

AI Ready Data의 개념적 구조와 한국의 AI Ready 공공데이터 정책 분석

이 정 윤* · 김 학 래**

요약

AI 기술의 확산에 따라 데이터 관리 패러다임이 단순한 '양적 확대'에서 'AI 활용 적합성 확보'로 전환되고 있다. 본 연구는 AI Ready Data(AIRD) 개념을 학술적으로 정립하고, 한국의 'AI Ready 공공데이터' 정책 현황 분석을 통해 향후 개선 방향을 제시하는 데 목적이 있다. 선행 연구와 정책 문서를 체계적으로 검토하여 산업계, 학계, 국제기구, 정부의 접근 방식을 분석한다.

본 연구는 AIRD의 5개 핵심 구성요소로 품질·신뢰성, 표준화·상호운용성, 거버넌스·책임성, 다양성·포용성, 지속적 운영 가능성을 도출하고, 이를 바탕으로 통합 프레임워크를 설계한다. 한편, 공공데이터 관리체계를 원천데이터(TL0), 기관 내부 가공데이터(TL1), 개방데이터(TL2)로 구분하는 전환계층 모델을 제시한다. 분석 결과, 한국 정책은 긍정적 방향성을 보이나, 국제 기준 대비 표준화·거버넌스·포용성 측면에서 초기 단계에 머물러 있는 것으로 나타난다. 본 연구는 개방 데이터 중심의 품질 개선 체계 구축과 생애주기 전반의 순환적 관리 체계 확립을 주요 정책 대안으로 제안한다.

주제어 : AI Ready Data, 공공데이터 정책, 데이터 품질관리, 상호운용성, 데이터 거버넌스

Conceptualizing AI Ready Data and Analyzing South Korea's AI Ready Public Data Policy Framework

Lee, JeongYun* · Kim, Haklae**

Abstract

With the proliferation of AI technology, the data management paradigm shifts from quantitative expansion to ensuring suitability for AI utilization. This study establishes a scholarly definition of AI Ready Data (AIRD) and proposes future directions by analyzing the current state of South Korea's 'AI Ready Public Data' policies. It conducts a systematic review of literature and policy documents to examine the approaches of industry, academia, international organizations, and governments.

The analysis identifies five core components of AIRD: Quality & Reliability, Standardization & Interoperability, Governance & Accountability, Diversity & Inclusivity, and Sustainability & Maintainability. Based on these, an integrated framework is designed. Furthermore, the study presents a transition layer model that categorizes public data management into Source Data (TL0), Internally Processed Data (TL1), and Open Data (TL2). The findings indicate that while South Korean policies show a positive direction, they remain in the nascent stages regarding standardization, governance, and inclusivity compared to international standards. This study proposes the establishment of an open-data-centric quality improvement system and a circular management framework throughout the data lifecycle as key policy alternatives.

Keywords : AI Ready data, public data policy, data quality management, interoperability, data governance

접수: 2025. 10. 2; 최종수정: 2025. 12. 4; 게재확정: 2025. 12. 4.

* 주저자, 중앙대학교 문헌정보학과 박사과정(cathx618@gmail.com, <https://orcid.org/0009-0000-5105-9764>)

** 교신저자, 중앙대학교 문헌정보학과 교수(haklaekim@cau.ac.kr, <https://orcid.org/0000-0002-2616-421X>)

I. 서론

인공지능(AI) 기술이 산업과 공공 영역 전반에서 혁신을 주도함에 따라 데이터 중심 사회의 패러다임은 기존의 '대규모 데이터(Big Data)' 중심에서 '데이터의 질과 적합성(AI Ready Data, AIRD)' 중심으로 전환되고 있다(EY, 2024). AI 모델의 성능과 신뢰성은 학습과 추론 과정에서 활용되는 데이터의 품질에 의존하며, 불완전하거나 편향된 데이터는 환각(hallucination), 오류 추론 등 사회적 신뢰를 심각하게 훼손할 수 있다(Gartner, 2025). 이에 따라 데이터 품질 관리의 패러다임이 전환되고 있다. 전통적으로 강조되어 온 정확성, 완전성 등을 넘어 대표성, 편향 완화, 독성 관리 등 AI 환경에서 요구되는 품질 요소가 핵심 요인으로 부상하고 있다(Gartner, 2024a; McKinsey & Company, 2023). 이에 따라, AIRD는 단순한 기술적 개념을 넘어 산업계, 학계, 국제기구, 정부 등 다양한 이해관계자로서 다른 목적을 가지고 접근하는 복합적 개념으로 확장되고 있다(Piontek, 2024).

한국의 공공데이터 정책은 이러한 글로벌 패러다임 전환 속에서 중요한 도전에 직면해 있다. 한국은 OECD OURdata Index 평가에서 지속적으로 최상위권을 유지하며 양적 개방 측면에서 우수한 성과를 기록해 왔지만(OECD, 2023), 활용성 측면에서 구조적 한계가 존재한다. 원천데이터와 개방데이터 간 분리된 관리 체계, 기존 품질 기준의 한계, 기계 판독성과 표준화의 부족 등은 공공데이터 활용성을 제약하는 주요 요인으로 지적되어 왔다(김학래, 2020). 이러한 문제의식 속에서 정부는 2025년 'AI Ready 공공데이터' 정책을 공식 선언하며 공공데이터 정책의 무게중심을 단순 개방에서 AI 기반 활용성과 품질 확보로 전환하는 노력을 진행하고 있다(행정안전부, 2025). 그러나 이 정책은 선행 연구와 실증 도구 개발보다 정책 추진 동력이 먼저 형성된 정책 선행(policy-first) 접근의 성격을 지니고 있어, 실효성 확보를 위해 AIRD의 학술적 정의와 실행 기준이 정교하게 마련될 필요가 있다.

본 연구는 AI 시대의 새로운 데이터 패러다임인 AIRD를 학술적으로 정립하고, 한국의 'AI Ready 공공데이터' 정책을 평가하기 위한 분석틀을 제시하는 것을 목적으로 한다. AIRD는 아직 국제적으로 합의된 정의나 기준이 존재하지 않는 초기 개념이기 때문에, 본 연구는 개념적 고찰(conceptual review)을 통해 다양한 이해관계자의 접근 방식을 종합적으로 분석한다. 이를 통해 AIRD의 구성요소를 체계적으로 도출하고, 품질·신뢰성, 표준화·상호운용성, 거버넌스·책임성, 다양성·포용성, 지속적 운영 가능성의 다섯 가지 핵심 요소로 구조화한 통합적 분석틀을 제시한다. 이 분석틀을 한국의 'AI Ready 공공데이터' 정책에 적용하여 정책의 현황과 한계를 평가하고, 국제적 기준과의 정합성을 기반으로 정책 발전 방향을 제시한다.

본 연구에서 'AI Ready Data(AIRD)'와 'AI Ready 공공데이터'는 범위와 성격에서 구분된다. AIRD는 AI 학습·추론·의사결정에 적합한 데이터의 구조적·관리적 조건을 지칭하는 국제적 개념으로, 민간·공공·연구 영역 전반의 데이터를 포괄한다. 반면 'AI Ready 공공데이터'는 공공기관이 생성·관리하는 데이터를 AI 기반 공공서비스와 정책 의사결정에 활용 가능한 형태로 전환하는 것을 목표로 한다(행정안전부, 2025). 행정안전부는 이후 발간한 가이드라인(행정안전부, 2026)에서 '인공지능 친화적 공공데이터'를 공식 용어로 채택하였다. 본 연구는 AIRD의 일반 개념을 학술적으로 정립한 뒤, 이를 분석틀로 활용하여 한국의 'AI Ready 공공데이터' 정책을 평가하는 구조를 갖는다.

본 연구는 세 가지 핵심 문제의식을 중심으로 전개된다. 첫째, AIRD에 대한 산업계, 학계, 국제기구, 그리고 정부 정책의 접근 방식은 각각 고유한 특징과 관점을 보여준다. 둘째, 이러한 다양한 접근 방식으로부터 AIRD를 구성하는 핵심 요소들을 추출할 수 있다. 셋째, 도출된 AIRD의 구성요소를 기준으로 한국의 'AI Ready 공공데이터' 정책을 체계적으로 검토하여 정책적 시사점을 도출할 수 있다.

본 논문의 구성은 다음과 같다. 제II장은 공공데이터

정책, 데이터 품질 관리, FAIR 원칙, 데이터 거버넌스 등 관련 선행연구를 검토한다. 제Ⅲ장은 개념적 고찰 방법론을 설명하고 문헌 선정 기준을 제시한다. 제Ⅳ장은 산업계·학계·국제기구·정부 등 이해관계자별 AIRD 접근 방식을 비교 분석하여 통합적 구성요소를 도출한다. 제Ⅴ장은 도출된 AIRD 분석틀을 적용하여 한국의 'AI Ready 공공데이터' 정책을 체계적으로 평가한다. 제Ⅵ장은 정책 개선을 위한 구조적 과제와 프레임워크 적용 방향을 제시하며, 제Ⅶ장은 연구의 학술적·정책적 기여, 한계, 향후 연구 방향을 논의한다.

II. 선행연구

1. 인공지능 시대의 공공데이터 정책 변화

인공지능 시대의 공공데이터 정책은 단순한 개방의 양적 확대를 넘어, 품질·표준화·상호운용성을 포함한 질적 전환을 요구받고 있다. 한국은「공공데이터의 제공 및 이용 활성화에 관한 법률」제정(2013)과 공공데이터포털 구축을 통해 비교적 빠르게 개방 인프라를 확충해 왔으나, 개방된 데이터가 실제 활용으로 이어지지 않는 구조적 간극이 지속적으로 지적된다. 이러한 간극은 데이터 품질의 불균형, 메타데이터의 미흡한 정합성, 기관·지자체 간 관리 역량 격차, 갱신 주기의 불규칙성 등과 연동되어 나타난다.

국내 연구는 이러한 한계를 품질, 표준화, 상호운용성 측면에서 분석해 왔다. 김학래(2020)는 개방 표준데이터의 완전성·정확성을 정량 평가하고 품질 개선 방향을 제시하였으며, 후속 연구는 메타데이터 표준화(김현지·김학래, 2022), 값 수준 품질 및 식별자 정비(박하람·김학래, 2024), 지식그래프 기반 데이터맵·검색 프레임워크 구축(송채은·김학래, 2025)으로 확장되었다. 거버넌스 연구에서는 PDCA 기반 품질관리 성숙도 모델(김선호 외, 2015), 지자체 역량이 개방 품질에 미치는 영향 분석(송채은·김학래, 2022)이 제시되었다. 차오신과 오세환(2024)은 공공데이터포털의 대규모 파

일데이터 분석을 통해 설명의 정량성, 오픈 포맷, 사용자 평가, 정기적 업데이트가 활용에 영향을 미치는 요인임을 분석했다. 황성수(2025)는 공공 AX 정책을 분석하며 공공데이터가 AI 정부의 핵심 기반임을 강조하고 있다.

다만, 대한민국 공공데이터 품질·표준화 정책은 여전히 행정기관 내부 원천 데이터베이스(DB) 중심에 머물러 있으며, 개방 과정에서 발생하는 품질 저하·구조적 정보 손실·메타데이터 불일치가 제도적으로 관리되지 않는다(행정안전부, 2020). 이러한 한계는 값(value) 수준 품질, 시계열 정합성, 식별자·참조코드 정비, 계보(lineage) 추적 등 AI 활용에 필수적인 요소에서 제약을 초래하며, AI Ready 상태를 위한 기준 정립은 아직 초기 단계이다.

2. 데이터 품질과 거버넌스 연구

데이터 품질 논의는 전통적으로 정확성, 완전성, 일관성, 적시성과 같은 내재적 지표에 기반해 발전해 왔다. 그러나 디지털 전환과 대규모 데이터 통합, AI 기반 의사결정 확산은 품질 개념을 구조적·맥락적·사회적 품질로 확장시키는 흐름으로 강화하고 있다. 최근 품질 관리 연구는 표준화, 재사용성, 사회적 편향, 기계 판독성, 출처·계보 관리 등 다차원적 요소를 핵심 영역으로 제시한다.

Vetrò et al.(2016)은 오픈 정부데이터 품질을 구조·문맥·메타데이터 기반의 다층적 프레임워크로 제시하고, Lehmann et al.(2016)은 식별자·표준 어휘·연결성을 중심으로 링크드 데이터 품질을 체계화하였다. Sadiq and Indulska(2017)는 개방 정책의 초점을 양적 공개에서 질 중심으로 이동해야 한다고 강조하고, Zhang and Xiao(2020)은 최신성·사용성·기계 판독성 중심의 품질 재편 경향을 분석하고 있다.

국내 연구도 유사한 확장 경향을 찾을 수 있다. 정하나 외(2022)는 도메인 규칙 기반 품질진단 모델을 제안하고, 김용환 외(2023)는 품질관리를 특정 시점 점검이

아닌 업무 프로세스에 내재된 지속적 개선 구조로 전환하였다. 윤건(2023)은 거버넌스 구조가 데이터기반 행정과 융합에 미치는 영향을 실증하였으며, 정기홍과 김훈(2025)은 표준 공공데이터 사례 분석을 통해 구조적 품질지표의 필요성을 제기하였다.

AI 환경에서 품질 요건은 더욱 강화된다. 대표성, 편향 완화, 독성 관리, 설명가능성, 출처·계보 추적 등은 AI 모델 신뢰성과 직접적으로 연계된다. FAIR 데이터 원칙(Jati et al., 2022)은 탐색성·접근성·상호운용성·재사용성을 제시하며 국내에서도 적용이 확산되고 있다(김학래, 2021; 박하람 외, 2024; 성혜정 외, 2024). 한편 FAIR for AI, FAIR², FAIR-R(Ravi et al., 2022; Huerta et al., 2023; Verhulst et al., 2025)은 기계 판독성, 문맥 정보, 책임성을 중심으로 품질 기준을 재해석한다.

데이터 품질 개념의 확장은 거버넌스 구조 변화로 이어진다. AI 환경에서 편향 관리, 계보 추적, 문서화, 사용 이력, 설명가능성 등이 추가 요소로 논의되고 있으며, 국제 사회 역시 이를 제도화하고 있다. 미국「AI 위험관리 프레임워크(AI RMF 1.0)」(2023), 유럽연합「인공지능 규제(AI Act)」(2024), OECD AI 원칙(2019), G7 히로시마 프로세스(2023)는 데이터 책임성·편향 검증·감사 가능성 등을 핵심 요건으로 포함하고 있다.

대한민국은「공공데이터의 제공 및 이용 활성화에 관한 법률」과「데이터기반 행정 활성화에 관한 법률」을 통해 개방·공동활용 체계를 구축했으나, 편향 관리, 계보 추적, 사용이력 관리 등 AI 시대의 핵심 요건은 충분히 반영되지 않았다. 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(2025)은 신뢰 기반 조성을 규정하지만, 학습데이터 검증·품질 보증·계보 관리 등 세부 기준은 향후 정교화가 필요하다.

3. 이해관계자별 AIRD 접근 동향

AIRD는 산업계, 학계, 국제기구, 정부 정책 등 다양한 이해관계자가 서로 다른 목표와 필요에 따라 접근하

는 복합적 개념이다(Sallam, 2024). 산업계는 AI 프로젝트 성과와 투자 회수(ROI; Return on Investment)를 중심으로 목적 적합성, 데이터 다양성, 자동화된 파이프라인을 강조하며(Gartner, 2025), 학술·연구기관은 FAIR/ 확장된 FAIR 원칙 기반의 재현성·책임성·편향 완화를 중시한다. 국제기구와 표준화 기구는 상호운용성, 발견 가능성, 투명성, 규범 준수를 중심으로 AIRD 관련 기준을 제시하고 있다(World Bank, 2025).

AIRD는 국가 전략 목표와도 밀접하게 연계된다. 영국은 National Data Library(NDL) 구상과 함께 AI Ready Data 개념을 공식화하여 표준 포맷, 시맨틱 상호운용성, 편향 설명, 버전 관리 등을 제시하고(Majithia & Simperl, 2024; Oderinwale & Longpre, 2024; ODI, 2024a), 유럽연합과 미국은 데이터 품질·편향·문서화 요건을 규제·위험관리 체계로 법제화하고 있다(European Commission, 2025; U.S. Department of Commerce, 2024). 대한민국은 「AI Ready 공공데이터」라는 정책 용어를 도입하며 공공데이터 중심의 AIRD 제도화를 추진하는 초기 단계에 있다(행정안전부, 2025).

이해관계자별 접근은 차이가 존재하지만, 품질·신뢰성, 표준화·상호운용성, 거버넌스·책임성, 다양성·포용성, 지속적 운영 가능성이라는 공통 구조가 반복적으로 등장한다. 본 연구는 이러한 공통 요소를 AIRD의 핵심 구성요소로 정립하고, IV장에서 이해관계자별 접근을 비교·분석한다.

III. 연구 방법

1. 연구 접근법

AIRD는 국제적으로 합의된 정의나 측정 기준이 확립되지 않았으며, 한국에서도 정책적 선언이 학술적·실무적 논의를 앞선 초기 개념으로 등장하고 있다(행정안전부, 2025; ODI, 2024a; World Bank, 2025). AIRD

는 이해관계자별 해석이 상이하고 구성요소가 다층적이기 때문에 단일 지표나 특정 사례 기반 접근으로 포착하는 데 한계가 존재한다. 이러한 특성을 고려하여 정량적 분석보다 개념적 틀을 정립하는 접근이 적합하다.

본 연구는 AIRD라는 새로운 개념의 구조와 구성요소를 도출하기 위해 개념적 고찰(conceptual review)을 적용한다. 개념적 고찰은 초기 정책 개념의 지형을 정리하고 핵심 속성을 구조화하는 데 적합한 방식으로, 개념의 복잡성과 다층성을 설명하는 데 유효한 방법론이다(Stebbins, 2001). 이에 따라 학술 문헌뿐 아니라 산업계 보고서, 국제기구 문서, 국가 정책 프레임워크까지 범위를 확장하여 AIRD 논의의 공통 구조를 파악하고 개념적 구성요소를 도출한다.

2. 문헌 수집 및 선정 기준

문헌은 학술 연구, 정책·법제 문서, 국제기구 보고서, 산업계 기술 백서 등 다양한 출처를 포괄적으로 수집한다. 데이터·AI 관련 법제와 정책 문서를 검토하여 AIRD의 제도적 기반을 파악하였고, 국제기구의 AI·데이터 규범(NIST AI RMF, EU AI Act, OECD AI 원칙, G20(2019)·G7(2023) 문서 등)을 통해 국제기준의 흐름을 분석한다. 산업계 자료는 실제 활용사례와 요구조건을 보완적으로 확인하기 위한 참고 자료로 활용한다.

문헌 선정은 다음 기준을 따른다: 1)AIRD, 데이터 준비성, 품질, 편향 관리, 표준화 등과 직접적으로 연관된 문헌, 2)공공데이터 품질·표준화·거버넌스·국가 데이터 전략 등 분석틀 구성에 기여하는 문헌, 3)개념 정의, 프레임워크, 지표 등 구조적 논의를 포함한 문헌. 한편, 최근 정책과 기술 변화 반영을 위해 2015년 이후 문헌을 대상으로 하나, 연구방법(Stebbins, 2001), 데이터 프레임워크(Janssen et al., 2012)와 같이 해당 분야의 개념 형성에 핵심적 역할을 한 기초 문헌은 예외적으로 포함한다. 반면, AIRD와의 개념적 연계가 약한 일반적 빅데이터·오픈데이터 문헌, AI 알고리즘·모델 구조 등

데이터 준비성과 무관한 기술 중심 문헌, 특정 조직 사례에 국한되어 일반화 가능성이 낮은 문헌, 비학술적 또는 검증되지 않은 출처의 자료는 제외한다. 이러한 기준을 통해 최종적으로 65편의 문헌을 분석 대상으로 확정한다.

3. 분석 관점의 설정

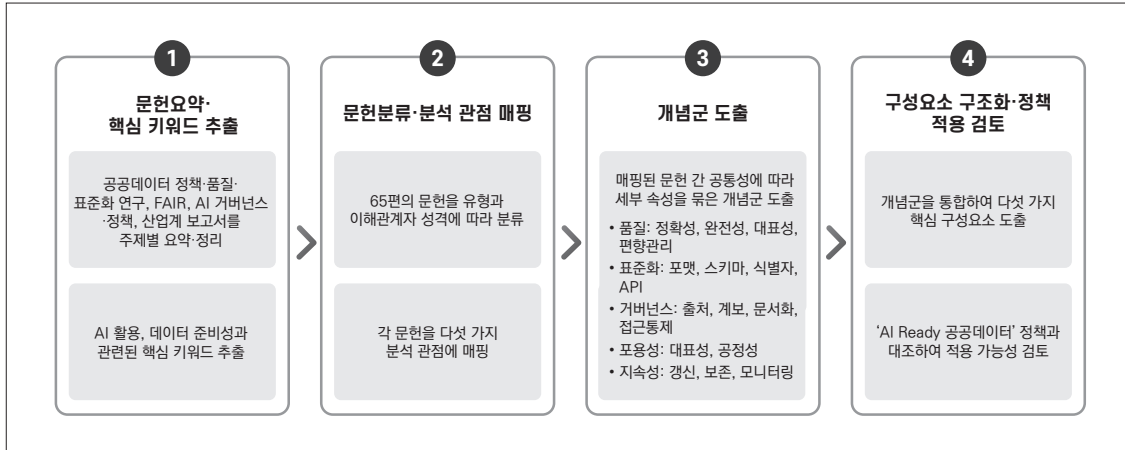
AIRD는 인공지능 학습·추론에 적합한 데이터의 상태를 의미하지만, 이해관계자에 따라 강조점이 다르게 나타난다. 산업계는 실용성과 목적 적합성, 학계는 재현성과 책임 있는 데이터 관리, 국제기구와 정부 정책은 규범 준수와 투명성을 중시한다. 이러한 다양한 관점을 통합적으로 분석하기 위해 선행연구와 정책 문헌을 기반으로 다섯 가지 핵심 관점을 도출한다.

첫째, 품질·신뢰성은 정확성, 완전성, 대표성, 편향 관리 등 데이터 자체의 품질 속성을 포함한다. 둘째, 표준화·상호운용성은 포맷, 메타데이터, 시맨틱 구조 등 데이터가 이질적 시스템 간에 결합·재사용될 수 있는 정도를 의미한다. 셋째, 거버넌스·책임성은 출처·계보 관리, 문서화, 접근통제, 규제 준수 등 책임 기반 데이터 운영 요소를 포함한다. 넷째, 다양성·포용성은 데이터가 다양한 집단과 맥락을 적절히 대표하며 특정 집단을 배제·왜곡하지 않는 속성을 의미한다. 다섯째, 지속적 운영 가능성은 갱신, 보존, 모니터링 등 데이터 생애주기 전반에서 이루어지는 운영 체계의 유지·관리를 포함한다.

이 다섯 가지 관점은 AIRD 관련 산업·학술·정책·국제기구 문헌에서 반복적으로 등장하는 구조적 요소를 통합한 상위 범주이다. IV장에서 이러한 관점을 기반으로 이해관계자별 AIRD 접근을 비교·분석한다.

4. 개념 도출과 통합 절차

〈그림 1〉에서 보듯이, AIRD 구성요소는 문헌과 정책 자료의 체계적 분석을 통해 네 단계로 도출하였다.



〈그림 1〉 AI Ready Data(AIRD) 개념 도출과 통합 절차

〈표 1〉 AI Ready Data(AIRD)의 5대 구성요소 요약

구성요소	핵심 내용	세부 속성
품질·신뢰성 (Quality & Reliability)	데이터 정확성, 신뢰 가능한 상태	• 정확성(accuracy) • 완전성(completeness) • 일관성(consistency) • 시계열 연속성(temporal continuity) • 대표성(representativeness) • 편향 관리(bias mitigation)
표준화·상호운용성 (Standardization & Interoperability)	다양한 시스템·도메인 간 결합·연계 가능한 구조	• 표준 포맷(format) • 메타데이터 스키마(metadata schema) • 참조코드·식별자(reference codes, identifiers) • 시맨틱 어휘(ontologies, vocabularies) • API 기반 접근성과 시스템 연계성
거버넌스·책임성 (Governance & Accountability)	데이터의 출처·계보, 보안·책임성 관리	• 출처·계보(lineage) 관리 • 접근통제(access control) • 개인정보·보안 요건 • 문서화(documentation) • 감사 가능성(auditability)
다양성·포용성 (Diversity & Inclusivity)	데이터의 사회적 다양성을 반영, 편향 최소화	• 사회적 대표성(social representativeness) • 과소·과대표집 방지 • 편향 완화(fairness, bias reduction)
지속적 운영 가능성 (Sustainability & Maintainability)	데이터의 장기적 운영·갱신·재사용	• 정기 갱신(update cycle) • 모니터링(monitoring) • 버전관리(versioning) • 장기 보존(long-term preservation) • 재사용 가능성(reusability)

첫째, 공공데이터 품질, 표준화, 상호운용성, FAIR/확장 FAIR 논의, AI 거버넌스·정책, 산업계 보고서를 주제별로 정리하고 핵심 키워드를 추출한다. 둘째, 선정된 65편의 문헌을 문헌 유형과 이해관계자 관점에 따라 분류한 뒤, 각 문헌에서 강조하는 요건을 다섯 가지 분석 관점과 연계하였다. 셋째, 연계된 문헌의 공통 속성을 비교하여 품질, 표준화, 거버넌스, 포용성, 지속성에 해당하는 개념군을 도출하였다. 넷째, 도출된 개념군을 통합하여 AIRD의 다섯 가지 상위 구성요소로 구조화하고, 행정안전부의 'AI Ready 공공데이터' 정책 방향(2025), 국내 공공데이터 품질·표준화 연구와 대조하여 적용 가능성을 검토하였다.

이와 같은 절차를 통해 AIRD는 품질·신뢰성, 표준화·상호운용성, 거버넌스·책임성, 다양성·포용성, 지속적 운영 가능성의 다섯 요소로 구성되는 다층적 개념으로 정리된다. <표 1>은 이러한 AIRD의 개념적 구성요소를 요약한 것이다.

IV. 이해관계자별 AIRD 접근 방식 분석

AIRD는 다양한 이해관계자가 서로 다른 목표와 맥락에서 접근하는 개념으로, 고유의 역할과 필요에 따라 구성요소를 상이하게 해석한다. 산업계는 활용성과 비즈니스 가치, 학계는 과학적 재현성과 책임성, 국제기구

화와 공공가치를 각각 우선한다. <표 2>는 Janssen et al.(2012)과 Zuiderwijk and Janssen(2014)이 제시한 오픈데이터 이해관계자 분류를 AI 시대의 데이터 환경에 맞게 확장한 것이다. 이와 같은 구분을 전제로 하여, 이해관계자별 AIRD 접근을 구조적으로 비교한다.

1. 산업계

산업계는 AIRD를 AI 프로젝트 성과를 좌우하는 실천적 자원으로 이해한다(Gartner, 2023, 2024a; EY, 2024; McKinsey & Company, 2022, 2023). 전통적 품질 기준을 넘어, 산업 현장에서 가장 크게 부각되는 요소는 목적 적합성, 데이터 다양성, 시계열 연속성과 같이 모델의 실제 성능과 직결되는 속성이다. Gartner(2024a)가 지적하듯, 데이터의 완전성보다 현실적 변동성이나 이상치 정보가 모델의 일반화 성능에 기여하는 경우도 많다. 한편, 산업계는 표준화·상호운용성을 기술적 요구로 명확히 제시하며, 포맷 통일성, 자동화된 메타데이터 생성, API 기반 접근성 등 실제 개발·운영 환경에서 필요로 하는 요소를 AIRD의 핵심 기준으로 본다. 이러한 요구에 대응하기 위해 주요 클라우드 기업은 데이터 레이크, 카탈로그, 거버넌스를 통합한 운영 구조를 제공하고 있다.

지속적 운영 관점에서 보면, 데이터 파이프라인의 자동화가 핵심 요소로 논의되고 있다. 수집-정제-검증-모

<표 2> AIRD 이해관계자 유형과 주요 특성

이해관계자	주요 목적	대표 주체	주요 산출물	강조점
산업계	비즈니스 가치 창출·성능 개선	Gartner, EY, McKinsey, AWS, Azure	실행 가이드, 플랫폼, 기술 백서	목적 적합성, ROI, 데이터 다양성
학술·연구기관	과학적 재현성·책임성	FAIR 커뮤니티, 대학·연구소	학술논문, 개념 프레임워크	투명성, 출처·계보, 편향 완화
국제기구·표준화	규범·표준 설정	OECD, G7, NIST, EU, World Bank	권고안, 표준, 지침	상호운용성, 투명성, 책임성
정부 정책	제도화·국가 경쟁력·데이터 주권	영국, 미국, EU, 한국	법령, 정책, 가이드라인	규제기준, 공공성, 지속 운영 체계

니터링이 순환적으로 자동화되지 않으면 모델 성능 저하(model drift)가 발생하기 때문에, 산업계는 운영 자동화와 품질 모니터링을 AIRD의 필수 요건으로 인식한다(Gartner, 2024b, 2024c). 여기에 고객군의 대표성·공정성을 포함한 데이터 다양성·포용성 요구가 규제·평판 위험과 맞물려 점점 더 강조되는 추세이다(Kidwai-Khan et al., 2024).

2. 학술·연구기관

학계는 AIRD를 재현성·책임성·투명성을 보장하는 지식 인프라로 이해하며, FAIR 원칙(Wilkinson et al., 2016)을 AI 시대에 맞게 확장하는 데 주력하고 있다(Huerta et al., 2023; Verhulst et al., 2025). AIRD의 핵심은 데이터가 과학적 검증을 견딜 수 있는 구조와 문서화를 갖추는 데 있으며, 레이블 품질, 계보·출처 정보, 메타데이터 수준, 편향 검증이 주요 논점으로 논의된다. Raza et al.(2024)는 FAIR 기반 데이터가 모델의 독성 감소와 정확도 향상에 기여함을 실증하며 이러한 관점을 뒷받침한다.

한편, 학계는 기계가 이해할 수 있는 상호운용성 기반을 중시한다. 시맨틱 기술(RDF, SKOS), 영구식별자(PID), 표준 메타데이터 스키마 등은 데이터의 연결성과 확장성을 확보하는 핵심 조건으로 간주된다. 국내 연구에서도 DCAT-AP-KR 기반 표준화(박하람·김학래, 2022), 데이터맵 기반 지식그래프 연구(송채은·김학래, 2025) 등 상호운용성 중심 접근이 확장되고 있다.

윤리적·사회적 관점에서도 학계는 책임성, 편향 완화, 대표성 확보 등을 중요하게 다룬다. FAIR-R, FAIR²(Senscience, 2025)와 같은 프레임워크는 커뮤니티 기반 운영과 지속 가능한 연구 데이터 생태계를 구축하기 위한 확장된 AIRD 요소를 제시하며(Verhulst et al., 2025; Hiniduma et al., 2025), 데이터 품질을 기술적 요건을 넘어선 사회적·윤리적 구성요소로 이해하도록 요구한다.

3. 국제기구·표준화 기관

국제기구와 표준화 기관은 AIRD를 기술적 최적화보다 규범·책임·상호운용성의 관점에서 이해하는 경향이 있다. 이들은 AIRD라는 용어를 정의하지 않지만, AI 신뢰성·책임성·안전성을 확보하기 위한 규범을 제시하는 과정에서 AIRD의 핵심 구성요소와 사실상 동일한 기준을 요구한다.

EU AI Act와 NIST AI RMF는 대표성, 편향 검증, 문서화, 지속적 모니터링 등 데이터 품질과 책임성을 법적·제도적 요건으로 규정한다. 이러한 규범은 데이터 계보(lineage), 감사 가능성(auditability), 문서화와 같은 거버넌스 요소를 국제적 기준으로 정착시키고 있다. 상호운용성 관점에서도 세계은행, ODI(Open Data Institute) 등은 표준 메타데이터 구조, API 기반 접근성, 발견 가능성을 강조하며, 인간과 AI가 모두 활용할 수 있는 데이터 환경의 구축을 핵심 목표로 제시한다. 한편, 국제기구는 AIRD의 사회적 역할을 고려하여 포용성·대표성을 기본 원칙으로 포함한다. EU AI Act는 대표성 부족이 초래하는 위험을 명확히 규정하고 편향 완화를 의무화하며, OECD는 인간 중심 가치와 다양성을 데이터 관리 원칙으로 제시한다. 지속성 측면에서도 국제기구는 AIRD를 일회성 프로젝트가 아닌 지속 가능한 공공 인프라로 이해하며, 장기 보존·갱신·접근성 확보를 국가 수준의 데이터 정책 요구로 제시한다.

4. 정부 정책

정부 정책은 AIRD를 데이터 주권, AI 경쟁력, 공공서비스 혁신과 같은 국가 전략 목표와 연계해 해석한다.

영국은 National Data Library(NDL)와 Multilayer Interoperability Framework(MIF)를 통해 AIRD 개념을 정책 문서에서 명시적으로 사용한 대표적 사례다(UK Government, 2024, 2025). 표준 포맷(Parquet, RDF), 버전 관리와 문서화, 편향 설명, 시맨틱 기반 상호운용성 등의 구체적 요건을 제시하며, 공공·민

간·연구 데이터를 AI 활용 가능한 형태로 통합하는 국가 데이터 인프라를 구축하려는 전략을 추진한다. 예를 들어, Hill(2024), King's College London(2023), ODI(2024b), Meroño-Peñuela et al.(2025)는 영국 NDL·AIRD 논의를 구체화하고 있다. 미국은 NIST AI RMF를 통해 책임성 중심 프레임워크를 제공하고, 위험 기반 규율 구조에 집중한다. NIST AI RMF는 데이터 품질, 편향 검증, 계보 추적 등을 AI 시스템 위험 관리의 핵심 요소로 제시하며, 자발적 준수(voluntary compliance)를 권장한다.

한국은 2025년 행정안전부를 중심으로 처음으로 'AI Ready 공공데이터'라는 정책적 개념을 제도화하고 있다(행정안전부, 2025). 그러나 정책은 초기 단계로, 일반적인 데이터 품질개선 중심의 접근에 머물며, 메타데이터·표준화의 구체적 기준, 포용성·지속성 등 AIRD 5개 요소 대비 부족한 영역 등이 과제로 남아 있다.

5. 소결

AIRD 5개 구성요소는 이해관계자에 따라 강조점과 목적에서 차이를 보이지만, 전 분야에서 반복적으로 나타나는 구조적 공통 요소이다(〈표 3〉 참고). 이러한 공통 구조는 AIRD를 단일 기술 개념이 아닌, AI 시대의

데이터 관리·운영 전반을 설명하는 종합적 프레임으로 위치시키며, 한국 공공데이터 정책 분석을 위한 기준으로 활용된다.

V. 한국의 'AI Ready 공공데이터' 정책 분석

1. 공공데이터 전환계층 모델(Public Data Transition Layer Model)

한국의 공공데이터는 원천 생성(TL0), 기관 내부 가공(TL1), 대국민 개방(TL2)의 세 단계로 구성된다. 이러한 구조를 Transition Layer 0·1·2로 유형화하고, 전체를 공공데이터 전환계층 모델(Public Data Transition Layer Model)로 정의한다. 전환계층 모델은 데이터의 구조, 정의, 품질, 표준, 계보가 각 단계를 거치며 어떻게 형성되고 변형되는지를 설명하는 분석틀이며, 공공데이터가 AI 환경에서 활용되기 위한 구조적 조건을 평가하는 데 사용된다. 〈그림 2〉는 TL0(원천데이터), TL1(기관 내부 가공데이터), TL2(개방데이터)의 3단계 구조와 생성·가공·개방의 하향 흐름과 TL2에서 TL1, TL0로 이어지는 역방향 피드백 흐름을 함께 제시하며, TL2를 품질개선의 실행 계층, TL1을 전달·조정 계층, TL0을 품질 기원 계층으로 위치시킨 순환적 품질관리

〈표 3〉 이해관계자별 AIRD 5대 요소 강조점 비교

요소	산업계	학계	국제기구	국가 정책
품질·신뢰성	목적 적합성, 다양성, 규모	재현성, 레이블링 품질, 편향 검증	법적 의무화, 고위험 시스템 품질 기준	기준 설정·검증체계 구축
표준화·상호운용성	API·플랫폼 이동성, 메타데이터 자동화	메타데이터 스키마, 시맨틱 기술, PID	국제 표준, 발견성, 통합 인터페이스	상호운용 프레임워크, 국가 표준
거버넌스·책임성	위험관리, 감사·계보, 접근통제	투명성, 커뮤니티 책임성, 출처 추적	규범·안전성, 국제 협력 기준	법령·제도 기반, 문서화 의무
다양성·포용성	시장 대표성, 비즈니스 위험 관리	편향 완화, 윤리성, 공정성	대표성·차별 방지 규정, 포용적 성장	공공가치 반영, 사회적 책임
지속적 운영 가능성	자동화 라이프라인, 모니터링	생애주기 관리, 버전·보존	지속 가능한 인프라, 공공재 관점	장기 운영 체계, 제도적 지원



〈그림 2〉 한국 공공데이터 관리의 전환계층 구조(Transition Layer Model)와 AI Ready 관점에서 본 구조적 한계

구조를 나타낸다.

기존 공공데이터 정책은 행정정보시스템 내부의 데이터베이스 품질관리와 형식 점검에 초점을 두어 왔다(김학래, 2020). 그러나 원천데이터는 업무 시스템 구조와 법적 절차에 종속되어 있어 사후 수정이나 표준화가 쉽지 않고, 기관 내부 가공 단계는 품질관리 체계와 분리된 채 운영되어 왔다(송채은·김학래, 2022). 이러한 구조적 요인은 TL0·TL1 단계의 불일치와 편차가 TL2 개방데이터에 그대로 반영되는 결과를 만든다.

전환계층 모델은 이러한 현실을 반영하여 TL0을 품질의 기초가 형성되는 계층으로 이해하고, 실질적 품질개선의 실행지점은 TL2이며, TL2의 결과가 TL1과 TL0로 순환적으로 전달되는 구조로 재정립된다. 이 관점에서 TL0은 행정업무 수행 과정에서 생성되는 운영데이터로, 값 수준 품질, 정의·용어 체계, 참조코드 적용, 시계열 연속성 등 근본적 속성이 최초로 형성되는 계층이다. 다만 업무 절차와 법적 규정에 결합된 구조적

특성 때문에 사후적 정비나 표준화 활동을 독립적으로 수행하기 어렵다는 제약이 존재한다.

TL1은 기관 내부에서 TL0 데이터를 분석, 보고, 집계, 관리 등을 위해 가공하는 계층이다. 이 과정에서 형식 변환, 항목 축약, 비공식적 처리, 구조 재배열이 발생하며 계보가 소실되거나 표준과의 정합성이 약화되는 문제가 반복된다. 동시에 TL1은 개방 직전의 사전 정비 단계를 포함하고 있어, 개방데이터의 구조와 품질이 실제로 형성되는 중간 계층이라는 특성을 갖는다. 그러나 이러한 역할은 제도적으로 명확히 규정되지 않아 품질개선 효과는 제한적 수준에 머물러 왔다.

TL2는 데이터가 외부에 공개되는 단계로, 품질·표준·정합성을 실질적으로 강화할 수 있는 개입 지점을 제공한다. 개방 단계는 형식·구조·참조코드의 자동 검증, 국가 표준 규칙의 일괄 적용, 이상치 탐지와 오류 교정, RDF·JSON-LD 등 시맨틱 구조 부여, 메타데이터 보강, 문서화, 변경이력·버전관리 제공 등 다양한 품질

관리 활동이 가능하다. 이러한 기능은 TL0·TL1 단계에서 발생한 품질·정합성 문제를 보완할 수 있는 현실적 기제이며, 개정된 전환계층 모델은 TL2를 품질개선의 핵심 실행계층으로 위치시킨다.

전환계층의 흐름은 생성-가공-개방으로 이어지는 전통적 하향(downstream) 구조와, 개방 단계에서의 품질관리·표준화 결과가 다시 TL1과 TL0로 전달되는 상향(upstream) 구조가 결합된 순환적 메커니즘이다. TL2에서 도출된 오류 보고서, 표준 변환 규칙, 참조코드 매핑 등의 결과가 기관 내부 데이터 구조와 운영데이터 품질 개선을 유도하고, 반복적 순환을 통해 장기적으로 구조적 편차를 감소시키는 효과를 가져올 수 있다. 이러한 구조는 원천데이터의 직접적 수정이 어려운 제도적 환경 속에서도 품질개선이 가능한 현실적 경로를 제시한다.

공공데이터 전환계층 모델의 의의는 다음과 같다. 첫째, 품질·정합성·표준화가 TL2에서 실질적으로 구현될 수 있는 구조적 조건을 명확히 제시한다. 둘째, TL2에서의 품질개선 활동이 TL1과 TL0로 전달되는 순환구조를 통해 전체 공공데이터 생태계의 품질을 점진적으로 향상시킬 수 있는 경로를 제도적으로 설계할 수 있다. 셋째, AIRD 5개 요소를 계층별로 배치함으로써 'AI Ready 공공데이터' 구축을 위한 구조적 기반을 구체화한다. 전환계층 모델은 이러한 점에서 공공데이터 품질관리 체계의 재구조화와 목적 기반 AI 활용체계 설계의 이론적 기반으로 활용할 수 있다.

2. 'AI Ready 공공데이터'의 조작적 정의

AIRD 개념은 국제적으로 주로 대규모 언어모델(LLM) 학습을 위한 비정형 데이터셋을 중심으로 논의되지만, 공공데이터 환경은 정형 구조(CSV, API), 표준코드 기반 값 데이터, 행정절차에서 생성되는 표 형식 데이터 등 서로 다른 특성을 가진다. 또한 'AI 활용'은 LLM 학습에 국한되지 않으며, 기계학습·통계분석·규칙 기반 추론·의미기반 검색·요약·추천뿐 아니라 지

식그래프 기반 추론, 온톨로지 기반 시맨틱 확장 등 다양한 형태의 AI 기능을 포함한다.

공공데이터를 위한 AIRD의 조작적 정의는 비정형 학습데이터 중심의 국제적 논의를 그대로 적용하기보다, AI 기반 분석과 추론이 가능하도록 만드는 구조적·관리적 준비성(readiness)에 초점을 두어 재구성할 필요가 있다. 전환계층 모델과 AIRD 5개 요소를 반영하여 'AI Ready 공공데이터'를 다음과 같이 정의한다.

'AI Ready 공공데이터'란, 공공기관이 생성·관리하는 정형·반정형·비정형 데이터 중, AI 기반 분석·추론·예측·의사결정을 지원할 수 있도록 품질·표준·거버넌스·포용성·지속성의 구조적·관리적 요건을 갖춘 데이터 상태(state)를 의미한다. AI Ready 공공데이터의 구조적·관리적 요건은 <표 1>에 제시한 기준을 따른다. 조작적 정의에서 강조할 점은 'AI Ready 공공데이터'가 특정 AI 모델(예: LLM)의 학습데이터로 직접 사용됨을 전제로 하는 개념이 아니며, 다양한 형태의 AI 활용이 가능하도록 준비된 데이터의 상태(state)를 지칭한다는 점이다.

3. 한국 정부의 'AI Ready 공공데이터' 정책 방향과 한계

한국 정부의 정책 방향은 2025년 9월 19일 발표된 보도자료와 「AI Ready 공공데이터 관리 가이드라인」(비공식 초안)을 분석 대상으로 한다. 해당 가이드라인은 최종 배포에 이르기까지 여러 차례의 수정·보완을 거쳤으며, 2026년 4월 24일 「공공데이터의 인공지능 친화적 관리 가이드라인」으로 공식 배포되었다(행정안전부, 2026). 초안과 배포본은 공공데이터의 품질관리와 메타데이터 중심으로 전체적인 방향과 맥락은 동일하게 유지된다.

행정안전부는 최근 공공데이터 정책에서 인공지능 활용을 고려한 데이터 관리 필요성을 강조하고 있다. 보도자료와 가이드라인은 공공데이터가 인공지능에 의해

자동 처리·해석될 수 있도록 구조적 조건을 갖추어 제공되어야 한다는 점을 반복적으로 언급한다. 이러한 접근은 기존의 양적 개방 중심 정책과 대비되는 문제의식으로, 데이터 품질, 표준화, 메타데이터 관리, 변경 이력 제공 등 AI 활용 가능성을 높이기 위한 관리 기준을 정비하려는 방향으로 이해할 수 있다.

가이드라인에서 확인되는 정책 방향의 핵심은 공공데이터의 품질관리 체계 정비와 메타데이터 강화에 있다. 표준 포맷 적용, 참조데이터 관리, 메타데이터의 기계 판독성 확보, 변경 이력과 문서화 요구사항은 비교적 구체적으로 제시된다. 한편 정확성, 대표성, 편향 가능성 같은 품질 요소에 대해 검토 여부와 한계 명시 요구함으로써, 인공지능 활용 과정에서 데이터의 해석 가능성과 책임성을 제고하려는 문제의식이 드러난다.

다만 이러한 정책 방향은 데이터값(value-level) 자체를 적극적으로 개선하는 방식이 아닌, 데이터의 상태와 한계를 명시하고 관리 정보를 강화하는 방식에 가깝다. 가이드라인에 포함된 품질 점검 항목들은 오류·이상치·편향 가능성을 탐지하거나 표시하는 데 초점을 두고 있으며, 원천데이터의 값 자체를 수정하거나 생성 단계로 환류하는 절차는 제도적으로 명확히 규정되어 있지 않다. 이와 같은 방식은 행정정보시스템에 종속된 원천데이터 구조와 부처 간 권한 분산이라는 현실적 제약을 반영한 설계 선택으로 해석할 수 있다.

한편, 보도자료에 언급된 인공지능 활용성이 높은 데이터를 중심으로 기준을 우선 적용한 후 점진적으로 확대하는 운영 방향은 가이드라인에 명시적으로 제시되어 있지 않다. 보건·의료, 재난·안전, 교통 등 특정 분야를 초기 적용 대상으로 설정하는 내용 역시 공식 기준으로 규정되지 않고 정책 담론과 사례 설명 수준에 머물러 있다.

가이드라인은 데이터 활용 주체와 인공지능 활용 맥락의 다양성을 인식하고 있으나, 활용 목적에 따라 데이터 제공 수준을 구분하는 목적 적합성 기반의 다층적 제공체계를 정책적으로 명시하고 있지 않다. 이러한 제공

방식은 가이드라인의 문제의식을 바탕으로 향후 보완이 요구되는 과제로 해석하는 것이 적절하다.

종합하면, 본 연구가 분석한 ‘AI Ready 공공데이터’ 정책은 인공지능 전환을 직접적으로 설계·추진하는 전략 문서라기보다, 인공지능 활용을 염두에 둔 공공데이터 관리 기준을 정비하고 메타데이터 중심의 품질 가시성을 높이기 위한 초기 제도화 단계로 평가할 수 있다. 이러한 접근은 공공데이터의 구조적 투명성과 책임성을 강화하는 데 기여하지만, 데이터값 수준의 품질 개선, 전환계층 간 환류 구조, 활용 목적에 따른 제공 수준 구분 등은 아직 정책적으로 구체화되지 않은 과제로 남아 있다.

4. AIRD 5개 요소 기준 정책 진단

공공데이터 전환계층 모델에서 확인되는 단계별 제약과 정부 정책의 실행 가능성, 그리고 국제 기준과의 정합성을 중심으로 현행 정책체계를 검토하였다. 평가 결과, 정책 방향은 AI 기반 활용을 지향하는 긍정적 전환을 제시하고 있으나, 구조적 성숙도는 초기 단계에 머물러 있다.

품질·신뢰성 측면에서 현행 정책은 개방 단계의 기본적 형식 점검에 머물러 있다. TL0에서 결정되는 값 수준 품질, 정의·용어 체계, 시계열 연속성 등에 대한 사후 수정이 제약되지만, TL2에서의 자동검증·표준규칙 적용·이상치 탐지를 통해 상위 단계의 불일치를 탐지·보고·보완할 수 있는 구조가 필요하다. 그러나 현재의 TL2는 품질개선 실행계층으로 충분히 작동하지 못하며, 값 수준 품질, 시계열 정합성, 대표성·편향성 검증 등 AIRD가 요구하는 품질 기준이 제도적으로 명확히 규정되어 있지 않다.

표준화·상호운용성은 TL0·TL1·TL2 전 단계에 걸친 이질적 포맷·식별자·참조코드·시맨틱 어휘 문제로 인해 제약을 받는다. 특히 TL1 단계의 비공식 정제, 항목 축약, 형식 변환이 TL2 개방 시점의 구조적 단절을 만들 수 있다. TL2가 표준규칙의 일괄 적용과 시맨

터 구조 부여를 통해 상호운용성의 출발점을 마련할 수 있으나, 현행 정책은 이러한 기능을 제도적으로 규정하지 않는다. 영국의 NDL, EU의 DCAT/SEMIC 지침과 비교하면 단층적 표준화 수준에 머물러 있으며, 참조어휘·식별자·스키마를 통합적으로 관리하는 국가 표준 체계의 제도화가 미흡하다.

거버넌스·책임성에서 가장 두드러진 문제는 TL1 단계에서 가공 이력과 변환 정보가 체계적으로 기록되지 않는다는 점이다. 따라서 TL2에서의 품질개선·표준화 결과가 상위 계층으로 불완전하게 전달되는 원인이 될 수 있다. 또한, 정부 정책이 거버넌스 강화의 필요성을 언급하고 있지만, 계보 추적, 변경이력 관리, 출처 투명성 확보를 위한 제도적 요건은 구체화되지 않았다. EU AI Act의 학습데이터 계보 의무나 NIST AI RMF의 위험 기반 문서화 요구와 비교할 때, 한국의 계보 관리체계는 초기 단계이다.

다양성과 포용성은 가장 큰 정책적 공백이 나타나는 영역이다. 사회적 대표성, 편향성, 소수 집단의 과대표·과소 반영 여부 등을 체계적으로 점검하거나, 편향을 식별·완화하기 위한 정책적 절차가 존재하지 않는다. 공공데이터가 AI 모델 학습에 활용되는 상황에서 이러한 공백은 알고리즘 편향·불평등 재생산 가능성을 확대시키는 요인으로 작용한다. EU AI Act에서 다양성과 공

정성을 핵심 요건으로 규정하는 것과 비교하면, 한국의 정책체계는 관련 요소가 제도적으로 반영되지 않은 상태이다.

지속적 운영 가능성 측면에서도 데이터 생애주기 전체를 포괄하는 체계가 미비하다. 갱신주기, 버전관리, 변경 이력의 자동화, 모니터링 체계가 기관별로 편차가 크고, TL2에서 수행되는 품질개선 결과가 TL1과 TL0에 순환적으로 전달되기 위한 구조적 기반이 마련되지 않았다. 데이터가 정적 자원으로 취급되는 경향이 강하며, 장기적 갱신·보존·정합성 유지를 위한 표준 절차는 부족하다. 국제기구가 데이터 인프라를 지속 가능한 공공재로 접근하는 흐름과 비교하면 제도적 보완의 필요성이 크다.

이러한 평가 결과는 <표 4>에 요약하였다. 종합적으로 보면, 품질·신뢰성은 상대적으로 높은 수준이나 TL0·TL1·TL2를 연계하는 다층적 품질관리 기준이 마련되지 않았고, 표준화·상호운용성, 거버넌스, 지속성은 중하 수준으로 평가된다. 다양성과 포용성은 가장 취약한 영역으로 확인된다. 국제 기준과 비교하면 방향성은 긍정적이지만, 개방 단계 중심 접근에서 벗어나 데이터 생성-가공-개방 전 과정을 포괄하는 다층적 AIRD 체계를 마련해야 한다.

〈표 4〉 AIRD 5요소 기준 한국 정책 종합 평가

요소	현재 수준	주요 한계	개선 방향
품질·신뢰성	중상	TL0 직접 수정의 제약, 시계열·대표성·편향 검증 부재	TL2 기반 자동검증·표준규칙 적용 강화 및 결과의 TL1·TL0 순환 반영
표준화·상호운용성	중하	TL1의 비공식 가공 및 스키마 단절, 참조코드 불일치	TL2 시맨틱 변환·코드정합성 적용, 다층 상호운용성 체계 도입
거버넌스·책임성	중하	TL1의 계보 소실, 변경이력 미흡	TL2→TL1 기반 계보 자동기록 시스템 제도화
다양성·포용성	하	대표성·편향성 점검 절차 부재	TL2에서의 편향측정·대표성 검증 강화 및 공개 의무화
지속적 운영 가능성	중하	갱신·버전관리 체계의 기관별 편차	TL2 중심 자동화된 모니터링·버전관리·정합성 검사 체계 도입

VI. 한국의 'AI Ready 공공데이터' 정책 개선 방향

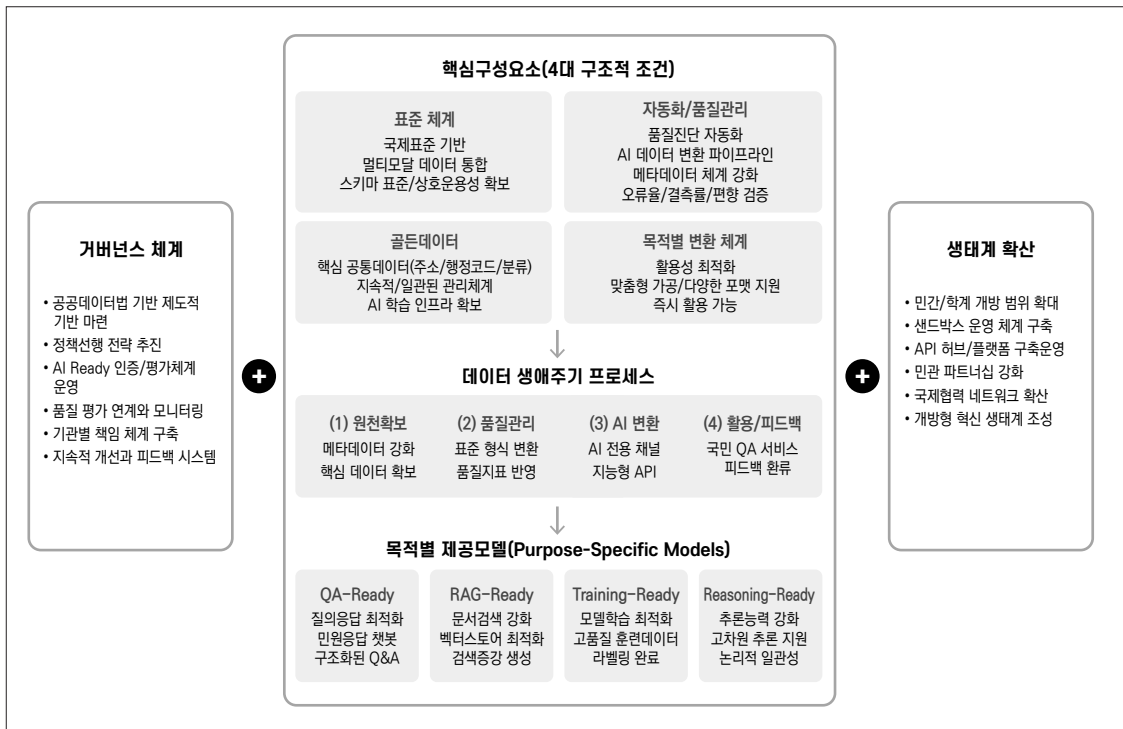
1. 기본 원칙과 전략적 전환 방향

공공데이터 체계가 AI 환경에서 즉시 활용하는 형태로 발전하기 위해 데이터 생애주기 전체를 포괄하는 구조적 접근이 요구된다. AIRD 5개 요소는 개별 항목이 아니라 생애주기 전반을 관통하는 구조적 속성으로 이해할 필요가 있다. 데이터 구조·정의·문서화·계보·편향·갱신주기 등 모든 요소를 하나의 통합 구조 안에서 관리해야 AI가 데이터를 일관된 방식으로 해석·활용할 수 있다. 또한 목적 적합성에 기반한 제공 체계로의 전환이 필요하다. 질의응답(QA-Ready), 검색증강생성(RAG-Ready), 기계학습(Training-Ready), 추론(Reasoning-Ready) 등 활용 목적별로 요구되는 품질

과 문서화 수준이 상이하므로, 단일 개방 기준만으로는 AI 활용 수요를 충족하기 어렵다.

이러한 기본 원칙에 따라, 'AI Ready 공공데이터' 구현을 위한 구조적 요구사항을 종합한 시각적 모델이 <그림 3>에 제시되어 있다. <그림 3>은 공공데이터를 AI 활용에 적합한 형태로 전환하기 위해 필요한 핵심 구성요소, 생애주기 기반 처리 프로세스, 목적별 제공체계를 통합적으로 제시한 구조이다.

'AI Ready 공공데이터' 프레임워크는 'AI Ready 공공데이터' 구축을 위해 필요한 네 가지 구조적 조건(표준체계, 자동품질·정합성 관리, 국가 참조데이터기반 통합, 목적별 변환)을 중심으로, 데이터 생애주기 전체에서 AI Ready 기준을 내재화하는 프로세스(데이터 확보-품질관리-AI 변환-활용)를 제시한다. 목적별 제공모델(QA-Ready, RAG-Ready, Training-Ready, Reasoning-Ready)은 AI 활용 맥락에서 요구되는 데



<그림 3> 'AI Ready 공공데이터' 프레임워크

이터 구조·품질·문서화 기준의 차이를 반영하는 구성 요소다. 거버넌스 체계, 생태계 확산은 AI 전환과 활용을 위한 핵심 요소다.

2. 구조적 개선 과제와 프레임워크 적용 방향

AI Ready 공공데이터로의 전환은 개별 단계의 단편적 개선이 아닌, 원천데이터의 신뢰성 환류, 가공 단계의 투명성 확보, 개방 단계의 실행력 강화를 유기적으로 결합하여 데이터 생애주기 전반의 품질을 상향 평준화하는 과정이다.

전환계층 기반 프레임워크에서 TL0·TL1·TL2는 이러한 순환 구조를 기능적으로 분담한다. TL0은 상위 계층에서 도출된 검증 결과를 수용하는 구조적 기반 마련에 집중한다. 개방 단계에서 식별된 오류 패턴과 표준 규칙 위반 사항을 원천데이터 관리자에게 정기적으로 공유하고, 참조코드 매핑 테이블과 표준 어휘를 원천 단계에서 직접 참조할 수 있는 체계를 구축한다. 시계열 불연속성이나 값 범위 이상 등 자동 탐지 결과는 데이터 생성 절차 개선의 근거로 활용하며, 환류 결과 반영 수준은 기관 평가 지표에 포함한다. TL1은 기관 내부 가공 계층으로, 전환계층 간 연결을 매개하는 기능을 한다. 가공 절차의 투명성 확보와 데이터 계보 정보 축적이 핵심 과제이다. 모든 가공 공정은 메타데이터로 기록·버전 관리되어야 하며, 원천데이터에서 가공데이터로의 변환 규칙을 명시적으로 정의한다. 참조코드와 식별자 매핑의 자동 검증을 실시하고, 개방 단계의 품질 개선 요구사항을 가공 절차에 즉각 반영한다. 마지막으로, TL2는 품질개선과 표준 적용이 실제 집행되는 실행 계층이다. 개방 시점에서의 표준 포맷, 참조코드, 메타데이터 스키마 정합성 자동 검증을 의무화하고, 그 결과를 하위 계층으로 환류한다. 시맨틱 웹 표준 적용을 통해 기계 판독성과 상호운용성을 확보하며, 모든 데이터셋에 대한 버전 관리 및 계보 정보를 자동 제공한다.

이러한 전환계층 구조를 기반으로 AI 활용 목적에 따른 제공모델은 차별화된 형태로 구성된다. QA-Ready

는 메타데이터와 요약 정보를 강화하여 질의응답 처리 효율성을 제고한다. RAG-Ready는 청크 단위 분할과 임베딩 최적화를 통해 검색증강생성을 지원한다. Training-Ready는 레이블, 품질 지표, 편향 정보를 포함하여 기계학습 모델 학습에 활용된다. Reasoning-Ready는 온톨로지, 규칙, 관계 정보를 제공하여 추론 기반 AI 활용을 가능하게 한다.

AI Ready 공공데이터의 실질적 구현을 위해 국가 차원의 통합 메타데이터 스키마, 표준 참조어휘, 영구식별자(PID) 체계 구축이 선행되어야 한다. 국제 표준과의 정합성을 유지하면서도 국내 공공데이터의 특수성을 반영한 확장 프로파일 개발이 요구되며, 규칙 기반 검증, 이상치 탐지, 참조데이터 정합성 검사, 시계열 연속성 모니터링 등 자동화된 품질·정합성 관리 체계가 필요하다. 이러한 검증 결과는 관리 도구를 통해 실시간 공유되어 상위 계층의 구조적 개선으로 연결된다. 이와 함께 주소, 행정구역, 기관코드 등 핵심 참조데이터의 국가 참조데이터 통합 관리가 수행되어야 한다. 모든 공공데이터가 동일한 참조코드 체계를 공유하며, 시맨틱 레이어 구축을 통해 데이터 간 의미적 연결성을 확보하고, 도메인 온톨로지를 활용하여 데이터셋 간 관계성을 명시적으로 정의하는 것이 필수적이다.

국제 기준은 한국형 AIRD 체계 구축의 핵심 준거점으로 기능한다. 특히 데이터 계보 관리, 위험 기반 문서화, 메타데이터 중심 관리 원칙은 전환계층 기반의 순환 구조를 제도적으로 뒷받침하는 필수 요소로서, 개방 단계의 품질개선과 기관 내부 가공 단계의 계보 관리를 정합적으로 연계하고 정책 신뢰성을 확보하는 중추적 역할을 수행한다. 이러한 기준을 실질적으로 구현하기 위해서는 제도, 조직, 기술적 기반의 병행 구축이 필요하다. 제도적 측면에서는 AI Ready 공공데이터 가이드라인의 법제화와 책임 체계 명확화가 선행되어야 하며, 조직적으로는 국가 참조데이터 전담 관리 기능을 강화하고 부처별 역할을 재정립해야 한다. 기술적으로는 자동 품질관리 체계와 데이터 계보 추적 엔진 구축을 통해 관리의 효율성과 정책 실행력을 제고해야 한다.

결론적으로, AI Ready 공공데이터 구축의 본질은 원데이터의 전면적 개편에 머무는 것이 아니라, 개방 단계 중심의 품질개선과 기관 내부 가공 단계의 구조적 조정, 그리고 원천데이터 단계로의 점진적 환류가 결합된 순환적 관리체계의 확립에 있다. 본 연구에서 제시한 정책 개선 방향은 공공데이터를 AI 기반 분석과 추론, 지능형 의사결정에 최적화된 국가 데이터 인프라로 고도화하기 위한 전략적 경로를 제시한다는 점에서 의의를 갖는다.

VII. 결론

AI 시대에 데이터는 단순한 정보 자원을 넘어 기계 학습과 추론, 의사결정이 가능한 구조적 조건을 갖춘 ‘AI Ready’ 상태로 제공되어야 한다. 본 연구는 AI Ready Data 개념을 학술적으로 정립하고, 한국의 ‘AI Ready 공공데이터’ 정책을 체계적으로 평가하기 위한 분석틀을 제시하였다. 65편의 학술논문, 정책문서, 국제기구 보고서, 산업계 자료를 분석한 결과, AIRD는 품질·신뢰성, 표준화·상호운용성, 거버넌스·책임성, 다양성·포용성, 지속적 운영 가능성의 다섯 가지 요소로 구성된 복합 개념임을 확인하였다. 산업계·학계·국제기구·정부의 접근 방식은 차이가 있으나, 이 다섯 요소는 공통적으로 요구되는 구조적 기반으로 나타났다.

한국 공공데이터 관리체계를 TL0(원천데이터)→TL1(기관 내부 가공데이터)→TL2(개방데이터)의 전환계층 모델로 분석한 결과, 기존 체계는 TL0와 TL1에서 발생한 구조적 편차가 TL2에서 단순히 전달되는 단방향 구조를 보여 왔다. 정부가 2025년 제시한 ‘AI Ready 공공데이터’ 정책은 생애주기 전반과 목적별 제공모형을 포함한다는 점에서 긍정적 전환으로 평가된다. 그러나, TL2의 품질·정합성·표준화 기능이 충분히 제도화되어 있지 않아 TL1·TL0로 환류되는 순환적 품질관리 체계는 아직 초기 단계에 머물러 있다. AIRD 5개 요소 기준으로 평가하면 품질 관리 수준은 비교적 높은 편이지만, 표준화·상호운용성·거버넌스·포용성·지속

성은 국제 기준(NIST, EU AI Act, World Bank, UK NDL)에 비해 미성숙한 수준으로 파악되었다.

본 연구의 학술적 기여는 AIRD 개념을 국내 최초로 체계화하고, 전환계층 모델을 통해 한국 공공데이터 체계의 구조적 단절 지점을 가시화한 데 있다. 정책적 측면에서는 TL2 중심의 품질 개선과 TL1 단계의 계보(lineage) 관리 강화, 그리고 TL0의 구조 개선으로 이어지는 순환적 우선순위를 도출하였다는 점에 의의가 있다.

그럼에도 본 연구는 AIRD 요소에 대한 정량적 측정 지표 결여, 정책 시행 초기에 따른 성과 추적의 한계, 다양성·포용성 요소에 대한 실증 분석 부족이라는 한계를 지닌다. 향후 연구에서는 AIRD 수준 평가를 위한 조작적 정의와 정량 지표 개발, 그리고 데이터 생애주기 전반에 걸친 종단적 연구가 수행되어야 한다.

인공지능 기술의 확산은 데이터 패러다임의 전환을 요구한다. 양적 개방을 넘어, 데이터를 AI 적합 구조로 전환하는 다층적 AIRD 체계의 확립과 이를 뒷받침하는 순환적 관리 체계의 제도화가 한국 공공데이터 정책의 핵심 과제가 될 것이다.

■ 참고문헌

- 김선호·이창수·정승호·김학철 (2015). 공공데이터 품질관리를 위한 조직 성숙도 평가 모델. <정보화정책>, 22권 1호, 28-46.
- 김용환·이진희·윤영진·이상일·김성진 (2023). 공공데이터 품질 수준 분석을 이용한 데이터품질관리 연구. <한국정보통신학회논문지>, 27권 12호, 1465-1472. <https://doi.org/10.6109/jkiice.2023.27.12.1465>
- 김학래 (2020). 공공데이터 개방표준 데이터의 품질평가. <한국콘텐츠학회논문지>, 20권 9호, 439-447.
- 김학래 (2021). FAIR 원칙: 데이터 관점의 디지털 아카이브 구현을 위한 고려사항. <한국기록관리학회지>, 21권 2호, 155-172. <https://doi.org/10.14404/JKSARM.2021.21.2.155>
- 김현지·김학래 (2022). 행정표준코드의 기관코드를 활용한 공공데이터의 품질 개선 방법 제안. <디지털콘텐츠학

- 회논문지>, 23권 3호, 481-488. <https://doi.org/10.9728/dcs.2022.23.3.481>
- 과학기술정보통신부 (2025). 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법. 법제처국가법령정보센터. [https://www.law.go.kr/%EB%B2%95%EB%A0%B9/%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5%20%EB%B0%9C%EC%A0%84%EA%B3%BC%20%EC%8B%A0%EB%A2%B0%20%EA%B8%B0%EB%B0%98%20%EC%A1%B0%EC%84%B1%20%EB%93%B1%EC%97%90%20%EA%B4%80%ED%95%9C%20%EA%B8%B0%EB%B3%B8%EB%B2%95/\(20676,20250121\)](https://www.law.go.kr/%EB%B2%95%EB%A0%B9/%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5%20%EB%B0%9C%EC%A0%84%EA%B3%BC%20%EC%8B%A0%EB%A2%B0%20%EA%B8%B0%EB%B0%98%20%EC%A1%B0%EC%84%B1%20%EB%93%B1%EC%97%90%20%EA%B4%80%ED%95%9C%20%EA%B8%B0%EB%B3%B8%EB%B2%95/(20676,20250121))
- 박하람·강지우·송채은·안지은·이정운·장소정·조하연·차화준·김학래 (2024). FAIR 데이터 원칙을 적용한 국내 디지털 아카이브 평가와 개선 방향. <한국기록관리학회지>, 24권 4호, 113-134. <https://doi.org/10.14404/JKSARM.2024.24.4.113>
- 박하람·김학래 (2022). DCAT-AP-KR: 국내 데이터 포털의 상호운용을 위한 애플리케이션 프로파일. <디지털콘텐츠학회논문지>, 23권 11호, 2249-2258. <https://doi.org/10.9728/dcs.2022.23.11.2249>
- 박하람·김학래 (2024). 공공데이터를 활용한 주소 기반 데이터 분석과 제언. <디지털콘텐츠학회논문지>, 25권 9호, 2617-2626. <https://doi.org/10.9728/dcs.2024.25.9.2617>
- 성혜정·정예진·임시영 (2024). 데이터 개방에서 활용으로, 공간정보가치 증대를 위한 FAIR 원칙 도입방안. <국토정책 Brief>, 990. 국토연구원.
- 송채은·김학래 (2022). 지방자치단체의 공공데이터 개방 현황 분석과 개선방안. <디지털콘텐츠학회논문지>, 23권 4호, 697-705. <https://doi.org/10.9728/dcs.2022.23.4.697>
- 송채은·김학래 (2025). 데이터 포털의 상호운용성을 위한 데이터템 검색 프레임워크 개발과 검증. <디지털콘텐츠학회논문지>, 26권 4호, 921-930. <https://doi.org/10.9728/dcs.2025.26.4.921>
- 윤건 (2023). 데이터 거버넌스 실증연구: 구성요소 간 구조적 관계와 영향을 중심으로. <정보화정책>, 30권 3호, 29-48.
- 정하나·김재웅·정영석 (2022). 공공데이터 품질관리 수준 평가 도메인 규칙 매핑 모델 제안. <한국컴퓨터정보학회논문지>, 27권 12호, 189-195.
- 정기흥·김훈 (2025). 표준 공공데이터의 품질 수준 평가: 공공데이터포털의 공공주차장 데이터 사례 연구. <전자공학회논문지>, 62권 7호, 37-43. <https://doi.org/10.5573/ieie.2025.62.7.37>
- 차오신·오세환 (2024). 데이터 플랫폼 활용 영향요인에 대한 탐색적 연구: 한국 공공데이터 포털 사례를 중심으로. <정보화정책>, 31권 4호, 87-103.
- 행정안전부 (2013). 공공데이터의 제공 및 이용 활성화에 관한 법률. 법제처 국가법령정보센터. <https://www.law.go.kr/법령/공공데이터의제공및이용활성화에관한법률>
- 행정안전부 (2020). 공공데이터 관리지침. 법제처 국가법령정보센터. <https://www.law.go.kr/행정규칙/공공데이터관리지침>
- 행정안전부 (2025). AI가 활용하기 좋은 공공데이터 만든다! 'AI-Ready' 기준 최초 도입. https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_0000000000008&nttlId=120651
- 행정안전부 (2026). 공공데이터의 인공지능 친화적 관리 가이드라인 배포. https://www.data.go.kr/bbs/rcr/selectRecsroom.do?pageIndex=1&originId=PDS_0000000001260+++&atchFileId=FILE_000000003631663&searchCondition3&searchCondition2&cndCtgrYLaword=Y&cndCtgrYEdc=Y&cndCtgrYBigdata=Y&cndCtgrYStd=Y&cndCtgrYNews=Y&cndCtgrYContest&cndCtgrYEtc=Y&cndCtgrYCardNews&bindCndCtgrY&sort-post=2&searchKeyword1&Laword=PDY01&Edc=PDY02&Bigdata=PDY03&Std=PDY04&News=PDY05&Etc=PDY06&fbclid=IwY2xjawRnq3VleHRuA2FlbQIxMQBzcnRjBmFwcf9pZBAyMjIwMzkxNzg4MjAwODkyAAEeVrB5ygBd0KvrZzfCBiMbK7Dgq6deivZquPcr9Mt3gm6wASS3xongswCQZCK_aem_0IxrABDdO5G-oQ4EAgaSsQ
- 황성수 (2025). AI 활용 정부 서비스의 사례와 유형 분석: 새 정부 AI 정부 서비스 성공을 위한 전략 설정 제언. <정보화정책>, 32권 3호, 3-20.
- European Commission (2025). European approach to artificial intelligence. <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of

- 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). *Official Journal of the European Union*. <http://data.europa.eu/eli/reg/2024/1689/oj>
- EY (2024). Knowledge assets enable AI to make informed decisions. Ernst & Young Global. https://www.ey.com/en_us/services/emerging-technologies/knowledge-assets-enable-ai-to-make-informed-decisions
- G20 (2019, June 9). G20 ministerial statement on trade and the digital economy. Ministry of Foreign Affairs of Japan. https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf
- G7 (2023, October 30). G7 leaders' statement on the Hiroshima AI process. Ministry of Foreign Affairs of Japan. https://www.mofa.go.jp/ecm/ec/page5e_000076.html
- Gartner (2023). A journey guide to delivering AI success through 'AI-ready' data. <https://www.gartner.com/en/documents/5845247>
- Gartner (2024a). AI-ready data essentials to capture AI value. Gartner. <https://www.gartner.com/en/articles/ai-ready-data>
- Gartner (2024b). Follow these five steps to make sure your data is AI-ready. Gartner. <https://www.gartner.com/en/documents/5848047>
- Gartner (2024c). Your organization is ready for AI, but is your data?. Gartner. <https://www.gartner.com/en/webinar/577575/1296498>
- Gartner (2025). Lack of AI-ready data puts AI projects at risk. Gartner. <https://www.gartner.com/en/newsroom/press-releases/2025-02-26-lack-of-ai-ready-data-puts-ai-projects-at-risk>
- Hill, S. (2024, March 5). AI-ready FAIR data: Accelerating science through responsible AI and data stewardship. Medium. https://medium.com/@sean_hill/ai-ready-fair-data-accelerating-science-through-responsible-ai-and-data-stewardship-3b4f21c804fd
- Hiniduma, K., Byna, S. & Bez, J. L. (2025). Data readiness for AI: A 360-degree survey. *ACM Computing Surveys*, 57(9), 1-39. <https://doi.org/10.1145/3722214>
- Huerta, E. A., Blaiszik, B., Brinson, L. C., Bouchard, K. E., Diaz, D., Doglioni, C., ... & Zhu, R. (2023). FAIR for AI: An interdisciplinary and international community building perspective. *Scientific data*, 10(1), 487. <https://doi.org/10.1038/s41597-023-02298-6>
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, 29(4), 258-268.
- Jati, P. H. P., Lin, Y., Nodehi, S., Cahyono, D. B. & van Reisen, M. (2022). FAIR versus open data: A comparison of objectives and principles. *Data Intelligence*, 4(4), 867-881. https://doi.org/10.1162/dint_a_00176
- Kidwai-Khan, F., Wang, R., Skanderson, M., Brandt, C. A., Fodeh, S. & Womack, J. A. (2024). A roadmap to artificial intelligence (AI): Methods for designing and building AI-ready data to promote fairness. *Journal of Biomedical Informatics*, 154, 104654. <https://doi.org/10.1016/j.jbi.2024.104654>
- King's College London (2023, July 18). King's scientists' plans for AI-ready National Data Library could turbocharge economy. <https://www.kcl.ac.uk/news/kings-scientists-plans-for-ai-ready-national-data-library-could-turbocharge-economy>
- Lehmann, J., Zaveri, A., Rula, A., Maurino, A., Pietrobon, R. & Auer, S. (2016). Quality assessment for Linked Data: A Survey. *Semantic Web* (1570-0844), 7(1).
- Majithia, N., & Simperl, E. (2024). Methods to assess the UK government's current role as a data provider for AI. arXiv. <https://doi.org/10.48550/arXiv.2412.09632>
- McKinsey & Company (2022). *Accelerating AI impact*

- by taming the data beast. McKinsey Global Institute. <https://www.mckinsey.com/~media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/Accelerating%20AI%20impact%20by%20taming%20the%20data%20beast/Accelerating-AI-impact-by-taming-the-data-beast.pdf>
- McKinsey & Company (2023). *The state of AI in 2023: Generative AI's breakout year*. McKinsey Global Institute. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year>
- Meroño-Peñuela, A., Massey, J., Newman, A. & Simperl, E. (2025). How an AI-ready National Data Library would help UK science. arXiv. <https://doi.org/10.48550/arXiv.2501.17013>
- NIST (2023). AI Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- Oderinwale, H. & Longpre, S. (2024). Future and AI-ready data strategies: Response to DOC RFI on AI and open government data assets. arXiv. <https://doi.org/10.48550/arXiv.2408.01457>
- OECD (2019). Recommendation of the Council on Artificial Intelligence. Organisation for Economic Co-operation and Development. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- OECD (2023). 2023 OECD Open, Useful and Re-usable data (OURdata) Index: Results and key findings, OECD Public Governance Policy Papers, No. 43, OECD Publishing, Paris. <https://doi.org/10.1787/a37f51c3-en>
- Open Data Institute (2024a). A framework for AI-ready data. <https://theodi.org/insights/reports/a-framework-for-ai-ready-data/>
- Open Data Institute (2024b). The UK government as a data provider for AI. https://theodi.cdn.ngo/media/documents/UK_government_as_a_data_provider_for_AI.pdf
- Piontek, D. (2024, May 8). How AI-ready data can help drive AI strategy success. EY. https://www.ey.com/en_us/services/emerging-technologies/knowledge-assets-enable-ai-to-make-informed-decisions
- Ravi, N., Chaturvedi, P., Huerta, E. A., Liu, Z., Chard, R., Scourtas, A., ... & Foster, I. (2022). FAIR principles for AI models with a practical application for accelerated high energy diffraction microscopy. *Scientific Data*, 9(1), 657. <https://doi.org/10.1038/s41597-022-01712-9>
- Raza, S., Ghuge, S., Ding, C., Dolatabadi, E. & Pandya, D. (2024). FAIR enough: Develop and assess a FAIR-compliant dataset for large language model training?. *Data Intelligence*, 6(2), 559-585.
- Sadiq, S. & Indulska, M. (2017). Open data: Quality over quantity. *International journal of information management*, 37(3), 150-154.
- Sallam, R. (2024, June 14). *What is AI-ready data? And how to get yours there*. Gartner Research. <https://www.gartner.com/en/articles/ai-ready-data>
- Senscience (2025, February 3). From open to reusable: How FAIR² data portals multiply scientific impact. Substack. <https://senscience.substack.com/p/from-open-to-reusable-how-fair-data>
- Stebbins, R. A. (2001). Exploratory research in the social sciences. SAGE Publications. <https://doi.org/10.4135/9781412984249>
- U.S. Department of Commerce (2024, April 17). Request for information: AI-ready open government data assets. U.S. Department of Commerce Blogs. <https://www.commerce.gov/news/blog/2024/04/request-information-ai-ready-open-government-data-assets>
- UK Government (2024). National Data Library. GOV.UK. <https://www.gov.uk/government/publications/national-data-library/national-data-library>
- UK Government (2025). AI opportunities action plan. GOV.UK. <https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai>

opportunities-action-plan

- Verhulst, S., Zahuranec, A. J. & Chafetz, H. (2025). Moving toward the FAIR-R principles: Advancing AI-ready data. SSRN. <https://ssrn.com/abstract=4956182>
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R. & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- World Bank (2025, January 15). From open data to AI-ready data: Building the foundations for responsible AI in development. World Bank Blogs. <https://blogs.worldbank.org/en/opendata/from-open-data-to-ai-ready-data--building-the-foundations-for-re>
- Zhang, H., & Xiao, J. (2020). Quality assessment framework for open government data: Meta-synthesis of qualitative research, 2009-2019. *The Electronic Library*, 38(2), 209-222.
- Zuiderwijk, A. & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government information quarterly*, 31(1), 17-29.