

2022년 제3호 (통권22호)



데이터산업 동향 이슈 브리프

ISSUE BRIEF

2022.3

유럽의 데이터 주도 혁신 촉진을 위한
샌드박스 사례 분석

유럽의 데이터 주도 혁신 촉진을 위한 샌드박스 사례 분석

I. 검토 배경	1
II. 공공 데이터와 시민 데이터의 결합-지리정보 데이터	2
III. 데이터 공급을 위한 민관 협력-도시 지역 대기질 데이터	10
IV. 결론 및 시사점	18

요 약

- EU는 유럽 데이터 전략(European strategy for Data)을 통해 데이터 주도 혁신을 촉진하고 경제와 사회를 위한 가치 창출 추구
 - 특히 다양한 출처의 데이터 조합 및 결합은 유럽 공통 데이터 공간(European data space)을 성공적으로 구축하는 데 가장 중요한 요소로, 데이터의 공유와 사용을 위한 첨단 기술과 혁신적 접근법이 이를 가속화
 - 첨단 기술과 다양한 도구를 사용하는 일련의 실험을 가능하게 하는 샌드박스 접근법에 기초한 사례를 고찰하는 것이 중요
- 동 보고서는 국가가 관리하는 신뢰 가능한 데이터(authoritative data)와 시민 생성 데이터(citizen-generated data)를 결합한 사례로 주소 데이터셋을 결합하는 실험을 통해 기술적 가능성과 제도적 한계점을 검토
 - 실험 결과, 신뢰 가능한 데이터의 정확성과 시민 생성 데이터의 빠른 업데이트라는 장점을 취할 수 있는 기술적 가능성을 확인
 - 그러나 정부 기관의 신뢰 가능한 데이터의 라이선스와 민간 데이터의 라이선스가 상이함에 따라 데이터 활용에 있어 법적·제도적 장애물이 존재
- 더불어 IoT 센서를 이용한 도시 지역 대기질 개선을 위한 민관 데이터 협력 사례를 거버넌스, 운영, 과학, 기술 및 인재의 역량이라는 분석 프레임워크에 기초해 분석한 결과, 거버넌스와 이를 관리하는 역량 구축 필요성 등이 제기
 - 수집된 데이터의 투명한 관리 및 접근, 데이터의 수집과 분석에 필요한 파트너십 구축과 분석 결과의 공유라는 요건들은 분석 대상 사례 모두에서 관찰됨
 - 그러나 거버넌스, 운영, 과학적 요건이나 기술적·인적 역량 측면에서 모두 한 가지 이상의 한계를 가지고 있는 것으로 나타남
 - 정책 측면에서 보다 일반적인 IoT 거버넌스 프레임워크의 개발, 프로젝트 참여자들의 역량 구축, 데이터 관리 체제의 확립 및 모범 사례 관련 연구의 확대가 요구됨
- 우리나라는 '데이터3법' 개정 시행('20.8월)으로 이중 데이터의 결합이 활성화되고, '22.4월부터 시행될 '데이터 산업진흥 및 이용촉진에 관한 기본법' 제정으로 데이터 활용 및 관련 산업진흥이 본격화
 - 이를 통해 새로운 데이터 산업 생태계가 형성될 수 있는 제도적 기반이 마련되었으나, 특히 도시 문제 해결에서 큰 효용성과 혁신성을 보이는 시민 생성 데이터 활용의 측면에서는 보다 많은 정책적 노력이 필요
 - 본 보고서의 유럽 데이터 샌드박스 사례와 데이터 결합 실험, 데이터 협업사례 등을 국내 민간·공공데이터 결합 및 협업 활성화를 위한 제도 마련에 참고할 수 있음

PART

I

검토 배경

- ▶ 유럽 경제와 사회의 디지털 전환은 EU 집행위원회의 가장 중요한 의제로, '20년 유럽 데이터 전략(European strategy for data)을 통해 유럽 공통 데이터 공간 구축을 추진함으로써 그 기틀이 마련됨
- 유럽 데이터 전략은 유럽에서 데이터의 자유로운 이동을 보장하는 유럽 단일 데이터 시장(European single market for data)을 형성함으로써 데이터 주도 혁신을 자극함과 동시에 경제와 사회를 위한 가치 창출을 추구
 - 환경, 농업, 산업, 보건 및 수송 등의 전략 분야별로 특화된 데이터 공간(domain-specific data spaces)을 기반으로 하는 유럽 공통 데이터 공간(European data space)을 구축하는 것이 목표
 - 다양한 출처의 데이터 조합 및 결합은 유럽 공통 데이터 공간을 성공적으로 구축하는 데 가장 중요한 요소로서, 데이터의 공유와 사용을 위한 첨단 기술과 혁신적 접근법이 이를 가속화하는 주요한 동인으로 작용
- ▶ 동 보고서*는 유럽 공통 데이터 공간의 성공적인 실행과 부가가치 제고를 위한 정책적 의제를 지원하는 데이터 주도 혁신과 관련된 신기술과 새로운 접근 방법에 대한 이해를 도모하고 가능성을 모색
 - * Joint Research Center(JRC), Emerging approaches for data-driven innovation in Europe, 2022.
 - 오늘날의 기술 지형이 매우 역동적으로 변화한다는 점을 감안할 때, 정책적 의제를 설정하고 사회 기술적 요인을 조사하기 위해서는 첨단 기술과 다양한 도구를 사용하는 일련의 실험을 가능하게 하는 샌드박스 접근법에 무게 중심을 두고, 다음의 세 가지를 고찰
 - 기술적 관점에서 서로 다른 정보 출처의 데이터를 수집, 결합, 공유하는 새로운 접근법 제시
 - 지역 사회와 공공 부문 데이터 주도 혁신의 주요한 조직적·사회적 특징
 - 첨단 기술과 아키텍처, 기술 표준 및 접근법의 성격을 사회·경제·기술적 관점에서 다각도로 접근하는 분석을 통해 정책의 범위와 방향을 제시
- ▶ 본 호에서는 데이터의 수집과 이를 결합하는 샌드박스의 두 가지 사례를 살펴보고, 정책적 시사점을 도출
 - 첫 번째로, 시민들이 자발적으로 구축하고 사용하는 OpenStreetMap (OSM) 데이터와 핀란드, 네덜란드의 행정 지리 정보 데이터를 결합하는 실험을 통해 데이터의 개선 가능성과 효율적인 데이터 사용을 저해하는 요인을 고찰
 - 두 번째로, 최근 빠르게 발전하고 있는 IoT 센서를 이용하여 대기질에 영향을 미치는 다양한 오염 물질 관련 데이터를 축적하고 사용하는 사례 분석을 통해 앞으로 대기질 관리에 IoT 기술을 사용할 경우 어떤 요건들이 필요한지를 검토



공공 데이터와 시민 데이터의 결합-지리정보 데이터

1 공공 데이터와 시민 데이터의 결합 필요성과 연구의 배경

- ▶ EU는 정책 결정을 개선하는 시민 생성 데이터(citizen-generated data)*의 잠재력을 인정하고, 특히 지리공간 영역에서 현대적인 지리공간 데이터 생태계 형성에 기여하는 중요한 요소로 인식
 - 시민 생성 데이터는 유럽 데이터 전략에서 언급된 데이터 이타주의(data altruism)** 개념에서 데이터 축적뿐 아니라 EU 일반개인정보보호규정(General Data Protection Regulation, GDPR, '16년 제정) 준수의 측면에서도 중요한 역할을 수행
 - * 개인이 제공하거나 생성한 데이터
 - ** 정보주체가 공익 목적을 위해 보상 없이 개인정보 또는 비개인정보 사용을 허락하는 것
- ▶ 오픈스트리트맵(OpenStreetMap, OSM)은 전 세계 지리 데이터베이스를 만들기 위한 시민이 주도하는 협력 프로젝트의 성공 사례로 꼽힘
 - '04년부터 시작된 OSM은 '21.11월 기준 170만 명에 달하는 정보 기여자를 확보하고 있음. 공개 데이터베이스 라이선스(Open Database License, ODbL)를 통해 자유로운 사용과 접근이 가능하며, 내용이 자세하고 풍부해 각국 정부는 물론 민간 기업, 비영리 단체에서 광범위하게 사용
 - OSM을 운영하는 기관은 비영리 단체인 오픈스트리트맵재단(OpenStreetMap Foundation, OSMF)으로, 서버와 같은 인프라와 서비스를 관리하고 프로젝트와 관련된 지부와 다양한 워킹 그룹들과의 협업을 진행
- ▶ 특히 도로, 건축물, 토지 사용 용도와 같은 지리 정보 분야에서 다양한 연구 및 실험 결과가 축적되고 있으며, 세계 각국에서 OSM 데이터와 각국 정부기관의 공공 데이터를 동시에 사용하거나 결합하는 추세
 - 일반적으로 데이터 결합에서는 데이터 생성/업데이트, 인코딩(encoding), 시맨틱 콘텐츠(semantic contents), 품질이 중요하며, 이 중 특히 품질에 관한 연구가 선행되어야 함
- ▶ 따라서 OSM데이터와 이질적으로 구축된 데이터, 주로 각국 정부의 지도제작기구(National Mapping Agencies, NMAs)의 신뢰 가능한 데이터(authoritative data)*의 결합 문제에 관해서는 OSM데이터의 품질 측면에서 다수의 연구가 수행됨
 - * 정부 기관이나 제공자가 제공하거나 생성한 (지리) 데이터
 - Du et al.(2012), Abdolmajidi et al.(2014), Brovelli et al.(2017) 등은 OSM 데이터를 신뢰 가능한 데이터와 비교함으로써 OSM 데이터의 품질을 평가하는 방법론을 개발

- Barron et al.(2014), Minghini and Frassinelli(2019) 등은 OSM 데이터 자체의 이력(업데이트 빈도, 데이터 기여자의 수 및 특성 등)을 고찰
- 그러나 이러한 연구들은 로컬 또는 지역적으로 소규모일 뿐만 아니라 특정 신뢰 가능한 데이터셋에만 제한되어 있으며 특정 데이터 모델에 종속적인 절차를 기반으로 하는 경우가 많기 때문에 광범위한 활용을 위해 일반화하고 복제하기에는 한계가 있음
- ▶ 상대적으로 NMA의 신뢰 가능한 데이터와 OSM 데이터의 결합 가능성 및 효용성을 고찰한 연구는 소수에 불과하지만 다음과 같은 연구들이 두 데이터셋 간 결합이 데이터셋을 더욱 풍부하게 만든다는 결론을 도출
 - OSM 데이터와 영국 육지 측량부의 벡터 지도(British Ordnance Survey's Vector Map) 도로망을 비교한 Pourabdollah et al.(2013) 등의 연구는 OSM 데이터와 신뢰 가능한 데이터의 결합이 데이터를 보다 풍부하게 만든다는 결론을 도출
 - 브라질의 도로망에 관한 신뢰 가능한 데이터에서 매핑된 지형지물에 대한 OSM 데이터 결합을 분석한 Silva et al.(2021)의 연구에서도 데이터가 오래되거나 누락된 지역에 대한 OSM 데이터의 잠재력을 확인
 - 이외에도 베트남과 파키스탄과 같은 인근 국가의 OSM 데이터를 동태적으로 결합하는 광범위한 연구 결과를 제시한 Zhou et al. (2015)의 연구, 특정 건물 용도 관련 태그가 달린 OSM 데이터에서 정보를 추출하여 신뢰 가능한 데이터의 내용을 풍부하게 만드는 가능성을 탐구한 Kunze and Hecht(2015)의 연구 등이 있음
- ▶ 본 장에서는 유럽 각국의 NMA 데이터셋과 OSM 데이터셋의 결합을 촉진하는 요인과 저해하는 요인을 식별하기 위해 핀란드와 네덜란드의 전국 주소 데이터셋 결합 실험 사례를 고찰
 - 궁극적으로는 의미론적인 측면 뿐 아니라 기술적·조직적·법적인 측면의 상호 운용성에 대한 정책적 권고를 도출하는 동시에 유럽 공통 데이터 공간 확립에 대한 가이드를 제시하는 것이 목표
 - 이 목표를 달성하기 위해 핀란드와 네덜란드의 NMA와 OSM의 전국 주소 데이터셋 결합 가능성을 시험하기 위한 '지리공간용 무료 오픈 소스 소프트웨어(Free and Open Source Software for Geospatial, FOSS4G)'에 기반한 실험 결과를 통해 기술적 시사점과 데이터 결합 과정의 장·단점을 식별

2 데이터 결합 실험을 위한 데이터셋의 특성과 접근법 및 실험 결과

- ▶ OSM과 결합할 신뢰 가능한 데이터셋을 선택하는 것은 신뢰 가능한 데이터의 전환과 데이터 재사용 단계에서 중요한 역할을 하는데, 본 연구에서는 다음과 같은 이유로 주소 데이터를 선택
 - 주소 데이터는 단순한 데이터 모델을 사용해 점(points)으로 모델링된다는 점 외에도 다양한 애플리케이션에 대한 레퍼런스 데이터셋을 대표함

- 정부가 생산하고 유지하는 핵심적인 데이터셋일 뿐 아니라 OSM 기반의 길 찾기나 응급 애플리케이션의 수를 감안할 때 OSM 생태계 내에서도 가장 중요한 데이터셋임
- 아울러, 주소 데이터는 행정정보 데이터 중 업데이트에 가장 많은 비용이 들고 자주 변경되지 않기 때문에 OSM 데이터와 결합함으로써 가장 높은 효용을 기대할 수 있다는 점도 고려됨
- ▶ OSM 데이터는 기하학적 구성요소(geometric component)와 시맨틱 구성요소(semantic component)를 결합한 단순한 개념 모델을 사용하여 구성됨
 - 기하학적 구성요소는 노드(nodes), 웨이(ways), 관계(relations)라는 3가지 속성을 가지고 있음
 - (노드) 관심 지점(point of interest), 나무, 도로 표지판, 벤치 등과 같은 독립된 위치의 피처(feature)¹⁾를 나타내며 해당 위치의 위도와 경도를 나타내는 점(point)으로 표현됨
 - (웨이) 도로, 강과 같은 최대 2,000개의 노드가 정렬된 선형 피처와 건물, 대지와 같은 영역 또는 다각형 형태의 피처를 모두 나타냄
 - (관계) 2,000개 이상의 노드를 사용해 호수와 같은 선형적 또는 영역적 피처 모두를 모델링하거나 교통망과 같은 두 개 이상의 기하학적 유형 간의 관계를 묘사하는 데 사용되는 데이터 구조
 - 시맨틱 구성요소는 하나 이상의 데이터 속성과 태그로 구성되며, 본 연구에서 사용되는 주소 데이터와 관련된 OSM 데이터 태그는 [표 1]과 같음
 - 주소를 식별하기 위해 사용되는 모든 태그의 키(key)²⁾는 'addr:'이라는 프리픽스(prefix)를 사용
 - 본 연구에서는 OSM 데이터와 핀란드 및 네덜란드의 주소 및 건축물에 대한 NMA 데이터를 기초로 분석

| 표 1 | 실험에 사용된 주소 관련 OSM 태그

OSM 태그	주요 내용
addr:country	주소의 국가 번호
addr:city	주소의 도시 이름
addr:street	주소에 면한 도로의 이름
addr:housenumber	주소의 건축물 번호

- ▶ 실험에 사용된 핀란드의 전국 토지 조사(National Land Survey, NLS)는 NMA에서 관리하는 데이터로 최근 새롭게 구축된 OGC API³⁾를 통해 접근이 가능
- OGC API는 웹상에서 지리공간 벡터 데이터를 노출하고 사용하기 쉽고, 개발자 친화적으로 개발된 표준이며, 최근 EU공간정보인프라(Infrastructure for Spatial Information in Europe, INSPIRE)*가 규정한 주소 데이터 사양을 준수

1) 관찰 대상의 개별적이고 측정 가능한 특징을 의미하는 용어

2) 데이터베이스에서 조건에 만족하는 자료 구조(튜플, tuple)을 찾거나 순서대로 정렬할 때 기준이 되는 속성

3) 개방형 공간정보 컨소시엄(Open Geospatial Consortium, OGC)은 공간정보 콘텐츠와 서비스, 데이터 처리와 교환 관련 표준 개발과 지원을 하는 비영리국제기구로서, 북미, EU 등 많은 국가에서 OGC의 공간정보 표준을 공간정보 인프라 개발에 사용하고 있음. OGC API는 OGC가 규정한 컴퓨터나 컴퓨터 소프트웨어 간의 지리정보 교환 방식 규약을 의미

* 지속 가능한 개발을 지원하기 위해 지리공간정보에 보다 쉽게 접근하고 상호운용할 수 있도록 설계된 유럽의 공간정보인프라를 구축하기 위한 EU의 이니셔티브

- 또한 크리에이티브 커먼즈(Creative Commons)⁴⁾의 CC BY 4.0 라이선스* 하에서 사용 가능

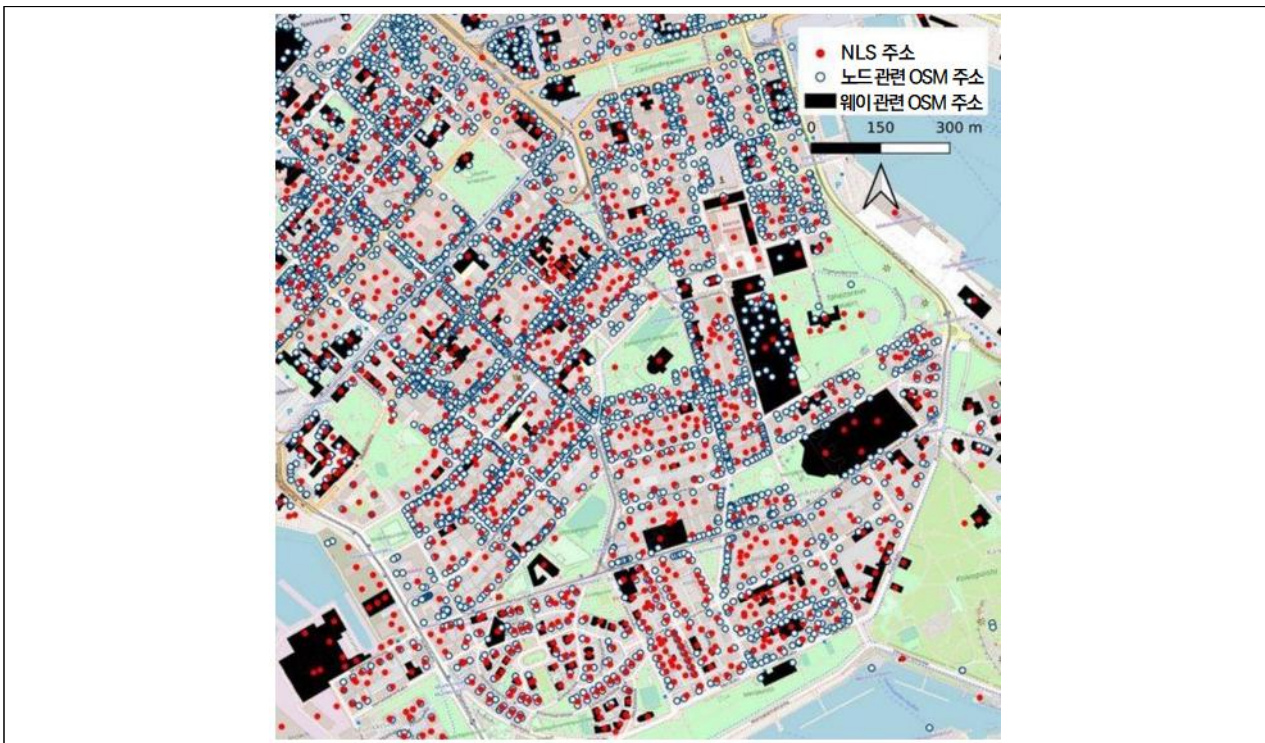
* CC BY 4.0은 저작자 표시 조건 하에서 복제, 배포, 전시, 공연 및 공중 송신이 가능하도록 하는 라이선스로, 리믹스나 변형, 2차 저작물의 창작, 영리 목적으로 사용가능한 수정 권한을 부여('13.11월 개정)

| 표 2 | 주소 관련 NLS의 속성

NLS 속성	주요 내용
component_ThoroughfareName_name fin	핀란드 주소의 거리 이름
component_ThoroughfareName_name swe	스웨덴 주소의 거리 이름
component_ThoroughfareName_name sme	스칸디나비아 북부 Sami 지역 주소의 거리 이름
locator_designator_addressNumber	해당 주소의 건축물 번호
component_AdminUnitName_4	주소가 속한 도시의 코드
component_AdminUnitName_1	주소의 국가명

- 핀란드 헬싱키의 OSM 데이터와 NLS 데이터를 지도에 표시한 결과에 따르면, 해당 지역에서 NLS 기반 주소보다 OSM 데이터의 주소가 훨씬 많다는 점을 알 수 있음

| 그림 1 | 헬싱키 일부 지역의 OSM 데이터와 NLS 데이터의 분포



- ▶ 실험에 사용된 네덜란드의 주소 및 건축물 대장(Basisregistratie Adressen en Gebouwen, BAG)은 다음과 같은 특징을 가지고 있음

4) 크리에이티브 커먼즈(Creative Commons)는 저작권의 부분적 공유를 목적으로 설립된 미국의 비영리 조직으로, 크리에이티브 커먼즈가 발행한 라이선스의 종류에 따라 요구 조건을 준수하면 저작물 배포를 허용

- 모든 지자체들은 이 BAG에 대한 정보 출처를 책임지고 있으며, 주소 및 건축물 관련 정보는 국립 BAG 기구(National Facility BAG, LV BAG)에 의해 데이터화됨
 - 본 연구와 관련된 네덜란드의 BAG 데이터는 INSPIRE의 Geoportal에서 원본 BAG 데이터와 INSPIRE 조화 주소 데이터(harmonised address data)라는 두 가지 데이터로 조회 및 다운로드 가능⁵⁾
 - * 원본 BAG 데이터는 주소의 변동이력까지 모두 포함한 확장형 XML 형태의 데이터로 매우 복잡하며, INSPIRE의 조화 주소 데이터는 INSPIRE의 사양에 부합하지만 관련 인프라의 보호를 위해 속도를 제한
 - 오픈소스 프로젝트인 NLExtract로 BAG 데이터는 텍스트 파일인 CSV 형태의 데이터로 다운로드 가능
 - * NLExtractor는 네덜란드의 BAG 지리 벡터 데이터를 공간 데이터베이스로 전환하는 소프트웨어로, 오픈소스 공개 웹사이트인 Github에 공개되어 있음

| 표 3 | 주소 관련 BAG의 속성

BAG 속성	주요 내용
openbareruimte	공공장소
huisnummer	주택 번호
huisletter	주택 문자
huisnummertoevoeging	주택 추가 번호
postcode	우편번호
woonplaats	지역
gemeente	지자체 이름
provincie	지역 이름

- ▶ OSM 데이터와 신뢰 가능한 주소 데이터셋(NLS와 BAG 데이터셋)에서 관련 정보를 추출하여 하나의 데이터셋으로 결합하기 위해 다음과 같은 전처리 단계를 수행
 - INSPIRE 호환 NLS 주소 데이터셋과 BAG 데이터셋은 OSM 데이터에 비해 풍부하기 때문에 이들을 OSM 데이터 모델로 변환할 수 있는 가장 간단한 결합 접근 방식을 선택
 - 핀란드와 네덜란드의 NMA가 제공하는 신뢰 가능한 데이터와 OSM 데이터의 결합은 QGIS⁶⁾ 그래픽 모델러 툴(QGIS Graphical Modeler tool) 기반 처리 알고리즘을 적용
 - 핀란드의 NLS 데이터와 OSM 데이터를 결합하는 과정에서 OSM의 핀란드 주소 객체에 'addr:city' 키(key)가 포함되지 않은 경우가 많아 EU 통계국(Eurostat)에서 제공하는 지역 행정 단위(Local Administrative Units, LAU) 데이터셋에서 필요한 정보를 추출
 - LAU 데이터셋과 결합한 OSM 데이터를 이용하여 다음과 같은 데이터 매핑 관계를 구축함

| 표 4 | INSPIRE/NLS 데이터와 OSM 데이터 모델 데이터 매핑

일반명	INSPIRE/NLS 속성	OSM 속성	비고
Street name	component_ThoroughfareName_name	addr:street	가능한 핀란드 지명을 주로 사용하고, 그렇지 않을 경우 스웨덴 지명을 사용
Address number	locator_designator_addressNumber	addr:housenumber	-

5) <https://inspire-geoportal.ec.europa.eu/results.html?country=nl&view=details&theme=ad>

6) 무료 오픈소스 소프트웨어를 기반으로 만든 전문 지리정보시스템 애플리케이션(<https://qgis.org/ko/site/about/index.html>)

일반명	INSPIRE/NLS 속성	OSM 속성	비고
City name	component_AdminUnitName_4	addr:city	숫자는 LAU 코드 ID
Country	component_AdminUnitName_1	addr:country	-

- 네덜란드의 BAG 데이터는 OSM 커뮤니티의 정기적인 BAG 데이터 결합으로 핀란드보다 적은 전처리를 거쳤으며, OSM 데이터의 품질과 완결성이 제고됨
 - BAG 데이터의 속성과 OSM 데이터 모델을 연계하기 위해 다음과 같이 매핑함

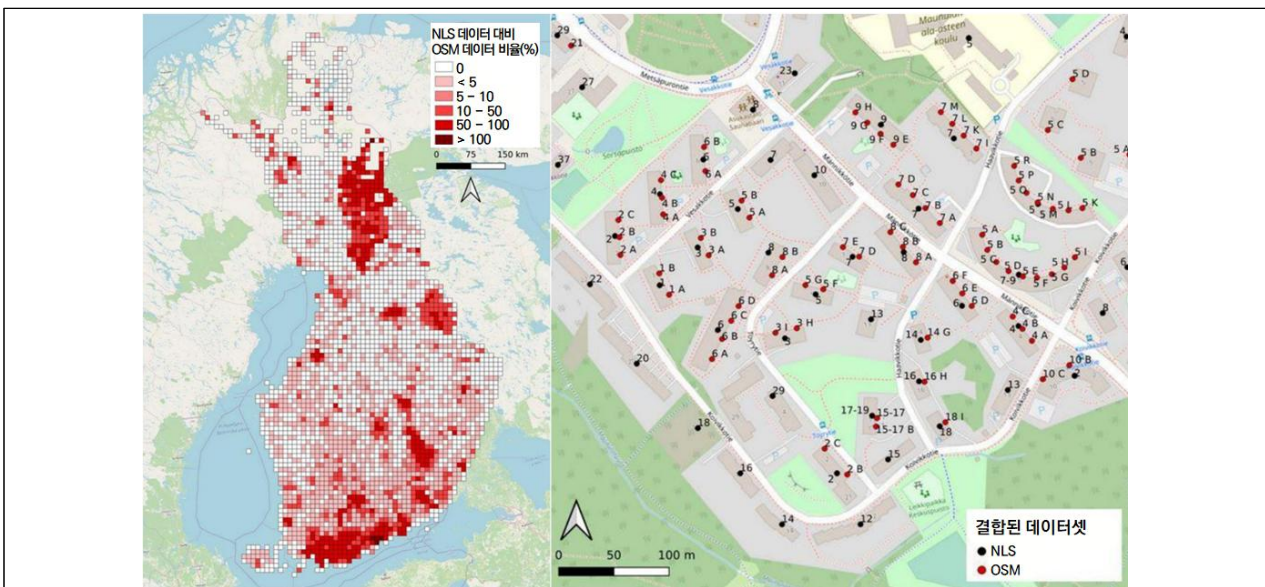
| 표 5 | BAG 데이터와 OSM 데이터 모델 데이터 매핑

일반명	BAG 속성	OSM 속성
Street name	openbareruimte	addr:street
Address number	Huisnummer + huisletter + huisnummertoevoeging	addr:housenumber
City name	woonplaats	addr:city

- ▶ 핀란드의 NLS 데이터와 네덜란드의 BAG 데이터를 OSM 데이터와 결합한 결과는 다음과 같이 도출됨

- 핀란드의 NLS 데이터를 OSM 데이터 모델에 적용한 데이터셋에는 약 330만 개의 주소가 포함된 반면, OSM 데이터에는 주소 수가 50만 개였음
 - 10x10 km 크기의 EEA 레퍼런스 그리드 셀⁷⁾로 표현된 지도상에서 두 데이터를 비교한 결과에 따르면 NLS 데이터셋에서 하나 이상의 주소를 가진 격자의 약 63%가 OSM 데이터셋 상에서 주소를 포함하지 않은 것으로 나타남(그림2(좌))의 하얀색 네모로 표시된 부분)

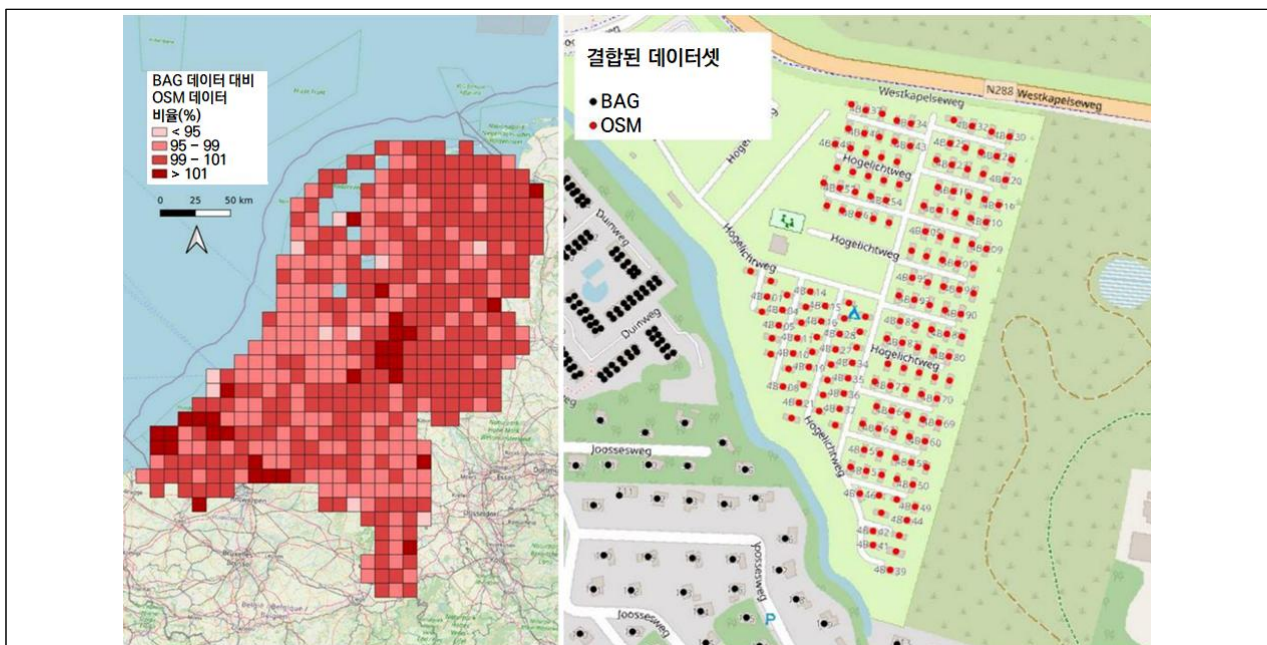
| 그림 2 | 핀란드의 NLS 데이터와 OSM 데이터의 비교 결과(좌)와 헬싱키 지역의 데이터 결합 결과(우)



7) 그리드 셀은 지리정보시스템 또는 원격탐사 영상 값이나 속성을 가지는 단위격자로서 도형을 임의의 격자 형태로 나타내기 위한 최소 단위(사용자가 임의로 크기 조절 가능)이며, 유럽환경청(EEA) 레퍼런스 그리드는 INSPIRE 지리 그리드 시스템의 권고에 기초해 국경, 버퍼(점이나 선 혹은 면으로부터 특정한 거리 이내에 포함되는 영역), 또는 배타적 경제 수역 등을 포함

- 핀란드의 6대 대도시 중 4대 도시(Helsinki, Espoo, Vantaa, Turku)에서 OSM 데이터의 완결성은 75~97%에 달하는 것으로 나타남
 - 이러한 결과는 OSM 매퍼(mapper)들이 인구 밀집 지역을 방문할 때 정보를 추가하고 업데이트하는 경향이 있음을 보여주는 기존의 연구 결과를 재확인
- OSM 데이터의 경우 NLS 데이터에 비해 도로나 도시의 이름이 잘못 입력되어 있는 경우가 상당히 많다는 것이 약점으로 흔히 지목되지만, OSM 데이터가 보다 정밀하고 최신의 정보를 가지고 있기 때문에 NLS 데이터를 개선하는 경우도 많음
- 네덜란드의 OSM 사용자 커뮤니티는 정기적으로 BAG 데이터를 OSM 데이터베이스에 결합하는 작업을 진행하고 있기 때문에 네덜란드의 신뢰 가능한 데이터인 BAG가 940만 개인 것에 비해 OSM 데이터는 8만 개 정도로 격차가 작은 편임
 - BAG 데이터셋과 OSM 데이터셋의 주소 수를 10x10 km 크기의 EEA 레퍼런스 그리드 셀에서 비교한 결과, BAG 주소 수 대비 OSM 주소 수 비율이 95% 이하의 비율을 보이는 셀은 약 2%에 불과했으며, 99~101%의 셀은 55% 이상인 것으로 나타남
 - 네덜란드의 경우에는 두 데이터의 결합한 데이터셋은 960만 개의 주소를 포함하며, 이 중 18만개가 OSM 데이터베이스에만 존재한 것으로 나타남
 - BAG보다 OSM에서 더 많은 주소를 표시한다는 사실은 OSM 매퍼들이 등록한 캠핑 지역의 오두막 주소가 BAG에서는 데이터베이스 개정 시 삭제되었기 때문

| 그림 3 | 네덜란드의 BAG 데이터와 OSM 데이터의 비교 결과(좌)와 캠핑 지역의 데이터 결합 결과(우)



3 법과 제도 측면의 문제점: 데이터 라이선스의 문제

- ▶ 서로 다른 데이터셋을 사용하고 결합할 때, 데이터의 작성자나 저작권자의 지침에 따라 데이터를 올바르게 재사용할 수 있으려면 데이터 정책 및 라이선스가 가장 중요한데, 라이선스 비호환성 문제가 발생하기도 함

- INSPIRE 지침은 EU회원국이 제공한 데이터의 공개 여부나 라이선스 종류를 규정하지는 않으나, INSPIRE Geoportal에서 구할 수 있는 일부 데이터셋은 크리에이티브 커먼즈 라이선스, 특히 CC-BY 4.0 라이선스 하에서 제공됨
- OSM 데이터가 갖고 있는 ODbL 라이선스는 데이터 사용자가 동일한 라이선스 하에서만 데이터를 공유해야 한다는 것을 요구 조건으로 내세움
- 비호환성으로 인해 CC-BY 4.0 데이터를 OSM에서 직접 재사용할 수 없는 문제가 발생
- ▶ 공공기관과 OSM 간의 완전한 상호 협력을 위해서는 양방향 정보 교환이 촉진되어야 하는데, CC-BY 4.0과 ODbL의 경우 쉽게 해결되는 문제가 아님
- OSMF의 라이선스 워킹그룹은 CC-BY 데이터베이스 및 데이터의 라이선스 제공자로부터 ‘명시적인’ 허락을 추가로 얻어야 OSM에서 데이터를 사용할 수 있다는 입장
- 반면 각국의 NMA들은 CC-BY 4.0 데이터와 OSM 데이터의 결합 과정에서 ODbL 규약을 준수하는 데이터를 ‘자유롭게 삽입하거나 제거’하기를 원해 OSMF와 입장 차이가 존재함
- ▶ 신뢰 가능한 데이터와 OSM 데이터의 결합을 추진하고 있는 스페인, 프랑스, 이탈리아에서 발생한 법제도적 문제의 사례와 전문가들이 제시한 해결방안은 다음과 같음
- (스페인) 국립지리정보센터(Centro Nacional de Información Geográfica, CNIG)가 보유한 교통망 데이터와 OSM 데이터 결합을 통해 도로 방향에 관한 데이터를 강화할 수 있었으나, 두 데이터셋 간의 라이선스 규약 간에 불일치성이 발견됨
 - 특히 스페인에는 OSMF를 대표하는 지부가 없기 때문에 스페인의 주요 이해관계자들과 OSM 사용자 커뮤니티 간의 지속적인 교류가 어렵다는 점도 문제로 작용
 - CNIG는 향후 CNIG 데이터에 대한 접근성과 사용 가능성을 직접 홍보하기 위해 OSM과 같이 개방적이고 글로벌한 데이터베이스에서 CNIG 데이터의 가용성을 제고할 예정
- (프랑스) 국립지리산림정보연구소(French National Institute for Geographic and Forestry Information, IGN)데이터와 OSM 데이터의 결합 과정에서 라이선스 비호환성 문제가 발생
 - 이러한 라이선스 규약 간의 불일치성을 해소하기 위한 방법으로 NMA 데이터에 ODbL 라이선스를 채택하거나 OSMF 측과 OSM 데이터에 ODbL 외에도 CC-BY 4.0 라이선스를 이중으로 적용하는 방안을 논의하는 두 가지 방법이 제시됨
- (이탈리아) 피에몬테 지역 환경보호지역기구(Regional Agency for the Protection of the Environment - ARPA Piedmont)에서도 행정 지리 정보의 업데이트를 위해 OSM과 지역 지리 관련 기구 간의 협력 가능성을 논의 중
 - 이탈리아에서도 OSM의 ODbL과 CC-BY 4.0 라이선스 간의 불일치성이 중요한 문제로 부상
 - 라이선스 간의 불일치성을 없애기 위해 건축물의 건축 연도나 다른 특성 등 특정 관심 분야에 존재하는 데이터들 간의 인덱스를 별도로 계산하는 방안을 제시

PART III

데이터 공급을 위한 민관 협력- 도시 지역 대기질 데이터

1 배경

- ▶ 스마트 디바이스와 IoT 센서는 팬데믹 상황에 대한 실시간 정보 수집, 기온의 변화에서 화물 운송 추적에 이르기까지 사회의 각 부문들이 서로 긴밀하게 연계되고, 상황 변화에 신속하게 대응하는 방법으로서 급부상
 - 특히 팬데믹 상황에 대응하기 위한 정책 수립과 대응에 있어서 IoT 기술이 생성하는 실시간 데이터는 데이터의 축적과 사용 측면에서 공유가 우선시되는 새로운 방향을 제시
 - 또한 책임성 있게 설계된 데이터 협업은 오늘날의 데이터 공급자와 수요자 간의 불균형을 시정할 가능성도 제시
 - 과거에는 통신 데이터, 위성 사진, 소셜 미디어 데이터, 금융 데이터와 같은 다양한 데이터들이 서로 고립되어 축적된 반면, 오늘날에는 관련 기관이나 개인들에게 자유로운 사용과 분석을 가능하게 만드는 환경이 조성되면서 새롭고 혁신적인 사회문제의 해결방안 도출의 가능성을 제고
 - 협력적인 데이터의 수집과 사용에 대해 다양한 운영 방향과 거버넌스 방식이 제시되면서 데이터의 개방성이나 협력의 수준에 대해 데이터 협력, 데이터 중개자의 신뢰성, 데이터 연구 파트너십, 데이터 포털의 보완성, 지능형 제품에 의한 접근 가능성 등 새로운 논의점들이 제시되는 상황
- ▶ 한편 전 세계적으로 도시화가 진행되고 대기질이 나빠짐에 따라 도시를 관할하는 지자체는 물론 시민 단체들은 대기 오염을 줄이기 위해 다양한 역할을 모색 중
 - 도시 지역은 지구의 표면에서 차지하는 비중이 2%에도 미치지 못하지만 도시 지역에서 발생하는 교통 수요, 산업 및 에너지 수요로 인해 전 세계 온실가스 배출의 70%를 차지하는 것으로 나타남
 - 최근 더 깨끗한 대기를 위한 인프라 투자와 서비스가 활성화되면서 대기질 모니터링과 진단 결과 간의 격차가 크게 축소
 - 데이터 과학자들이 대기질과 관련된 다양한 데이터에 대해 접근하기가 용이해지고, 대기질 관련 데이터의 품질이 높아지면서 관리의 효율성이 증대됨
- ▶ IoT센서는 도시의 대기질을 개선하려는 정책 입안자들에게 실시간으로 수집된 신뢰도 높은 데이터를 제공
 - IoT 센서의 발전으로 대기 중의 미세먼지, 유해 가스와 같은 여러 오염원의 모니터링이 가능해졌으며, 대기질 관련 데이터 외에 대기압이나 기상 데이터, 소음 및 진동 등의 다양한 데이터를 수집 가능
 - 최근에는 카메라 기술이 발전되면서 영상으로도 데이터를 축적하는 기술이 개발되고 있음

- 이러한 기술 발전에 힘입어 도시 지역의 대기질 뿐 아니라 글로벌 기후 변화를 광범위하게 추적하고 분석하기 위한 공공-민간의 IoT 대기질 관련 계획들이 추진되고 있음
 - 디지털 유럽 프로그램(Digital Europe Programme, DEP)은 기후변화 관련 데이터 공간의 설립을 지원하고 보다 스마트한 사회를 형성하는 방안을 고려하고 있으며, 이러한 노력의 결과로 현재 유럽 그린딜 데이터 공간의 구축이 추진되고 있음
 - 대기질 센서의 가격이 낮아짐에 따라, 유럽에서는 저렴한 IoT 센서를 이용하는 모니터링 장비가 비영리단체들에 의해 보급되면서 클라우드 소싱이나 시민들의 적극적인 참여에 기반한 대기질의 모니터링이 확산되는 추세

▶ IoT 기반 센서의 확산으로 대기질과 관련된 데이터의 수집은 크게 개선되었지만, 이렇게 수집된 데이터의 사용 방법에 대한 다양한 문제와 개선 필요성이 제기됨

- 센서들이 수집하는 데이터들은 종종 개인이 저장하거나, 특정 데이터베이스에 고립되어 저장되어 접근이 어려울뿐더러 개인의 프라이버시 보호나 데이터와 관련된 윤리 및 책임 등의 문제가 존재
- 데이터 협업과 관련되어 제기되는 중요한 이슈들은 다음과 같이 정리될 수 있음
 - (의사결정을 위한 데이터의 정밀도와 정확성) 대기질 관련 데이터의 다양성과 필요한 정보 간의 격차를 상쇄하는 것은 환경과 도시 관련 계획 및 서비스에 대한 정책 수립 과정에서 정부가 직면하는 다양한 문제를 해결하는 방안이 될 수 있지만, IoT 데이터들이 대부분 민간 영역에 집중되어 있어 공공 부문의 접근성 확대가 필요
 - (데이터 기반 연구와 예측 및 전망) 다양한 출처에서 생성되는 데이터의 결합과 데이터 접근성 제고를 통해 보다 많은 이해관계자들이 문제 해결을 위한 솔루션을 제시하거나 적극적인 정책 수립에 참여하게 됨
 - (데이터 주도의 모니터링과 평가) 많은 이해관계자가 반복적이고 신속하면서 효율적으로 환경 정책의 결과를 모니터링하고 평가할 수 있도록 IoT 센서가 생성하는 데이터의 공유가 필요
 - (이해관계자 간의 협력적 문화) 데이터 협업 구축은 기업과 정부, 연구기관과 시민 간의 상호 신뢰와 협력을 기반으로 대중의 이해를 제고하고 새로운 정책 도입을 가능하게 함
- IoT 기반의 데이터 협업 구축과 이에 기반한 실험에 대한 문헌 검토 결과, 지속가능성 측면에서 제시되는 문제는 다음과 같음
 - (비용과 규모의 확대) 많은 도시에서 민간 디바이스 제조업체, 클라우드 서비스 공급업체, 데이터 플랫폼 등의 다양한 IoT 시장 참여자들 간 파트너십을 형성하기 위해 노력하고 있지만, 재정적 지속가능성을 담보하지 못하거나 혁신 기술의 도입을 확대하는 전략이 부재
 - (상호운용성의 결여와 데이터의 파편화) 시장 참여자들과 도시의 정책 입안자들은 상호운용성의 결여로 인한 데이터의 파편화 때문에 데이터의 수집·저장·공개 과정에서 다른 이해관계자들과 상호 작용하고, 미래에 대한 통찰을 공유하는 데 현실적인 어려움을 경험
 - 결국 이러한 어려움은 다양한 출처를 가진 데이터에 기반하여 정책을 수립해야 하는 정책 입안자들의 역량을 제한하는 결과를 야기
 - (데이터의 품질과 선택) 저가 IoT 센서들이 출시됨에 따라 데이터의 품질 관리에 대한 문제가 발생

- 공식 데이터가 어떻게 이러한 센서 데이터로 보완되어야 하는지에 대한 문제 역시 존재함. 대기질과 관련된 IoT 센서 데이터의 수집 과정에서 EU대기질지침(EU Air Quality Directives)⁸⁾을 준수해야 하는데, 이를 위해서는 공공 부문의 적극적 참여와 데이터 기준의 수립을 위해 파트너십 구조를 형성해야 함
- (포용적 거버넌스와 투명성) 도시 거주자들과 다른 이해관계자들은 IoT 기기를 통해 데이터를 수집하는 과정에서 어떤 일이 발생하는지를 거의 알지 못하는 불투명성이 존재할 수 있는데. 이러한 문제는 IoT의 가치사슬 전체 뿐 아니라 도시의 감시 및 데이터 정의와 관련하여 시민들의 우려를 증폭하고 있음. 따라서 IoT에 기반하는 데이터의 수집과 사용 과정들이 본연의 목적에 충실할 수 있도록 지속적인 시민들의 감시와 책임 소재를 분명하게 밝히는 메커니즘이 존재해야 함

2 도시의 대기질 모니터링을 위한 IoT 데이터 협업

- ▶ IoT 센서를 사용해 대기질을 모니터링하기 위한 데이터 협업의 기반을 마련하기 위해 분석 프레임워크는 거버넌스, 운영, 과학, 기술 및 인재의 역량이라는 4가지 측면에서 설계될 수 있음
- (거버넌스) 책임성, 참여, 투명성, 문제 정의 등을 통해 위험을 규명하고 완화하며 정당성을 제공하는 거버넌스 절차와 구조

| 표 6 | 거버넌스 측면의 요건

구분	설계 요건과 성공 요인
1. 문제의 정의	- 가장 바람직한 데이터 협업 구축을 위해 공공부문의 긍정적인 영향과 투입(input)에 대한 고려가 있었는가? - IoT 센서가 수집하는 데이터에서 적절한 가치를 산출하기 위한 프로젝트의 목표와 혜택의 매핑, 예상되는 문제에 대한 언급과 핵심 사용자의 규명 등에 초점이 맞추어졌는가? - 성공적인 데이터 협업을 위해 이해관계자 간의 협의가 이뤄졌는가?
2. 참여적 의제 설정	- 해당 지역의 거주자나 해당 영역의 전문가, 비영리 단체 등 핵심 이해관계자들이 규명되었는가? - 중요한 문제를 설정하는 데 있어서 이러한 핵심 이해관계자들의 참여가 존재했는가?
3. 위험성 평가	- 특정한 경우에 적용되는, 데이터 사용에 따르는 위험성이나 데이터의 민감도를 등급별로 분류하고 평가하는 체계가 존재하는가? · 1등급: 관련 정보 자체가 규명되지 않음, 2등급: 상황에 따른 역량의 규명, 3등급: 수집된 민감 데이터를 사용하기 전에 윤리적 측면이나 법적 측면의 검토를 거치는가? 등 - 표준화된 리스크 평가와 적절한 원칙이 존재하는가? - 적절한 IoT 시스템 규정이 존재하는가?

8) EU 대기질 지침(EU Air Quality Directives)은 1970년대 후반부터 주요한 정치적 의제로 대두된 대기 오염 문제에 대처하기 위해 미세먼지와 질소산화물 관련 오염 물질에 대한 규제를 담은 Air Quality Framework Directive 96/62/EC를 '96년에 제정한 이래, 다양한 기준을 강화하여 발전하고 있는 EU의 대기질 관련 규제 프레임워크임. 현재 EU는 대기 오염에 대한 WHO의 권고와 '19년의 자체 평가 결과를 반영하여 '22년에 이 지침의 개정을 준비하고 있음

구분	설계 요건과 성공 요인
	- 데이터의 책임감 있는 수집과 사용을 통지하는 표준과 정책에 대한 정기적 검토가 존재하는가? - 데이터 생애 전 주기에 걸쳐서 투명하고 설명 가능한 의사결정 과정이 존재하는가?
4. 책무성	- 데이터 생애 전 주기에 걸쳐서 존재하는 위험을 감시하고 교정하는 과정이 존재하는가? - 강력한 데이터 보안 과정이 존재하는가? - 설계 측면에서 프라이버시를 보호하는 데이터 수집 방법이 있는가? - 수집된 데이터를 공공 부문과 공유하는 과정에서 프로젝트 추진 이전에 위험성을 평가하는가?
5. 투명성	- 데이터가 어떻게, 왜 그리고 어디에서 수집되는지를 분명하게 밝히는 공공 캠페인이 존재하는가? - 투명성과 신뢰도를 강화하는 업계의 규칙이 존재하는가?
6. 이해관계자의 참여	- 주요 이해관계자 및 전문가들이 다양한 집단들과 내부적 협의를 하는가? - 외부 전문가들의 견해를 청취하는 노력을 기울이고 있는가? - IoT 혁신과 협력 허브를 위한 공모가 자주 개최되었는가?

- (운영적 요건) 데이터로부터 도출한 분석 결과의 보급 방안과 데이터의 접근이나 품질 및 상호운용성과 같은 당초의 목적에 부합하는 운영상의 접근법을 개발할 수 있는 방안과 관행을 수립하고 실행

| 표 7 | 운영 측면의 요건

구분	설계 요건과 성공 요인
7. 데이터 품질	- 사전에 도시의 인구학적인 분석이 되었는가? - 통계적으로 소수자 집단과 취약계층에 대한 잠재적인 편향을 식별하고 수정할 수 있는가?
8. 데이터 접근성	- 데이터 접근을 감시하는 정책과 시스템이 갖추어져 있는가? - 데이터 공급망 전체에 걸친 조직 차원의 보안 설계가 존재하는가? - 도시의 투명성과 보고 기준이 채택되어 있으며, 행정 기관 간 상호 운용성이 확보되어 있는가?
9. 데이터 검증	- 비례성, 대표성, 평등성과 같은 관점에서 제기되는 문제들에 기초한 데이터 관련 감사 기능이 있는가? - 데이터 보호의 측면에서 데이터의 결합과 익명화 필요성이 있는가?
10. 데이터 보유	- 데이터에 대한 니즈나 비례성의 원칙 하에서 데이터 저장 및 사용 시간을 규정했는가? - 데이터의 공유 관련 협약에 유효기간을 규정하고 있는가?
11. 파트너십	- 이해관계자들의 참여 동기나 관심을 일목요연하게 정리할 수 있는 오픈 데이터 평가 템플릿이 사용되는가? - IoT 혁신이나 협력 허브와 같은 지식 공유 네트워크에 대한 신뢰가 존재하는가?
12. 결과의 보급	- 이해자들에게 알려주는 형태의 분석 결과와 이에 기반한 전망을 공유하는가? - 프로젝트와 데이터가 주는 영향과 가치를 평가하는가?

구분	설계 요건과 성공 요인
13. 상호운용성	- IoT 기기에서 생성하는 데이터를 공유하고, 전송하며, 재사용할 수 있는 지역의 기준이 존재하는가? - 데이터를 다양하게 사용할 수 있도록 설계된 유연성 있는 플랫폼이 존재하는가? - 수집된 데이터 및 분석 결과가 운송, 생물학적 다양성, 도시 계획 등 대기질과 관련된 다른 산업과의 시너지가 존재하는가?

- (과학적 요건) 적절한 질문의 설정, 데이터 분석 결과 활용에 대한 이해 등의 영향을 결정하는 과학적 요건

| 표 8 | 과학적 측면의 요건

구분	설계 요건과 성공 요인
14. 적절한 질문 설정	- IoT의 데이터가 해결할 수 있는 질문이나 의사 결정이나 행동에 정보를 제공할 수 있는 상황 분석, 인과 관계, 영향 평가, 전망 등과 관련된 질문들이 설정되어 있는가?
15. 분석결과 활용에 대한 이해	- 의사 결정에 인사이트를 어떻게 사용할지에 대한 프로젝트 개시 전이나 실행 도중의 계획이 존재하는가? - 의사 결정에 분석 결과를 사용하기 위한 대중의 인식도나 동의가 있는가? - 다양한 의사소통 방법을 사용할 수 있는가?

- (기술 및 인재의 역량) 환경 및 재정적 지속가능성과 데이터의 저장 등의 요소와 기술 및 인재 측면의 역량

| 표 9 | 기술적/인간적 측면의 요건

구분	설계 요건과 성공 요인
16. 분석 및 IoT 보안 역량	- 데이터 협업을 위한 조달, 구축, 실행, 유지를 할 수 있는 조직 자원이 충분한가? - 데이터 공유 문화가 목적 지향적이고 사용자 중심적이며, 참여와 피해 방지에 기반하는가? - 책임감 있고 투명한 데이터 처리를 위한 의사결정 출처 확인 과정이 존재하는가?
17. 적절한 기술	- 시스템이나 클라우드 보안, 데이터 공유에 대한 감사가 존재하는가? - 데이터 접근 허가 절차가 존재하는가? - 제3자 감사인에게 신뢰성을 제공하기 위한 변경 불가능한 감사 시스템이 유지되는가?
18. 환경적 지속가능성	- 환경 영향 평가가 이뤄졌는가? - 환경친화적인 방식으로 IoT 제품을 생산하거나 배치하는가?
19. 재정적 지속가능성	- 민·관·학 파트너십에 의해 강화되는 데이터 협업 설계를 위해 사회적 기업 등 사회적으로 유익한 혁신적인 비즈니스 모델을 가지고 있는가? - 공공 IoT 혁신 자금과 같이 자원이 적은 분야나 중소기업에서 IoT 기술의 사용을 지원하는 새로운 메커니즘이 존재하는가? - IoT 관련 제품에 신뢰할 수 있는 제품이라는 인증 마크를 획득하기 위한 표준화된 운영 요건이 존재하는가?

구분	설계 요건과 성공 요인
20. 적시성	- 인증을 받은 당사자들이 실시간으로 데이터에 접근할 수 있는가? - 지역에서 데이터를 저장하고 처리하며 마이닝할 수 있는 역량을 가지고 있는가?

3 제시된 프레임워크에 기반한 데이터 협업 사례 분석

▶ 대기질 데이터 협업을 위한 IoT 운영 운영 모델은 다음과 같은 3가지 유형으로 나누어볼 수 있음

- (민-관 파트너십) 특정 지역 혹은 광역 지자체 수준에서 데이터를 수집하는 민간 기업의 센서 기술을 활용하는 프로젝트
- (학술 연구 허브) 학술 기관이 다른 연구 그룹이나 정부와 협력하여 수행하는 프로젝트
- (공동체 주도 프로젝트) 주로 지역 차원에서 대기질을 모니터링하기 위해 공동체 그룹이 추진

| 표 10 | 3가지 운영 모델에 따른 대기질 관련 IoT 데이터 협업 사례

구분	사례
민-관 파트너십	(사례연구 1) 지역 수준의 민-관 파트너십 : Breeze Technologies - '15년에 설립된 Breeze Technologies는 공공장소의 구조물에 부착할 수 있는 소형 모니터링 기기를 제조하는 회사 - 이 회사가 제조하는 센서는 온도, 습도, 일산화탄소, 이산화탄소, 질소산화물, 암모니아, PM10 및 PM2.5의 미세먼지, 이산화황, 휘발성 유기화합물에 대한 정보를 수집 - 인공지능이 이 센서가 수집한 정보를 실시간으로 분석하고, Breeze Environmental Intelligence Cloud에 저장해 데이터에 기반한 예측적 분석을 수행하고, 이 정보는 도시의 대기질을 개선하는 '스마트 행동'을 전개하는 데 필요한 정보를 제공 - 시 전역에 설치된 이 센서들은 데이터 기반 정책의 수립을 가능하게 하며, 시민들이 적극적인 대기 오염 방지 행동을 할 수 있도록 정보를 제공하는 역할을 수행 - Breeze Technologies는 일반 대중에게 공개된 Citizen Portal을 통해 센서 데이터 정보를 발표 (사례연구 2) 민-관 및 여러 지자체 공동 파트너십: AirThings - AirThings는 '효율적 대기질 모니터링'을 통해 대기 중의 특정 물질과 가스 형태의 화학물질을 측정하는 네트워크를 호스팅하는 업체 - 불가리아, 그리스, 키프로스, 알바니아, 북마케도니아 5개국이 참여하는 발칸반도-지중해 프로그램(Balkan-Mediterranean Program)*을 통해 EU에서 자금을 지원받고 있음 * '07~'13년 운영된 남동부 유럽 프로그램을 계승한 새로운 형태의 국제 협력 프로그램으로, 발칸반도와 지중해 지역 국가들의 협력을 증진하기 위한 목적으로 추진됨. 주요 지원 프로그램은 지역 내 창업 및 혁신 촉진과 환경 개선에 목표를 두고 있으며, 기금 규모는 2천 8,33만€에 달함 - 불가리아의 Sofia시(市)가 이 프로젝트를 선도적으로 운영하고 있으며, 이외에도 그리스의

구분	사례
	Thessaloniki시, 키프로스의 Nicosia시, 알바니아의 Tirana시, 북 마케도니아의 Skopje시가 프로젝트에 참여하고 있음 - Airthings는 91개의 센서를 설치하고 실시간으로 대기질 데이터를 수집 및 모니터링하고 있으며, 대기오염 완화 정책 수립과 대중의 경각심을 제고하기 위해 필요한 정보를 제공 - 수집된 데이터는 Open Data System에 의해 호스팅되며, 다양한 정보를 시각화하여 다양한 개인과 조직들이 사용할 수 있도록 제공하고 있음
학술 연구 허브	(사례연구 3) 미국 시카고 Array of Things Project - 미국 일리노이 대학의 Northwestern-Argonne 과학엔지니어링연구소와 시카고 대학에 의해 추진된 Array of Things 프로젝트는 미국 Chicago의 교통부와 국립과학재단과 파트너십을 구축하여 시카고시의 주요 교차로에 130개의 센서를 설치함 - 이 센서들은 Argonne 국립 연구소가 개발한 오픈소스 시스템인 Waggle에 연결되어 있으며, 대기질, 온도, 광량, 진동 및 기압 관련 정보를 실시간으로 수집 및 제공 - '16년부터 설치되기 시작한 이 센서들은 화상으로 정보를 수집하기 때문에, 개인정보 보호를 위해 프라이버시와 관련된 영상은 삭제됨 - 이 프로젝트는 투명성과 책임성 원칙 하에서 운영되며, 다양한 이해관계자들이 프라이버시관련 정책을 주기적으로 검토 - 센서가 수집하는 정보는 30초마다 Argonne 국립 연구소에 저장되고, 5분마다 API 서비스에 의해 업로드되는 한편, 데이터를 관리하는 웹사이트는 24시간마다 갱신됨 - 센서에 의해 수집된 데이터는 보다 나은 도시의 삶을 위한 솔루션 제공에 사용되기 위해 무료로 개방되고 있음
커뮤니티 주도 프로젝트	(사례연구 4) Stadslab Air Quality - 대기질을 개선하고자 하는 네덜란드 Rotterdam시의 공모 결과에 따라 '14년에 'Stadslab Air Quality Lab'이 설립되었으나 현재는 더 이상 시 정부의 재정지원을 받지 않음 - Stadslab Air Quality Lab은 다양한 분야에 걸친 학제적인 지식의 공유와 지역 공동체의 참여에 중점을 둠 - Stadslab Air Quality Lab은 대기질과 관련된 데이터의 수집뿐 아니라 지역 사회에서 대기질과 관련된 다양한 컨퍼런스나 행사, 워크숍을 주최하고 대기 오염을 저감할 수 있는 모범사례를 찾고, 대중의 관심을 환기 - 최근 Stadslab Air Quality Lab은 자연에 존재하는 이끼를 이용한 공기 정화 필터에 의한 대기질 개선 방법 등을 연구 중

- ▶ 온라인 상의 자료 조사를 통해 4가지 사례를 데이터 협업 분석 프레임워크에 따라 평가한 결과, 4가지의 사례 모두 거버넌스 구조, 리스크 관리 프레임워크, 재정적 지속가능성이 부족하다는 결과를 도출
- (거버넌스) 확립된 책임성이라는 측면에서 광범위한 거버넌스 프레임워크는 다소 혼란스러운 모습을 보이는 것으로 나타남
 - (운영) 제시된 모든 사례에서 수집된 데이터에 대한 접근이나 투명성이 준수되고 있으나, 데이터의 품질 관리와 정보에 대한 감사 방법에서는 한계를 보이고 있음. 데이터 품질과 관련하여 Airthings가 골격을 갖추고 있고, 'Chicago Array of Things academic hub' 사례에서 감사 메커니즘이 실행되고 있음

- (과학적 요건) 일부 프로젝트에서 지역의 운동가들이 정의한 질문과 중요한 문제를 프로젝트 운영에 반영하는 것으로 나타났지만, 데이터의 분석 결과에서 제시되는 시각이나 통찰들이 뒤이은 시민들의 행동에 어떻게 반영되는지는 명확하지 않음
- (기술 및 인재 역량) 4가지 사례 모두 강력한 파트너십 네트워크와 이해 관계자들 사이의 데이터 수집 및 분석 결과에 대한 공유가 이루어지고 있음. 한편으로 4가지 사례 모두 환경적으로 지속가능한 IoT 기술을 채택하고 있지만 데이터 수집 과정에서 비영리 기관의 참여에 따른 경제적 지속가능성이 결여될 가능성은 존재함

PART IV

결론 및 시사점

- ▶ EU는 유럽 데이터 전략(European strategy for data, '20년)을 통해 유럽 공통 데이터 공간 구축을 추구하면서 시민 생성 데이터의 정책 결정 개선 잠재력을 인정
 - 다양한 출처의 데이터 조합 및 결합은 유럽 공통 데이터 공간을 성공적으로 구축하는 데 가장 중요한 요소로서, 데이터의 공유와 사용을 위한 첨단 기술과 혁신적 접근법이 이를 가속화하는 주요한 동인으로 작용
 - 동 보고서는 오늘날의 기술 지형이 매우 역동적으로 변화한다는 점을 감안할 때, 정책적 의제를 설정하고 사회 기술적 요인을 조사하기 위해서는 첨단 기술과 다양한 도구를 사용하는 일련의 실험을 가능하게 하는 샌드박스 접근법에 무게 중심을 두고 다음의 두 가지 사례를 분석
 - 시민 주도 지리공간데이터 협력 프로젝트인 오픈스트리트맵(OSM) 데이터셋과 각국 정부의 지도제작기구(NMA)의 신뢰 가능한 데이터를 결합한 사례 분석
 - IoT 센서를 이용하여 대기질에 영향을 미치는 다양한 오염 물질 관련 데이터를 수집하고 사용하는 사례 분석
- ▶ 주소 데이터 관련 공공 데이터와 시민 생성 데이터의 결합 실험 결과, 데이터 결합이 두 가지 데이터셋 모두를 개선하면서 기술적 상호운용성을 입증했으나 조직이나 법·제도적 상호운용성은 확대될 필요가 있음
 - 핀란드의 경우, 신뢰 가능한 데이터는 지역의 포괄 정도가 균일하고 상대적으로 높은 위치 정밀도를 가지고 있으나 OSM 데이터는 불균등한 분포 정도를 보이고 있음
 - OSM 데이터는 최신 정보를 업데이트하는 속도가 빠르지만, 행정정보 데이터의 갱신주기는 짧다는 문제가 있음
 - 네덜란드의 경우, OSM 커뮤니티의 정기적인 BAG 데이터 결합으로 기술적 문제가 적게 발생
 - BAG 데이터의 복잡도가 높기 때문에 데이터 결합 과정에서 상당히 깊은 데이터 구조에 대한 이해가 필요하지만, OSM 데이터는 이에 비해 직관적이고 상대적으로 편리하게 데이터에 접근할 수 있는 구조를 가지고 있어 행정정보 데이터와의 결합이 가능
 - 그러나 행정 기관과 OSM 커뮤니티 간의 조직간 상호운용성의 확보를 통해 데이터의 결합 과정에서 공통 및 표준 관련 정책을 수립할 필요성 존재
 - 두 가지 데이터의 결합 과정에서 나타난 라이선스의 사용 조건과 관련된 상호운용성 측면의 제약은 결합된 데이터셋의 활용이나 데이터 결합 과정의 효과적인 거버넌스 수립에 장애요인으로 작용
 - 일부 NMA들은 신뢰 가능한 데이터의 라이선스를 CC-BY 4.0에서 ODbL로 전환하는 방안을 제시하고 있음

· 또 다른 방안으로 OSM 데이터에 기존의 ODbL을 유지하면서 CC-BY 4.0 라이선스도 동시에 부여하는 방안이 고려되고 있음

- 현재 EU 차원에서 공개 라이선스의 사용과 관련하여, '19년에 오픈데이터지침(Open Data Directive)*이 공개되었으며, '22년 중 '높은 경제 및 사회적 가치를 가진 데이터'의 공개 라이선스에 대한 법적 근거가 마련될 것으로 예상

* 'Directive on open data and the re-use of public sector information'의 약칭. 기존의 공공부문정보지침(Public Sector Information (PSI) Directive)을 대체하여 '19.6월부터 발효된 규정으로, 높은 가치를 가진 지리 공간, 환경, 우주, 통계, 기업 소유 관련 정보, 운송 등의 데이터를 API 등 기계가 읽을 수 있는 형태로 공개하는 것이 골자

▶ IoT 센서로 다양한 대기 오염 물질 관련 데이터 수집·사용에 관한 데이터 협업 사례 분석 결과, IoT 거버넌스 개발, 역량 구축, 시민 참여 및 연구의 확대 등의 정책적 시사점 도출 가능

- (일반적인 IoT 거버넌스 프레임워크의 개발) 공공과 기업 및 시민사회의 이해관계자들은 IoT 기술에 의해 신뢰도가 높고 재사용이 가능한 대기질 데이터 거버넌스 프레임워크를 구축해야 하며, 이러한 프레임워크에는 데이터 공개 정책, 투명성 요건 및 안전 장치, 책임성 메커니즘 등이 포함되어야 함
- (역량의 구축) 중앙 및 지방정부는 IoT 대기질 관련 데이터의 운영 역량을 확보하고 사전에 준비할 필요가 있음. 이러한 준비에는 관련 정책 입안자 및 행정 담당자들을 대상으로 하는 교육 훈련 등이 있으며, IoT 데이터의 분석을 통해 답을 얻어낼 수 있는 정식화된 질문 사항들을 수립해야 함
- (데이터 관리 체제의 수립) IoT 대기질 데이터를 효과적으로 운용하기 위해서는 기업이나 공공 및 시민사회 모두 데이터 관리 체제를 만들고 운용해야 함. 이러한 데이터 관리 체제는 공공의 이익에 부합하면서도 민감한 정보를 보호하고 데이터 분석에 의해 의미있는 시사점을 도출하는데 필수적임
- (시민의 참여 확대) 시민의 참여를 통해 IoT 데이터의 수집을 확대하고 대기질 관리 목표를 정식화할 수 있음, 이를 위해서 IoT에 의해 가능한 혜택이 무엇인지에 대해 시민들에게 투명하게 공개하는 노력이 지속되어야 함
- (관련 연구의 확대) IoT에 의한 데이터 수집에 대한 관심이 높아지고 있지만, 어떤 접근이 가장 좋은지에 대한 연구는 부족한 것이 현실임. 따라서 IoT 데이터 수집과 관련하여 모범 사례의 발굴과 관련된 다양한 연구들이 진행될 필요가 있음

▶ 우리나라의 경우, '20.8월 '데이터3법' 개정 시행으로 이중 데이터 간 결합이 허용되고, '22.4월부터 시행될 '데이터 산업진흥 및 이용촉진에 관한 기본법' 제정으로 데이터 활용 및 관련 산업진흥이 본격화되면서 데이터 산업 생태계를 활성화하는 기반이 마련됨

- '20.12월부터 공공기관이 결합전문기관으로 지정받아 해당 기관이 보유한 공공데이터를 결합해 제3자에게 제공 가능하며, 네이버의 '네이버 클라우드 데이터 샌드박스' 출시를 통해 스타트업, 공공기관, 대학연구소가 데이터와 분석도구를 제공해 자유롭게 데이터를 분석·가공해 활용 가능
- 본 보고서의 유럽 데이터 샌드박스 사례와 데이터 결합 실험, 데이터 협업사례 등을 국내 민간·공공데이터 결합 및 협업 활성화를 위한 제도 마련에 참고할 수 있음

참 고 문 헌

- Granell. C. et al., Emerging approaches for data-driven innovation in Europe, JRC Technical Report, European Commission, 2022.
- Academy of Europe Law, How to handle Court Proceedings Invoking Non-compliance with EU Air Quality and Noise Legislation, 2012.
- <https://ec.europa.eu/environment/air/quality/index.htm>
- https://ec.europa.eu/regional_policy/EN/atlas/programmes/2014-2020/cyprus/2014tc16m4tn003#:~:text=The%20Balkan%2DMediterranean%20Programme%20is, promote%20cooperation%20in%20the%20area
- <https://creativecommons.org/>
- <https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data>
- <https://wiki.openstreetmap.org/wiki/Foundation>
- <https://github.com/nlextract/NLExtract>
- <https://inspire.ec.europa.eu/>



데이터산업 동향 이슈 브리프

ISSUE BRIEF

| 발행일 2022년 3월 31일

| 발행처 **K data** 한국데이터산업진흥원

서울시 중구 세종대로 9길 42, 부영빌딩 8층

| 기획 및 편집 데이터산업본부 산업기획팀

| 문의처 Tel: 02-3708-5363, 5364

* 본 지에 실린 내용은 한국데이터산업진흥원의 공식 의견과 다를 수 있습니다.
본 내용은 무단전재를 금하여, 가공/인용할 경우 반드시 출처를 밝혀주시기 바랍니다.