

2023년 제7호 (통권3호)



데이터산업 동향 이슈 브리프

ISSUE BRIEF

2023.7

유럽 데이터스페이스 -
데이터 공유 및 활용에 대한 과학적 통찰

유럽 데이터스페이스 - 데이터 공유 및 활용에 대한 과학적 통찰

I. 개요	1
II. 유럽 데이터스페이스 구축 배경	3
III. 공통 유럽 데이터스페이스에 대한 요구사항	7
IV. 데이터스페이스 가이드	10
V. 결론	23

요약

- 본 데이터산업 동향 이슈브리프 2023년 제7호는 유럽연합집행위원회(European Commission)가 "유럽 데이터스페이스 - (규모에 따른) 데이터 공유 및 활용에 대한 과학적 통찰(EUROPEAN DATA SPACES / Scientific insights into data sharing and utilisation at scale)"이라는 명칭으로 2023년 발간한 보고서를 요약·정리함
- 원문 보고서는 유럽연합집행위원회가 "디지털 시대에 적합한 유럽"이라는 목표를 세우고 공정하고 신뢰할 수 있는 데이터 공유를 보장하기 위해 사회 전반 및 공공 영역에서 공통 유럽 데이터스페이스(Common European data spaces)를 구축해 나가는 과정에 대해서 서술함
- 유럽연합은 2030년까지 유럽의 디지털 전환을 위해 디지털 10년(Digital Decade) 정책 프로그램을 운영하고 있음
- 데이터스페이스 구축을 지원하기 위해 데이터 '거버넌스법(The Data Governance Act)', '데이터법(The Data Act)', '고부가가치 데이터세트 개방을 위한 시행법(The Implementing Act on High Value Datasets)', '디지털 시장법(The Digital Markets Act)' 등 여러 부문에서 법안이 도입되고 있음
- 공통의 유럽 데이터스페이스를 구축하는 과정에서 적용할 수 있는 단일 기술적 또는 조직적 접근 방식은 존재하지 않으므로 데이터 공유를 위한 중앙 집중식 거버넌스 모델을 적용할 수 없는 한계점이 있었음
- 이를 해결하기 위해서는 데이터 유형, 데이터 흐름, 비즈니스 모델 및 이해관계자 요구 측면 등 부문별 특수성을 고려한 공동의 접근 방식을 마련해야 함
- 원문 보고서("EUROPEAN DATA SPACES")는 이에, 유럽연합 법률을 기반으로 공통의 유럽 데이터스페이스에서 필수적인 핵심 원칙 및 요구사항에 대해서 서술함
- 데이터 공유 과정에서 이러한 핵심 원칙은 데이터 가치 사슬 전체에 적용되어야 하며, 이해관계자는 핵심 원칙을 기반으로 공통 유럽 데이터스페이스의 생성, 유지 및 관리에 참여해야 함
- 원문 보고서는 유럽 데이터스페이스에서 공통으로 필요한 사항을 분석하여 기능 및 비기능적 범주로 분류하였으며, 공통 유럽 데이터스페이스를 운영할 때 발생할 수 있는 문제를 정리하고, 이러한 문제를 이해관계자가 어떻게 해결할 수 있을지 기술적, 거버넌스 및 조직적 측면에서 서술함

- ▶ 유럽연합집행위원회(European Commission)는 “디지털 시대에 적합한 유럽”이라는 목표를 세우고 공정하고 신뢰할 수 있는 데이터 공유를 보장하기 위해 사회 전반 및 공공 영역에서 **공통 유럽 데이터스페이스(Common European data spaces)**을 구축하는 비전을 제시함
 - 유럽연합은 2030년까지 유럽의 디지털 전환을 위해 디지털 10년(Digital Decade) 정책 프로그램을 운영하고, 데이터스페이스 구축을 지원하기 위해 데이터 ‘거버넌스법(The Data Governance Act)’, ‘데이터법(The Data Act)’, ‘고부가가치 데이터세트 개방을 위한 시행법(The Implementing Act on High Value Datasets)’, ‘디지털 시장법(The Digital Markets Act)’ 등 여러 부문에서 법안을 도입하고 있음
- ▶ 그러나 공통의 유럽 데이터스페이스를 구축하는 과정에서 적용할 수 있는 단일 기술적 또는 조직적 접근 방식은 존재하지 않으므로 데이터 공유를 위한 중앙 집중식 거버넌스 모델을 적용할 수 없음
 - 부문별로 데이터 유형, 데이터 흐름, 비즈니스 모델 및 이해관계자 요구 측면에서 모두 차이가 있으므로 부문별 특수성을 고려하는 공동 설계 접근 방식이 유일한 해결책임
 - 마찬가지로, 단일 아키텍처 및 표준 스택을 모든 부문에 보편적으로 적용할 수 없으므로 웹을 통해 분산적으로 활용할 수 있는 최소한의 프로토콜 및 사양 스택을 설정하는 것이 바람직함
 - 관할 기관인 유럽 데이터 혁신 위원회(The European Data Innovation Board)는 이러한 기술과 표준을 선택하는 과정에서 중심적인 역할을 할 예정임
- ▶ 원문 보고서(“EUROPEAN DATA SPACES, JRC SCIENCE FOR POLICY REPORT”)는 앞서 언급한 유럽연합 법률을 기반으로 공통의 유럽 데이터스페이스에서 필수적인 핵심 원칙 및 요구사항을 설명하고자 함
 - 데이터 공유 과정에서 이러한 핵심 원칙은 데이터 가치 사슬 전체에 적용되어야 하며, 이해관계자는 핵심 원칙을 기반으로 공통 유럽 데이터스페이스의 생성, 유지 및 관리에 참여해야 함
 - 또한, 공통 유럽 데이터스페이스에서 필요한 사항을 분석하여 기능 및 비기능적 범주로 분류함
 - 기능적 요구사항은 데이터스페이스가 제공해야 하는 요소를 지칭하며, 비기능적 요구사항은 데이터스페이스가 작동하는 방식을 지칭함

- 한편, 공통 유럽 데이터스페이스를 운영할 때 발생할 수 있는 문제를 정리하고, 이러한 문제를 이해관계자가 어떻게 해결할 수 있을지 기술적, 거버넌스 및 조직적 측면에서 해답도 제시함
 - 예를 들어, “데이터스페이스에 가장 적합한 소프트웨어 스택을 선택하는 방법은 무엇인가?”, “이해관계자가 데이터스페이스에서 합성 데이터를 활용할 수 있는 방법은 무엇인가?”, “데이터스페이스에서 데이터를 쉽게 검색하는 방법은 무엇인가?”와 같은 질문을 하고 이에 대한 답변을 제시함

④ 공통 유럽 데이터스페이스는 데이터의 투명한 접근, 공유 및 사용을 보장하기 위해 적절한 거버넌스 구조를 마련해야 함

- 데이터스페이스 지원센터(The Data Spaces Support Centre)¹⁾는 거버넌스에 대해서 조직을 지시하고 관리하기 위한 규칙, 관행 및 프로세스 시스템으로 정의함
- 거버넌스는 다양한 이해관계자가 데이터스페이스에 참여하는 데 있어 필수적인 요소로, 데이터스페이스 내에서 서로 다른 수준의 거버넌스가 필요함
 - 공통 유럽 데이터스페이스의 전략과 가치를 설정하고 개방적인 참여를 보장하는 측면에서는 광범위한 거버넌스 구조가 필요한 반면, 표준 설정과 프로세스 확립과 같은 보다 기술적인 운영 측면에서는 더 면밀하고 집중된 거버넌스 구조가 필요함
- 데이터 거버넌스는 각 이해관계자의 명확한 의무 및 책임을 개략적으로 설명해야 함
 - 또한, 데이터 공유 및 이동성을 향상하기 위해 개방성을 지원하는 동시에 유럽연합 법률 및 표준에 따라 데이터가 적절히 보호되도록 보장해야 함

⑤ 원문 보고서는 총 5개의 장으로 구성됨

- 1장은 서론으로서 유럽 데이터스페이스의 원칙, 요구 사항 및 기능 등에 대해 간략하게 설명함
- 2장은 유럽연합의 디지털 시대에 적합한 유럽 비전을 이루기 위한 정책의 개요, 특히 유럽의 데이터 전략을 제시함
 - 구체적으로 공통 유럽 데이터스페이스의 개념 및 주요 이해관계자를 설명하고, 이러한 데이터스페이스의 핵심 원칙 및 요구사항, 특징에 관해 주로 다룸
- 3장은 공통 유럽 데이터스페이스에 대한 요구사항을 제시함
 - 유럽연합 정책 조항을 기반으로 한 요구사항을 기능적 및 비기능적 범주로 분류하여 설명함
- 4장은 데이터스페이스 가이드로 구성되며, 데이터스페이스 구축 및 운영 과정에서 발생 가능한 시나리오 및 해결책을 제공함
- 5장은 공통 유럽 데이터스페이스와 관련한 전체적인 내용을 요약하고 주요 결과를 제공함

1) 디지털 유럽 프로그램(Digital Europe Program)의 일환으로 유럽연합 집행위원회(European Commission)가 자금을 지원하는 조직으로, 주권 데이터 공간을 만들고자 하는 공공 부문 및 기업을 대상으로 지원(<https://dssc.eu>)

PART II

유럽 데이터스페이스 구축 배경

- ▶ 유럽연합은 데이터 주도 혁신의 잠재력을 달성하기 위해 공통의 유럽 데이터스페이스를 구축하고자 함
 - 공통 유럽 데이터스페이스는 다양한 이해관계자가 공통 거버넌스, 조직, 규제 및 기술 메커니즘에 따라 안전하고 신뢰할 수 있는 방식으로 데이터를 공유, 접근 및 사용할 수 있는 분산형 인프라임
 - 공통 데이터스페이스는 수요 중심 프로세스를 통해 다양한 데이터 에코시스템을 연결하며, 여러 이해관계자가 데이터를 문제없이 교환하는 데 핵심적인 역할을 할 것으로 기대됨
 - 유럽연합 집행위원회는 2030년까지 유럽의 디지털 전환 비전을 제시했으며, 이는 데이터와 기술, 인프라에 대한 구체적인 조치를 통해 디지털 주권 강화를 목표로 함
 - 또한, 2023년에 발간된 ‘The Annual Single Market Report’는 데이터스페이스의 상호 운용성이 보장되는 데이터 기반 단일시장을 통해 단일 유럽연합 데이터 경제를 창출하려는 목표를 재차 강조함²⁾
- ▶ 유럽연합은 전반적인 거버넌스 프레임워크를 마련하고 각 부문, 이해관계자 및 도메인 간의 데이터 공유 문제를 해결하기 위해 여러 법안을 도입함
 - ‘데이터 거버넌스법’은 데이터 중개자에 대한 신뢰를 높이고 유럽연합 전체에 걸쳐 다양한 데이터 공유 메커니즘을 강화함으로써 데이터의 더 많은 재사용을 촉진하는 프레임워크 제공을 목적으로 하는 법안임
 - 데이터 거버넌스법은 보건 데이터와 같이 민감한 공공 부문 데이터의 재사용³⁾ 촉진, 공통 유럽 데이터스페이스 내에서 데이터 중개자의 역할 규제, 기업의 편의성 향상을 통해 공통 유럽 데이터스페이스의 구축 및 개발을 지원함
 - 데이터 거버넌스법은 2020년에 제안되어 2022년 6월에 발효(into force)되었으며, 2023년 9월부터 본격 시행될(will be applicable) 예정임⁴⁾

2) European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, “A European strategy for data”, COM(2020) 66 final, (2020),

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN>

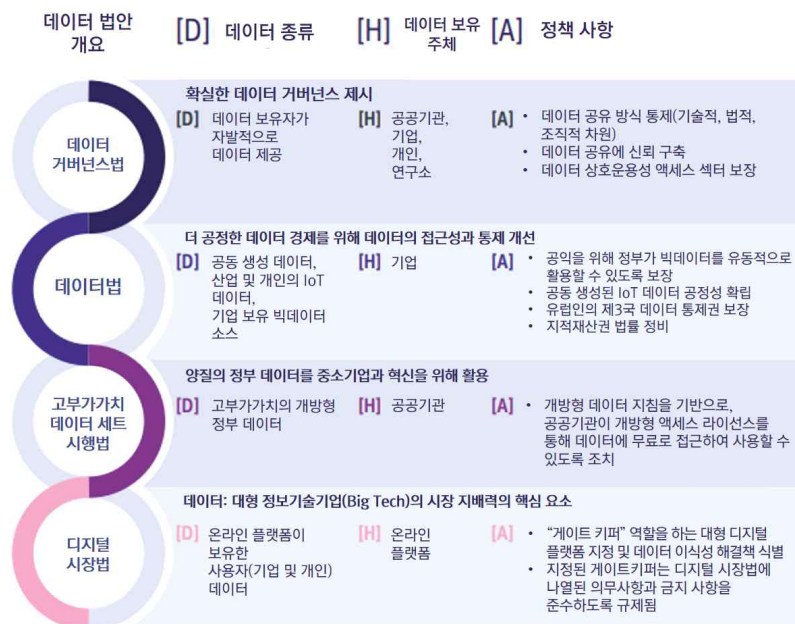
3) 본 보고서는 생성된 데이터에 대해 여러 부문에서 접근해서 공유 또는 사용하도록 허용되는 것을 데이터의 ‘재사용’으로 일괄 표현함

4) European Commission, “Regulation of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)”, Official Journal of the European Union, L 152, (2020), pp. 1-44

<http://data.europa.eu/eli/reg/2022/868/oj>

- ‘데이터법’은 유럽연합에서 생성된 데이터를 어떤 주체가 접근하고 사용할 수 있을지에 대해 구체적인 규칙을 제정하여 여러 산업 부문에서 더 많은 비즈니스 데이터를 재사용할 수 있도록 하는 법안임
 - 데이터법은 사용자가 자신이 생성한 데이터에 접근하고 이 데이터를 제3자와 공유할 수 있도록 허용하며, 비상시 등 특정 상황에서는 공공기관이 민간 기업이 보유한 데이터에 접근하여 사용할 수 있도록 함
 - 또한, 데이터 공유 계약에서 불공정한 계약으로부터 EU 중소기업(SME)을 보호하고자 이들의 재조정 협상력을 높이고, 고객이 서로 다른 클라우드 데이터 처리 서비스 공급자 간에 자유롭게 전환할 수 있도록 하여 시장 경쟁을 촉진함
- ‘오픈데이터 지침(The Open Data Directive)’을 기반으로 하는 ‘고부가가치 데이터 세트 개방을 위한 시행법’은 공공 부문이 보유한 고부가가치 데이터 세트를 정의하는 법안으로, 사회, 경제 및 환경 분야에서 가치를 창출할 수 있는 잠재력이 있음
 - ‘고부가가치 데이터 세트 개방을 위한 시행법’은 공공기관이 개방형 액세스 라이선스를 통해 무료로 제공할 데이터 세트 목록을 지정하고, 애플리케이션 프로그래밍 인터페이스(API, Application Programming Interface)를 통해 접근할 수 있도록 ‘오픈데이터 지침’을 구현하는 법안임
- 디지털 시장법은 디지털 부문의 시장을 보다 공정하고 경쟁적으로 만들기 위해 제정된 법률로, “게이트키퍼” 역할을 하는 대형 디지털 플랫폼을 식별하여 이들의 권한을 포괄적으로 규제하는 도구임

[그림 1] 유럽 데이터 전략 구현을 위한 법적 프레임워크 개요⁵⁾

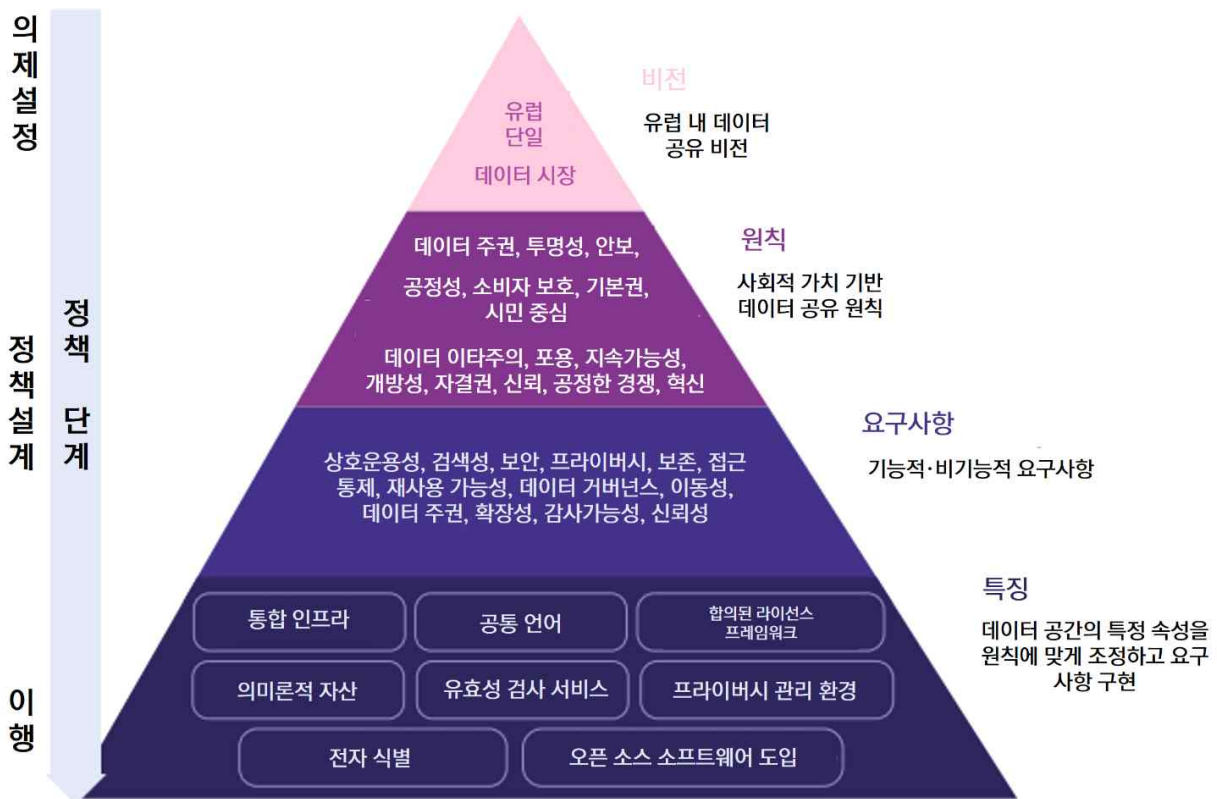


5) European Commission, Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance(Data Governance Act), 2020.11.25.

<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020SC0295&rid=6>

- 유럽연합 집행위원회는 공통 데이터스페이스의 필요조건을 명시한 관리자 작업 문서(SWD, Staff Working Document)를 발행함
 - 공통 유럽 데이터스페이스는 데이터 풀링 및 공유를 용이하게 하기 위해 관련 데이터 인프라 및 거버넌스 프레임워크를 한데 모은 것임
 - SWD에 따르면, 공통 유럽 데이터스페이스는 (i)효율적이고 신뢰 가능한 클라우드 서비스, (ii)데이터 접근 및 처리에 관한 유럽연합 법률을 준수하는 데이터 거버넌스 구조, (iii)부문 간 데이터 가용성, 품질 및 상호 운용성 향상이 필요함
 - 또한, 유럽연합 집행위원회는 데이터 공유를 위한 핵심 원칙을 제시함
 - 그림 2는 다양한 정책 결정 단계(의제 설정, 정책 설계, 이행 단계)에 해당하는 원칙, 요구사항 및 특징을 비롯한 주요 특성을 표현함

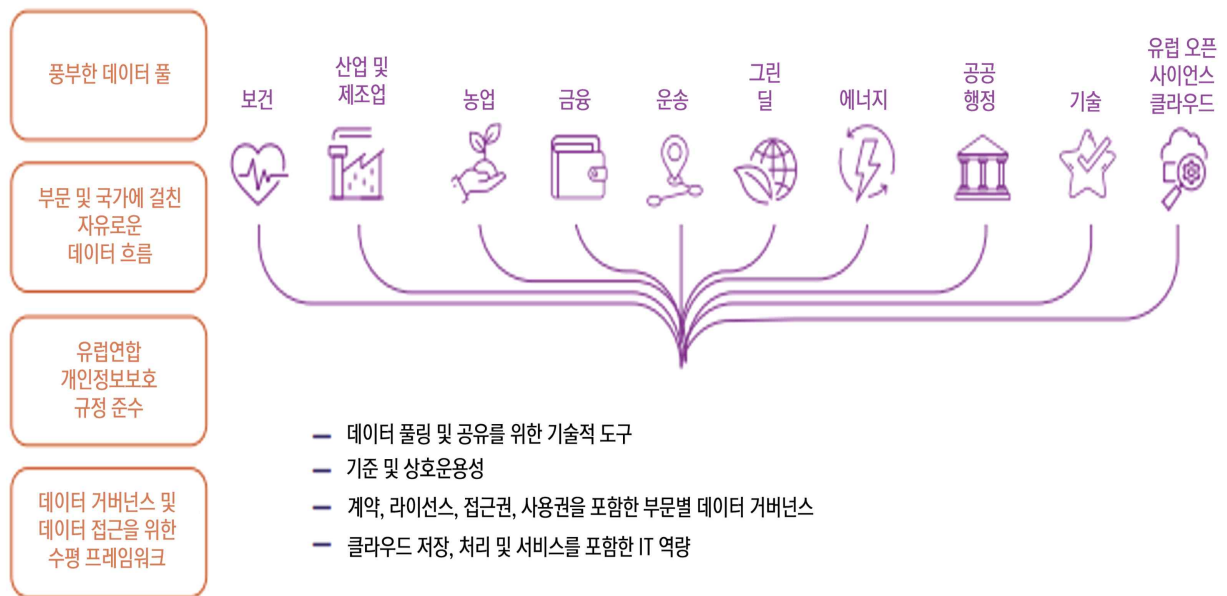
| 그림 2 | 공통 유럽 데이터스페이스 특성



- 유럽연합 집행위원회는 10가지의 부문별 공통 유럽 데이터스페이스를 제시함
 - 10가지 공통 유럽 데이터스페이스는 아래와 같음
 - 질병의 예방, 감지 및 치료 개선뿐만 아니라 의료 시스템을 개선하기 위한 정보 기반 의사 결정에 필수적인 보건 데이터스페이스

- 유럽연합의 산업 경쟁력을 높이기 위한 산업 및 제조업 데이터스페이스
- 데이터 분석을 통해 농업 부문의 지속가능한 성과와 경쟁력을 높이는 농업 데이터스페이스
- 혁신, 시장 투명성, 지속 가능한 금융뿐만 아니라 유럽 기업의 금융 접근성 개선을 위한 금융 데이터스페이스
- 유럽연합을 지능형 교통 시스템 개발의 선두에 올려놓기 위한 운송 데이터스페이스
- 기후 변화, 순환 경제, 오염, 생물 다양성 및 산림 벌채와 같은 문제에 대한 그린딜⁶⁾ 우선 조치를 지원하기 위해 데이터 잠재력을 활용하는 그린딜 데이터스페이스
- 고객 중심적이고 안전하며, 신뢰할 수 있는 방식으로 부문 간 데이터 공유를 촉진하는 에너지 데이터스페이스
- 공공 지출의 투명성과 책임성을 개선하고 유럽연합의 부패 퇴치를 위한 공공 행정 데이터스페이스
- 교육 및 훈련 시스템과 노동 시장의 요구 간 불일치를 줄이기 위한 기술 데이터스페이스
- 신뢰할 수 있고 개방적인 분산 데이터 환경을 조성하고, 관련 서비스를 통해 유럽 연구자, 혁신가, 기업 및 개인에게 연구 데이터 접근과 재사용이 가능하도록 하는 유럽 오픈 사이언스 클라우드(EOSC, European Open Science Cloud) 데이터스페이스

| 그림 3 | 부문별 공통 유럽 데이터스페이스의 분석 개요



6) 미국 IRA, 일본 그린성장전략, 인도 생산연계 인센티브 등 주요국의 친환경 산업 육성 정책에 대응하기 위해 유럽 자체적으로 제시한 친환경 산업전략

PART III

공통 유럽 데이터스페이스에 대한 요구사항

- ▶ 원문 보고서는 유럽연합 정책 문서를 기반으로 공통 유럽 데이터스페이스에 대한 요구사항을 분석함
 - 공통 유럽 데이터스페이스는 기술적 인프라로만 고려해야 할 것이 아니라 기술적, 조직적, 법적 및 거버넌스 측면의 포괄적 계획으로 간주되어야 함
 - 공통 유럽 데이터스페이스는 관리되는 데이터 자산의 특성과 활용되는 애플리케이션의 특성에 따라 설계·실증 과정에서 맞춤형 접근 방식을 취해야 함
 - 아울러, 정책 입안자와 공공기관, 민간 조직, 과학계 및 일반 대중을 아우르는 광범위한 이해관계자에게 맞는 기능을 제공해야 함
 - 공통 유럽 데이터스페이스는 유럽연합이 추진하는 신규 계획이지만, 기존 공공기관 및 민간 조직의 데이터 공유 공간 개발 사례의 노하우를 반영함
 - 따라서, 원문 보고서는 관리자 작업 문서와 같은 유럽연합 정책 보고서를 참고하여 공통 유럽 데이터스페이스에 대한 요구사항을 정리함
 - 원문 보고서에서는 공통 유럽 데이터스페이스에 대한 요구사항을 기능 및 비기능적 범주로 분류함
 - 기능적 요구사항은 데이터스페이스가 제공해야 하는 요소를 지칭하며, 비기능적 요구사항은 데이터스페이스가 작동하는 방식을 지칭함

【표 1】 공통 유럽 데이터스페이스에 대한 기능적/비기능적 요구사항 목록

정책 조항	기능적 요구사항	비기능적 요구사항
<ul style="list-style-type: none"> ● 데이터 통합, 접근, 처리, 사용 및 공유를 위한 개인정보 보호 IT 인프라를 개발해야 함 	데이터 전송 및 교환 데이터 저장 데이터 처리 및 분석 데이터 통합 및 공동 작업	보안 비밀 보호
<ul style="list-style-type: none"> ● 데이터 소유자는 데이터스페이스 내, 특정 개인 또는 비개인 데이터에 대한 접근을 허용하거나 공유할 수 있음 	신원, 인증, 접근 통제 사용 통제 정책	비밀 보호
<ul style="list-style-type: none"> ● 검색·접근·상호운용·재사용 가능(FAIR, Findable, Accessible, Interoperable and Reusable) 원칙과 공통 	데이터 전송 및 교환 신원, 인증, 접근	상호 운용성 감사 가능성

정책 조항	기능적 요구사항	비기능적 요구사항
<p>개방형 표준 개발을 위해 모든 유형의 데이터를 통합, 접근, 처리, 사용 및 공유할 수 있는 도구 개발을 촉진함</p> <ul style="list-style-type: none"> 데이터 소유자는 이러한 도구를 활용하여 데이터스페이스에 데이터를 쉽게 업로드하고 데이터에 대한 권한을 부여하거나 철회하며, 접근 권한을 변경하고 시간이 지남에 따라 접근·재사용 조건을 변경할 수 있음 사용 가능한 데이터는 보상을 통해 재사용하거나 무료로 사용할 수 있음 참여자들은 데이터스페이스를 효율적이고 조율된 방식으로 구축할 수 있는 공통 기술 인프라 및 구성요소를 사용함 공통 기술 인프라는 사이버 보안 원칙을 통합해야 함 	통제 사용 통제 정책	
<ul style="list-style-type: none"> 참여자들은 데이터스페이스를 효율적이고 조율된 방식으로 구축할 수 있는 공통 기술 인프라 및 구성요소를 사용함 공통 기술 인프라는 사이버 보안 원칙을 통합해야 함 	거래 계량 및 명세	유지관리성 가변성
<ul style="list-style-type: none"> 참여 가능한 조직/개인의 수를 보장해야 함 	신원, 인증, 접근 통제	보안 확장성
<ul style="list-style-type: none"> 공통 유럽 데이터스페이스는 국제 표준, INSPIRE* 및 FAIR 원칙에 따라 개발되며 상호 운용성을 보장해야 함 유럽연합 내 컴퓨팅 인프라의 데이터 활용 및 상호 연결, 상호 운용성을 보장해야 함 <p>* 유럽연합 환경 정책 및 환경에 영향을 미칠 수 있는 정책이나 활동을 위하여 유럽연합 데이터스페이스 인프라를 생성하도록 하는 지침으로, 공공기관 간 환경 공간 정보의 공유를 촉진하고 유럽 내 공간 정보에 대한 공공 접근을 활성화하고자 하는 목표가 있음</p>	데이터 상호 운용성 기능 데이터 처리 및 분석	상호 운용성 성능
<ul style="list-style-type: none"> 개인 데이터 보호, 소비자 보호 법률 및 경쟁법과 같은 유럽연합의 규칙을 존중해야 함 	규정 준수 모니터링 및 감사 데이터 보호	감사 가능성
<ul style="list-style-type: none"> 유럽연합의 새로운 데이터 기반 제품 및 서비스 개발을 통해 상호 연결되고 경쟁력 있는 유럽 데이터 경제를 구성함 (데이터스페이스 미들웨어*) 고성능 컴퓨팅과 원거리 컴퓨팅의 미래 통합을 반영하여 에지에서 운영할 수 있는 기본 서비스가 포함된 전체 클라우드 스택을 제공함 공통 서비스 및 기능을 애플리케이션에 제공하여 개발자와 운영자가 애플리케이션을 더욱 효율적으로 구축하고 배포하도록 돕는 소프트웨어 및 클라우드 서비스임 (데이터스페이스 미들웨어) 	데이터 처리 및 분석	자주권
<ul style="list-style-type: none"> 고성능 컴퓨팅과 원거리 컴퓨팅의 미래 통합을 반영하여 에지에서 운영할 수 있는 기본 서비스가 포함된 전체 클라우드 스택을 제공함 공통 서비스 및 기능을 애플리케이션에 제공하여 개발자와 운영자가 애플리케이션을 더욱 효율적으로 구축하고 배포하도록 돕는 소프트웨어 및 클라우드 서비스임 (데이터스페이스 미들웨어) 	다중 지원 연합 및 조정	이동성 성능
<ul style="list-style-type: none"> 모든 공통 유럽 데이터스페이스에서 사용할 기술 기준을 제공하여 중복 수행을 방지하고 다양한 접근 방식을 조정함 (데이터스페이스 미들웨어) 	-	유지관리성 가변성
<ul style="list-style-type: none"> 클라우드와 에지* 간에 최첨단 데이터 관리를 통해 클라우드와 에지 간에 초고속 데이터 워크로드 밸런싱을 지원하고 중앙 집중식 데이터 인프라와 분산형 데이터 인프라 간 지능형 데이터 변환을 가능하게 함 여러 클라우드 및 에지 공급업체의 애플리케이션을 실행할 	데이터 전송 및 교환 다중 지원, 연합 및 조정 신원, 인증, 접근 통제	이동성 성능 보안 유지관리성

정책 조항	기능적 요구사항	비기능적 요구사항
<p>때, 성능 및 서비스 품질을 보장함</p> <ul style="list-style-type: none"> • 내장된 식별 및 보안 관리 서비스를 통해 멀티 클라우드 오케스트레이션 솔루션을 제공함 * 분산 컴퓨팅 환경을 구현하는 데 발생하는 여러 문제점을 해결하기 위해 등장한 소프트웨어로, 공통 서비스 및 기능을 애플리케이션에 제공하여 개발자와 운영자가 애플리케이션을 더 효율적으로 구축하고 배포하도록 돕는 클라우드 서비스임 (데이터스페이스 미들웨어) 		
<ul style="list-style-type: none"> • 데이터 매핑 서비스, 데이터 익명화 및 마스킹* 서비스를 제공함 * 카메라로 촬영되거나 온라인으로 전송되는 이미지에서 민감한 정보가 되는 부분을 암호화하여 식별될 수 없도록 처리하거나, 문자나 숫자로 구성된 정형 데이터에서 특정 항목의 일부 또는 전부를 공백, 문자(특수 문자 포함)로 대체함으로써 개인정보가 노출되는 사고를 방지하는 기술 (데이터스페이스 미들웨어) 	개인정보 보호 구조 데이터 상호 운용성 기능	비밀보호 상호 운용성
<ul style="list-style-type: none"> • 교차 편집, 저전력 및 소프트웨어 지원 서비스를 기반으로 하는 다중 사용 비즈니스 인텔리전스 서비스를 제공함 (데이터스페이스 미들웨어) 	데이터 처리 및 분석	-
<ul style="list-style-type: none"> • 환경 추적 성능 시스템을 통합하여 서비스가 저전력 모드에서 운영되도록 보장함 (데이터스페이스 미들웨어) 	-	에너지 효율
<ul style="list-style-type: none"> • 안전하고 자원 효율적인 데이터 저장 서비스를 제공함 (데이터스페이스 미들웨어) 	데이터 저장	-
<ul style="list-style-type: none"> • 서비스형 고성능 컴퓨팅(High Performance Computing as a service) 연결 장치를 제공하여 클라우드 연합 사용자가 고성능 컴퓨팅 자원에 접근할 수 있도록 지원함 (데이터스페이스 미들웨어) 	다중 지원, 연합 및 조정	성능 이동성
<ul style="list-style-type: none"> • 인공지능 솔루션이 미들웨어 플랫폼에서 작동하도록 보장하고, 지속가능하고 대기 시간이 매우 짧은 디지털 트윈의 비즈니스 애플리케이션을 지원함 • 복잡한 비즈니스 활동 시뮬레이션, 예측 및 모델링을 위한 전문화된 도구 호스팅을 허용함 (데이터스페이스 미들웨어) 	데이터 처리 및 분석	성능
<ul style="list-style-type: none"> • 안전한 통신, 생산성 및 협업 서비스를 제공함 워크플로우 관리 서비스를 제공함 • 클라우드 투 엣지 서비스와 마켓플레이스 통합을 촉진함 	데이터 통합 및 공동 작업	이동성
<ul style="list-style-type: none"> • 데이터 통합 및 공유를 촉진하기 위해 관련 데이터 인프라 및 거버넌스 프레임워크를 통합하는 공통 유럽 데이터스페이스를 제공함 • 공정하고 투명하며, 차별 없는 방식으로 데이터에 접근하여 사용할 수 있는 실용적인 데이터 거버넌스 구조를 제시함 	데이터 통합 및 공동 작업 사용 제어 정책 개인정보 보호 구조 데이터 보호 데이터 거버넌스	포용 공정성 지속 가능성 신뢰성 투명성

PART IV

데이터스페이스 문제 해결을 위한 가이드

- 원문 보고서는 데이터스페이스를 개념화하고 운영하는 과정에서 발생할 수 있는 문제를 해결하기 위한 데이터스페이스 가이드를 제시함
- 해당 데이터스페이스 가이드는 데이터를 공유하고 데이터에 접근하는 데 필요한 기술적 및 조직적 측면을 자세히 설명함
- 보고서는 크게 기술적 측면 10가지 시나리오와 조직적 측면 7가지 시나리오를 설정하여 이에 대한 해결책을 제시함

표 2 | 데이터스페이스 가이드 시나리오 목록

	시나리오 목록	데이터스페이스 주제
기술적	1. 이해관계자가 데이터스페이스에서 합성 데이터를 활용할 수 있는 방법은 무엇인가?	합성 데이터
	2. 데이터스페이스에 가장 적합한 소프트웨어 스택을 선택하는 방법은 무엇인가?	소프트웨어 스택
	3. 데이터스페이스에서 데이터 세트에 대한 명확한 접근 및 사용 조건을 보장하는 방법은 무엇인가?	인허가
	4. 데이터스페이스에서 서로 다른 이해관계자가 공유하는 데이터 세트를 함께 사용할 수 있도록 보장하는 방법은 무엇인가?	상호 운용성
	5. 기술 요구사항과 표준이 준수되고 있는지 확인하는 방법은 무엇인가?	데이터 인증
	6. 데이터스페이스에서 데이터를 쉽게 검색할 수 있는 방법은 무엇인가?	데이터 검색 가능성
	7. 데이터스페이스에 가장 적합한 표준을 선택하는 방법은 무엇인가?	데이터 기준
	8. 데이터스페이스에서 디지털 자원과 데이터가 참조 표시되도록 하는 방법은 무엇인가?	데이터 기록부
	9. 데이터스페이스에서 데이터에 대한 접근을 제공하는 방법은 무엇인가?	데이터 접근을 위한 응용프로그램 인터페이스
	10. 데이터스페이스에서 개인정보와 개인 데이터 및 비즈니스 데이터를 보호하는 방법은 무엇인가?	개인정보 보호 제고 기술
조직적	1. 데이터스페이스에서 어떤 이해관계자가 어떤 유형의 데이터를 제공하는가?	데이터 이해관계자
	2. 데이터스페이스에서 사람 중심의 데이터 접근 방식을 육성하는 방법은 무엇인가?	개인 데이터
	3. 데이터스페이스에서 데이터를 공유함으로써 비즈니스에 미칠 수 있는 장점은 무엇인가?	데이터스페이스에서의 비즈니스 이익
	4. 정부가 공익을 위해 어떻게 민간 부문 데이터에 접근할 수 있는가?	B2G 데이터 접근
	5. 데이터스페이스에서 인공지능 시스템의 데이터 투명성을 어떻게 높일 수 있는가?	데이터스페이스의 인공지능 데이터 투명성
	6. 데이터스페이스에서 자발적인 데이터 공유를 활용하는 방법은 무엇인가?	자발적인 데이터 공유
	7. 데이터스페이스에서 새로운 데이터 중심 솔루션을 생성, 제공 또는 사용할 때 고려해야 할 법적 측면은 무엇인가?	법적 측면

(기술적 측면)**1. 이해관계자가 데이터스페이스에서 합성 데이터를 활용할 수 있는 방법은 무엇인가?**

- 공통 유럽 데이터스페이스는 여러 데이터 소스를 활용하기 때문에 개인정보 보호 문제, 인허가 문제, 데이터 대표성 부족, 데이터 분절화 및 누락, 데이터 부족 등의 문제가 발생할 수 있음
 - 예를 들어, 의료계는 개인정보보호를 위해 암 환자의 기록을 공유할 수 없으나 환자 기록을 공유하면 의학계와 연구 기관에서는 환자 데이터를 얻고 의료 서비스를 개선할 수 있음
 - 정책의 경우 평균적인 인구 통계를 기반으로 수립되기 때문에 소수 집단이나 취약 계층은 정책 수립 과정에서 배제되는 경우가 발생할 수 있음
 - 자율 주행 차량의 인공지능 알고리즘을 훈련하는 시나리오의 경우 차량의 센서, 카메라, 레이저 등의 시스템을 통해 얻는 데이터가 불충분해 모든 사고 가능성에 대비할 수 없음
- 해당 경우, 원본 데이터 세트에서 무작위로 생성된 데이터인 합성 데이터를 통해 해결할 수 있음
 - 개인정보 보호를 위해 환자 기록을 공유할 수 없는 경우, 합성 데이터를 통해 원본 데이터의 특성만을 복제해 개인정보를 보호하는 동시에 의료 연구를 수행할 수 있음
 - 인구 조사 데이터와 추가 데이터 소스(행동 조사, 이동성 패턴 조사, 전염병 접촉 조사 등)를 교차하여 합성 데이터를 도출하면 정책 수립 과정에서 소수 집단을 반영할 수 있음
 - 게임 엔진을 활용한다면 시뮬레이션을 통해 합성 데이터를 생성할 수 있으며, 필요한 데이터를 추가적으로 확보할 수 있음
- 합성 데이터조차도 재식별⁷⁾의 위험이 있기 때문에 올바른 설계가 필요함
 - 데이터가 실제 모집단에 가까울수록 분석 목적에 대한 가치가 높지만, 재식별의 위험이 커짐
 - 따라서 합성 데이터를 생성할 수 있지만 높은 비용이 발생할 수 있기 때문에 합성 데이터 생성 이전에 경제성 평가를 수행하는 것이 중요함

2. 데이터스페이스에 가장 적합한 소프트웨어 스택을 선택하는 방법은 무엇인가?

- 데이터스페이스와 같은 데이터 공유 인프라는 소프트웨어 스택이라고 하는 여러 소프트웨어 요소로 구성됨
 - 소프트웨어 스택은 데이터 저장, 게시, 데이터 보호 및 접근 제어, 데이터 검증, 처리 및 분석, 조정, 이동성과 같은 필요한 기능을 제공함
 - 이러한 기능 수행을 위해 수많은 소프트웨어 도구가 존재하므로 소프트웨어 스택이 분절될 위험이 있으며, 특정 데이터스페이스에 가장 적합한 소프트웨어 스택을 선택하는 것이 어려워짐
 - 일례로, 환경 영향 평가를 수행하기 위한 유럽 그린딜 데이터스페이스에서는 시민의 데이터, 코페르니쿠스 위성 이미지, 공공 데이터와 같은 여러 범주의 데이터에 인공지능 알고리즘을 사용하여 접근, 결합 및 처리해야 함
 - 이를 위해서는 추출·변환·로딩(Extract, Transform, Load) 도구와 인공지능 알고리즘

7) 개인정보를 다른 정보와 조합, 분석 및 처리하여 특정 개인을 다시 식별할 수 있도록 하는 과정

교육 및 테스트를 위한 특정 라이브러리를 사용해야 함

- 이때 응용프로그램 인터페이스, 메타데이터 편집기 이용 및 국제 표준 소프트웨어 구현이 필요하며, 민감한 개인 데이터에 대한 접근 제어 및 개인정보 보호 구조도 마련해야 함

● 이때 오픈소스 소프트웨어⁸⁾가 해결책으로 이용될 수 있음

- 데이터스페이스에 가장 적합한 소프트웨어 스택을 선택하기 위한 일반적 해결책이 없기 때문에 오픈소스 소프트웨어를 통해 데이터스페이스의 요구에 맞춰 문제를 해결할 수 있음
- 오픈소스 소프트웨어는 대개 무료로 기존 소스 코드에 대한 접근성을 제공함
- 유럽연합 집행위원회는 스마트 미들웨어 플랫폼(SIMPL, Smart Middleware Platform)을 발표하여 데이터스페이스 구축을 위한 오픈소스 기술 지원 플랫폼을 제공할 예정임

● 소프트웨어 스택을 선택할 때 아래와 같은 기준을 활용할 수 있음

- 국제 표준, 깃허브(Github)와 같은 코딩 플랫폼의 유무, 소프트웨어 코드 베이스 품질 확인을 위한 지침, 유지보수의 용이성 및 기술 발전 가능성

🟡 3. 데이터스페이스에서 데이터 세트에 대한 명확한 접근 및 사용 조건을 보장하는 방법은 무엇인가?

● 데이터 세트는 대부분 저작권 정보가 불분명하거나 누락된 상태로 제공되기 때문에 데이터스페이스에서 데이터 사용자가 데이터를 어떻게 사용할 수 있을지 불확실하며, 데이터의 재사용이 비효율적으로 이루어짐

- 데이터스페이스 사용자가 파생 데이터 세트 또는 서비스를 생성하기 위해 기존 데이터 세트를 결합하거나 인공지능 알고리즘을 사용하려는 경우 저작권이 장애물로 작용할 수 있음
- 예를 들어, 데이터스페이스 사용자가 서로 다른 조직에서 생성한 데이터 세트를 결합하여 국경 간 환경 위험 평가를 수행하려고 하는 경우 불분명한 데이터 세트 재사용 설명으로 인해 데이터 세트를 사용하지 못할 수 있음
- 저작권에 대한 명확한 정보가 부족하면 저작권 침해 또는 데이터 접근 조건을 위반할 가능성이 있음

● 여기에는 크리에이티브 커먼즈(CC, Creative Commons)⁹⁾ 라이선스 제품군과 같은 데이터 세트를 활용하는 방안이 있음

- 크리에이티브 커먼즈는 전 세계에서 널리 활용되고 있으며, CC0 또는 CC-BY 라이선스(저작권자를 명시해야 함)를 사용하면 파생 제품 및 서비스에 포함되는 데이터 세트 라이선스를 확인할 수 있음
- 아울러, 자동화된 데이터 교환 및 사용을 촉진하기 위해 기계 판독 가능한 라이선스(검색 엔진과 같이 기계에 의한 자동 검색을 허용하는 코드를 포함한 라이선스)가 제공되어야 함

🟡 4. 데이터스페이스에서 서로 다른 이해관계자가 공유하는 데이터 세트를 함께 사용할 수 있도록

8) 최종 사용자가 소스 코드를 합법적으로 사용할 수 있도록 특별한 라이선스를 통해 공개되는 소프트웨어

9) 기존의 저작권과 다른 대안적 저작권을 발전시키기 위한 비영리 조직. 상업적 목적을 위해 재사용하는 것이 아니라면 합법적으로 활용할 수 있는 창작의 범위를 넓히고 다른 사람과 공유할 수 있게 하기 위해 노력하고 있음

보장하는 방법은 무엇인가?

- 상호운용이 가능한 데이터를 통해 정보를 쉽게 생성, 공유 및 재사용할 수 있으나 현실적으로는 데이터 세트를 결합하는 것이 매우 어렵기 때문에 상호 운용성이 낮다는 제약이 있음
 - 데이터 세트는 다양한 형식(개방형/비개방형)으로 제공될 수 있고 라이선스 정보가 누락될 수 있으며, 서로 관련이 없는 데이터 세트일 수 있기 때문에 여러 소스의 데이터를 결합하는 것은 매우 어려움
 - 따라서 데이터 상호 운용성을 달성하는 것은 매우 중요한 과제임
 - 일례로, 공공기관에서 보행 가능한 녹지를 늘리려는 새로운 도시 계획을 진행하기 위해 도로 네트워크, 시민 접근 지점, 도시 내 인기 구역 등 여러 데이터 세트를 찾더라도 이러한 데이터 세트가 상호운용 가능한 방식으로 제공되지 않는 경우 이용하기 어려움
- 이를 위해 데이터의 문서화, 모델링, 인코딩 및 공유와 관련한 국제 표준이 제시됨
 - 스마트시티 계획을 위한 최소 상호 운용성 메커니즘(MIM, Minimal Interoperability Mechanisms)이 대표적임
 - 유럽 상호 운용성 프레임워크(EIF, European Interoperability Framework)는 12가지 원칙과 개념 모델을 통해 법적, 조직적, 의미론적 및 기술적 4가지 차원을 중심으로 데이터 공유의 상호 운용성을 개선하기 위한 체계를 제시함
 - 또한, 유럽연합 집행위원회는 공공 행정을 혁신하고 디지털 전환에 도움을 주기 위한 상호운용 유럽(Interoperable Europe) 이니셔티브에 착수함

5. 기술 요구사항과 표준이 준수되고 있는지 확인하는 방법은 무엇인가?

- 데이터 세트의 결합과 효율적인 데이터 공유를 보장하는 것보다 데이터 인코딩 및 공유 과정에서 기술 요구사항 및 표준을 준수하는 것이 더욱 중요함
 - 몇몇 유럽 공공 데이터 제공자는 유럽연합 환경 지침과 관할 국가 법률에 따라 농산물 배송 과정에서 위치 데이터를 공유하도록 의무가 부과됨
 - 해당 지침은 국제 표준을 기반으로 하며, 국경을 넘어 다른 애플리케이션 도메인에서 데이터를 검색, 접근 및 사용할 수 있도록 함
 - 이러한 지침에 따라 데이터 제공자는 관련 데이터 세트를 게재하기 위한 맞춤형 인프라를 개발해야 하나 국가마다 모두 소프트웨어 툴이 다르고 구현 환경이 다르기 때문에 상호 운용성을 보장하기에 충분하지 않음
- 다양한 데이터 소스의 결합은 확립된 국제 표준에 기반한 데이터 분류 방식을 통해 가능함
 - 데이터 공유 표준의 기술 사양과 법적 요구사항을 준수하도록 보장하면 데이터 경제에서 이해관계자 간의 투명성과 신뢰성이 향상됨
 - 그러나 데이터 소스 결합은 객관적이고 정량화 가능한 정보를 기반으로 하는 자동화된 기술을 활용해서 검증될 때 효과가 있음
 - 데이터 소스 검증은 전문 검증 도구를 사용하여 오류를 수정하며, 반복적으로 테스트를 실행했을 때 가장 효과가 있음

6. 데이터스페이스에서 데이터를 쉽게 검색할 수 있는 방법은 무엇인가?

- 데이터 검색에 사용되는 수단은 검색 엔진, 전용 데이터 카탈로그 및 웹 애플리케이션 인터페이스 등 다양하기 때문에 사용자가 데이터를 찾고 활용하는 것이 복잡할 수 있음
 - 예를 들어, 사용자가 여러 국가에 분산된 여러 도시의 도시 계획 애플리케이션에 대한 이동성 데이터를 식별하는 과정에서 검색 엔진, 오픈 데이터 포털 등 여러 수단을 통해 기존 데이터 세트를 검색함
 - 검색 엔진 및 웹 애플리케이션 인터페이스에서 쉽게 찾을 수 있는 데이터 세트는 활용할 수 있으나, 재사용이 불가능한 데이터 세트거나 쉽게 검색이 되지 않는 데이터 세트의 경우는 활용할 수 없음
- 표준화된 메타데이터를 기계나 인간이 읽을 수 있는 형식으로 제공하고, 애플리케이션 인터페이스를 통해 데이터 및 메타데이터를 공유하면 데이터의 검색 가능성이 크게 향상됨
 - 또한, HTML과 같은 웹 네이티브(개발) 형식을 포함하여 여러 인코딩을 통해 데이터를 제공하면 검색 엔진에서 데이터를 색인할 수 있음
 - ISA2 프로그램의 SEMIC Action¹⁰⁾, 디지털 유럽(Digital Europe) 프로그램 및 schema.org와 같은 이니셔티브에서 개발된 것과 같이 일반적으로 합의된 언어를 사용해서 데이터에 주석을 달면 웹에서 데이터를 쉽게 찾을 수 있음

7. 데이터스페이스에 가장 적합한 표준을 선택하는 방법은 무엇인가?

- 표준은 공통 유럽 데이터스페이스에서 상호 운용성과 효율적인 데이터 공유를 위한 핵심 구성요소지만, 구체적인 데이터 및 서비스에 참조될 표준이 큰 숙고 없이 선택되고 있음
 - 표준을 잘못 선택하면 데이터 공유 프로세스에 지장이 생기며, 궁극적으로 데이터스페이스 전체를 방해할 수 있음
 - 표준의 수가 많으면 최적의 표준을 선택하는 것이 어려우며, 사용자 기반이 정립되지 않은 표준이나 소프트웨어 지원이 원활하게 이루어지지 않는 표준이 있을 수도 있음
 - 개발 단계에 있는 표준을 선택할 경우, 최종 표준이 변경되면 데이터스페이스에 영향을 미침
- 예를 들어, 교통 관련 프로젝트를 진행하는 기업의 경우 기계 학습 분석을 수행하여 도시 내 여러 지역에 걸쳐 위치한 녹지 접근성을 평가해야 함
 - 데이터를 찾는 과정에서 데이터 인코딩 및 공유 표준이 각기 다를 수 있으며, 다른 표준으로 변환할 수 없는 데이터도 존재함
- 따라서 공공 부문, 민간 부문 및 학계를 포함한 여러 다양한 이해관계자가 참여하는 거버넌스 프로세스를 통해 확립되고 신뢰받는 국제 표준 개발 조직(SDO, Standards Development Organizations)이 개발한 표준을 우선시해야 함
 - 위의 교통 관련 프로젝트 예시의 경우 1994년에 설립된 국제 표준화 기구인 개방형 공간 정보 컨소시엄(OGC, Open Geospatial Consortium)의 표준을 활용할 수 있음

10) 유럽 공공 행정부가 국경 간 및 도메인 간 데이터 교환을 원활하고 의미 있게 수행할 수 있도록 솔루션을 제공하는 이니셔티브

- 표준의 성숙도를 고려해 아직 개발 단계에 있거나 공개되지 않은 표준은 피하고, 이미 대규모 사용자 기반을 보유하고 있으며 기존 소프트웨어 툴을 지원하는 표준을 선택해야 함

8. 데이터스페이스에서 디지털 자원과 데이터가 참조 표시되도록 하는 방법은 무엇인가?

- 서로 다른 데이터 공급자로 인해 데이터의 인코딩 및 공유 과정에서 용어, 코드리스트¹¹⁾, 식별자 및 기타 디지털 자산 문제가 발생함
 - 서로 다른 자연어를 사용하는 데이터를 웹을 통해 공유하면 해당 문제가 더욱 복잡해짐
 - 예를 들어, 서로 다른 국가 출신의 데이터 공급자는 공통적으로 합의된 언어를 사용하지 않고 다른 데이터 모델과 자체 언어를 통해 데이터를 인코딩함
 - 이런 경우 여러 국가에 걸친 하천 유역의 홍수 위험 평가와 같이 데이터를 결합해야 하는 상황이 발생하면 데이터를 결합하기 어려움
- 공통적으로 합의된 데이터 모델, 항목 및 언어를 사용하면 데이터 상호 운용성이 증대됨
 - 공통적으로 합의된 언어가 코드리스트를 통해 지속적으로 웹에 노출될 경우, 다른 데이터 공급자는 자연어와 관계없이 코드리스트 항목의 고유한 리소스 식별자를 제공하여 데이터를 인코딩할 수 있음
 - 코드리스트 항목과 기술은 RDF, JSON(JSON- LD 포함), CSV, XML과 같은 다양한 인코딩을 통해 사용할 수 있음
 - 유명한 솔루션은 Re3gistry 오픈소스 소프트웨어로, 고유 식별자를 통해 참조할 수 있는 항목을 노출하는 데 유용함

9. 데이터스페이스에서 데이터에 대한 접근을 제공하는 방법은 무엇인가?

- 데이터스페이스는 데이터스페이스 내에 위치한 서로 다른 데이터 세트의 보완성을 활용하여 데이터 가치를 최대로 실현하는 것을 목표로 함
 - 데이터 세트에 효과적으로 접근하고 디지털 환경에서 이해관계자와 시스템을 원활하게 통합하는 것은 기술, 표준 및 아키텍처의 차이로 인해 어려움
 - 예를 들어, 애플리케이션 개발자는 대기 오염 물질 모니터링 애플리케이션을 구축하기 위해 대기 오염 센서, 교통 및 도로 인프라의 데이터를 활용함
 - 서로 다른 데이터 공급자가 제공하는 데이터는 모든 데이터 소스에 대한 접근이 보장되는 경우에만 계획된 애플리케이션을 구현할 수 있음
- 웹 기반 응용프로그램 인터페이스는 디지털 환경에서 서로 다른 이해관계자와 시스템을 연결하는 역할을 수행함
 - 응용프로그램 인터페이스는 현대 디지털 인프라를 연결하는 표준으로서, 접근할 수 있는 데이터와 접근 방법 및 조건에 대해 설명함
 - OpenAPI와 같은 광범위한 표준을 활용하면 디지털 가치 사슬에서 응용프로그램 인터페이스 구성요소를 쉽게 통합할 수 있음

11) 입력을 제한하고 오류를 방지하기 위해 문서에 사용되는 코드 목록

10. 데이터스페이스에서 개인정보를 보호하고 개인 데이터와 비즈니스 데이터를 보호하는 방법은 무엇인가?

- 많은 데이터 세트는 비밀 유지 조건, 유럽 법률에서 제시하는 개인정보 보호 조항, 윤리적 제약 등과 같은 다양한 이유로 인해 공유되지 않음
 - 이러한 데이터는 개인과 비즈니스 데이터를 포함하며, 기업의 귀중한 내부 자산일 수 있음
 - 따라서, 개인정보를 보호하면서 이러한 데이터 세트를 활용할 수 있다면 여러 사회적, 경제적 이점을 창출할 수 있음
- 예를 들어, 여러 병원에서는 환자의 질병 발생을 예측할 수 있는 기계학습 알고리즘 개발에 관심이 있음
 - 그러나 희귀 질환에 대한 데이터를 보유하고 있더라도 기계학습 알고리즘으로 활용할 정도로 충분한 데이터 세트를 가지고 있는 병원은 없으며, 필요한 기술적 용량도 갖추고 있지 않음
 - 기존 데이터를 결합하면 기계학습 알고리즘으로 활용할 수 있지만, 개인정보 보호를 고려하여 환자 데이터를 공개할 수 없는 상황임
- 민감한 개인 및 비즈니스 데이터를 보호하면서 개인정보보호 규정(GDPR, General Data Protection Regulation)에 따라 재사용할 수 있도록 하는 개인정보 보호 강화 기술(PET, Privacy Enhancing Technologies)이 존재함
 - 그러나 이는 왜곡된 출력, 입력 데이터 균질화 필요 등 결점이 있어 개인정보 보호를 위한 보편적인 해결책이 될 수 없음
 - 또한, 사용되는 아키텍처와 기술에 따라 입력물 보호, 출력물 보호, 관련 알고리즘 보호 등 개인정보 보호 대상이 다를 수 있으며, 데이터 보관 중, 전송 중, 사용 중 등 개인정보 보호 시점에도 차이가 있음
- 그러므로 여러 개인정보 보호 강화 기술을 고려하여 활용하는 것이 중요함
 - **안전한 다자간 계산(SMPC, Secure Multi-Party Computation)** : 서로 신뢰하지 않는 데이터 공급자가 출력물 외에 아무것도 공유하지 않고 공동으로 데이터를 산출해내는 기술
 - **차등 개인정보 보호(DP, Differential Privacy)** : 평균, 합계 및 분산과 같은 집계 함수를 통해 데이터를 쿼리할 때 개별 기록의 식별을 피하기 위해 입력물에 노이즈를 추가하는 기술
 - **동형 암호(Homomorphic Encryption)** : 원시 데이터 대신 암호화된 데이터를 기반으로 작업을 수행하는 일련의 암호화 기술
 - **연합 학습(Federated Learning)** : 모델이 저장된 위치에 데이터를 남기지 않고 각각 자신의 데이터만 포함하는 여러 노드¹²⁾를 통해 훈련하는 분산 컴퓨팅의 한 형태
 - **영지식 증명(Zero-Knowledge Proof)** : 증명자가 자신이 알고 있는 정보를 공개하지 않고도 확인자에게 그 정보를 알고 있다는 것을 증명할 수 있는 암호화 체계

12) 네트워크 또는 데이터 구조를 구성하는 각각의 개체

- 신뢰 실행 환경(TEE, Trusted Execution Environment) : 프로세서 내에 일반 영역과 보안 영역을 나누어 제공함으로써 두 영역 간의 정보 교환을 통제하여 보안성을 유지하는 방식
- 가명화(Pseudonymisation) : 개인정보를 직접적으로 공개하거나 실체 간의 관계를 설정할 수 있는 식별자를 변경, 삭제 또는 인코딩하는 절차를 사용하여 데이터를 보호하는 모델
- 각 기술은 일정한 비용이 수반됨
 - 일부 기술은 추가적인 연산 리소스 또는 하드웨어를 필요로 하며, 다른 기술은 데이터 기반 분석의 출력물을 왜곡하는 경우가 있음
 - 많은 기술이 아직 연구 중이므로 특정 기술을 선택할 때 기술 준비 수준(TRL, Technology Readiness Level)을 고려하는 것이 중요함

(조직적 측면)

1. 데이터스페이스에서 어떤 이해관계자가 어떤 유형의 데이터를 제공하는가?

- 공공 부문의 데이터 공급자와 학계, 민간 부문 및 시민에 이르기까지 점점 더 다양한 이해관계자에 의해 매일매일 엄청난 양의 데이터가 생성되고 있음
 - 방대한 양의 데이터가 통합된다면 의미 있는 데이터 세트를 도출할 수 있음
 - 사용 가능한 데이터의 범위를 이해하기 위해서는 데이터스페이스와 관련이 있는 모든 데이터 세트와 데이터를 생성·배포하는 이해관계자를 식별해내야 하며, 여기에는 많은 시간이 소요됨
- 예를 들어, 지난 10년간의 국가 내 토지 이용 영향을 평가하는 기후 변화 연구에서 유럽연합 회원국 기관이 활용할 수 있는 자료는 정부 토지 사용 디지털 지도가 유일함
 - 정부 토지 사용 디지털 지도의 경우 6년마다 제작되거나 갱신되며, 해상도가 낮음
 - 그러나 정부 토지 사용 디지털 지도 이외에도 각 국립 대학의 회원국 토지 이용 연구 지도, 민간 기업의 토지 이용 조사, 유럽연합 기관 위성 이미지에서 파생된 토지 이용 지도, 시민 참여 과학 프로젝트의 토지 이용 표본(지리적 위치 사진 및 설명)을 이용할 수 있으나 활용되지 않음
 - 이는 데이터 탐색 및 수집 단계가 짧은 시간 내에 수행되기 때문으로, 다양한 데이터 세트가 사용되었다면 기후 변화 영향을 평가하는 기계학습 모델을 훈련하고 적용할 뿐만 아니라 정부 토지 사용 디지털 지도를 보완·검증하는 과정에서 활용할 수 있음
- 그러므로 공통 유럽 데이터스페이스를 개념화하고 기술 인프라 설계를 시작하기 전에 데이터스페이스 내에서 데이터를 생산하거나 배포하는 이해관계자를 탐색하고 식별하는 데 충분한 시간을 투자해야 함
- 유럽연합에서는 개방형 데이터 지침(지침 2019/1024)과 고부가가치 데이터세트 시행법을 통해 데이터 세트의 가용성을 증대시키고자 함
- 데이터 거버넌스법과 데이터법은 시민이 생성한 데이터뿐만 아니라 비공개 데이터의 접근 및 사용을 촉진하는 것을 목표로 함

- 과학계에서도 검색·접근·상호운용·재사용 가능(FAIR) 원칙을 따르는 추세임

2. 데이터스페이스에서 사람 중심의 데이터 접근 방식을 육성하는 방법은 무엇인가?

- 정책, 관계, 사회, 기술적 의사 결정 구조로 데이터 수집 방법을 정의하는 데이터 거버넌스를 통해 사람 중심의 접근 방식을 보장할 수 있음
 - 데이터 거버넌스는 데이터스페이스 내에서 데이터를 공유하고 교환할 수 있는 신뢰도 높은 환경을 조성하는 데 중요한 요소임
 - 데이터 수집자와 데이터 공급자를 포함한 이해관계자 간에는 비대칭성이 존재하기 때문에 데이터 거버넌스를 통해 데이터 공급자 및 데이터 기반 결정으로 인해 영향을 받을 수 있는 이해관계자의 권리와 이익을 보호해야 함
- 예를 들어, 시의회가 대중교통, 폐기물 수집 분석 및 공공 주택의 에너지 소비 모니터링 등 여러 공공 서비스를 디지털화하는 과정에서 주민들에게 알리지 않고 주민들의 데이터를 수집함
 - 시의회가 독단적으로 데이터 관리 서비스를 기술 기업에게 맡기는 경우, 주민들은 이러한 과정에 전혀 참여하지 않게 되어 사람 중심의 데이터 접근 방식이 취해지지 않음
- 데이터 수집, 공유 및 사용 과정에서 사람 중심의 데이터 접근 방식을 육성하는 방법은 다음과 같음
 - 대중의 목소리와 관심사가 데이터 거버넌스 내에서 잘 표현되도록 보장함
 - 일반 대중 사이에서 디지털 사용 능력과 기술 이용 능력을 높이고 디지털 권리에 대한 인식을 제고함
 - 대중이 자신의 데이터 접근, 공유 및 사용 방식을 통제할 수 있도록 함
 - 공공 이익을 위한 데이터 수집에 참여하고, 데이터 제공을 통해 지역 내 의사 결정에 참여할 수 있는 기회를 대중에게 제공함
- 데이터스페이스 내에서 신뢰할 수 있는 환경과 사람 중심의 데이터 접근 방식을 구축하면 데이터 공유 및 교환을 증대시키는 데 도움이 될 수 있음

3. 데이터스페이스에서 데이터를 공유함으로써 비즈니스에 미칠 수 있는 장점은 무엇인가?

- 신뢰할 수 있는 데이터 공유를 촉진하는 기업은 주요 기업 가치 지표에서 다른 기업보다 더 좋은 성적을 보일 것임
 - 그러나, 많은 기업에서는 여전히 데이터 공유의 여러 이점에 대해서 알지 못함
 - 공통 유럽 데이터스페이스는 기업이 데이터를 안전하게 공유할 수 있는 개인정보 보호 인프라를 제공하고자 하며, 기업에서는 이를 통해 간접적인 이익을 누릴 수 있음
 - 기업은 데이터스페이스에서 상호 보완적인 데이터 세트를 통합함으로써 활용 가능한 데이터의 크기, 품질 및 범위를 늘릴 수 있음
- 예를 들어, 전기차 충전소는 전기차 제조업체가 활용할 수 있는 데이터를 수집함

- 충전소는 유럽연합 이동성 데이터스페이스에 데이터를 마케팅할 수 있으며, 전기차 제조업체가 이를 구매하여 비즈니스 파트너십을 체결할 수 있음
- 이러한 예시는 공통 유럽 데이터스페이스를 통해 데이터를 공유하면 데이터 기반 혁신을 촉진하고 수익성 있는 파급 효과를 창출할 수 있음을 보여줌
- 기업은 공유 데이터에 접근함으로써 공급망을 최적화하고 여러 지표를 비교할 수 있으며, 많은 마케팅 기회를 누릴 수 있음

4. 정부가 공익을 위해 어떻게 민간 부문 데이터에 접근할 수 있는가?

- 정부는 기업이 수집한 여러 데이터를 활용하여 보건, 사회 및 환경 문제를 해결할 수 있음
 - 이러한 관계를 B2G(Business-to-Government) 데이터 공유 관계라고 함
 - 예를 들어, 공공기관은 코로나19 바이러스의 확산을 방지하기 위해 통신사가 수집한 위치 정보에 접근하고자 함
- 그러나 데이터 수집 및 접근 과정에서 정부, 기술 기업 및 기타 민간 기업 간에 상당한 정보 비대칭성이 존재함
 - 지방 정부를 포함한 공공기관은 정보 측면에서 취약하므로 민간 부문 데이터에 접근하기 어려움
 - 또 다른 문제는 기업에서 잠재적인 개인정보 보호 위험과 데이터 민감성 때문에 공공기관과 데이터 공유를 꺼린다는 점임
- 예를 들어, 한 차량 공유 서비스가 지방 도시에서 공유 서비스를 운영할 때 지방 정부는 정보의 비대칭성으로 인해 서비스 운영 사항을 모두 이해하고 점검하기 어려움
 - 지방 정부에서 개별 기업과 민관 데이터 공유 계약을 체결하려고 해도 기업에서는 서비스 제공 과정에서 수집된 데이터를 공유하지 않는 경우가 많음
 - 또한, 이러한 계약에는 과도한 비용이 소요되고 데이터가 분절되어 최적의 해결책이 아님
- 정부가 기업이 수집한 민간 부문 데이터에 접근할 수 있는 솔루션은 아래와 같음
 - 공공기관은 공공 서비스를 제공할 때 기업이 수집한 데이터가 개인정보 보호를 준수한다는 조건 하에 접근을 허용하도록 요구하는 **데이터 주권 조항**을 계약에 포함시킬 수 있음
 - 공공기관은 상호 이익을 위해 공동으로 데이터를 분석할 의향이 있는 기업과 **파트너십**을 체결할 수 있음
 - 정부는 데이터 공유와 관련한 기업의 **사회적 책임 이니셔티브(Corporate Social Responsibility Initiative)**를 활용해 기업이 공공의 이익을 위해 공공기관에 데이터를 제공하도록 할 수 있음¹³⁾

5. 데이터스페이스에서 인공지능 시스템의 데이터 투명성을 어떻게 높일 수 있는가?

13) 유럽연합 데이터법은 긴급한 상황에서 기업이 정부에게 데이터를 무료로 제공하도록 요구하고, 다른 예외적인 상황에서 데이터 공유에 대한 보상을 청구할 수 있도록 하는 사회적 책임 이니셔티브 조항을 제시함

- 데이터스페이스의 데이터는 인공지능 시스템 개발에 사용될 수 있는데, 이때 올바른 데이터 세트를 선택하는 것이 중요함
 - 데이터 세트의 품질이 인공지능 시스템의 성능과 결과에 큰 영향을 미치기 때문에 실질적으로 활용되기 전에 품질 기준이 충족되는지 확인해야 함
 - 데이터가 인공지능 시스템에 잘못 반영될 경우 사회적 편견을 증폭시키고 인권을 침해할 수 있음
 - 정책 입안자와 대중 모두 데이터를 공유하고 관리하는 조직의 투명성과 책임성 증대를 요구하고 있으며, 유럽연합 인공지능 법(The EU AI Act)¹⁴에서 데이터 품질, 투명성 및 거버넌스 의무를 규정하고 있음
- 예를 들어, 공공기관 내 복지 부서에서 인공지능 기반 사기 탐지 시스템을 개발할 때 편향된 결과를 도출할 수 있음
 - 알고리즘 훈련에 사용되는 데이터 세트에는 저소득자에 대한 편향이 포함되어있을 수 있어 사기 탐지 시스템이 저소득자에 대한 편향적 결과를 낼 수 있음
- 데이터 세트용 데이터 표, 데이터 세트 뉴트리션 라벨(Dataset Nutrition Label)¹⁵, 모델 카드¹⁶, 인공지능 팩트 시트(AI Factsheet)¹⁷ 등 여러 도구 및 프레임워크를 활용해 데이터 세트, 인공지능 모델 및 알고리즘을 문서화 하면 편향성을 줄이고 데이터를 처리하는 공공기관의 신뢰도를 높일 수 있음
 - 인공지능 모델 개발자가 데이터 세트의 출처, 범위, 정확성, 대표성 등에 대해 알고 있으면 데이터 세트의 편향성을 줄이기 위한 해결책을 마련할 수 있음
 - 이처럼 데이터 세트 및 인공지능 모델을 문서화하면 인공지능 모델 개발자에게 도움이 될 뿐만 아니라 인공지능 시스템 사용자와 대중의 신뢰도를 높일 수 있음

6. 데이터스페이스에서 자발적인 데이터 공유를 활용하는 방법은 무엇인가?

- 오늘날의 사회, 경제 및 환경 문제를 해결하기 위해서는 데이터 공유를 증대시켜야 하나 이를 위한 안전한 장소가 부족하여 많은 이해관계자가 데이터에 접근하지 못하고 있음
 - 데이터 공유를 원하지 않는 주요 원인은 보통 개인정보 보호 문제와 경쟁 우위 유지임
- 예를 들어, 환자와 병원이 자발적으로 건강 데이터를 공유한다면 희귀질환에 대한 연구를 도울 수 있음
 - 공인된 데이터 이타주의 조직(RDAOs, Data Altruism Organisations Recognised in the EU)¹⁸이 공익을 위해 이러한 데이터를 수집, 관리, 처리 및 재배포한다면 희귀질환 연구가 능률화될 수 있음
- 이러한 배경에서 공통 유럽 데이터스페이스는 공공 이익을 위해 데이터를 공유할 수 있는 보호된

14) 인공지능 시스템의 위험도를 평가하고 데이터 거버넌스, 투명성 및 감독과 관련한 의무를 준수하도록 하는 법안

15) 데이터 세트 품질 평가를 위한 표준 품질 프레임워크

16) 기계 학습 모델이 수행하는 작업, 대상 고객, 유지 관리 방식을 설명하는 요소

17) 데이터 및 인공지능 수명 주기를 관리하고 자동화하기 위한 도구

18) 의료, 기후변화 해결 등 사회적 이익을 위해 데이터 주체가 데이터를 자발적으로 공유하도록 하는 조직을 의미함

환경을 제공하고자 함

- 다양한 이해관계자들 간에 데이터를 자발적으로 공유하도록 하기 위해 데이터 거버넌스법은 데이터 이타주의 조직의 새로운 범주를 도입함
- 데이터 이타주의 조직은 데이터 소유자와 데이터 주체가 데이터 유형, 데이터 공유 목적, 사용 연한 및 데이터 사용자의 범위를 선택할 수 있도록 함
- 데이터 이타주의 확산을 통해 사회 전반에 걸쳐 데이터의 가치를 재분배하고 공공 이익을 증대시킬 수 있음

7. 데이터스페이스에서 새로운 데이터 중심 솔루션을 생성, 제공 또는 사용할 때 고려해야 할 법적 측면은 무엇인가?

- 디지털 서비스를 제공하는 모든 주체는 정보 사회 서비스와 관련된 법적 의무를 준수해야 함
 - 유럽연합 지침에서는 새로운 데이터 기반 솔루션 제공업체가 전체 데이터 가치 사슬에 걸쳐 데이터, 소프트웨어 제품 및 서비스 사용에 대한 법적 측면을 모두 준수할 것을 요구함
 - 데이터와 데이터베이스, 데이터에 접근할 수 있도록 하는 소프트웨어는 모두 저작권 및 지적 재산 보호권에 의한 보호 대상이 될 수 있음
 - 중간 데이터 공급기업에서는 서비스 약관(TOS, Terms of Service) 또는 서비스 수준 계약(SLAs, Service Level Agreements)¹⁹⁾의 적용을 받는 서비스를 통해 데이터를 제공함
 - 이때 서비스 약관 및 서비스 수준 계약에는 기본 데이터 라이선스, 사용 약관, 인터페이스 사용에 대한 조건이 포함되어야 함
 - 또한, 중간 데이터 공급기업은 현재 유럽 데이터 공유 프레임워크 및 유럽연합 계약법, 경쟁법 및 기타 규제를 준수해야 함
- 예를 들어, 지속 가능한 스마트시티 애플리케이션이 운송 수단에서 발생하는 이산화탄소 배출량을 조절하고자 하는 경우 여러 데이터 소스가 필요함
 - 교통, 대중교통 네트워크 및 승차 공유 서비스 데이터, 이산화탄소 센서 및 이산화탄소 배출 데이터, 이동 패턴 및 개인의 위치 정보 등 여러 데이터 소스의 통합이 필요함
 - 이러한 모든 데이터에는 특정 라이선스와 사용 약관이 적용되어 있으며, 데이터 통합 과정에서 이러한 약관을 위반해서는 안 됨
 - 애플리케이션에서 필요한 데이터 세트가 특정 사용을 제한하는 경우 해당 사항을 통합 과정에 반영하고 사용하지 않도록 해야 함
- 각 데이터스페이스는 데이터스페이스 내 여러 이해관계자가 새로운 데이터 중심 솔루션을 사용할 때, 법률을 준수하면서 운용할 수 있도록 지원해야 함
 - 데이터스페이스는 법적 의무 관련 법률에 대한 설명을 제시하여 각 이해관계자가 더 원활히 소통하고 디지털 생태계 내에서 법적 의무를 수행할 수 있도록 해야 함

19) 서비스 수준 계약(SLA)은 서비스 제공자와 클라이언트 간의 일종의 계약 또는 계약의 일부임

PART V

결론

- ④ 원문 보고서("EUROPEAN DATA SPACES, JRC SCIENCE FOR POLICY REPORT")는 공통 유럽 데이터스페이스 구축과 관련한 법적, 사회적, 경제적 인사이트를 제공함

 - 공통 유럽 데이터스페이스는 다양한 이해관계자가 공통 거버넌스, 조직, 규제 및 기술 메커니즘에 따라 안전하고 신뢰할 수 있는 방식으로 데이터 공유, 접근, 사용이 가능한 분산형 인프라임

 - 유럽연합은 데이터스페이스 구축을 지원하기 위해 데이터 거버넌스법, 데이터법, 고부가가치 데이터세트 시행법, 디지털 시장법 등 여러 부문에서 법안을 도입하고 있음
 - 유럽 데이터 전략 실행을 위한 입법 프레임워크를 비교함으로써 인사이트를 도출하였으며, 유럽연합 정책 문서를 기반으로 공통 유럽 데이터스페이스에 대한 요구사항을 분석함

 - 해당 요구사항을 기능적, 비기능적 범주로 구분하였으며, 데이터스페이스 가이드를 통해 심층적으로 요구사항을 어떻게 구체화할 수 있을지 설명함
 - 공통 유럽 데이터스페이스 구축 과정에서 고려해야 하는 중심 키워드도 다름

 - 여기에는 합성 데이터, 소프트웨어 스택, 인허가, 상호 운용성, 데이터 인증, 데이터 검색 가능성, 데이터 기준, 데이터 기록부, 데이터 접근을 위한 응용프로그램 인터페이스, 개인정보 보호 기술, 데이터 행위자, 개인 데이터, B2G 데이터 접근, 자발적인 데이터 공유 등이 포함됨
 - 데이터스페이스 구축에 참여하는 다양한 이해관계자를 위해 데이터 공유를 촉진하기 위한 여러 해결책을 제시함

 - 데이터 이타주의 조직은 데이터 소유자와 데이터 주체가 데이터 유형, 데이터 공유 목적, 사용 연한 및 데이터 사용자의 범위를 선택할 수 있도록 하는 조직으로서, 사회 전반에 걸쳐 데이터의 가치를 재분배하고 공공 이익을 증대시킬 수 있음
 - 공통 유럽 데이터스페이스는 단일 데이터 시장이기 때문에 모든 이해관계자가 참여하는 포괄적인 프로세스를 통해서만 실현될 수 있음

 - 데이터스페이스의 핵심은 탈중앙화로, 데이터스페이스를 단독으로 운영하거나 다른 모든 이해관계자가 따르도록 하향식 결정을 내릴 수 있는 단일 조직은 없음
 - 동시에 여러 이해관계자가 거버넌스 규칙을 세워 자체 데이터스페이스를 마련하는 것은 피해야 함

참 고 문 헌

- European Commission(2023), “EUROPEAN DATA SPACES, JRC SCIENCE FOR POLICY REPORT”, Publication Office of the European Union,
<https://op.europa.eu/en/publication-detail/-/publication/dcac6aee-0e7a-11ee-b12e-01aa75ed71a1/language-en/format-PDF/source-288689101>
- Joint Research Center(2023), “European Data Spaces”, Farrel, E, et al.,
- European Commission “Regulation of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)”, Official Journal of the European Union, L 152, (2020), pp.1-44,
<http://data.europa.eu/eli/reg/2022/868/oj>https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj



데이터산업 동향 이슈 브리프

ISSUE BRIEF

| 발행일 2023년 7월 31일

| 발행처 **K data** 한국데이터산업진흥원

서울시 중구 세종대로 9길42, 부영빌딩 7, 8, 11층

| 기획 및 편집 산업기반본부 산업기획팀 하진희 팀장, 안선빈 주임

| 문의처 Tel: 02-3708-5364

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