지방자치단체 데이터세트의 서비스 방안 연구*

- 서울 열린 데이터 광장 서비스를 중심으로 -

A Study on the Services of Data-sets in the Local Government: Based on the Cases of Seoul Open Data Portal Services

안 대 진 (Dae-Jin An)**
이 해 영 (Hae-young Rieh)***

목 차

1. 서 론
2. 이론적 배경
3. 데이터세트 정보서비스 사례
4. 서울시 데이터세트 서비스 현황
5. 데이터세트의 서비스 방안
6. 결 론

〈초 록〉

최근 각국은 정부 데이터 개방을 위한 데이터 플랫폼을 구축하고, 행정정보데이터세트 등 다양한 데이터를 민간에서 웹을 통해 자유롭게 활용하도록 하고 있다. 이 연구는 전 세계적으로 가장 많이 활용되는 대표적인 오픈 데이터 플랫폼인 소크라타와 CKAN의 기능을 분석하고, 각국의 활용 사례를 살펴보았다. 그리고 서울시의 데이터세트 서비스 현황을 사례로 검토하여, 오픈 데이터 플랫폼을 적용한 지방자치단체의 데이터세트 관리 및 서비스 방안을 개념적으로 제시하였다. 이를 통해 기록으로서의 데이터세트 서비스 제공을 위한 지방자치단체의 장기적인 보존과 관리체제 방안을 제도적 측면과 데이터의 생산 및 관리 측면, 서비스 측면 및 플랫폼의 선택 등 다양한 측면에서 제시하였다.

주제어: 데이터세트, 기록정보서비스, 오픈데이터, 오픈데이터 플랫폼, 서울 열린데이터광장

<ABSTRACT>

Recently many countries have established data platforms to disclose government-owned data that include administrative data-sets and provide free access to the public via Web. This research analyzes the "Socrata" and "CKAN", the most popular representative open data platforms in the world, and reviews functions and their practical cases in operation in several cities of various nations. It also examines the current status of the data-set services in the City of Seoul to provide conceptual bases for management and service of the local governments' data-sets using open data platform. Then it suggests measures that ensure the long-term preservation and management of data-sets as archives for services, which includes the aspects of preparing systems, creating and managing data, providing services, and selecting platforms.

Keywords: data-sets, information services, open data, open data platform, Seoul open data portal

---

* 본 연구는 안대진의 한국기록관리학교육원 논문(2013)인 「지방자치단체 데이터세트 정보서비스 방안: 서울 열린데이터광장 서비스를 중심으로」를 요약·수정한 것임.

** 명지대학교 디지털아카이빙연구소 선임연구원(ahndj@rikar.org)
*** 명지대학교 기록정보과학전문대학원 교수(hyrieh@mju.ac.kr) (교신저자)

■ 접수일: 2013년 7월 18일 ■ 최초심사일: 2013년 7월 26일 ■ 게재확정일: 2013년 8월 20일

www.kci.go.kr

1. 서 론

1.1 연구의 배경과 목적

컴퓨터가 처리하거나 분석할 수 있는 형태로 존재하는 관련 정보의 집합체를 말하는(기록학 용어사전 2009) 데이터세트는 중앙정부와 지방자치단체에서 다양한 형태로 생산하고 있다.[1] 업무시스템에 축적되는 데이터세트는 업무활동의 증거이자 시민들의 삶을 반영하고 있는 공적 트랜잭션의 결과로서 기록으로 관리하고 장기적으로 보존할 필요가 있다. 국가기록원은 행정데이터를 수집, 보존, 활용하기 위해 2007년 '행정정보시스템 데이터세트 기록관리방안 연구용역'을 추진하여 법령에 의해 데이터세트를 전자기록으로 관리할 수 있는 체계를 마련하였다(국가기록원 2010).

주요 선진국들은 데이터세트를 중요 기록물로 인식하고 별도의 아카이브를 구축하여 관리하고 있으며, 최근에는 정부 데이터 개방을 위한 데이터 포털을 개설하고 이를 통해 교통, 통신, 안전, 의료, 고용 등 다양한 분야의 정보 서비스를 제공하고 있다. 데이터 포털은 정부 데이터 활용의 역할 외에도 개인이나 기업 등 민간이 보유하고 있는 공공기록을 자발적으로 제공하는 창구의 역할을 하며 기록 수집정책의 새로운 패러다임을 제시하고 있다.

인터넷 기술의 발달과 열린 정부를 지향하는 세계적 흐름은 그 동안의 기록관리 전략에 새로운 변화를 요구하고 있다. 기록정보 서비스의 시점이 아카이브즈 단계를 포함해 기록관이나 처리과 단계로 다변화되고, 전자 형태의 비정형 데이터세트가 원자료(raw data)나 Open API[2] 형태로 제공되며, 정보 활용의 방식도 변화하고 있기 때문이다(서울특별시 2012). 선별과 평가 전략 측면에서는 기능, 활동, 주제의 관점 이외에도 해당 데이터세트의 가용성, 활용성, 유용성, 오픈 라이선스, 시각화의 용이성, 데이터 공개로 인한 경제적 가치 등이 새로운 고려 요소로 부각되고 있다.[3]

이러한 정보 활용 행태의 변화는 각국이 추구하고 있는 거버먼트 2.0 기반의 오픈 거버넌스, 또는 오픈 거버먼트 데이터 전략에 기인한다. 2012년 5월 서울시 또한 오픈 거버먼트 전략의 일환으로 '서울 열린 데이터 광장'을 개설했다.[4] '열린 데이터 광장'이 기록 정보서비스 측면에

1) 예를 들면 현재 서울 열린 데이터 광장에서 파일형태로 제공하는 데이터세트 중 조회 수가 많은 데이터는 유동인구정보, 공중화장실 공간정보, 금연구역정보, 방재시설현황, 자전거도로 공간정보, 자전거편의시설 공간정보, 2011년 서울서베이 도시정책지표조사 도서관 공간정보, 5678역 출구별 관광지정보 등이 있다. [cited 2013.7.18]. 〈http://data.seoul.go.kr/openinf/dataset/datasetlist.jsp〉.

2) Open API(Open Application Programming Interface)는 인터넷 이용자가 일방적으로 웹 검색 결과 및 사용자 인터페이스(UI) 등을 제공받는 데 그치지 않고 직접 응용 프로그램과 서비스를 개발할 수 있도록 공개된 API를 말한다. 지도 서비스 및 다양한 서비스에서 시도되고 있으며 누구나 접근하여 사용할 수 있다는 장점이 있다(네이버 지식백과). [cited 2013.7.20]. 〈http://terms.naver.com/entry.nhn?cid=2955&docId=20297&categoryId=2955〉.

3) DATA.GOV 운영지침(DATA.GOV Concept of Operations)에서는 데이터 서비스의 성공여부를 가늠하는 기준으로 데이터의 가용성, 활용성, 유용성을 들고 있다. 또한 서울시는 공개 대상 공공정보를 선정하는 기준으로 선호도 및 공개로 인한 경제적 가치 등을 이용하였다(Office of E-Government and IT. et al. 2010).

4) 서울 열린 데이터 광장. [cited 2013.7.20]. 〈http://data.seoul.go.kr/〉.

www.kci.go.kr

서 중요한 이유는 민원청구 형태로 진행되는 정보공개 사이트와 분리 및 차별화되었다는 점이다. 서울시는 '열린 데이터 광장'이라는 플랫폼을 통해 공공정보 데이터세트를 다양한 형태로 제공함으로써 기록관리의 목표라 할 수 있는 설명책임성과 지식의 유통을 충실히 달성해 나가고 있다.

이 연구는 지방자치단체가 보유하고 있는 공공정보 데이터세트가 효과적 관리와 활용을 통해 소요 인력과 자원을 최소화시키고, 기관의 설명책임성을 강화하며, 시민들의 삶의 질을 향상시킬 수 있다는 전제하에, 이의 효과적 관리와 활용방안을 도출하고자 한다. 이를 통해 무엇보다 데이터세트가 기록학적 관리나 활용의 측면에서 다루어지지 않고 전산부서의 업무로 인식됨으로써 발생할 수 있는 문제를 지적하고 바람직한 개선방안을 제안하고자 한다. 아울러 행정정보시스템 내에서 전자기록으로서의 진본성을 유지하기 위한 방안 또한 고려해 보고자 한다. 또한 오픈 거버먼트를 위한 대표적인 오픈 데이터 플랫폼인 소크라타(Socrata), CKAN 등을 서울시 사례와 비교 분석하여 우리나라의 지방자치단체 수준에 적용할 수 있는 최선의 데이터세트 관리 및 서비스 방안을 제안해 보고자 한다.

1.2 선행연구

기록관리 분야의 데이터세트에 대한 연구는 기록물관리법이 개정된 2007년을 전후로 시작되었다. 초기에는 행정정보시스템에서의 데이터세트 획득이나 이관 등 기록으로 관리하기 위한 기본적인 연구가 이루어졌고, 이후 각 행정정보시스템의 데이터세트 유형별 기록관리 방안에 대한 연구로 세분화되었다. 데이터세트의 기록화를 직접적으로 다루지는 않았지만 오픈 거버먼트의 측면에서 각국 정부의 공공기록 활용 사례를 분석하는 연구 또한 활발히 이루어지고 있다.

우선 데이터세트 아카이빙에 대한 전반적인 연구로는 현문수(2005), 김포옥, 윤수영(2007), 조은희, 임진희(2009)와 임진희, 조은희(2010a/2010b), 김남경(2012) 등을 들 수 있다. 현문수(2007)는 다양한 유형의 데이터세트의 획득과 등록, 저장, 조직 및 기술, 접근, 장기보존에 이르기까지 관리를 위한 각 고려 요소를 제안했다. 이를 통해 데이터세트 기록 관리의 중요성에 대한 인식을 확산시킨 측면이 있다. 김포옥, 윤수영(2007)은 전자기록물의 개념으로부터 데이터세트의 의미와 특성을 정리하고, 데이터세트의 각 유형별 데이터 추출 방법을 제시했다. 조은희, 임진희(2009)와 임진희, 조은희(2010b)는 행정정보시스템에서 건 단위 데이터세트 기록물을 선별하고 이관하는 현실적 관리 프로세스의 측면에서 연구를 수행했다. 임진희, 조은희(2010a)는 이관하는 과정을 지원하는 도구가 갖춰야 할 기능요건을 데이터세트 유형별로 제시함으로써 본격적인 전자기록의 이관을 앞둔 시기에 유용한 성과를 제공했다. 김남경(2012)은 미국, 영국, 호주와 국내의 아카이빙 사례를 비교하고 제도나 전담기관, 시스템적 측면부터 보존과 활용 서비스까지 데이터세트 기록관리를 위해 고려해야 할 요소를 정리했다.

두 번째로 각 행정정보시스템의 데이터세트 기록화에 관한 연구를 들 수 있다. 이은별(2008), 이순한(2008), 한철희(2007), 최정윤(2009), 남

www.kci.go.kr

영준, 서만덕(2009) 등의 연구에서는 디지털예산회계시스템, 조달업무시스템, 교육행정정보시스템, 지리정보시스템, KSDC 여론조사자료 등 행정정보시스템별로 기록화 방안이 제시되었다. 각 연구에서는 개별 행정정보시스템의 특성과 생산되는 데이터세트의 유형을 분석하고 해당 업무의 가치와 기록화 대상의 선별에 중점을 두어 체계적 관리 방안을 제시하고 있다.

세 번째로 공공기록 활용에 관련된 연구이다. 이영숙(2007)은 국가기록원과 미국의 NARA, 영국의 TNA의 기록서비스 현황을 통해 검색, 정보공개, 학습지원, 온라인 전시, 견학 등 전통적 의미의 기록정보서비스를 분석하며 인구통계를 비롯한 각종 통계수치 데이터세트의 유용성에 대해 언급했다. 김유승(2010)과 윤은하(2012)는 정보공개제도 중심으로, 제공되고 있는 공공정보 서비스의 문제점을 지적하고 있다. 김유승, 전진한(2011)은 거버먼트 2.0 개념과 정보공개제도의 상관성을 논하면서 공공정보의 공개와 재사용(Reuse)을 위해 정보공개제도가 데이터세트 서비스로 확장되어야 함을 주장하였다.

해외에서는 디지털 데이터의 재사용에 관한 연구가 진행되었다. 사회과학 분야 연구자들이 정성적 데이터로부터 정량 데이터를 추출하고, 쉽게 맥락을 획득하고, 다양한 데이터세트를 연결하고 조합하는 방식에 대한 연구가 진행되었다(Faniel, Kriesberg & Yakel 2012). Faniel, Yakel(2011)은 DIPIR(Dissemination Information Packages for Information Reuse)프로젝트를 통해 데이터 생산자, 재사용자, 데이터 큐레이터, 또는 데이터세트 자체 등 각 주체의 목적에 따라 데이터의 가치가 달라지며, 이러한 차이를 극복하기 위해 신뢰적 데이터세트 제공 기반을 마련해야 한다고 주장했다. 공공정보 데이터세트를 재사용할 때 공공기관과 시민, 아키비스트 등 각기 다른 세 주체가 새로운 가치를 창출하기 위한 공동의 목적을 인식하고 신뢰성 있는 정보를 제공하고, 올바르게 사용하는 하나의 모델로 삼을 수 있을 것이다.

이상 살펴본 연구들은 데이터세트의 기록화나 서비스를 위한 구체적인 프로세스까지 제시하고 있지는 않다. 본 연구는 공공정보 데이터세트의 지속적 재사용을 위한 방안으로 장기적 보존체계인 데이터세트 아카이브의 설립을 우선 과제로 삼았으며, 이를 전제로 실제 데이터세트를 활용하기 위한 프로세스와 데이터 플랫폼이 갖추어야 할 요소 등을 살펴보았다.

1.3 연구 설계 및 방법

본 연구는 기록 정보서비스의 관점에서 서비스 대상 데이터세트의 효과적인 선별과 획득, 시각화와 분석 등을 위한 플랫폼 기능요건, 그리고 효과적 활성화를 위한 제도적 개선점을 제안하여, 지방자치단체가 생산하고 있는 데이터세트를 좀 더 서비스 중심적, 사용자 중심적 관점에서 선별, 보존, 확산시키는 방안을 제시하는 것으로 연구의 범위를 설정하였다. 이를 위해 지방자치단체 중 서울시의 '서울 열린 데이터 광장' 사례를 중심으로 공개된 데이터세트의 유형과 서비스 특성을 분석한다. 이를 바탕으로 우선 데이터세트의 이론적 개념과 공공정보 데이터세트의 유형 및 특징, 공공정보의 재사용이 갖는 의미와 중요성을 알아보고 확산 플랫폼을 개발하여 활용하고 있는 미국과 영국, 유럽 등의 선진 사례를 서울시와 비교하여 서비스 플

www.kci.go.kr

랫폼이 갖추어야 할 각 단계별 기능을 도출한다. 이 과정에서 제시된 시사점을 분석하여 제도, 생산 및 관리, 서비스, 플랫폼의 관점에서 서울열린데이터광장의 서비스 개선방안을 제시하겠다.

2. 이론적 배경

2.1 기록으로서 데이터세트의 개념

데이터세트는 컴퓨터의 잦은 갱신으로 인해 기존 데이터를 새로운 환경에서 읽을 수 없는 문제를 해결하기 위해, 데이터베이스에서 그 데이터만을 추출해 보관하는 원래의 데이터 묶음으로 정의된다(Shepherd 2000). SAA용어집에서는 데이터베이스를 보다 조직화하고 구조화된 형태로 저장시켜 관리시스템에 의해 접근 또는 업데이트가 가능한 정보로서 다양한 목적에 부합되도록 조작되고 추출될 수 있다고 정의하고 있다. 즉 데이터세트는 반복적으로 발생되는 사건이나 이에 대한 정보가 관계를 가지고 데이터베이스에 누적되어 컴퓨터로 처리할 수 있게 구조화된 것을 의미한다(Pearce-Moses [2013]).

공공기록물 관리에 관한 법률 시행령에서는 데이터세트의 범위를 한정하여 기록화 대상으로서의 행정정보 데이터세트를 언급하고 있다. 시행령에서 말하는 '행정정보 데이터세트'란 각급 행정기관에서 업무상 사용하고 있는 행정정보시스템에서 생산되는 문자, 숫자, 도형, 이미지 및 그 밖의 데이터를 말한다(제2조 11항).

데이터세트의 일반적 특징으로는 첫째, 모체인 데이터베이스 의존성을 들 수 있다. 데이터베이스의 유형에 따라 해당 데이터세트의 아카이빙 전략 또한 변화한다. 데이터베이스의 구성요소는 콘텐츠, 데이터베이스 관리 시스템(DBMS), 어플리케이션이다(Dutch National Archives and the Dutch Ministry of the Interior and Kingdom Relations 2003). 기록의 3대 구성요건을 내용, 구조, 맥락이라고 했을 때, 위 세 가지 구성요소 중 콘텐츠는 데이터세트의 내용에 해당하고, DBMS와 어플리케이션은 구조와 맥락의 유지에 관여하는 요소라 할 수 있다. 따라서 계층형, 관계형, 객체형, 네트워크형 등 각 데이터베이스의 유형에 따라 데이터세트의 내용을 선별하고 구조와 맥락을 획득하는 전략 또한 달라질 수밖에 없다(Ashley 2004).

둘째, 데이터세트는 데이터베이스에 비해 상대적으로 기록화 대상의 식별이 쉽고 특정 대상의 임의적 선택 및 보존이 가능하다.[5] 스냅샷 등의 추출방법을 통해 임의의 테이블이나 컬럼만을 보존할 수 있는데, 이러한 특성은 아키비스트가 좀 더 유연하게 보존 대상을 선정할 수 있도록 한다.

셋째, 데이터세트 기록물은 건(item)과 철(file)의 구분이 모호하다. 비록 데이터세트의 아카이빙이 데이터베이스보다 수월하다고 해도, 건과 철 단위의 구분이 분명하지 않기 때문에 장기

5) 관계형 데이터베이스의 보존은 복잡한 문제를 지니고 있어, TNA나 NARA에서 구축하고 있는 데이터세트 아카이브와 같이 데이터베이스 안에 저장되었던 데이터세트를 대상으로 보존 및 이용을 보장하는 방향이 우선적으로 취해지고 있다(현문수 2005. 104).

www.kci.go.kr

보존을 위한 파일시스템 구성이나 이관을 위한 철 편성의 어려움은 여전히 남아 있다(임진희, 조은희 2010a). 행정정보시스템 안에서는 여러 단위업무가 하나의 데이터베이스를 참조하는 경우가 일반적이다. 이처럼 복합적인 속성을 가진 데이터베이스 및 데이터세트의 구조 때문에 건 수준을 식별하고 철로 묶어내는 작업은 단위과제 중심의 일반적 기록관리 프로세스와 상당한 개념적 차이를 드러낸다(국가기록원 2010). 각 행정정보시스템별로 데이터세트 기록을 식별하기 위한 기준을 구체화하는 추가적인 연구가 필요한 상황이다.

2.2 공공정보의 재사용

공공정보의 재사용은 오픈데이터의 가치를 통해 실현될 수 있다. 오픈 데이터란 활용, 재사용, 재배포가 누구에게나 허락된 데이터를 말한다. Open Definition에서는 오픈 데이터의 요건으로서 여러 가지를 제시하고 있다.[6] 대표적인 내용들을 정리해보면, 먼저 사용가능성과 접근성이 확보되어야 한다는 것이다. 데이터 전체의 사용이나 합당한 재생산에 비용이 들지 않아야 하며, 편리하게 사용하고 수정할 수 있는 형식이어야 한다. 또한 재사용과 재배포가 가능해야 한다. 여러 데이터세트를 혼합하여 사용하는 것을 포함하여 데이터의 재사용과 재배포를 허용하는 조항을 제공해야 한다. 그리고 보편적 참여가 가능해야 한다. 개인이나 단체에 상관없이 누구나 차별 없이 사용할 수 있어야 한다. '상업적인' 사용에 대한 제한이나 일정한 목적을 위해서만 사용하도록 하는 제한(예컨대 교육 용도로만 사용)인 '비상업적인' 제한은 금지된다.

유럽연합은 2003년 11월 「공공정보의 재사용에 관한 지침」을 제정하고 각 회원국으로 하여금 적극적으로 공공정보의 상업적 재사용에 관한 법률을 제정할 것을 권고하고 있다. 영국은 유럽연합의 지침에 의거하여 2005년 6월 「공공정보 재사용에 관한 규정 2005」(Re-use of Public Sector Information Regulation 2005)를 제정하였다.[7] '재사용'의 정의 또한 지침의 내용과 거의 비슷하다. "공공기관이 보유하고 있는 공공정보가 해당 공공기관의 공무상 업무의 최초 목적과는 다른 목적으로 개인 또는 법인에 의해 사용되는 것"을 말한다.

3. 데이터세트 정보서비스 사례

3.1 오픈 데이터 플랫폼 소크라타(Socrata)

미국 정부 및 각 주 정부들은 오픈 데이터 전략을 실현하기 위한 도구로 민간 플랫폼인 소크라타[8]를 사용하고 있다. 소크라타는 공공기관이 데이터를 저장・관리・공개・활용하는 모든 프로세스를 제공한다. 미국 정부 기관들은 공통 플랫폼을 이용함으로써 데이터 공개 프로세스의 효율성 및 예산절감을 꾀하고, 기관끼리

6) Open Definition. [cited 2013.7.13]. 〈http://opendefinition.org/〉.
7) The National Archives. Public sector information Directive and Regulations. [cited 2013.7.7]. 〈http://www.nationalarchives.gov.uk/information-management/legislation/directive-and-regulations.htm〉.
8) Socrata, Inc. [cited 2013.7.13]. 〈http://www.socrata.com〉.

www.kci.go.kr

의 유연한 데이터 연계를 실현하고 있다. 소크라타 플랫폼의 기능 및 특징을 분석하고, 대표적 활용사례인 DATA.GOV 및 시카고와 시애틀 시 정부의 데이터 서비스 현황을 살펴 본 후 시사점을 분석해 보았다.

3.1.1 기능 및 특징

1) 등록 및 저장

소크라타 플랫폼으로의 데이터세트 등록 과정은 관련된 데이터세트의 조사 및 식별, 데이터 정리 및 포맷 설정, 그리고 메타데이터 처리 및 등록의 3단계를 거친다. 조사 및 식별 단계에서는 시민들에게 유용할 만한 데이터세트를 선별하게 된다. 국가 안보와 관련된 자료를 제외하고 전부 공개 대상으로 식별하는 것이 일반적이다. 정리 및 포맷 설정 단계에서는 데이터베이스 테이블과 뷰(view)[9]를 스프레드시트로 처리할 수 있는 CSV나 웹에서 처리 가능한 XML 등의 다양한 형태로 추출하는 작업이 이루어지며, 이후 데이터 자체와 데이터세트의 제공 목적 등을 기술하게 된다. 이러한 세 단계 과정을 통해 원천 데이터를 등록하면 외부 이용자들에게 공개할 준비를 마치게 된다.

소크라타 플랫폼에서 데이터세트의 분류는 각 기관의 분류체계에 따라 생산부서 및 주제분야로 나누어 설정되고, 활용성을 높이기 위한 방편으로 소크라타의 주제어 설정과 소셜 태깅이 추가된다. 등록된 데이터세트는 플랫폼 패키지에 포함된 스토리지나 기관에서 구축한 별도의 리포지터리에 저장되어 변경이력이 관리된다.

2) 활용

정부 및 기관이 보유한 데이터세트는 중앙 리포지터리로 옮겨진 후 다운로드, 검색 및 조합, 분석 및 시각화, 커뮤니티 참여, 소셜미디어를 통한 공유 등으로 활용된다. 데이터세트의 분석 및 시각화 기능은 벤치마크 스터디 당시 시민들의 요구사항을 반영한 것으로 SPSS 등 연구자들이 데이터 마이닝을 위해 사용하는 도구 이상의 기능을 온라인 환경에서 제공한다. 이용자들은 웹사이트에서 대용량 데이터세트를 다운로드 받지 않아도 적절한 그래픽 형태를 선택하여 즉각적으로 데이터세트 요소들을 분석하고 그래프로 표현할 수 있다. 또한 특별한 프로그램 없이도 웹 브라우저를 통해 스프레드시트, 지도, 도표, 캘린더 등 다양한 형식으로 데이터세트를 볼 수 있고, 자신만의 필터나 효과를 적용하여 저장한 후 다른 사이트에 공유할 수 있다.

해당 데이터세트에 대하여 개인이나 커뮤니티의 댓글 참여와 리뷰 또한 가능하다. 시민들의 적극적인 참여는 정부가 보유한 데이터세트의 가치를 재발견하고 추가적 공개를 촉진하는 기능을 한다. 트위터나 페이스북 등 소셜 미디어 사이트로의 데이터세트 공유는 데이터세트의 이용을 쉽게 늘리고 기관 웹사이트의 사용자 트래픽을 완화할 수 있는 좋은 수단이 되기도 한다.

소크라타는 접근성을 다양화하기 위해 아이폰과 블랙베리 등 대표적 모바일 플랫폼에서의 접근을 허용하고 있다. GPS 기능이 내장된 스마트폰의 확산과 관련 어플리케이션 사용의 증

9) 데이터베이스에서 뷰(view)는 정보를 테이블(표)이라는 개념으로 관계시킨 관계형 데이터베이스에서의 가상적인 테이블을 말한다(네이버 지식백과). [cited 2013.7.8].
〈http://terms.naver.com/entry.nhn?cid=2959&docId=841055&mobile&categoryId=2959〉.

www.kci.go.kr

〈그림 1〉 소크라타 공개 프로세스 및 솔루션

출처: Socrata. 2012. Unlock Government Data with Social Data Discovery

가, 고속 네트워크 접속 등을 고려하여 사용자의 현재 접속 위치에 따른 맞춤형 데이터세트를 제공하고 있다.

소크라타의 분석 도구들은 데이터세트 제공자가 커뮤니티 내의 활동적인 사용자를 식별하는 등 이용자들과의 소통 및 웹을 통한 이용 행태 분석을 가능하도록 해 준다. 데이터세트 제공자는 대시 보드를 통해 각 데이터세트의 접속 현황과 포맷별 이용 현황 등을 그래픽과 통계 도표 형태로 제공받을 수 있다. 이러한 분석 결과는 가장 이용이 활발한 웹사이트의 식별뿐만 아니라, 재배포된 제 3의 웹사이트에서의 이용 현황까지 보여 준다.

3) 자동화 도구

데이터 서비스를 제공하고자 하는 기관은 보유 데이터를 시민들과 개발자들이 이용할 수 있는 포맷으로 변환해 주는 도구가 필요하다. 소크라타 플랫폼에서는 공개된 모든 유형의 데이터세트를 Open API로 자동 변환해 주는 솔루션을 제공한다.[10] 이 솔루션을 사용하여 스프레드시트 형태의 원천 데이터를 SODA(Socrata Open Data API)로 변환하는 데는 채 몇 분이 걸리지 않는다. 공공기관에서 원천 데이터를 Open API로 개발하기 위한 비용과 노력이 상당히 투입됨을 고려할 때, 데이터 플랫폼이 자동화 도구를 제공하는 것은 제공기관, 사용자, 개발자

10) Socrata API Foundry. [cited 2013.7.13]. 〈http://www.socrata.com/api-foundry/〉.

www.kci.go.kr

모두에게 상당히 중요하다.

GIS 형태의 데이터세트를 지도와 매쉬업 해주는 도구 또한 제공한다.[11] 이러한 도구는 지리정보 Open API를 활용한 응용 프로그램이나 서비스 개발자들이 구글 맵 등 공개된 지도 API와 매쉬업하여 지도와 결합된 정보를 제공하고자 할 때 부딪치게 되는 기능 제한 문제나 사용료 부담을 완화해 준다.

3.1.2 소크라타 활용 사례

1) DATA.GOV

미국의 DATA.GOV는 모든 연방 기관의 데이터세트가 업로드되는 정부 데이터 포털의 역할을 하며, 다양한 수준의 메타데이터 구조와 정의를 통해 데이터세트를 쉽게 찾을 수 있도록 해 준다. DATA.GOV는 2009년 데이터 서비스를 시작했고 현재는 데이터 공개의 범위를 확장하고 효과적으로 활용하는 것에 집중하고 있다. DATA.GOV 사이트에서는 초기에 연방정부와 각 부처의 데이터만을 제공했으나, 최근에는 별도의 사이트를 통해 발티모어, 시카고, 뉴욕 등 12개 시 정부가 보유하고 있는 데이터세트의 통합서비스를 제공하고 있다.[12] 데이터세트는 원천 데이터와 GIS 데이터로 나누어 서비스하고 있으며, 각 기관에서 제공하는 어플리케이션의 목록과 링크 또한 제공한다. 정부 차원의 데이터 포털인 만큼 참여 기관의 각 부처별 공개 데이터세트 목록과 그에 대한 설명, 통계 등도 제공하고 있다. 블로그와 주제별 커뮤니티를 통해 개인화된 서비스 창구를 제공함으로써 국민들의 참여를 장려하고 있다.

DATA.GOV는 2013년 7월 현재 183,656개의 데이터세트, 176,074개의 지리정보 데이터세트, 349개의 시민 개발 어플리케이션, 137개의 모바일 어플리케이션 등을 서비스하고 있다. 농림부, 상무부 등 총 172개 정부 부처가 공개에 참여하고 있으며, 농업, 금융, 기업정보 등 다양한 주제 분야에서 데이터세트가 제공되고 있다.

2) 시카고

시카고 시 정부는 2012년 12월 시카고 시의 모든 정부 기관이 보유한 공공 데이터를 공개하기로 하는 등 오픈 데이터 운동의 선두에 서 왔다.[13] 시카고 시는 2010년 2월부터 소크라타 플랫폼을 통해 데이터 세트 서비스를 제공하고 있다.[14] 이 사이트에는 2013년 7월 현재 943개의 데이터세트가 공개되어 있다. 이 중 원천 데이터세트 형태는 275개, 외부 데이터세트 4개, 파일이나 문서 201개, 필터링된 뷰 359개, 도표 26개, 지도 73개 등이다. 주제 분야는 행정 및 재정, 건축, 지역과 경제개발, 교육, 환경 및 지속가능 개발, 인종, 사건, 정보공개 등 16개이다. 1년간 가장 많이 접근된 데이터세트는 시카고 공무원 이름・직함・급여 데이터, 교통카드 판매점 지도, 범죄발생 지도, 건축 허가, 경찰서 위치, 식품위생검사 결과 등 주로 생활 밀착형 데이터들이다.

11) Socrata Mondara. [cited 2013.7.13]. 〈http://www.socrata.com/mondara/〉.
12) DATA.GOV Communities. [cited 2013.7.15]. 〈http://www.data.gov/cities/community/cities〉.
13) About data.cityofchicago.org. [cited 2013.7.5]. 〈http://www.socrata.com/customer-spotlight/city-of-chicago/〉.
14) City of Chicago Data Portal. [cited 2013.7.5]. 〈https://data.cityofchicago.org/〉.

www.kci.go.kr

3) 시애틀

시애틀 역시 2010년 2월 시카고, 뉴욕 등과 함께 소크라타 플랫폼을 통해 데이터세트 서비스를 시작했다.[15] 시애틀 시 정부의 각 부처들은 기계 가독형 데이터세트를 개발하면서 데이터세트를 활용한 정보 서비스의 가치를 높이고자 노력하고 있다.[16] 2013년 7월 현재 공개된 699개의 데이터세트의 종류는 원천 데이터 세트 83개, 도표 32개, 지도 267개, 필터링된 뷰 575개, 파일과 문서 21개, 캘린더 5개 등이다. 데이터세트 소유주는 예산청, 기획개발부 등 14개 소속기관이며, 주제 분야는 커뮤니티, 교육, 재정, 국토, 허가, 공공 안전, 교통 등 7개이다. 이용이 활발한 데이터세트는 911 화재신고 실시간 데이터, 911 범죄신고, 법규위반 데이터, 경찰서 보고 사고 등 치안 및 안전 관련 정보들이다.

3.2 오픈소스 데이터 허브 플랫폼 CKAN[17]

CKAN은 공공데이터 개방 및 확산을 지원하는 오픈 플랫폼으로, 비영리 단체인 Open Knowledge Foundation에 의해 설립되었다. 유럽 지역을 중심으로 약 40개의 글로벌 데이터 허브가 CKAN을 기반으로 구축되었다.[18] 2012년 10월 서비스가 시작된 한국의 Korea Data Hub 역시 CKAN 플랫폼 1.7.1 버전을 사용하여 개발되었다.[19]

CKAN은 국가나 지방자치단체 수준의 공공기관 및 기업이나 단체 등의 데이터 공급자들이 보유하고 있는 데이터를 공개하여 활용되도록 하는 것을 목표로 개발되었다. CKAN은 기술적으로 오픈 소스에 의존하는 모듈화된 구조를 가지고 있으며, 데이터 수집이나 추가 기능 등이 필요할 경우 사용자별로 확장이 가능하다. 또한 데이터세트를 참조하는 응용 프로그램의 접근을 위해 풍부한 API를 제공한다.

3.2.1 기능 및 특징

1) 등록 및 저장

데이터세트의 등록은 특정 포털 사이트가 CKAN 플랫폼을 사용하는 지방자치단체로부터 데이터세트를 획득하거나, 여러 CKAN 소스에 존재하는 특정 주제의 데이터세트만을 발췌하여 가져오는 것이다. CKAN은 데이터 색인 메타데이터로 DCAT[20] 표준을 사용하므로, CKAN 플랫폼을 사용하지 않는 색인으로부터 자료를 획득할 수 있다. 기관마다 구축된 기존 리포지터리로부터 자동 획득도 가능하다. 이를 위해 CKAN은 지리정보 CSW서버, 웹 카탈로그, HTML 인덱스 페이지, 기타 CKAN 개체들로부터의 수집을 지원한다. 데이터 제공자들에게는 대시보드가 제공되어 다운로드 수나, 댓

15) Data.Seattle.Gov. [cited 2013.7.5]. 〈https://data.seattle.gov/〉.
16) About Data.Seattle.Gov. [cited 2013.7.7]. 〈http://www.socrata.com/customer-spotlight/city-of-seattle/〉.
17) CKAN. [cited 2013.7.7]. 〈http://ckan.org/〉.
18) 2013년 2월 미국 정부는 DATA.GOV 플랫폼에 CKAN의 검색, 브라우징, 메타데이터 지원, 수집 시스템, 머신 인터페이스 등의 요소를 통합하기로 공식 발표했다(CKAN news). [cited 2013.7.8]. 〈http://ckan.org/2013/02/04/us-data-gov-to-use-ckan/〉.
19) Korea Data Hub. [cited 2013.7.13]. 〈http://thedatahub.kr/〉.
20) DCAT은 데이터세트의 웹 상호운용성을 촉진하기 위한 W3C의 RDF 구문 표준이다(W3C). [cited 2013.7.13]. 〈http://www.w3.org/TR/vocab-dcat/〉.

www.kci.go.kr

글 수 등 데이터세트 이용 현황에 대한 리포트를 제공한다.

데이터세트를 공개하기 위해서는 웹 인터페이스를 통해 직접 데이터세트를 등록하거나 CKAN의 인터페이스나 자체 개발한 스프레드시트 추출 프로그램 등을 사용할 수 있다. 제공하는 메타데이터 요소는 제목, 고유 식별자, 기술요소, 갱신 내역, 미리보기, 라이선스, 태그, 제공 그룹, 포맷, API key, 기타 요소 등 11개 필드로 구성된다.

등록된 데이터세트는 자동으로 'CKAN DataStore'에 입고되고, DataStore는 사용자가 등록한 파일로부터 데이터를 추출하여 독자적 포맷으로 변환한 후 저장하게 된다. CKAN은 사용자가 업로드한 파일을 로컬 파일 시스템에 저장하거나 클라우드 저장소에 보존하도록 미리 설정하여 저장 프로세스를 자동화할 수도 있다.

2) 활용

데이터세트가 업로드되면, 사용자들은 API를 통해 다운로드받지 않은 상태로 검색하거나 임의로 사용할 수 있게 된다. CKAN은 시각화 툴을 제공하므로 검색 결과에 데이터세트에 대한 미리보기를 보여 준다. 데이터세트 검색과 이용을 위해 자체 인터페이스를 통해 응용 프로그램의 질의와 이용자의 전문 확장 검색, 다운로드 링크를 포함한 데이터세트의 모든 정보, 제공자의 주제별 데이터세트 보유 목록, RSS/Atom 피드를 비롯한 최근의 활동 내역, 구글 분석도구를 이용한 다운로드 수 등의 데이터세트 사용 통계 등을 제공한다. 데이터세트 제공자를 포함한 등록 사용자에게는 메타데이터의 업데이트 내역을 저장할 수 있는 인터페이스가 따로 제공된다.

데이터세트의 편리한 활용을 위해서는 다양한 미리보기 툴과 시각화 툴이 제공된다. CSV

〈그림 2〉 CKAN 플랫폼 구조

www.kci.go.kr

포맷이나 파일의 테이블을 표현하기 위한 테이블 뷰, 수치 데이터세트를 표현하기 위한 가로축과 세로 축의 그래픽 데이터, 테이블에 포함된 칼럼, 좌표 등을 표현하기 위한 매핑 데이터, 표준 이미지 파일을 표현하기 위한 이미지 데이터 등이 제공되는 것이다. 또한 DataStore의 API를 활용하면 자신만의 데이터세트 미리보기 도구를 생성할 수 있다. 데이터세트의 시각화 외에도 서비스 사이트의 디자인을 사용자별로 변형할 수 있도록 CSS 등의 표준 테마를 제공한다.

사용자들은 CKAN 커뮤니티 포털을 통해 데이터세트별로 댓글이나 리뷰를 작성할 수 있다. 특정 데이터세트를 '팔로우'하여 변경사항이나 업데이트, 새로운 사용자 활동들을 전달받을 수 있다. 또 '플래그' 기능을 통해 잘못된 사항이나 오류 등을 데이터세트 제공자에게 전달할 수 있는데, 이러한 기능은 데이터세트의 품질을 향상시키는 수단이 된다. 데이터세트마다 소셜 네트워크 확장 API를 적용하여 트위터나 페이스북, 구글 플러스 등 자신의 소셜 네트워크 페이지로 특정 데이터세트를 쉽게 공유할 수도 있다.

CKAN은 오픈 소스이므로 누구나 무료로 다운로드하여 설치하고, 소스를 수정할 수 있으며 기능을 사용하는 데 기간의 제한이 없다. 총 60개 이상의 확장 프로그램이 제공되므로 자신의 필요에 따라 선택하여 사용하거나 추가 기능을 만들 수도 있다. 또한 공개 표준이므로 등록한 데이터세트의 호스팅이 끊겨 접근이 불가능해지는 일은 발생하지 않는다. 하지만 개인 수준의 사용에 비해 지방자치단체 등 공공기관들의 복잡한 프로젝트의 경우 맞춤형 개발 및 호스팅을 위한 비용이 발생한다.

3) 자동화 도구

CKAN에서도 사용자의 원천 데이터세트를 Open API로 변환해 주는 기능을 제공한다. 제목, URL, 라이선스 종류, 기술정보, 그룹 등 기술 정보를 입력하는 간단한 과정을 거쳐 Open API가 생성된다. 특정 데이터세트를 지도와 매쉬업하는 것 또한 가능하다. 소크라타의 경우처럼 사용자가 쉽게 여러 데이터세트를 매쉬업할 수 있는 인터페이스를 제공하는 것은 아니지만, 데이터세트에 포함된 지리정보 메타데이터를 시스템이 자동으로 인식하여 지리정보 필터를 통해 걸러지도록 목록화해 준다.

3.2.2 CKAN 활용사례

1) 영국

DATA.GOV.UK는 영국의 모든 공공기관 데이터가 통합 서비스되는 데이터 포털이다. DATA.GOV.UK는 오픈소스 플랫폼인 CKAN과 Drupal[21] 엔진을 사용하여 개발자 중심으로 디자인되었고, 데이터의 시각화, 접근성 개선 등을 위해 새로운 버전을 개발 중이다.[22]

DATA.GOV.UK의 가장 큰 특징은 링크드 데이터를 이용하여 영국 정부가 공개한 데이터세트를 모든 국민이 어디서든 유용하게 활용할

21) Drupal은 오픈 소스 컨텐츠관리시스템(CMS, Content Management System)으로, 사용자의 컨텐츠를 쉽게 조직하고 관리하도록 해 주는 다양한 사용자별 환경을 제공한다(Drupal). [cited 2013.7.17]. 〈https://drupal.org/about〉.

22) DATA.GOV.UK. [cited 2013.7.13]. 〈http://data.gov.uk/〉.

www.kci.go.kr

수 있도록 서비스를 개발해 나가고 있다는 것이다. DATA.GOV.UK는 2013년 7월 기준 9,736개의 정부 공공기관의 데이터세트를 공개하고 있다. 2013년 7월 현재 가장 이용이 활발한 데이터세트는 2010년 영국 실직에 대한 색인과 가구의 생활수준을 나타내 주는 통계 등 주로 국립통계청의 인구 센서스에 기반한 데이터세트이다. 대표적인 응용 서비스로는 주택 가격 동향을 지역별, 연대별로 시각화한 'UK House Prices', 가까운 개인병원 위치를 검색해 주는 'Find GPs', 분산되어 있는 연구 투자 정보를 통합하여 하나의 뷰로 서비스해 주는 'Research Funding Explorer' 등이 있다.

2) EU

PublicData.eu는 범 유럽 데이터세트 카탈로그를 제공하는 데이터 포털로 2011년 6월 구축되었다. 포털의 백엔드[23]가 CKAN으로 구성되며 CKAN 수집 프레임워크를 통해 유럽 지역 25개 정부 기관의 공공데이터 카탈로그로부터 데이터세트 메타데이터를 수집했다. 수집 대상 기관은 프랑스의 Data Publica, 스웨덴의 OpenGov, 이탈리아의 CSI Piemonte, 그리고 런던, 파리, 비엔나를 포함한 여러 시 단위 카탈로그이다. 2013년 7월 현재 전체 20,210 데이터세트가 공개되었고, 주제 분야별 숫자는 재정・예산 436개, 사회 분야 226개, 환경 213개, 교통 193개, 교육・커뮤니케이션 187개, 농업・수산업・임업 176개, 인구 140개, 경제・산업 114개, 건강 79개, 고용 77개이다. 제공 포맷별로는 스프레드시트 포맷 CSV 3,683개, 엑셀 XLS 2,943개, HTML 1,576개, PDF 1,224개 등이 공개되어 있다.[24]

3) 헬싱키

HRI(Helsinki Region InfoShare)는 국민들에게 지역 정보를 쉽고 빠르게 제공하는 것을 목표로 개발되었다. 기능과 데이터 카탈로그를 위해 CKAN을 핵심 코어로 두었고, 워드프레스는 뉴스와 블로그를 위해 사용된다. 핀란드어 사이트에서는 2013년 7월 현재 1,030개의 데이터세트가 서비스되고 있다. HRI는 헬싱키 시민뿐만 아니라, 기업, 대학, 연구기관, 시 공무원 등의 상업적・비상업적 이용을 목표로 구축되었으며 모든 데이터세트는 무료로 제공된다. 개발 초기에는 삶의 질, 경제와 복지, 고용과 교통 등 국가나 도시 차원의 광범위한 통계 데이터세트 위주로 공개되었으나 활용도가 높은 GIS 데이터세트의 비중이 점점 높아지고 있다.[25]

3.3 시사점

미국과 영국, EU의 사례에서 살펴보았듯이 대부분의 선진국에서는 소크라타와 CKAN 등

23) 일반적으로 프론트엔드(front-end)와 백엔드(back-end)라는 용어는 프로세스의 처음과 마지막 단계를 가리킨다. '프론트엔드'는 사용자로부터 다양한 형태의 입력을 받아 '백엔드'가 사용할 수 있는 규격을 따라 처리할 책임을 진다. '프론트엔드'를 '백엔드'로 연결하는 것은 인터페이스의 일종이다(위키백과). [cited 2013.7.14]. 〈http://ko.wikipedia.org/wiki/%ED%94%84%EB%A1%A0%ED%8A%B8%EC%97%94%EB%93%9C%EC%99%80_%EB%B0%B1%EC%97%94%EB%93%9C〉.

24) publicdata.eu. [cited 2013.7.12]. 〈http://publicdata.eu/〉.

25) HELSINKI REGION INFOSHARE. [cited 2013.7.14]. 〈http://www.hri.fi/fi/〉.

www.kci.go.kr

오픈 데이터 플랫폼을 통해 국가 전체 기관의 공공정보를 통합 및 연계하여 서비스하고 있다. 위 사례들을 통해 국내 지방자치단체에서 서비스를 준비할 때 참고할 만한 세 가지 시사점을 얻을 수 있었다.

우선 첫째로 동일한 플랫폼의 사용이다. 공공영역 데이터세트의 가치를 높이려면 활용 가능한 데이터를 최대한 많은 시민들이 사용할 수 있도록 해야 한다. 예를 들면 우리나라도 서울시가 구축한 데이터 플랫폼은 중앙정부 차원의 데이터 허브인 공공데이터포털과 기능을 연계하는 방안이 고려되어야 할 것이다.

둘째로, 데이터세트가 실제 서비스되기 위한 다양한 솔루션이 필요하다. 서비스 담당자를 위한 Open API 변환도구와 일반 사용자를 위한 온라인 시각화 도구가 대표적이다. 데이터세트 제공자를 위한 사용현황 통계 및 분석도구가 제공되면 더욱 바람직하다. 소크라타나 CKAN 등의 패키지를 국내 기관이 도입하는 것도 하나의 방법이 될 수 있다. 소크라타의 경우 시각화나 분석, 자동화 도구 등 풍부한 기능을 제공하지만 민간 플랫폼으로 도입시 많은 예산이 필요하며, CKAN은 오픈 소스인 장점이 있으나 커스터마이징을 위한 인력이 투입되어야 하는 장단점이 있다. Open API로 개발하는 것이 아니라면, 가능한 한 재가공하지 않고 그대로 공개하여 시간과 비용을 아끼는 방법을 고민해 봐야 할 것이다. 일반 시민들이 웹페이지를 통해 활용하기 위한 솔루션 또한 필요하다. 데이터세트를 XML 파싱하여 웹으로 출력하거나 스프레드시트를 그래프로 변환해 주는 등 데이터세트의 활용 범위를 응용 프로그램에서 일반 사용자까지 넓혀 가려는 노력이 요구된다.

셋째, 데이터 서비스 결과는 지방자치의 훌륭한 도구가 될 수 있다. 시카고 시에서는 주로 교통이나 관공서 위치 등 생활 밀착형 정보가 많이 이용되었던 데 반해 시애틀에서는 이용률 상위 13개 데이터세트가 범죄나 화재 등 공공안전 카테고리에 해당되었다. 지방자치단체마다 고유한 정책 요구가 있음을 알 수 있다. 시민들의 정보 요구를 면밀히 분석하여 그에 부합하는 지역 밀착형 공공데이터를 개발하는 것이 중요하다. 기록관리 차원에서는 데이터세트 활용도를 근거로 단위업무별 보존기간의 책정이 달라질 수 있으며 서비스 활성화에 따라 공개 대상 기록물의 수가 증가할 수도 있을 것이다.

4. 서울시 데이터세트 서비스 현황

서울시는 2012년 5월 데이터 서비스 포털인 '서울 열린 데이터 광장'을 구축하였다. 열린 데이터 광장은 Web 2.0을 기반으로 시민이 참여해 공공 데이터를 자유롭게 이용할 수 있는 서비스이다. 열린 데이터 광장을 통해 제공되고 있는 데이터세트 유형 및 서비스 현황을 제도, 생산 및 관리, 서비스, 플랫폼 측면에서 분석해 보았다.

4.1 제도

서울시는 정보공개법 등에 의거하여 '정보소통광장'을 운영하고 있다. 하지만 서비스가 시작된 시점에는 열린 데이터 광장 서비스의 근거로 삼을 만한 공공 데이터의 민간 개방에 관한 제도적 기반이 마련되지 않았다.[26] 이에 따라

www.kci.go.kr

서울시는 2012년 5월 '서울특별시 열린 데이터 광장 서비스 이용약관'을 자체적으로 제정하고 서비스 방침을 수립했다. 제5조 '열린 데이터 광장 서비스의 이용'부터 제7조 '인증Key의 이용 및 관리'에서는 Open API 서비스의 이용을 위한 과정과 제한사항, 인증키 사용에 관한 규칙을 명시하고 있다. 제8조에서는 서비스 운영 전반과 이용시 주의사항, 제10조부터 제13조까지는 저작권, API 및 데이터파일 이용허락 조건, 책임의 제한, 손해배상 등의 원칙을 정의하고 있다.[27]

서울시는 공공데이터의 민간 활용 관련 해외 법률과 가용한 라이선스 유형을 참고하여 실행계획을 수립하고, 서비스 이용약관을 마련했다. 향후 공공데이터 개방 및 이용 활성화를 위한 조례를 마련하고, 2013년 '공공데이터 활용 지원체계 및 운영방안'을 수립・시범운영하는 등 꾸준한 제도적 정비를 진행하고 있다(서울특별시 2012). 또한 개발자 커뮤니티의 활성화를 통해 Open API 제공 등과 관련된 문제점을 파악하여 서비스 지침이나 제도에 지속적으로 반영하는 체계를 갖추고 있다. 그러나 서울시의 노력만으로는 제한이 있을 수밖에 없다. 지방자치단체와 국가적인 차원에서 생성되는 데이터세트를 지속적이고 안정적으로 보존하고 활용할 수 있는 제도적 기반이 마련되는 것이 시급한 실정이다.

4.2 생산 및 관리

서울시는 출연기관과 25개 자치구가 보유하고 있는 행정정보시스템을 통해 공공정보 데이터세트를 생산 및 보유하고 있다. 2012년 '행정데이터 통합 및 열린 데이터 광장 확장사업'을 통해 공공데이터 활용을 위한 통합 관리체계를 마련하였다. 또한 분류체계 기준 현황에 관한 내부 조사와 연구용역 결과를 종합하여 보유 데이터세트 목록을 상세히 조사하고 서비스를 위한 품질 점검 작업을 실시했다. 열린 데이터 광장의 공공정보는 서울시 BRM을 근거로 공개 가능한 행정 데이터베이스를 맵핑하여 분류체계가 구성되었다. 이는 일반 행정, 환경, 문화관광, 산업경제, 도시 관리, 복지, 교통, 안전, 보건, 교육 등의 10개 상위 분야에 49개 하위 분류체계로 이루어져 있다. 정의된 분류체계를 기준으로 수요조사를 실시하고, 이를 바탕으로 서비스를 위한 선별 즉, 개방 우선순위를 선정했다. 설문조사 결과 교통 20%, 문화 17%, 교육 12% 순으로 공공정보 영역별 공개 우선순위가 설정되었다.

메타데이터는 더블린 코어와 해외 사례를 참조하여 27개로 정의하고 있다. 이 중 공통 메타데이터는 서비스 유형과 상관없이 동일하게 적용되는 항목으로 분류체계, 태그, 제공기관, 제공부서, 담당자명, 담당자연락처, 활용성평가, 원본시스템, 저작권자명, 저작권자 연락처, 적재주기, 등록일자, URL 복사하기 등 14개 요소이

26) 2013년 7월 30일 '공공데이터의 제공 및 이용 활성화에 관한 법률'이 제정되어 2013년 10월 31일 시행 예정이다. 이 법률은 공공기관이 보유, 관리하는 데이터의 제공 및 그 이용 활성화에 관한 사항을 규정함으로써 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지하는 것을 목적으로 한다.

27) 서울 열린 데이터 광장 서비스이용약관. [cited 2013.7.11]. 〈http://data.seoul.go.kr/openinf/openplaza/agreement.jsp〉.

www.kci.go.kr

다. 서브 메타데이터는 Open API 서비스의 유형에 따라 달라지는 항목으로 크게 요청주소, 요청인자, 출력 값으로 구분된다. 서비스 유형이 Open API인 경우에만 서브 메타데이터 항목이 추가된다.

그럼에도 불구하고 개방을 위한 우선순위 선정 및 데이터세트의 관리 범주 등의 설정과 대상 식별 등에 객관적이 기준이 제시되지 않아 수요조사 등에 의존하고 있으며, 명확한 프로세스가 제시되고 있지 못하다는 한계가 있다. 전반적인 생산 및 관리 프로세스의 정립이 필요한 것으로 판단되었다.

4.3 서비스

열린 데이터 광장은 해외 각국의 선진 서비스에 대한 케이스 스터디를 통해 서비스 방향이 설정되었다. 미국, 영국, 호주, 뉴욕시, 그리고 국내의 공공데이터포털 등의 데이터세트 유형과 서비스의 장단점을 분석하여 서울시 보유 데이터에 적합한 서비스 모델을 설정하였다. 2011년 12월 클로즈드 베타를 시작으로 5개월 간의 테스트 기간을 거쳐 2012년 5월 시범 서비스를 시작하였다.

데이터세트는 여섯 개의 시각화 유형으로 제공되고 있다. 2013년 7월 기준으로 공개된 전체 1,085건의 데이터세트 중 외부 사이트 링크 664건을 제외하면, Open API 유형이 392건으로 가장 큰 비중을 차지하고 있다. Open API는 다른 유형에 비해 많은 개발 자원이 소요되지만 '서울 버스' 등의 응용 프로그램으로 개발되는 등 민간의 활용도가 가장 높은 서비스이다. 그 외의 유형으로, 스프레드시트는 334건, 도표는 1건, 지도는 142건, 파일은 172건이 제공된다.

그러나 서비스 유형이나 분석도구는 양적으로나 질적으로 아직 해외의 사례만큼 다변화되어있지 못하고, 편리한 도구가 제공되고 있지 못하다. 이러한 한계를 뛰어넘을 수 있는 체계적인 데이터 분석이 가능한 서비스를 개발할 수 있는 기술적인 지원 방안이 필요한 것으로 나타났다.

4.4 플랫폼

열린 데이터 광장 서비스 아키텍처는 원천데이터의 관리로부터 DB 품질관리, 오픈 아키텍처, 오픈데이터마트로 이루어진다(서울특별시 정보시스템담당관 2012). 서울시는 이러한 기본 구조 위에 기준정보 관리, 분류체계 관리, 정보공개 관리를 위한 오픈 아키텍처 DB와 데이터 서비스 목록을 위한 오픈데이터마트 DB를 구축했다. 오픈데이터마트 DB는 데이터 서비스를 위해 구축되었다. 서울시의 모든 소속기관과 출연기관에 산재되어 있는 행정정보 시스템의 데이터베이스로부터 서비스 대상 데이터세트를 통합하여 저장하는 리포지터리의 역할을 수행한다. PC, 모바일, 태블릿 등 다양한 기기의 OS와 브라우저 환경에 맞는 웹 서비스를 제공한다.

서울시는 데이터 서비스 제공을 위한 기본 아키텍처와 별도로 서비스 담당 공무원을 위한 '서울 열린 데이터 광장 내부업무관리시스템'을 구축하여 운영하고 있다. 이 시스템은 공공정보명 관리, 서비스 관리, 부서 및 계정 관리, 분류체계 관리, 인증키 관리 등의 서비스 관리 기능

www.kci.go.kr

과 서비스 현황을 도표 등 다양한 시각화 형태로 제공하는 통계 분석 기능으로 구성되어 있다. 이러한 서비스 관리 시스템은 서비스 담당 공무원이 통계자료를 분석하여 문제점을 파악하거나 전략 설정에 반영하도록 하여 서비스의 안정적 운영 및 발전을 돕는다.

이러한 노력에도 불구하고, 외국의 정부에서 활용하고 있는 소크라타나 CKAN플랫폼과 비교해보면 큰 한계를 볼 수 있다. 또한 다른 지방자치단체와의 연계성 또한 고려되어 서비스가 진행되어야 할 것이다.

4.5 시사점

서울시의 사례를 통해 살펴본 우리나라 데이터세트에 관한 현행 서비스 체계에 있어 가장 큰 문제점은 열린 데이터 광장 자체의 문제보다는 서비스의 근간이 되는 관련된 법이나 제도 등을 비롯하여, 전반적으로 데이터세트 기록물의 관리 체제가 마련되지 않은 것을 들 수 있다. 구체적으로는 먼저 장기적인 데이터세트 기록물의 보존 전략이 마련되지 않은 점과, 공개 대상 선정의 기준이 일반인 수요 조사에 한정되어 있다는 점, 그리고 데이터세트 기록의 식별부터 등록까지 이어지는 서비스 준비 단계의 절차나 방법이 명확하지 않은 점 등을 들 수 있다. 시각화나 분석 등 편리한 활용을 위한 기능의 구현 또한 중요하지만 이는 데이터세트 기록물의 관리체제와 철저한 서비스 준비 단계를 바탕으로 진행되어야 할 것이다. 그러므로 관련 법률 및 서비스 지침에서 데이터세트의 아카이빙과 서비스를 위한 요소들이 복합적으로 고려되어야 할 것이다.

또한 생산 및 관리 단계에서 데이터세트를 식별하고 관리계층을 구성하는 작업이 적절하게 이루어져야만 서비스 레벨에서 해결할 수 없는 문제들의 발생을 방지할 수 있다. 예를 들어 특정 기록물은 데이터베이스의 전체 단위가 하나의 데이터세트로 식별되어야 효과적으로 활용될 수 있고, 또 어떤 데이터세트는 테이블 단위로 추출되어야만 할 경우도 있을 것이다. 서비스 이전에 생산과 관리 단계부터 서비스를 전제로 한 프로세스를 밟아 관리하는 것이 전제되어야 할 것이다.

이러한 제도적인 환경이 준비되고, 이용자들 입장에서 다양하게 데이터들을 활용할 수 있는 도구가 제공되며, 효율적인 프로세스와 플랫폼이 마련되어 쉽고 편리한 서비스가 제공되기 위해서는 여러 가지 해결해야 할 일들이 산적해 있다. 다음 장에서는 서울시의 서비스 현황 분석을 통해 도출한 시사점을 바탕으로 데이터세트의 관리와 민간 활용을 위한 제도적 미비점을 지적하고 생산・관리 단계에서 고려해야 할 기록관리 측면에서의 방안을 제시하도록 한다. 데이터세트의 등록부터 실제 활용 단계에서의 시각화 및 분석 기능들은 소크라타, CKAN 등 해외 서비스 플랫폼의 사례를 비교 분석하여 정리해 본다.

5. 데이터세트의 서비스 방안

5.1 제도

데이터세트 기록물의 지속적인 서비스를 위해서는 먼저 제도적인 정비가 필요하다. 이를

www.kci.go.kr

위해 첫째로 데이터세트 아카이브 설립이 필요할 것이다. 데이터세트 아카이브는 현용 단계의 활용이 끝난 데이터세트 중 영구적 보존 가치가 있는 대상을 이관받아 보존 및 관리하는 기능을 수행한다. 미국의 데이터세트 아카이브인 AAD(Access to Archival Database)는 1834년부터 현재까지의 데이터세트를 보존하고 있으며, TNA의 NDAD(The National Digital Archive of Datasets) 역시 1963년부터 현재까지의 데이터세트를 관리하고 있다.[28] 이와 같이 국가기록원 산하에도 데이터세트 아카이브를 별도로 설립하여 데이터세트의 관리와 보존을 전담하도록 하는 것이 바람직하다. 또는 장기적 관점에서 NDAD의 경우처럼 웹 아카이브에 데이터세트를 포함하는 전략 또한 고려해 볼 수 있다. 웹 아카이브 구축은 웹 기술의 빠른 변화, SNS 등 형태의 다양성, 웹 크롤링 기술의 한계로 인하여 상당한 기술 개발이 요구되므로 공공데이터의 활용이 본격화되고 있는 우리나라의 현실을 고려할 때 데이터세트 아카이브를 별도로 구축한 이후 웹 아카이브로 통합하는 전략이 유용할 것이다. 이를테면 서울시가 데이터서비스를 위해 구축한 오픈 데이터마트는 현용 가치가 있는 데이터세트의 활용을 위한 신뢰성 있는 저장소로 인식되어야 하며, 데이터세트 아카이브로 이관되기 전까지 기록관의 관리 아래 진본성과 무결성을 유지하며 서비스되어야 할 것이다.

둘째, 2013년 7월 30일 제정된 '공공데이터의 제공 및 이용 활성화에 관한 법률(이하 "공공데이터 활성화법")'에 근거한 서비스체계 확립이 시급하다. 구체적으로는 정책 및 관리기구인 '공공데이터 활용 지원센터'와의 협력, 통합제공시스템인 공공데이터 포털로의 목록 등록 및 데이터세트 목록정보 공표, 데이터세트 제공, 그리고 품질관리 · 표준화 · 제공기반 구축 · 공공데이터 관련 교육 · 훈련시책 마련 등이다.

셋째, 정보공개법과 공공데이터활성화법의 관계를 명확히 하여 데이터세트 서비스를 제공할 때 법의 해석과 적용에 혼선이 없도록 해야 한다. 우선 공공정보로서의 데이터세트를 두 법령의 측면에서 어떻게 해석할 것인지가 중요하다. 행정정보 데이터세트는 기록물관리법상 "전자기록물"에 해당하며, 정보공개법상 "정보"에 포함되므로 정보공개 청구의 대상이 된다.[29] 또한 데이터세트가 포함하고 있는 내용이 정보공개법 제7조제1항 각호에 해당할 경우 행정정보 공표의 대상이 될 수 있다.[30] 하지만 데이터세트의 구조 및 내용적 특성으로 인해 정보공개 청구나 행정정보 공표의 형태로 서비스를 제공하기는 어려운 실정이다. 해외 사례에서는 공공

28) NDAD는 1997년부터 정부 기관의 데이터세트를 선별하여 맥락정보와 함께 물리적 이관을 수행해 왔으나 2010년 이후 모든 정부기관의 웹페이지 아카이브인 'UK Government Web Archive'의 일반적인 아카이빙 스케줄에 데이터세트를 포함시켰다. TNA는 웹 아카이브 구축을 위해 2004년부터 2009년까지 'UK Web Archive Consortium'을 조직하여 5개 협력기관과 함께 웹 아카이빙 기술을 개발하였다(TNA Information on Web Archiving [cited 2013.8.1]. 〈http://nationalarchives.gov.uk/webarchive/〉).

29) "정보"란 공공기관이 직무상 작성 또는 취득하여 관리하고 있는 문서(전자문서를 포함한다. 이하 같다), 도면, 사진, 필름, 테이프, 슬라이드 및 그 밖에 이에 준하는 매체 등에 기록된 사항을 말한다(정보공개법 제2조제1항).

30) 정보공개법 제7조에서는 국민생활에 매우 큰 영향을 미치는 정책정보, 대규모 예산투입사업, 행정감시, 그 밖에 공공기관의 장이 정하는 정보를 공표하도록 하고 있다.

www.kci.go.kr

데이터의 재사용에 관한 법률을 제정하여 제도적 기반을 마련하고 있으며,[31] 데이터세트를 적절한 형태로 시각화하고 활용할 수 있는 소크라타와 CKAN 등의 데이터 플랫폼을 통해 서비스를 제공하고 있다. 서울시 또한 이러한 사례를 벤치마킹하여 정보공개청구 사이트와 데이터세트 서비스 사이트인 열린 데이터 광장을 분리하고 있다.

공공데이터 활성화법에서는 "공공데이터"의 범위를 정보공개법 제9조 각항의 비공개대상정보와 저작권법 등의 법령에서 보호하고 있는 제3자의 권리가 포함된 것으로 정당한 이용허락을 받지 아니한 정보 등을 제외한 정보로 규정하며 기존 법률과의 정합성을 고려하고 있다. 나아가 기술적으로 분리가 가능한 경우에는 정보공개법 9조 각항에 해당하는 부분을 제외하고 제공할 수 있도록 하고 있으며 제3자의 권리가 포함된 공공데이터에 대하여 안전행정부장관이 이용허락 확보를 위한 방안을 제시할 수 있도록 하고 있다. 즉 기록물관리법상 보존 대상인 행정정보데이터세트는 정보공개법에 의해 서비스 대상이 되며 구체적으로는 공공데이터 활성화법에 기반하여 등록 등의 제공기반이 정의되는 것으로 이해할 수 있다.

넷째, 데이터세트 서비스 과정에서의 진본성과 무결성 확보 방안이 필요하다. 이번 법 제정을 계기로 가장 우려되는 점은 기록물관리법상 이관 대상인 데이터세트가 공공데이터 활성화법에 의해 서울시의 통합행정정보시스템인 오픈데이터마트나 공공데이터포털에 제공되는 과정에서 진본성 및 무결성 확보 방안이 고려되지 않은 점이다. 이러한 법령상의 미비점을 개선하기 위해서는 데이터세트가 기록관의 관리체제 하에서 표준 준수 및 품질관리를 통해 진본성, 무결성을 훼손하지 않도록 프로세스가 명시되어야 한다. 그리고 공공데이터포털로 데이터세트 제공 시 목록정보와 해당 데이터세트의 링크 및 메타데이터만을 제공하는 것이 바람직하다. 영국의 경우 DATA.GOV.UK는 기관 리포지터리로부터 해당 데이터세트의 링크를 연계하여 카탈로그 서비스만을 제공하고 있으며,[32] 미국의 경우 DATA.GOV에 제공된 연방정부 기관의 각 데이터세트에 적용된 NARA의 보유스케쥴을 반드시 준수하도록 하고 있다.[33]

다섯째, 기준정보 연계 방안의 마련이 필요하다. 데이터베이스로부터 추출되어 서비스를 위한 저장소로 이관되는 데이터세트 서비스의 특성상 생산 시스템의 원본 데이터세트와 서비스 시스템의 사본 데이터세트가 동일한 기준정보를 유지하는 것 또한 중요하다. 노스캐롤라이나주에서는 지역의 지리정보 데이터세트인 'NC OneMap'의 처분보유 스케쥴을 생산시스템인 NC OneMap 데이터베이스와 주 아카이브가 동일하게 유지하도록 하기 위해 반드시 하나의 시스템에서만 기준정보를 관리하도록 하고 있다. 또한 매년 해당 데이터세트의 보존 가치를 평가하고 갱신사항을 모니터링한 후 공익에 필요하다고 판단될 경우 '프레임워크 데이터세트'

31) 미국의 경우 공공정보의 활용에 관한 개별법은 없으나, FOIA에 의하여 연방정부가 보유하고 있는 공공정보를 자유롭게 이용할 수 있도록 함으로써 사실상 공공정보 이용의 법적 근거가 되고 있다(고동원 2010).

32) TNA Archiving Datasets. [cited 2013.8.1]. 〈http://www.nationalarchives.gov.uk/webarchive/archiving-datasets.htm〉.

33) DATA.GOV Data Policy Statements. [cited 2013.8.1]. 〈http://www.data.gov/data-policy〉.

www.kci.go.kr

로 구분하여 주 아카이브로 이관하여 관리하도록 지침을 정하고 있다.[34] 데이터세트의 보존기간, 폐기일정 등의 기준정보를 동일하게 유지하기 위해서는 각 사례들을 참고하여 적합한 방안을 채택해야 할 것이다.

마지막으로 데이터세트 서비스에 적합한 매체와 그렇지 않은 매체를 분리하여 적합한 서비스 형태를 개발해야 한다. 응용 프로그램이나 웹 서비스로 쉽게 개발될 수 있는 데이터세트 형태는 Open API로 제공하고, 텍스트 문서는 정보공개 사이트로, 멀티미디어 등은 기관 소셜미디어 등으로 구분하여 제공하는 것이 혼선을 줄일 수 있는 방법일 것이다. 그리고 활용도 기반으로 서비스가 평가되고 운영될 수 있도록, 이용 수요조사, 이용 현황조사 등의 절차가 포함되어야 한다. 서울시는 보유 공공정보 목록을 대상으로 민간 수요조사를 통해 공개 우선순위를 선정하였는데, 이러한 서비스의 효율성이 지속적으로 담보될 수 있는 프로세스를 법이나 규정에 명시하는 것이 공공데이터 서비스의 주요한 성공 요인이 될 것이다.

5.2 생산 및 관리

데이터세트의 생산 및 관리 단계는 효과적인 서비스 전략을 수행하는 발판이 된다. 데이터세트 기록물이 전자기록으로서의 진본성과 무결성을 유지하며 적합한 방법으로 서비스되기 위해서는 데이터세트 기록의 선별(selection) 및 공개대상 선정 등, 서비스 사이트 등록 이전의 각 서비스 준비 단계별로 고려해야 할 기록관리 측면의 요소들이 포함되어야 한다. 그래야 이들이 데이터로서가 아니라 행정업무의 결과물인 기록으로 가치를 가지고 서비스 될 수 있을 것이다.

민간에서의 활용 가능성과 경제적 가치 등은 선별에 있어 중요한 기준이 될 수 있다. 데이터세트의 선별 과정은 서비스 대상 데이터세트의 결정, 식별, 관리계층 구성, 등록, 획득 단계로 이루어진다(조은희, 임진희 2009). 소크라타 플랫폼에서의 데이터세트 등록 과정 역시 서비스 대상 결정, 데이터 정화, 메타데이터 처리의 단계를 거친다. 이러한 프로세스를 거쳐 오픈 데이터 플랫폼에 데이터세트가 등록되어야 데이터세트 기록물의 관리와 서비스를 동시에 도모할 수 있다.

서울시의 경우를 보면 '오픈데이터마트'로 통합된 행정데이터들은 '열린 데이터 광장' 내부 업무관리시스템 안에서 목록화되어 서비스 대상 데이터세트가 결정된다. 내부 업무관리시스템에서는 공공 데이터 현황조사, 분류체계 결정, 서비스 유형 재분류 등의 프로세스를 거친다. 여기에서 고려해야 할 기록 관리 측면의 요소들은 다음과 같다.

먼저 기록관리시스템과 연계된 공개 및 선별 지침의 마련이 필요하다. 세부 사항으로는 수요도 조사 및 인터뷰의 주기적 실행, 데이터세트별 활용도를 반영할 수 있는 공개기준의 적용, 기록관리시스템과 열린 데이터 광장 내부업무관리시스템의 기준정보 연계 방안 등이 있을 것

34) North Carolina Geographic Information Coordinating Council. Guide to the Records Retention and Disposition Schedule for NC OneMap Datasets. [cited 2013.8.1]. 〈http://www.ncgicc.org/Portals/7/documents/NCOneMap_RecordsRetentionSchedule_20120302.pdf〉.

www.kci.go.kr

이다. 다음으로 각 행정정보시스템의 시스템 구조, 메뉴 구성, 업무의 범위와 처리 방식 등에 대해 상세히 기술하여 이를 데이터세트의 기록화 및 공개 여부를 판단하는 근거 자료로 활용해야 할 것이다.

또한 데이터세트를 기록으로 관리하고 활용하기 위해서는 데이터세트의 어느 범주를 하나의 기록 건으로 볼 것인지 식별하여 확정하는 것이 필요하다. 데이터세트의 식별 기준으로는 기록관리 생애 주기 및 관리단계에 따라 통제 및 관리가 가능한 단위여야 하며, 기술・분류・처분 등을 수행할 때 계층별로 논리적 연관성을 가져서 하위 계층에서 상위계층의 속성을 상속받는 등 설계 및 수행에 현실적인 편의성을 제공할 수 있어야 한다. 또한 업무적・논리적・물리적으로 연관성을 가진 데이터세트가 범주화될 수 있어야 한다(조은희, 임진희 2009).

데이터세트의 관리 계층 설정이 끝나면 물리적 실체로 변환하는 획득 과정을 거쳐야 할 것이다. 데이터세트는 단위과제의 특성별로 해당 데이터세트의 추가, 갱신, 종결의 시점이 달라지므로 각각에 맞는 획득전략이 수행되어야 한다(Ashley 2004). 시카고 시 정부에서는 경찰국으로 접수된 12년간의 범죄신고 데이터를 1년 단위로 구분하여 추가로 제공하고 있다. 시애틀 시 정부는 911 화재신고 데이터를 기간에 따라 지난 24시간, 실시간, 연도별 등으로 제공하고, 발생지역이나 신고 유형에 따라 시내, 외곽, 화재신고, 엘리베이터 구조 등으로 세분하여 제공하고 있다. 이러한 서비스가 가능하도록 하기 위해서는 하나의 데이터세트를 다양한 유형의 기록 건으로 식별하고 획득하는 전략이 요구된다. 두 사례들을 통해 데이터세트와 테이블의 중간 계층으로 서브 데이터세트가 추가되는 개념을 생각해 볼 수 있다. 5백만 개가 넘는 12년 동안의 누적 레코드를 하나의 데이터세트로 제공받기 원하는 사용자도 있을 것이고, 특정 연도의 데이터만 분석하려는 사용자도 있을 것이기 때문이다.

민간에서 보유하고 있는 데이터세트의 획득 또한 고려해야 한다. 이를 위해서는 해당 데이터세트의 품질 개선을 위한 도구를 제공하고, 재배포와 재사용을 위해 해당 기관의 저작권 및 라이선스 정책에 동의하는 과정이 포함되어야 할 것이다. 또한 동일한 분류체계와 선별 기준을 적용하여 활용될 수 있도록 프로세스를 마련하고, 획득하는 시점의 결정 또한 그에 따라 이루어져야 할 것이다.

아카이브에서의 선별 과정이 끝나면 서비스 사이트로 등록을 하게 된다. 등록은 선별된 데이터세트를 서비스 가능한 새로운 형태의 데이터세트로 생성하여 업로드하는 것을 의미한다. 등록된 데이터세트에 대해서는 사용자나 서비스 담당자가 데이터세트 변환 시 특정 행이나 열만을 추출하거나, 복수의 데이터세트를 병합하는 기능이 제공되어야 한다. 등록 과정에는 데이터세트의 오류사항을 점검하고 형식을 결정하는 작업, 메타데이터나 기술요소를 점검・관리하는 작업 등의 처리과정이 포함되어야 한다. 데이터세트의 선별 과정은 전문적인 지식과 많은 노력이 필요하지만, 생성과 등록 과정은 오픈 데이터 플랫폼을 통해 자동화할 수 있는 솔루션을 제공하는 것이 가능하다. 각 기관의 기록관은 데이터세트의 논리적 식별을 통한 선별 작업을 주관하고, 선별 이후의 등록 과정은 서비스를 담당하는 정보화 부서 등에서 수행하

www.kci.go.kr

여도 무방하다.

또한 자치단체가 추구하는 정책의 방향과 시민들이 원하는 공공정보의 수요가 조화된 맞춤형 데이터세트 개발이 필요하다. 해외 사례에서 보았듯이 시카고는 교통이나 관공서 위치, 시애틀은 공공안전, 헬싱키는 GIS정보 등 자치구의 지리적, 문화적 특성별로 요구되는 서비스 또한 다르다. 지속적인 수요조사와 데이터세트 활용 현황 분석 결과가 신규 서비스에 반영되는 체계를 갖추어야 할 것이다.

5.3 서비스

데이터세트 활용 서비스를 위한 도구로는 시각화, 분석 및 관리도구, 자동화, 공유 및 참여를 위한 도구를 들 수 있다. 데이터세트의 시각화는 일반 사용자들이 웹을 통해 좀 더 편리한 방식으로 공공 데이터를 이용할 수 있도록 해 준다. 행정정보시스템이나 서비스 전용 리포지터리에 물리적으로 존재하는 데이터세트는 XML 형태로 웹에 파싱하여 출력하는 방식으로 구현된다. 일반적인 시각화 유형으로는 스프레드시트, 도표, 지도, 캘린더 형태가 있다. 소크라타 플랫폼에서는 데이터세트의 속성에 따라 시각화 유형이 활성화된다. 예를 들어 스프레드시트에 '날짜(date)'나 '텍스트(text)' 컬럼이 포함된 경우 캘린더 형태로 변경할 수 있다. 주소나 위도, 경도 등의 컬럼이 포함된 데이터세트의 경우 지도 유형의 시각화가 적합하다. 다양한 지도 API를 확보하고, 포인트 맵, 바운더리 맵, 히트 맵 등 데이터세트의 내용을 가장 잘 표현하는 방식으로 변경할 수 있는 효과적인 시각화 도구가 제공되는 것이 바람직하다.

분석 및 관리 도구는 공개한 데이터세트의 사용 현황을 파악하고 새로운 데이터세트를 개발하는 데 사용될 수 있다. 기관 서비스 담당자, 개발자, 일반 사용자를 위해서는 각각 다른 관리 도구가 제공되어야 할 것인데, 예를 들면 기관 서비스 담당자나 데이터세트 제공자에게는 서비스 현황을 분석하고 공개목록을 관리하는 기능이 필요할 것이다. 그래서 예를 들어 서울시의 열린 데이터 광장이 플랫폼으로 기능하려면 내부업무시스템의 관리 인터페이스를 민간 데이터세트 제공자에게도 제공할 필요가 있다. 개발자를 위한 관리 도구는 API 테스트 및 응용프로그램 관리와 분석을 위한 것이다. 서울시의 열린 데이터 광장 역시 개발자를 위한 사이트를 별도로 운영하고 있으나, 소크라타나 CKAN처럼 체계적인 API 명세 제공 및 테스트, 응용프로그램 등록 등 실질적인 지원보다는 질문답변 게시판 형태의 개발자 커뮤니티를 제공하는 수준에 머물러 있다. 개발자들이 데이터세트를 창의적으로 활용하여 수준 높은 응용 서비스로 개발되도록 적극 지원할 필요가 있다. 또 일반 사용자가 데이터세트를 열람하는 과정에도 분석 도구가 제공되어야 한다. 여러 데이터세트를 효과적으로 필터링하고 상관관계를 분석하는 기능을 제공할 수 있다. 해외 데이터 플랫폼들은 SPSS 등 통계 패키지가 제공했던 분석도구의 기능을 웹상에서 제공하여 사용자로 하여금 데이터를 다른 관점으로 쉽게, 또한 선택적으로 추출하도록 하고 있다.

지방자치단체에서 보유 데이터세트를 공개하고 서비스하기 위해서는 기술적 지원이 필요하다. 즉 데이터세트의 오류를 체크하여 정제하고, 메타데이터나 기술정보를 검토하는 단계를 거

www.kci.go.kr

쳐 Open API로 변환하는 기능을 제공해야 한다. 또한 GIS데이터 등 비정형 데이터를 축적한 경우, 지도 형태로 매쉬업해 주는 도구 또한 유용할 것이다. 국가적 데이터 플랫폼인 공공데이터포털 등이 솔루션을 개발하여 지방자치단체에 보급하는 방식이 현실적일 것이다.

Web 2.0의 개념과 도구들은 공공정보 서비스에 접목되어 사용자들이 쉽게 공유하고 편리한 방식으로 참여할 수 있도록 해 준다. 해외 사이트에서 데이터세트의 효과적인 활용을 위해 제시되는 방식은 Open API 제공을 통한 응용프로그램 개발, 댓글 및 리뷰, 위젯, RSS, 소셜 네트워크 사이트의 API와 결합한 소셜 연계 서비스, 데이터 포털에 개설된 주제별 커뮤니티, 링크드 데이터 구축 등이 있다. 오픈 데이터 플랫폼은 다양한 형태와 창구를 통해 공공정보가 유통될 수 있도록 풍부한 API 기능을 제공하여 재사용의 효율성을 높여야 한다.

5.4 플랫폼

미국의 연방정부 및 지방자치단체들은 공공정보 확산 플랫폼인 소크라타를 사용하여 효율적으로 서비스를 연계하고 있다. 또한 미국 정부는 각국의 오픈 거버먼트 운동을 지원하기 위해 2013년 1월 오픈소스 플랫폼인 OGPL (Open Government Platform) 베타버전을 공개하였다.[35] OGPL 베타버전은 정부 기관에 산재되어 있는 공공정보를 모으고 기본적인 서비스 체계를 제공하는 데에는 무리가 없는 수준이다. 영국 역시 오픈소스 플랫폼인 CKAN을 DATA.GOV.UK에 적용하였고 유럽을 중심으로 40여개 정부가 플랫폼으로 채택하여 사용하고 있다.

서울시를 비롯한 우리나라 지방자치단체는 공개된 오픈소스 플랫폼을 도입하는 방안, 유료 패키지로 도입하는 방안, 독자적 플랫폼을 개발하는 방안 등을 검토할 수 있겠다. 서비스 페이지는 독자적으로 개발하고, 데이터세트를 수집하고 공유하는 기능만을 CKAN으로 개발하는 커스터마이징하는 복합적 전략 또한 고려할 수 있을 것이다.

〈표 1〉은 데이터세트의 등록 및 저장, 분석 및 시각화, 공유 및 재배포, 사용자 참여 및 웹 출판 분석, 자동화, 접근성, 확장성 등의 기능과 데이터세트 유형별로 각 플랫폼의 제공 기능을 비교한 것이다. 소크라타는 가장 많은 기능을 제공하고 있으며, 시각화나 온라인 분석 툴, 매쉬업 등 자동화 도구 기능이 잘 정비되어 있다. CKAN은 CMS(컨텐츠관리시스템)와 연계가 가능하고 링크드 데이터를 지원하여 확장성이 뛰어나다. OGPL은 데이터의 등록 및 저장, 분석 및 시각화 기능 등 서비스에 꼭 필요한 기본 기능에 충실한 것을 알 수 있다.

공공데이터 부문의 PaaS(Platform as a Service), 즉 '서비스로서의 플랫폼'이라는 것은 정부가 제공하는 데이터 시장에 공공정보 제공자와 수요자가 모여 마음껏 활용하고 참여할 수 있는 정보 생태계를 실현하는 것이다. 이러한 환경이 이루어지려면 〈표 1〉의 각 서비스 단계별 기능뿐만 아니라 제도의 정비, 정부 기관들 간의 연계 등을 종합적으로 고려해야 한다.

35) Open Government Platform - Beta. [cited 2013.7.19]. 〈https://github.com/opengovplatform/opengovplatform-beta〉.

www.kci.go.kr

〈표 1〉 오픈 데이터 플랫폼 기능 비교

			소크라타 (Socrata)	CKAN	OGPL	서울시
기능	등록 및 저장	데이터세트 정제 및 서식 변경	○	○		
		메타데이터 처리	○	○	○	
		데이터세트 생성	○		○	
	분석 및 시각화	편집	○	○	○	
		필터링	○	○	○	○
		시각화	○	○	○	○
		온라인 분석	○	○	○	
	공유 및 재배포	웹페이지 삽입	○	○	○	
		댓글, 리뷰	○	○	○	
		웹 퍼블리싱	○	○	○	
		컨텐츠 재배포	○	○	○	
		소셜 미디어 통합	○	○	○	
	사용자 참여 및 웹 출판 분석	데이터세트 제공자용 분석도구	○	○	○	○
		개발자용 분석도구	○	○	○	
		사용자 커뮤니티 관리도구	○	○	○	
	자동화	Open API 자동 생성도구	○	○	○	
		매쉬업	○			
	접근성	확장성	○	○		
		검색엔진 친화성(SEO-Friendly)	○			
		장애인 접근성 제공	○			
		RSS	○	○		
		위젯	○	○		
		링크드 데이터		○		
		로그인 필요				○
		사용자 승인 필요				○
	확장성	CMS(컨텐츠관리시스템) 호환성		○	○	
		소스 코드 라이브러리	○	○	○	
		플랫폼 다운로드		○	○	
유형	데이터세트 유형	스프레드시트(Sheet)	○	○	○	○
		도표(Chart)	○	○	○	○
		지도(Map)	○	○	○	○
		파일(File)	○	○	○	○
		링크(Link)	○	○	○	○
		일정(Calendar)	○			
		서식(Forms)	○			
		Open API	○	○	○	○

www.kci.go.kr

6. 결 론

각국 정부는 공공정보를 개방하여 새로운 경제적·사회적 가치를 만들어 내고 있다. 또한 중앙정부 및 각 정부 기관들이 동일한 공유 플랫폼을 사용하여 개발 비용을 절감하고, 공공데이터의 효율적인 수집 및 확산을 도모하고 있다. 우리나라도 안전행정부의 공공데이터포털이 공공자원 확산의 포털 역할을 하고 있다. 하지만 민간 영역에서 재사용하기 위한 제도적 기반이 뒤늦게 마련되었고 공개된 데이터의 활용 가능성 부족 및 이용의 불편함 등으로 서비스의 효율성은 부족한 실정이다. 서울시는 국내 지방자치단체 중 선도적으로 공공정보 서비스를 제공하고 있다. 열린 데이터 광장은 선진국의 사례를 벤치마킹하여 개발된 자체 서비스 시스템으로 원천 데이터세트를 스프레드시트, 도표, 지도, Open API 등 다양한 형태로 제공하고 있다. 서울시 뿐만 아니라 경상남도, 전주시, 대구시 등의 지방자치단체들도 보유하고 있는 공공데이터의 원문을 공개하거나 Open API로 개발하며 데이터세트 활용 서비스에 관심을 갖고 있다(경상남도 2012).

이러한 상황에서 본 연구는 지방자치단체에서 데이터세트 기록물을 효과적으로 선별하고 활용하기 위한 방안을 제시하였다. 외국에서 데이터세트 공공서비스를 위해 가장 많이 활용되는 오픈 데이터 플랫폼인 소크라타와 CKAN을 통해 서비스 플랫폼의 역할과 기능을 살펴보았고, 서울시의 데이터 서비스 현황과 비교하여 시사점을 분석해 보았다. 분석 과정을 통해 제도, 생산 및 관리, 서비스, 플랫폼의 측면에서 데이터세트 서비스 방안을 제안하였다.

그 방안으로는 우선 공공데이터활성화법에 기반을 둔 서비스 체계 정비와 데이터세트 아카이브 설립의 필요성을 제시했다. 그리고 데이터세트의 선별부터 실제 활용되기까지 각 단계별로 필요한 기능과 관리사항들을 명시하였다. 플랫폼의 측면에서는 데이터세트 관리를 기록관의 관리체계로 편입시켜 전자기록으로서의 진본성과 무결성을 확보해야 함을 강조했다. 그리고 현용 단계 이후에도 영구적 보존 가치를 지닌 데이터세트가 데이터세트 아카이브로 이관되어 관리가 이루어지는 통합 체제가 필요하다는 것을 역설하였다.

본 연구의 의의는 대표적인 오픈 데이터 플랫폼의 기능을 분석하고 확산 플랫폼으로 발전하기 위해 이루어져야 할 내용을 서비스 단계별로 제시함으로써, 데이터세트의 관리와 서비스에 대한 통합적 방안을 개념화한 것이다. 본 연구의 한계로는 데이터세트의 아카이빙에 대한 구체적 방법론을 제시하지 못한 점을 들 수 있다. 향후 데이터세트 서비스 과정에서 진본성과 무결성을 확보하기 위한 구체적 방안이 연구되어야 할 것이다. 또한 보존기간 이후에도 꾸준한 Open API 호출이 발생하는 데이터세트의 기준정보 변경이나 데이터세트 아카이브 이관 후의 서비스 연계 방안도 제시되어야 할 것이다.

www.kci.go.kr

참 고 문 헌

경상남도. 2012.「지방자치단체 공공정보 개방과 민간활용 증진방안에 관한 연구 - 경상남도 보유 정보를 중심으로」. [cited 2013.8.15]. 〈http://www.mopas.go.kr/gpms/view/jsp/download/userBulletinDownload.jsp?userBtBean.bbsSeq=1012663&userBtBean.ctxCd=1002&userBtBean.orderNo=4〉.

고동원, 권영준. 2010. 공공정보 이용의 활성화를 위한 법적 개선 방안.「성균관법학」23(2), 263-287.

국가기록원. 2007.「행정정보시스템 데이터세트 기록관리」. 대전: 국가기록원. [cited 2013.8.15]. 〈http://www.mopas.go.kr/gpms/view/jsp/download/userBulletinDownload.jsp?userBtBean.bbsSeq=1012663&userBtBean.ctxCd=1002&userBtBean.orderNo=4〉.

국가기록원. 2010. 행정정보시스템 전자기록의 체계적 관리 체계 마련.「e-기록 속으로」제4호.

김남경. 2012. 공공정보 데이터세트의 아카이빙 방안 연구. 명지대학교 기록정보과학전문대학원 석사학위논문.

김유승. 2010. 거버먼트 2.0 기반의 정보공개제도 개선방안에 대한 연구.「기록학연구」25, 197-231.

김유승, 전진한. 2011. 거버먼트 2.0 기반의 공공정보서비스에 관한 연구.「한국기록관리학회지」11(1), 47-66.

김포옥, 윤수영. 2007. 데이터세트 기록물의 기술요소에 관한 연구.「한국비블리아학회지」18(2), 39-59.

남영준, 서만덕. 2009. 정부 여론조사자료 아카이브 구축방안에 관한 연구.「한국기록관리학회지」9(1), 175-196.

서울특별시. 2012.「개방, 공유, 소통의 열린 시정 2.0 종합 추진계획(안)」. [cited 2013.7.4]. 〈http://gov20.seoul.go.kr/files/2012/12/50c1df078cb5b7.44633035.pdf〉.

서울특별시 정보시스템담당관. 2012.「서울 열린 데이터 광장 사례를 중심으로 본 서울시 오픈 데이터 전략과 숙제」. [cited 2013.7.4]. 〈http://webst.kaist.ac.kr/2012opendata/pdf/yim.pdf〉.

서울특별시. 서울 열린 데이터 광장. [cited 2013.7.4]. 〈http://data.seoul.go.kr/〉.

서울특별시. 서울 열린 데이터 광장 서비스 이용약관. [cited 2013.7.11]. 〈http://data.seoul.go.kr/openinf/openplaza/agreement.jsp〉.

윤은하. 2012. 기록정보 서비스와 정보공개.「한국기록관리학회지」12(2), 163-179.

이순한. 2008. 조달업무의 설명책임성 확보를 위한 조달정보 기록관리 요건 연구. 한국외국어대학교 대학원 석사학위논문.

이영숙. 2007. 공공기록물 정보서비스 발전방안에 관한 연구.「한국기록관리학회지」7(2), 73-94.

이은별. 2008. 국가 재정정보의 기록학적 관리방안 연구. 명지대학교 기록정보과학전문대학원 석사학위논문.

임진희, 조은희. 2010a. 데이터세트 기록의 이관도구 기능요건 연구.「한국정보관리학회 학술대회 논문

www.kci.go.kr

집」, 155-162.
임진희, 조은희. 2010b. 행정정보 데이터세트 기록 이관 시 데이터 보정 및 품질 개선 방법 연구.「기록학연구」 25, 91-129.
조은희, 임진희. 2009. 행정정보 데이터세트 기록의 선별 기준 및 절차 연구.「기록학연구」 19, 251-291.
최정윤. 2009. GIS데이터의 기록관리 방안 연구. 명지대학교 기록정보과학전문대학원 석사학위논문.
한국기록학회. 2008.「기록학 용어 사전」, 서울: 역사비평사.
한철희. 2007. NEIS 교무업무시스템 데이터의 기록화방안 연구(학교생활기록부를 중심으로). 명지대학교 기록정보과학전문대학원 석사학위논문.
현문수. 2005. 데이터세트 기록의 관리 방안.「한국기록관리학회지」 5(2), 103-124.
Ashley. 2004. The preservation of databases. *VINE* 34(2), 66-70.
City of Chicago Data Portal. [cited 2013.7.5]. 〈https://data.cityofchicago.org〉.
CKAN. [cited 2013.7.7]. 〈http://ckan.org/〉.
DATA.GOV Communities. [cited 2013.7.15]. 〈http://www.data.gov/cities/community/cities〉.
DATA.GOV.UK. [cited 2013.7.13]. 〈http://data.gov.uk/〉.
DATA.GOV Data Policy Statements. [cited 2013.8.1]. 〈http://www.data.gov/data-policy〉.
Data.seattle.gov. [cited 2013.7.5]. 〈https://data.seattle.gov〉.
Drupal. [cited 2013.7.17]. 〈https://drupal.org/about〉.
Dutch National Archives and the Dutch Ministry of the Interior and Kingdom Relations. 2003. Digital Preservation Testbed. From Digital Volatility to Digital Permanence: Preserving Databases. [cited 2013.7.2]. 〈http://www.ltu.se/cms_fs/1.83816!/file/Preserving%20Databases.pdf〉.
Faniel, Ixchel M., & Yakel, Elizabeth. 2011. "Significant Properties as Contextual Metadata." *Journal of Library Metadata* 11(3-4), 155-165.
Faniel, Ixchel, Kriesberg, Adam, & Yakel, Elizabeth. 2012. Data Reuse and Sensemaking among Novice Social Scientists. ASIST 2012 Annual Meeting.
HELSINKI REGION INFOSHARE. [cited 2013.7.14]. 〈http://www.hri.fi/fi/〉.
Korea Data Hub. [cited 2013.7.13]. 〈http://thedatahub.kr/〉.
North Carolina Geographic Information Coordinating Council. 2012. Guide to the Records Retention and Disposition Schedule for NC OneMap Datasets.
Office of E-Government and IT, Office of Management and Budget, Federal Chief Information Officers Council, 2010. Data.gov Concept of Operations. 14 November 2010 - Version 1.0. [cited 2013.7.3]. 〈http://www.data.gov/sites/default/files/attachments/data_gov_conops_v1.0.pdf〉.

www.kci.go.kr

Open Government Platform - Beta. [cited 2013.7.19]. 〈https://github.com/opengovplatform/opengovplatform-beta〉.

Pearce-Moses, Richard. [2013]. A Glossary of Archival and Records Terminology. Society of American Archivists. [cited 2013.7.1]. 〈http://www2.archivists.org/glossary〉.

Publicdata.eu. [cited 2013.7.12]. 〈http://publicdata.eu/〉.

Shepherd, Elizabeth, & Smith, Charlotte. 2000. "The Application of ISAD(G) to the Description of Archival Datasets." *Journal of the Society of Archivists* 21(1), 55-86.

Socrata, Inc. [cited 2013.7.13]. 〈http://www.socrata.com〉.

Socrata API Foundry. [cited 2013.7.13]. 〈http://www.socrata.com/api-foundry/〉.

Socrata Mondara. [cited 2013.7.13]. 〈http://www.socrata.com/mondara/〉.

The National Archives. Public sector information Directive and Regulations. [cited 2013.7.7]. 〈http://www.nationalarchives.gov.uk/information-management/legislation/directive-and-regulations.htm〉.

The Open Knowledge Foundation. CKAN. [cited 2013.7.7]. 〈http://ckan.org〉.

TNA Archiving Datasets. [cited 2013.8.1]. 〈http://www.nationalarchives.gov.uk/webarchive/archiving-datasets.htm〉.

KS X ISO 15489-1:2007. 문헌정보 - 기록관리 - 제1부: 일반사항.

KS X ISO 15489-1:2007. 문헌정보 - 기록관리 - 제2부: 지침.

[관련법령]

「공공기관의 정보공개에 관한 법률」.

「공공기록물관리에 관한 법률 시행령」.

「공공데이터의 제공 및 이용 활성화에 관한 법률」.

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

Cho, Eun Hee, & Yim, Jin-Hee. 2009. "A Study on Record Selection Strategy and Procedure in Dataset for Administrative Information." *The Korean Journal of Archival Studies* 19, 251-291.

Choi, Jong Youn. 2009. *Archival Management Scheme for GIS Data*. Master's thesis, Myoungji University.

www.kci.go.kr

Gyeongsangnam-do. 2012. *A Study on Disclosure and Improvement of Private Use of Public Sector Information in Local Government- Focused on Gyeongsangnam-do owned data*. [cited 2013.8.15]. 〈http://www.mopas.go.kr/gpms/view/jsp/download/userBulletinDownload.jsp?userBtBean.bbsSeq=1012663&userBtBean.ctxCd=1002&userBtBean.orderNo=4〉.

Han, Chul Hee. 2007. *A Study on Documentation of NEIS School Affairs System Related with Cumulative Records*. Master's thesis, Myoungji University.

Hyun, Moonsoo. 2005. "A Study on the Management of Dataset as Records." *Journal of Korean Society of Archives and Records Management* 5(2), 103-124.

Information System Manager of Seoul Metropolitan Government. 2012. *A Challenges of the Seoul Metropolitan Government Open Data Strategy focused on Seoul Open Data Portal*. [cited 2013.7.4]. 〈http://webst.kaist.ac.kr/2012opendata/pdf/yim.pdf〉.

Kim, Nam Kyung. 2012. *A Study on Archiving Plans for Dataset of the Public Sector Information*. Master's thesis, Myoungji University.

Kim, Po-Ok, & Yun, Soo-Young. 2007. "A Study on the Description of Archival Datasets." *Journal of the Korean Biblia Society for Library and Information Science* 18(2), 39-59.

Kim, You-Seung, & Jeon, Jin-Han. 2011. "A Study on Public Sector Information Services in the Government 2.0 Era." *Journal of Korean Society of Archives and Records Management* 11(1), 47-66.

Kim, You-Seung. 2010. "A Study on Freedom of Information in the Government 2.0 Era." *The Korean Journal of Archival Studies* 25, 197-231.

Ko, Dong-Won, & Kwon, Young-June. 2010. "Legal Review on Re-use of Public Sector Information in Korea." *Sungkyunkwan Law* 23(2), 263-287.

Korean Society of Archival Studies. 2008. *The Glossary of Archival Terms*. Seoul: Yeoksabipyongsa.

Lee, Eun Byol. 2008. *Archival Management Scheme for the National Financial Information*. Master's thesis, Myoungji University.

Lee, Soon-Han. 2008. *A Study on Recordkeeping Requirements of Procurement Information for Securing Accountability of Procurement Business*. Master's thesis, Hankuk University of Foreign Studies.

Lee, Young-Sook. 2007. "A Study on Development Directions for Public Archival Information Service." *Journal of Korean Society of Archives and Records Management* 7(2), 73-94.

Nam, Young-Joon, & Seo, Man-Deok. 2009. "A Study on Archiving of Government Survey Data." *Journal of Korean Society of Archives and Records Management* 9(1), 175-196.

www.kci.go.kr

National Archives of Korea. 2007. *The Research Report for Management of Dataset Record in Administrative Information System*. [cited 2013.8.15]. 〈http://www.mopas.go.kr/gpms/view/jsp/download/userBulletinDownload.jsp?userBtBean.bbsSeq=1012663&userBtBean.ctxCd=1002&userBtBean.orderNo=4〉.

National Archives of Korea. 2010. [Webzine] "Development of Organized Management System for Electronic Records in Administrative Information System." e-Girok Soguro 4.

Seoul Metropolitan Government. Seoul Open Data Portal. [cited 2013.7.4]. 〈http://data.seoul.go.kr/〉.

Seoul Metropolitan Government. User Agreement of Seoul Open Data Portal. [cited 2013.7.11]. 〈http://data.seoul.go.kr/openinf/openplaza/agreement.jsp〉.

Seoul Metropolitan Government. 2012. A Comprehensive Plan for the Advancement of Government 2.0 with Openness, Sharing, Communication. [cited 2013.7.4]. 〈http://gov20.seoul.go.kr/files/2012/12/50c1df078cb5b7.44633035.pdf〉.

Yim, Jin-Hee, & Cho, Eun Hee. 2010a. "A Study on the Functional Requirements of Migration Tool for Dataset Records." *Korea Society for Information Management Symposium Journal*, 155-162.

Yim, Jin-Hee, & Cho, Eun Hee. 2010b. "A Study on Data Adjustment and Quality Enhancement Method for Public Administrative Dataset Records in the Transfer Process." *The Korean Journal of Archival Studies* 25, 91-129.

Youn, Eunha. 2012. "Archival Reference Service and Freedom of Information in Korea." *Journal of Korean Society of Archives and Records Management* 12(2), 163-179.

www.kci.go.kr