

디지털 아카이브의 위키데이터 활용방안 연구

A Study on Wikidata Utilization for Digital Archives

한상은(Sangeun Han)¹, 박희진(Heejin Park)²

E-mail: silver86eun@gmail.com, papermod@hansung.ac.kr



¹ 제 1저자 성균관대학교 일반대학원 문헌정보학과 박사과정 수료

² 교신저자 한성대학교 디지털인문정보학트랙 부교수

논문접수 2022-01-25

최초심사 2022-02-03

게재확정 2022-02-11

ORCID

Sangeun Han
https://orcid.org/0000-0002-0759-5487

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.

*이 연구는 한성대학교 교내 학술연구비 지원과제임.

초 록

본 연구의 목적은 위키데이터를 활용하여 디지털 아카이브에서 소장하고 있는 기록 자원에 대한 접근을 높이고 구축한 데이터를 공유, 재사용할 수 있는 방안을 모색하는 데 있다. 이를 위해 먼저 기록학 분야에서 논의되고 있는 디지털 환경에서 기록 자원의 기술(description) 표준과 디지털 아카이브의 데이터 관리 및 공유와 관련된 연구들을 통해 국내외 디지털 아카이브의 기록 데이터의 현황을 파악하였다. 위키데이터의 구조와 특성을 파악하고 위키데이터를 적용한 해외 디지털 아카이브의 사례들을 조사하였다. 사례 분석을 통해 위키데이터 기반의 기록 데이터 개방, 공유와 재사용을 위한 시사점을 도출하고 위키데이터를 디지털 아카이브에 적용할 때 고려할 사항을 제안하였다. 본 연구의 결과는 디지털 아카이브 구축에 있어서 최근 데이터 개방, 공유 관련 기술의 동향을 이해하고 실제 아카이브 현장에서 위키데이터를 활용한 서비스 개발의 기초자료로 활용할 수 있을 것이다.

ABSTRACT

This study aims to use Wikidata to increase access to archival resources and see ways to share and reuse constructed data. To this end, the status of digital archives locally and abroad was identified using the description standard of archival resources in the digital environment discussed in the archival science field and the research related to data management and sharing of digital archives. The structure and characteristics of Wikidata were identified, and cases of digital archives applying Wikidata were investigated. The case analysis suggested implications for data sharing and reuse as well as the considerations for applying Wikidata to digital archives. The results of this study can be used as preliminary data for developing services using Wikidata in the actual archive field by understanding the recent trends in data open and sharing related technologies in constructing digital archives.

Keywords: 위키데이터, 디지털 아카이브, 기록 기술, 링크드데이터, 오픈 아카이브
Wikidata, digital archives, archival description, linked data, open archives

1. 서론

디지털 아카이브는 초창기 아날로그 형태의 소장 자료를 디지털화하여 디지털 컬렉션을 구축하고 온라인으로 서비스하는 정보시스템의 개념에서, 현재는 축적된 디지털 컬렉션과 자원을 데이터화하여 광범위한 공유와 활용을 통해 새로운 가치를 창출할 수 있는 디지털 플랫폼으로서 변화하고 있다. 디지털 아카이브의 서비스 또한 전문가뿐 아니라, 일반 대중을 대상으로 이용자들이 보다 쉽게 자원에 접근, 발견할 수 있도록 데이터를 개방하고 공유하는 방식을 모색하고 있다. 이에 디지털 아카이브에 온톨로지, 링크드 데이터 기술 등 시맨틱 웹 기술을 적용하여 어느 누구나 쉽게 데이터를 공유하고 활용할 수 있는 오픈 아카이브, 시맨틱 아카이브에 대한 연구가 꾸준히 진행되고 있다. 그러나 디지털 아카이브의 실무 현장에서는 시맨틱 웹 기술들에 대한 인식과 경험의 부족, 인력자원과 관련 시스템 구축에 대한 부담 등 여러 문제로 구현이 쉽지 않은 실정이다 (Hawkins, 2021).

많은 문화유산 기관에서는 아카이브 데이터의 구축과 운영 관점에서 제기되는 인프라 문제를 극복할 수 있는 방안으로 웹 상에서 데이터를 공유하고 협력할 수 있는 개방형 데이터 모델인 위키데이터(wikidata)를 제시하고 있다. 위키데이터는 위키미디어 재단에서 운영하고 있는 플랫폼으로 어느 누구나 협력하여 데이터를 구축할 수 있으며, 무료로 자유롭게 데이터를 재사용할 수 있는 라이선스를 부여한다. 스코틀랜드 국립 미술관, 현대 미술관, 미국 의회 도서관, WorldCat 등 국제적인 문화유산 기관들은 아카이브에 위키데이터를 활용하고 데이터 구축과 서비스에 대한 연구들을 진행하고 있다. OCLC는 2017년부터 2년 간 대학 연구도서관들과 협력하여 위키데이터 프로토타입을 구현하였고, 2019년 IFLA는 위키데이터 워킹그룹 연구를 통해 도서관 데이터를 확장시킬 수 있는 새로운 아카이브의 데이터 구축 방향을 제시하였다(Bianchini, Bargioni, & Pellizzari di San Girolamo, 2021).

최근 국내 기록관리학 분야에서도 위키데이터를 활용하여 디지털 아카이브의 지식그래프 구축하는 연구가 발표되었다(박하람, 김학래, 2021). 이 연구는 디지털 아카이브의 온톨로지 구축 시, 외부 데이터와의 연계와 수집한 데이터의 정제 방법으로 위키데이터를 활용하고 아카이브의 데이터 모델링과 일치하는 개체에 대해 위키데이터에 축적된 속성 정보를 추가함으로써 보다 풍부한 정보를 제공할 수 있음을 보여주었다. 그러나 아직 기록관 영역에서 위키데이터의 도입과 이를 활용한 서비스는 초기의 단계로 위키데이터에 관한 연구는 거의 찾아보기 어렵다. 디지털 아카이브의 기록 자료의 공유와 재사용을 위한 위키데이터의 구조와 특성 등을 살펴보고 그 활용 방안을 검토해볼 필요가 있다. 위키데이터를 통해 외부 자원과 연계하고 이를 활용한 서비스의 확장방안과 실제 위키데이터를 도입하고자 할 때 고려해야 할 사항이 무엇인지에 대한 이론적 고찰과 실무적인 검토가 필요한 시점이다.

본 연구의 목적은 위키데이터를 활용하여 디지털 아카이브에서 소장하고 있는 기록 자료에 대한 접근을 높이고 구축한 데이터를 공유, 재사용할 수 있는 방안을 모색하는 데 있다. 이를 위해 먼저 기록학 분야에서 논의되고 있는 디지털 환경에서 기록 자료의 기술(description)과 검색과 관련된 국제 표준과 디지털 아카이브의 데이터 관리 및 공유와 관련된 연구들을 통해 디지털 아카이브의 기록 데이터의 현황을 파악하였다. 기록 기술과 검색에 활용할 수 있는 위키데이터의 구조와 특성을 파악하고 위키데이터를 적용한 해외 디지털 아카이브의 사례들을 조사하였다. 사례 분석을 통해 위키데이터 기반의 기록 데이터 개방, 공유와 재사용을 위한 시사점을 도출하고 위키데이터를 디지털 아카이브에 적용할 때 고려할 사항을 제안하였다. 본 연구의 결과는 디지털 아카이브 구축에 있어서 최근 데이터 개방, 공유 관련 기술의 동향을 이해하고 실제 아카이브 현장에서 위키데이터를 활용한 서비스 개발의 기초자료로 활용할 수 있을 것이다.

2. 이론적 배경

2.1 디지털 환경에서 기록 기술(archival description)

기록 기술이란 기록관에 소장 중인 대량의 기록물을 효과적으로 관리하고 검색·활용할 수 있도록 도와주는 국제적으로 표준화된 기록물 관리방법이다. 기록물의 내용, 구조, 맥락에 대한 정보를 표준화된 형식(기술 규칙)에 따라 서술하여 기록물 검색도구(finding Aids)를 마련하는 일련의 과정을 의미한다(국가기록원, 2013). 기록 기술을 통해 기록관이 소장하고 있는 기록물에 대해 이용자가 원하는 기록물을 쉽게 찾고 활용할 수 있도록 한다. 디지털 환경에서는 기록이 유래된 기원이나 생산자의 관계뿐 아니라, 새로운 관점에서 기록을 접근할 수 있는 다차원적인 기록 기술 표준들이 필요하게 되었다. 이러한 요구에 등장한 것이 바로 국제기록협의회(International Council on Archives)에서 개발한 보존기록 기술표준인 RiC(Record in Context)이다. RiC은 기록의 출처 개념을 확장하고 생산, 관리, 배포되는 맥락의 복잡한 관계를 효과적으로 표현할 수 있도록 설계되었다. RiC은 개념 모델(RiC-CM, Records in Contexts-Conceptual Model)과 온톨로지(RiC-O, Records in Contexts-Ontologies)로 구성된다. RiC-CM은 기록 기술에 필요한 개체(entities)와 상호관계(interrelations)를 표현하는 개념 모델로 기존의 기록 표준 ISAD(G), ISDIAH, ISAAR(CPF), ISDF를 통합하고 있다. 국내에서도 2016년 RiC가 처음 소개된 이후, RiC 중심의 기록물 기술에 대한 연구들이 꾸준히 수행되어 왔다. 주로 초창기의 RiC 관련 연구들은 새로운 표준의 구조와 특성을 파악하고 국내 기록물에 RiC-CM을 적용할 수 있는지 이론적 검토가 주를 이루었다(박지영, 2016; 2017; 박선희, 2019; 신미라, 김익한, 2019).

위의 연구들에서 주목하는 RiC-CM의 특징은 기록의 기술 단위에서 기록과 기록 집합을 구분한다는 점이다. RiC-CM은 기록과 기록 집합이 많은 속성을 공유하지만 개별 기록의 기술과 기록 집합의 기술 방식이 다르다는 것을 강조한다. 이는 기존의 다계층 중심이었던 기록 기술에서 다차원 중심으로 기록 기술의 변화가 반영된 것이다. ISAD(G)는 기록의 상위 계층인 품부터 하위 계층의 아이템이나 컴포넌트까지의 계층 구조를 제시하고 있다. 그러나 RiC은 품을 기반으로 한 기존의 단일 다중계층 구조 뿐 아니라, 출처에 대한 폭넓은 이해를 위해 다른 품과의 관계 속에서 기록의 맥락을 파악하고자 하였다. 이를 통해 기록의 다른 개체와의 상호 관계를 네트워크로 표현하고 다각적인 관점에서 데이터를 표현할 수 있는 것이다.

RiC-O은 개념 모형인 RiC-CM을 기계가독형으로 표현하기 위한 기술 표준이다. 2019년 12월 RiC-O v0.1이 발표되면서, RiC-O을 활용하여 온톨로지를 모델링하여 국가기록원(정희명, 이성숙, 2021), 외환위기 아카이브(이유경, 김학래, 2020) 등 여러 기관과 아카이브의 실제 기록물에 적용해보고 그 가능성과 문제점을 점검하는 연구들이 진행되고 있다. 이들 연구는 사례를 통해 RiC-O가 기록을 둘러싼 맥락과 개체와의 관계에 초점을 두고 정확하게 상세하게 표현할 수 있는 점에 주목하는 한편, 실제 데이터모델링 시 구조와 어휘의 복잡성을 한계점으로 지적하였다. RiC-O를 기반으로 기록 정보 온톨로지를 구축하는 데 있어서 기록 분야 외 다른 어휘의 재사용을 고려하지 않고 있어 어휘 수준의 연계와 상호운용성에 제한이 따른다(이유경, 김학래, 2020). 또한 RiC-O 적용에 대한 가이드라인이나 예시가 없어서, 온톨로지 작성자에 따라 기록맥락 기술을 위해 사용하는 요소에 관한 판단과 구성요소의 계층에 대한 이해가 다를 수 있는 문제점도 제기되었다(정희명, 이성숙, 2021).

2.2 디지털 아카이브의 기록 데이터 구축과 공개

디지털 아카이브의 기록 자료를 데이터로 구축하여 활용하는 방안으로 가장 많이 논의되는 기술 중 하나는 링크드 오픈 데이터(LOD)라 할 수 있다. LOD는 웹에서 누구나 사용할 수 있도록 무료로 공개되는 링크드 데이터(Linked data)와 오픈 데이터(Open data)가 합쳐진 개념이다. LOD는 웹에 게시되는 데이터에 식별자(URI)를 부여

하고 관련 정보를 구조적으로 제공하는 연계 데이터로 저작권 없이 무료로 제공하여 사용자가 정보를 다양하고 효율적으로 활용할 수 있다.

LOD를 활용한 기록학 분야 연구들은 주로 링크드 데이터의 필요성을 살펴보고 이를 활용한 서비스 방안을 고찰하는 데 초점을 두고 있다. 2010년대 초반 링크드 데이터 클라우드를 위한 선행연구로 국가기록원 전거데이터 세트를 RDF/OWL로 모델링하는 과정을 거쳐 링크드 데이터를 시범 구축하는 연구가 수행되었다(박옥남, 2012). LOD를 활용한 한국사 콘텐츠 서비스 시스템 구축 연구(윤소영, 2013)와 대통령기록관의 기록정보서비스의 확장을 위한 연구(이정현 외, 2015) 등을 통해 이용자의 다양한 목적과 맥락에서 아카이브의 기록 데이터를 연계할 수 있는 방안들이 제시되었다. 그러나 다수의 LOD에 대한 연구들은 대부분 이론적 검토나 실험적 시도에 그치고 있으며, 2018년에 수행된 국내 기록관의 지능화 기술 현황 연구에서는 실제 기록관의 정보서비스에 LOD가 적용되는 사례가 거의 없는 것으로 조사되었다. 이는 투입되는 예산과 인력자원의 문제 뿐 아니라 LOD 발행을 위한 메타데이터나 통제어휘, 온톨로지 설계가 어렵기 때문이다. 기록관의 LOD 서비스 확산을 위해서 기록관에서 소장 기록물에 대한 링크드 데이터를 발행하고 공유할 수 있는 공개 플랫폼을 구축하고, 이를 활용한 온라인 아카이브 시범 서비스가 제안되었다(김태영 외, 2018).

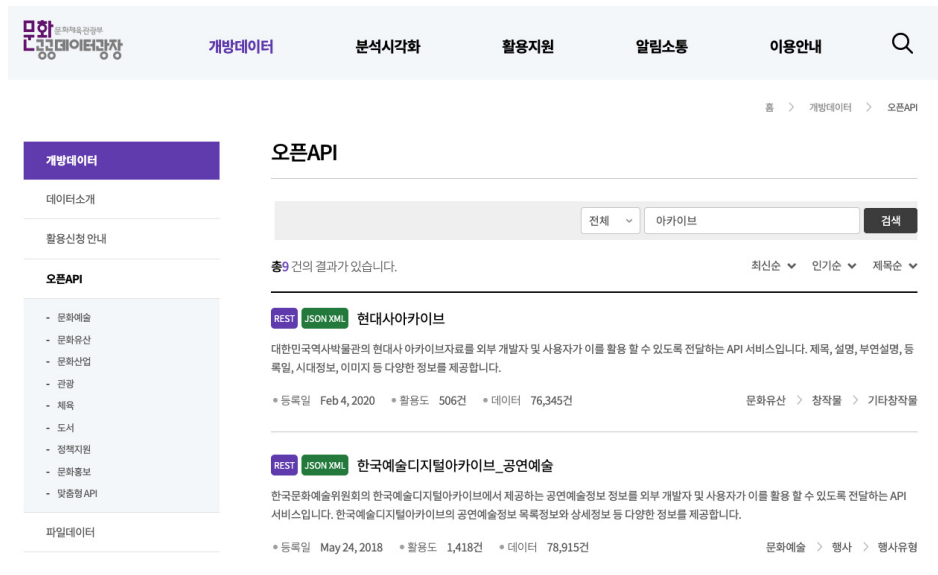
최근에는 디지털 아카이브에서 공공데이터로서 개방, 공유할 수 있는 자원 구축과 서비스 개발에 많은 관심을 보이고 있다. 역사학, 예술 아카이브 등 여러 영역에서 아카이브의 소장된 기록 자료의 데이터 셋을 링크드 데이터로 발행하고 다양한 시각화 도구 등을 통해 이용자들에게 제공하는 서비스 연구가 진행되고 있다. 박하람, 김학래(2021)는 일본군 위안부 디지털 기록을 연계하기 위한 링크드 데이터 모델을 제안하고, 이를 통해 관련 기록을 통합하여 지식그래프의 방식으로 서비스를 제안하였다. 국립극단에서는 2020년 국립극단 연극 디지털 아카이브의 링크드 데이터 서비스를 통해 공연/활동, 인물/단체, 작품, 장소, 공연자료, 학술/보도, 사건, 등장인물 등 공연의 다양한 영역의 정보가 유기적 관계에서 제공할 수 있는 데이터 모델을 제시하였다(정주영, 2021). 이 연구들은 디지털 아카이브의 기록 관련 데이터셋을 링크드 데이터로 제공함으로써 이용자들이 다양한 맥락을 탐색할 수 있으며 궁극적으로 아카이브 기록자료에 대한 다각적인 접근과 이해를 도모할 수 있음을 강조하고 있다.

2013년 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’이 제정되면서, 공공기관의 디지털 아카이브에서는 데이터 개방과 활용을 위한 서비스 확대에 주목하고 있다(이성숙, 2020; 정주영, 2021). 공공기관에서 생산하는 데이터를 누구에게나 공유하여 활용할 수 있도록 개방을 의무화하고 있으며, 공공데이터 포털(<https://data.go.kr>)에서 국가에서 생산한 다양한 영역의 데이터를 제공하고 있다. 국가기록원에서는 나라기록물 검색을 위해 OpenAPI를 홈페이지를 통해 제공하고 있으며 기록물에 대한 검색 결과를 RSS 형식으로 제공하고 질의 건수를 제한하고 있다. 대한민국 역사박물관의 근현대사 아카이브, 국립예술 자료원의 한국디지털아카이브, 국립국악원의 국악 아카이브, 한국학중앙연구원의 장서각디지털아카이브를 비롯하여 대부분 국내 디지털 아카이브 시스템을 운영하는 기관에서는 자체적으로 데이터를 제공하는 대신, <그림 1>과 같이 문화공공데이터광장(www.culture.go.kr) 또는 공공데이터포털(data.go.kr)과 같은 오픈데이터포털을 통해 Open API 뿐만 아니라 디지털 아카이브 데이터를 CSV, JSON_XML, XML, XLSX 등의 파일 형식으로 제공하고 있다.

이처럼 디지털 아카이브에서 데이터를 공개하고 서비스에서 재사용하는데 있어 가장 문제가 되는 것은 바로 데이터 품질에 있다. 데이터를 활용하는 이용자의 다양한 목적을 충족시켜줄 수 있는 수준으로 데이터가 구축되고 제공되어야 한다. 개방된 웹 환경에서 다양한 데이터와 연계와 통합을 염두에 두고 디지털 아카이브에서 제공하고 있는 데이터 요소와 형식, 규칙에 대한 원칙을 수립하고 지속적인 데이터의 품질 관리와 평가가 필요할 것이다. 그러나 디지털 아카이브의 데이터에 관한 연구는 주로 데이터 개방과 공유 방식을 검토하는 것들로 데이터 구축과 운영 단계에서 데이터의 품질 관리와 평가에 관한 연구들은 많지 않다.

디지털 아카이브의 데이터 공유와 활용을 위해서 김학래(2021)는 FAIR 데이터 원칙을 기반으로 한 데이터 평가 프레임워크를 제안하였다. FAIR 데이터 원칙은 데이터를 찾을 수 있으며(Findable), 접근 가능하고(Accessible), 상호운용성이 확보되며(Interoperable), 재사용할 수 있는(Reusable) 위한 지침을 말한다. 이를 디지털 아카이브

데이터에 적용하여 식별자, 내용중심의 메타데이터(F), 데이터 접근 여부를 판단할 수 있는 메타데이터(A), 데이터 형식, 어휘(I), 라이선스(R)을 평가 기준으로 하고 각각에 대한 세부 지표를 제시하였다. 또한 FAIR 기반 평가지표를 기록의 생애주기 관점에서 포괄적으로 적용하여 메타데이터 표준, 데이터 형식, 데이터의 상호운용성을 위한 프레임워크로 데이터를 구축, 운용할 것을 제안하였다.



<그림 1> 문화체육관광부 문화공공데이터광장 디지털아카이브 데이터 제공 사례

2.3 위키데이터와 기록 기술 프로젝트

2006년 위키미디어 프로젝트를 처음 발표한 이후 위키피디아는 280여 개 언어로 위키데이터를 제공하고 있으며, 2012년 위키미디어 독일 지부(Wikimedia Deutschland)에 의해 위키피디아의 데이터를 기계 가독 데이터로 정리하고 누구나 활용할 수 있는 위키데이터 프로젝트가 본격적으로 시작되었다. Wikidata.org에 사용자가 직접 참여하여 데이터를 생성하고 편집할 수 있는 인터페이스를 제공하고 불특정 다수가 참여하여 사물, 인물, 개념 항목에 대한 거대한 데이터베이스를 구축하였다. 위키데이터는 다국어 지식베이스로 위키백과 등 위키미디어 프로젝트에서 공통으로 사용할 정보를 제공하는 것을 목적으로 한다. 위키데이터 인프라는 무료 소프트웨어를 사용하여 생성되며 모든 콘텐츠는 다른 이용자들이 데이터를 자유롭게 재사용할 수 있도록 라이선스가 부여된다. 즉, 위키데이터는 다국어 기반으로 누구나 편집할 수 있는 자유로운 지식저장소로, 전 세계에 공개되는 지식 그래프를 형성하며 누구나 모든 지식에 접근하고 공유할 수 있도록 지원하고, 외부식별자 속성을 통해 다른 오픈 데이터베이스와 웹 자원을 연결하는 메커니즘을 제공하고 있다.

위키데이터 저장소는 주로 문서가 설명하고자 하는 대상인 아이템(item, 항목)으로 구성된다. 아이템은 접두사 Q 다음에 숫자로 구성된 고유한 ID를 통해 식별된다. 위키데이터는 사람과 기계가 모두 읽고 편집할 수 있도록 구조화된 RDF 형식의 데이터로만 입력할 수 있으며, 각 아이템에는 레이블, 설명, 별칭 등을 가진다. 서술(statements)은 아이템의 세부적인 특성을 설명하며 속성과 값으로 구성된다. 속성 또한 아이템과 같은 형식으로 접두사 P 다음에 숫자로 구성된 고유한 ID를 가진다. 위키데이터는 기본적으로 기계가독형 데이터로 구성되지만 실제 이용자를 위해서 레이블을 중심으로 화면에 표시되게 된다. 예를 들어, 조선왕조실록(Veritable Records of the Joseon Dynasty, Q497030)인 경우 <그림 2>와 같은 페이지와 정보를 가지게 된다. 조선왕조실록은 속성 “instance

of (P31)”을 통해 실록(Veritable Records, Q852215)의 값을 가지는 것을 확인할 수 있다. 속성을 통해 해당 아이템에 대해 설명을 제공하고 이 속성들의 가운데 여러 값을 가지거나 각주 등 여러 정보를 가지는 속성을 서술목록이라고 한다.

레이블 — Veritable Records of the Joseon Dynasty (Q497030) — **고유식별자**

설명 — annals — **다른이름**

Annals of the Joseon Dynasty

→ In more languages

Language	Label	Description	Also known as
English	Veritable Records of the Joseon Dynasty	annals	Annals of the Joseon Dynasty
Korean	조선왕조실록	조선 왕조의 역사를 기록한 책	조선실록 이조실록 리조실록
Japanese	朝鮮王朝実録	No description defined	
Chinese	朝鮮王朝實錄	No description defined	朝鮮實錄

All entered languages

서술 속성 — Statements

Instance of: Veritable Records — **값**

country: South Korea, North Korea

서술목록

<그림 2> 위키데이터 예- 조선왕조실록
(<https://www.wikidata.org/wiki/Q497030>)

위키데이터의 데이터를 활용하기 위해 이용자가 접근할 수 있는 방법은 두 가지로 덤프데이터 다운로드와 위키데이터 쿼리 서비스(Wikidata Query Service)가 있다. 덤프데이터 다운로드의 경우 위키데이터 덤프사이트(<https://dumps.wikimedia.org/wikidatawiki/entities/>)에서 주기적으로 업데이트 되는 최신 버전으로 다운로드 가능하다. 위키데이터 쿼리 서비스(<https://query.wikidata.org/>)는 위키데이터의 모든 데이터를 SPARQL을 통해 조회하고 그 결과를 JSON, TSV, HTML 등의 파일 형식으로 다운로드 받을 수 있도록 도와준다. 더불어 조회 결과에 따라 Table, Image Grid, Map, Tree, Timeline과 같은 다양한 형식으로 시각화할 수 있는 기본 기능을 제공하고 있어 위키데이터를 활용하여 손쉽게 데이터 시각화를 구현할 수 있다.

또한 위키데이터를 통해 데이터를 간편하게 반입, 반출할 수 있으며 편집과 수정이 자유롭다. 위키데이터는 개별 아이템을 따로 생성할 수 있으며 기관 아카이브로부터 데이터 전체를 이관할 수도 있다. 확장 가능한 플러그인 플레이 솔루션으로 별도의 사용자 조작이나 프로그램 설치 없이 바로 사용할 수 있는 장점을 가지고 있다. 다른 Wikibase 소프트웨어와 같이 위키데이터는 아카이브의 자원들을 검색하고 접근하기 위한 새로운 기술 플랫폼을 처음부터 구축할 필요가 없으며 특정 도메인 분야의 온톨로지를 도입할 필요도 없다. 위키데이터는 직관적이고 사용하기 쉬운 편집 인터페이스를 통해 자동으로 식별자 생성하여 데이터를 게시할 수 있으며, 바로 사용 가능한 질의 기능을 제공하고 있다.

2018년부터 벨기에, 프랑스, 스위스의 기록관을 중심으로 기록 풍과 컬렉션의 거대한 데이터베이스를 구축하고 위키데이터를 활용하여 검색, 활용할 수 있는 Archival Description(위키데이터 기록 기술) 프로젝트가 시작되었다(wikidata, n.d.). 위키데이터 기록 기술 프로젝트는 기록물의 위키데이터 활성화를 위해 아카이브 및 문화유산 컬렉션에 대해서 가장 포괄적이고 고품질의 데이터베이스를 구축하는 것을 목표로 한다. 기록 기술 프로젝트에서는 다음의 <표 1>과 같은 활동을 진행하고 있다.

〈표 1〉 기술 기록(archival description) 활동

기록 기술(archival description) 활동
기록관 및 문화유산 컬렉션에 대한 정보를 포함하는 모든 공공 데이터베이스 목록작성
기록관 및 문화유산에 대한 설명과 관련된 온톨로지 및 다국어 사전 구현 및 유지관리
기록물 검색도구와 위키데이터간 상호 연결, 기록컬렉션 목록과 위키데이터간 상호 연결
기록관 및 문화유산 컬렉션에 대한 데이터를 위키데이터로 수집 및 데이터 향상
Wikipedia에서 유지관리 목록을 위한 데이터 사용
수집한 데이터를 Wikipedia 및 관련 프로젝트의 정보상자(info boxes)와 리스트에 적용
수집한 데이터를 활용한 애플리케이션 구축

현재는 캐나다, 미국을 비롯하여 독일, 네덜란드 등 여러 나라의 기록관에서 기록 데이터를 위키데이터 플랫폼에서 공유하고 서비스하고 있으며, 위키데이터를 기반으로 새로운 서비스를 시도하고 있다. 위키데이터 기록 기술 프로젝트는 참여하고 있는 기록관의 협력 뿐 아니라 다른 도메인과의 협력을 통한 위키데이터 확산을 꾀하고 있다.

3. 위키데이터 활용 사례 분석

본 연구는 위키데이터 기록 기술 프로젝트에서 활발하게 활동하고 있는 대표적인 디지털 아카이브를 중심으로 위키데이터 활용 목적, 데이터 연계 방법, 데이터 모델, 서비스에 대하여 조사하였다. 조사 대상으로는 캐나다의 Discover Archives, Canadian Archives of Women in STEM, 독일의 20th Century Press Archives를 선정하였다.

3.1 간접 연계: 캐나다의 Discover Archives

캐나다 Discover Archives는 토론토 대학교(University of Toronto)를 포함하여 다른 대학도서관들이 소장하고 있는 보존 기록들을 중심으로 토론토 대학의 역사 기록물과 관련 정보, 인물, 가족, 기업 및 조직에 대한 정보를 제공하는 검색시스템이다. 토론토 대학도서관에서는 2017년부터 꾸준히 대량의 도서관 데이터를 공유하고 활용하기 위해 Wikipedia에 관심을 갖고 연구를 지속해 왔다. 2020년 본격적으로 온라인 상에서 이용자가 더 쉽게 Discover Archives에 접근하고 기록자원을 검색하는 것을 목표로 위키데이터 프로젝트에 참여하게 되었다.

Discover Archives는 위키데이터 플랫폼에 “archives at” 속성을 추가하여 참여 기관의 설명과 식별 정보를 입력함으로써 아카이브의 접근점을 확장하고자 하였다. “archives at” 속성으로 입력되는 위키데이터는 연계된 아카이브의 기록레코드로 이용자를 안내할 뿐 아니라, 다른 위키미디어 프로젝트인 wikipedia등의 엔티티에 해당하는 문서에 연결 되는 사이트 링크를 제공함으로써, 이용자가 기록물을 이해하는데 필요한 다양한 정보를 발견할 수 있다. 이와 같은 위키데이터와 기록 데이터의 간접적인 연계를 통해, Discover Archives는 해당 기관 사이트에 대한 접근을 향상시키고 궁극적으로 참여 기관들의 소장 자료의 이용률을 높이고자 하였다.

이를 위해 Discover Archives는 위키데이터에 “archives at” 속성을 추가하고 아카이브의 참여 기관들과 기록에 대한 항목을 우선 생성하였다. “archives at” 속성은 OpenRefine을 사용하여 해당 식별자와 간단한 정보를 입력하는 것으로, 이를 통해 위키데이터 내에 이미 존재 하는 아이템과 연계가 가능해진다. 새롭게 추가하는 아이템에 대해서는 기관의 전거레코드를 입력하는 대신 간단한 데이터 모델을 만들어 위키데이터에 추가한 후 “archives

at” 속성 정보를 입력하였다(Wong & Babcock, 2021). <그림 3>은 Discover Archives의 “archives at” 속성 정보 추가로 위키데이터가 Urshlar Fraklin의 위키피디아 문서와 연결되고 WikiMedia를 통해 데이터를 공유하고 있는 다른 프로젝트로도 연결되는 것을 보여주고 있다. 위키데이터를 통해 아카이브의 접근점을 확장하고 관련된 연계 정보를 충분히 제공함으로써 이용자로 하여금 특정 기록 검색에만 머물지 않고 또 다른 유용한 정보를 발견할 수 있도록 하는 것이다.



<그림 3> Discover Archives, 위키데이터, Wikipedia-Ursula Franklin 예

Discover Archives는 관련 기관들과 협업을 위해 대학 도서관 사서와 아카이비스트에게 위키데이터 교육과 훈련을 제공하고, 캐나다 아카이비스트들과 위키데이터를 이용하는 다양한 커뮤니티들과의 토론에도 활발히 참여하고 있다. 이들은 아울러 아카이브의 협력목록을 함께 작성하는 PCC(The Program for Cooperative Cataloging) 위키데이터 Pilot 프로젝트도 진행하고 있다.

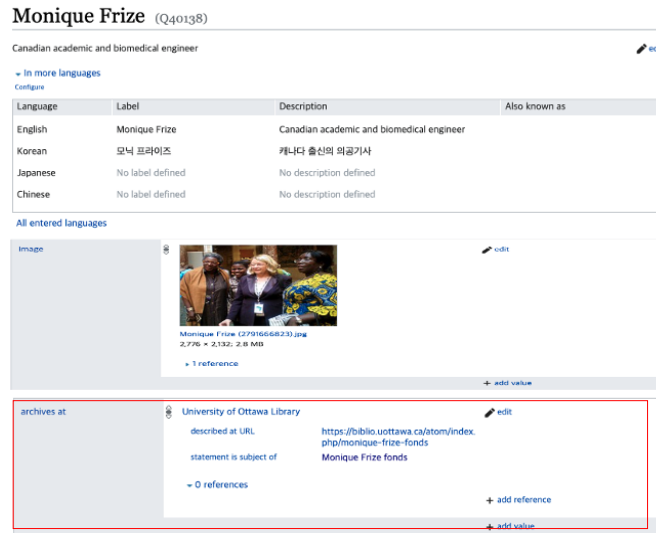
3.2 직접 연계: 캐나다의 Canadian Archives of Women in STEM

Canadian Archives of Women in STEM(Science, Technology, Engineer, and Mathematics)은 캐나다 오타와 대학 도서관을 중심으로, 캐나다 도서관 및 기록관(LAC, Library and Archives), 국제 여성 엔지니어 및 과학자 네트워크-교육 및 연구기관들이 협력해 구축한 지식 프로젝트이다. 이 프로젝트는 STEM 기록물에 대해서 이용자들이 원하는 기록을 찾고 그 의미를 이해할 수 있도록 기록 풍(fonds)을 식별하고 발견할 수 있는 검색포털 개발을 목표로 시작되었다.

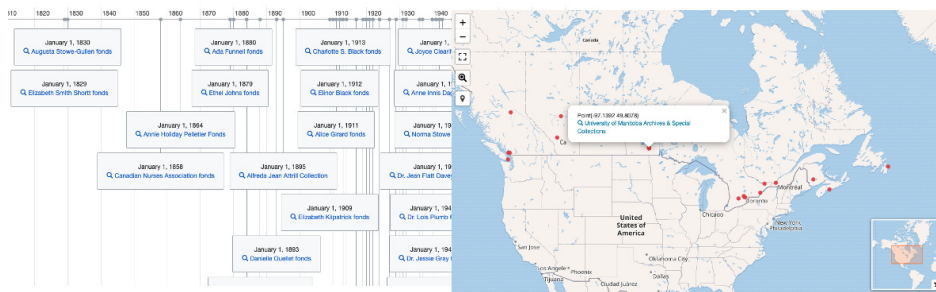
이를 위해, 아카이브는 STEM의 아카이브 기록 기술 필드와 위키데이터 속성간의 매핑을 통해 위키데이터의 다양한 응용서비스와 연동하고 확장시켜, STEM의 캐나다 여성들에 대한 관심을 높이고 시스템의 검색성을 향상시키고자 하였다(Lee, Mairelys, & Marnia, 2019; 2020). STEM 포털에서 이용자는 캐나다 전역의 여성과 기관에 대한 정보를 검색할 수 있지만 API와 같이 아카이브 데이터를 활용할 수 있는 서비스를 제공하지는 않는다. 이러한 한계를 극복하기 위해 아카이브는 STEM 데이터를 위키데이터에 매핑시킴으로써 위키데이터 플랫폼에서 제공하는 다양한 응용프로그램과 상호작용하며 데이터를 활용한 새로운 서비스를 제공할 수 있게 되었다.

<그림 4>와 <그림 5>는 STEM의 아카이브 전거데이터와 매핑 되는 다른 위키데이터를 활용한 시계열과 지도

기반의 데이터 시각화 서비스를 보여준다. 이처럼 디지털 아카이브는 위키데이터를 통해 기록 자원과 관련된 다양한 서비스와 연결하고 해당 서비스에 의해 데이터 추출 및 필터링을 통한 도표, 그래프 등 시각화 기능을 제공할 수 있다. 위키데이터를 활용하여 디지털 아카이브는 서비스를 향상시키고 이용자들의 기록 자원에 대한 이해도와 접근성을 높일 수 있는 것이다.



〈그림 4〉 STEM-위키데이터 전거레코드의 예 (<https://www.위키데이터.org/wiki/Q40138>)



〈그림 5〉 위키데이터 서비스를 이용한 시계열 및 지도 기반 시각화 예

3.3 통합과 기여: 독일의 20th Century Press Archives

20세기 언론 아카이브(20th Century Press Archives, 이하 PM20)는 독일을 포함한 전 세계에서 발행된 1500개 이상의 신문을 수집한 대규모의 공공 신문 스캔 아카이브이다. 이 아카이브는 1908년~2005년까지의 신문을 대상으로 인물(Persons), 일반 주제와 사건(general subjects and events), 회사(companies), 제품(products)과 관련된 주요 주제를 선정하여, 각 주제별 정보를 제공하고 있다. 현재 PM20는 독일 국립 경제 도서관(Deutsche Zentralbibliothek für Wirtschaftswissenschaften, 이하 ZBW)에서 독일연구 재단(German Research Foundation, DFG)의 지원을 받아 1960년까지의 자료를 디지털화하고, 1949년까지의 2백만 페이지가 넘는 25,000개의 주제 정보를 온라인에서 무료로 제공하고 있다.

ZBW Leibniz-Informationszentrum
Wirtschaft Leibniz Information Centre
for Economics Press Archives

en | de

Historical press archives of ZBW. Inherited from the "Hamburgisches Welt-Wirtschafts-Archiv" and the "Wirtschaftsarchiv" of the Kiel Institute for the World Economy. Partially digitized with funding from the German Research Foundation (DFG).

- > Persons archive
- > Companies/Institutions archive
- > Countries-Subjects archive - by country category system
- > Countries-Subjects archive - by subject category system
- > Wares archive
- > Digitized films (only metadata on the web - full content from ZBW reading room)
- > About ZBW Press Archives

Feedback Legal Imprint Data protection All metadata under CC0 license

<그림 6> 20th Century Press Archives 홈페이지 메인 화면
(https://pm20.zbw.eu)

ZBW는 많은 디지털 아카이브들이 겪고 있는 기록 기술을 위한 막대한 비용과 시간, 인력 소모의 문제 등을 해결하기 위한 방법으로, 위키데이터를 LOD로 적용하여 기록 자원의 검색과 접근을 향상시키고자 하였다. 2019년부터 2021년까지 ZBW는 PM20 메타데이터에 CCO 라이선스를 부여하여 어느 누구나 사용할 수 있게 하고 PM20 데이터를 위키데이터에 통합(데이터 기여)하는 작업을 진행하였다. ZBW는 위키데이터에 단순히 아카이브의 기록 데이터를 연계하는 것이 아니라, Wikimedia 프로젝트, 오픈사이언스를 포함하여 일반 대중들이 PM20 기록 자원을 위키데이터를 통해 통합 검색할 수 있는 것을 목표로 하였다. 이를 위해 ZBW에서 관리하고 있는 개인, 회사 기록 자원의 메타데이터를 위키데이터에 추가하고 이용자들이 쉽게 사진 기록을 접근, 재사용할 수 있도록 IIIF(International Image Interoperability Framework) 표준에 따라 디지털 이미지로 저장, 공개하였다. 또한 ZBW는 더 많은 사람들이 PM20의 기록자원을 접근하고 이용할 수 있도록 소장 기록물에 대해 통합된 메타데이터를 위키데이터에 공유하고 정적 랜딩 페이지(Static Landing Page)로 제공하였다.

위키데이터로 데이터를 통합하기 위해 ZBW는 기록 자원에 대한 위키데이터의 Q-id 식별체계를 구축하고 위키데이터 내에 이미 생성되어 있는 아이템들과의 매핑을 위해 Mix-nMatch와 OpenRefine을 활용하였다. 이를 통해 60명 이상의 위키데이터 이용자가 참여하여 위키데이터의 인물 항목과 PM20 인물 항목을 매칭시키는 데 기여할 수 있었다. 이용자들은 <그림 7>과 같이 PM20 홈페이지를 통해 아카이브의 기록 자원들을 찾아 볼 수 있으며, 위키데이터와 통합된 자원의 경우 검색 결과 상세보기에서 해당 위키데이터 페이지로 이동하여 더 많은 정보를 획득할 수 있게 되었다. 이용자가 위키데이터로 정보를 검색 시, PM20 아카이브의 정보가 노출됨으로써 이용자가 아카이브에 접속할 수 있는 접근점을 제공하고 아카이브의 이용률을 높일 수 있는 것이다.

The screenshot displays two interconnected data points. On the left, the ZBW PM20 interface for 'Gandhi, Mohandas Karamchand' (1869-1948) shows 443/669 documents and a 'PM20 folder ID' of 'pe/005823'. On the right, the Wikidata entry for 'Mahatma Gandhi' (Q1001) lists various aliases and a table of works accessible online.

Language	Label	Description	Also known as
English	Mahatma Gandhi	pre-eminent leader of Indian nationalism during the British Raj	Mahatma Mohandas Karamchand ... M. K. Gandhi Mohandas K. Gandhi M. K. Gandhi Mohandas Gandhi Gandhi Mohandas K. Gandhi

PM20 folder ID	pe/005823
number of works	669
number of works accessible online	377
named as	Gandhi, Mohandas Karamchand
→ 0 references	

<그림 7> PM20와 위키데이터-Gandhi 예

<표 2>는 디지털 아카이브의 위키데이터 활용 사례들을 분석한 내용을 종합하여 정리한 것이다. 사례 분석을 통해 디지털 아카이브의 위키데이터 활용 목적은 무엇보다 기관에서 소장하고 있는 기록물 데이터의 접근 및 검색을 통한 이용 활성화에 중점을 두고 있음을 알 수 있었다. Discover Archives는 기관의 위키피디아 페이지에 ‘archives at’ 속성을 추가하여 아카이브 내부의 기록 목록과 검색 도구로 연결함으로써 접근점을 확장하여 검색서비스를 향상시켰다. Canadian Archive of Women in STEM은 아카이브의 기록 데이터와 위키데이터를 매핑함으로써 위키데이터 플랫폼에서 제공하는 다양한 응용프로그램과 상호작용하며 데이터를 활용한 새로운 서비스를 제공할 수 있었다. 더 나아가 ZBW의 PM20 아카이브는 자관의 데이터를 위키데이터와 통합하여 이용자가 이를 활용하여 가치있는 정보를 재생산할 수 있는 서비스까지 확장할 수 있음을 보여주었다.

<표 2> 디지털 아카이브의 위키데이터 사례 분석

구분	Discover Archives	Canadian Archive of Women in STEM	20th Century Press Archives
목적	아카이브 접근 및 검색	아카이브 데이터 이용 및 활용 활성화	기록자료의 검색과 접근 향상
연계방법	“archives at”을 통한 간접연계	직접연계	통합과 기여
데이터 모델	사람, 기관을 중심으로 한 간단한 데이터 모델	전거데이터, 기록물 품을 중심으로 한 데이터 모델	사람, 기관, 주제 등을 중심으로 한 데이터 모델
이용 도구	OpenRefine	확인 불가	OpenRefine, Mix-nMatch
Wikiproject 페이지	제공	미제공	제공
융합서비스	Wikidata Query 서비스 - 단순 쿼리 예시 제공	Wikidata Query 서비스 Timeline, Map 시각화 예시 제공	Wikidata Query 서비스 - 다양한 활용사례 제공

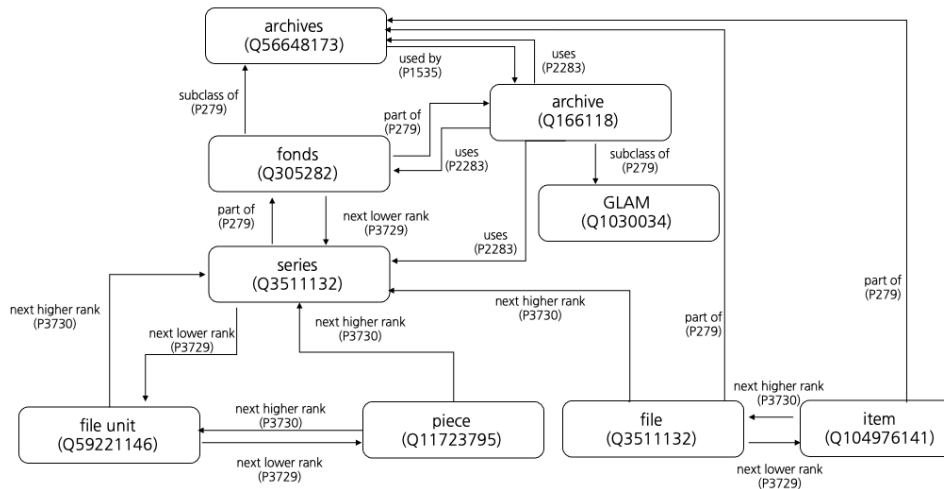
4. 디지털 아카이브의 위키데이터 활용방안

4.1 위키데이터 기록 기술 모델

본 연구는 디지털 아카이브에서 위키데이터의 어휘(vocabulary)를 기록 기술에 적용되고 있는 계층적 관리의 원칙(다계층 기술의 원칙)에 적용시킬 수 있도록 다음 <그림 8>과 같이 위키데이터 기록 기술 모델을 제시하였다. 이는 기록 기술의 ‘기록물풍(fonds)’ - ‘기록물시리즈(series)’ - ‘기록물철(file)’ - ‘기록물건(item)’ 다계층적 구조를 중심으로 위키데이터에 나타나는 어휘들 즉, 아이템의 속성을 확인하고 아이터간의 관계를 분석하여 계층과 관계로 표현할 수 있도록 구조화한 것이다.

그동안의 기록 기술은 다계층으로 분류하고 기록물을 정리하며, 계층별로 조직된 기록물의 내용, 구조, 맥락정보를 확인하고 설명하는 것으로, 기록 기술의 신뢰도와 일관성을 확보하여 기록관 간의 정보 교환을 가능토록 하였다. 위키데이터는 아카이브에서 사용하고 있는 이러한 기록 기술 계층을 각각 객체와 속성, 관계로 정의함으로써, 기록 데이터와 전체적인 매핑을 통해 데이터를 반입, 반출할 수 있다.

본 연구에서 제시한 위키데이터의 기록 기술 모델을 통해 디지털 아카이브는 추가적인 온톨로지 설계 없이 기록 데이터를 위키데이터로 매핑하거나 새롭게 생성할 수 있을 것이다. 또한 제시된 위키데이터 기록 기술 계층 구조를 활용해 기록관의 다른 맥락에서 생산되거나 수집된 데이터를 기존 아카이브의 데이터와 매핑함으로써, 관련 아카이브 데이터와 유기적으로 연계할 수 있을 것이다.



<그림 8> 위키데이터 기록 기술(Archival Description) 모델

4.2 기대효과

우리가 살아가는 정보세계는 매우 거대하고 복잡하며, 많은 것을 포함하고 있다. 대부분의 문화유산기관은 메타 데이터를 통해 소장 자료를 기술하지만 이러한 메타데이터나 목록정보는 해당 기관의 웹 사이트 내에서만 검색이 가능하다. 이러한 상황에서 디지털 아카이브 자료가 고립되지 않고 웹에서 더 많은 사람들이 정보를 찾고 효과적으로 정보를 이용, 재사용하기 위한 방안으로 위키데이터를 고려해 볼 수 있다. 디지털 아카이브의 위키데이터 기반 기록 데이터의 연계 및 통합 사례 분석을 바탕으로 위키데이터의 기대효과를 정리하면 다음과 같다.

무엇보다 위키데이터는 기관에 투입되는 예산과 인력자원의 부족으로 링크드 데이터를 구현하기 어려운 디지털 아카이브에서 링크드 데이터의 허브로 활용될 수 있다. 위키데이터는 구조화된 데이터를 기계가독형으로 생성하여 다른 저장소로 쉽게 연결할 수 있을 뿐 아니라, 위키데이터를 활용하여 전문가가 아니더라도 도서관 사서, 박물관 아키비스트들이 쉽게 데이터 공유를 위한 링크드 데이터를 발행할 수 있다. 링크드 데이터가 구조화된 데이터 발행을 통해 관련 데이터들이 서로 연결되어 풍부한 정보를 제공하는 것처럼, 위키데이터는 1300개 이상의 시스템으로부터 다른 식별자를 매칭하여 데이터들을 연결시켜줌으로써 보다 유용한 데이터 활용 환경을 구축할 수 있다.

디지털 아카이브에서는 위키데이터를 아카이브의 기록 자원의 전거레코드와 연계하여 검색서비스에 활용할 수 있다. 전거레코드는 “기록물 목록에서 접근점으로 선정된 단체·인물·사건 등을 표준화된 형식으로 상세기술하고 다른 전거레코드와 연결하여 상호 참조할 수 있도록 관련 정보를 조직화한 레코드”를 말한다(국가기록원, 2011, 4). 아카이브의 전거레코드는 기록물의 주요 접근점으로 기록 생산자와 같이 변경이 있을 시 효율적 검색을 위해 신속하고 주기적인 갱신이 필수적이다. 따라서 전거레코드는 보통 별도의 데이터베이스로 구축해 연계시키는 구조로 개발, 유지되는데, 이를 위키데이터를 활용하여 연계한다면 기존의 전거레코드를 확장하고 연계 레코드를 통해 갱신된 정보를 자동적으로 획득할 수 있다. 위키데이터는 미국 의회, NARA, SNAC(Social Networks and Archival Context)의 전거데이터와 연결되어 이용자는 방대한 위키피디아를 통해 다양한 관련 정보를 접근하고 탐색할 수 있을 것이다.

더불어 위키데이터는 데이터를 탐색하고 새로운 서비스 도구에서 재사용할 수 있다는 점을 감안할 때 디지털 아카이브의 데이터를 많이 노출 시킬 수 있다는 장점이 있다(Lemus-Rojas, Mairelys, & Yoo, 2019). 이를 통해 아카이브의 접근과 활용 범위를 넓힐 수 있으며, 컬렉션 간의 연계를 통해 기록정보서비스를 향상시킬 수 있다.

위키피디아를 비롯한 거대한 위키프로젝트에서 수집되는 위키데이터는 구조화된 데이터로 입력되어 지식그래프를 구성하기 용이하며, 구글의 지식그래프의 검색 결과에도 영향을 주어 이용자들에게 아카이브 데이터를 웹상에 쉽게 노출함으로써 서비스를 확장할 수 있다.

또한 위키데이터를 활용해 디지털 아카이브의 기록 데이터의 품질 관리 차원에서 데이터 형식, 데이터의 상호운용성을 제고할 수 있다. 위키데이터는 데이터를 찾을 수 있고, 접근가능하며 상호운용성이 확보되고 재사용하기 위한 지침인 FAIR 원칙을 부합한 거대한 지식베이스(Knowledge base)라 할 수 있다. FAIR 데이터 원칙은 디지털 자원을 기계가 읽고 처리할 수 있는 데이터에 대한 지침으로 웹에 존재하는 모든 자원을 대상으로 적용할 수 있다. 위키데이터는 누구나 편집할 수 있도록 오픈되어 있으며 Wikipedia와 마찬가지로 그 범위가 책, 역사적 사건, 인물, 지리정보 등 다양한 주제를 다루고 있으며 거의 무한하다고 할 수 있다. Wikipedia가 텍스트 기반의 문서로 지식을 제공하고 있다면, 위키데이터는 문서가 아닌 데이터의 구조화된 형식으로 지식을 표현하는 것에 중점을 둔다. 따라서 위키데이터 플랫폼을 통해 아카이브의 데이터가 공유, 연계하는 것은 궁극적으로 FAIR 원칙에 따라 데이터의 검색과 인용, 재사용을 가능하게 하며 아카이브의 도메인과 커뮤니티의 합의된 메타데이터를 준수하는 것으로, 디지털 아카이브의 데이터의 일관성과 정확성을 유지할 수 있을 것이다.

4.3 디지털 아카이브를 위한 고려사항

Babcock et al.(2021)은 디지털 아카이브에서 위키데이터를 적용하고자 할 때 무엇보다 위키데이터라는 지식공유 플랫폼을 통해서 데이터를 공유하고 상호작용하게 되는 생태계를 잘 이해하는 것이 중요하다고 하였다. 위키데이터를 적용하고자 할 때 아카이브에서 활용할 수 있는 데이터 모델링, 온톨로지 도구, 기록자원 서비스의 윤리적 및 법적 검토와 함께 기관과 기관이 봉사하는 커뮤니티가 추구하는 목표, 표준 등을 우선 결정하여야 한다. 본 연구의 위키데이터 사례 분석을 통해 아카이브의 목표에 따라 위키데이터를 적용하는데 있어서 연계방법, 데이터 모델, 도구, 제공하는 융합서비스 등이 다양함을 확인할 수 있었다. 사례 분석 결과를 토대로 향후 위키데이터를 디지털아카이브에 적용할 때 고려해야 할 사항을 종합하면 다음과 같다.

먼저, 아카이브는 기관과 조직의 환경에 따른 고유한 각자의 기록 방침과 목표가 있기 마련이다. 개방형 데이터 환경에서 기록 데이터를 구축함으로써 구현하고자 하는 아카이브의 가치는 무엇인지, 그 의의와 범위를 설정하여야 한다.

둘째, 위키데이터의 공개 플랫폼을 통해서 공유하는 아카이브의 기술 수준에 대해 결정하여야 한다. 위키플랫폼을 통해 접근하게 되는 일반 커뮤니티를 위한 기술 수준과 고려할 만한 기술 표준은 어떠한 것들이 있는지, 또한 아카이브 특정 영역의 전문가들을 위한 기술의 특정성 수준도 결정하여야 한다.

셋째, 일반적으로 기록물 기술 표준을 따르고 있지 않는 외부 시스템의 데이터 연계 시, 아카이브 기록 기술 규격에 따라 기 구축된 데이터를 처리하기 쉬운 지, 기록관, 박물관, 도서관 등의 문화유산기관 이외에도 전혀 종류가 다른 데이터세트와의 연결도 고려할 필요가 있다.

넷째, 위키데이터를 적용할 기준은 무엇인지, 아카이브의 기록 자원과 연결할 외부 데이터의 범위는 어디까지 한정시킬 것인지 결정하여야 한다. 로컬시스템을 통해 메타데이터를 수확하는 방법으로 데이터를 반입하거나 기관 자체적으로 오픈데이터로 편집하는 방법을 선택할 수 있다.

그 외에도 실제 디지털 아카이브에서 위키데이터를 아카이브의 데이터 관리와 공유 서비스에 적용하는 실무 차원에서 고려해야 할 사항들이 있다. 위키데이터 기록 기술 프로젝트의 참여자들은 그동안의 경험을 바탕으로 위키데이터의 문제점과 앞으로 개선이 필요한 내용들을 공유하고 있다. 이를 종합해보면, 아카이브의 기록 데이터와 위키데이터를 통해 외부 데이터를 연계할 때 일괄 방식으로 데이터가 반입, 반출할 때 오류가 가장 많이 발생하였다. 이는 대량의 데이터를 처리할 때 생겨날 수 있는 기술적 오류와 함께 외부의 전자데이터와 기록물을 대상으

로 한 전거데이터의 구조가 상이하여 생겨날 수 있는 문제이기도 하다. 따라서 데이터 연계 시 어느 분야의 전거데이터를 기준으로 할 것인지에 대한 기관 차원에서의 결정과 분야별 전거레코드의 구조와 특성에 대한 이해가 우선 되어야할 것이다.

이와 함께 위키데이터 품질에 있어서, 다양한 기관에서 자발적인 참여로 데이터가 구축되므로 지속적으로 변경되는 데이터, 부정확한 데이터, 악의적으로 변경된 데이터로 인한 문제도 고려하지 않을 수 없다. 위키데이터는 자발적 참여로 이루어지는 비영리재단인 위키미디어 재단에서 운영, 관리되고 있으므로 축적된 위키데이터의 지속가능성과 영속성을 보장할 수 없는 기본적인 한계점을 지닌다 할 것이다.

5. 결론 및 제언

위키데이터는 전 세계 사용자가 협업하여 데이터를 공유, 재사용할 수 있도록 구축한 오픈 지식플랫폼이다. 본 연구는 디지털 아카이브에서 위키데이터를 활용해 양질의 기록 데이터를 확보하고 구조화된 데이터와 연계를 통해 가치 있는 정보를 재생산하고 축적된 아카이브 데이터의 상호운용성을 보장할 수 있는 방안을 제안하고자 하였다.

본 연구는 디지털 아카이브의 데이터 관리와 활용 관점에서, 위키데이터가 가지는 이점은 어느 누구나 쉽게 참여하여 편집할 수 있는 인터페이스, 데이터 식별자 생성의 자동화, 데이터 게시 즉시 필요한 데이터간의 조합과 질의가 가능한 점 등을 들었다. 이러한 위키데이터를 디지털 아카이브의 기록 데이터의 연계와 구축에 활용하여 이용자에게 풍부한 정보를 제공하고 데이터를 효율적으로 관리할 수 있을 것이다. 위키데이터를 활용해 디지털 아카이브의 데이터 정제(cleaning), 표준화와 같은 엄격한 수준의 데이터 품질 관리의 장벽을 낮추고 고품질의 기록 데이터를 관리, 공유, 재사용할 수 있다.

본 연구에서는 위키데이터가 가지는 개념적 특징이 실제 아카이브의 데이터 공유와 재사용에서는 어떻게 활용되고 있는지 사례분석을 통해 논의하고자 하였다. 위키데이터 기록 기술 프로젝트에 참여하고 있는 대표 기관을 선정하여 위키데이터와의 간접적 연계를 통해 아카이브 기록자원의 접근점을 추가하고 직접적인 연계와 통합을 통해 데이터 기반 시각화서비스, 이용자 참여 서비스를 제공하는 방법을 분석하였다. 사례 분석 결과를 토대로, 디지털 아카이브에서 위키데이터의 어휘를 기록 기술에 적용할 수 있는 위키데이터 기술 모델을 제시하고 위키데이터의 활용방안과 향후 적용을 위한 고려사항을 논의하였다.

위키데이터는 오픈 플랫폼으로 많은 데이터가 축적되고 공유될수록 그 효과가 크다. 많은 기관들이 서로 협력하여 활발하게 데이터를 공유하고 상호작용할 때 위키데이터는 거대한 지식베이스로 그 가치가 발휘될 수 있을 것이다. 일반적으로 기록은 집합적 기술의 원칙에 따라 개별 기록의 내용보다 전체적인 구조와 맥락을 중심으로 기술된다. 하지만 보다 풍부한 데이터들이 서로 연결되어 유용한 데이터 활용 환경을 조성하기 위해서는 기록 자체 혹은 데이터의 실질적인 내용이 중요하다. 이를 위해서는 기록물에 포함된 다양한 키워드 및 메타데이터 정보가 필요하며, 기록물의 내용 정보에 대한 기술이나 메타데이터 확보가 시급하다. 지금까지 도메인이나 학문적 주제에 국한되었던 기록 자원을 디지털인문학 연구자나 보다 광범위한 이용자들이 접근, 재사용할 수 있도록 데이터의 공간으로 아카이브의 개념을 확장할 필요가 있다. 위키데이터를 통해서 아카이브의 기록 자원에 대한 접근을 높이고 다양한 관점과 목적의 데이터와 연결하고 이용자에게 새로운 정보를 제공하고 관련된 다양한 맥락을 탐색할 수 있는 기록정보서비스를 모색하여야 할 것이다.

본 연구는 이론적 검토와 사례 조사를 통해 위키데이터를 아카이브의 링크드 데이터와 함께 기록 데이터의 공유와 활성화 방안으로 제안하고 있다. 추후 본 연구에서 제안한 위키데이터 기록 기술 모델을 실제 기록 데이터에 시범적용해보고, 위키데이터를 바탕으로 한 아카이브 자원에 대한 접근성과 이용자 검색이 개선될 수 있는지

실증연구가 뒷받침되어야 할 것이다. 또한 디지털 아카이브의 이용자의 구체적인 요구를 파악하고 기록학 분야 외 IT 전문가들이 참여하여 구축된 위키데이터의 연계와 통합에 있어서 데이터의 신뢰성과 무결성을 검증하는 연구가 진행되어야 할 것이다.

참고문헌

- 국가기록원 (2011). 국가기록원 전거레코드 지침. 대전: 국가기록원.
- 국가기록원 (2013). 영구기록물 정리·기술 지침. 대전: 국가기록원.
- 김태영, 강주영, 김진, 오효정 (2018). 지능형 기록정보서비스를 위한 선진 기술 현황 분석 및 적용 방안. *한국기록관리학회지*, 18(4), 149-182. <https://doi.org/10.14404/JKSARM.2018.18.4.149>
- 김하래 (2021). FAIR 원칙: 데이터 관점의 디지털 아카이브 구현을 위한 고려사항. *한국기록관리학회지*, 21(2), 155-172. <https://doi.org/10.14404/JKSARM.2021.21.2.155>
- 박선휘 (2019). 기록물 맥락정보 향상 및 통합시스템 개발에 관한 연구. RiC-CM 및 RiC-O를 중심으로. *한국외대 정보기록학연구소 기록과 정보·문화 연구*, 9, 55-96.
- 박옥남 (2012). 기록물 전거통제 기반 Linked Data 구축에 관한 연구. *한국비블리아학회지*, 23(2), 5-25. <https://doi.org/10.14699/kbiblia.2012.23.2.005>
- 박지영 (2016). 차세대 기록물 기술표준에 관한 연구. ICA EGAD의 Record In Context를 중심으로. *한국기록관리학회지*, 16(1), 223-245. <http://dx.doi.org/10.14404/JKSARM.2016.16.1.223>
- 박지영 (2017). ISAD(G)에서 RiC-CM으로의 전환에 관한 연구. *한국기록관리학회지*, 17(1), 93-115. <https://doi.org/10.14404/JKSARM.2017.17.1.093>
- 박하람, 김하래 (2021). 일본군 ‘위안부’ 지식그래프: 파편화된 디지털 기록의 연결. *한국기록관리학회지*, 21(3), 61-78. <https://doi.org/10.14404/JKSARM.2021.21.3.061>
- 신미라, 김익한 (2019). RiC를 적용한 아카이브 시스템 데이터 모델링 연구. *한국기록관리학회지*, 19(1), 23-67. <https://doi.org/10.14404/JKSARM.2019.19.1.023>
- 윤소영 (2013). LOD 기반 한국사 콘텐츠 서비스 구축에 관한 연구. *정보관리학회지*, 30(3), 297-315. <https://doi.org/10.3743/KOSIM.2013.30.3.297>
- 이성숙 (2020). 국내 도서관 링크드 오픈 데이터 구축과 발행의 개선방안 연구. *정보관리학회지*, 37(2), 145-169. <https://doi.org/10.3743/KOSIM.2020.37.2.145>
- 이유경, 김하래 (2020). 1997 외환위기 지식그래프: 디지털 아카이브의 관계 중심적 접근. *한국기록관리학회지*, 20(4), 1-17. <https://doi.org/10.14404/JKSARM.2020.20.4.001>
- 이정현, 이윤용, 방기영, 김용 (2015). Linked Open Data 기반 대통령기록관 기록정보 서비스 확장에 관한 연구. *한국기록관리학회지*, 15(2), 55-82. <https://doi.org/10.14404/JKSARM.2015.15.2.055>
- 정주영 (2021). 남산예술센터 시맨틱 디지털 아카이브 구축에 관한 연구. *한국연극학*, 1(77), 211-248. <https://doi.org/10.18396/ksa.2021.1.77.006>
- 정희명, 이성숙 (2021). 디지털 환경에서 기록물 맥락 기술을 위한 Records in Contexts-Ontology(RiC-O) 적용 연구. *한국기록관리학회지*, 21(2), 23-48. <https://doi.org/10.14404/JKSARM.2021.21.2.023>
- Babcock, K., Heberlein, R., McCormick, A., Roke, E., Suiter, G., & Tillman, R. (2021). The Power of Parallel Description: Wikidata and Archival Discovery. In Matienzo, M. A., & Handel, D. (eds). *The Lighting the Way Handbook: Case Studies, Guidelines, and Emergent Futures for Archival Discovery and Delivery*, 99-115.
- Bianchini, C., Bargioni, S., & Pellizzari di San Girolamo, C. C. (2021). Beyond VIAF. *Information Technology and Libraries*, 40(2). <https://doi.org/10.6017/ital.v40i2.12959>.

- Hawkins, A. (2021). Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. *Arch Sci* (2021). <https://doi.org/10.1007/s10502-021-09381-0>
- Lee, Y. Y., Mairelys, L. R., & Marina, B. (2019). Using Wikidata to Provide Visibility to Women in STEM. Paper presented at DCMI 2019 conference. Available: <https://depapers.dublincore.org/pubs/article/view/4255.html>
- Lee, Y. Y., Mairelys, L. R., & Marina, B. (2020, August). Bring the Canadian Archive of Women in STEM to Wikidata: How wikidata can be used in an archival context. Presented at ACA 2020 Virtual Workshop.
- Leibniz-Informationszentrum Wirtschaft Ihr Partner für Forschung und Studium [n.d.]. 20th Century Press Archives. Available: <https://pm20.zbw.eu/>
- Wikidata [n.d]. Wikidata: WikiProject Archival Description. Available: https://www.wikidata.org/wiki/Wikidata:WikiProject_Archival_Description
- Wong, A. & Babcock, K. (2021, April 20). Exploring Wikidata “Archives At” at U of T. Presented at the LD4 Wikidata Affinity Group community.

• 국문 참고자료의 영어 표기

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

- Jeong, Hoemyeong & Lee, Sungsook (2021). A study on the application of records in contexts-ontology(RiC-O) for the description of archives contexts in a digital environment. *Journal of Korean Society of Archives and Records Management*, 21(2), 23-48. <https://doi.org/10.14404/JKSARM.2021.21.2.023>
- Jung, Joo Young (2021). A study on establishment of semantic digital archive of the namsan arts center. *Journal of Korean theatre studies association*, 1(77), 211-248. <https://doi.org/10.18396/ksa.2021.1.77.006>
- Kim, Haklae (2021). FAIR principles: considerations for implementing digital archives from a data perspective. *Journal of Korean Society of Archives and Records Management*, 21(2), 155-172. <https://doi.org/10.14404/JKSARM.2021.21.2.155>
- Kim, Tae-Young, Gang, Ju-Yeon, Kim, Geon, & Oh, Hyo-Jung (2018). A study on the current status and application strategies for intelligent archival information services. *Journal of Korean Society of Archives and Records Management*, 18(4), 149-182. <https://doi.org/10.14404/JKSARM.2018.18.4.149>
- Lee, Jeong Hyeon, Lee, Youn Yong, Bang, Ki Young, & Kim, Yong (2015). A study on the extension of archival information service based on linked open data in the presidential archives. *Journal of Korean Society of Archives and Records Management*, 15(2), 55-82. <https://doi.org/10.14404/JKSARM.2015.15.2.055>
- Lee, Sung Sook (2020). A study on the improvement plan of publishing in library linked open data. *Journal of The Korean Society for Information Management*, 37(2), 145-169. <https://doi.org/10.3743/KOSIM.2020.37.2.145>
- Lee, Yu-kyeong & Kim, Haklae (2020). A knowledge graph of the Korean financial crisis of 1997: a relationship-oriented approach to digital archives. *Journal of Korean Society of Archives and Records Management*, 20(4), 1-17. <https://doi.org/10.14404/JKSARM.2020.20.4.001>
- National Archives of Korea (2011). Guide for Authority Record of National Archives of Korea. Daejeon: National Archives of Korea.
- National Archives of Korea (2013). Guide for Archives Arrangement & Description. Daejeon: National Archives of Korea.
- 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, Ok-Nam (2012). The design and development of linked data from authority data in national archives of Korea. *Journal of the Korean Bibla Society for Library and Information Science*, 23(2), 5-25.

<https://doi.org/10.14699/kbiblia.2012.23.2.005>

- Park, Sun-hee (2019). A study on improving record contextual information and developing integrated system: focusing on RiC-CM and RiC-O. *The Korean Journal of Archival, Information and Cultural Studies*, 9, 55-96.
- Park, Ziyong (2016). Anlayzing the next-generation archival deacription standard: “record in context” of ICA EGAD, *Journal of Korean Society of Archives and Records Management*, 16(1), 223-245.
<http://dx.doi.org/10.14404/JKSARM.2016.16.1.223>
- Park, Zi-young (2017). Transition of archival description from ISA D(G) to record in context conceptual model. *Journal of Korean Society of Archives and Records Management*, 17(1), 93-115.
<https://doi.org/10.14404/JKSARM.2017.17.1.093>
- Shin, Mira & Kim, Ikhan (2019). A study in the data modeling for archive system applying RiC. *Journal of Korean Society of Archives and Records Management*, 19(1), 23-67. <https://doi.org/10.14404/JKSARM.2019.19.1.023>
- Yoon, So-Young (2013). A study on the implementation of korean history contents service based on linked open data. *Journal of Korean Society of Archives and Records Management*, 30(3), 297-315.
<https://doi.org/10.3743/KOSIM.2013.30.3.297>