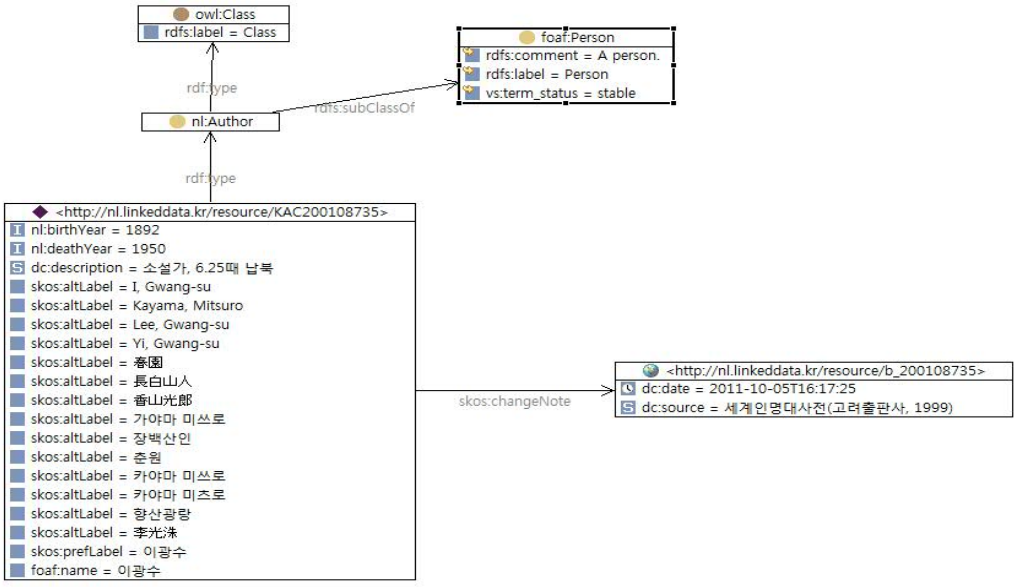
```
@prefix dc:      <http://purl.org/dc/elements/1.1/> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf:    <http://xmlns.com/foaf/0.1/> .
@prefix owl:     <http://www.w3.org/2002/07/owl#> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos:    <http://www.w3.org/2004/02/skos/core#> .
@prefix nl:      <http://nl.linkeddata.kr/ontology/> .

<http://nl.linkeddata.kr/resource/KAC200108735>
        a          nl:Author ;
        nl:birthYear "1892"^^<http://www.w3.org/2001/XMLSchema#int> ;
        nl:deathYear "1950"^^<http://www.w3.org/2001/XMLSchema#int> ;
        dc:description "소설가, 6.25때 납북"^^<http://www.w3.org/2001/XMLSchema#string>
```

```
;
        skos:altLabel "가야마 미쓰로" , "카야마 미쓰로" , "장백산인" , "춘원" , "I, Gwang-su" , "長白山人" , "Yi, Gwang-su" , "春園" , "李光洙" , "향산광랑" , "카야마 미츠로" , "Kayama, Mitsuro" , "香山光郎" , "Lee, Gwang-su" ;
        skos:changeNote <http://nl.linkeddata.kr/resource/b_200108735> ;
        skos:prefLabel "이광수" ;
        foaf:name "이광수" .

<http://nl.linkeddata.kr/resource/b_200108735>
        dc:date "2011-10-05T16:17:25"^^<http://www.w3.org/2001/XMLSchema#dateTime> ;
        dc:source "세계인명대사전(고려출판사, 1999)"^^<http://www.w3.org/2001/XMLSchema#string> .
```

<표 13> 저자명 데이터에 대한 변환된 트리플 예시



<그림 47> 저자명 데이터에 대한 변환된 트리플 Diagram

주제명 전거데이터는 시소러스 관계를 활용하여 각각의 resource, 그리고 resource 간의 관계를 표현하도록 하였다. 시소러스 관계에서 사용하는 용어는 SKOS 어휘와 유사하게 사

용되고 있음으로 이를 수용하여 사용하였으며, 그 외 필요에 의해 Property를 생성하여 데이터를 변환하였다. BT, NT, RT 용어는 skos의 broader, narrower, related 로 각각 매칭하여 사용하였으며 NTI는 협의어의 구체적인 예시이기 때문에 skos의 Collection을 이용하여 skos:member로 각각 매칭하였다. UF는 USE의 반대 개념으로서 실제로 UF는 그 용어의 다른 이름을 나타내고 USE는 그 용어의 선호이름을 나타낼 때 사용한다. 그러나 온톨로지의 매칭에서는 UF만 사용하여 skos:altLabel의 Literal 값으로 매칭하였다.

국립중앙도서관 주제명 전거데이터 용어	사용 어휘
BT	skos:broader
NT	skos:narrower
RT	skos:related
UF	skos:altLabel
NTI	skos:narrower (skos:Collection의 skos:member)
ENG	skos:altLabel{@en}
JPN	skos:altLabel{@ja}
ESP	skos:altLabel{@es}
FRA	skos:altLabel{@fr}
GER	skos:altLabel{@de}
CHI	skos:altLabel{@zn-Hant}
USE	사용하지 않음(UF의 역관계)
KEN	사용하지 않음(USE의 역관계)

<표 14> 주제명 용어어 대한 온톨로지 사용어휘 매핑

SKOS에서는 특정 Concept이 어떤 ConceptScheme에서 사용되는지 나타내기 위하여 skos:inScheme 어휘를 사용하여 표현하고 있다. 국립중앙도서관의 데이터를 Linked Data로 발행하기 위해서는 이러한 개념을 수용하여 국립중앙도서관의 분류체계임을 명시하기 위해

```
nls:KSHxxx skos:inScheme nls:subjectTermScheme .
nls:subjectTermScheme a skos:ConceptScheme .
```

와 같이 표현하였다.

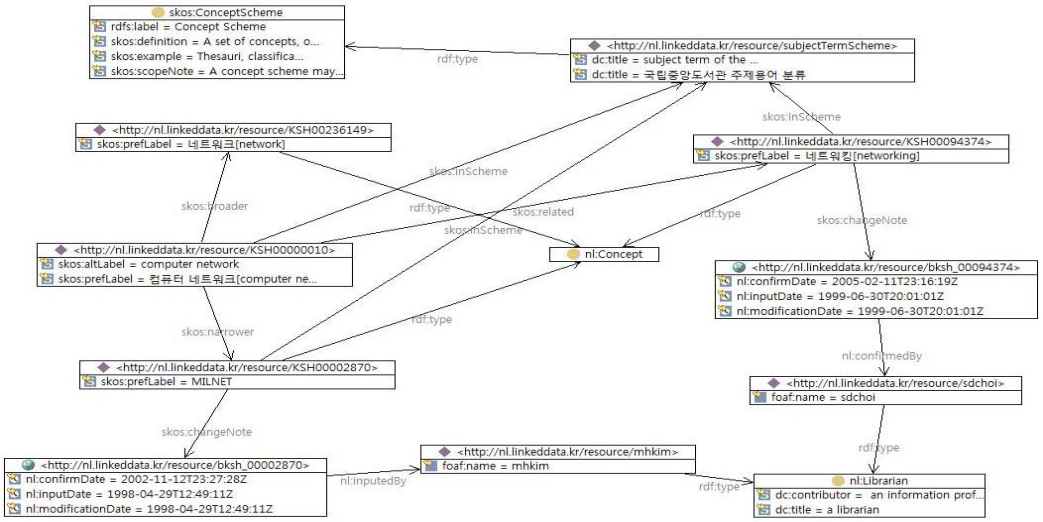
```
@prefix dc:      <http://purl.org/dc/elements/1.1/> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf:    <http://xmlns.com/foaf/0.1/> .
@prefix owl:     <http://www.w3.org/2002/07/owl#> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos:    <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .
@prefix nl:      <http://nl.linkeddata.kr/ontology/> .
@prefix nls:     <http://nl.linkeddata.kr/resource/> .

nls:KSH00000010
        a          nl:Concept ;
        skos:altLabel "computer network"@en ;
        skos:broader nls:KSH00236149 ;
        skos:changeNote nls:bksh_00000010 ;
        skos:inScheme nls:subjectTermScheme ;
        skos:narrower nls:KSH00001410 , nls:KSH00000017 , nls:KSH00003579 ,
nls:KSH00034021 , nls:KSH00047151 , nls:KSH00480163 , nls:KSH00003596 ,
nls:KSH00001524 , nls:KSH00000018 , nls:KSH00001520 , nls:KSH00047020 ,
nls:KSH00002089 , nls:KSH00001719 , nls:KSH00002870 , nls:KSH00030422 ,
nls:KSH00001921 , nls:KSH00002548 , nls:KSH00002347 , nls:KSH00003546 ,
nls:KSH00263199 , nls:KSH00000007 , nls:KSH00001956 , nls:KSH00003109 ,
nls:KSH00000095 , nls:KSH00002989 , nls:KSH00000019 ;
        skos:prefLabel "컴퓨터 네트워크[computer network]"@ko ;
        skos:related nls:KSH00094374 , nls:KSH00003218 , nls:KSH00002967 ,
nls:KSH00034388 , nls:KSH00023645 , nls:KSH00001235 .

nls:subjectTermScheme
        a          skos:ConceptScheme ;
        dc:title "subject term of the Nation Library of Korea"@en , "국립중앙도서관
주제용어 분류"@ko .
```

```
nls:bksh_00000010
        nl:confirmDate "2003-11-17T20:24:10Z"^^xsd:dateTime ;
        nl:confirmedBy nls:sdchoi ;
        nl:inputDate "1998-04-29T12:45:28Z"^^xsd:dateTime ;
        nl:inputedBy nls:mhkim ;
        nl:modificationDate "1999-06-30T19:59:21Z"^^xsd:dateTime ;
        nl:modifiedBy nls:sdchoi .
```

<표 15> 주제명 데이터에 대한 변환된 트리플 예시 1



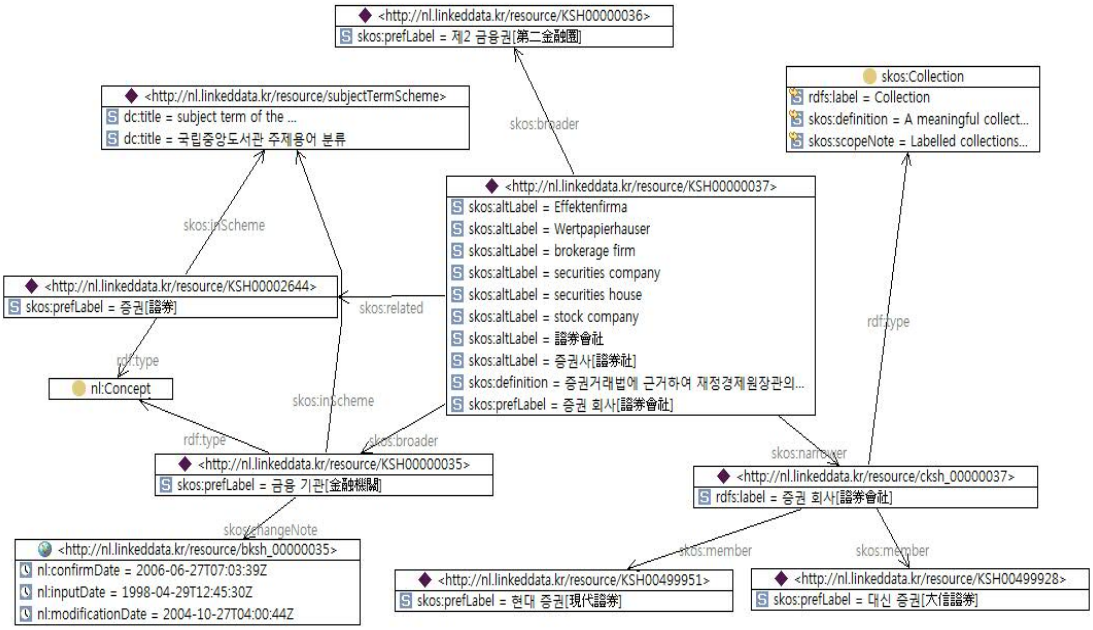
<그림 48> 주제명 데이터에 대한 변환된 트리플 Diagram 1

NTI 용어로 연결된 데이터는 위에서도 언급하였다시피 어떠한 개념에 대한 구체적인 사례를 나타내는 용도로 사용하고 있다. 따라서 SKOS에서는 이를 skos:narrower 프로퍼티를 이용하여 skos:Collection으로 연결하고, 다시 하나의 Collection에서 skos:member 프로퍼티를 이용하여 어떤 개념에 대한 구체적인 사례 개념들을 연결하도록 하고 있다. 따라서 이를 반영하여 국립중앙도서관 주제명 전거 데이터에서 나타나는 NTI 관계를 아래와 같이 표현하였다

nls:KSH00000037
a nl:Concept ;
skos:altLabel "securities house"@en , "stock company"@en , "증권사[證券社]"@ko , "Effektenfirma"@de , "證券會社"@ja , "brokerage firm"@en , "securities company"@en , "Wertpapierhauser"@de ;
skos:broader nls:KSH00089298 , nls:KSH00000035 , nls:KSH00000036 ;
skos:changeNote nls:bksh_00000037 ;
skos:definition "증권거래법에 근거하여 재정경제원장관의 허가를 받고 증권업을 영위하는 주식회사."@ko ;
skos:inScheme nls:subjectTermScheme ;
skos:narrower nls:KSH00028678 , nls:KSH00028679 , nls:KSH00501097 , nls:KSH00000032 , nls:KSH00028680 , nls:KSH00031218 , nls:KSH00001139 , nls:cksh_00000037 , nls:KSH00031203 ;
skos:prefLabel "증권 회사[證券會社]"@ko ;
skos:related nls:KSH00084024 , nls:KSH00575558 , nls:KSH00270564 , nls:KSH00012152 , nls:KSH00593538 , nls:KSH00084030 , nls:KSH00599048 , nls:KSH00002644 .
nls:cksh_00000037
a skos:Collection ;
rdfs:label "증권 회사[證券會社]"^^xsd:string ;
skos:member nls:KSH00500439 , nls:KSH00500724 , nls:KSH00499930 , nls:KSH00500732 , nls:KSH00499928 , nls:KSH00499951 .
nls:bksh_00000037
nl:confirmDate "2006-04-03T01:18:45Z"^^xsd:dateTime ;
nl:confirmedBy nls:sdchoi ;
nl:inputDate "1998-04-29T12:45:30Z"^^xsd:dateTime ;
nl:inputedBy nls:mhkim ;
nl:modificationDate "2002-11-24T00:00:00Z"^^xsd:dateTime ;
nl:modifiedBy nls:sdchoi .
nls:KSH00500439
a nl:Concept ;
skos:changeNote nls:bksh_00500439 ;
skos:inScheme nls:subjectTermScheme ;
skos:prefLabel "LG 증권[--證券]"@ko .
nls:KSH00084024

```
        a          nl:Concept ;
        skos:changeNote  nls:bksh_00084024 ;
        skos:inScheme  nls:subjectTermScheme ;
        skos:prefLabel  "증권 거래소 거래원[證券去來所去來員]"@ko .
nls:KSH00028678
        a          nl:Concept ;
        skos:changeNote  nls:bksh_00028678 ;
        skos:inScheme  nls:subjectTermScheme ;
        skos:prefLabel  "대형 증권사[大型證券社]"@ko .
```

<표 16> 주제명 데이터에 대한 변환된 트리플 예시 2



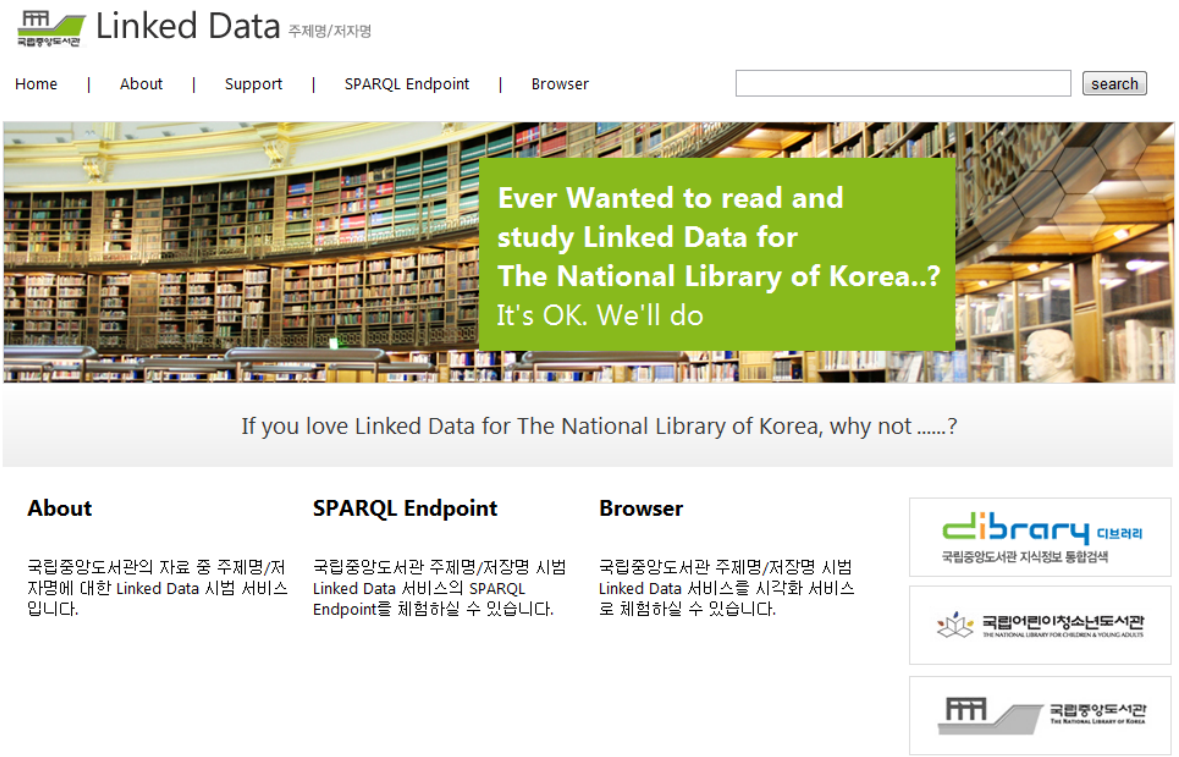
<그림 49> 주제명 데이터에 대한 변환된 트리플 Diagram 2

3.5 Linked Data 발행 서비스

Linked Data로의 발행을 위해 생성한 국립중앙도서관 저자명, 주제명 트리플 데이터는 전체 3,769,571 건으로 이를 활용하기 위해 Web Page, SPARQL Endpoint, Browser 등을 지원하는 웹서비스를 구축하여 운영하고 있다. Linked Data for The National Library of Korea 웹서비스는 http://nl.linkeddata.kr/home/에서 제공하고 있다.

이 서비스는 SPARQL Endpoint를 제공하여 사용자로 하여금 국립중앙도서관에서 발행하는 트리플 데이터를 직접 핸들링하도록 지원하며 Search API를 제공함으로 트리플 데이터에 대한 검색도 지원하고 있다. 또한 FLEX 기반의 Visualizing Browser를 제공하여 사용자로 하여금 가시적으로 탐색을 가능하도록 지원하고 있다.

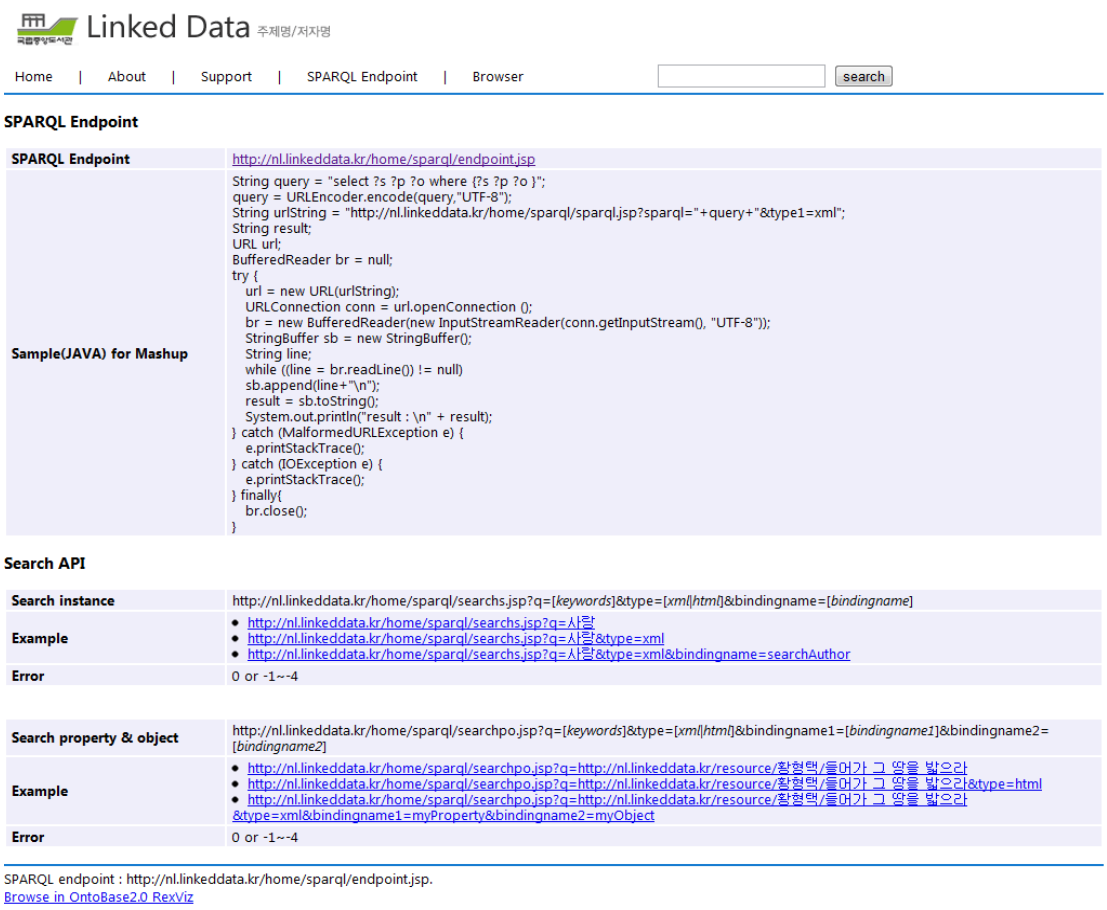
그림 48은 Linked Data 발행 서비스의 메인 화면이다.



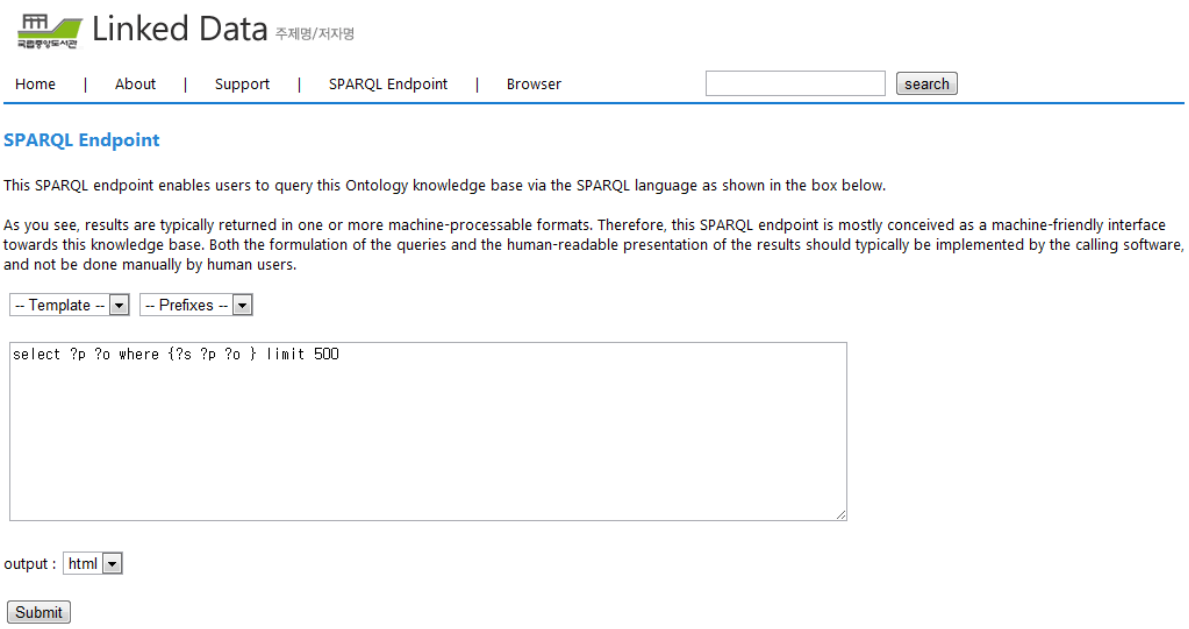
<그림 50> 국립중앙도서관 Linked Data 발행 서비스 메인 화면

Linked Data 서비스가 제공하는 기능은 다음과 같다.

- URI에 대한 Text Browsing 서비스
- SPARQL Endpoint 서비스
- RDF, HTML Content Negotiation 서비스
- 키워드 검색 서비스
- Facet 기반의 Graphical Visualization 서비스
- REST based API 서비스



<그림 51> Linked Data 발행 서비스의 Support 화면



<그림 52> Linked Data 발행 서비스의 SPARQL Endpoint 화면

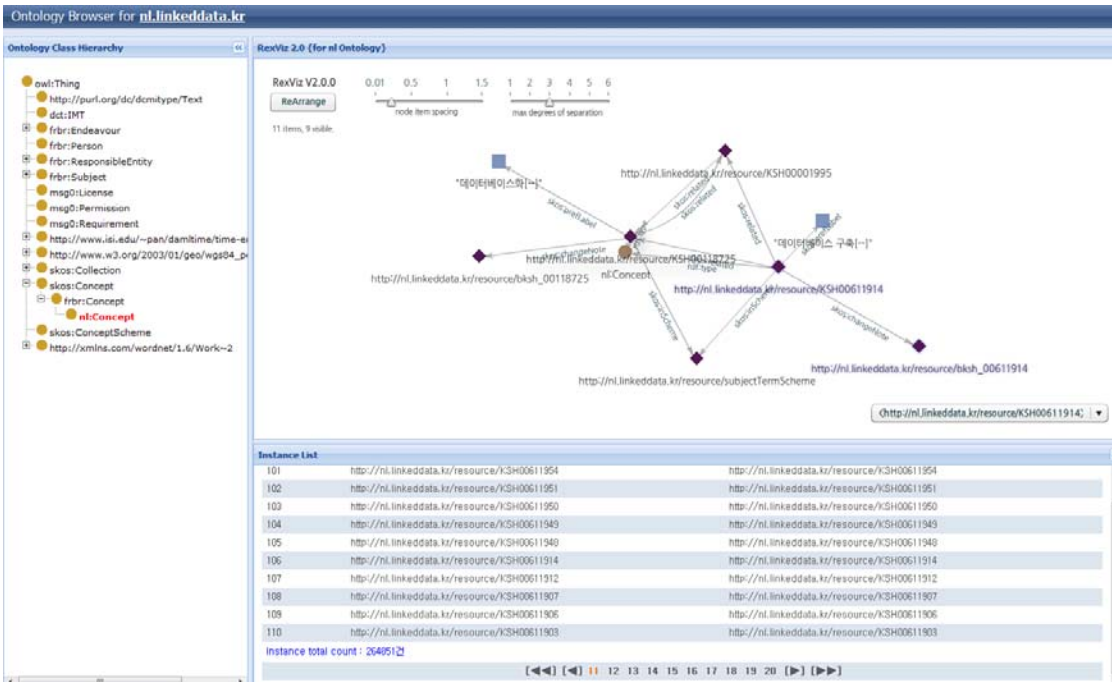
또한, 위의 결과를 RDF, JSON, HTML, N-Triples, XML 형식으로 다운로드 받을 수 있는 기능을 제공한다.

아래는 SPARQL Endpoint에 대한 것으로 결과는 XML, RDF, HTML, JSON 형식으로 반환받을 수 있다. SPARQL 질의에 대한 HTML 형식의 결과 예시는 다음과 같다.

nl.linkeddata.kr/home/sparql/index.jsp?query=select+%3Fs+%3Fp+%3Fo+where+%7B%3Fs+%3Fp+%3Fo+%7D+limit+500&type1

s	p	o
file:///var/www/vocab.org/www/htdocs/frbr/	http://web.resource.org/cc/license	http://creativecommons.org/licenses/by/2.0/uk/
file:///var/www/vocab.org/www/htdocs/frbr/	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://web.resource.org/cc/Work
http://bruce.darcus.name/about#me	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
http://bruce.darcus.name/about#me	http://xmlns.com/foaf/0.1/name	Bruce D'Arcus
http://creativecommons.org/licenses/by/2.0/uk/	http://web.resource.org/cc/permits	http://web.resource.org/cc/DerivativeWorks
http://creativecommons.org/licenses/by/2.0/uk/	http://web.resource.org/cc/permits	http://web.resource.org/cc/Distribution
http://creativecommons.org/licenses/by/2.0/uk/	http://web.resource.org/cc/permits	http://web.resource.org/cc/Reproduction
http://creativecommons.org/licenses/by/2.0/uk/	http://web.resource.org/cc/requires	http://web.resource.org/cc/Attribution
http://creativecommons.org/licenses/by/2.0/uk/	http://web.resource.org/cc/requires	http://web.resource.org/cc/Notice
http://creativecommons.org/licenses/by/2.0/uk/	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://web.resource.org/cc/License
http://iandavis.com/id/me	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Person
http://iandavis.com/id/me	http://xmlns.com/foaf/0.1/name	Ian Davis
http://nl.linkeddata.kr/ontology/	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Ontology
http://nl.linkeddata.kr/ontology/	http://www.w3.org/2002/07/owl#imports	http://purl.org/dc/elements/1.1/
http://nl.linkeddata.kr/ontology/	http://www.w3.org/2002/07/owl#imports	http://purl.org/vocab/frbr/core
http://nl.linkeddata.kr/ontology/	http://www.w3.org/2002/07/owl#imports	http://www.w3.org/2004/02/skos/core
http://nl.linkeddata.kr/ontology/	http://www.w3.org/2002/07/owl#imports	http://xmlns.com/foaf/0.1/
http://nl.linkeddata.kr/ontology/	http://www.w3.org/2002/07/owl#versionInfo	Created with TopBraid Composer
http://nl.linkeddata.kr/ontology/Author	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Class
http://nl.linkeddata.kr/ontology/Author	http://www.w3.org/2000/01/rdf-schema#subClassOf	http://xmlns.com/foaf/0.1/Person
http://nl.linkeddata.kr/ontology/Concept	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#Class
http://nl.linkeddata.kr/ontology/Concept	http://www.w3.org/2000/01/rdf-schema#subClassOf	http://purl.org/vocab/frbr/core#Concept
http://nl.linkeddata.kr/ontology/Librarian	http://purl.org/dc/elements/1.1/contributor	an information professional trained in library
http://nl.linkeddata.kr/ontology/Librarian	http://purl.org/dc/elements/1.1/title	a librarian

<그림 53> SPARQL Endpoint에서 질의한 결과 화면



<그림 54> 국립중앙도서관 온톨로지 데이터에 대한 Visualization Browser 화면

4. 결론

4. 결론

MARC는 시스템간의 레코드 교환을 위한 표준적인 목록 레코드 형식으로서 도서관 분야에서 유용하게 사용되고 있다. 반면에 도서관 분야에서만 유통되는 표준이기 때문에 도서관 관련 시스템 사이의 정보 유통에만 활용될 뿐, 웹을 중심으로 하는 정보 유통 환경에서 도서관 이외의 분야와 정보를 매쉬업 한다거나 유통하는 과정에서는 활용이 어렵다.

많은 도서관들이 이미 SKOS 포맷을 활용해서 기존의 시소러스나 어휘 등을 웹에서 Linked Data형태로 표현하고 있다. 따라서 기존의 지식어휘체계를, 의미적으로 풍부하게 시맨틱 웹에 표현하게하고, 재사용성과 상호운용성을 증진할 수 있도록 하고 있다.

Linked Data로 표현하고 발행을 하게 되면 MARC를 사용하여 데이터를 구조화, 관리하는 것 보다 많은 이점과 기대효과를 가져올 수 있다. Web이라는 형태의 특성을 잘 이용하고 그 생태에 참여하면 도서관에서만 사용하는 제한적인 데이터보다 더 많은 데이터와의 연결을 고려할 수 있으며 도서관 이외의 다른 분야와 융복합을 통해 데이터의 가치가 더 상승하는 계기가 된다. Linked Data를 통해 데이터를 웹으로 발행했을 때의 기대효과는 아래와 같다.

첫째, 국가적 정보 융합의 기틀을 만들 수 있다. Linked Data는 원천데이터를 표준 기술을 활용하여 자유롭게 활용이 가능하도록 하는 웹 생태계에 새로운 데이터 계층을 만드는 것이다. 이것은 새로운 정보 유통 패러다임이라고 볼 수 있다. 이러한 변화는 도서관과 같은 기존의 정보유통관리, 보존 기관간의 데이터 융합이 가능하도록 하며, 연구자 등 개인까지도 자유롭게 원천데이터를 활용하고 새로운 정보를 형성해 나갈 수 있는 기틀이 된다.

둘째, 데이터의 가치를 상승시킬 수 있다. 우리나라의 도서관 데이터를 세계 각국 LOD(Linking Open Data)의 데이터 셋과 함께 연계함으로써 전 세계 누구나 자유롭게 활

용 가능하고 이로 인한 외부의 접근과 데이터 결합이 증가함에 따라 정보의 가치 상승을 도모할 수 있다. 또한 도서관내의 정보가 더 다양한 관련 정보, 또는 기타 다양한 정보(지리정보, 인물 정보 등)와 연계되고, 유통되기 때문에 공유 자원으로써 가치의 상승을 가져올 수 있다.

셋째, 관련된 정보의 활용도를 제고할 수 있다. 도서관 데이터가 Linked Data 형태로 발행하면 세계 각국의 LOD와 자유로운 양방향 정보 유통이 가능하므로 다양한 데이터들이 서로 융합될 수 있으며, 데이터의 재사용성이 증가한다. 웹을 통해 누구나 접근할 수 있으며 타 분야의 정보 융합이 용이하므로 국가 차원에서의 지식 활용도 제고도 가능하며, Linked Data 유통 인프라를 통해 향후, 타분야 데이터로 확장, 적용이 용이하므로 관련 정보 융합을 통한 서비스 발굴 및 연구 발전에 기여할 수 있다.

넷째, 창의적 연구 개발이 가능하도록 한다. Linked Data 질의 언어인 Sparql을 통해 LOD에 포함된 모든 데이터셋에 대해 마치 하나의 글로벌 데이터베이스처럼 질의가 가능하므로 창의적 융합 연구 발전에 기여할 수 있다.

다섯째, 소프트웨어 산업 발전에 기여할 수 있다. 자유롭게 활용가능한 데이터 셋들의 존재는 이를 기반으로 새로운 서비스 모델, 소프트웨어 개발을 촉진시킬 수 있을 뿐만 아니라 다양한 정보환경 및 및 채널(웹, 스마트폰, IPTV 등)에서의 다양한 서비스(정보제공, 분석, 통계, 추천, SNS 등) 발굴에 기여할 수 있다.

정보기술환경이 네트워크 중심으로 이동하고, 컨텐츠의 유형이 매우 다양해지고 있다. 그런데 도서관의 데이터는 아직도 도서관내에서만 공유가 가능한 형태로, 도서관을 벗어난 외부기관과 이용자들의 이용이 불가능한 상황이다. 이는 도서관 데이터가 주로 텍스트로 표현되어 있고, 연결 가능한 URI로 표현되지 않았기 때문이다. 이런 면에서 MARC는 표현 모델 자체의 한계를 가지고 있다. 이것을 극복하기 위해 도서관 데이터에 연결성(connection or relation)과 식별성(identification)을 부여하도록 해야 한다.

전통적으로 도서관에서는 통제어휘를 사용해서 관계가 있으리라고 생각되는 것을 묶어주

고 정보검색에서의 어휘의 모호성 문제를 해결해 왔다. 주제명표목을 통한 전거활동이 그 일부이다. 현재 정보시스템의 문제는 이용자가 찾고 있는 개념과 정보검색시스템에서 의도했던 개념 사이에 큰 간격이 존재한다는 것이다. 이 문제를 해결하기 위해서는 다양한 주제분야에서 시소러스를 만들어 사용했으나 이용자의 요구에 즉각적인 반응을 할 수 있는 시스템이 없었다. 즉, 두 개념 사이의 간격을 좁혀줄 수 있는 보다 이용자 중심의 해석을 제공할 수 있는 무언가가 필요하다.

SKOS를 활용한 시소러스가 그 대안이 될 수 있는데, 이것을 Linked Data로 발행하면 상호운용성 확보에 유리하다. 웹에 의미를 부여해서 개념과 개념 사이의 내포 및 외연을 이용자가 적절하게 조절할 수 있게 해주면 정보 탐색 및 발견의 효율성을 향상하게 도와줄 것이다. 정보의 '정확한' 의미를 정의하기는 어렵지만, 추상화 수준을 높여 '느슨한' 구조만이라도 담아내면서, 온톨로지에서의 복잡한 모델링 없이도 시맨틱하게 보여주자는 시도이다. SKOS 바탕의 시맨틱 시소러스를 구체적으로 만들고 이것을 웹에서 Linked Data로 발행하여 의미 있는 링크를 통해 연계하면 마치 하나의 글로벌 지식베이스와 같이 통합적으로 질의, 검색함으로써 정보의 활용성을 높이고 이용자에게는 통합적인 데이터 뷰를 제공할 수 있고, 정보를 쉽게 접근할 수 있는 체계를 확립할 수 있다.

또한 정보 접근 체계와 더불어 도서관 데이터를 Linked Data로 발행할 때 기대효과는 다음과 같다.

□ LOD 클라우드에 고품질의 데이터 제공

- o 전거 데이터, 분류체계, 시소러스, 메타데이터
  - 개별 목록에서 이를 이용하여 데이터의 품질 관리
  - 웹 데이터의 품질 관리
- o 데이터의 상호운용성 지원
  - 시맨틱 웹 기술 기반의 의미적 상호운용성

□ 도서관 데이터 공개와 연계

- o 구조화된 데이터의 글로벌 정보 공간
- o 다른 컨텍스트에서 사용하는 도서관 데이터
- o 혁신적인 사용자 서비스

□ Linked Data 기술을 이용한 데이터 노출

- o 도서관 간의 상호 협력 증진
  - 상호 협력을 위한 기술적 부담이 적음
  - 웹 프로토콜을 이용한 데이터 교환이 용이
  - 애플리케이션 개발 비용이 낮음
- o 외부 커뮤니티와의 협력 증진
  - 도서관 종속적인 프로토콜과 포맷을 모르더라도 손쉽게 도서관 데이터를 사용할 수 있음
  - 사회의 지식 자원인 도서관 데이터를 공유 자원화

이러한 효과들은 웹 데이터의 품질 향상, 도서관 데이터의 활용 증가, 도서관의 협력 증진 및 지식자원 공유 확대를 가져온다. 기존에는 아무 의미없이 흩어져 있던 정보들을 의미있게 묶어서 보여주고 뜻을 명확하게 구별할 수 있게 해주자는 시도가 도서관 뿐만 아니라 모든 계층에서 활발히 진행되고 있다. 국내도서관은 연계를 위한 충분한 네트워크를 가지고 있다. 이제 웹을 통한 온톨로지 기반의 데이터 공유와 데이터의 연결은 필수적이다. Linked Data를 통해서 도서관을 웹에 연결하여 보다 가치있고 유용한 도서관의 데이터로 만들어 나가야한다.

오늘날 정보기술의 발전에 따라 기존의 도서관서비스는 많은 변화를 거듭해 왔다. 도서관에 컴퓨터기술이 도입되면서 전통적인 도서관의 서비스는 자동화된 서비스체계를 갖추게 되었고, 이러한 서비스의 자동화체계에 디지털기술과 네트워크기술이 응용되어 새로운 개념의 '디지털도서관'으로 발전하게 되었다. 디지털도서관이란 분산 네트워크로 연결된 이질적인 정보저장소의 정보들을 이용자중심 인터페이스를 통해 시·공간에 구애됨이 없이 탐색·접근·이용할 수 있도록 설계된 멀티(하이퍼)미디어 정보시스템으로서 디지털정보를 수집·조직·축적·탐색·검색하여 원격지로 배포하고 저작권을 관리할 수 있는 시스템으로 정의할 수 있다. 디지털도서관의 대전제는 정보의 '분산과 공유'이므로 시스템간의 상호운용성이 높은 효율적인 디지털도서관 서비스를 위해서는 관련된 각종 기술과 데이터형식, 서비스방식 등에 대한 표준화가 선행되어야 한다.

디지털 도서관은 단일 실체가 아닌 인터넷을 포함하여 전 세계에 분산되어 있는 수많은 디지털 형태의 정보자원을 검색 대상으로 하기 때문에 복수의 실체들로 이루어져 있다. 디지털 도서관은 많은 정보자원들을 연결시킬 수 있는 기술을 필요로 한다. 수많은 디지털 도서관 및 정보서비스 사이의 연결은 이용자들에게 편리하게 제공돼야 한다. 즉, 이용하기 쉽고 편리한 이용자 인터페이스를 통해 많은 디지털 도서관에서 제공하는 다양한 정보자원들에 끊김 없이 접속하여 원하는 정보를 활용할 수 있어야 한다.

디지털 도서관과 정보서비스에 대한 범세계적인 접근이 중요한 목표가 된다. 즉, 기존의 도서관 자동화 시스템은 특정 도서관이 소장하고 있는 자료에 대한 검색 서비스를 목표로 하고 있는데 비하여 디지털 도서관 환경 하에서는 전 세계에 분산되어 있는 다양한 형태의 디지털 객체들을 이용자의 검색 단말기에서 손쉽게 검색 할 수 있다. 이러한 가능성으로의 문을 더 쉽고 효율적으로 열 수 있는 기술이 Linked Data이며 이미 도서관 데이터는 Linked Data로 들어와서 진행 중이며 앞으로 더 많은 Linking을 통해 확장해 나갈 것이다.