ISSN 1229-1889(Print)
ISSN 2287-9005(Online)
J Korean Soc Qual Manag Vol. 50, No.4: 811-830, December 2022
https://dx.doi.org/10.7469/JKSQM.2022.50.4.811

철도 산업의 공기 질 데이터베이스 연합형 통합을 위한 지능형 데이터 거버넌스

김민정* · 원종운* · 박상찬** · 박가영*†

* 한국철도기술연구원
** 한국뉴욕주립대학교, 기술경영학과

Intelligent Data Governance for the Federated Integration of Air Quality Databases in the Railway Industry

Minjeong Kim* · Jong-Un Won* · Sangchan Park** · Gayoung Park*†

* Korea Railway Research Institute
** Department of Technology and Society, The State University of New York Korea

ABSTRACT

Purpose: In this paper, we will discuss 1) prioritizing databases to be integrated; 2) which data elements should be emphasized in federated database integration; and 3) the degree of efficiency in the integration. This paper aims to lay the groundwork for building data governance by presenting guidelines for database integration using metrics to identify and evaluate the capabilities of the UK's air quality databases.

Methods: This paper intends to perform relative efficiency analysis using Data Envelope Analysis among the multi-criteria decision-making methods. In federated database integration, it is important to identify databases with high integration efficiency when prioritizing databases to be integrated.

Results: The outcome of this paper aims not to present performance indicators for the implementation and evaluation of data governance, but rather to discuss what criteria should be used when performing 'federated integration'. Using Data Envelope Analysis in the process of implementing intelligent data governance, authors will establish and present practical strategies to discover databases with high integration efficiency.

Conclusion: Through this study, it was possible to establish internal guidelines from an integrated point of view of data governance. The flexiblity of the federated database integration under the practice of the data governance, makes it possible to integrate databases quickly, easily, and effectively. By utilizing the guidelines presented in this study, authors anticipate that the process of integrating multiple databases, including the air quality databases, will evolve into the intelligent data governance based on the federated database integration

● Received 22 November 2022, 1st revised 29 November 2022, accepted 2 December 2022
† Corresponding Author(gayoung.park@sunykorea.ac.kr)

* 본 연구는 한국철도기술연구원 주요사업 "지능형 철도·교통 기술개발을 위한 인공지능 지원 플랫폼 개발"의 지원을 받아 수행된 연구입니다.

when establishing the data governance practice in the railway industry.

Key Words: intelligent data governance, Data Envelopment Analysis, railway industry, air quality database, federated database integration

1. 서 론

4차 산업혁명은 빅데이터를 중심으로 성장한다. 인공지능 알고리즘, 빅데이터와 같은 첨단 기술이 연결되어 통합, 적응, 최적화, 서비스 지향, 상호 운용 가능한 프로세스를 통해 다양한 구성요소가 상호 혁신하는 형태로, 4차 산업혁명의 중심에는 디지털 데이터를 통한 다양한 파트너 사이의 소통을 기반으로 한 새로운 가치 창조가 존재한다(Slusarczyk, 2018).

급증하는 빅데이터의 효과적인 활용을 지원하고자 한국의 경우에는 공공분야의 공공빅데이터 발굴 및 활성화 방안에 대해 중점을 두고 있으나(Choi and Yoon, 2018) 급격한 공공 빅데이터 관련 인프라 및 전문 인력 부족으로 인해 수요를 충족하지 못하고 있다(Cheon and Kim, 2017). 데이터 생애 주기(Data Life Cycle)를 크게 데이터 수집 및 생성, 데이터 저장, 데이터 가공 및 활용, 데이터 배포 등의 네 단계로 놓고 보았을 때, 현재 우리의 수준은 빅데이터를 수집 및 생성, 저장하는 단계를 활발하게 진행하고 있으며 가공 단계와 배포 단계로 넘어가고 있는 수준이다. 그리고 공공 빅데이터 개방 요구의 급격한 증가에 공공기관들은 빅데이터 관련 인프라 또는 전문 인력의 부족으로 인해 그 수요를 충족하지 못하고 있는 실정이다

빅데이터 거버넌스를 위해서는 빅데이터 발굴 등의 데이터의 양적인 확대 뿐 아니라 데이터를 가공하고 배포하여 활용도를 높이기 위한 데이터베이스 간 통합의 중요성이 강조되어야 한다. 이를 위해 데이터의 수직적 수평적 통합을 위한 노력에 대해 논의가 필요하다. 이것이 먼저 적용될 수 있으며 적용되어야 하는 부분은 공공분야이다. 공공데이터의 경우에는 다방면의 데이터를 '수집'하는 단계에서도 어려움을 겪고 있는데, 데이터의 의미 있는 활용을 위해서 동 시간에 다양한 시·공간을 연결하고, 여러 출처 및 분야에서의 수집되는 자료를 통합하는 방향성에 대한 논의가 필요하다.

이러한 맥락에서 본 논문에서 중점을 두고 논의하고자 하는 것은 데이터베이스 통합 차원에서의 데이터 거버넌스이다. 데이터 통합을 논하는 데 있어서 데이터 거버넌스 대상을 정의 하는 것이 중요하다. 데이터 거버넌스의 대상은 양적 뿐 아니라 질적으로도 확장되고 있다. 디지털 전환 시대에는 소비자 만족, 소비자 경험등의 새로운 품질에 대한 기대가 등장하고 있기 때문이다(Park et al., 2021). 공공 데이터 거버넌스 영역 중 하나인 철도 산업의 경우, 철도 산업은 사회-기술적 시스템으로 철도 서비스 관리의 다양한 주체와, 다양한 출처의 데이터에서 오는 정보에 기반하고 있기 때문에 복잡할 수밖에 없다(Kans et al., 2016; Rotter et al., 2016). 기존에는 데이터를 수집해야 할 대상이 철도 시설 및 차량 운행에서 산출되는 데이터, 즉 물적 자원(physical asset)이 주 대상이었다. 현재는 이러한 물적 자원에 대한 데이터 뿐 아니라 사이버 자원(cyber asset)에 대한 데이터를 고려해야 하며, 철도를 운행하는 종사자와 철도를 이용하는 사용자 등에 대한 인적 데이터를 포함해야 할 필요성이 대두되고 있다.

포괄성을 위한 대상의 양적인 확대 뿐 아니라 질적인 측면에 대한 논의 역시 주목해야 한다. 승객 서비스 질 향상을 예로 들자면, 철도 시설이 첨단화 고속화됨으로 인해 밀폐되어 있는 차량 시설 내의 공기 질에 대한 이슈도 부각되고 있다. 철도 기술의 고도화와 더불어 모니터링 및 개선되어야 하는 분야로 철도 실내 및 철도 역사, 철도 터널 등에서의 공기 질 문제는 꾸준히 대두되어 왔다. 그럼에도 불구하고 한국의 철도를 포함한 대중교통의 경우 실내 공기 측정에 기준 권고치는 초미세먼지(PM-2.5)와 이산화탄소만 제시하고 측정하도록 하고 있으며 통합적이고 전략

적 수준의 공기 질 데이터베이스의 구축에 대한 노력은 미비하다.

철도 데이터 거버넌스를 논할 때 대상 데이터 측면에서는 물적 자원을 넘어 사이버 자원과 인적 자원으로, 시설, 차량, 운행 및 안전에 관한 데이터에서 질적인 요소를 추가하는 방향으로 나아가고 있다고 할 수 있다. 뿐만 아니라, 해당되는 데이터를 지역적으로 잘 관리할 수 있게 하는 모든 요소들이 데이터 거버넌스의 대상이 될 수 있다. 이러한 데이터 거버넌스 대상의 확장 및 포괄성은 글로벌 이라는 도전적인 환경 안에서 더 다양하고 다층적인 정성적, 포괄적 측면에서 유연한 측정지표를 요구한다. 그럼에 불구하고, 철도의 경우 시설 상황, 운행에 대한 데이터에서 안전과 인적 요소에서의 퀄리티 등을 추가적인 논의에 따라 측정지표가 변화할 수밖에 없다. 글로벌 환경에서 데이터가 수집 되는 경우 어느 영역, 지역 등에서는 데이터가 수집 되지 않은 경우도 발생하게 된다.

물적 요소를 넘어서 사이버 요소와 인적 요소, 데이터의 양적인 확충에서 질적인 요소 고려 등을 포함한 포괄적인 측정지표를 논의하기 위해서 우리는 본 논문에서 지능형 데이터 거버넌스(Intelligent Data Governance, 이하 IDG) 도입의 중요성을 강조한다.

지능형 데이터 거버넌스란 데이터 거버넌스가 자동화를 통해 확장되어 인적 입력을 극대화함으로써 결과를 가속화됨을 의미한다. 4차산업혁명에서는 인공지능과 머신러닝이 데이터 관리에 대한 생각을 전환시키기 시작했다(Informatica). 4차 산업혁명 시대에 품질 4.0의 핵심요소로서 데이터 획득과 분석기술, 연결과 통합이 강조된다(Soh et al., 2021). 아울러 빅데이터 활용을 통한 맞춤형 대민 서비스가 등장하면서 행정 데이터 간의 공동 활용을 연계할 수 있는 기준데이터의 중요성 제기 된다(Choi et al., 2015). 기준 데이터를 기반으로 잘 연계 되었을 때에는 자료가 손쉽게 검색되고, 횡적으로 같은 시각 타임라인에 관련한 자료를 통해 정보가 생성되고, 시간의 추세에 대한 자료가 축적됨으로 포괄적인 정보 활용이 가능하게 된다. 이에 따라 데이터 거버넌스를 통합이라는 관점에서 논의하고, 이를 위한 내부적인 가이드라인을 제시하고자 하는 것이 본 논문의 목적이다. 데이터 거버넌스 구축을 위한 데이터베이스 통합이라는 관점에서 연합형 접근(federated approach)을 통한 유연한 접근 방법을 취하기 때문에 이니셔티브가 빠르고, 통합 과정에 빠르게 적용해 볼 수 있게 된다.

다양한 출처의 많은 데이터가 횡적(cross-sectional)이며 종적(longitudinal)인 연결이 가능하며, 포괄성도 우수하며 데이터 타임라인에 대한 일관성을 더하기 위해서 가장 좋은 형태는 하나의 중앙 집중형(centralized approach) 데이터 거버넌스에 의한 통합 시도 일 것이다. 하지만 중앙 집중형 시도가 실패하기 쉬운 이유는. 통합 당시에 특정한 목적에 따라 부합되게 생성되다 보니, 이후 확장되는 연계 부분을 생각을 안 하고 만들 수밖에 없는 태생적인 문제가 있기 때문이다. 그렇기 때문에 본 논문에서는 연합형 접근 방식(federated approach)을 취하는 것을 고려해야 한다고 주장한다. 연합형이기 위해서는 일일이 수동적으로 통합을 수행하는 것이 어렵기 때문에 지능적인(intelligent) 접근을 취할 수밖에 없다. 지능적이기 위해서는 통합의 대상이 정해 져야 하고, 통합 대상 순서 및 기준점이 제시 되어야 한다. 그러나 구체적으로 연합형 접근방법을 취하는 지능형 데이터 거버넌스에 대한 실증적인 사례를 찾기는 어려움이 많다.

이 논문에서는 산재되어 있는 기존의 데이터베이스들 중에서 1) 통합 대상 데이터베이스의 우선순위를 정하고, 2) 데이터베이스 연합형 통합에 있어서 어떤 데이터 요소를 강조해야 하는 것인지 3) 통합 효율성은 어느 정도인지 논의하고자 한다. 불행히도, 한국의 경우에는 데이터베이스 간의 통합 사례를 찾아보기 힘든 상황이기 때문에 국내 사례 분석을 통한 데이터와의 연계 및 통합을 위한 가이드라인을 제시하기 어렵다. 이러한 맥락에서 본 논문에서는 영국의 사례를 기준으로 논의하고자 한다. 영국의 경우는 공기 질과 관련한 데이터베이스를 통합하려는 시도가 있어 왔다. 공기 질 향상을 위한 국제적인 조직인 INSPIRE의 제시안에 의거한 중앙 집중형 통합이 될 수 있도록 시도했으나 하나의 기준에 내부적으로 통합을 이룰 수 있는 구체적인 가이드라인을 제시 하지 못했고, 데이터베이스의 존재 여부와 해당 데이터베이스가 어떤 역량을 가지고 있는지 파악 되어있는 상황이다(Monteith et al., 2010). 그럼에도

불구하고 각 데이터베이스의 역량을 파악 및 평가하기 위한 지표 값이 제시되었기 때문에 데이터 통합을 위한 가이드라인을 제공할 수 있다. 향후 이 방법을 따르게 되면, 공기 질 관련 데이터베이스와 기상관련 데이터 통합에 있어서 로드맵 또는 가이드라인을 역시 제시할 수 있을 것으로 기대한다.

지능형 데이터 거버넌스(Intelligent Data Governance, 이하IDG)에 따라 데이터베이스의 연합형 통합 자동화를 수행하려면, 통합 우선순위를 결정하기 위하여 데이터베이스 성과평가지표가 투입 요소와 산출 요소로 구별할 필요가 있다. 영국의 공기 질 관련 데이터베이스를 평가한 성과평가지표를 분석해 본 결과 사용 용이성 및 데이터 가공과 관련한 과정적인 지표를 투입요소로 보고, 데이터가 얼마나 포괄적이고 일관적 인지 등에 대한 데이터 자체의 수준을 나타내는 결과적인 지표를 산출요소로 나누어 볼 수 있었다. 특히 산출 요소가 하나가 아니라 여러 개라는 측면을 고려한다면 다 기준 의사결정론(Multiple Criteria Decision Making) 방법을 고려해 볼 수 있다.

본 논문은 다 기준 의사결정론 방법론 중에서 자료포락분석(Data Envelope Analysis)을 이용하여 상대적 효율성(efficiency)분석을 수행하려 한다(Charnes et al., 1987). 데이터베이스의 연합형 통합에 있어서 통합 우선순위 결정시 통합 효율성이 높은 데이터베이스를 파악하는 것이 중요하다. 본 연구에서는 우선, 영국의 전략적 공기 질 측정 데이터베이스 성과평가 지표 10개를 PRISMA(Matthew J.P. et al., 2021) 방법을 통해 선별적으로 추려냈다.

본 논문의 목적은 데이터 거버넌스의 구현과 평가를 위한 성과지표를 제시하는 것에 있는 것이 아니라, '연합형 통합' 수행 시 어떤 기준으로 통합되어야 하는지에 대한 논의이다. 국내에 구축되어 있는 공기 질 데이터베이스에 대한 평가지표가 전무하기 때문에, 전략적 수준의 데이터베이스가 존재하고 해당 데이터베이스 통합을 위해 평가 지표가 마련되어 있는 해외사례 분석을 통해서 함의를 발견하고자 한다. 아울러, 자료포락분석 (DEA)을 사용하여 지능형 데이터 거버넌스 구현을 위한 벤치마킹 대상인 효율성이 높은 데이터베이스를 발견하기 위한 실천 전략을 수립하여 제시할 것이다.

본 연구는 다음과 같이 구성된다. 제 2장에서는 데이터베이스 통합에 관한 사례를 논의하고 이를 통해 데이터베이스 통합을 위한 핵심사항을 고찰한다. 제 3장에서는 본 연구에서 사용된 연구 방법인 자료포락분석과 분석에 사용된 데이터베이스에 대해 기술한다. 제 4장에서는 연구결과로써, 영국 철도 공기 질 데이터베이스 통합을 위한 성과평가 자료를 자료포락분석으로 분석하여 연합형 통합의 기준점을 제시한다. 마지막으로 5장 결론 및 함의에서는 연구 결과가 한국의 철도분야 데이터 거버넌스 구현에 어떠한 함의가 있는지, 향후 데이터 거버넌스는 어떤 방향성을 띄어야 하는지 논의한다.

2. 문헌고찰

데이터베이스 통합 구축 및 관리에서 문제점으로 제기되는 것은, 모든 데이터베이스를 일률적인 방식으로 처리하려는 시도, 로드맵, 정책, 표준 목표 등이 모호하게 제시되는 점, 데이터의 소유 및 사용권에 대한 불명확한 설정, 어떤 것이 우선적으로 문서화, 데이터화 되어야 하는지 우선순위를 정하지 않는 점 등이 있다(Alen and Cervo, 2015). 본 장에서는 교통 분야와 철도 분야의 데이터베이스 통합에 관한 해외사례를 논의하며 데이터 통합에 관한 핵심사항을 고찰하여 보고자 한다.

2.1. 교통 분야 데이터베이스 통합에 관한 미국 사례

미국은 교통 통합 데이터 구축에 있어서 연합형 접근(Federated approach)을 취하고 있는 대표적인 사례라고 볼

수 있다. 미국의 경우에는 교통 분야에서 연방 정부와 주정부로 구성되는 연방 정치 사회 시스템의 특성상 중앙집권적인 접근 혹은 하향식 접근(Top-down)을 취하기가 힘들며, 데이터베이스 통합하기 위해 각각의 주정부 교통국과 외부 파트너 기간 관의 신중한 계획 및 관리 및 조정을 수행하고 있다.

미국의 경우 여러 주정부 교통국(Department of Transportation, 이하 DOT) 중 모범 사례를 보유하는 DOT를 중심으로 교통 데이터 거버넌스 구축을 시도 하고 있다(NOCoE, 2021; Transportation Research Board of the National Academies, 2015). 참여 주정부 교통국은 다음과 같다:

- 알래스카 교통국(DOT) 공공시설국 PF),
- 아이다호 교통국(ITD),
- 아이오와 DOT,
- 루이지애나 교통개발국(DOTD),
- 메릴랜드 교통국(DOT) 주 고속도로 관리국(SHA),
- 미시간 교통국(DOT),
- 몬태나 교통국(DOT),
- 오하이오 교통국(DOT),
- 로드아일랜드 교통국(DOT),
- 워싱턴 교통국(DOT).

교통 데이터 거버넌스와 관련하여 각 주정부 교통국은 4가지 주요 주제에 초점을 맞췄다:

1. 데이터 거버넌스를 위한 비즈니스 사례,
2. 데이터 거버넌스의 필수 요소,
3. 교통국(DOT)에서의 데이터 거버넌스 운영,
4. 데이터 거버넌스를 사용하여 데이터 공유 및 통합 향상.

교통 데이터 거버넌스와 관련하여 각 주정부 교통국이 파악한 6가지 연구 주제는 다음과 같다:

1. 성공적인 데이터 거버넌스를 위한 지표 결정,
2. 교통 데이터 및 정보를 자산으로 관리하는 경영진의 시각에 대한 상호 의견교환,
3. 교통 기관의 비즈니스 의사 결정을 촉진하기 위한 데이터 통합의 필수 요소,
4. 중앙정보기술의 맥락에서의 데이터 거버넌스,
5. 충돌 데이터 수명 주기 실증 사례,
6. 위치 정확도가 안전 데이터 분석에 미치는 영향: 데이터 거버넌스를 위한 사례.

참여 주정부 교통국은 다음과 같은 공식적인 데이터 거버넌스에 대한 노력을 기울이고 있다:

- 장거리 운송 계획(LRTP),
- 전략 고속도로 안전 계획(SHSP),
- 고속도로 안전 개선 프로그램(HSIP),
- 교통기록조정위원회(TRCC),
- 교통자산관리계획(TAMP),
- 안전 관리 시스템(SMS),

- 충돌 데이터 관리,
- 자산 관리 데이터 관리.

2.2. 철도분야 데이터베이스 통합에 관한 유럽 사례

유럽의 경우는 유럽연합 (European Union,이하 EU)을 중심으로 철도 분야의 경제 발전, 통합 및 지속 가능성을 위한 지속적인 표준화 작업을 진행해 왔다. 철도 데이터 거버넌스와 관련하여 초국가차원의 EU 가이드라인이 존재하고 있으며 이에 맞춰 데이터 거버넌스가 진행되고 있다. 유럽 연합 국가의 경우 철도의 상호 운용의 복잡성을 띄고 있어 상호 통합 운영을 위한 기술 기준 및 유럽 기준(European Norm)이 제정되어 왔다(ITF, 2021). 영국뿐만 아니라 스웨덴도 활발하게 철도 데이터 거버넌스 활동을 전개하여 왔다. 스웨덴의 경우 철도 운영 및 사용과 관련한 포괄적인 데이터 수집 및 관리에 대한 모범적인 사례를 찾아 볼 수 있다(Kans & Ingwald, 2021).

영국은 미국 보다 데이터 거버넌스 구축 및 관리 측면에서 앞서 나가고 있다. 그 중 철도 공기 질 과 관련한 데이터베이스 구축, 관리 및 통합에 있어 선도적이다. 이는 영국이 철도 데이터의 질적인 측면, 즉 철도 환경 및 서비스의 질적인 측면을 고려하고 있는 흔적이라고 볼 수 있다. 반면 공기 질 측정 데이터베이스가 다수 구성이 되어 있음으로 인해 데이터베이스 간 호환이 잘 되지 않는다는 문제점이 있다. 무엇보다 관련 데이터를 수집, 정렬 및 분석하여 활용하는 데 있어 어려움이 있다(Monteith et al., 2010). 이에 영국은 INSPIRE 국제 조직의 제시안을 따라 하향식 (Top-down approach) 통합을 택하여 하나의 기준을 제시 하여 공기 질 과 관련한 데이터 거버넌스를 시도했다. 그러나 하나의 기준에 맞춰 통합하는 과정이 순조롭지 않았고, 통합을 위한 내부 가이드라인을 제시 하지는 못했다. 현재 통합 수준은 산재하는 공기질 데이터베이스를 파악하고, 어떠한 역량을 가지고 있는지를 평가하는 수준이다. 따라서 해당 데이터베이스 통합 시도가 성공적이라고 평가하기 어렵지만, 공기 질 데이터베이스의 역량평가가 이루어졌다는 면에서 영국의 사례는 데이터 거버넌스 구축의 기준점을 제시한다는 측면에서 활용 가치가 높다.

해외 사례를 살펴보았을 때 상호 운영의 복잡성을 띄는 유럽이나 주정부의 자치권이 보장되어 있는 미국의 경우 중앙 집중 형 데이터 거버넌스 접근은 현실적으로 어려움이 많다는 것을 알 수 있다. 데이터 통합을 위한 한 가지 기준이나, 하나의 모델을 적용 하는 것은 현실적으로 어려울 뿐 아니라, 데이터 통합을 위해 필요한 요소들로 제시되는 것들은 다양한 목적과 필요성에 의해 구성된 여러 가지 데이터베이스를 고려해야 하기 때문에 중앙 집중 형 데이터 거버넌스 접근은 지양되어야 한다.

데이터 거버넌스를 위한 평가 연구는 특히나 미비하다. 데이터 거버넌스를 정량적으로 측정할 수 있도록 데이터베이스 성과평가모형 개발이 시급하다. 뿐만 아니라, 성과평가모형의 타당성과 유효성 검증이 필요하며 지속적인 사례 연구가 필요하다(Jang and Kim, 2016)

3. 연구방법

3.1. 활용 데이터

본 논문에서 활용할 데이터는 영국의 공기 질 데이터베이스 통합을 위해 개발된 평가지표로 평가된 주요 데이터베이스의 평가 지수 매트릭스이다. 영국의 경우 공기 질 관련한 데이터베이스 간 호환의 어려움을 겪고 있다. 영국은 전략적 수준의 공기 질 데이터베이스의 통합을 위해서 일대일 전문가 인터뷰 및 워크숍을 통해 평가지표를 개발하

고, 평가된 수치를 이용하여 영국의 공기 질 측정 및 관리와 관련한 주요 데이터베이스를 정성적으로 평가하고 수치화해서 점수표(Scoring matrix)로 제시하였다. 이 평가는 이전 데이터베이스는 준수할 필요가 없었던 새로운 기준을 제시한 것이기 때문에, 평가 점수가 낮다는 것은 데이터베이스가 목적에 적합하지 않게 설계 되었다는 의미가 아니라, 데이터를 통합하기 위해 더 많은 변환이 필요하다는 것을 의미한다(Monteith et al., 2010).

한국의 경우에도 대기 질 및 실내공기 측정에 관한 데이터는 존재하지만, 영국과 같은 수준의 전략적이고 통합적인 데이터베이스 및 통합을 위한 평가지표 체계를 구축하고 있지 않다. 철도 실내 및 철도 관련 환경 공기 데이터 측정 및 시스템에 대한 연구 및 공기질 비교 분석에 관한 연구 활발히 진행 되어 왔다(Lee, 2005; Soh and Yoo, 2008a; 2008b; Ryu, 2018; Lee, 2019; Jin et al., 2022). 그럼에 불구하고 한국의 철도를 포함한 대중교통의 경우 실내 공기 측정에 기준 권고치는 초미세먼지(PM-2.5)와 이산화탄소만 제시하고 측정하도록 하고 있으며 통합적이고 전략적 수준의 공기 질 데이터베이스의 구축에 대한 노력은 미비하다.

<Table1>은 영국의 공기 질 측정 관련 주요 데이터베이스와 설명을 요약해 놓은 것이다. 여섯 가지 주요 데이터베이스로는 NAEI(The National Atmospheric Emissions Inventory(NAEI); PCM(Pollution Climate Mapping); AURN(The Automatic Urban and Rural Network); LAQN(The London Air Quality Network); NPL: Non-automatic networks(managed by NPL); CEH(CEH Deposition monitoring and modeling data)가 있다.

Table 1. Major Air Quality Databases of England

Database	Description
NAEI(The National Atmospheric Emissions Inventory (NAEI)	· Provides comprehensive estimates of annual air emissions of pollutants and hazardous air pollutants in the UK · The purpose of the data is to comply with EU and internal reporting requirements (Greenhouse Gases(GHSGs) Monitoring Mechanism, NEC Directive and LRTAP Convention)
PCM(Pollution Climate Mapping)	· Provides a collection of models to report on he concentrations of particular pollutants in the atmosphere. · Models per pollutant (NOx, NO2, PM10, PM2.5, SO2, CO, benzene, ozone, As, Cd, Ni, Pb and B[a]p) · Provides outputs on a 1x1 km grid of background conditions and around 9,000 representative road side values. · Data compiled for reporting under EU ambient air quality directives and air quality policy developments
AURN(The Automatic Urban and Rural Network)	· Recordes NOx, SO2, CO, O3, PM2.5 and PM10 concentrations at around 130 air quality monitoring stations across the UK · Data is updated hourly and can be accessed at www.airquality.co.uk
LAQN(The London Air Quality Network)	· A group of quality air quality monitoring stations in London, Essex, Kent and Surrey. · The monitoring is owned and funded by local authorities
NPL(Non-automatic networks (managed by NPL)	· Provides an average of he Hydrocarbon, UK Heavy Metals, and UK Black Carbon and Black Smoke networks · Contains data on atmospheric NOx, SO2, CO, O3, PM2.5 and PM10 concentrations measured at automated monitoring sites in Scotland, Wales and Northern Ireland. · Measures data using non-automated methods at daily, weekly, biweekly, etc. intervals across the UK
CEH(CEH Deposition monitoring and modeling data)	· Data measured through Rural Heavy Metals Monitoring · Provides 26 individual atmospheric heavy metals (PM10)concentrations in suburban areas and time information of rainwater concentrations collected from 15 sites.

<Table2>와 <Table 3>은 영국 공기 질 관련 데이터베이스 통합을 위한 평가지표와 평가결과표이다. 영국은 INSPIRE 국제조직이 제시한 하향식 (Top-down approach) 방식을 채택하여 중앙 집중형 기준을 제시하고 있다. 해당 기준에 맞춰 데이터베이스를 통합하기 위해 공기 질과 관련한 여섯 가지 주요 데이터베이스를 평가하였다. 평가 지표로는 데이터 검색 용이성, 데이터 타임라인, 데이터 다운로드 용이성, 데이터 추세, 데이터 포괄성, 데이터 정확성, 데이터 일관성, 메타데이터, 데이터 형식 및 표준화를 사용하였다.

Table 2. Scoring Matrix for Integrating Air Quality Databases of the UK

Criteria	Description
Data Searchability	How easily searchable and understandable the dataset is to the user 0: Data cannot be found, i.e. 5: The dataset can be found after typing in key words related to that dataset
Data Timeline	The dataset contains up-to date data. 0: Data are rarely updated if at all after it has original been published 5: Data are frequently updated, and there is an indication of when the last update took place/when future updates will take place.
Data Downloadability	How easy the data are to download and the usability of the data for analysis 0: The data cannot be downloaded 5: The entire dataset can be downloaded and used for other applications for analysis
Data Historical Trendings	Does the dataset contain record from past years? 0: Contains no previous data 5: Contains over 20 years worth of available data
Data Comprehensiveness	The detail and the area that the data are collected over. 0: Data are collected disparately. 5: Data are collected at regular intervals, and with in-depth detail.
Data Accuracy	The correctness of the data and the methods that are in-process to ensure as little data inaccuracy occurs as possible. 0: Data are inaccurate, out of date and are presented in variety of different forms. 5: The data are kept regularly updated, and there are automated checks to ensure that the data recorded are reasonable and acceptable.
Data Consistency	Are the data kept in the same format, and the same methodology used to capture the data. 0: Data are kept in a variety of formats, and different methods have been used to record the data. 5: The data are recorded through a defined standard, designed to reduce erroneous data, the same approach has been used to collect the data and if it has altered, then previous data have been modified accordingly.
Meta Data	The amount of details that are given regarding the dataset; data about the data. 0: No information about the data is given 5: There is detailed structured list regarding the details of the dataset, which is stored in a human readable format such as XML which is complaint with INSPIRE regulation.
Data Format and Standards	Doe the data comply with INSPIRE standards? 0: The data are not in any shape or form compliant with INSPIRE standards, or other EU standards. 5: The data are fully compliant with INSPIRE regulations.

Source: Montieth et al.(2010)

Table 3. Scoring Metrics and Their Values for Integrating Air Quality Databases of the UK

	Dataset					
Scoring Metrics	NAEI	PCM	AURN	LAQN	NPL	CEH
Data Searchability	2	1	1	1	1	2
Data timeline	2	2	3	5	2	1
Data Donwloadability	1	3	2	2	2	2
Data Historical Trending	5	3	5	2	1	1
Data Comprehensiveness	4	4	3	3	2	2
Data Accuracy	3	3	3	3	2	3
Data Process	2	2	1	4	2	2
Metadata	3	1	2	3	3	2
Data Format and Standards	3	1	2	3	3	2
INSPIRE	1	1	1	1	1	1
Total	28	26	27	28	23	22

<Table 4>는 전 처리 결과 데이터를 나타낸다. 데이터 프로세스를 위한 전 처리 과정으로 가장 보편적으로 사용되는 서열척도의 역 코딩(reverse scoring)을 사용했다. 이를 통해 투입 변수인, 검색 용이성, 타임라인, 다운로드 용이성, 데이터 추세 변수를 전 처리하였다. 예를 들어 타임라인의 경우 전처리 전에는 NAEI는 2의 값을 가졌지만, 역 코딩을 통해 4의 값을 가지게 되었다. 이는 방어적 측정(defensive measure)으로 값이 클수록 더 좋은 것이 아니라, 값이 클수록 방어 성능이 더 나쁜 것을 의미하는 역 투입(reverse input)이다.

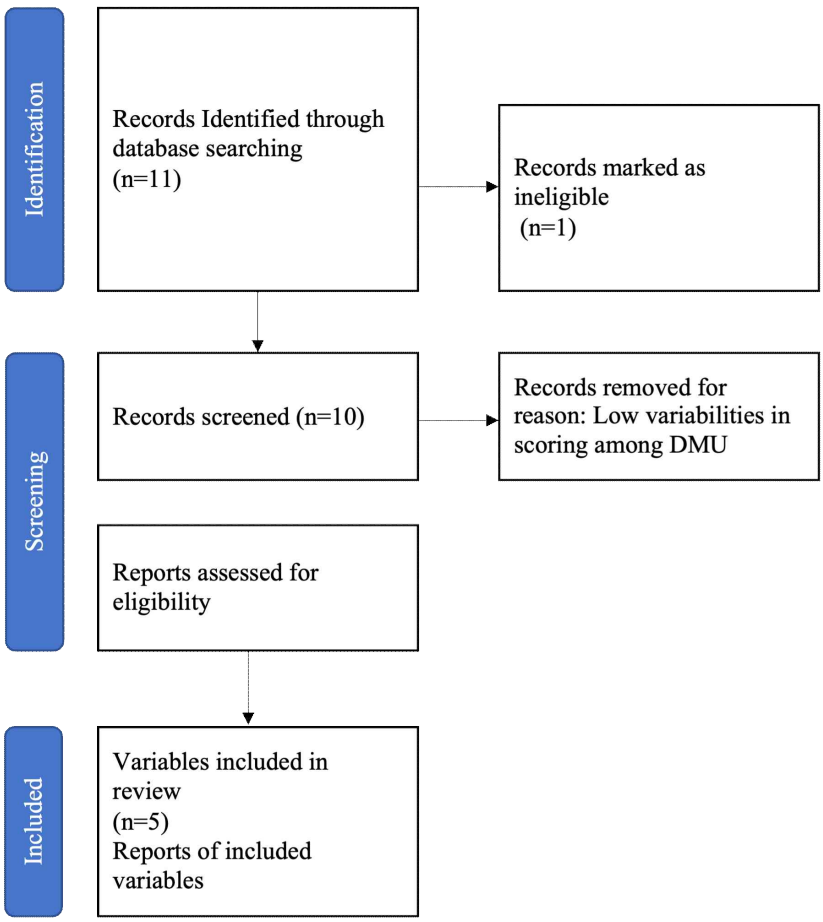
Table 4. Preprocessing Results Data

	Data Searchability	Data Timeline	Data Downloadability	Data Timeline	Data Comprehensiveness	Data Accuracy	Data Consistency	Data Processing	Metadata	Data Format and Standards
NAEI	4	4	5	1	4	3	3	2	3	2
PCM	5	4	3	3	4	3	4	2	1	2
AURN	5	3	4	1	3	3	4	1	2	2
LAQN	5	1	4	4	3	3	2	4	3	2
NPL	5	4	4	5	2	2	5	2	3	2
CEH	4	5	4	5	2	3	5	2	2	1

우리는 해당 데이터를 가공하여 분류하고, PRISMA 방법을 통한 데이터 전 처리 과정을 거쳐 의미 있는 발견을 하고자 한다. 본 연구에서 사용한 자료포락분석(Data Envelope Analysis,이하 DEA) 방법은 의사결정단위(Decision-making units, 이하 DMU) 간의 상대적인 효율성을 결정하는 방법이기 때문에, 사용하는 투입 및 산출 요소가 증가 할수록 효율적인 DMU의 수가 증가하는 경향이 있기 때문에 가능하다면 최소한의 투입 및 산출 요소를 사용하여 설명하는 것이 바람직하다(Nyhan & Martin, 1999). 따라서 본 논문에서 우리는 PRISMA 방법을 이용하

여 최소한의 투입 및 산출 요소를 선택하였다. <Figure 1>은 PRISMA 방법을 통해 데이터 전 처리 한 과정을 다이어그램으로 도식화 한 것이다. 변수 INSPIRE 은 분석 대상이 아님으로 제외 했다. 본 논문에서는 DEA 분석을 위한 투입 및 산출 요소로 선정되기 위한 중요한 자격 기준 (eligibility criteria)으로 DMU간에 되도록 높은 변동성을 가진 것으로 제안한다. 이로써 해당 변수가 DMU 사이에 평가 값 사이의 차이가 크지 않은 경우는 제외했다. 예를 들어, 데이터 검색용이성의 경우 6개의 DMU가 4 또는 5 의 값만을 가진다. 반면에, 데이터 타임라인의 경우는 DMU가 1, 3, 4, 5 다양한 값을 가진다. 이렇게 DMU간 높은 변동성을 가진 것으로 투입 요소 변수로는 데이터 타임라인, 데이터 추세, 산출 요소 변수로는 데이터 포괄성, 데이터 일관성, 데이터 처리를 선정하였다.

Figure 1. Data Screening Through the PRISMA Method



3.2. 분석 방법

3.2.1. DEA 모형

본 논문에서 사용하고 있는 자료포락분석 (Data Envelope Analysis,이하 DEA) 방법은 여러 산업 분야에 널리 사용되고 있는 평가 방법이다. (Jomthanachai et al., 2021; Misiunas et al., 2016). DEA는 각 의사결정단위

(Decision-making units, 이하 DMU)의 효율성을 평가하는데 있어서 다수의 투입과 다수의 산출을 개별적인 효율성 점수로 변환하여 보여주기 때문에 상대적 효율을 평가하는데 용이하다(Charnes et al., 1997; Bryce et al., 2000). 이 기법은 Farrell(1967)에 의해 처음 연구되었으며, 다수의 투입과 다수의 산출을 동시에 고려할 수 있고, 모델을 개발하기 위해서 특정한 통계적 가정이 필요하지 않은 대표적인 비모수적 접근방법이다. 무엇보다 투입이나 산출물을 계량적으로 측정하기 어려운 경우에도 효율성을 비교적 쉽게 평가할 수 있기 때문에 정부기관, 교육기관, 기업 등의 효율성을 평가하는데 사용되어 왔다.

3.2.1.1 기본 DEA 모형

평가해야 할 공기 질 데이터베이스 DMU가 n 개가 있고, 각 데이터베이스는 m개의 투입 요소 을 이용하여 s개의 산출물을 생성한다고 가정하자. 개별 DMU는 하나 이상의 투입 요소를 사용하여 하나 이상의 산출 요소를 생산한다고 가정한다. DMU가 n 개 있다고 가정할 때, j번째 DMU는 투입 요소 m을 이용하여 산출 요소 s를 생산한다. 기본적인 산출 지향 CCR(Charnes, Cooper and Rhodes: CCR) DEA 모델의 수식은 다음과 같다. 아래의 수식은 목적함수의 투입물의 가중 합을 1로 고정하여 제약조건식을 변형한 선형 계획법이다(Cooper et al., 2007).

$$(\max) h_k = \sum_{r=1}^{s} u_r y_{rk}$$

$$st \tag{1}$$

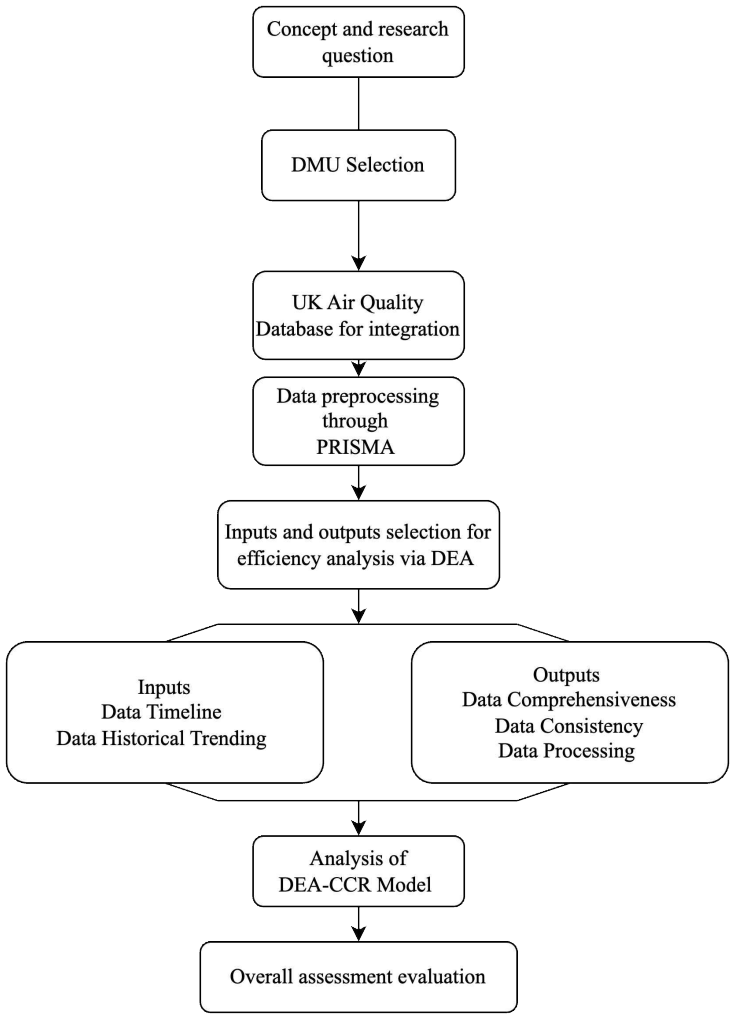
$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1, \ldots, n$$

$$u_r \geq 0, \ r = 1, 2 \ldots, \ s,$$

$$v_i \geq 0; i = 1, 2, \ldots, \ m,$$

여기서 점수는 해당 DMU의 상대적인 효율성을 나타낸다. 이론적으로 DEA 모형은 상대적인 정도를 비교하는 방법이다. 각각의 DMU가 산출물 대비 최소의 투입이 사용되는 효율성 프런티어(efficiency frontier)와 얼마나 떨어져 있는가를 측정하기 때문이다(Ko et al., 2011). 이때 DMU가 가질 수 있는 가장 큰 효율성 값은 1이다. 만약 DMU의 값이 1이 된다면 해당 DMU는 효율성 프런티어(efficiency frontier)로 판단한다. 최대 가능 효율 점수(Maximum possible efficiency score)를 계산하기 위해서 산출 요소와 투입요소에 각각 가중치를 둔다. u_r, v_i 는 각각 r번째 산출 요소($r = 1, 2, \ldots s$)와 i번째 투입 요소($i = 1, 2, \ldots m$)에 가중치를 둔 것을 나타낸다. x 와 y 는 각각 투입 요소와 산출 요소를 의미한다. 효율성 값은 다른 DMU와 비교 했을 때 최대로 얻을 수 있는 점수를 뜻한다. 이상의 연구방법을 요약하여 도식화 하면 아래 <Figure 2>와 같다.

Figure 2. The Summary of the Research Methodology



3.2.1.2 연구 문제

영국의 전략적 공기 질 측정과 관련된 데이터베이스DMU의 효율성을 평가하는 데 있어서 투입 변수 4개와 산출 변수 6개를 고려하였다. 개별 DMU의 효율성을 100%로 가정할 때, 나머지 5개 DMU의 효율성이 100%에 못 미치는 경우에, 효율성이 못 미치는 DMU에 대한 아래와 같은 귀무가설을 기각하는 방식으로 검증하였다.

가설 1. NEIH데이터베이스는 투입 변수 값 대비 산출 변수 값 비율이 효율적이다.
가설 2. PCM 데이터베이스는 투입 변수 값 대비 산출 변수 값 비율이 효율적이다.
가설 3. AURN 데이터베이스는 투입 변수 값 대비 산출 변수 값 비율이 효율적이다.
가설 4. LAQN데이터베이스는 투입 변수 값 대비 산출 변수 값 비율이 효율적이다.

가설 5. NPL데이터베이스는 투입 변수 값 대비 산출 변수 값 비율이 효율적이다.
가설 6. CEH데이터베이스는 투입 변수 값 대비 산출 변수 값 비율이 효율적이다.

3.2.2. 변수의 조작적 정의

데이터 거버넌스와 관련한 효율성을 분석하는 데 있어서 투입 및 산출 변수를 어떤 것을 설정하느냐에 따라서 효율성 점수가 달라질 수 있다. 일반적으로 DEA에서 고려하는 투입 변수는 조직의 비용, 산출 변수는 조직의 편익을 의미한다. 데이터 거버넌스를 위한 통합 데이터 구성에 있어서 투입 변수 4개는 데이터 검색 용이성, 데이터 타임라인, 데이터 다운로드 용이성, 데이터 추세이다. 산출 변수 6개는 데이터 포괄성, 데이터 정확성, 데이터 처리, 메타데이터, 데이터의 형식과 표준화이다. 이 중 의사결정단위 (DMU) 간의 높은 변동성을 고려하고, PRISMA 기법을 통해 선별적으로 투입 및 산출 변수를 선정하였다. <Table 5>는 영국의 전략적 수준의 공기 질 데이터 평가 지표를 통합 데이터베이스 구축을 위해 투입 변수 및 산출 변수로 조정 분류한 것이다.

Table 5. Adjusted Input and Output Variables and Their Meanings for the Integrated Database

	Variables	Descriptions
Input Variables	Data Timeline	· The data contains up-to-date data and there is an indication of when the last update took place/when future updates will take place. · Data and relevant data are frequently updated with regular interval.
	Data Historical Trending	· The database contains records past years which thus well reflect the trends
Output Variables	Data Comprehensiveness	· Data are collected at regular intervals, and with in-depth detail.
	Data Consistency	· The data are recorded through a defined standard, designed to reduce erroneous data · The same approach has been used to collect the data and if it has altered, then previous data have been modified accordingly
	Data Processing	· The data is uploaded and processed in a way that can well be linked to meaningful information (i.e. there is representative data linked to a specific process)

투입 변수와 산출 변수 사이의 상관관계를 분석 결과가 <Table 6>에 나와 있다. 표본의 통계치를 중심으로 상관관계 계수가 높은 변수는 타임라인과 일관성의 상관계수 0.807으로 매우 강한 음(-)의 상관관계를 갖고 있는 것으로 나타난다. 포괄성과 추세는 0.731으로 다소 강한 양(+)의 상관관계를 가지고 있다. 이는 오랜 과거 기록을 잘 보유하고 있는 데이터 세트, 즉 추세를 잘 반영하고 있는 데이터는 포괄적인 내용을 담고 있을 수 있다고 유추해 볼 수 있다. 횡적으로 잘 연결되어 있는 데이터 일수록 다양한 곳에서 수집 기록되어 있다고 볼 수 있기 때문에 데이터의 일관성은 떨어질 수 있음을 예상해 볼 수 있다. 그럼에도 불구하고, 횡적으로 잘 연결되어 있는 데이터 일수록 데이터 처리를 통해 의미 있는 정보로 연결 될 수 있다는 점을 유추할 수 있다.

Table 6. The Corelation Coefficients Between Input and Output Variables

Variables	Data Comprehensiveness	Data Consistency	Data Processing
Data Timeline	0.162	-0.807	0.664
Data Trending	0.731	-0.388	-0.425

4. 연구 결과

4.1 공기질 데이터베이스의 효율성 분석

영국의 전략적 공기질 데이터베이스의 효율성 분석을 위하여 역 코딩 (reverse-scoring) 투입 변수 2개와 산출 변수 3개를 이용하여 DEA분석을 한 결과가 <Table7> 에 제시되어 있다. DEA 적용 결과 1의 효율성을 갖는 DMU (의사결정단위)는 3개이다. NAEI, ARUN, LAQN이다.

이에 따라서 가설 2,5,6을 기각 한다:

가설 2. PCM데이터베이스는 투입 변수값 대비 산출 변수 값 비율이 효율적이다;
가설 5. NPL데이터베이스는 투입 변수값 대비 산출 변수 값 비율이 효율적이다;
가설 6. CEH데이터베이스는 투입 변수값 대비 산출 변수 값 비율이 효율적이다.

Table 7. The Results of DEA Application

Master Data	Efficiency
NAEI	1
AURN	1
PCM	0.8060
LAQN	1
NPL	0.8333
CEH	0.6875

효율이 가장 좋은 공기질 데이터베이스 DMU는 총 세가지 NAEI, AURN, LAQN이다. 효율이 가장 좋은 프런티어에 있는 세 가지 공기질 데이터베이스 DMU는 어떤 특성을 가지고 있는 지 해당 DMU의 투입 요소, 산출 요소를 분석하여 어떠한 방향성을 보이고 있는지 살펴 볼 수 있다. 이를 통해 세 가지 DMU가 프런티어에 선정된 이유를 논하고, 같은 해당 DMU가 다른 DMU와 같은 결을 가지고 있는지 분석할 수 있다.

4.2. 효율이 가장 좋은 프런티어에 있는 공기질 데이터베이스 DMU에 대한 분석

DEA 분석 결과 1의 효율성을 갖는 DMU는 동일하게 값이 1이 나왔다고 하더라도 그 안에서 효율성을 갖게 된 이유의 차이가 존재할 수 있다. DEA 분석 결과 1의 효율성을 갖는 공기질 데이터베이스 DMU는 NAEI, AURN,

LAQN 총 세 가지이다. <Table 8>은 1의 효율성을 보이는 DMU가 왜 프런티어에 선정되었는지, 투입 요소와 산출 요소를 기준으로 비슷한 프런티어의 성향으로 분류 한 것이다. 이를 통해 효율적인 여러 공기 질 데이터베이스 DMU의 기준점의 투입 요소와 산출 요소간의 관계를 살펴 볼 수 있고, 통합 데이터 구축에 있어서 강조 되어야 할 부분이 어디에 있는지 유추해 볼 수 있다.

Table 8. DMU Located at the Frontier Sharing the Similar Characteristics

DMU	Superior Inputs	Inferior Inputs	Superior Outputs	Inferior Outputs
NAEI	Data Trending	Data Timeline	Data Comprehensiveness	Data Processing
AURN				
LAQN	Data Timeline	Data Trending	Data Processing	Data Comprehensiveness

우선 NAEI와 AURN은 상대적으로 효율적이라 판단된 이유가 동일하다. NAEI와 AURN은 강한 데이터 추세를 보이고 있다. 이는 해당 데이터가 과거 년도에 대한 데이터를 포함하고 있다는 것이고, 데이터베이스가 최신의 데이터도 포함될 수 있도록 잘 관리 되고 있다는 것이다. 데이터의 추세가 잘 반영되어 있다는 것은 데이터의 종적인 연결이 좋다는 것을 의미하며 측정하려고 하는 세부 내용에 대해 포괄적으로 잘 반영하고 있을 가능성이 높다. 이러한 의미에서 해당 데이터베이스는 포괄성 측면에서 우수한 것을 볼 수 있다. 반면에 NAEI 와 AURN은 데이터 처리 수준 정도가 비교적 낮은 것을 볼 수 있다. 이것은 수많은 년도의 데이터를 포함함으로 데이터가 양적으로 커졌기 때문에 사용 가능한 정보로 변환하는 것에 어려움이 있을 수 있다는 것을 의미한다.

LAQN의 경우에는 데이터 타임라인이 우수하여 프런티어 선상에 있음을 알 수 있다. 데이터 타임라인이 우세하다는 것은 데이터의 횡적인 연결(cross-sectional)이 우수하다는 것을 의미한다. 데이터의 횡적인 연결이 잘 되어 있다는 것은, 데이터베이스의 업데이트 주기 간격이 일정하고, 같은 시간대에 관련 데이터가 동일하게 업데이트 된다는 것을 의미하기 때문에 상호 운용성을 높일 수 있다.

본 연구를 통하여 데이터 거버넌스를 위해 고려할 사항으로, 충실하게 구성된 종적 연결 데이터를 기반으로 횡적 연결을 세밀하게 할 때 유용한 데이터를 구축하는 것이 중요하다는 점을 파악할 수 있었다. 하지만 실질적으로 종적이면서 동시에 횡적으로도 잘 연결된 데이터를 구축하는 것은 도전적인 일이다. 한 데이터를 긴 기간 동안 정밀하고 세밀하게 축적하는 것에 비해 하나의 대상과 관련된 다수의 다른 성향의 데이터를 동 시간대에 업데이트하여 축적하는 것에는 어려움이 따르기 때문이다. 하지만 영국의 공기 질 측정 데이터 세트를 DEA로 분석한 결과에서 볼 수 있는 것처럼 과거의 데이터를 포함하고 있어서 추세를 파악 할 수 있는 것과, 관련 데이터베이스들의 업데이트가 같은 주기로 이루어져 횡적인 연결이 좋은 것이 같은 방향성을 보이지 않아도 해당 데이터 세트는 상대적으로 효율적일 수 있음을 알 수 있었다.

이러한 분석 결과는 본 논문에서 주장하는 연합형 방식의 통합 데이터 구축 시 시사점을 제공한다. 데이터베이스 간 연합을 위한 통합 데이터 선정 시, 데이터 추세를 강화하면서 상대적은 많은 양의 정보를 공유하는 포괄적인 데이터베이스를 구축하는 방향으로 갈 것인지, 아니면, 데이터 업데이트 주기를 동기화 하여 횡적으로 잘 연결 되어 있는 데이터베이스를 구축하여 데이터 처리를 통해 의미 있는 정보로 연결 될 수 있도록 구현 할 것인지에 대한 데이터 거버넌스 정책 방향을 제시한다

4.3. 효율성이 프런티어에 못 미치는 공기 질 데이터베이스 DMU 분석

DEA 분석 결과 나머지 세 개 공기 질 데이터베이스DMU인, PCM, NPL, CEH 는 모두 효율성이 프런티어에 못 미치는 것으로 나타났고, 세부사항을 검토하면 다음과 같다. DEA 분석 결과를 논의할 때 유의해야 하는 점은 각 데이터 세트가 비효율적으로 수집, 처리 되고 있다는 것이 아니라, 데이터 거버넌스 성과 측정변수 관점에서 데이터 통합 시 상대적으로 비효율성을 보이고 있다는 의미로 해석해야 한다.

PCM을 NAEI 나 AURN과 비교하여 볼 때, 투입 요소인 데이터 추세 및 데이터 타임라인 모두가 좋지 않다. 산출 요소를 보았을 때 데이터 포괄성은 비교적 좋으나 데이터 처리가 좋지 않다. PCM을 LAQN에 비교해 보았을 때에는 현전하게 데이터 타임라인이 열세 한 것을 알 수 있다. NPL의 경우는 NAEI나 AURN에 비교해 보았을 때 데이터 추세가 좋지 않고, 포괄성 역시 떨어진다. 일관성은 높다. NPL을 LAQN과 비교했을 때 두드러지는 점은 타임라인이 좋지 않고 처리가 떨어진 다는 점이다. CEH는 NAEI나 AURN에 비교했을 때 데이터 추세가 좋지 않다.

효율성 값은 본 논문에서 주장하는 연합형 방식의 통합 데이터 구축을 고려했을 때 효율성이 프런티어에 못 미치는 공기 질 데이터베이스에 대하여 연합형 데이터 거버넌스를 시행하기 위해서는, 세부적으로 효율성 증진 요소를 데이터 추세, 데이터 타임라인, 데이터 처리 가운데 어느 부분에 초점을 두어야 하는 것이 보다 용이하게 통합되는지 판단하는 기준을 제공한다고 볼 수 있다.

4.4. DEA 분석 결과 요약

DEA 분석 결과 어떤 데이터베이스를 중심으로 통합할 때, 투입 요소를 강조함으로 효율성을 높일 수도 있고, 산출 요소를 강조함으로 효율성을 높일 수도 있음을 알 수 있었다. 통합 데이터베이스를 구축한다는 의미는 모든 데이터베이스들을 연결하겠다는 것에 있지 않다. 그것은 현실적으로 협업 업무가 지나치게 증가 할 문제가 있을 뿐 아니라, 다양한 데이터 세트의 특성을 고려했을 때 모든 투입 및 산출 요소가 우세한 데이터 세트의 구성 및 구축이 불가능하기 때문이다.

DEA 분석을 통해서 효율성 프런티어 라인에 위치하고 있는 여러 데이터베이스를 확인해 볼 수 있고, 그 데이터베이스를 기준으로 통합 시 필요한 통합데이터의 우선순위를 정해 볼 수 있다. 어떤 데이터베이스DMU가 100퍼센트의 효율을 보인다고 하는 것은 특정 투입 요소나 산출 요소가 우세했기 때문이다. 그렇다면, 공기 질 데이터베이스DEA 분석 결과가 현실적으로는 어떤 의미를 가지는 것일까? 투입 요소를 고려해 보았을 때 데이터 추세를 강화하거나, 데이터 타임라인을 강화하는 방법에 대해서 고려해 볼 수 있다.

데이터 추세를 강화하기 위해서는 측정하고자 하는 해당 대상의 센싱을 강화하는 것이 한 가지 방법이다. 투입할 센서의 양과 종류를 강화해서 한 가지 대상에 대한 데이터 추세 즉 시간대를 강화하여 종적으로 포괄적인 데이터를 구축할 수 있게 된다. 데이터 타임라인을 강화하기 위해서는 횡적으로 같은 시간대에 발생하는 여러 관련 대상의 측정 데이터들을 엮어야 한다. 실제로는 타임라인 별 자료를 구성하여 횡적으로 연결 하는 것이 더 어렵다고 할 수 있다.

5. 결론 및 함의

본 논문은 지능형 데이터 거버넌스 구축을 위한 데이터 통합의 기준점을 제시하기 위하여 DEA방법을 이용해 영

국의 전략적 공기 질 측정 데이터베이스의 통합을 위한 효율성을 검증하였다.

이 방법을 통해 우리는 프런티어 라인에 위치하여 데이터 거버넌스을 위한 데이터베이스 통합 시 효율과 효과를 높일 수 있는 공기 질 데이터베이스가 무엇인지 알게 되었다. 효율적인 프런티어에 위치한다고 판단되더라도 각 DMU의 강조점이 다르다는 것을 파악 하였다. 무엇보다 효율적인 공기질 데이터베이스 DMU와 비효율적인 공기 질 데이터베이스 DMU의 위치를 파악하게 되어, 데이터 거버넌스 구축을 위하여 기존 데이터베이스를 통합함에 있어서 어떤 평가요소에 초점을 맞춰야 하는 가를 도출할 수 있게 되었다.

DEA 분석 결과 NAEI, AURN, LAQN 총 세 가지 공기 질 데이터베이스가 효율적으로 평가 되었다. 나머지 공기 질 데이터베이스 DMU의 효율성은 대체로 80%에 머물고 있고 (PCM, NPL), 60% 정도의 효율성에 머무는 DMU도 있었다 (CEH). 분석 결과 고려해야 할 투입 요소로서 데이터 타임라인이나 데이터 추세가 우세하여, 어떤 공기 질 데이터베이스를 통합의 기준으로 삼고, 어떤 통합 데이터를 선정하여 통합할 것인지, 알아볼 수 있었다. 이러한 방향으로 연합형 데이터 거버넌스가 수행된다면 산출요소 관점에서 데이터의 포괄성이나 데이터 처리의 용이성을 확보할 수 있다는 점 역시 파악하였다.

데이터 거버넌스를 위하여 하나의 최적화된 기준을 제시하는 것이 아니라 연합형 통합 방식을 취하는 것이 유리하다는 것이 본 논문의 관점이다. 효율성 분석을 했을 때 공기 질 데이터베이스가 프런티어에 있을 수 있는 이유는 고려된 성과 측정 변수에 따라 달라 질 수 있다. 그렇기 때문에 데이터 거버넌스에서 중앙 집중형 통합이 항상 좋은 결과를 가져 온다고 볼 수 없다고 본 연구는 제시한다.

본 논문에서 주장하는 '지능형' 데이터거버넌스 (IDG)는 데이터 거버넌스를 위한 데이터 통합을 좀 더 용이하게 할 수 있는 여러 가지 기준점, 특별히 연합형 관점에서의 다수의 기준점을 제시 했다는 점에서 지능형 데이터 거버넌스의 방향성을 제시한다고 할 수 있다. 본 연구는 정성적인 평가를 정량적인 수치로 평가한 결과를 DEA방법을 이용하여 평가 결과에 대한 심층적인 평가가 가능함을 보여주었다는 점에서 학문적 기여를 한다. 본 연구는 무엇보다 데이터 거버넌스 구축을 위한 내부적인 가이드라인 및 방향성을 제시하고 있다. 이것은 향후 철도 데이터 거버넌스 뿐 아니라 타 분야에서 데이터 거버넌스 구축을 위한 내부적인 가이드라인은 어떤 방향성을 띄어야 하는지에 대한 함의를 제공한다는 의미이며, 이것이 본 연구의 실무적인 기여라 할 수 있다.

본 연구가 한국 공기 질 평가 데이터를 기반 하였다면 한국 공기 질 평가 데이터의 연합형 통합을 위한 직접적인 제안이 될 수 있었을 것이라 기대할 수 있다. 하지만, 한국의 경우에는 공기 질 측정 항목 역시 포괄적이지 않기 때문에 영국의 전략적 공기 질 평가 결과를 사용하여 간접적으로 시사점을 도출해야 했던 데이터 선택의 지역적 한계가 있다. 하지만 이러한 한계는 본 논문의 목적으로 상쇄가 가능하다. 앞서 밝힌 것처럼 본 연구의 목적은 데이터 거버넌스의 구현과 평가를 위한 성과 지표 개발 및 제시에 있는 것이 아니라, 연합형으로 데이터를 통합할 때 어떤 기준으로 통합되어야 하는지에 대한 논의에 있기 때문이다.

그럼에도 불구하고, 데이터 선택의 지역적 한계를 극복하기 위해서 향후 연구 과제로 한국표준협회 및 한국표준협회의 공기 질 인증 지표 등(예: 한국표준협회 실내 공기 질 인증제도)과 연계하여, 지표 개선 방향을 논의하는 연구 진행을 제안한다. 공공기관에 ESG경영에 대한 요구가 부상하고 있다(Cho and Pyun, 2022). 데이터 통합은 공공서비스 품질관리와 맥락을 같이 한다. 공기 질과 관련한 데이터 통합에 대한 논의는 공공기관에 부상하고 있는 ESG 경영에 대응 하는 것에 있어서도 시의적절하다. 또한 해당 데이터베이스를 어떻게 관리할 것인지, 이를 위해서 한국표준협회와 어떤 논의가 이루어질 수 있는 지에 대한 연구를 기대해 볼 수 있다.

본 연구를 통해 데이터 거버넌스의 통합적 관점에서 내부적인 가이드라인을 세울 수 있게 되었다. 무엇보다 데이터 거버넌스 통합 관점에서 유연한 접근성을 취하기 때문에, 빠르고, 쉽고, 효과적으로 통합 할 수 있게 되었다. 본 연구에서 제시한 가이드라인을 통해 철도 산업의 데이터 거버넌스 구축 시 공 기질 데이터베이스를 포함한 여러 데

이터베이스의 통합과정이 연합형을 기반으로 하는 지능적인 데이터 거버넌스 형태로 발전할 수 있기를 기대해 본다.

REFERENCES

Allen, M, and Cervo, D. 2015. Strategy, Scope, and Approach In Allen, M., and Cervo (Eds.), Multi-Domain Master Data Management: Advanced MDM and Data Governance in Practice, (pg.244). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-800835-5.00001-4.

Bryce, CL., Engberg, JB., and Wholey, DR. 2000. Comparing the agreement among alternative models in evaluating HMO efficiency. Health Serv Res. 35(2):509-28. PMID: 10857474; PMCID: PMC1089131.

Charnes, A., Cooper, W., Lewin, A. Y., and Seiford, L. M., 1997. Data envelopment analysis theory, methodology and applications. Journal of the Operational Research society 48(3):332-333.

Charnes, A., Cooper,. WW, and Rhodes, E. 1987, Measuring the efficiency of decision-making units. European Journal of Operational Research 2:429-44.

Cheon, BJ., and Kim, HW. 2016. An Exploratory Study on the Sharing and Application of Public Open Big Data Informatization Policy 24(3):27-41.

Cho, JH. and Pyun, JB 2022. A Study on the Implementation Plan for Public Service Quality Management Applying the ISO 18091 Framework. J Korean Soc Qual Manag 50(1): 1-19.DOI: https://doi.org/10.7469/JKSQM.2022.50.1.1

Cho, YM., Park, KS., Park, DS., Koo, HY., Bin, HK., and Kim, HM. 2010. Strategies for Improvement of Air Quality in Subway Stations, The Korean Society for Railway 2010 Spring Conferences Proceedings pp. 2117-2121.

Choi, B. and Yoon, JJ 2018. A Study on Policies to Revitalize the Public Big Data in Seoul. The Seoul Institute Policy Report. The Seoul Institute. p.147.

Choi, YH., Kim, TJ., Soh, SH., and Jeong, SH. 2015. A Study on the Improvement Plan for Data Management System. Ministry of the Interior and Safety (MOIS) Research Report.

Cooper, S. A., Smiley, E., Finlayson, J., Jackson, A., Allan, L., Williamson, A., & Morrison, J. 2007. The prevalence, incidence, and factors predictive of mental ill-health in adults with profound intellectual disabilities. Journal of Applied Research in Intellectual Disabilities 20(6):493-501.

Informatica, Best Practices for Intelligent Data Governance. White Papers. Retreived from https://www.informatica.com/kr/lp/best-practices-azure-data-governance_4237.html.

ITF, 2021, Reporting Mobility Data: Good Governance Principles and Practices, International Transport Forum Policy Papers, No. 101, OECD Publishing, Paris

Joen, S., Lim, H., Park, S., & Jung, H. 2022. Indoor Air Data Meter and Monitoring System. Journal of the Korea Institute of Information and Communication Engineering 26(1):140-145.

Jomthanachai, S., Wong, W.-P., & Lim, C.-P. 2021. A Coherent Data Envelopment Analysis to Evaluate the Efficiency of Sustainable Supply Chains. IEEE Trans. Eng. Manag. 1-18.

Jomthanachai, S., Wong, WP., Soh, KL, and Lim, CH. 2021. A global trade supply chain vulnerability in COVID-19 pandemic: An assessment metric of risk and resilience-based efficiency of CoDEA method. Research in Transportation Economics 93(4):101166. https://doi.org/10.1016/j.retrec.2021.101166.

Kans, M. and Ingwald, A. 2021, Service-based business models in the Swedish railway industry, Journal of Quality in Maintenance Engineering, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JQME-06-2021-0051.

Kans, M., Galar, D., and Thaduri, A. 2016, Maintenance 4.0 in railway transportation industry, Proceedings of the 10th World Congress on Engineering Asset Management (WCEAM 2015), Springer. pp.317-331.

Kim, JH., Jeong, SY., Lee, SB., Choi, JA. 2020. Agenda and Key Issues in Data Governance. Korea Institued of Public Administration. Sep 2020 Issue No.5.

Kim, MJ., and Sagong, HS. 2015. Era of Data Technology: Geospatial Information Establishment and Management Plan. KRIHS Policy Brief. 530:1-6.

Ko, DW., Park, HU, and Park, JW. 2011. An Analysis of the efficiency of public sport facilities in local governments. Korean Society of Sport Policy 9(3):1-11.

Lewis, H.F., and Sexton, T.R. 2004. Data Envelopment Analysis with Reverse Inputs and Outputs. Journal of Productivity Analysis 21:113-132. https://doi.org/10.1023/B:PROD.0000016868.69586.b4.

Matthew J.P et al. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 2021; 372 doi: ttps://doi.org/10.1136/bmj.n71.

Misiunas, N., Oztekin, A., Chen, Y., & Chandra, K. 2016. DEANN: A healthcare analytic methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status. Omega 58:46-54.

Monteith, A, Cronk, O., Yardley, R., Willis, P., and Xiao, X. 2010. Air Quality Data Management and Integration System. AEA Group.

NOCoE Data Governance for TSMO 2021, Virtual Peer Exchange Proceeding Report, National Operations Center of Excellence (NOCoE).

Nyhan, R. C., and Martin, L. L. 1999. Assessing the Performance of Municipal Police Services Using Data Envelopment Analysis: An Exploratory Study. State and Local Government Review 31(1):18-30. https://doi.org/10.1177/0160323X9903100102.

Park, MS., Bae KM., and Kim, YS. 2021. How to Apply the New Quality Dimensions to the New Business in the Digital Transformation Era?J Korean Soc Qual Manag. 49(4):609-622.

Pieriegud, J. 2018. Digital Transformation of Railways. Siemens Sp.Z 0.0., Poland.

Rotter, M., Hoffmann, E., Pechan, A. and Stecker, R. 2016. Competing priorities: how actors and institutions influence adaptation of the German railway system, Climatic Change 137(3):609-623.

Ryu, SM 2018. A study on the comparative analysis of air quality in tunnels according to road bed of underground section in subway tunnel. [Master's thesis, University of Seoul]

Ślusarczyk, B. 2018. Industry 4.0: Are we ready? Polish Journal of Management Studies 17(1). DOI: 10.17512/pjms.2018.17.1.19.

Soh, HJ., Byun JH., and Kim, DH. 2021. Quality 4.0: Concept, Elements, Level Evaluation and Deployment Direction. J Korean Soc Qual Manag 2021; 49(4):447-466. DOI: https://doi.org/10.7469/JKSQM.2021.49.4.447.

Soh, JS., and Yoo, SY. 2008a. Measurement and Analysis of Indoor Thermal Environment in Passenger Car. International Journal of Railway 11(2):120-125.

Soh, JS., and Yoo, SY. 2008b. A prediction of CO2 Concentration and Measurement of Indoor Air Quality in the EMU. International Journal of Railway 11(4):378-383.

Transportation Research Board of the National Academies. 2015. Improving Safety Programs Through Data Governance and Data Business Planning, Transportation Research Circular, Number E-C196.

저자소개

김민정 경희대학교 환경공학 학사(2010년), 석사(2012년), 박사(2016년) 학위를 취득하였다. 현재는 한국철도기술연구원에서 선임연구원으로 근무 중이며, 주요 관심분야는 AI, 실내 공기질, 예측 제어, 최적화 등이다.

원종운 한국해양대학교 제어계측공학 학사(1996년), 동 대학원 제어계측공학 석사(1998년), 경북대학교 전자공학 박사(2004년) 학위를 취득하였다. 현재는 한국철도기술연구원에서 책임연구원으로 근무 중이며, 주요 관심분야는 AI, 영상처리, 유통물류, 블록체인 등이다.

박가영 이화여대학교 영어영문학 학사(2008년), 고려대학교 국제대학원 국제개발협력석사(2011), 뉴욕주립대학교 Stony Brook, 기술혁신정책 공학박사(2022년) 학위를 취득하였다. 현재는 한국뉴욕주립대학교에서 겸임교수로 근무 중이며, 주요 관심 분야는 AI, 디지털 전환, 글로벌 융복합 혁신, 디지털 휴머니티, 에너지-수자원 넥서스, 섹터커플링 등이다.

박상찬 서울대학교 경영학 학사(1984년), 미국 미네소타대학 MBA(1985년), 미국 일리노이대학교 경영학 박사(1991년) 학위를 취득하였다. 현재는 한국뉴욕주립대학교 기술경영학과에서 학과장으로 근무 중이며, 주요 관심분야는 AI, 배터리, 신재생 에너지, Robotics, 유통물류, 의료경영, 데이터 거버넌스 등이다.