

# 국가전거를 활용한 단위 도서관의 전거데이터 소급 구축 연구\*

## An Exploratory Study on the Construction of Retrospective Authority Data with National Authorities

노 지 현 (Jee-Hyun Rho)\*\*

### 목 차

- |                         |                          |
|-------------------------|--------------------------|
| 1. 서론                   | 4. 시범사업의 방법론적 한계와 정책적 제언 |
| 2. 전거데이터의 구축과 활용        | 5. 결론                    |
| 3. 국가전거를 활용한 소급전거 시범 구축 |                          |

### 초 록

최근의 목록 환경에서 전거데이터의 기능이 특히 중요해지고 있다. 이에 따라 국내에서는 2019년부터 국가전거데이터의 공개와 공동 활용을 위한 시범사업을 추진하면서 국가전거를 활용한 단위 도서관의 전거데이터 소급 구축 방안을 모색하고 있다. 이 연구에서는 지난 2019년 국립중앙도서관과 KERIS에서 공동으로 추진했던 전거데이터 소급 구축을 위한 시범사업의 방식과 결과를 분석하고, 이 과정에서 나타난 방법론적 한계와 문제점을 분석한 후 보다 많은 도서관들이 국가전거를 활용하여 전거데이터의 소급 구축을 효율적으로 추진할 수 있도록 정책적 제언을 도출하는데 목적을 두었다. 이를 위해 시범사업에 참여한 A대학도서관을 사례로 하여, 전거데이터의 자동 생성률과 정확률, 그리고 자동 생성 과정에서 나타난 오류와 누락의 사례와 원인을 분석하였다. 연구결과, 국가전거를 이용한 전거데이터의 자동생성률은 20% 남짓에 불과했고, 전거형의 형식에 따라 편차가 큰 것으로 나타났다. 이러한 낮은 생성률과 오류, 누락은 종합목록이나 단위 도서관의 서지레코드 품질에서 비롯된 것이었고, 이에 따라 기존 데이터에 대한 보다 세밀한 검토와 함께 정교한 알고리즘을 개발, 적용할 필요가 있음을 확인할 수 있었다.

### ABSTRACT

In the recent cataloging environment, the function of authority data is becoming more important. In Korea, a pilot project for sharing and joint use of the national authority data has been promoted since 2019, and a plan to automatically construct retrospective authority data using national authorities is in progress. This study aims to (1) analyze the results of the pilot project for retrospective authority data jointly promoted by the National Library of Korea and KERIS in 2019, (2) explore methodological limitations and problems in the process, and (3) derive policy proposals for constructing authority data more efficiently. To this end, a university library participating in the pilot project was investigated as a case study. Based on it, the automatic generation rate and accuracy of authority data, and errors and omissions cases and causes were analyzed in detail. As a results of the study, it was found that the quality of union catalogs and bibliographic records had a great influence on automatically constructing authority data, and supplementary methods were needed to minimize these limitations.

키워드: 전거제어, 전거데이터, 국가전거제어, 소급 전거데이터

Authority Control, Authority Data, National Authority Data, Retrospective authority data

\* 이 과제는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

\*\* 부산대학교 문헌정보학과 교수(jhrho@pusan.ac.kr / ISNI 0000 0004 6484 8385)

논문접수일자: 2023년 1월 16일 최초심사일자: 2023년 2월 7일 게재확정일자: 2023년 2월 8일

한국문헌정보학회지, 57(1): 161-182, 2023. <http://dx.doi.org/10.4275/KSLIS.2023.57.1.161>

© Copyright © 2023 Korean Society for Library and Information Science

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.

## 1. 서론

### 1.1 연구배경과 목적

최근 몇 년간 도서관계는 도서관 데이터의 활용성 강화를 위해 링크드 데이터와 시맨틱 웹 기술을 어떻게 활용할지 깊이 있는 논의를 전개하고 있다. 지금까지의 논의를 살펴보면, 링크드 데이터 환경에서 가장 강점으로 활용될 수 있는 도서관 데이터는 전거데이터라는데 이견이 없는 듯하다(Lampron & Wacker, 2019; Zhu, 2019). 전거데이터는 특정 개체와 관련되는 자원을 정확하게 식별하고 검색할 수 있도록 함으로써 자원에 대한 효율적인 접근을 제공하고, 다른 개체나 서비스가 연결된 거대한 지식 웹을 만드는데 있어 근간(backbone)이 되는 데이터로 인식되기 때문이다. 그러나 전통적인 전거데이터는 사람을 위해 만들어진 것으로 시맨틱 웹에 통합될 수 없는 구조로 되어 있다. 이러한 이유로 기존 데이터 형식을 시맨틱 웹에서 활용할 수 있는 링크드 데이터 형식으로 변환하고, 국제적으로 상호 연결된 데이터 세트를 지향하는 움직임이 이루어지고 있다.

이러한 국제적 추세와 달리, 국내에서는 국립중앙도서관과 일부 대학도서관을 제외하고 대다수 도서관에서 여전히 전거데이터를 구축하지도, 활용하지도 못하고 있다. 원인은 다양하다. 시스템 측면에서 애초에 서지데이터와 전거데이터 간의 연계 관계를 제대로 확립하지 않은 채 서지데이터 구축에만 집중해왔고, 제도적 측면에서 전거데이터 구축을 위한 표준을 정비하지 못했으며, 여기에 상당한 시간과 노력이 요구되는 전거업무에 충분한 인력과 예산을 투입

할 수 없었던 도서관의 여건이 결합되면서 전거데이터 문제는 계속 미루어져 왔다. 나아가, 전거데이터의 가치나 효용성에 대한 부정적 인식도 지금까지 이를 등한시해 온 원인으로 작용하였던 것이 사실이다. 그러나 검색효율을 제고하고, BIBFRAME과 같은 새로운 서지모델의 도입에 선제적으로 대비해야 하며, 무엇보다 도서관 데이터의 확장성과 상호운용성을 확보해야 하는 상황에서 전거데이터 문제는 더 이상 미룰 수 없는 과제가 되었다.

그동안 국내 연구자들은 이를 해결하기 위한 현실적인 방안으로 국가기관 주도의 협력체제 구축과 국가전거데이터의 공동 활용을 꾸준히 주장하여왔다(오동근, 2000; 심경, 2006; 이미화, 2012). 이에 대한 가시적인 움직임은 비교적 최근에서야 나타나고 있다. 국립중앙도서관에서 2017년부터 인명, 단체명을 시작으로 ‘국가선거’ 데이터를 확대 구축하고, 2018년 KERIS 및 대학도서관 실무진과 선거TF를 구성한 후 2019년부터 국가선거 반입을 위한 시스템 개발과 소급 선거 구축을 위한 시범사업을 순차적으로 진행해 나가고 있기 때문이다(국립중앙도서관, 2019). 비록 협력에 기반한 전거데이터의 ‘공동구축’ 체제를 갖춘 것은 아니지만, 시범사업의 성과가 어느 정도 축적되면 국내에서도 전거데이터 공유를 위한 생태계가 정착될 것으로 보인다. 그러나 국가선거의 공동 활용과는 별개로, 개별 도서관 차원에서 추진해 나가야 할 과제가 많이 남아 있다. 자관 서지데이터에 국가선거를 연결하고, 자관의 요구에 맞춰 데이터를 보완하거나 커스터마이징(customizing) 하면서 이를 활용한 서비스 모델도 창출해야 하기 때문이다.

이 연구는 국가선거데이터를 활용하여 단위 도서관에서 선거데이터를 효율적으로 '소급' 구축하는 과정을 탐색하는데 목적을 두고 있다. 국가선거의 관점에서 다양한 유형의 데이터를 병합하면서 데이터의 품질을 강화하기 위한 노력이 지속적으로 이루어져야 할 것이지만, 이와 동시에 단위 도서관 차원에서도 국가선거를 활용하여 소급 서지에 선거를 어떻게 구축 혹은 연결할지에 대한 현실적인 방법을 마련해야 한다. 특히, 선거 구축은 단위 도서관에서 단독으로 해결할 수 있는 단순한 문제가 아니기 때문에 이를 위한 합리적인 방안을 공동으로 모색할 필요가 있다. 이에 이 연구에서는 첫째, 국립중앙도서관과 KERIS의 협력으로 선거데이터의 일괄 구축을 추진한 시범사업의 방법론적 한계와 문제점을 분석하고, 둘째, 시범사업의 성과를 안정적으로 정착시키고 앞으로 국내의 많은 도서관들이 국가선거를 활용하여 선거데이터의 소급 구축을 추진할 수 있도록 방법론적 제언을 하고자 한다. 연구에 필요한 데이터는 문헌연구와 국립중앙도서관의 국가선거 공유 시범사업에 참여한 A대학도서관 실무진과의 면담, 그리고 A대학도서관에서 일괄 반입을 통해 구축한 선거데이터에 대한 분석 등을 활용하였다.

## 1.2 선행연구

선거제어의 방향이나 선거 기술규칙을 다룬 연구를 제외하고, 선거데이터의 구축에 관한 연구를 살펴보면 다음과 같다.

오동근(2000)과 심경(2006)은 선거데이터의 구축에 따르는 개별 도서관의 어려움을 고려하

여 도서관 상호협력에 의한 선거 구축 방안을 제안하였다. 이를 위해 협력 기반의 국가선거 또는 공유 선거파일을 구축하고, 로컬도서관에서 이를 다운로드하여 사용하거나 필요한 경우 자관의 실정에 맞춰 변환하여 사용할 것을 제안하였다. 김미향(2009)은 대학도서관 중심의 협력적 선거데이터 구축 방안을 제안하였다. 특히, 이 연구에서는 기관별로 상이한 선거제어 규칙을 적용하는 현실을 고려하여 통일된 선거형의 사용 대신에 모든 접근점을 대상으로 하는 '선거 클러스터링'을 구축하는 방안을 제안하였다. 이어, 김미향과 김태수(2010)는 기구축된 선거데이터를 활용하여 선거형 개념을 배제한 접근점제어 데이터를 구축하고, 기관 리포지터리 시스템에서 이를 활용하는 방안을 제안하였다. 이미화(2012)는 국가서지기관에서 자국의 선거제어에 대한 책임을 가져야 하며, 개별 도서관은 자관에서 구축한 폐쇄적인 선거데이터의 사용에서 벗어나 협력적으로 선거데이터를 운용할 필요가 있다고 제안하였다. 오삼균 등(2018)은 VIVO 온톨로지와 ISNI를 활용한 국가선거 데이터 모델과 한국 상황에 적합한 협력형 선거 구축 모델을 제안하였다.

이처럼 국내 관련 연구에서는 선거데이터의 공동 사용을 전제로 하는 국가선거 혹은 협력형 선거파일의 구축 방안을 비중 있게 다루어 왔다. 이에 반해 단위 도서관에서 국가선거나 협력형 선거파일을 활용하여 엄청난 양의 소급 데이터를 어떻게 구축할지에 대한 방법론적 탐구는 지금까지 거의 이루어지지 못하였다. 국가선거데이터의 공동 활용에 대한 사서들의 인식을 조사한 이성숙(2021)의 연구가 있지만, 이 연구는 국가선거를 활용하여 선거업무를 수행하는 과정에서

경험한 어려움을 분석하는데 중점을 두었다.

한편, 국외 연구에서는 단위 도서관에서 시도해 온 다양한 전거 처리 사례를 찾아볼 수 있다. Zhu와 Von Seggern(2005)은 벤더가 제공하는 자동화된 전거 처리 과정을 소개하고, 자동화를 통해 실행 가능한 부분과 로컬에서 처리할 부분을 구분하여 분석하였다. Ashmann(2002)은 벤더에 의한 자동 처리와 자관의 전거 제어 팀이 수행했던 업무를 비교 분석한 후 아웃소싱이 반드시 시간절감의 효과가 있었던 것이 아님을 밝혔다. 또한, 기 구축된 전거데이터의 품질향상을 위한 후속 처리 과정을 자세히 다룬 연구도 다수 있는데, Williams(2010)는 자동화된 방법으로 자관 전거데이터를 정제(cleanup)한 사례를 소개하였다. Kreyche와 Lisius, Park(2010)은 국가선거에 새로 추가되는 개인의 사망일자 정보를 자관 전거데이터에 자동 반입하는 방법을, Mak(2013)은 RDA 표준을 적용하여 변경된 LC의 이름 전거레코드 형식을 자관에 일괄 적용하는 방법을 연구하였다. Wolf(2020)는 종합목록시스템인 OCLC Connexion과 도서관 자료관리시스템(ILS), 외부 벤더, 무료 소프트웨어인 MarcEdit 등을 활용하여 UNT(University of North Texas) 도서관의 전거 제어 업무를 자동화한 사례를 소개하였다.

이렇듯 국외 연구에서는 단위 도서관 차원에서 국가전거를 활용하여 자관 전거를 효율적으로 구축하고 운영하는 방법을 다양하게 모색하고 있음을 알 수 있다. 특히, 전거데이터의 구축 이후에도 데이터의 변경이나 수정사항을 효율적으로 반영하기 위한 방법이 지속적으로 연구되고 있어 전거 처리가 일회성으로 끝나는 간단한 문제가 아님도 확인할 수 있다. 그러나 국

내의 경우 가용자원이나 환경 조건이 많이 다르고, 특히 전거데이터가 거의 구축되지 않은 상황에서 막대한 양의 데이터를 일괄 처리할 방법을 찾아야 하므로 국외의 사례를 단순 참고하기에는 한계가 있다.

## 2. 전거데이터의 구축과 활용

### 2.1 전통적 환경에서 전거데이터의 구축과 활용

전거제어(authority control)는 목록에 사용되는 접근점 형식을 제어하거나 관련된 접근점을 상호 연결하는 과정을 말한다. 이러한 전거 제어의 목적은 개체를 명확히 식별하고 상이한 이름을 가진 동일한 개체를 모이거나 동일한 이름을 가진 상이한 개체를 배타적으로 구분함으로써 검색의 효율을 높이는 데 있다. 전통적으로 전거제어의 목적이 접근점 제어를 통해 정확하고 정교한 검색을 지원하는데 있었다면, FR 개념모형(FRBR, FRAD)의 개발 이후에는 검색뿐 아니라 개체의 '식별'과 다른 개체와의 '연결' 기능이 강조되고 있다. 전거데이터에 관한 개념모형인 FRAD에서는 이를 명확히 구분하여 탐색(find), 식별(identify), 관계부여(contextualize), 근거제시(justify)를 전거데이터의 이용자과업으로 정리하였다.

일반적으로 전거제어는 동일한 이름이 여러 상이한 이름이나 이름 형식을 가질 때 이 가운데 하나를 '전거형'(전거형접근점)으로 선정하고, 나머지를 '이형'(이형접근점)으로 연결하며, 그 외 '식별정보'와 '정보원'(출처), 그리고

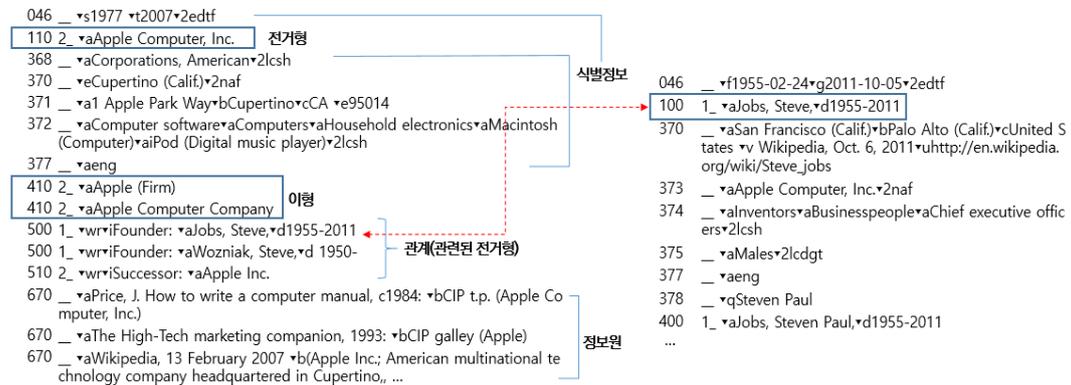
다른 개체와의 ‘관계’ 등 개체를 식별하거나 이해하는데 필요한 추가적인 메타데이터를 기록하는 과정으로 이루어진다. <그림 1>은 “Apple Computer, Inc.”에 대한 전거레코드의 예시이다. 여기에는 또 다른 전거형접근점인 “Jobs, Steve, 1955-2011” 등과의 관계도 표현되어 있다. 이러한 전거레코드는 서지레코드의 접근점에 전거형접근점을 기술하는데 근거자료로 활용되며, 서지레코드와 독립적으로 존재하지만 전거형접근점(또는 고유식별자)을 통해 서지레코드와 상호 연결되는 구조를 갖는다. 뿐만 아니라 관계를 통해 다른 접근점과도 연결되므로 탐색의 범위를 확장하거나 관련 정보를 식별하는 데에도 유용하게 활용할 수 있다.

전거레코드의 구축 모델은 (1) 기관에서 직접 구축하거나(로컬 전거파일), (2) 지역이나 국가 혹은 국제적으로 여러 기관이 협력하여 공동으로 구축하거나(공동 전거파일), (3) 권위 있는 국가서지기관에서 국가전거를 주도적으로 구축하여 공동으로 활용하는 방식(국가 전거파일)으로 구분할 수 있다. 분명한 것은, 현 시점

에서 로컬 전거파일보다는 협력이나 공유 기반의 전거파일이 전 세계적으로 선호되고 있으며, 이에 따라 전거제어 업무에 대한 국가서지기관의 책무가 더욱 강화되고 있다는 점이다. 이는 전거업무에 투입되는 인력과 예산을 절감하면서 양질의 데이터를 확보함과 동시에 로컬데이터가 가지는 고립(silo)을 막고자 하는데 주된 이유가 있다. 이러한 환경을 기반으로 하여, 국외 도서관들에서는 다양한 도구를 사용하여 전거데이터를 보다 효율적으로 구축하기 위한 방법을 개발하거나, 벤더에 의해 정해진 기간마다 배치처리(batch processing)하는 방식 등을 적용하면서 로컬도서관의 전거업무를 처리하고 있다(Wiederhold & Reeve, 2021).

## 2.2 링크드 데이터 환경에서 전거데이터의 활용 가치

전거데이터의 미래 활용 가치를 이해하려면 링크드 데이터 환경에 대한 이해가 선행되어야 한다. MARC 형식으로 관리되어 온 전거데이



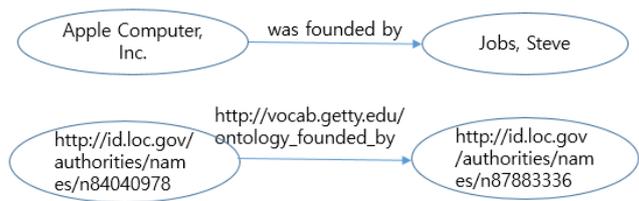
<그림 1> 전거레코드의 속성과 관계 기술 예시  
(출처: LC Authorities)

터는 도서관계에서만 사용되는 표준으로, 상당한 정보적 가치에도 불구하고 웹에서 활용이 어렵다는 단점을 가진다. 또한, 전거형접근점을 통해서만 관련 개체가 연결되고, 관련된 접근점 간의 관계 또한 전거레코드 내에 표현되어 있어 활용성이 떨어진다는 한계를 가진다. 이에 웹 중심의 데이터로 변모하기 위해 링크드 데이터의 데이터 모델인 RDF 형태로 변환하기 위한 시도가 이어지고 있다.

RDF는 데이터를 주어, 술어, 목적어의 트리플 형식으로 표현하고, 이러한 트리플의 집합을 그래프로 모델링한다. 가령, 앞서 <그림 1>의 “Apple Computer, Inc.”와 “Jobs, Steve, 1955-2011”의 관계는 다음 <그림 2>와 같이 표현된다. 전통적으로 사용해진 전거형은 개체를 명명하는 레이블로 사용될 뿐 주어, 술어, 목적어는 고유식별자인 URI나 IRI로 식별되며, 한 자원의 목적어는 다시 자원이 되어 속성과 속성값을 가지고 이 속성값이 다시 자원이 되면서 개체 간에 무한 연결이 이루어진다. RDF 트리플은 같은 식별자를 가지는 동일한 노드를 쉽게 병합할 수 있기 때문에 링크드 데이터 환경에서 상당한 유연성과 확장성을 가질 수 있다(노지현, 이미화, 이은주, 2023, 446-447). 따라서 도서관의 서지데이터나 전거데이터를 RDF 형태로 변환하는 시도는 “고립된 서비스 모델을 광범위한

커뮤니티 기반 모델로 대전환”하는 것을 의미하며(Schreur & Carlson, 2020), 이에 따라 거대한 웹 생태계에서 데이터의 가치를 엄청나게 상승시킬 것으로 예측되고 있다.

이러한 이유로, RDF 형식으로 표현된 전거 데이터는 링크드 데이터 환경으로 나아가기 위한 시작점(starting point)에 비유되기도 한다(Haffner, Young, & Neubert, 2011). 전거데이터가 웹의 다른 신뢰할 수 있는 자원에 연결되거나 통합되는 것은 물론이고, 박물관, 기록관, 상업기관(뉴스 에이전시, 서점 등) 등 다양한 기관에서 재사용될 수 있기 때문이다. 이렇게 될 경우 전거데이터는 도서관 안팎의 다양한 검색에 활용되면서 활용도가 크게 증가할 것이고, 웹에서도 양질의 데이터에 접근하기 위한 핵심 데이터로서 기능할 수 있게 된다. 또한, 데이터의 통합과 재사용을 통해 혁신적인 이용자 서비스를 개발하는 데에도 다양하게 활용될 수 있다. 이에 각국의 국가도서관이나 국가서지기관에서는 전거데이터를 링크드 데이터 형식으로 발행하고 이에 대한 개방된 접근을 허용하고 있다. 국내에서도 이미 국립중앙도서관에서 저자명 전거데이터를 rdfs, schema, dcterms, foaf 등을 활용하여 RDF로 변환하여 제공하고 있어, 개별 도서관의 입장에서 국가 전거데이터를 활용하게 되면 이러한 RDF 형



<그림 2> 전거데이터의 RDF 트리플 구문 예시

식으로의 변환이나 상호연계도 보다 손쉽게 해결할 수 있다.

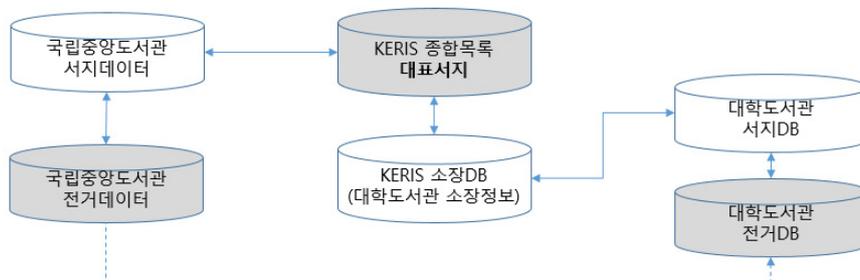
### 3. 국가선거를 활용한 소급선거 시범 구축

#### 3.1 시범사업의 개요

2022년 12월을 기준으로, 국립중앙도서관에서 구축한 선거데이터는 개인명 146만 건, 단체명 10만 건으로 저자 선거만 총 156만 건에 이르며(국립중앙도서관 국가선거 누리집), 향후 주제와 표제 선거데이터도 지속적으로 확충할 계획을 가지고 있다. 국립중앙도서관에서는 국가선거데이터의 공동 활용과 효율적인 선거데이터 구축 프로세스를 정립하기 위해 2017년부터 대학도서관들과 본격적인 논의를 시작하였다. 그 결과, 2019년부터 국립중앙도서관에서 구축한 저자 선거데이터(ISNI가 부여된 선거데이터)를 공개하고, 대학도서관에서 서지데이터 작성 시 이를 활용하거나 소급 서지에 선거데이터를 일괄 구축할 수 있도록 시범사업을 추진하였다(국립중앙도서관, 2019).

시범사업은 KERIS를 통해 진행되었는데, 각 도서관에서 KERIS 종합목록을 통해 국가선거데이터를 다운로드하거나 이를 직접 자관 서지데이터에 연결하여 사용하는 것으로 사업방향이 결정되었다. 이를 위해 KERIS 종합목록의 대표서지와 국립중앙도서관 서지를 연결하는 작업이 우선 진행되었다. 국립중앙도서관은 2019년 상반기에 선거가 연결된 서지레코드 100만 건(선거레코드 45만 건)을 KERIS에 일괄 전송하고, 추후 생성되거나 수정된 데이터를 주기적으로 KERIS에 전송하였다. 대학도서관에서의 소급 서지에 대한 선거 구축은 KERIS 종합목록의 대표서지에 연결된 각 대학도서관의 소장정보를 이용하는 것으로 결정되었다. 즉, KERIS 종합목록의 대표서지에 연결된 소장DB에서 대학도서관의 소장정보(도서관 코드와 자관 제어번호)를 추출하고, 대표서지에 연결된 국립중앙도서관의 선거데이터를 각 대학도서관의 서지데이터에 자동 연결한다는 계획이었다(〈그림 3〉 참조).

이러한 계획을 추진하기 위해 국립중앙도서관은 대학도서관 자료관리시스템에서 국가선거데이터를 조회하고 활용할 수 있는 '선거모듈'의 개발과 보급도 지원하였다. 선거모듈은



〈그림 3〉 국가선거데이터 활용을 위한 시스템 연계도

국가선거데이터 검색 기능, 선거데이터 반입 및 수정 기능, 선거와 서지 연결기능, 기 구축된 선거와 국립중앙도서관 선거 비교 기능 등을 구현할 수 있도록 설계되었다. 2019년 시범사업에서 선거도플 개발은 대학도서관에서 가장 많이 사용하고 있는 2개 시스템을 대상으로, 그리고 실질적인 선거데이터 구축 사업은 3개 대학도서관을 시범으로 하여 각각 진행되었다.

### 3.2 시범사업의 참여도서관 현황

A대학도서관은 시범사업에 참여하기 전까지 선거데이터를 구축하지 않았던 도서관으로, 2019년 시범사업을 통해 46만 건의 서지에 매칭되는 선거데이터 약 17만 건을 소급 구축하였다. 소급 구축 이후에는 신규 서지데이터 구축 시 KERIS 종합목록에서 국가선거데이터를 '건'별로 검색하여 반입하여 사용하고 있는데, 월평균 약 4,000건의 신규 서지에 대해 국가선거데이터를 이용하여 약 1,000건의 선거데이터를, 그리고 국가선거에서 검색되지 않는 약 400건의 선거데이터를 자관에서 직접 생성하고 있다. 실무자가 체감하는 바로는 신규 서지 중 약

70%가 국가선거로부터 활용 가능한 수준이다. 다만, 건별로 검색해야 하고, 동일한 접근점을 가진 선거데이터가 상당히 많고 개별 데이터의 식별이 어려워 이를 일일이 확인하는데 따른 업무부담이 매우 높은 편이다.

A대학도서관의 경우 별도의 선거형접근점 규칙을 가지고 있지 않다. 다만, 언어 표기 형식을 보면, 중국인명과 일본인명은 한자의 한글음을 적용하고 성과 이름 사이를 쉼표(,)로 구분한다. 중국인명과 일본인명을 제외한 모든 외국인명(로마자, 비로마자)의 선거형접근점은 원어로 표기한다. 그러나 실제 서지레코드에는 이러한 언어 형식이 일관성 있게 적용되어 있지 않고, 원어의 한글음이나 한자, 비로마자에 대한 로마자 이름 등이 혼재되어 있는 상태이다. A대학도서관에서 공식적으로 규정하는 접근점 형식을 국가선거데이터의 '개인명' 기술 지침(국립중앙도서관 국가서지과, 2018)과 비교하면, <표 1>과 같은 차이가 있다. 따라서 A대학도서관에서는 국가선거데이터를 반입하되, 선거형접근점을 자관 규정에 맞춰 수작업으로 수정하는 작업을 하고 있다.

A대학도서관에서 2019년 시범사업을 통해 반입한 선거데이터 통계는 <표 2>와 같다. 단순

<표 1> 선거형접근점 형식 비교

	A대학도서관	국가선거데이터
일본인명	한자의 한글음(성과 이름 ,로 구분) (예) 천단, 강성	일본어의 한글 표기 형식 (예) 가와바타 야스나리
중국인명	한자의 한글음(성과 이름 ,로 구분) (예) 모, 택동	중국어의 한글 표기 형식 (예) 마오쩌둥
로마자 서양인명	원어 (예) Poe, Edgar Allan	로마자 형식 (예) Poe, Edgar Allan
비로마자 외국인명	원어 (예) Беляков, Илья (러시아인) Nguyễn, Ngọc Từ (베트남인)	영어 정보원에 있는 로마자 형식 (예) Belyakov, Ilya Nguyễn, Ngọc Từ

〈표 2〉 시범사업을 통한 A대학도서관의 전거 구축 현황

시기	총 서지레코드	전거가 연결된 서지레코드	전거레코드 구축 건 수
2019 시범사업	1,650,000건(추정치)*	460,770건 (27.9%)	169,999건
2022.12.30. 기준	1,740,037건	354,409건 (20.4%)	183,016건

\* 시범사업 직후의 총 서지레코드 수는 현시점에서 산출이 불가능하여 2022년 기준치(174만건)에서 2년치를 제외한 추정치임.

통계로 보면, 시범사업 직후 전체 서지레코드 중 27.9%에 전거데이터가 1건 이상 구축된 것으로 나타난다(서지레코드 1건당 전거데이터 0.36건 비율). 그러나 자동생성된 전거데이터 중에서 실제 A대학도서관의 서지레코드에 적용되지 않은 사례가 다수 있었다. 이러한 이유로 시범사업 이후에 전거데이터를 꾸준히 구축해 왔음에도 불구하고 2년이 지난 시점에 전거가 연결된 서지레코드는 오히려 10만 건 이상이 줄어들어, 2022년 12월말 기준으로 전체 레코드 중 20.4%를 차지하고 있다. 구축된 전거데이터는 향후 다양한 연구자 정보(ISNI, ORCID, ResearchID 등)와 연계하여 폭넓게 활용할 예정이다. 아직은 적극적으로 검색서비스에 활용하고 있지 않다.

### 3.3 전거 구축 결과

소급 서지에 대한 전거 일괄구축의 성과를 분석하기 위해, 이 연구에서는 A대학도서관에서 시범사업을 통해 구축한 전거데이터를 사례로 분석하였다. 시범사업 직후 생성된 데이터에 대한 기초정보는 A대학도서관으로부터 수집하였으며, 이를 토대로 서지와 전거의 연결이 적절하게 이루어졌는지 확인하는 절차를 거쳤다. 확인은 A대학도서관으로부터 수집한 데이터에 있는 서지 제어번호로 검색한 후 각 서

지에 연결된 전거데이터를 확인하는 방식으로 진행하였다. 이 과정에서 원천 데이터와의 비교가 필요하다고 판단되면 해당 서지에 연결된 국립중앙도서관의 서지 및 전거 제어번호를 이용하여 확인하는 절차를 거쳤다.

분석결과는 당초 전거데이터의 정확률이나 누락률 등으로 산출할 계획이었으나 앞서 언급하였듯이 시범사업을 통해 구축한 전거데이터가 실제로 A대학도서관의 서지에 연결되지 않은 사례가 상당수 있었고, 또 지난 2년 동안 오류 데이터에 대한 삭제나 수정이 이미 어느 정도 진행된 상태여서 정량적 통계는 크게 의미가 없다고 판단하였다. 또한, 사전점검 결과에서 사례에 따라 오류 발생빈도의 편차가 크고 그 원인도 다양한 것으로 나타나 이번 연구에서는 전거 구축 과정에서 발생한 오류나 누락의 '사례'와 '원인'을 파악하는데 중점을 두었다. 개인 저자를 대상으로 임의로 sampling하여 분석한 결과를 (1) 전거데이터의 생성률과 정확률, (2) 전거데이터의 오류 사례, (3) 전거데이터의 누락 사례로 나누어 보면 다음과 같다.

#### 3.3.1 전거데이터의 생성률과 정확률

일괄 소급 구축된 전거데이터를 토대로 ① 관련 서지레코드 중에서 실제 전거데이터가 생성된 비율과 ② 자동생성된 전거데이터 중 오류가 없는 정확한 전거데이터의 비율로 구분하

여 살펴보았다. 이를 위해 먼저 특정 개인과 관련된 서지레코드를 A대학도서관의 저자검색을 통해 파악하는 과정을 거쳤다. 관련 서지레코드는 학위논문과 전자자료, 시청각자료 등을 제외하고 ‘단행본’으로 제한하였다. 가령, <표 3>에서 A대학도서관에서 ‘김대성’과 관련된 서지레코드를 모두 추출한 결과 총 75건으로 파악되었으며, 시범사업에서 ‘김대성’에 대한 전거데이터는 모두 14건의 서지레코드에 생성되어 ‘김대성’에 대한 전거데이터의 생성 비율은 18.7%(14/75)로 볼 수 있다. 이어서, 전거데이터와 연결된 서지레코드를 직접 확인하여 해당 저자와 관련된 자료가 아니거나(오류) 전거데이터가 실제로 반영되지 않은 경우(누락)를 제외한 정확률을 산출하였다.

사례는 앞서 <표 1>에서 살펴본 전거형접근점의 언어 형식에 따라 한국인명, 일본인명, 중국인명, 기타 외국인명(베트남인, 미국인명)으로 구분하여 살펴보되, 한국인명의 사례는 동명이인이나 저자 역할에 따른 차이를 살펴보기 위해 의도적으로 다수의 동명이인이 존재하는 사례(국가선거의 전거형접근점을 기준으로 ‘김

대성’은 125명, ‘김갑수’는 44명)와 유일한 접근점을 가진 사례(‘노승영’)로 나누어 조사하였다. 추가적으로 확인이 필요하다고 판단되면, <표 3>의 사례 외에 대상을 확대하면서 재확인하는 과정을 거쳤다.

먼저, A대학도서관의 서지레코드 중에서 전거데이터가 자동생성되지 않은 사례가 적지 않은 것으로 나타났다. 전거데이터가 자동생성되었으나 이 가운데 명백한 오류로 확인된 사례를 제외하면, 정확하게 전거데이터가 생성된 비율은 전체적으로 더 떨어졌다(<표 3> 참조). 전거데이터의 자동생성률은 사례에 따라 편차가 컸는데, 일본인명과 중국인명의 전거데이터 생성이 특히 낮은 것으로 나타났다. 이는 <표 3>의 사례 이외의 일본인명과 중국인명에서 대부분 유사한 것으로 확인되었다. 전거 자동생성률과 정확률이 비교적 높은 편에 해당되었던 ‘노승영’의 사례는 동명이인이 존재하지 않고, 단독 번역자로 참여한 저작만 존재하여 서지레코드에 접근점이 비교적 균일하게 적용되었기 때문으로 파악되었다. 이외에, 다수의 동명이인이 존재하는 사례에서는 해당 인명에 대

<표 3> 전거데이터의 생성률과 정확률

접근점 사례	서지레코드 (A)	전거 자동생성 서지레코드(B)	전거자동생성률 (B/A)	오류/누락 (C)	전거데이터 정확률 (B-C/A)
김대성	75	14	18.7%	3	14.7%
김갑수	113	43	38.1%	9	30.1%
노승영	91	58	63.7%	2	61.5%
천단, 강성 (가와바타 야스나리)	74	14	18.9%	10	5.4%
여, 화 (위화)	128	21	16.4%	20	0.8%
Nhật Hạnh, Thích (Thích Nhật Hạnh)	69	28	40.6%	17	15.9%
Aaker, David A.	29	16	55.2%	6	34.5%

한 전거데이터가 비교적 정확하게 반영되어 있었다.

A대학도서관에서 전거가 자동생성되지 않은 원인은 해외 자료나 보고서 등과 같이 국립중앙도서관 미소장자료이거나 국립중앙도서관 서지레코드에 전거데이터가 존재하지 않기 때문인 ‘당연한’ 결과도 있었지만, 그보다는 이유를 파악하기 어려운 경우가 훨씬 더 많았다. 접근점의 언어 형식에 따른 차이에서 비롯된 것일 수도 있다고 생각했으나, 다음 <표 4>와 같이 A대학도서관 서지레코드의 접근점에 기술된 동일인의 상이한 이름 형식 중에서 전거데이터가 자동생성된 사례와 미생성된 사례 간에 뚜렷한 차이는 발견할 수 없었다. 뿐만 아니라 국립중앙도서관 서지레코드 및 국가전거데이터의 전거형, 이형과 비교하여도 전거데이터가 생성되지 않은 원인은 찾기 어려웠다.

### 3.3.2 전거데이터의 오류 사례

A대학도서관에 자동생성된 전거데이터 중 오류 사례는 3가지 유형으로 구분되었다. 먼저,

A대학도서관의 서지레코드에 국가전거데이터가 연결되었으나, 해당 전거와 관련이 없는 자료이거나 해당 자료가 존재하지 않는 경우가 있었다. <표 5>는 전거형접근점 “Aaker, David A., 1938- ”가 자동생성된 서지레코드로서 사례 ①과 같이 전혀 관련이 없는 자료이거나 사례 ②와 같이 실제 존재하지 않는 서지레코드에 전거데이터가 자동생성된 경우이다. 이에 대한 원인은 KERIS 종합목록과 A대학도서관 서지레코드 간의 연결 오류에서 비롯된 것으로 유추할 수 있다. 즉, KERIS 종합목록에 연결된 A대학도서관의 소장정보 오류(제어번호 오류)로 인해 <표 5>와 같이 관련이 없거나 실제로 없는 서지에 전거데이터가 자동생성된 것이다.

두 번째는 동일한 서지레코드에 전거가 중복 구축된 사례가 있었다. KERIS 종합목록의 대표서지 1건에 대해 A대학도서관의 상이한 두 개의 서지 제어번호가 존재하여 이에 대한 전거데이터가 각각 생성되었지만, 결과적으로 상이한 두 개의 서지 제어번호는 제어번호 오류

<표 4> A대학도서관의 전거 자동생성 및 미생성 접근점 비교

전거 생성 접근점	전거 미생성 접근점	비교: 국가전거데이터	
700 1 ▼a천단강성 700 1 ▼a가와바타 야스나리 700 1 ▼a川端, 康成	100 ▼a천단, 강성 700 1 ▼a가와바타, 야스나리	전거형	가와바타 야스나리
		이형	川端康成 천단강성 등
100 1 ▼aThich Nhat Hanh 100 0 ▼aNhat Hanh, Thich 100 1 ▼aThich, Nhat Hanh	100 1 ▼aNhat Hauh, Thich 100 1 ▼aNhat Hanh, Thich 100 1 ▼aNhất-Hạnh, Thích	전거형	Thích Nhất Hạnh
		이형	틱낫한 Nhất Hạnh Thích Thich, Nhat Hanh 등
100 1 ▼aЛотман, Юрий Михайлович 100 1 ▼aLotman, Yuri M	100 1 ▼aLotman, Yuri M 700 1 ▼aLotman, Iuri 100 1 ▼aLotman, Iurii	전거형	Lotman, Youri Mikhailovich
		이형	Lotman, Iuri M. Lotman, Yuri Лотман, Юрий Михайлович 등

〈표 5〉 전거데이터 자동생성 오류 사례

구분	사례
사례 ①	245 00 ▼a事務管理·不當利得 / ▼c松坂佐一 著 250 ▼a新版 260 ▼a서울: ▼b大學書院 , ▼c1992
사례 ②	<b>검색 결과가 없습니다.</b>

로 인해 하나의 서지레코드인 것으로 확인되었다. 비록 관련 사례가 많은 것은 아니었지만 이러한 중복도 명백히 오류에 해당한다고 볼 수 있다.

마지막으로, 전거데이터의 연결에는 문제가 없지만 전거데이터가 중복 생성된 사례가 있었다. 〈표 6〉과 같이 동일한 전거형접근점이 적용된 사례는 100, 700필드에서 다양하게 발견되었다. 이 중에는 〈표 6〉의 사례 ②와 같이 전거형접근점의 상이한 형식이 하나의 서지레코드에 중복 생성된 경우도 있었다. 이러한 오류는 전거데이터 자동생성 시 발생한 기술적인 문제로 짐작된다.

### 3.3.3 전거데이터의 누락 사례

KERIS 종합목록을 통해 A대학도서관의 서

지에 전거데이터가 자동 생성되었지만, 최종적으로 반영되지 않은 경우는 모두 '누락'으로 간주하였다. 이러한 누락 중에는 A대학도서관 데이터의 오류나 국립중앙도서관과 KERIS 종합목록, A대학도서관 간 데이터 기술 방식의 차이에서 비롯된 것도 있었다. 가령, 〈표 7〉의 사례 ①은 A대학도서관의 서지레코드에 저자 'Aaker, David A.'에 대한 접근점 데이터가 아예 존재하지 않는 오류에 해당한다(번역자에 대한 접근점만 있음). 따라서 전거데이터가 자동생성되었지만, 적용될 접근점 데이터의 부재로 인해 최종 반영되지 않은 것이다. 사례 ②와 ③은 4인 이상 저자 중 공저자나 기타 다양한 역할을 가진 저자에 대한 접근점이 생략된 사례이다. 오류에 해당하지는 않지만, 국립중앙도서관 서지와 비교해 보면 대표저자 외에 다양

〈표 6〉 전거데이터 중복 생성 오류 사례

구분	사례
사례 ①	245 00 ▼a이슬람 사회의 여성/▼d김대성...[등]저 700 1 ▼a김대성, ▼g金大成, ▼d1955- ▼0[자관전거번호] 700 1 ▼a김대성, ▼g金大成, ▼d1955- ▼0[자관전거번호]
사례 ②	100 1 ▼a박범신, ▼g朴範信, ▼d1946- ▼0[자관전거번호] 100 1 ▼a박범신 ▼0[자관전거번호] 245 10 ▼a(朴範信 長篇小說) 풀잎처럼 눕다 / ▼d朴範信 著 260 ▼a서울: ▼b고려원, ▼c1986
사례 ③	100 0 ▼aBstan-'dzin-rgya-mtsho, ▼cDalai Lama XIV, ▼d1935- ▼0[자관전거번호] 100 1 ▼aBstan-'dzin-rgya-mtsho, ▼cDalai Lama XIV, ▼d1935- ▼0[자관전거번호] 245 20 ▼a(달라이 라마의) 행복 찾기 / ▼d달라이 라마 지음: ▼e제프리 홉킨스 엮음: ▼e우충환 옮김

〈표 7〉 전거데이터 자동생성 누락의 원인(1)

구분	데이터 내용	
사례 ① 접근점 누락	245 00 ▼a [중략] /▼d데이비드 아커 [외]저; ▼e김고명 옮김 700 1 ▼a김고명▼0[자관전거번호]	
사례 ② 4인 이상 저자의 공저자 생략	A대학	245 ▼a [중략] /▼d존 켈치 ...[등] 지음 700 ▼aQuelch, John A.▼0[자관전거번호]
	국립중앙 도서관 (국가전거)	245 00 ▼a [중략] /▼d존 켈치 [외] 지음; ▼e현대경제연구원 옮김 700 1 ▼aQuelch, John A.,▼d1951- 700 1 ▼aJoachimsthaler, Erich,▼d1956- 700 1 ▼aAaker, David A.,▼d1938- 700 [이하 중략]
사례 ③ 특정 역할의 저자에 대한 접근점 미생성	A대학	100 1 ▼a왕, 건문 245 20 ▼a [중략] /▼d왕건문 지음; ▼e이재훈; ▼e은미영 [공] 옮김 700 1 ▼a이재훈▼0[자관전거번호] 700 1 ▼a은미영▼0[자관전거번호]
	국립중앙도서관 (국가전거)	245 20 ▼a /▼d왕건문 지음; ▼e이재훈, ▼e은미영 옮김 500 ▼a감수: 김갑수 700 1 ▼a왕젠원 700 1 ▼a이재훈 700 1 ▼a은미영 700 1 ▼a김갑수, ▼g金甲秀, ▼d1961-

한 저자에 대한 접근점이 생성되어 있는 것을 확인할 수 있다. 이러한 경우 해당 서지레코드에 전거데이터가 자동생성되었다더라도 A대학 도서관에는 최종 반영되지 않았다.

이러한 사례는 역의 관계에서도 성립한다. 가령, 국립중앙도서관 서지데이터에 존재하지 않는 접근점이 A대학도서관에 존재한다면, 이 역시도 동일한 이유로 전거데이터가 자동생성되지 않았다(〈표 8〉 참조).

뿐만 아니라 권차에 따른 기술방식의 차이도 전거데이터의 생성에 많은 영향을 미치는 것으로 나타났다. 이러한 차이는 A대학도서관과 KERIS 종합목록 간에, 그리고 KERIS 종합목록과 국립중앙도서관의 서지레코드 간 데이터 기술방식의 차이에서 비롯된다. 〈표 9〉의 사례에서 보듯이, 총 17권으로 구성된 자료에

대해 A대학도서관은 종합기술방식으로 1개의 서지레코드를, KERIS 종합목록은 각 권별 분할기술방식과 종합기술방식 둘 다에 해당하는 복수의 서지레코드를, 국립중앙도서관은 본문에 해당하는 15권과 부록 2권을 구분하여 2개의 서지레코드를 각각 생성하고 있다. 이러한 상이한 기술방식에서 만일 각 권의 저자가 모두 다르다면 정확한 전거데이터의 생성은 더욱 어렵게 된다. 이는 향후 저작 접근점이나 표현형 접근점의 생성에서도 동일한 문제를 일으킬 수 있는데, 현재 구축된 서지레코드에서 이러한 유형의 사례가 적지 않게 발견되었다.

이상 살펴본 오류와 누락의 유형별 발생빈도와 비율을 산출하면 다음 〈표 10〉과 같다.

〈표 8〉 전거데이터 자동생성 누락의 원인(2)

구분	데이터 내용	
A대학도서관	100 1	▼a이, 시진, ▼d1518-1593
	245 10	▼a본초강목, ▼n2, ▼p본초강목 제3권 / ▼d이시진 지음; ▼e민족의학연구원 옮김
	700 1	▼a김갑수
	700 1	▼a박상리
	700 1	▼a안은수
	700 1	▼a은석민
	700 1	▼a김정선
	710	▼a민족의학연구원
국립중앙도서관 서지레코드	245 00	▼a본초강목, ▼n1-2 / ▼d이시진 지음; ▼e민족의학연구원 옮김
	505 10	▼n1, ▼t본초강목 제1권 서례상(序例上), 본초강목 제2권 서례하(序例下) -- ▼n2, ▼t본초강목 제3권 백병주치약상(百病主治藥上)
	700 1	▼a이시진, ▼g李時珍, ▼d1518-1593
	710	▼a민족의학연구원

〈표 9〉 권차 기술방법에 따른 차이

A대학도서관	KERIS 종합목록 대표서지	국립중앙도서관 서지
(新註解) 本草綱目 第1卷-第17卷	(新註解) 本草綱目. 第1卷, 序例	(新註解) 本草綱目. 第1卷-第15卷
	(新註解) 本草綱目. 第2卷, 百病主治上·下	
	...	本草綱目 附圖. 上卷, 下卷

〈표 10〉 자동생성 전거데이터의 오류/누락 사례

사례	전거데이터 자동 생성(B)	오류			누락	오류/누락 계(C)	오류/누락 비율(C/B)
		연결 오류	서지 중복	데이터 중복			
김대성	14	1	1	1	0	3	21.4%
김갑수	43	7	0	0	2	9	20.9%
노승영	58	0	2	0	0	2	3.4%
천단, 강성 (가와바타 야스나리)	14	2	0	0	8	10	71.4%
여, 화(위화)	21	2	2	0	16	20	95.2%
Nhật Hạnh, Thích	28	2	0	3	12	17	60.7%
Aaker, David A.	16	1	3	0	2	6	37.5%

#### 4. 시범사업의 방법론적 한계와 정책적 제언

국내 도서관의 상황에서 전거 소급 구축을

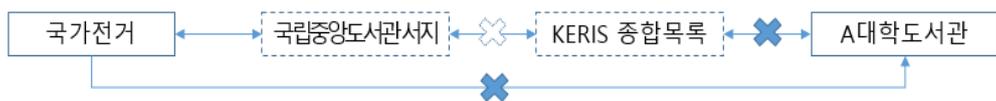
위한 가장 효율적인 방법은 현재 구축된 ‘국가 전거’를 적절히 활용하는 것이란데 의심의 여지가 없다. 국내 최대의 전거데이터가 축적되어 있고, 이미 공동 활용을 전제로 한 전거데

이터의 구축과 공개가 진행되고 있어 단위 도서관에서 중복적으로 구축하는 것보다 내용적으로나 비용적으로 훨씬 효율적이기 때문이다. 문제는 이 데이터를 단위 도서관에서 어떻게 활용할 것인가에 있다. 활용은 다시 두 가지로 구분하여, 지금까지 구축되지 못했던 서지레코드에 '소급' 적용하는 방법과 신규 서지레코드에 '상시적으로' 활용하는 방법으로 나누어 동시에 접근하는 것이 바람직할 것이다. 어떤 방식의 접근이든 국내 도서관이 처한 여건을 고려할 때 인력에 전적으로 의존해야 하는 수작업은 최소화하고 가능한 기계적인 방법으로 고효율을 추구하는 방법이 마련되어야 할 것이다.

이런 측면에서 볼 때, top-down 방식으로 단위 도서관의 선거데이터를 일괄구축하고자 했던 시범사업의 접근은 분명 타당하다고 할 수 있다. 나아가, 종합목록을 매개로 선거데이터를 구축하고자 했던 계획 또한 '과급력'이나 '효율성' 측면에서 상대적으로 더 큰 효과를 기대할 수 있는 방법임이 틀림없다. 문제는 A대학교도서관에서 저자 선거의 구축율이 20% 남짓에 불과할 정도로 매우 낮게 나타났고, 그마저도 접근점 형식에 따른 편차가 컸다는 데 있다. 일반적으로 하나의 서지레코드에 공저자나 역할이 다른 저자가 다수 존재한다고 보면 실제로 선거데이터 구축율은 이보다 훨씬 떨어진다고 볼 수 있다. 이러한 낮은 선거 구축율은 A대학교도서관에서 수작업으로 구축한 선거데이터

와 비교하여도 분명히 드러난다. A대학교도서관에서 시범사업 이후에 건별로 선거데이터를 직접 구축한 결과 국가선거데이터를 활용할 수 있는 비율이 70% 수준으로 가늠되지만, 시범사업을 통해 자동생성한 선거데이터의 비율은 이에 훨씬 못 미치기 때문이다. 이러한 사실은 시범사업의 추진 방법에 한계가 있다는 의미로 해석할 수 있다.

시범사업은 국립중앙도서관의 서지와 KERIS 종합목록의 대표서지를 연결하고, 다시 KERIS 대표서지에 연결된 대학도서관의 소장정보를 이용하여 국립중앙도서관 서지에 연결된 국가선거를 개별 대학도서관의 서지레코드에 적용하는 방식으로 진행되었다. 이러한 방식의 적용에 있어 관건은 각 서지 간에 정확한 '연결'(매핑)이라 할 수 있다. 어느 쪽에서건 연결이 제대로 이루어지지 않은 경우 선거데이터의 자동생성은 원천적으로 불가능해지기 때문이다(<그림 4> 참고). 이번 연구에서 살펴본 일부 사례로 일반화하기에 무리가 있지만, KERIS 종합목록과 A대학교도서관 소장정보 간의 연결 오류는 일부 사례만으로도 충분히 확인할 수 있었다. 뿐만 아니라 종합목록에 자관 서지가 중복 생성되어 있거나 반대로 누락된 경우가 종종 발견된다는 사실을 실무자 면담을 통해서도 확인할 수 있었다. 시범사업에서는 종합목록에 존재하는 이러한 문제점이 전혀 고려되지 않았다는 점에서 방법론적 한계를 가진다.



<그림 4> 선거 자동생성 요건

한편, 이번 연구에서 국립중앙도서관 - KERIS 종합목록 - 단위 도서관의 서지레코드를 연결하여 국가선거데이터를 제대로 추출하였더라도, 최종적으로 단위 도서관에 선거데이터가 생성되지 않은 경우도 적지 않게 있음을 확인할 수 있었다. 주된 이유는 시범사업에 참여한 도서관의 서지레코드에 선거를 생성할 접근점 데이터가 애초에 존재하지 않기 때문이기도 했고, 정확한 이유를 파악하기 어려웠지만 접근점의 상이한 형식이나 일관되게 적용되어 있지 않은 데이터에 원인으로 있을 가능성도 배제할 수 없었다. 이러한 사실은 시범사업의 성과를 좌우하는 또 다른 중요한 요인이 단위 도서관의 서지레코드 품질에 있음을 의미하는 것이라 할 수 있다. 즉, 서지레코드(특히, 접근점)에 문제가 있는 경우 이를 토대로 하는 선거데이터의 구축과 연결에도 문제가 생기기 때문이다. 이러한 점에서 시범사업은 개별 도서관에서 구축한 서지레코드의 품질에 따라 편차가 크게 나타날 수 있다는 방법론적 한계를 가진다.

이외에도, 시범사업의 추진과정에서 비롯되는 문제도 있다. 시범사업이 참여도서관을 대상으로 ‘단기간’에 ‘일회성’으로 진행되다보니 사업의 성과가 제대로 분석되지 못했고, 후속 사업을 통해 더 많은 도서관으로 확산되지도 못했다. 이번 연구의 대상이었던 A대학도서관에서도 시범사업이 끝난 후 업무를 수행하는 과정에서 오류를 발견하거나 미구축의 정도를

체감하는 정도였을 뿐 시범사업 이후 성과를 제대로 파악하지 못하고 있었다. 이는 시범사업에 적용된 알고리즘에 어떠한 문제가 있었는지 혹은 선거구축 도서관의 어떤 문제로 알고리즘이 당초 계획에 미치지 못한 것인지가 전혀 분석되지 않았고, 심지어 시범사업 이후에 필요한 후속 조치에 대한 파악도 제대로 되지 않았음을 의미하는 것이라 할 수 있다. 이러한 문제는 사업주체와 시행 도서관 간에 커뮤니케이션이 적절히 이루어지지 않았다는 데서도 원인을 찾을 수 있다.

이론적으로나 개념적으로 아무리 우수한 방법이라 할지라도 실제 적용과정에서 다양한 방법론적 한계가 있을 수 있다. 따라서 추진과정에서 드러나는 문제와 원인을 즉각적으로 파악하고 그에 따른 적절한 대응책을 마련하여 반복적으로 적용해 보는 과정은 필수적이다. 특히, 선거데이터 소급 구축에서는 데이터의 오류나 도서관별 데이터 기술방식의 차이 등에 대한 세심한 분석이 선행되어야 한다. 이번 연구에서도 나타났듯이 A대학도서관에서 일본인명과 중국인명 등 외국인명의 선거데이터 생성에서 발생된 문제를 바로 감지하였다면 그 원인을 분석하여 다른 보완적인 방법을 시도할 수도 있었을 것이다.<sup>1)</sup> 이러한 조치는 개별 도서관에서 선거데이터 일괄구축 후 처리해야 할 후속 업무, 특히 수작업 처리의 범위를 최소화한다는 측면에서도 필수적이고 중요한 과정이

1) 선거 관련 규정이 존재하지 않고 선거레코드를 관리하지 않았던 국내의 많은 도서관들은 일반적으로 접근점의 상이한 형식을 다른 방식으로 서지레코드에 표현해 왔다. 즉, 접근점에 기술된 이름이나 이름 형식과 다른 접근점 서지레코드의 로컬필드(9XX)에 임의로 기술하고 이를 검색에 활용하고 있다. 그러나 시범사업에서는 1XX, 7XX에 기술된 접근점만을 대상으로 선거데이터 생성을 시도하였다. 시범사업에서 외국인명의 처리를 비롯하여 권차 기술방식에 따른 차이, 부존재 접근점에 대한 처리 등에 대해 별도의 해결책을 마련하여 적용하였다면 선거데이터 자동생성 비율은 더욱 증가하였을 것이다.

라 할 수 있다.

시범사업에서 드러난 이러한 한계를 최소화 하면서 사업의 효과를 안정적으로 정착시키기 위해서는 추진방식이나 내용의 개선책을 마련할 필요가 있다.

먼저, 추진방식에 있어 획일화된 방식을 모든 도서관에 적용하기보다는 개별 도서관의 특성에 부합하는 소급선거 구축 방향과 방법을 결정하여 적용하는 방안을 고려할 필요가 있다. 국내 도서관의 서지데이터는 이미 높은 수준의 표준화가 이루어져 있으므로 도서관들 간에 엄청난 차이를 보이지 않을 것으로 예측된다. 다만, 접근점에 해당하는 데이터는 선거 관련 표준의 부재가 오랜 기간 이어지다보니 데이터의 생성에서부터 기술형식에 이르기까지 적지 않은 차이를 가지고 있다(이미화, 2012). 뿐만 아니라 종합목록과 연계한 선거데이터의 구축 방식을 적용하더라도, 종합목록이나 개별 도서관의 서지레코드 품질에 따라 접근 방식이 다소 상이해야 할 수도 있다. 가령, KERIS 종합목록을 이용한 동일한 방식을 KOLIS-NET을 매개로 하여 공공도서관의 선거데이터 구축에 활용할 경우 결과는 완전 다를 수 있다. 따라서 선거데이터를 구축하고자 하는 기관의 데이터 특성을 파악하고 이를 공통되는 특성에 따라 유형화하여 그에 따른 선거 알고리즘을 차별화하여 적용한다면 선거데이터의 자동 구축율은 현재보다 훨씬 높아질 것으로 판단된다. 이를 위해 사업추진에 앞서, 사업 주체와 선거구축 도서관 간에 충분한 협의를 의무화하는 방안도 고려할 필요가 있다.

다음으로, 선거데이터 일괄구축이 완료되더라도 유지관리와 상시적으로 선거데이터를 구

축할 수 있는 체제를 마련할 필요가 있다. 시범사업에 참여한 A대학도서관은 시범사업이 끝난 이후 건별로 일일이 국가전거를 검색하여 선거데이터를 구축하고 있다. A대학도서관이 국내에서 규모가 상당히 큰 도서관에 해당됨에도 불구하고 인력구조나 업무량을 고려할 때 이러한 방식으로 업무를 지속하는 것은 분명 어려울 수 있다. 국외 대학도서관(University of North Texas)에서 선거업무를 전담하는 사서 7명으로도 부족하여 자동처리 방식으로 업무를 개선한 사례(Wolf, 2020)를 참조하면 이는 충분히 짐작할 수 있다. 특히, A대학도서관처럼 일괄소급에 의한 선거데이터 구축율이 낮게 나타난 상황에서는 상시적으로 선거데이터를 자동구축할 수 있는 체제를 마련하는 것이 일회성 사업성과 못지않게 중요하다고 할 수 있다.

이와 관련하여 방법적인 측면에서는 이미 국외 도서관들이 시도하고 있는 방법을 참조할 수 있다. 가령, 선거데이터가 미생성되었거나 신규 구축된 서지레코드를 대상으로, 월별, 분기별 또는 연간 등 주기적으로 선거데이터를 자동생성하여 자관에 반입하는 절차를 시행할 수 있다. 이 과정은 선거데이터 소급구축이 어느 정도 완료된 이후에 진행될 것이므로 1차적으로는 자관에 구축된 선거데이터를 대상으로 자동생성하고, 이를 통해 선거데이터가 생성되지 않은 경우 다시 국가선거데이터를 활용하는 방식으로 순차적으로 적용하는 것이 단위 도서관의 업무부담을 줄이는데 보다 효과적일 것이다. 국외 도서관의 경우 이러한 절차를 주로 벤더를 통해 진행하고 있지만, 시범사업을 활용하여 국가 주도의 선거 구축 방법을 잘 정착시

킨다면 국내 도서관에서는 선거데이터의 유지 관리에도 동일한 방식을 충분히 활용할 수 있을 것으로 보인다.

구축된 선거데이터의 유지관리와 관련하여서는 국가선거데이터를 생산하는 국립중앙도서관의 추가적인 역할도 필요하다. 선거데이터는 완벽한 상태의 고정적인 것이 아니라 지속적으로 수정, 변경, 삭제가 이루어지므로 이를 기구축된 선거데이터에 반영하는 절차도 반드시 필요하다. 이러한 작업의 수월성을 위해 LC나 OCLC 종합목록에서는 선거레코드에서 새로운 상호참조가 추가되거나 필드 삭제, 개인의 사망일자 추가 등과 같은 선거데이터의 수정이 이루어진 경우 이에 대한 정보를 별도로 모아서 주기적으로 제공하고 있고(〈그림 5〉 참조), 각 도서관에서는 수정된 선거데이터를 반영하기 위해 절차를 별도로 진행하고 있다. 시범사업과 같이 국가선거데이터를 활용하여 자관 선거데이터를 구축한 경우 이 부분은 손쉽게 해결할 수 있을 것이다. 선거데이터 생성 과정에서 반입된 국가선거데이터 제어번호를 활

용한다면 효율적으로 변경할 수 있기 때문이다. 따라서 국립중앙도서관에서는 수정된 국가선거데이터에 대한 정보를 별도로 제공하고, 개별 도서관에서는 이를 기구축된 선거데이터에 반영하는 시스템도 마련할 필요가 있다. 더불어, 개별 도서관에서 선호하는 접근점의 형식이 상이하다는 점을 고려하여 국가선거데이터에 접근점의 형식을 식별하는 데이터를 보완하는 방안(오삼균 외, 2018)도 적극 고려할 필요가 있다.

## 5. 결론

최근의 목록 환경에서 선거데이터의 기능이 특히 중요해지고 있다. 링크드 데이터 환경에서 선거데이터는 핵심 데이터로 인지되며, 새로운 서지모델로 확장하기 위해서도 선거데이터가 핵심적인 기능을 담당할 것으로 예측되고 있다. 그러나 국내의 많은 도서관에서는 아직까지 선거데이터를 구축하지도, 제대로 활용하

### Authority records for 20221225

Authority Records updates posted after January 19, 2008 require Arial Unicode font to display software packages or can be purchased separately. ([Learn more about Arial Unicode fonts](#))

n 92080422

Old: 100 1 Anderson, Thornton, Id 1915-

New: 100 1 Anderson, Thornton, Id 1915-1997

n 89280913

Old: 100 0 Banzragch, N. Iq (Namsaraïn), Id 1925-

New: 100 0 Banzragch, N. Iq (Namsaraïn), Id 1925-2003

no2003061571

Old: 100 1 Bara, Margit, Id 1927-

New: 100 1 Bara, Margit, Id 1928-2016

n 88193750

Old: 100 1 Barā, Mahima, Id 1926-

New: 100 1 Barā, Mahima, Id 1924-2016

100 Affleck family [Not Subd Geog] [sp2021011957] (C)  
072 H 1120  
400 UF Auchinleck family

150 African American linguists [May Subd Geog] [sp2021009055] (A)  
450 UF Linguists, African American  
550 BT Linguists--United States

150 African American political candidates [May Subd Geog] (A)  
[sp2021008149]  
450 UF Political candidates, African American  
550 BT Political candidates--United States

100 'Alawī dynasty, 1666- [Not Subd Geog] [sp2021008617] (C)  
072 H 1120  
400 UF Alauite dynasty, 1666-  
400 UF 'Alawid dynasty, 1666-

(OCLC의 저자전거 업데이트 공개 화면)

(LC 주체선거 업데이트 정보 공개 화면)

〈그림 5〉 국가선거데이터 업데이트 정보 공개를 위한 참조사례

지도 못하고 있다. 그나마 2019년부터 국가선거데이터가 공개되고 공동 활용을 위한 논의가 본격적으로 전개되면서 국가선거데이터를 활용하여 자관 선거데이터를 구축할 수 있는 환경이 점차 만들어져 가고 있는 상황이다. 그러나 국가선거데이터의 구축과 공개와 별도로, 막대한 양의 선거데이터 소급 구축은 여전히 단위 도서관에서 개별적으로 해결해야 할 과제로 여겨지고 있다. 이에 이 연구에서는 단위 도서관의 관점에서 국가선거데이터를 활용하여 자관 선거데이터를 효율적으로 구축할 수 있는 방법론을 탐색해 보고자 하였다.

이를 위해 이 연구에서는 종합목록을 이용하여 선거데이터의 소급 구축을 시도한 시범사업의 주요 내용과 방법을 살펴보고, 시범사업의 참여한 3개 도서관 중 하나인 A대학도서관을 사례로 하여 사업의 성과를 분석하였다. A대학도서관의 데이터를 분석한 결과, 시범사업을 통해 단기간에 구축된 선거데이터는 양적으로 결코 적지 않다고 볼 수 있지만, 전체 서지레코드와 비교하여 구축된 선거데이터는 20% 남짓에 불과한 것으로 나타났다. 뿐만 아니라 샘플 데이터를 기초로 하여 선거데이터의 생성률과 정확률, 그리고 선거데이터의 오류나 누락의 사례를 분석한 결과, 생성되어야 할 전체 데이터의 양과 비교하여 실제 자동구축된 선거데이터는 매우 낮은 편이었고, 그마저도 오류 등으로 인해 실제 반영되지 않은 경우가 많았다는 사실을 확인할 수 있었다. 이에 대한 원인으로 종합목록과 단위 도서관의 서지레코드 품질이 이러한 선거데이터의 자동구축에 큰 영향을 미치고 있음을 확인하였고, 시범사업의 추진과정에서 비롯된 방법론적 한계도 파악할 수 있었다.

시범사업은 말 그대로 '시범'에 불과하므로 이를 확대하고 안정적으로 정착시키기 위해서는 이를 개선하여 반복적으로 추진하는 것이 중요하다. 이에 더해, 선거데이터 구축과 관련된 다양한 주체들(국립중앙도서관, KERIS, 개별 도서관 등)의 역할을 명확하게 설정하고 공동으로 노력하는 것도 중요하다고 볼 수 있다. 국내 도서관의 여건은 국외 사례와 같이 선거처리를 위해 벤더에 엄청난 예산을 투입할 수 있는 것도, 내부에 전문인력을 충분히 둘 수 있는 상황도 아니다. 따라서 선거데이터 소급 구축은 시범사업과 같이 국가 차원에서 진행하되, 도서관 네트워크별 또는 개별 도서관과의 긴밀한 소통 하에 공동으로 추진하는 것이 바람직하다고 볼 수 있다.

후속연구에서는 최근 국가선거 또는 공유 선거과일을 이용하되 python, R, MarcEdit, 상용 스프레드시트 등 다양한 툴을 활용하여 새로운 방법으로 선거데이터 구축을 시도하고 있는 외국 도서관의 사례를 세밀하게 분석하고, 이를 참조하여 국내 도서관에 적용할 수 있는 실행방법론을 제안하고자 한다. 이 과정에서, 국가선거데이터를 명실상부한 '국가'적 데이터로 공동 활용하기 위해 선거데이터(형식, 내용)에 대한 단위 도서관의 요구를 분석하여 이를 수용하기 위한 방안과, 국가선거에 없는 선거데이터를 단위 도서관에서 구축하게 되는 경우 이를 통합하는 실질적인 '공유' 방안도 제안할 것이다. 개별 도서관 차원에서 선거데이터를 효율적으로 구축할 수 있는 환경과 실질적인 방법을 개발하는데 도서관 현장은 물론이고 학계에서의 학술적인 관심이 이어지기를 촉구한다.

## 참 고 문 헌

- 국립중앙도서관 (2019. 11. 7). 국가선거 공유 및 활용 확산을 위한 추진 계획(안) [내부문서].
- 국립중앙도서관 국가서지과 (2018). 국립중앙도서관 선거데이터 기술 지침: 개인명. 서울: 국립중앙도서관. 출처: <https://librarian.nl.go.kr/afile/fileDownloadById/20111631227oxSIV>
- 국립중앙도서관 국가선거 누리집. 출처: <https://librarian.nl.go.kr/LI/contents/L20103010000.do>
- 김미향 (2009). 대학도서관의 협력형 선거제어 방안 연구. 국립대학도서관보, 27, 41-70.
- 김미향, 김태수 (2010). 기관 리포지터리의 검색기능 향상을 위한 인명 접근제어 시스템 구축 연구. 정보관리학회지, 27(3), 125-146.
- 노지현, 이미화, 이은주 (2023). 목록이론의 이해와 적용. 서울: 한국도서관협회.
- 심경 (2006). 국내 공공도서관의 인명 선거제어의 현황 및 발전 방향. 한국문헌정보학회지, 40(4), 221-244.
- 오동근 (2000). 한국형 선거데이터베이스시스템의 개발에 관한 연구. 한국도서관·정보학회지, 31(4), 21-47.
- 오삼균, 한상은, 손태익, 심성훈 (2018). VIVO를 활용한 국가적 선거구축모델에 관한 연구. 정보관리학회지, 35(3), 165-187.
- 이미화 (2012). 대학도서관 선거제어 현황분석을 통한 선거제어 방안 모색. 한국도서관·정보학회지, 43(3), 5-26.
- 이성숙 (2021). 국가선거데이터 공동활용에 대한 사서들의 인식에 관한 질적 탐구. 한국비블리아학회지, 32(1), 443-467.
- Ashmann, A. (2002). The lowdown on automated vendor supplied authority control. Technical Services Quarterly, 20(3), 33-44.
- Haffner, A., Young, J., & Neubert, J. (2011). Cluster Authority Data. Available: [https://www.w3.org/2005/Incubator/ld/wiki/Cluster\\_Authority\\_data](https://www.w3.org/2005/Incubator/ld/wiki/Cluster_Authority_data)
- Kreyche, M., Lisius, P. H., & Park, A. (2010). The DeathFlip project: automating death date revisions to name headings in bibliographic records. Cataloging & Classification Quarterly, 48(8), 684-695.
- Lampron, P. & Wacker, M. (2019). Name authority work in the linked data environment. Journal of Library Metadata, 19(1-2), 137-140.
- Mak, L. (2013). Coping with the storm: automating name authority record updates and bibliographic file maintenance. OCLC System & Services, 29(4), 235-245.
- Monaco, M. (2020). Methods for in-sourcing authority control with MarcEdit, SQL, and regular

- expressions. *Journal of Library Metadata*, 20(1), 1-27.
- Schreur, P. E. & Carlson, A. J. (2020). Bridging the worlds of MARC and linked data: transition, transformation, accountability. *The Serials Librarian*, 78(1-4), 48-56.
- Wiederhold, R. & Reeve, G. (2021). Authority control today: principles, practices, and trends. *Cataloging & Classification Quarterly*, 59(2-3), 129-158.
- Williams, H. (2010). Cleaning up the catalogue. *Library & Information Update*, Jan/Feb., 46-48.
- Wolf, S. (2020). Automating authority control processes. *The Code4Lib Journal*, 47.  
<https://journal.code4lib.org/articles/15014>
- Zhu, L. (2019). The future of authority control: issues and trends in the linked data environment. *Journal of Library Metadata*, 19(3-4), 215-218.

• 국문 참고자료의 영어 표기

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

- Kim, Mi-Hyang & Kim, Tae-Soo (2010). A study on developing a name access point control system to improve the performance of information retrieval from institutional repositories. *Journal of Korean Society for Information Management*, 27(3), 125-146.
- Kim, Mi-Hyang (2009). Cooperative authority control in university libraries. *Journal of National University Libraries*, 27, 41-70.
- Lee, Mi-Hwa (2012). A Study on direction of authority control by surveying the authority control in university libraries. *Journal of Korean Library and Information Science Society*, 43(3), 5-26.
- Lee, Sung-Sook (2021). A qualitative study on librarians' recognition of the joint utilization of national authority data. *Journal of Korean Biblia Society for Library and Information Science*, 32(1), 443-467.
- National Library of Korea (2018). Guidelines for Describing Authority Data: Personal names. Seoul: National Library of Korea. Available:  
<https://librarian.nl.go.kr/afile/fileDownloadById/20111631227oxSIV>
- National Library of Korea (2019. 11. 7). Plan for sharing and joint use of National Authorities (draft).
- National Library of Korea. National Authorities. Available:  
<https://librarian.nl.go.kr/LI/contents/L20103010000.do>
- Oh, Dong-Geun (2000). A study on the development for the Korean authority database system.

- Journal of Korean Library and Information Science Society, 31(4), 21-47.
- Oh, Sam-Gyun, Han, Sang-Eun, Son, Tae-Ik, & Kim, Seong-Hun (2018). Constructing a model for national authority control utilizing VIVO. *Journal of Korean Society for Information Management*, 35(3), 165-187.
- Rho, Jee-Hyun, Lee, Mihwa, & Lee, Eun-Ju (2023). *Cataloging Theory and Practice*. Seoul: Korean Library Association.
- Shim, Kyung (2006). Personal name authority control in Korean public libraries. *Journal of the Korean Society for Library and Information Science*, 40(4), 221-244.