

데이터의 윤리적 활용을 위한 메타데이터 및 거버넌스

임성준 한국산업지능화협회 전략사업팀 팀장



1. 머리말

인공지능, 클라우드 컴퓨팅, 빅데이터 등 데이터 기반 기술이 사회 전반으로 활용되어 데이터 기술은 산업 분야와 더불어 일상생활의 모든 측면에 영향을 주고 있다. 이에 따라 윤리적인 도전이 새롭게 생겨나기도 했는데, 본고에서는 데이터에 기반한 윤리 문제와 데이터 활용의 윤리성을 고찰하고자 한다. 또한 비윤리적 데이터 활용으로 야기되는 문제를 미연에 방지하는 기준을 제시하는 표준에 대해 살펴보고자 한다.

2. 데이터의 윤리적 문제

데이터 윤리와 관련된 대표적 문제는 개인정보 보호이다. 개인정보 유출로 인한 사생활과 인권침해, 기업 신용도 및 가치하락, 국가경제적 손해 등 그 피해 사례는 다양하고 폭넓다. 데이터 조작도 중요한 문제이다. 연구·실험 데이터를 조작해 유명 국제학술지에 논문을 실거나, 임상시험 결과를 조작하기도 한다. 그뿐만이 아니다. 산업 현장에서도 데이터를 조작해 사회경제적인 피해를 일으키기도 한다. 자동차 배기가스 데이터 조작사건이 그 예다. 이 밖에 데이터를 수집할 때 발생하는 편향성과 주관성 문제, 데이터를 소수의 기업이 일방적으로 독점하는 문제, 선진국과 개발도상국 사이의 데이터 양극화 문제처럼 데이터 윤리 문제에는 다양한 사례가 있다.

이러한 문제 사례를 종합적으로 유추해보면, 데이터를 윤리적으로 활용한다는 말은 데이터를 신뢰성 높은 방법으로 저장하고, 투명하게 관리하며, 정당하고 합리적으로 활용하는 것을 의미한다. 본고에서는 데이터의 저장, 관리, 활용 각 단계별로 메타데이터, 데이터 거버넌스, 데이터 사용

분야의 표준을 통해 데이터를 윤리적으로 활용하는 데 필요한 기준에 대해 알아보하고자 한다.

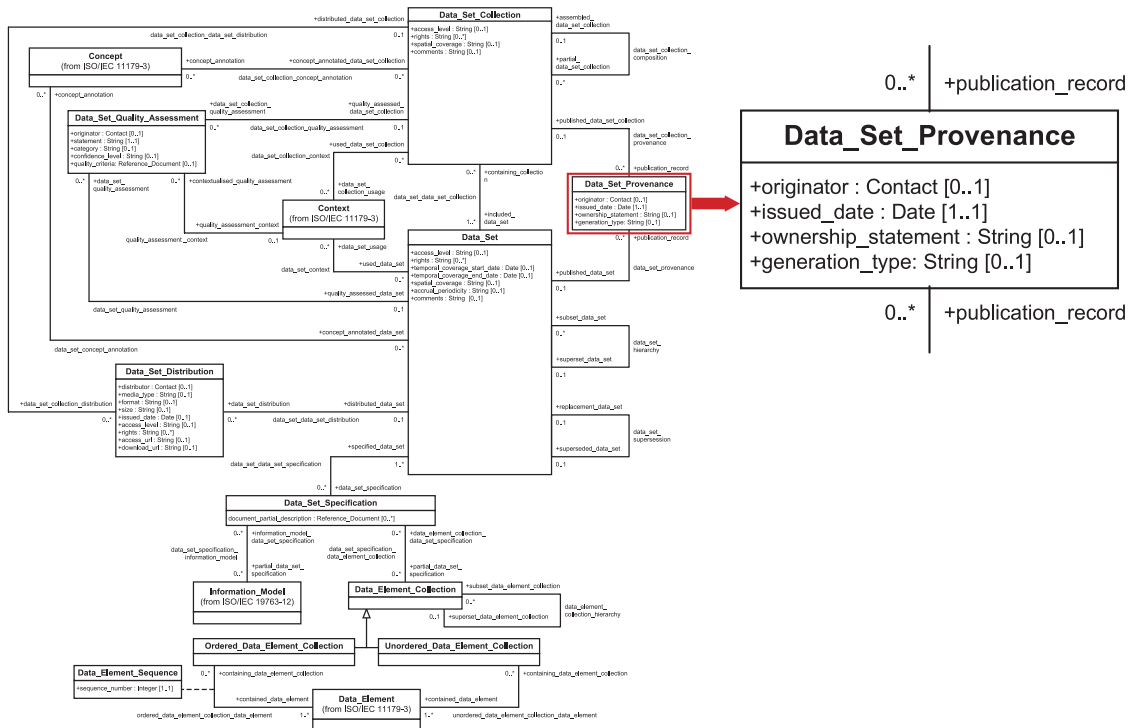
3. 데이터의 윤리적 활용을 위한 표준

3.1 데이터 저장 관점 - 메타데이터 표준

ISO/IEC 11179 시리즈는 메타데이터 레지스트리(Metadata registry) 시리즈다. 메타데이터를 관리하기 위한 프레임워크, 분류체계, 레지스트리 메타모델 및 기본 속성, 데이터 정의 공식 표시방법, 명명규칙, 등록, 데이터 세트를 등록하는 메타모델을 정의한다. 현재 출판된 시리즈 목록은 다음과 같다.

- ISO/IEC 11179-1:2015 Metadata registries(MDR) — Part 1: Framework
- ISO/IEC TR 11179-2:2019 Metadata registries(MDR) — Part 2: Classification
- ISO/IEC 11179-3:2013 Metadata registries(MDR) — Part 3: Registry metamodel and basic attributes
- ISO/IEC TS 11179-30:2019 Metadata registries(MDR) — Part 30: Basic attributes of metadata
- ISO/IEC 11179-4:2004 Metadata registries(MDR) — Part 4: Formulation of data definitions
- ISO/IEC 11179-5:2015 Metadata registries(MDR) — Part 5: Naming principles
- ISO/IEC 11179-6:2015 Metadata registries(MDR) — Part 6: Registration
- ISO/IEC 11179-7:2019 Metadata registries(MDR) — Part 7: Metamodel for data set registration

데이터는 일반적으로 데이터 세트의 형태로 저장된다. 데이터 세트란 하나 이상의 형식으로 액세스하거나 다운로드할 수 있는 식별 가능한 데이터 모음이다. 이때 데이터 세트를 설명하는 메타데이터가 등록된 메타데이터 레지스트리가 있으면 적절한 데이터 세트를 쉽게 찾을 수 있다. 그뿐만 아니라 데이터 세트를 메타데이터를 통해 투명하게 관리함으로써 데이터를 신뢰성 있게 활용할 수 있다. ISO/IEC 11179-7 데이터 세



※출처: ISO/IEC 11179-7 Figure 1

[그림 1] 데이터 세트 메타모델 중 Data_Set_Provenance 클래스

트 등록을 위한 메타모델은 이러한 데이터 세트의 메타데이터를 등록할 때 갖춰야 할 메타모델을 제시한다.

특히 데이터 세트의 출처(Provenance) 정보를 제공할 것을 명시한다. 이는 출처 정보로 데이터의 계보를 파악할 수 있기 때문이다. 또한 생성기록, 데이터의 진위, 과거 소유권과 관련된 투명성도 확보할 수 있어 데이터를 윤리적으로 활용하는 기본 정보로 쓰인다.

ISO/IEC 11179-7에서 정의하는 Data_Set_Provenance 클래스는 originator, issued_date, ownership_statement, generation_type의 속성을 띤다.

3.2 데이터 관리 관점 - 데이터 거버넌스 표준

본 장에서는 데이터 거버넌스의 윤리적 활용

을 살펴보고자 한다. 데이터 거버넌스에 대한 표준은 TTA표준으로 5개 파트로 제정된다.

- TTA.KO-10.0555-Part1 데이터 거버넌스 - 제1부 : 데이터 관리를 위한 기능
- TTA.KO-10.0555-Part2 데이터 거버넌스 - 제2부 : 거버넌스 역할
- TTA.KO-10.0555-Part3 데이터 거버넌스 - 제3부 : 거버넌스 프로세스
- TTA.KO-10.0555-Part4 데이터 거버넌스 - 제4부 : 데이터 베이스 보안
- TTA.KO-10.0555-Part5 데이터 거버넌스 - 제5부 : 프레임 워크

위 표준에 따르면 데이터 거버넌스의 궁극 목적은 고품질 데이터를 확보하고 유지해 기업에 제공하는 정보의 활용을 극대화하는 것이다. 또한 데이터를 이용해 얻은 정보를 기초로 기업의 가치 창출에 기여하는 것이다(TTA.KO-10.0555-Part1).

데이터 거버넌스는 프라이버시, 보안성, 관리 규정을 비롯한 데이터 관리에 신뢰성을 더하는 기본 항목을 다룬다는 점에서 데이터를 윤리적으로 활용하기 위한 필수 요소이다. 파트별로 해당 요소를 정리하면 <표 1>과 같다.

데이터 관리 관점에서 데이터 거버넌스 표준을 적용할 때 이러한 요소에 중점을 두고 표준을 적용한다면 데이터 관리를 더 신뢰성 있게 수행할 수 있다.

3.3 데이터 활용 관점 - 데이터 사용 표준

데이터 사용에 대한 표준은 ISO/IEC JTC 1/SC 32 데이터 관리서비스 위원회 WG6 데이터 사용 워킹그룹에서 개발 중이다. 현재 개발 중인 표준 목록은 다음과 같다.

- ISO/IEC AWI 5207 Data usage — Terminology and use cases
- ISO/IEC WD 5212 Data usage — Guidance for data usage

ISO/IEC 5212 Guidance for data usage에서는 주로 민감정보나 개인정보를 활용할 때 지켜야할 프로세스를 정의한다. 이와 더불어 호주 정부에서 개발 중인 PIF도구(Personal Information Factor, 개인정보요소)를 표준개발 기고서를 통해 소개한다.

PIF도구는 민감성과 개인정보요소를 조합해 개인 식별정보를 활용하는 프로젝트가 얼마나

안전한지 측정하는 도구이다. 민감성 평가에는 다음 사항이 포함된다.

데이터로 포착된 민감한 주제

- 분석에 사용되는 데이터의 품질 문제(정확성, 해상도, 범위, 편향)
- 산출물에 대한 신뢰도(정확성, 정밀도, 일관성, 설명 가능성, 편향)
- 산출물(통찰력 또는 데이터 기반 결정)이 사용되는 방식
- 통찰력을 얻기 전에 사람의 판단이 적용될지에 대한 문제
- 산출물의 사용 또는 재사용으로 인해 발생할 수 있는 피해
- 부정적인 당혹감으로 이어지는 분석 결과
- 분석 결과를 적절하게 해석하는 데 필요한 전문 지식 또는 정보
- 데이터 또는 통찰력이 우발적으로 공개되었을 때에 대한 문제
- 데이터 시점에 대한 문제(또는 조사한 적이 없는 데이터가 사용된 경우 등)

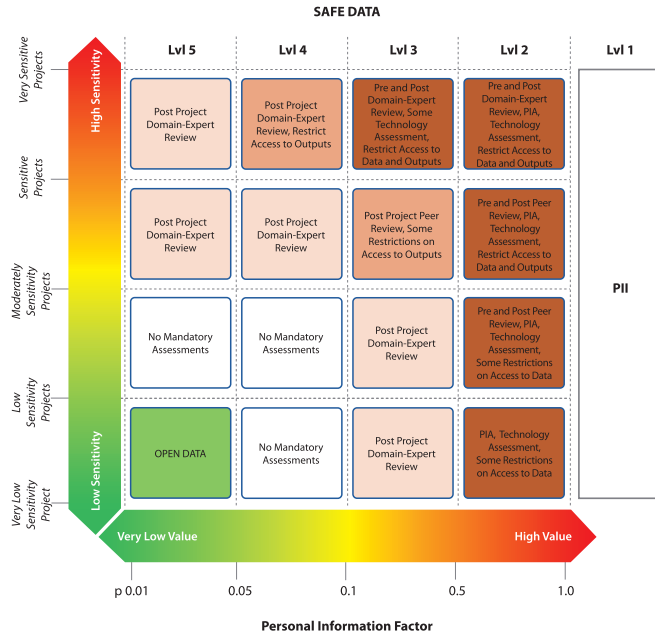
또한 WG6에서는 데이터 사용 시 나타날 수 있는 잠재 위험을 다음과 같이 언급한다.

- 개인정보 데이터 공개
- 데이터에서 인사이트에 대한 독점적 접근의 상실
- 데이터의 적절한 사용에 대한 우려
- 데이터에서 얻은 통찰력의 적절한 사용에 대한 우려
- 데이터 분석에서 나타나는 예상치 못한 결과
- 사용자가 데이터를 분석하거나 해석할 전문 지식 부족
- 데이터 시점 및 품질
- 수집 또는 데이터 품질에 대한 정보 없이 데이터 사용

WG6에서는 아직 데이터 사용에 대한 표준을 제정·공표하지는 않았다. 그러나 데이터 사용에 대한 여러 위험과 문제사항을 종합적으로 제시하고, 국가 간 토론의 장을 제공하면서 데이터를 윤리적으로 활용하기 위한 지침으로 제공할 수

<표 1> 데이터 거버넌스의 주요 요소

파트	제목	데이터 윤리에 관한 주요 요소
1	데이터 관리를 위한 기능	데이터 형상 관리, 데이터 이력관리, 데이터 검증, 데이터 승인, 데이터 접근 권한 관리 등
2	거버넌스 역할	데이터 거버넌스 자문위원회, 데이터 스튜어드 등
3	거버넌스 프로세스	데이터 정책, 데이터 규제 준수, 데이터 이슈 관리, 데이터 관리 계획 및 통제 등
4	데이터베이스 보안	데이터 보안관리, 보안체계 등
5	프레임워크	규제준수, 이슈관리, 데이터 표준 및 관리절차 등



※출처: ISO/IEC JTC 1/SC 32/WG 6 N 42

[그림 2] PIF와 민감성의 거버넌스 프레임워크

있는 표준을 개발하고자 많은 노력을 기울인다.

4. 맺음말

지금까지 데이터를 윤리적으로 활용하기 위해 데이터 저장·관리·활용 단계에서 적용할 수 있는 기준을 제시한 대표적 표준에 대해 살펴보았다. 데이터 윤리의 문제는 복잡하고 다양하며, 문제를 해결하기 위한 방안 역시 단순하지 않다. 기

계적이거나 자동화된 방식으로 접근해서는 데이터 윤리로 발생하는 문제를 해결하기 쉽지 않다. 다만 데이터를 다루는 조직에서 데이터를 투명하게 관리하고 합리적이고 정당하게 활용한다면 데이터의 윤리성을 어느 정도 확보할 수 있을 것이다. 본고에서 소개한 표준들이 데이터를 윤리적으로 활용하고자 하는 조직의 활동에 기본 지침으로 사용되기를 기대해본다. TTA

※ 본 연구는 2021년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(‘20011766’)

참고문헌

- [1] ISO/IEC 11179-7:2019 Metadata registries(MDR) — Part 7: Metamodel for data set registration
- [2] TTAK.KO-10.0555-Part1 데이터 거버넌스 - 제1부 : 데이터 관리를 위한 기능
- [3] ISO/IEC JTC 1/SC 32/WG 6 N 42 Content for working draft of ISO/IEC 5212 for review
- [4] 송선영, 김항인(2016). 정보화시대의 빅 데이터(Big Data) 활용에 대한 윤리적 논쟁과 전망. 윤리연구, 1(108), 227- 248.