

공공데이터 포털을 통해 개방된 도서관 관련 데이터 분석*

A Study about Library-Related Open Data through Public Data Portals

조 재 인 (Jane Cho)**

초 록

본 연구는 공공데이터 포털을 통해 개방된 도서관 관련 데이터의 현황을 다각도로 살펴보고, 공개된 데이터의 수준과 공개 주체에 따라 데이터의 활용 정도에 어떠한 차이가 나타나는지 분석하였다. 또한 데이터의 공개 주체를 지방자치단체와 국가 및 공공기관으로 구분하여 데이터에 부여된 키워드를 활용해 PathFinder Network 분석을 수행함으로써 양 주체가 공개한 데이터의 주제 군집과 중심성에 어떠한 차이를 보이는지 분석하였다. 그 결과 국가 및 공공기관이 공개한 데이터로부터는 연구자료 등을 중심으로 4개의 주제군집이, 지방자치단체가 공개한 데이터로부터는 소장자료, 평생교육프로그램 등을 중심으로 8개의 주제 군집이 형성되었으며, 국가 및 공공기관이 공개한 데이터가 학교 및 공공도서관의 단순 현황을 주로 공개하고 있는 지방자치단체 공개 데이터보다 높은 활용도를 나타냈다.

ABSTRACT

This study examines the current state of library related data opened through public data portals, and analyzes how much data is being utilized according to the type of releasing organization, and open level. In addition, we analyzed the subject cluster of data and the centrality of data by performing PathFinder Network analysis using keywords assigned to data by dividing the releasing subject into local government and national/public institutions. Based on this, the subject area of library - related data disclosed by local governments and national/public organizations is understood. And identify the main open body that should be opened first by linking with data utilization analysis result and then suggest implications for future improvement in connection with library big data business.

키워드: 공공데이터, 오픈사이언스, 데이터 개방, 공공데이터 포털, 도서관
Public Open Data, Open Science, Data Open, Data Portal, Library

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017067726).

** 인천대학교 문헌정보학과 부교수(chojane@inu.ac.kr)

논문접수일자 : 2018년 5월 18일 논문심사일자 : 2018년 6월 6일 게재확정일자 : 2018년 6월 7일
한국비블리아학회지, 29(2): 35-56, 2018. [<http://dx.doi.org/10.14699/kbiblia.2018.29.2.035>]

1. 서론

공공데이터 개방이란 중앙정부, 지방자치단체 및 공공기관이 보유 관리하는 공공데이터를 대중이 자유롭게 활용할 수 있도록 다양한 형태로 개방, 제공하는 것이다. 미국, 영국, 일본 등 선진국의 정부에서는 데이터의 민주화, 민간과의 투명한 관계구축 및 재활용을 도모할 수 있다는 점에서 오픈데이터 정책을 적극적으로 추진하고 있으며 지방자치단체 역시 데이터 개방에 활발한 움직임을 보이고 있다. 미국은 2009년 정부가 보유한 공공데이터의 개방원칙 담고 있는 열린 정부 이니셔티브를 시작으로 일원화된 공공데이터 개방 포털인 data.gov를 설치하였고 영국은 data.gov.uk라는 공공데이터 포털을 통해 2014년부터 1,171개 기관에서 14,020개의 데이터셋을 개방할 뿐 아니라 데이터 품질 관리에도 많은 관심을 기울이고 있다(유재영 2015). 우리나라에서는 2013년 공공데이터법을 제정 및 시행하고 그 이후 공공데이터관리지침을 제정 고시(2014), 국가중점데이터(2015)를 선정하여 공공데이터 포털(data.go.kr)을 통해 공개하고 있다. 개방건수는 13년부터 16년까지 3.9배, 이용건수는 120배 증가하여 OECD 공공데이터 개방지수 1위를 달성하고 있으며(행정자치부 2016), 2018년 현재 고가치 고수요 국가 중점데이터 45개 분야를 선정해 개방함으로써 활용사례도 지속적으로 증가하고 있다.

한편, 이러한 시류에 의해 도서관 분야에서도 공공데이터 포털을 통해 많은 데이터가 개방되고 있다. 행정자치부 고시 도서관 및 평생학습 분야 데이터 개방 표준에 의해 자치단체와 단위 도서관이 시설, 소장자료, 평생학습 관련 데이

터를 공개하고 있을 뿐 아니라, 국가 및 공공기관에서도 전국 및 지역 단위 조사 통계 및 각종 연구 자료 관련 데이터를 개방하고 있다.

본 연구는 공공데이터 포털을 통해 개방된 도서관 관련 데이터의 구축 및 활용 현황을 다각도로 살펴보고, 공개된 데이터의 수준과 공개 주체에 따라 데이터의 활용 정도에 어떠한 차이가 나타나는지 분석하고자 한다. 또한 데이터의 공개 주체를 지방자치단체와 국가 및 공공기관으로 구분하여 데이터에 부여된 키워드를 활용하여 PathFinder Network 분석을 수행함으로써 양 주체가 공개한 데이터의 주제 근접과 중심성에 어떠한 차이를 보이는지 분석해 본다. 이를 기반으로 지방자치단체와 국가 및 공공기관이 공개한 도서관 관련 데이터의 주제적 실체를 이해하고 데이터 활용도와 연계 해석함으로써 우선적 개방 필요성이 있는 데이터 개방 주체를 파악해 보는 것이 본 연구의 목적이다.

2. 이론적 배경

2.1 공공데이터 개방 및 표준 형식

우리나라는 2013년 공공데이터 민간 개방을 총괄하는 제도적 장치를 마련하기 위해 '공공데이터의 제공 및 이용 활성화에 관한 법률(법률 제 14839호)'을 발의하여 시행하고 있다. 더불어 공공데이터 전략위원회를 발족하고 공공데이터 관리 지침과 개방 표준을 제정 고시하고 있다. 이를 본격 추진하기 위하여 정책을 총괄하는 행정자치부 산하에 한국정보화진흥원과 공공데이터활용지원센터가 설립되어 실무 지원을 총괄하

고 있으며, 공공데이터전략위원회, 공공데이터제
공분쟁조정위원회, 공공데이터전략실무위원회, 분
야별 공공데이터전문위원회를 설치하고 있다. 더
불어 범정부 공공데이터 개방 통합 단일 창구로
서 공공데이터 포털(data.go.kr)을 구축 및 운영
하고 있어, 2018년 현재 2만여 건의 파일데이터
와 2천5백여 건의 오픈API를 제공하고 있다.

공공데이터란 데이터베이스, 전자화된 파일
등 공공기관이 법령 등에서 정하는 목적을 위
하여 생성 또는 취득하여 관리하고 있는 자료
또는 정보를 의미한다(법 제2조 제2호). 데이
터베이스는 버스운행정보DB, 기상관측DB, 공
간정보DB 등과 같이 각 기관이 생성·취득한
다양한 형태의 데이터를 테이블 형식으로 구조
화하여 DBMS에서 관리하고 있는 유형을 의미
하며, 전자화된 파일은 각 기관이 생성·취득
한 다양한 형태의 데이터를 데이터베이스가 아
닌, 개별 파일 형태로 관리하고 있는 유형을 의
미한다(행정자치부 2016). 이들 데이터는 기계
판독이 가능한 형태로 공개되어야 하는데, 기
계 판독이 가능한 형태란 소프트웨어로 데이터
의 개별 내용 또는 내부 구조를 확인하거나 수
정, 변환, 추출 등 가공할 수 있는 상태를 의미
한다(법 제2조 제3호).

행정자치부는 공공데이터의 품질 관리를 강
화하기 위하여 품질관리 수준 평가제를 실시해
분류 평가하고 있으며 개방표준을 확대해 같은
종류의 데이터는 동일한 항목과 형식으로 개방
하도록 규정하고 있다(공공데이터포털 2016).
도서관의 경우 도서관과 평생학습 정보가 개방
표준으로 제정되어(행정안전부 고시 제 2018-1
호, 17.12.29. 7차 개정) 도서관명, 시도명, 도서
관 유형, 휴관일, 열람좌석수, 강좌명, 강사명, 강

좌내용 등을 표준 속성으로 제시하고 있다. 더불
어 모든 프로그램에서 활용 가능한 형태인 CSV,
JSON, XML 등을 오픈 포맷으로 정의하여 이
를 채택하도록 규정하고 있다. 행정자치부에서
는 팀 버너스리가 제안한 'Five Stars of Linked
Data(<http://5stardata.info/ko/>)' 모델을 기
반으로 한 오픈 데이터 측정 지표를 참조하여 공
공 데이터의 공유 단계를 5가지 형태로 정의하
고 있으며 3단계 이상의 포맷을 채택하도록 규
정하고 있다. 1단계는 사용자가 데이터를 인쇄
하거나 보조기억 장치에 저장될 수 있는 형태로
제공되는 수준으로 데이터가 수록된 문건을 스
캐닝해 PDF와 같은 이미지 형태로 공개되고 있
는 단계를 의미할 수 있겠다. 2단계는 사용자가
특정 소프트웨어를 사용하여 처리할 수 있는 구
조화된 상태로 엑셀파일, 아래한글과 같은 포맷
으로 작성하여 공유하는 단계를 말한다. 이는 사
용자가 데이터를 처리하거나 다른 포맷으로 가
공할 수 있다는 점에서 PDF 형태로 공개된 1단
계에 비하여 공개 수준이 높다고 말할 수 있다.
3단계는 특정 소프트웨어에 한정되지 않고 비독
점적 오픈 포맷으로 데이터가 개방되는 방식이
다. 이용자가 원하는 방법으로 데이터를 가공하
거나 처리할 수 있도록 CSV나 XML, JSON과
같은 비독점 포맷으로 개방되는 형태를 의미한
다. 4단계는 데이터 의미를 기계적으로 해석함
으로써 재사용성을 제고하기 위하여 URI가 부
여되어 있는 상태로 데이터 항목의 재사용성을
제고할 수 있도록 RDF와 같은 방식의 구조화가
이루어지는 단계이다. 마지막 5단계는 기능적 상
호운용이 가능하도록 항목별 데이터를 제어하여
데이터의 문맥과 배경을 제공하는 형태로 링크
드 오픈 데이터 방식으로 데이터 레지스트리를

통해 의미와 연관된 지식까지도 획득할 수 있는 단계라고 말할 수 있겠다(〈표 1〉 참조).

본 연구에서는 공공데이터 포털에 구축된 도서관 관련 데이터의 구축 및 활용에 관한 일반 현황과 데이터의 주제 분포를 주로 분석하지만, 도서관 관련 데이터가 위의 ‘Five Stars of Linked Data’ 모델을 근간으로 한 행정자치부 고시를 어느 정도 준수하고 있으며, 공개 수준이 데이터 활용도에 영향을 미치고 있는지 여부도 살펴볼 것이다.

2.2 선행연구

공개된 공공데이터를 분석한 연구는 다음과 같이 수행된 바 있다.

김지혜(2016)는 지방자치단체와 공공기관 데이터를 수집한 후, 개방 현황, 파일 형식 등을 기반으로 평가를 수행하였으며, 이현정과 남영준(2014)도 광역시도와 기초자치단체에서 보유한 공공데이터를 수집해 문제점을 분석한 바 있다. 그들은 지방자치단체의 소극적 공개와 특정 소프트웨어에 의존적인 공개 형식 문제를 지적하였으며 지역 공공데이터 활용성을 제고하기 위해서 링크드 오픈 데이터 형태의 개방이

필요함을 시사하였다. 한편, 박옥남(2015)은 공공데이터가 지방정부 공공행정 데이터에 치우치고 있어 앞으로 중앙정부 및 공공기관으로 확대되어야 한다고 주장하였으며, 윤소영(2013)도 공공데이터 활성화를 위해서는 보유 기관의 적극 참여, 일관성 있는 URL 생성 관리, 도메인 온톨로지 구축과 결합을 통한 링크드 데이터 발행이 필요함을 제안하였다. 또한 도서관 분야로 한정해 개방데이터 현황을 분석한 연구도 수행된 바 있다. 김혜선과 김완중(2016)은 도서관 분야에서 개방된 데이터 현황을 살펴보고 오픈 데이터의 특성과 데이터 재활용성 제고를 위한 방안을 제안하였다. 특히 단순 운영, 평가성 통계 데이터 개방을 지양해 실효성을 제고해야 한다고 하였다.

해외에서 공개된 공공데이터의 문제점과 발전 방향에 대해서는 다음과 같이 논의되고 있다. 영국에서는 data.gov.uk에 공개된 많은 데이터가 CSV 및 스프레드 시트로 구성되어 있어 Linked Data 방식의 구조화가 시급하며, 더불어 라이선스 제한 문제, 광범위한 메타데이터 채택 문제, LDW(Linked Data Web) 버전 관리 문제 등이 해결되어야 한다고 논의된 바

〈표 1〉 오픈데이터 발달 단계 측정 지표(행정안전부 2016)

| 구분 | 1단계 | 2단계 | 3단계 | 4단계 | 5단계 |
|--------------|------------------------------------|---|-------------------------|--|-----------------------|
| 기계판독이 가능한 형태 | 미충족포맷 | 최소충족포맷 | 오픈포맷 | | |
| 특징 | 특정 소프트웨어에서 읽을 수만 있는 데이터로 수정, 변환 불가 | 특정 소프트웨어에서 읽고 수정, 변환 가능 | 모든 소프트웨어에서 읽고 수정, 변환 가능 | URI를 기반으로 데이터 속성 특성 관계를 기술하고 있는 데이터 구조 | 웹상의 다른 데이터와 연결, 공유 가능 |
| 예시 | PDF | HWP, XLS, JPG, PNG, WMV, MPEG, MP3, SWF | CSV, JSON, XML | RDF | LOD |

있다(Hadbolt et al. 2012). 미국 data.gov 경우에도 시민과 개발자가 각기 다른 정부 기관의 데이터를 매쉬업하여 활용할 수 있도록 상호호환성을 높여야 한다고 논의되었다. 더불어 대규모 OGD(Open Government Data) 데이터 세트를 처리할 수 있는 효과적인 인프라와 필요한 인적 자원을 각계에 분산 배치하는 더 나은 사회 메커니즘이 필요하다고 제안되었다(Hendler et al. 2012).

한편 스페인(Clabo and Ramos-Vielba 2015)과 호주(Hardy and Maurushat 2016)에서도 개방된 공공데이터의 저작권과 상호운용성의 문제가 논의되었으며, 중동지역에서도 개방형 데이터의 구현과 성숙도에 대한 문제가 연구되고 있어(Alanazi and Chatfield 2012), 전 세계 각 국가에서 공공데이터 공개를 둘러싼 현상과 발전 방향에 대하여 논의가 지속되고 있음을 확인할 수 있다.

3. 연구의 방법

본 연구는 도서관 관련 공공 데이터의 개방 및 활용에 관한 일반 현황을 살펴보고 데이터 수준과 공개 주체에 따라 활용 정도에 차이가 나타나는지 분석하였다. 또한 공개 주체의 유형에 따라 어떠한 주제적 특성의 데이터가 개방되고 있는지 PathFinder Network 분석을 수행하였다. 이를 위해 공공데이터 포털에 개방된 도서관 관련 공공데이터를 추출하여 분석하였으며, 개방 및 활용 현황과 주제 분석 방법은 다음과 같은 절차로 수행되었다.

도서관 관련 공공데이터의 개방 및 활용 현

황, 그리고 공개 주체와 수준이 활용 정도에 미치는 영향에 대한 검증은 다음과 같은 방법으로 수행하였다. 첫 번째, 2018년 1월을 기준으로 공공데이터 포털사이트에서 도서관이라는 키워드로 검색하여 도출된 560건을 샘플데이터로 추출하였다. 이를 기반으로 개방기관, 공개수준, 다운로드 정도 등을 기술통계 분석하였다. 두 번째, 데이터의 공개 주체와 수준에 따라 활용 정도에 차이가 나타나는지 분석하였다. 도서관 관련 데이터를 공개하고 있는 기관은 지방자치단체, 단위도서관, 그리고 국가 및 기타 공공기관으로 구성되고 있어 이들 주체가 각각 공개한 데이터의 활용 정도에 유의미한 차이가 나타나는지 살펴보았다. 한편 공개 수준은 'Five Stars of Linked Data' 모델에 근거해 개별 데이터를 평가하였으며 그에 따라 다운로드 건수에 차이가 나타나고 있는지 확인하였다. 영향력 검증은 Kruskal Wallis 분석을 이용하였으며 모든 통계 분석과 시각화는 SPSS 23과 R을 활용하였다.

한편, 공공데이터의 주제 분석은 다음과 같이 키워드의 Pathfinder Network 분석 방법을 통해 수행하였다. 지역 단위 데이터와 국가 수준 데이터의 주제 범주를 살펴보기 위하여 공개 주체를 자치단체와 국가 및 공공기관으로 양분해 구축된 1,674개의 키워드를 추출하였다. 두 번째, Knowledge Matrix를 사용해 추출된 키워드의 유사 개념 통합 작업을 수행하고 고출현빈도를 보인 키워드를 자치단체가 공개한 데이터에서 31개(4회이상출현), 국가 및 공공기관이 공개한 데이터에서 14개(3회이상출현), 총 45개 추출하였다. 세 번째, 추출한 키워드를 대상으로 이재윤이 개발한 COOC 프로그램을

사용해 동시출현행렬을 산출하고 피어슨상관계수를 사용해 PathFinder Network 분석을 수행하였다. 더불어 병렬최근접 이웃 클러스터링(PNNC: Parallel Nearest Neighbor Clustering)을 수행해 키워드 군집을 형성하였으며(이재윤 2006) 가중 네트워크의 중심성(Weighted Network Centralities)을 파악하였다. 이는 모두 이재윤이 개발한 WNET(<http://cafe.daum.net/wnets>)을 사용해 처리하였다. PathFinder 알고리즘은 중요한 링크가 존재하는 노드들만을 남기는 방식으로 네트워크를 표시하므로 전체 구조를 이해하는데 매우 유용하다. 한편, 가중 네트워크 분석에서는 기존의 네트워크 분석에서 사용되는 중심성 지수와 다른 방법으로 중심성 지수를 산출하므로(이재윤 2013), 여기에서는 노드들간을 매개시켜주는 정도를 측정하는 상대적삼각매개중심성($rTBC$)과 한 노드가 다른 노드들과 가진 가중치의 평균값인 평균연관성(AVGSIM)을 중심성 지수로 사용하였다.

났다. 특별, 광역, 기초 및 교육청을 포함하는 지방 자치단체가 개방한 데이터는 396건으로 전체 도서관 관련 데이터의 70% 가량을 차지하고 있는 것으로 나타났다. 그 다음은 단위도서관으로 76건(13.%)의 데이터를, 도서관리공단, 시설관리공단과 같은 위탁기관이 27건(4.8%)의 데이터를 공개하고 있는 것으로 나타났다. 한편, 국가 및 공공기관은 60건(10.7%)의 도서관 관련 데이터를 공개하고 있는 것으로 나타났는데 한국문화정보원, 한국생명과학연구원, 한국학중앙연구원, 한국교육학술정보원, 국립중앙도서관, 국방부, 국제교류재단 등을 주요 공개기관으로 예시할 수 있겠다.

〈표 2〉 공개 주체에 따른 데이터 현황

| 공개 주체 | 사례수 | 비율 |
|-----------|-----|------|
| 국가 및 공공기관 | 60 | 10.7 |
| 단위도서관 | 76 | 13.5 |
| 위탁기관 | 27 | 4.8 |
| 자치단체 | 396 | 70.8 |
| sum | 559 | 100 |

4. 도서관 관련 공공데이터 공개 및 활용

4.1 공개 현황

먼저 공공데이터 포털에 개방된 도서관 관련 데이터의 개방 및 활용 현황을 살펴보면 다음과 같다.

첫 번째, 데이터 공개 주체를 국가 및 공공기관, 지방자치단체, 위탁기관, 단위도서관으로 구분해 살펴본 결과 가장 압도적으로 많은 데이터를 공개하고 있는 곳은 지방자치단체로 나타

두 번째, 공공데이터 포털 사이트에 개방된 도서관 관련 데이터는 〈표 3〉과 같이 서울특별시가 84건(15%)로 가장 많고 경기도가 75건(13.4%)으로 그 다음으로 많은 것으로 나타났으며 세종특별자치시, 제주특별자치도 등의 공개건수가 가장 저조한 것으로 나타났다.

세 번째, 데이터 공개 시점을 보면 〈표 4〉와 같이 2014년 최초 공개된 데이터가 202건(36%)로 가장 많은 것으로 나타났으며, 데이터의 공개는 대부분 행정업무 관련 규정, 도서관법, 지방자치단체 조례나 기관 관련법을 근거로 하고 있는 것으로 조사되었다.

〈표 3〉 지역에 따른 공개 데이터 현황

| 지역 | 사례수 | 비율 |
|---------|-----|-------|
| 강원도 | 23 | 4.11 |
| 경기도 | 75 | 13.41 |
| 경상남도 | 51 | 9.12 |
| 경상북도 | 52 | 9.30 |
| 광주광역시 | 13 | 2.32 |
| 대구광역시 | 35 | 6.26 |
| 대전광역시 | 25 | 4.47 |
| 부산광역시 | 50 | 8.94 |
| 서울특별시 | 84 | 15.02 |
| 세종특별자치시 | 3 | 0.53 |
| 울산광역시 | 11 | 1.96 |
| 인천광역시 | 23 | 4.11 |
| 전라남도 | 38 | 6.79 |
| 전라북도 | 29 | 5.18 |
| 제주특별자치도 | 7 | 1.25 |
| 충청남도 | 21 | 3.75 |
| 충청북도 | 18 | 3.22 |
| sum | 559 | 100 |

〈표 4〉 연도에 따른 공개 데이터 현황

| 연도 | 사례수 | 비율 |
|------|-----|-------|
| 2011 | 5 | 0.89 |
| 2012 | 2 | 0.35 |
| 2013 | 43 | 7.69 |
| 2014 | 202 | 36.13 |
| 2015 | 125 | 22.36 |
| 2016 | 111 | 19.85 |
| 2017 | 71 | 12.70 |
| sum | 559 | 100 |

네 번째, 공개된 데이터 파일의 유형을 ‘Five Stars of Linked Data 모델’에서 제시한 데이터 공유 단계와 매핑해 살펴보면 〈표 5〉와 같다. 행정자치부 고시 데이터 관리 지침에 의하면 공공데이터는 ‘Five Stars of Linked Data 모델’의 3단계 이상의 공유 수준을 유지해야 하는 것으로 제시되어 있다. 3단계는 비독점적 오픈

포맷으로 이용자가 원하는 방법으로 데이터를 가공하거나 처리할 수 있는 포맷으로 개방된 상태를 의미한다. 분석된 데이터의 대부분이 3단계에 해당하는 CSV, XML, TXT 등을 채택하고 있는 것으로 나타났으나 응용프로그램 의존 형식인 2단계로 개방된 경우도 다수 존재하는 것으로 나타났다. 구체적으로 살펴보면 CSV가 448건으로 80%를 차지하고 있는 것으로 나타났고, 2단계 포맷인 XLS나 HWP도 각각 30건(5.3%), 9건(1.6%)이 존재하는 것으로 나타났다. 데이터가 수록된 문건을 스캐닝해 PDF와 같은 이미지 형태로 개방함으로써 재 활용성을 저해하는 1단계와 데이터 의미를 구조화한 4단계 개방 형식은 부재한 것으로 나타났다. 한편, 5단계 공개 수준을 가지는 도서관 관련 데이터도 존재하였는데 이는 한국문화정보원이 LOD 기반으로 구축한 박물관, 미술관, 도서관, 기록관 데이터인 것으로 조사되었다. 이 데이터는 문화용성을 위한 융합·오픈DB(GLAM) 사업을 통해 구축된 문화융합DB로 박물관, 미술관, 도서관, 기록관 등에서 소장하고 있는 문화예술 분야의 자료를 개방형 데이터 형태로 구축한 자료이다. 그 밖에 분석 대상 데이터 범위에는 포함되지 않았지만, 국가서지 LOD와 NDSL LOD도 공개되어 있는 것으로 나타났다.

분석 대상이 된 공공데이터 포털에는 LOD로 구축된 데이터가 많지 않은 반면 오픈API 형태로 공개된 데이터는 다수 존재하였다. 오픈 API는 이용자가 웹 검색 결과 및 인터페이스 등을 제공받는 데 그치지 않고 직접 응용 프로그램과 서비스를 개발할 수 있도록 공개된 API이다. 누구나 접근하여 사용할 수 있다는 장점이 있으며 앱 개발에 바로 활용될 수 있으나 LOD에

〈표 5〉 공개 데이터 형식과 ‘Five Stars of Linked Data 모델’에 의한 단계

| 형식 | 단계* | 사례수 | 비율 |
|----------|--|-----|-------|
| csv | 3 | 448 | 80.14 |
| xml | 3 | 36 | 6.440 |
| xls | 2 | 22 | 3.93 |
| json+xml | 3 | 11 | 1.96 |
| hwp | 2 | 9 | 1.610 |
| xlsx | 2 | 8 | 1.43 |
| html | 2 | 1 | 0.17 |
| jpg | 1 | 1 | 0.17 |
| marc | 3 | 1 | 0.17 |
| tfl | 5 | 1 | 0.17 |
| txt | 3 | 1 | 0.17 |
| sum | | 559 | 100 |
| openAPI | 경기도7, 경상남도2, 경상북도2, 광주광역시3, 대구광역시1, 대전광역시7, 부산광역시14, 서울특별시17, 세종특별자치시1, 울산광역시3, 전라남도2, 전라북도3, 제주특별자치도1 | | |

* Tim Berners-Lee's five stars of linked data based stage

비하여 데이터의 접근과 활용 그리고 매쉬업에 용이한 구조를 가지진 않는다. 공공데이터 포털을 구축할 때부터 LOD를 고려하여 추진한 영국(data.gov.uk)의 사례와 달리 한국의 공공데이터 포털은 LOD 개방 및 활용 플랫폼으로서의 인프라가 미비하다고 평가되고 있다(이재원 2016). 여하튼 여기에 개방된 도서관 관련 오픈 API는 국회도서관, 국립중앙도서관, 한국과학기술원, 한국교육학술정보원, 국방부, 기계연구원 등 국가 및 공공기관에서 23개, 부산지역 단위도서관에서 13개, 자치단체에서 27건 총 63건이 존재하는 것으로 나타났는데, 자치단체는 대부분 관내 문화시설이나 도서관 현황을 공개하고 있으며, 서울시, 대전, 광주 등과 같은 광역자치단체를 중심으로 하는 것으로 나타났다.

4.2 활용 현황

공개된 데이터의 활용 정도를 파악해 본 결

과를 제시하면 다음과 같다.

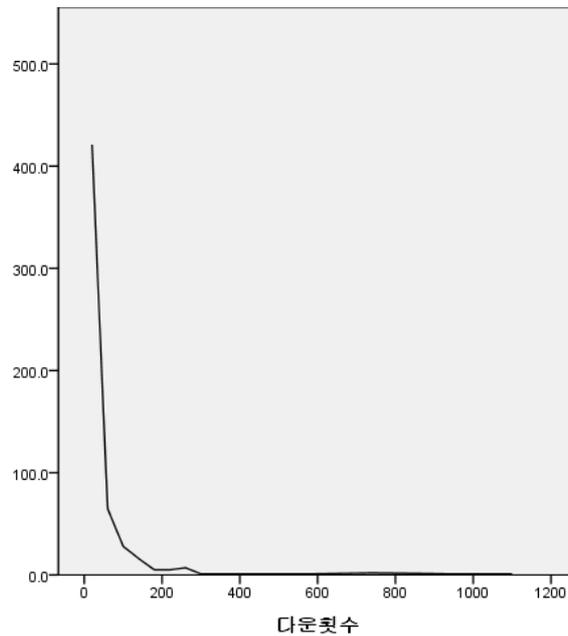
첫 번째, 비정상적으로 다운로드 횟수가 많은 2개 기관의 데이터를 제외하고 공공 데이터의 다운로드 평균을 계산하니 43회를 나타냈다. 다운로드 횟수에 대한 그래프를 생성하면 다음 〈그림 1〉과 같이 긴꼬리 모양을 보이고 있어 소수의 자료에 이용이 집중된 현상을 나타내고 있다. 다운로드 횟수가 100건이 넘는 경우는 단지 55개 데이터 뿐이며, 300여개의 공공데이터가 20회 미만의 다운로드 횟수를 보이고 있다. 10건 이하로 아주 저조하게 활용된 데이터가 전체의 36%(203건)를 차지하고 있었으며, 해당 데이터는 대부분 지방자치단체에서 공개한 공공도서관 일반 현황에 관한 데이터로 나타났다. 일부 지방자치단체나 교육청이 공개한 데이터를 제외하면 대부분의 공공도서관 데이터는 매우 저조한 활용도를 보이고 있다. 반면, 국가 및 공공기관이 공개한 데이터는 분석 대상 데이터 중 18%만이 10건 이하의 다운로드 횟

〈표 6〉 10건 미만 이용데이터의 공개 기관

| 공개 주체 | 기관수 | 전체 사례수 |
|-----------|-----|--------|
| 국가 및 공공기관 | 11 | 60 |
| 단위도서관 | 36 | 76 |
| 위탁기관 | 4 | 27 |
| 지방자치단체 | 152 | 396 |
| sum | 203 | 559 |

〈표 7〉 고효용 데이터

| 제목 | 공개 기관 | 공개 주체 구분 | 지역 | 유형 | 다운로드횟수 |
|------------------|-----------|-----------|---------|----------|--------|
| 혁신도시 인근 시설 현황 | 전라북도 | 자치단체 | 전라북도 | csv | 1,080 |
| 인천광역시 교육 및 문화 정보 | 인천광역시 | 자치단체 | 인천광역시 | csv | 1,018 |
| 대학도서관통계 | 한국교육학술정보원 | 국가 및 공공기관 | 대구광역시 | csv | 922 |
| 도시계획시설도 | 국토교통부 | 국가 및 공공기관 | 세종특별자치시 | json+xml | 745 |
| 사회조사결과 | 부산광역시 강서구 | 자치단체 | 부산광역시 | csv | 739 |
| 장애인복지시설정보 | 행정안전부 | 국가 및 공공기관 | 서울특별시 | csv | 564 |



〈그림 1〉 도서관 관련 공공데이터 다운로드 횟수별 사례 수

〈표 8〉 공공 데이터 다운로드 통계

| 사례수 | 최소값 | 최대값 | 평균 | 표준편차 |
|-----|-----|-------|-------|---------|
| 557 | 0 | 1,080 | 43.80 | 102.064 |

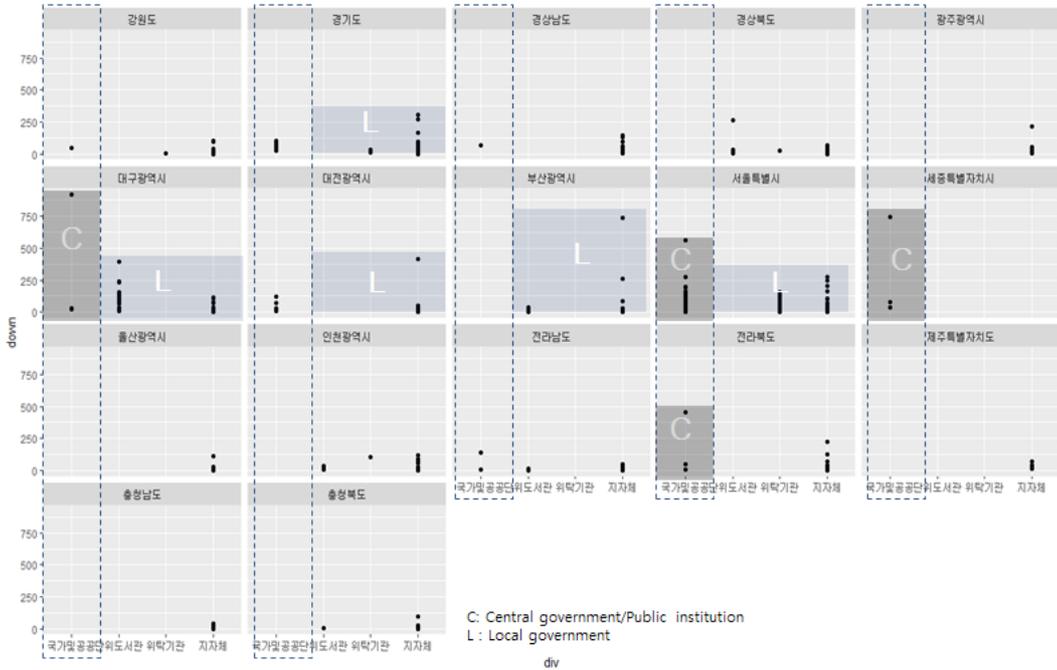
수를 보이고 있어, 저이용도를 보이는 데이터가 많지 않은 것으로 분석되었다. 한편 국가 및 공공기관이 공개한 데이터 중 고효용빈도를 나타낸 데이터는 한국교육학술정보원의 대학도서관 통계, 한국출판문화산업진흥원의 해외독서관련단체, 행정안전부의 장애인복지시설 정보, 국회도서관 자료 검색서비스, 한국문화정보원의 박물관, 미술관, 도서관, 기록관 소장자료, 한국학중앙연구원 국제교과서도서관 소장 교과서 정보 등으로 나타났다.

두 번째, 데이터 공개 주체를 지역별로 구분하여 활용 정도를 살펴보면 <표 9>와 같다. 다운로드 평균이 가장 높은 지역은 세종(287건), 대구(107건) 순으로 나타났는데, 도서관등 기반 시설을 포함하는 국토교통부(세종)의 도시계획 시설(openAPI)이 높은 활용신청건수(745회)를 보였으며, 한국교육학술정보원(대구)의 대

학도서관통계(922회)도 높은 다운로드 건수를 보여 지역 평균을 높이고 있는 것으로 나타났다. 평균 다운로드 수치는 시도군보다 특별/광역시 높은 것으로 나타났는데, 실제 평균을 산정해 비교한 결과 시도군이 29.58회, 특별/광역시가 63.86으로 두배 이상의 수치 차이를 보이는 것으로 나타났다. 한편, 17개 시도에 개방된 데이터의 다운로드 분포를 공개 기관의 유형에 따라 구분한 후 <그림 2>와 같이 R을 통해 시각화해 보았다. 그 결과, 관내 국가 및 공공기관에서 높은 다운로드 분포를 보이는 지역은 앞에서 언급한 세종과 대구이외에 서울과 전북인 것으로 드러났다. 서울의 경우, 국립중앙도서관과 국회도서관의 데이터가, 전북의 경우 한국출판문화산업진흥원의 해외 독서관련 단체/기관 및 독서장려프로그램에 관한 데이터가 높은 활용도를 보이고 있는 것으로 나타났다. 한편, 단위

<표 9> 지역별 다운로드 건수

| 지역구분 | 지역 | 평균 | 사례수 | 평균 | 계 |
|-------|----|--------|-----|-------|-----|
| 시도군 | 강원 | 21.57 | 23 | 29.58 | 326 |
| | 경기 | 39.84 | 76 | | |
| | 경남 | 20.65 | 51 | | |
| | 경북 | 19.37 | 52 | | |
| | 전남 | 14.38 | 37 | | |
| | 전북 | 79.38 | 29 | | |
| | 충남 | 14.29 | 21 | | |
| | 충북 | 11.89 | 18 | | |
| 특별광역시 | 광주 | 40.08 | 13 | 63.86 | 231 |
| | 대구 | 107.97 | 35 | | |
| | 대전 | 43.28 | 25 | | |
| | 부산 | 32.02 | 50 | | |
| | 서울 | 64.19 | 84 | | |
| | 세종 | 287.00 | 3 | | |
| | 울산 | 23.27 | 11 | | |
| | 인천 | 77.39 | 23 | | |
| | 제주 | 31.67 | 6 | | |



〈그림 2〉 공개주체(지역, 기관유형)에 따른 다운로드 건수 분포

도서관, 위탁기관, 지방자치단체가 공개한 데이터의 다운로드 건수가 상대적으로 높게 분포된 지역은 대구, 대전, 부산, 서울 등의 특별/광역시와 경기도 정도인 것으로 분석되었다.

4.3 공개 주체와 수준에 따른 활용도 차이

한편, 공개 주체와 수준에 따라 데이터 활용 정도에 차이가 나타나는지 통계적으로 검증한 결과를 제시하면 다음과 같다.

첫 번째, 공개 주체가 다운로드 횟수에 유의미한 영향을 주고 있는지 파악하기 위하여 Kruskal Wallis 분석을 수행한 결과를 제시해 보도록 한다. 먼저 공개 기관의 유형을 변수로 검증한 결과, 평균 순위가 <표 10>과 같이 국가 및 공공기관(384.02) > 위탁기관(368.85) > 단위도서

관(261.04) > 자치단체(260.31) 순으로 나타났다. 지방자치단체, 단위도서관, 위탁기관의 데이터보다 국가 수준의 특화 데이터를 공개하고 있는 국가 및 공공기관의 데이터에 대한 활용 수준이 높은 평균 순위를 보이는 것으로 검증되었다. 한편 공개 주체를 지역에 따라 구분하여 평균 순위를 분석한 결과, 세종(485.5) > 대구(408.7) > 서울(353.1) 순을 나타냈으며, 지역을 특별/광역시와 시도군으로 구분하여 다시 한번 차이를 검증한 결과에서도 특별광역시(319.95) > 시도군 (249.98) 순을 나타내, 전반적으로 특별/광역시에 소재하거나 국가 및 공공기관이 공개한 데이터가 높은 활용도를 보인다고 해석되었다. 공개 주체가 다운로드 횟수에 미치는 영향은 모든 분석에서 유의미한 수준($p < .001$)인 것으로 검증되었다.

〈표 10〉 공개 주체에 따른 다운로드 차이에 대한 Kruskal Wallis 검정 결과

| 공개 주체 | 사례수 | 평균 순위 | Kruskal Wallis Test | |
|-----------|-----|--------|---------------------|------|
| | | | Chi-Square | df |
| 국가 및 공공기관 | 60 | 384.02 | 40.261 | 3 |
| 단위도서관 | 76 | 261.04 | Asymp. Sig. | .000 |
| 위탁기관 | 27 | 368.85 | | |
| 자치단체 | 394 | 260.31 | | |
| 합계 | 557 | | | |

〈표 11〉 공개 지역에 따른 다운로드 차이에 대한 Kruskal Wallis 검정 결과

| 지역 | 사례수 | 평균 순위 | 지역 | N | 평균 순위 | Kruskal Wallis Test | |
|--------|-----|----------|-------|-----|----------|---------------------|---------------------|
| | | | | | | Chi-Square | df |
| 강원 | 23 | 241.8043 | 세종 | 3 | 485.5 | 112.554 | Asymp. Sig. .000 |
| 경기 | 76 | 332.2697 | 울산 | 11 | 251.4545 | 16 | |
| 경남 | 51 | 220.3627 | 인천 | 23 | 314.1957 | | |
| 경북 | 52 | 213.6827 | 전남 | 37 | 186.027 | | |
| 광주 | 13 | 331 | 전북 | 29 | 269.2759 | | |
| 대구 | 35 | 408.7429 | 제주 | 6 | 352.0833 | | |
| 대전 | 25 | 318.26 | 충남 | 21 | 217.0238 | | |
| 부산 | 50 | 210.59 | 충북 | 18 | 148.0278 | | |
| 서울 | 84 | 353.1964 | Total | 557 | | | |
| 시도군 | | | | 326 | 249.98 | 25.581 | |
| 특별/광역시 | | | | 231 | 319.95 | .000 | |

두 번째, 공개 수준에 따라 다운로드 횟수에 차이가 나타나는지 통계적으로 검증한 결과를 제시하면 다음과 같다. 앞에서 설명한 바와 같이 공공데이터의 공개 수준은 1단계에서 5단계로 구분되는데, 행정자치부에서는 공개데이터의 수준을 3단계 이상으로 유지할 것을 권고하고 있다. 데이터의 공개 수준이 높아지면 데이터의 재사용성이 높아지면서 다른 데이터 소스와의 링크를 통해 연관지식 획득 및 활용효과를 극대화할 수 있을 것으로 기대되기 때문이다. 데이터의 공개 수준이 다운로드 횟수에 영향을 미치는 지 확인하기 위하여 openAPI를 포함해 데이터 공개 수준을 독립변수로 Kruskal Wallis 분석을 수행한 결과 평균 순위는 5단계

(535) > 2단계(348.41) > openAPI(312.13) > 1단계(304) > 3단계 (268.15) 순을 보였으며, 유의수준은 p=.007로 나타나 유의한 차이가 존재하는 것으로 검증되었다. 그러나 오픈포맷 형식으로 공개된 3단계의 데이터보다 엑셀과 같은 응용 프로그램 형식으로 개방된 2단계 수준에 대한 평균 순위가 더욱 높은 것으로 나타나, 공개 수준이 높다고 해서 활용도가 상대적으로 높아지는 않는다는 것을 유추할 수 있었다. 이러한 추측을 확인하기 위하여 openAPI를 4단계로 설정해 Spearman 분석을 실시해 본 결과, r = -0.02를 나타내 변수간의 관계에는 상관성이 존재하지 않는 것으로 검증되었다. 공개 수준과 활용도 간에 상관성이 존재하지 않는 이

〈표 12〉 공개 수준에 따른 다운로드 차이에 대한 Kruskal Wallis 검정 결과

| 단계 | 사례수 | 평균순위 | Kruskal Wallis | |
|---------|-----|--------|----------------|--------|
| | | | Chi-Square | df |
| 1 | 1 | 304 | Asymp. Sig. | 14.185 |
| 2 | 37 | 348.41 | | |
| 3 | 455 | 268.15 | | |
| 5 | 1 | 535.00 | | |
| openAPI | 63 | 312.13 | | |
| 합계 | 557 | | | |

〈표 13〉 공개 수준과 다운로드 횟수간의 Spearman 검정 결과

| Spearman's rho | | 공개 단계 | 다운횟수 |
|----------------|-------------------------|-------|-------|
| 공유단계 | Correlation Coefficient | 1.000 | -.002 |
| | Sig. (2-tailed) | . | .955 |
| | N | 557 | 557 |
| 다운횟수 | Correlation Coefficient | -.002 | 1.000 |
| | Sig. (2-tailed) | .955 | . |
| | N | 557 | 557 |

유는 범용적으로 활용되는 엑셀과 아래한글 파일로 공개된 2단계 데이터가 공공데이터를 열람하는 이용자의 단순 참고 및 분석에 장애가 되지 않았기 때문에 추측된다. 또한 공공데이터 포털을 통해서도 도서관 관련 데이터가 가공 또는 융합되어 새로운 서비스로 재활용되는 시도가 상대적으로 많지 않기 때문에 3, 4단계 데이터에 대한 활용도가 높게 나타나지 않았다고 설명해 볼 수 있겠다.

5. 데이터 공개 기관의 유형에 따른 공공데이터의 주제 분야 분석

앞에서는 국가 및 공공기관이 공개한 데이터가 지방자치단체, 단위도서관, 위탁기관이 공개한 데이터에 비해 높은 활용도를 보이는 것으로

분석되었다. 여기에서는 지방자치단체, 단위도서관, 위탁기관을 하나의 분석 단위로 묶고 전국단위 특화데이터를 공개하고 있는 국가 및 공공기관과 비교해 어떠한 주제 분포의 차이를 보이는지 분석하였다.

5.1 고출현 키워드의 차이

고출현 키워드의 차이를 살펴보기 위하여 우선 양쪽 공개 주체로부터 추출한 키워드의 유사 단어를 통합하는 과정을 수행하였다. KISTI가 개발한 Knowledge Matrix를 활용해 서적/책/도서, 마을문고/작은도서관, 도서목록/서지/장서목록 등과 같은 유사 용어를 통합하였다. 그 후에 고출현빈도를 계산한 결과, 지방자치단체 데이터에서는 서적(238회), 마을문고(181회), 평생학습(63회), 교육(56회) 등의 키워드가 고출

현키워드로 나타났으며 국가 및 공공기관에서 공개한 데이터에서는 도서(17회), 전자도서관(10회), 국방(6회) 등의 키워드가 출현빈도가 높은 키워드로 나타났다. 지방자치단체 데이터에서는 강좌, 신착, 문화 등의 키워드 역시 다수 출현하는 것으로 나타나 공공도서관과 학교도서관, 작은도서관의 신착자료나, 문화/평생교육 프로그램, 도서관 시설 현황이나 대출 규정과 같은 단순 도서관 현황에 대한 공개가 주로 이루어지고 있는 것으로 추측할 수 있었다. 반면, 소장 도서별 대출량이라던가 평생교육 프로그램별 등록건수 그리고 이들 정보와 회원의 특성을 연계 해석할 수 있는 정보의 제공은 미약한 것으로 파악되었다. 한편, 국가 및 공공기관에서 공개된 데이터는 지방자치단체 개방 데이터에 비해 개방된 데이터량이 미약해 키워드

의 출현빈도가 전반적으로 높지 않았으나 도서, 전자도서관 이외에도 연구자료, 해외도서, 학술정보, 국방, 기계기술, 한국학 등 지방자치단체 분석 결과에서는 드러나지 않았던 특수 분야의 연구, 학술 관련 개념을 반영하는 키워드가 등장하였다(〈표 14〉 참조).

5.2 PFNET 분석 결과 차이

네트워크 분석을 통해 양 주체가 공개한 데이터에 부여된 키워드를 대상으로 COOC을 통해 동시출현행렬을 산출하였다. 피어슨 상관계수 값을 가지고 WNET을 통해 패스파인더 네트워크와 병렬최근접 이웃 클러스터링을 수행하고 가중 네트워크의 중심성도 파악하였다. 그 결과 지방자치단체가 공개한 데이터와 국가

〈표 14〉 지방자치단체와 국가 및 공공기관이 공개한 데이터를 기반으로 한 고출현키워드

| 지방자치단체 | | | | 국가 및 공공기관 | |
|--------|-------|----|---------|-----------|-------|
| 빈도 | 키워드 | 빈도 | 키워드 | 빈도 | 키워드 |
| 238 | 서적 | 7 | 소장자료 | 17 | 도서 |
| 181 | 마을문고 | 7 | 자료검색 | 10 | 전자도서관 |
| 63 | 평생학습 | 7 | 프로그램 | 6 | 국방 |
| 56 | 교육 | 6 | 추천도서 | 5 | 연구자료 |
| 52 | 강좌 | 5 | 평생교육강좌 | 5 | 색인 |
| 24 | 목록 | 5 | 도서대출 | 4 | 해외 도서 |
| 23 | 신착 | 5 | 휴관일 | 4 | 한국학 |
| 20 | 공공도서관 | 5 | 어린이도서관 | 4 | 학술정보 |
| 20 | 도서정보 | 4 | 학교도서관 | 4 | 대학도서관 |
| 11 | 신간 | 4 | 장난감 도서관 | 4 | 기계기술 |
| 10 | 서지 정보 | 4 | 도서관명 | 3 | 저자 |
| 9 | 문화 | 4 | 희망도서 | 3 | 자료 |
| 9 | KOLAS | 4 | 박물관 | 3 | 목록 |
| 9 | 시립도서관 | 4 | 시설 | 3 | 고문서 |
| 8 | 열람실 | 4 | 구립도서관 | | |
| | | 4 | 공연 | | |

및 공공기관이 공개한 데이터는 상이한 키워드의 중심성을 보였으며, 형성된 주제 군집에도 차이를 보였다. 지방자치단체 개방 데이터 중 상대적삼각매개중심성(rTBC)이 높은 키워드는 마을문고(0.583), 서적(0.627), 시립도서관(0.72)으로 나타났으며 그중 서적(0.030)은 평균연관성(AVSIM)까지 높은 키워드로 분석되었다. 삼각매개중심성은 노드가 다른 노드들 간을 매개시켜주는 정도를 측정하는 지표로 네트워크상에서 전역적인 영향력을 발휘하는 키워드라고 말할 수 있다. 다시 말해 시립도서관과 마을문고, 그리고 서적이라는 주제 개념이 지방자치단체가 개방한 공공데이터에 있어 가장 전반적인 영향력을 행사하는 핵심 주제라고 말할 수 있다. 한편, 노드가 다른 노드들과 가진 가중치의 평균값인 평균연관성(AVSIM)이 높은 키워드는 네트워크 전체에서 다른 분야들과 상당히 많이 연관되는 분야임을 나타낸다. 앞에서 언급한 서적은 상대적삼각매개중심성 뿐 아니라, 평균연관성까지 높은 주제 개념으로 나타났으며, 그 외에도 목록(0.032), 신착(0.026)이 높은 수치를 나타냈다.

반면에 국가 및 공공기관이 공개한 데이터를 대상으로 한 분석에서는 전혀 다른 결과를 나타냈다. 마을문고, 서적, 시립도서관이 가장 높은 상대적삼각매개중심성(rTBC)을 보인 지방자치단체 데이터의 분석 결과와 달리 고문서(0.576), 자료(0.461), 목록(0.435)이 전역적인 영향력을 갖는 키워드로 나타났다. 자료라는 키워드는 지방자치단체가 개방한 데이터에서는 고출현빈도로 등장하지 않았으나 국가 및 공공기관이 개방한 데이터에서는 영향력이 높은 키워드로 드러났는데, 전국 단위의 특수 분야를

다루는 이들 기관이 지방자치단체와는 달리 주제자료나 학술자료, 발간자료 등을 공공 데이터로 다수 공개하고 있기 때문인 것으로 보인다. 한편, 색인(0.01), 전자도서관(0.05)과 같은 키워드가 평균연관성(AVSIM)이 높은 키워드로 나타났는데 이는 특수 분야의 전자도서관과 기사색인에 대한 정보를 공개하고 있는 국가 및 공공기관의 특성을 반영하고 있다.

한편, NodeXL을 통하여 네트워크 분석결과를 시각화하고 PNNC로 산출된 군집을 적용해 나타낸 결과는 <그림 3>, <그림 4>와 같다. 지방자치단체가 공개한 데이터는 <표 15>와 같이 8개의 군집(C1: 도서관 및 문고, C2: 평생교육, C3: 소장자료, C4: KOLAS, C5: 문화, C6: 열람실, C7: 장난감도서관 프로그램, C8: 학교도서관및기타)으로 구분되었다. 공개된 데이터는 평생교육, 장서현황, 문화, 열람실, 학교도서관 등 비교적 지방자치단체에서 서비스하고 있는 공공도서관과 학교도서관의 현황을 반영하고 있는 것으로 종합할 수 있겠다. 반면 국가 및 공공기관이 공개한 데이터에서는 4개의 군집(C1: 색인 및 목록, C2: 전자도서관, C3: 연구자료, C4: 대학도서관)이 형성되었다. 여기에서는 특수 분야 소장 자료 현황이나 구축된 기사 색인, 전자도서관이나 학술 연구 자료가 주요 군집을 이루어 지방자치단체가 공개한 데이터와는 주제적 속성에서 명확한 차이를 나타냈다.

국가 및 공공기관이 공개한 데이터는 앞의 <표 10>에서 분석한 바와 같이 지방자치단체가 공개한 데이터보다 훨씬 활용도가 높은 데이터로 검증되었다. 전국 단위 특수 전문 자료를 다루고 있기 때문에 지역 단위 도서관의 일반적인 현황보다 회소가치가 높을 뿐 아니라, 활용

〈표 15〉 지방자치단체와 국가 및 공공기관이 공개한 데이터를 기반으로 한 PFNET 분석 결과

| 지방자치단체 | | | | 국가 및 공공기관 | | | |
|--------|---------|----------------|----------------|-----------|-------|----------------|----------------|
| CL | NODE | rTBC | AVGSIM | CL | NODE | rTBC | AVGSIM |
| C1 | 마을문고 | 0.58391 | 0.02206 | C1 | 도서 | 0.35897 | 0.06143 |
| | 서적 | 0.62759 | 0.03013 | | 색인 | 0.25641 | 0.01978 |
| | 시립도서관 | 0.72644 | -0.0418 | | 지자 | 0.34615 | -0.04096 |
| C2 | 강좌 | 0.08736 | 0.00389 | C2 | 목록 | 0.4359 | -0.05964 |
| | 교육 | 0.15172 | 0.0041 | | 한국학 | 0.14103 | -0.06775 |
| | 평생교육강좌 | 0.34023 | -0.03538 | | 전자도서관 | 0.35897 | 0.05986 |
| C3 | 평생학습 | 0.04598 | 0.00503 | C3 | 국방 | 0.32051 | -0.01426 |
| | 공공도서관 | 0.56092 | -0.03804 | | 기계기술 | 0.42308 | -0.03241 |
| | 도서대출 | 0.4046 | -0.02685 | | 학술정보 | 0.17949 | -0.02084 |
| C4 | 목록 | 0.1954 | 0.03217 | C4 | 연구자료 | 0.28205 | -0.03048 |
| | 서지 정보 | 0.11034 | -0.0205 | | 자료 | 0.46154 | -0.10136 |
| | 소장자료 | 0.30115 | -0.02164 | | 고문서 | 0.57692 | -0.12512 |
| C5 | 신간 | 0.34023 | -0.01223 | C5 | 대학도서관 | 0.20513 | -0.07321 |
| | 신착 | 0.21379 | 0.02645 | | 해외 도서 | 0.28205 | -0.10004 |
| | 어린이도서관 | 0.22299 | -0.04533 | | | | |
| C6 | 자료검색 | 0.17011 | -0.0045 | | | | |
| | 추천도서 | 0.22759 | -0.00089 | | | | |
| | 희망도서 | 0.31264 | -0.00895 | | | | |
| C7 | KOLAS | 0.32414 | -0.02791 | | | | |
| | 도서정보 | 0.45287 | -0.00743 | | | | |
| C8 | 공연 | 0.54943 | -0.04267 | | | | |
| | 문화 | 0.34943 | -0.01878 | | | | |
| C9 | 박물관 | 0.54943 | -0.04267 | | | | |
| | 도서관명 | 0.12414 | -0.031 | | | | |
| C10 | 열람실 | 0.26897 | -0.01993 | | | | |
| | 휴관일 | 0.02069 | -0.03015 | | | | |
| C11 | 장난감 도서관 | 0.41609 | -0.04882 | | | | |
| | 프로그램 | 0.29885 | -0.02938 | | | | |
| C12 | 구립도서관 | 0.51954 | -0.03419 | | | | |
| | 시설 | 0.41149 | -0.04615 | | | | |
| C13 | 학교도서관 | 0.29425 | -0.04257 | | | | |

가치가 광범위하다고 할 수 있겠다. 따라서 국가 및 공공기관은 우선적으로 데이터를 공개함으로써 수혜 범위를 넓히고 재활용이 용이한 오픈포맷 형식으로 제공함으로써 융합 콘텐츠 생산을 유도해야 할 것이다.

한편, 앞에서 살펴본 바와 같이 지방자치단

체 데이터에는 공공도서관의 시설이나 보유 장서 현황, 평생학습 정보 등이 주요 주제 군집을 형성하고 있다. 이용자의 대출이나 평생학습프로그램 등록에 대한 정보는 개방 표준 항목에 누락되어 있어, 이를 통해 연령이나 지역별 조건에 따른 인기 대출도서, 평생교육프로

그램을 파악하거나 사서의 장서 개발과 프로그램 기획과 같은 의사결정에 유용하게 활용되기 어렵다. 그러나 도서관 빅데이터 분석 활용 체계 구축 사업의 일환으로 국립중앙도서관과 한국과학기술정보원이 운영하는 도서관 정보 나무(<https://www.data4library.kr/openCenter>)를 살펴보면 유사한 공공도서관의 데이터로 상이한 활용 가치가 만들어지는 것을 확인할 수 있다. 여기에서는 콘텐츠 융합과 신규 서비스 개발에 유용하게 활용될 수 있는 장서, 대출, 회원 정보 등이 개방되어 있으며 600여개의 단위 공공도서관의 데이터가 모여 빅데이터를 형성하고 있다. 이를 기반으로 다양한 분석 서비스를 제공하고 있어 사서, 민간에 의해 높은 다운로드 및 API 호출 건수로 이어지고 있다.

도서관 공공데이터 포털에 개방되어 있는 단순 현황에 대한 데이터는 민간에게 그렇게 매력적으로 다가가지 않는다. 양적 지표 충족을 위해 반기에 한번씩 의무적으로 개방되고 있는 공공도서관 데이터는 앞으로도 활용도가 크게 성장할 것으로 예측되지 않는다. 따라서 민간분야에서 필요로 하는 데이터를 발굴해 우선적으로 개방하거나 도서관 빅데이터 사업과 연계해 개방 방식을 효율화할 필요가 있을 것이다. 더불어 행정자치부에서도 공공기관 데이터 개방에 있어 정량 지표가 아니라 얼마나 실효성 있는 데이터가 전방위로 개방되어 활용되고 있는지를 주목해 사업의 방향성을 재정립할 필요가 있을 것이다.

6. 결론

정부 3.0사업의 일환으로 시작된 공공데이터

개방 정책은 국민의 공공데이터에 대한 이용권을 보장하고, 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지하기 위해 추진되고 있다. 2017년까지 2만건 이상의 데이터가 공개되고 있으나 아직까지 공개된 공공데이터를 민간에서 활용해 개발한 서비스는 데이터량 대비 5% 수준이다(디지털타임즈 2017). 이는 당초 지난 정부가 2013~2017년간 누적 14.4조원의 생산, 5.2조원의 부가가치 및 7.4만명의 고용 유발 효과를 예상했었던 바와 달리(허필선의 2013) 민간의 개발 참여도가 기대에 못 미치는 수준이라는 것이다.

본 연구에서 분석한 도서관 관련 공공데이터를 살펴봐도, 전체의 70%를 차지하는 지방자치단체와 단위도서관이 공개한 단순 도서관 현황 데이터의 평균 다운로드 횟수는 국가 및 공공기관이 개방한 전국단위 특화데이터에 비해 매우 저조한 수준이다. 또한 다양한 활용 사례가 나타나고 있는 문화체육관광부의 도서관 빅데이터 사업의 결과와 비교해 볼 때, 지방자치단체나 단위도서관이 공개한 단순 도서관 현황에 대한 데이터는 민간에게 그다지 매력적으로 다가가지 못한다.

우리나라의 공공데이터 등록은 현재 양적 측면에서는 크게 증가해 월드와이드웹재단(World Wide Web Foundation)이 평가한 오픈데이터 지표(Open Data Barometer: opendatabarometer.org)에서 71.19점으로 8위를 보이고 있다. 세부 평가 결과를 보면, 정부정책측면 98점, 정부시행력측면 98점으로 매우 우수한데 반해, 경제적 영향력 62, 사회적 영향력 45점, 정치적 영향력 60점으로 상대적으로 저조한 수치를 보이고 있다. 이와 같이 우리나라 공공데이터 개방은 매

우 편향된 평가를 받고 있는데 이는 실무적 차원에서 공공데이터 개념 정립이 미흡하고 수요자 관점에서 실효성 있는 데이터의 공개가 부족하기 때문이다(손주연 2014). 공공데이터 개방사업을 통해 2만여 건의 공공데이터가 개방됐지만, 정작 민간에서 이를 활용하여 개발한 사업은 853건에 불과한 것으로 나타나 활용도 5%의 저조한 실적에 그치고 있다. 이와 관련해 사업의 방향성 정립이 필요하다고 지적되고 있기도 하다(국회뉴스 2016).

지방자치단체나 단위도서관에서는 행정자치부의 정책에 의해 의무적으로 데이터 공개에 임하고 있으나 공공도서관의 경우 도서관 정보

나루를 통해서도 유사 데이터를 개방하고 있기 때문에 개방 방식의 일원화, 효율화가 필요하다고 본다. 또한 저조한 활용 실적 개선을 위해 민간분야에서 필요로 하는 공공데이터가 무엇인지 파악하고 이에 맞는 데이터를 우선적으로 개방하는 노력이 필요할 것이다. 한편, 전국단위의 특화 데이터를 보유한 국가 및 공공기관에서는 좀 더 적극적으로 민간의 수요가 높은 도서관 관련 데이터를 지속적으로 발굴하여 재 활용이 용이한 오픈포맷 형식으로 제공함으로써 융합 콘텐츠 생산 및 이를 기반으로 한 신규 아이디어 창출을 유도해야 할 것이다.

참 고 문 헌

- 공공데이터전략위원회. 2016. 제2차('17~'19) 국가 중점개방 데이터개방계획(안). 『제8회 공공데이터전략위원회 회의』. [online]. [cited 2018.1.17].
 <<http://www.odsc.go.kr/common/attachfile/attachfileDownload.do?attachNo=00000924>>.
- 『국회뉴스』. 2016. 이재정 “공공데이터 개방, 활용은 5%에 불과.” 7월 11일. [online]. [cited 2018.1.17].
 <<http://www.a-news.co.kr/news/articleView.html?idxno=88146>>.
- 김혜선, 김원중. 2016. 도서관 분야 데이터 개방 현황과 개선방안 연구. 『한국정보관리학회 학술대회 논문집』, 77-80.
- 김지혜. 2016. 『공공데이터 포털 분석을 통한 서비스 수준평가』, 석사학위논문. 충북대학교 대학원.
- 『디지털타임즈』. 2017. 공공데이터 개방 2만건 ‘정부3.0’ 탄력받나. 2월 1일. [online]. [cited 2018.1.17].
 <http://www.dt.co.kr/contents.html?article_no=2017020202100351041001>.
- 박옥남. 2015. 공공데이터 개방현황 및 질적 성장을 위한 고찰. [online]. [cited 2018.1.15].
 <<http://dlps.nanet.go.kr/DlibViewer.do?cn=KINX2015203503&sysid=nhn>>.
- 손주연. 2014. 공공데이터 개방 정책의 현황과 보완 방안. 『이슈와 논점』, 940호. [online]. [cited 2018.5.19]. <<https://m.nars.go.kr>>.
- 유재영. 2015. 해외 오픈데이터 정책 현황과 한국의 오픈데이터 정책 발전방안. 『한국통신학회 학술대

- 회논문집』, 529-530.
- 윤소영. 2013. 공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축에 관한 연구. 『정보관리학회지』, 30(1): 259-284.
- 이재원. 2016. 공공데이터포털 국내외 동향 분석. [online]. [cited 2018.1.17].
〈https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=25699&bcIdx=17905&parentSeq=17905〉.
- 이재원. 2006. 지적 구조 분석을 위한 새로운 클러스터링 기법에 관한 연구. 『정보관리학회지』, 23(4): 215-231.
- 이재원. 2013. tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. 『정보관리학회지』, 30(4): 241-264.
- 이현정, 남영준. 2014. 우리나라 공공데이터의 이용활성화 방안에 관한 연구: 링크드 오픈 데이터화 전략을 중심으로. 『정보관리학회지』, 31(4): 249-266.
- 한국정보화진흥원. 2013. 오픈데이터 플랫폼과 국가데이터 전략방향. 『IT & Future Strategy』, 16. [online]. [cited 2018.1.19]. 〈<http://www.nia.or.kr>〉.
- 행정자치부. 2016. 공공데이터 관리지침(행정자치부 고시 제2016-42호, 2016. 11. 22. 개정). [online]. [cited 2018.1.19].
〈https://www.data.go.kr/information/PDS_000000000000238/recsroom.do〉.
- 허필선, 박광만, 박원, 조기성, 류원. 2013. 공공정보 민간 활용 시장 및 파급효과. 『전자통신동향분석』, 28(4): 118-131.
- Alanazi, J. M. and A. T. Chatfield. 2012. "Sharing government-owned data with the public: A cross-country analysis of open data practice in the Middle East." *AMCIS*, 335-344.
- Clabo, N. and I. Ramos-Vielba. 2015. "The re-use of open data in the Spanish public administration and the application of public copyright licenses." *Revista Espanola de Documentacion Cientifica*, 38(3).
- Hadbolt, N. et al. 2012. "Linked open government data: lessons from data.gov.uk." *IEEE Intelligent Systems*, 27(3): 16-24.
- Hardy, K and A. Maurushat. 2016. "Opening up government data for Big Data analysis and public benefit." *Computer Law & Security Review*, 33(1): 30-37.
- Hendler, J., J. Holm, C. Musialek, and G. Thomas 2012. "US government linked open data: semantic.data.gov." *IEEE Intelligent Systems*, 27(3): 25-31.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Congress News. 2016. Lee Jae-jung, "Opening and Utilizing Public Data, Only 5%." 11th July. [online]. [cited 2018.1.17].
 <<http://www.a-news.co.kr/news/articleView.html?idxno=88146>>.
- Digital Times. 2017. 2million public data open 'Government 3.0' resilient. February 1. [online]. [cited 2018.1.17].
 <http://www.dt.co.kr/contents.html?article_no=2017020202100351041001>.
- Hur, P. S., K. M. Park, W. S. Park, K. S. Cho, and W. Ryu. 2013. "The private market of public information and its effects." *Analysis of Electronic Telecommunications Trends*, 28(4): 118-131.
- Kim, H. S. and W. J. Kim. 2016. "A study on the current state of data opening in library sector and improvement plan." *Proceedings of The Korea Information Management Society*, 77-80.
- Kim, J. H. 2016. *Service level assessment through public data portal analysis*. Master thesis. Graduate school, Chungbuk National University.
- Korea Information Society Agency. 2013. "Open data platform and national data strategy direction." *IT & Future Strategy*, 16. [online]. [cited 2018.1.19]. <<http://www.nia.or.kr>>.
- Lee, H. J. and Y. J. Nam. 2014. "A study on revitalizing the use of korean public data: focused on linked open data strategy." *Journal of Information Management*, 31(4): 249-266.
- Lee, J. W. 2016. Analysis of trends in public data portals at domestic and overseas. [online]. [cited 2018.1.17].
 <https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=25699&bcIdx=17905&parentSeq=17905>.
- Lee, J. Y. 2006. "A novel clustering method for examining and analyzing the intellectual structure of a scholarly field." *Journal of Information Management*, 23(4): 215-231.
- Lee, J. Y. 2013. "A comparison study on the weighted network centrality measures of tnet and wnet." *Journal of Information Management*, 30(4): 241-264.
- Ministry of Government Administration and Home Affairs. 2016. Guidelines for management of public data (Ministry of Government Administration and Home Affairs Notification No. 2016-42, Revised November 22, 2016). [online]. [cited 2018.1.19].
 <https://www.data.go.kr/information/PDS_000000000000238/recsroom.do>.

- Park, O. N. 2015. Public data opening status and review for qualitative growth. [online]. [cited 2018.1.15]. <<http://dlps.nanet.go.kr/DlibViewer.do?cn=KINX2015203503&sysid=nhn>>.
- Public Data Strategy Committee. 2016. The 2nd ('17~'19) national key open data opening plan (draft). [online]. [cited 2018.1.17]. <<http://www.odsc.go.kr/common/attachfile/attachfileDownload.do?attachNo=00000924>>.
- Sohn, J. Y. 2014. "Status and complement of public data opening policy." *Issue and Point*, 940. [online]. [cited 2018.5.19]. <<https://m.nars.go.kr>>.
- Yoo, J. Y. 2015. "International Open Data Status and Open Data Development Direction." *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, 529-530.
- Yoon, S. Y. 2013. "A study on national linking system implementation based on linked data for public data." *Journal of Information Management*, 30(1): 259-284.