

FAIR 원칙에 따른 한국 고문헌 메타데이터 관리체계 진단: KORCIS를 중심으로

Assessing the Metadata Management System for Korean Classical Documents Based on FAIR Principles: A Case Study of KORCIS

신은선(EunSun Shin)

E-mail: librarian1027@gmail.com

한국학중앙연구원 인문정보학 박사과정생



논문접수 2026-04-14
최초심사 2026-04-23
게재확정 2026-05-12

ORCID

EunSun Shin
<https://orcid.org/0009-0004-9329-2678>

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

초 록

본 연구는 한국고문헌종합목록시스템(KORCIS) 서지 데이터 567,212건을 대상으로 FAIR 원칙 준수 현황을 실증적으로 진단하고 개선 방향을 제시하는 것을 목적으로 한다. 본 연구에서 메타데이터 관리체계란 고문헌 서지 데이터의 생산·기술·식별·접근·연계에 이르는 전 과정을 규율하는 표준, 시스템 구조, 운영 정책의 총체를 의미하며, KORCIS는 142개 참여 기관을 포괄하는 국가 단위 고문헌 메타데이터 관리체제로 기능하고 있다. 분석은 KORMARC 주요 필드 출현 빈도 및 완성도 측정, 전거 제어 및 식별자 체계 실태 분석, FAIR 원칙 15개 세부 항목별 준수도 평가의 세 단계로 수행하였으며, 필드 유형(필수·해당시 필수·재량) 구분과 데이터 구조적 특성·시스템 기능 지원 여부·정량 분석 결과를 종합하여 판정하였다. 분석 결과 필드 필드의 완성도는 높으나 국제 영구 식별자·전거 연계·표준 어휘 등 시맨틱 웹 환경의 핵심 요건이 전반적으로 미충족 상태임이 확인되었으며, FAIR 15개 항목 평가 결과 적합 0개(0%), 부분 적합 5개(33.3%), 미흡·결여 10개(66.7%)로 나타났다. Europeana, 일본 NDL, 중국 중화고적자원고와의 비교를 통해 RDF 기반 온톨로지 전환, 국제 전거 연계, 라이선스 체계 정비가 시급한 과제임을 논증하였으며, 분석 스크립트와 FAIR 항목별 판정 기준은 공개 리포지터리 (<https://github.com/SHINEUNSUN1027/korcis-fair-analysis>)에 공개하였다.

ABSTRACT

This study aims to empirically assess the current state of compliance with the Findable, Accessible, Interoperable, Reusable (FAIR) principles in Korean classical document metadata and propose directions for improvement, targeting 567,212 bibliographic records from the Korean Old and Rare Collection Information System (KORCIS). In this study, the metadata management system refers to the comprehensive framework of standards, system architecture, and operational policies governing the entire process of producing, describing, identifying, accessing, and linking bibliographic data for classical documents, while KORCIS serves as a national-level metadata management infrastructure encompassing 142 participating institutions. The analysis proceeded in three stages: measurement of Korean Machine Readable Cataloging (KORMARC) field occurrence rates and completeness, examination of authority control and identifier systems, and evaluation of compliance with the 15 FAIR subcriteria—with ratings determined by integrating data structural characteristics, system-level functional support, and quantitative results, while accounting for KORMARC field types (mandatory, mandatory if applicable, and optional). The findings confirm that while mandatory fields exhibit high completion rates, core semantic web requirements, including persistent identifiers, international authority linkage, and controlled vocabulary usage, are largely unmet, with FAIR evaluation yielding 0 compliant items (0%), 5 partially compliant (33.3%), and 10 noncompliant (66.7%). Comparative analysis with Europeana, Japan's National Diet Library, and China's Classical Literature Resource Repository identifies RDF-based ontology conversion, international authority linkage, and license framework development as urgent priorities. All analysis scripts and FAIR evaluation criteria are publicly available at: <https://github.com/SHINEUNSUN1027/korcis-fair-analysis>.

Keywords: 한국 고문헌, KORCIS, FAIR 원칙, 메타데이터 품질, 연결 데이터, KORMARC, 전거 제어, Korean classical documents, KORCIS, FAIR principles, metadata quality, linked data, KORMARC, authority control

<https://jksarm.koar.kr>

1. 서론

1.1 연구 배경 및 목적

디지털 기술의 발전과 함께 문화유산의 디지털 자산화 및 온라인 서비스가 확대되고 있다. 디지털 인문학(Digital Humanities)은 디지털로 표현하고 소통하는 현대 사회에서 인문 지식이 보다 의미 있게 탐구되고 가치 있게 활용되도록 하려는 노력이며, 연구 방법의 혁신을 통해 인문학 본연의 학술 연구에 기여하는 것을 일차적인 과제로 삼는다(김현, 2016). 이와 맥락을 같이하여 문헌정보학(Library and Information Science)은 인문 지식 자산이 체계적으로 조직되고 안정적으로 유통될 수 있도록 메타데이터 표준화와 지식조직(Knowledge Organization) 체계를 구축하는 방법론적 기반을 제공한다(강순애, 2014). 특히 고문헌과 같은 역사적 자산의 디지털화 과정에서 문헌정보학적 서지 제어와 데이터 품질 관리는 디지털 인문학 연구의 신뢰성과 지속가능성을 확보하는 전제 조건이다.

한국의 고문헌은 유교적 통치 철학, 역사적 기록, 사상적 정수가 집약된 지식 자산으로서 국내외 연구의 대상이 되고 있다. 그러나 현행 고문헌 메타데이터는 개별 소장 기관별로 분산 관리되고 있으며, 데이터 표준과 기술 형식의 불일치로 인해 통합적 검색 및 고도화된 데이터 활용에 한계가 존재한다(임소정, 손계영, 2023). 이러한 데이터 분절 현상은 학술 자원의 공유와 확산을 저해하는 요인으로 지적된다.

이러한 상황을 배경으로 전 세계적으로 FAIR 원칙(Findable, Accessible, Interoperable, Reusable)이 연구 데이터 및 문화 자산 관리의 핵심 지침으로 확산되고 있다(Wilkinson et al., 2016). FAIR 원칙은 데이터의 발견 가능성, 접근 가능성, 상호운용성, 재사용 가능성을 지향하며, 특히 인간뿐만 아니라 컴퓨터가 데이터를 자동으로 처리·연계할 수 있는 기계 가독성(machine-actionability)을 강조한다(Wilkinson et al., 2016). 이는 지식의 표준적 기술 체계를 통해 개별 연구 성과가 공동의 지식 인프라로 결집되도록 하는 디지털 인문학의 지향점인 동시에, 문헌정보학이 추구하는 정보의 전 생애주기 관리 및 공유 가치와도 부합한다.

국내 고문헌 메타데이터는 한국고문헌종합목록시스템(KORCIS)을 중심으로 통합 관리되고 있다. KORCIS는 2007년 서비스를 개시한 이래 지속적으로 확충되어, 2025년 기준 국내 91개 및 국외 53개를 포함한 총 144개 참여 기관으로부터 약 51만 건의 서지 데이터와 6만 5천여 건의 원문 이미지를 제공하는 국가 단위의 서지 인프라로 기능하고 있다(국립중앙도서관, n.d.b). 그러나 KORCIS는 전통적인 도서관 서지 표준인 KORMARC에 기반하고 있어, 시맨틱 웹 환경에서 요구되는 RDF(Resource Description Framework) 형식의 연결 데이터(Linked Data)로의 전환이 충분히 이루어지지 않은 상태이다. 또한 외부 전거 파일(Authority File)과의 연계 부족으로 인해 글로벌 지식 네트워크와의 상호운용성 확보에 어려움이 있다(이미화, 배정현, 2025).

이에 본 연구는 한국 고문헌 메타데이터의 현황을 FAIR 원칙에 의거하여 체계적으로 진단하고, 국제적 호환성 확보와 데이터 활용성 강화를 위한 방향을 제시하고자 하며, 구체적인 연구 목적은 다음과 같다.

첫째, KORCIS의 실증 데이터를 FAIR 원칙에 의거하여 체계적으로 진단함으로써 메타데이터 품질 현황과 문제점을 정량적·정성적으로 도출한다. 둘째, Europeana, 일본 국립국회도서관(NDL), 중화고적자원고(中華古籍資源庫) 등 문화유산 메타데이터 LOD 구현의 성격이 상이한 세 유형의 해외 기관을 대상으로 메타데이터 표준화 및 LOD 구현 사례를 비교 분석하여 국내 표준화 방향에 대한 시사점을 도출한다. 셋째, 진단 결과를 바탕으로 RDF 기반 시맨틱 표준화의 필요성과 방향을 제시하며, 구체적인 온톨로지 설계 및 변환 가이드라인은 후속 연구에서 다룬다.

1.2 선행연구

선행연구는 크게 세 영역으로 구분하여 검토한다. FAIR 원칙의 문헌정보학적 적용 연구, 고문헌 디지털화 및

메타데이터 표준화 연구, 그리고 시맨틱 웹·LOD 기반 서지 표준화 연구가 그것이다.

FAIR 원칙은 Wilkinson et al.(2016)이 과학 데이터 관리를 위한 지침으로 제안한 이후, 문헌정보학 및 기록관리학 분야로 적용 범위가 확산되고 있다. 국내에서는 박하람, 김학래(2023)가 1997년 외환위기 아카이브를 대상으로 DCAT 기반 지식그래프를 구축하고 RDA FAIR 성숙도 모델로 평가한 결과, 표준 어휘를 활용한 지식그래프가 기존 아카이브보다 상호운용성과 재사용성 측면에서 향상된 결과를 보임을 확인하였다. 박하람 외(2024)는 국내 디지털 아카이브 69개를 ARDC FAIR Self Assessment Tool로 평가하여 상호운용성과 재사용성이 전반적으로 낮은 수준임을 실증하였으며, 표준 어휘 미사용과 라이선스 정보 부재를 주요 원인으로 지목하였다. 이 두 연구는 국내 디지털 아카이브 전반의 FAIR 준수 실태를 진단하였다는 점에서 의의가 있으나, 고문헌 메타데이터에 특화된 분석은 다루지 않았다.

고문헌 디지털화 분야에서는 임소정, 손계영(2023)이 국내 대학도서관 고문헌 디지털화 현황을 분석하여 기관 간 표준 불일치와 메타데이터 품질 격차를 지적하였다. 강순애(2014)는 문서 목록규칙의 표준화 방안을 제시하며 KORMARC 기반 고문헌 서지 기술의 체계화 방향을 논의하였다. 허철(2024)은 한국 고문헌 디지털 데이터의 연구 발전을 위한 과제를 포괄적으로 검토하며 데이터 품질 관리와 국제 표준 정렬의 필요성을 제기하였다. 박민정, 이승민(2023)은 고문헌 기술을 위한 LRM 기반 서지구조 구축을 시도하여 에이전트, 장소, 시간 개체를 중심으로 한 관계 표현 방안을 제안하였다.

KORCIS의 현황과 문제점을 직접 다룬 연구로는 김영진(2024)이 있다. 이 연구는 해방 이후 한국 고문헌 목록화의 전반적 흐름을 검토하고, KORCIS의 구조적 문제점을 실증적으로 지적하였다. 구체적으로 한자 검색을 저조, 동일서명 소장처 통합 검색 기능 부재, 미참여기관 다수 존재, 참여기관의 데이터 부분 제공 등을 개선 과제로 제시하였다. 그러나 이 연구는 서지학적·목록학적 관점에서의 현황 진단에 집중하였으며, FAIR 원칙을 평가 기준으로 적용한 정량적 분석은 수행하지 않았다.

시맨틱 웹 기반 서지 표준화와 관련하여 이미화, 배정현(2025)은 국내 주요 도서관의 BIBFRAME 전환 현황을 분석하며 한국 고문헌 서지의 시맨틱 웹 전환이 시급한 과제임을 논증하였다. 박지영(2014)은 링크드 데이터 환경에서의 서지기술형식 BIBFRAME과 그 활용을 고찰하며 국내 도서관의 LOD 전환 방향을 검토하였다. 해외에서는 Barbuti(2021)가 디지털 도서관의 보존 관점에서 FAIR 원칙의 R 영역을 중심으로 문화유산 데이터의 장기 보존 방안을 논의하였다.

이상의 선행연구를 종합하면, FAIR 원칙을 문화유산 및 디지털 아카이브 평가에 적용한 연구는 국내외에서 축적되고 있으나, KORCIS의 실증 데이터를 직접 수집하여 FAIR 원칙에 의거한 정량적 진단을 수행한 연구는 아직 이루어지지 않았다. 본 연구는 이러한 공백을 채우는 것을 목적으로 한다.

1.3 연구 방법 및 범위

본 연구는 국립중앙도서관이 공공데이터포털을 통해 공개한 KORCIS 전체 서지 데이터를 수집하여 분석하였다. 수집 데이터는 2025년 7월 31일 기준으로 공개된 파일로, 총 567,212건의 서지 레코드와 142개 참여 기관(국내 91개, 해외 51개)의 데이터를 포함한다. 이는 KORCIS가 서비스하는 전체 데이터 규모(51만여 건, 144개 기관)에 상응하는 수치로, 본 연구의 분석은 공개된 전체 데이터를 대상으로 수행한 전수 분석에 해당한다. 데이터는 레코드키, MARC(KORMARC 기반 서지데이터), 도서관번호, 도서관명의 4개 필드로 구성되어 있으며, KORCIS XML DTD 정의서에 따른 태그 및 식별기호 체계를 따른다.

분석은 다음 세 단계로 수행하였다. 첫째, MARC 태그별 출현 빈도 및 완성도 분석이다. KORCIS XML DTD 정의서에 정의된 KORMARC 필드를 기준으로 주요 태그의 출현 빈도와 기재 완성도를 전체 데이터를 대상으로 측정하였다. 본 연구에서 필드 완성도는 전체 레코드 중 해당 필드가 기재된 레코드의 비율(출현율)로 정의하며,

필드 내 데이터 요소의 내용적 정확성은 분석 범위에서 제외한다. 분석 대상은 서명저자사항(245), 간사사항(260), 형태사항(300), 기본표목-개인명(100), 부출표목-개인명(700), 일반주제명(650), 지명(651), 비통제 색인어(653), 전자적 위치 및 접속(856) 등 서지 기술 및 전자·주제 관련 핵심 필드를 포함한다. 특히 300 필드의 형태코드(b 식별기호)에 수록된 광곽(코드 02), 계선(코드 04), 판구(코드 07), 어미(코드 08) 등 고문헌 특화 물리 기술 정보의 입력 현황을 별도로 분석하였다.

둘째, 전자 제어 및 식별자 체계 분석이다. 100·700·650 등 전자 관련 필드의 0 식별기호(전자레코드 제어번호) 기재 현황을 확인하고, 외부 전자 시스템(VIAF, Wikidata, 국립중앙도서관 인명전자)과의 연계 여부를 점검하였다. 또한 레코드키의 식별자 체계와 856 필드 URI의 영구 식별자 적용 여부를 분석하였다.

셋째, FAIR 원칙 항목별 준수도 평가이다. Wilkinson et al.(2016)이 제시한 FAIR 원칙의 15개 세부 항목을 평가 기준으로 삼아 각 항목별 준수 여부를 적합, 부분 적합, 미흡·결여의 3단계로 분류하였다. 평가는 전체 데이터의 구조적 특성과 시스템 차원의 기능 지원 여부를 종합하여 수행하였다.

해외 사례 비교 분석은 Europeana, 일본 국립국회도서관(NDL), 중화고적자원고(中華古籍資源庫)를 대상으로 각 기관의 공식 문헌 및 웹 서비스를 통해 수행하였다. 비교 사례 기관은 문화유산 메타데이터 LOD 구현의 유형적 다양성을 고려하여 선정하였다. Europeana는 다기관 이질 데이터를 단일 온톨로지로 통합하는 연합 플랫폼 모델을 대표하며, NDL은 JAPAN/MARC 기반에서 BIBFRAME으로의 단계적 전환을 추진하는 단일 국가도서관 모델로서 KORMARC 기반의 KORCIS와 가장 유사한 구조적 조건을 공유한다. 중화고적자원고는 국가 주도의 고전적 전문 플랫폼으로서 한국과 유사한 한자 문화권의 고문헌 디지털화 현황을 비교하는 데 적합한 사례이다.

본 연구의 범위는 다음과 같이 한정한다. 첫째, 분석 대상은 고문헌 메타데이터의 서지 정보에 한정하며, 디지털 이미지 및 원문 데이터는 제외하였다. 둘째, FAIR 원칙 준수도 평가는 공개된 전체 KORCIS 데이터와 시스템 차원의 기능을 기준으로 수행하며, 개별 소장 기관의 내부 시스템은 분석 범위에 포함하지 않았다. 셋째, 시맨틱 표준화 방향은 RDF 기반 온톨로지 설계의 필요성과 핵심 방향을 제시하는 수준으로 다루며, 구체적인 온톨로지 설계 및 변환 가이드라인의 실증적 구현은 후속 연구에서 별도로 다룬다.

2. FAIR 원칙과 문화유산 메타데이터

2.1 FAIR 원칙의 개요 및 15개 세부 항목

FAIR 원칙은 2016년 Wilkinson et al.이 Scientific Data에 발표한 “The FAIR Guiding Principles for scientific data management and stewardship”을 통해 제시된 데이터 관리 지침이다(Wilkinson et al., 2016). FAIR는 Findable(발견 가능성), Accessible(접근 가능성), Interoperable(상호운용성), Reusable(재사용 가능성)의 약자로, 현재 유럽 연합의 Horizon Europe, 미국 NIH 등 주요 연구 기금 기관들이 준수를 권장하거나 요건으로 명시하는 국제 지침으로 확산되고 있다.

FAIR 원칙의 핵심 특징은 기계 가독성(machine-actionability)을 강조한다는 점이다. 이는 인간뿐만 아니라 컴퓨터가 데이터를 자동으로 발견하고, 접근하고, 통합하고, 재사용할 수 있어야 함을 의미하며, 그 구현 수단으로 RDF(Resource Description Framework)와 같은 시맨틱 웹 기술의 활용이 논의되고 있다(Wilkinson et al., 2016).

FAIR 원칙은 4개 영역에 걸쳐 15개의 세부 항목으로 구성된다. 발견 가능성(F) 4개 항목은 식별자 부여 및 메타데이터 등록과 관련되며, 접근 가능성(A) 4개 항목은 표준 프로토콜을 통한 검색 및 장기 접근성을 다룬다.

상호운용성(I) 3개 항목은 공식 언어·어휘·외부 참조 사용을 요구하고, 재사용 가능성(R) 4개 항목은 라이선스·출처·커뮤니티 표준 준수를 규정한다. 각 항목의 구체적 내용은 <표 1>과 같다.

<표 1> FAIR 원칙의 15개 세부 항목

영역	항목	설명
Findable (발견 가능성)	F1	전 지구적 고유 식별자 부여 (예: DOI, HTTP URI)
	F2	풍부한 메타데이터로 데이터 기술
	F3	메타데이터에 데이터 식별자 명시
	F4	검색 가능한 외부 인덱스 서비스 등록
Accessible (접근 가능성)	A1	표준 통신 프로토콜로 식별자를 통해 검색 가능 (예: HTTP, SPARQL)
	A1.1	프로토콜은 무료이고 개방적이며 보편적으로 구현 가능
	A1.2	필요한 경우 인증과 권한 부여 절차를 허용하는 프로토콜
	A2	데이터가 이용 불가능한 경우에도 메타데이터는 접근 가능
Interoperable (상호운용성)	I1	지식 표현을 위한 공식 언어 사용 (예: RDF, OWL)
	I2	FAIR 원칙 준수 어휘 사용 (예: Dublin Core, SKOS)
	I3	정규화된 외부 참조 포함 (예: VIAF, LCSH URI 연결)
Reusable (재사용 가능성)	R1	정확하고 관련성 높은 복수의 속성으로 메타데이터 기술
	R1.1	명확하고 접근 가능한 데이터 이용 라이선스와 함께 공개
	R1.2	상세한 데이터 출처 및 변동 이력 기록
	R1.3	도메인별 커뮤니티 표준 준수 (예: BIBFRAME, EDM)

이 15개 항목은 독립적으로 작동하기보다 상호 연계된 구조를 이룬다. 예를 들어 F1(고유 식별자 부여)이 충족되지 않으면 A1(프로토콜 기반 접근)과 I3(외부 참조 연결)의 구현이 구조적으로 제약되며, I1(공식 지식 표현 언어 사용)이 이행되지 않으면 R1.3(커뮤니티 표준 준수)의 실질적 효과를 기대하기 어렵다. 따라서 FAIR 원칙의 준수도 평가는 항목별 개별 진단과 함께 영역 간 연계 구조를 고려한 종합적 해석이 필요하다.

2.2 문화유산·도서관 분야 적용 동향

FAIR 원칙은 과학 데이터 분야에서 출발하였으나, 이후 도서관·박물관·아카이브(LAM: Libraries, Archives, Museums) 분야로 적용 범위가 확대되고 있다. 문화유산 데이터의 경우 기관 간 상호운용과 장기 보존이 핵심 과제로 부각되면서, FAIR 원칙이 메타데이터 표준화와 LOD 구축의 방향성을 제시하는 지침으로 활용되고 있다 (Barbuti, 2021).

국제 기관의 실무 적용 측면에서는 메타데이터에 CC0 라이선스를 적용하고 REST API와 SPARQL 엔드포인트를 통한 기계적 접근 환경을 구축하는 방향이 확산되고 있다. 유럽연합의 Horizon Europe은 연구 데이터 관리 계획(DMP)에 FAIR 준수를 의무화하고 있으며, 문화유산 분야에서도 디지털 객체에 영구 식별자를 부여하고 국제 전자 시스템과 owl:sameAs로 연결하는 실천이 FAIR 이행의 핵심 조건으로 자리잡고 있다. 구체적인 기관별 구현 사례-Europeana의 EDM 기반 LOD, 일본 국립국회도서관(NDL)의 BIBFRAME·VIAF 연계, 중국 중화고적자원고의 현황-는 4장에서 상세히 검토한다.

국내에서는 FAIR 원칙의 문화유산·기록관리 분야 적용이 최근 들어 본격화되고 있다. 박하람 외(2024)는 국내 디지털 아카이브 69개를 대상으로 FAIR 준수도를 평가한 결과 상호운용성과 재사용성 영역의 준수 수준이 전반적으로 낮음을 확인하였으며, 표준 어휘 미사용과 라이선스 정보 부재를 주요 원인으로 지목하였다. 박하람, 김학래(2023)는 RDA FAIR 성숙도 모델을 적용하여 표준 어휘 기반 지식그래프가 기존 아카이브 대비

상호운용성과 재사용성을 실질적으로 향상시킴을 실증하였다. 서지 표준화 측면에서는 이미화, 배정현(2025)이 국내 주요 도서관의 BIBFRAME 전환 현황을 분석하며 한국 고문헌 서지의 시맨틱 웹 전환이 시급한 과제를 논증하였다.

이상의 국내외 적용 동향은 KORCIS 메타데이터의 FAIR 이행 실태를 진단하는 배경적 맥락을 형성한다.

3. KORCIS 메타데이터 현황 진단

3.1 KORCIS 개요 및 구조적 특성

한국고문헌종합목록시스템(KORCIS)은 국립중앙도서관이 국내외 고문헌 소장 기관과 협력하여 구축한 국내 유일의 고문헌 종합목록 포털이다. 2005년 정보통신부로부터 국가지식정보자원으로 지정받아 규장각 등 28개 참여기관으로부터 고서 목록 37만 건을 제공받아 한국고전적종합목록시스템으로 서비스를 시작하였으며, 2022년 1월 시스템 재구축과 함께 현재의 명칭으로 개편되었다(김영진, 2024). 이후 국내외 도서관, 박물관, 교육기관, 사찰, 문중, 개인 소장 고문헌에 대한 지속적인 조사를 통해 연간 3,000건 이상의 서지 및 원문 데이터베이스를 확충하고 있다. 2025년 현재 국내 91개, 국외 53개 등 총 144개 기관의 서지 데이터베이스 51만 건과 원문 이미지 6만 5천여 건을 제공하며, 한국학을 비롯한 다양한 학문 분야의 기초 자료 조사에 활용되는 연구 인프라로 기능하고 있다(국립중앙도서관, n.d.b).

KORCIS의 메타데이터 표준은 KORMARC(Korean Machine Readable Cataloging)에 기반한다. KORMARC은 한국 도서관계의 서지 정보 교환을 위해 제정된 표준으로, MARC21을 기반으로 하되 한국 고문헌의 특수성을 반영한 필드를 추가하였다. 예를 들어 형태사항(300) 필드의 형태코드(b 식별기호)에는 광곽(코드 02), 계선(코드 04), 판구(코드 07), 어미(코드 08) 등 고문헌의 물리적 특징을 기술하는 코드가 정의되어 있으며, 판사항(250) 필드의 판종류코드(a 식별기호)를 통해 목판본, 필사본 등 판본 유형을 구분하여 기술할 수 있다(강순애, 2014). 이러한 고문헌 특화 필드의 존재는 KORMARC이 전통적인 서지학 정보를 체계적으로 수용하고 있음을 보여준다.

그러나 KORMARC 기반의 현행 KORCIS 구조는 시맨틱 웹 환경에서 요구되는 데이터 연계와 기계 가독성 측면에서 다음과 같은 구조적 특성을 지니며, 이는 이후 3장의 실증 진단에서 구체적인 수치로 확인된다.

첫째, 레코드 단위의 분절 구조로 인해 저작·인물·소장처 간의 복합적인 관계를 표현하기 어렵다. 예를 들어 퇴계 이황(退溪 李滉)이 저술한 여러 저작들과 그 이본들 간의 체계적인 연결이 구현되지 않아, 동일 저작의 저작군을 통합적으로 탐색하는 데 제약이 따른다. 이는 FRBR(Functional Requirements for Bibliographic Records) 모델이 요구하는 저작-표현형-구현형-개별자료 간의 계층 관계 표현과 근본적으로 상충된다(임소정, 손계영, 2023).

둘째, 텍스트 기반의 평면 구조로 인해 의미적 추론이 불가능하다. 주희(朱熹)와 주자(朱子)가 동일 인물임을 시스템이 자동으로 인식할 수 없으며, 동명이인 통제도 전거 연계 없이는 구현하기 어렵다. 이는 FAIR 원칙의 I1 항목이 요구하는 기계 가독성과 직접적으로 충돌한다(Wilkinson et al., 2016).

셋째, 외부 전거 파일과의 연계 메커니즘이 부재하여 글로벌 지식 네트워크와의 통합이 제한된다. VIAF, Wikidata 등 국제 전거 시스템과의 연계가 이루어지지 않아, 동일 인물이나 저작에 대한 국제적 식별이 가능하지 않다(이미화, 배정현, 2025). 예를 들어 KORCIS에 기재된 저자명 이황(李滉)은 VIAF나 Wikidata의 해당 식별자와 연결되지 않아, 해외 도서관 시스템에서 동일 인물로 인식되지 않는다.

넷째, 영구 식별자 체계가 부재하여 데이터의 지속적인 접근성과 인용 가능성이 보장되지 않는다. KORCIS의

내부 제어번호(001 필드)는 시스템 내부에서만 유효한 식별자로, DOI나 HTTP URI와 같이 전 지구적으로 고유하게 식별 가능한 식별자 조건을 충족하지 못한다. 이는 FAIR 원칙의 F1 항목이 명시하는 요건을 충족하지 못하는 상태이다(Wilkinson et al., 2016).

또한 KORCIS는 웹 검색 인터페이스만을 제공하며 REST API나 SPARQL 엔드포인트가 부재하여 기계적 접근이 구조적으로 불가능하다. 동일 서명에 대한 소장처 및 서명의 동시 제한 검색 기능이 제공되지 않으며, 동일 서명을 한글과 한자로 각각 입력하였을 때 검색 결과에 차이가 발생하는 유니코드 비일관성 문제도 확인된다(김영진, 2024). 이는 FAIR 원칙의 A1 항목과 직결되는 시스템 차원의 제약이다.

이러한 구조적 한계는 고문헌 메타데이터가 단순한 검색 목록의 역할을 넘어 학술 연구 및 문화유산 활용을 위한 지식 인프라로 발전하는 데 주요 제약 요인으로 작용하고 있다. 다음 절에서는 실증 데이터를 토대로 KORCIS 메타데이터의 품질 현황을 구체적으로 진단한다.

3.2 데이터 수집 및 분석 방법

분석 대상 데이터는 국립중앙도서관이 공공데이터포털을 통해 공개한 KORCIS 전체 서지 데이터를 수집하였다. 수집 파일은 2025년 7월 31일 기준으로 공개된 총 10개 파일이며, 합산 레코드 수는 567,212건이다. 참여 기관은 총 142개로 국내 91개 기관(551,606건, 97.25%)과 해외 51개 기관(15,606건, 2.75%)으로 구성되어 있다. 이는 KORCIS가 공식적으로 밝힌 144개 참여 기관, 51만여 건 규모와 거의 일치하며, 본 연구의 분석은 공개된 전체 데이터를 대상으로 한 전수 분석에 해당한다.

공개 데이터는 레코드키, MARC, 도서관부호, 도서관명의 4개 필드로 구성되어 있다. MARC 필드에는 KORCIS XML DTD 정의서에 따른 KORMARC 기반 서지 데이터가 수록되어 있으며, 레코드키는 전체 567,212건에서 중복이 없어 100% 유일성을 확보하고 있다. 도서관부호와 도서관명은 142개 기관에 대해 1:1 매핑 관계를 유지하고 있다.

기관별 소장 건수 분포는 <표 2>와 같다. 국립중앙도서관이 104,957건(18.50%)으로 가장 많은 데이터를 보유하고 있으며, 계명대학교 동산도서관(72,271건, 12.74%), 전남대학교 도서관(30,887건, 5.45%), 서울대학교 규장각한국학연구원(27,611건, 4.87%), 동국대학교 중앙도서관(25,279건, 4.46%) 순으로 나타났다. 상위 5개 기관이 전체의 46.02%, 상위 10개 기관이 62.72%, 상위 20개 기관이 78.21%를 점유하여 소장 데이터의 상위 기관 집중도가 뚜렷하게 확인된다. 해외 소장 기관 51개의 데이터는 15,606건으로 전체의 2.75%를 차지하며, 일본·미국·중국·프랑스·영국·러시아 등 다양한 국가의 기관이 포함되어 있다.

다만 공개 데이터의 기관별 수치는 각 기관이 KORCIS에 실제로 제공한 서지 건수이며, 기관의 실제 고문헌 소장 규모와는 차이가 있을 수 있다. 김영진(2024)에 따르면 일부 참여기관은 전체 소장 자료의 일부만 서지를 KORCIS에 제공하고 있어, 공개 데이터 수치가 해당 기관의 실제 소장 규모를 반영하지 않는 경우가 있다.

<표 2> KORCIS 데이터 참여 기관 현황(소장 건수 상위 20개)

순위	기관명	기관유형	레코드수	비율	누적비율
1	국립중앙도서관	국가기관	104,957	18.50%	18.50%
2	계명대학교 동산도서관	사립대학교	72,271	12.74%	31.25%
3	전남대학교 도서관	국립대학교	30,887	5.45%	36.69%
4	서울대학교 규장각한국학연구원	국립대학교	27,611	4.87%	41.56%
5	동국대학교 중앙도서관	사립대학교	25,279	4.46%	46.02%
6	한국국학진흥원 도서관	공공기관	22,659	3.99%	50.01%
7	고려대학교 도서관	사립대학교	22,327	3.94%	53.95%

순위	기관명	기관유형	레코드수	비율	누적비율
8	영남대학교 도서관	사립대학교	20,273	3.57%	57.52%
9	충남대학교 도서관	국립대학교	15,103	2.66%	60.18%
10	대구가톨릭대학교 중앙도서관	사립대학교	14,365	2.53%	62.72%
11	한국학중앙연구원 도서관	공공기관	12,346	2.18%	64.89%
12	성균관대학교 존경각	사립대학교	10,287	1.81%	66.71%
13	경상국립대학교 도서관	국립대학교	9,999	1.76%	68.47%
14	경기대학교 도서관	사립대학교	9,072	1.60%	70.07%
15	(사)나라얼연구소	연구소	8,361	1.47%	71.54%
16	서울대학교 중앙도서관	국립대학교	8,016	1.41%	72.96%
17	한양대학교 백남학술정보관	사립대학교	7,956	1.40%	74.36%
18	부산대학교 도서관	국립대학교	7,521	1.33%	75.68%
19	동국대학교 경주캠퍼스 도서관	사립대학교	7,370	1.30%	76.98%
20	원광대학교 도서관	사립대학교	6,969	1.23%	78.21%
	기타 122개 기관		123,583	21.79%	100%
	합 계		567,212	100%	

분석은 다음 세 단계로 수행하였다. 첫째, MARC 태그별 출현 빈도 및 완성도 분석이다. KORCIS XML DTD 정의서에 정의된 KORMARC 필드를 기준으로 주요 태그의 출현 빈도와 기재 완성도를 전체 데이터를 대상으로 측정하였다. 분석 대상 필드는 서명저자사항(245), 간사사항(260), 형태사항(300), 기본표목-개인명(100), 부출표목-개인명(700), 일반주제명(650), 지명(651), 비통제 색인어(653), 전자적 위치 및 접속(856)을 포함한다. 특히 300 필드의 형태코드(b 식별기호)에 수록된 고문헌 특화 물리 기술 정보인 광곽(코드 02), 계선(코드 04), 판구(코드 07), 어미(코드 08)의 입력 현황을 별도로 분석하였으며, 250 필드의 판종류코드(a 식별기호)를 통한 판본 유형 기재 현황도 함께 검토하였다.

둘째, 전거 제어 및 식별자 체계 분석이다. 100·700·650 등 전거 관련 필드의 0 식별기호(전거레코드 제어번호) 기재 현황을 확인하고, 외부 전거 시스템(VIAF, Wikidata, 국립중앙도서관 인명전거)과의 연계 여부를 점검하였다. 또한 레코드키의 식별자 체계와 856 필드 URI의 영구 식별자 적용 여부를 분석하였다.

셋째, FAIR 원칙 항목별 준수도 평가이다. Wilkinson et al.(2016)이 제시한 FAIR 원칙의 15개 세부 항목을 평가 틀로 삼아 각 항목별 준수 여부를 적합, 부분 적합, 미흡·결여의 3단계로 분류하였다. FAIR 원칙 자체는 원칙적 지침이므로, 각 항목의 준수 여부를 판정하기 위한 구체적 기준을 별도로 설정하였다. 판정은 ① 전체 데이터의 구조적 특성, ② 시스템 차원의 기능 지원 여부(API·SPARQL 엔드포인트·영구 식별자 체계 등의 존재 여부), ③ 앞선 두 분석 단계에서 산출된 정량 지표(필드 출현율·전거번호 기재율·영구 식별자 적용 건수 등)를 종합하여 수행하였으며, 개별 레코드의 내용적 정확성은 판정 범위에 포함하지 않았다. 판정 등급의 정의와 항목별 판정 기준은 다음과 같다. 적합한 항목의 요건을 데이터 구조 및 시스템 기능 차원에서 실질적으로 충족한 경우, 부분 적합은 일부 요건을 충족하나 핵심 요건이 제한적으로 이행된 경우, 미흡·결여는 요건 충족을 위한 구조적 조건이 부재하거나 이행 수준이 극히 낮은 경우로 정의한다.² 또한 각 항목의 판정 시에는 KORMARC 필드 유형을 함께 고려하였다. KORMARC 필드는 반드시 기재해야 하는 필수(M) 필드, 해당 자료에 관련 정보가 있는 경우에만 기재하는 해당시 필수(MA) 필드, 기재 여부가 기관의 재량에 따르는 재량(O) 필드로 구분된다. 따라서 필드 출현율은 필드 유형에 따른 해석이 필요하며, 재량 필드의 낮은 출현율을 단순히 데이터 품질의 결여로 환원하지 않는 것을 원칙으로 하였다.¹⁾

1) 15개 항목별 세부 판정 기준은 본 연구의 분석 데이터 및 스크립트와 함께 다음 리포지터리에 공개하였다.
<https://github.com/SHINEUNSUN1027/korcis-fair-analysis>

본 분석은 공공데이터포털을 통해 공개된 KORCIS 데이터에 한정되며, 개별 소장 기관의 내부 시스템은 분석 범위에 포함하지 않았다. 공개 데이터(567,212건, 142개 기관)는 KORCIS 전체 서비스 규모와 거의 일치하나, 앞서 언급한 바와 같이 일부 기관의 데이터가 부분적으로만 제공된 결과물임을 고려하여 해석하여야 한다. 또한 MARC 데이터의 세부 내용 분석은 필드 및 식별기호 단위의 구조적 특성을 중심으로 수행하였으며, 개별 레코드의 내용적 정확성에 대한 검증은 본 연구의 범위를 벗어난다.

3.3 메타데이터 필드 완성도 및 기관별 품질 분석

3.3.1 주요 필드 출현 빈도 분석

전체 567,212건을 대상으로 KORMARC 주요 필드의 출현 빈도를 분석한 결과는 <표 3>과 같다. 서명저자사항(245)은 567,210건(100.00%)으로 전 레코드에 존재한다. 245는 KORMARC 필수(M) 필드로서 미기재 시 시스템에 저장되지 않는 구조이므로, 이 수치는 필수 조건이 충족되어 있음을 확인하는 의미를 지닌다. 형태사항(300)은 567,177건(99.99%), 간사사항(260)은 565,668건(99.73%)으로 기본 서지 기술 필드의 완성도는 매우 높은 수준이다. 판사항(250)은 523,708건(92.33%)에 기재되어 있다. 250은 해당시 필수(MA) 필드로서 판사항 정보가 존재하는 자료에만 적용되므로, 출현율만으로 기재 충실도를 직접 판단하기는 어렵다. 다만 동일한 판본 유형이 木板本·목판본·木版本 등으로 혼용되는 비일관성 문제가 관찰되어, 통제 어휘 미적용에 따른 기술 형식의 불일치가 확인된다.

반면 재량(O) 필드의 완성도는 현저히 낮다. 비통제 색인어(653)가 288,111건(50.79%)으로 절반 수준에 그치며, 부출표목-개인명(700)은 253,425건(44.68%), 일반주제명(650)은 155,148건(27.35%)에 불과하다. 기본표목-개인명(100) 필드는 220,428건(38.86%)으로 전체의 절반에도 미치지 못한다. 다만 이 수치를 단순히 저자 정보의 누락으로 해석하는 것은 적절하지 않다. 100은 재량(O) 필드로서, 현행 한국목록규칙 제4판(KCR4)은 기본표목 설정을 규정하지 않으므로 100 필드를 사용하지 않는 것이 원칙이며, 해당 기관의 저자 정보는 245 필드의 d·e 식별기호에 기재되어 있는 경우가 일반적이다. 따라서 100 필드 출현율은 저자 정보의 기재 여부보다 기관별 목록 규칙 적용 방식의 차이를 반영하는 지표로 해석하여야 한다. 통제 주제명 관련 필드인 주제명부출-개인명(600)은 8,385건(1.48%), 지명주제(651)는 1,411건(0.25%)으로 극히 저조한 수준이며, 이들은 모두 재량(O) 필드이다. 전자적 위치 및 접속(856) 필드는 8,467건(1.49%)에 그쳐 원문 접근 정보 제공이 매우 제한적임을 보여준다.

<표 3> KORCIS 주요 메타데이터 필드 출현 빈도

태그	필드명	출현건수	출현율	비고
245	서명저자사항	567,210	100.00%	필수
300	형태사항	567,177	99.99%	필수
260	간사사항	565,668	99.73%	필수
500	일반주기	447,299	78.86%	선택
250	판사항(판본유형)	523,708	92.33%	선택
653	비통제 색인어	288,111	50.79%	선택
700	부출표목-개인명	253,425	44.68%	선택
100	기본표목-개인명	220,428	38.86%	선택
650	일반주제명	155,148	27.35%	선택
856	전자적 위치 및 접속	8,467	1.49%	선택
600	주제명부출(개인명)	8,385	1.48%	선택
651	지명주제	1,411	0.25%	선택

3.3.2 고문헌 기술 정보 입력 현황

KORCIS XML DTD 정의서는 형태사항(300) 필드의 b 식별기호(형태코드)에 고문헌의 물리적 특징을 세분하여 기술하도록 광곽(코드 02), 계선(코드 04), 판구(코드 07), 어미(코드 08) 등의 코드를 정의하고 있다. 이는 KORMARC이 고문헌 서지학의 전통적 기술 항목을 디지털 환경에 수용한 고문헌 특화 설계에 해당한다(강순애, 2014).

그러나 실제 데이터를 분석한 결과, 전체 567,212건 중 형태코드가 기재된 사례는 광곽(코드 02) 7건, 어미(코드 08) 12건, 삽화(코드 01) 25건 등으로 사실상 입력이 이루어지지 않고 있음이 확인되었다. 판사항(250) 필드를 통한 판본 유형 기재는 상대적으로 충실하여 목판본(木板本) 107,415건(18.94%), 필사본(筆寫本) 106,411건(18.76%), 목활자본(木活字本) 41,789건(7.37%) 등의 분포가 확인된다. 그러나 판본 유형은 자유 텍스트 형식으로 입력되는 경우가 다수여서 동일한 판본 유형이 木板本, 목판본, 木版本 등 다양한 표기로 혼용되는 비일관성 문제가 관찰된다. 이는 KORCIS XML DTD 정의서가 판종류코드를 별도 정의하고 있음에도 현장에서는 통제 어휘 대신 자유 텍스트 입력이 관행화되어 있음을 보여준다.

3.3.3 전거 제어 및 외부 연계 실태

전거 제어는 동일 인물·단체·개념에 대한 이표기를 통합하여 검색의 정확성과 데이터 연계 가능성을 확보하는 핵심 메커니즘이다. KORCIS의 전거 제어 실태를 전거 관련 필드(100·700·650)의 0 식별기호(전거레코드 제어번호) 기재 현황을 중심으로 분석하였다. 다만 기관에 따라서는 시스템 내부에서 전거가 연결되어 있더라도 서지 MARC의 0 식별기호에 해당 번호가 기재되지 않는 경우가 있다. 따라서 본 분석 결과는 전거 연계의 실제 적용 범위가 아닌, 공개된 서지 레코드 차원에서의 전거 연계 가시성(visibility)을 측정하는 것으로 해석하여야 한다.

기본표목-개인명(100) 필드를 보유한 220,428건 중 전거레코드 제어번호가 기재된 레코드는 1,600건(0.73%)에 불과하며, 나머지 218,831건(99.27%)은 전거번호 없이 자유 텍스트 형태의 인명만 기재되어 있다. 부출표목-개인명(700) 필드의 경우도 253,425건 중 전거번호가 연결된 레코드는 1,720건(0.68%)에 그치며, 일반주제명(650) 필드에서도 전거번호가 기재된 사례는 해당 필드 보유 레코드(155,148건) 대비 1,199건(0.77%)에 불과하다.

전거번호가 기재된 레코드의 형식을 확인한 결과, KAC로 시작하는 국립중앙도서관 인명전거 식별자(예: KAC200203290, KAC201834611)가 사용되고 있었다. 그러나 VIAF, Wikidata, LCNAF 등 국제 전거 시스템의 URI는 단 1건도 확인되지 않았다. 국가전거데이터(KAC)는 국가서지 LOD를 통해 공개되어 있으며(국립중앙도서관, n.d.a), VIAF 등 국제 전거 시스템과의 간접 연계가 추진되고 있다. 그러나 본 분석에서 확인한 것은 KORCIS 서지 레코드 단위에서 국제 전거 URI가 직접 기재된 사례가 없다는 점이다. 서지 레코드 자체에 URI가 명시되지 않으면 외부 시스템이 해당 연계를 자동으로 탐지·활용하기 어려우며, 이는 FAIR 원칙의 I3(정규화된 외부 참조)이 요구하는 레코드 단위의 명시적 연계가 이루어지지 않고 있음을 의미한다. 결과적으로 KORCIS에 입력된 인명은 동일 인물의 이표기(예: 朱熹·朱子·주희·주자, 李滉·退溪·이황·퇴계)가 각기 독립된 문자열로 처리되어 시스템이 동일 인물로 인식하거나 통합 검색하는 것이 구조적으로 불가능한 상태이다. 이는 FAIR 원칙의 I3(정규화된 외부 참조) 항목과 직접적으로 충돌하는 구조적 문제이다(Wilkinson et al., 2016).

서지 레코드의 식별자 체계를 살펴보면, KORCIS는 001 필드에 내부 제어번호를 부여하고 있으나 그 체계가 통일되어 있지 않다. OLDP로 시작하는 국립중앙도서관 방식(예: OLDP1000000222)이 3,754건, 각 기관의 고유 코드를 앞에 붙인 기관코드형이 183,039건, 그 외 기타 형식이 380,419건으로 가장 많았다. 더 근본적인 문제는 이 제어번호들이 어느 것도 HTTP URI, DOI, ARK, PURL과 같은 국제적으로 통용되는 영구 식별자 형식을 따르지 않는다는 점이다. 따라서 KORCIS의 서지 레코드는 외부 시스템에서 식별자를 통해 직접 접근하거나 인용하는 것이 불가능하며, 이는 FAIR 원칙의 F1(전 지구적 고유 식별자 부여) 항목을 충족하지 못하는 상태이다(Wilkinson et al., 2016).

전자적 위치 및 접속(856) 필드에 기재된 URL 8,467건의 도메인을 분석한 결과, dcollection.kmu.ac.kr(계명대, 7,590건), lib.aks.ac.kr(한국학중앙연구원, 2,412건), dcollection.ewha.ac.kr(이화여대, 2,315건) 등 각 기관 내부 디지털 컬렉션 서버를 가리키는 URL이 대부분을 차지하였다. IP 주소를 직접 사용한 URL(예: http://168.188.11.60 /...)도 확인되어 서버 이전 시 링크 무효화 가능성이 높은 구조임을 알 수 있다. DOI, PURL, ARK 등 영구 식별자 형식을 따르는 URL은 단 1건도 확인되지 않았다. 이는 FAIR 원칙의 AI(표준 프로토콜을 통한 식별자 기반 접근) 항목의 실질적 이행이 이루어지지 않고 있음을 보여준다.

3.3.4 기관별 메타데이터 품질 격차

소장 건수 상위 10개 기관을 대상으로 주요 선택 필드의 출현율을 비교한 결과, 기관 간 메타데이터 품질에 상당한 격차가 존재함이 확인되었다. 분석 결과는 <표 4>와 같다.

<표 4> 상위 10개 기관별 주요 필드 출현율

기관명	총 건수	100저자	250판본	650주제	653색인어	700부출저자	856URL
국립중앙도서관	104,957	37.2%	90.5%	22.8%	93.4%	45.9%	2.4%
계명대 동산도서관	72,271	64.5%	99.9%	0.0%	34.6%	52.5%	1.8%
전남대 도서관	30,887	91.8%	100.0%	78.8%	99.7%	16.6%	3.8%
서울대 규장각	27,611	0.0%	99.4%	99.7%	0.0%	45.6%	0.0%
동국대 중앙도서관	25,279	57.0%	98.8%	83.3%	92.3%	70.8%	8.2%
한국국학진흥원	22,659	0.0%	90.9%	0.0%	0.0%	49.7%	0.0%
고려대 도서관	22,327	68.9%	97.6%	0.0%	99.6%	41.6%	0.6%
영남대 도서관	20,273	0.7%	56.1%	0.0%	0.2%	77.7%	0.0%
충남대 도서관	15,103	40.5%	99.9%	68.0%	45.1%	60.1%	2.2%
대구가톨릭대 도서관	14,365	60.5%	33.3%	0.1%	100.0%	28.7%	0.0%

분석 결과에서 주목되는 점은 다음과 같다. 첫째, 기본표목-개인명(100) 필드의 출현율이 기관에 따라 0.0%에서 91.8%까지 극단적인 편차를 보인다. 이는 KCR4가 기본표목 설정을 규정하지 않아 100 필드를 사용하지 않는 것이 원칙임에도 불구하고, 기관에 따라 100 필드를 실무적으로 활용하는 경우와 그렇지 않은 경우가 혼재하는 현실을 반영한다. 100 필드를 사용하지 않는 기관의 저자 정보는 대개 245 필드의 d·e 식별기호에 기재되어 있으므로, 이 필드의 출현율 차이를 곧바로 데이터 품질의 우열로 환원하기 어렵다. 다만 기관별로 동일한 정보가 상이한 필드에 분산 입력되는 구조는 시스템 간 데이터 통합 활용의 일관성을 저해하는 요인으로 지적될 수 있다.

둘째, 주제 접근 필드의 패턴이 기관별로 상이하다. 서울대학교 규장각한국학연구원은 650(일반주제명) 출현율이 99.7%로 가장 높지만 653(비통제 색인어)은 전혀 기재하지 않는 반면, 고려대학교 도서관은 650이 0%이고 653이 99.6%로 정반대의 패턴을 보인다. 이는 단순히 기관별 입력 지침의 차이라기보다, 기관이 활용할 수 있는 통제 어휘 인프라의 구비 여부, 해당 시소러스에 적용 가능한 용어가 수록되어 있는지 여부 등 어휘 환경의 차이에서 비롯된 것으로 해석된다. 통제 어휘를 활용할 수 있는 기관은 650 필드를 우선 적용하고, 해당 용어가 없거나 통제 어휘 인프라가 갖추어지지 않은 경우 비통제 색인어(653)를 사용하는 것이 일반적 관행이다.

셋째, 영남대학교 도서관은 100(0.7%), 250(56.1%), 650(0.0%), 653(0.2%)으로 선택 필드 전반의 완성도가 낮아 주요 기관 중 가장 미흡한 수준을 보인다. 대구가톨릭대학교 중앙도서관은 250(판본) 출현율이 33.3%로 유독 낮아, 판본 정보 입력이 체계적으로 이루어지지 않고 있음을 알 수 있다.

이러한 기관별 품질 격차는 KORCIS가 통합 목록 포털로서 표준화된 서지 기술 환경을 지향하고 있음에도 불구하고, 기관별 입력 관행이 일관되게 수립되지 않고 있음을 보여준다.

3.4 FAIR 원칙 준수도 평가

이상 3.3절의 분석 결과를 종합하여 FAIR 원칙 15개 세부 항목별 준수도를 평가하면 다음과 같다. 판정 기준은 3.2절에서 제시한 바와 같이 적합, 부분 적합, 미흡·결여의 3단계로 분류하였다. 개별 레코드의 내용적 정확성은 판정 범위에 포함하지 않았다.2)

〈표 5〉 KORCIS FAIR 원칙 준수도 평가 결과

영역	항목	평가	근거
F	F1 전 지구적 고유 식별자	미흡·결여	내부 제어번호만 존재, HTTP URI·DOI 없음
F	F2 풍부한 메타데이터 기술	부분 적합	필수 필드 충실, 선택 필드 기관별 편차
F	F3 메타데이터에 식별자 명시	미흡·결여	영구 식별자 미부여로 명시 불가
F	F4 외부 인덱스 등록	미흡·결여	국제 메타데이터 인덱스 미등록
A	A1 표준 프로토콜 기반 접근	부분 적합	웹 검색 가능, API·SPARQL 미제공
A	A1.1 무료·개방 프로토콜	부분 적합	웹 접근 무료, 기계적 접근 미지원
A	A1.2 인증·권한 허용 프로토콜	미흡·결여	접근 제어 정책 미명시
A	A2 데이터 불가 시 메타데이터 보존	미흡·결여	메타데이터 장기 보존 정책 미명시
I	I1 공식 지식 표현 언어 사용	미흡·결여	RDF·OWL 미사용, KORMARC 텍스트 형식
I	I2 FAIR 준수 어휘 사용	미흡·결여	Dublin Core·SKOS 등 표준 어휘 미적용
I	I3 정규화된 외부 참조	미흡·결여	VIAF·Wikidata 연계 전무, 국내 전거 0.73%
R	R1 복수 속성 메타데이터 기술	부분 적합	서지 기술 충실, 관계·맥락 정보 미흡
R	R1.1 라이선스 정보 제공	미흡·결여	데이터 이용 라이선스 미명시
R	R1.2 데이터 출처 및 변동 이력	미흡·결여	원천 기관 정보 외 변동 이력 미기록
R	R1.3 도메인별 커뮤니티 표준 준수	부분 적합	KORMARC 준수, 국제 표준(BIBFRAME 등) 미적용

평가 결과 미흡·결여 10개(66.7%), 부분 적합 5개(33.3%), 적합 0개(0%)로 나타났다. 영역별로는 발견 가능성(F) 4개 항목 중 3개 미흡·결여, 1개 부분 적합이며, 접근 가능성(A) 4개 항목 중 2개 미흡·결여, 2개 부분 적합이다. 상호운용성(I) 3개 항목은 전부 미흡·결여이며, 재사용 가능성(R) 4개 항목 중 2개 미흡·결여, 2개 부분 적합이다.

부분 적합으로 평가된 항목들도 실질적 이행 수준은 낮다. F2(풍부한 메타데이터 기술)의 경우 필수 필드의 완성도는 높으나 선택 필드의 기관별 편차가 크고 국제 표준 어휘를 사용하지 않아 기계 가독성이 낮다. A1(표준 프로토콜 기반 접근)은 웹 검색 인터페이스가 제공되나 API나 SPARQL 엔드포인트가 부재하여 기계적 접근이 불가능하다. R1.3(커뮤니티 표준 준수)은 KORMARC이라는 국내 표준을 준수하고 있으나 BIBFRAME, EDM 등 국제 표준과의 정렬이 이루어지지 않아 국제적 상호운용성을 충족하지 못한다.

가장 심각한 항목은 I 영역 전체이다. RDF와 같은 공식 지식 표현 언어를 사용하지 않고(I1), FAIR 준수 표준 어휘도 적용되지 않으며(I2), 전거번호를 통한 외부 연계는 전체 레코드의 1% 미만으로 사실상 부재한 상태이다(I3). 이는 KORCIS 데이터가 글로벌 지식 네트워크와 단절된 채 국내 시스템 내부에서만 유통되고 있음을 의미한다.

이상의 진단 결과는 KORCIS 메타데이터의 시맨틱 웹 전환과 국제 전거 연계가 시급한 과제를 실증적으로 보여준다. 이에 대한 구체적인 표준화 방향은 5장에서 논의한다.

2) 항목별 세부 판정 기준은 본 연구의 분석 스크립트와 함께 다음 리포지토리에 공개되어 있다.
<https://github.com/SHINEUNSUN1027/korcis-fair-analysis>

4. 해외 고문헌 메타데이터 시맨틱 표준화 사례

4.1 Europeana: EDM 기반 LOD 구현

Europeana는 유럽 각국의 도서관·박물관·아카이브(LAM)로부터 수집한 문화유산 객체를 단일 플랫폼에서 제공하는 유럽의 대표적 문화유산 LOD 플랫폼이다. 2008년 서비스를 개시한 이래 현재 약 4,500만 건 이상의 디지털 문화유산 객체를 제공하고 있으며, 유럽연합의 문화유산 디지털 인프라 핵심 사업으로 운영되고 있다(Europeana, n.d.b). Europeana의 주요 특징은 이질적인 기관 데이터를 통합하기 위해 독자적인 데이터 모델인 Europeana Data Model(EDM)을 개발하고 이를 LOD 형식으로 공개하고 있다는 점이다(Europeana, n.d.a).

EDM은 Dublin Core, OAI-ORE, SKOS, OWL, FOAF 등 기존 국제 표준 어휘를 재사용하는 방식으로 설계되었다. 이는 새로운 어휘를 독자적으로 개발하는 대신 이미 널리 통용되는 표준을 결합하여 상호운용성을 확보하는 전략이다(Isaac, 2013). EDM의 핵심 클래스 구조는 다음과 같다. edm:ProvidedCHO(제공된 문화유산 객체)는 실제 문화유산 객체를 표현하며, ore:Aggregation은 디지털 표현과 메타데이터를 집합적으로 관리하고, edm:WebResource는 디지털 이미지·원문 등 웹 접근 가능한 자원을 표현한다. 각 객체에는 HTTP URI 기반의 고유 식별자가 부여되어 FAIR 원칙의 F1 항목을 충족한다.

EDM은 인물·장소·시간·개념 등 맥락 정보를 별도의 개체로 분리하여 기술하는 구조를 채택하고 있다. 예를 들어 특정 회화 작품에 대한 메타데이터는 작품 자체(edm:ProvidedCHO), 작가(edm:Agent), 제작 시기(edm:TimeSpan), 소장 장소(edm:Place)를 각각의 URI로 식별하고 이들 간의 관계를 명시적으로 기술한다(Isaac, 2013). 이러한 구조는 FAIR 원칙의 I1·I2 항목을 충족하며, 단순 텍스트 기반 서지 레코드로는 불가능한 관계 중심의 데이터 탐색을 가능하게 한다.

Europeana는 VIAF, GeoNames, Getty Thesaurus of Geographic Names(TGN), DBpedia 등 국제 전거 시스템과 owl:sameAs 속성을 통해 연결되어 있다(Europeana, n.d.b). 예를 들어 특정 인물의 전거는 Europeana 내부 URI와 함께 VIAF URI, DBpedia URI, Wikidata URI가 owl:sameAs로 연결되어, 서로 다른 플랫폼에서 동일 인물 데이터가 단일한 연결망 내에서 통합 탐색될 수 있는 구조를 갖추고 있다. 이는 FAIR 원칙의 I3 항목의 실질적 이행 사례에 해당한다.

Europeana는 메타데이터에 CC0 라이선스를 적용하여 누구나 제한 없이 재사용할 수 있도록 하고 있으며, 디지털 객체에는 저작권 상태에 따라 CC BY, CC BY-SA, Rights Reserved 등 다양한 라이선스를 명시하고 있다(Europeana, n.d.b). 또한 data.europeana.eu를 통한 LOD 공개와 함께 REST API를 통한 JSON-LD 형식의 데이터 제공을 지원하여 기계적 접근 환경을 갖추고 있다.

Europeana 사례는 국제 표준 어휘 재사용, 개체 중심의 관계 기술, 라이선스 명시와 기계적 접근 환경 구축이 FAIR 이행의 실질적 조건임을 보여주며, KORCIS는 이 조건들을 현재 충족하지 못하고 있다.

4.2 일본 국립국회도서관(NDL): BIBFRAME · VIAF 연계

일본 국립국회도서관(国立国会図書館, 이하 NDL)은 국립국회도서관 디지털 컬렉션(国立国会図書館デジタルコレクション)을 통해 고전 자료를 포함한 방대한 디지털 자원을 제공하며, 국제 LOD 환경과의 연계를 위해 전거 데이터의 RDF 공개 및 시맨틱 표준화를 단계적으로 추진하고 있다. NDL의 시맨틱 표준화 전략은 크게 세 축으로 구성된다. 전거 데이터의 LOD 공개, VIAF와의 owl:sameAs 연결, 그리고 BIBFRAME 기반 서지 데이터 전환이 그것이다.

NDL은 Web NDL Authorities(国立国会図書館典拠データ検索・提供サービス)를 통해 인명·단체명·지명·주제명 선거 데이터를 RDF 형식으로 공개하고 있다(国立国会図書館, n.d.). 각 선거 개체에는 HTTP URI 형식의 고유 식별자(예: <https://id.ndl.go.jp/auth/entity/00271847>)가 부여되어 외부 시스템에서 직접 참조하거나 인용하는 것이 가능하며(F1), SPARQL 엔드포인트(<https://id.ndl.go.jp/auth/ndla>)를 통해 프로그래밍 방식의 데이터 조회도 지원한다(国立国会図書館, n.d.).

NDL 선거 데이터의 핵심 특징은 국제 선거 시스템과의 연계이다. 인명 선거 데이터는 동일 인물에 대한 VIAF URI와 owl:sameAs 속성으로 연결되어 있어 단일한 연결망 내에서 통합 탐색될 수 있는 구조를 갖추고 있다. 예를 들어 松尾芭蕉(마쓰오 바쇼)의 NDL 선거 URI는 VIAF URI(<https://viaf.org/viaf/5232080>) 및 Wikidata URI와 owl:sameAs로 연결되어 국제 도서관 시스템에서 동일 인물로 인식되고 다국어 데이터와 통합 탐색이 가능하다. 이는 FAIR 원칙의 I3 항목의 실질적 이행 사례에 해당하며, KORCIS에서 국내 선거번호 연계 레코드가 전체의 1% 미만에 그치고 VIAF·Wikidata 연계는 전무한 상황과 대조된다.

서지 데이터 표준화 측면에서 NDL은 BIBFRAME 전환을 단계적으로 추진하고 있다. 기존 JAPAN/MARC 기반 데이터를 RDF로 변환하는 시범 사업을 병행하고 있으며, 저작(Work)·인스턴스(Instance)·개별자료(Item)의 계층 구조로 서지 개체를 재편하는 방향을 모색하고 있다. BIBFRAME은 MARC의 평면적 레코드 구조와 달리 저작·표현형·구현형·개별자료 간의 관계를 RDF 트리플로 명시하여, 동일 저작의 이본(異本)과 번역본 간의 연결을 기계 가독성 기반으로 표현할 수 있다.

다만 NDL의 시맨틱 표준화도 완전히 구현된 상태는 아니다. 디지털 컬렉션 전체의 서지 레코드가 BIBFRAME으로 전환된 것은 아니며, 고전 자료에 특화된 서지 기술 항목의 RDF 표현 방식은 여전히 정비 과정에 있다. 또한 데이터 이용 라이선스가 선거 데이터와 서지 데이터에 따라 상이하게 적용되어 있어 R1.1 항목의 일관성도 개선 과제로 남아 있다.

자국 선거 식별자와 VIAF URI를 owl:sameAs로 연결하는 NDL의 방식은 국내 선거 체계를 유지하면서도 글로벌 지식 네트워크와의 상호운용성을 확보하는 경로를 보여주며, KORCIS의 KAC 식별자를 VIAF 및 Wikidata와 연계하는 방향을 설계하는 데 직접적인 참조 모델이 된다.

4.3 중국 중화고적자원고(中華古籍資源庫)의 현황과 한계

중국의 고전 문헌 디지털화는 1980년대 탐색기를 거쳐 1990년대 상업화·대형화 단계로 발전하였으며, 2007년 중국 정부가 ‘중화고적보호계획(中華古籍保護計劃)’을 수립한 이후 국가 주도의 공공 디지털화 사업이 본격화되었다(장염, 2022). 그 핵심 성과가 2016년 9월 정식 서비스를 개시한 중화고적자원고(中華古籍資源庫)이다. 중국 국가도서관이 주관하고 국내외 도서관·연구기관이 참여하는 이 플랫폼은 현재 33,000건 이상의 희귀 고문헌을 무료로 공개하고 있으며, 국내 14개 지방 도서관과 프랑스 국가도서관·하버드대학 옌칭도서관·도쿄대학 동양문화연구소 등의 해외 기관 소장 자료도 포함한다(장염, 2022).

중화고적자원고의 서지 메타데이터는 서명, 소장처, 판본 유형, 수록 연대 등의 기본 서지 사항을 제공하나, 고문헌 원문은 대부분 이미지 형태로만 제공되며 원문 텍스트에 대한 다기능 검색은 지원하지 않는다(장염, 2022). 허철(2024)은 중국의 고문헌 디지털화에서도 한국과 공통적으로 상세한 메타데이터의 부재, 데이터 이력 관리 미흡, 구축 표준 및 평가 도구의 부재 문제가 나타나고 있음을 지적한다. 결과적으로 중화고적자원고의 서지 레코드는 기본적인 탐색 도구로서의 기능은 수행하나, 기계 가독성과 의미 추론을 위한 구조화된 데이터로서의 기능은 갖추지 못하고 있다.

중화고적자원고는 현재 RDF 기반의 LOD 형식으로 데이터를 공개하지 않으며, SPARQL 엔드포인트나 API를 통한 기계적 접근도 지원하지 않는다. 서지 레코드에 HTTP URI 형식의 영구 식별자가 부여되지 않아 외부 시스

템에서 개별 레코드를 직접 참조하는 것이 불가능하며, VIAF · Wikidata · LCNAF 등 국제 전거 시스템과의 연계도 이루어지지 않아 FAIR 원칙의 I1 · I2 · I3 항목 전반에서 미충족 상태이다.

중국의 고전 문헌 디지털화는 공공 기관 외에 베이징 애여생(愛如生) 디지털화 기술 연구 센터, 서동문(書同文) 디지털 기술 유한회사, 중화서국(中華書局) 등 기업과 출판사가 대형 상업 데이터베이스를 병행 구축하는 구조를 갖고 있다. 이들 상업 데이터베이스는 원문 이미지와 텍스트를 함께 제공하고 다기능 검색을 지원하는 등 서지 기능 면에서는 공공 플랫폼보다 우수하나, 이용료 기반 운영으로 전면적인 데이터 공개가 이루어지지 않고 있다(허철, 2024). 또한 개별 사업자 단위의 독립적 데이터 구축으로 동일 문헌의 중복 디지털화와 기관 간 메타데이터 이질성 문제가 발생하고 있으며, 시맨틱 표준화 측면에서도 중화고적자원고와 마찬가지로 미구현 상태이다(허철, 2024).

이러한 중국 사례는 LOD 전환과 국제 전거 연계가 동아시아 고문헌 메타데이터 전반의 공통 과제를 보여준다. 특히 국가 차원의 통일된 계획 없이 기관별로 독립적으로 데이터베이스를 구축할 경우 중복 디지털화와 메타데이터 이질성이 심화된다는 점은 KORCIS의 기관별 품질 격차 문제와 구조적으로 동일한 맥락에서 이해될 수 있다.

4.4 비교 분석 및 시사점

앞서 검토한 Europeana, 일본 국립국회도서관(NDL), 중화고적자원고의 사례를 FAIR 원칙 15개 항목을 기준으로 비교하면 <표 6>과 같다.

<표 6> 해외 주요 기관 FAIR 원칙 준수도 비교

항목	Europeana	NDL	중화고적자원고	KORCIS
F1 고유 식별자	HTTP URI	HTTP URI	미충족	미충족
F2 풍부한 메타데이터	충족	충족	부분 충족	부분 충족
F3 식별자 명시	충족	충족	미충족	미충족
F4 외부 인텍스 등록	충족	부분 충족	미충족	미충족
A1 표준 프로토콜 접근	REST API / SPARQL	SPARQL	미충족	미충족
A1.1 무료·개방 프로토콜	충족	충족	부분 충족	부분 충족
A1.2 인증·권한 프로토콜	충족	충족	미명시	미명시
A2 메타데이터 보존	충족	충족	미명시	미명시
I1 공식 지식 표현 언어	RDF / OWL	RDF(전거)	미충족	미충족
I2 표준 어휘 사용	DC / SKOS / OAI-ORE	BIBFRAME(추진)	미충족	미충족
I3 외부 참조 연계	VIAF / DBpedia / Wikidata	VIAF / Wikidata	미충족	미충족
R1 복수 속성 기술	충족	충족	부분 충족	부분 충족
R1.1 라이선스 제공	CC0(메타데이터)	부분 명시	미명시	미명시
R1.2 출처·변동 이력	충족	부분 충족	미충족	미충족
R1.3 커뮤니티 표준 준수	EDM / BIBFRAME	JAPANMARC / BIBFRAME	자체 기준	KORMARC

비교 분석 결과, 네 기관은 FAIR 이행 수준에 따라 두 그룹으로 명확히 구분된다. Europeana와 NDL은 HTTP URI 기반 고유 식별자 부여, RDF 형식의 데이터 공개, 국제 전거 시스템 연계, 기계적 접근 환경 구축이라는 네 가지 조건을 상당 수준 충족하고 있는 반면, 중화고적자원고와 KORCIS는 텍스트 기반 서지 기술에 머물러 있어 F·I 영역 전반에서 미충족 상태이다. 특히 상호운용성(I) 영역에서의 격차가 두드러지는데, 이는 단순한 기술 선택의 문제가 아니라 데이터 설계 단계에서의 표준 정렬 여부에서 비롯된 구조적 차이이다.

세 해외 사례의 비교에서 도출되는 공통 원리와 기관별 특징을 정리하면 다음과 같다.

첫째, FAIR 이행의 출발점은 영구 식별자 체계의 확립이다. Europeana와 NDL은 모두 HTTP URI를 개별 서지 개체와 전거 개체에 부여하는 것을 LOD 구축의 전제 조건으로 삼고 있다. 식별자가 없으면 외부 시스템에서의 참조와 인용이 불가능하고, 이는 연쇄적으로 A1(프로토콜 기반 접근)과 I3(외부 참조 연계)의 구현을 막는다. KORCIS의 내부 제어번호와 중화고적자원고의 식별자 체계가 공통적으로 국제 통용 영구 식별자 형식을 갖추지 못하고 있다는 점은, 두 시스템 모두 FAIR 이행의 기초 조건에서 출발 단계에 머물러 있음을 보여준다.

둘째, 독자적 표준 개발보다 기존 국제 표준 어휘의 재사용이 상호운용성 확보에 효과적이다. Europeana의 EDM은 Dublin Core, OAI-ORE, SKOS, OWL 등 이미 통용되는 표준 어휘를 결합하는 방식으로 설계되었으며 (Isaac, 2013), NDL 역시 BIBFRAME이라는 국제 표준을 전향 목표로 삼고 있다. 이에 반해 중화고적자원고는 자체 기준에 따른 메타데이터 구조를 유지하면서 국제 표준과의 정렬을 이루지 못하고 있다. KORCIS의 KORMARC 역시 국내 표준으로서 체계적으로 운영되고 있으나, BIBFRAME이나 EDM과의 정렬이 이루어지지 않아 국제적 상호운용성을 기대하기 어렵다.

셋째, 전면적 시스템 재구축 없이도 단계적 전환이 가능하다는 점을 NDL 사례가 보여준다. NDL은 기존 JAPAN/MARC 기반 서지 시스템을 유지하면서 전거 데이터만 우선적으로 RDF 형식으로 공개하고 SPARQL 엔드포인트를 구축하였으며, 이를 통해 F1·A1·I3 항목을 단계적으로 충족하는 경로를 택하였다. 이는 대규모 데이터 변환 작업이 수반되는 BIBFRAME 전면 전환에 앞서 현실적으로 실행 가능한 중간 단계를 설정한 사례로, 자원과 인력이 제한된 환경에서의 전환 전략으로 참고할 수 있다.

넷째, 중화고적자원고 사례는 국가 주도의 대규모 디지털화 사업도 LOD 전환과 국제 전거 연계 없이는 FAIR 원칙 이행에 한계가 있음을 보여준다. 허철(2024)이 지적하듯 한국과 중국 모두 다양하고 상세한 메타데이터의 부재, 데이터 이력 관리 미흡, 구축 표준의 부재라는 공통 과제를 안고 있다. 이는 동아시아 고문헌 메타데이터 전반의 구조적 문제임과 동시에, 국제 표준과의 정렬을 처음부터 설계 원칙으로 내재화하지 않은 데서 비롯된 결과이다.

이상의 비교 분석을 통해 KORCIS의 시맨틱 표준화 방향에 대한 시사점을 세 가지로 정리할 수 있다. 첫째, HTTP URI 기반 영구 식별자를 서지 레코드와 전거 개체에 부여하는 것이 FAIR 이행의 선결 조건이다. 둘째, 전거 데이터를 우선적으로 RDF 형식으로 공개하고 SPARQL 엔드포인트를 구축하는 NDL 방식은 전면적 시스템 재구축에 앞서 실행 가능한 단계적 전환 경로로 검토할 수 있다. 셋째, 국내 전거 식별자(KAC)를 VIAF 및 Wikidata URI와 owl:sameAs로 연결하는 방식은 국내 전거 체계를 유지하면서도 글로벌 지식 네트워크와의 상호운용성을 확보하는 현실적인 방법이다. 이를 구체적으로 구현하기 위한 온톨로지 설계 및 변환 방안은 후속 연구에서 다룬다.

5. 결론

본 연구는 한국 고문헌 메타데이터의 국가 단위 통합 인프라인 KORCIS를 대상으로 FAIR 원칙에 의거한 실증적 진단을 수행하고, 해외 주요 기관의 시맨틱 표준화 사례와 비교하여 국내 고문헌 메타데이터의 구조적 문제와 개선 방향을 제시하였다.

KORCIS 전체 서지 데이터 567,212건을 대상으로 수행한 전수 분석 결과, FAIR 원칙 15개 세부 항목 중 적합 판정을 받은 항목은 단 하나도 없었으며 미흡·결여 10개(66.7%), 부분 적합 5개(33.3%)로 나타났다. 서명저자사항(245), 형태사항(300), 간사사항(260) 등 필수 서지 필드의 완성도는 99% 이상으로 높은 수준이나, 기본표목-개인명(100) 38.86%, 일반주제명(650) 27.35%, 전자적 위치 및 접속(856) 1.49% 등 선택 필드의 완성도는 현저히 낮았다. 기관별로는 동일 정보를 상이한 필드에 입력하는 관행의 차이로 인해 기관 간 메타데이터 품질 격차가

뚜렷하게 확인되었다.

전자 제어 실패는 더욱 심각하다. 인명 필드(100·700)를 보유한 레코드 중 전자레코드 제어번호(0 식별기호)가 기재된 레코드는 전체의 1% 미만에 그쳤으며, VIAF·Wikidata·LCNAF 등 국제 전자 시스템의 URI는 단 1건도 확인되지 않았다. 서지 레코드의 내부 제어번호는 HTTP URI·DOI 등 국제적으로 통용되는 영구 식별자 형식을 따르지 않아 외부 시스템에서의 직접 참조와 인용이 불가능하며, API나 SPARQL 엔드포인트도 제공되지 않아 기계적 접근 환경이 전혀 구축되어 있지 않다. 이러한 결과는 KORCIS 데이터가 글로벌 지식 네트워크와 단절된 채 국내 시스템 내부에서만 유통되고 있음을 실증적으로 보여준다.

해외 사례 비교에서는 Europeana, 일본 국립국회도서관(NDL), 중화고적자원고를 검토하였다. Europeana는 EDM을 통해 이질적인 기관 데이터를 단일 온톨로지로 통합하고 HTTP URI 기반 고유 식별자, CC0 라이선스, REST API·SPARQL 엔드포인트를 구축하여 FAIR 원칙의 주요 항목을 충족하고 있다. NDL은 전면적인 시스템 재구축 없이 전자 데이터를 우선적으로 RDF 형식으로 공개하고 VIAF URI와 owl:sameAs로 연결하는 단계적 전환 전략을 채택하고 있으며, 이는 자원과 인력이 제한된 환경에서의 현실적 참조 모델이 된다. 중화고적자원고는 국가 주도의 대규모 디지털화 사업임에도 LOD 전환과 국제 전자 연계를 실현하지 못하고 있으며, 허철(2024)이 지적하듯 다양하고 상세한 메타데이터의 부재, 데이터 이력 관리 미흡, 구축 표준의 부재라는 문제가 한국과 공통적으로 나타나고 있다.

이상의 진단과 비교 분석을 통해 KORCIS의 시맨틱 표준화를 위한 선결 과제를 세 가지로 정리할 수 있다. 첫째, 서지 레코드와 전자 개체에 HTTP URI 기반 영구 식별자를 부여하는 것이 FAIR 이행의 기초 조건이다. 둘째, 전자 데이터를 우선적으로 RDF 형식으로 공개하고 SPARQL 엔드포인트를 구축하는 단계적 전환 경로를 설정할 필요가 있다. 셋째, 국내 전자 식별자(KAC)를 VIAF 및 Wikidata URI와 owl:sameAs로 연결함으로써 국내 전자 체계를 유지하면서도 글로벌 지식 네트워크와의 상호운용성을 확보할 수 있다.

이상의 진단 결과는 한국 고문헌 메타데이터 관리체계 전반의 재정비 필요성을 실증적으로 뒷받침한다. 여기서 메타데이터 관리체계란 서지 데이터의 생산·기술·식별·접근·연계에 이르는 전 과정을 규율하는 표준, 시스템 구조, 운영 정책의 총체를 의미하며, 본 연구의 진단은 이 관리체계의 핵심 요소-식별자 체계, 전자 연계 구조, 접근 프로토콜, 라이선스 정책-가 국제 표준 수준에서 전반적으로 미흡함을 확인시켜 준다.

본 연구는 KORCIS 전체 공개 데이터를 대상으로 한 정량적 분석을 통해 국내 고문헌 메타데이터의 FAIR 이행 실패를 실증적으로 규명하였다는 점에서 의의가 있다. 다만 분석 대상이 공개 데이터에 한정되어 개별 소장 기관의 내부 시스템은 검토 범위에 포함되지 않았으며, 메타데이터의 내용적 정확성에 대한 검증은 수행하지 않았다. 이는 본 연구의 한계이자 후속 연구의 과제이다.

그러나 이러한 한계에도 불구하고, 본 연구가 수행한 실증 진단은 단순한 현황 기록에 그치지 않는다. KORCIS의 구조적 결함을 항목별로 규명하고 해외 사례와의 격차를 수치로 확인한 이 작업은, 한국 고문헌의 서지적 특수성을 반영한 RDF 기반 온톨로지 설계를 위한 실질적 요건 목록으로 기능한다. 후속 연구에서 설계할 KoDEX (Korean Old Document Extended schema)는 한국 고문헌의 서지적 특수성을 반영한 RDF 기반 온톨로지, 별도 후속 논문에서 상세히 다룬다. KoDEX는 본 연구에서 도출된 미충족 항목-영구 식별자 부재, 국제 전자 연계 전무, 표준 어휘 미적용, 라이선스 정보 부재-을 직접적인 설계 근거로 삼는다. KORMARC의 고문헌 특화 필드를 RDF 클래스 및 속성으로 변환하고, KAC·VIAF·Wikidata URI를 owl:sameAs로 연결하며, BIBFRAME 2.0 및 Europeana EDM과의 개념적 정렬을 통해 국제 서지 표준과의 호환성을 확보하는 것이 핵심 목표이다. 본 연구는 그 설계의 출발점이자 논거적 토대로서, 고문헌 메타데이터의 LOD 전환이라는 장기 과제의 첫 단계에 해당한다.

참고문헌

가. 국내 문헌

- 강순애 (2014). 문서 목록규칙 표준화(안) 연구. 국립중앙도서관.
국립중앙도서관 (발행년불명a). 국가서지 링크드 오픈 데이터(LOD) 서비스. 출처: <https://lod.nl.go.kr>
국립중앙도서관 (발행년불명b). 한국고문헌종합목록(KORCIS). 출처: <https://www.nl.go.kr/korcis>
김영진 (2024). 한국 고문헌 목록의 현황과 문제점, 향후 방향. *대동문화연구*, 128, 237-274.
<https://doi.org/10.18219/ddmh..128.202412.237>
김현 (2016). 디지털 인문학이란 무엇인가. 한국학중앙연구원 디지털인문학연구소.
출처: <http://digerati.aks.ac.kr/paper/I2DH/I2DH0101.pdf>
박민정, 이승민 (2023). 고문헌 기술을 위한 LRM 기반 서지구조 구축: 에이전트, 장소, 시간 개체를 중심으로. *정보관리학회지*, 40(3), 197-219. <https://doi.org/10.3743/KOSIM.2023.40.3.197>
박지영 (2014). 링크드 데이터 환경에서의 서지기술형식 BIBFRAME과 그 활용에 대한 고찰. *한국비블리아학회지*, 25(4), 235-263.
<https://doi.org/10.14699/kbiblia.2014.25.4.235>
박하람, 강지우, 송채은, 안지은, 이정윤, 장소정, 조하연, 차화준, 김학래 (2024). FAIR 데이터 원칙을 적용한 국내 디지털 아카이브 평가와 개선 방향. *한국기록관리학회지*, 24(4), 113-134. <https://doi.org/10.14404/JKSARM.2024.24.4.113>
박하람, 김학래 (2023). 디지털 기록의 상호운용을 위한 지식그래프의 평가. *한국기록관리학회지*, 23(4), 159-178.
<https://doi.org/10.14404/JKSARM.2023.23.4.159>
이미화, 배정현 (2025). BIBFRAME 구축을 위한 국가도서관의 실행방안에 관한 연구. *한국도서관·정보학회지*, 56(2), 87-108.
<https://doi.org/10.16981/kliss.56.2.202506.87>
임소정, 손계영 (2023). 대학도서관 고문헌 디지털화 현황과 활성화 방안. *서지학연구*, 94, 145-172.
<https://doi.org/10.17258/jib.2023..94.145>
장엽 (2022). 중국 내 고전 문헌의 디지털화 현황 및 전망. *고전문학과 교육*, 49, 45-93.
허철 (2024). 한국 고문헌 디지털 데이터 관련 연구 발전을 위한 제언. *개념과 소통*, 34, 7-38.
<https://doi.org/10.15797/concom.2024..34.001>

나. 일본 문헌

- 国立国会図書館 (n.d.). 国立国会図書館典拠データ検索・提供サービス(Web NDL Authorities). Available:
https://www.ndl.go.jp/jp/data/data_service/authorities.html

다. 영문 문헌

- Barbuti, N. (2021). Thinking digital libraries for preservation as digital cultural heritage: By R to R4 facet of FAIR principles. *International Journal on Digital Libraries*, 22, 309-318. <https://doi.org/10.1007/s00799-020-00291-7>
Europeana (n.d.a). Europeana Data Model (EDM). Available: <https://pro.europeana.eu/page/edm-documentation>
Europeana (n.d.b). Linked Open Data. Available: <https://pro.europeana.eu/page/linked-open-data>
Isaac, A. (2013). Europeana Data Model Primer. Europeana. Available:
https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
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>

라. 웹사이트

공공데이터포털 (n.d.). 한국고문헌종합목록시스템(KORCIS) 서지데이터. 출처: <https://www.data.go.kr>
 Library of Congress (n.d.). LC Name Authority File (LCNAF). Available: <https://id.loc.gov/authorities/names.html>
 Wikidata (n.d.). Available: <https://www.wikidata.org>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Heo, Cheol (2024). Suggestions for the development of research on Korean classical document digital data. *Concept and Communication*, 34, 7-38. <https://doi.org/10.15797/concom.2024..34.001>
- Im, Sojeong & Son, Gyeyeong (2023). Current status and activation plans for digitization of classical documents in university libraries. *Journal of Bibliography*, 94, 145-172. <https://doi.org/10.17258/jib.2023..94.145>
- Jang, Yeom (2022). Current status and prospects of digitization of classical literature in China. *Classical Literature and Education*, 49, 45-93.
- Kang, Sunae (2014). A Study on the Standardization of Document Cataloging Rules. National Library of Korea.
- Kim, Hyeon (2016). What is Digital Humanities? Academy of Korean Studies, Digital Humanities Research Institute. Available: <http://digerati.aks.ac.kr/paper/I2DH/I2DH0101.pdf>
- Kim, Yeongjin (2024). Current status, problems, and future directions of Korean classical document cataloging. *Daedong Munhwa Yeongu*, 128, 237-274. <https://doi.org/10.18219/ddmh..128.202412.237>
- Lee, Mihwa & Bae, Jeonghyeon (2025). A study on the implementation plan of national libraries for BIBFRAME construction. *Journal of the Korean Society for Library and Information Science*, 56(2), 87-108.
<https://doi.org/10.16981/kliiss.56.2.202506.87>
- National Library of Korea (n.d.a). National Bibliography Linked Open Data (LOD) Service. Available: <https://lod.nl.go.kr>
- National Library of Korea (n.d.b). Korean Classical Document Integrated System (KORCIS). Available:
<https://www.nl.go.kr/korcis>
- Park, Haram & Kim, Hakrae (2023). Evaluation of knowledge graphs for the interoperability of digital records. *Journal of Korean Society of Archives and Records Management*, 23(4), 159-178.
<https://doi.org/10.14404/JKSARM.2023.23.4.159>
- Park, Haram, Kang, Jiwoo, Song, Chaeun, An, Jieun, Lee, Jeongyun, Jang, Sojeong, Jo, Hayeon, Cha, Hwajun, & Kim, Hakrae (2024). Evaluation and improvement of domestic digital archives applying FAIR data principles. *Journal of Korean Society of Archives and Records Management*, 24(4), 113-134. <https://doi.org/10.14404/JKSARM.2024.24.4.113>
- Park, Jiyeong (2014). A study on BIBFRAME as a bibliographic description format in the linked data environment and its application. *Journal of the Korean Biblia Society for Library and Information Science*, 25(4), 235-263.
<https://doi.org/10.14699/kbiblia.2014.25.4.235>
- Park, Minjung & Lee, Seungmin (2023). Building an LRM-Based bibliographic structure for classical document description: focusing on agent, place, and time entities. *Journal of the Korean Society for Information Management*, 40(3), 197-219.
<https://doi.org/10.3743/KOSIM.2023.40.3.197>