

박사학위과정 예비발표

한국 고문헌 메타데이터의 시맨틱 표준화 연구

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

신은선

한국학중앙연구원 인문정보학 박사과정

2026년 8월 28일

발표 순서

PART 1

문제 제기

연구 배경과 KORCIS의
구조적 한계, 연구 질문

PART 2

선행연구와 연구 설계

선행연구 검토와
박사학위논문 3단계 설계

PART 3

세부 연구 내용

지금까지 수행한
진단·설계 연구와 앞으로
수행할 연구

PART 4

향후 계획과 의의

연구 방법 종합, 향후 일정,
연구의 의의

PART 1

문제 제기

한국 고문헌 메타데이터는 국제 시맨틱 웹 기준을 얼마나 충족하는가

한국 고문헌 메타데이터와 디지털 인문학

고문헌은 국내외 학술 연구의 핵심 1차 자료이며, 문헌정보학은 메타데이터 표준화와 지식조직을 통해 이 자산이 안정적으로 유통되도록 하는 방법론적 기반을 제공.

국내 고문헌 메타데이터는 한국고문헌종합목록시스템 (KORCIS)을 중심으로 통합 관리.

144개

참여 기관 (국내 91 · 국외 53)

약 56만여 건

서지 데이터

약 6.5만여 건

원문 이미지

2007년~

서비스 개시 후 지속 확충

KORCIS는 한국 고문헌 메타데이터를 논의할 때 가장 먼저 검토되어야 할 국가적 규모의 인프라다(국립중앙도서관, n.d.).

KORCIS의 구조적 한계

- × **RDF 미전환** 전통적 KORMARC 서지 표준에 기반해, 시맨틱 웹이 요구하는 **연결 데이터 (Linked Data)로의 전환이 충분히 이루어지지 않음**
- × **전거 연계 부족** **외부 전거 파일 (Authority File)과의 연계가 부족**해 글로벌 지식 네트워크와의 상호운용성 확보에 어려움
- × **기관 간 분산·불일치** 개별 소장 기관별 분산 관리 구조와 **데이터 표준·기술 형식의 불일치**가 통합 검색과 고도화된 활용을 저해

전 세계적으로는 FAIR 원칙(Findable·Accessible·Interoperable·Reusable)이 연구 데이터 관리의 핵심 지침으로 확산되고 있다 (Wilkinson et al., 2016). KORCIS는 이 국제 기준을 얼마나 충족하는가 — 이 물음이 본 연구 전체의 출발점이다.

진단 · 설계 · 구현의 세 가지 질문

한국 고문헌 메타데이터를 국제 시맨틱 웹 환경에서 발견·접근·연계·재사용 가능하게 하기 위한 방안을 진단·설계·구현의 세 단계에 걸쳐 규명하는 것을 총괄 목적으로 한다.

진단

연구질문 1

현재 KORCIS 메타데이터는 **FAIR** 원칙을 얼마나 충족하고 있으며, 무엇이 부족한가?

설계

연구질문 2

국제 표준 어휘와 한국 **고문헌**의 서지적 특수성을 함께 담아내는 온톨로지는 어떻게 설계하는가?

구현

연구질문 3

설계된 온톨로지로 실제 데이터를 변환하고, 이를 **LOD** 서비스로 구현하는 방법은 무엇인가?

PART 2

선행연구와 연구 설계

재사용 가능한 국제 표준과 남아 있는 연구 공백, 그리고 3단계 연구 설계

서지 온톨로지 및 LOD 표준화 연구

BIBFRAME 2.0

미국 의회도서관(LC) 주도.
저작·인스턴스·개별자료 3계층 구조로 서지
개체 관계를 RDF화(Library of Congress, n.d.).
국내 연구도 2014년 이후 지속 축적
(박옥남·오정선, 2014; 이미화, 2017;
이미화·배정현, 2025).

CIDOC-CRM

박물관·아카이브 문화유산 정보의 참조
온톨로지. 사건(Event) 중심 기술이 고문헌의
간행·전래·소장 이력 표현에 적합(Doerr,
2003).

Europeana(EDM)

Dublin Core·OAI-ORE·SKOS·OWL 등 기존
표준을 결합해 이질적 기관 데이터를 통합.
독자 표준 개발보다 어휘 재사용이
상호운용성에 효과적임을 보여주는 사례
(Isaac, 2013).

국내에서는 BIBFRAME 적용을 위한 세부 방법론 연구도 활발(박진호, 2025; 송민선 외, 2026; 이미화 외, 2025; 박진호 외, 2026). 이들은 속성 세분화·변환 규칙·조직 거버넌스 등 실무 절차 정교화에 초점을 두는 반면, 본 연구의 KoDEX(Korean Old Document Extended Schema)는 한국 고문헌이라는 특정 자료군의 물리 기술 요소와 외부 전거 연계에 특화된 확장 어휘를 다룬다는 점에서 차별화.

한국 고문헌 디지털화 연구와 연구 공백

박민정·이승민(2023) LRM 기반 서지구조의 고문헌 적용 · 임소정·손계영(2023) 기관 간 표준 불일치와 품질 격차 · 강순애(2014) KORMARC 기반 고문헌 서지 기술 체계화 · 허철(2024) 국제 표준 정렬의 필요성 · 김영진(2024) KORCIS 구조적 문제점의 목록학적 실증

연구 공백

국제 표준 어휘를 활용한 서지 LOD 연구, BIBFRAME의 국내 실무 적용 방법론 연구, 그리고 한국 고문헌 디지털화 연구는 각각 축적되어 왔으나, 한국 고문헌의 서지적 특수성을 반영한 RDF 온톨로지를 독자적으로 설계한 연구는 아직 충분히 이루어지지 않음.

본 연구는 이 지점 - 실증적 진단에 기반한 온톨로지 설계와 그 구현 경로 - 을 연구 대상으로 삼는다.

3단계 연구 설계: 진단 - 설계 - 구현

하나의 총괄 연구 질문에 대한 답을 세 편의 개별 논문으로 나누어 단계적으로 규명하며, 최종적으로 하나의 박사학위논문으로 통합·확장될 예정.

1 진단

FAIR 원칙에 따른 한국 고문헌
메타데이터 관리체계 진단

게재 완료 · 한국기록관리학회지 26(2),
93-111(2026.5)

2 설계

한국 고문헌 메타데이터 시맨틱
표준화를 위한 KoDEX 온톨로지
설계

최종본 작성중 · 2026년 10월 투고

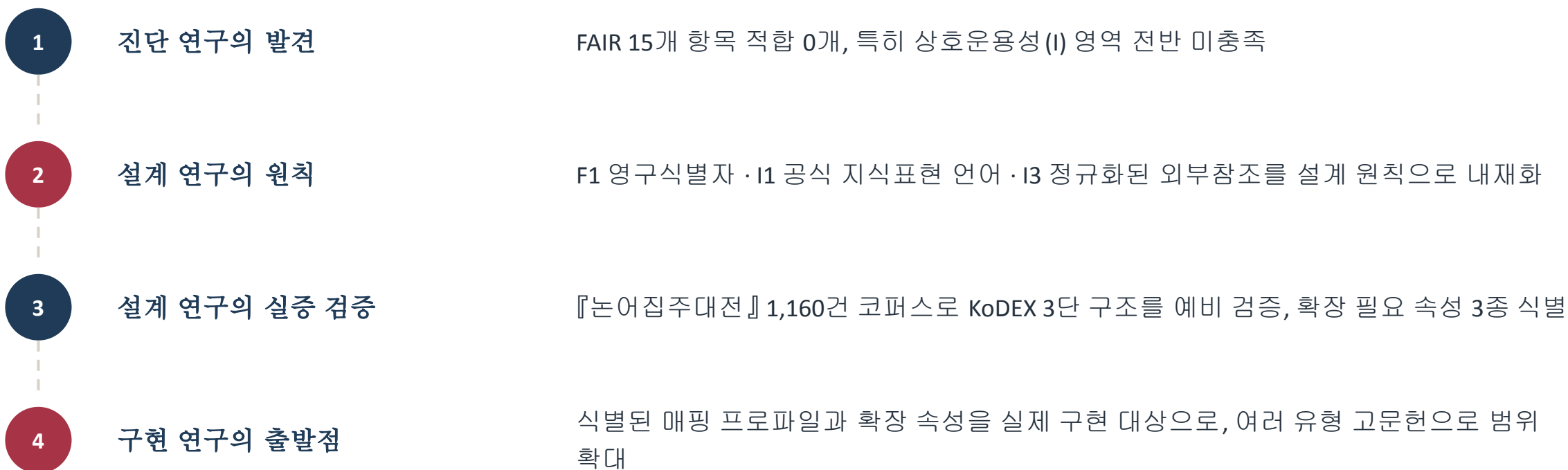
3 구현

KoDEX 온톨로지 기반 KORMARC
데이터의 RDF 변환 가이드라인 및
LOD 구현 방안

목차 구상 단계 · 착수 예정

연구 간 연계 논리

각 단계는 이전 단계의 결과에서 직접 도출 : 독립된 세 주제가 아니라, 하나의 문제의식이 진단에서 설계, 구현으로 심화되는 단일한 연구 궤적.



PART 3

세부 연구 내용

지금까지 수행한 진단·설계 연구를 정리하고, 앞으로 수행할 구현 연구의 방향을 제시한다

메타데이터 진단에서 온톨로지 설계까지

지금까지는 두 단계의 연구를 순차적으로 수행. KORCIS 서지 데이터 전체를 대상으로 FAIR 원칙 준수 실태를 실증적으로 진단하고, 이 결과를 출발점 삼아 한국 고문헌에 특화된 RDF 온톨로지 KoDEX를 설계한 뒤 단일 저작 사례로 예비 검증.

가

FAIR 원칙에 따른
KORCIS 메타데이터 진단

KORCIS 서지 데이터 567,212건 전체를 대상으로
필드 완성도, 전거 제어, FAIR 15개 세부 항목
준수도를 정량적으로 진단.

<https://github.com/SHINEUNSUN1027/korcis-fair-analysis>



나

KoDEX 온톨로지 설계 및
예비 실증 검증

국제 표준 어휘 재사용과 한국 고문헌 특화 확장을
결합해 KoDEX를 설계하고,
『[논어집주대전](#)』 1,160건으로 예비 실증 검증.

Ⅳ-가. 지금까지 수행한 연구 ① 진단 연구

FAIR 원칙에 따른 KORCIS 메타데이터 진단

「FAIR 원칙에 따른 한국 고문헌 메타데이터 관리체계 진단 — KORCIS를 중심으로 —」

567,212건

KORCIS 서지 데이터 전체를 대상으로 한
실증 진단

한국기록관리학회지 26(2), 93-111 게재 (2026.5)

신은선(2026) · 필드 유형(필수·해당시필수·재량) 구분과 데이터 구조적 특성·시스템 기능 지원 여부·정량 분석
결과를 종합하여 판정

①

KORMARC 주요 필드
출현빈도·완성도 측정

②

전거 제어 및 식별자 체계 실태
분석

③

FAIR 15개 세부 항목별 준수도
평가

자료 비판: KORCIS는 어떻게 형성되었는가

KORCIS는 단일 기관이 일관된 원칙으로 구축한 균질적 데이터셋이 아니라, 144개 참여기관(국내 91 · 해외 53)이 각자의 목록 환경에서 독립적으로 생성한 레코드를 국립중앙도서관이 사후 수집·통합한 연합형 통합목록(union catalog)이다(국립중앙도서관, n.d.).

① 원 소장기관 목록층

144개 참여기관이 각자의 목록 규칙과 인력 수준에 따라 독립적으로 KORMARC 레코드를 생성

② KORCIS 통합 수집층

국립중앙도서관이 표준 인터페이스로 사후 수집·색인
: 규칙의 사전 통일이나 사후 정합성 검증 근거는 공개 문서에서 미확인

이 이질적 형성 구조는 진단 결과에 이미 흔적을 남김.

필수 필드 완성도는 99% 이상으로 균질하지만, 선택 필드는 1.49%~38.86%까지 큰 편차를 보임.

이는 단순한 입력 누락이 아니라 참여기관마다 다른 목록 상세화 수준(level of cataloging)과 서지 기술 관행을 시사하며, 이 이질성의 구체적 양상은 뒤의 『논어집주대전』 실증 사례(계명대학교 청구기호 재집계)에서 다시 확인.

Ⅳ-가. 진단 연구

주요 결과: FAIR 15개 항목 중 적합 0개

99%+

필수 서지 필드(245·300·260 등) 완성도

1.49~38.86%

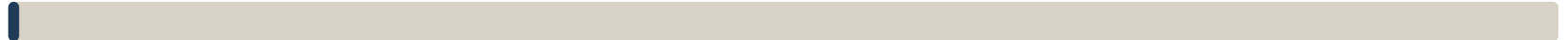
선택 필드(100·650·856 등) 완성도 편차

1% 미만

인명 필드 중전거레코드 제어번호 기재 비율
(국제전거 URI 0건)

FAIR 15개 세부 항목 준수도 평가

적합 0개 (0%)



0%

부분 적합 5개 (33.3%)



33.3%

미흡·결여 10개 (66.7%)



66.7%

Ⅳ-가. 진단 연구

FAIR 15개 세부 항목 준수도 세부 평가

발견 가능성(F)·접근 가능성(A)·재사용 가능성(R)은 부분 적합 항목이 섞여 있는 반면, 상호운용성(I) 3개 항목은 전부 미흡·결여로 판정되었다(신은선, 2026).

발견 가능성 Findable	접근 가능성 Accessible	상호운용성 Interoperable	재사용 가능성 Reusable
F1 전 지구적 고유 식별자	A1 표준 프로토콜 기반 접근	I1 공식 지식 표현 언어 사용	R1 복수 속성 메타데이터 기술
미흡·결여	부분 적합		부분 적합
F2 풍부한 메타데이터 기술	A1.1 무료·개방 프로토콜	미흡·결여	R1.1 라이선스 정보 제공
부분 적합	부분 적합		미흡·결여
F3 메타데이터에 식별자 명시	A1.2 인증·권한 허용 프로토콜	미흡·결여	R1.2 출처 및 변동 이력
미흡·결여	미흡·결여		미흡·결여
F4 외부 인덱스 등록	A2 메타데이터 보존	I3 정규화된 외부 참조	R1.3 커뮤니티 표준 준수
미흡·결여	미흡·결여	미흡·결여	부분 적합

Ⅳ-가. 진단 연구

해외 주요 기관과의 FAIR 준수도 비교

Europeana·NDL은 HTTP URI 식별자·RDF 공개·국제 전거 연계·기계적 접근이라는 네 조건을 상당 수준 충족하는 반면, 중화고적자원고와KORCIS는 텍스트 기반 서지 기술에 머물러 상호운용성(I) 영역 전반이 미충족이다(신은선, 2026).

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 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	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

〈표〉 해외 주요 기관 FAIR 원칙 준수도 비교(신은선, 2026, 100쪽 표6)

KoDEX 온톨로지 설계 및 예비 실증 검증

「한국 고문헌 메타데이터 시맨틱 표준화를 위한 KoDEX 온톨로지 설계」

연구 목적

- ① HTTP URI 기반 영구 식별자 체계 수립
- ② BIBFRAME 2.0 중심의 SKOS·FOAF·CIDOC-CRM 혼합 어휘 체계 확립
- ③ 핵심 클래스·속성 정의 및 물리기술 확장 어휘의 OWL 형식화
- ④ VIAF·Wikidata·국중 전거와의 외부 전거 연계 구조 설계
- ⑤ 단일 저작 심층사례를 통한 설계안의 예비 실증 검증

설계 원칙

- 1 국제 표준 어휘 재사용 우선
- 2 FAIR 원칙 정렬 (F1·I1·I3)
- 3 한국 고문헌 특수성 반영
- 4 개방성 및 확장성
- 5 단계적 구현 가능성

설계 원칙의 이론적 근거

- ① 어휘 재사용

온톨로지 공학의 상호운용성 확보 핵심 전략(Isaac, 2013)
: 진단에서 확인된 국제 전거 연계 0건 공백을 근본적으로 해소
- ② FAIR 정렬

진단 결과를 직접 반영
: F1(영구 식별자)·I1(공식 지식표현 언어)·I3(정규화된 외부참조)를 설계 시점부터 구조적 요건으로 내재화
- ③ 한국 고문헌 특수성

BIBFRAME 2.0·CIDOC-CRM은 서구 자료를 준거로 발전
: 광곽·어미·판구 등 동아시아 물리적 특징 표현에 태생적 한계
- ④ 개방성·확장성

실증 적용에서 kodex:possibleDuplicateOf 등 예상 못한 확장 속성의 필요성이 실제로 확인
: 모듈형 구조로 설계
- ⑤ 단계적 구현

이미 56만여 건을 운영 중인 국가 인프라라는 현실적 제약
: KORMARC 필드 매핑으로 레코드·기관 단위 점진적 전환 지원

논어집주대전 코퍼스로 KoDEX를 검증

공공데이터포털 KORDIS 전체 서지데이터(567,212건)와 국립중앙도서관「공공도서관 고문헌 서지데이터」(7,170건)를 전수 검색하여 『논어집주대전』 관련 레코드를 추출하고, 이를 대상으로 KoDEX의 저작-판본-소장본 3단 구조를 예비 검증.

1,160건

추출 레코드

70개 기관

북미 3개 기관 포함 분산 소장

3단 구조

저작·판본·소장본 예비 검증

판본 유형 구성비

목판본 계열 (921건)



79.4%

활자본 계열 (97건)



8.4%

필사본 계열 (51건)



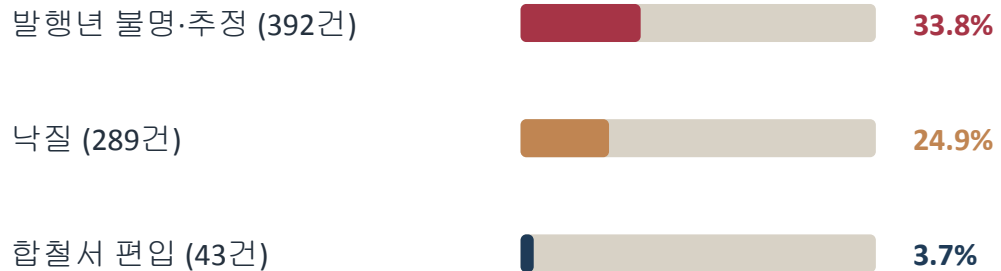
4.4%

설계 단계에서 드러나지 않은 세 가지 발견

중복 등록 사례 → `kodex:possibleDuplicateOf` 제안

1820년 내각장판본, 국립중앙도서관 한 기관 내에서 표제·간기·형태사항이 실질적으로 동일한 레코드가 서로 다른 제어번호로 3건 독립 등록된 사실을 확인 : 설계 단계에서 드러나지 않았던 요구사항

잔여 정보(residue) 3유형



KORMARC에는 있으나 현재 설계로는 손실 없이 흡수되지 않는 정보: 각각 대응 확장 속성을 제안

소장기관 집중도 (70개 기관 중)



계명대학교 동산도서관 256건 · 국립중앙도서관 87건 · 전남대학교 도서관 57건 ·
대구가톨릭대학교중앙도서관 52건 · 고려대학교 도서관 48건

같은 세트, 다른 서지 단위: 계명대학교 사례

256건

계명대학교 동산도서관 소장 레코드

↓ 청구기호 (090) 기준 재집계

32개

서로 다른 청구기호에 대응 (평균 8건, 최대 34건)

동일한 소장 세트를 여러 책(冊) 단위로 나누어 책마다 별도의 MARC 레코드를 생성하는 카탈로깅 관행 때문으로 추정되며, 다른 기관이 한 세트 전체를 레코드 1건으로 기술하는 관행과 뚜렷이 대비.

KORCIS 참여 기관마다 레코드가 표상하는 서지 단위(전체 세트 대 개별 책)가 다르다는 것을 확인.

`kodex:Edition`과 `kodex:HoldingItem` 사이에 레코드가 어느 단위로 대응하는지를 명시하는 속성(`kodex:catalogingUnit`)이 추가로 필요함을 시사.

국제 전거 연계와 실제 해외 소장의 정합

4.2절 1820년 내각장판본 사례와 동일 판본으로 보이는 하버드엔칭도서관 소장본, 합철서 구성부로 편목된 UC버클리도서관 사례, 판사항에 낙질을 명시한 컬럼비아대학교도서관 사례를 포함.

소장기관	서명	권차	판사항	발행년	수량	비고
하버드엔칭도서관	論語集註大全	(전체)	木板本	純祖20(1820)	20卷 7冊	4.2절 사례와 동일 판본 추정
하버드엔칭도서관	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	—

컬럼비아대학교도서관이 판사항에 缺帙(낙질)을 명시한 사례처럼, 해외 소장 레코드 역시 국내 레코드와 동일한 잔여 정보 유형을 포함한다.

전거 통제 공백의 정량적 재확인

43.3%

저자 전거(700 필드) 미기재 (502/1,160건)

42.1%

245\$d엔 이름 있으나 700
미기재로 미구조화 (313/743건)

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(칙찬) 포함

700 필드가 채워진 658건에서도 표기는 통일되지 않음.
편찬자 胡廣을 지칭하는 표기만 28가지 변이가 525건에 걸쳐 나타났으며, 한글·한자 표기가 스크립트 층위에서부터 갈라져 단순 문자열 매칭으로는 동일 인물로 묶이지 않음.

인코딩 정규화 문제의 재현과 물리기술 비정형성

발행지(260\$a) 미상 표기 : NFKC 정규화 전후 비교



"發行地不明"처럼 "不"자가 호환용 한자 이체자로 입력된 레코드가 200건 이상 추가로 발견.

서명 필드에서 확인했던 문제(정규화 전 337건 → 후 1,160건)가 이 필드에 국한된 우연이 아님을 보여준다.

향후 변환 파이프라인에서 NFKC 정규화를 전체 레코드 처리의 표준 절차로 포함해야 함.

592종

300\$c(크기) 필드의 사실상 서로 다른 자유텍스트 표기 (1,160건 중) — "33.8x22.6 cm.", "35.0 x 23.0 cm", "32.5 × 21.3 cm" 등 구분자·소수점 자릿수·공백이 제각각. 폭(width)·높이(height)를 별도 수치 속성으로 분리 기술할 필요를 뒷받침.

하버드옌칭도서관 소장 1820년 내각장판본

저작(ClassicalWork)·판본(Edition)·소장본(HoldingItem)·기관(Organization)·편찬자(Person) 다섯 클래스가 하나의 그래프로 맞물리는 방식.
판본 유형은 SKOS 통제 어휘(kodex:MokpanBon), 소장 수량은 확장 속성(kodex:heldVolumes), 편찬자는 위키데이터·LC 전거와 owl:sameAs로 연결.

```
<.../work/NGJJJDJ> a kodex:ClassicalWork ;  
  dcterms:title "논어집주대전"@ko, "論語集註大全"@zh ;  
  dcterms:contributor <.../person/Hu-Guang> .  
  
<.../edition/NGJJJDJ-1820-Naegak> a kodex:Edition ;  
  bf:instanceOf <.../work/NGJJJDJ> ;  
  kodex:hasEditionType kodex:MokpanBon ;  
  dcterms:date "1820"^^xsd:gYear .  
  
<.../item/HYL-NGJJJDJ-1820> a kodex:HoldingItem ;  
  bf:itemOf <.../edition/NGJJJDJ-1820-Naegak> ;  
  kodex:heldBy <.../org/HYL> ; kodex:heldVolumes 7 .  
  
<.../person/Hu-Guang> a kodex:Person ; foaf:name 胡廣"@zh ;  
  owl:sameAs <wikidata.org/entity/Q6728408>,  
    <id.loc.gov/authorities/names/nb2023001408> .
```

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

SPARQL을 통한 소장기관 분포 재현

70개 기관의 레코드가 모두 동일한 저작 URI로 연결되면, 지금까지 별도 파싱·집계 스크립트로 수행한 소장기관 분포 집계를 SPARQL 질의 한 줄로 즉시 얻을 수 있음.

```
PREFIX kodex: <https://w3id.org/kodex/>
PREFIX bf: <.../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 <.../work/NGJJDJ> .
    ?org foaf:name ?기관명 .
}
GROUP BY ?기관명
ORDER BY DESC(?소장건수)
```

이 질의가 반환할 값은 심화 분석 상위 5개 기관과 정확히 일치한다

계명대 256 · 국중 87 · 전남대 57 ·
대구가톨릭대 52 · 고려대 48

며칠에 걸친 수작업 집계가 표준 SPARQL 질의 한 줄로 대체되는 것이 LOD 전환의 실질적 효용이다.

※ 실제 LOD 서비스에서는 Accept 헤더에 따라 HTML·JSON-LD·Turtle 응답을 함께 제공하는 콘텐츠 협상(content negotiation)이 표준적이며, SPARQL 엔드포인트는 아직 실제 배포되지 않은 설계안이다.

KORMARC-RDF 변환 가이드라인 및 LOD 구현 방안

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

설계 연구의 투고를 마친 이후 본격적으로 착수할 계획.

설계 연구가 『논어집주대전』 단일 저작으로 예비 검증하며 제안한 확장 속성(kodex:possibleDuplicateOf 등)과 인물 전거 연계 방식을 실제 구현 대상으로 삼아, 대상 범위를 여러 유형의 고문헌으로 넓히는 것이 이 구현 연구의 출발점. 현재는 목차를 구상하는 단계.

- 1 서론
- 2 KORMARC-to-RDF 변환 가이드라인 — 변환 원칙·매핑 체계, 필드별 변환 규칙, 전거·식별자 처리, 미상·불확실 데이터 처리
- 3 실증 적용 및 검증: 논어집주(論語集註) — 실증 대상 및 절차, 변환 결과, 국중 LOD 데이터와의 구조적 비교, Competency Question 검증
- 4 LOD 서비스 구현 방안 — 아키텍처 설계 모델 비교, 현실적·이상적 구현 방안, SPARQL·REST API 및 국제 포털 연계, 단계적 구현 로드맵
- 5 결론

PART 4

향후 계획과 의의

박사학위논문 전체를 관통하는 연구 방법과 향후 일정, 그리고 연구가 갖는 의의

정량 분석 · 설계 방법론 · 실증 적용을 결합한 혼합 연구

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 방법.

진단 연구

정량적 데이터 분석

KORCIS 공개 서지 데이터 567,212건 전체를 대상으로 한 전수 조사. 소장 기관 내부 시스템·내용적 정확성 검증은 범위에서 제외.

설계 연구

온톨로지 설계 방법론

국제 표준 어휘 문헌 검토와 설계 방법론을 적용하고, 『논어집주대전』 1,160건의 단일 저작 심층사례로 설계안을 예비 검증.

구현 연구

실증 적용 및 시스템 구현 방안 검토

예비 검증에서 식별된 확장 필요 지점을 바탕으로, 여러 유형의 고문헌으로 대상 자료를 확대한 실증 적용을 중심으로 진행 예정.

설계 연구 투고에서 학위논문 종합 집필까지



다섯 가지 기대효과

1

실증적 진단에서 출발

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

2

국제 표준 재사용과 특화 확장의 균형

BIBFRAME 2.0 등 검증된 국제 표준 어휘를 최대한 재사용하고, 필요한 영역에 한해 한국 고문헌 특화 어휘를 확장

3

글로벌 지식 네트워크와의 연계

VIAF·Wikidata 등 국제 전거 시스템과의 연계 구조를 설계해 세계 학술 네트워크로 개방될 수 있는 기반 마련

4

단계적으로 실현 가능한 설계

전면적 시스템 재구축을 전제하지 않고, 현행 KORCIS 위에서 단계적으로 구현 가능한 경로를 제시

5

실증 데이터와 코드 공개를 통한 재현 가능성 확보

『논어집주대전』 1,160건의 추출·파싱 코드와 결과 데이터셋을 GitHub으로 공개해 개방적 연구 실천에 기여

감사합니다

신은선 · 한국학중앙연구원 인문정보학 박사과정

한국 고문헌 메타데이터의 시맨틱 표준화 연구