

한국 고문헌 메타데이터의 시맨틱 표준화 연구

FAIR 원칙 진단에서 KoDEX(Korean Old Document Extended Schema) 온톨로지 설계,
LOD 구현까지

신은선

한국학중앙연구원 인문정보학 박사과정생

2026년 8월 28일

목차

I. 연구의 배경 및 목적

1. 연구 배경
2. 문제 제기
3. 연구 목적 및 연구 질문

II. 선행연구 검토

1. 서지 온톨로지 및 LOD 표준화 연구
2. 한국 고문헌 디지털화 및 메타데이터 연구
3. 선행연구 종합 및 연구의 위치

III. 박사학위논문 전체 연구 설계

1. 3단계 연구 설계 : 진단 – 설계 – 구현
2. 연구 간 연계 논리

IV. 세부 연구 내용

1. 지금까지 수행한 연구 : 메타데이터 진단에서 온톨로지 설계까지
2. 앞으로 수행할 연구 : KORMARC-RDF 변환 가이드라인 및 LOD 구현 방안

V. 연구방법 및 범위 종합

VI. 향후 연구 일정

VII. 연구의 의의 및 기대효과

참고문헌

I. 연구의 배경 및 목적

1. 연구 배경 : 한국 고문헌 메타데이터와 디지털 인문학

한국 고문헌은 유교적 통치 철학, 역사적 기록, 사상적 정수가 집약된 지식 자산으로서 국내외 학술 연구의 핵심 1차 자료로 활용되고 있다. 디지털 기술의 발전과 함께 문화유산의 디지털 자산화 및 온라인 서비스가 확대되면서, 디지털 인문학(Digital Humanities)은 인문 지식이 디지털 환경에서 보다 의미 있게 탐구되고 가치 있게 활용되도록 하는 것을 과제로 삼고 있다(김현, 2016). 이러한 흐름 속에서 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 조건이 된다.

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 56만여 건의 서지 데이터와 6만 5천여 건의 원문 이미지를 제공하는 국가 단위의 서지 인프라로 기능하고 있다(국립중앙도서관, n.d.). 이는 한국 고문헌 메타데이터를 논의할 때 가장 먼저 검토되어야 할 국가적 규모의 인프라라 할 수 있다.

2. 문제 제기 : KORCIS의 구조적 한계

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 연계가 부족하여 글로벌 지식 네트워크와의 상호운용성 확보에 어려움을 겪고 있다(이미화·배정현, 2025). 개별 소장 기관별로 분산 관리되는 구조와 데이터 표준·기술 형식의 불일치는 통합적 검색과 고도화된 데이터 활용을 저해하는 요인으로 지적되어 왔다(임소정·손계영, 2023).

전 세계적으로는 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 인간뿐 아니라 컴퓨터가

데이터를 자동으로 처리·연계할 수 있는 기계가독성(machine-actionability)을 강조한다. 그렇다면 KORCIS는 실제로 이러한 국제적 기준을 얼마나 충족하고 있는가 - 이 물음에 대한 실증적 답을 구하는 것이 본 연구 전체의 출발점이다.

3. 연구 목적 및 연구 질문

본 연구는 한국 고문헌 메타데이터를 국제 시맨틱 웹 환경에서 발견(Findable)·접근(Accessible)·연계(Interoperable)·재사용(Reusable) 가능하게 하기 위한 방안을 진단·설계·구현의 세 단계에 걸쳐 규명하는 것을 총괄 목적으로 한다. 이를 구체화한 세 가지 연구 질문은 다음과 같다.

- 연구질문 1 (진단): 현재 KORCIS 메타데이터는 FAIR 원칙을 얼마나 충족하고 있으며, 무엇이 부족한가?
- 연구질문 2 (설계): 국제 표준 어휘와 한국 고문헌의 서지적 특수성을 함께 담아내는 온톨로지는 어떻게 설계하는가?
- 연구질문 3 (구현): 설계된 온톨로지로서 실제 데이터를 변환하고, 이를 LOD 서비스로 구현하는 방법은 무엇인가?

II. 선행연구 검토

선행연구는 크게 두 영역으로 구분하여 검토한다. 서지 온톨로지 및 LOD 기반 표준화 연구, 그리고 한국 고문헌 디지털화 및 메타데이터 연구가 그것이다.

1. 서지 온톨로지 및 LOD 표준화 연구

서지 온톨로지 분야에서 BIBFRAME 2.0은 미국 의회도서관(LC)이 주도하여 기존 MARC 기반 서지 데이터를 RDF로 전환하기 위해 개발한 온톨로지로서, 저작(Work)·인스턴스(Instance)·개별자료(Item)의 3계층 구조를 통해 서지 개체 간의 관계를 RDF 트리플로 명시한다(Library of Congress, n.d.). 국내 BIBFRAME 연구는 2014년 이후 지속적으로 축적되어 왔다. 박옥남·오정선(2014)은 BIBFRAME의 구조와 MARC21-BIBFRAME 어휘 간 매핑을 분석하여 국내 도서관 LOD 전환에의 적용 가능성을 검토하였다. 이미화(2017)는 BIBFRAME 2.0의 특징을 변경사항 중심으로 분석하고 국내 특화 기술요소를 반영한 응용프로파일 개발의 필요성을 제기하였으며, 이미화·배정현(2025)은 LC와 스웨덴 국립도서관의 구축 사례를 분석하여 KORMARC이 장기간 유지될 가능성을 전제로 한 국내 적용 방안을 구체적으로 제시하였다.

CIDOC-CRM은 박물관·아카이브 분야의 문화유산 정보 표현을 위한 참조 온톨로지로서, 사건(Event) 중심의 기술 방식이 고문헌의 간행·전래·소장 이력을 표현하는 데 적합하다(Doerr, 2003). Europeana는 EDM(Europeana Data Model)을 통해 Dublin Core, OAI-ORE, SKOS, OWL 등 기존 국제 표준 어휘를 결합하는 방식으로 이질적인 기관 데이터를 통합하고 있으며, 이는 독자적 표준 개발보다 기존 어휘의 재사용이 상호운용성 확보에 효과적임을 보여주는 사례이다(Isaac, 2013).

국내에서는 BIBFRAME의 실제 적용을 위한 세부 방법론 연구도 최근 활발히 이루어지고 있다. 박진호(2025)는 BIBFRAME이 재사용하는 OWL 속성의 특성을 분석하여 객체 속성(object property) 세분화 기준을 제시하였으며, 송민선·이미화·박진호(2026)는 MODS 레코드를 BIBFRAME으로 변환하기 위한 구체적 매핑 규칙을 도출하였다. 이미화·박진호·이종욱·송민선(2025)은 KORMARC을 BIBFRAME으로 변환할 때 고려해야 할 사항을 체계적으로 정리하였고, 박진호·송민선·이미화(2026)는 BIBFRAME 도입 과정에 필요한 내규 기반 조직 거버넌스 설계를 다루었다. 이들 연구는 속성 세분화, 변환 규칙, 조직 거버넌스 등 BIBFRAME 도입의 실무적·기술적 절차를 정교화하는 데 초점을 두는 반면, 본 연구가 설계하는 KoDEX(Korean Old Document Extended Schema)는 이러한 일반 BIBFRAME 적용 논의를

전제로 하되 한국 고문헌이라는 특정 자료군의 물리 기술 요소와 외부 전거 연계에 특화된 확장 어휘를 다룬다는 점에서 차별화된다.

2. 한국 고문헌 디지털화 및 메타데이터 연구

박민정·이승민(2023)은 LRM(Library Reference Model) 기반의 서지구조를 고문헌에 적용하여 에이전트·장소·시간 개체를 중심으로 한 관계 표현 방안을 제안하였다. 임소정·손계영(2023)은 국내 대학도서관 고문헌 디지털화 현황을 분석하여 기관 간 표준 불일치와 메타데이터 품질 격차를 지적하였으며, 강순애(2014)는 KORMARC 기반 고문헌 서지 기술의 체계화 방향을 논의하였다. 허철(2024)은 한국 고문헌 디지털 데이터의 연구 발전을 위한 과제를 포괄적으로 검토하며 국제 표준 정렬의 필요성을 제기하였다. 김영진(2024)은 KORCIS의 구조적 문제점을 목록학적 관점에서 실증적으로 지적하고, 한자 검색을 저조·동일서명 통합 검색 기능 부재 등을 개선 과제로 제시하였다.

3. 선행연구 종합 및 연구의 위치

연구 공백

국제 표준 어휘를 활용한 서지 LOD 연구, BIBFRAME의 국내 실무 적용을 위한 방법론 연구, 그리고 한국 고문헌 디지털화 연구는 각각 축적되어 왔으나, 한국 고문헌의 서지적 특수성을 반영한 RDF 온톨로지를 독자적으로 설계한 연구는 아직 충분히 이루어지지 않았다. 본 연구는 이 지점 - 실증적 진단에 기반한 온톨로지 설계와 그 구현 경로 - 을 연구 대상으로 삼는다.

Ⅲ. 박사학위논문 전체 연구 설계

1. 3단계 연구 설계 : 진단 - 설계 - 구현

본 박사학위논문은 하나의 총괄 연구 질문 - 한국 고문헌 메타데이터를 국제 시맨틱 웹 환경에서 발견·접근·연계·재사용 가능하게 하려면 무엇이 필요한가 - 에 대한 답을 세 편의 개별 논문으로 나누어 단계적으로 규명하는 구조를 취한다.

단계	연구 제목	진행 상태
진단	FAIR 원칙에 따른 한국 고문헌 메타데이터 관리체계 진단: KORCIS를 중심으로	완료 · 한국기록관리학회지 26(2), 93-111 게재(2026.5)
설계	한국 고문헌 메타데이터 시맨틱 표준화를 위한 KoDEX(Korean Old Document Extended Schema) 온톨로지 설계	최종본 작성중
구현	KoDEX 온톨로지 기반 KORMARC 데이터의 RDF 변환 가이드라인 및 Linked Open Data 구현 방안	목차 구상 단계 · 착수 예정

〈표〉 박사학위논문 3단계 연구 설계

세 편의 논문은 최종적으로 하나의 박사학위논문 「한국 고문헌 메타데이터의 시맨틱 표준화: 진단, 설계, 구현」으로 통합·확장될 예정이다.

2. 연구 간 연계 논리

각 단계는 이전 단계의 결과에서 직접 도출된다. 진단 연구는 KORCIS 서지 데이터 567,212건에 대한 FAIR 15개 세부 항목 준수도를 전수 분석하여, 적합 판정 항목이 하나도 없고 특히 상호운용성(I) 영역이 전반적으로 미충족 상태임을 실증하였다. 이 진단에서 도출된 핵심 문제 - HTTP URI 형식의 영구 식별자 부재, 국제 전거 시스템(VIAF·Wikidata)과의 연계 전무, RDF 기반 데이터 공개 미구현 - 가 이어지는 설계 연구의 설계 원칙(F1 영구식별자 부여, I1 공식 지식표현 언어 사용, I3 정규화된 외부참조)으로 직접 이어진다. 이렇게 설계한 KoDEX 온톨로지는 다시 마지막 구현 연구에서 KORMARC 실제 데이터의 RDF 변환 규칙과 LOD 서비스 구현의 근거 구조로 활용된다. 이처럼 세 단계는 독립된 주제가 아니라, 하나의 문제의식이 진단에서 설계, 구현으로 심화되는 단일한 연구 궤적을 이룬다.

이 연계는 최근 구체화되었다. 설계 연구는 설계안을 제시하는 데 그치지 않고, 『논어집주대전』(論語集註大全) 한 저작을 KORCIS 전체 서지데이터에서 전수 추출한 1,160건 코퍼스로 삼아 KoDEX 3단 구조(저작-판본-소장본)를 예비 검증하는 실증 적용 절을 자체적으로 포함하게 되었다. 이 예비 검증은 기관 내 중복 레코드, 낙질, 함철서 편입이라는 세 가지 확장 필요 지점을 데이터로 식별하고, 인물(Agent) 수준의 외부 전거 연계까지 실물 식별자로 예시하였다. 이어질 구현 연구는 이 단일 저작 심층 사례에서 확인된 매핑 프로파일과 확장 필요 속성을 출발점으로 삼아, 대상 범위를 여러 유형의 고문헌으로 넓히고 실제 RDF 변환·LOD 서비스 구현으로 완성하는 단계이다.

IV. 세부 연구 내용

1. 지금까지 수행한 연구 : 메타데이터 진단에서 온톨로지 설계까지

지금까지는 두 단계의 연구를 순차적으로 수행하였다. 먼저 **KORCIS** 서지 데이터 전체를 대상으로 **FAIR** 원칙 준수 실태를 실증적으로 진단하였고, 이 진단 결과를 출발점으로 삼아 한국 고문헌에 특화된 **RDF** 온톨로지 **KoDEX**를 설계한 뒤 단일 저작 사례로 예비 검증까지 마쳤다. 두 연구의 구체적인 내용은 다음과 같다.

가. **FAIR** 원칙에 따른 **KORCIS** 메타데이터 진단

「**FAIR** 원칙에 따른 한국 고문헌 메타데이터 관리체계 진단 - **KORCIS**를 중심으로 -」

연구 목적 및 방법

KORCIS 서지 데이터 567,212건 전체를 대상으로 **FAIR** 원칙 준수 현황을 실증적으로 진단하고 개선 방향을 제시한 연구이다. 분석은 ① **KORMARC** 주요 필드 출현빈도·완성도 측정, ② 전거 제어 및 식별자 체계 실태 분석, ③ **FAIR** 15개 세부 항목별 준수도 평가의 3단계로 수행하였으며, 필드 유형(필수·해당시필수·재량) 구분과 데이터 구조적 특성·시스템 기능 지원 여부·정량 분석 결과를 종합하여 판정하였다.¹

자료 비판: **KORCIS**는 어떻게 형성되었는가

KORCIS는 처음부터 단일 기관이 일관된 원칙에 따라 구축한 균질적 데이터셋이 아니다. 2007년 서비스 개시 이후 지속적으로 확충되어 2025년 기준 국내 91개·국외 53개, 총 144개 참여기관이 각자의 목록 환경과 인력 수준에 따라 독립적으로 생성한 **KORMARC** 레코드를, 국립중앙도서관이 표준 인터페이스로 수집·색인하여 통합 검색을 제공하는 연합형 통합목록(**union catalog**)이다(국립중앙도서관, n.d.). 이 형성 방식은 최소 두 개의 서로 다른 층위를 남긴다. 하나는 개별 참여기관이 자체 목록 규칙에 따라 레코드를 생성하는 원 소장기관 목록층이며, 다른 하나는 그 레코드를 사후적으로 수집·통합하는 **KORCIS** 통합 수집층이다. 두 층위 사이에 목록 규칙의 사전 통일이나 사후 정합성 검증이 이루어졌다는 근거는 **KORCIS**의 공개 문서에서 확인되지 않는다.

이 이질적 형성 구조는 본 진단의 결과에서 이미 그 흔적을 드러낸다. 필수 서지 필드의 완성도는 99% 이상으로 기관을 막론하고 균질한 반면, 선택 필드의 완성도는 항목에 따라 1.49%에서 38.86%까지 큰 편차를 보였다. 이는 단순한 입력 누락이라기보다, 참여기관마다 적용하는 목록

¹ 분석결과 및 판정 근거 : <https://github.com/SHINEUNSUN1027/korcis-fair-analysis>

상세화 수준(level of cataloging)과 서지 기술 관행이 애초에 다르다는 사실을 시사한다. 이 이질성의 구체적인 양상 - 동일한 소장 세트를 서로 다른 서지 단위로 분할해 기술하는 기관별 관행의 차이 - 은 이 장 뒷부분의 『논어집주대전』 실증 사례에서 청구기호 기준 재집계를 통해 보다 구체적으로 확인된다(나. 참조). KORCIS를 하나의 평평한(flat) 데이터셋으로 전제하고 완성도만을 측정하는 접근은 이 자료의 형성 과정 자체가 남긴 구조적 이질성을 간과할 위험이 있으며, 본 연구는 이 점을 진단 결과의 해석 틀에 명시적으로 반영하고자 하였다.

실제로 소장 건수 상위 10개 기관을 대상으로 주요 선택 필드의 출현율을 비교하면, 이러한 이질성이 수치로도 뚜렷하게 드러난다. 기본표목-개인명(100) 필드의 출현율은 기관에 따라 0.0%에서 91.8%까지 극단적인 편차를 보이며, 주제 접근 필드의 활용 양상도 기관마다 상반된다. 서울대학교 규장각한국학연구원은 일반주제명(650)을 99.7% 활용하는 반면 비통제 색인어(653)는 전혀 사용하지 않는데, 고려대학교 도서관은 정반대로 653을 99.6% 활용하면서 650은 0%에 그친다. 이는 개별 기관이 활용할 수 있는 통제 어휘 인프라의 차이가 서로 다른 입력 관행으로 나타난 결과로 해석된다.

기관명	총 건수	100(저자)	250(판본)	650(주제)	653(색인어)	700(부출저자)	856(URL)
국립중앙도서관	104,957	37.2%	90.5%	22.8%	93.4%	45.9%	2.4%
계명대학교 동산도서관	72,271	64.5%	99.9%	0.0%	34.6%	52.5%	1.8%
전남대학교 도서관	30,887	91.8%	100.0%	78.8%	99.7%	16.6%	3.8%
서울대학교 규장각한국학연구원	27,611	0.0%	99.4%	99.7%	45.6%	0.0%	0.0%
동국대학교 중앙도서관	25,279	57.0%	98.8%	83.3%	92.3%	70.8%	8.2%
한국학진흥원 도서관	22,659	0.0%	90.9%	0.0%	0.0%	49.7%	0.0%
고려대학교 도서관	22,327	68.9%	97.6%	0.0%	99.6%	41.6%	0.6%
영남대학교 도서관	20,273	0.7%	56.1%	0.0%	0.2%	77.7%	0.0%
충남대학교 도서관	15,103	40.5%	99.9%	68.0%	45.1%	60.1%	2.2%
대구가톨릭대학교 도서관	14,365	60.5%	33.3%	0.1%	100.0%	28.7%	0.0%

〈표〉 상위 10개 기관별 주요 필드 출현율

주요 결과

필수 서지 필드(서명저자사항 245, 형태사항 300, 간사사항 260 등)의 완성도는 99% 이상으로 높은 수준이나, 선택 필드(기본표목-개인명 100: 38.86%, 일반주제명 650: 27.35%, 전자적 위치 및 접속 856: 1.49%)의 완성도는 현저히 낮았다. 전거 제어 실태는 더욱 심각하여, 인명 필드 보유 레코드 중 전거레코드 제어번호가 기재된 비율은 1% 미만에 그쳤고 VIAF·Wikidata·LCNAF 등 국제 전거 시스템 URI는 단 1건도 확인되지 않았다. FAIR 15개 세부 항목 평가 결과 적합 0개(0%), 부분 적합 5개(33.3%), 미흡·결여 10개(66.7%)로 나타났으며, Europeana·일본 국립국회도서관(NDL)·중국 중화고적자원고와의 비교를 통해 RDF 기반 온톨로지 전환, 국제 전거 연계, 라이선스 체계 정비가 시급한 과제임을 논증하였다.

FAIR 원칙 준수도 세부 평가

FAIR 15개 세부 항목을 항목별로 평가한 결과는 다음과 같다. 발견 가능성(F) 4개 항목 중 3개가 미흡·결여, 1개가 부분 적합이며, 접근 가능성(A) 4개 항목 중 2개가 미흡·결여, 2개가 부분 적합이다. 상호운용성(I) 3개 항목은 전부 미흡·결여이며, 재사용 가능성(R) 4개 항목 중 2개가 미흡·결여, 2개가 부분 적합이다. 특히 상호운용성(I) 영역의 전면적 미충족은 RDF와 같은 공식 지식 표현 언어를 사용하지 않고(I1), FAIR 준수 표준 어휘도 적용되지 않으며(I2), 전거번호를 통한 외부 연계가 전체 레코드의 1% 미만으로 사실상 부재한(I3) 세 가지 결여가 겹친 결과다.

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%

영역	항목	평가	근거
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 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

〈표〉 KOCIS FAIR 원칙 준수도 평가 결과

해외 주요 기관과의 비교

Europeana, 일본 국립국회도서관(NDL), 중국 중화고적자원고와 KOCIS를 FAIR 15개 항목 기준으로 비교하면 다음과 같다(신은선, 2026). 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 KOCIS는 텍스트 기반 서지 기술에 머물러 있어 상호운용성(I) 영역 전반에서 미충족 상태다.

항목	Europeana	NDL	중화고적자원고	KOCIS
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 인덱스 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 허용 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

항목	Europeana	NDL	중화고적자원고	KORCIS
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

〈표〉 해외 주요 기관 FAIR 원칙 준수도 비교

이 비교는 KORCIS의 시맨틱 표준화를 위한 세 가지 선결 과제 - ① HTTP URI 기반 영구 식별자를 서지 레코드와 전거 개체에 부여하는 것, ② 전거 데이터를 우선적으로 RDF 형식으로 공개하고 SPARQL 엔드포인트를 구축하는 단계적 전환 경로, ③ 국내 전거 식별자(KAC)를 VIAF·Wikidata URI와 owl:sameAs로 연결하는 것 - 를 시사하며, 이는 앞서 제시한 KoDEX의 설계 원칙과 직접 연결된다.

목차 (진단 연구)

1. 서론

1.1 연구 배경 및 목적

1.2 선행연구

1.3 연구 방법 및 범위

2. FAIR 원칙과 문화유산 메타데이터

2.1 FAIR 원칙의 개요 및 15개 세부 항목

2.2 문화유산·도서관 분야 적용 동향

3. KORCIS 메타데이터 현황 진단

3.1 KORCIS 개요 및 구조적 특성

3.2 데이터 수집 및 분석 방법

3.3 메타데이터 필드 완성도 및 기관별 품질 분석

3.4 FAIR 원칙 준수도 평가

4. 해외 고문헌 메타데이터 시맨틱 표준화 사례

4.1 Europeana

4.2 일본 NDL

4.3 중국 중화고적자원고 /

4.4 비교 분석 및 시사점

5. 결론

나. KoDEX 온톨로지 설계 및 예비 실증 검증

「한국 고문헌 메타데이터 시맨틱 표준화를 위한 KoDEX(Korean Old Document Extended Schema) 온톨로지 설계

: 국제 표준 어휘 재사용과 고문헌 특화 확장 어휘 체계 구축을 중심으로」

연구 목적

앞선 진단 연구의 결과를 출발점으로 삼아, 한국 고문헌 메타데이터의 시맨틱 표준화를 위한 RDF 기반 온톨로지 KoDEX(Korean Old Document Extended Schema)를 설계하는 것을 목적으로 한다. 세부적으로는 ① HTTP URI 기반 영구 식별자 체계 수립, ② BIBFRAME 2.0을 주축으로 SKOS·FOAF·CIDOC-CRM을 보완하는 혼합 어휘 체계 확립, ③ 한국 고문헌의 서지적 특수성을 반영한 핵심 클래스·속성 정의 및 물리기술 특화 확장 어휘의 OWL 형식화, ④ VIAF·Wikidata·국립중앙도서관 전거 시스템과의 외부 전거 연계 구조 설계, ⑤ 단일 저작 심층사례를 통한 설계안의 예비 실증 검증을 목표로 한다.

설계 원칙

- 국제 표준 어휘 재사용 우선 : BIBFRAME 2.0·SKOS·FOAF 등을 최대한 재사용하고, 반드시 필요한 경우에 한해 확장 어휘를 정의
- FAIR 원칙 정렬 : F1(영구식별자), I1(공식 지식표현 언어), I3(정규화된 외부참조) 등을 설계 원칙으로 내재화
- 한국 고문헌 특수성 반영 : 광곽·어미·판구·계선 등 물리적 특징, 판본 유형, 저자 미상 등 불확실 정보 표현을 위한 전용 어휘
- 개방성 및 확장성 : OWL로 형식화하여 새로운 고문헌 기술 항목·국제 표준 변화에 대응해 확장 가능한 구조
- 단계적 구현 가능성 : 전면적 시스템 재구축 없이 KORCIS KORMARC 데이터를 단계적으로 RDF 전환할 수 있도록 설계

설계 원칙의 이론적 근거

① 국제 표준 어휘 재사용을 우선하는 것은 온톨로지 공학에서 상호운용성을 확보하는 핵심 전략으로 받아들여진다. 새 어휘를 독자적으로 만들 때마다 기존 시스템과의 매핑 비용과 유지보수 부담이 발생하며, 이런 개별 어휘가 누적되면 오히려 상호운용성을 저해하는 "표준의 난립"으로 이어진다는 것이 **Europeana**와 같은 대규모 통합 사례가 보여주는 공통된 교훈이다(Isaac, 2013). 진단 연구에서 **KORCIS**의 국제 전거 시스템 연계가 전무하다는 사실이 확인된 만큼(VIAF·Wikidata·LCNAF URI 0건), **KoDEX**는 처음부터 광범위하게 검증된 국제 어휘를 기반으로 삼아 이 공백을 근본적으로 해소하고자 하였다.

② **FAIR** 원칙 정렬은 진단 연구의 결과를 직접 반영한다. **FAIR** 15개 세부 항목 중 적합 판정이 하나도 없었고 특히 상호운용성(I) 영역의 결여가 두드러졌으므로, 설계 단계에서부터 **F1**(전역적으로 고유하고 영구적인 식별자 부여), **I1**(공식적이고 접근 가능하며 공유되고 널리 적용 가능한 지식표현 언어의 사용), **I3**(정규화된 외부참조 포함)을 명시적 설계 요건으로 내재화하였다. 이는 완성된 시스템을 사후에 **FAIR** 기준으로 점검하는 것이 아니라, 설계 시점부터 **FAIR**를 구조적 요구사항으로 채택하는 접근이다.

③ 한국 고문헌 특수성 반영이 필요한 이유는 재사용 대상 어휘들의 태생적 한계에 있다. **BIBFRAME 2.0**과 **CIDOC-CRM**은 서구의 인쇄본·필사본·박물관 자료를 주요 준거로 발전해 왔기 때문에, 광곽(匡郭)·어미(魚尾)·판구(版口)·계선(界線) 등 동아시아 고문헌 특유의 물리적 구성요소나 판본 유형의 구분, 저자 미상·간행연도 불명과 같은 불확실 정보의 체계적 표현에는 한계를 지닌다. 따라서 국제 표준의 전면 재사용이 아니라 재사용을 전제로 한 선택적 확장이 **KoDEX** 설계의 핵심 전제가 되었다.

④ 개방성 및 확장성을 설계 원칙으로 둔 것은, 설계 당시 예상하지 못한 요구사항이 실제 데이터에 적용하는 과정에서 계속 발견되리라는 전제에서 출발한다. 이 전제는 이후 실증 적용에서 실제로 확인되었다. 단일 저작 심층사례를 검증하는 과정에서만도 **kodex:possibleDuplicateOf** 등 세 가지 확장 속성의 필요성이 새로 드러났다(아래 실증 적용 주요 결과 참조). 클래스·속성 정의를 처음부터 완결된 것으로 고정하지 않고 점진적으로 확장 가능한 모듈형 구조로 설계한 것은 이 때문이다.

⑤ 단계적 구현 가능성을 고려한 것은 **KORCIS**가 이미 56만여 건의 서지 데이터를 운영 중인 국가 인프라라는 현실적 제약 때문이다. 전면적 재구축은 사실상 불가능하므로, **KoDEX**는 기존 **KORMARC** 필드와의 명시적 매핑 관계를 각 속성 정의에 포함시켜 전체 시스템을 한 번에 교체하지 않고도 레코드 단위·기관 단위의 점진적 **RDF** 전환이 가능하도록 설계하였다. 이는

앞으로 수행할 구현 연구가 특정 기관이나 특정 문헌 유형부터 시작해 범위를 넓혀가는 전략을 취할 수 있는 기반이 된다.

연구 방법 및 범위

연구는 ① 문헌 검토 및 표준 어휘 분석(BIBFRAME 2.0·SKOS·CIDOC-CRM·FOAF·Schema.org 등 검토), ② KoDEX 온톨로지 설계(핵심 클래스·속성 정의, 확장 어휘 OWL 형식화, 외부 전거 연계 구조 설계), ③ 실증 적용(KORCIS 전체 서지데이터에서 『논어집주대전』 관련 레코드 1,160건을 전수 추출해 KoDEX 3단 구조를 예비 검증하고, 설계 단계에서 드러나지 않은 확장 필요 지점을 데이터로 식별)의 3단계로 수행한다. 연구 범위는 KORCIS 서지 데이터에 수록된 고문헌 메타데이터로 한정하며, 디지털 이미지·원문 텍스트 데이터의 표현은 포함하지 않는다. 실증 적용은 『논어집주대전』이라는 단일 저작에 대한 예비적 검증이며, 여러 유형의 고문헌으로 범위를 넓힌 전체 규모의 KORMARC 변환 규칙 수립과 실제 LOD 서비스 구현 방안은 이후 수행할 구현 연구의 과제로 이관한다.

목차 (설계 연구)

1. 서론

1.1 연구 배경 및 목적

1.2 선행연구

1.3 연구 방법 및 범위

2. 관련 표준 어휘 및 온톨로지 검토

2.1 서지 온톨로지: BIBFRAME 2.0과 BIBO

2.2 지식조직 어휘: SKOS

2.3 문화유산 온톨로지: CIDOC-CRM

2.4 인물·기관 기술: FOAF와 Schema.org

2.5 검토 결과 종합 및 KoDEX 설계에의 시사점

3. KoDEX 온톨로지 설계

3.1 설계 원칙 및 네임스페이스

3.2 핵심 클래스 정의

3.3 속성 정의: 국제 표준 어휘 재사용 및 확장

3.4 한국 고문헌 특화 확장 어휘

3.5 외부 전거 연계 설계

4. 실증 적용: 논어집주대전 사례

4.1 사례 데이터 및 추출 방법

4.2 저작-판본-소장본 매핑: 1820년 내각장판본 사례

4.3 잔여 정보(residue)의 정량화

4.4 실증 적용의 시사점과 한계

5. 결론

위와 같이 설계를 마친 뒤, 이 설계가 실제 데이터 위에서도 작동하는지를 확인하기 위해 『논어집주대전』 코퍼스에 대한 예비 실증 적용을 함께 수행하였다. 그 결과는 다음과 같다.

사례 데이터 및 추출 방법 [X 논어집주대전_전체서지레코드_KORCIS.xlsx]

실증 적용에 사용한 『논어집주대전』 관련 레코드는 두 가지 공개 데이터 출처에서 추출하였다. 첫째는 공공데이터포털이 제공하는 「한국고전종합목록 서지데이터」로 KORCIS 서지 데이터 567,212건 전체를 담고 있으며, 둘째는 문화체육관광부·국립중앙도서관이 제공하는 「한국고문헌종합목록 공공도서관 고문헌 서지데이터」(2025년 7월 31일 기준, 7,170건)이다. 두 출처의 MARC 001~740 필드 전체를 대상으로 "論語集註大全"(정자)·"论语集注大全"(간체)·"논어집주대전"(한글) 문자열을 검색하였다. 이때 단순 문자열 검색만으로는 호환용 한자 이체자(예: U+F941 論)로 입력된 레코드가 다수 누락되므로, 유니코드 NFKC 정규화를 적용해 재검색하는 절차를 병행하였다. 이 절차를 통해 KORCIS에서 1,150건, 공공도서관 고문헌 데이터에서 10건, 총 1,160건을 추출하였다.

추출된 레코드는 매치 유형에 따라 두 가지로 구분된다. 245 필드의 서명에 논어집주대전이 직접 기재된 '제목(245) 직접일치'가 1,117건(96.3%)이며, 서명은 『어제성리사서대전(御製性理四書大全)』과 같은 다른 합철서이지만 500 주기 또는 740 관련서명 필드에 논어집주대전이 부기된 '주기·추가표목 등 간접일치'가 43건(3.7%)이다. 간접일치 레코드는 서명만으로는 포착되지 않는 이본·합철·이명 관계를 확인할 수 있게 해준다는 점에서 별도로 구분해 관리할 가치가 있다.

매치유형	건수	비율
제목(245) 직접일치	1,117건	96.3%
주기·추가표목 등 간접일치	43건	3.7%
합계	1,160건	100.0%

〈표〉 매치유형 분포

70개 소장기관 가운데 소장 건수 상위 30개 기관의 분포는 다음과 같다.

소장기관	건수	소장기관	건수
계명대학교 동산도서관	256	성균관대학교 존경각	24
국립중앙도서관	87	연세대학교 학술정보원	23
전남대학교 도서관	57	경상국립대학교 도서관	20
대구가톨릭대학교중앙도서관	52	용인대학교 중앙도서관	20
고려대학교 도서관	48	원광대학교 도서관	17
경기대학교 도서관	47	충남대학교 도서관	16
영남대학교 도서관	44	안동대학교 도서관	14
국립중앙박물관 도서관	43	건국대학교 상허기념도서관	14
숙명여자대학교 도서관	39	경희대학교 도서관	12
동국대학교 중앙도서관	34	단국대학교 퇴계기념도서관	12
부산대학교 도서관	31	사우당종택	12
한양대학교 백남학술정보관	28	한밭도서관	10
서울대학교 규장각한국학연구원	28	국립중앙도서관 고문헌과(수탁)	10
동신대학교 도서관	27	춘호재	8
동국대학교 경주캠퍼스 도서관	24	단국대학교 율곡기념도서관	7

〈표〉 소장기관별 분포 상위 30개 기관(전체 70개 기관 중)

판사항(250 필드)을 세분화하여 살펴보면, 앞서 제시한 목판본·활자본·필사본의 대분류 안에서도 표기가 다양하게 나뉜다. 특히 목판본 계열 안에서만도 중간(重刊)·후쇄(後刷)·복각판(覆刻版)·번각(飜刻) 등 간행 이력을 나타내는 세부 표기가 병존하여, 동일 대분류로 묶이는 레코드들 사이에도 서지적으로 구분되는 정보가 담겨 있음을 보여준다. 상위 20개 표기는 다음과 같다.

판사항	건수	판사항	건수
木板本	690	木板本(後刷)	13
木版本	79	木板本(丁酉字翻刻)	7
(미기재)	67	活字本	7
筆寫本	45	木板本(乙丑重刊)	7
木板本(重刊)	23	木板本(嶺營藏板)	7
木活字本	19	傳匏活字本	6
木板本(未詳)	18	木板本(中國)	5
金屬活字本(戊申字)	16	木板本(淸)	5
木板本(丁酉字覆刻版)	16	木板本(壬午新刊)	5
金屬活字本(丁酉字)	15	壬辰字	4

〈표〉 판사항(250 필드) 분포 상위 20개(미기재 포함)

이 추출본을 해석할 때는 다음의 유의사항을 함께 고려해야 한다. 첫째, 도서관 간 동일 서명의 중복 소장이 다수 포함되어 있어 위 수치는 '고유 판본 수'가 아니라 '소장 레코드 수'를 의미한다. 둘째, 일부 레코드는 특정 권차만 소장한 낙질 사례이므로 완질 여부는 별도로 확인이 필요하다. 셋째, 원본 데이터 자체에 인코딩 손상으로 보이는 문자가 일부 포함되어 있는데, 이는 추출 과정에서 발생한 오류가 아니라 원본 공공데이터 자체의 품질 문제이며, 자료의 현황을 있는 그대로 보여주기 위해 수정 없이 보존하였다. 이러한 자료 자체의 한계는, 앞서 자료 비판에서 지적한 KORCIS의 이질적 형성 구조가 개별 저작 단위에서도 구체적으로 재현됨을 보여주는 사례라 할 수 있다.

실증 적용 주요 결과

공공데이터포털 KORCIS 전체 서지데이터(567,212건)와 국립중앙도서관 「공공도서관 고문헌 서지데이터」(7,170건)를 전수 검색하여 『논어집주대전(論語集註大全)』 관련 레코드 1,160건을 추출하였다. 70개 기관(북미 3개 기관 포함)에 분산 소장되어 있으며, 목판본 계열 921건(79.4%)·활자본 계열 97건(8.4%)·필사본 계열 51건(4.4%)이 혼재한다.

이 코퍼스로 KoDEX의 저작-판본-소장본 3단 구조를 예비 검증한 결과, 1820년 내각장판본 사례에서 국립중앙도서관 한 기관 내에 표제·간기·형태사항이 실질적으로 동일한 레코드가 서로 다른 제어번호로 3건 독립 등록된 사실을 확인하였다. 이는 설계 단계에서는 드러나지 않았던 요구사항으로, kodex:possibleDuplicateOf 속성 신설을 제안하는 근거가 되었다. 아울러 발행년

불명·추정(392건, 33.8%), 낙질(289건, 24.9%), 합철서 편입(43건, 3.7%) 등 KORMARC에는 있으나 현재 KoDEX 설계로는 손실 없이 흡수되지 않는 잔여 정보(residue) 세 유형을 정량화하고, 각각에 대응하는 확장 속성을 제안하였다.

저작 수준뿐 아니라 편찬자 胡廣에 대한 인물(Agent) 전거 연계도 위키데이터·미국 의회도서관 전거파일(LC/NACO)을 owl:sameAs로 연결하는 실물 예시로 제시하여, 3장에서 설계한 외부 전거 연계 메커니즘이 여러 개체 수준에서 동일하게 작동함을 확인하였다.

예비발표 준비 과정에서 소장기관별 분포를 추가로 분석한 결과, 70개 기관 중 상위 5개 기관(계명대학교 동산도서관 256건, 국립중앙도서관 87건, 전남대학교 도서관 57건, 대구가톨릭대학교중앙도서관 52건, 고려대학교 도서관 48건)이 전체 1,160건의 43.1%를, 상위 10개 기관이 60.9%를 차지하는 뚜렷한 편중 구조를 확인하였다.

특히 최다 소장 기관인 계명대학교 동산도서관의 256건을 청구기호(090) 기준으로 재집계하면 실제로는 32개의 서로 다른 청구기호에 대응하며, 청구기호 하나당 평균 8건, 많게는 34건의 개별 레코드가 팔려 있었다. 이는 동일한 세트를 여러 책(冊) 단위로 나누어 각 책마다 별도의 MARC 레코드를 생성하는 카탈로깅 관행 때문으로 추정되며, 다른 기관이 한 세트 전체를 레코드 1건으로 기술하는 관행과 뚜렷이 대비된다. 즉 KORCIS 참여 기관마다 레코드가 표상하는 서지 단위(전체 세트 대 개별 책)가 다르다는 것으로, 이는 3장에서 정의한 kodex:Edition과 kodex:HoldingItem 사이에 레코드가 어느 단위로 대응하는지를 명시하는 속성(예: kodex:catalogingUnit)이 추가로 필요함을 시사하는 새로운 발견이다.

국제 전거 연계뿐 아니라 실제 해외 소장 현황도 세 기관 모두에서 구체적으로 확인된다. 아래 다섯 건 중에는 4.2절의 1820년 내각장판본 사례와 동일 판본으로 보이는 하버드옌칭도서관 소장본, UC버클리도서관이 『論語集註大全』을 독자 저작이 아니라 합철서 『三經四書大全』의 구성부로 편목한 사례(간접일치의 실제 해외 소장 사례), 컬럼비아대학교도서관이 판사항에 缺帙(낙질)을 명시한 사례가 모두 포함되어 있다.

소장기관	서명	권차	판사항	발행년	수량	비고
하버드옌칭도서관	論語集註大全	(전체)	木板本	純祖 20[1820]	20卷 7冊	4.2절 1820년 내각장판본과 동일 판본으로 추정
하버드옌칭도서관	論語集註大全	(전체)	木板本	[發行年不明]	2卷 1冊	
UC버클리도서관	三經四書大全	卷1-90	木板本(翻刻)	純祖 20(1820)	90卷 50冊	간접일치: 합철서 구성부로 편목
컬럼비아대학교도서관	論語集註大全	卷5-12,16-18	匏活字本 印	[肅宗·英祖年間]	11卷 5冊(缺帙)	낙질: 판사항에 缺帙 명시
컬럼비아대학교도서관	論語集註大全	卷1-20	金屬活字本 印	[英祖年間]	20卷 3冊	

〈표〉 북미 소장기관 사례

심화 분석 ① : 선거 통제 공백의 정량적 재확인

소장 현황 확인에서 한 걸음 더 나아가, 이 코퍼스가 앞선 진단 연구에서 확인한 문제들을 실제로도 재현하는지를 두 가지 축에서 정밀 분석하였다. 첫 번째 축은 선거 통제 공백이다. 예비발표를 준비하며 1,160건 코퍼스 자체에 대해 저자 선거(700 필드) 기재 실태를 다시 정밀 분석하였다. 그 결과 700 필드가 전혀 기재되지 않은 레코드가 502건(43.3%)에 달했다. 더 주목할 점은, 700 필드가 없는 레코드 가운데 313건은 245\$d(표제책임표시)라는 자유텍스트 필드에는 편찬자 胡廣의 이름이 이미 적혀 있었다는 사실이다. 즉 이름 정보 자체는 레코드 안에 존재하지만, 검색·선거연계가 가능한 구조화된 접근점으로는 전혀 전환되지 않은 것이다. 245\$d에 편찬자명이 등장하는 743건 가운데 42.1%(313건)가 이런 상태였다. 이는 앞선 진단 연구가 KORCIS 전체 567,212건을 대상으로 제시한 "선거통제 기재율 1% 미만"이라는 진단이, 이 저작 하나의 코퍼스 안에서도 훨씬 구체적인 형태로 재현됨을 보여준다.

700 필드가 채워진 나머지 658건에서도 표기는 통일되어 있지 않았다. 편찬자 胡廣을 지칭하는 표기만 해도 28가지 서로 다른 문자열 변이가 525건에 걸쳐 나타났으며, 한글 표기(호광)와 한자 표기(胡廣)가 스크립트 층위에서부터 갈라져 있어 단순 문자열 매칭으로는 동일 인물로 묶이지 않는다. 아래는 상위 8개 변이이다. 이 28가지 변이를 하나의 `kodex:Person` 개체로 묶고 `owl:sameAs`로 위키데이터·LC 선거와 연결하는 것이 바로 4.2절에서 제시한 인물 선거 연계 설계가 실제로 풀어야 하는 문제의 크기다.

표기 형태	건수	비고
호광	176	가장 단순한 형태(구두점 없음)
호광,	95	후행 쉼표 포함
호광(명),	85	왕조명(明) 병기
호광, 주희,	49	주희와 복수 전거 병기
주희, 호광,	47	전거 나열 순서 반대
주희 호광	19	쉼표 없는 병기
胡廣	11	한자 표기(스크립트 상이)
호광(명) 등봉칙찬	4	역할어(직찬) 포함

〈표〉 편찬자 胡廣 표기 변이 상위 8개

심화 분석 ② : 인코딩 정규화 문제의 재현과 물리기술 필드의 비정형성

두 번째 축은 인코딩 정규화 문제이다. 제목 필드에서 확인했던 CJK 호환용 한자 이체자 문제(정규화 전 337건 → 정규화 후 1,160건)가 이 필드에 국한된 우연이 아님도 함께 확인되었다. 발행지(260\$a) 필드를 NFKC 정규화 없이 "未詳"·"不明" 키워드로만 집계하면 미상 표기가 343건(29.6%)으로 나타나지만, 코퍼스 전체에 NFKC 정규화를 적용해 재집계하면 584건(50.3%)으로 늘어난다 — 즉 "發行地不明"처럼 "不"자가 호환용 한자 이체자로 입력된 레코드가 200건 이상 추가로 존재했다. 서명 필드와 발행지 필드라는 서로 다른 두 필드에서 동일한 유형의 인코딩 문제가 재현되었다는 것은 이 문제가 KORCIS 전체 코퍼스에 걸쳐 구조적으로 존재할 가능성을 시사하며, 향후 KORMARC-RDF 변환 파이프라인 설계 시 NFKC 정규화를 개별 필드 단위가 아니라 전체 레코드 처리 단계의 표준 절차로 포함해야 함을 뒷받침한다.

물리기술 정보의 비정형성도 뚜렷했다. 크기를 기술하는 300\$c 필드는 1,160건 중 592종의 사실상 서로 다른 자유텍스트 표기("33.8x22.6 cm.", "35.0 x 23.0 cm", "32.5 × 21.3 cm" 등 구분자·소수점 자릿수·공백이 제각각인 표기가 병존)를 담고 있어, 레코드의 절반 이상이 사실상 레코드마다 고유한 표기를 갖는 셈이다. 이는 3.4절에서 제안한 물리 기술 특화 확장 속성이 단순 참고용 자유텍스트가 아니라, 폭(width)·높이(height)를 별도 수치 속성으로 분리해 기술해야 할 실질적 필요를 뒷받침하는 근거가 된다.

통합 온톨로지 예시 : 하버드옌칭도서관 소장 1820년 내각장판본

4.2절의 저작-판본-소장본 매핑 예시를 실제 북미 소장 사례로 확장하면, KoDEX의 핵심 클래스가 하나의 그래프 안에서 어떻게 서로 맞물리는지 더 온전한 형태로 볼 수 있다. 아래는 하버드옌칭도서관이 소장한 1820년 내각장판 목판본(論語集註大全, 20卷 7冊)을 예로, 저작(ClassicalWork)·판본(Edition)·소장본(HoldingItem)·기관(Organization)·편찬자(Person) 다섯 개 클래스를 하나의 통합 그래프로 나타낸 것이다. 판본 유형은 3.4절의 SKOS 통제 어휘(kodex:MokpanBon)를 그대로 사용하고, 소장 수량은 앞서 제안한 확장 속성(kodex:heldVolumes)으로 구조화했으며, 편찬자는 위키데이터·미국 의회도서관 전거와 owl:sameAs로 연결했다.

```
# 저작(Work)
<https://w3id.org/kodex/work/NGJJDDJ>
  a kodex:ClassicalWork ;
  dcterms:title "논어집주대전"@ko, "論語集註大全"@zh ;
  dcterms:contributor <https://w3id.org/kodex/person/Hu-Guang> .

# 판본(Edition) - 1820년 내각장판 목판본
<https://w3id.org/kodex/edition/NGJJDDJ-1820-Naegak>
  a kodex:Edition ;
  bf:instanceOf <https://w3id.org/kodex/work/NGJJDDJ> ;
  kodex:hasEditionType kodex:MokpanBon ;
  dcterms:date "1820"^^xsd:gYear ;
  skos:note "刊記: 庚辰(1820)新刊內閣藏板"@zh .

# 소장본(HoldingItem) - 하버드옌칭도서관 (구조화된 수량 속성 적용)
<https://w3id.org/kodex/item/HYL-NGJJDDJ-1820>
  a kodex:HoldingItem ;
  bf:itemOf <https://w3id.org/kodex/edition/NGJJDDJ-1820-Naegak> ;
  kodex:heldBy <https://w3id.org/kodex/org/HYL> ;
  kodex:heldVolumes 7 ;
  dcterms:extent "20卷 7冊" .

# 소장 기관(Organization)
<https://w3id.org/kodex/org/HYL>
  a kodex:Organization ;
  foaf:name "하버드옌칭도서관"@ko, "Harvard-Yenching Library"@en .

# 편찬자(Person) - 국제 전거 연계
<https://w3id.org/kodex/person/Hu-Guang>
  a kodex:Person ;
  foaf:name "胡廣"@zh ;
  owl:sameAs <http://www.wikidata.org/entity/Q6728408>,
               <https://id.loc.gov/authorities/names/nb2023001408> .
```

※ 위 URI는 3장의 설계 원칙에 따른 예시이며, 기관·판본·수량 정보는 실제 KORCIS 레코드 값을 반영하였다.

LOD 구현 예시: SPARQL을 통한 소장기관 분포 재현

위와 같이 70개 기관의 레코드가 모두 동일한 저작 URI(<https://w3id.org/kodex/work/NGJJDDJ>)를 통해 연결되면, 현재는 원본 KORMARC 레코드를 하나씩 내려받아 별도 파싱·집계 스크립트를 돌려야만 얻을 수 있는 정보 - 어느 기관이 이 저작을

몇 건 소장하고 있는가 - 를 SPARQL 질의 하나로 즉시 얻을 수 있다. 아래는 심화 분석에서 수작업으로 집계한 소장기관 상위 목록을, KoDEX로 변환된 데이터에 대한 SPARQL 질의로 재현한 예시이다.

```
PREFIX kodex: <https://w3id.org/kodex/>
PREFIX bf:    <http://id.loc.gov/ontologies/bibframe/>
PREFIX foaf:  <http://xmlns.com/foaf/0.1/>

SELECT ?기관명 (COUNT(?item) AS ?소장건수)
WHERE {
  ?item bf:itemOf ?edition ;
        kodex:heldBy ?org .
  ?edition bf:instanceOf <https://w3id.org/kodex/work/NGJJDJ> .
  ?org foaf:name ?기관명 .
}
GROUP BY ?기관명
ORDER BY DESC(?소장건수)
```

※ 위 SPARQL 엔드포인트는 아직 실제 배포되지 않았으며, 질의 형태를 예시하기 위한 설계안이다.

이 질의를 1,160건 코퍼스 전체에 대해 수작업으로 실행한 결과(즉 위 질의가 반환할 것으로 기대되는 값)는 심화 분석에서 이미 확인한 상위 5개 기관과 정확히 일치한다: 계명대학교 동산도서관 256건, 국립중앙도서관 87건, 전남대학교 도서관 57건, 대구가톨릭대학교중앙도서관 52건, 고려대학교 도서관 48건. 즉 KoDEX로의 변환이 완료되면, 지금까지 이 예비발표를 위해 별도의 추출·파싱 스크립트로 며칠에 걸쳐 수행한 집계 작업이 표준 SPARQL 질의 한 줄로 대체된다는 것이 LOD 전환의 실질적 효용이다.

아울러 실제 LOD 서비스에서는 저작 URI에 대한 HTTP 요청 시 Accept 헤더에 따라 사람이 보는 HTML 화면과 기계가 처리하는 JSON-LD·Turtle 응답을 함께 제공하는 콘텐츠 협상(content negotiation)을 적용하는 것이 표준적이며, 이는 하나의 영구 식별자가 연구자와 애플리케이션 모두에게 동일하게 서비스되도록 하는 LOD 발행의 핵심 요건이다.

다만 이 실증 적용은 단일 저작·유교 경전류라는 제한된 범위의 예비적 검증이며, 추출·파싱에 사용한 코드와 결과 데이터는 재현 가능성 확보를 위해 GitHub 저장소로 별도 공개할 예정이다.

2. 앞으로 수행할 연구: KORMARC-RDF 변환 가이드라인 및 LOD 구현 방안

「KoDEX 온톨로지 기반 KORMARC 데이터의 RDF 변환 가이드라인 및 Linked Open Data 구현 방안」

이 연구는 앞서 설계한 KoDEX 온톨로지를 실제 KORMARC 데이터에 적용하여 변환 규칙을 수립하고, 이를 LOD 서비스로 구현하는 방안을 다룰 예정으로, 설계 연구의 원고작성을 마친 이후

본격적으로 착수할 계획이다. 설계 연구가 『논어집주대전』 단일 저작으로 예비 검증하며 제안한 확장 속성(kodex:possibleDuplicateOf 등 중복·낙질·합철서 편집 대응 속성)과 인물 전거 연계 방식을 실제 구현 대상으로 삼아, 대상 범위를 여러 유형의 고문헌으로 넓히는 것이 이 구현 연구의 출발점이다. 현재는 본문 집필 전 목차를 구상하는 단계이며, 지금까지 구상된 목차는 다음과 같다.

구상 목차(안)

1. 서론

2. KORMARC-to-RDF 변환 가이드라인

2.1 변환 원칙 및 매핑 체계

2.2 주요 필드별 변환 규칙

2.3 전거 및 식별자 처리 방안

2.4 미상·불확실 데이터 처리 방안

3. 실증 적용 및 검증: 논어집주(論語集註)

3.1 실증 대상 및 적용 절차

3.2 KoDEX 변환 결과

3.3 국립중앙도서관 LOD 데이터와의 구조적 비교

3.4 Competency Question 기반 검증

4. LOD 서비스 구현 방안

4.1 아키텍처 설계 모델 비교 : 국립중앙도서관 통합 운영 vs. 기관 분산+연계 허브

4.2~4.3 현실적·이상적 구현 방안

4.4 SPARQL 엔드포인트·REST API 및 국제 포털 연계

4.5 단계적 구현 로드맵

5. 결론

V. 연구방법 및 범위 종합

본 박사학위논문은 정량적 데이터 분석(진단 연구)과 온톨로지 설계 방법론(설계 연구), 실증 적용 및 시스템 구현 방안 검토(구현 연구)를 결합한 혼합 연구 방법을 취한다. 진단 연구는 KORCIS 공개 서지 데이터 567,212건 전체를 대상으로 한 전수 조사이며, 개별 소장 기관의 내부 시스템 및 메타데이터의 내용적 정확성 검증은 범위에서 제외하였다. 설계 연구는 국제 표준 어휘 문헌 검토와 온톨로지 설계 방법론을 적용하고, 여기에 더해 『논어집주대전』 1,160건을 대상으로 한 단일 저작 심층사례로 설계안을 예비 검증하였다. 이 예비 검증에서 식별된 확장 필요 지점과 대규모·다유형 자료로의 확장 검증은 구현 연구로 이관하였다. 구현 연구는 이를 바탕으로 여러 유형의 고문헌으로 대상 자료를 확대한 실증 적용을 중심으로 진행할 예정이다.

VI. 향후 연구 일정

시기	내용
2026년 10월	설계 연구 학술지 투고
2026년 하반기	설계 연구 심사·수정 대응, 구현 연구 착수(변환 가이드라인 수립·실증 적용)
2027년 상반기	구현 연구, 학위논문 종합 집필 완료

〈표〉 향후 연구 일정

VII. 연구의 의의 및 기대효과

실증적 진단에서 출발

인상이나 추정이 아닌 **KORCIS** 서지 데이터 567,212건에 대한 전수 정량 분석을 통해 문제를 구체적으로 규명하였다.

국제 표준 재사용과 특화 확장의 균형

독자적 표준을 새로 만드는 대신 **BIBFRAME 2.0** 등 검증된 국제 표준 어휘를 최대한 재사용하고, 반드시 필요한 영역에 한해 한국 고문헌 특화 어휘를 확장함으로써 국제 호환성과 국내 특수성을 동시에 확보하고자 하였다.

글로벌 지식 네트워크와의 연계

VIAF·Wikidata 등 국제 전거 시스템과의 연계 구조를 설계함으로써, 한국 고문헌 메타데이터가 국내 시스템 내부에 머무르지 않고 세계 학술 네트워크로 개방될 수 있는 기반을 마련하고자 하였다.

단계적으로 실현 가능한 설계

전면적인 시스템 재구축을 전제하지 않고, 현행 **KORCIS** 위에서 단계적으로 구현 가능한 경로를 제시함으로써 실행 가능성을 높이하고자 하였다.

실증 데이터와 코드의 공개를 통한 재현 가능성 확보

『논어집주대전』 1,160건을 추출·파싱하는 데 사용한 코드와 결과 데이터셋을 **GitHub** 저장소로 공개할 예정이다. 이는 **KORCIS** 공개 데이터를 이용한 연구의 재현 가능성을 높이고, 후속 연구자가 다른 저작이나 자료 유형으로 동일한 절차를 적용해 검증·확장할 수 있는 기반을 제공한다는 점에서 개방적 연구 실천(open research practice)에 기여하고자 한다.

참고문헌

가. 국내 문헌

강순애. (2014). 문서 목록규칙 표준화(안) 연구. 국립중앙도서관.

공공데이터포털. (n.d.). 한국고전적종합목록 서지데이터.

<https://www.data.go.kr/data/15096231/fileData.do>

국립중앙도서관. (n.d.). 국가서지 링크드 오픈 데이터(LOD) 서비스. <https://lod.nl.go.kr>

국립중앙도서관. (n.d.). 한국고문헌종합목록(KORCIS). <https://www.nl.go.kr/korcis>

김영진. (2024). 한국 고문헌 목록의 현황과 문제점, 향후 방향. 大東文化研究, 128, 237–274.

<https://doi.org/10.18219/ddmh..128.202412.237>

김현. (2016). 디지털 인문학이란 무엇인가. 한국학중앙연구원 디지털인문학연구소.

문화체육관광부, 국립중앙도서관. (2025). 한국고문헌종합목록 공공도서관 고문헌

서지데이터 (2025. 7. 31. 기준). 공공데이터포털.

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>

박옥남·오정선. (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>

박진호. (2025). OWL 속성 특성에 기반한 BIBFRAME 객체 속성의 세분화.

한국비블리아학회지, 36(2), 31–66. <https://doi.org/10.14699/kbiblia.2025.36.2.031>

박진호·송민선·이미화. (2026). BIBFRAME 도입을 위한 내규 기반 조직 거버넌스 설계.

한국비블리아학회지, 37(1), 269–286. <https://doi.org/10.14699/kbiblia.2026.37.1.269>

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>

송민선·이미화·박진호. (2026). MODS to BIBFRAME 변환 규칙 연구. 정보관리학회지, 43(1), 1–28. <https://doi.org/10.3743/KOSIM.2026.43.1.001>

신은선. (2026). FAIR 원칙에 따른 한국 고문헌 메타데이터 관리체계 진단: KORCIS를 중심으로. 한국기록관리학회지, 26(2), 93-111.
<https://doi.org/10.14404/JKSARM.2026.26.2.093>

이미화. (2017). BIBFRAME 2.0 특징 분석 및 BIBFRAME 구축시 고려사항에 관한 연구. 한국도서관·정보학회지, 48(4), 107–127. <https://doi.org/10.16981/kliss.48.4.201712.107>

이미화·박진호·이종욱·송민선. (2025). KORMARC to BIBFRAME 변환을 위한 매핑 시 고려사항에 관한 연구. 한국도서관·정보학회지, 56(1), 307–326.
<https://doi.org/10.16981/kliss.56.1.202503.307>

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>

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. 외국 문헌

Bekiari, C., Bruseker, G., Doerr, M., Ore, C., Stead, S., & Athanasios, T. (2022). CIDOC Conceptual Reference Model (CRM), version 7.1.2. ICOM CIDOC.

Brickley, D., & Miller, L. (2014). FOAF vocabulary specification 0.99.
<http://xmlns.com/foaf/spec/>

Doerr, M. (2003). The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata. AI Magazine, 24(3), 75–92.

Isaac, A. (Ed.). (2013). Europeana data model primer. Europeana.

Library of Congress. (n.d.). BIBFRAME 2.0 specification. <https://www.loc.gov/bibframe/>

Miles, A., & Bechhofer, S. (Eds.). (2009). SKOS simple knowledge organization system reference. W3C Recommendation.

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>