도서관 공공데이터의 품질에 관한 연구: 도서관 정보나루의 도서 상세 조회 API를 중심으로*

Quality Diagnosis of Library-Related Open Government Data: Focused on Book Details API of Data for Library

양수완 (Suwan Yang)**

초 록

공공데이터의 개방과 제공의 활성화와 함께, 공공도서관이 업무 중에 생산한 서지 데이터와 대출 이력과 같은 데이터가 도서관 공공데이터로 제공되고 있다. 본 논문은 도서관 공공데이터의 품질을 진단하고, 그 결과를 바탕으로 도서관 공공데이터의 품질을 높일 개선방안을 제안하고자 한다. 먼저, 문헌정보학 영역에서 공공데이터에 관해 이루어진 연구를 개괄한다. 그다음으로, 도서관 공공데이터 개방 플랫폼인 도서관 정보나루의 오픈 API를 통해 확보한 도서관 공공데이터의 완전성과 정확성을 진단한다. 마지막으로, 데이터 품질 진단 결과에 바탕을 개선방안을 도출한다. 완전성을 진단한 결과, 도서의 식별과 검색을 위 필수적인 서지 요소에서 다수의 공백이 확인되었다. 정확성을 진단한 결과, 값의 유형, 값의 범위, 제한조건을 따르지 않는 부정확한 서지 요소가 확인되었다. 본 연구는 데이터 품질 진단 분석 결과를 바탕으로, 도서관 정보나루의 데이터 수집 절차 개선, 데이터별 스키마 구축, 데이터 수집과 데이터 처리에 관한 안내 제공, 원자료 공개를 제언하였다.

ABSTRACT

With the popularization of open government data, Library-related open government data is also open and utilized to the public. The purpose of this paper is to diagnose the quality of library-related open government data and propose improvement measures to enhance the quality based on the diagnosis result. As a result of diagnosing the completeness of the data, a number of blanks are identified in the bibliographic elements essential for identifying and searching a book. As a result of diagnosing the accuracy of the data, the bibliographic elements that are not compliant with the data schema have been identified. Based on the result of data quality diagnosis, this study suggested improving the data collection procedure, establishing data set schema, providing details on data collection and data processing, and publishing raw data.

키워드: 공공데이터, 개방 데이터, 데이터 품질, 완전성, 정확성
open government data, open data, data quality, completeness, accuracy

* 본 연구는 2018년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음.
** 중앙대학교 문헌정보학과 박사과정 수료(marma909@gmail.com)
■ 논문접수일자 : 2020년 11월 23일 ■ 최초심사일자 : 2020년 12월 3일 ■ 게재확정일자 : 2020년 12월 15일
■ 정보관리학회지, 37(4), 181-206, 2020. http://dx.doi.org/10.3743/KOSIM.2020.37.4.181

1. 서 론

오늘날 스마트폰을 통해 코로나바이러스감염증-19 지역별 감염자 수, 버스 도착 예정 시각, 미세먼지 경보, 아파트 매매 가격과 같은 정보를 확인하는 일이 일상화되었다. 이러한 정보의 골자가 되는 데이터는 공공기관이 생산하고 공개하는 데이터인 공공데이터이다. 공공데이터는 '공공기관이 법령에 따라 직무를 수행하는 과정에서 작성하거나 취득한 전자 자료'이다(공공데이터 제공 및 이용활성화에 관한 법률, 제2호 제2항, 2017). 행정업무의 효율화를 위한 정보통신기술이 공공기관의 업무에 도입되어 공공기관의 자료가 전자 형태로 생산되고 관리되자, 이용의 제약이 없는 자유로운 데이터 이용을 목표로 하는 개방 데이터 운동과 정부가 생산한 모든 자료에 대한 접근권을 보장하는 열린 정부 정책을 기반으로 삼아, 공공데이터가 등장하였다(Chignard, 2013; Schrock, 2016; Ayre & Craner, 2017).

공공데이터의 등장과 공공데이터를 가공한 정보 이용의 일상화로 인해, 공공데이터에 대한 일반의 관심이 커지고 있다. 구글(Google)에서 검색된 질의어와 주제의 관심도를 시계열로 보여주는 구글 트렌드(Google Trends)를 이용하여 대한민국의 공공데이터 주제에 대한 관심도를 살펴보면, 공공데이터 주제의 관심도는 2010년 3월 27의 관심도로 처음 발생한 이래 지속해서 상승하여 2020년 8월 기준으로 100의 관심도에 이르렀다(*공공 데이터 - 탐색 - Google 트렌드*, 2020). 구글 트렌드에서 관심도 100은 해당 질의어나 주제의 최대 검색량을 나타내는 지표이므로, 이 관심도의 추세는 공공데이터에 대한 일반의 관심이 지속해서 늘어왔다는 점을 보여준다.

일반의 관심과 동시에 공공데이터의 개방과 활용도 증대되어왔다. 민간의 공공데이터 활용을 통한 국민의 편익 증대와 일자리 창출을 도모하기 위해 2013년 「공공 데이터 제공 및 이용활성화에 관한 법률(이후, 공공데이터법)」이 제정된 이후, 6년간 공공데이터의 공개와 활용은 매우 증가하였다. 대한민국 공공기관별로 산재한 공공데이터를 통합하여 제공하는 공공데이터포털의 공공데이터의 개방 및 이용 현황을 살펴보면, 파일 데이터와 오픈 API(open Application Programming Interface; open API)를 포함한 공공데이터의 개방 건수는 2014년 13,157건에서 2019년 33,600건으로 늘었고, 같은 기간 공공데이터 활용 건수는 153,320건에서 13,141,413건으로 증가하였다(*공공데이터 개방 및 활용*, 2020).

공공데이터의 확산과 함께, 도서관 관련 공공데이터 또한 개방·활용되고 있다. 도서관 관련 공공데이터는 공공데이터포털, 서울특별시의 서울 열린데이터 광장과 경기도의 경기데이터드림을 비롯한 지방자치단체의 공공데이터 통합 제공 시스템, 국립중앙도서관이 운영하는 공공도서관 데이터 통합 제공 시스템인 도서관 정보나루를 통해 개방되고 있다.[1] 도서관 관련 공공

1) 공공데이터포털은 정부의 개방 데이터 활용 기조에 따라 개설되었다. 이와 달리, 도서관 정보나루는 도서관 종합계획에 따라 빅데이터 기술에 기반을 둔 공공도서관 데이터 활용의 목적으로 만들어졌다. 양자의 기원에 차이가 있으나, 도서관 정보나루에 참여하는 도서관의 관종이 공공도서관과 작은도서관으로 한정된다는 점, 도서관 정보나루가 공개하는 데이터가 공공기관인 공공도서관이 업무를 수행하는 과정에서 생산하는 이용자 데이터와 장서 목록과 대출 데이터에 기반을 둔다는 점을 고려하여, 이 논문에서는 도서관 정보나루를 공공도서관에 특화된 공공데이터 통합 제공 시스템으로 간주한다.

데이터는 생산 주체에 따라 공공도서관의 관계 기관이나 상급 기관이 도서관 현황을 조사한 데이터와 공공도서관이 업무를 수행하는 과정에서 생산한 공공데이터로 구분할 수 있다. 이 중 후자의 데이터는 도서관 정보나루를 통해 공개되고 있다.

도서관 정보나루는 제2차 도서관발전종합계획에 '빅데이터 기반의 도서관 서비스 기술 연구・개발'이 포함된 이후, 도서관 종합 계획의 시책을 담당하는 국립중앙도서관이 한국과학기술정보연구원과 협력하여 개발한 도서관 데이터 통합 제공 시스템이자 서비스이다(대통령 소속 도서관 정보정책위원회, 2014). 도서관 정보나루는 2020년 11월 16일 기준으로 1,177개 도서관, 122,452,521건의 장서, 29,074,714명의 회원, 1,595,580,679건의 대출 이력에 관한 데이터를 보유하고 있다(국립중앙도서관, 2020a). 도서관 정보나루는 축적한 데이터를 바탕으로 공공도서관의 장서별 대출 빈도 데이터를 개방하고, 수집한 데이터의 분석 서비스를 제공하고 있다.

도서관 정보나루는 도서관 관련 데이터에 관심이 있는 연구자, 개발자, 도서관, 민간사업자 등을 대상으로 공공도서관의 데이터와 도서관 데이터 분석을 제공한다. 공공도서관에서는 도서관 정보나루를 통해 이용이 가능한 사서 의사결정 지원서비스 솔로몬(Solomon)을 통해 공공도서관 데이터의 분석 결과를 활용하여 도서관 이용자에게 책을 추천하거나 최근 대출 동향을 분석해 인기 있는 주제의 도서를 구매하고 있다(국립중앙도서관, 2019). 비단 도서관뿐만 아니라, 민간에서도 도서관 정보나루를 통해 도서관 데이터를 활용한다. 예를 들어, 검색 엔진 네이버(Naver)는 도서관 정보나루의 도서 대출 통계 오픈 API를 이용하여, 검색된 도서관의 대출 인기도서나 요일별 또는 시간대별 대출 건수를 검색 결과와 함께 제공하고 있다(과학기술정보연구원, 2018).

다방면의 공공데이터 개방과 활용이 늘어나면서, 공공데이터의 품질에 관한 문제가 제기되고 있다. 입력자의 실수, 처리 미흡, 데이터 표준 부재로 인한 널(null) 값 오류, 형식 오류, 규칙 오류 등 다양한 공공데이터의 오류가 보고되고 있다(이원재, 김휘강, 2020). 행정안전부가 스키마를 지정해 관리하는 개방 표준 데이터 세트조차 완전성과 정확성 측면에서 품질 개선 필요성이 제기되고 있다(김학래, 2020). 공공데이터의 품질은 이용 만족과 정책에 대한 신뢰와 인과 관계를 갖기에, 공공데이터의 품질이 담보되지 않으면 민간의 공공데이터 이용 만족과 정부에 대한 신뢰를 확보할 수 없다(김현철, 2014).

이에 정부는 공공데이터법과 「공공데이터의 제공 및 이용 활성화에 관한 법률 시행령(이후, 공공데이터법 시행령)」을 비롯한 관련 법령에 근거하여 공공데이터의 효율적인 관리와 제공을 담보하는 조치를 하고 있다. 행정안전부는 공공데이터법 제11조 '공공데이터 관리지침'에 따라 『공공데이터 관리지침』을 고시하고 지속해서 개정하고 있으며, 공공데이터법 시행령 제17조 '공공데이터의 품질 진단 및 기준'에 근거하여 『공공데이터 관리지침』 내 공공데이터 품질 관리 단계별 기준을 포함함으로써 각 공공기관이 공공데이터의 품질을 확보하도록 하고 있다. 이에 더해, 행전안전부는 고시를 통해 『공공데이터 개방 표준』을 발표하여 활용도가

높은 공공데이터를 표준 데이터로 지정하여 관리하고 있다. 개방 표준을 통해 표준 데이터의 스키마를 밝히고, 공공데이터포털을 통해 표준 데이터 세트로 공개하고 있다. 또한, 행정안전부와 한국정보화진흥원은 공공데이터의 품질 향상을 위해 공공기관을 대상으로 품질 관리 수준 평가를 시행하고 있다.

그러나 문제는 도서관 관련 공공데이터가 품질 기준과 평가가 미진하다는 점이다. 대한민국의 공공데이터 통합 제공 시스템을 통해 공개되는 도서관 관련 공공데이터의 유형은 도서관 시설・운영 현황, 장서 목록, 서지 데이터, 도서관 프로그램 현황, 도서관 통계로 구분할 수 있다. 이 중 데이터 스키마가 공개된 데이터는 도서관 시설・운영 현황에 관한 표준 데이터 세트인 전국도서관표준데이터뿐이다.[2] 전국도서관표준데이터의 데이터 스키마는 해당 데이터 세트를 구성하는 26개 속성을 밝히고, 각 속성의 필수 여부, 설명, 도메인과 자료형을 지정하고 있다. 전국도서관표준데이터는 데이터 스키마가 공개되어 있으므로, 이를 품질 기준으로 삼아 품질 평가를 할 수 있다. 예를 들어, 전국도서관표준데이터의 속성 중 '도서관 유형'은 자료형이 'text'이고 열거형 값이 지정되어 있으므로 지정되지 않은 값의 유무를 통해 정확성을 확인할 수 있다. 전국도서관표준데이터 외 도서관 관련 공공데이터는 데이터와 함께 데이터 스키마가 공개되어 있지 않으므로, 데이터 품질 평가 기준이 부재하여 데이터 품질 평가를 할 수 없고 품질도 담보할 수 없다.

데이터 스키마가 없다고 하더라도, 도서관 관련 공공데이터 중 서지 데이터는 편목 규칙과 그 인코딩 방식에 따라 작성되는 서지 레코드로 구성이 되므로 공공데이터로 구성되기 이전에 서지 레코드 단위에서 품질이 담보되어야 한다. 그러나 편목의 외주화로 인해 사서는 자료와 함께 구매한 서지 레코드를 검수하게 되었으며, 전국 도서관 운영평가에서 서지 레코드의 품질이 평가 요소로 포함되지 않는다(박지영, 2016). 이로 인해, 서지 레코드의 품질 평가가 이루어지고 있지 않아, 서지 데이터의 품질을 알 수 없다.

도서관 관련 공공데이터는 데이터 생산 주체와 데이터의 유형을 불문하고, 품질의 사각지대에 놓여있으며, 이러한 상황은 현장과 학계를 가리지 않는다. 여러 학문 영역에서 공공데이터의 정량적 품질, 정성적 품질, 품질 평가 기준, 품질 관리 절차 등 품질에 관한 다양한 주제에 관한 연구가 이루어지고 있다. 그런데도, 문헌정보학 영역에서 도서관 공공데이터 품질에 관한 연구는 도서관 공공데이터 품질에 관한 연구는 확인하기 어렵다. 최근 10년간 국내 문헌정보학 영역에서 이루어진 공공데이터에 관한 연구는 도서관 관련 공공데이터의 개방 현황(김혜선, 김완종, 2016; 조재인, 2018; 한희정, 황성욱, 이정민, 오효정, 2020), 도서관의 공공데이터 활용(이정미, 2013; 표순희, 김윤형, 김혜선, 김완종, 2015; 김태영, 백지연, 오효정, 2018;

2) 전국도서관표준데이터는 행정안전부의 『공공데이터 개방 표준』에 따라 데이터 스키마가 지정된 122개 표준 데이터 중 하나로서 공공데이터포털을 통해 공개되며, 전국 공공도서관, 대학도서관, 어린이도서관, 작은도서관, 전문도서관, 학교도서관의 시설과 운영 현황에 관한 데이터를 담고 있는 데이터 세트이다(공공데이터 개방 표준 고시, 행정안전부고시 제2020-54, 2020).

온정미, 박성희, 2020; Ostler, Norlander, & Weber, 2020), 공공데이터 법제(김유승, 2014), 공공데이터의 매개로서 도서관의 역할(Robinson & Mather, 2017)을 중심으로 진행되었다. 문헌정보학 영역에서 서지 레코드의 품질을 중심으로 데이터 품질에 관한 연구가 전통적으로 이루어져 왔으나, 근래에 들어서는 연구의 명맥이 미약하다. 최근 5년간 국내 문헌정보학 영역에서 데이터 품질에 관해 이루어진 연구는 비도서 자료의 서지 레코드에 관한 연구(김우정, 이지원, 조용완, 2017)와 도서관 공공데이터 품질 평가 모형에 관한 연구(박진호, 2018) 이외에 찾아보기 어렵다.

민간에서 도서관 관련 공공데이터를 활용할 뿐만 아니라, 공공도서관에서도 도서관 정보나루와 솔로몬을 활용하여 데이터에 기반을 둔 의사결정과 서비스 제공을 하고 있다는 점은 고려하면, 도서관 관련 공공데이터의 품질은 비단 공공데이터를 이용하는 민간과 공공도서관만이 아니라 도서관 이용자 일반에 영향을 미치고, 종국에는 도서관에 대한 신뢰를 좌우한다고 할 수 있다. 도서관 관련 공공데이터의 개방과 활용을 높이고, 이용자에게 신뢰를 형성하기 위한 전제로서 도서관 관련 공공데이터의 품질에 관한 연구가 필요하다. 이에 이 논문은 선행연구로서 문헌정보학 영역에서 이루어진 공공데이터에 관한 연구와 현황을 일람하며, 도서관 관련 공공데이터 중 도서관에서 생산된 데이터를 종합하여 공공데이터로 제공되는 데이터를 도서관 공공데이터라 명명하고 도서관 공공데이터 중 도서관 정보나루에서 오픈 API로 제공하는 '도서 상세 조회' 데이터의 품질을 진단하여, 도서관 공공데이터의 품질을 개선하는 방향을 제언하고자 한다.

2. 선행연구 개관

이 글에서는 종래의 데이터 처리 도구로써 분석할 수 없었으나 컴퓨터 성능의 증대와 새로운 분석 처리 기법의 등장으로 인해 분석할 수 있게 된 대규모 데이터나 그러한 데이터를 처리하는 기술을 지칭하는 용어로 빅데이터를 언급하는 경우를 제외하고, 공공데이터를 지칭하기 위해 쓰인 '빅데이터'는 공공데이터로 치환하여 선행연구를 개관한다. 또한, 선행연구에서 사용된 '오픈 데이터', '개방 데이터', '개방형 데이터', '공개 데이터'가 민간 영역의 데이터를 포함하지 않거나 공공 영역의 데이터만을 지칭하는 경우, 해당 용어와 표현을 공공데이터로써 지칭한다.

최근 5년간 문헌정보학 영역에서는 공공데이터와 관련하여, 공공데이터 활용, 공공데이터 품질 모형, 공공데이터를 통한 새로운 도서관 모색을 위한 연구가 이루어져 왔다. 공공데이터 활용에 관한 연구는 도서관 서비스를 개선하고 의사결정을 지원하기 위한 수단으로서 공공데이터의 가치에 주목하여 이루어져 왔다. 이 연구는 공공데이터의 활용 현황과 활용 방법에 관한 연구로 이분할 수 있다. 김혜선과 김완종(2016)은 정부의 공공데이터 개방 기조에 따라 공개되고 있는 도서관 관련 공공데이터의 개방 현황을 조사하고, 조사 결과에 근거를 두어 도서관 관련 공공데이터의 활용을 촉진하는 방안을 모색하였다. 저자는 서울특별시의 서울열린데이터광장, 행정안전부의 공공데이터포털,

미국 연방 정부의 Data.gov에서 도서관과 공공도서관에 관한 검색어를 입력하여, 도서관 관련 공공데이터의 개방 현황을 조사하였다. 조사 결과, 데이터 개방과 이용 범위의 명확화, 시스템의 사용성 제고, 공공데이터의 다각화, 공공데이터의 주기적 갱신, 공공데이터 명명 규칙의 제정, 공공데이터 품질에 관한 연구가 제안되었다. 조재인(2018)은 공공데이터로 제공되는 도서관 데이터의 개방과 활용 현황을 확인하고, 데이터의 수준과 제공 주체에 따른 활용도의 차이를 분석하였다. 분석 결과, 공개 주체에 따라 활용도의 차이가 있다는 점이 확인되었다. 개별 도서관의 데이터보다 전국 단위 특화 데이터를 개방하는 국가 및 공공기관의 데이터가 활용도가 높다는 점이 드러났다. 연구자는 도서관 데이터의 이용 활성화를 위해 개방 방식의 일원화가 필요하며, 민간의 수요를 고려하여 새로운 도서관 데이터를 발굴할 필요가 있다는 점을 제언하였다. 온정미와 박성희(2020)는 도서관 정보나루와 솔로몬 등 한국과학기술정보연구원이 개발하고 국립중앙도서관이 운영하는 도서관 공공데이터 분석・활용 체계인 '도서관 빅데이터 플랫폼'의 활용 사례를 확인하고, 한밭도서관 통합 도서관 시스템과 '도서관 빅데이터 플랫폼'을 비교하여, '도서관 빅데이터 플랫폼'의 개선 방향을 도출하였다. 연구 결과, 공공도서관 외 다른 관종의 '도서관 빅데이터 플랫폼' 참여, 도서관 공공데이터와 외부 데이터를 결합한 분석 서비스 제공, 개별 도서관 맞춤 분석 서비스 기능 추가가 '도서관 빅데이터 플랫폼'의 개선방안으로 제안되었다. Ostler, Norlander, Weber(2020)는 이용자의 특성과 정보 요구가 지역마다 크게 다른 대도시에서 공공도서관의 분관이 이용자의 요구에 대응하기 위한 수단으로 개방 데이터와 오픈 소스 소프트웨어 주목하였다. 이 연구를 통해 저자는 시애틀 공공 도서관(Seattle Public Library)의 분관을 위한 시작형(試作形) 분석 도구를 개발하고, 도서관 프로그램과 도서관 서비스를 개선하기 위한 수단으로서 개방 데이터의 가치를 제시하였다.

공공도서관의 활용은 일선 공공도서관의 개선책과 연구자의 연구대상으로만이 아니라, 국가적인 과제로도 대두되고 있다. 이는 영국 정부의 공공데이터를 활용한 공공도서관 역량 강화 시도에서도 잘 드러난다. 영국의 공공도서관 정책을 총괄하는 영국 디지털・문화・매체・체육부(Department for Digital, Culture, Media, And Sports; DCMS)는 2014년 12월 『영국을 위한 독립 도서관 보고서(Independent Library Report for England)』를 발간하였다. 이 보고서는 급격한 예산 절감과 변화된 이용자 요구로 인해, 공공도서관이 더는 종래의 시설과 서비스를 유지할 수 없다는 결론을 내리고, 공공도서관이 존립하는 방법을 모색하기 위해 대책 위원회의 결성을 제언한다(Department for Culture, Media, & Sport, 2014). 이 제언에 따라, 2015년 3월 DCMS 산하에 도서관 대책 위원회(Libraries Taskforce)가 결성되었다. 도서관 대책 위원회는 공공도서관 홍보, 도서관 데이터 제공, 지속 가능한 도서관 서비스 지원, 도서관의 미래 통찰과 기술 습득 보조를 주요 업무 목표로 삼았다(*Libraries Taskforce*, 2020). 2016년 12월 도서관 대책 위원회는 『도서관 제공: 영국 내 공공도서관을 위한 포부(Libraries Deliver: Ambition for Public Libraries in

England 2016 to 2021)』라는 보고서를 발표하고 도서관에 핵심 데이터 세트의 구축이 필요하다고 주장한다. 도서관 대책 위원회는 도서관이 핵심 데이터 세트를 통해 이용자 요구 파악, 전략적 계획 수립, 미래 투자 확보와 이용 촉진을 위해 사용할 수 있는 정보 개발, 개선이 필요한 영역 식별, 효율적이고 시기적절한 방식의 일상 업무 관리를 수행할 수 있다고 보았다(Libraries Taskforce, 2016). 2017년 7월 도서관 대책 위원회는 핵심 데이터 정의 작업을 진행하여 개별 도서관(individual libraries), 이용자 정보(information on user), 행사(events), 방문(visits), 지원(staffs), 자원봉사자(volunteers), 물리적 도서와 전자책에 관한 공공 대출 보상권 정보(public lending right information for physical books and e-books), 장서(stock), 재정정보(financial), 영향의 10가지 핵심 데이터와 그 속성을 발표한다(Libraries Taskforce, 2017). 그러나 데이터 세트 중 장서와 재정정보의 속성은 정의되지 않았으며, 8개 데이터 세트는 속성의 목록만 밝힌 수준에 지나지 않았다. 도서관 대책 위원회가 업무를 종료한 2020년 3월 이후에 핵심 데이터 중 일부 데이터의 스키마가 공개되었다(Back, 2020). 이러한 영국의 도서관 역량 강화를 위한 공공데이터 활용의 도정에서 흥미로운 점은 데이터를 개방하기에 앞서, 데이터 표준 구축 작업에 역량을 집중했다는 점이다. 즉 도서관 공공데이터의 스키마와 가이드를 마련하여, 도서관 공공데이터의 품질 보장하고, 도서관 공공데이터 이용자의 소요를 가볍게 하였다. 이는 대한민국의 도서관 공공데이터 통합 제공 시스템인 도서관 정보나루가 막대한 양의 데이터를 수집하여 제공하고 있음에도 불구하고, 데이터 스키마를 공개하지 않은 바와 대조된다.

도서관 공공데이터의 개방과 활용의 확산과 함께, 도서관 공공데이터의 품질 모형에 관한 연구도 발표되었다. 박진호(2018)는 공공데이터 품질에 관한 연구가 이루어지고 있음에도 불구하고, 도서관 공공데이터의 품질에 관한 연구가 부족함을 지적하고, 델파이 기법을 통해 도서관 공공데이터 품질 측정 모형을 개발하고, 도서관 공공데이터 이용자를 대상으로 모형의 타당도와 신뢰도를 검증하였다. 품질 모형은 이용자 서비스 품질, 데이터 품질, 지원 체계 품질의 3개 차원, 18개 요인, 133개 측정 요소로 구성되며, 이 중 3개 차원, 15개 요인, 56개 측정 요소가 타당성을 확보했으며, 신뢰도는 모두 기준치인 0.6을 상회한 것으로 확인되었다.

공공데이터 활용과 품질에 관한 연구와 같이 데이터라는 시대적 화두에 대응하고 수용하려는 연구 경향과 함께, 공공데이터를 통한 공공도서관의 새로운 정체성 수집과 역량 확장에 관한 연구 또한 이루어지고 있다. Robinson과 Mather(2017)는 시민의 정보매개자(civic infomediaries)라는 개념을 통해, 공공도서관이 일반에게 공공데이터를 제공하는 새로운 소임을 수행을 수 있는지 고찰하였다. 저자는 공공도서관이 비당파적인 공공기관이라는 점에 주목하여, 공공데이터와 시민을 연결하는 구실을 할 수 있다고 보았다. 공공데이터는 시민에게도 유용하나, 시민은 공공데이터를 활용한 기술을 가지고 있지 않다. 연구자는 공공도서관은 이용자에게 공공데이터를 이용하는 데 필요한 기술에 대한 접근을 제공하여, 시민과 공공데이터를 연결하는 시민의 정보매개자이자 디지털 활동의 중심지가

될 수 있다고 주장하였다. Ayre와 Craner(2017)는 공공데이터를 이용하여 도서관 이용자의 요구를 충족하고, 공공데이터의 개방과 활용을 도우며, 직접 공공데이터를 개방하는 도서관의 특수한 위치에 주목하여, 공공도서관이 공공데이터 개방과 이용의 선도자로 자리매김할 수 있다고 주장하였다. 데이터는 정보로, 정보는 지식으로, 지식은 더 낳은 의사결정으로 이어진다. 저자는 사서가 이용자, 학생, 사업자, 지역 구성원, 공무원이 데이터와 정보를 이용할 수 있도록 돕고, 이를 통해 지역 사회, 기업, 삶에서 더 나은 결정이 이루어지도록 도움으로써, 공공도서관과 사서의 성공도 도모할 수 있다고 결론을 내렸다.

도서관과 공공데이터의 새로운 관계 정립에 대한 전망은 비단 학계만이 아니라 도서관 단체와 일선 도서관에서도 확인할 수 있다. 2020년 3월 6일 국제도서관협회연맹(International Federation of Library Associations and Institutions; IFLA)은 공식 블로그에 제10회 개방 데이터의 날(Open Data Day)을 기념하는 글을 게시했다. 개방 데이터의 날은 개방 데이터의 이점을 피로하고, 정부, 기업, 시민 사회가 개방 데이터 정책을 수용하도록 권장하기 위해 매년 개최되는 행사이다(*WHAT IS OPEN DATA DAY?*, 2020). 이 글에서 IFLA는 가능한 많은 사람에게 정보 접근을 제공하고자 하는 개방 데이터의 목표와 공중의 정보 제공을 목표로 하는 도서관 이념은 친밀하며, 공공도서관은 전통적인 업무와 서비스에 걸친 역량을 활용하여 개방 데이터의 확산을 돕고 개방 데이터의 가치를 완전한 실현에 핵심적인 역할을 할 수 있다고 이야기하였다(*Libraries and Open data*, 2020). 에드먼턴 공공도서관(Edmonton Public Library)에서는 2014년 개방 데이터의 날을 기념하여 일정 시간 안에 소프트웨어나 하드웨어를 개발하는 행사인 해커톤(hackathon)을 개최함으로써, 개방 데이터 이용의 시민 참여 기회를 부여하고, 지방 정부와 관계를 강화하며, 개방 데이터 운동을 지원하고, 도서관 데이터의 활용 기회를 만들며, 개방 데이터 공동체의 관심과 요구에 대응할 수 있게 되었다(Carruthers, 2014).

선행연구를 통해 살펴본 바와 같이 공공데이터의 등장은 공공도서관의 서비스와 의사결정을 강화할 기회일 뿐 아니라, 공공데이터의 이용자, 생산자, 중개자, 가공자로서 도서관의 정체성을 새롭게 수립할 기회이기도 하다. 이 기회를 통해 도서관은 매체의 형태와 정보의 종류에 상관없이 정보 일체를 수집, 보존, 조직함으로써 이용자에게 제공하는 문헌정보학과 이념을 구현할 수 있다. 이를 위해서 비단 종이책 중심의 문헌만이 아니라 데이터를 아우르는 문헌의 전역에 대응할 수 있는 도서관의 역량을 마련하여야 한다.

그러나 도서관 자체의 자원으로서 도서관이 생산된 데이터의 품질을 담보할 역량이 없다면, 도서관의 공공데이터 대응은 언제나 피동적인 영역에 멈출 수밖에 없다. 이에 이 논문은 공공데이터 일반의 정책이나 서비스 방식이 아니라, 도서관 공공데이터 자체의 품질을 연구대상으로 한다. 완전성과 정확성 같은 공공데이터의 정량적 품질 요소에 기반을 두어, 도서관 공공데이터의 품질 확인하고, 도서관 공공데이터 품질을 담보하는 방안을 도출하는 바가 이 글의 목적이다.

3. 연구 방법

3.1 분석 대상

이 논문은 도서관 공공데이터의 품질 진단을 위해 대한민국의 도서관 공공데이터 통합 제공 시스템인 도서관 정보나루의 데이터 중 오픈 API로 제공되는 '도서 상세 조회' 데이터를 분석 대상으로 삼았다. 그 이유는 '도서 상세 조회' 데이터가 서지 기술 요소를 포함하고 있으며, 개별 도서관의 서지 데이터가 아니라 도서관 정보나루가 축적한 서지 데이터를 확인할 수 있기 때문이다. 서지 기술 요소를 포함한 데이터는 『한국목록규칙 제4판(Korean Cataloguing Rules 4th edition; KCR4)』과 그 인코딩 방식인 『한국문헌자동화목록형식(KORean Machine Readable Cataloging format; KORMARC)』을 참조하여 품질 진단이 가능하다.

도서관 정보나루가 제공하는 데이터의 유형은 〈표 1〉과 같다. 도서관 정보나루가 제공하는 데이터 중 서지 레코드를 포함하는 데이터는 '장서/대출 데이터', '인기대출도서', '도서별 이용분석', '대출 급상승 도서', '오픈 API' 중 '도서 상세 조회' 데이터이다. 이 중 '도서별 이용분석'과 '대출 급상승 도서'는 개별 도서의 서지 레코드를 표시하며, '인기대출도서'는 단위 기간 특정 인구통계학적 집단이 많이 대출한 상위 200권의 서지 레코드만을 담고 있다. 도서관 정보나루의 데이터 중 10,000건 이상의 서지 레코드를 확보할 수 있는 데이터는 '장서/대출 데이터'와 '도서 상세 조회' 데이터뿐이다. 이 중 전자는 개별 도서관의 서지 레코드와 해당 자료의 대출 누적 건수를 병합한 데이터이고, 후자는 도서관 정보나루가 국립중앙도서관, 공공도서관, 서점 등 외부 협력 기관으로부터 수집한 서지 데이터라는 차이가 있다. 개별 도서관의 데이터보다 공공기관의 전국 단위 데이터의 활용도가 높으므로 이 글에서는 '도서 상세 조회' 데이터를 대상으로 품질 진단을 진행한다(조재인, 2018).

〈표 1〉 도서관 정보나루 제공 데이터 유형

대분류	소분류
공개 데이터	참여 도서관 목록
	장서/대출데이터
	인기대출도서
	도서별 이용분석
	대출 급상승 도서
테마 데이터	이슈별 테마
	지역별 테마
	이용자별 테마
	테마 데이터 요청
데이터 활용	오픈 API

'도서 상세 조회' 데이터는 API를 통해 요청한 ISBN에 해당하는 도서의 서지 레코드를 담은 데이터이다. '도서 상세 조회'는 도서관 정보나루의 서지 데이터에 포함된 레코드를 반환하며, 도서관 정보나루의 서지 데이터는 국립중앙도서관, 공공도서관, 서점 등 외부 협력 기관에서 서지 레코드를 수집하여 구축된다. '도서 상세 조회' API의 요청 변수와 응답 메시지 필드는 〈표 2〉와 같다.

'도서 상세 조회' API는 'authKey'와 'isbn13'을 필수 요청 변수로 한다. 응답 메시지는 서지에 관한 11개 필드와 대출에 관한 3개 필드로 구성된다. 'loaninfoYN'를 입력하지 않은 경우, 대출에 관한 필드는 응답 메시지에 포함되지

〈표 2〉 도서 상세 조회 API의 요청 변수와 응답 메시지 필드

요청 변수	설명
authKey	인증키
isbn13	13자리 ISBN
loaninfoYN	대출 필드 포함 여부
displayinfo	대출 조회 대상
format	응답 유형
응답 메시지 필드	설명
no	순번
bookname	도서명
publication_date	출판 일자
authors	저자명
publisher	출판사
class_no	분류기호
publication_year	출판연도
bookImageURL	이미지 URL
isbn	ISBN
isbn13	13자리 ISBN
description	책 소개

않는다. 응답 메시지는 'format' 요청 변수를 통해 XML 형식이나 JSON 형식으로 반환된다. 응답 메시지 필드의 'no'는 응답 메시지의 순번이다. 예를 들어, 요청 필드에 세 개 ISBN을 입력한 경우, 순서에 따라 응답 메시지의 순번이 정해진다. 그러나 '도서 상세 조회' API에서 요청 변수 'isbn13'의 값으로 복수의 ISBN과 변수 구분자(delimiter)를 입력하면, 오류가 반환된다. 그러므로 'no'는 1 이외의 값을 갖지 않는다. 응답 메시지 필드 'bookImageURL'은 도서 앞표지의 이미지 파일 URL이다.

3.2 데이터 확보

'도서 상세 조회' 데이터를 확보하기 위해 요청 메시지의 필수 요청 변수로서 13자리 ISBN이 필요하다. ISBN의 변수 목록을 작성하기 위해 도서관 정보나루의 '테마 데이터 요청'을 통해 공공도서관의 대출 데이터를 요청하였다. 대출 데이터에서 ISBN을 추출한 이유는 대출의 요건 때문이다. 공공도서관에서 도서가 대출되기 위해서는 공공도서관이 실물 자료를 확보한 후 서지 레코드를 작성하고 청구기호에 따라 배가하여야 하며, 이용자가 온라인 열람 목록을 통해 서지 레코드를 확인하거나 서가를 훑어보고 자료에 접근하여 실물 자료를 확인하여야 한다. 즉, 도서의 대출은 온전한 서지 레코드가 존재함을 나타내기 때문에, 서지 기술 요소의 품질 진단 과정에서 완전성을 논할 수 있다. 예를 들어, 실제 대출된 도서의 ISBN에 해당하는 서지 레코드에 분류번호가 없다면, 이는 해당 서지 레코드의 품질에 문제가 있다는 방증이 된다.

'테마 데이터 요청'은 도서관 정보나루가 이용자의 요청에 따라 데이터를 추출하여 제공하는 서비스이다. 대출 데이터를 요청한 8개 공공도서관은 국가도서관통계시스템의 2018년 공공도서관 통계에서 연간 대출자 수사 가장 많은 10개 도서관 중 도서관 정보나루에 참여하는 8개 도서관이다(*공공도서관 통계 보기*, 2019). 2019년 11월 12일에 8개 도서관의 대출 데이터를 요청하여, 2019년 12월 2일에 8개 공공도서관의 대출 데이터를 받았다.[3)]

대출 데이터에서 8개 도서관의 명칭, 도서 식별자(book identifier), 이용자 식별자(user identifier)는 비식별처리 되어 있다. 데이터는

3) 해당 대출 데이터는 도서관 정보나루의 테마 데이터 요청 게시글에서 확인할 수 있다. https://www.data4library.kr/userThemaCallInfo?thema_no=6

〈표 3〉 도서관별 대출 데이터 구성

도서관	대출자	대출도서	대출	대출일 범위	대출일수
1	31,087	90,701	563,465	2017-10-20~2018-12-31	405
2	61,396	246,520	3,558,791	2009-01-02~2018-12-31	3,008
3	22,435	64,954	822,120	2016-06-22~2018-12-31	896
4	53,798	125,686	2,420,156	2010-02-07~2018-12-31	2,611
5	16,401	77,076	380,181	2018-01-23~2018-12-31	301
6	35,514	91,717	1,275,386	2015-10-28~2018-12-31	1,023
7	122,042	299,126	5,295,245	2009-01-02~2018-12-31	3,385
8	111,505	283,463	4,852,377	2009-01-02~2018-12-31	3,428

CSV(Comma-Separated Value) 형식의 71개 텍스트 파일로 분절되어 있어, ISBN을 추출하기 위해서는 텍스트 파일을 병합하고 행렬 구조로 변형하는 작업이 필요하다. 이에 데이터 정제(data wrangling)를 위한 오픈 소스 애플리케이션인 오픈리파인(OpenRefine)을 이용하여 데이터를 병합하였다. 대출 데이터의 도서관별 구성은 〈표 3〉과 같다. ISBN을 추출하기 위해 통계 분석을 위한 프로그래밍 언어인 R을 이용하여 19,177,721건의 대출 이력에서 453,136개의 고유한 ISBN을 추출하였다.

'도서 상세 조회' API가 제공하는 응답 메시지 형식 XML과 JSON 중 행렬 형식으로 변환하기 쉬운 JSON 형식을 응답 유형으로 지정하고 453,136개 ISBN을 포함한 요청 URL을 만들었다. R의 JSON 파일 형식 분석 패키지인 jsonlite를 이용하여 '도서 상세 조회' API로부터 받은 453,136개 응답 메시지를 연결해 행렬 형태로 만들었다.

4. 데이터 품질 분석

박진호(2018)의 품질 측정 모형 내 품질 요인 중 완전성과 정확성을 데이터 품질 진단 요소로 선정하였다. 품질 요인은 이용자 서비스 품질, 데이터 품질, 지원 체계 품질의 3개 차원으로 구성되며, 각각의 차원은 이용, 데이터, 정책에 대응한다. 이용자 서비스 품질 차원은 공공데이터 플랫폼을 통해 제공되는 이용자 편의 기능과 상세 정보의 유무와 수준에 관한 측정 요소로 구성된다. 데이터 품질 차원은 데이터의 재사용과 재배포 정책, 데이터 자체의 품질, 데이터의 최신성에 관한 측정 요소로 이루어진다. 지원 체계 품질 차원은 데이터 제공 책임 주체, 보안 제도, 이용 참여의 여부와 수준을 진단하는 측정 요소를 포함한다. 세 개 차원 중 데이터 정량적 품질을 확인할 수 있는 데이터 품질 차원이다. 다른 차원의 요인과 측정 요소는 공공데이터 제공 절차와 체계를 아우르는 정성적 품질과 관련되어, 객관적 품질 진단이 제한된다. 데이터 품질 차원의 요인은 재사용성, 완전성, 적시성, 정확성이며, 이 중 완전성과 정확성은 확보한 데이터의 값과 특성을 통해 품질을 진단할 수 있다. 완전성의 측정 요소 중 '데이터값이 누락 없이 입력되어 있는가?'를 확인하기 위해, 측정 데이터 내 공백인 셀과 공백이 아닌 셀의 개수를 확인하여 완전성 지수

를 도출하였다. 완전성 지수는 전체 셀 중 공백이 아닌 셀의 비중으로 계산된다(김학래, 2020). 정확성은 측정 요소인 '엔티티와 속성에 맞는 정확한 값이 입력되었는가?'와 '데이터 모순 없이 일관된 속성을 적용하였는가?'는 각 속성의 자료형, 값의 범위, 제한규칙 등의 도메인을 통해 확인하였다.

데이터 품질 진단은 R을 이용하여 진행하였다. 진단 대상인 데이터는 453,136개 레코드와 12개 속성으로 구성된다. 각 레코드는 '도서 상세 조회' API의 응답 메시지에 해당한다. 12개 속성 중 첫 번째 속성은 요청 메시지의 필수 요청 변수 'isbn13'이며, 그 외 11개 속성은 '도서 상세 조회' 응답 메시지의 서지 관련 11개 응답 메시지 필드이다. 요청 변수 'isbn13'을 응답 메시지 필드의 'isbn13'과 구별하기 위해 'isbn13_request'라고 명명하였다. 표현의 편의를 위해, 이후 'isbn13_request'는 'A1'으로, 'no'는 'A2'로, 'bookname'은 'A3'로, 'publication_date'는 'A4'로, 'authors'는 'A5'로, 'publisher'는 'A6'로, 'class_no'는 'A7'으로, 'publication_year'는 'A8'으로, 'bookImageURL'은 'A9'으로, 'isbn'은 'A10', 'isbn13'은 'A11'로, 'description'은 'A12'로 치환하여 기술한다.

전체 453,136개 레코드 중 2,674개 레코드는 오류 메시지이다. 응답 메시지에 "ISBN을 확인해 주시기 바랍니다."라는 오류 메시지 외에 오류 코드가 포함되어 있지 않아, 오류 메시지가 응답한 정확한 원인은 파악할 수 없다. 요청 변수 'isbn13'의 값으로 문자열을 넣으면 같은 오류 메시지가 반환되므로, '도서 상세 조회'와 연결된 데이터베이스나 테이블에 해당 ISBN 값을 가진 레코드가 없으므로 오류 메시지가 반환된 것으로 추정된다. 오류 메시지가 반환된 ISBN 중에는 도서관 정보나루에서 검색할 수 있는 ISBN이 포함되어 있다. 예를 들어, 오류 메시지를 반환한 ISBN '9788957363192'는 도서관 정보나루의 '도서별 이용분석' 서비스에서 『(만화로 보는) 그리스 로마 신화』 19권의 ISBN으로 확인된다(국립중앙도서관, 2020b).

이에 관해 도서관 정보나루를 운영하는 도서관 빅데이터 사무국에 문의한 결과, 도서관 정보나루의 서지 데이터베이스는 국립중앙도서관의 국가자료종합목록을 중심으로 하여 공공도서관과 서점 등의 서지 데이터를 수용하였으며, 도서관 정보나루의 모두 서비스에 공통되게 사용된다는 답을 받았다. 이 답변을 고려하면, 오류 메시지를 반환한 것은 해당 API의 일시적인 오류로 추정할 수 있으나, 시간을 두고 오류 메시지를 반환한 ISBN을 '도서 상세 조회' API로 조회한 결과 같은 오류 메시지가 반환되었다. 현상과 도서관 정보나루의 답변이 상이하여, 오류가 반환된 원인을 확인할 수 없다. 이에 오류 메시지가 담긴 레코드를 제외하고, 450,461개 레코드를 품질 진단 대상으로 삼아 분석을 진행하고자 한다.

4.1 데이터 완전성

데이터의 완전성을 측정하기 위해 공백의 수를 확인하였다. 속성별 공백의 개수는 〈표 4〉와 같다. 데이터의 전체 5,405,532개 셀 중 공백인 셀은 236,462개이며, 공백이 아닌 셀은 5,169,070개이다. 확보한 '도서 상세 조회' 데이터의 완전성 지수는 0.95625555449이다. 12개 속성 중 A1, A2, A10, A11에는 공백이 포함되지 않는다. A2

는 요청의 순차를 나타내는 속성으로 모든 레코드가 숫자 '1'의 값을 갖는다. A1, A2, A11은 ISBN에 해당하는 속성이다. 반응 메시지가 오류를 반환하지 않은 경우, A1에 응답 메시지에 요청 변수의 ISBN이 포함된다. A10은 10자리 ISBN이며 A11은 13자리 ISBN이다.

〈표 4〉 속성별 공백의 개수

속성	공백	속성	공백	속성	공백
A1		A5	2,610	A9	47,833
A2		A6	9,491	A10	
A3	15	A7	69,833	A11	
A4	21,916	A8	21,916	A12	62,847

공백이 포함된 속성은 A3, A4, A5, A6, A7, A8, A9, A12이다. A9은 표지의 사진 파일의 URL이다. A3, A4, A5, A6, A8은 서지 요소와 관련되며, A7은 분류기호이다. 각 속성의 대응하는 서지 요소를 살펴보면, A3는 표제, A5는 책임표시, A6는 발행처, A8은 발행년이다. KCR4와 KORMARC에서 서지 요소로서 발행일을 다루지 않기 때문에, 발행일을 나타내는 A4는 국립중앙도서관이나 공공도서관에서 작성한 서지 레코드에서 비롯되었다고 볼 수 없다. A4는 서점 등 외부 협력 기관의 데이터를 수집한 과정에서 추출한 속성으로 추정된다. A7은 KORMARC의 '한국십진분류기호'나 '듀이십진분류기호' 필드에 해당한다.

한국의 국가도서관이나 공공도서관의 서지 레코드는 KCR4와 KORMARC에 따라 작성된다. KORMARC에서 A1, A3, A4, A5, A6, A7, A8, A10, A11에 대응하는 필드의 적용수준은 '필수' 또는 '해당시필수'이다. 그러므로 도서관정보나루의 '빅데이터 분석 플랫폼'에서 수집한 국립중앙도서관과 공공도서관의 서지 레코드에 해당하는 속성이 반드시 포함된다. 도서관정보나루의 '빅데이터 분석 플랫폼'이 서지 레코드를 수집하는 과정에서 해당 속성이 빠지어 공백이 발생할 가능성은 적다. 그런데도 해당 속성에 공백이 포함된 이유는 '빅데이터 분석 플랫폼'이 공백이 포함된 불완전한 서지 레코드를 수집한 바로 추정된다. 예를 들어 〈표 5〉와 같이 서명이 공백인 15개 레코드를 살펴보면, 서명만이 아니라 여러 속성의 값이 공백이며, 실질적으로 서지 레코드로서 역할을 할 수 없는 레코드다. 이는 국립중앙도서관은 물론이고 공공도서관에서 생산되었다고 보기 어려운 레코드다.

도서관 정보나루가 제공하는 데이터에 불완전한 서지 요소가 포함된 원인은 데이터를 수집하는 절차에 문제가 있기 때문으로 판단된다. 예를 들어, ISBN이 '9791170262213'인 레코드는 ISBN을 제외하고 모든 속성의 값이 공백이지만, 국립중앙도서관과 국가자료종합목록에서 해당 ISBN의 MARC 레코드를 확인할 수 있기 때문이다. 이 점을 고려하면, 서지 레코드의 수준을 고려하지 않고 '빅데이터 분석 플랫폼'의 데이터 수집 절차를 구성하여, 불완전한 서지 레코드를 먼저 수집한 후 완전 수준의 서지 레코드를 수집하지 않은 바로 추정된다.

4.2 데이터 정확성

데이터 정확성을 확인하기 위해 12개 속성 중 A1, A4, A7, A8, A10, A11의 도메인 준수 여부를 확인하였다. 이 6개 속성의 정확성은 값

〈표 5〉 서명이 공백인 15개 레코드

	A1	A3	A4	A5	A6	A7	A8	A11
1	9791170262213							9791170262213
2	9791162338858					31		9791162338858
3	9791158511142							9791158511142
4	9791186951125		2018	최용호		991.184	2018	9791186951125
5	9791187790297		2017	이혜승 외 7명			2017	9791187790297
6	9791186324554		2017	서울역사박물관[편]		238.2	2017	9791186324554
7	9791186463369							9791186463369
8	9791196196721							9791196196721
9	9780544743366			written and illustrated by Brian Lies		843.6		9780544743366
10	9788968735950							9788968735950
11	9788968735967							9788968735967
12	9788968736025							9788968736025
13	9789997311252		2015	Маурис Сендак		843	2015	9789997311252
14	9791157234233							9791157234233
15	9791188454235							9791188454235

자체의 진위를 확인할 수 없어도, 각 속성의 길이, 자료형, 제약조건 등의 도메인 준수 여부를 통해 진단할 수 있기 때문이다. 예를 들어, A11은 13자리 ISBN을 값으로 가져야 하므로, 자릿수는 13이어야 하고, 숫자 이외의 문자는 포함되어서는 안 된다. 그러므로, A11이 13자리의 숫자만을 값으로 갖는지 확인하여, 해당 속성의 정확성을 확인할 수 있다.

국제표준도서번호의 자릿수는 10자리와 13자리로 고정되므로, ISBN 관련 3개 속성의 정확성을 진단하기 위해 각 값의 자릿수를 확인하였다. A1, A10, A11의 3개 속성에서 10자리와 13자리 이외의 자릿수를 갖는 값은 확인되지 않았다. A1은 10자리 1개와 13자리 450,460개, A10은 10자리 450,461개, A11은 13자리 450,461개 값으로 구성된다.

A1과 A11은 10자리 ISBN 1개 값을 제외하고 13자리인 450,560개 값 모두 일치한다. A1의 10자리 ISBN '8952701267'은『모자 사세요!』라는 동화책의 10자리 ISBN이며, 이 도서의 13자리 ISBN은 '9788952701268'이다. 해당 레코드의 A11 값은 '9788952701268'이므로, 해당 레코드에서 A1과 A11을 같은 도서의 ISBN을 나타낸다.

GS1이 관리하는 상품의 바코드와 숫자 부여 표준인 국제상품번호 체계 EAN-13에 따라, 13자리 ISBN의 앞 세 자릿값은 '978'이나 '979'이다. 확보한 데이터 내에서 13자리 ISBN을 갖는 A1과 A11은 같은 자료의 ISBN을 가지기 때문에, A11의 앞 세 자리가 '978'이나 '979'를 갖는지 점검하였다. 앞 세 자릿값이 '978'이나 '979'인 값은 450,461개 값 중 447,055개이며, 3,406개 값은 다른 세 자릿값을 가진다. 세 자릿값의 빈도는 〈표 6〉과 같다. 빈도가 10 이상인 세 자릿값은 '978', '979', '880', '893', '977', '200', '919'이다.

〈표 6〉 ISBN 앞 세 자리 코드별 빈도

코드	000	004	005	007	009	078	080	082	111	112
빈도	1	2	1	6	2	1	7	1	2	3
코드	119	140	191	200	201	202	209	250	330	372
빈도	1	3	1	16	4	1	1	1	1	3
코드	386	390	456	479	505	516	642	788	839	862
빈도	1	1	1	1	4	1	1	1	1	1
코드	877	879	880	885	893	894	898	899	918	919
빈도	1	5	3,121	3	154	2	1	1	4	15
코드	934	943	972	977	978	979	992			
빈도	2	1	3	17	401,981	45,074	5			

상기한 바와 같이 '978'과 '979'는 ISBN을 나타내며, '880'과 '893'은 GS1 국가 코드로서 각각 대한민국과 베트남을 의미한다. A11의 값이 '880'으로 시작하는 레코드는 영화나 애니메이션의 DVD, 오디오북 CD 등의 레코드이다. A11의 앞 세 자리가 '893'인 레코드는 베트남판 『백설 공주와 일곱 난쟁이(Nàng Bạch Tuyết và bảy chú lùn)』를비롯한 베트남 자료에 해당한다. '977'은 국제표준연속간행물번호(International Standard Serial Number; ISSN)를 나타내는 GS1 코드이다. '200'은 RCN(Restricted Circulation Number) 중 하나로 'GS1 회원기관이 지정한 제한된 환경에서 특정 목장을 위해 사용'하는 코드로서 단행본이나 연속간행물을 위한 코드가 아니다(*GS1 표준 사용자 매뉴얼*, 2015, p. 53).

A1, A10, A11의 값에 숫자 이외 문자가 포함되었는지 점검하였다. 10자리 ISBN에서 체크 기호(check digit) 10을 대신해 사용되는 X를 제외하고, ISBN은 모두 숫자로 구성되어야 한다. A1과 A12에서 숫자 이외의 값은 확인되지 않았다. A11의 레코드 중 409,564개 레코드는 숫자로만 이루어지며, 40,897개 레코드는 숫자와 문자로 이루어진다. 숫자와 문자로 이루어진 40,879개 레코드에서 문자는 10번째 자리에만 위치하며, 문자의 종류와 개수는 'X' 40,819개, 'x' 54개, '*' 15개, ':' 9개이다.

ISBN의 오류를 검출을 위해 체크 기호를 계산하고, 마지막 자리 숫자와 비교하였다. 10자리 ISBN을 '$d_1d_2d_3d_4d_5d_6d_7d_8d_9d_{10}$'으로 표현할 때, 10자리 ISBN의 체크 기호 d_{10}은 d_1부터 d_9까지 수에 각각 10부터 2까지의 양의 정수를 곱하고 모두 더한 후, 이를 11로 나누고, 그 나머지를 11에서 뺀 후, 그 값을 11로 나눈 나머지이다(*International Standard Book Number*, 2020). 10자리 ISBN의 체크 기호 d_{10}을 계산하는 수식은 아래와 같다.

$$d_{10} = \left\{11 - \left[\left(\sum_{i=1}^{9} d_i \times (11-i)\right) \bmod 11\right]\right\} \bmod 11$$

'$d_1d_2d_3d_4d_5d_6d_7d_8d_9d_{10}d_{11}d_{12}d_{13}$'으로 13자리 ISBN 표현할 때, ISBN의 체크 기호 d_{13}은 마지막 자리를 제외한 d_1부터 d_{12}까지 수 중 홀수 번째 수에 각각 1을 곱하고 짝수 번째 수에 각각 3을 곱하여 모두 더한 후, 이를 10으로 나눈 나머지로 10을 뺀 값이다(*International Standard*

Book Number, 2020). 13자리 ISBN의 체크 기호 d_{13}을 구하는 수식은 아래와 같다.

$$d_{13} = 10 - \left\{ \left[\sum_{i=1}^{6} (d_{2i-1} \times 1) + \sum_{i=1}^{6} (d_{2i} \times 3) \right] \bmod 10 \right\}$$

ISBN 관련 세 개 속성의 체크 기호를 계산하고, 마지막 자리의 체크 기호와 비교하였다. 'x'는 'X'와 같게 취급하였다. A1은 10자리 ISBN '8952701267'을 제외하고 모두 13자리 ISBN을 갖는다. A1의 13자리 ISBN의 체크 기호 계산 결과, 13번째 자리의 체크 기호와 모두 일치하는 바가 확인되었다. 10자리 ISBN이 기재된 A10의 450,461개 값 중 450,089개 값의 마지막 자리에 있는 체크 기호와 ISBN 체크 기호 계산 결과가 같다. 계산 결과와 체크 기호가 다른 값은 372개이다. 국립중앙도서관의 서지정보유통지원시스템에 ISBN 통보서 신청 과정에서 체크 기호는 자동 계산되므로, A10의 부정확한 체크 기호는 편목자의 오기로 추정된다. A11의 13자리 ISBN의 체크 기호와 체크 기호 계산 결과는 모두 일치한다. 10자리 ISBN의 체크 기호를 제외하고 A1과 A11의 체크 기호 종류와 빈도는 같다.

A4는 발행일자를 나타내는 속성이며, A8은 발행년을 나타내는 속성이다. 두 속성의 자릿수를 확인하였다. A4에 날짜와 시간의 데이터 표준인 ISO 8601에 따라 발행일자가 기재되는 경우, YYYY-MM-DD 형식에 따라 자릿수가 10이어야 한다. 곽 괄호([ ]) 부기가 없으면, 발행년은 대부분 4자리 아라비아 숫자이다. A4의 공백을 제외한 428,546개의 값 중 1,487개 값은 자릿수가 10이고, 427,059개 값은 자릿수가 4이다. A8의 공백을 제외한 428,546개 값은 모두 자릿수가 4이다. 4자리 값은 '2001'과 같이 아라비아 숫자로 이루어진 서력 기년 형식이고, 10자리 값은 '2008-04-15'와 같이 YYYY-MM-DD 형식이다.

두 속성의 값에 숫자와 '-' 이외 다른 문자가 포함되었는지 확인하였다. 두 속성은 서력 기년 형식과 ISO 8601의 일자 형식을 따르기 때문에, '년'이나 'December'와 같은 문자가 포함되지 않아야 한다. 한 레코드를 제외하고 A4와 A8의 모든 값은 숫자와 '-'로 구성되어 있다. 예외인 레코드는 13자리 ISBN인 '9780451527295'인 가진 양서이며, A4와 A8의 값으로 문자열 'June'을 갖는다.

두 속성의 공백 개수가 일치하며, 두 속성은 대부분 자릿수가 4인 값으로 구성되어 있다. 이 점을 고려하면 두 속성은 같은 발행년을 갖고, 때에 따라, A4에 발행일자가 덧붙여 있을 가능성이 크다. 이에 두 속성의 값 앞 네 자리가 일치하는지 확인하였다. A4에서 10개 자리로 이루어진 값의 앞 4자리를 분리한 후, A4와 A8의 값을 비교한 결과, A4의 앞 네 자릿값과 A8의 값이 모두 같다는 점을 확인하였다. 즉, 두 속성은 같은 발행년을 가지며 때에 따라 A4에 발행일자가 덧붙여 있는 형태이다.

이어서 값의 형식에 따라 서력 기년 형식 발행년과 YYYY-MM-DD 형식 발행일자의 기술 통계량을 확인하였다. 기술 통계량은 최솟값, 제1사분위수, 중앙값, 평균, 제3사분위 , 최댓값을 산출하였다. 기술 통계량은 〈표 7〉과 같다. 발행년과 발행일자는 시계열에 따라 나열되므로, 6가지 기술 통계량을 확인하여 이상치를 확인할 수 있다. 기술 통계량을 확인하기 위

해 발행일의 자료형을 문자열에서 정수로 변경하고 발행일자의 자료형을 문자열에서 날짜로 변경하였다. 문자열을 정수로 변경하는 과정에서 자리를 채우기 위한 숫자가 사라졌다. 예를 들어, 0000은 0이 되었다. 확인 결과 이상치가 검출되었다. 이 글의 데이터는 2009년 1월 2일부터 2018년 12월 31일까지의 대출 이력에서 ISBN을 추출하여 '도서 상세 조회' API에서 데이터를 확보하였으므로, 발행년이 2020년인 자료의 레코드가 있을 수 없다. 또한, 0년이나 서력 원년에 발행된 자료가 공공도서관에서 대출되었을 가능성도 극히 낮다.

〈표 7〉 발행년과 발행일자의 기술 통계량

기술 통계량	발행년	발행일자
최댓값	2020	2018-08-21
제3사분위수	2014	2015-01-22
평균	2008	2005-03-25
중앙값	2009	2013-01-03
제1사분위수	2005	2009-08-29
최솟값	0	0001-01-01

발행년 빈도를 〈표 8〉로 정리하고, 발행년이 광복 이전인 282개 레코드와 발행년이 2019년과 2020년인 166개 레코드를 검토하고 이상치 발생의 원인을 추정하였다. 정확하지 않은 발행년이 확인된 이유 중 하나는 대상 자료의 발행년이 아니라 원전의 발행년이 기재되어 있기 때문이다. 예를 들어, ISBN이 '9780060821418'인 도서는 하퍼토치(HarperTorch) 출판사가 2005년 발행한 올더스 헉슬리(Aldous Huxley)의 『멋진 신세계(Brave New World)』인데, 해당 ISBN의 레코드에는 발행년으로 『멋진 신세계』의 최초 발행년인 '1932'이 기재되어 있다. 두 번째 이유는 ISBN과 다른 서지 요소가 일치하지 않기 때문이다. 예를 들어, ISBN이 '9788934900177'인 레코드에 기재된 서명은 '스토브리그 = 이신화 대본집 /Stove league'이다. 이신화의 『스토브리그 대본집』은 1권과 2권이 있으며, 각각의 ISBN은 '9788934900184'와 '9788934900191'이다. 두 ISBN 모두 서지 레코드의 ISBN과 일치하지 않는다. 즉, 서지 레코드에 ISBN을 잘못 입력한 것이다.

마지막으로 분류기호를 나타내는 A7을 확인하였다. 국립중앙도서관과 공공도서관에서

〈표 8〉 발행 연도별 빈도

발행년	0000	0001	1900	1902	1903	1930	1932	1950	1952	1953
빈도	1	122	154	1	1	2	1	2	1	2
발행년	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973
빈도	4	6	5	5	7	3	9	9	8	11
발행년	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
빈도	22	21	31	41	68	51	98	103	141	237
발행년	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
빈도	3,975	5,531	5,839	7,588	9,256	11,088	13,613	15,896	16,977	21,131
발행년	2012	2013	2014	2015	2016	2017	2018	2019	2020	
빈도	23,131	22,652	24,607	25,305	26,168	21,935	15,766	95	71	

동양서의 분류기호는 한국십진분류법(Korean Decimal Classification; KDC)에 근거하여 할당된다. 한국십진분류법은 주류, 강목, 요목의 순으로 전개되는 십진식 분류법이므로, 최소 3의 자릿수를 가진다. 또한, 세목을 전개하는 경우 소수점을 포함하므로, 자릿수가 4인 분류기호는 있을 수 없다. 그런데도, 공백을 제외한 A7의 380,629개 값 중 자릿수가 3 미만이거나 4인 값이 31개가 있다. 'T'나 '아'와 같이 별치기호로 추정되는 값이나, '808/'이나 '410'과 같이 분류기호에 오자가 포함된 값이 이에 해당한다.

4.3 소결

완전성을 측정한 결과, 450,461개 레코드의 5,405,532개 셀 중 공백인 셀 236,462개로 확인되었다. 이는 전체 셀 중 공백이 아닌 셀의 비중인 완전성 지표로 0.9563에 해당한다. 행정안전부의 고시에 따라 데이터 세트와 스키마가 함께 제공되는 개방 표준 데이터의 완전성 지표 평균 0.88이다(김학래, 2020). 이 점을 고려하면, 스키마가 부재한 '도서 상세 조회' 데이터의 완전성 지표가 높다고 할 수 있다. 하지만, 이 데이터가 서지 레코드에 기초하고 있음을 고려하면 완전성이 양호하다고 볼 수 없다. 일례로 450,461개 레코드 중 69,832개 레코드에 분류기호가 공백이라는 결과는 정상적이지 않다. 이 연구에서 확보한 데이터는 공공도서관에서 실제로 대출된 도서의 ISBN을 사용하여 요청하였다. 일반적으로 공공도서관에서 서지 요소가 불완전한 도서는 대출되지 않는다. 청구기호가 없는 도서는 사서에 의해 배가될 수 없고, 서명이 없는 도서는 이용자에게 식별되고 검색될 수 없기 때문이다. 달리 말하자면, 공공도서관에서 도서의 대출 이력이 있다는 바는 완전성이 높은 서지 데이터가 있다는 의미이다. 실제로, ISBN이 '9791170262213'인 자료는 '도서 상세 조회' 데이터에서 ISBN을 제외한 모든 서지 요소가 공백이지만, 도서관 정보나루에서 확보한 대출 이력에서 해당 ISBN을 가진 5개 대출 레코드를 확인할 수 있으며, 이 레코드에는 서명과 저자명이 기재되어 있다. 즉, 도서관 정보나루의 데이터베이스 내에 서지 요소의 완전성이 더 뛰어난 레코드가 축적되어 있음에도 불구하고, 완전성이 낮은 레코드를 이용자에게 제공하고 있다. 이 문제의 원인은 데이터 수집 절차로 판단된다. 도서관 정보나루는 국립중앙도서관과 공공도서관으로부터 도서관 데이터를 수집하며, 국립중앙도서관은 '수집되는 모든 자료에 대해 완전수준의 서지 레코드'를 작성한다(국립중앙도서관, n.d.). 서지 데이터를 구축하는 과정에서, 국립중앙도서관의 서지 레코드를 먼저 수집하거나, 서지 레코드의 입력 수준이나 완전성을 비교하여 더 나은 레코드를 서지 테이블에 수용할 필요가 있다.

정확성 측면에서 문제는 도서관 정보나루의 데이터베이스가 관계형 데이터베이스 관리 시스템에 의해 운영됨에도 불구하고, '도서 상세 조회' API와 연결된 데이터베이스나 테이블이 자료형, 값의 길이, 값의 범위, 제약조건 등을 아우른 스키마가 부재하다는 점이다. 예를 들어, 13자리 ISBN은 숫자만으로 가지므로 자료형을 8바이트 숫자형으로 지정하면, 13자리의 숫자로 이루어질 수 있고 문자열이 포함될 수 없다. 또한, 13자리 ISBN의 앞 세 자리 숫자는

'978'이나 '979'로 시작하므로, 체크 제약조건으로 범위를 지정해 '978'과 '979' 이외의 코드가 포함되지 않게 할 수 있다. 완전성 측면에서도 무결성 제약조건 지정하면, 서명, 저자명, 출판사, 발행년과 같은 서지 요소가 공백인 레코드는 서지 테이블에 저장되지 않는다. 값의 진위를 판단하지 않더라고, 값이 갖는 형식을 데이터의 정확성을 확인할 수 있다. 행정안전부의 개방 표준 데이터 세트나 영국의 도서관 공공데이터 스키마와 예와 같이, 도서관 정보나루의 주요 데이터 세트에 대해 스키마 구축 작업을 진행하고 이를 내부적으로 데이터 품질 검사에 사용하고, 외부적으로는 데이터 이용자에게 데이터를 확인할 수 있게 제공해야 한다.

5. 결 론

공공도서관이 업무 과정에서 생성하고 축적한 데이터로서 도서관 공공데이터의 품질 진단을 하였다. 8개 공공도서관의 19,177,721개 대출 레코드에서 453,136개의 고유한 ISBN을 추출하여, 도서관 공공데이터 플랫폼인 도서관 정보나루의 '도서 상세 조회' API로부터 453,136개 레코드를 확보하였다. 확보한 데이터에서 오류 메시지가 포함되지 않은 450,461개 레코드의 완전성과 정확성을 분석하였다. 품질 진단 결과, 데이터의 완전성과 정확성에 문제가 있음이 드러났다. 서명, 저자명, 발행년, 발행처, 분류기호와 같이 자료의 식별과 검색에 핵심적인 서지 요소에서 공백이 포함되어 있다는 점이 확인되었다. 또한, ISBN, 발행년, 분류기호에서 값의 유형, 값의 범위, 제약조건에 따르지 않는 값이 있음이 드러났다. '도서 상세 조회'라는 명칭과 달리 확보한 데이터에는 DVD나 오디오북과 같은 다수 비도서 자료가 포함되어 있음이 확인되었다. 완전성과 정확성 진단 결과에 따라 다음과 같은 도서관 정보나루의 도서관 공공데이터 품질 개선방안을 제안한다.

첫째, 데이터 수집 절차를 검토해야 한다. 도서관 정보나루의 데이터는 국립중앙도서관, 공공도서관, 서점 등 외부 기관으로부터 수집된다. 세 유형의 기관 중 서지 데이터의 입력 수준이 가장 높은 기관은 입수하는 모든 자료의 서지 레코드를 완전 수준으로 작성하는 국립중앙도서관이다. 그러므로 국립중앙도서관의 서지 레코드를 먼저 수집하여 서지 데이터에 탑재해야 한다. 이후 레코드의 완전성과 정확성에 따라 국립중앙도서관의 서지 레코드를 보완하거나 대체하는 순서로 데이터 수집 절차를 정비하여, 이미 불완전한 레코드가 서지 데이터에 포함되어 있어 완전한 레코드가 수집되지 않는 일을 방지해야 한다.

둘째, 도서관 정보나루가 제공하는 주요 데이터와 API의 데이터 스키마를 작성해야 한다. 도서관 공공데이터의 핵심은 서지 레코드이며, 서지 레코드의 작성은 편목 규칙과 인코딩 방식에 따른다. 관련 규정에 근거하여 서지 요소에 대응하는 속성의 값의 유형, 값의 범위, 제약조건을 지정한 데이터별 스키마를 통해 내부적으로는 데이터 품질 진단을 시행하고 외부적으로는 민간 연구자와 개발자의 데이터 활용의 소요를 덜어 줘야 한다.

셋째, 도서관 정보나루의 데이터 수집 방식, 데이터 처리 방식에 관한 방식에 관한 안내를 제공해야 한다. 도서관 정보나루에 데이터의

품질에 관해 문의한 결과, 국립중앙도서관의 국가자료종합목록시스템의 서지 데이터를 중심으로 수집을 한다는 원론적인 답변을 받았다. 그러나 확보한 레코드의 15.50%에서 분류기호가 공백이라는 점은 도서관 정보나루의 데이터 수집과 처리 과정에 문제가 있다는 추정을 가능하게 한다. 연구자나 개발자가 데이터를 이해할 수 있도록 어떻게 데이터가 수집되어 제공되는지를 설명하는 상세한 자료가 필요하다.

넷째, 다양한 분석과 연구에 이용될 수 있도록 공공도서관의 원자료를 공개해야 한다. 가공된 자료는 원자료로부터 도출될 수 있고, 연구자와 개발자는 원자료를 통해 더 폭넓은 연구를 할 수 있다. 그러므로, '장서/대출 데이터'와 같이 공공도서관의 서지 데이터와 대출 이력을 결합하여 제공하는 데이터는 각각의 원자료를 함께 제공해야 한다.

이 연구는 모든 도서관 공공데이터를 대상으로 한 연구가 아니다. 도서관 정보나루가 소장한 122,452,521건의 장서 데이터 중 오직 '도서 상세 조회' API로 확보한 450,461개의 레코드만으로 데이터의 완전성과 정확성을 진단하였다는 한계를 갖는다. 또한, 데이터 품질 중 오직 데이터 완전성과 정확성만을 분석하였다. 도서관 공공데이터의 데이터 품질 진단을 위해서는 데이터의 규모와 데이터 품질 요소를 확장할 필요가 있으며, 이는 향후 연구의 과제로 남는다.

참 고 문 헌

공공데이터 - 탐색 - Google 트렌드 (2020. 11. 17). Google trends. Retrieved from https://trends.google.com/trends/explore?date=all&geo=KR&q=%EA%B3%B5%EA%B3%B5%EB%8D%B0%EC%9D%B4%ED%84%B0

공공데이터 개방 및 활용 (2020). e-나라지표. Retrieved from http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2844

공공데이터 개방 표준 개정 고시, 행정안전부고시 제2020-54호 (2020). Retrieved from https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000016&nttId=80780

공공데이터의 제공 및 이용 활성화에 관한 법률 시행령, 대통령령 제28211호 (2017). Retrieved from https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률시행령

공공데이터의 제공 및 이용 활성화에 관한 법률, 법률 제14839호 (2017). Retrieved from https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률

공공도서관 통계 보기 (2019). 국가도서관통계시스템. Retrieved from https://www.libsta.go.kr/libportal/libStats/publicLib/unitStats/getUnitStatsPop.do?gubun

＝STEP0000000001&libGubun＝LIBTYPE002
과학기술정보연구원 (2018). 도서관 빅데이터 활용사례집. Retrieved from https://www.data4library.kr/downloadCaseBook?c＝2018+도서관+빅데이터+활용사례집.pdf
국립중앙도서관 (2014). 한국문헌자동화목록 - 통합서지용. Retrieved from http://www.nl.go.kr/common/jsp/kormarc_2014/index.html
국립중앙도서관 (2019). 도서관 빅데이터 활용사례집. Retrieved from https://www.data4library.kr/downloadCaseBook?c＝2019+도서관+빅데이터+활용사례집.pdf
국립중앙도서관 (2020a). 도서관 정보나루: 데이터 수집현황. Retrieved from https://www.data4library.kr
국립중앙도서관 (2020b). 도서관 정보나루: 도서별 이용분석. Retrieved from https://data4library.kr/bookV?seq＝2785982
국립중앙도서관 (n.d.). 목록규칙적용세칙. Retrieved from https://www.nl.go.kr/nation/c3/page3_2.jsp
국립중앙도서관 자료정리규정, 국립중앙도서관규정 제582호 (2020). Retrieved from https://www.law.go.kr/LSW/admRulLsInfoP.do?chrClsCd＝&admRulSeq＝2100000189784
김우정, 이지원, 조용완 (2017). 대학도서관의 DVD 자료 목록레코드 품질에 관한 연구. 한국비블리아학회지, 28(4), 77-100. http://dx.doi.org/10.14699/kbiblia.2017.28.4.077
김유승 (2014). 기록으로서 공공데이터 관리를 위한 제도적 고찰:『공공데이터의 제공 및 이용 활성화에 관한 법률』분석을 중심으로. 한국기록관리학회지, 14(1), 53-73. http://doi.org/10.14404/JKSARM.2014.14.1.053
김태영, 백지영, 오효정 (2018). 빅데이터 로그 기반 도서관 이용자 및 대출 현황 분석: 국립세종도서관을 중심으로. 한국도서관·정보학회지, 49(2), 357-388.
김학래 (2020). 공공데이터 개방표준 데이터의 품질평가. 한국콘텐츠학회논문지, 20(9), 439-447. https://doi.org/10.5392/JKCA.2020.20.09.439
김현철 (2014). 공공데이터 품질 요인이 공공데이터 개방정책의 신뢰에 미치는 영향에 관한 연구. 박사학위논문, 숭실대학교 대학원, 경영학과.
김혜선, 김완종 (2016). 도서관 분야 데이터 개방 현황과 개선방안 연구. 제23회 한국정보관리학회 학술대회 논문집, 77-80.
대통령 소속 도서관정보정책위원회 (2014). 제2차 도서관발전종합계획(2014~2018). Retrieved from https://www.korea.kr/archive/expDocView.do?docId＝37665
박지영 (2016). 서지프레임워크를 활용한 공공도서관 서지데이터와 서비스 데이터의 연계. 한국정보관

리학회지, 33(1), 293-316.
박진호 (2018). 도서관의 오픈 데이터 품질측정모델 개발. 정보관리학회지, 35(1), 33-59.
온정미, 박성희 (2020). 도서관 빅데이터 플랫폼을 활용한 공공도서관 빅데이터 분석 연구: 대전한밭도서관을 중심으로. 정보관리학회지, 37(3), 25-50. http://dx.doi.org/10.3743/KOSIM.2020.37.3.025
이원재, 김휘강 (2020). 공공데이터 품질환경 내 데이터 오류의 발생원인별 보안기술 대응방안에 관한 연구. 정보보호학회지, 30(4), 77-89.
이정미 (2013). 빅데이터의 이해와 도서관 정보서비스에의 활용. 한국비블리아학회지, 24(4), 53-73. http://dx.doi.org/10.14699/kbiblia.2013.24.4.053
조재인 (2018). 공공데이터 포털을 통해 개방된 도서관 관련 데이터 분석. 한국비블리아학회지, 29(2), 35-56. http://dx.doi.org/10.14699/kbiblia.2018.29.2.035
표순희, 김윤형, 김혜선, 김완종 (2015). 도서관 빅데이터 서비스 모형 개발에 관한 연구: 공공도서관을 중심으로. 정보관리학회지, 32(2), 63-86. http://dx.doi.org/10.3743/KOSIM.2015.32.2.063
한국도서관협회 목록위원회 (2003). 한국목록규칙 제4판. 서울: 한국도서관협회.
한희정, 황성욱, 이정민, 오효정 (2020). 공공데이터포털 이용자 서비스 현황 분석 및 개선방안. 한국도서관·정보학회지, 51(1), 255-279. http://dx.doi.org/10.16981/kliss.51.1.202003.255
Ayre, L. B., & Craner, J. (2017). Open data: What it is and why you should care. Public Library Quarterly, 36(2), 173-184. https://doi.org/10.1080/01616846.2017.1313045
Back, C. (2020. 8. 3). Library open data: An update. GOV.UK blogs. Retrieved from https://dcmslibraries.blog.gov.uk/2020/08/03/library-open-data-an-update/
Carruthers, A. (2014). Open data day hackathon 2014 at edmonton public library. Partnership: The Canadian Journal of Library and Information Practice and Research, 9(2), 1-13. https://doi.org/10.21083/partnership.v9i2.3121
Chignard, S. (2013). A brief history of open data. paris innovation review. Retrieved from http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data
Department for Culture, Media and Sport (2014. 12. 18). Independent library report for England. Retrieved from https://www.gov.uk/government/publications/independent-library-report-for-england
GS1 Company Prefix (n.d.). GS1. Retrieved from https://www.gs1.org/standards/id-keys/company-prefix
GSI 표준 사용자 매뉴얼 (2015). 대한상공회의소 유통물류진흥원. Retrieved from http://www.gs1kr.org/File/New/Data01/GS1%20표준사용자%20매뉴얼_20151218.pdf
International Standard Book Number (2020. 9. 3). In Wikipedia. Retrieved from

https://en.wikipedia.org/wiki/International_Standard_Book_Number

Libraries and Open Data (2020. 3. 6). IFLA. Retrieved from https://blogs.ifla.org/faife/2020/03/06/libraries-and-open-data/

Libraries Taskforce (2016. 12. 1). Libraries deliver: Ambition for public libraries in England 2016-2021. Retrieved from https://www.gov.uk/government/publications/libraries-deliver-ambition-for-public-libraries-in-england-2016-to-2021/libraries-deliver-ambition-for-public-libraries-in-england-2016-to-2021

Libraries Taskforce (2017. 7. 20). List of contents for the libraries core dataset for England. Retrieved from https://www.gov.uk/government/publications/list-of-contents-for-the-core-dataset-for-libraries/list-of-contents-for-the-libraries-core-dataset-for-england

Libraries Taskforce (2020). GOV.UK. Retrieved from https://www.gov.uk/government/groups/libraries-taskforce

Library open data (n.d.). Retrieved from https://schema.librarydata.uk/

Ostler, K. R., Norlander, B., & Weber, N. (2020). Using open data to inform public library branch services. Public Library Quarterly, 1-13. Preprint. https://doi.org/10.1080/01616846.2020.1798206

Robinson, P., & Mather, L. W. (2017). Open data community maturity: Libraries as civic infomediaries. Journal of the Urban & Regional Information Systems Association, 28(1), 31-38

Schrock, A. R. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. New Media & Society, 18(4), 581-599. https://doi.org/10.1177/1461444816629469

WHAT IS OPEN DATA DAY? (2020). Open data day. Retrieved from https://opendataday.org/

What is Open Data? (n.d.). Open data handbook. Retrieved from http://opendatahandbook.org/guide/en/what-is-open-data/

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

Act on Promotion of the Provision and Use of Public Data, Act No. 14839 (2017). Retrieved from https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률

Cho, Jane (2018). A study about library-related open data through public data portals. Journal of the Korean BIBLIA Society for library and Information Science, 29(2), 35-56. http://dx.doi.org/10.14699/kbiblia.2018.29.2.035

Enforcement Decree of The Act on Promotion of the Provision and Use of Public Data, Presidential Decree No. 28211 (2017) Retrieved from https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률시행령

GSI Standard User Manual (2015). Korea chamber of commerce and industry, institute of distribution & logistics. Retrieved from http://www.gs1kr.org/File/New/Data01/GS1%20표준사용자%20매뉴얼_20151218.pdf

Han, Hui-Jeong., Hwang, Sung-Wook., Lee, Jung-Min., & Oh, Hyo Jung (2020). Analysis of current status and improvement plans of the user service in open data portal - Focusing on citizen participation data portal -. Journal of Korean Library and Information Science Society, 51(1), 255-279. http://dx.doi.org/10.16981/kliss.51.1.202003.255

Jung, Bo Ra (2013. 2. 24). "Open data, let's use it like this." BLOTER. Retrieved from http://www.bloter.net/archives/144788

Kim, Haklae (2020). Quality evaluation of the open standard data. The Journal of the Korea Contents Association, 20(9), 439-447. https://doi.org/10.5392/JKCA.2020.20.09.439

Kim, Hye-Sun, & Kim, Wan-Jong (2016). A study on library data open status and improvement strategies. Proceedings of the Korean Society for Information Management 23th Conference, 77-80.

Kim, Hyun Chul (2015). A study on public data quality factors affecting the confidence of the public data open policy (Doctoral dissertation, Sungsing University, Seoul, Republic of Korea).

Kim, Tae-Young., Baek, Ji-Yeon., & Oh, Hyo Jung (2018). An analysis of library user and circulation status based on bigdata logs: A case study of national library of Korea, Sejong. Journal of Korean Library and Information Science Society, 49(2), 357-388.

Kim, Woo-Jeong., Lee, Ji-Won., & Cho, Yong-Wan (2017). A study on quality of bibliographic records for DVDs in university libraries. Journal of Korean BIBLIA Society for Library and Information Science, 28(4), 77-100. http://dx.doi.org/10.14699/kbiblia.2017.28.4.077

Kim, You-Seung (2014). A study on legal issues of public data management as records: Focused on analysis of the act on provision and use of public data. Journal of Korean Society of Archives and Records Management, 14(1), 53-73. http://doi.org/10.14404/JKSARM.2014.14.1.053

Korea Institute of Science and Technology Information (2018). Casebook on Using Big Data in Library. Retrieved from https://www.data4library.kr/downloadCaseBook?c=2018+도서관+빅데이터+활용사례집.pdf

Korean Library Association Catalogue Committee (2003). Korean cataloguing rules 4th edition. Seoul: Korean Library Association.

Lee, Jeoung-Mee (2013). Understanding big data and utilizing its analysis into library and information services. Journal of the Korean BIBLIA Society for Library and Information Science, 24(4), 53-73. http://dx.doi.org/10.14699/kbiblia.2013.24.4.053

Lee, Su-Sang (2014). library and big data. KLA Journal, 55(8), 14-25.

Lee, Won-Jae, & Kim Huy-Kang (2020). A study on security technology countermeasures by cause of data error in open government data quality environment. Review of KIISC, 30(4), 77-89.

Library Information Policy Committee under the jurisdiction of the President (2014). Second library comprehensive development plan. https://www.korea.kr/archive/expDocView.do?docId=37665

National Library of Korea (2014). Korean machine readable cataloging format - Integrated format for bibliographic data. Retrieved from http://www.nl.go.kr/common/jsp/kormarc_2014/index.html

National Library of Korea (2019). Casebook on Using Big Data in Library. Retrieved from https://www.data4library.kr/downloadCaseBook?c=2019+도서관+빅데이터+활용사례집.pdf

National Library of Korea (2020a). Data for library: Current Status of Collecting Library Data. Retrieved from https://www.data4library.kr

National Library of Korea (2020b). Data for library: Analysis on Book Usage. Retrieved from https://data4library.kr/bookV?seq=2785982

National Library of Korea (n.d.). Cataloguing Rules Application Details. Retrieved from https://www.nl.go.kr/nation/c3/page3_2.jsp

Notification for Revision in Open Government Data Standard, Notification No. 2020-54 of Ministry of Interior and Safety (2020). Retrieved from https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000016&nttId=80780

On, Jeong-Mee, & Park, Sung-Hee (2020). Big data analysis for public libraries utilizing big data platform: A case study of Daejeon hanbat library, Journal of the Korean Society for

information Management, 37(3), 25-50. http://dx.doi.org/10.3743/KOSIM.2020.37.3.025

Open Government Data - Search - Google Trends. Google Trends. Retrieved from https://trends.google.com/trends/explore?date=all&geo=KR&q=%EA%B3%B5%EA%B3%B5%EB%8D%B0%EC%9D%B4%ED%84%B0

Open Government Status (2020). e-National index Retrieved from http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2844

Park, Jin Ho (2018). Developing an assessment model of library open data quality. Journal of Korean Society for Information Society, 35(1), 33-59.

Park, Zi-Young (2016). Linking bibliographic data and public library service data using bibliographic framework. Journal of the Korean Society for information Management, 33(1), 293-316.

Public Library Statistics (2019). National library statistics system. Retrieved from https://www.libsta.go.kr/libportal/libStats/publicLib/unitStats/getUnitStatsPop.do?gubun=STEP0000000001&libGubun=LIBTYPE002

Pyo, Soon-Hee., Kim, Yun-Hyung., Kim, Hye-Sun., & Kim, Wan-Jong (2015). A study on the developing of big data services in public library. Journal of the Korean Society for information Management, 32(2), 63-86. http://dx.doi.org/10.3743/KOSIM.2015.32.2.063