

지방자치단체의 공공데이터 개방 현황 분석과 개선방안

송 채 은¹ · 김 학 래^{2*}¹중앙대학교 문헌정보학과 석사과정^{2*}중앙대학교 문헌정보학과 교수

Improvements of public data policy through data portal analysis of local governments

Chaeun Song¹ · Haklae Kim^{2*}¹Master's Course, Department of Library and Information Science, Chung-Ang University, Seoul 06974, Korea^{2*}Professor, Department of Library and Information Science, Chung-Ang University, Seoul 06974, Korea

[요 약]

한국 정부는 2013년부터 공공데이터를 적극적으로 개방하고 있고, 세계적으로 리더 국가로 평가받고 있다. 공공데이터는 공공기관이 보유하고 있는 데이터를 시민에게 개방한 것으로, 다양한 공공기관이 대규모 데이터 세트를 제공하고 있다. 특히 지방자치단체는 공공데이터 포털의 약 50% 데이터 세트를 제공하고 있다. 그러나, 지방자치단체는 공공데이터의 개방과 활용을 위한 체계적인 정책을 수립하고 집행하는 데 여러 가지 제한이 있고, 지방자치단체에 따라 공공데이터의 개방, 데이터 관리 수준이 매우 다른 것이 현실이다. 본 논문은 지방자치단체의 효율적인 공공데이터 관리와 정책 추진을 위한 현황을 분석한다. 지방자치단체별 공공데이터 예산, 개방 규모, 전문 인력, 데이터 관리현황 등 다양한 특성을 정량적으로 평가하고, 지방자치단체의 행정·운영적 특징을 종합적으로 분석한다. 이를 통해 지방자치단체의 데이터 개방정책과 관리체계에 대한 개선방안을 제안한다.

[Abstract]

The Korean government has been actively opening public data since 2013, and it is evaluated as a world leader. Public data refers to data held by public institutions open to citizens, and various public institutions are included. In particular, local governments, which have occupied about 50% provided by the public data portal, play a very important role in providing public data. However, local governments have various limitations in establishing and executing systematic policies for the disclosure and use of public data, and the reality is that the level of disclosure and data management of public data is very different depending on the local government. This paper analyzes the current status of local governments for efficient public data management and policy implementation, and analyzes the characteristics of local governments from a management and operational perspective. Through this, we suggest ways to improve the public data policy and management system of local governments.

색인어 : 공공데이터, 데이터 포털, 데이터 거버넌스, 지방자치단체, 데이터 분석**Keyword** : Public data, Data portal, Data governance, Local government, Data analysis<http://dx.doi.org/10.9728/dcs.2022.23.4.697>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 25 February 2022; **Revised** 16 March 2022**Accepted** 07 March 2022***Corresponding Author, Haklae Kim****Tel:** +82-2-820-5561**E-mail:** haklaekim@cau.ac.kr

I. 서론

「공공데이터의 제공 및 이용 활성화에 관한 법률」(이하 공공데이터법)에 따르면 공공데이터는 공공기관이 보유, 관리하는 데이터를 제공하여 국민이 자유롭게 이용할 수 있게 하는 것을 목적으로 한다. 한국은 2013년에 공공데이터법을 제정하고, 다양한 분야의 공공데이터를 적극적으로 제공하고 있다. 특히, 2년마다 측정하는 OECD 공공데이터 지수(Open-Useful-Reusable; OURdata Index)에서 3회 연속으로 공공데이터 개방지수 1위를 차지하였다.

공공기관은 공공데이터법 제2조의 국가기관, 지방자치단체와 지능정보화 기본법 제2조 제16호에 따른 공공기관을 의미한다. 공공데이터 포털의 제공기관은 ‘국가행정기관’, ‘자치행정기관’, ‘교육행정기관’, ‘입법기관’, ‘헌법기관’, ‘공공기관’, ‘위원회’, ‘교육기관’ 8개 유형으로 구분하고 있다. 광역 지방자치단체와 기초지방자치단체를 포함한 ‘자치행정기관’은 공공데이터의 약 50%를 차지한다(2021년 02월 04일 기준 총 39,282건 중 19,544건).

한국의 지방자치단체는 공공데이터법 제7조에 따라 3년마다 ‘공공데이터의 제공 및 이용 활성화에 관한 기본계획(이하 기본계획)’을 수립한다. 기본계획은 공공데이터 제공을 위한 기본목표와 추진 방향, 공공데이터의 등록 및 이용 현황 등의 사항을 포함한다. 그리고 기본계획에 따라 매년 ‘공공데이터 제공 및 이용 활성화에 관한 시행계획(이하 시행계획)’을 수립한다. 시행계획은 직전 연도의 공공데이터 개방 수량과 해당 연도의 공공데이터 관련 예산 운용계획을 포함한다. 개방하는 데이터는 행정안전부가 주관하는 공공데이터 포털(<http://data.go.kr>)에 공개하는 것이 원칙이다.

공공데이터 정책에 있어 지방자치단체의 역할은 매우 중요하지만, 데이터를 통합적으로 관리하기 위한 연구는 미흡한 실정이다. 예를 들어, 서울특별시 데이터의 일부는 공공데이터 포털에서 제공하고, 다른 일부는 서울열린데이터광장에 제공하여 데이터가 일관성 있게 제공되지 않는다. 세종특별자치시는 자체 데이터 포털의 운영이 원활하지 않아 공공데이터 포털과 통합 운영을 검토하고 있다. 한편, 각 지방자치단체가 운영하는 데이터 포털의 분류체계, 개별 데이터 세트의 기술적 메타데이터(descriptive metadata) 항목도 일치하지 않는다.

국내의 공공데이터 연구는 대부분 공공데이터법과 정책, 공공데이터 포털 중심의 개방 현황, 데이터 품질 연구에 집중되어 있다. 지방자치단체 관점의 공공데이터 연구는 일부 지방자치단체를 대상으로 한 연구이거나, 데이터 관리보다는 이용 활성화 정책을 위한 연구다. 지방자치단체가 공공데이터에 투자하는 예산, 전문인력 등과 같은 행정적 요소에 따른 차이점 또는 지방자치단체가 운영하는 데이터 포털 연구는 미흡하다. 향후 지방자치단체의 데이터 관리와 활성화를 위해 행정 현황과 데이터 관리에 대한 정량적 연구방안을 함께 고려해야 한다 [1]-[3].

본 연구는 17개 지방자치단체의 공공데이터 운영현황을

조사하고, 개선점을 제안한다. 2장은 국내의 공공데이터 연구 동향과 지방자치단체의 공공데이터 현황을 소개한다. 3장은 지방자치단체의 공공데이터 운영현황을 점검하기 위한 연구 대상과 범위를 정의한다. 4장은 연구 결과를 요약하고, 5장은 연구의 시사점과 결론을 정리하고 향후 연구로 마무리한다.

II. 관련 연구

세계 각국 정부는 공공데이터 정책을 적극적으로 추진하고 있다. 미국의 오바마 정부는 열린 정부 정책을 통해 투명성을 강조하고, 대규모 데이터를 데이터 포털(<http://data.gov>)에 공개하고 있다. 영국은 2010년에 투명성과 책임성, 시민 주도형, 정보의 경제적·사회적 가치, 시맨틱웹 총 4가지 의제를 제안하고, 국가 데이터 포털(<http://data.gov.uk>)을 통해 공공데이터를 제공하고 있다. 영국 정부는 2020년에 데이터 활용 증진을 위해 국가 데이터 전략(National Data Strategy)을 수립하였다[4].

그러나 데이터의 개방이 활용을 직접 이끌어내는 것은 아니다. Marijn Janssen et al.(2012)은 공공데이터 자체의 본질적인 가치보다 활용할 수 있게 지원하는 것이 중요함을 주장한다[5]. 데이터 사용은 목적에 맞는 데이터의 정제가 필요하고, 데이터의 탐색과 접근을 위해 충분한 메타데이터를 제공해야 한다. Vijay Khatri, Carol V. Brown(2010)은 데이터 거버넌스가 데이터에 대한 권한과 조직의 의사 결정에 대한 책임을 의미한다고 말한다. 데이터 거버넌스를 위한 프레임워크는 데이터 원칙, 데이터 품질, 메타데이터와 데이터 접근, 데이터 생애 주기가 포함된다[6]. Rene Abraham et al.(2019)은 데이터 거버넌스 관련 문헌 145개를 기반으로 개념적 프레임워크를 개발하였다. 데이터 거버넌스를 위한 의사 결정 과정에서 데이터 품질, 데이터 보안, 데이터 구조, 데이터 생애 주기, 메타데이터, 데이터 저장 시설을 중점 요소로 분류하였다[7].

공공데이터가 개방되면서 품질관리에 대한 문제는 꾸준히 제기되었다[8]-[9]. Anastasija Nikiforova, Keegan McBride(2021)는 데이터 포털이 데이터 관리자와 데이터 사용자 간의 주요 상호작용 지점으로 작용하는 것에 주목했다. EU의 41개 데이터 포털을 대상으로 사용성 평가 연구를 진행하고, 포털 사용성 평가를 위한 평가 프레임워크를 설계했다. 이를 통해 포털 간 유용성을 검증하고 사용성을 향상하는 방법을 제시하고 있다[10].

공공데이터 정책이 효과적으로 추진되기 위해 지방 정부의 참여와 역할이 매우 중요하다. Peter Conradie, Sunil Choenni(2014)은 공공데이터가 주로 지역 수준에서 수집되는 것을 주목하여, 인터뷰와 설문 조사를 통해 로테르담의 데이터개방 과정을 연구하였으며[1], Akemi Takeoka Chatfield, Christopher G. Reddick(2017)은 데이터 세트 수, 기계가 읽을 수 있는 데이터 형식, 데이터 정책, 데이터 포

털 서비스 총 4가지 지표를 설정한 후, 호주 내 20개의 지방 정부 데이터 포털을 분석하였다[11]. Faizura Haneem et al.(2019)는 조직에서 가장 우선적으로 관리해야 하는 데이터인 마스터 데이터 관리(Master Data Management, MDM)를 지방 정부 관점으로 연구한다. 데이터 품질과 데이터 거버넌스가 지방 정부에서 MDM 채택을 결정하는 두 가지 요소임을 강조한다[12]. 데이터 포털은 선별된 개방데이터를 관리하고 이용자에게 제공한다. Di Wang et al.(2018)은 지방 정부가 데이터 포털 개발 시 우선순위 지침을 마련하기 위해 데이터 규모 (Data Quantity), 데이터 접근성 (Data Accessibility), 데이터 품질 (Data Quality) 지표를 이용한 평가 프레임워크를 설계했다. 중국 내 데이터 포털에 평가 프레임워크를 적용한 결과, 데이터 규모와 비교하여 데이터 접근성과 품질이 더 중요하다는 것을 강조한다[13].

윤상오, 현지우(2018)는 공공데이터 개방정책을 개선하기 위해 공공데이터 포털의 국가중점데이터를 중심으로 연구를 진행했다. 가용성(Availability), 이용용이성(Accessibility), 활용도(Utilization)를 분석한 결과, 국민 생활 밀착형 데이터를 적극적으로 개방하고 데이터 이용방식을 수요자 중심으로 전환하기를 강조한다[14]. 한희정 외(2020)는 공공데이터 포털 이용자 서비스 현황을 분석하여 시민참여 활성화를 위한 개선방안을 제안한다. 데이터 포털 간 공유서비스, 시각화·분석 서비스, 의사소통을 위한 다양한 커뮤니티 조성을 개선 방안으로 주장한다[15]. 김학래(2019)는 공공데이터 간 표준화 연구를 위해 공공데이터의 행정표준어휘 적용을 평가하였다. 정부에서 권장하는 표준어휘사용 활성화를 위해서 공공데이터 규정과 지침을 개선하고, 표준어휘를 검증할 수 있는 자동화 도구를 마련하는 것을 제안한다[16].

한편, 지방 정부 수준에서 공공데이터를 개발하고 데이터 포털을 통해 공개하는 것은 중요하지만, 기술적 한계, 데이터 정책 부족, 관리자의 미흡한 이해 등의 한계점이 있다[17]-[18]. 김신곤(2013)은 전자정부 초기, 지방자치단체의 빅데이터 도입을 위해서 국내외 공공데이터 활용사례를 선정하여 지역 정보화 관점의 준비 사항을 제안한다[19]. 송효진, 황성수(2014)와 서형준, 명승환(2015)은 지방자치단체의 데이터 포털을 중심으로 개방 현황을 조사한 결과, 지역적 상황을 고려한 법·제도적 기반과 공공데이터 담당자의 인력과 예산 지원을 강화하여 지방자치단체의 주체적인 참여의식이 필요함을 주장한다[20]-[21]. 김동준 외(2020)는 공공데이터 포털 내 지방자치단체 데이터의 활용도를 평가하고 개선방안을 제안한다[22].

본 연구는 지방자치단체의 공공데이터 운영을 행정적인 부분부터 데이터 포털 관리현황까지 정량적으로 파악하여 공공데이터 정책 개선을 위한 연구를 목적으로 한다.

III. 연구 범위와 대상

표 1. 공공데이터 정책 개선을 위한 평가 지표

Table 1. Evaluation indicators for public data policy improvement

Dimension	Indicators
Administrative operation	Budget, Data quantity, Staff in charge
Status of Open data	Data matching rate, Metadata, Category

지방자치단체의 공공데이터 현황을 알아보기 위해 지방자치단체별 행정운영과 개방데이터의 관리현황을 조사하고 연구대상을 선정한다. 표 1은 데이터 거버넌스 요소인 데이터 관리 조직, 메타데이터 관리를 중심으로 구성된 연구방법 지표를 정리한 것이다[6]-[7]. 개방데이터 규모, 예산, 인력의 정량적 평가는 데이터를 관리하는 조직의 의사 결정에 도움을 줄 수 있고, 알맞은 공공데이터 정책 연구의 근거가 된다[20]. 개방된 데이터의 데이터 포털 현황과 메타데이터, 분류체계 분석은 데이터 품질관리의 개선요소가 될 수 있다.

지방자치단체의 ‘공공데이터의 제공 및 이용 활성화에 관한 기본계획’, ‘공공데이터 제공 및 이용 활성화에 관한 시행계획’을 기준으로 연구대상과 방법을 선정한다.

- 데이터와 예산: 공공데이터 시행계획은 중앙행정기관과 지방자치단체의 공공데이터 이용 활성화 추진 실적과 계획이 포함되어 있다. 기본계획과 시행계획을 통해 5년 동안의 공공데이터 운영현황과 관련 예산을 조사한다. 다만, 2019년 공공데이터의 제공 및 이용 활성화 시행계획은 해당연도의 개방데이터 수량이 명시되어 있지 않기 때문에 ‘범정부 공공데이터 중장기(‘19~’21) 개방계획’을 참조한다. 공공데이터의 개방 규모와 예산의 관계를 분석하기 위해 상관관계 분석을 수행한다.
- 전문인력: 공공기관은 공공데이터의 전문적인 관리를 위해 인력을 확보해야 한다. 공공기관은 공공데이터법 제12조에 따라 공공데이터에 관한 업무를 총괄하는 책임관과 실무담당자를 임명하고 해당 기관의 홈페이지를 통해 공표하여야 한다. 지방자치단체가 공표한 공공데이터제공 책임관과 실무담당자를 전문인력으로 집계한다. 개별 공공기관은 행정안전부가 고시한 ‘공공데이터 관리지침’에 따라 공공데이터를 공공데이터 포털에 등록하는 것이 원칙이다. 다만 공공데이터 관리지침은 별도의 데이터 포털을 운영하는 제공기관에 대한 예외사항이 규정하고 있다(“기관이 개별로 운영하는 데이터 포털(이하 개별 포털)이 있는 경우 해당 개별 포털에서 제공하는 데이터는 공공데이터 포털과 일치되게 등록”). 지방자치단체가 운영하는 데이터 포털과 공공데이터 포털에서 제공하는 지방자치단체의 데이터 세트를 수집해서 일치율을 분석한다.
- 데이터 포털 현황: 공공데이터 포털에서 제공하는 데이터 세트는 제공기관 유형이 ‘자치행정기관’으로 설정하고, 파이썬(Python)으로 파일 데이터 목록을 수집한다. 공공데

이더 포털과 지방자치단체의 데이터 포털에서 제공하는 데이터 세트의 데이터명을 기준으로 비교하고 특수문자, 띄어쓰기, 숫자 등 불필요한 문자열은 제거한다.

- 메타데이터와 분류체계: 데이터 포털은 데이터를 관리하기 위해 메타데이터를 제공한다. 개별 데이터 세트는 여러 개의 기술적 메타데이터를 갖고 있고, 주제를 구분하기 위한 분류체계를 적용하고 있다. 그러나 개별 데이터 포털이 적용하고 있는 메타데이터와 분류체계가 다르므로, 전체 항목을 수집해서 분석한다.

IV. 연구 결과

4-1 예산과 인력

그림 1과 그림 2는 2016년부터 2020년까지 ‘공공데이터 제공 및 이용 활성화에 관한 시행계획’에 공개된 지방자치단체별 예산과 데이터 수량을 직접 집계하여 그래프로 나타낸 것이다.

그림 1에서 2019년까지 가장 많은 예산을 투자한 지방자치단체는 경기도와 서울특별시다. 2020년에 광주광역시와 약 1,000억 원의 예산을 데이터 산업에 투자하면서 평균 예산이 예년보다 5배가량 증가하였다. 광주광역시는 2020년부터 2024년까지 ‘인공지능 중심 산업융합 직접 단지 조성’ 사업에 약 4,061억 원을 투자하여 인공지능 융합 생태계를 조성하는 것을 추진하고 있다. 반면 세종특별자치시는 평균 약 6,700만 원으로 가장 적은 예산을 투자하고 있다.

공공데이터 예산은 세부적으로 ‘개방 확대’, ‘활용 활성화’, ‘품질 개선’, ‘기반 강화’로 구분하여 계획한다. 2020년부터는 세부 요소가 변경되어 ‘데이터 연계·분석’, ‘개방 및 융합’, ‘신산업 육성 및 생태계 조성’, ‘추진기반’으로 구분한다.

세부분야별 기준이 같은 2016년부터 2019년까지의 예산안을 분석하였다. 연도별 추이를 조사한 결과, 2017년까지는 ‘개방 확대’ 분야에 예산을 투자했지만, 2018년 이후로는 ‘활용 활성화’ 분야에 가장 많은 예산을 운용한다.

지방자치단체별 예산 운용 현황은 다음과 같다. 경기도는 ‘개방 확대’, ‘활용 활성화’, ‘기반 강화’ 분야에서 다른 지방자치단체보다 많은 예산을 투자하고, 서울특별시는 ‘품질개선’ 분야에 가장 많은 예산을 투자한다.

경상남도, 경상북도, 광주광역시, 제주특별자치도는 ‘활용 활성화’에 많은 예산을 투자하고, 인천광역시, 충청북도는 예산 대부분을 ‘개방 확대’에 투자한다. 지방자치단체의 예산 운용은 추진 계획에 따라 차이를 보인다.

그림 2는 공공데이터 시행계획의 지방자치단체별 개방예정 수량을 기준으로 누적 데이터 수량을 나타낸 그래프이다. 경기도는 공공데이터 정책 초기부터 적극적으로 개방한 지방자치단체로 2020년까지 가장 많은 수량의 데이터를 개방했다. 서울특별시는 경기도 다음으로 많은 데이터를 개방했고,

세종특별자치시는 2016년부터 2020년까지 매년 누적 개방 데이터 수량이 가장 적은 지방자치단체다. 광주광역시 2020년 예산이 급격히 증가했지만 개방데이터 수량 추이에 큰 변화가 없다. 대구광역시는 2016년부터 2019년까지 ‘개방 확대’ 분야에 예산을 투자한 적이 없지만, 개방데이터 수량은 매년 꾸준히 상승세를 보인다. 경기도는 예산 투자와 데이터 개방이 비교적 비례하지만, 대구광역시는 예산 편성이 없어도 데이터를 개방한다. 즉, 지방자치단체마다 ‘개방 확대’ 예산과 개방데이터 수량 간의 관계가 명확하지 않다. 16개 지방자치단체를 대상으로 2016년부터 2019년까지의 ‘개방 확대’ 예산과 데이터개방의 상관관계를 파악하기 위해 지방자치단체별 개방 확대 예산 평균값과 개방데이터 수량 간 상관분석을 진행했다. 상관계수는 약 0.71, 유의도 p값은 약 0.0014로 통계적으로 유의미한 양의 상관관계를 보인다.

지방자치단체에서 투자한 개방 확대 예산이 많을수록 데이터개방에 영향을 미친다는 것을 알 수 있다.

2020년 기준 공공데이터 전문인력은 부산광역시는 6명, 서울특별시와 인천광역시, 전라북도, 충청남도는 각 4명, 경상남도과 대구광역시, 대전광역시, 충청북도, 경상남도는 각 3명, 강원도와 경기도, 경상북도, 광주광역시, 세종특별자치시, 울산광역시, 전라남도, 제주특별자치도는 각 2명이다. 2018년 시행계획의 예산 상세내용에 따르면, ‘기반 강화’에 전문인력과 관계된 내용이 포함된다. 2020년 전체 지방자치단체 기준으로 전문인력 수와 평균 ‘기반 강화’ 예산의 상관분석을 진행한 결과, -0.08의 값이 도출됐다. 즉, 상관분석에 의한 예산과 인력의 상관관계는 거의 없다.

4-2 데이터 포털 현황

17개 지방자치단체가 운영하는 데이터 포털 현황은 표 2와 같다. 현재 운영되고 있는 데이터 포털은 독립형(Self-build), 호스트형(Link to the government data portal), 기능형(Embedded government data portal) 3가지 유형으로 구분할 수 있다.

독립형은 ‘서울열린데이터광장’과 같이 지방자치단체가 자체적으로 데이터 포털을 구축하고 운영하는 유형이다. 호스트형은 ‘광주 빅데이터 플랫폼’과 같이 데이터를 공공데이터 포털에 제공하고 바로가기로 연계해 서비스하는 유형이다.

기능형은 ‘대전광역시 공공데이터개방’와 같이 데이터 포털에 공공데이터 포털 기능을 삽입해서 공공데이터 포털의 검색과 이용 서비스를 제공하는 유형이다. 기능형은 이용자가 지방자치단체의 데이터 포털을 벗어나지 않고 이용할 수 있고, 호스트형은 바로가기로 연계하기 때문에 기존 플랫폼을 떠나 공공데이터 포털에서 다시 탐색을 시작한다는 차이점이 있다. 표 2에서 서울특별시(Seoul)와 부산광역시(Busan)는 현재 제공하고 있는 공공데이터가 가장 많다.

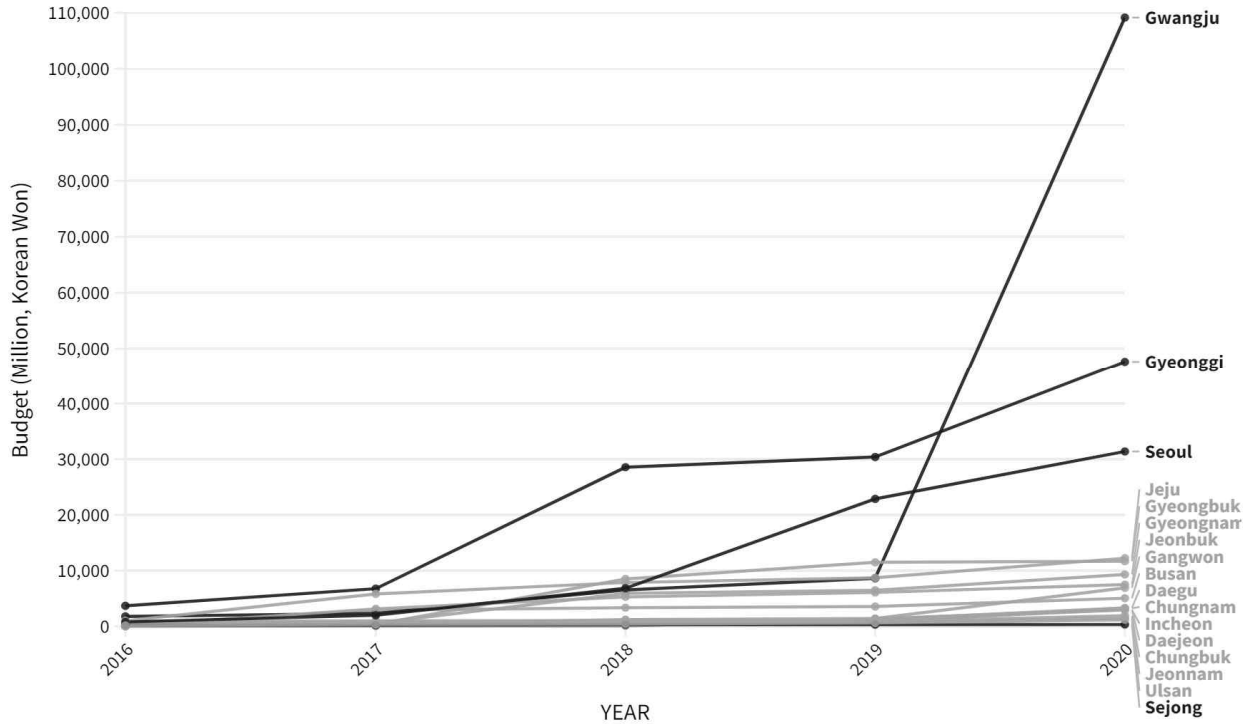


그림 1. 지방자치단체별 누적 공공데이터 예산 추이
 Fig. 1. The cumulative trend of public data budget by local government

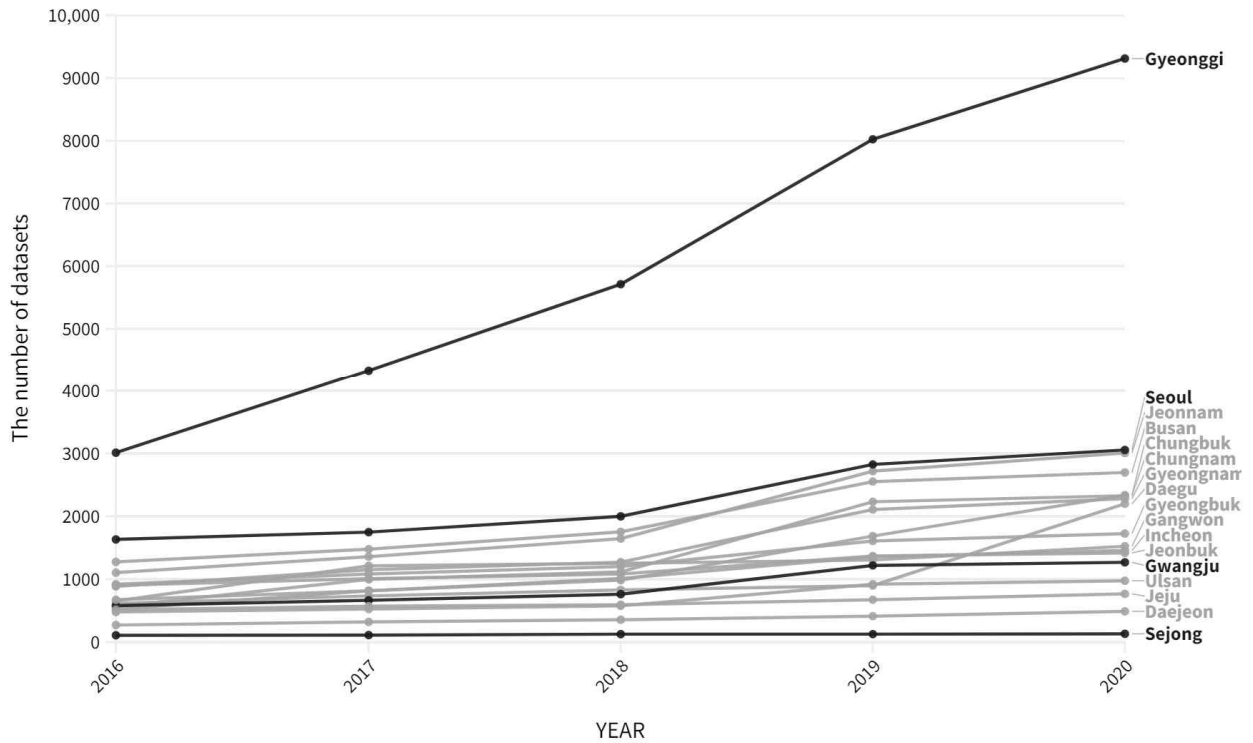


그림 2. 지방자치단체별 누적 공공데이터 개방 추이
 Fig. 2. The cumulative trend of opening public data by local government

표 2. 지방자치단체별 공공데이터 현황

Table 2. Status of open data by local government

local government	Central government (The number of datasets)	Local government (The number of datasets)	Data portal	Type of portal	URL
Busan	1,744	2,117	O	Self-build	https://data.busan.go.kr/index.nm
Chungbuk	728	0	X	Link to the government data portal	
Chungnam	813	0	X	Link to the government data portal	
Daegu	599	1,673	O	Self-build	http://data.daegu.go.kr/open/main.do
Daejeon	585	0	O	Embedded government data portal	https://www.daejeon.go.kr/pubc/publicDataPortalList.do?menuSeq=5147
Gangwon	485	501	O	Self-build	https://data.gwd.go.kr/index
Gwangju	806	0	O	Link to the government data portal	https://bigdata.gwangju.go.kr/index.gwangju
Gyeongbuk	939	251	O	Self-build (Out of operation)	http://data.gb.go.kr/
Gyeonggi	2,162	1,438	O	Self-build	https://data.gg.go.kr/portal/mainPage.do
Gyeongnam	1,405	1,640	O	Self-build	https://bigdata.gyeongnam.go.kr/index.gn
Incheon	1,546	1,548	O	Self-build	https://www.incheon.go.kr/data/index
Jeju	1,081	690	O	Self-build	https://www.jejudatahub.net/data/list
Jeonbuk	1,206	1,824	O	Self-build	http://www.bigdatahub.go.kr/index.jeonbuk
Jeonnam	1,591	0	X	Link to the government data portal	
Sejong	77	44	O	Self-build (Out of operation)	https://www.sejong.go.kr/stat/stats/public/list.do?key=2004075258108#
Seoul	3,379	6,571	O	Self-build	https://data.seoul.go.kr/
Ulsan	398	680	O	Self-build	http://data.ulsan.go.kr/index.ulsan

반면, 광주광역시(Gwangju), 대전광역시(Daejeon), 전라남도(Jeonnam), 충청남도(Chungnam), 충청북도(Chungbuk)는 0건이다. 즉, 독립형 데이터 포털을 운영하는 지방자치단체는 자체적으로 관리하는 데이터 세트가 있지만, 호스트형과 기능형은 자체적으로 관리하는 데이터 세트가 없고, 동시에 개방한 데이터 세트가 많지 않다는 특징을 갖는다.

4-3 데이터 일치율

‘공공데이터 관리지침’에 의하면, 공공기관이 개별적으로 데이터 포털을 운영해도 데이터 목록은 공공데이터 포털과 일치해야 한다. 서로 다른 데이터 포털에서 제공하더라도 데이터 수량과 내용이 일관성 있어야 한다. 반면, 표 2에서 공공데이터 포털 내의 ‘서울특별시’ 데이터와 ‘서울열린데이터광장’의 데이터는 각각 3,379건, 6,571건으로 각 포털에 개방하는 서울특별시 데이터 수량이 다르다. 서울특별시(Seoul)처럼 공공데이터 포털보다 자체 포털에서 데이터를 더 많이 제공하는 곳은 강원도(Gangwon), 경기도(Gyeonggi), 경상남도(Gyeongnam), 대구광역시(Daegu), 부산광역시(Busan), 울산광역시(Ulsan), 전라북도(Jeonbuk)다. 대구광역시와 서울특별시는 통계청 데이터, 자치구 데이터를 모두 개방하여 공공데이터 포털보다 많은 데이터를 보유하고 있다. 경상북도(Gyeongbuk), 세종특별자치시(Sejong), 인천광역

시(Incheon), 제주특별자치도(Jeju)는 지방자치단체 데이터 포털의 수량이 더 적다.

경상북도와 세종특별자치시는 공공데이터 포털과 통합 예정으로 데이터 포털 운영을 중지하여 새로운 데이터를 개방하지 않고 있다. 제주특별자치도에서 운영하는 ‘제주데이터 허브’는 2021년 1월 18일에 개편하여 운영자의 허가를 받으면 누구나 데이터 등록을 할 수 있다. 공공데이터가 아닌 외부 데이터가 포함되기 때문에, 공공데이터 포털에 개방한 데이터와 수량과 종류가 일치하지 않는다. 강원도와 울산광역시는 공공데이터 포털보다 많은 공공데이터를 개방하고 있지만, 2020년 이후 1년 이상 새로운 데이터를 개방하지 않고 있다.

공공데이터 포털과 지방자치단체의 개별 포털은 개방 수량뿐 아니라 데이터 세트의 목록을 다르게 제공하고 있다. 예를 들면, 강원도 데이터 포털에서 제공하는 ‘강원도 축산판매업 인허가 정보’는 공공데이터 포털 내 강원도 데이터에서 찾을 수 없다.

공공데이터 포털과 지방자치단체 데이터 포털의 데이터 일치율을 분석하기 위해 각 데이터 포털의 데이터 목록을 수집하고, 공공데이터 포털의 데이터 세트명을 기준으로 분석하였다. 높은 일치율을 보인 지방자치단체는 부산광역시(99.25%, 1,744건 중 1,731건), 경상남도(97.93%, 1,405건 중 1,376건), 대구광역시(95.16%, 599건 중 570건), 인천광역시(85.77%, 1,546건 중 1,326건), 울산광역시(78.26%, 398건 중 288건)다. 제주특별자치도(45.70%, 1,081건 중 494

건)는 절반가량 일치했고, 그 외 전라북도(14.59%, 1,206건 중 176건), 서울특별시(11.96%, 3,379건 중 404건), 경기도(4.58%, 2,162건 중 99건), 강원도(0.82%, 485건 중 4건), 경상북도(0.43%, 939건 중 4건), 세종특별자치시(0%, 77건 중 0건)는 일치율이 낮았다.

강원도, 경상북도, 세종특별자치시는 지방자치단체 데이터 포털의 개방을 중지하고 공공데이터 포털을 통해 새로운 데이터 개방하기 때문에 낮은 일치율을 보인다. 울산광역시도 운영 중단 이전의 데이터는 일치율이 높았지만, 약 1년간 새로운 데이터를 개방하지 않아 일치율이 낮다.

경기도는 개방데이터 수량이 많은 지방자치단체이지만, 경기도의 데이터 포털인 ‘경기데이터드림’은 자치구를 제외한 경기도청에서 제공하는 공공데이터만 개방한다. 반면, 공공데이터 포털은 경기도청과 자치구의 공공데이터를 모두 개방하여 저조한 일치율을 보인다. 서울특별시는 공공데이터 포털보다 2배가량 많은 데이터를 ‘서울열린데이터광장’에 개방하지만 낮은 일치율을 보인다. 그 이유는 경기도와 마찬가지로 서울특별시의 자치구 데이터는 ‘서울열린데이터광장’에서 제공하지 않고 있으며, 일치의 기준이 되는 데이터명에 ‘서울특별시’와 ‘서울시’를 혼용해서 사용하는 등 공공데이터 포털과 데이터명에 대한 표준화가 부족하다.

모든 지방자치단체의 데이터 포털은 공공데이터 관리 지침의 내용과 다르게 공공데이터 포털과 연계되지 않아 일관된 데이터 소재 관리가 어렵다는 문제가 있다.

4-4 메타데이터와 분류체계

표 3. 데이터 포털별 메타데이터 항목 수량

Table 3. Number of metadata by data portal

Public	Busan	Daegu	Gangwon	Gyeonggi	Gyeongbuk
24	14	11	14	12	17
Gyeongnam	Incheon	Jeju	Jeonbuk	Seoul	Ulsan
15	17	6	15	14	10

표 3, 표 4는 메타데이터와 분류체계를 제공하는 총 12곳의 데이터 포털을 기준으로 각 항목의 수량을 집계한 것이다.

데이터 세트의 메타데이터는 사용자가 데이터의 맥락을 이해하는 중요한 요소다. 공공데이터 포털은 24개의 메타데이터 항목을 정의하고 있다. 지방자치단체 데이터 포털은 각기 다른 메타데이터로 데이터를 관리하고 있어 항목의 개수와 표현하는 어휘가 다르다. 예를 들어, 공공데이터 포털의 ‘등록’ 항목은 ‘등록일’, ‘데이터개방일’, ‘데이터등록일’ 등 서로 다른 어휘로 표현한다.

공공데이터 포털을 기준으로 메타데이터 항목을 종합적으로 분석한 결과, ‘분류체계’, ‘제공기관’ 항목은 모든 데이터 포털이 사용하고, ‘등록’, ‘수정’, ‘관리부서명’, ‘업데이트주기’, ‘이용허락범위’ 항목도 비교적 공통으로 사용하고 있다.

각 데이터 포털의 메타데이터 기입률을 조사한 결과, ‘분류체계’, ‘등록일’, ‘수정일’, ‘키워드’, ‘이용허락범위’는 모든 포털에서 높은 수치다. 반면, ‘관리부서명’은 총 12곳의 데이터 포털 중 10곳에서 사용되는 메타데이터 요소임에도 경기도(24.76%), 전라북도(0.44%), 경상남도(0%), 대구광역시(0%), 인천광역시(0%) 5곳에서 낮은 수치를 보인다. 경상남도과 대구광역시는 모든 데이터에서 ‘관리부서명’이 공백값이고, 인천광역시는 ‘관리부서명’에 분류체계에 해당하는 값을 기재하여 관리부서명에 대한 정보가 없다.

표 4. 데이터 포털별 분류체계 항목 수량

Table 4. Number of categories by data portal

Public	Busan	Daegu	Gangwon	Gyeonggi	Gyeongbuk
16	16	16	16	9	16
Gyeongnam	Incheon	Jeju	Jeonbuk	Seoul	Ulsan
15	10	10	16	12	15

공공데이터 포털은 16개(공공행정, 과학기술, 교육, 교통물류, 국토관리, 농축수산, 문화관광, 법률, 보건의료, 사회복지, 산업고용, 식품건강, 재난안전, 재정금융, 통일외교안보, 환경기상) 항목으로 분류체계를 정의하고 있다. 부산광역시, 대구광역시, 강원도, 전라북도, 경상남도는 공공데이터 포털과 같은 분류체계를 사용하고, 이외 지방자치단체는 다른 분류체계를 사용한다. 예를 들어, ‘공공행정’, ‘교육’, ‘교통물류’, ‘문화관광’, ‘사회복지’, ‘재난안전’ 항목은 모든 데이터 포털에서 사용한다. 반면, 환경기상(10개), 농축수산(9개), 보건의료(9개), 국토관리(8개), 과학기술(7개), 통일외교안보(7개), 법률(6개), 식품 건강(6개)과 같이 일부 지방자치단체만 사용하는 항목도 존재한다. 지역 데이터 특성에 따라 서울특별시의 ‘인구/가구’, ‘주택/건설’, 경기도의 ‘도시주택’, 인천광역시의 ‘주택/건설’, 제주도의 ‘관광/여행’, 경상북도의 ‘국방’, ‘통신’은 해당 지방자치단체만 사용한다.

데이터 포털의 분류체계는 명확한 정의와 기준이 존재하지 않고, 분류항목 간 경계가 모호한 한계가 있다. ‘서울특별시 공동체주택 지원사업 정보’ 데이터는 서울열린데이터광장에서 ‘주택/건설’로 분류하지만, 공공데이터포털은 ‘국토관리’로 분류하고 있다.

분류항목의 표기 방식 또한 데이터 포털마다 다른 어휘를 사용하여 일관된 해석이 쉽지 않다. 유사한 의미임에도 공공데이터 포털은 ‘사회복지’, 서울특별시는 ‘복지’, 경기도는 ‘가족보건복지’, 인천광역시는 ‘보건복지’를 사용한다.

V. 결 론

한국의 공공데이터는 공공데이터 포털을 통해 개방하고 있으며, 적극적인 개방정책으로 OECD가 평가하는 공공데이터

개방지수에서 3회 연속으로 1위를 차지하였다. 공공데이터 포털의 데이터 중 과반수는 지방자치단체 데이터로 지방자치단체의 공공데이터 정책은 중요한 역할을 한다. 지방자치단체는 공공데이터 이용 활성화를 위한 기본계획과 시행계획을 주기적으로 수립한다. 본 연구는 지방자치단체의 공공데이터 정책을 개선하기 위해 공공데이터 행정과 관리현황을 정량적으로 분석한 연구다. 행정적인 관점으로 공공데이터 예산, 개방데이터 수량, 전문인력을 조사하였고, 운영관점에서 지방자치단체의 데이터 포털과 데이터 관리 현황을 분석하였다.

연구 결과는 다음과 같다. 공공데이터 예산을 가장 많이 확보한 지방자치단체는 경기도와 서울특별시이고, 가장 적은 예산을 투자한 곳은 세종특별자치시다. 경기도와 세종특별자치시의 경우, 최근 5개년 평균 데이터 수량과 예산 모두 약 140배 이상 차이가 난다. 예산의 세부분야 중 ‘개방 확대’와 데이터 수량의 상관분석을 진행한 결과, 양의 상관관계를 보였지만, 예산 투자 규모부터 세부분야별 예산 운용까지 지방자치단체 간 편차가 크다. 지방자치단체별 극심한 예산 분배의 차이를 극복하기 위해 공공데이터 예산 지원에 대한 정책적 논의와 투명한 공개가 필요하다.

공공데이터를 담당하는 전문인력은 지방자치단체의 공식 홈페이지에서 ‘공공데이터 책임담당관’, ‘실무담당자’로 표기된 인원을 집계한 결과, 부산광역시가 6명으로 가장 많은 인력을 보유하고 있다. 한편, 기관의 업무분담은 공공데이터 담당자가 아님에도 데이터 관련 업무를 하는 인력이 존재한다. 직급에 의한 소수 인력뿐 아니라 데이터 관리에 관여하는 모든 인력을 공공데이터 전문인력으로 고려할 수 있는 조직 체계를 마련해야 한다.

지방자치단체 중 12곳은 자체적으로 데이터 포털을 구축하였다. 개별 데이터 포털을 구축한 경우, 공공데이터 포털과 같은 데이터를 개방해야 한다는 지침이 있다. 공공데이터 포털과 지방자치단체의 데이터 포털을 비교한 결과, 모든 지방자치단체는 공공데이터 포털과 다른 데이터를 개방하고 있다. 또한, 각 데이터 포털마다 메타데이터와 분류체계는 서로 다른 항목과 어휘를 사용하고 있다. 데이터 포털의 표준이 되는 체계가 없고, 데이터 품질관리에 대한 거버넌스가 제대로 이루어지고 있지 않다. 이러한 한계점은 웹의 정보를 의미적으로 연계하여 표현하는 RDF 어휘를 통해 개선할 수 있다. W3C(World Wide Web Consortium) 표준어휘인 DCAT(Data Catalog Vocabulary)은 데이터 목록과 데이터 세트의 메타 정보를 연계할 수 있다. SKOS(Simple Knowledge Organization System)는 데이터의 분류체계를 개념체계(Concept Scheme)와 개념(Concept)의 관계를 표현하여 서로 다른 분류체계를 사용하더라도 의미적 연계가 가능하도록 표현할 수 있다.

향후 모든 공공기관의 분산된 공공데이터로 연구대상을 확장하고, 개별 포털 간 의미적 연계를 구축하여 통합적으로 관리하는 방안에 관한 연구가 필요하다.

감사의 글

이 논문은 2021년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음

참고문헌

- [1] Peter Conradie, Sunil Choenni, “On the barriers for local government releasing open data”, *Government Information Quarterly*, Vol. 31, pp. 10-17, June 2014. <https://doi.org/10.1016/j.giq.2014.01.003>
- [2] Vishanth Weerakkody, Zahir Irani, Kawal Kapoor, Uthayasankar Sivarajah and Yogesh K. Dwivedi, “Open data and its usability: An empirical view from the citizen's perspective”, *Information Systems Frontiers*, Vol. 19, No. 2, pp. 285-300, April 2017. <http://dx.doi.org/10.1007/s10796-016-9679-1>
- [3] Maxat Kassen, “A promising phenomenon of open data: A case study of the Chicago open data project”, *Government Information Quarterly*, Vol. 30, No. 4, pp. 508-513, October 2013. <https://doi.org/10.1016/j.giq.2013.05.012>
- [4] UK. National Data Strategy [Internet]. Available:<https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>.
- [5] Marijn Janssen, Yannis Charalabidis, Anneke Zuiderwijk, “Benefits, Adoption Barriers and Myths of Open Data and Open Government”, *Information Systems Management*, Vol. 29, No. 4, pp. 258-268, October 2012. <https://doi.org/10.1080/10580530.2012.716740>
- [6] Vijay Khatry, Carol V. Brown, “Designing data governance”, *Commun. ACM*, Vol. 53, No. 1, pp. 148-152, January 2010. <https://doi.org/10.1145/1629175.1629210>
- [7] Rene Abraham, Johannes Schneider, Jan vom Brocke, “Data governance: A conceptual framework, structured review, and research agenda”, *International Journal of Information Management*, Vol. 49, pp. 424-438, December 2019. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- [8] Mayernik, Matthew S, “Open Data: Accountability and Transparency.” *Big Data & Society*, Vol. 4, No. 2, pp. 1-5, December 2017. <https://doi.org/10.1177/2053951717718853>
- [9] Richard Y. Wang, Diane M. Strong, “Beyond Accuracy: What Data Quality Means to Data Consumers”, *Journal of Management Information Systems*, Vol. 12, No. 4, pp. 5-33, December 1996. <https://doi.org/10.1080/07421222.1996.11518099>

- [10] Anastasija Nikiforova, Keegan McBride, "Open government data portal usability: A user-centred usability analysis of 41 open government data portals", *Telematics and Informatics*, Vol. 58, May 2021. <https://doi.org/10.1016/j.tele.2020.101539>
- [11] Akemi Takeoka Chatfield, Christopher G. Reddick, "A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments", *Government Information Quarterly*, Vol. 34, No. 2, pp. 231-243, April 2017. <https://doi.org/10.1016/j.giq.2017.02.004>
- [12] Faizura Haneem, Nazri Kama, Nazim Taskin, David Pauleen, and Nur Azaliah Abu Bakar, "Determinants of master data management adoption by local government organizations: An empirical study", *International Journal of Information Management*, Vol. 45, pp. 25-43, April 2019. <https://doi.org/10.1016/j.ijinfomgt.2018.10.007>
- [13] Di Wang, Chuanfu Chen, Deborah Richards, "A prioritization-based analysis of local open government data portals: A case study of Chinese province-level governments", *Government Information Quarterly*, Vol. 35, No. 4, pp. 644-656, October 2018. <https://doi.org/10.1016/j.giq.2018.10.006>
- [14] Sangoh Yun, Jiwoo Hyun, "An Analysis of Open Data Policy in Korea: Focused on National Core Data in Open Data Portal", *Korean Public Management Review*, Vol. 33, No. 1, pp. 219-247, March 2019. <https://doi.org/10.24210/kapm.2019.33.1.010>
- [15] Huijeong Han, Sungwook Hwang, Jungmin Lee, and Hyojung Oh, "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 (JKLISS)*, Vol. 51, No. 1, pp. 255-279, March 2020. <https://doi.org/10.16981/kliiss.51.1.202003.255>
- [16] Haklae Kim, "Analysis of standard vocabulary use of the open government data: the case of the public data portal of Korea.", *Quality & Quantity* Vol. 53, No. 3, pp. 1611-1622, May 2019. <http://doi.org/10.1007/s11135-018-0829-z>
- [17] Cory L. Armstrong, "Providing a clearer view: An examination of transparency on local government websites", *Government Information Quarterly*, Vol. 28, No. 1, pp. 11-16, January 2011. <https://doi.org/10.1016/j.giq.2010.07.006>
- [18] Ebrahim, Z. and Irani, Z, "E-government adoption: architecture and barriers", *Business Process Management Journal*, Vol. 11 No. 5, pp. 589-611, October 2005. <https://doi.org/10.1108/14637150510619902>
- [19] Shinkon Kim, Jae-Hee Cho, "A Proposal for the Successful Introduction of Big Data to a Local Governments", *The Korean Association For Regional Information Society*, Vol. 16, No. 3 pp. 1-29, September 2013. <https://doi.org/10.22896/karis.2013.16.3.001>
- [20] Hyojin Song, Sungsoo Hwang, "Seeking Strategies for Local Governments to Prepare for Public Data Act", *The Korean Association For Regional Information Society*, Vol. 17, No. 2 pp. 1-28, June 2014. <https://doi.org/10.22896/karis.2014.17.2.001>
- [21] Hyungjun Seo, Seunghwan Myeong, "Policy Alternatives for Public Data Disclosure by Local Government: Focusing on Operation of Public Data Portal Site", *The Korean Association For Regional Information Society*, Vol. 18, No. 4, pp. 1-27, December 2015. <http://doi.org/10.22896/karis.2015.18.4.001>
- [22] Dongjun Kim, Hyeonji Kim, Chaeun Song, Jiwoo Yang and Haklae Kim, "Methods for Utilising Local Government's Public Data Released to The Public Data Portal", *Journal of Digital Contents Society(JDCS)*, Vol. 22, No. 3, pp. 445-452, March 2021. <http://doi.org/10.9728/dcs.2021.22.3.445>



송채은(Chaeun Song)

2021년 : 중앙대학교 문헌정보학사

2016년~2021년: 중앙대학교 문헌정보학과

2021년~현 재: 중앙대학교 문헌정보학과 정보확진공 석사과정

※관심분야 : 지식그래프, 공공데이터, 데이터 포털 등

김학래(Haklae Kim)



2010년 : 아일랜드 국립대학교 (공학박사)

2004년~2009년: Digital Enterprise Research Institute, Ireland

2009년~2016년: 삼성전자

2017년~2019년: 한국과학기술정보연구원

2019년~현 재: 중앙대학교 문헌정보학과 교수

※관심분야 : 지식그래프, 인공지능, 데이터 사이언스 등