

朱

박사학위논문 예비발표

조선총독부박물관 문서의 데이터 모델링과 편찬

불국사 관련 기록의 편중과 결락

정송이 · 한국학중앙연구원 인문정보학

2026. 08. 28.



『朝鮮古蹟圖譜』 4, 朝鮮總督府, 1916, p.462.

1 한 사찰의 기록이 어디에 있는지 알 수 없다

공개 목록과 실제 편철의 거리

592

건

공공데이터포털이 공개한 문서철 목록

41,913

건 · 116,983면

상세페이지를 전수 수집해 확보한 서지

👁️ 조선총독부박물관 문서 서지 목록

<https://github.com/Song-yiJung/jggk-museum-catalog>

1,033

면

불국사 관련 기록 · 26개 문서철

공개 목록 592건 가운데 제목에 「불국사」가 나오는 것은 두 건이다.

실제로는 26개 문서철에 걸쳐 있다. 제목만 보고는 나머지 24개에 무엇이 들어 있는지 알 수 없다.

1,033면은 현 시점의 범위이며, 서지에 드러나지 않는 기록이 더 있을 가능성은 남는다.

판독의 단위인 면에는 적을 자리가 없다

자료의 네 계층과 식별자

계층	단위	무엇인가	식별과 기술
1	卷 문서철	쇼와 4~8년 고적보존공사	pseq=4297
2	案件 문건	불국사 사리탑 복구 공사	문건명
3	文書 문서	기안 한 건, 전보 한 통	A059-024-078
4	面 면	스캔 이미지 한 장 · 판독의 단위	기술할 자리 없음

면은 이미지 파일 이름으로 구별되지만, 그 면에 관하여 적을 자리가 목록에 없다.
어느 면이 어느 면의 정서본인지, 어느 면이 사진이고 어느 면이 도면인지는 문서 단위 서지에 기록되지 않는다.

이 사실이 III장에서 면 간 관계를 설계하는 근거가 된다.

같은 대상을 어디에서 보았는가

선행연구는 세 자리에서 보았다. 이 연구는 넷째 자리에 선다

사찰의 편

한상길 2019

근현대 불교사 · 제도사

사리탑의 귀환과 대규모 법회, 보물 공인을 계기로 불국사가 사격(寺格)을 회복해 본산 승격 청원 운동을 전개한 과정을 실증했다.

→ 같은 시기 행정이 그 사찰의 무엇을 기록했는가를 본다.

대상의 편

신은영 2018

탈식민주의 미술사 · 담론비평

반출·반환·재이전의 경로를 추적해 반환에 작동한 논리를 드러내고, 1970년대 원위치 이전 결정에도 식민 사관이 잔존함을 지적했다.

→ 경위를 다시 쓰지 않고, 그 서술이 어느 면에 근거하는지를 되짚는다.

제도와 자료의 편

오영찬

개념사 · 기록관리론

「‘고적’의 탄생」 — 1914년경 원고의 史蹟을 붉은 선으로 지우고 古蹟으로 고친 사실을 밝혔다.

「분류 체계 시론」 — 기능적 출처에 따른 4단계 분류를 제안했다.

→ 분류는 나무이고 이 연구는 그래프다.

그 밖에 검토 — 이현일·이명희(같은 문서철, [고적조사] 57권·[지정] 47권) · 허형욱 2012 · 이성시 · 김신재

여덟 편 모두 이 문서를 근거로 삼는다. 그러나 문서 자체를 다시 쓸 수 있는 데이터로 만든 연구는 없다.

업무별로 편철되어 흩어져 있는 조선총독부박물관의 불국사 관련 기록을 **대상 단위로 연계하여 활용할 수 있는 방안** 을 수집·설계·편찬의 세 단계로 제시한다.

물음 1

행정은 불국사의 어느 대상에 기록을 남기고 어느 대상에는 남기지 않았는가

물음 2

남은 기록에서 무엇이 기재되지 않았는가

물음 3

선행연구의 어느 진술이 어느 면에 근거하는가

없음을 확인하려면 현재 전하는 불국사의 유산 목록과 견주어야 한다. 대조의 기준을 함께 밝힌다.

연구의 구상

자료의 상태가 방법을 정하고, 방법이 물을 수 있는 것을 정한다

연구자료	본 연구의 내용	본 연구의 활용
업무별로 편철되어 한 사안이 여러 문서철에 흩어져 있다	면마다 식별자를 부여하고 면과 면, 면과 대상의 관계를 적는다	어느 대상에 기록이 물리고 어느 대상에는 남지 않았는가
붓글씨와 등사본이 섞여 있고 판독의 근거가 다른 면에 있다	확인한 글자와 추정한 글자를 나누고 추정의 근거가 된 면을 가리킨다	선행연구의 어느 진술이 어느 면에 근거하는가
인쇄된 서식이 적을 자리를 미리 정해 두었다	기입된 칸과 비어 있는 칸과 이 면에 없는 칸을 갈라 적는다	행정이 적을 자리를 두고도 적지 않은 것은 무엇인가
사진과 도면에는 기입된 글자가 없다	무엇을 담은 면이고 어느 대상에 관한 것인지를 적는다	문자가 아닌 자료도 같은 질의에 걸리는가

가운데가 이 연구가 하는 일이다. 가운데를 건너뛰면 오른쪽 물음은 인상에 머문다.

판독과 해석을 갈라 적는다

근거의 범위가 다른 것을 같은 층에 두지 않는다

sourceDoc

종이에 있는 것

보이는 글자만 적는다. 해석 태그를 두지 않는다.

```
<line xml:id="l_061_05"
      facs="#z_base">1.0</line>
```

text / body

그 글자가 무엇인가

연구자의 판단. 단위와 대상이 여기서 붙는다.

```
<measure type="height"
          unit="尺" quantity="1.0">
  1.0</measure>
```

standOff

면과 대상 사이

면이 무엇을 찍고 무엇을 설계한 것인지.

```
<relation name="設計"
          active="#...061-001"
          passive="#obj_tapgak"
          cert="high"/>
```

도면에서 종이에 적힌 것은 1.0이라는 숫자뿐이다. 그것이 콘크리트 층의 두께라는 것은 연구자의 판단이다. 두 층은 xml:id와 @corresp로 이어져 있어, 나중에 판단만 고쳐도 판독은 그대로 남는다.

무엇을 말하고, 어떤 종이이고, 어떻게 이어지는가

세 갈래로 나눈 통제 어휘

taxonomy_term.xml

개념어 45

무엇을 말하는가

행위·처분 · 공간 이동·설치
가치·보존 · 의례 · 행정
대상의 성격 규정

taxonomy_form.xml

서식 12 · 구획 8

어떤 종류의 종이인가

起案紙 本文紙 電報紙 記文
申込書 評價書 目錄 書簡
印刷物 圖面 寫眞 收受電報

taxonomy_relation.xml

관계 9

어떻게 이어지는가

寫眞 設計 參照 載置 保護
奉安 寫譯 淨寫 樣式
원문 용어 7 · 연구자 조어 2

어휘 계열 — 45개 개념어 가운데 41개에 부여



15 admin 행정 절차의 말



9 aesthetic 감상·수집의 말



8 religious 종교적 성격의 말



8 neutral 어느 쪽도 아닌 말



1 misidentified 잘못 규정한 말

misidentified 계열에 속한 유일한 말은 石燈이다. 이 분류는 자료에 있는 사실이 아니라 연구자의 판단이므로, 개념어마다 정의와 실제 용례의 출처를 함께 적어 판단의 근거를 추적할 수 있게 했다.

방법의 시연이며 연구 결과가 아니다

16건 47면

인코딩을 마친 표본

12종

한 표본 안에 섞인 서식

오류 0

2026-08-27 스키마 검증

5면

배포 자료에 담은 샘플

서식	면	서식	면	서식	면
本文紙	17	評價書	2	收受電報	1
起案紙	7	目錄	2	書簡	1
記文	6	申込書	2	印刷物	2
電報紙	4	圖面	2	寫眞	1

무작위 표본이 아니다. 행정문서·서간·전보·평가서·조사서·도면·사진이 모두 들어가도록 구성했다.
이질적인 자료를 하나의 스키마로 다룰 수 있는지 확인하는 것이 표본의 목적이었다.

9

한 대상이 불린 이름들

같은 석조물을 문서마다 다르게 부른다

시점	명칭	출처	시점	명칭	출처
1904-02	石燈	韓國建築調查報告	1933-08	佛舍利石塔	A059-024-053-003
1933-02	石燈形浮屠	A059-024-003-001	1933-08-15	佛國寺舊藏舍利石塔	A059-024-053-001
1933-02-08	佛國寺浮屠	A059-024-001-001	1933-09	寶塔・古塔	A059-024-077-001
1933-07-21	佛國寺基塔	A059-024-004-001	1934-08-27	佛國寺舍利塔	H007-002-086-001
1933-07-21	舍利塔	A059-024-023-007	1963-01-21	경주 불국사 사리탑	국가유산청
1933-07-22	古塔	A059-024-022-005			

1904년의 石燈은 등(燈)이라는 규정이며 사리를 담는 탑과 성격이 다르다.
음영은 1933년 — 반년 사이에 일곱 이름이 오갔다. 1934년 조사서에서 이름이 굳고, 현재의 국가 지정 명칭은 그 굳은 이름을 잇는다.

명칭으로만 존재하는 대상

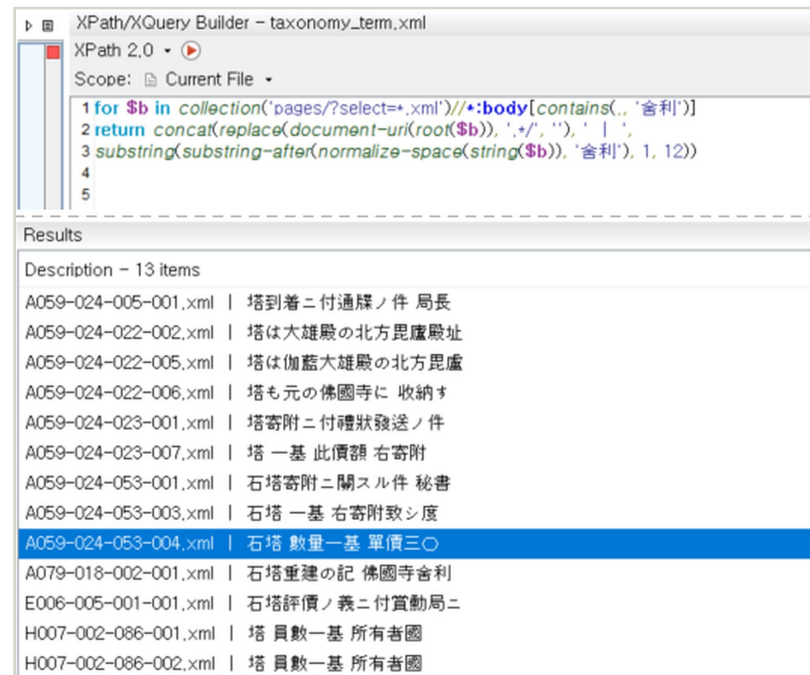
舍利는 47면 중 14면에 26회 나타난다

행정이 기재한 항목	그 항목이 가리키는 것
一基 — 수량	탑
價額・單價 — 가격	탑
所有者 — 소유자	탑
到着・收納 — 이동	탑
舍利	어느 항목에도 오르지 않는다

26회 모두舍利塔・舍利石塔이라는 대상의 명칭이며,
사리 자체를 가리킨 곳은 없다.

이는 「사리에 관한 기록이 없다」와 다른 진술이다. 언급이 없는 것이 아니라,
명칭의 일부로만 존재하고 행정 절차의 어느 항목에도 대상으로 오르지 않는다.

이 구별이 V-2 「기록의 결락」이 하려는 일이다.



oXygen XPath 2.0 — 본문에舍利가 있는 면 조회, 13면

인쇄된 칸이 비어 있다는 사실 자체가 데이터

기안지 A059-024-001-001의 서식란

未完結	完結	種	施行月日番號	接受	鈐印
關係番號	起案	發送	記帳	校合	淨寫
決裁	部局長	課長	主任	件名	總督・政務

☐ 값이 기입된 칸

☐ 칸은 있으나 비어 있음

☐ 이 면에는 그 칸이 없음

184

표본 47면에서 서식이 요구한 칸

103

값이 기입된 칸

81

칸은 있으나 비어 있는 칸

빈칸은 작성자가 쓰지 않아 생긴 것이 아니라 종이에 미리 인쇄되어 있던 자리다.

기안지의 열아홉 칸은 1912년 훈령 제36호 「조선총독부 및 소속관서 공문서규정」에 근거한 관제 용지다.

읽지 못한 것과 적히지 않은 것을 가른다

표본 47면의 판독 상태

gt-confirmed 원본 대조 확정



32면

ocr-only 판독 미확정



12면

도면 영역 구획 후 기입 문자 기록



2면

no-text 적을 것이 없음



1면

211

unclear — 불확실한 판독 (cert="low"
4)

40

gap — 남아 있으나 읽어 내지 못함

3

gap — 물리적으로 잘려 나감

5

supplied — 다른 면에서 보충

확신도를 적지 않으면 못 읽은 것과 안 적힌 것이 데이터에서 똑같이 「없음」으로 보인다.
그러면 「행정이 기록하지 않았다」는 결론이 사실은 판독 실패의 결과일 수 있다.

같은 데이터에 두 가지로 묻는다

XPath는 한 면 안에서 글자를 찾고, SPARQL은 면을 넘어 대상을 따라간다

Active ontology x Entities x Classes x Individuals by class x OntoGraf x SPARQL Query x

SPARQL query:

```
PREFIX jg: <https://github.com/Song-yiJung/jgk-museum-catalog/sarira/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT ?date ?form ?source WHERE {
  jg:obj_sarira jg:attestation ?a .
  ?a jg:form ?form ; dct:date ?date ; dct:source ?source .
}
ORDER BY ?date
```

date	form	source
"1904-02"	"石燈"@ja	"韓國建國調查報告"
"1933-02"	"石燈形浮屠"@ja	"A059-024-003-001"
"1933-02-08"	"佛國寺浮屠"@ja	"A059-024-001-001"
"1933-07-21"	"佛國寺基塔"@ja	"A059-024-004-001"
"1933-07-21"	"舍利塔"@ja	"A059-024-023-007"
"1933-07-22"	"古塔"@ja	"A059-024-022-005"
"1933-08"	"佛舍利石塔"@ja	"A059-024-053-003"
"1933-08-15"	"佛國寺蓋藏舍利石塔"@ja	"A059-024-053-001"
"1933-09"	"寶塔"@ja	"A059-024-077-001"
"1933-09"	"古塔"@ja	"A059-024-077-001"
"1934-08-27"	"佛國寺舍利塔"@ja	"H007-002-086-001"
"1935-12-18"	"佛國寺旧藏 / 舍利石塔"@ja	"E006-005-001-001"
"1963-01-21"	"Stupa of Bulguksa Temple, Gyeongju"@en	"국기유산형"
"1963-01-21"	"慶州 佛國寺 舍利塔"@ko	"국가유산형"
"1963-01-21"	"경주 불국사 사리탑"@ko	"국가유산형"

Protégé SPARQL Query — 실제 실행 화면. 앞 화면의 명칭 변천표가 이 질의의 출력이다.

질의문 — 옮겨 적은 것

```
SELECT ?date ?form ?source
WHERE {
  jg:obj_sarira jg:attestation ?a .
  ?a jg:form ?form ;
    dct:date ?date ;
    dct:source ?source . }
ORDER BY ?date
```

TEI — 판단과 근거를 함께 적는다

```
<supplied resp="#JSY"
  source="#...-022-005" cert="low">
  長尾欽弥氏別邸に</supplied>
<gap reason="illegible"
  quantity="2"/>
```

왼쪽 화면은 표본 47면에서 만든 RDF 1,873 트리플에 질의를 돌린 결과다. 오른쪽 TEI 발체를 보면, 이 글자는 이 면에서 읽은 것이 아니라 정서본 022-005에서 가져왔고 확신이 낮다는 사실이 함께 적혀 있다.

한 벌인 것을 한 벌로 적는다

봉투 · 사진 · 설명이 각각 다른 문서로 편철되어 있다



붉은 점선 — 아직 데이터에 없는 관계

078은 사진이지만 무엇을 찍은 것인지 자기 파일 안에서는 알 수 없다. 답은 076 봉투의 표제에 있는데 둘을 잇는 줄이 없다. 표제가 사진을 명시하므로 078은 medium, 설명을 명시하지 않으므로 077은 low로 나누었다.

관계 어휘 아홉 종에는 봉투와 내용물을 잇는 말이 없어 새로 세워야 한다.

구축한 데이터로 무엇을 하는가

V장 구축 데이터의 활용

1

기록의 분포와 편중

대상·시기·서식의 세 축으로 집계하여 1,033면이 어느 대상에 몰리고 어느 대상에는 남지 않았는지 확인한다.

2

기록의 결락

문서가 한 장도 남지 않은 대상, 빠지거나 잘려 나간 면, 서식이 기입란을 두었음에도 비워 둔 칸의 세 층위에서 확인한다.

3

사리탐 사안 기록의 소재

흩어진 낱장을 면 단위로 배열하여 선행연구의 어느 진술이 어느 면에 근거하고 어느 대목에는 근거가 없는지 확인한다.

행정 기록은 대상의 가치에 비례하지 않고 수리·이전·반출과 같은 사무가 발생한 대상에만 남는다.
따라서 면수의 많고 적음이 무엇을 가리키는지를 함께 밝힌다.

I 서론

1. 연구 목적 및 배경
2. 선행연구 검토
3. 연구 방법

II 자료의 수집과 고증

1. 자료의 수집
2. 자료의 계층과 식별자
3. 연구 대상의 고증

III 온톨로지 설계

1. 선행모델 검토
2. 클래스 설계
3. 관계 설계
4. 기술 요소 설계

IV 데이터 모델링

1. 모델링 범위와 절차
2. TEI 인코딩
3. RDF 변환

V 구축 데이터의 활용

1. 기록의 분포와 편중
2. 기록의 결락
3. 사리탑 사안의 재구성

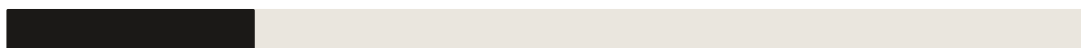
VI 결론

1. 연구 결과
2. 연구의 의의
3. 한계와 후속 과제

II에서 V까지가 수집 · 설계 · 편찬 · 활용의 네 단계에 대응한다. 항(項)은 자료가 확정된 뒤에 붙인다.

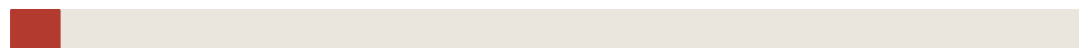
판독을 기다리지 않고 진행할 수 있는 층위부터

236 / 1,033면



판독 — 하루 5면, 남은 기간 약 160일

46 / 1,033면



서식 판별 — 판독 없이 진행 가능

서식 판별 전수

V-1의 구성이 여기서 정해진다. 가장 먼저 착수한다.

우선 처리 대상

「경북 경주군 내동면 불국사」 342면 · AI OCR 14면 · 유리건판 전거 정리

수치의 정리

대상목록과 OCR 대장의 차이, 석굴암 단독 9건 20면의 처리

공개 계획

서지 목록 선공개, 판독·TEI·시맨틱 데이터는 제출 후

표본 16건 47면은 IV장 방법의 시연이며 V장의 결과가 아니다.

단행본

국립중앙박물관, 『국립중앙박물관 보관 고문서목록』, 국립중앙박물관, 1996.

국립중앙박물관, 『광복이전 박물관 자료목록집』, 국립중앙박물관, 1997.

關野貞, 『韓國建築調査報告』, 1904.

이순자, 『일제강점기 고적조사사업 연구』, 경인문화사, 2009

朝鮮總督府, 『朝鮮古蹟圖譜』 1권, 조선총독부, 1915.

朝鮮總督府, 『朝鮮古蹟圖譜』 4권, 조선총독부, 1916.

朝鮮總督府, 『佛國寺と石窟庵』, 조선총독부, 1938.

황수영편, 『일제기 문화재 피해 자료-고고미술자료 제22집』, 고고미술동인회, 1973.

황수영 편, 이양수·이소령 증보, 강희정·이기성 해제, 『일제기 문화재 피해자료』, 국외소재문화재단, 2014.

학위논문/학술논문

김계원, 「유리건판 사진으로 보는 고미술: 세키노 타다시(關野貞)의 고적조사사업을 중심으로」, 대동문화연구 제114호, 성균관대학교 대동문화연구원, 2021, 7~36쪽.

김도형, 「일제하 총독부박물관 문서와 관리 체계」, 기록학연구 제3호, 한국기록학회, 2001, 115~137쪽.

김대길, 「국사편찬위원회의 고문서 조사·수집의 성과와 과제」, 고문서연구 제31호, 한국고문서학회, 2007, 15~47쪽.

김대환, 「조선총독부고적조사 사업에서 후지타 료사쿠(藤田亮策)의 역할국립중앙박물관 소장 조선총독부박물관 공문서의 검토-」, 한국상고사학보 제91호, 한국상고사학회, 2016, 121~141쪽.

_____, 「일제강점기 조선고적조사사업과 한국고고학사」, 한국상고사학보 제97호, 한국상고사학회, 2017, 79~100쪽.

김동하, 「일제강점기 경주지역 불교유적조사와 경주고적보존회의 '慶州古蹟及遺物調書」, 불교미술사학 15, 불교미술사학회, 205-233쪽.

김바로, 「제도와 인사의 관계성 데이터 아카이브 구축과 활용 : 근대 학교 자료(1895~1910)를 중심으로」, 한국학중앙연구원 한국학대학원 박사학위논문, 2017.

_____, 「역사기록의 전자문서 편찬방법 탐구-역사요소를 중심으로」, 열상고전연구 제51호, 열상고전연구회, 2016, 203~223쪽.

강희정, 「일제강점기의 조선 불교미술 조사와 복원 -한국 불교미술 연구의 출발점-」, 미술사와 시각문화9, 미술사와 시각문화학회, 2010, 146-173쪽.

- 김 현, 「한국古典籍의 전산화의 성과와 과제: 《조선왕조실록 CD-ROM》개발 사업의 경과와 발전 방향」, 민족문화 18, 한국고전번역원, 1995, 70~102쪽.
- _____, 「고문헌 자료 XML 전자문서 편찬 기술에 관한 연구」, 고문서연구 제29권, 한국고문서학회, 2006, 183~230쪽.
- 김 현·강혜원, 「고고학술정보 데이터 아카이빙 전략과 구현 방안-한국고고학사전의 디지털 전환을 위한 시맨틱 데이터 아카이브 제안-」, 고고학술정보디지털대전환 국제학술심포지엄, 국립문화유산연구원, 2022, 126~147쪽.
- 권혁산, 「국립중앙박물관 소장 조선총독부박물관 자료의 공개와 정리」, 미술사연구 제40호, 미술사연구회, 2021, 7~24쪽.
- 오영찬, 「조선총독부박물관 문서의 분류 체계에 대한 시론」, 미술자료 제96호, 국립중앙박물관, 2019, 181~208쪽.
- _____, 「‘고적’의 탄생-일제 시기 문화재의 제도화와 식민주의」, 한국문화연구 45, 이화여자대학교 한국문화연구원, 2023, 109-139쪽.
- _____, 「조선총독부박물관 문서의 분류 체계에 대한 시론」, 미술자료 96, 국립중앙박물관, 2019, 181-208쪽.
- 이기성, 「국립중앙박물관 일제강점기 자료 공개의 향후 사업 방향에 대한 제언」, 고고학지 제28집, 국립중앙박물관, 2022, 5~27쪽.
- 이현일·이명희, 「조선총독부박물관 공문서로 본 일제강점기 문화재 등록과 지정」, 미술자료 85호, 국립중앙박물관, 2014, 95~111쪽.
- 신은영, 「불국사 사리탑의 流轉과 식민주의」, 미술사학연구 298, 한국미술사학회, 2018, 127-153쪽.
- 정송이·김바로, 「조선총독부박물관 문서 XML 데이터 설계 및 활용 시론 —XML 기반 다중 분류 체계 및 내·외부 데이터 연결을 중심으로—」, 인문과학 96, 성균관대학교 인문학연구원, 2025, 181-227쪽.
- 차윤정, 「근세 이후의 불국사 연구: 건축, 상, 인식」, 서울대학교 대학원, 2026.
- 허형욱, 「佛國寺 石造舍利塔의 造成時期와 性格 考察」, 동원학술논문집 13, 국립중앙박물관, 한국고고미술연구소, 2012, 53-80쪽.
- 한상길, 「근현대 불국사의 사격」, 대각사상 32, 대각사상연구원, 2019, 137-173쪽.
- Roth, H.D. (1993) “Text and edition in early Chinese philosophical literature”, *Journal of the American Oriental Society*, 113(2), pp. 214–227. doi: 10.2307/603026.
- Thorpe , K. (2024) “Designing Indigenous-Led Archival Futures: The Application of Indigenous Research Methodologies Within Archival Research and Practice”, *Archives & Manuscripts*, 52(1), pp. 29-44. doi: 10.37683/asa.v52.10991.

표준 · 데이터

정송이, 조선총독부박물관 문서 서지 목록, 2025. <https://github.com/Song-yiJung/jggk-museum-catalog>

국립중앙박물관, 문화체육관광부 국립중앙박물관_조선총독부박물관 문권정보, 공공데이터포털, 2021 등록·2025 수정. <https://www.data.go.kr/data/3070537/fileData.do>

TEI Consortium, TEI P5: Guidelines for Electronic Text Encoding and Interchange 4.12.0, 2026. <https://tei-c.org/Vault/P5/4.12.0/doc/tei-p5-doc/en/html/>

Chrysoula Bekiari 외 편, Definition of the CIDOC Conceptual Reference Model 7.1.3, ICOM/CIDOC CRM Special Interest Group, 2024. <https://cidoc-crm.org/>

Syncro Soft, <oXygen/> XML Editor 17.0 (build 2015072912), 2015. <https://www.oxygenxml.com/>

Stanford Center for Biomedical Informatics Research, Protégé 5.5.0, 2019. 3. 14. <https://protege.stanford.edu/>

본 발표물의 디자인 및 이미지는 클로드(Opus5)를 활용해 제작·편집되었습니다. 발표자는 AI가 생성한 결과물을 직접 검수하였으면 최종 편집 및 내용에 대한 책임은 전적으로 발표자에게 있습니다.