

개방형 공공데이터의 공간정보 활용을 위한 Geo-Data Appliance 구현방안

A Study on Geo-Data Appliance for Using Geospatial Information of Public Open Data

연성현* · 김현덕** · 이인수***

Yeon, Sung-Hyun · Kim, Hyeon-Deok · Lee, In-Su

초 록

최근 국가에서 공공데이터를 적극 개방하여 민간의 활용을 장려하고 있는데, 이는 정부의 국정운영 패러다임인 '정부 3.0'에 근거한다. 이러한 흐름에 따라 공간정보를 활용하여 비즈니스 모델을 수립하고, 사업화하기 위한 데이터 플랫폼이 요구되고 있다. 그러나 아직까지 텍스트 기반의 공공데이터를 공간정보 분야에서 활용하기 위한 체계 수립이 부족한 실정이다. 본 연구는 공간참조유형의 공공데이터를 효율적으로 제공하고 활용할 수 있는 방안을 모색하기 위하여 텍스트 기반의 공공데이터를 활용한 공간정보데이터 제공 시스템을 구축하였다. 이를 통해 위치정보를 가진 공공데이터에 대한 사용자의 접근성과 편의성을 향상시킬 수 있었다. 또한 생산기관이 상이한 공공데이터를 연계하여 공간데이터로 가공한 후 완제품 형식으로 제공하는 어플라이언스(Appliance) 시스템의 구성요건을 제시하였다.

주요어 : 공공데이터, 데이터포털, 공간정보데이터 어플라이언스

ABSTRACT

Recently, the South Korean government actively opens the public data to encourage people to use it in private sector. It is based on 'Government 3.0' that is the paradigm for government operation. According to this trend, a data platform is required for establishing and commercializing

* 한국국토정보공사 공간정보연구원 국토정보연구실 연구원(전자우편 : yeon@lx.or.kr), 주저자.

** 한국국토정보공사 공간정보연구원 국토정보연구실 연구원(전자우편 : hdkim@lx.or.kr), 교신저자.

*** 한국국토정보공사 공간정보연구원 국토정보연구실 책임연구원 (전자우편 : les05@lx.or.kr), 공동저자.

business models that utilizing geospatial data. However, it is currently insufficient to establish the geospatial data system using the text-based public open data. This study constructs a geospatial data supply system using the public data for the purpose of providing and using the public data of spatial reference type efficiently. It improves the accessibility of user and the usability to the public data having location information. Besides, this study suggests that the components of the appliance system that connects the public data from different public institutions for different purposes with producing the geospatial data in the type of a finished product.

KEYWORDS : Public Open Data, Data Portal, Geospatial data Appliance

1. 서론

공공데이터의 민간 개방 및 활용이 세계적으로 크게 증가하고 있다. 공공데이터는 막대한 자원이 투입된 국가자산으로, 경제·지리·통계·교통·기상·식품·문화 등 국민 생활 전반에 걸쳐 오랜 기간 방대한 자료가 축적되어 있다. 따라서 공공데이터는 이용방법에 따라 생산될 수 있는 부가가치가 무궁무진하며 생활전반에 걸쳐 과학적 의사결정, 비즈니스 정책 수립 등 다양한 분야에 의사결정을 위한 기초자료로 활용될 수 있다(박원준, 2014). 국내에서도 경제성장을 촉진시킬 추진 동력으로 공공데이터 활용을 장려하고 있지만, 정보를 보유한 공공기관들 중 일부는 데이터의 품질이나 민감성으로 인해 다소 부담을 느끼며 개방을 꺼려하고 있다. 그러나 정부는 민간에서 공공데이터 활용이 활성화 될 수 있도록 수요자 중심의 데이터 개방을 지속적으로 추진하고 있다(송인국, 2015). 이에 따라 한국정보화진흥원이나 여러 지자체를 중심으로 ‘공공데이터포털’¹⁾이나 ‘서울 열

린데이터 광장’²⁾ 등과 같이 민간에서 이용할 수 있는 공공정보 개방 채널을 통해 공공데이터를 제공하고 있다.

이러한 공공데이터 중 많은 정보가 공간정보와 연관되어 있지만 대부분 텍스트 기반의 정보이며, 데이터 제공방식의 제약으로 인해 활용하는데 한계가 있다. 공공데이터를 지리적 시각화나 맵 쉬업 서비스로 제공하기 위해서는 필요한 주제의 공공데이터를 제공 방식별로 분석하여 개별적으로 수집하고, GIS 소프트웨어나 포털사이트의 지도API를 함께 이용하는 절차가 일반적이다. 그러나 비용 문제는 차치하더라도 앞서 기술한 바와 같이 사용 절차가 복잡하고 전문성이 요구되어 대중들이 쉽게 활용하기 어렵다. 또한 개방 채널들의 데이터 제공 방법으로 활용되는 오픈API(Open Application Programming Interface)나 파일다운로드(File Download) 방식은 단순 참조로 이용되기에 편하나, 다차원 관점에서 업무분석 용도로 활용하기에는 한계가 있다(강영욱 등, 2014).

1) 공공데이터포털, <http://www.data.go.kr>.

2) 서울 열린데이터 광장, <http://data.seoul.go.kr>.

국의 공공데이터 제공 플랫폼은 대부분 LOD(Linked Open Data) 기반의 데이터를 제공하도록 구조화하여 데이터를 웹 환경에서 확장성 있게 연계하였다(이현정 등, 2014). 링크드 데이터(Linked Data)는 2006년 팀 버너스 리(Tim-Berners Lee)가 제안한 개념으로, 시맨틱 웹(Semantic Web)이 표방하는 데이터 웹(Data Web)을 구체적으로 표현하는 방법이다. 이것은 웹에 존재하는 다양한 정보를 의미적으로 연결하기 위해 데이터를 구조화하고 웹 환경에서 확장성 있게 연계하기 위한 기술이다. 따라서 이만재(2011), 윤소영(2013)은 개방된 공공데이터를 활용하기 위한 도구로 LOD를 제안하였다. 하지만 국내 개방된 공공데이터 중 극히 일부 분야의 데이터만이 이를 표방하고 있다(유현경 등, 2015). 더불어 백인수(2013)는 LOD를 지원하기 위해 오픈소스 데이터포털 소프트웨어인 CKAN³⁾을 도입할 필요성을 제기하였으나 공공데이터포털의 경우, 아직까지는 독자적으로 개발한 CMS⁴⁾를 유지하고 있다. 결국 현재까지 공개된 국내 공공데이터는 대부분 오픈API 혹은 파일다운로드 방식으로 제공된다고 볼 수 있다. 이러한 두 가지 유형의 연동 방식으로 수집할 수 있는 공공데이터는 단일 정보를 가져와 해석하기는 쉽지만 복수 항목을 수집하고 정보를 재가공하여 다른 서비스를 개발하기에는 용이하지 않다(김청림 등, 2010). 따라서 다양한 데이터의 수집 체계를 표준화할 필요가 있다. 이를 공간정보와 융합하여 데이터베이스를 제공함으로써 공간정보 수요자의 창조적 생산 활동을 지원할 수 있을 것이다. 특히 여러 채널에서 분산되고 단절

되어 있는 정보를 추출, 변환, 통합하는 과정을 거쳐 저장 및 축적함으로써 사용자가 원하는 분석정보를 취득할 수 있다.

이와 같이 공공데이터를 수집하여 공간정보화하기 위해 가공하는 과정은 데이터웨어하우스(Data Ware House)의 일반적인 기능과 같다(정용완, 2012). 본 연구에서는 공간정보로 가공한 데이터 결과물을 수요처의 요구에 맞게 제공해주는 일련의 과정을 결합하여 공공데이터를 활용한 공간정보데이터용 어플라이언스를 제안하고자 한다. 이를 위해 선행연구에서 제시한 연계서버 모형(연성현 등, 2014)을 기반으로 다양한 채널의 데이터를 수집하는 수집기와 수집한 데이터에 좌표를 부여하고 가공하는 전환기를 구성하여 텍스트 기반의 공공정보를 활용한 공간정보데이터 제공 시스템을 구축하였다. 그리고 이를 활용하여 기존에 공공데이터를 수집하고, 재가공하여 공간정보서비스로 배포하는 일련의 절차에 소요되는 비용을 줄일 수 있는지 검토하고, 공간참조유형의 공공정보를 효율적으로 제공하고 활용할 수 있는 방안을 살펴보고자 하였다.

2. 국내의 공공데이터 제공방식 분석

미국, 영국 등에서는 공간정보데이터를 각 국가별 공공데이터 포털을 통해 제공하고 있다. 이 장에서는 국외의 공공데이터 포털과 우리나라의 공공데이터포털의 지리정보 제공 현황을 비교하고, 특히 공간정보데이터를 제공하는 기술에 대해 인프라 측면에서 분석하였다. 이를 통해 국내 공공데이터를 공간정보서비스로 배포하는데 필요한

3) CKAN(Comprehensive Knowledge Archive Network), 데이터 관리를 위한 웹 기반의 오픈소스 시스템.

4) CMS(Contents Management System), 일반적인 데이터 관리 전용 시스템.

기술 요소로 ‘데이터 수집과 변환 기술’, ‘텍스트 정보의 지리적 시각화 기술’, ‘일체형 서비스 제공 기술’을 도출할 수 있었다.

2.1. 공공데이터 포털 현황

2009년부터 미국 정부는 열린 정부(Open Government)를 표방하며 ‘Open Data Action Plan’을 통해 거버넌스 혁신을 달성하고 공공데이터를 개방하기 위해 교통, 교육 등의 정보자원에 대한 창구로 Data.gov⁵⁾를 구축하였다. ‘Data.gov’의 목적은 정부에서 생성한 데이터셋(dataset)의 가치를 높이고, 기계 판독이 가능하도록 제공함으로써 대중들의 접근을 향상시키는 것이다(서형준 등, 2014). 최근에는 클라우드 컴퓨팅 기술을 검색 부문 등에 접목하여 외부에서도 자동 수집, 가공이 가능한 형태로 적용하였다(송효진, 2013). 2015년 현재 농업, 비즈니스, 기후, 교육 등의 카테고리

성된 125,200건의 데이터셋을 제공하고 있다. 또한 다양한 형식의 데이터를 제공받을 수 있는 링크를 제공하며, 시(city), 연방정부(federal), 대학(university), 주(state), 국가(country) 등의 범위에 해당하는 데이터로 표현하여 보여준다. 특히 데이터를 LOD 형태로 관리하고 있으며 Data.gov를 통해 일원화 된 지리정보 데이터를 제공한다. 기반 플랫폼도 데이터 시각화나 온라인 분석 툴 및 자동화 도구 기능을 갖추고 있어 활용이 용이하다.

한편 국내 ‘Data.go.kr’은 여러 기관에 분산되어 있는 데이터를 한곳으로 모아 통합창구를 만들기 위한 목적으로 구축되었다(김단비, 2015). 우리나라의 공공데이터포털은 <표 1>에서 보는 바와 같이, 제공하는 데이터셋의 확장자 종류가 국외의 공공데이터 제공 사이트보다 제한적임을 알 수 있다. 특히 대부분 텍스트 기반의 행정정보를 제공하고 있고, 웹상에 링크(link) 되어있는 일부 정보에

<표 1> 공공데이터셋의 확장자별 분류 현황

Portal	File extensions based on Korea open data portal										*
	XLS	XLSX	CSV	XML	HTML	JSON	RDF	TXT	JPG	HWP	
	Other Extensions										
Data.gov	O	O	O	O	O	O	O	O	O	X	49
	ACCESS, TIF, GML, WMS, ZIP and so on										
Data.gov.uk	O		O	O	O	X	O	X	X	X	15
	WMS, PDF, ZIP, ODS, SPARQL										
Data.gov.au	O	O	O	O	O	O	X	O	O	X	49
	ZIP, SHP, KML, WMS, PDF, WFS and so on										
Data.go.kr	O	O	O	O	O	O	O	O	O	O	14
	Nothing										

출처: 김단비(2015), 「정부 3.0 시대의 개방형 공공데이터포털의 효과적인 접근을 위한 사례 연구」, p.21.

5) Data.gov, <http://www.data.gov>, 미국의 공공데이터포털.

〈표 2〉 국내 기관 별 주요 공공정보 개방서비스

제공기관	사이트	주요 제공정보	주요기능
행정자치부	공공데이터포털	저작권 이슈에서 자유로운 텍스트 기반의 행정정보	• 국가가 보유하고 있는 다양한 공공정보를 국민에 개방 및 중계
국토교통부	국가공간정보통합체계	공공부문에서 생산한 공간 정보	• 수평/수직의 데이터 연계·취합 체계를 기반으로 방대한 양의 공간데이터 확보하고 오픈API, 관련 라이브러리 제공
국토교통부	공간정보오픈플랫폼	고해상도 3차원 공간정보	• 2D, 3D 지도, 연속지적도 등 범정부적으로 생산된 다양한 공간정보 제공 • 데이터 및 응용시스템 개발 API 제공

한해서 공간데이터 파일 형태로 제공한다. 국내 공간정보데이터는 오히려 국가공간정보통합체계나 브이월드(공간정보오픈플랫폼) 같은 다른 채널을 통해서 제공 받을 수 있다. 〈표 2〉에서 제시한 국내 기관별 공공정보 제공 서비스와 같이 이원화 된 체계는 공공데이터의 연계 비용을 상승시키고, 데이터를 융합하여 사용하기 위한 부수적인 가공 작업을 수반한다.

2.2. 공간정보데이터 제공 틀 현황

1) 정보공개플랫폼 CKAN(The Comprehensive Knowledge Archive Network)

CKAN은 데이터의 저장과 관리, 검색, 메타데이터, 시각화, 지도표시 등의 기능을 가지고 있으며 미국의 'Data.gov', 영국의 'Data.gov.uk' 와 같이 전 세계적으로 공공플랫폼과 국가 데이터 포털 구축 등에 많이 사용되고 있다. CKAN 플랫폼은 데이터를 개방하고 연결하는 일들이 쉽게 이루어질 수 있도록 하였다. CKAN은 키워드, 태그, 연관 데이터셋 등 구글 스타일의 검색을 지원한다. 이 도구는 오픈데이터의 등록을 간편하게 할 수 있도록 데이터셋의 등록에 있어서 Title, URL, Description, Tags, License, Organization 등의 정보를 입

력하도록 하고 있고, 데이터셋을 그룹화 할 수 있도록 하는 편의 기능을 제공한다(백인수, 2013).

2) OGPL(The Open Government Platform)

OGPL은 인도와 미국의 합동 프로젝트로 정부 데이터 등을 공개함으로써 정부의 투명성과 국민의 참여도를 향상시키는 것을 목표로 시작되었다. 특히 인도의 'India.gov.in'과 미국의 'Data.gov' 사이트의 장점을 결합 및 확장하였으며, 유용한 기계판독가능 형식(Machine-readable Format)으로 데이터를 변환함으로써 더 나은 의사결정을 위한 정보 제공을 가능하게 한다. OGPL의 주요 기능은 국가 내 여러 부서들의 행정문서 등을 발행(Publish)하고 데이터셋(Dataset) 관리 및 승인을 위한 내부 워크플로우를 처리하며, 국가의 주요 주제에 관한 풍부한 정보를 제공하는 커뮤니티 공간을 생성할 수 있도록 하고 시, 구 단위 등 다양한 레벨별 데이터 링크 기능과 오픈소스 아키텍처를 지원한다(백인수, 2013).

3) 구글 퓨전테이블(Fusion Tables)

구글 문서도구(Google Docs) 중 하나인 퓨전테이블은 클라우드 기반의 스프레드시트(Spread-

sheet) 프로그램이다. 단순히 데이터를 입력하고 관리하는 기능뿐만 아니라 자신이 가지고 있는 데이터를 이용하여 시각화 기능을 통한 지도제작 서비스를 제공한다(전보애, 2012). 시각화하기 위한 데이터가 주소와 같은 위치 참조 속성을 가지고 있을 경우, 퓨전테이블의 지오코드 기능이 이를 자동으로 지리적 좌표와 연결해주고, 밀도맵(Density Map), 점묘도(Dot Map) 등으로 데이터의 공간적 분포를 분석할 수 있도록 지원한다. 또한 지도데이터와 데이터테이블 간의 병합 기능을 지원하여 새로운 데이터를 만들 수 있고, 데이터를 KML(Keyholes Markup Language)로 변환하는 기능도 제공하고 있으므로 구글어스와도 연동 가능하다.

4) Data Appliance for ArcGIS

ESRI사에서는 ArcGIS를 활용한 데이터 어플라이언스 서비스를 하고 있으며, 정제되어 바로 사용할 수 있도록 준비된 GIS 데이터들을 제공한다. arcgis.com을 통해 제공되는 서비스로서 사용자는 웹브라우저 접속만으로 다양한 GIS 데이터를 조회하고 활용할 수 있다.

2.3. 시사점

국내 공공데이터포털은 양적으로는 단기간에 공공기관이 보유하고 있는 데이터를 확보하였으나 데이터 제공 형태가 제한적이어서 선진국 대비 산업적 활용과 서비스 다양성은 미흡한 수준이다. 공공데이터를 효율적으로 활용하기 위해서는 공

공데이터의 품질을 제고하고, 법·제도를 정비하여 공공데이터의 공유 문화가 조성될 수 있는 기반을 구축하는 것도 중요하지만, 기술적으로 상호 데이터를 연계하고, 변환하는 기술, 그리고 데이터를 검색하고, 시각화 하는 부분 역시 중요하다(강희중 2014). 또한 국내에서 제공하는 공공데이터 중 공간데이터 포맷으로 제공하는 정보가 해외에 비해 많지 않은 실정이다.

이러한 현 상황의 원인을 우선 시스템 측면에 국한하여 살펴보면, CKAN과 같은 지도서비스를 지원하는 오픈소스 기반의 데이터 공유 전용 플랫폼을 활용하지 않고, 자체 CMS를 구축한 것에 기인한다. 또한 민간 부문의 활용 수요가 해외에 비해 낮은 이유는 다수의 데이터가 텍스트 기반의 일차원적 자료이기 때문이다.

이에 대한 해결방안을 찾기 위해 CKAN과 OGPL에서 공공데이터를 관리하는 매커니즘을 살펴해보았다. 국내에서는 공간데이터를 제공하는 채널이 따로 분리되어 있어 접근이 어려운 측면도 있으므로, 공공데이터포털과 국가공간정보유통채널 혹은 브이월드를 LOGD⁶⁾형태로 통합하는 것도 대안이 될 수 있다. 그러나 부처 간 협업이나 비용문제 등으로 인해 현실적인 어려움이 따른다. 따라서 우선적으로 텍스트 기반의 공공데이터를 공간정보로 전환하여 민간에서 사용하기 쉬운 형태로 제공해 준다면, 원천 데이터의 상품성을 높여서 활용 수요 증대로 이어질 수 있을 것이다. 이를 위해서는 구글의 퓨전테이블과 ESRI의 'Data Appliance for ArcGIS' 와 같이 일반데이터와 공간데이터를

6) LOGD(Linked Open Government Data), 공공정보 배포를 위해 팀버너스리의 LOD를 확장한 개념.

다루고 배포하는 기능들을 결합하여 해당 기술을 차용할 필요가 있다.

3. 시범시스템 구축 및 활용예시

3.1. 시범시스템 구축 목적 및 범위

공공데이터를 활용한 주제도 제작의 편의를 돕고 Geo-Data Appliance 플랫폼의 최소 요구 기능을 도출하기 위하여 프로토타입(prototype)을 개발하였다. 프로토타입으로 개발한 ‘공공DB 연계 서버’는 [그림 1]과 같이 공공데이터포털 및 그 외 공공데이터를 제공해주는 플랫폼을 통해 오픈API 혹은 파일링크 방식으로 데이터를 수집한 후, 공간참조유형 정보를 이용하여 좌표를 부여한 데이터로 전환하고 이를 제공하는 시스템이다(연성현 등, 2014). 주요 기능은 공간참조유형의 공공데이터 수집, 데이터 ETL⁷⁾ 변환, 데이터 제공 서비스 부분으로 구성된다.

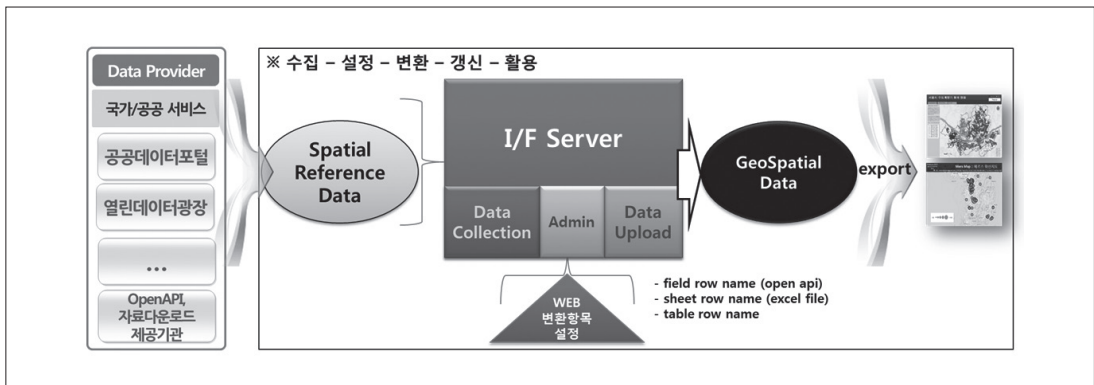
시범시스템을 구축하기 위해 먼저 공공데이터포털을 중심으로 국내외 공공데이터 개방서비스의 정보 공개 및 정보 유통 구조를 분석하고 이를

공간정보로 변환하는 기본적인 기능을 구현하였다. 특히 공공데이터를 공간정보 분야에서 손쉽게 유통시킬 수 있는 장기적 비전을 세울 수 있도록 정보 수집, 정보 가공 등 세부적 분야부터 실행 가능성을 확인하는 전략을 수립하였고, 구체적인 시스템 설계 및 모델 검증을 실행하였다.

시범시스템 구축 범위는 선행연구를 통해 제시된 모델인 ‘공공DB 연계서버’를 바탕으로 공공데이터포털의 데이터를 수집하여 가공하는 기술요소를 일부 구현하고, 그에 따른 운영 시나리오를 점검할 수 있는 내용으로 한정하였다. 즉, 공공데이터를 연계하는 기법 구현, 이를 공간정보데이터로 전환하는 기술에 초점을 맞추었고, 데이터를 확인할 수 있는 최소한의 지도 가시화 기능만을 적용하였다. 또한 경제적 효율성을 확보하고, 특정 벤더에 대한 종속성을 극복할 수 있도록 공개 소프트웨어를 활용하였다.

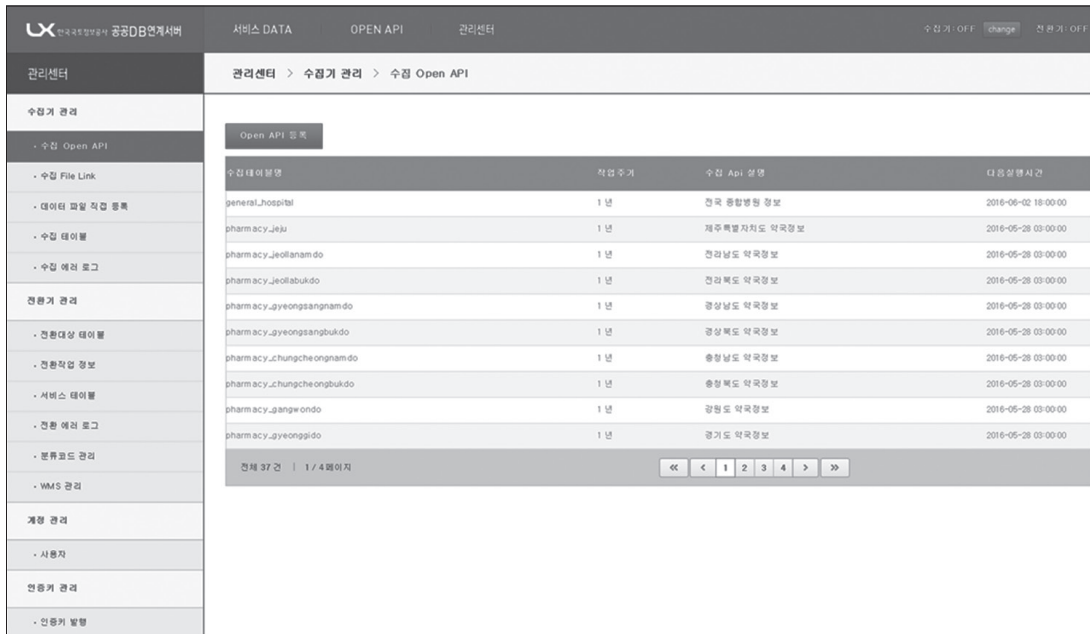
3.2. 주요기능 및 방법

주요 기능은 [그림 2]와 같이 공공정보 수집 부

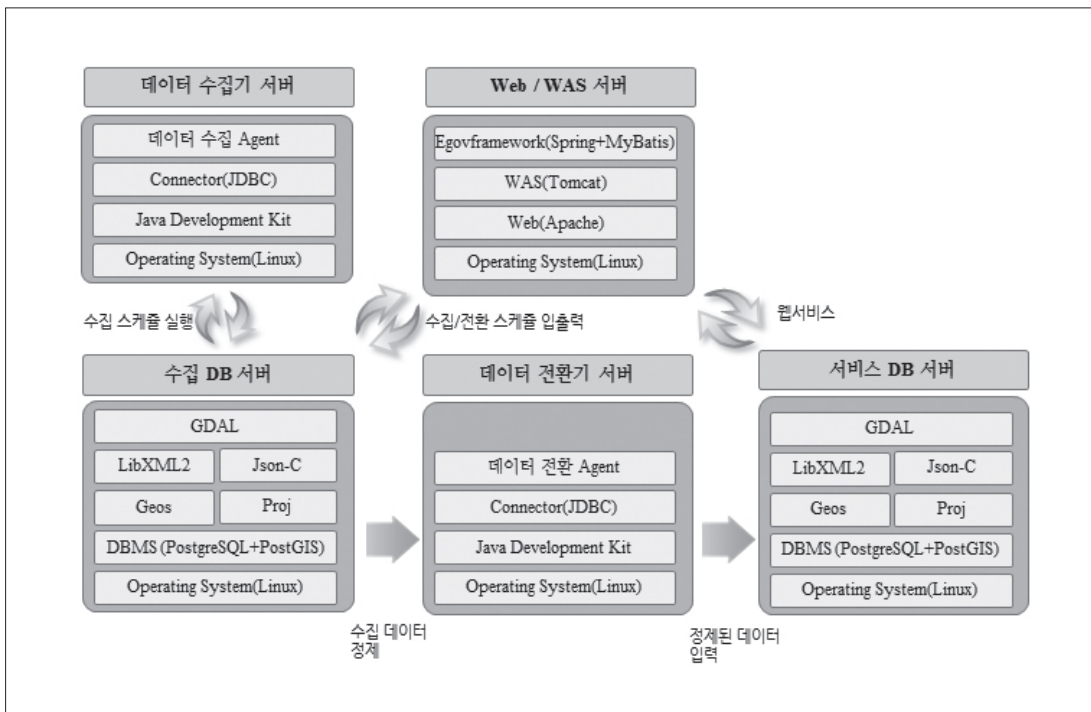


[그림 1] 공공DB 연계서버 Prototype 시스템(논리적 구성도)

7) ETL(Extraction, Transformation, Loading), 데이터를 추출하여 가공한 후 적재하는 일련의 과정.



[그림 2] 공공DB 연계서버 기능(수집, 전환, 인증키, 계정) 관리 화면



[그림 3] 공공DB 연계서버 Prototype 시스템 (물리적 구성도)

문과 공공데이터 전환 부분, 사용자 서비스 부분과 관리 부분으로 나눌 수 있다. 수집 부분에서는 공공데이터포털을 자료제공자로 한정하여 오픈 API 방식과 파일다운로드 방식으로 데이터를 가져온다. 오픈API는 인증키를 신청하여 해당 자료를 제공받는데 링크된 파일 다운로드와는 달리 인증 과정과 주기적 연동 처리 등은 관리페이지에 정보 입력 후 자동화 될 수 있도록 구성하였다.

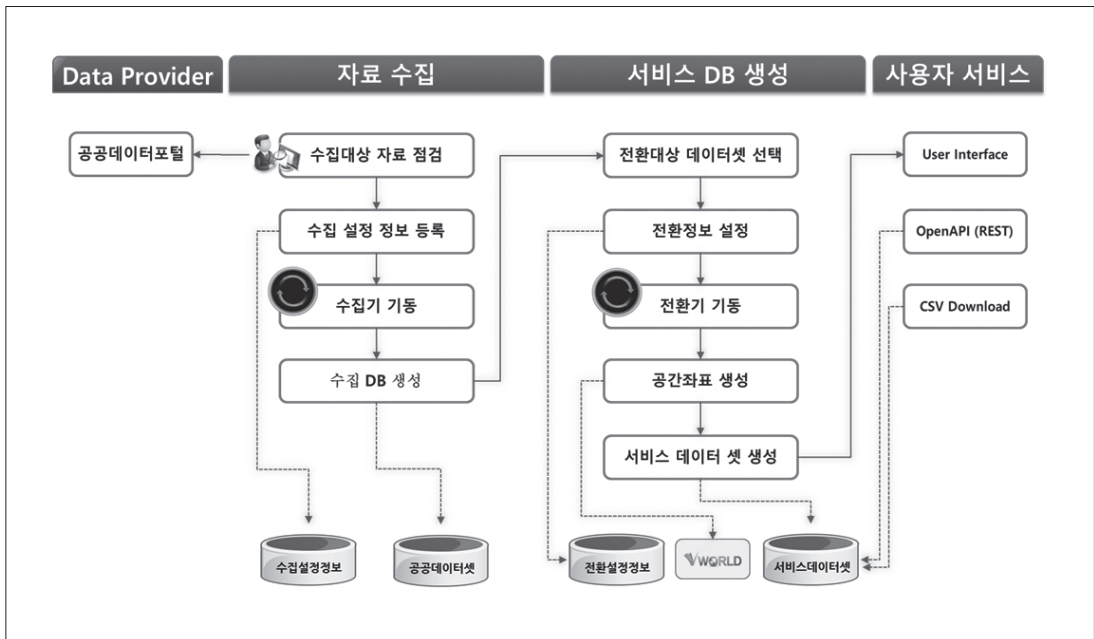
전환 부분에서는 다양한 형식의 데이터 중 도로명주소나 지번 등 위치 참조 유형을 포함하고 있는 데이터에 대해 좌표를 부여하고, 실패 기록을 남긴다. 이때 1차적으로 브이월드가 제공하는 지오코딩 API를 활용하고, 좌표를 찾지 못할 경우 2차적으로 포털사이트에서 제공하는 지오코딩 API를 호출한다. 전환된 모든 공공데이터는 웹 기반의 브이월드 배경지도 위에서 확인할 수 있고, 관

리 부분에서는 인증키를 발급하여 외부에서 KML 포맷으로 받을 수 있는 URL을 제공한다. 또한 접속 권한을 관리하여 유료 사용자만이 서비스 리스트를 확인할 수 있도록 구성하였다.

시범시스템의 물리적 시스템 구성은 컴퓨팅 자원과 물리적인 특성들을 분리하는 방법으로 가상화 기술을 차용하였다. [그림 3]에서와 같이 전자정부프레임워크를 적용하였고, 브이월드 배경지도와 오픈API를 사용하였으며 정보표출도구로서 오픈레이어스(OpenLayers)를 확장 정의하여 제공하였다.

3.3. 서비스 시나리오

시범시스템은 [그림 4]와 같이 여러 기관의 공공 데이터를 취합한 후, 다양한 수요목적에 따라 가공하여 데이터베이스 형식으로 제공하는 기능으



[그림 4] 공공DB 연계서버 Work Flow

로 구현했고, 이를 통해 분석용 정보를 쉽게 취득할 수 있다.

시범시스템을 활용한 구체적인 시나리오는 다음과 같다. 먼저 ① 생산기관이 다른 다양한 공공 데이터를 수집하여 관리용(수집용) DBMS에 적재하고, ② 데이터 활용 분야 및 품질기준에 따라 재가공한다. 이 때 연계서버의 클라우드 운영환경에서 ③ Virtual DBMS 인스턴스(Instance)를 제공한다. 이렇게 수집기에서 ④ 흩어져 있는 다양한 데이터 소스를 수집하여 관리용 저장소에 저장하면 전환기에서는 ⑤ 관리용 DB의 데이터를 정보수요자들의 필요에 맞도록 변환(가공)한 후, 제공 시점에 공간DBMS 인스턴스로 전환할 수 있도록 클라우드 템플릿(Cloud Template) 형식을 사용한다. ⑥ 인터넷 환경에서 데이터 요청 시, 요청자의 수요 목적에 맞는 가상공간 DBMS 를 생성하여 제공하고, ⑦ 요청자는 서비스가 제공하는 웹기반으로 이루어진 경량의 공간정보 시각화 도구를 이용하여 정보를 조회하거나, 자신이 보유한 분석도구

(예 : ArcGIS, Quantum GIS, SAS 등)를 이용하여 사용 목적에 맞게 활용할 수 있다.

3.4. 활용예시

2015년 6월 메르스(MERS)로 인해 질병정보 지도에 대한 수요가 증가함에 따라 활용사례 연구를 위하여 ‘메르스 정보지도’를 시범시스템으로 활용하여 제작하였고, 이를 통해 공공데이터 활용의 효율성을 분석하였다. 제작 과정은 우선 종합병원, 보건소, 약국 정보를 공공데이터포털을 통해 소싱(Sourcing) 하고, 확진자 감염장소, 확진자 발생병원, 경유병원, 안심병원, 확진자 이동 경로 등 추가 데이터를 기사에서 수집하여 엑셀로 작성하였고, 연계서버의 전환기를 통해 공간정보파일로 변환하였다. 이러한 방식으로 시범시스템에 제작된 서비스 데이터는 배경지도 상에 중첩해서 확인할 수 있었다.

[그림 5]와 같이 ‘공공DB 연계서버’를 활용하여 메르스 정보지도 서비스를 구성하는데 걸린 시간



[그림 5] 공공DB 연계서버 Work Flow

과 비용은 전문적인 GIS 소프트웨어를 이용하여 시스템을 구축하는 것에 비해 줄어들었는데, 구글 퓨전테이블을 통한 지리적 시각화와 비슷한 수준으로 작업 시간을 개선할 수 있었다. 이러한 시간 절감은 GIS 소프트웨어에 익숙하지 않은 사용자일수록 더 크게 나타날 것으로 예상된다. ‘메르스 정보지도’와 같은 유형의 시스템을 구축하기 위한 일반적인 작업 절차는 크게 세 단계로 나뉘는데, 공공데이터포털의 공간참조유형 데이터를 연계하고, 지오코딩 과정을 통해 공간정보데이터로 전환한 후, 수집한 여러 데이터를 수요자가 원하는 시점에 지도상에 가시화하고 KML 형태의 파일로 변환하여 제공하는 것이다. 절차상 구체적인 부분에서는 차이가 있을 수 있지만, 사용자가 개방된 공공데이터를 공간정보화할 수 있도록 특화된 시스템을 활용한다면 원하는 공간분석 기초자료를 쉽고 빠르게 얻을 수 있음을 확인할 수 있었다.

3.5. 시사점

공공데이터를 활용함으로써 행정의 효율성을 향상시키고 국민 편의와 사업화 가치를 증대시킬 수 있다. 그러나 제공하는 정보를 단편적으로 활용해서는 수익 창출이 어렵다. 따라서 새로운 서비스를 발굴하기 위하여 다양한 분야의 공공데이터를 연계할 필요가 있으나, 각각의 목적을 위해 일회성으로 공공데이터를 찾고 연계하는 것은 비효율적이다. 또한 기술적인 측면에서 수정, 추출, 변환 등의 가공 작업이 어려워 활용이 용이하지 않

고, 텍스트 기반의 공공데이터가 많아 공간정보를 위한 특정 포맷의 파일로 추출하기 어려우며 재사용도 어려운 실정이다.

〈표 3〉 공공데이터 활용에 대한 BOCR 분석 요소

주요 요소	하부 요소
Benefit (혜택)	공공서비스 행정효율성 향상
	공공정보 품질향상 및 업무혁신
	서비스를 통한 국민편의 증진
Opportunity (기회)	공공정보 사업화 가치 증대
	일자리 창출
	고용효과
Cost (비용)	개방 및 공유에 따른 전환비용
	공공데이터 가공 및 분석 비용
	활용을 위한 교육비용
Risk (위험)	민감한 개인정보 유출 우려
	부적절한 분석결과 활용 우려
	독점 및 선점으로 형평성 우려

출처 : 송인국(2015), 「인터넷기반 공공데이터 활용방안 연구」, p.135.

비용 측면에서도 〈표 3〉의 BOCR⁸⁾ 분석 항목에서 제시한 바와 같이 전환, 가공 및 분석 등 데이터를 직·간접적으로 핸들링하는 부분에서 큰 비용이 발생한다(송인국, 2015). 특히 데이터 핸들링을 제외하더라도 품질의 한계 때문에 데이터의 질을 높이기 위한 정보 검증이나 재가공 비용이 수반된다.

따라서 생산기관이 다른 공공데이터를 연계하여 공간데이터로 가공한 후 완제품 형식으로 제공해 주는 어플라이언스의 개발이 요구된다. 활용 예시와 같이 공공데이터를 가공하여 주제도로 제공

8) BOCR, 의사결정의 기초자료 모형으로 BCR(Benefit, Cost, Risk)에서 잠재적인 기회(Opportunity)를 추가.

할 수 있는 기초 데이터셋을 프로토타입 시스템을 통해 빠른 시간에 제공할 수 있다면 비용 절감을 통한 공간정보 서비스 창출에 기여할 수 있을 것이다. 이러한 시사점을 토대로 공공데이터를 활용한 'Geo-Data Appliance'의 요구사항 및 기술 구성요소를 도출하고자 한다.

4. Geo-Data Appliance의 구성요건

4.1. 정의 및 요구사항 도출

본 연구에서 제안하는 '공간정보를 위한 Data Appliance'란 다양한 공공기관이 생산하는 공공데이터를 공간정보와 통합 가공하여, 가상화 DBMS 서버에 적재하여 제공하는 클라우드(cloud) 기반 공공데이터 에이전시(Agency) 서비스라고 정의할 수 있다. 또한 부가적으로 공간정보로 전환되어 제공되는 공공 데이터셋의 활용을 위한 공간정보 분석도구, 개발도구 등 활용 도구를 제공하는 플랫폼 서비스라고도 정의할 수 있다.

특히 데이터의 추출, 변환, 적재, 분석을 위해 하나의 디바이스에 하드웨어(H/W), 데이터베이스(Database), 저장장치(Storage)를 통합해 일체형으로 구성하는 것을 DW어플라이언스(Database Warehouse Appliance)라고 하는데 일반적으로 소프트웨어 시장에서 어플라이언스는 특정 서비스에 종속되어서 하드웨어(H/W)와 소프트웨어(S/W)를 하나로 묶은 전용 시스템을 의미한다. 가트너(Gartner)에 따르면 DW어플라이언스는 미리 구성된 하드웨어와 소프트웨어를 묶어서 제공하고, 유지보수를 총체적으로 제공해야 하고, 서버(Server), 메모리(Memory), 저장장치, 입출력채널(Input/Output Channel) 등 하드웨어 구성 요소와 운영체

제, 데이터베이스와 관리 소프트웨어 등의 요소들이 사전에 구성되고, 고가용성을 충족시킬 수 있도록 이중화 구성을 제공해야 한다(Gartner IT Glossary). 어플라이언스의 장점은 비용 절감과 솔루션 최적화인데 기업의 이윤추구가 아니라 공익 목적의 플랫폼에 도입할 예정이기 때문에 더욱 도입이 필요한 부분이다. 이런 기준에 맞게 공간정보를 위한 Data Appliance 서비스 요구사항을 <표 4>와 같이 정리하였다.

<표 4> 공공데이터 연계한 공간정보 제공용 Data Appliance 요구사항

구분	주요 요구사항
수집	공급자가 아닌 수요자 및 이용목적에 맞는 자료수집 체계 요구
	생산기관이 다른 동일한 주제의 dataset에 대한 일관된 관리체계 요구
	시계열 특성을 가진 데이터에 대한 최신성 유지를 위한 운영정책 수립
	자료제공기관의 제공 특성에 따른 최적의 공공데이터 수집 방안 마련
	수요자의 피드백을 통한 지속적 수집체계 개선
공급	공공데이터 수요자의 이용목적 및 수준에 따른 다양한 데이터 제공 방법
	수요목적에 따른 다양한 공간정보 융합 dataset 생성 방법 및 절차
	수요발생시점에 수요자가 원하는 dataset을 적시에 제공하는 방법 및 절차
	비즈니스 목적 달성을 위하여 제공받은 데이터의 사전처리작업 최소화
	제공된 데이터에 대한 최신성 유지를 위한 수요자의 작업 최소화
인프라	다양한 공간정보 수요자의 이용 목적에 따라 능동적으로 대응하는 플랫폼
	지속적 데이터 증가에 유연하게 대응하는 Computing Architecture
	공간정보 및 공간정보 이용 도구를 동시에 제공하는 서비스 모델 지원
	SW 생명주기 및 향후 재사용을 지원하는 Computing Architecture

공공데이터의 개방과 활용이 급속히 확대되고 있음에도 불구하고 오픈데이터(Open Data)에서 기대하는 데이터 융합과 그에 따른 고부가가치 창출을 위해서는 아직 해결해야 할 문제들이 남아있다. 이러한 문제를 해결하기 위한 방안들 중에는 데이터 제공 방법 및 포맷의 다양성, 데이터 탐색의 신속성 및 정확성, 목적에 따른 데이터 품질의 적합성 등이 포함되기 때문에 공공데이터를 연계한 공간정보 제공용 데이터 어플라이언스 플랫폼을 ① 데이터 수집 측면, ② 공급 측면, 그리고 ③ 인프라 환경 구성 측면으로 구분하였다.

4.2. 기술구성요소

제안한 시스템의 주요 기능은 지리적 위치를 참조하는 다양한 공간 어플리케이션에 대하여 사용자의 목적에 따라 최적화된 컴퓨팅 인프라 및 일체의 필요한 데이터를 신속하게 제공하는 것이다. 이를 위해 일반인이 공간정보데이터를 쉽게 취득할 수 있도록 하는 구글의 퓨전태이블 매키니즘을 참고하고, 공공데이터를 연동하여 가공하며 분류가 다른 여러 데이터를 묶어서 제공하는 DW 기술을 결합하였다. 더불어 공간정보서비스를 수요 시점에 즉시 제공할 수 있도록 시스템 일체를 클라우드(cloud) 인프라에서 제공해주는 어플라이언스 개념을 도입하였다. 중점적으로 추진할 기술 요소는 ① 생산기관이 다른 공공데이터의 일관된 품질 확보를 위한 관리 기술 ② 이기종의 데이터를 통합하고 관리하는 기준으로서 지리적 위치를 활용하는 기술 ③ 활용 분야별로 요구되는 공공데이터

식별 및 데이터마트(Datamart) 생성 기술 ④ 제공된 데이터의 최신성 유지를 보장하기 위한 시스템 운영 기술 ⑤ 공공데이터 수요시점에 필요한 가상화 자원 및 데이터 집합을 즉시 제공하는 기술 등으로 요약할 수 있다.

5. 결론

「공공데이터의 제공 및 이용 활성화에 관한 법률」⁹⁾ 시행 이후 공공데이터가 다량 축적되었고, 오픈데이터의 활용이 크게 증가하고 있다. 이러한 개방형 공공데이터는 저비용으로 공급이 가능하고 활용에 따른 기대효과도 매우 크다. 또한 스마트 환경이 일상화되고 데이터 관리 및 분석에 필요한 빅데이터, 클라우드 기술 등이 발달하여 활용하기 위한 여건 또한 빠르게 성숙하고 있다. 그러나 민간에서는 더욱 많은 데이터의 개방을 요구하고 있고, 따라서 일반 사용자가 다양한 종류의 수많은 공공데이터에 쉽게 접근하고 활용할 수 있는 체계 구축이 필요하다. 본 연구의 제안시스템인 Geo-Data Appliance를 미래를 예측할 수 있는 도구로 사용하기 위해서는 Data Appliance 서비스를 제공하기 위한 공간정보 클라우드 센터(Cloud Center)가 먼저 구축되어 있어야 한다. Geo-Data Appliance의 기능 보완 및 기술적 타당성 검토에 대해서는 추후 연구가 필요할 것이다.

본 연구에서 제작한 프로토타입만으로도 ‘공공 DB 연계서버’의 활용예시와 같이 개방된 공공데이터를 공간정보 기반으로 빠르게 융합하여 웹 환경에 가시화 할 수 있었다. 다양한 공공데이터를

9) 국가법령정보센터, 법률 제11956호(시행 2013. 10. 31., 제정 2013. 7. 30).

용·복합한다면 공공정보 주제도를 만들 수도 있고, 이를 브이월드(공간정보오픈플랫폼)에서 서비스 할 수도 있다. 브이월드의 서비스 목적이 국가가 보유하고 있는 공개 가능한 공간정보를 모든 국민이 자유롭게 활용할 수 있는 기반이 되는 것이므로, 공공정보 주제도를 브이월드에서 서비스 하는 것은 이러한 목적에 부합한다. 우선 데이터 수집 측면에서 공공데이터포털이나 서울 열린데이터 광장에서 제공하는 공공정보 중 단기간에 공간 정보화가 가능한 항목과 고부가가치를 창출할 수 있는 항목을 선별하여 적용해야 한다. 많은 공공데이터를 연계하여 표출하는 것도 중요하지만 수요자가 필요로 하는 데이터를 파악하여 서비스 하는 것이 더 중요하기 때문이다(김정옥 등, 2014). 향후 민간이 보유한 공간정보와 정형, 비정형 정보들을 연계하여 활용할 수 있는 방안이 마련된다면, 일반 사용자들이 쉽게 공공데이터를 활용하여 부가가치를 창출하는 기반이 조성될 것이고, 이는 산업 활성화로 이어질 수 있을 것이다.

감사의 글

본 연구는 국토교통부 도시건축개발사업 연구비지원(13도시건축A02)에 의해 수행되었음.

【참고문헌】

- 강영옥·김현덕(2014), 「구글 API를 활용한 공공데이터의 지리적 시각화 연구」, 한국지도학회지, 14권, 1호, pp.1-15.
- 강희종(2014), 「공공데이터의 효율적 활용을 위한 정책 과제」, STEPI Insight, (156), pp.1-30.
- 김단비(2015), 「정부 3.0시대의 개방형 공공데이터포털의 효과적인 접근을 위한 사례 연구」, 석사학위논문, 서울시립대학교, pp.14-33.
- 김정옥·김지영·차득기(2014), 「공공데이터의 위치정보를 이용한 지오태깅 방법론 분석」, 한국지적학회 학술대회 논문집, pp.119-129.
- 김청림·권용진·박재휘·이상구(2010), 「시맨틱 데이터 클라우드 : 시맨틱 기술 기반 공공정보 배포 프레임워크」, 한국정보과학회 학술발표논문집 27(2C), pp.20-24.
- 박원준(2014), 「창조경제 기반조성을 위한 공공데이터 개방과 활용 사례」, 동향과 전망 : 방송·통신·전파, 제72호, pp.30-43.
- 백인수(2013), 「오픈데이터 플랫폼과 국가 데이터 전략방향」, 한국정보화진흥원, IT & Future Strategy, 제16호, pp.1-23.
- 서형준·명승환(2014), 「수요자 중심의 공공 데이터 민간 활용 방안」, 한국지역정보학회지, 17(3), pp.61-86.
- 송인국(2015), 「인터넷기반 공공데이터 활용방안 연구」, 인터넷정보학회논문지, 16(4), pp.131-139.
- 송효진(2013), 「정보속으로 : 특집 ; 공공데이터 개방 및 활용지원을 위한 지방자치단체의 조례 제정의 필요성」, 지역정보화, pp.18-21.
- 연성현·이인수·차득기(2014), 「공공데이터와 오픈플랫폼의 효율적 자료연계를 위한 연계서버 모형」, 지적, 제44권, 제1호, pp.113-125.
- 유현경·육혜인·한희정·김용(2015), 「LOD기반의 문화콘텐츠 정보서비스 확장에 관한 연구」, 정보관리학회지, 31(4), pp.109-134.
- 윤소영(2013), 「공공데이터 활용을 위한 링크드

- 데이터 국가 연계체계 구축에 관한 연구」, 정보관리학회지, 30(1), pp.259-284.
- 이만재(2011), 「빅 데이터와 공공 데이터 활용」, Internet&information security, 2(2), pp.47-64.
 - 이현정·남영준(2014), 「우리나라 공공데이터의 이용활성화 방안에 관한 연구 : 링크드 오픈 데이터화 전략을 중심으로」, 정보관리학회지 31(4), pp.249-266.
 - 전보애(2012), 「지리교육에서 GeoCloud의 활용」, 한국사진지리학회지, 제22권, 제4호, pp.75-88.
 - 정용완(2012), 「데이터웨어하우스(DW) 어플라이언스 기술 동향 분석」, 한국정보기술학회지, 제10권, 제2호 pp.107-111.
 - Data.gov.uk(2015), <http://www.data.gov.uk/> (접속일 : 2015.9.12.).
 - Data.gov.kr(2015), <http://www.data.go.kr/> (접속일 : 2015.8.20.).
 - Data.gov(2015), <http://www.data.gov/> (접속일 : 2015.9.12.).
 - Data.gov.au(2015), <http://data.gov.au/> (접속일 : 2015.9.13.).
 - Data Appliance for ArcGIS(2015), <http://www.esri.com/software/arcgis/data-appliance> (접속일 : 2015.1.9.).
 - Data.seoul.go.kr(2015), <http://data.seoul.go.kr/> (접속일 : 2015.7.10.).
 - IT Glossary(2015), <http://www.gartner.com/it-glossary/> (접속일 : 2015.1.9.).
 - Google Fusion Tables(2014), <https://www.google.com/fusiontables/data?dsrclid=implicit> (접속일: 2014.8.19.).
 - India.gov.in(2015), <http://india.gov.in/> (접속일 : 2015.9.13.).