

2022년 제6호 (통권25호)



데이터산업 동향 이슈 브리프

ISSUE BRIEF

2022.6

빅데이터와 SDG: 모니터링을 위한 혁신과 파트너십

빅데이터와 SDG: 모니터링을 위한 혁신과 파트너십

I. 개요	1
II. 빅데이터의 정의 및 전망	3
III. SDG 모니터링을 위한 빅데이터 활용 사례	6
IV. SDG 모니터링을 위한 빅데이터 활용 파트너십	11
V. 빅데이터를 활용한 SDG 모니터링 혁신방안	17
VI. 결론	20

요 약

- 2030년 지속가능개발 의제(2030 Agenda for Sustainable Development)의 채택으로 경제, 사회, 정책 영역의 기준 개발 목표(SDG, Sustainable Development Goal) 모니터링에서 지표 진행 현황 파악을 위해 데이터 수집과 활용이 중요해짐
 - 그러나, 2020년 12월 기준 개발 목표(SDG)의 40% 이상이 정기적으로 확인되지 않고 있으며, 사용 가능한 데이터 세트가 업데이트되지 않은 경우가 많아 효과적인 SDG 모니터링 구현에 어려움이 있음
- 이를 개선하기 위해 지난 10년간 통계를 바탕으로 SDG 모니터링을 지원하기 위해 빅데이터를 실험하는 국가의 최근 사례를 제시하고, 새로운 데이터 세트 생성과 같은 방법론적 발전으로 가능해지고 있다는 것을 증명함
- 위의 사례를 통해 빅데이터를 활용하는 데 있어 여러 국가가 겪은 최근의 경험과 교훈을 탐구하고, 국가 간의 파트너십이 어떻게 SDG 모니터링을 지원할 수 있는지에 대한 지침을 제공하고자 함
- SDG 모니터링을 위한 빅데이터 파트너십에 대한 유형과 역할에 대해 명시하고, 통계청을 위한 빅데이터 성숙도 및 파트너십 모델을 추가로 제안함
- 또한 빅데이터 소스 사용을 위한 파트너십 프로세스와 SDG 모니터링을 위한 방법에 대한 통계청의 실질적인 지침을 제시함
- 결론적으로 SDG와 관련한 모니터링 과제를 해결하는 데 있어 빅데이터의 가치를 강조하고 이에 관한 활용 사례와 캐즘(Chasm)¹⁾을 줄여 데이터 혁신을 실현하기 위한 파트너십 유형과 방안을 제시하고 있음
- 최근 SDG에 관한 관심도가 높아지면서 우리나라도 빅데이터 기반 SDG 과제 실현을 위한 국제 협력, 데이터 세분화, 데이터 혁신 등을 추진 중임
 - 정부는 개발도상국가(베트남, 요르단 등)의 캐즘을 줄이기 위해 성숙한 빅데이터 활용 국가통계시스템을 전수하고 있음
 - 국내에서는 SDG 데이터 혁신 포럼을 개최하여 데이터 기반 SDG 모니터링과 빅데이터 활용 등 데이터 혁신 등의 의제가 주로 논의되었음
 - 교통 빅데이터 등 SDG 과제별 주요 데이터 접근법 연구 결과가 발표되었으며, 향후 확대 적용 방안에 대해 논의 예정

1) 데이터 활용 능력 격차가 발생하는 현상

PART I

개요

- ▶ 2030년 지속가능개발 의제(2030 Agenda for Sustainable Development)²⁾의 채택으로 인해 광범위한 경제, 사회 및 환경 정책 영역에서 각종 지표 진행 상황을 검토하기 위해 데이터의 수집과 활용이 중요해짐
- 그러나, 2020년 12월 기준 지속가능 개발 목표(SDG, Sustainable Development Goals)의 40% 이상이 정기적으로 확인되지 않고 있으며, 사용 가능한 데이터 세트가 업데이트되지 않은 경우가 많아 효과적인 SDG 모니터링 구현에 어려움이 있음
- 데이터 수집, 조정, 검증 및 품질 보증에 중추적 역할을 하는 각국의 통계청(NSO, National Statistical Offices)은 SDG 실현을 위한 모니터링 과정에서 중점적인 역할을 수행함
 - 최근 통계청이 사용자의 데이터 수요 증가, 예산 감소 및 데이터 수집 비용 증가라는 상황에 직면하면서 국가 ‘데이터 생태계’에서 데이터 활용에 관한 관심이 높아지고 있음
 - 데이터 생태계에는 - 통계청 및 부처, 시민 사회 조직, 학계 및 민간 부문의 협력자가 포함되어 있으며, 새로운 데이터 소스, 지식 및 기술 또는 국가적 SDG 모니터링을 지원하기 위한 컴퓨팅 인프라 등도 포함됨
- 최근 몇 년간 SDG 모니터링을 지원하는 빅데이터의 잠재력에 대해 각종 논쟁이 있어왔음, 빅데이터는 비용 절감 및 효율성을 제공한다는 장점이 있지만, 방법론적 문제, 기술 및 기능 격차, 지속 가능성 및 접근성 문제, 데이터 개인 정보 보호 및 보안 문제 등 여러 문제를 수반함
- 코로나19로 인해 각국 통계청은 기존 데이터 수집 방식의 효율성을 높여야 했으며, 데이터 격차를 해소하기 위해 대체 데이터 소스를 활용하는 새로운 협력 관계를 구축함
 - 통계청은 파트너십 체결을 통해 협력자로부터 빅데이터 프로젝트를 진행하는 데 필요한 데이터 소스, 기술, 인프라, 지식 및 자금을 받을 수 있어 상당한 기회 요인이 존재하지만, 데이터 개인 정보 보호 및 보안 문제 등 여러 가지 과제와 장벽으로 인해 효과적인 파트너십 체결에 어려움이 있음
 - 여기에는 프로젝트 시작에 대한 인식 부족과 어려움, 법률 및 개인정보보호 문제, 기술적 능력과 방법론적 문제가 포함되며, 국가와 통계청의 빅데이터 성숙도에 따라 해결 방식에 차이가 있음
 - 그럼에도 국가 통계 작업에서 다양한 국가의 통계청들이 빅데이터를 활용하며, SDG 모니터링을 위한 빅데이터 파트너십 체결도 계속되고 있음
- 이러한 빅데이터 혁신을 진행하는 통계청의 성숙도를 평가하는 성숙도 모델은 조직 특성과 기술적 역량, 거버넌스 등 핵심 영역을 가짐

2) 2015년 9월 유엔 총회와 특별정상회담에서 채택된 새로운 개발목표는 2016부터 2030년까지 15년간 유엔을 비롯한 국제사회의 빈곤퇴치와 지속 가능한 발전 실현을 위한 큰 틀을 제시함

- 조직 특성은 빅데이터 인식, 사용 능력 및 빅데이터 사용 사례 지원을 포함하며, 기술적 역량은 인적 자원 능력 및 인프라를 포함하고, 거버넌스는 법적 및 규제 환경을 포함함

- 빅데이터 성숙도는 초기, 도입 전, 도입 초기, 발달, 성숙 단계로 구분되며, 성숙도에 따라 통계청은 다양한 협력자 및 파트너십 모델이 필요함

- 효과적인 빅데이터 활용 파트너십을 위해 성숙도 단계에 따라 빅데이터 인식 및 리터러시, 인적 자원 역량 및 인프라, 거버넌스 및 규제 프레임워크 개발 등 여러 요소가 고려되어야 함

- ▶ 동 보고서*는 이러한 파트너십이 어떻게 국가의 SDG 모니터링을 지원할 수 있는지 데이터 분야의 정부 대표자 및 전문가와의 인터뷰, 관련 문헌 검토 등을 통해 사례를 제시하고자 함

* Thematic Research Network on Data and Statistics (TReNDS), Big Data and the Sustainable Development Goals: Innovations and Partnerships to Support National Monitoring, 2021

- 이를 통해 빅데이터를 활용하는 데 있어 여러 국가가 겪은 최근의 경험과 교훈을 탐구하고, 파트너십이 어떻게 SDG 모니터링을 지원할 수 있는지에 대한 지침을 제공하고자 함

- 2장에서는 SDG 모니터링을 위한 빅데이터의 개념과 정의를 명확히 하고, 가치 제안을 요약함
- 3장에서는 빅데이터의 유형과 소스, SDG 모니터링을 위해 빅데이터를 활용한 국가의 최근 경험을 제시하고 빅데이터 소스에서 SDG 지표를 도출하는 새로운 혁신에 대한 체계적인 검토 결과를 제시함
- 4장에서는 빅데이터 파트너십 환경을 검토하고 협력자의 파트너십 유형과 역할에 대해 논의하며, 통계청을 위한 빅데이터 성숙도 및 파트너십 모델을 제안함
- 5장에서는 혁신적인 빅데이터 소스 사용을 위한 파트너십 프로세스와 SDG 모니터링을 위한 방법에 대한 통계청의 실질적인 지침을 요약함

PART II

빅데이터의 정의 및 전망

1 빅데이터 정의 및 유형

- ▶ 빅데이터는 정보의 수집, 저장, 배포, 관리 및 분석을 지원하는 고급 기술이 필요한 대량의 가변 데이터를 설명하는 용어임³⁾
- 빅데이터의 개념은 2000년대 이후에 도입되었으며, 2011년부터 본격적으로 활용됨
 - 빅데이터의 발전은 컴퓨팅 파워, 유비쿼터스 및 모바일 컴퓨팅, 네트워크 스토리지, 새로운 형태의 데이터베이스 설계, 소프트웨어 매개 통신 및 트랜잭션의 새로운 모드, 머신 러닝을 활용하는 데이터 분석의 발전으로 가능해짐
- 빅데이터는 다른 데이터 소스와 구별되는 여러 가지 고유한 속성을 가지고 있음
 - 다음 표는 SDG 지표 모니터링에 대한 빅데이터의 기여도를 이해하기 위해 중요함

| 표 1 | 빅데이터의 10가지 속성

속성	정의
양(Volume)	- 데이터의 양 (데이터 기록 수, 속성 및 연결 장치)
속도(Velocity)	- 데이터의 생산 속도 (데이터 기록 생산, 수신, 처리 및 이해 속도)
다양성(Variety)	- 데이터의 종류 및 형식에 대한 다양성 (데이터 소스, 형식, 미디어 및 콘텐츠의 다양성)
변동성(Volatility)	- 데이터의 변동성 (변화하는 기술 및 데이터 스토리지)
진실성(Veracity)	- 데이터의 출처에 따른 진실성 (출처 및 가용성의 신뢰성)
유효성(Validity)	- 데이터의 정확도 (데이터의 정확성, 신뢰성 및 품질)
가치(Value)	- 데이터의 가치 (수집된 데이터의 비즈니스 가치)
가변성(Variability)	- 환경에 따른 데이터의 가변성 (데이터 의미의 지속적 변화)
취약성(Vulnerability)	- 데이터의 유출 가능성 (데이터의 개인적 특성과 개인정보 보호 및 보안의 필요성)
시각화(Visualization)	- 데이터의 여러 속성으로 인한 시각화가 어려운 상황 (낮은 확장성 및 기능성)

3) Gadoni, A. & Haider, M. Beyond the hype: Big data concepts, methods, and analytics. International journal of information management, 2015

- 빅데이터 유형과 소스를 분류하는 보편적인 접근 방식은 없지만, 유엔 통계위원회(The United Nations Statistical Commission)는 빅데이터의 유형을 다음과 같이 분류함

| 표 2 | 빅데이터의 유형

유형	내용
행정 데이터 (Administrative Data)	- 정부 또는 비정부 프로그램 관리에서 발생하는 데이터 소스 (전자 의료 기록, 병원 방문, 보험 기록 및 은행 기록)
상업 데이터 (Commercial Data)	- 두 기업 간의 거래에서 발생하는 상업 또는 거래 데이터 소스 (신용카드 및 온라인 거래, 전자 판독 데이터)
센서 데이터 (Sensor Data)	- 센서 네트워크 소스 (위성 영상, 도로 센서 및 실내 온도 센서)
추적 또는 모바일 데이터 (Tracking or Mobile Data)	- 추적 장치 소스 (모바일 및 위성 위치 확인 시스템의 데이터 추적)
행동 및 의견 데이터 (Behavioral and Opinion Data)	- 행동 데이터 소스(제품, 서비스 또는 온라인 페이지 방문 횟수에 대한 검색) 및 사용자 의견 데이터 소스(소셜 미디어 내의 댓글)

2 공식 통계에서의 빅데이터 전망

- ▶ 궁극적으로 빅데이터에 대한 투자는 투자 대비 효용성을 고려하여 신중하게 비교되어야 하며, 이를 위해 서로 다른 소스의 데이터를 결합해 공식 통계의 가치를 높이고 데이터 가치 사슬(Data Value Chain)의 관리자를 통해 표준화시켜야 함

- SDG와 관련해 빅데이터는 기존 데이터 방식과 비교했을 때 더 나은 해결 방안을 제공할 수 있음
 - 국가통계시스템(NSS, National Statistical Systems)에서 빅데이터는 다음과 같은 사항을 포함해 여러 목적으로 활용될 수 있으므로 높은 가치를 지님

- SDG를 포함한 기존 및 새로운 측정 기준에서 통계 통찰력의 범위, 폭 및 품질 향상
- 데이터 제품이 시계열의 공백을 메우거나 실시간 데이터 제품에 대한 새로운 요구를 충족할 수 있도록 시기적절하게 지원
- 소규모 면적 추정치를 사용하여 기존 데이터 세트의 세분성 향상
- 기존 설문 조사와 관련된 비용 및 보고 부담 감소
- 새로운 방법론의 혁신 추진 및 새로운 인재 유치

- 빅데이터와 관련된 기회를 이용하기 위해서는 범위, 적시성 및 해결 방법 제고뿐 아니라 통계를 리소스화하고 정리하는 데 있어 효율성을 높이고 비용을 절감해야 함
 - 빅데이터의 발전은 급격한 혁신을 수반하기 때문에 국가통계시스템에 여러 과제와 위험을 안길 수 있으며, 여기에는 방법론적 문제, 법률 및 규제 문제, 지속 가능성 및 접근권 문제, 데이터 개인정보보호,

보안 및 평판 문제가 포함됨

- 빅데이터의 기회, 과제 및 위험은 빅데이터의 유형(오픈 소스 또는 사유), 법률 및 규제 맥락, 기관 역할 설정에 따라 크게 달라질 수 있음

| 표 3 | 공식 통계에서의 빅데이터 기회, 과제 및 위험

기회	과제	위험
· 기존의 데이터 세트 보완, 교체, 개선 및 추가 · 기존 설문 조사의 비용 절감 · 결과물의 증가 · 보고 부담 및 설문 조사 부담 감소 · 소규모 및 세부적인 분석 보완 및 확장 · 품질 및 지상 실측 정보 개선 · 기존 통계 구성 개선 · 관할 구역 간 비교 용이 · 다른 데이터 세트에 대한 연결성 향상 · 데이터 분석을 통해 새롭고 더 나은 통찰력 창출 · 생산 과정의 작업 방식 및 효율성 향상 최적화 · 가치 높은 작업에 인력 재배치 · 계산 과학, 데이터 과학 및 데이터 산업과의 협업 강화 · 공식 통계의 사용 및 가시성 향상	· 빅데이터 기업과 전략적 제휴 체결 · 데이터 및 메타데이터에 대한 접근 권한 획득 · 데이터 세트의 출처 설정 · 법률 및 규제와 윤리적 문제 · 목적에 대한 적합성 설정 · 진실성(정확도), 불확실성, 오류, 편향, 신뢰성 및 교정과 관련된 데이터 세트 품질 설정 · 데이터 보안 보장 · 데이터 분석 및 모델링 기능 개발 · 기관의 변경 관리 · 관리 기관 간 협업 및 공통 표준 보장 · 대규모 데이터 세트의 처리, 저장 및 전송 · 데이터 변동성	· 미션 드리프트⁴⁾ · 평판 손상 및 국민의 신뢰 상실 · 개인정보 침해 및 데이터 보안 · 일관성 없는 접근권 및 연속성 · 빅데이터 기업 및 대중의 반발 · 국가 간 접근 방식의 단편화 · 자원활용의 제약 및 절감 · 민영화 및 경쟁

4) 비영리 단체 내에서 초기 조직 목표와 사명이 변질되는 상황



SDG 모니터링을 위한 빅데이터 활용 사례

1 SDG 모니터링을 위한 국가통계시스템 혁신

- ▶ 각국의 국가통계시스템은 시스템을 현대화하기 위해 빅데이터 소스와 응용 프로그램을 사용해 혁신을 강화하고 있으며, SDG 모니터링을 위한 많은 혁신 및 활용 사례가 존재함
- 연구 기관, 학술 기구, 국제 기구, 지구과학 및 우주기관, 민간 부문 및 시민 사회 조직을 포함한 다양한 주체를 통해 데이터 생태계 내에서 혁신이 일어나고 있음
 - SDG 모니터링을 지원하기 위해 국가데이터시스템과 빅데이터 혁신 주체 사이의 데이터 격차를 좁힐 필요가 있으며, 이를 위해 각 국가의 데이터 소유권을 보장하고 국가별 통계청(National Statistical Office)에 대한 지원이 활성화되고 있음
- 공식 통계의 목적과 의사결정을 지원하기 위해 빅데이터를 사용하는 다수의 국가별 빅데이터 활용 사례도 있음
 - ① 가격 통계를 위한 웹 스크래핑* 스캐너 데이터 ② 소비자 신뢰 지수에 사용되는 소셜 미디어 데이터 ③ 이동 및 관광 통계를 위한 휴대전화 데이터 ④ 농업 통계를 위한 위성 데이터와 같은 분야에서 수년간 프로젝트가 진행됨
 - * 인터넷 웹 페이지에 나타나는 데이터 중에서 필요한 데이터만을 추출하도록 만들어진 프로그램
 - 이러한 프로젝트는 주로 OECD 국가들의 발전된 국가통계시스템에 의해 주도되어왔으며, 전문화된 공동체를 넘어 데이터 사용을 확대하기 위해 기관별 의사결정자를 지원할 필요가 있음
- 각 국가의 SDG 모니터링을 지원하기 위해 빅데이터 소스를 실험하는 국가의 사례 연구도 다수 보고됨
 - 필리핀 통계청(PSA, The Philippine Statistics Authority)은 농촌 접근 지수와 빈곤율 및 지역 GDP와 같은 각자 다른 사회 및 경제 지표와의 관계를 측정하기 위해 연구를 수행함
 - 또한, 필리핀 통계청은 아시아 개발 은행(The Asian Development Bank)의 통계 및 데이터 혁신부 및 세계 데이터 연구소(World Data Lab)와 협력하여 위성 이미지 및 관련 지리 공간 데이터를 사용한 빈곤 매핑의 실현 가능성을 검토함
 - 필리핀 첨단과학기술원(ASTI, Advanced Science and Technology Institute)도 필리핀 통계청과 해양·해운 이용 교육과 위성 데이터 등 차기 농어업 인구 조사를 위한 인공지능 활용을 모색하는 내용의 양해각서를 체결함

| 표 4 | 국가 공식 통계의 빅데이터 활용 사례

기관	내용
이탈리아 통계청 (ISTAT, Italian National Institute of Statistics)	- 웹 인텔리전스 (기업 웹사이트의 웹 스크래핑, 가격 스크래핑, 일자리 스크래핑) - 소셜 미디어 데이터 (트위터 데이터에 기초해 경제 지수에 대한 사회적 분위기 파악) - 공공 센티넬 2(Sentinel 2)* * 높은 공간 해상도로 지구 관측 광학 이미지를 얻는 위성 - 위성 이미지 사용 및 생산을 위한 자동 파이프라인을 통한 딥 러닝 등 빅데이터 프로젝트 진행
영국 통계청 (UK-ONS, United Kingdom Office for National Statistics)	- 가격 데이터 웹 스크랩, 스마트 미터 데이터, 주소 매칭을 위한 머신러닝 및 텍스트 설문 응답 코딩에 대한 자연어 처리 탐색 프로젝트 진행 - 영국 정부는 2017년 데이터 사이언스 캠퍼스를 설립해 빅데이터와 행정 데이터 등 새로운 데이터 소스를 조사하고 데이터 사이언스 역량 구축을 도움 - 2019년 기준 영국 외무부(Foreign Office)에서는 빅데이터 도구 개발, 국제 개발 전문가를 위한 데이터 과학 역량 구축, 파트너 국가(르완다, 가나, 케냐 등) 통계청 직원 멘토링 등을 포함한 글로벌 공공재를 위한 데이터 과학 공동 팀을 출범시킴
네덜란드 통계청 (Netherlands Statistics)	- 2015년 기준 세계 최초로 빅데이터로 제작된 공식 교통 통계를 발표함 - 실시간 통계의 생성과 기존 통계의 세분화를 위한 빅데이터 통계 센터를 설립하고 사회 통계(안전, 건강, 주택), 경제 통계(물가, 노동 시장), 이동 통계(교통, 관광)와 SDG 및 스마트 시티 등을 담당함
호주 통계청 (ABS, Australian Bureau of Statistics)· 뉴질랜드 통계청 (Statistics New Zealand)	- 네덜란드의 초기 성공을 바탕으로 아시아 태평양 지역의 통계청은 공식 통계를 작성하기 위해 기존과 차별화된 데이터를 사용해왔음 - 소비자물가지수를 조사하기 위해 호주 통계청과 뉴질랜드 통계청은 소매업체에서 바코드를 통해 얻은 데이터와 온라인 데이터를 통합해 상당한 관심을 받음

- 많은 국가에서는 SDG 모니터링을 위해 지리 공간 및 지구 관측 데이터를 활용하고 있음
 - 일본 우주국(JAXA, Japan's Space Agency)은 새로운 일본 위성 히마와리(Himawari)를 이용해 지표면 에어로졸 농도를 측정하고 있으며, 토지 소비 비율, 도시 개방 공간, 산지 등 다른 SDG 지표에 대한 방법론도 개발하고 있음
 - 중국 과학원 빅어스 데이터 엔지니어링 프로그램(CASEarth, The Chinese Academy of Science Big Earth Data Engineering Program)은 다양한 SDG 지표를 모니터링하기 위한 지구 관측 데이터의 기여도를 연구함
 - 멕시코 국립통계지리연구소(INEGI, The National Institute of Statistics and Geography)는 최근 오픈데이터큐브(ODC, The Open Data Cube)*를 활용한 수상 생태계 모니터링 애플리케이션 등을 통해 지리 공간 역량을 강화하고 있음
- * 위성 데이터를 더 잘 관리하고 활용하기 위해 호주 지형연구원(Geoscience Australia)이 개발한 비영리 오픈 소스 프로젝트

2 SDG 모니터링을 위한 빅데이터 연구 커뮤니티 혁신

- ▶ 클라우드 컴퓨팅 인프라 및 알고리즘, 위성 이미지 유비쿼터스 환경에 의해 지리 공간 데이터 관련 SDG 애플리케이션이 발달됨
- 학계에서는 SDG 모니터링을 지원하기 위한 빅데이터 소스의 활용에 대한 혁신이 빠르게 진행되고 있음
 - SDG 분야 15가지 항목 중 빅데이터 세트의 주요 소스 활용도는 1부터 21로 정량화되었으며, 높은 숫자일수록 활용도가 높음
 - 조사 결과 빅데이터 소스 활용도가 가장 높은 빅데이터 소스 종류는 위성 및 지구 관찰 데이터이며, 이는 위성 및 지구 관찰 분야가 1970년대부터 진행되었기 때문에 세계적으로 수용도가 높기 때문임

표 5 | SDG 분야의 빅데이터 유형 및 소스 활용도

SDG 분야	빅데이터 유형(숫자는 활용도를 의미함)	데이터 소스(숫자는 활용도를 의미함)
빈곤 종식	행정 데이터(1), 의견 데이터(3), 위성 및 지구 관찰 데이터(10), 추적 데이터(3)	CDR(2), CHIRPS(1), Facebook(2), GHS(1), Google(2), Lands(1), MODIS(2), NTL(6), OSM(2), QuickBird(1), SRTM(1), Sentinel(1), WorldClim(1), WorldPop(1)
기아 종식	행정 데이터(1), 위성 및 지구 관찰 데이터(5), 센서 데이터(4)	CHIRPS(2), MODIS(1), Sentinel(1)
건강과 웰빙 증진	행정 데이터(4), 위성 및 지구 관찰 데이터(6), 센서 데이터(3), 추적 데이터(2)	CDR(4), Google(1), OSM(2), SRTM(1), WorldPop(2)
양질의 교육 보장	위성 및 지구 관찰 데이터(1)	NTL(1), WorldPop(2)
성 평등	행동 데이터(1), 의견 데이터(4), 추적 데이터(1)	Facebook(4), Google(2), Twitter(1)
깨끗한 물과 위생 보장	행정 데이터(1), 위성 및 지구 관찰 데이터(14)	GPW(1), Google(1), Landsat(8), MODIS(3), SRTM(1), Sentinel(2)
모두를 위한 에너지 보장	위성 및 지구 관찰 데이터(5), 추적 데이터(3)	GHS(3), Google(1), MODIS(1), NTL(4), OSM(3), SRTM(2), WorldPop(2)
경제성장과 양질의 일자리	행정 데이터(1), 위성 및 지구 관찰 데이터(2)	GHS(1), Google(1), NTL(1)
사회기반 시설 산업화 및 혁신	위성 및 지구 관찰 데이터(1)	Landsat(1), Sentinel(1)
불평등 감소	행정 데이터(1), 위성 및 지구 관찰 데이터(1)	NTL(1), SRTM(1),
지속가능한 도시와 주거지	위성 및 지구 관찰 데이터(11), 센서 데이터(2), 추적 데이터(1)	GHS(2), GPW(2), Landsat(4), MODIS(2), NTL(3), OSM(2), QuickBird(1), Sentinel(2)
기후변화 대응	행정 데이터(1), 위성 및 지구 관찰 데이터(1), 센서 데이터(2), 추적 데이터(1)	CHIRPS(1), MODIS(1), WorldClim(1)
해양 생태계 보존	위성 및 지구 관찰 데이터(3), 센서 데이터(1)	Landsat(1), Sentinel(1)
육상 생태계 보호	행정 데이터(1), 위성 및 지구 관찰 데이터(21), 추적 데이터(1)	Google(1), Landsat(9), MODIS(6), NTL(1), OSM(1), SRTM(2), Sentinel(4), WorldClim(1)
평화·정의·강력한 기관	의견 데이터(1), 위성 및 지구 관찰 데이터(1)	NTL(1), Twitter(1)

- 위성 및 지구 관찰 분야에 속하는 구글 어스 엔진(Google Earth Engine) 플랫폼은 유럽우주국의 센티넬(Sentinel) 1과 2의 레이더 이미지를 통합해 무료로 사용할 수 있도록 하며, 미국의 우주 기관과의 협력을 통해 일반인에게도 지리 공간 데이터를 제공할 수 있게 되어 지난 10년간의 가장 중요한 혁신으로 꼽힘

| 표 6 | SDG 모니터링을 위한 빅데이터 소스에 대한 설명

데이터 세트	내용
CDRs	- 통화 시간, 기간, 완료 상태, 발신 번호 및 수신처를 포함해 통화에 대한 정보를 제공하는 데이터 세트(Call Detail Records)임
CHIRPS	- 기후 위험 그룹 적외선 강수량 관측소 데이터(Climatic Hazards Group InfraRed Precipitation with Station data)는 35년 이상 수집된 강수량 데이터 세트임
GHSL	- 인간의 정착지에 대한 글로벌 공간 정보를 제공하며, 인구 밀도 지도, 정착 지도, 미세한 규모의 위성 이미지, 인구 조사 데이터 및 자원 정보 등을 제공함
Landsat	- 나사(NASA)의 Landsat 프로그램은 1975년에 발사된 위성이 촬영하는 지구 위성 사진을 제공함
MODIS	- 나사의 중간 해상도 이미징 분광계(Moderate Resolution Imaging Spectroradiometer)는 2,330km 폭의 판독 범위를 가지고 있으며, 생물권의 대규모 변화를 모니터링함
NTL	- 야간 조명 데이터(Night-time Light Data)는 1990년대 이후의 지구 표면의 야간 위성 사진을 제공하며, 이는 도시화, 경제 활동 및 인구를 포함한 사회 경제적 매개 변수와 상관관계가 있는 것으로 나타남
OSM	- 오픈스트리트맵(OpenStreetMap)은 오픈 클라우드소싱 세계 지도로, 휴대용 GPS 장치를 통해 지도 데이터가 수집됨 - 누구나 데이터에 접근, 사용 및 공유할 수 있는 개방형 라이선스로 게시됨
Sentinel	- 유럽의 코페르니쿠스 프로그램을 위해 방대한 양의 데이터와 이미지를 전달하도록 설계된 유럽우주국(ESA, European Space Agency)의 위성임 - 센티넬 1과 2는 육상 및 해양 모니터링을 위한 주야간 레이더 영상과 다중 스펙트럼 고해상도 영상을 제공함
SRTM	- 고해상도 지형 데이터를 제공하고 전 세계적인 규모의 디지털 지형 모델을 제공함
WorldPop	- 인구 분포에 대한 현대 데이터를 이용해 지역 인구 분포, 구성, 특성, 성장 및 역학을 정확하게 측정하는 서비스를 제공함
WorldClim	- 높은 공간해상도(1km)를 지닌 날씨 및 기후 정보의 데이터베이스임

3 SDG 모니터링을 위한 빅데이터 혁신의 과제

- ▶ 많은 빅데이터 소스와 데이터 세트가 새롭게 등장함에 따라 이를 국가별 모니터링 및 보고 과정에 최적으로 적용하는 방법을 명확히 할 필요성이 증가하고 있음
- 여러 데이터 세트의 주체는 단순히 학계를 넘어 국제 기구, 우주 기관, 시민 사회 조직 등 다양하지만 통계청과 관련된 사례는 거의 없고 각국 정부가 통계 보고를 지원하거나 보완하기 위해

새로운 데이터 세트를 어떻게 사용할 수 있는지 명확하지 않은 상황이며 이에 따라 개선이 필요함

- 개방형 데이터 저장소와 오픈 소스 소프트웨어 등 확장 가능한 애플리케이션과 툴을 사용하면 글로벌 연구와 빅데이터에서 파생된 새로운 데이터 세트 간의 격차를 해소하고, 각국 정부가 통계 보고를 위해 데이터 세트를 활용할 수 있도록 하는 데 도움이 될 수 있음
 - 유엔 글로벌 빅데이터 워킹 그룹(The UN Global Working Group on Big Data)이 개발한 마켓플레이스(The Marketplace)는 각국 정부가 빅데이터에서 파생된 새로운 데이터 세트를 통합, 표준화 및 배포할 수 있도록 하는 플랫폼으로 기능하며, SDG 모니터링에 대한 지침도 제공함
 - 그러나 각국 정부에서 데이터 세트의 해석, 검증, 이해 및 활용 등과 관련한 문제가 남아 있어 국제 기구 및 국제 정보 시스템의 관여가 필요함
- 빅데이터에서 파생된 데이터 세트는 현상이 발생할 특정 가능성을 제시하는 추정치이기 때문에 지금의 결과물은 잘 설계된 조사 데이터에 비해 품질이 떨어지며 대표성과 정확성, 신뢰성이 보장되지 않음
- 전문가들이 검토한 과학적인 방침을 따라 통계 데이터를 수집하고 기존 데이터 소스와 통합하여 정확성과 신뢰성을 보장하는 동시에 세부적인 부분을 조정하고, 적용 범위 및 적시성을 개선할 수 있음
- 데이터 세트는 적용되는 국가의 상황이 고려되지 않고, 교정 및 검증 절차가 투명하게 이루어지지 않기 때문에 각국 정부가 모니터링 및 보고 과정에 적용하는 데 어려움을 겪을 수 있음
- 일부 국가는 국가 우주 기관 및 지구 과학 연구 기관을 통해 정교한 지리 공간 및 데이터 과학 기능을 보유해도 통계청에서 엄격한 표준을 적용하고 데이터 소스 채택을 거부해 데이터 활용이 어려울 수 있음을 보여줌
 - 따라서 SDG 모니터링 및 의사 결정을 지원하기 위해 지구 관측 데이터를 사용하는 데 도움이 되는 도구와 지침을 제공하는 국토개발목표 모니터링을 지원하기 위한 토지 열화 매핑(Mapping Land Degradation to Support National SDGs Monitoring)과 같은 글로벌 이니셔티브가 진행 중임
 - 이러한 접근 방식은 SDG에 대한 글로벌 모니터링의 격차를 좁히고 자체 데이터 세트를 생성하고 검증하는 국가 역량을 구축하는 이중 목표를 충족하도록 보장하는 데 유용함
 - 국제 및 지역 조직[유엔(The United Nations), 개발 은행(Development Bank) 및 유럽 위원회(The European Commission), 우주 기관, 국가 정부 기관, 연구 기관, 비정부 기구 및 싱크탱크[국제 자연 보전 연맹(International Union for the Conservation of Nature) 등], 민간 부문[구글(Google), 오렌지 텔레콤(Orange Telecom) 등] 등에서 통계청이 빅데이터를 SDG 모니터링에 활용할 수 있도록 지원하는 것이 매우 중요함

PART **IV**

SDG 모니터링을 위한 빅데이터 활용 파트너십

1 빅데이터 프로젝트의 협력자 유형

- ▶ SDG 모니터링을 지원하기 위한 빅데이터 프로젝트에는 정부, 민간 기업, 학계, 국제 기구, 시민 사회 및 기술 서비스 제공업체 등 여러 이해관계자와 협력자가 참여함
- 이러한 새로운 데이터 생태계는 기존 데이터 모델을 개선하며, 통계청이 새로운 데이터 소스를 활용하고 새로운 플랫폼을 활용하며, 새로운 협력자와 협력하도록 해야 함
 - 데이터 및 통계에 대한 파트너십의 중요성이 증가하는 현상은 공공-민간 파트너십, 비즈니스 모델 및 민간 부문 인센티브, 스마트 데이터 전략 및 역량, 빅데이터 통계 프로젝트에서의 파트너십 구축 및 활용과 같이 최근의 연구 및 지침에서 분명하게 드러나고 있음
- 협력자 유형, 파트너십에 대한 이익, 접근 및 비즈니스 모델, 빅데이터 소스 사용에 대한 고려가 필요하며, 빅데이터 파트너십과 프로젝트를 개발하기 위한 여러 가지 주요 과제와 장벽도 존재함
 - 이는 전반적으로 빅데이터의 유형, 데이터의 소유권, 작업 환경, 기능 및 법적 상황에 따라 크게 차이가 있으며, 국가마다 데이터의 성숙도에 차이가 있음
 - 효과적인 파트너십을 구축하기 위해서 파트너십 환경, 주요 장벽 및 과제, 효과적인 완화 조치를 이해하는 것이 중요함
- 빅데이터를 통한 정보의 수집, 저장, 배포 및 분석이 가능하도록 하기 위한 대상 간 파트너십 환경은 매우 다양하고 복잡함
 - 통계 기관에서는 사내에 빅데이터 기술 전문가가 없을 때 데이터 제공자 외에도 기술 서비스 및 기술 제공자와 협력해야 할 필요가 있을 수 있으며, 초기에 빅데이터 소스를 탐색하기 위해 지식 중개자 및 기관과 상의해야 함
 - 빅데이터 프로젝트를 시작하기 위해서 추가 자금이 필요할 경우가 많기에 경영진 후원자 및 자금을 제공하는 외부 후원자의 도움이 필요함
 - 법률 및 데이터 개인정보보호 담당자의 역할도 규제 환경을 탐색하는 과정에서 필요함

| 표 7 | 빅데이터 프로젝트의 협력자 유형

협력자 유형	역할	예시
데이터 사용자	- 국가 통계의 관리자와 사용자	통계청, 정부/정책 입안자, 비정부 기구, 국제기구
데이터 공급자	- 빅데이터 소스의 공공 및 개인 소유자로, 오픈 액세스 및 영리 목적으로 데이터를 제공함	우주국, 통신사, 소셜 미디어 기업, 정부 기관, 비정부 기구
분석 협력자	- 정부 또는 통계청이 수행하지 못하는 설계 및 분석(알고리즘 및 프로그램 개발 등)과 같은 고급 기술을 제공함 - 기본 데이터를 획득하고 원시 데이터를 의도한 데이터 구조로 변환하며, 최종 결과를 생성하는 기술을 적용하는 노하우를 제공함	플로우마인더(Flowminder), 구글(Google), 학술 기관, 연구 기관, 지구과학 기관, 도메인 전문가
기술 협력자	- 독립 실행형 소프트웨어, 전체 플랫폼 또는 하드웨어를 통해 저장공간, 컴퓨팅 및 분석을 위한 제품을 제공함 - 클라우드 기반 솔루션이 포함되며, 분석을 용이하게 하기 위해 솔루션을 교육 및 컨설팅 서비스와 결합하는 경우가 많음	구글, 마이크로소프트(Microsoft), 아마존(Amazon), 에스리(ESRI), 클라우드 컴퓨팅 제공자
지식 중개자/협의회	- 다양한 행위자를 소집하고 조정하며, 협력자 간의 관계를 구축하고 사용 사례에 대한 인식을 제고함	GPSDD, TrendDS, PARIS21, CEPEI, P4R 및 국제기구
기부자/자금 제공자	- 빅데이터 프로젝트에 대한 자금을 지원함	자선 재단, 비정부 기구, 국제기구, 재정부
지도자/후원자	- 인지도 제고, 협력, 자금 지원 도모	정치 지도자, 간부, 민간 부문 지도자
규제 기관	- 데이터 사용을 규제하고 데이터 보호법을 시행하며, 빅데이터 프로젝트를 승인함	개인정보보호위원회 및 정부 규제 기관
법적/프라이버시 협력자	- 데이터 사용, 개인정보보호, 계약 체결 및 데이터 공유에 관한 법률 자문 제공	정부 개인정보보호 담당자, 법률 부서

- 빅데이터를 활용할 수 있는 통계청의 역량과 적절한 파트너십 구축은 주로 데이터 거버넌스 및 기술 수준과 성숙도에 따라 결정되며, 이러한 차이로 인해 운영 환경의 다양한 모델이 등장함
 - 데이터 공급자에 의한 통계의 생산, 제3자에게 데이터 전송 및 처리, 통계청 또는 제3자에게 데이터 접근권 부여 등을 통해 빅데이터를 활용할 수 있음
 - 그러나 이러한 모델은 데이터 처리 작업에 대한 책임 할당이 적절히 수행되고 사용 사례가 잘 이해되며, 기타 초기 장벽이 극복되어야 한다는 조건이 있음
 - 빅데이터 프로젝트에 대한 인식 부족과 어려움, 법률 및 개인정보보호 문제, 기술적 능력 등 빅데이터 소스 사용에 대한 장벽이 있기 때문에 국가와 통계청마다 파트너십 유형에 있어 차이가 발생함

2 빅데이터 성숙도 및 파트너십 모델

- ▶ 빅데이터 성숙도 모델을 통해 각국 통계청의 빅데이터 혁신 상황과 성숙도 다음 단계로 이동하는 데 필요한 단계를 파악할 수 있음
- 해당 프레임워크는 초기 단계에서 성숙 단계까지 기술하며, 성숙도 모델은 SDG 지표, 통합 국가 데이터 시스템 및 조직의 빅데이터 성숙도를 위한 통계청에 관해 고안됨⁵⁾
- 통계청의 속성을 구성하는 데 사용되는 3가지 자질은 빅데이터 사례 인식, 리터러시 및 지원을 의미하는 조직 특성과 인적 자원 능력 및 인프라를 의미하는 기술적 역량, 법적 및 규제 환경을 의미하는 거버넌스로 구성됨

[그림 1] 빅데이터 성숙도 및 파트너십 모델



- 조직 특성의 경우 정부와 통계청 내 빅데이터의 잠재적 적용과 더불어 빅데이터의 강점, 약점 및 한계에 대한 이해 부족까지 다양한 인식을 수반함
- 통계청은 빅데이터 및 사용 사례를 이해하기 위해 사용자 및 전문가와 협력해야 하며, 빅데이터 도입 과정에서 발생하는 신뢰도 문제를 해결하기 위해 데이터 소스의 잠재력에 대한 신뢰를 구축해야 함
- 또한 사용자에게 잠재적 사용 사례를 숙지시키며, 모든 이해당사자를 빅데이터 활용 과정에 참여시켜야 함
- 기술적 역량의 경우 데이터 사이언스, 고성능 데이터베이스 시스템 및 컴퓨팅 장비, 클라우드 컴퓨팅 플랫폼의 기술 세트와 같은 인프라와 빅데이터를 전송, 저장, 처리 및 분석할 인적 자원의 역량이 해당되며, 저소득 국가의 경우 빅데이터 솔루션 이용이 어렵고 특히 재정 자원도 부족함

5) Marcovecchi, Capability Maturity Models as a Means to Standardize Sustainable Development Goals Indicators Data Production, 2018

- 이미 빅데이터 소스를 수집, 저장, 처리, 분석할 수 있는 공공·민간 사업자는 빅데이터에서 통찰력을 도출할 수 있는 인프라와 노하우를 갖추고 있을 가능성이 높음
- 따라서 이러한 데이터 공급자와 분석 협력자 및 기술 협력자는 저소득 국가와 고소득 국가 간 역량 격차를 해소하고 교육을 제공하며, 장기적인 프로젝트를 지속할 수 있는 역량을 구축하는 데 도움을 제공할 수 있음
- 거버넌스의 경우 데이터 수집, 저장 및 공유 방법뿐만 아니라 데이터 공유 준비 및 위험 완화 조치의 적합성을 결정하는 다양한 행위자, 표준 및 규범으로 구성됨
- 일부 국가에서는 통계청이 데이터를 수집하고 공식 통계를 작성하는 광범위하고 명확한 권한을 가지고 있으며, 개인정보보호 및 기밀성의 범위 내에서 공공 및 민간 부문의 데이터에 접근할 수 있는 권한이 부여됨
- 이와는 대조적으로, 일부 통계청은 공공 기관의 데이터에 접근할 수 있는 권한만 갖거나 적절한 입법 권한이 없을 수 있음
- 많은 경우 입법 프레임워크가 기술 발전을 따라잡지 못해 빅데이터 소스와 관련한 법과 제도가 마련되지 않았으며, 데이터 처리와 사용을 다루는 법적 프레임워크에 결함이 있을 수 있음
- 이러한 법적 안전장치가 없으면 빅데이터 소스의 적합성에 문제가 발생할 수 있으므로 각국의 통계청에서 주도해 공식 통계와 SDG 지표를 도출할 수 있도록 먼저 입법 프레임워크를 강화할 필요가 있음

▶ 통계청의 빅데이터 성숙도는 초기, 도입 전, 도입 초기, 발달, 성숙기의 5단계로 분류되며, 이러한 단계를 거치며 통계청은 다양한 협력자 및 파트너십 모델을 필요로 함

- 초기 단계는 빅데이터 이전 환경을 나타내며, 해당 단계에서 통계청은 빅데이터에 대한 인식, 잠재적 활용 사례, 공식 통계 및 SDG 모니터링을 지원하기 위한 준비가 미흡함
- 정부 내에서는 빅데이터의 잠재적 가치에 관심이 있으나, 빅데이터 활용을 위한 데이터 분석 능력이 갖추어져 있지 않음
- 도입 전 단계에서 통계청은 글로벌 파트너십 참여와 후원자 모집, 빅데이터 적용을 위한 조직 역량 개발 등 빅데이터에 대한 탐구를 시작하나, 아직까지 지도부에서 폭넓은 지지를 얻지 못함
- 도입 전 단계에서 통계청은 빅데이터 분석, 프로그램 감독 운영위원회 도입, 개인정보보호 및 윤리적 요구사항 도입 등 실험을 중심으로 빅데이터 혁신을 시작함
- 가나, 필리핀, 콜롬비아와 같은 국가에서는 빅데이터 도입 과정에서 데이터보호법 미준수, 빅데이터에 대한 인식 부족 등으로 문제를 겪고 구글, 플로우마인더 등 민간 부문 데이터 제공업체와 분석 협력자, GPSDD(Global Partnership for Sustainable Development Data)와 같은 지식 중개자, 법률 협력자 등의 도움으로 문제를 해결할 수 있었음
- 이탈리아 통계청은 2013년부터 빅데이터에 관한 탐구를 시작했으며, 유럽 통계 공동체의 여러 지역 협약과 협력 프로젝트에 참여해 성숙 단계로 진전할 수 있었음

- 도입 초기 단계는 일반적으로 획일화된 1-2개의 개념 증명(POC, Proof of concept)⁶⁾을 지니고 있으며, 한 명 이상의 지도자 및 후원자가 있음
 - 해당 단계에서 통계청은 빅데이터에 대한 도입을 적극 추진하며, 해당 범위를 늘리기 위한 계획 및 전략 수립을 진행할 수 있는 팀을 구성함
 - 해당 단계에서 통계청은 다양한 데이터 공급자로부터 데이터를 수집하고, 증가하는 데이터 용량을 활용하기 위해 인프라에 투자하거나 기술 협력자에게 아웃소싱함
 - 통계청은 데이터 접근 및 공유 계약을 체결해 프라이버시 보호와 데이터 공유를 가능하게 하고 데이터 분석 기능을 구축하며, 사내 역량을 보완하기 위해 기술 지원 서비스 및 교육을 제공함
 - 필리핀의 경우 미국의 소프트웨어 기업인 에스리(Esri)를 통해 지리 정보 체계(ArcGIS) 소프트웨어와 PSA⁷⁾교육을 받았으며, 가나의 경우 보다폰(Vodafone)과 플로우마인더와 같은 기술 협력자와 파트너십을 체결해 교육을 받고 빅데이터 기술을 지원받았으며, 국민의 이동 데이터를 제공 받을 수 있었음
- 통계청이 도입 초기 단계에서 발달 단계로 넘어감에 따라 캐즘(Chasm)⁸⁾이 존재함⁹⁾
 - 인식 향상, 빅데이터 사용 능력 및 사용 사례 확인 등 초기 과제는 해결이 되었으나, 지속적인 자금 확보를 포함한 역량 개발 및 거버넌스 맥락에서 캐즘이 지속될 수 있음
 - 설문 조사에 따르면 많은 통계청에서 여전히 빅데이터로 작업할 수 있는 기술적 능력이 부족하며, 급여가 매력적이지 않아 기술 전문가를 유치하기가 어려움
 - 발달 단계로 넘어가기 위해서 빅데이터 프로젝트가 다른 통계 생산 및 데이터 통합 프로그램과 동등하게 취급되고 예산이 할당되어야 하나, 각국 통계청 사례를 보면 발달 단계로 넘어가는 과정에 많은 어려움이 수반됨
 - 이러한 캐즘을 해소하기 위해서 통계청은 올바른 거버넌스, 데이터 아키텍처¹⁰⁾, 데이터 라이프사이클 관리 및 조직 구조를 갖추어야 하며, 프로젝트 시작부터 생산 프로세스 구축을 시작하고 시작부터 마무리까지 전체 가치 사슬을 설정하는 것이 중요함
- 발달 단계와 성숙 단계에 도달한 통계청은 계획된 이니셔티브를 가지고 잘 구축된 빅데이터 프로그램을 보유하고 있음
 - 통계청은 보안, 복구 및 백업, 성능 관리를 통해 분석을 지원하기 위해 다양한 인프라 및 기술을 통합적으로 사용할 수 있으며, 전체 데이터 가치 사슬에 통합된 실시간 데이터를 비롯한 다양한 형태의 빅데이터를 활용함
 - 데이터 공유는 강력한 데이터 거버넌스 정책을 통해 관리되며, 통계청 내에 데이터 분석 진행 상황을 감독 기관을 포함한 거버넌스가 유지됨

6) 제품, 기술, 정보 시스템 등이 조직의 특수 문제 해결을 실현할 수 있다는 증명 과정

7) 정보 기술 서비스 및 컨설팅 회사의 서비스 관련 프로세스와 자원을 관리하고 프로젝트의 효율적인 수행을 지원하는 애플리케이션

8) 데이터 활용 능력 격차가 발생하는 현상

9) Halper & Krishan, TDWI Big Data Maturity Model Guide. The Data Warehousing Institute, 2013

10) 데이터 측면에서 기업 시스템을 처음부터 끝까지 조망하여 시스템의 본질인 데이터를 체계적, 구조적으로 관리하고 설계하는 전 과정

- 이탈리아와 같은 국가의 통계청은 성숙 단계에 해당하며, 이미지 처리를 위한 딥 러닝 기법, 센티넬 2 위성 이미지, 마이크로소프트의 클라우드 컴퓨팅 플랫폼을 활용해 완전히 자동화된 파이프라인을 개발함
 - 이탈리아 통계청은 기업 웹사이트의 웹 스크래핑 데이터, 소셜 미디어 공용 API, 위성 이미지 등을 포함한 개방형 액세스 데이터 소스에 의존함
 - 영국의 빅데이터 팀과 데이터 사이언스 캠퍼스는 명확한 업무 프로그램, 전담 직원 및 리소스, 지속적인 자금 지원을 통해 성숙한 빅데이터 역량을 보여줌
- 그러나, 여전히 데이터 개인정보보호 고려 사항 및 민간 부문 데이터에 접근하는 측면에서 방법론적 문제가 남아있기에 유럽에서는 표준화된 머신러닝 방법, 훈련 데이터 세트 공유, 내외부 검증 방법 등을 개발하기 위해 노력하고 있음
 - 유럽연합은 미가공 데이터에 대한 접근을 용이하게 하기 위한 규제와 프라이버시 보호 기술과 같은 방안을 검토하고 있으며, 유럽 일반개인정보보호법(GDPR, General Data Protection Regulation)을 통해 개인 프라이버시와 집단 이익 사이의 균형을 맞추고자 함

PART V

빅데이터를 활용한 SDG 모니터링 혁신방안

- ▶ 4장의 사례를 통해 배울 수 있는 점은 각 국가와 통계청이 빅데이터 인식 및 사용 능력, 인적 역량 및 인프라, 거버넌스 및 규제 프레임워크 측면에서 성숙도 수준이 다르다는 것임
- 동 보고서는 통계청이 빅데이터 성숙도 및 혁신 과정에 따라 파트너십을 활용할 수 있는 단계를 제시하고자 함

[그림 2] 빅데이터 혁신을 위한 파트너십 활용 단계



- ▶ 빅데이터 혁신을 시작하지 않은 통계청의 경우 일반적으로 빅데이터 사용 능력, 기회, 한계 및 위험에 대한 인식이 부족하기에 글로벌 및 지역 통계 회의 및 데이터 포럼에 참여하고, 글로벌 및 중개자 또는 조직과 협력하여 문제를 해결해야 함
- 통계청은 SDSN, TrenDS, GPSDD, Data For Now, P4R, GIZ Data Lab, GEO, 개발 데이터 파트너십(The Development Data Partnership)과 같은 협력자를 통해 파트너십을 구축하고, 빅데이터에 대해 이해하며 고위 관료 또는 경영진의 초기 투자를 이끌어낼 수 있음
- 통계청의 빅데이터 전문가 또는 후원자, 정부 또는 민간 부문의 빅데이터 전문가가 빅데이터

혁신 과정에 참여하는 것도 빅데이터 혁신 과정에 대한 인식 제고에 많은 도움이 됨

▶ 초기 단계를 거쳐 도입 전 단계로 나아가기 위해서 국가 통계를 위한 빅데이터 적용을 시험하고 사용 사례를 구체화해야 함

- 통계청, 데이터 제공자, 규제 기관, 기술 서비스 제공자, 개인정보보호 담당자, 법률 자문 등 다양한 협력자는 사용 사례 개발에 필요한 경우가 많으므로 협력자 간에 인센티브 조정을 위한 설득력 있는 가치 제안과 각 협력자가 지닌 책임의 명확한 설명이 필요함
- 공공-민간 빅데이터 파트너십에 대한 인센티브에는 상호주의, 연구 통찰력, 평판 및 공공 관계 구축, 규제 준수 및 새로운 수익창출원이 포함될 수 있음
- 여러 국가의 사례를 확인하면 빅데이터로 인한 사회적 가치를 입증하는 것이 민간 부문 파트너가 지금까지 빅데이터 프로젝트에 참여하도록 동기를 부여하는 데 큰 역할을 수행했음을 알 수 있음
 - 따라서 시간이 지남에 따라 기업의 사회적 책임과 장기간에 걸쳐 프로젝트를 지원할 수 있는 지속 가능한 비즈니스 모델의 균형을 맞추는 민간 기업의 필요성이 증가할 것으로 보임
 - 각 이해관계자의 인센티브와 운영 상황을 이해하는 것은 가치 제안을 구축하고 협력자 간의 인센티브 조정을 진행하는 것이 중요함

▶ 성공적인 프로젝트를 수행하기 위해 통계청은 법률 및 규제 환경에 집중해야 함

- 예를 들어 휴대폰의 데이터 공유는 민간 기업의 빅데이터 소스이므로 공식화된 데이터 공유 계약이 필요하며, 이해관계자 간 데이터 전송, 저장 및 공유 과정에서 데이터 무단 공개 가능성이 있어 빅데이터 프로젝트에서 개인정보 유출 및 보안 위험이 발생할 수 있음
- 양해각서 또는 계약을 통해 체결되는 파트너십 계약은 데이터 제공자와의 협상이 필요하며, 분석 및 기술 제공자를 포함한 다른 형태의 파트너십 프레임워크를 제공하기 위해 보다 광범위하게 사용될 수 있음
- 규제 환경 및 개인정보보호 요건을 탐색할 수 있도록 데이터 개인정보보호 기관, 규제 기관과 담당자는 물론 통계청 또는 정부 내 법률 상담 부서와 사전에 계약을 체결해야 함
- 이러한 계약 예시로는 데이터 공동 작업을 위한 계약(C4DC, The Contracts for Data Collaboration)이 있으며, 이를 통해 효과적인 데이터 공동 작업을 가능하게 하는 법정 조건에 대한 내용을 확인할 수 있음
 - 데이터 공유 계약(Data Sharing Agreement)은 당사자 간에 데이터를 공유하는 조건을 정리한 문서로, 파트너십 관계를 관리하고 위험을 감독하는 데 필수적임
 - 그러나 데이터 공유 계약은 체결 기간이 길어지는 경우가 잦아 C4DC를 통해 데이터 공유 계약이 형성되는 과정을 문서화하는 정책 리소스를 사용하여 부담을 줄임
 - 온라인 C4DC 라이브러리에는 다양한 SDG 관련 응용프로그램에 사용되는 데이터 공유 계약 자료와 함께 사례 연구 및 보고서가 포함되어 있음

- ▶ 빅데이터 혁신 초기 단계에서는 빅데이터 적용에 필요한 인적 역량 및 인프라 구축을 시작할 기회가 있음
 - 분석 및 기술 협업자는 이 단계에서 사내 역량을 보완하고 확장시킬 수 있도록 기술 지원 서비스와 교육을 제공하는 것이 중요함
 - 지도 제작 기관, 우주 기관, 학술 및 연구 기관, 도메인 전문가와의 관계 구축은 국가 역량 개발을 위해 필수적이며, 통계청은 빅데이터 소스를 활용하기 위해 빅데이터 전문가, 데이터 과학자, 통계학자, 컴퓨터 과학자 등으로 구성된 자문 그룹과 협력해야 함
- ▶ 빅데이터 혁신 초기 단계에서 빅데이터 혁신의 성숙도를 높이기 위해서는 지속적인 투자와 협력이 필요함
 - 빅데이터 소스를 수집, 처리 및 분석하고 새로운 데이터 수집 및 생산 방법을 작업 프로그램에 통합하기 위해서는 숙련된 데이터 과학자가 필요하며, 이를 위해서 지속적인 역량 구축과 교육, 방법론적 발전과 자금 조달이 필요함
 - 데이터 공급자, 분석 협력자, 기술 협력자 및 후원자는 통계청이 빅데이터 숙련도의 격차를 메울 수 있도록 지원해야 함
 - 정부는 또한 데이터 관리자와 관련 정부 기관의 역할과 관련해 빅데이터 접근 및 사용에 대한 규칙을 정하는 프레임워크를 확립하는 데 노력을 기울여야 함
 - 개인정보를 보호하는 동시에 데이터 공유를 가능하게 하는 효과적인 입법 프레임워크를 도입하는 데는 시간이 많이 소요되기 때문에 법률 고문 및 개인정보보호 기관과의 지속적인 소통이 필요함

PART VI

결론

- ▶ 동 보고서는 SDG와 관련한 모니터링 과제를 해결하는 데 있어 빅데이터의 가치를 강조하고 이에 관한 활용 사례와 파트너십을 통한 통계청의 혁신방안에 대해 제시함
 - 동 보고서는 지난 10년간 통계를 바탕으로 SDG 모니터링을 지원하기 위해 빅데이터를 실험하는 국가의 최근 사례를 제시했으며, 이러한 새로운 데이터 세트 생성과 같은 방법론적 발전으로 가능해지고 있다는 것을 증명함
 - 특히 접근 가능한 위성 또는 지구 관측 데이터 소스를 사용하면 다양한 SDG 모니터링 목표를 달성할 수 있음
 - SDG 모니터링을 지원하기 위해 데이터 격차를 메워야 할 필요성을 고려했을 때, 데이터 생태계 전반에서 혁신을 일으키고 활용하는 것이 중요함
 - 최근 SDG 모니터링을 지원하기 위한 빅데이터 소스의 글로벌 및 국가 적용 사례는 정부, 민간 기업, 학계, 국제 조직, 시민 사회 및 기술 서비스 제공업체가 참여하는 복잡한 생태계를 나타냄
 - 이러한 생태계에서 통계청은 기존의 데이터 생산 모델과 다른 새로운 데이터를 활용하고, 새로운 협력자와 협력해야 함
 - 그러나 빅데이터에서 파생된 데이터 세트는 완벽한 데이터가 아니고 추정 데이터이기 때문에 기존 데이터 소스를 완벽하게 대체할 수 없음을 인지하고 기존 데이터 소스와 통합을 통해 정확성과 신뢰성, 정교함, 적용 범위 개선을 목표로 해야 함
 - SDG 모니터링을 위한 빅데이터 소스의 광범위한 효용성을 고려할 때 통계청이 빅데이터 혁신을 조속히 진행하고 성숙도 단계를 높여나가야 함
 - 각 통계청은 빅데이터 혁신 과정에서 다른 단계에 위치하며, 빅데이터 인식 및 사용 능력, 인적 역량 및 인프라, 거버넌스 및 규제 프레임워크 측면에서 성숙도 수준이 상이함
 - 따라서 빅데이터 성숙도 및 협력 모델을 참고해 각 통계청에 맞는 협력자와 파트너십을 체결하고 성숙도를 높일 필요가 있음
- ▶ 우리나라도 빅데이터 기반 SDG 과제 실현을 위한 국제협력, 데이터 세분화, 데이터 혁신 등을 추진 중임
 - 정부는 개발도상국가(베트남, 요르단 등)와의 캐즘을 줄이기 위해 성숙한 빅데이터 활용 국가통계시스템을 전수하고 있음
 - 국내에서는 SDG 데이터 혁신 포럼을 개최하여 데이터 기반 SDG 모니터링과 빅데이터 활용 등 데이터 혁신에 대한 의제가 주로 논의되었음
 - 교통 빅데이터 등 SDG 과제별 주요 데이터 접근법 연구 결과가 발표되었으며, 향후 확대 적용 방안에 대해 논의 예정

참 고 문 헌

- Thematic Research Network on Data and Statistics (TReNDS), Big Data and the Sustainable Development Goals: Innovations and Partnerships to Support National Monitoring, 2021
- Gadomi & Haider, Beyond the hype: Big data concepts, methods, and analytics. International journal of information management, 2015
- Marcovecchi, Capability Maturity Models as a Means to Standardize Sustainable Development Goals Indicators Data Production, 2018
- Halper & Krishan, TDWI Big Data Maturity Model Guide. The Data Warehousing Institute, 2013
- 연합뉴스, 韓, 개도국 디지털 격차 줄이려 'K-통계' 전수 나선다, 2022
- 제1회 SDG 데이터 혁신 포럼. www.sdgdataforum.kr



데이터산업 동향 이슈 브리프

| 발행일 2022년 6월 30일

| 발행처 **K data** 한국데이터산업진흥원

서울시 중구 세종대로 9길 42, 부영빌딩 8층

| 기획 및 편집 데이터산업본부 산업기획팀

| 문의처 Tel: 02-3708-5363, 5364

ISSUE BRIEF

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