

국채보상운동 디지털 아카이브 기록물의 활용을 위한 위키데이터 연계 방안에 대한 연구*

A study on Wikidata linkage methods for utilization of digital
archive records of the National Debt Redemption Movement

도슬기(Seulki Do)¹, 박희진(Heejin Park)²

Email: sinhwask@gmail.com, papermod@hansung.ac.kr



¹ 제1저자 한성대학교 디지털인문정보학트랙 강사
² 교신저자 한성대학교 디지털인문정보학트랙 부교수

논문접수 2023.04.25
최초심사 2023.05.08
게재확정 2023.05.22

ORCID

Seulki Do
https://orcid.org/0000-0001-6473-9240
Heejin Park
https://orcid.org/0000-0002-7217-9869

© 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.

*본 논문은 2022년 국채보상운동 디지털
아카이브의 세계화를 위한 위키데이터
(Wikidata) 연계 사업의 연구비를 지원
받아 수행되었음.

초 록

본 연구는 세계기록유산으로 등재된 국채보상운동 디지털 아카이브 기록물의 활용성을
높이기 위한 방안으로 위키데이터와 연계하는 데이터 모델을 설계하고, 적용가능성을 검토
하였다. 국채보상운동 기록물 메타데이터, 시소러스, 시맨틱 네트워크 그래프를 분석하여
시사점을 도출하고, 기록물의 내용 분석을 통해 위키데이터와 연계를 위한 데이터 모델의
클래스인 기록물건, 행위주체, 시간, 장소, 사건을 도출하였다. 또한 클래스 간 연결을
위한 관계 속성들을 파악하여, 설계한 데이터 모델을 실제 기록물 사례에 적용해봄으로써,
속성을 중심으로 한 객체 간의 이동을 통해 풍부한 관련 정보들을 획득할 수 있는 가능성을
확인하였다. 본 연구의 결과는 소규모 로컬 아카이브에서 위키데이터를 활용한 데이터의
공유와 활용 체계 마련에 활용될 수 있을 것이다.

ABSTRACT

This study designed a data model linked to Wikidata and examined its
applicability to increase the utilization of the digital archive records of the
National Debt Redemption Movement, registered as World Memory Heritage,
and implications were derived by analyzing the existing metadata, thesaurus,
and semantic network graph. Through analysis of the original text of the
National Debt Redemption Movement records, key data model classes for
linking with Wikidata, such as record item, agent, time, place, and event,
were derived. In addition, by identifying core properties for linking between
classes and applying the designed data model to actual records, the
possibility of acquiring abundant related information was confirmed through
movement between classes centered on properties. Thus, this study's result
showed that Wikidata's strengths could be utilized to increase data usage
in local archives where the scale and management of data are relatively
small. Therefore, it can be considered for application in a small-scale
archive similar to the National Debt Redemption Movement digital archive.

Keywords: 위키데이터 연계, 위키데이터 모델링, 국채보상운동 기록물, 매뉴스크립트 컬렉션
아카이브, 아카이브 데이터 활용

Wikidata linkage, Wikidata modeling, National Debt Redemption Movement
record, manuscript collection archive, archive data utilization

1. 서론

1.1 연구의 필요성 및 목적

국채보상운동 기록물은 국권침탈의 시기인 1920년대 일어난 국채보상운동의 전개 과정과 그 과정에서 일어난 다양한 사건, 국채보상운동과 연관된 수많은 인물 등 더할 나위 없이 중요한 내용을 온전히 품고 있는 역사적 사료로서 가치를 지니고 있다. 국채보상운동 기록물은 2017년에 세계 기록유산에 등재되어 전 세계적으로 역사적 가치를 인정받았고(박승희, 2020), 국채보상운동기념사업회는 국채보상운동의 정신을 세계에 널리 알리고, 누구나 국채보상운동 기록물에 접근하여 활용할 수 있도록 디지털 아카이브를 구축하였다. 국채보상운동 디지털 아카이브를 통해 국가기록원, 국립고궁박물관, 한국연구원, 개인소장자 등 각 기관에 산재된 기록물들을 디지털 원문, 원문해제, 영문 번역본으로 웹상에서 검색하여 활용할 수 있도록 정보서비스를 제공하고 있다(국채보상운동기념사업회, 2020; 2021b).

최근의 디지털 아카이브는 누구나 아카이브의 자원을 검색하고 활용할 수 있도록 정보의 접근성을 높이고, 데이터 수준에서 기록 자원을 개방하고 공유하는 방식을 모색하고 있다(한상은, 박희진, 2022). 이를 위해 많은 문화유산 기관들은 자관이 보유한 디지털 아카이브에 시맨틱 기술을 적용하여 오픈 아카이브, 혹은 시맨틱 아카이브화 하는 방법을 시도하고 있다. 소장자료나 기록물의 디지털화 및 홈페이지를 통한 자료제공에서 나아가 더 많은 사람들에게 기록 자원을 개방하여 그 활용성을 높이는 것을 목표로 하고 있는 것이다. 그러나 이러한 작업은 기관의 규모 및 인력, 시스템 구축에 대한 비용 문제가 수반된다. 국가 단위의 디지털 아카이브나 범국가적 디지털 아카이브와 달리 국채보상운동 디지털 아카이브처럼 소규모의 로컬 아카이브는 예산상의 한계와 운영 인력, 시스템 규모의 제한으로 보다 비용 효과적인 데이터 개방 방법에 대해 고민해 볼 필요가 있다.

이러한 관점에서 현재 로컬 아카이브 내에서만 이용할 수 있는 국채보상운동 기록물 데이터를 위키데이터를 통해 외부의 다양한 기관들과 연결함으로써 이용자들에게 다양한 접근점을 제공하고 자원을 효율적으로 관리하는 방안을 고려할 수 있다. 위키데이터는 웹상에서 데이터를 공유하고 협력할 수 있는 개방형 데이터 모델로 많은 문화유산 기관에서 아카이브 데이터의 구축과 운영 관점에서 제기되는 인프라 문제를 해결하는 대안으로 제시되고 있다. 위키데이터는 누구나 구축할 수 있으며, 무료로 자유롭게 데이터를 재사용할 수 있도록 라이선스를 부여한다. 위키데이터에 구축된 데이터는 다른 위키미디어의 주요 자원들과 함께 연계되어, 웹상에 노출될 가능성이 높아 소규모의 로컬 아카이브를 알리고 컬렉션에 대한 접근성을 높일 수 있다. 일반적인 시스템은 단순한 문자열 값으로 구성된 메타데이터를 바탕으로 특정 자원에 관한 설명을 제공하는 데 그친다. 그러나 위키데이터로 구축된 데이터는 데이터의 연계와 공유를 통해 위키 플랫폼에서 공유하고 있는 수많은 데이터를 지식베이스로 활용하며 해당 자원과 관련된 다양한 개념으로 정보 탐색을 이어 나갈 수 있게 한다.

미국 의회도서관이 이름과 주제 전거데이터에 위키데이터를 연계하여 데이터를 강화하고 웹상에 공개된 관련 정보를 제공하고 있으며(Ferriter, 2019), 독일과 캐나다, 그리고 네덜란드의 문화예술 아카이브 분야에서도 로컬 아카이브 메타데이터 속성과 위키데이터 속성 정보를 연계하여 이용자에게 아카이브 자원과 관련된 풍부한 데이터와 정보를 제공하고 있다(Evenhaugen, 2022; Faraj & Micsik, 2019; Richard, 2022). 중국에서도 도서관 사서들이 전 세계에 있는 중국 관련 자료들을 위키데이터와 연계하여 서비스 하는 프로젝트를 수행하고 있다(Heng et al., 2022). 국내의 경우 박하람과 김학래(2021)는 위키데이터의 속성 정보를 활용하여 디지털 아카이브 온톨로지 구축 시 기록의 맥락에 대한 정보를 풍부하게 할 수 있음을 보여주었다. 한상은과 박희진(2022)은 디지털 아카이브에서 위키데이터와 연계한 사례를 분석하고, 기록 데이터와 위키데이터를 연계하기 위한 기록 기술 모델을 제안하였다.

앞선 연구들을 통해 확인한 위키데이터의 강점은 아카이브 기관에서는 데이터 연결과 공유를 위한 자원 식별자의 히브로 사용할 수 있으며, 웹상의 검색 과정에서 연계된 위키데이터를 통해 아카이브의 접근성을 높이고, 이용자에게

게는 관련된 풍부한 정보들을 활용해 탐색을 확장할 수 있다는 점을 들 수 있다. 국채보상운동에 관한 배경지식이 부족한 일반 이용자의 경우, 아카이브를 이용할 때 원하는 자원이 무엇이며, 어떻게 찾아야 하는지 잘 모를 수 있다. 하지만 위키데이터처럼 여러 데이터가 연결된 형태의 시스템은 한 데이터에서 연결된 또 다른 데이터로 이동하는 브라우저를 통해 기존에 알지 못했던 새로운 지식의 발견을 가능하게 한다. 전문 이용자 또한 자신이 검색하여 얻은 결과 외의 사실을 추가로 살펴보면서 더 많은 정보를 획득할 수 있을 것이다.

본 연구는 국채보상운동 디지털 아카이브의 기록물에 대한 접근성을 높이고, 데이터의 개방 및 활용을 위해 위키데이터와의 연계를 위한 모델링을 수행하였다. 앞선 연구를 통해 데이터나 기록물의 위키데이터 연계의 강점에 대해 확인하였지만, 아직 국내에서는 기록데이터를 기반으로 한 위키데이터를 연계한 사례 연구가 없다. 이에 본 연구는 국채보상운동 디지털 아카이브 기록물의 내용, 맥락, 구조를 분석하여 기록데이터를 위키데이터로 연계하기 위한 데이터 모델을 설계하고 기록물 검색에 적용가능성을 확인하고자 한다.

1.2 연구의 범위 및 방법

본 연구는 국채보상운동 디지털 아카이브의 기록물에 대한 접근 및 검색성을 높이기 위한 방법으로 위키데이터와 연계하는 데이터 모델을 설계하는데 초점을 두었다. 첫째, 국외 문화유산기관의 위키데이터 연계 사례를 조사하여, 문화유산 데이터 구조의 핵심 객체 및 속성, 연계 방법 등을 분석하였다. 둘째, 국채보상운동 디지털 아카이브의 관리 시스템과 서비스 현황을 파악하고, 기록물 기술과 검색에 사용되는 메타데이터, 시소러스, 네트워크 모델을 분석하여 위키데이터로 연계하기 위한 데이터 모델링에 반영할 사항들을 파악하였다. 셋째, 국채보상운동 기록의 내용분석을 통해 국채보상운동 기록의 맥락을 파악할 수 있는 시간, 장소, 사건 객체를 도출하고, 객체들 간의 관계 속성을 정리하였다. 넷째, 설계한 위키데이터 연계모델의 객체에서 기록물건과 행위주체, 즉 인물 객체에 실제 국채보상운동 기록물 사례를 적용하여 데이터 모델의 적용가능성을 확인, 검토하였다(〈표 1〉 참조).

〈표 1〉 연구의 범위 및 방법

연구 절차	연구 내용과 범위	연구 방법
위키데이터 연계 사례 조사 및 분석	문화유산 기관의 위키데이터 연계 프로젝트 사례를 중심으로, 프로젝트의 개요, 목적, 위키데이터와의 연계 방법(직접 연계 혹은 데이터 기여 등), 데이터 구조의 핵심 객체 및 속성들을 조사	- 연계 사례 프로젝트 웹페이지, 서비스 웹페이지 조사 - 프로젝트 관련 연구논문 및 웹자원(블로그, 홈페이지 등) 조사
국채보상운동 디지털 아카이브 현황 분석	국채보상운동 디지털 아카이브 기록물의 구조 및 내용을 분석하여 시사점 도출 및 위키데이터와 연계하기 위해 필요한 고려사항 정리 1) 국채보상운동 디지털 아카이브의 관리 시스템 및 웹서비스 현황 조사 2) 기록물 기술과 검색에 사용되는 메타데이터 조사 3) 일반주제 및 단체명, 지명, 인명 시소러스 조사 4) 1907 KOREA 네트워크 모델 및 서비스 분석	- 국채보상운동 디지털 아카이브의 웹서비스 현황, 메타데이터 요소, 시맨틱 네트워크 모델 현황 분석 - 국채보상운동기념사업회의 기존 프로젝트 관련 문헌조사 - 메타데이터 분석 - 시소러스의 구조 및 내용 분석
국채보상운동 기록물 내용 분석	국채보상운동 기록물의 제목, 한글번역본 원문의 내용을 분석하여 데이터 모델의 핵심 객체(인물 및 단체, 시간, 장소, 사건) 도출	- EDM, DPLA, 국가기록원 등의 데이터 모델 관련 문헌조사 - 기록 원문의 텍스트 데이터 추출 후 빈도 분석, 용어 동시출현 분석(엑셀 및 네트워크 분석 도구 사용)
기록물, 인물 데이터 모델링	데이터 모델의 객체들 간의 관계속성을 도출하고 기록물, 인물 모델링을 수행하고 적용가능성 파악	설계한 모델에 국채보상운동 데이터를 적용하고, 위키데이터 페이지에 입력하여 적용가능성 파악

2. 위키데이터와 위키데이터 연계 사례

2.1 위키데이터

위키데이터는 위키미디어 재단(Wikimedia foundation)이 운영하는 지식 베이스로, 인간과 기계 모두 읽고 편집할 수 있는 구조화된 개방형 다국어 지식 서비스를 제공한다. CC0 1.0 라이선스를 바탕으로 데이터를 배포함으로써 누구나 다양한 목적으로 자유롭게 활용할 수 있다. 또한 위키백과(Wikipedia), 위키미디어 커먼즈(Wikimedia Commons)을 비롯한 각종 위키미디어 프로젝트에서 공통으로 사용할 수 있는 구조화된 데이터의 수집 및 제공에 목적을 둔 중앙 저장소의 역할을 하고 있다(위키데이터, n.d.)

저장소의 기본 단위는 항목(item)이며, 위키데이터는 각 항목을 식별하고, 설명하는 데 초점을 맞춘 문서 지향형(document-oriented) 데이터베이스이다. 한 항목은 접두어 Q와 숫자의 두 요소로 구성된 식별기호(QID 혹은 Q number)로 기술되고 연결된다. 아래의 <그림 1>은 ‘국채보상운동’ 항목에 대한 위키데이터 페이지이다. 최상단에는 대표 레이블인 ‘국채보상운동’과 식별자(Q623859)가 있으며, 바로 아래 항목에 대한 설명과 다른 이름(별칭)이 표시된다. ‘국채보상운동’에 대한 상세한 정보들은 속성(Property)과 값으로 이루어진 서술(statement) 부분에 기술되고, PID 혹은 P number의 고유 식별기호를 갖는다. 속성의 값은 식별기호를 갖는 항목이나 식별기호가 없는 일반 문자열로 기술할 수 있다.

The screenshot displays the Wikidata page for '국채보상운동' (Q623859). The page is structured as follows:

- 레이블 (Label):** 국채보상운동 (Q623859)
- 설명 (Description):** 대한제국 대구 광문사터 시민 모금 운동
- 다른 언어 (Other Languages):**

언어	레이블	설명	다른 이름
한국어	국채보상운동	대한제국 대구 광문사터 시민 모금 운동	국채 보상 운동
영어	National Debt Repayment Movement	movement (1907-1908) in the Korean Empire to restore national sovereignty by paying for government bonds with public fundraising	National Debt Redemption Move...
- 속성 (Properties):**
 - 시작된 날 (Start Date):** 2월 1907 (대한제국)
 - 참고문헌: 1 개의 참고문헌
 - 참고문헌 URL: http://encykorea.aks.ac.kr/Contents/I...tem/E0006527
 - 종료된 날 (End Date):** 7월 1908 (대한제국)
 - 참고문헌: 1 개의 참고문헌
 - 출발지 (Origin):** 대구광역시
 - 참고문헌: 0 개의 참고문헌

<그림 1> 위키데이터 ‘국채보상운동(Q623859)’ 예시
출처: <https://wikidata.org/wiki/Q623859>

2.2 위키데이터 연계 사례

캐나다, 미국을 비롯하여 독일과 네덜란드, 중국 등 문화유산기관에서 기관의 문화유산자원을 위키데이터와 연계하여 데이터를 풍부하게 만들고, 위키데이터 플랫폼에서 공유하는 새로운 아카이브 서비스를 시도하고 있다. 본 연구는 위키데이터를 기반으로 데이터를 연계하고 공유하여 활용하고 있는 문화유산기관의 사례들을 위키데이터와 연계 대상, 데이터 구조, 연계 방법 중심으로 살펴보았다. COURAGE(Cultural Opposition - Understanding the cultural Herit AGE of Dissent in the Former Socialist countries)는 EU의 Horizon 2020에서 자금을 지원하여 2016년부터 2019년까지 3년간 진행된 프로젝트의 결과물이다. 불가리아, 체코, 독일, 헝가리, 폴란드 등 동유럽 사회주의의 문화적 반대에 관심이 있는 학계의 연구자들이 모여 사회주의 시대의 문화적 탄압과 관련된 기록들을 수집하여 컬렉션을 구축하였고, 2023년 4월 현재, 디지털 아카이브를 통해 563개의 컬렉션, 1,055개의 아이টে를 제공하고 있다. COURAGE는 위키데이터를 통해 COURAGE의 컬렉션과 개별항목 정보들을 더 풍부하게 하고 접근, 활용성을 높이기 위해서 인물(persons)과 기관 및 단체(groups/organizations) 객체를 위키데이터와 연계하였다(Faraj & Micsik, 2019). 인물과 단체는 컬렉션과 관련된 행위주체이기도 하며, 컬렉션을 구성하고 있는 기록물건(item)의 행위주체이다. COURAGE project 디지털 아카이브의 인물(people) 메뉴와 기관 및 단체(groups&org) 메뉴에서 행위주체를 검색할 수 있으며, 상세보기 페이지에서 위키데이터로 이동하는 링크를 확인할 수 있다.

20세기 언론 아카이브(20th Century Press Archives; 이하 PM20)는 독일에서 1908년부터 2005년 사이에 발행된 신문자료 1,500개 이상, 페이지로는 2백만 페이지 가량을 디지털 아카이빙하여 서비스하고 있다. PM20은 인물(persons), 일반 주제와 사건(general subjects and events), 회사(companies), 제품(products)과 관련된 주요 주제를 선정하여, 각 주제별 정보를 디지털 아카이브를 통해 무료로 제공하고 있다. PM20은 위키데이터를 활용하여 자원에 대한 검색과 접근을 용이하게 하고, PM20의 데이터를 어느 누구나 활용할 수 있도록 위키데이터에 통합하는 작업을 수행하였다. 신문자료에서 추출한 개인, 회사 혹은 기관명, 그리고 주제가 될 수 있는 장소나 주제명 데이터를 위키데이터에 통합(데이터 기여)하였고, PM20 디지털 아카이브에서는 인물, 회사/기관, 지역명, 주제명으로 검색 및 브라우징 할 수 있으며 상세정보에 위키데이터로 이동할 수 있도록 링크로 연계하였다.

네덜란드 왕립도서관(Koninklijke Bibliotheek)은 중세시대의 매뉴스크립트 컬렉션을 디지털화하여 제공하고 있다. 컬렉션에서 채색이 되어 있는 매뉴스크립트의 400개를 선별하여 위키미디어 커먼즈 페이지에 업로드 하고, 위키데이터와 연결하고 각 매뉴스크립트에 대한 위키백과 페이지를 작성하여 컬렉션을 연계하였다. 이 컬렉션은 네덜란드 왕립도서관 홈페이지에서 위키미디어 커먼즈 페이지와 링크를 통해 연결되며, 위키미디어 커먼즈에서도 위키데이터의 해당 페이지로 이동하게 하여 다양한 경로로 관련 정보를 접근, 활용할 수 있게 하였다.

중국의 고대인물 초상화 - 위키데이터 연계 프로젝트는 스미소니언 도서관 및 아카이브가 공동목록활동의 일환으로 파일럿으로 시작하였다(Richard, 2022). 이 프로젝트를 통해 Freer 아트 갤러리와 국립 아시아 예술품 박물관의 Arthur M. Sackler 갤러리 컬렉션에 있는 90개의 중국 고대인물 초상화와 위키데이터를 연계하였다. 연계의 핵심은 초상화의 대상이 되는 인물(sitter)에 대한 상세한 정보를 위키데이터와 연계하여 자원에 대한 접근성을 강화하는 것이었다. 초상화에 대한 면밀한 분석을 바탕으로 사람(human)의 주요 역할(creator or depicted people, contributor)과 초상화의 시대적, 공간적 범위가 되는 맥락정보를 위키데이터와 연계하였다. 각 초상화의 인물별로 개별 위키미디어 커먼즈 페이지를 생성하고, 위키데이터와 위키백과 등에 존재하는 다른 개별항목 페이지와도 모두 연결하였다. 이 프로젝트는 초상화의 이미지를 위키 네트워크에 업로드하여 위키미디어에 기여하였다.

스미소니언 도서관 및 아카이브의 위키데이터 연계 프로젝트는 파일럿 프로젝트에서 시작하여 기관이 보유한 예술자료와 예술가 데이터 컬렉션을 위키데이터에 업로드하는 것으로 발전하였다(Evenhaugen, 2022). 기관이 보유한 예술가 데이터는 약 6만 건으로 예술가의 개인에 대한 정보, 예술가 집단, 갤러리의 정보 등으로 구성되었는데, 도서관 목록과 같이 표준화할 수 있는 기술 정보가 부족하여, 데이터를 의미 있게 연결하고 데이터의 활용성을

높이기 위해 위키데이터와 연계하는 방식을 채택하였다. 기관이 정제한 아티스트 3,797건을 위키데이터에 업로드함으로써 위키미디어 네트워크에 기여하고, 위키데이터를 통해 기관 예술가에 대한 다양하고 유용한 데이터를 보다 많은 사람들이 검색하여 활용할 수 있게 하였다.

지금까지 살펴본 문화유산기관에서 위키데이터와의 연계 사례의 주요 특징들을 정리하면 아래의 <표 2>와 같다. 초창기의 위키데이터 연계 프로젝트는 기관에서 보유한 자원에 대한 웹상의 접근 및 검색성을 높이기 위한 방법으로 위키데이터를 적용하였다. 기관의 아카이브 자원과 관련된 인물이나 단체, 그리고 주제와 같은 객체들을 위키데이터와 연계하기 위해 데이터를 모델링하고, 위키데이터의 항목들과 매핑하였다. 위키데이터 연계 사례는 파일럿 혹은 협동 연구들로 점차 늘어나게 되고, 단순히 위키데이터와의 매핑을 통해 데이터를 연결하는 수준에서 벗어나, 기관의 데이터를 위키플랫폼에 제공하여 데이터를 기여하는 것까지 발전하게 되었다. 본 연구에서 조사한 PM20, 네덜란드 왕립도서관 중세시대 매뉴스크립트, 중국 고대인물 초상화 프로젝트는 기관에서 수집한 유용한 아카이브 데이터를 위키데이터로 공유함으로써, 의미 있는 데이터 연결을 가능하게 하고, 많은 사람들이 데이터를 활용할 수 있도록 기여하고 있다.

중국 고대인물 초상화 프로젝트는 위키데이터를 전통적인 도서관 목록작업에 적용하고자 시작된 프로젝트라고 밝히고 있다(Heng et al., 2022). 도서관목록의 표준화에 있어서 핵심적인 역할을 수행하는 미국 의회도서관도 이름전거(Library of Congress Name Authority File, LCNAF)와 주제명표목(Library of Congress Subject Headings, LCSH)을 위키데이터와 연계하고 있다. 의회도서관의 LCNAF은 인물, 단체, 사건, 장소, 자원의 제목으로, LCSH 주제, 지리적 명칭, 이름, 주제의 하위분류로 사용되는 용어와 이름으로 구성되어 있는데, 이들은 해당되는 위키데이터 항목이 존재하는 경우 직접 이동이 가능하도록 전거 페이지에 링크로 이동이 가능하도록 연계하였다. 미국 의회도서관의 전거 파일을 전 세계의 많은 나라와 기관에서 활용하고 있다는 점에서 미국 의회도서관의 위키데이터 연계는 시사하는 바가 크다. 또한 미국의 디지털공공도서관(Digital Public Library of America)과 Europeana에서도 위키데이터에 회원기관의 데이터를 기여하기 위한 방법론과 도구들을 연구 개발하고 있으며(DPLA, n.d.; Freire et al., 2019) 이러한 국가기관이나 협동네트워크의 기여는 위키미디어 네트워크를 더욱 더 활성화시킴으로써 소규모 기관에서 위키데이터를 활용할 수 있는 토대를 마련할 수 있다는 점에서 의미가 있다고 볼 수 있다.

<표 2> 위키데이터 연계 사례 - 주요 내용 요약

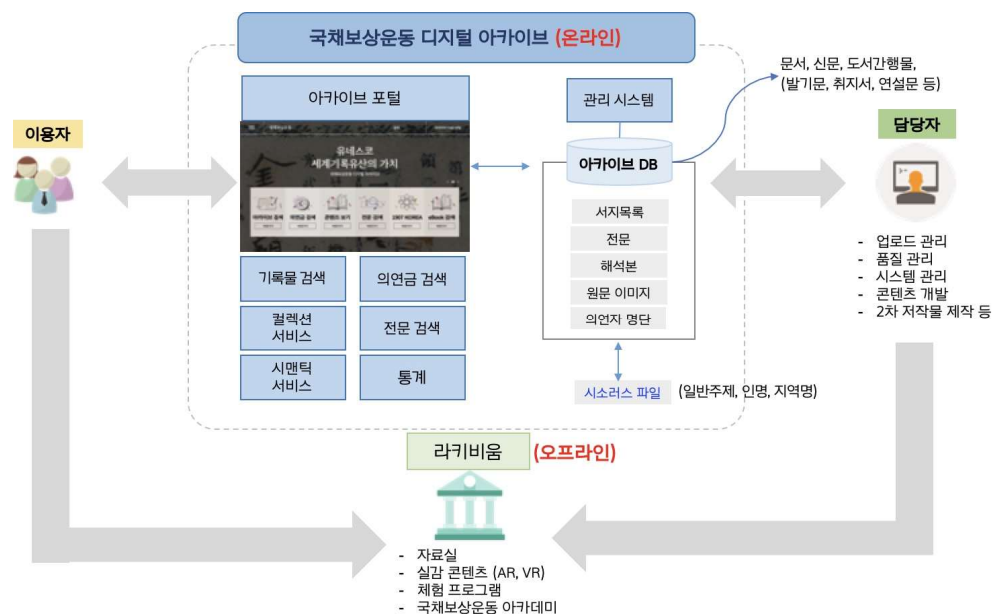
프로젝트명 구분	COURAGE	PM20	MMDC	Chinese Ancestor Portrait	Smithsonian Artist file
연계대상	사회주의 시대의 문화 탄압에 대한 모든 유형 의 콘텐츠 매뉴스크립 트 컬렉션	전세계에서 1908-2005 년에 발행된 신문자료 매뉴스크립트 컬렉션	네덜란드 왕립도서관의 중세시대 매뉴스크립트 컬렉션 중 채색된 400개 의 매뉴스크립트 컬렉션	스미소니언 Arthur M. Sackler 갤러리 컬렉션 에 있는 90개의 중국 고 대인물 초상화 컬렉션	스미소니언이 보유한 2 개의 예술가 DB에서 정 제한 3,797명의 예술가
기록 계층	collections - featured items	archives - collection - file - item	collections - item	collection - item	item
연계대상의 주요 객체	events, people, groups&orgs	persons, general subjects and events, companies, products	location	human, subject, location	human
위키데이터 연계 객체	people, groups&orgs	persons, companies/institution s, places, subjects	creator, place	creator, location, point in time, time period, significant event	people, place, collection
연계방법	데이터 기여 (donation), 링크 연계	데이터 기여 (donation), 링크 연계	데이터 기여 (donation), 링크 연계	데이터 기여 (donation), 링크 연계	데이터 기여 (donation),
예시	http://cultural-opposition.eu/registry/?uri=http://courage.btk.mta.hu/courage/individual/n249444&type=people	https://pm20.zbw.eu/category/subject/i/144570/about.en.html	https://www.kb.nl/ontdekken-bewonderen/topstukken/beatrijs	https://www.wikidata.org/wiki/Q296367	https://www.wikidata.org/wiki/Q4733608

3. 국채보상운동 디지털 아카이브 현황 분석

3.1 기록물의 유형 및 서비스 분석

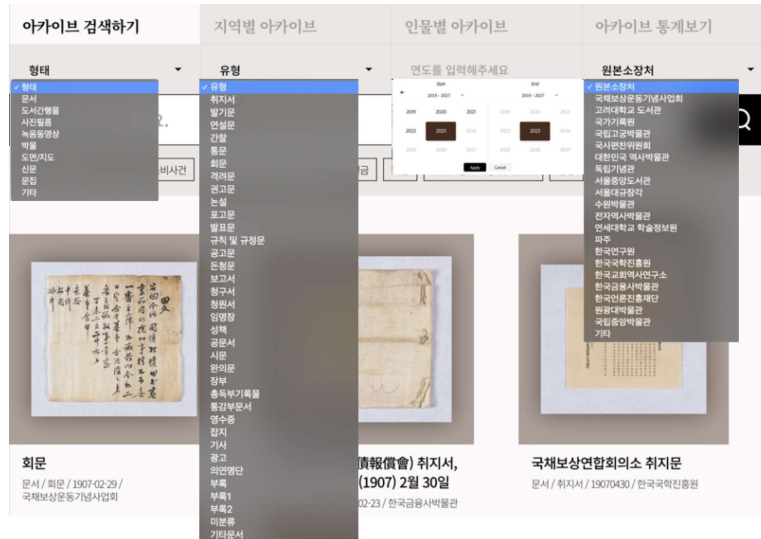
국채보상운동 디지털 아카이브에는 국채보상운동과 관련된 기록물이 총 4,134건이 구축되어 있다(2022년 12월 기준). 그 중 세계기록유산에 등재된 기록물은 2,475건이고, 나머지는 미등재 기록물이다. 실제로 기념사업회가 소장하고 있는 기록물은 세계기록유산 등재기록물 2,475건과 미등재기록물 약 4천여 건이다(김지옥, 2017). 기록물은 형태에 따라 문서, 도서관행물, 사진필름, 녹음동영상, 박물, 도면·지도, 신문, 문집, 기타로 구분되며, 국채보상운동 디지털 아카이브에서는 문서 275건, 신문 3,822건과 문집 35건과 국채보상운동기념사업회가 발간한 도서관행물이 제공된다. 기록물의 유형은 취지서, 발기문, 연설문, 간찰, 통문, 회문, 격려문, 권고문, 논설, 포고문, 발표문, 규칙 및 규정문, 공고문, 돈청문, 보고서, 청구서, 청원서, 임명장, 정책, 공문서, 시문, 완의문, 장부, 총독부기록물, 통감부문서, 영수증, 잡지, 기사, 광고, 의연명단, 부록, 미분류, 기타문서로 분류된다. 디지털화한 기록물의 원본은 국립고궁박물관이 1,189건, 한국연구원이 1,300건을 소장하고 있다. 그 외 원본 소장처는 서울중앙도서관, 연세대학교학술정보원, 국사편찬위원회 등 20개 이상의 기관과 개인소장자, 그리고 출처미상의 자료로 구분되며 국채보상운동기념사업회는 이들을 모두 수집하여 디지털화한 매뉴스크립트를 보유하고 있다(김지옥, 2017).

국채보상운동기념사업회는 다른 기관에 산재되어 있는 국채보상운동 기록물들을 세계기록유산 공동등재기관(자료소장기관)과의 협약을 통해 수집하여 디지털화 작업을 진행한 후 매뉴스크립트를 디지털 아카이브로 구축하였고, 이용자가 검색하여 활용할 수 있도록 웹 기반의 포털 서비스를 제공하고 있다. 디지털 아카이브는 기록물을 관리하는 데이터베이스와 이용자 서비스를 제공하는 아카이브 포털로 구성되어 있다(〈그림 2〉 참조). 기록물 데이터베이스는 기록물 원본이미지와 국·영문 해석본과 기록물 메타데이터, 검색도구인 시소러스(일반주제, 인명, 지역명), 의연자 명단, 인물에 대한 메타데이터를 포함하고 있다. 아카이브 포털에서는 구축이 완료된 기록물 4,134건과 의연금명단, 인물, 도서관행물에 대한 검색과 시맨틱 네트워크 그래프 서비스를 제공하고 있다(국채보상운동기념사업회, 2020; 2021b).



〈그림 2〉 국채보상운동 디지털 아카이브 시스템 개념도

2023년 4월 현재 디지털 아카이브에서 검색하여 접근 가능한 기록물은 4,118건으로 취지서 637건, 발기문 8건, 간찰 1건, 권고문 1건, 논설 58건, 공고문 2건, 보고서 12건, 기사 997건, 광고 366건, 의연명단 416건, 미분류 1,324건이다. 디지털 아카이브의 주된 검색 서비스인 ‘아카이브 검색하기’에서 기록물에 대한 검색이 가능하며, 지역별, 인물별, 형태(예: 문서, 신문 등), 유형(예: 발기문, 통감부문서 등), 연도, 원본 소장처에 따라 검색조건을 제한하여 상세 검색할 수 있으며(<그림 3> 참고), 브라우징은 지역명, 인물명으로만 지원된다.



<그림 3> 국채보상운동 디지털 아카이브 검색 페이지 예시
출처: <http://www.gukchae-archive.org/gp/page/itemList>

3.2 기록물 기술과 검색 도구 분석

국채보상운동기념사업회는 국채보상운동 기록물 검색과 관리를 위하여 메타데이터 스키마, 시소러스와 데이터 시각화를 위한 네트워크 그래프 서비스를 개발하여 사용하고 있다. 기록물의 등록 및 보존, 그리고 서비스에 사용하는 메타데이터는 식별 영역, 배경 정보, 추가 정보의 세 가지 영역으로 구분되어, 총 58개의 요소로 구성되어 있다(국채보상운동기념사업회, 2019). 필수항목은 식별 영역 8개, 배경 정보 5개, 추가 정보 4개로 총 17개 요소이다.

<표 3> 국채보상운동 기록물 메타데이터 요소 (*: 필수 요소)

영역	식별 영역	배경 정보	추가 정보
요소명	기록물번호	생산처 위치구분*	기념관 복제본 소장여부*
	종류*	생산처 구분*	박물관코드
	원문언어*	생산기관	유네스코등재여부*
	원문제목*	생산기관 조직명	유네스코등재번호
	원문	생산자명	크기*
	번자제목	원본소장처 수집정보 중 수집방법	쪽수
	번자내용	원본소장처 수집정보 국내외구분	수량
	해석제목(국문/영문)	원본소장처 수집정보 기관구분	관련문서번호
	해석내용(국문/영문)	원본소장처 수집정보 중 기관정보	비고1
	내용요약	원본소장처 수집정보 중 소장자명	비고2
	키워드	원본소장처 수집정보 중 수집일자	비고3
	생산일자 확인여부*	원본소장처 수집정보 중 기증자	문서유형*
	생산일자	원본소장처 수집정보 중 기증일자	구 문서번호
	생산종료일자	비용	수신자
	생산일자 기타	열람방식*	발신자

국채보상운동 디지털 아카이브의 검색 결과 상세보기 화면에서 제공되는 메타정보는 기록물의 유형에 상관없이 아카이브 유형(예: 문서, 신문 등으로 등록 과정에서 식별영역의 ‘종류’의 값), 생산일자, 소장처, 복제본 소장 여부, 크기, 유네스코 등재여부, 문서유형(예: 발기문, 통감부문서 등), 아카이브 설명과 동일한 정보를 제공하고 있다. 검색 결과 화면에서는 원문과 국문, 영문 해석본 뿐 아니라 원본이미지의 열람 또한 가능하다.



〈그림 4〉 국채보상운동 디지털 아카이브: 검색 결과 상세보기 화면 예시
출처: <http://www.gukchaebosangundong.org/gp/page/itemInfo>

국채보상운동 메타데이터의 식별영역, 배경정보, 추가정보의 영역 구분과 기술항목명은 ISAD(G) 2판을 기준으로 설계되어 있다(김경남, 2022). 국채보상운동 디지털 아카이브 관리 시스템의 메뉴에 ‘기록물명칭(건)’과 ‘기록물명칭’, 그리고 기록물 시리즈를 구성하는 메뉴를 제공하고 있는데 아래의 <그림 5>에서와 같이, 기록물명칭의 값과 기록물명칭의 값이 모두 동일하여, 기록물의 계층구조를 파악하기 어려우며 기록물건(item) 단위로 관리하고 있음을 알 수 있다.



〈그림 5〉 국채보상운동 디지털 아카이브: 기록물건(item) 단위 등록 화면

시소러스는 국채보상운동기념사업회가 2020년에 국채보상운동 기록물의 활용성 증대를 위해 개발한 것으로, 일반주제 및 단체명, 지명, 인명으로 나누어있다. 각 시소러스의 우선어는 일반주제 및 단체명 557개, 지명 634개, 인명 264개이다. 시소러스는 세계기록유산에 등재된 2,475건의 기록물과 국채보상운동 관련 논문 92건, 단행본 14책, 그리고 추가적으로 발굴한 국채보상운동 관련 기록물 1,674건에 출현한 단어를 언어학적 처리 과정을 거쳐 표준화 및 상호관계를 설정한 것이다(국채보상운동기념사업회, 2020). 아래의 <그림 6>은 일반주제 및 단체명의 시소러스에서 ‘국채보상금’ 용어의 예시이다. ‘국채보상금’은 국채보상운동 기록물 내용 분석에서 가장 많이 출현한 단어이다. 일반주제 및 단체명 시소러스에는 개념(용어)의 뜻을 설명하는 “의미”와 “특성”, 관련된 문서를 보여주는 “출처”와 등장한 기록물의 유형인 “문서종류”로 관계어를 표현하고 있는데, 이는 지명과 인명에는 나타나지 않는 관계속성이다.

<일반주제 및 단체명> 국채보상금 의미 1300만원 UF 국채보상전 BT 국채보상운동 BT 1300만원 RT 국채 한자어 國債報償金 특성 사물명 출처 공립신보, 만세보, 대한매일신보, 대한민보, 매일신보, 황성신문, 미상, 수기자료, 통감부 문서, 활자자료, 총독부 기록물, 경향신문, 경남일보 문서종류 기탁문서, 기사, 격려문, 간찰, 보고문, 청원서, 권고문, 취지서, 신문기사, 통문, 광고문, 광고, 총독부 기록물, 사고, 잡보, 공문서, 신문 국채보상번호 0029, 0091, 0092, 0096, 0101, 0104, 0150 포함 총 366건

<그림 6> 국채보상운동 기록물 시소러스 - 일반주제 및 단체명 ‘국채보상금’ 예시

지명 중 가장 많은 기록물에서 출현하는 용어는 ‘대구’이다. <그림 7>을 보면 지명과 관련된 관계어는 일반주제 및 단체명과 비교하였을 때, 용어 관계가 단순하며 마찬가지로 국채보상운동 기록물 번호와 연계하고 있다. <그림 8>은 가장 많은 기록물에서 출현하는 ‘인명’인 국채보상운동 발기인 ‘서상돈’ 용어의 예시이다. 인명 시소러스는 일반주제 및 단체명, 지명과는 달리 상위어, 하위어, 우선어, 비우선어와 같은 개념 간의 관계가 없는 인명 데이터베이스라 할 수 있다.

<지명> 대구 UF 대구부, 대구군, 대구읍, 달성 BT1 경상북도 NT 남일동 RT 현풍 한자어 大丘 국채보상번호 0002, 0015, 0047, 0067, 0072 포함 총 47건

<그림 7> 국채보상운동 기록물 시소러스 - 지명 ‘대구’ 예시

<인명>

서상돈

한자명 徐相敦

이칭 아우구스티노

출생일 1850.10.17.

사망일 1913.06.30.

거주지 대구

이력 광문사 부사장

국채보상운동참여 국채보상운동 발기인, 대구단연상채소 회원

국적 대한민국

훈장 건국훈장 애족장

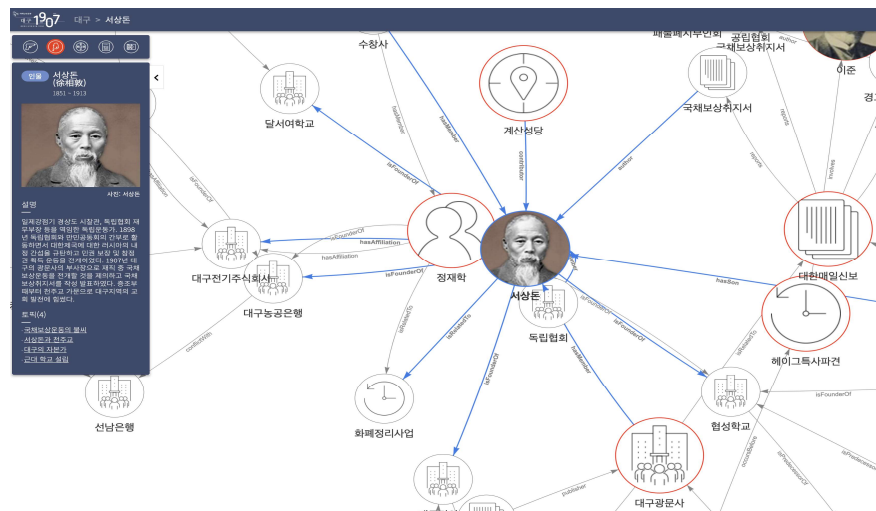
국채보상번호 0002, 0005, 0086, 0090, 0196 포함 총 38건

<그림 8> 국채보상운동 기록물 시소러스 - 인명 '서상돈' 예시

구축된 시소러스는 검색에 활용된다. 예를 들어, 국채보상운동 디지털 아카이브의 검색창에 '국채보상금'을 키워드로 입력하게 되면, 용어 '국채보상금'과 연결되어 있는 국채보상번호에 해당하는 기록물 366건이 검색 결과로 제시된다. 국채보상운동기념사업회는 기록물 분석을 통해 다섯 개의 핵심 테마를 추출하고, 테마와 관련된 핵심 객체와 객체 간의 관계를 중심으로 설계한 온톨로지 시맨틱 모델을 개발하고, 네트워크 그래프 서비스를 제공하고 있다. 국채보상운동 기록물은 1907년부터 1910년까지 이어진 역사적 배경 속에서 생산된 기록물들이기에 시대적 배경, 그리고 기록물과 관련된 인물 및 사건과 같은 맥락들을 가지고 있는데, 디지털 아카이브에서는 이러한 기록물의 배경 및 맥락정보의 브라우징을 지원하지 못하였다. 이에 국채보상운동기념사업회와 한국학중앙연구원의 연구팀이 국채보상운동 온톨로지 기반 시맨틱 데이터 모델을 설계하고, 객체-관계 간의 브라우징이 가능하도록 네트워크 그래프를 구현하였다.

국채보상운동 시맨틱 모델의 중심 클래스는 기록물(record)과 행위자(actor)이다. 기록물의 하위 클래스는 문헌(book), 문서(document), 신문기사(media)이며, 행위자의 하위 클래스는 개인(person), 기관단체(organization), 출판사(publisher)로 구성된다. 기록물은 사건(event) 클래스와 연결되어 기록물, 행위자와 연관성이 있는 모든 행사, 사건, 행위, 크고 작은 에피소드로 정의되어 있다. 사건의 목적은 개념(concept) 클래스와 연결된다(김지명, 2018). 추후 장소(place) 클래스가 추가되었다. 관계를 통해 클래스와 클래스 간 브라우징이 가능한데, 기록물과 인물 간에는 creator, editor와 같은 관계를 통해 연결되고, 기록물과 장소는 isLocatedIn과 같은 관계로 연결된다. 디지털 아카이브의 네트워크 그래프 이용 방법 페이지에 명시된 관계의 수는 약 40여개이다.

이용자들은 국채보상운동 디지털 아카이브의 '1907 KOREA' 메뉴에서 네트워크 그래프를 통해 '저항', '대구', '경제', '언론', '외교' 테마(theme)를 브라우징 할 수 있다. 예를 들어 '대구' 테마를 선택하면, 테마에 대한 설명과 하위 토픽, 그리고 관련된 클래스들의 개별 인스턴스가 나타나며, 관계를 통해 클래스 간의 맥락 파악이 가능하다. 아래의 <그림 9>는 인물 클래스의 하나의 인스턴스인 '서상돈'을 중심으로 한 네트워크 그래프의 예시이다. 관계를 중심으로 한 네트워크 그래프를 중심으로 인물 '서상돈'은 기록물 '국채보상취지서'의 저자(author)이고, '대구광문사'의 구성원(hasMember)이자, 단체 '대구민회', '대구공농은행', '달서여학교'를 설립(isFounderOf) 하였음을 보여준다. 뿐만 아니라 관계를 통해 연결된 다른 클래스의 인스턴스와 연결된 또 다른 클래스인 기록물, 인물, 기관 등의 클래스의 인스턴스를 자유롭게 브라우징 할 수 있다. 해당 인스턴스와 연결된 다른 클래스의 인스턴스들과의 관계가 그래프로 시각화하여, 이용자들이 기록물, 인물, 기관 등의 클래스를 브라우징할 수 있다.



〈그림 9〉 1907 KOREA 네트워크 그래프 서비스 예시: 인물 ‘서상돈’

3.3 시사점

국채보상운동 디지털 아카이브는 여러 기관에서 소장하고 있는 원본 기록물을 디지털화한 국채보상운동 매뉴스크립트 컬렉션을 제공하고 있다. 유형은 문서와 신문자료가 대부분이다. 국채보상운동 디지털 아카이브 포털 서비스는 이용자들이 기록물에 대한 접근 및 검색을 위해 메타데이터, 시소러스를 개발하고, 온톨로지 기반의 시맨틱 모델을 구현하여 기록물과 관련된 맥락 객체와의 관계성을 파악할 수 있도록 하였다. 본 연구는 이러한 국채보상운동 기록물의 기술과 검색 도구 분석을 통해, 위키데이터 연계를 위한 데이터 모델링을 위해 고려할 사항들을 파악하고자 하였다.

국채보상운동 기록 메타데이터를 분석한 결과, 식별영역, 배경정보, 추가정보 등 ISAD(G)와 국가기록원 영구기록물 보존 기술을 따르고 있으나, 기록물의 계층 정보가 없으며, 배경정보와 기록물 원본이나 관련 자료들에 대한 기술 요소가 선택 사항으로 되어 있어 기록의 맥락에 대해 충분한 정보를 파악하기 어려웠다. 현재 메타데이터에서는 국채보상운동 기록의 출처, 행위자 등 관련 정보를 상세하게 기술하고 있지 않으므로, 메타데이터와의 매핑을 통한 위키데이터와의 연계는 불가능하였다. 따라서, 본 연구의 위키데이터 연계를 위한 데이터 모델에서는 국채보상운동 기록물의 내용과 구조를 파악하여, 배경을 폭넓게 이해하고 다른 기록물과의 관계를 이해할 수 있는 핵심 클래스와 클래스 간의 관계를 파악하여 연결할 수 있는 속성 정보를 도출하는 것이 필요하였다.

국채보상운동 기록 시소러스는 디지털 아카이브에 구축된 기록물에 대한 검색을 돕기 위한 목적으로, 일반주제와 단체명, 지명, 인명의 세 가지 시소러스로 개발되었다. 시소러스는 기록물에 대한 내용분석을 바탕으로 키워드의 출현 빈도가 높은 단어들을 처리하여 용어를 표준화하여 상호관계를 제시하고 있어, 국채보상운동 기록물 위키데이터 연계모델의 핵심 개념과 그 관계를 파악하는 기초 자료로 고려할 수 있을 것이다. 그러나 <그림 6>의 예시와 같이 ‘국채보상금’의 상위어가 ‘국채보상운동’이고, 의미가 ‘1300만원’인데, 상위어가 ‘1300만원’으로 되어 있듯이 개념 간의 상호관계가 명확하지 않는 문제점이 있었다. 또한, 국채보상운동 기록 시소러스에서 인물과 관련된 용어들이 단체명, 인명의 시소러스 파일로 따로 구축되어 있어, 연계데이터 모델의 인물 객체와 관련된 개념들을 구조화하고 다른 주제명의 시소러스의 개념들과의 관계를 파악하기 위해서 해당 기록물의 내용을 다시 분석하여야 했다.

‘1907 KOREA’는 이용자들이 국채보상운동의 배경을 이해하고 관련 기록물을 브라우징할 수 있도록 개발된 시맨틱 네트워크 서비스이다. 이는 국채보상운동 기록물의 제목, 초록, 원문 등의 내용을 분석하여 기록물과 행위자(인물 및 기관, 단체)를 중심으로 한 사건, 장소 등의 핵심 클래스를 도출하고, 클래스 간의 관계를 통해 이용자들이 기록물과 관련된 다양한 관점의 맥락을 파악할 수 있도록 구조화 하였다는 점에서 의미가 있다. 그러나 현재 국채보상운동 디지털 아카이브에서 ‘1907 KOREA’는 실제 기록물 혹은 인물 페이지와 연계되어 있지 않아 기록물 검색

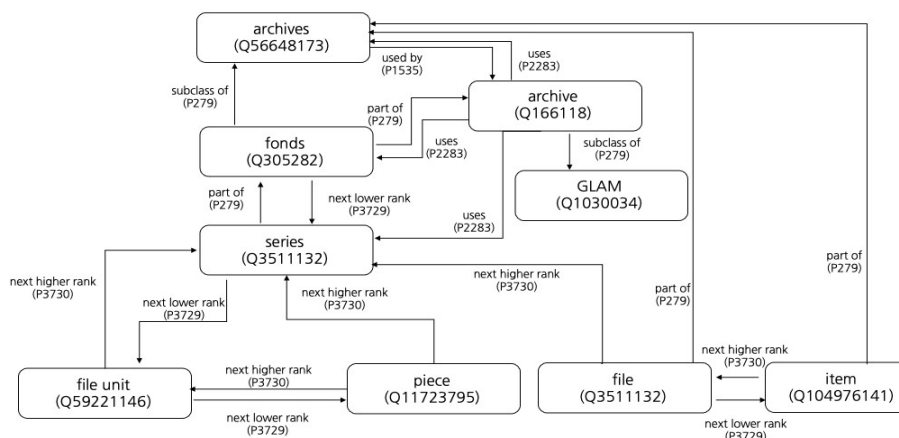
도구가 아닌 국채보상운동에 관한 개념들의 관계를 파악하기 위한 시각화 도구로서만 사용되고 있는 한계를 가지고 있다. 또한 시맨틱 웹의 온톨로지 모델링에서는 응용 시스템 간의 어휘 재사용이 매우 중요한데(김수경, 안기홍, 2008), ‘1907 KOREA’ 온톨로지는 네임스페이스를 갖지 않는 자체적인 어휘를 사용하고 있다는 점에서 자원의 연결과 확장성을 고려한다면 이에 대한 개선이 필요하다.

4. 국채보상운동 기록물과 위키데이터 연계를 위한 기록 기술 모델

4.1 다계층 기록 구조로의 매핑

국채보상운동 기록물과 위키데이터를 연계하기 위해서는 국채보상운동 기록물과 현황을 분석한 내용을 토대로 한 데이터 모델이 필요하다. 데이터 모델은 위키프로젝트에서 매우 중요한 단계로, 위키데이터 항목과의 일관성을 유지하여 데이터를 더 풍부하게 만들어주기 위함이다(Heng et al., 2022). 특히 기록물의 계층 구조와 기록물의 맥락을 파악할 수 있게 하는 행위주체, 장소, 사건이나 주제와 같은 객체들을 파악하여 관계성을 정립해야 한다. 현재 국채보상운동 기록 기술이 기록물을 중심으로 한 행위주체, 그리고 사건, 장소, 시간 등의 맥락 객체와의 완전한 매핑은 어려우므로, 기록물의 내용 분석을 바탕으로 기록의 계층 구조를 파악하고, 기록에 포함된 맥락을 이해할 수 있는 핵심 개체를 도출하고자 하였다.

보존 기록의 계층은 일반적으로 기록군(records group) 혹은 컬렉션(collection) - 기록물시리즈(records series) - 기록물철(records file) - 기록물건(item)의 다계층으로 구성된다(한국기록관리학회, 2020). 한상은과 박희진(2022)이 제안한 위키데이터 기록 기술 모델은 이러한 기록의 다계층적 구조를 바탕으로 위키데이터의 핵심 객체와 속성을 연결한 것이다(<그림 10> 참고). 위키데이터 기록 기술 모델에서 기록물 계층은 fonds(Q3052382) - series(Q3511132) - file(Q3511132) - item(Q10496141) 구조로, fonds(Q3052382)은 ‘어떤 법인 또는 물리적인 인물이나 단체가 업무상 작성하여 접수한 문서의 전체’를 보존하고 있는 archives(Q56648173)의 하위 클래스로 구성된다. 하나의 개별 archive(Q166118)는 동일한 작성자의 기록, 문서의 집합인 fonds(Q3052382)를 사용하여 구체화 될 수 있다. fonds(Q3052382)은 series(Q3511132)를 하위 계층, 혹은 클래스, 부분으로 구조화할 수 있다. series(Q3511132)는 개별 아카이브의 하위 계층이나 클래스 일 수도 있고, 품의 하위 계층이나 클래스 일 수도 있다. 시리즈는 하위 계층으로 file(Q3511132)이나 file unit(Q59221146)을 가질 수 있는데, 파일 유닛으로 구조화할 경우 piece(Q11723795)와 상/하위 관계를 갖는다. file로 계층 관계를 갖는 경우 파일의 하위에 item(Q10496141)을 둘 수 있다. 기록물의 계층 관계는 아카이브의 기록물의 관리 수준에 따라 달라질 수 있으며, 각각의 계층은 모두 가장 최상위 archives(Q56648173)의 부분이자 하위 계층이 될 수 있다. 기록물의 계층적 구조에서 모든 계층 사이에 임의의 중간 계층이 포함될 수 있다.



<그림 10> 위키데이터 기록 기술(Archival Description) 모델 (한상은, 박희진, 2022)

국채보상운동 디지털 아카이브의 기록물 계층을 다계층 구조와 매핑하여 위키데이터와 연계할 수 있는 계층 구조로 구성해 볼 수 있다. 국채보상운동 디지털 아카이브의 기록물은 국채보상운동기념사업회에 의해서 수집된 매뉴스크립트 컬렉션으로 쉘렌버그의 구분에 따르면 각급 기관과 개인의 소장 기록을 인위적으로 수집한 ‘인위적 매뉴스크립트 컬렉션’으로 볼 수 있다(김경남, 2022). 생산과 출처의 맥락에서 구분하면 품 레벨은 대한민국 국내와 국외로 구분할 수 있을 것이며, 기록시리즈는 기능뿐만 아니라 유형, 매체 등 다양한 기준에 따라 구성하고, 공공기록물이나 민간기록물, 혹은 더 하위시리즈로 의연금 시리즈, 연설문 시리즈 등으로 구성할 수 있을 것이다. 국채보상운동 기록물은 기록물건으로 정리되어 있기에 기록물건은 기록물철로 묶을 수 있고, 파일을 다시 시리즈로, 시리즈를 품으로 구성할 수도 있다. 이러한 다계층 구조로의 기록데이터와 위키데이터와의 매핑을 통해 기록물건 수준에서 관리되고 있는 국채보상운동 기록물과 관련된 다른 기록물과의 접근 및 검색성을 높일 수 있을 것이고, 품 레벨에서는 원생산지나 출처와의 연결을 통해 기록의 맥락을 이해할 수 있으며, 개념 간의 관계를 통해 다양한 기록의 집합체(시리즈)를 구성하여 이용자의 접근성을 높일 수 있을 것이다.

4.2 위키데이터와의 연계를 위한 기록물 모델링

국채보상운동 디지털 아카이브는 인명 시소러스와 함께, 인물에 대한 상세정보 제공 등 인물 위주의 검색서비스를 제공하고 있으므로, 본 연구는 인물 정보를 중심으로 데이터 모델링을 전개하였다. 기록물 중 유네스코에 등재된 기록물 2,475건 중 2명 이상의 인물이 원문에 등장하는 기록물 233건에 대한 메타데이터와 원문을 텍스트로 수집하였다. 기록물의 제목과 원문 한글 해석본을 언어학적으로 처리하여 키워드를 추출하고, 키워드의 빈도와 동시 출현 여부를 분석하여 키워드 간 관계성이 있는지를 확인하였다. 마지막으로 인명 시소러스의 용어와 비교하여 각 키워드가 속할 수 있는 핵심 클래스를 도출하였다.

앞서 위키데이터와 연계한 사례 분석에서 자원의 행위 주체가 되는 인물이나 단체, 주제와 같은 객체들을 위키데이터와 연계하고, 유로피아나의 EDM 모델과 DPLA의 데이터 모델에서도 문화유산을 사건(event), 행위자(agent), 장소(place), 시간범위(timeSpan) 객체와 연결하여 맥락 정보를 제공하고 있음을 확인하였다(DPLA, 2017; Europeana, 2017). 뿐만 아니라 한국정보통신기술협회의 ‘TTAK.KO-10.0850: 문화유산 DB 연계를 위한 맥락 기반 메타데이터 요소 및 형식’ 모델의 데이터 표준은 맥락을 표현할 수 있는 클래스를 행위주체, 사물, 사건, 시간, 장소로 제시하고 있다. 본 연구에서는 국채보상운동 기록물 위키데이터 연계 모델의 클래스를 ‘인물’, ‘기관/단체’, ‘개념’, ‘장소’, ‘사건’의 다섯 가지로 구분하였다.

〈표 4〉 기록물 중심 핵심 클래스 도출 절차 요약

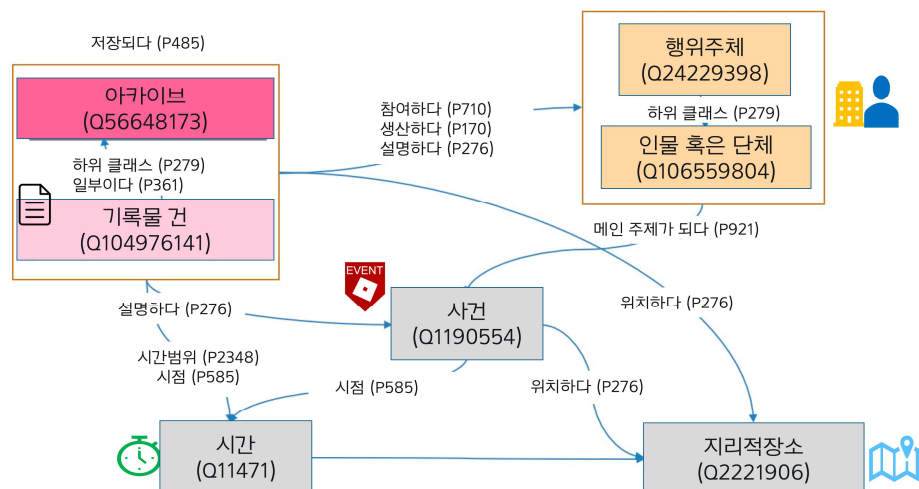
	절차	상세 연구 내용
1	인물 중심의 기록물 선정 - 233건	1) 유네스코에 등재된 기록물 2,475건 대상 2) 국채보상운동연구총서 19권 - 인물편에 등장하는 인물 추출 3) 인물별 아카이브에 속한 인물 추출 4) 인물이 2번 이상 등장하는 기록물 선정
2	원문 텍스트 분석을 통해 주요 키워드 추출 - 약 5,000건	1) 233건의 기록물 제목, 원문 한글 해석본의 텍스트 추출 2) 불용어 처리 3) 키워드의 빈도 분석 및 동시출현 여부 분석 4) 높은 빈도를 보이는 키워드 순으로 정렬(빈도 10 이상)
3	키워드별 클래스 도출	1) 키워드와 시소러스의 용어 비교 2) 수동으로 클래스 도출

아래의 <표 5>는 동시출현 빈도가 가장 높은 상위 20개 키워드에 대한 맥락 클래스 매핑의 결과를 요약한 것이다. 예를 들어 ‘부인’, ‘어머니’, ‘아내’와 같은 키워드는 인물과 관련된 키워드로 매핑하였고, ‘의연금’, ‘국채보상금’, ‘빚’과 같은 단어들은 개념(concept) 클래스와 매핑하였다. 기존에 구축된 시소러스에 우선어가 있을 경우 시소러스를 참고하여 클래스를 부여하였다. 매핑은 수작업으로 진행하였으며 공동 연구자들이 매핑 결과를 상호검토하였다. <표 5>는 상위 20개의 키워드를 나타낸 결과로 사건과 시간에 대한 값이 보이지 않지만, 실제로 ‘광무’는 시간, ‘소비사건’은 사건 클래스와 매핑하였다.

<표 5> 기록물 내용분석을 바탕으로 한 핵심 클래스 도출: 상위 20개 키워드

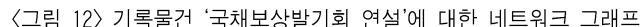
	키워드	빈도수	클래스	불용어		키워드	빈도수	클래스	불용어
1	사람	1,991	인물	○	11	베델	466	인물	
2	의연금	1,849	개념		12	아내	445	인물	
3	부인	1,497	인물		13	주사	395	인물	
4	국채보상	936	개념		14	참봉	381	인물	
5	어머니	833	인물		15	군	363	장소	
6	국채	654	개념		16	의연	363	개념	
7	국채보상금	570	개념		17	보상	360	개념	
8	돈	519	개념		18	리	321	장소	
9	원전리	479	장소		19	국채보상의무금	318	개념	
10	빚	473	개념		20	담배	318	개념	

이러한 과정을 거쳐 국채보상운동 기록물의 계층 구조와 맥락을 반영한 기록물과 맥락 클래스와의 객체 - 관계 모델은 다음의 <그림 11>과 같다. 우선 기록 계층은 국채보상운동기념사업회에 의해서 인위적으로 구성된 매뉴스크립트 컬렉션을 보유하고 있는 아카이브(Q56648173)와 기록물건(Q104976141)으로 구성하였다. 계층 간의 관계는 하위클래스 혹은 부분으로 설정하였다. 기록물을 설명하거나 생산자, 참여자의 행위주체(Q24229398)는 인물 혹은 단체(Q106559804)와 연계하도록 구성하였다. 기록물이 위치한 지리적장소(Q2221906), 기록물과 관련된 시간이나 시점은 시간(Q11471), 기록물이나 행위주체와 관련된 사건들은 사건(Q1190554) 객체와 연계하였다. 상-하위 클래스로 구성된 기록계층과 행위주체는 속성을 상속받을 수 있게 되어 기록물과 관련된 맥락정보인 행위주체, 시간, 장소, 사건에 대한 정보를 관계를 통해 브라우징 할 수 있도록 연계하였다. 그리고 핵심 객체들은 모두 하나의 개념(concept)이자 주제어(subject, keyword)로 관리하게 된다.



<그림 11> 국채보상운동 디지털 아카이브와 위키데이터 연계를 위한 기록물 객체 - 관계 모델

이동이 가능해지면서 다양한 관련 정보들을 탐색할 수 있게 된다.



4.3 위키데이터와의 연계를 위한 인물 모델링

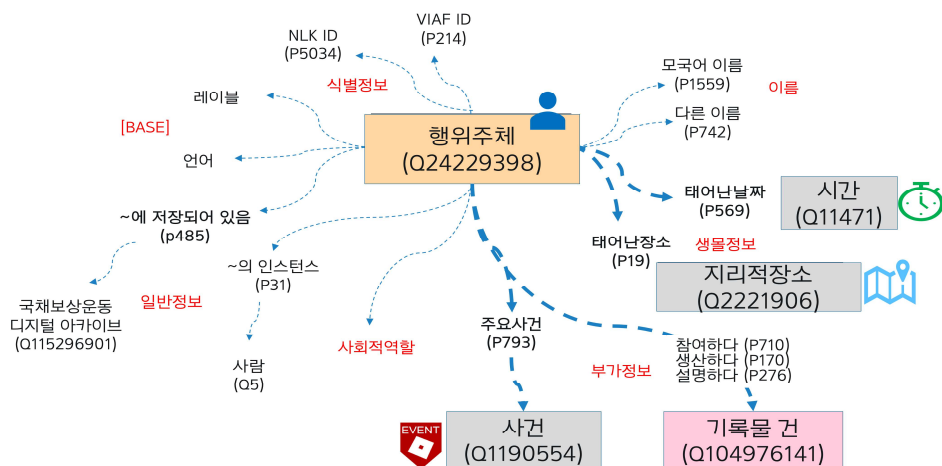
획득할 수 있게 된다.

관계 속성을 정리하였다.

〈표 6〉 인물(person) 객체와 위키데이터 연계를 위한 모델링 절차 요약

	절차	상세 연구 내용
1	분석 대상 인물 선정 - 25인	1) 디지털 아카이브 - 인물별 아카이브 인물 191명 대상 2) 국채보상운동연구총서 19권 - 인물편에 등장하는 인물 선정
2	인물 기술 정보 분석	1) 25인에 대한 인물별 아카이브 메타데이터 요소 분석 2) 25인의 직업군인 정치인, 사업가, 언론인에 해당하는 인물의 속성 분석 3) 비슷한 속성들을 범주화 ex) 이름: 모국어 이름, 이름, 성, 본명, 태명, 이명 등
3	주요 속성과 맥락 클래스와의 연결	기록물건, 사건, 시간, 장소와 관련된 관계 속성 추출

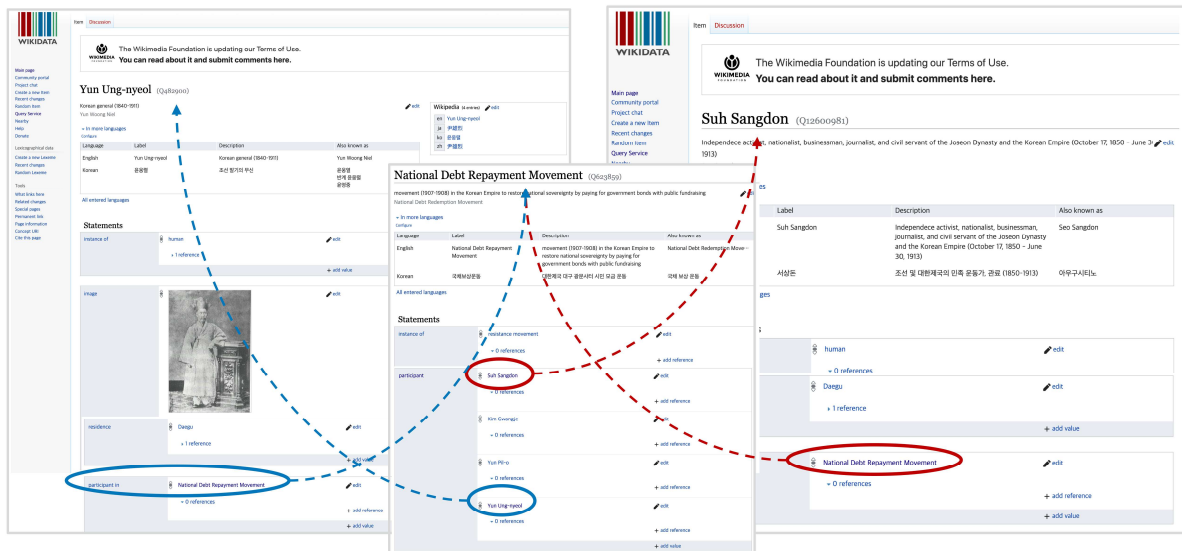
이러한 과정을 거쳐 국채보상운동 기록물의 행위주체(인물)와 맥락정보와의 객체 - 관계 모델을 다음의 <그림 13>과 같이 설계하였다. 행위주체(인물)를 설명하는 핵심 속성은 정치인이나 사업가 등의 사례에서 추출한 속성 약 100여개에서 필수로 입력해야 하는 속성들을 선정하였다. 인물을 설명하는 핵심 속성은 비슷한 속성들끼리 그룹화 하였는데, 그룹은 [BASE], 그리고 식별정보, 이름, 생몰정보, 부가정보, 사회적역할, 일반정보로 범주화하였다. [BASE]는 위키데이터의 최상단에 위치하여 개별항목의 레이블, 그리고 언어와 같은 기본적인 사항들을 기술한다. 식별정보는 개별항목과 관련된 식별자정보로 VIAF ID(P214), 국립중앙도서관 식별자(P5034) 등을 기술한다. 이름은 모국어이름(P1559)을 대표형으로 기술하는 경우가 많으며, 다른이름(P742)을 기술할 수 있다. 생몰정보는 태어난 날짜(P569)나 태어난장소(P19)의 속성을 기술할 수 있는데, 이들은 각각 시간(Q11471)과 지리적장소(Q2221906) 객체와 연결할 수 있다. 사건(Q1190554) 객체와는 주요사건(P793) 속성을 통해 연결할 수 있다. 이 외에도 인물의 상위객체, 그 상위의 객체와 연결하기 위해서 인스턴스(P31) 속성을 활용할 수 있다. 그리고 이 모든 정보들은 국채보상운동 디지털 아카이브(Q115296901)에 저장되어 있음(P485) 속성을 활용하여 연계할 수 있다.



〈그림 13〉 국채보상운동 디지털 아카이브의 인물(person) 객체와 위키데이터 연계를 위한 모델

아래의 <그림 14>는 인물 클래스와 맥락정보들을 연계한 위키데이터 페이지의 예시를 보여준다. 인물 “윤응렬 (Q482900)”의 주요사건(P793)은 “국채보상운동(Q623859)”으로, 실제로는 ‘참여하다(participant in)’ 속성으로 연결되어 있다. 사건 “국채보상운동” 페이지에 들어가 보면, ‘참여자(participant)’로 인물 “서상돈(Q12600981)”이 연결되어 있다. 참여자 중에는 인물 “김광제(Q16172930)”도 찾아볼 수 있는데, 앞서 기록물건과 연계한 사례에서 볼 수 있듯 국채보상운동 기록물을 포함한 맥락정보 객체 간의 다양한 브라우징이 가능하다. 이러한 관계를 통한 객체 간 브라우징은 기존의 디지털 아카이브에서는 제공할 수 없는 서비스이다. 국채보상운동 기록물과 위키데

이터를 연계함으로써 이용자들은 기록물과 맥락정보 간의 다양한 탐색이 가능하며, 연계된 객체와 관계 속성을 통해 링크를 타고 디지털 아카이브로 접근하여 관련 정보를 찾아봄으로써, 로컬 아카이브로의 유입 효과 또한 기대할 수 있을 것이다.



〈그림 14〉 위키데이터 연계 예시 - 인물 “윤응렬” - 사건 “국채보상운동” - 인물 “서상돈”

5. 결론 및 제언

본 연구는 국채보상운동 디지털 아카이브의 기록물 접근과 활용을 높이기 위해, 위키데이터와 연계하는 데이터 모델을 설계하는데 초점을 두었다. 위키데이터는 관리적 측면에서 엄격한 수준의 데이터 품질관리의 장벽을 낮추고 고품질의 기록 데이터를 관리, 공유, 재사용할 수 있다는 장점을 가지고 있다. 이용의 측면에서는 데이터의 관계성을 통해 연결된 객체를 탐색하는 과정에서 기존에 알지 못했던 새로운 지식의 발견을 가능하게 한다.

국채보상운동 디지털 아카이브는 여러 곳에 산재되어 있는 기록물을 수집하여 구축한 매뉴스크립트 컬렉션을 서비스하고 있다. 본 연구는 국채보상운동 디지털 아카이브의 기록물에 대한 접근성을 높이고, 데이터의 개방 및 활용을 위해 위키데이터와의 연계를 위한 모델 설계를 목적으로 한다. 이를 위해, 국채보상운동 디지털 아카이브의 관리 시스템과 서비스 현황을 파악하고, 기록물 기술과 검색에 사용되는 메타데이터, 시소러스, 네트워크 모델을 분석하여 위키데이터 연계를 위한 데이터 모델링에 반영할 사항들을 파악하였다.

국채보상운동 기록물 233건의 원문 내용 분석을 통해 기록물의 핵심 객체를 도출하였다. 국채보상운동 기록물은 1907~1910년(시간), 대한민국(물리적 장소)에서 일어난 국채보상운동(사건)을 기록한 것으로 기록물건(item) 단위로 관리되고 있다. 따라서 핵심 객체를 기록물건 - 행위주체 - 시간, 장소, 사건 객체로 구성하고, 객체의 주요 속성들을 관계로 맺어주어 객체 간의 자유로운 브라우징이 가능하도록 하였다. 인물은 기존의 인물 아카이브의 속성과 정치인, 사업가 등의 위키데이터 페이지 사례를 분석하여 주요 속성들을 도출하고, 그룹화하는 작업을 수행하였다. 그리고 앞서 기록물 모델링 과정에서 도출한 사건, 장소, 시간과 관련된 주요 속성을 도출하여 연계하였다. 설계한 데이터 모델에 실제 기록물 사례를 적용하여 국채보상운동의 행위주체인 인물, 그리고 사건, 시간, 장소 등의 객체들의 관계를 통해 이용자의 탐색이 확장됨을 확인하였다.

위키데이터는 거대한 지식베이스로 누구나 웹상에서 구축하고 활용할 수 있다는 점에서 소규모의 로컬 아카이브

에서 기록물의 활용성을 높이는 방법으로 적용해 볼 수 있는 플랫폼이다. 국채보상운동 디지털 아카이브의 경우에도 기존의 메타데이터만으로는 기록의 내용이나 구조, 맥락을 충분히 파악하기 어려웠으나, 위키데이터를 통해 네트워크와 연계함으로써 기록물의 정보를 더 풍부하게 하고, 나아가 국채보상운동 디지털 아카이브를 웹상에 노출시켜 아카이브로의 유입효과를 기대할 수 있을 것이다.

본 연구에서 설계한 국채보상운동 위키데이터 모델은 몇 가지 한계점을 갖는다. 우선 기록물의 계층 구조를 2개의 계층으로만 구성하였는데, 시리즈나 기록물군 등의 다양한 계층 구조에 활용할 수 있는 모델에 대한 연구가 필요하다. 또한 행위주체 모델링에서 인물의 속성은 기존의 국채보상운동 시소러스와 인물 총서의 개인 인물 정보에 기초했기 때문에, 추후 연구를 통해 관련된 단체(기관)와 속성 분석에 따른 보완이 이루어져야 할 것이다. 후속 연구에서는 다양한 기록물의 계층 구조와 핵심 객체들의 속성들을 면밀히 분석하여 위키데이터 모델에 반영한다면, 국채보상운동과 관련된 웹상의 다양한 데이터와 연결해 보다 풍부한 정보를 제공할 수 있을 것이다. 또한 기존의 사례연구에서 확인한 것처럼 데이터 기여의 모범 기관이 될 수 있도록 위키 커먼즈와 같은 위키 네트워크를 적극적으로 활용하고, 양방향의 직접 연계가 가능하도록 국채보상운동 디지털 아카이브의 인터페이스 개선도 함께 이루어져야 할 것이다.

참고문헌

- 국채보상운동기념사업회 (2019). 국채보상운동 기록물 등록 가이드v1.9. 대구: 국채보상운동기념사업회.
- 국채보상운동기념사업회 (2020). 유네스코 세계기록유산 국채보상운동 기록물 시소러스.
- 국채보상운동기념사업회 (2021a). 국채보상운동 인물사. 국채보상운동 연구총서, 19. 서울: 도서출판 선인.
- 국채보상운동기념사업회 (2021b). 2020년도 국채보상운동 기록물 디지털 아카이브 구축.
- 국채보상운동기념사업회 [발행년불명]. 디지털 아카이브. 출처: <http://www.gukchae-archive.org/gp/#intro>
- 국채보상운동기념사업회 [발행년불명]. 국채보상운동 위키데이터 페이지. 출처: <https://www.wikidata.org/wiki/Q623859>
- 김경남 (2022). 세계기록유산 국채보상운동기록물의 수집현황과 기술규칙 제안. 기록학연구, 71, 91-130.
<https://doi.org/10.20923/kjas.2022.71.091>
- 김수경, 안기홍 (2008). 시맨틱 웹 응용을 위한 웹 온톨로지 구축기법. 정보처리학회논문지D, 15(1), 47-60.
<https://doi.org/10.3745/KIPSTD.2008.15-D.1.47>
- 김지명 (2018). 기록문화유산의 디지털 큐레이션 모델 연구 - 국채보상운동 기록물을 중심으로. 디지털인문학연구총서, 4. 파주: 보고서.
- 김지옥 (2017). 유네스코로 간 국채보상운동 기록물. 기록인, 41, 69-77.
- 박승희 (2020). 국채보상운동 기록물과 인문 콘텐츠 연구. 대구경북연구, 19(3), 129-147.
<https://doi.org/10.23029/jdgs.2020.19.3.129>
- 박하람, 김학래 (2021). 일본군 '위안부' 지식그래프: 파편화된 디지털 기록의 연결. 한국기록관리학회지, 21(3), 61-78.
<https://doi.org/10.14404/JKSARM.2021.21.3.061>
- 위키데이터 [n.d.]. 위키데이터 소개 페이지. 출처: <https://www.wikidata.org/wiki/Wikidata:Introduction/ko>
- 한국기록관리학회 (2020). 기록관리의 이론과 실제. 서울: 조은글터.
- 한상은, 박희진 (2022). 디지털 아카이브의 위키데이터 활용방안 연구. 한국기록관리학회지, 22(1), 201-217.
<https://doi.org/10.14404/JKSARM.2022.22.1.201>
- COURAGE [n.d.]. connecting collections. Available: <http://cultural-opposition.eu/>
- DPLA [n.d.]. DPLA + Wikimedia project. Available: <https://pro.dp.la/projects/dpla-wikimedia>
- DPLA MAP Working Group (2017). Introduction to the DPLA Metadata Application Profile, version 5.0.
Available: <https://pro.dp.la/hubs/metadata-application-profile>
- Europeana (2017). Definition of the Europeana Data Model v5.2.8. Available:

- <https://pro.europeana.eu/page/edm-documentation>
- Evenhaugen, A. (2022, May 18). Smithsonian Libraries and Archives & Wikidata: Adding Artist Files to Wikidata. Available: <https://blog.library.si.edu/blog/2022/05/18/wikidata-artists-files/>
- Faraj, G. & Micsik, A. (2019). Enriching wikidata with cultural heritage data from the COURAGE Project. In Garoufallou, E., Fallucchi, F., William De Luca, E. eds. *Metadata and Semantic Research*, 407-418. https://doi.org/10.1007/978-3-030-36599-8_37
- Ferriter, M. (2019, May 22). Integrating Wikidata at the Library of Congress. Available: <https://blogs.loc.gov/thesignal/2019/05/integrating-wikidata-at-the-library-of-congress/>
- Freire, N., Isaac, A., Charles, V., Manguinhas, H. & Wyatt, L. (2019). Wikidata's linked data for cultural heritage digital resources: an evaluation based on the Europeana Data Model. *International Conference on Dublin Core and Metadata Applications*, 59-68. <https://doi.org/10.5281/zenodo.3384377>
- Heng, G., Xu, A., Deng, S., Zhu, L. & Li, X. (2022, June 25). Enhance the discovery and interoperability of culturally rich information: the Chinese Women Poets WikiProject. The Chinese American Librarians Association (CALA) Annual Conference 2022 poster session, Washington DC, USA.
- Koninklijke Bibliotheek [n.d.]. *Medieval Manuscripts in Dutch Collections*. Available: <https://www.mmdc.nl/static/site/>
- Library of Congress [n.d.]. *Library of Congress Name Authority File*. Available: <https://id.loc.gov/>
- Richard, K. (2022, March 30). Smithsonian Libraries and Archives & Wikidata: Chinese Ancestor Portrait Project. Available: <https://blog.library.si.edu/blog/2022/03/30/smithsonian-libraries-and-archives-wikidata-chinese-ancestor-portrait-project/>
- ZBW [n.d.]. *20th Century Press Archives*. Available: <https://pm20.zbw.eu/>

● 국문 참고자료의 영어 표기

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

- Han, Sangen & Park, Heejin (2022). A Study on Wikidata Utilization for Digital Archives. *Journal of Korean Society of Archives and Records Management*, 22(1), 201-217. <https://doi.org/10.14404/JKSARM.2022.22.1.201>
- Kim, Ji-myung (2018). *Developing a Digital Archive and Curation Model for Historical Document: Focusing on the National Debt Redemption Movement in Korea*. A Research Series of Digital Humanities, 4. Paju: bogosa.
- Kim, Jiwook (2017). Records of the National Debt Redemption Movement that went to UNESCO. *RecordIn*, 41, 69-77.
- Kim, Kyungnam (2022). Survey of the Archives of NDRM, Memory of the World and a Proposal of their Rules for Archival Description. *The Korean Journal of Archival Studies*, 71, 91-130. <https://doi.org/10.20923/kjas.2022.71.091>
- Kim, Sukyoung & Ahn, Kee-hong (2008). Web Ontology Building Methodology for Semantic Web Application. *The KIPS Transactions: Part D*, 15(1), 47-60. <https://doi.org/10.3745/KIPSTD.2008.15-D.1.47>
- Korean Society of Archives & Records Management (2020). *Records and Archives Management: Theory and Practice*. Seoul: Joeungeulter.
- Park, Haram & Kim, Haklae (2021). A Knowledge Graph on Japanese "Comfort Women": Interlinking

- Fragmented Digital Archival Resources. *Journal of Korean Society of Archives and Records Management*, 21(3), 61-78. <https://doi.org/10.14404/JKSARM.2021.21.3.061>
- Park, Seung hee (2020). A Study of the National Debt Redemption Movement and Digital Humanities Contents. *Journal of Daegu Gyeongbuk Studies*, 19(3), 129-147. <https://doi.org/10.23029/jdgs.2020.19.3.129>
- The Commemorate the National Debt Redemption Movement (2019). A registration guide for the National Debt Redemption Movement records, v1.9. Daegu: The Commemorate the National Debt Redemption Movement.
- The Commemorate the National Debt Redemption Movement (2020). A Thesaurus of National Debt Redemption Movement.
- The Commemorate the National Debt Redemption Movement (2021a). The Historical Figures for the National Debt Redemption Movement. A Research Series of the National Debt Redemption Movement, 19. Seoul: Suninbook.
- The Commemorate the National Debt Redemption Movement (2021b). A building a digital archive of record of the National Debt Redemption Movement 2020.
- The Commemorate the National Debt Redemption Movement [n.d.]. Digital archive of the National Debt Redemption Movement. Available: <http://www.gukchae-archive.org/gp/#intro>
- Wikidata [n.d.]. The National Debt Redemption Movement. Available: <https://www.wikidata.org/wiki/Q623859>
- Wikidata [n.d.]. Wikidata Introduction. Available: <https://www.wikidata.org/wiki/Wikidata:Introduction/ko>