

도서관 분야 데이터 개방 현황과 개선방안 연구

A Study on Library Data Open Status and Improvement Strategies

김혜선, 한국과학기술정보연구원, hskim@kisti.re.kr

김완종, 한국과학기술정보연구원, wgkim@kisti.re.kr

Hye-Sun, Kim, Korean Institute of Science and Technology Information

Wan-Jong, Kim, Korean Institute of Science and Technology Information

국내의 공공기관들은 정부 3.0 정책에 의해 다양한 분야의 데이터를 민간에 개방하고 있다. 본 연구는 국내외의 주요 데이터 개방 사이트를 중심으로 도서관 분야 데이터 개방현황을 살펴보았다. 조사 대상 사이트는 서울 열린데이터 광장, 공공데이터 포털, Data.gov 이다. 이 사이트들에서 나타난 문제점을 중심으로 오픈 데이터의 활용성 제고를 위한 개선방안을 제안하였다.

1. 서론

1.1 연구목적과 내용

2009년 미국 오바마 정부의 Open Government 추진 이후 우리나라에서도 2013년부터 공공기관을 중심으로 정부 3.0 정책을 활발하게 추진하고 있다. 교육, 환경, 기상 등 다양한 주제분야의 데이터들이 데이터셋, 오픈API 등의 형식으로 개방되고 있다.

도서관 분야도 예외는 아니다. 도서관 소장 정보, 소장목록, 신착도서, 문화프로그램 등을 민간에 개방하고 있다. 이런 상황에서 도서관 분야에서는 어떤 데이터들을 어떻게 개방하고 있으며, 향후 도서관 데이터를 활용한 새로운 서비스와 산업 창출을 위해 개선할 부분은 무엇인지 점검해 볼 필요가 있다.

본고는 국내외의 대표적인 데이터 개방 사이트를 대상으로 첫째, 각 사이트의 전반적 특징을 조사하였다. 둘째, 도서관 분야 개방 데이

터의 유형, 최신성, 제공방식 등을 살펴보고, 셋째, 데이터 활용 촉진을 위한 개선방안을 제안하였다.

1.2 연구대상 및 방법

조사 대상은 서울시의 서울 열린데이터 광장, 한국정보화진흥원의 공공데이터포털, 미국의 Data.gov 이다. 이 사이트들에서 “도서관” 혹은 “공공도서관”으로 검색했을 때 제공하는 데이터를 살펴보았다. 각 사이트들을 2016년 초반부터 2~3개월의 시차를 두고 2회 이상씩 조사하였다. 제조사시, 개방 데이터의 양은 조금씩 늘었지만 특징은 유사하였다. 따라서 초기 방문시 조사결과를 중점 기술하였고, 데이터 개방 형식 중에서 데이터셋을 주로 살펴보았다.

2. 도서관 데이터 개방현황 조사

2.1 서울 열린데이터 광장

2.1.1 사이트 개요

서울특별시가 운영하는 사이트로써, 공공 데이터를 민간에 개방하고 소통함으로써 공익성, 업무효율성, 투명성을 높이고 시민의 자발적 참여로 새로운 서비스와 공공의 가치를 창출하도록 지원하데 목적을 둔다. 2016년 8월 16일 기준 4,510건의 데이터셋, 3,230건의 오픈API를 개방하고 있다(서울열린데이터광장, 2016).

데이터는 일반행정, 문화관광, 교통 등 주제별로 제공하며, 시민의 관심을 끌 만한 데이터를 분석하여 그 결과를 시각화하여 제공한다.

2016년 1월 13일 “도서관”으로 검색시 총 70건이 검색되었다. 70건은 데이터셋 35건, 카탈로그 9건, 활용갤러리 13건, 이용활용 문의 13건으로 이루어졌다. 이후 2016년 4월말, 8월 중순 재방문하여 동일 질의어로 검색시 각각 75건과 77건의 데이터가 검색되었다.

데이터셋은 sheet, 오픈API, map, file 등 여러 형식으로 개방되며 sheet, 오픈API가 다수를 차지하고 있다. sheet는 xls, csv, json 방식으로 다운로드 가능하다.

2.1.2 도서관 데이터 개방 현황

데이터셋 35건을 한건씩 살펴본 결과 최신성, 지속성, 원본데이터 출처, 유사데이터 중복 제공 등이 문제로 나타났다.

구체적으로 보면 첫째, 서울특별시가 데이터 개방주체가 되어 서울시 공공도서관 전반에 대해 안내하는 경우가 가장 많았고, 특정 도서관에 대한 데이터가 그 다음 순 이었다. 둘째, 개방데이터는 도서관 소재지 및 현황 안내, 통계, 장서목록, 게시판, 소장정보 검색 API에 대한 것이었다. 도서관 소재지 및 현황

안내가 15건으로 가장 많았고, 도서관 장서목록은 5건이고 일회성 제공인 경우가 많았다. 즉, 특정시점의 목록을 한번 제공하고 그치는 경우가 대부분이었다. 소장정보 검색 API 2건은 특정 도서관 장서에 대한 것이었다. 통계는 단순 성과통계로 활용성이 낮아 보였다. 셋째, 데이터 등록일은 2012년 6건, 2013년 12건, 2014년 15건, 2015년 2건으로, 2013년과 2014년에 등록된 경우가 많았다. 넷째, 최종갱신일은 2015년 하반기 갱신된 경우가 8건, 2016년 갱신된 경우가 10건이었다. 2014년 이전에 갱신된 데이터는 17건으로 이는 갱신되었다고 보기 힘들다. 다섯째, 적재주기는 비정기(수시, 자료변경시)가 24건으로 가장 많았다. 데이터가 일회성 공개에 그치고, 적재주기가 비정기인 경우가 많아서 데이터의 활용성과 최신성을 담보하기 어려워 보였다. 향후 데이터의 특성에 따라 일간, 월간 등 주기적인 갱신이 필요할 것이다. 여섯째, 데이터는 신뢰성과 진본성이 중요한데 원본시스템이 특정 기관 홈페이지로 연결되어 해당 사이트내 어느 메뉴에서 출처가 나왔는지 찾기 힘든 경우가 존재하였다. 일곱째, 커버리지, 제공필드는 다소 다르지만, 유사한 내용의 데이터가 중복 공개되는 경우가 존재하였다. 이 경우 이용자가 임의로 판단할 수 밖에 없으므로 이에 대한 대책이 필요해 보였다.

2.2 공공데이터포털

2.2.1 사이트 개요

정부차원에서 운영하는 국내 최대 규모의 데이터 개방 사이트로써 2016년 8월 16일 14,986건의 파일데이터, 1,979건의 오픈API를 제공하고 있다. 교육, 재정금융, 문화관광 등 16개 분야별로 데이터를 제공하고 있다. 또한, 건축, 상권 등 민간에 개방이 시급한 국가 중점개방 데이터 36개 분야를 선정 제공하

고 있다(공공데이터포털, 2016).

2016년 2월 11일 “도서관”으로 검색시 파일데이터 266건, 오픈API 49건, 게시물 51건 등 총 366건 데이터가 검색되었다. “도서관”으로 검색된 파일데이터의 경우, 공공행정, 교육, 농축수산, 통일외교안보 등 12개 분야와 중점데이터에 분산 분류되어 있었다. 정확도순, 날짜순, 제목순, 조회순, 다운로드순 등 정렬기능과 연관 데이터셋을 제공하고 있다. 연관 데이터셋은 내용상 연관성 보다는 키워드 매칭 수준의 결과 제공으로 보였다.

2.2.2 도서관 데이터 개방 현황

데이터 개방량은 많지만, 단일도서관 목록과 도서관 현황정보가 대부분이었다. 이 경우 도서관 현황과 목록의 커버리지, 최신성이 중요한데 일회성 개방이 다수였다. 주요 조사 결과는 다음과 같다.

첫째, 266건의 데이터 중에서 2015년 이후 데이터는 총 90건으로, 데이터의 최신성이 높지 않았다. 이 중에서 2016년 이후 등록된 데이터는 17건이고, 도서관 소재와 운영 현황이 대부분이었다. 소장목록 제공은 5건이었다. 목록의 경우 ISBN을 제공하지 않아서 활용성이 높지 않았고, 일회성으로 산발적으로 제공하였다. 둘째, 개방데이터의 네이밍에 일관성이 없어서 제공데이터의 내용과 커버리지에 대한 예측이 어려웠다. 셋째, 상세정보에서 제공하는 업데이트 주기, 차기 등록예정일, 비용부과 유무, 비용부과 기준 및 단위, 다운로드 횟수 등의 정보와 미리보기 기능은 유용하였다. 넷째, 단일 데이터가 csv, xls 등 여러 파일 형식으로 제공되는데, 파일 형식에 따라 제공 필드가 달라지는 문제점이 나타났다. 다섯째, 간략페이지에 나오는 파일 제공형태와 상세페이지의 파일 제공형태가 일치하지 않는 경우도 존재하였다. 여섯째, 단일도서관 혹은 구 단위 도서관 현황정보 제공이 많았다. 이 경

우 데이터의 최신성, 연속성, 망라성이 보장되지 않아 큰 의미가 없어 보였다. 여섯째, 데이터는 거의 같지만 분류는 상이한 경우가 존재하였다. 예를 들어, 국방분야 연구소의 소장정보는 통일외교안보, 특정시의 소장정보는 교육으로 분류되었다. 이런 현상은 데이터 공개기관의 특성에 따라 주제가 분류되기 때문으로 추정된다. 한편, 서울열린데이터 광장의 데이터가 이 사이트에도 존재하는지 보기 위해 10건 정도 무작위 검색했으나 동일한 검색결과를 없었다.

2.3 data.gov

2.3.1 사이트 개요

미국정부에서 운영하는 데이터 개방 사이트로써, 185,881건의 데이터셋을 제공하고 있다(2016년 8월 17일 기준). 농학, 기상, 해양 등 14개 분야별로 데이터를 제공하며, 중앙부처 및 산하기관, 주 및 지방정부 등 600개 이상이 참여중이다. 미국상무성은 72,618건의 데이터셋을 제공하며, 상무성 산하 미국해양대기국은 70,429건의 최다 데이터 제공하고 있었다. 이 사이트는 데이터셋을 html, csv, rdf, json, xml, api, excel, pdf, txt 등 다양한 형식으로 제공하고 있다(Data.gov, 2016).

2016년 3월 7일 “library”로 검색시 3,003건의 데이터셋을 제공하며, 도서관과 무관한 결과를 다수 포함하고 있었다. 따라서, “public library”로 구문검색하여 나온 114건 중점적으로 살펴보았다. 114건 중에서 2015년 이후 갱신된 것이 108건, 2014년 이전이 6건 이었다. 2016년 4월 19일 재접속시에도 유사한 상황이었고, 데이터 갱신은 비교적 잘 되고 있었다.

2.3.2 도서관 데이터 개방 현황

이 사이트는 연방정부 차원의 통계조사정

보를 제공하여 유용하지만 수년치 자료를 한꺼번에 공개하는 단점이 있었다. 국내 개방사이트에서는 도서관 소재지, 목록정보가 많은데 이 사이트는 도서관 운영성과에 대한 연방, 주정부 차원의 조사결과를 다수 제공하고 있었다. 따라서, 데이터를 활용한 서비스와 앱 개발보다는 연구용으로 더 적절해 보였다. 단, 연구용으로 활용되기 위해서는 다양한 조사결과와 방법, 근거자료 등이 지속적으로 공개되어야 할 것이다. 주요 결과는 다음과 같다.

첫째, 114건 분석결과, 통계정보 30건, 조사결과/통계 27건, 소재정보 10건, 과거 인기도서목록 4건, 장서목록 1건, 전자자원목록 1건, 참고질의 1건, 설문조사 1건, 운영통계 1건을 제공하고 있었다. 둘째, 간략화면에서는 데이터셋 제목, 개요, 데이터 제공 방식, 연방/주 등 레벨을 표시하고 있다. 셋째, 상세화면에서는 메타데이터 제공주체, 담당자, 메타데이터 수정일자, 데이터 개요, 접근 및 이용정보, 다운로드 방식, 메타데이터 생성일/최근 갱신일, 메타데이터 소스 등 데이터에 대한 세부정보를 제공하고 있었다. 넷째, 도서관 프로그램수, 참석자 수 등 단순 통계를 제공하는 경우도 있었다. 다섯째, 단일 도서관으로는 시카고 공공도서관 데이터가 가장 많았다.

3. 결론 및 제언

이상의 결과를 중심으로 오픈 데이터의 특성과 데이터 재활용성 제고를 위한 방안을 제안하면 다음과 같다.

첫째, 데이터의 개방과 이용의 범위를 명확하게 제시할 필요가 있다.

둘째, 시스템의 사용성을 높여야 한다. 제공 데이터 및 검색결과 다운로드, 정렬기능 등을 원활하게 제공해야 할 것이다.

셋째, 개방데이터 종류의 다양화가 필요하다. 도서관 소재지, 장서목록, 인기대출도서, 문화프로그램 등을 균형있게 개방하며, 단순 운영/평가성 통계는 지양해야 할 것이다.

넷째, 일회성 데이터 개방보다는 데이터 개방의 연속성, 주기성, 최신성을 보장해야 데이터로서 활용가치가 높아질 것이다.

다섯째, 개방데이터 종류에 관계없이 개방되는 데이터에 해당 도서관명, 주소, 연락처 등 도서관 관련 정보를 필수적으로 제공하면 더 유용할 것이다.

여섯째, 데이터셋 제목에서 개방데이터의 내용을 파악할 수 있는 표준화된 네이밍 규칙을 고민해볼 필요가 있다. 데이터 활용 여부를 좀 더 정확하고 빠르게 판단하는데 도움이 될 것이다.

도서관 데이터 개방은 매우 유용하고 지속되어야 한다. 그러나, 개방 자체가 목적이 되어서는 안되며 데이터에 대한 지속적인 품질관리와 이를 위한 연구가 뒷받침 되어야 할 것이다.

참 고 문 헌

- 도서관 정보 나루(2016, August 17). Retrieved from <http://data4library.kr/>
- 서울 열린데이터 광장(2016, August 17). Retrieved from <http://data.seoul.go.kr/>
- 신은자(2015) 디지털 시대 오픈데이터 정책의 현황과 과제 정보관리학회지 32(3): 49-68.
- 공공데이터포털(2016, August 17). Retrieved from <https://www.data.go.kr/>
- 현미환 등(2015). 『과학기술 오픈 데이터 평가 지표』. 서울 : 한국과학기술정보연구원. 126p.
- Data.gov(2016, August 17). Retrieved from <https://www.data.gov/>