

## 미세먼지 연구를 위한 공공 데이터베이스 현황 비교 분석 및 제언: 한국과 미국의 사례를 중심으로\*

### Comparative Analysis of Administrative Databases for Particulate Matter Air Pollution Between South Korea and the United States

김주희\*\* · 김선영\*\*\*

JooHui Kim · Sun-Young Kim

**요약:** 국가 기관이 운영하는 공식 측정망에서 얻어진 미세먼지 데이터는 미세먼지 관련 연구를 수행하고 정책을 수립하는 데에 필수적이다. 본 논문은 한국과 미국의 미세먼지 공공 데이터베이스를 비교·분석함으로써 한국 데이터베이스의 활용도 제고 방안을 제시한다. 본 연구는 미세먼지 관련 연구를 수행하는 연구자들의 입장에서 공공 데이터베이스의 “데이터 품질”과 “서비스 품질”을 살펴볼 수 있는 항목을 각각 3개씩 총 6개를 선정하여 한국의 에어코리아와 정보공개포털, 미국의 에어데이터를 분석하였다. 그 결과 양국의 데이터베이스는 서비스 품질에서는 비슷한 수준을 보인 반면, 데이터 품질에 해당하는 메타데이터의 제공 여부에서 가장 큰 차이를 보였다. 한국 데이터베이스의 미세먼지 데이터에 대한 정확한 이해를 바탕으로 관련 연구 및 정책 수립이 더욱 활발하게 이루어질 수 있도록 하려면, 데이터가 얻어진 배경에 대한 설명(측정 기구, 측정 방법, 측정 시 특이사항, 이상값 발생 여부 등)을 제공할 필요가 있다.  
**핵심주제어:** 미세먼지, 공공 데이터베이스, 에어코리아, 정보공개포털, 에어데이터

**Abstract:** Particulate matter (PM) data collected in regulatory monitoring networks are commonly used to assess PM distribution and its health effects. The assessments are then relied on to establish environmental and public health policies. This paper aims to offer constructive suggestions to improve the South Korean PM database system for research purposes by comparing two Korean systems with the U.S. system: “Air Korea” and “Open Data” in South Korea, and “Air Data” in the U.S. Our comparative analysis utilizes two criteria—“data quality” and “service quality”—which were developed based on previous studies. In terms of data quality, the major difference in the database systems between South Korea and the U.S. is found in metadata. Regarding service quality, the database systems in both countries provide services, such as data selection tools and user support systems. Our study suggests that the provision of relevant information supplementing measurement data can maximize the benefit of utilizing the Korean PM databases for future scientific investigation and policy recommendations.

**Key Words:** Particulate Matter, Administrative Database, Air Korea, Open Data, Air Data

\* 본 연구는 한국연구재단의 지원을 받아 수행된 기초연구사업(과제번호: 2018R1A2B6004608, 2013R1A6A3A04059017)과 국립암센터의 기관고유사업(NCC-1810220-01)의 일환으로 수행되었습니다.

\*\* 주저자, 서울대학교 협동과정 과학사 및 과학철학 박사과정

\*\*\* 교신저자, 국립암센터 국제암대학원대학교 암관리학과 부교수

## I. 서론

중앙 정부나 지방 정부에서 생산한 데이터를 공공에 공개하는 것은 시민들의 알 권리를 보장하고 투명한 의사결정이 이루어질 수 있게 하는 기반으로 여겨진다(Lathrop and Ruma, 2010; Janssen et al., 2012). 정보 공개는 시민들의 참여를 이끌어냄으로써, 국정 운영의 민주성과 개방성을 높일 수 있을 것으로 기대된 것이다. 하지만 정보 공개 그 자체를 좋은 것으로 보고, 공공 데이터가 얼마나 공개되고 있는지에만 초점을 맞추는 경향은 최근 들어 변화하는 모양새를 보이고 있다. 정보의 공개 여부보다는, 정보가 ‘어떻게’ 공개되고 있는지의 문제가 더 중요하다는 주장이 제기되고 있는 것이다. 이러한 주장의 기저에는 정보 공개가 그 자체로 “만병통치약”이 될 수는 없으며, 경우에 따라서는 부정적인 효과를 낼 수도 있다는 문제의식이 깔려 있다(Janssen et al., 2012; Levin and Leonelli, 2017). 정보의 공개 목적이나 공개 방식에 따라서, 정보 공개는 오히려 비효율이나 불평등을 가져올 수도 있다는 것이다.

이러한 문제의식을 바탕으로 본 논문은 최근 커다란 쟁점으로 떠오른 미세먼지와 관련하여 한국의 공공 데이터베이스가 어떻게 구축되어 있는지 살펴보고자 한다. 데이터가 공개되고 있는지 아닌지의 여부뿐만 아니라, 미세먼지 관련 데이터가 구체적으로 어떻게 공개되고 있는지를 검토하고자 하는 것이다. 한국에서는 보통 미세먼지를 그 크기에 따라 미세먼지와 초미세먼지로 나누어 부르는데(국립환경과학원, 2019), 본 논문에서는 미세먼지와 초미세먼지를 통칭하는 말로 미세먼지라는 용어를 사용하였으며, 그 둘을 구분해야 할 경우에는  $PM_{10}$ ,  $PM_{2.5}$ 로 표기하였다. 여기서  $PM_{10}$ 은 직경이 10 마이크로미터 이하인 입자상 물질(particulate matter less than or equal to 10 micrometers in diameter)로, 건강에 악영향을 미친다는 연구 결과가 전 세계적으로 보고되었다(Brook et al., 2010; Pope and Dockery, 2006; U.S. EPA, 2019). 최근 연구에서는 직경이 2.5 마이크로미터보다 작은  $PM_{2.5}$ 가 건강에 미치는 악영향이 더욱 큰 것으로 드러났

다. 이러한 연구 결과를 바탕으로 2014년 세계보건기구(World Health Organization) 산하 국제암연구소(International Agency for Research on Cancer)에서는 미세먼지를 “인간에게 발암성이 있는(carcinogenic to humans)” 1군(Group 1) 발암물질로 분류하기도 했다.

미세먼지의 건강 영향에 대한 이와 같은 연구들은 대부분 국가에서 대기오염을 관리할 목적으로 운영하는 측정망에서 측정한 미세먼지 자료를 이용했다(U.S. EPA, 2009; WHO, 2004). 우리나라에서 수행되는 미세먼지에 대한 연구들 역시 대부분 국가에서 운영하는 측정소의 측정 자료에 기반을 두고 있다(한혁 등, 2017; 박순애·신현재, 2017; 백정희·송재민, 2019; Lee et al., 2000; Hong et al., 2002; Kim et al., 2016; Kim et al., 2017; Kim and Song, 2017; Kim et al., 2019). 따라서 대기오염측정망의 측정 자료를 축적한 공공 데이터베이스의 품질은 미세먼지 연구의 정확성과 관련 정책의 방향에 큰 영향을 미칠 수 있다.

이에 본 논문은 “데이터 품질”, “서비스 품질”에 해당하는 여러 가지 기준들을 바탕으로 한국의 미세먼지 데이터베이스를 살펴봄으로써, 미세먼지 데이터가 연구자들에게 어떻게 제공되고 있는지를 검토해보고자 한다(한국데이터베이스진흥센터, 2000). 특히 오래전부터 방대한 미세먼지 데이터를 공개해오고 있는 미국 데이터베이스와의 비교를 통해, 한국 데이터베이스의 향후 발전 방안과 관련하여 정책적인 시사점을 제시하고자 한다. 정보 공개에 대한 기존의 연구들이 주로 일반 대중과 시민들의 필요에 맞게 정보가 공개되고 있는지의 여부에 초점을 맞추었다면(Lathrop and Ruma, 2010), 본 논문은 미세먼지와 관련된 연구를 수행하는 연구자들의 입장에서 한국의 미세먼지 데이터베이스를 검토해보고자 한다.

미세먼지는 오염원, 형성 과정, 건강 영향, 사회·경제적 효과 등 다양한 차원을 지닌 문제로, 환경과학, 의학, 정책 등 다양한 분야에서 개별 연구와 협력 연구를 활발하게 진행하고 있는 주제이다. 이처럼 다양한 분야에서 미세먼지 데이터를 활용하는 연구들이 증가하고 있는 만큼, 본 논문은 각기 다른 학문 분야에서 서로 다른 수준의 연구자들이 미세먼지 공공 데

이터베이스를 접할 때의 어려움과 개선방안에 대해서도 고민해보고자 한다. 미세먼지 데이터가 다양한 연구자들의 필요에 맞게 공개되는 것은 미세먼지에 대한 과학적 연구가 활발하게 이루어질 수 있도록 하고, 나아가 과학적 증거에 기반을 둔 정책을 수립할 수 있도록 하는 밑바탕이라는 점에서 중요하기 때문이다(김용표, 2017).

## II. 선행연구

미세먼지와 같은 대기 오염을 측정한 환경 정보(environmental data)는 대기 오염이 공중 보건에 어떤 영향을 미치는지 연구하고 그에 걸맞은 정책을 수립하는 데에 필수적이다. 이에 많은 연구들은 대기 오염과 같은 환경 오염이 사람들의 건강에 어떤 영향을 끼치는지 체계적으로 살펴볼 수 있도록, 환경 정보와 공중의 건강에 관련된 정보를 통합적으로 관리할 수 있는 시스템을 고민해왔다. 예를 들어 Thacker et al.(1996)은 그러한 시스템을 구축하기 위해서 환경 “위해(hazards)”와 그에 대한 “노출(exposure),” 그리고 위해에 대한 노출이 사람들의 건강에 미친 “결과(outcome)”라는 세 가지 요소에 대한 자료가 필수적이라고 보았다(Thacker et al., 1996). Litt et al.(2004)은 미국에서 환경 오염과 건강 영향을 통합적으로 살펴볼 수 있는 국가적 시스템을 만들 때 극복해야 할 장벽이 무엇인지를 파악하기 위해 공중의 건강과 관련된 여러 데이터베이스의 현황을 살피고 있다(Litt et al., 2004). 실제로 미국 질병관리본부(Centers for Disease Control and Prevention)는 2001년부터 환경 정보와 건강 관련 정보가 담긴 데이터베이스를 통합할 수 있는 시스템을 만들 고자 노력했는데, 이에 대해 Kyle et al.(2006)과 Strosnider et al.(2014)은 향후 과제를 제시하면서, 부처 간 협력이나 학계 전문가들과의 협업을 강조하기도 한다(Kyle et al., 2006; Strosnider et al., 2014). 이처럼 해외의 선행연구들은 주로 환경 오염과 건강 영향을 함께 살펴볼 수 있는 시스템

을 고안하는 과정에서, 대기 오염에 대한 데이터베이스가 어떻게 마련되어 있는지를 점검했다.

반면 한국에서 대기 오염과 같은 환경 오염과 관련된 데이터베이스의 현황을 검토한 연구는 부족한 실정이지만, 공공 데이터 개방에 있어 한국이 선도 국가로 제시될 만큼 공공 데이터베이스의 수가 증가함에 따라(OECD, 2017) 관련 연구들도 늘어나는 양상을 보인다. 많은 연구자들은 한국의 공공 데이터 개방을 공개되는 데이터의 양으로만 평가하는 것은 한계가 있다고 지적하면서, 다양한 방식의 공개를 위한 정책이나 제도, 법령 등을 검토해왔다. 예를 들어 한국의 공공 데이터 포털에서 제공되는 국가중점데이터의 조회수 및 활용도 등을 살펴본 연구에서는, 데이터의 양은 증가하고 있으나 가공이나 활용이 제한되는 경우가 많다고 지적했다(윤상오·현지우, 2019). 주요 국가들의 공공 데이터 개방 현황을 비교한 다른 연구에서도, 공공 데이터를 활용하기 어려운 이유를 제시하며 앞으로의 데이터 공개 정책이 “활용도” 중심으로 전환되어야 함을 제안하였다(김태엽, 2018).

이처럼 자료의 양보다는 이용자들의 활용도를 더 중점적으로 살펴봐야 한다는 문제의식을 바탕으로, 여러 선행연구들은 공공 데이터베이스의 이용자들을 대상으로 실시한 설문 조사 결과를 바탕으로 분석을 수행하고 있다. 예를 들어 이견직(2008)은 국가복지정보포털의 이용자들 50명을 무작위로 선정하여 “전반적인 만족도,” “타인에게 권유 정도,” “업무 활용의 유용성”을 평가하였다(이견직, 2008). 변대호(2005)는 한국의 전자 정부 웹사이트에서 얻을 수 있는 정보를 바탕으로 학부생들에게 문제를 내고 설문을 통해 그 정답률을 계산함으로써 해당 웹사이트들의 유용성을 점검하기도 했다(변대호, 2005). 공공 데이터의 산업적 활용이 중요하다는 입장을 바탕으로 산업적 활용 가능성이 높은 데이터들의 공개에 주목한 연구들도 있다(김지혜 등, 2016; 한국정보화진흥원, 2017; 김구, 2017). 특히 한국정보화진흥원에서는 공공 데이터를 기반으로 산업 생태계를 조성하는 것이 중요하다는 입장을 바탕으로, 공공 데이터를 활용하는 기업들을

대상으로 설문조사와 간담회를 실시하여 그 결과를 분석하고 있다(한국정보화진흥원, 2017).

과학기술 분야의 공공 데이터베이스를 살펴본 연구들 역시 데이터의 양보다는 질이 중요하다고 강조하면서 이용자 중심의 평가를 수행했다. 예를 들어 이제환(1997)은 KORDIC(연구개발정보센터, 현 한국과학기술정보연구원(KISTI))의 과학기술 분야 서지 데이터베이스 “KRISTAL DB”를 평가하기 위해 초점 집단(focus group) 인터뷰를 수행하고, 자체적으로 개발한 기준을 토대로 데이터베이스의 현황을 점검하고 있다(이제환, 1997). 이응봉 등(2001) 역시 KORDIC에서 구축한 과학기술 분야 7개의 데이터베이스를 평가하기 위해 이용자의 입장에서 직접 “검색 실험”을 실시하여 검색된 데이터의 건수나 검색에 소요된 시간 등을 비교하고 있다(이응봉 등, 2001). 김석영(2002)은 총 50개의 데이터베이스를 대상으로 저자가 “컨텐츠,” “기능성,” “디자인”과 관련된 기준들에 대해 예(1점)/아니오(0점)로 평가하여 점수를 매김으로써, 전기·전자 분야 데이터베이스들의 점수가 가장 높고 생명과학 분야 데이터베이스들의 점수가 가장 낮았다고 설명한다(김석영, 2002).

본 논문은 미세먼지 공공 데이터베이스를 중심으로 이와 같은 선행연구들의 분석에서 한 걸음 더 나아가고자 한다. 본 논문은 미세먼지 데이터가 게시되어 있는 데이터베이스의 차원에서만 분석을 수행하는 것이 아니라, 직접 미세먼지 데이터를 다운로드하여 한국과 미국의 데이터에서 어떤 차이가 드러나는지의 측면까지 비교하고자 한다. 이러한 비교 분석은 어느 쪽이 더 나은지를 판가름하기 위한 것이 아니라, 연구자들의 입장에서 한국 데이터베이스의 활용도를 제고하기 위해 필요한 것이 무엇인지를 구체적으로 따져보기 위한 것이다. 공공 데이터베이스를 평가한 기존 논문들에서는 데이터베이스의 이용자를 주로 일반인으로 보고 일반인과 연구자에 대한 구별 없이 연구를 진행해왔으며, 연구자 집단에 주목한 논문들도 데이터를 직접 다운로드 하여 분석한 내용을 담고 있지는 않다. 이에 본 논문은 최근 다양한 연구 분야에서 활용되고 있는 미세먼지 공공 데이터

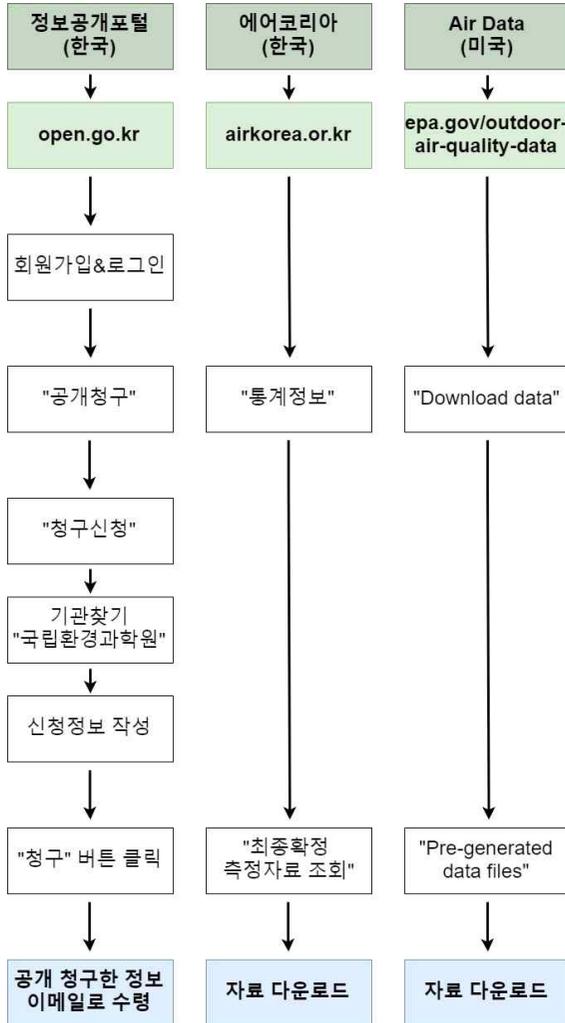
베이스에 주목하여, 미세먼지 데이터베이스가 향후 연구에 기여하기 위해 필요한 구체적인 개선 방안을 연구자들의 입장에서 찾아보고자 한다.

### III. 연구 대상

현재 한국과 미국에서는 미세먼지 데이터를 크게 두 가지 방식으로 공개하고 있다. 첫째, 미세먼지 측정치와 예보치를 실시간으로 공개함으로써 미세먼지 수준에 대한 정보를 제공하고 시민들로 하여금 미세먼지 피해에 대비할 수 있도록 하고 있다. 한국의 경우 에어코리아(Air Korea, airkorea.or.kr), 미국의 경우 에어나우(Air Now, airnow.gov)가 이러한 목적에 초점을 맞추어 운영되고 있다. 둘째, 축적된 미세먼지 원시 자료를 시간, 공간 등에 따라 공개하여 시민들의 알 권리를 보장하고 환경 행정의 투명성을 높일 수 있도록 하고 있다. 한국의 경우 정보공개포털(open.go.kr)을 통해 국립환경과학원(NIER, National Institute of Environmental Research, nier.go.kr)의 미세먼지 데이터를 받아볼 수 있으며, 미국의 경우 에어데이터(Air Data, epa.gov/outdoor-air-quality-data)에서 미세먼지 데이터를 다운로드 할 수 있다. 한편 한국의 에어코리아 역시 2018년부터 실시간 미세먼지 정보뿐만 아니라 장기간 축적된 미세먼지 데이터를 함께 제공하고 있다.

미세먼지의 실시간 자료와 장기간 축적된 원시 자료 중에서 대부분의 미세먼지 연구자들이 필요로 하는 데이터는 특정 지역의 실시간 자료라기 보다는 측정된 모든 시간과 지역을 포함하는 원시 자료이다. 이에 본 논문은 한국의 정보공개포털과 에어코리아, 미국의 에어데이터를 중점적으로 살펴보고자 한다. 각 데이터베이스에서 미세먼지 원시 자료를 얻는 방법은 <그림 1>과 같다.

〈그림 1〉 정보공개포털, 에어코리아, 에어데이터에서 미세먼지 데이터를 얻을 때의 흐름도

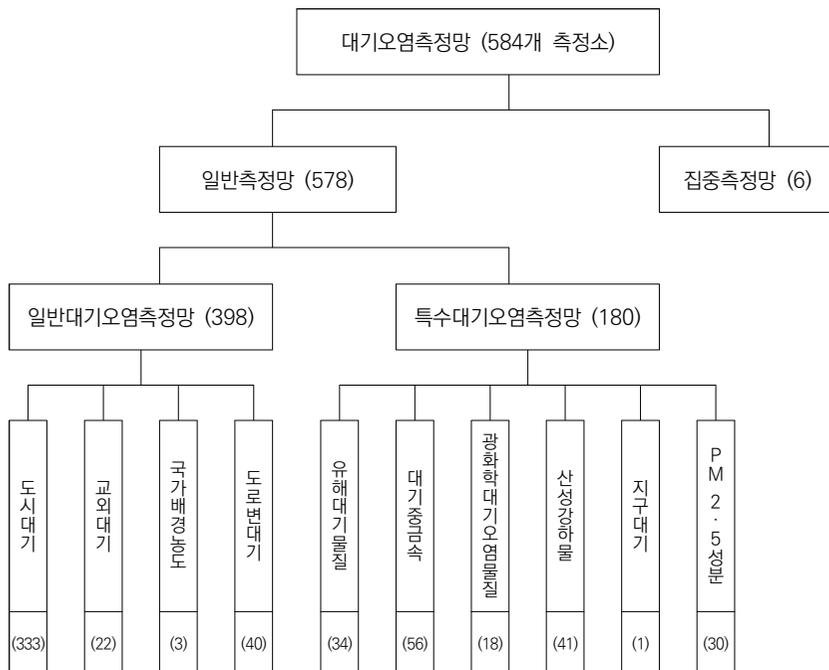


〈그림 1〉에서 알 수 있듯이, 에어코리아와 에어데이터에서는 미세먼지 원시 자료를 직접 다운로드할 수 있는 반면에, 정보공개포털을 통해 국립환경과학원의 데이터를 구할 때에는 정보 공개 청구를 해야 한다. 그럼 본격적으로 데이터베이스를 비교·평가하기에 앞서, 각각의 데이터베이스에 대해 간략히 살펴보도록 하겠다.

## 1. 한국

한국은 2018년 12월말을 기준으로 측정 목적에 따라 총 11가지 종류의 측정망을 바탕으로 중복을 포함하여 총 584개의 측정소를 운영하고 있다(국립환경과학원, 2019). 이 중에서 미세먼지를 측정하고 있는 측정소는 일반대기측정망인 398개 측정소(도시대기 333개; 도로변대기 40개; 교외대기 22개; 국가배경농도 3개), 특수대기오염측정망인 PM<sub>2.5</sub> 성분 측정망 30개 측정소, 그리고 집중측정망 6개 측정소이다(〈그림 2〉).

〈그림 2〉 한국의 대기오염측정망 현황



자료: 환경부·국립환경과학원(2018), 국립환경과학원(2019)을 재구성

각 측정소들을 관할하고 있는 수도권대기환경청, 한국환경공단, 지방자치단체 보건환경연구원과 같은 기관들은 미세먼지 측정치를 1차로 확정된 후, 이를 국가대기오염 정보관리시스템(National Ambient air quality

Monitoring Information System, 이하 NAMIS)으로 전송한다. 이후 NAMIS의 데이터는 국립환경과학원의 전산 시스템으로 전송되며, 통계 처리와 최종 확정 과정을 거쳐 데이터베이스의 형태로 저장된다(국립환경과학원, 2019). 이렇게 저장된 데이터는 정보공개포털에서 공개를 청구하면 받아볼 수 있게 되어 있다.

정보공개포털은 1996년에 제정·공포되어 1998년부터 시행된 “공공기관의 정보공개에 관한 법률(정보공개법)”에 전자적 정보 공개의 법적 근거가 마련됨에 따라, 2006년에 개설되어 오늘에 이르고 있다(정진임·김유승, 2014). 한국의 정보공개법은 아시아에서 최초로 입법된 정보공개법으로서, 이에 기반을 두고 있는 전자적 정보공개제도 역시 다른 국가들에 비해 비교적 이른 시기에 도입되어 안정적으로 운영되고 있다(홍강훈, 2012; 전술비·강순애, 2016). 이러한 전자적 정보공개제도의 중심축을 담당하고 있는 정보공개포털은 중앙행정기관 및 지방자치단체들이 가진 여러 정보들을 공개하도록 청구할 수 있는 사이트로, 일반 시민과 행정 기관들을 대개해주는 역할을 수행하고 있다. 국립환경과학원에 저장된 미세먼지 데이터 역시 이를 필요로 하는 사람이라면 누구나 정보공개포털을 통해 공개를 청구하여 받아볼 수 있다. 이전까지는 연구자들이 직접 연구기관 명의의 공문을 국립환경과학원에 발송해야 했지만, 정보공개포털이 생긴 이후에는 누구나 공문 없이도 데이터를 청구할 수 있다.

에어코리아는 2002년 한·일 월드컵의 성공적인 개최를 위해 경기장 주변의 대기오염도를 공개했던 것이 계기가 되어, 한국환경공단이 구축한 NAMIS의 자료를 실시간으로 시민들에게 공개하고 있는 웹사이트이다. 2020년 2월 현재 에어코리아는 실시간 미세먼지 데이터뿐만 아니라 국립환경과학원의 확정을 거친 2001-2019년 PM<sub>10</sub>, 2015-2019년 PM<sub>2.5</sub>의 시간별 자료를 공개하고 있다. 이렇게 에어코리아에서 미세먼지 원시 자료를 직접 다운로드 할 수 있다는 점은 정보공개포털과 비교했을 때 가장 큰 차이를 보이는 부분이다. 에어코리아와 달리 정보공개포털을 이용하는 경우, 연구자들은 정보 공개 청구에 앞서 자신의 필요에 맞는 자료가 국립

환경과학원에 있는지 문의를 통해 확인해야 하기 때문이다.

## 2. 미국

미국의 미세먼지 공공 데이터베이스에 대한 법적 근거 마련과 시스템 구축은 한국보다 20년 이상 앞서 시작되었다. 1970년에 제정되어 1990년에 개정된 청정대기법(Clean Air Act)에 따라 미국은 6가지 대기오염 기준 물질((오존(ozone, O<sub>3</sub>), 이산화질소(nitrogen dioxide, NO<sub>2</sub>), 일산화탄소(carbon monoxide, CO), 이산화황(sulfur dioxide, SO<sub>2</sub>), 납(Pb), 미세먼지(PM))에 대한 국가 대기환경기준(National Ambient Air Quality Standard, NAAQS)을 마련하였으며, 국립과학원(National Academy of Sciences, NAS)의 권고 등에 발맞추어 측정망을 발전시켜왔다. 현재 미국 전역에서는 3000개 이상의 측정소에서 여러 종류의 가스상, 입자상 물질을 측정하고 있으며, 미세먼지 역시 여러 종류의 네트워크에서 측정이 이루어지고 있다(CENRS, 2013). 미세먼지는 6가지 대기오염 기준 물질 중 하나이기 때문에, 대기오염 기준 물질을 측정하는 약 300여 개 주·지역 정부의 측정소(SLAMS, State and Local Air Monitoring Stations)에서 기본적인 측정이 이루어지고 있다. 1997년에 PM<sub>2.5</sub>의 대기환경기준이 정해진 후에는 PM<sub>2.5</sub>의 질량 농도를 측정하기 위해 약 1500여 개의 측정소로 이루어진 네트워크가 구축되기도 했다. 그뿐만 아니라 기존의 측정 네트워크를 보완하기 위해 2006년 이래 구축된 NCore(National Core) 네트워크에서도 PM<sub>10</sub>과 PM<sub>2.5</sub>의 질량 농도 및 구성 성분에 대한 측정을 수행하고 있다(Demerjian, 2000, 2007; 한국환경공단, 2010; CENRS, 2013).

이처럼 여러 측정소에서 측정된 미세먼지 자료는 모두 AQS(Air Quality System)로 모이게 되며, AQS Data Mart라는 데이터베이스에 저장된다. AQS Data Mart는 규제와 관련된 영역이나 학계, 보건 연구 커뮤니티 등 대기 관련 자료를 필요로 하는 다양한 분야의 분석가들을 위해 만들어진 것으로, 미국 환경보호청(Environmental Protection Agency, 이하 EPA) 웹사이트의 에어데이터에서 직접 다운로드 할 수 있다. 에어데이터는 일

반 시민들뿐만 아니라 전문가들에게 필요한 아주 구체적이고 기술적인 정보들을 모두 제공하는 웹사이트로서 2011년부터 서비스를 시작했다. 이와 같은 정보 공개의 흐름은 2009년 버락 오바마 전 미국 대통령이 “정보 공개(open data)”를 최우선 과제로 천명하고 내무부(Department of the Interior)와 EPA의 최고 책임자들로 하여금 정보 공개 웹사이트(data.gov)를 구축하여 시민들에게 정보를 투명하게 공개하도록 했던 것과는 궤를 같이 한다(Peled, 2011; Coglianese, 2009).

본 논문에서는 AQS Data Mart의 정보가 공개되어 있는 EPA의 에어데이터를 중점적으로 살펴보고자 한다. 미국 정부의 정보 공개를 향한 움직임의 중심에 EPA가 있었던 만큼, EPA는 정보 공개 웹사이트(data.gov)에도 여러 정보를 게시했다(Peled, 2011). 하지만 해당 웹사이트를 통하지 않고서도 에어데이터 웹사이트에서 모든 자료를 바로 다운로드 할 수 있는 만큼, 미국의 경우에는 에어데이터에만 초점을 맞추고자 한다.

#### IV. 연구 방법

미세먼지 공공 데이터베이스와 같은 웹 기반 데이터베이스를 평가할 수 있는 기준과 방법은 다양하다. 먼저 한국데이터베이스진흥센터(2000)가 정리한 기준들을 중심으로 국내 선행연구들이 어떤 기준을 사용했는지를 요약하면 아래 <표 1>과 같다.

<표 1>에 정리된 평가 기준들 중에서 네 개 이상의 선행연구가 채택한 것으로는 정확성, 완전성, 최신성, 검색성, 사용 용이성, 사용자 지원성이 있다. 이중 정확성, 완전성, 최신성은 “데이터 품질”을 살펴볼 수 있는 평가 기준이며, 검색성, 사용 용이성, 사용자 지원성은 “서비스 품질”을 살펴볼 수 있는 평가 기준이다(한국데이터베이스진흥센터, 2000).

〈표 1〉 웹 기반 데이터베이스 평가 기준 정리

	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ <sup>1</sup>
수록 범위		● ("포괄성")	●	●		
정확성	●	● ("적합성")	●	●	● ("적합성")	●
일관성	●	●				●
완전성	●		●	●		●
최신성	●	●	●	●	●	●
통합성						●
내용				●		
권위			●	● ("전문성")	●	●
검색성	●	●		●	●	●
사용 용이성	●	●	●	●	●	●
사용자 지원성	●	●	●	●	●	●
디자인					●	●
비용			●	●		●
하드웨어 네트워크			●	●		

ㄱ. 한국데이터베이스진흥센터(1996) ㄴ. 이제환(1997) ㄷ. 한국데이터베이스진흥센터(2000) ㄹ. 이응봉 등(2001) ㅁ. 김석영(2002) ㅂ. 홍현진(2005)

주: 1) "수록 범위," "정확성," "일관성" 등 선행 연구에서 공통적으로 쓰인 총 14 개의 기준은 한국데이터베이스진흥센터(2000, p.20)에서 "출력" 대신 "디자인"을 넣어 구성한 것이다. 이제환(1997), 이응봉 등(2001), 김석영(2002)은 모두 과학기술분야의 데이터베이스의 품질을 분석한 연구이며, 한국데이터베이스진흥센터(1996), 한국데이터베이스진흥센터(2000), 홍현진(2005)은 데이터베이스 일반을 평가할 수 있는 품질 평가 기준을 제시하고 있는 연구이다

2) <sup>1</sup>홍현진(2005)은 그 외에도 객관성, 유일성, 정보적합성, 정보탐색 충족률, 정보 유용성, 정보 영향력, 비용 대비 효과성, 이용자 만족도를 평가 기준으로 제시함

본 논문에서도 한국과 미국의 데이터베이스의 데이터 품질과 서비스 품질을 살펴보기 위해, 선행연구들이 활용한 평가 항목들을 미세먼지 관련 연구자의 입장에서 재구성하였다. 여기서 데이터 품질을 평가하기 위한 항목들은 데이터 자체가 어떻게 마련되어 있는지, 연구자들이 데이터를 받았을 때 그것을 잘 이해할 수 있도록 되어있는지를 살펴보는 것들로 구

성하였다. 서비스 품질을 살펴보는 항목들은 연구자들이 자료를 쉽게 받을 수 있는지, 그것을 잘 활용할 수 있도록 되어있는지를 살펴보는 것들로 구성하였다. 비교 분석에 활용한 항목은 아래 <표 2>와 같다.

<표 2> 미세먼지 데이터베이스 비교 분석 기준

데이터 품질	어떤 데이터를 제공하는가?
	메타데이터가 있는가?
	데이터는 얼마나 자주 갱신되는가?
서비스 품질	사용자가 시·공간 스케일을 선택할 수 있는가?
	자료를 얻는 데에 얼마만큼의 비용과 시간이 드는가?
	사용자 안내서, 연락 정보 등 사용자 지원 시스템이 잘 갖추어져 있는가?

먼저 데이터 품질의 측면에서는 어떤 종류의 미세먼지 데이터를 제공하는지, 메타데이터가 있는지, 데이터는 얼마나 자주 업데이트 되는지를 살펴보고자 한다. 이는 많은 선행연구들이 활용한 기준인 정확성, 완전성, 최신성을 살펴볼 수 있게 하는 항목들이기도 하다. 서비스 품질의 측면에서는 연구자가 미세먼지 데이터의 시·공간 스케일을 선택할 수 있는지, 자료를 얻는 데에 얼마만큼의 시간과 비용이 드는지, 사용자 안내서나 연락 정보 등 사용자 지원 시스템이 잘 갖추어져 있는지를 검토하고자 한다. 이러한 항목들은 검색성, 사용 용이성, 사용자 지원성을 살필 수 있는 것들이기도 하다.

대부분의 선행연구들은 <표 1>에 제시된 기준들을 정량적 방법을 통해 살펴봄으로써 데이터베이스를 평가하였다. 예를 들어 <표 1>에 정리된 기준들을 측정할 수 있는 질문을 따로 만들어 이용자들에게 설문조사를 실시함으로써, 점수를 계산한 연구가 있다(홍현진, 2005). 혹은 각 항목을 예/아니오로 구성하여 점수를 계산하거나(김석영, 2002), 데이터베이스에서 제공되는 정보를 바탕으로 문제를 만들어 정답률을 계산함으로써 데이터베이스의 유용성을 검토한 연구도 있다(변대호, 2005). 이와 같은 평가는 이용자들의 경험을 수치화하여 포착해내고 서로 다른 데이터베이스 사이의 정량적인 비교를 가능하게 한다는 점에서 중요하다.

하지만 본 논문은 점수화하기 어렵지만 연구자들에게 중요한 특징들에 주목하여 미세먼지 데이터베이스의 현황을 분석하고자 한다. 한국과 미국의 미세먼지 데이터베이스에 점수를 매기고 둘 중 어느 것이 더 나은지를 평가하려는 것이 아니라, 양국 데이터베이스의 현황과 특성을 파악함으로써 연구자의 입장에서 데이터베이스의 발전 방안을 구체적으로 제시하는 것이 본 논문의 목적이기 때문이다. 이에 본 논문은 이제환(1997)과 이응봉 등(2001)의 연구에서 검색 실험을 수행했던 것처럼 한국과 미국의 데이터베이스가 어떻게 다른지 검색을 통해 파악해보고(이제환, 1997; 이응봉 등, 2001), 나아가 각각의 데이터베이스에서 다운로드 한 데이터가 어떤 공통점과 차이점을 보이는지를 살펴보고자 한다. 이러한 비교 분석은 <표 2>의 총 6가지의 평가 항목을 바탕으로 수행되었다. 이와 같은 분석 결과를 토대로 본 논문은 한국의 미세먼지 공공 데이터베이스의 발전을 위한 정책적 시사점을 보다 구체적으로 제시하고자 한다.

## V. 연구 결과

### 1. 데이터 품질

#### 1) 어떤 데이터를 제공하는가?

미국 EPA 에어데이터 웹사이트의 “자료 다운로드(Download Data)” 섹션은 크게 “예비 자료(Pre-generated Data)”, “일별 자료(Download Daily Data)”, “원자료(Download Raw Data)”의 세 가지 부분으로 나누어져 있다. 먼저 예비 자료 섹션에는 PM<sub>10</sub>과 PM<sub>2.5</sub>의 질량이나 구성 성분을 FRM(Federal Reference Method)이나 FEM(Federal Equivalent Method)으로 측정된 시간별·일별 자료가 모두 게시되어 있다. 여기서 FRM은 연방의 여러 측정 기관들이 대기오염을 측정할 때에 표준적으로 사용하는 방법을 일컬으며, FEM은 FRM과는 다른 샘플링 방법이나 분석 기술을 사

용하지만 대기질에 대한 의사 결정에 있어 FRM에 기반을 둔 자료와 동등한 영향력을 발휘할 수 있는 자료를 생산하는 측정 방법을 일컫는다. 예비 자료 섹션의 자료들 중에서 정부가 구축한 측정 네트워크에서 FRM으로 미세먼지를 측정한 자료로는 PM<sub>10</sub>의 경우 약 830개 측정소(2010년 기준)에서 측정한 일별 농도 자료(1988~2019년, 2020년 2월 기준), PM<sub>2.5</sub>의 경우 약 930개 측정소(2010년 기준)에서 측정한 일별 농도 자료(1997~2019년, 2020년 2월 기준)가 있다. 이와 같은 일별 농도 자료는 FRM에 따라 필터의 중량을 측정하여 미세먼지의 질량 농도를 파악한 것으로, 대부분 3일 간격으로 이루어진 측정에 기반을 두고 있다(〈표 3〉). 또한 예비자료 섹션에서는 미세먼지 외의 다른 기준 물질들의 시간별·8시간 평균·일별 자료, 기온이나 풍속과 같은 기상 자료 등을 다운로드 할 수 있다.

〈표 3〉 에어코리아와 에어데이터의 미세먼지 자료 현황 비교\*

		에어코리아 <sup>1</sup> (한국)	Air Data <sup>2</sup> (미국)
PM <sub>10</sub>	공간	약 270개 <sup>3</sup> (2015년 기준)	약 830개 <sup>4</sup> (2010년 기준)
	시간	2001~2019년	1988~2019년 <sup>4</sup>
	측정 간격	시간별 측정	일별 측정
	관련 정보 <sup>5</sup>	지역, 측정소 코드, 측정소명, 측정 일시, 주소	측정소 코드, 경위도, 샘플링 시간, 측정 일자, 대기질 지수 (Air Quality Index), 측정 방법, 최종 수정 일자 등 26개
PM <sub>2.5</sub>	공간	약 110개 <sup>3</sup> (2015년 기준)	약 930개 <sup>4</sup> (2010년 기준)
	시간	2015~2019년	1997~2019년 <sup>4</sup>
	측정 간격	시간별 측정	일별 측정, 3일 혹은 6일마다
	관련 정보 <sup>5</sup>	지역, 측정소 코드, 측정소명, 측정 일시, 주소	측정소 코드, 경위도, 샘플링 시간, 측정 일자, 대기질 지수 (Air Quality Index), 측정 방법, 최종 수정 일자 등 26개

주: 1) 에어코리아, 에어데이터 웹사이트의 2018년도 미세먼지 데이터를 다운로드 하여 비교

2) 국립환경과학원의 최종 확정을 거친 자료 <sup>2</sup>에어데이터에 게시되어 있는 자료 중에서 Federal Reference Method로 측정한 자료로 한정 <sup>3</sup>이수현 등(2018) <sup>4</sup>Kim et al.(2020) <sup>5</sup>자료 파일에 포함되어 있는 측정값 관련 메타데이터

일별 자료 섹션에서는 대기오염 기준 물질인 PM<sub>10</sub>, PM<sub>2.5</sub>, O<sub>3</sub>, SO<sub>2</sub>, CO, NO<sub>2</sub>, Pb 중에서 하나를 선택한 뒤 연도와 지역을 고르면, 사용자가 설정한 조건에 해당하는 일별 자료를 엑셀 파일의 형태로 바로 다운로드 할 수 있다. 원자료 섹션에서는 일별 자료 섹션에서보다 훨씬 더 세세하게 조건을 설정하여 자료를 다운로드 할 수 있다. 예를 들어 사용자의 필요에 따라 PM<sub>2.5</sub>의 구성 성분 중에서도 금속 물질들(metal pollutants)의 측정치만 따로 다운로드 하거나, 특정 측정소에서 측정된 자료들만 받을 수도 있으며, 이용 가능한 자료들을 모두 다 다운로드 하는 것 역시 가능하다.

한국의 경우, 일반대기오염측정망에서 측정되어 국립환경과학원의 확정을 거친 미세먼지 데이터는 정보공개포털을 통해 공개를 청구하거나 에어코리아에서 직접 다운로드 할 수 있다. 이들 자료는 전국 약 250개 측정소에서 측정한 2001-2019년의 PM<sub>10</sub> 시간별 농도를 포함하고 있다. PM<sub>2.5</sub>의 경우에는 법적으로 공개가 의무화된 시점이 2015년 1월 1일이기 때문에, 약 110개 측정소에서 측정한 2015-2019년 PM<sub>2.5</sub>의 시간별 농도 자료가 공개되어 있다. 앞서 언급했듯이 EPA에서는 필터의 중량을 측정하는 중량법(gravimetric method)을 이용하기 때문에 주로 일별 측정 농도 자료를 제공하는 반면, 한국은 베타선 방식( $\beta$ -ray Attenuation Method, BAM)의 자동측정장치를 사용하기 때문에 시간별 농도값을 제공하는 것이다(〈표 3〉). 에어코리아에 공개되어 있는 자료는 정보공개포털을 통해 국립환경과학원에 정보 공개를 청구하는 방식으로도 얻을 수 있다.

에어코리아에서 바로 다운로드가 가능한 일반대기오염측정망의 자료 외에, 집중측정망이나 특수대기오염측정망의 PM<sub>10</sub> 및 PM<sub>2.5</sub>의 성분 측정 자료는 정보공개포털을 통해 국립환경과학원 측에 공개를 요청하여 받아볼 수 있다. 그러나 자료제공시스템이 체계화되어 있는 일반대기오염측정망의 자료와는 달리, 이 경우에는 국립환경과학원에서 자료를 취합하고 가공하는 등 추가적인 작업이 필요한 상황이다. 이에 일반대기오염측정망이 아닌 다른 측정망의 자료를 필요로 하는 경우에는 정보공개포털을 통해 국립환경과학원 담당자와의 연락을 거쳐 공문을 발송했을 때 원하는 자료를 이메일로 받을 수 있다.

## 2) 메타데이터가 있는가?

메타데이터란 특정 데이터의 속성에 대한 정보를 담고 있는 데이터를 뜻한다. 예를 들어 미세먼지 측정 자료에서는 미세먼지 측정치가 어디서, 언제, 어떻게 측정된 것인지에 대한 정보를 메타데이터라고 할 수 있다. 미국의 경우 측정소 위치나 측정 기구, 측정 방법에 대한 정보는 에어데이터 웹사이트 예비 자료 섹션의 “측정소 및 측정기구 설명(Site and Monitor Descriptions)”에서 얻을 수 있다. 나아가 데이터의 포맷이나 측정 시 특이사항, 이상값 발생 여부 등의 각종 변수에 대한 설명은 같은 예비 자료 섹션의 “데이터와 포맷에 대한 설명(Description of Data and Formats)”에서 찾아볼 수 있다. 해당 섹션에서는 에어데이터에서 다운로드가 가능한 모든 자료들의 메타데이터를 제공하고 있다. 시간별, 8시간 평균, 일별, 연도별 자료 각각에 해당하는 CSV 형식의 자료 파일이 포함되어 있는 변수들의 이름과 의미를 따로 정리하여 제공하고 있는 것이다.

반면 한국의 경우에는 자료를 다운로드 받았을 때  $PM_{10}$ ,  $PM_{2.5}$ 의 측정값과 관련한 메타데이터로는 측정소에 대한 정보 외에는 별다른 정보가 주어지지 않는다(표 4). 측정 기기나 측정 방법, 측정 시 특이사항, 이상값 발생 여부 등은 미세먼지의 측정값 그 자체에 영향을 미치는 요인이며, 측정 자료를 바탕으로 한 노출 평가나 건강 영향 분석에도 영향을 미칠 수 있다. 예를 들어 평상시보다 훨씬 높게 측정된 미세먼지 측정값이 자료에 포함되어 있다면, 그것은 산불과 같은 특수한 상황이 발생하여 실제로 미세먼지 농도가 높았던 것일 수도 있고, 혹은 측정 기기에 이상이 생겼던 것일 수도 있다. 하지만 국립환경과학원 자료의 경우에는 측정값만 제공되기 때문에 측정값에 대한 정확한 이해가 어려운 상황이다. 각 측정소에 설치되어 있는 측정 기기에 대한 정보도 포함되어 있지 않아서, 측정소들 간 다른 기기 설치나 한 측정소 내에서 기기 교체에 따라 발생할 수 있는 측정값의 차이도 확인할 수 없는 상황이다. 국립환경과학원으로부터 자료를 받을 경우 자료의 활용이나 재배포와 관련한 유의 사항이 담긴 파일이 함께 오는데, 해당 파일에도 메타데이터는 포함되어 있지 않은 실정이다.

〈표 4〉 세 개의 데이터베이스에서 다운로드 한 자료 파일에 포함되어 있는 변수들의 목록

	정보공개포털	에어코리아	Air Data	
측정소 정보	시도	지역	State code	
	도시		State name	
	시군구 <sup>1</sup>	측정소명	County code	
	측정소명		County name	
	TMSID(측정소 코드)	측정소 코드	Site number	
	주소	주소		
			Latitude	
		Longitude		
샘플링 관련 정보	YYYYMMDDHH (측정 일시)	측정 일시	Date local	
			Time local	
			Date GMT <sup>2</sup>	
			Time GMT <sup>2</sup>	
				Parameter name
				Parameter code
				POC <sup>3</sup>
				Datum
				Method type
				Method code
				Method name
				Units of measure
				MDL <sup>4</sup>
Uncertainty				
Qualifier				
			Date of last change	
측정값	SO <sub>2</sub>	SO <sub>2</sub>	Sample measurement <sup>5</sup>	
	CO	CO		
	O <sub>3</sub>	O <sub>3</sub>		
	NO <sub>2</sub>	NO <sub>2</sub>		
	PM <sub>10</sub>	PM <sub>10</sub>		
	PM <sub>2.5</sub>	PM <sub>2.5</sub>		

주: <sup>1</sup>한국의 행정구역 단위 명칭 <sup>2</sup>Greenwich mean time <sup>3</sup>Parameter occurrence code <sup>4</sup>Method detection limit <sup>5</sup>한국의 경우 하나의 파일에 6개 오염물질의 측정값을 모두 제공하고 있는 반면, 미국 에어데이터의 경우 각 오염물질별로 파일을 따로 만들어 측정값 정보를 제공하고 있다

물론 메타데이터는 대기환경연보나 월보에서 일부 얻을 수 있다. 예를 들어 측정소 코드가 정리된 대기오염 측정망 제원이나 측정 방법에 대한 정보는 대기환경연보에 담겨있다. 에어코리아 웹사이트의 경우에는 최종 확정자료를 다운받는 곳에 “측정소 코드는 대기환경연보 [부록14. 대기오염 측정망 제원]을 참고”하라고 명시해 놓고 있기도 하다.

### 3) 데이터는 얼마나 자주 갱신되는가?

에어데이터 웹사이트에 명시되어 있는 바에 따르면, 미세먼지의 시간별, 일별 자료를 연도별로 다운로드 할 수 있는 EPA 에어데이터 예비자료 섹션은 1년에 두 번, 6월과 12월에 자료가 업데이트된다. AQS Data Mart의 원자료(pre-extracted data)가 매년 5월과 11월에 업데이트되고, 이를 바탕으로 정리된 자료가 6월과 12월에 에어데이터에 업데이트되는 것이다. 또한 AQS Data Mart의 자료가 에어나우에 실시간으로 공개되는 것처럼, 에어데이터의 일별 자료 섹션 역시 기본적으로 매일 측정 자료가 갱신되고 있다.

반면 에어코리아는 자료를 언제, 어떤 주기로 갱신하는지 명시해 놓지는 않고 있다. 갱신 주기에 대한 정보는 국립환경과학원의 웹사이트에도 게시되어 있지 않기 때문에, 자료가 언제 업데이트되는 것인지 알기는 어려운 실정이다. 하지만 비록 원자료는 아닐지라도 국립환경과학원의 가공을 거친 요약(summary) 자료는 매년 대기환경연보와 매월 대기환경월보를 통해 시민들에게 공개되고 있다. 환경부 산하의 국립환경과학원이 발간하는 대기환경 연·월보는 환경부 디지털 도서관이나 에어코리아의 통계정보 탭에서 찾아볼 수 있다.

## 2. 서비스 품질

### 1) 사용자가 시·공간 스케일을 선택할 수 있는가?

미국 EPA의 에어데이터의 경우, 일별 자료, 원자료 섹션에서 사용자의

필요에 맞게 시·공간 스케일을 선택하여 자료를 다운로드 할 수 있다. PM<sub>10</sub>, PM<sub>2.5</sub>을 포함한 법정 기준 오염물질들에 대한 자료를 연도와 지역을 선택하여 다운로드 할 수 있으며, 특히 원자료 섹션에서는 사용자의 필요에 맞게 직접 쿼리(query)를 구성하여 자료를 다운로드 할 수 있다. 앞서 살펴보았던 것처럼 PM<sub>2.5</sub>의 구성 성분 중에서도 금속 물질들의 측정치만 따로 다운로드 하거나, 특정 측정망에서 측정된 자료들만 받는 것 역시 가능하다. 예를 들어 지역명을 검색하여 해당 지역 근처의 측정소에서 측정된 6가지 오염물질의 시간별, 일별 자료를 확인할 수 있으며, 엑셀 파일 형태로 다운로드가 가능하다.

에어코리아 역시 이와 비슷한 기능을 갖추고 있다. 예를 들어 에어코리아는 통계정보 탭의 “측정소별 확정자료” 섹션에서 측정소별로 법정 기준 6개 오염물질의 시간별 자료와 일평균 자료를 다운로드 할 수 있도록 하고 있다. “측정망·항목별 확정자료” 섹션에서도 측정망의 종류(국가배경, 교외대기, 도시대기, 도로변대기), 지역(서울, 경기, 인천, 강원, 충남 등), 오염물질(PM<sub>10</sub>, PM<sub>2.5</sub>, CO, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>), 측정소(시·군·구, 읍·면·동 단위에 설치된 측정소)와 조회 기간을 사용자의 필요에 맞게 지정하여 자료를 조회할 수 있도록 하고 있다. 반면 “측정망·항목별 확정자료” 섹션의 경우 사용자의 필요에 맞게 측정망, 지역, 항목, 측정소, 조회 기간을 지정하여 자료를 조회할 수는 있으나, 엑셀 파일 형태로 다운로드 하는 것은 불가능하다.

한편 정보공개포털을 통해 국립환경과학원의 자료를 공개 청구하는 경우에는, 연구자가 연구 목적에 맞게 시·공간 스케일을 특정하여 자료를 청구하면 국립환경과학원 측이 제공할 수 있는 범위 내에서 자료를 제공하는 식으로 정보 공개가 이루어진다. 보통 국립환경과학원에서는 미세먼지의 시간별 자료를 제공하며, 연구자들은 연구 목적에 따라 시·공간 스케일을 조정하고 자료를 가공하여 연구를 수행하게 된다.

에어코리아에서 직접 다운로드 할 수 있도록 공개된 자료보다 장기간이나 다양한 오염물질에 대한 측정 자료는 정보공개포털을 통해 국립환경과

학원에 요청할 수 있다. 하지만 구체적으로 자료를 요청할 수 있는 시기나 오염물질에 대한 정보가 게시되어 있지 않기 때문에, 연구자들은 요청 시 담당자와 문의를 주고받아야 하는 상황이다.

## 2) 자료를 얻는 데에 얼마만큼의 비용과 시간이 드는가?

EPA 에어데이터의 자료와 에어코리아의 자료는 모두 무료로 바로 다운로드가 가능하다. 이용자의 인터넷 속도나 다운로드 하고자 하는 파일 크기에 따라 걸리는 시간은 다르더라도, 모든 자료를 곧바로 다운로드 할 수 있는 것이다.

정보공개포털을 통해 국립환경과학원의 자료를 받을 때에는 정보공개포털의 수수료 정책을 따르게 되며, 전자 파일로 자료를 받는 경우는 비용이 모두 무료이다. 또한 정보공개포털은 10일 이내에 관련 기관에서 자료 공개를 처리하도록 하고 있기 때문에, 늦어도 10일 안에는 원하는 자료를 받아볼 수 있다. 이에 국립환경과학원에서도 미세먼지 자료를 공개하는데에 10일 이상 소요될 것으로 예상되는 경우에는, 정보공개포털을 통하지 않고 국립환경과학원에 직접 공문을 발송하여 자료를 받아볼 것을 권장하고 있다.

## 3) 사용자 안내서, 연락 정보 등 사용자 지원 시스템이 잘 갖추어져 있는가?

에어데이터는 다양한 사용자 지원 시스템을 갖추고 있다. 예를 들어, 에어데이터에서는 쿼리를 구성하는 과정에 대한 설명을 가이드라인에 담아 제시함으로써, 사용자가 자신의 목적에 맞게 쿼리를 생성할 수 있도록 돕고 있다. 또한 사용자가 EPA의 자료를 활용하는 과정에서 가질 수 있는 의문을 “자주 하는 질문(Frequent Questions)”란을 통해 다루고 있다. 특히 “에어데이터 웹사이트에서 얻은 자료나 그래픽은 어떻게 인용해야 하나요?”, “CSV 파일에는 대기오염 농도에 영향을 미치는 예외적인 상황(exceptional event)이 표시되어 있나요?”와 같은 질문에 대한 답변들은

연구자들이 EPA의 자료를 활용하는 과정에서 꼭 필요한 정보를 담고 있다고 할 수 있다.

이에 더해 에어데이터는 미국 연방정부 홈페이지(USA.gov), 정보공개 웹사이트(data.gov), 규제 검색 사이트(Regulations.gov)와 함께 EPA의 뉴스 검색 사이트 등으로 통하는 링크를 제공하고 있다. 연락 정보도 다양하게 제공하고 있는데, EPA의 소셜 미디어 정보나 이메일 주소, 본부나 지역 사무실의 연락처, 특정 직원의 연락처 등을 검색할 수 있도록 하고 있다. 이에 더해 기름 유출과 같은 긴급 상황이나 환경 관련 법령을 위반한 사례를 신고하는 방법, 환경 관련 규제에 대한 의견을 남기는 방법 등을 제시함으로써 사용자들의 적극적인 참여도 이끌어내고 있다. 나아가 EPA는 “요약 리포트 만들기(Generate Summary Reports)”, “데이터 시각화하기(Visualize Data)”, “기술적 리포트 만들기(Generate Technical Reports)”와 같은 기능을 제공함으로써 사용자들로 하여금 자료를 손쉽게 가공할 수 있도록 돕고 있다.

앞서 살펴본 것처럼 정보공개포털은 국립환경과학원을 통해 미세먼지 자료를 제공하고 있으며, 자료를 사용하는 것과 관련하여 따로 매뉴얼을 제공하고 있지는 않다. 정보공개포털은 미세먼지 자료뿐만 아니라 다른 분야의 정부자료 또한 공개하는 포털이기 때문에, 자료의 내용에 대한 것 보다는 주로 수수료 납부와 같이 정보공개포털을 이용할 때 발생할 수 있는 기술적 문제와 관련된 질문들을 주로 다루고 있다. 미세먼지 자료와 관련된 사용자 지원 시스템은 국립환경과학원 웹사이트에서 일부 찾아볼 수 있다.

국립환경과학원 웹사이트에서는 정보공개를 가장 첫 번째 탭으로 두고, 정보공개제도란 무엇이며 정보공개 청구가 어떤 방식으로 이루어지는지를 안내하고 있다. 이에 더해 2020년 2월 현재 총 33건의 공공 데이터와 총 88건의 주요 문서 목록을 공개하고, 해당 정보와 관련된 담당 부서와 연락처를 명시하고 있다. 사용자들이 국립환경과학원의 조직도를 직접 검색하여 부서별로 직원들의 연락처를 알 수 있게 함으로써, 담당자와의 신

속한 연결이 이루어질 수 있도록 돕고 있다. 하지만 미세먼지 원자료와 관련된 내용은 찾아보기 어려운 실정이다.

에어코리아의 경우 자주 하는 질문 게시판을 통해 측정 자료와 관련하여 사용자들이 가질 수 있는 궁금증을 일부 다루고 있다. 측정 자료나 측정소 정보와 관련한 질문, OpenAPI를 사용하는 방법 등을 다루고 있는 것이다.

〈표 5〉 정보공개포털, 에어코리아, 에어데이터 비교 내용 요약 (2020년 2월 기준)

	정보공개포털	에어코리아	Air Data
데이터 종류 (〈표 3〉 참고)	게시되어 있는 자료 없음 (청구자가 자료를 지정해서 청구 후 수령 가능, 어떤 자료를 청구할 수 있는지에 대한 정보는 게시 X)	시간별 측정 자료 PM <sub>10</sub> : 2001~2019년 PM <sub>2.5</sub> : 2015~2019년	일별 측정 자료 PM <sub>10</sub> : 1988~2019년 PM <sub>2.5</sub> : 1997~2019년
데이터 품질	메타데이터 (〈표 4〉 참고) -측정 일시 -측정소 정보(측정소명, 측정소코드, 주소)		-측정 일시, 최종수정일 -측정소 정보(측정소 코드, 경도 및 위도) -그 외 에어데이터 웹사이트 이트에서 제공하는 정보: 측정소 및 측정 기구에 대한 설명, 데이터와 포 맷에 대한 설명(데이터 파일이 포함하고 있는 변수들의 이름과 의미를 정리하여 제공)
	갱신 주기	자료는 갱신되고 있으나 갱신 주기에 대한 정보는 X	1년에 두 번
서비스 품질	시·공간 스케일	행정구역 및 연도를 지정 하여 청구	측정망 종류, 지역, 오염물질, 측정소, 조회 기간 등 을 설정하여 다운로드
	비용 및 시간	전자 파일로 받는 경우 무료, 10일 이내에 자료를 받을 수 있음	무료로 바로 다운로드 가능
	사용자 지원 시스템	-자주하는 질문: 수수료 납부 관련 사항 소개 -담당 부서 및 직원 연락처 (국립환경과학원 웹사이트)	-자주하는 질문: 측정소 및 측정 자료, Open API 사용법 등 소개 -관련 링크

## VI. 결론 및 제언

본 논문은 미세먼지 문제에 대한 국민적 관심이 증가하고 있는 가운데 공공 데이터베이스를 이용한 연구가 점점 더 많이 이루어지고 있는 한국의 맥락에서, 연구자의 시점으로 미세먼지 공공데이터베이스를 비교·분석하였다. 공공 데이터베이스를 평가한 기존 연구들이 주로 일반 대중의 입장이나 산업체의 수요를 중심으로 하여 데이터베이스 시스템을 분석하는 것에 머물렀다면, 본 연구는 미세먼지 현황의 파악과 정책적 기여를 위해 노력하고 있는 연구자들의 입장에서 직접 데이터를 다운로드 하여 비교 분석을 수행함으로써 한국 데이터베이스의 발전 방향을 모색하였다.

보다 구체적으로 본 논문은 한국과 미국의 미세먼지 데이터베이스를 데이터 품질 기준과 서비스 품질 기준으로 나누어 살펴보았다. 그 결과 한국과 미국의 미세먼지 데이터베이스는 서비스 품질에서는 유사한 수준을 보인 반면, 데이터 품질, 특히 메타데이터의 측면에서 커다란 차이를 보인다는 점을 확인할 수 있었다. 에어데이터가 메타데이터를 상세하게 제공하고 있는 것과 달리, 에어코리아나 정보공개포털을 통해 획득한 자료에는 측정소에 대한 정보를 제외하고는 메타데이터가 포함되어 있지 않았다. 메타데이터는 측정 기구, 방법, 측정 시 특이사항, 이상값 발생 여부 등 측정값이 얻어진 배경에 대한 설명을 포함한다. 이러한 정보들은 미세먼지의 경향이나 건강 영향에 대한 연구 결과에 영향을 미칠 수 있어 연구를 설계할 때 중요하게 고려해야 하는 요소이다. 따라서 이미 공개되어 있는 기존 자료들부터 메타데이터를 보강함으로써, 한국 미세먼지 자료에 대한 이해를 돕고 나아가 활용도를 제고하고자 하는 노력이 필요하다.

이에 더해 미세먼지 데이터베이스의 활용도를 더욱 높이기 위해서는 정부에서 보유하고 있는 미세먼지 자료에 어떤 것이 있는지 자세한 목록을 작성하여 연구자 및 대중과 공유할 필요가 있다. 에어코리아에서 다운로드 할 수 있는 자료 외에도, 집중측정망이나 성분측정망 등에서 측정된 여러 원시 자료들은 정보공개포털에서 공개 청구를 통해 얻을 수 있다.

하지만 어떤 미세먼지 자료에 대해서 공개 청구가 가능한 것인지는 국립환경과학원 등 관련 기관에 따로 문의를 해야 하는 상황이다. 따라서 어떤 자료를 보유하고 있는지에 대한 정보를 제공하는 것이 필요하다. 나아가 집중측정망이나 성분측정망에서 측정된 원시 자료들 역시 사용자가 직접 다운로드 할 수 있도록 공개하는 방향을 고려해 볼 필요가 있다. 이러한 자료들은 측정과 관리 등에서 더 많은 작업을 요구하기 때문에 관련된 인력을 충원하거나, 자료를 먼저 공개한 뒤 연구자들의 피드백을 통해 자료의 품질을 개선하는 방법을 고려해볼 수 있다.

한국은 20여 년 전 전국적인 미세먼지 측정망을 구축한 이후, 상당히 빠른 시간 안에 모든 연구자와 일반인에게 주요 측정망의 측정 자료를 공개하는 공공 데이터베이스를 성공적으로 구축하였다. 이렇게 구축된 데이터베이스는 미세먼지 측정 자료를 이용한 다양한 분야의 활발한 연구에 기여하였고, 이러한 연구 결과들은 미세먼지와 관련된 정책의 근거로 사용되고 있다. 본 연구에서 제안하고 있듯이 측정 자료에 대한 메타데이터를 제공하고 일반대기오염측정망 외 기타 측정망의 자료도 제공하는 등의 부가적인 노력이 이루어진다면, 한국의 미세먼지 공공 데이터베이스는 미세먼지 현황을 보다 정확하게 파악하고, 나아가 미세먼지의 건강 악영향을 감소시킬 수 있는 연구와 정책을 활성화하는 데에 더욱 크게 기여할 수 있을 것으로 기대된다.

## ■ 참고문헌 ■

- 국립환경과학원, 2019, 『2018 대기환경연보』, (NIER-GP : 2019-053), 인천: 국립환경과학원.
- 김구, 2017, “공공데이터 개방정책에 관한 평가적 고찰: 공공데이터포털을 중심으로,” 『국가정책연구』, 31(2), pp.57-82, DOI: 10.17327/ippa.2017.31.2.003.
- 김석영, 2002, “과학기술분야 웹 정보원 평가 및 비교 연구,” 『한국도서관·정보학회지』, 33(3), pp.133-152.
- 김용표, 2017, “초미세먼지 문제 해결을 위한 연구 및 정책 방향,” 『한국대기환경학회지』,

- 33(3), pp.191-204, DOI: 10.5572/KOSAE.2017.33.3.191.
- 김지혜·조상우·이경희·조완섭, 2016, “공공데이터 개방 평가지표 개발을 통한 현황분석 및 가시화,” 『한국빅데이터학회지』, 1(1), pp.53-60.
- 김태엽, 2018, 『공공데이터 개방 정책의 현황 및 향후 과제』, (이슈와 논점: 제1455호), 서울: 국회입법조사처.
- 박순애·신현재, 2017, “한국의 초미세먼지(PM<sub>2.5</sub>)의 영향요인 분석,” 『환경정책』, 25(1), pp.227-248, DOI: 10.15301/jepa.2017.25.1.227
- 백정희·송재민, 2019, “지역 차원의 미세먼지 관리를 위한 권역 설정 및 특성에 관한 연구,” 『환경정책』, 27(3), pp.75-104, DOI: 10.15301/jepa.2019.27.3.75.
- 변대호, 2005, “대한민국 전자정부 웹사이트의 유용성 평가,” 『Information Systems Review』, 7(1), pp.1-20.
- 윤상오·현지우, 2019, “공공데이터 개방정책의 실태분석 및 개선방안에 관한 연구,” 『한국공공관리학보』, 33(1), pp.219-247, DOI: 10.24210/kapm.2019.33.1.010.
- 이건직, 2008, “전문 공공포털사이트의 유용성 평가요인 및 제고 방안 연구: 국가복지정보포털사이트를 중심으로,” 『보건사회연구』, 28(1), pp.27-50, DOI: 10.15709/hswr.2008.28.1.27.
- 이수현·김옥진·김선영·김호, 2018, “전국 시군구 PM<sub>2.5</sub> 연평균 농도 추정 방법 제안: 전국규모 국가대기오염측정망 측정 자료가 없는 2001~2014 년을 중심으로,” 『한국대기환경학회지』, 34(6), pp.806-821, DOI: 10.5572/KOSAE.2018.34.6.806
- 이용봉·조현양·류범중·최재황, 2001, “과학기술분야 데이터베이스의 품질향상을 위한 품질평가 연구,” 『한국문헌정보학회지』, 35(2), pp.109-132.
- 이제환, 1997, “과학기술분야 서지 DB의 품질관리 및 평가 방안: KORDIC 의 KRISTAL DB 를 중심으로,” 『한국문헌정보학회지』, 31(3), pp.109-134.
- 전슬비·강순애, 2016, “중앙행정기관의 정보공개제도 운영 현황 및 개선 방안 연구,” 『한국기록관리학회지』, 16(1), pp.61-88, DOI: 10.14404/JKSARM.2016.16.1.061.
- 정진임·김유승, 2014, “정부 3.0 시대, 정보공개시스템의 개선 과제,” 『기록학연구』, (39), pp.45-72.
- 한국데이터베이스진흥센터, 1996, 『최종(연도별)연구결과보고서: 데이터베이스 구축 및 유통진흥사업에 관한 연구』, 서울: 한국데이터베이스진흥센터.
- \_\_\_\_\_, 2000, 『데이터베이스 품질평가 항목』, (정보통신연구진흥원 출연사업 『데이터베이스 표준화 연구』의 세부 과제 연구보고서), 서울: 한국데이터베이스진흥센터.
- 한국정보화진흥원, 2017, 『공공데이터 기반 산업 생태계 조성을 위한 법제 개선 연구』, 대구: 한국정보화진흥원.
- 한혁·정창훈·금현섭·김용표, 2017, “미세먼지(PM<sub>10</sub>) 저감 정책의 비판적 검토,” 『환경정책』, 25(1), pp.49-79, DOI: 10.15301/jepa.2017.25.1.49.

- 홍강훈, 2012, “독일 정보자유법을 통해 본 한국 정보공개법의 개선방안,” 『헌법학연구』, 18(1), pp.267-302.
- 홍현진, 2005, “웹 기반 데이터베이스의 품질평가 기준 개발에 관한 연구,” 『한국문헌정보학회지』, 39(2), pp.211-235.
- 환경부·국립환경과학원, 2018, 『대기오염측정망 설치·운영 지침』, 세종: 환경부.
- Brook, R. D., S. Rajagopalan, C. A. Pope III, J. R. Brook, A. Bhatnagar, and A. V. Diez-Roux et al., 2010, “Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association,” *Circulation*, 121(21), pp.2331-2378, DOI: 10.1161/CIR.0b013e3181d8e1.
- Committee on Environment, Natural Resources, and Sustainability (CENRS) of the National Science and Technology Council, 2013, *Air quality observation systems in the United States*, Washington, D.C.: CENRS.
- Coglianesi, C., 2009, “The transparency president? The Obama administration and open government,” *Governance*, 22(4), pp.529-544, DOI: 10.2139/ssrn.1433815.
- Demerjian, K. L., 2000, “A review of national monitoring networks in North America,” *Atmospheric Environment*, 34(12-14), pp.1861-1884, DOI: 10.1016/S1352-2310(99)00452-5.
- \_\_\_\_\_, 2007, U.S. federal and state monitoring networks, In G. Visconti, P. Di Carlo, W. Brune, M. Schoeberl, and A. Wahner. (Eds.), *Observing systems for atmospheric composition - Satellite, aircraft, sensor web and ground-Based observational methods and strategies*, (pp.159-168), New York: Springer Science.
- Hong, Y. C., J. T. Lee, H. Kim, and H. J. Kwon, 2002, “Air pollution: A new risk factor in ischemic stroke mortality,” *Stroke*, 33(9), pp.2165-2169, DOI: 10.1161/01.STR.0000026865.52610.5B.
- Janssen, M., Y. Charalabidis, and A. Zuiderwijk, 2012, “Benefits, adoption barriers and myths of open data and open government,” *Information Systems Management*, 29(4), pp.258-268, DOI: 10.1080/10580530.2012.716740.
- Kim, H., H. Kim, and J. T. Lee, 2019, “Spatial variation in lag structure in the short-term effects of air pollution on mortality in seven major South Korean cities, 2006-2013,” *Environment International*, 125, pp.595-605, DOI: 10.1016/j.envint.2018.09.004.
- Kim, K. N., Y. H. Lim, H. J. Bae, M. Kim, K. Jung, and Y. C. Hong, 2016, “Long-term fine particulate matter exposure and major depressive disorder in a

- community-based urban cohort,” *Environmental Health Perspectives*, 124(10), pp.1547-1553, DOI: 10.1289/EHP192.
- Kim, O. J., S. Y. Kim, and H. Kim, 2017, “Association between long-term exposure to particulate matter air pollution and mortality in a South Korean national cohort: Comparison across different exposure assessment approaches,” *International Journal of Environmental Research and Public Health*, 14(10), p.1103, DOI: 10.3390/ijerph14101103.
- Kim S. Y. and I. Song, 2017, “National-scale exposure prediction for long-term concentrations of particulate matter and nitrogen dioxide in South Korea,” *Environmental Pollution*, 226, pp.21-29, DOI: 10.1016/j.envpol.2017.03.056.
- Kim, S. Y., M. Bechle, S. Hankey, L. Sheppard, A. A. Szpiro, and J. D. Marshall, 2020, “Concentrations of criteria pollutants in the contiguous U.S., 1979-2015: Role of prediction model parsimony in integrated empirical geographic regression,” *PLoS ONE*, 15(2), e0228535, DOI: 10.1371/journal.pone.0228535.
- Kyle, A. D., J. R. Balmes, P. A. Buffler, and P. R. Lee, 2006, “Integrating research, surveillance, and practice in environmental public health tracking,” *Environmental health perspectives*, 114(7), pp.980-984, DOI: 10.1289/ehp.8735.
- Lathrop, D. and L. Ruma, 2010, *Open government: Collaboration, transparency, and participation in practice*, Sebastopol, CA: O'Reilly Media, Inc..
- Lee, J. T., H. Kim, Y. C. Hong, H. J. Kwon, J. Schwartz, and D. C. Christiani, 2000, “Air pollution and daily mortality in seven major cities of Korea, 1991-1997,” *Environmental Research*, 84(3), pp.247-254, DOI: 10.1006/enrs.2000.4096.
- Levin, N. and S. Leonelli, 2017, “How does one “open” science? Questions of value in biological research,” *Science, Technology, & Human Values*, 42(2), pp.280-305, DOI: 10.1177/0162243916672071.
- Litt, J., N. Tran, K. C. Malecki, R. Neff, B. Resnick, and T. Burke, 2004, “Identifying priority health conditions, environmental data, and infrastructure needs: A synopsis of the pew environmental health tracking project,” *Environmental Health Perspectives*, 112(14), pp.1414-1418, DOI: 10.1289/ehp.7147.
- OECD, 2017, *Government at a glance 2017*, Paris: OECD Publishing.
- Peled, A., 2011, “When transparency and collaboration collide: The USA open data program,” *Journal of the American Society for Information Science and Technology*, 62(11), pp.2085-2094, DOI: 10.1002/asi.21622.

- Pope III, C. A. and D. W. Dockery, 2006, "Health effects of fine particulate air pollution: Lines that connect," *Journal of the Air & Waste Management Association*, 56(6), pp.709-742, DOI: 10.1080/10473289.2006.10464485.
- Strosnider, H., Y. Zhou, L. Balluz, and J. Qualters, 2014, "Engaging academia to advance the science and practice of environmental public health tracking," *Environmental Research*, 134, pp.474-481, DOI: 10.1016/j.envres.2014.04.039.
- Thacker, S. B., D. F. Stroup, R. G. Parrish, and H. A. Anderson, 1996, "Surveillance in environmental public health: Issues, systems, and sources," *American Journal of Public Health*, 86(5), pp.633-638, DOI: 10.2105/ajph.86.5.633.
- U.S. EPA., 2019, *Integrated science assessment for particulate matter*, (Report No. EPA/600/R-19/188), Washington, DC, USA: U.S. Environmental Protection Agency.
- 에어코리아, <https://www.airkorea.or.kr>.
- 정보공개포털, <https://www.open.go.kr/>.
- 한국환경공단, 2010, "미국 대기측정망 운영현황," <https://www.keco.or.kr/kr/business/climate/communityid/189/view.do?idx=124>, [2020.2.28]
- Airnow, <http://airnow.gov>.
- EPA Air data, <http://www.epa.gov/outdoor-air-quality-data>.
- World Health Organization (WHO), 2004, "Review of methods for monitoring of PM<sub>2.5</sub> and PM<sub>10</sub>," [http://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0020/143156/E85769.pdf?ua=1](http://www.euro.who.int/__data/assets/pdf_file/0020/143156/E85769.pdf?ua=1), [2019.2.28]

---

**김주희:** 서울대학교 협동과정 과학사 및 과학철학에서 과학기술학(STS) 전공으로 석사 학위를 취득했고, 현재 동 대학원 박사과정에 재학 중이다. 소송, 규제, 정책 평가 등의 영역에서 대기오염을 측정된 데이터와 모델링한 데이터가 어떤 긴장 관계를 형성하는지를 살펴보고, 그것이 대기오염에 대한 법적·정치적 책임의 문제와 어떻게 결부되어 있는지를 분석한 내용으로 박사 논문을 작성 중이다(joohuikim@snu.ac.kr).

**김선영:** 서울대학교 보건대학원에서 보건통계 전공으로 보건학 박사학위를 취득했고, 현재 국립암센터 국제암대학원대학교 암관리학과에 재직 중이다. 주요 관심 분야는 대기오염 모델링, 대기오염의 건강영향 평가, 대기오염역학 연구방법론이고, 박사과정 때부터 15년 이상 지속적으로 관심 분야에서 연구하고 있다(sykim@ncc.re.kr).