

FAIR 데이터 원칙을 적용한 국내 디지털 아카이브 평가와 개선 방향*

The Evaluation of Domestic Digital Archives Using the FAIR Data Principles and Their Improvement Directions*

박하람(Haram Park)¹, 강지우(Jiwoo Kang)², 송채은(Chaeun Song)², 안지은(Jieun Ahn)², 이정윤(JeongYun Lee)², 장소정(So Jung Jang)³, 조하연(Hayeon Cho)³, 차화준(Hwa June Cha)³, 김학래(Haklae Kim)⁴

E-mail: haram9553@gmail.com, kangjiwoo330@gmail.com, chaeunsong.hike@gmail.com, aje20010827@gmail.com, cathx618@gmail.com, sji@gmail.com, zoharyeon@gmail.com, hjcha@arko.or.kr, hakelaekim@cau.ac.kr

- 1 제1저자 중앙대학교 일반대학원 문헌정보학과 문헌정보학전공 박사과정
- 2 공동저자 중앙대학교 일반대학원 문헌정보학과 문헌정보학전공 석사과정
- 3 공동저자 중앙대학교 일반대학원 문헌정보학과 기록관리학전공 석사과정
- 4 교신저자 중앙대학교 사회과학대학 문헌정보학과 교수



논문접수 2024.10.15

최초심사 2024.10.23

게재확정 2024.11.22

ORCID

Haram Park

<https://orcid.org/0000-0002-2091-0613>

Jiwoo Kang

<https://orcid.org/0009-0003-8793-7831>

Chaeun Song

<https://orcid.org/0000-0002-1603-6425>

Jieun Ahn

<https://orcid.org/0009-0000-1209-9843>

JeongYun Lee

<https://orcid.org/0009-0000-5105-9764>

So Jung Jang

<https://orcid.org/0009-0007-2167-5110>

Hayeon Cho

<https://orcid.org/0009-0004-8638-2667>

Hwa June Cha

<https://orcid.org/0009-0002-8683-9271>

Haklae Kim

<https://orcid.org/0000-0002-2616-421X>

© 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 논문은 2023학년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음

초 록

디지털 아카이브는 온라인 환경에서 기록을 보존하고 접근할 수 있는 도구로, 문화예술, 과학, 비즈니스 등 다양한 분야로 확산되고 있다. 디지털 아카이브는 웹페이지, 게시판, 기록관리 시스템 등 여러 형태로 구현되어 운영되고 있다. 그러나, 디지털 아카이브에 대한 명확한 개념이 부재하고, 기록관리의 이론을 적용하는 데 한계가 있다. 본 연구는 탐색성, 접근성, 상호운용성, 재사용성의 관점에서 디지털 아카이브를 평가하는 방법론을 제시하고, 국내 디지털 아카이브 69개를 대상으로 FAIR 데이터 원칙 준수 여부를 분석한다. 연구 결과에 따르면, 국내 디지털 아카이브는 전반적으로 FAIR 데이터 원칙을 반영하지 못하고 있다. 261개의 디지털 아카이브에서 74%가 탐색과 접근에 제한이 있고, 평가 대상에 포함된 69개 아카이브의 FAIR 평균 점수는 50.43이다. 특히, 상호운용성과 재사용성은 전반적으로 매우 낮은 수준이다. 디지털 아카이브는 단순한 기록 제공 도구가 아닌 기록관리의 연장선상에서 중요한 역할을 해야 하며, 이를 위해 FAIR 데이터 원칙을 적용한 체계적인 연구와 개선이 필요하다.

ABSTRACT

Digital archives serve as tools for preserving and accessing records online, expanding across diverse domains, including arts, science, and business. While these archives are implemented and operated in various forms, such as web pages, bulletin boards, and records management systems, a lack of a clear conceptual framework for digital archives and limitations in applying archival theory exist. This study proposes a methodology for evaluating digital archives from the perspectives of findability, accessibility, interoperability, and reusability (FAIR), analyzing FAIR data principle adherence across 69 domestic digital archives. The findings indicate that domestic digital archives in South Korea generally fail to adequately reflect the FAIR data principles. Among the 261 digital archives examined, 74% demonstrate limitations in search and access, while the 69 archives selected for detailed evaluation achieved an average FAIR score of 50.43. Notably, the metrics for interoperability and reusability demonstrated significantly low performance levels across the evaluated archives. Digital archives must transcend their function as mere record provision tools to serve as essential extensions of records management. This transformation necessitates systematic research and improvements focusing on applying FAIR data principles to enhance their effectiveness.

Keywords: 디지털 아카이브, FAIR 데이터 원칙, 기록관리, 상호운용성, 프레임워크
Digital archives, FAIR data principles, Records management, Interoperability, Framework

1. 서론

문화유산, 학술자료, 공공기록 등 다양한 분야에서 디지털 아카이브 구축과 서비스가 활발히 이루어지고 있다. 디지털 아카이브의 확산은 인공지능으로 대표되는 디지털 기술의 발전, 정부의 디지털 전환 정책, 비대면 서비스 수요 증가 등 복합적 요인에 기인하고, 이러한 동향은 기록의 중요성을 제고하는 기회로 인식되고 있다.

기록에 대한 관점에 따라 디지털 아카이브의 역할과 범위는 다르게 해석된다. 기록관리(Records Management)는 조직, 개인 또는 기관이 기록물의 생성, 수집, 유지, 보존, 폐기 등 전체 수명 주기를 체계적으로 관리하는 프로세스를 의미한다. 전통적 기록 관리는 진본성, 무결성, 신뢰성, 이용가능성을 보장하고, 기록의 법적, 행정적, 역사적 가치 유지에 중점을 둔다. 이러한 맥락에서 기록관리 시스템(RMS: Records Management System)은 조직 내에서 생성, 수집되는 기록물의 전체 수명 주기를 체계적으로 관리하는 목표를 갖고 있다. 협의의 개념에서 디지털 아카이브는 기록관리 시스템의 일부로 해석할 수 있다. 디지털 아카이브는 기록물의 수명 주기 관리와 더불어 사용자들의 기록 탐색, 분석, 활용을 위한 다양한 기능을 포함하고 있다. 전통적 기록관리와 마찬가지로 디지털 아카이브는 특정 자료를 체계적으로 정리하고 보관하지만, 물리적 자료의 디지털화와 디지털 환경에서 생산된 원본 디지털 자료(born-digital materials)까지 포함한다.

그러나 전통적 기록관리 영역에서 디지털 아카이브의 특성과 범위에 대한 명확한 합의가 이루어지지 않고 있다. ISO 15489-1과 OASIS(Open Archival Information System) 모델은 기록 관리 원칙과 장기 보존을 위한 기록관리 시스템의 구조를 포괄하는 가이드라인을 제공하지만, 디지털 아카이브 시스템의 핵심 요소와 요구사항에 대한 구체적 지침은 제시하지 않는다. 디지털 기술의 발전으로 아카이브의 형태와 기능이 지속적으로 변화하고 있지만, 디지털 아카이브의 개념과 특성에 대한 이론적 합의가 미흡하고, 시스템 관점에서 공통 기능, 구축, 평가를 위한 기준이 부재한 것이 현실이다. 이러한 문제는 디지털 아카이브 운영 사례에서 더욱 두드러지게 나타난다. 대부분의 디지털 아카이브는 검색 기능의 한계, 메타데이터의 불완전성으로 인해 사용자의 효과적인 정보 접근이 어렵고, 데이터의 연계와 재사용, 상호운용이 제한적이다.

한편, 디지털 아카이브 평가를 위한 프레임워크는 체계적으로 연구되지 않고 있고, 디지털화된 자료의 수, 이용자 접속 횟수와 같은 제한적 지표에 의존하고 있다. 디지털 아카이브의 질적 가치와 실제 활용을 위해 객관적인 평가지표의 설계와 적용이 중요하다. 이러한 맥락에서 FAIR 데이터 원칙(Findable, Accessible, Interoperable, Reusable)(Wilkinson et al., 2016)의 적용은 디지털 아카이브 평가에 새로운 관점을 제시한다. FAIR 데이터 원칙은 데이터의 실질적인 재사용과 공유를 위해 개발된 가이드라인으로, 데이터의 탐색 가능성, 접근성, 상호운용성, 재사용성을 강조한다. FAIR 데이터 원칙은 별점 5 오픈 데이터(5 Star Open Data)에 기술한 데이터의 접근과 활용이라는 보편적 가치를 지향하며, 연구 데이터와 정부 데이터에 폭넓게 적용되고 있다. FAIR 데이터 원칙은 디지털 아카이브의 핵심 가치와 밀접하게 연관된다. FAIR 데이터 원칙에서 강조하는 표준의 활용, 메타데이터의 충실한 기술, 명확한 라이선스와 출처정보 제공 등은 디지털 아카이브의 기록을 신뢰할 수 있고 재사용 가능한 형태로 유지할 수 있는 가이드라인이 될 수 있다. 그러나 기록관리와 도서관 분야에서 이에 대한 연구가 부족한 것이 현실이다.

본 연구의 목적은 한국에서 운영되는 디지털 아카이브를 FAIR 데이터 원칙에 따라 평가하고, 이를 바탕으로 개선 방안을 제시하는 것이다. 연구 방법은 크게 두 단계로 구성된다. 첫째, 디지털 아카이브의 운영 현황을 검토하고 선정 기준을 마련하여 평가 대상을 선정한다. 둘째, 선정된 대상에 FAIR 데이터 평가 프레임워크를 적용하여 분석을 수행한다. 평가 결과는 FAIR 지수로 수치화되며, 이를 통해 구체적인 개선 방향을 도출한다. 본 연구는 국내 디지털 아카이브의 현재 상황을 객관적으로 파악하고, 국제적 기준에 부합하는 발전 방향을 제시하여 향후

관련 연구와 정책 수립의 중요한 기초 자료로 활용될 수 있다.

2. 관련연구

2.1 기록의 이해

기록은 개인이나 조직이 법적 의무 이행이나 업무 수행 과정에서 발생하는 증거나 정보로서 생산, 접수, 유지하는 정보를 말한다(International Organization for Standardization, 2016, ISO 15489-1). 국제기록관리협의회(International Council on Archives: ICA)는 ‘활동의 증거를 제공하는 데 충분한 내용, 맥락, 구조를 포함’하는 특징을 강조한다. 즉, 기록은 사건의 증거를 말한다.

전자기록은 디지털 형식으로 존재한다는 의미에서 디지털 기록과 같은 맥락으로 사용한다. 전자기록은 원래부터 디지털화된 기록(born digital)과 디지털로 변환된 기록(being digital)으로 구분할 수 있다. 두 개의 유형은 생성 방식과 특성에서 차이를 보이며, 각각의 관리와 보존에 있어 고유한 특징이 있다. 전자형태로 생산된 기록은 물리적 원본 없이 컴퓨터 소프트웨어와 같은 디지털 기술을 통해 생성되며, 단일한 원본 참조대상(singular original reference)이 없고 끝없이 복제될 수 있어 원본과 복제본의 경계가 존재하지 않는다(이상호, 2023; Carbajal & Caswell, 2021). 특히, 생산부터 보존까지 특정 하드웨어와 소프트웨어에 의존하고 있어 기술 변화에 따른 보존전략이 함께 필요하다(김유승, 2019). 한편, 기록의 디지털화는 물리적인 형태(예: 종이 기록물)로 존재하는 기록을 디지털 형식으로 변환하는 과정으로 기록의 보존, 관리, 검색, 공유 측면에서 효율성을 높이기 위해 필수적이다. 디지털화된 기록은 물리적 훼손 없이 오랫동안 보관 가능하며, 어디서나 빠르게 검색하고 접근할 수 있다. 아날로그 기록을 디지털화하는 과정에서 새로운 맥락이 추가되기 때문에 별도의 메타데이터(예: 스캔 날짜)가 추가되며, 원본과의 연계를 위한 보존방안이 별도로 필요하다(한국산업표준, 2020, KS X ISO TR 13028:2010).

국내 기록관리 분야에서 전자기록에 대한 다양한 연구가 진행되고 있다. 그러나 대부분의 연구는 주로 전자기록의 생성, 관리, 보존 방법론에 집중되어 있으며, 실제 기록 데이터의 연계와 활용을 고려한 연구는 상대적으로 미흡한 실정이다(김명훈, 2024; 김유승, 2019; 이상호, 2023; 이승역, 설문원, 2021). 디지털 형태의 기록은 존재하는 순간부터 변형 가능성, 복제의 용이성, 기술 환경에 대한 의존성이라는 공통적인 속성을 갖는다(박순철 외, 2018; Pacheco et al., 2023). 기록 메타데이터는 디지털 기록의 변경사항을 지속적으로 추적하고 디지털 기록의 무결성과 신뢰성을 보장하는 핵심 요소이다(Carbajal & Caswell, 2021; Pacheco et al., 2022; Zhang & Mauney, 2013). 기록관리의 대상과 영역이 확대되는 것을 고려하면, 기존의 생성, 관리, 보존 방법론에 대한 연구를 넘어 다양한 유형의 전자기록 간 연계와 통합적 활용 방안에 대한 심도 있는 이해와 연구가 요구된다. 특히, 기록의 보존과 더불어 기록된 데이터의 가치를 극대화하고 실제적인 활용을 촉진하는 방향에 대해 고려할 필요가 있다. 본 연구는 디지털 기록의 보존, 접근뿐만 아니라 기록 자원의 활용을 위한 공통적인 특징을 소개하고 이에 대한 가이드라인을 제시한다.

2.2 디지털 아카이브에 대한 개념

아카이브는 장기적 가치를 위해 보존된 인류 활동의 기록물이며, 텍스트, 사진, 비디오, 음성, 아날로그, 디지털

등 다양한 형식으로 존재한다. 아카이브는 개인과 기관이 보유하고 있으며, 기록물을 수용하는 건물을 종종 ‘아카이브’라고 지칭하기도 한다. 한편, 디지털 아카이브는 전자적으로 생성되거나 디지털화된 자료를 장기적으로 보존하고 지속 가능한 접근을 제공하는 시스템으로 정의한다(Hedstrom, 1997).

전통적인 기록관리 관점에서 디지털 아카이브는 ‘지속적인 가치를 가진 디지털 객체를 장기간 관리하는 활동(이소연, 2002)’ 또는 ‘디지털 자원의 보존, 관리, 접근을 가능하게 하는 포괄적 활동(설문원, 2005)’이라고 정의한다. 즉, 디지털 아카이브를 기록의 디지털 보존(digital preservation)에 초점을 둔 기록관리 시스템의 한 유형으로 이해하는 경향이 있다. 이와 같은 관점에서 OAIS(Open Archival Information System), ISO 15489-1 등 기록관리와 관련된 표준과 디지털 아카이브에 대한 연구(강민정, 장우권, 2021; 이소연, 2002; Lee, 2010), 응용 시스템의 개발(Caplan, 2010; Riasetiawan et al., 2010), 보존 메타데이터 표준 구축(박희진, 2020; Caplan, 2009)에 대한 연구가 있다. 한편, 디지털 아카이브는 단순히 디지털화된 기록물을 저장하는 데이터베이스의 개념으로 이해되기도 한다. 디지털 아카이브는 ‘디지털 자원의 수집, 저장, 백업, 아날로그 자원의 디지털화 등을 포함하는 활동과 그로 인해 구축된 데이터베이스(정희선 외, 2016)’로서 저장의 의미가 강조된다. 반면, 디지털 아카이브는 전통적인 기록물과는 맥락과 특성에서 본질적인 차이가 존재하므로 새로운 개념화가 필요하다는 주장이 있다(Klareld & Gidlund, 2017; Laermans & Gielen, 2007). 디지털 보존은 디지털 정보에 대한 접근성을 장기적으로 제공하는 데 초점이 있고 디지털 아카이브는 정보의 검색과 접근성 확보가 핵심이다(Abu, 2021). 이러한 개념을 기반으로 웹 기반 디지털 아카이브의 구축(Basuki et al., 2022), 디지털 컬렉션의 효과적인 인터페이스 설계(Padia et al., 2012) 등의 활용 중심의 연구가 이루어졌다. 디지털 아카이브의 높은 개방성과 접근성(Yadav, 2016), 상호운용성과 재사용을 위한 메타데이터 구축 방법론(Zhang & Mauney, 2013), 디지털 시대에서 아키비스트의 새로운 역할(Jensen, 2021)에 대한 논의도 진행되고 있다.

그러나 다양한 연구에도 불구하고 디지털 아카이브에 대한 현대적 관점에서의 명확한 정의와 그 특징에 대한 합의는 여전히 모호하다. 디지털 기록의 급격한 증가로 인해 다양한 디지털 자료에 대한 접근성이 확장되고 있고, 디지털 아카이브가 단순한 보존을 넘어서 상호운용성과 재사용성을 포함하는 확장된 개념으로 이해되어야 한다는 필요성이 제기되고 있다(Hawkins, 2022). Barbuti(2021)가 지적한 바와 같이 디지털 아카이브는 아날로그 기록에서 디지털화된 기록과 디지털로 생성된 기록을 모두 포괄하며, 이들 간의 서로 다른 특성을 고려해야 한다. 본 연구의 수행에 앞서 디지털 아카이브는 디지털 형태로 생성되거나 디지털화된 정보 자원을 체계적으로 수집, 정리, 보존하고 사용자에게 제공하는 정보 시스템으로 정의한다.

2.3 디지털 아카이브에 대한 평가 방안

디지털 아카이브의 평가는 대상, 목적에 따라 다양하며, 평가 방법론을 일반화하는 데 한계가 있다(Villanueva & Shiri, 2021). 일반적으로 디지털 아카이브의 평가는 사용성(Kilchenmann et al., 2019; Matusiak, 2022; Power et al., 2017; Villanueva & Shiri, 2021), 소프트웨어(박태연, 신동희, 2016; 전경선, 장운금, 2018; Kim et al., 2022), 메타데이터(하은미 외, 2022; Force & Smith, 2021; Zhang & Mauney, 2013) 관점으로 구분한다. 한편, 디지털 아카이브에서 메타데이터의 상호운용을 위해 온톨로지 어휘를 적용 방안을 제안하는 연구도 비교적 활발하게 연구되고 있다(도슬기, 박희진, 2023; 정성미, 2023; 정주영, 2021; 한상은, 박희진, 2022). 디지털 아카이브를 평가하는 다양한 방식이 존재하지만, 기록의 특성을 반영해 일관성 있게 디지털 아카이브를 평가하는 방법은 부족한 실정이다.

FAIR 데이터 원칙은 연구 데이터의 재사용성을 향상시키기 위한 지침으로, 연구 과정에서 생산되는 데이터를

효율적이고 장기적으로 관리할 수 있는 방법을 제공한다(Wilkinson et al., 2016). FAIR 데이터 원칙은 연구 데이터의 재사용성을 넘어 공공행정(Lia & Colella, 2023), 생물학(Corpas et al., 2018), 고고학(Nicholson et al., 2022), LLM(Sundaram & Musen, 2023)과 같은 다양한 분야에서 데이터 공유와 재사용을 위한 원칙으로 적용되고 있다. 문헌정보(Barbuti, 2021; Dobreski et al., 2022)와 기록관리(Calamai & Frontini, 2018; Park & Kim, 2023) 분야에서 디지털 기록물의 관리와 보존, 상호운용, 활용을 위해 FAIR 원칙을 소개하고 있지만, 디지털 아카이브 관점에서 평가 방법은 정립되지 않았다.

FAIR 데이터 원칙의 평가 도구는 다양하다. 일반적으로 평가 프레임워크로 말하는 평가도구는 온라인 자가평가 설문조사(online self-assessment survey), 온라인 (준)자동화(online (semi-)automated), 오프라인 자가평가(Offline self-assessment)와 같은 방식이 있다(Krans et al., 2022). 온라인 자가평가는 사용자가 직접 FAIR 데이터 원칙에 대해 평가할 수 있는 설문을 웹 페이지로 제공한다. FAIRdat(DANS, 발행년불명), ARDC(Australian Research Data Commons) FAIR Self Assessment Tool(ARDC, 2024), FAIR DataSet Maturity Assessment Tool(Welter et al., 2023)은 평가지표와 점수 계산방식에 대해 세부적인 차이가 있지만, 사용자가 온라인 환경에서 비교적 간단하게 FAIR 평가를 할 수 있다. 온라인 (준)자동화는 GUID, URL, DOI와 같은 온라인 식별자를 입력받아서 FAIR 평가 결과를 자동으로 측정한다. WGF FAIR assessment(Ammar et al., 2020), FAIR evaluator software(Wilkinson et al., 2019), F-UJI(Devaraju et al., 2021), FAIR-checker(Gaignard et al., 2023), CkanFAIR(Lia & Colella, 2023)는 웹에 있는 디지털 자원을 사용자의 추가적인 개입 없이 자동으로 평가하는 기능을 제공한다. 오프라인 자가평가는 문서 형식으로 구체적인 평가 지표와 매트릭스만 제공하므로 사용자가 직접 개별 지표를 평가하고 계산해야 한다. 이 방식의 평가도구는 RDA-SHARC Simple Grids(David et al., 2019), FAIRsFAIR Data Object Assessment Metrics(Devaraju et al., 2021), RDA FAIR Data Maturity Model(Bahim et al., 2020), FAIRplus DataSet Maturity Model(Welter et al., 2023)이 포함된다.

본 연구는 기록의 접근성, 활용성, 상호운용성 관점에서 디지털 아카이브를 평가하며, 이를 위해 FAIR 데이터 원칙과 평가 프레임워크를 적용한다. 다양한 평가 프레임워크가 존재하나, 확장성 있게 FAIR 원칙의 핵심을 평가할 수 있으면서 웹 환경에서 쉽게 평가가 가능한 ARDC FAIR Self Assessment Tool을 평가에 사용한다. 따라서 본 연구의 평가 방식은 전통적인 방법에서 사용된 사용성 평가와 차별화되며, 사용자의 주관적 해석이 아닌 객관적인 평가지표를 통해 정량적으로 분석한다.

3. 연구 방법

3.1 FAIR 데이터 원칙의 이해

FAIR 데이터 원칙은 데이터의 접근, 공유와 재사용을 강화하기 위해 제안된 가이드라인이다(Wilkinson et al., 2016). FAIR 데이터 원칙은 데이터를 탐색할 수 있고(findable), 접근 가능하고(accessible), 상호운용이 가능하며(Interoperable), 재사용(resuable)할 수 있는 15가지의 요소를 제공한다(FORCE11, n.d.b).

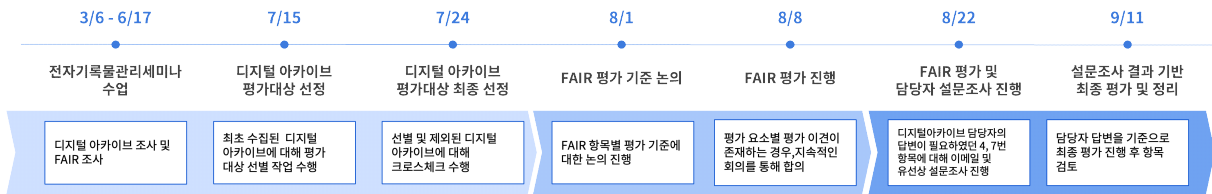
탐색성은 4개의 요소(F1, F2, F3, F4)로 구성되고, 데이터를 탐색할 수 있는 특성을 강조한다. 탐색성의 기본 원칙은 다음과 같다: F1) 데이터는 범용적으로(universally) 고유하고 영구적인 식별자(identifier)를 갖는다. F2) 데이터가 풍부한 메타데이터로 기술되어야 한다. F3) 메타데이터가 식별자를 포함해야 한다. F4) 메타데이터가 구글과 같은 검색엔진에 색인화될 수 있도록 구성되어야 한다. 접근성은 표준적이고 보편적인 프로토콜을 통해

데이터에 접근할 수 있는 특성으로, 사용자가 데이터를 재사용하기 위한 필수적인 조건이다: A1) 데이터와 메타데이터는 표준화된 통신 프로토콜(예: HTTP(S), FTP)을 사용해 식별자로 검색되어야 한다. A1.1) 이 프로토콜은 누구나 접근할 수 있도록 무료이고 개방형이어야 한다. A1.2) 데이터에 접근할 수 있는 정확한 인증, 권한 등의 조건을 명확히 제공해야 한다. A2) 데이터가 웹 환경에서 활용이 불가능한 상황에서도 메타데이터를 통해 접근할 수 있는 방법이 제공되어야 한다. 상호운용성은 서로 다른 데이터와 메타데이터를 쉽게 결합하고 사용할 수 있도록 지원하는 원칙이다: I1) 데이터와 메타데이터는 도메인과 관련된 커뮤니티에서 합의한 표준이나 광범위하게 적용되고 있는 어휘를 사용해야 한다. I2) 이때 통제 어휘, 시소러스, 온톨로지와 같이 FAIR 원칙을 준수하는 어휘를 사용해야 한다. I3) 데이터와 메타데이터가 다른 데이터 또는 메타데이터와 연결을 통해 의미 있는 정보를 제공해야 한다. 재사용성은 FAIR 데이터 원칙의 궁극적인 목표로서, 명확한 라이선스와 출처정보를 상세히 설명해야 한다는 특성을 강조하고 있다: R1) 데이터의 맥락을 이해할 수 있도록 데이터에 정확하고 관련성 높은 메타데이터가 충분히 기술되어야 한다. 이때 R1.1) 데이터의 이용조건을 명확히 알 수 있는 라이선스와 R1.2) 출처정보(예: 생산자, 생산 방식, 맥락)가 메타데이터에 포함되어야 한다. R2) 더불어 데이터와 메타데이터가 재사용 가능하도록 도메인과 관련된 커뮤니티 표준을 따라야 한다.

FAIR 데이터 원칙은 데이터의 재사용과 공유를 위해 고려해야 할 요소를 제공하고 있지만, 다양한 평가 도구가 존재한다. 이 중 ARDC에서 개발한 FAIR Data Self Assessment Tool은 총 12개의 세부 지표로 구성된 설문지 기반의 평가 도구로, 연구자와 데이터 사서, IT 직원 등이 FAIR 관점에서 데이터를 자가검진 할 수 있도록 설계되었다. 이 평가 도구는 웹상에서 사용자가 직접 설문을 진행할 수 있고, 탐색성, 접근성, 상호운용성, 재사용성 항목마다 핵심 설명과 함께 추가적인 개념을 설명하거나 관련 자원에 대한 링크를 제공하고 있다. 특히, 개별 지표에 대해 구체적인 선택지를 제공하고 있어 평가가 비교적 용이하다. 각 항목마다 정규화된 점수를 제공하여 항목 사이의 점수 비교가 가능하며 총 평가 점수도 제공한다. 본 연구는 평가 프레임워크로 FAIR 원칙의 핵심 요소를 객관적으로 정량화하여 평가할 수 있으며, 웹 인터페이스를 통해 쉽게 평가가 가능한 ARDC FAIR Self Assessment Tool을 선택하고, 국내 디지털 아카이브의 현황과 개선방안을 분석한다.

3.2 평가 대상의 선정

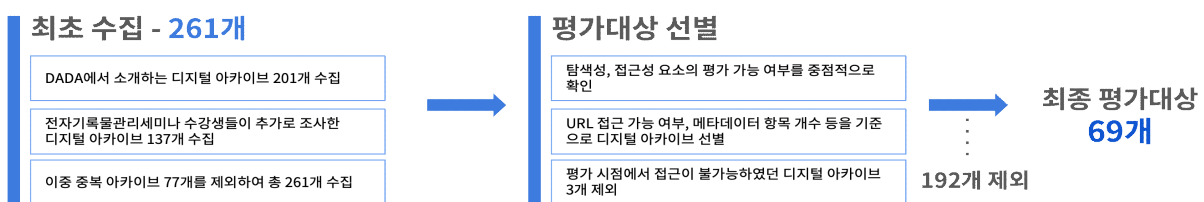
본 연구는 2단계로 구분해서 진행되었다. 먼저, 중앙대학교 일반대학원의 전자기록물관리세미나 수업(2024년 1학기)을 통해 261개의 디지털 아카이브 목록을 정리하고, 중간/기말 프로젝트를 통해 디지털 아카이브에 FAIR 데이터 평가 프레임워크를 적용했다. 이후 본 연구에 참여한 9명의 연구진이 디지털 아카이브 관점에서 평가 프레임워크의 해석, 디지털 아카이브의 재선정, 그리고 평가 작업을 수행했다. 그림 1에서 보듯이 디지털 아카이브의 평가 절차는 단계별로 진행되었다. 평가 대상의 선정은 제외 대상의 기준을 수립하고(7/15) 최종 대상을 목록으로 정리했다(7/24). FAIR 데이터 평가 프레임워크는 디지털 아카이브에 적용할 수 있는 범용성에 우선순위를 두었고, ARDC 평가도구가 최종 선정됐다. 이후 개별 지표에 대한 평가 기준을 상세화했다(8/1). 평가 작업은 연구진이 평가하는 대상을 선정하고 1차 평가를 수행한 후, 평가 결과를 검토하고 논의하는 과정으로 진행되었다. 이 단계에서 개별 연구원이 적용한 평가 지표가 모호한 경우 기준을 명확하게 재정의했다. 예를 들어, FAIR 평가 프레임워크의 2개 지표(F4, A3)는 디지털 아카이브 담당자의 확인이 필요하다는 의견이 있었고, 이에 따라 이메일 또는 유선으로 확인하는 과정이 포함되었다. 마지막으로 최종 평가결과는 1개의 디지털 아카이브에 2명 이상의 연구자가 교차검증을 진행했다.



<그림 1> 연구 진행 과정

디지털 아카이브의 선정은 그림 2와 같이 수행되었다. 최초 수집한 261개의 디지털 아카이브 목록은 다음과 같은 두 가지 방법으로 수집했다. 첫째, 2024년 3월 6일부터 6월 17일까지 구글 검색엔진에서 “디지털 아카이브”, “디지털기록관”, “온라인 아카이브”, “디지털 컬렉션”, “디지털 전시관”, “온라인 기록관” 키워드로 검색을 실시하여 총 137개의 디지털 아카이브를 조사했다. 둘째, 디지털 아카이브를 수집한 웹 사이트 DADA(Digital Archiving for Digital Archive)에서 2024년 6월 17일부터 7월 14일까지 웹 크롤링을 통해 201개의 아카이브 목록을 수집했다. 이 과정에서 중복으로 수집된 아카이브를 제외하고 총 261개의 디지털 아카이브 목록을 수집했다. 이 목록에서 평가할 디지털 아카이브를 선별하기 위해 최소한의 기준을 적용했다. FAIR 데이터 원칙에서 상호운용성과 재사용성은 탐색성과 접근성을 필수적으로 요구하므로 다음의 기준을 최소 선정 기준으로 정의한다. 이 기준으로, 온라인으로 접근이 불가능한 3개의 아카이브를 포함해 총 192개의 디지털 아카이브를 제외하고, 총 69개의 디지털 아카이브를 평가 대상으로 선정했다.

- 탐색성: 아이템(item) 수준의 디지털 기록물을 식별할 수 있는 웹 URL이 존재하는지 평가한다. 예를 들어, 한국전통문화 사진 디지털 아카이브와 같이 웹 URL로 개별 기록물에 접근할 수 없는 디지털 아카이브는 평가 대상에서 제외한다.
- 접근성: 첫째, 별도의 제한 없이 디지털 아카이브의 기록물에 접근해야 한다. 한국기록공보 디지털 아카이브와 같이 유료구독을 통해서만 모든 컬렉션과 아이템에 접근가능한 경우는 제외한다. 둘째, 디지털 기록을 기술하는 충분한 메타데이터(예: 제목, 설명 이상)가 제공해야 한다. 청주공예비엔날레 아카이브나 서울시 도시건축 위원회 아카이브와 같이 제공하는 메타데이터가 매우 적으면(예: 제목, 설명만 기술) 평가 대상에서 제외한다.



<그림 2> 디지털 아카이브의 선별 과정

3.3 평가 프레임워크

본 연구에서 적용한 ARDC의 프레임워크는 FAIR 원칙의 각 구성요소에 대해 가중치를 차등부여한 점수 체계를 적용하고 있다. 탐색성은 4개의 세부 지표로 구성되며 총 17점이 배정된다. 접근성, 상호운용성, 재사용성은 각각 3개(10점), 3개(8점), 2개(7점)의 지표로 구성되어 있다. 세부 지표는 2~7개의 선택항목을 갖고 있고, 각 선택항목에 따라 할당된 점수는 다르다. 최종 점수(Total Score)는 개별 항목에 부여된 점수 비율에 균등한 가중치를 부여한 총합(최대 100점)으로 계산된다. F 항목은 4개의 지표, A 항목은 3개의 지표, I 항목은 3개의 지표, R

항목은 2개로 구성되고, 개별 항목의 점수는 개별 지표를 합한 값의 평균 점수다. 총 평가 점수는 개별 항목에 대한 평균 점수에 25점의 가중치를 부여한 총합이다. 총 평가 점수의 계산식(Total Score)은 다음과 같다.

$$Total\ Score = \left(\frac{\sum_{i=1}^4 F_i}{17} + \frac{\sum_{i=1}^3 A_i}{10} + \frac{\sum_{i=1}^3 I_i}{8} + \frac{\sum_{i=1}^2 R_i}{7} \right) \times 25$$

ARDC 평가 프레임워크는 FAIR 데이터 원칙에 대한 범용성 있는 도구이지만, 기록관리 또는 디지털 아카이브의 특성을 반영하는 데 한계가 있다. 따라서, 본 연구는 디지털 아카이브의 특성을 반영할 수 있도록 평가지표의 세부 내용을 구체화하고 일부 항목은 재해석해서 적용하고 있다. 표 1은 ARDC 프레임워크의 평가지표와 재해석한 평가 기준을 요약하고 있다. 표 2는 기존의 ARDC 프레임워크의 평가지표에 대해 연구진들이 재해석한 일부 지표에 대한 예시다. 디지털 아카이브 평가의 재현가능성을 위해 연구진들이 정리한 세부 지표의 해석 전문은 깃헙에서 확인할 수 있다.

<표 1> ARDC 프레임워크의 지표와 재해석한 결과

항목	설명
탐색성	<ul style="list-style-type: none"> F1: 어떤 식별자가 데이터세트에 할당되었는가? (0~8점) 디지털 아카이브에서 데이터세트는 ‘디지털 기록물(아이템)’로 간주하고, 이 기록물에 할당된 식별자를 평가한다. 예를 들어, 개별 디지털 기록물이 특정한 웹 URL로 식별될 경우는 웹주소(URL)로 평가한다.
	<ul style="list-style-type: none"> F2: 데이터세트의 식별자가 데이터를 기술하는 모든 메타데이터 레코드/파일에 포함되어 있는가? (0~1점) 이 질문은 디지털 기록물의 메타데이터에 고유한 식별자 항목이 포함되는지 판단한다. 예를 들어, ‘식별자’, ‘순번’, ‘고유번호’, ‘finding aid ID’, ‘참조코드’는 디지털 기록물의 고유한 식별자를 포함한 메타데이터로 평가할 수 있다.
	<ul style="list-style-type: none"> F3: 데이터는 어떤 방식으로 메타데이터에 기술되었는가? (0~4점) 메타데이터의 평가 대상은 디지털 기록물 단위의 메타데이터를 평가한다. 기계가 처리할 수 있는 형식의 메타데이터는 RDF, XML, CSV, JSON, html의 meta tag와 같이 구조화된 형태를 의미한다.
	<ul style="list-style-type: none"> F4: 메타데이터 레코드가 어떤 유형의 레포지토리(repository)나 레지스트리(registry)에 저장되어 있는가? (0~4점) 이 질문은 개별 디지털 객체 자체가 아닌 디지털 객체의 ‘메타데이터’가 존재하는 저장소를 기준으로 평가한다. 이 질문은 디지털 아카이브의 담당자에게 설문한 결과를 토대로 평가한다. 보편적인 검색 엔진(예: 구글)에서 개별 아이템 단위의 검색이 가능한 경우는 가장 높은 점수로, 별도의 레포지토리/레지스트리 없이 내부 시스템에서만 관리할 경우는 가장 낮은 점수로 평가한다.
접근성	<ul style="list-style-type: none"> A1: 어떻게 데이터에 접근 가능한가? (0~5점) 디지털 아카이브의 데이터는 디지털 객체(아이템 또는 디지털 기록물)로 평가한다. 디지털 아카이브의 특성에 따라 대상의 형식은 다양할 수 있다(예: 공연 영상(youtube), 사진(jpg, png), 문서(pdf, hwp)). 한편, 이 질문은 기록물의 공개가 아닌 접근 방식에 대해 평가한다. 디지털 아카이브는 기록물의 내용에 따라 저작권 또는 상업적 이익의 문제로 접근이 제한될 수 있다. 이 질문의 평가는 직접 기록물을 확인할 수 없어도 명시적인 열람 조건과 절차가 안내되어 있는지를 판단한다.
	<ul style="list-style-type: none"> A2: 접근이 승인되면 특별한 프로토콜이나 도구 없이도 온라인에서 데이터를 이용할 수 있는가? (0~4점) 이 질문은 http(s)와 같은 표준 프로토콜로 디지털 기록물에 접근가능한지 평가한다. 예를 들어, 디지털 기록물이 특별한 제한 없이 API로 제공되거나 파일 다운로드가 가능한 경우를 포함한다.
	<ul style="list-style-type: none"> A3: 더 이상 데이터를 사용할 수 없는 경우에도 메타데이터 레코드에 접근할 수 있는가? (0~1점) 이 질문은 디지털 아카이브 서비스가 종료되어도 디지털 기록물의 메타데이터에 접근할 수 있는 방법을 제공하는지 평가한다. 이 질문은 디지털 아카이브의 담당자에게 직접 설문한 조사를 기반으로 평가한다.
상호운용성	<ul style="list-style-type: none"> I1: 데이터가 이용가능한 (파일) 형식은 무엇인가? (0~2점) 이 질문은 데이터의 형식이 구조화되고 개방형 표준을 사용해 기계가 읽을 수 있는 형식(structured, open standard, machine-readable format)에 가까울수록 높은 점수를 부여한다. 예를 들어, pdf/나 csv 형식은 특정 어플리케이션에 국한되지 않고 이용할 수 있으므로 높은 평가를 받는다. 반면, hwp나 xls는 특정 소프트웨어로만 열람이 가능하므로 낮은 점수를 받는다.

	<ul style="list-style-type: none"> I2: 데이터의 요소(elements)를 정의하는 데 사용되는 어휘/온톨로지/태깅 스키마의 유형을 가장 잘 설명하는 것은 무엇인가? (0~3점) 이 질문은 디지털 객체의 메타데이터를 표현할 때 표준화된 어휘, 온톨로지, 태깅 스키마를 사용했는지 평가한다. 표준 어휘는 ISO, DCMI 등의 표준 기관에서 공식적으로 출판된 것을 포함한다. 디지털 아카이브에서 명시적으로 표준 어휘의 사용에 대해 표기하지 않은 경우는 표준 어휘가 적용되지 않았다고 판단한다. I3: 메타데이터는 다른 데이터 또는 메타데이터와 어떻게 연결되는가? (0~3점) 디지털 기록물은 다른 기록물 또는 메타데이터와 많이 연결되어 있을수록 맥락적 지식이 풍부해질 수 있다. 이 질문은 디지털 기록물이 다른 기록물 또는 메타데이터를 참조하고 있는지 평가한다. 예를 들어, '관련 기록물', '관련 연표'와 같은 항목이 존재하는 경우, 다른 기록물과의 연결지점을 제공하는 것으로 판단한다.
재사용성	<ul style="list-style-type: none"> R1: 데이터의 라이선스/사용 권한을 가장 잘 설명하는 것은 무엇인가? (0~4점) 이 질문은 메타데이터 항목에서 디지털 기록물에 대한 라이선스를 제공하는지 평가한다. 크리에이티브 커먼스(creative commons)나 공공누리와 같이 표준 라이선스를 사용하고 해당 라이선스를 링크로 참조하는 경우는 높은 점수를 부여한다. 그러나 디지털 아카이브 수준에서 이용약관 또는 저작권 정책만 제공하거나 텍스트 기반의 라이선스를 제공하는 경우는 낮은 점수를 부여한다. R2: 데이터의 재사용을 용이하게 하기 위해 얼마나 많은 출처 정보가 확보되었는가? (0~3점) 이 질문은 디지털 기록물의 출처 정보의 기술이 얼마나 풍부하게 기술되었는지 평가한다. 이때 풍부함의 기준은 GO FAIR의 4가지의 조건(생산자/수집자, 획득/처리 방법, 발행처/출판 정보, 수정/가공정보 기술)을 충족한 경우로 한정한다. 이 기준을 일부 충족하는 경우는 부분적 기술(partially recorded)로 평가한다.

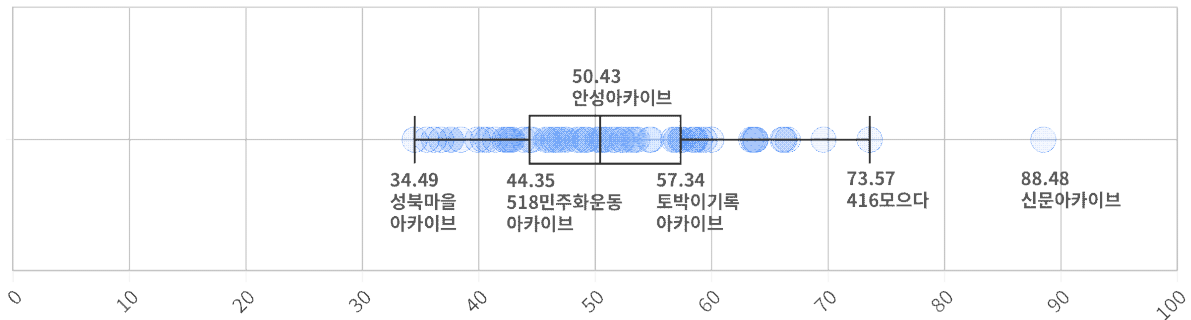
<표 2> ARDC 프레임워크의 세부항목별로 재해석한 예시

항목	ARDC 프레임워크의 설명	디지털 아카이브에 맞게 재해석한 설명
F2	개별 디지털 객체의 메타데이터에 식별자가 포함되는지 평가하는 지표이나, 기준을 명시하지 않았다.	디지털 아카이브의 개별 기록물에 대한 메타데이터에 `식별자`, `순번`, `고유번호`, `finding aid ID`, `참조코드` 등이 포함된 경우는 개별 지표를 Yes라 평가하고, 아닌 경우는 No라 평가한다.
F4	디지털 객체의 메타데이터가 저장되어 있는 형태를 질문하는 항목으로, 구글이나 구글 스칼라와 같은 검색 엔진에 색인화되는 것이 바람직하다고 설명하고 있다.	개별 기록물이 (1) 구글과 같은 범용적인 검색 엔진, (2) 공공데이터포털과 같이 일반적인 공공 레포지토리, (3) 문화공공데이터포털, 문화예술포털과 같이 특정 분야의 레포지토리/레지스트리, (4) 기관 단위의 레포지토리/레지스트리, (5) 내부 시스템에서만 검색이 가능한지 평가한다.
A1	데이터에 접근할 수 있는 방법을 평가하는 항목으로 데이터에 대한 오픈하고 등의 상황에서 데이터에 접근할 수 있는 방법을 평가한다.	디지털 아카이브에서 데이터는 디지털 기록물의 '원문'에 대응하며, 디지털 아카이브의 특성에 따라 원문의 특성이 다양할 수 있다. 예를 들어, 공연예술 분야는 해당 공연의 영상을 원문으로 인식할 수 있고, 건축 분야는 건축물의 사진을 원문이라고 할 수 있다. 따라서 디지털 아카이브의 다양한 원문에 대한 접근 제한을 기준으로 해당 지표를 평가한다.
I2	디지털 객체의 메타데이터를 표현할 때 어휘나, 온톨로지, 태깅 스키마 등을 사용했다면 어느 수준으로 사용했는지에 대해 평가하는 지표다. 표준화된 스키마의 예시는 DataCite Metadata Schema, PROV, Dublin Core 등이 있다.	디지털 아카이브 분야에서 사용되는 표준으로 Dublin Core, EAD, DACS, RiC-O 등을 고려하고, 해당 표준을 사용해 메타데이터가 기술되었는지 판단한다.
R1	디지털 객체의 라이선스가 기술되어 있는 방법을 평가하는 지표이며, 대표적으로 Creative Commons가 데이터의 재사용을 표현할 수 있는 기계가독형 라이선스다.	개별 기록물의 메타데이터 항목에 라이선스가 기술되어 있는지 평가하며, Creative Commons 또는 공공누리 라이선스로 기록물에 대한 라이선스를 제공하는지 평가한다.
R2	디지털 객체의 출처정보가 기술되어 있는 방법을 평가하는 지표로, 데이터가 어디서부터 생성되었으며 어떤 과정에 의해 생성되었는지에 대한 정보를 제공한다.	기록물에 대한 충분한 출처 정보는 GO FAIR의 R1.2 원칙을 기반으로 4가지에 해당하는 메타데이터를 포함한 경우에 가장 높은 점수를 부여한다. (1) 누가 기록물을 생산했는지(생산자, 수집자), (2) 어떻게 데이터를 처리했는지(획득방법, 처리방법), (3) 어떻게 발행되었는지(출판여부, 발행처, ISBN/ISSN), (4) 수정이나 변경사항에 대한 기술이 있는지(원본/사본 여부, 가공여부, 원본출처)가 포함되어야 한다.

4. 평가 결과

4.1 순위

디지털 아카이브의 FAIR 데이터 평가는 개별 아카이브가 탐색성, 접근성, 상호운용성, 재사용성을 종합적으로 평가한다. 평가 대상인 69개의 디지털 아카이브는 다양한 형태와 기능을 지니고 있으며, 이러한 아카이브들이 FAIR 원칙을 얼마나 충족하고 있는지 분석하였다. 그림 3은 FAIR 원칙으로 평가한 총 69개의 디지털 아카이브에 대한 평가점수를 박스플롯으로 시각화한 것이다. 평가된 디지털 아카이브의 총점은 0에서 100 사이에 분포하며, 평균 점수는 51.36점이다. 전체 아카이브 중 50%는 44.11점(Q1)에서 57.43점(Q3) 사이에 분포하고 있어 대다수가 중간 수준으로 평가되었다. 평가 대상은 최소한의 탐색성과 접근성을 갖춘 디지털 아카이브로, 개별 아이템 수준에서 식별 가능한 웹 URL을 제공하고, 특별한 권한 없이 메타데이터와 기록 원문에 접근이 가능한 아카이브들만 선별되었다. 이러한 요소가 반영되어 평균 50점을 넘는 결과가 도출되었다고 해석할 수 있다. 개별 디지털 아카이브의 F, A, I, R 점수와 총점은 공개된 깃헙에서 확인할 수 있다.



<그림 3> 디지털 아카이브의 평가 결과

디지털 아카이브의 평가는 최소 34.49점에서 최대 88.48점까지 매우 큰 편차를 보인다. 상위 25%의 점수(57.43~88.48점)는 하위 구간과 비교할 때 더욱 큰 차이가 있다. 상위 그룹에 속한 일부 디지털 아카이브가 FAIR 원칙을 상당히 잘 준수하고 있는 반면, 하위 그룹에 속한 아카이브는 이 원칙을 거의 충족하지 못하고 있음을 보여준다. 이러한 차이는 디지털 아카이브 간의 기능적, 구조적 격차가 크다는 것을 의미하며, 아카이브별로 FAIR 원칙 적용 수준의 불균형이 심각함을 보여준다. 상위 그룹의 아카이브는 탐색성, 접근성, 상호운용성, 재사용성 측면에서 뛰어난 성과를 보이는 반면, 하위 그룹은 이러한 측면에서 부족함을 보인다.

국립중앙도서관의 대한민국 신문 아카이브는 탐색성(22.05점), 접근성(20점), 상호운용성(25점), 재사용성(21.43점)에서 모두 높은 평가를 받았다. 이러한 평가는 웹 URL과 최소한의 메타데이터 제공을 넘어, 기록물의 재사용을 지원하는 다양한 기능을 포함하고 있기 때문이다. 특히, 대한민국 신문 아카이브는 MODS와 DCMI와 같은 표준 어휘를 사용해 풍부한 메타데이터를 기술하고 있으며, 원문 다운로드 기능과 기계 판독이 용이한 형식(예: XML)의 메타데이터를 제공함으로써 데이터의 상호운용성과 재사용성을 높이고 있다. 또한, 명확한 라이선스(예: nlk:license0)와 출처 정보(예: '한국언론진흥재단에서 원문 파일을 기증받아 이미지 파일로 변환하였음')를 통해 기록의 공유와 재사용을 촉진할 수 있다. '4.16모오다'(73.57점)와 '디지털장서각'(69.91점)은 공통적으로 기록에 대한 자유로운 접근과 기계가 처리하기 쉬운 구조화된 형식(예: XML, JSON)으로 기술된 메타데이터를 제공하고 있다. 이러한 특징들은 상호운용성과 재사용성 측면에서 높은 평가를 받을 수 있었던 주요 요인으로 작용한다. 간단히 요약하면, 대한민국 신문 아카이브와 같은 사례들은 디지털 아카이브의 탐색성, 접근성, 상호운

용성, 재사용성을 종합적으로 강화하는 것이 기록의 활용성과 데이터 가치 제고에 중요한지를 잘 보여준다.

하위 점수를 받은 디지털 아카이브들은 유사한 특징을 갖고 있다. ‘성북마을아카이브’(34.49점), ‘소리아카이브’(35.52점), ‘노동역사사료’(36.14점), ‘용인문화재단 디지털 아카이브’(36.77점), ‘우토로마을아카이브’(37.62점) 등 가장 낮은 평가를 받은 5개의 디지털 아카이브는 주로 기록물의 접근과 전시에 초점을 맞추고 있다. 이와 같은 아카이브는 웹 화면에서 제공되는 기록물에 대한 단순한 열람 기능을 제공하고, 텍스트 기반의 메타데이터 사용과 비구조화된 데이터 제공으로 인해 상호운용성과 재사용성이 매우 제한적이다.

특히, 표준화된 메타데이터를 사용하지 않고 자체적으로 정의한 메타데이터를 적용하는 것은 재사용에 있어 주요한 문제점으로 보인다. 표준 어휘를 사용하지 않는 메타데이터는 자동 색인화, 검색 엔진 시스템에서 기록물 탐색을 어렵게 하며, 기록물 간 상호운용성을 저해하는 중요한 요인으로 작용한다. 한편, 비구조화된 메타데이터는 기계 판독이 어렵기 때문에 기록물의 활용도와 재사용 가능성을 낮춘다. 그뿐만 아니라, 불명확한 라이선스와 출처정보의 부재는 기록물의 재사용을 제한하는 또 다른 요인으로 나타난다. 불분명한 라이선스는 기록물의 재사용 범위를 명확히 알 수 없게 하여, 사용자들이 기록을 재활용하는 데 있어 법적·윤리적 장애가 될 수 있다. 출처 정보가 부족하거나 제공되지 않는 경우, 기록물의 맥락을 파악하기 어렵게 만들어 기록의 신뢰성과 활용성을 저하시킨다.

4.2 평가지표별 특징

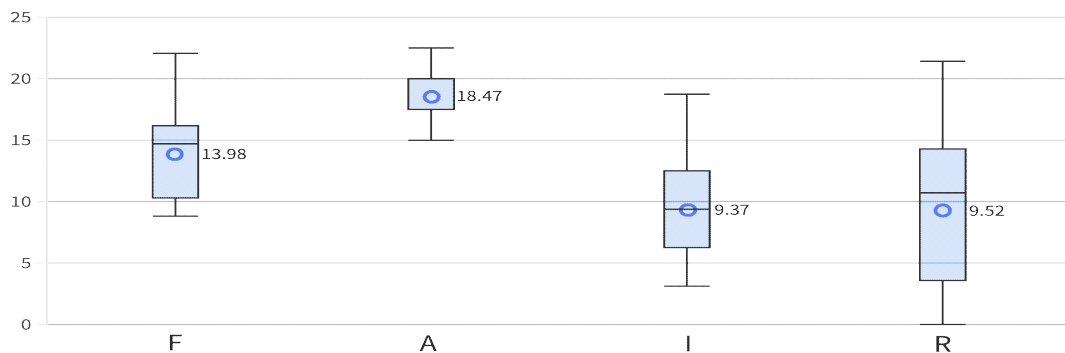
디지털 아카이브의 FAIR 데이터 원칙 평가 결과, 탐색성, 접근성, 상호운용성, 재사용성은 큰 차이가 나타났다. 그림 4에서 보듯이, F 점수는 13.98점, A 점수는 18.47점, I 점수는 9.37점, R 점수는 9.52점이다. 평가된 디지털 아카이브는 접근성과 탐색성에서 상대적으로 높은 점수를 받았지만, 상호운용성과 재사용성에서 낮은 점수를 기록했다. 평가대상의 디지털 아카이브는 개별 기록물에 대한 접근과 탐색은 원활하게 제공하고 있지만, 다양한 환경에서의 상호운용성과 기록물 재사용을 지원하는 데는 부족함이 있다.

탐색성은 상대적으로 높은 평균 점수인 13.98점을 받았고, 비교적 점수의 분포가 높게 위치해있다. 국내 디지털 아카이브에서 웹 URL은 개별 아이템을 식별하기 위한 방법으로 가장 많이 사용되고 있었다(F1). 대다수의 아카이브는 웹 URL과 함께 기관에서 개별적으로 정의한 고유 식별자를 제공하나, 웹 URL과 식별자가 서로 연관되어 있는 형태는 아니다. 이 웹 URL로 접속하면 기록물과 기록물의 메타데이터를 대부분 확인할 수 있다(F2). 대다수의 아카이브가 개별 아이템 단위의 충분한 메타데이터를 제공하지만, 기계가 찾기 어려운 텍스트 기반으로 메타데이터를 제공한다(F3). 그러나 다수의 아카이브는 구글과 같은 검색 포털의 검색 결과로 개별 기록물에 접근이 가능하다(F4). 이는 최소한의 메타데이터(예: 제목, 설명)가 검색 엔진이 색인화할 수 있는 형태(예: meta tag)로 표현된 결과다. 하지만, 메타데이터가 별도의 저장소나 레포지토리에 저장되지 않아 탐색이 제한된 경우도 많다.

접근성은 모든 항목 중에서 평균적으로 가장 높은 18.47점을 받았으며, 점수의 분포도 상당히 조밀하다. 이는 모든 디지털 아카이브에서 제한 없이 기록물과 메타데이터에 접근이 가능했기 때문이다(A1). 최소한의 접근성 기준을 준수한 디지털 아카이브만 평가한 결과이기도 하지만, 대부분의 아카이브는 기록물의 열람이나 접근에 대해 제한을 두고 있지 않다. 기록물 원문에 대한 접근은 파일 다운로드로 직접 원문을 확인할 수 있는 형태가 많다(A2). 그러나 특정 뷰어나 웹 화면에서만 열람이 가능해 원문에 접근이 제한된 경우도 우세하다. 기록물의 메타데이터는 기록 원문에 대한 접근이 불가능해도 별도의 저장소나 레지스트리에 저장한다는 응답이 많다(A3). 즉, 디지털 아카이브 서비스가 종료되어도 메타데이터에 접근할 수 있는 방안을 마련하고 있어 기록물의 장기적인 관리와 접근을 확보하고 있다.

상호운용성은 모든 지표 중에서 가장 낮은 평균 점수인 9.37점을 받았고, 그 범위도 상당히 낮게 분포되어 있다. 이 결과는 기록물과 메타데이터가 외부 기록물과 통합, 연계되기 어렵다는 사실을 의미한다. 기록 원문은 일반적으로 물리적인 원본을 디지털화한 사례가 많다. 이 경우는 기록물이 특정 뷰어 또는 특정 소프트웨어(예: 한글, 마이크로소프트 파워포인트)를 사용해야 하는 hwp, xls, ppt 등의 형식으로 제공된다(I1). 이 외에도 pdf, jpg, png와 같이 기계가 비교적 처리하기 어려운 형식으로 제공하는 사례가 많아 활용이 제한된다. 특히, 대부분의 메타데이터는 기관에서 자체적으로 정의한 메타데이터를 텍스트 기반으로 제공한다(I2). 즉, 개별 기관이 구축한 메타데이터는 다른 기관이 구축한 메타데이터와 상호호환되기 어려우며 기계가 탐색하거나 처리하기 어려운 방식으로 제공된다. 대다수는 메타데이터에서 외부 기록이나 자원과 연결된 링크를 제공하지 않아 외부 기록이나 자원과 연결되기 어렵다(I3).

재사용성은 상호운용성과 비슷하게 상대적으로 낮은 평균 점수인 9.52점을 받았으며, 가장 디지털 아카이브마 격차가 큰 항목이다. 이는 기록의 공유나 재사용에 대해 명확한 라이선스나 출처정보를 제공하는 정도가 디지털 아카이브마다 상당한 격차가 있다는 사실을 의미한다. 대부분의 디지털 아카이브는 명확한 저작권 정보를 제공하지 않는다(R1). 라이선스는 기록물의 활용 가능성을 규정하는 명시적인 정보다. 저작권에 민감한 기록물의 특성을 고려할 때 라이선스의 미제공은 기록의 재사용을 어렵게 만드는 대표적인 요인이다. 또한 대다수의 아카이브가 기록에 대한 출처 정보를 최소한으로 제공하고 있다(R2). 풍 존중의 원칙에서 강조하듯이, 출처 정보는 기록물을 완전히 이해하기 위해 기록물과 함께 관리되어야 하는 중요한 요소다. 기록물의 출처에 대한 불충분한 정보는 기록물의 맥락 파악을 어렵게 하고, 동시에 신뢰도와 활용 가치를 저하시킨다.



<그림 4> F, A, I, R 항목별 평균 점수

5. 토론

5.1 국내의 디지털 아카이브 현주소

국내 디지털 아카이브의 전반적인 현황을 살펴보면, 평가를 위한 261개의 디지털 아카이브 중 69개가 평가 대상으로 선정되었다. 즉, 디지털 아카이브라는 이름을 갖고 있지만, 많은 아카이브가 기본적인 탐색과 접근이 어렵거나, 운영되지 않는 문제를 갖고 있다. 평가 결과를 종합해 보면, 국내 디지털 아카이브의 상당수가 운영상의 기본 요건을 충족하지 못하거나, FAIR 원칙을 따르지 않고 있으며, 체계적인 개선이 필요한 상황이다.

69개 디지털 아카이브의 FAIR 데이터 원칙 준수 평가 결과, 전체 평균 점수는 50.43점으로, 디지털 아카이브들

이 전반적으로 FAIR 데이터 원칙을 충족하지 못하고 있는 것으로 나타났다. 이러한 결과는 디지털 아카이브를 구축하는 과정에서 데이터의 접근성, 상호운용성, 재사용성을 고려하지 않고, 필수적인 기술적·관리적 요소에 대한 검토가 미흡하다는 사실을 보여준다. 디지털 아카이브의 목적이 기록물의 보존과 활용에 있음에도 불구하고, 운영상 문제로 인해 실질적인 기능을 충분히 발휘하지 못하고 있다.

FAIR 데이터 평가 결과에 따르면, 국내 디지털 아카이브는 기록물의 열람과 전시에 초점을 맞추고 있어, 실질적인 기록물의 재사용과 상호운용성에 많은 한계가 존재한다. 데이터 관점에서 대부분의 기록물은 기계가 읽기 어려운 형식으로 제공되며, 기록의 재사용을 위해 별도의 처리 과정(OCR 등)이 필요하다. 특히, 특정 소프트웨어나 뷰어 프로그램을 통해 접근 가능한 기록물이 많아 다양한 환경에서 디지털 기록물의 접근성을 저해하고 있다. 메타데이터 관점에서 보면, 대부분의 디지털 아카이브가 표준 어휘를 사용하지 않고 동일한 요소를 각기 다른 용어로 표현하고 있다. 일부 아카이브는 표준 메타데이터를 적용했지만, 자체적으로 추가한 요소를 갖고 있다. 더불어, 디지털 아카이브가 메타데이터를 기계가 읽을 수 있는 형식(RDF, XML 등)이 아닌 텍스트 기반으로 제공하며, 기록물의 탐색과 활용을 제한하고 있다. 라이선스와 출처 정보의 부재 역시 디지털 아카이브의 재사용성을 저해하는 주요 요인이다. 대부분의 디지털 아카이브는 개별 아이템 단위의 라이선스를 명시하지 않고 이용약관과 같은 일반적인 정책만을 제공하는 경우가 많다. 이러한 문제는 기록물의 재사용 가능 여부를 판단하기 어렵게 만들며, 출처 정보의 미흡은 기록물의 신뢰성을 낮추는 결과를 만든다.

반면, 아르코 예술기록원의 한국예술디지털아카이브(DA-Arts)는 평가 점수에서 38.46점으로 비교적 낮은 성적을 기록했으나, 개별 기록에 대한 기술이 매우 풍부하다는 점에서 주목할 만하다. 이처럼 기록물에 대한 세부 기술이 충실한 아카이브는 평가지표의 원칙을 적절하게 적용하여 개선할 수 있는 여지가 크다. 즉, 구조적인 상호운용성이나 데이터 재사용성을 보완한다면, 디지털 아카이브로서의 가치가 크게 향상될 가능성이 있다.

5.2 디지털 아카이브의 개념 정립

디지털 아카이브는 학술적으로 통일된 개념이 부재한 상태이며 다양한 방식으로 구현되고 있다. 많은 연구가 디지털 아카이브에 대해 논의하고 있지만, 학계와 실무 영역에서 합의된 명확한 정의는 여전히 존재하지 않으며, 각기 다른 방식으로 이해되고 있다. 이러한 개념적 모호성은 기록관리와 디지털 아카이브 간의 학문적 단절을 만든다. 기록관리 분야에서 강조하는 기록의 원칙과 기록 메타데이터에 대한 기준은 디지털 아카이브에서 충분히 반영되지 않고 있다. 본 연구에서 목록으로 정리한 261개의 디지털 아카이브에서 기록관리의 이론적 틀을 적용할 수 있는 대상은 극히 적은 것이 사실이다. 이와 같은 현상은 현재의 디지털 아카이브에서 기록관리의 필수 요소를 충족하지 못한다고 해석할 수 있다. 반대로, 기록관리 분야에서 디지털 아카이브에 대한 이론적·시스템적 기반을 제공하지 못하고 있다는 비판도 가능하다.

디지털 아카이브는 필연적으로 디지털 기록을 대상으로 하며, 이런 맥락에서 디지털 아카이브는 데이터의 관점에서 이해하고 접근해야 한다. 아날로그 환경에서 진본성은 보존(recordkeeping) 과정에서 엄격한 통제에 의해 보장될 수 있었지만, 디지털 환경에서 기록은 시스템과 매체를 넘나들며 복제·변경되기 때문에 진본성과 신뢰성을 보장하기 어렵다(Pacheco et al., 2023). 이 관점에서 보존해야 할 대상은 문서가 아닌 기록에 담긴 정보(information)이며, 기록의 변화를 충실히 기록한 메타데이터가 기록의 신뢰성과 무결성을 보장할 수 있는 방안으로 강조되고 있다(Carbajal & Caswell, 2021; Pacheco et al., 2023; Zhang & Mauney, 2013). 국내의 디지털 아카이브는 게시판이나 단순한 웹페이지 형태로 운영되거나, 일부는 메타데이터 관리 시스템으로 구성되어 있다. 현재 운영되는 디지털 아카이브는 대부분 온라인에서 기록에 대한 단편적인 정보를 제공하거나, 화려한 시각화

기법을 적용하는 데 중점을 두는 경향이 있다. 이러한 방식은 사용자 경험을 개선하는 데 부분적으로 도움을 줄 수 있지만, 기록을 데이터로 관리하고 분석할 수 있는 체계적이고 구조화된 접근에는 적합하지 않다.

연구 결과에 따르면, 일부 디지털 아카이브는 기록 메타데이터를 적용하고 있다. 기록 메타데이터는 디지털 아카이브에서 기록을 탐색하고 관리하는 데 중요한 역할을 한다. 그러나 단순히 메타데이터를 적용하는 것을 넘어 메타데이터 표준을 준수하고 이를 기반으로 상호운용성을 확보하는 것이 중요하다. ISO 15489-1에 따르면, 기록 관리 시스템은 기록의 이용가능성을 보장하기 위해 타 시스템과의 상호운용성을 보장해야 한다. 그러나 현재 디지털 기록은 기록관리 시스템에 종속되어 해당 시스템을 벗어난 환경에서 기록의 상호운용과 이용가능성이 현저히 감소한다. 특히, 기록관리 시스템이 부재하거나, 연계되지 않은 상황에서 디지털 아카이브는 기록 간의 상호운용에 대한 최소한의 기준을 만족하기 어려울 수 있다. 이런 관점에서 FAIR 데이터 원칙은 기록 메타데이터의 충실한 기술, 기록 관련 표준의 적용, 메타데이터의 상호운용을 위한 가이드라인으로 적용할 수 있다.

5.3 ARDC의 FAIR 평가 지표

본 연구에서 사용한 ARDC의 FAIR 프레임워크는 평가를 위한 가이드라인과 평가지표별로 구체적인 선택 항목을 제공하는 장점이 있다. 그러나 FAIR 평가 프레임워크는 디지털 기록의 특성을 반영하기 위한 추가적인 연구가 요구된다. FAIR 데이터 원칙은 상호운용과 재사용을 위해 도메인과 관련된 표준 어휘의 사용을 강조한다. 기록관리와 문헌정보 분야에서 EAD, METS, DACS, RiC-O와 같이 기록을 기술하는 다양한 표준 어휘가 있지만, 표준에 따라 기술계층과 속성, 인코딩 방식에 차이가 있다. ICA EGAD에서 개발한 RiC-O는 기록기술을 위한 온톨로지 어휘이지만, 기존에 광범위하게 사용되는 다른 어휘의 재사용(예: PREMIS, Schema.org, DCAT)을 고려하지 않았다(Kim, 2024). 따라서, 기존의 RiC-O 어휘로 기술된 기록과의 상호운용이 어렵고, 외부 어휘와의 연계를 제한한다. ARDC의 평가 지표는 메타데이터 표준의 적용 여부를 판단하지만, 개별 표준에서 규정한 기록 메타데이터의 적합성은 판단하지 않고 있다. 따라서, 디지털 아카이브를 위한 FAIR 데이터 프레임워크는 기록관리의 이론적 배경과 선행 연구를 검토하여 평가 지표를 확대하거나 상세화시킬 필요가 있다. 본 연구에서 제안한 프레임워크는 디지털 아카이브의 평가를 위해 재해석한 부분이 있으며, 개별 지표에 대한 논의와 개선사항은 다음과 같다.

- 탐색성: ARDC의 평가 프레임워크는 식별자와 메타데이터 기술을 평가하는 항목을 포함하고 있지만 기록 아이템이 평가대상이다. 일반적으로 기록물은 문서군(fonds), 시리즈(series), 파일(file), 아이템(item)과 같은 계층적 구조로 구성되어 있다(ISAD(G), 2000). 대부분의 디지털 아카이브는 계층적 구조로 표현된 기록물을 디지털로 변환하거나, 서울 사진 아카이브(서울기록원)와 같이 계층구조를 표현한 디지털 아카이브도 존재한다. 따라서 디지털 아카이브를 평가하는 과정에서 아이템의 상위 단위를 식별하고 메타데이터를 함께 평가하는 방안을 고려해야 한다. 한편, F4는 디지털 아카이브의 메타데이터를 저장한 레포지토리를 평가하며, 공공 레포지토리(generalist public repository), 도메인에 특화된 레포지토리(domain specific repository) 등으로 구성되어 있다. 본 연구는 구글, 공공데이터포털, 문화공공데이터광장 등 레포지토리의 범위를 확대했지만, 향후 일관성 있는 평가를 위해 디지털 아카이브 도메인의 레포지토리 또는 레지스트리에 대한 구체적인 기준을 정의해야 한다.
- 접근성: 접근성의 A1, A2는 데이터의 접근 방식에 대한 평가이지만, 사실상 데이터의 획득에 대한 평가를 함께 포함한다. FAIR 데이터 원칙의 A1은 민감한 데이터(sensitive data)에 대해 완전한 접근을 허용하는 것이 어렵기 때문에 데이터의 접근 방법을 안내하는 연락처의 명시를 강조하고 있다. 그러나 A1, A2는 웹상에서 데이터를 열람할 수 없거나 메타데이터만 접근할 수 있는 경우, 하위 점수를 부여해야 한다. 특히, 저작권상

의 이유로 온전히 웹상에 디지털 기록물을 개방하기 어려운 사례가 존재하는 것을 고려하면 접근성의 평가는 데이터 획득이 아닌 데이터의 접근방식에 대한 평가로 수정될 필요가 있다. 접근성의 A2는 웹 서비스 API(web service API)에 대한 표준과 비표준의 기준을 사용했다. 그러나 웹 서비스 API의 표준과 비표준의 기준은 모호하기 때문에 웹상의 접근 방법에 대해 표준 프로토콜(예: tcp, http(s))로 접근가능한지 평가하는 방안으로 수정되어야 한다.

- ◆ 상호운용성: 상호운용성의 I1은 데이터 형식을 평가할 때 ‘구조적, 개방 표준, 기계가 읽을 수 있는 형식(structured, open standard, machine-readable format)’에 가장 높은 점수를 부여한다. 이 항목의 예는 PDF/A, HTML, JPEG 2000, MP3, CSV 가 있다. 그러나 PDF/A는 구조적인 데이터 형태에 포함되지 않아 그 기준이 모호하다. 한편, 디지털 아카이브는 이미지, 음성, 영상 등의 다양한 형식을 다루고 온라인 스트리밍 서비스(예: 유튜브)와 같이 새로운 형식의 매체를 포함하는 방향으로 변화하고 있다. 따라서 디지털 아카이브의 형식을 평가할 때 이와 같은 특징에 대한 모호함이 있기 때문에 상호운용성을 위한 적절한 데이터 형식을 구체적으로 정의해야 한다. 더불어, 메타데이터에 대한 상호운용성은 표준 어휘의 사용 여부와 함께 품질에 대한 평가도 이루어져야 한다. I2는 실질적으로 표준 어휘의 사용 여부만 평가하고, 기계로 읽을 수 있는 메타데이터가 충실히 기술되었는지 평가하지 않는다. 이러한 기준은 국내 디지털 아카이브의 평가에 있어서 한계점이 될 수 있다. 예를 들어, 국내 디지털 아카이브는 하나의 아이템에 대한 메타데이터 레코드를 다양한 형식으로 함께 제공하는 경우가 존재한다. 웹 페이지의 텍스트로 기술된 메타데이터는 자세하지만, 기계가 읽을 수 있는 형식으로 제공하는 메타데이터는 상당부분 생략되어 최소한으로 제공하는 경우가 많기 때문에 메타데이터의 품질이 좋지 않다. 따라서 기계가 읽을 수 있는 메타데이터의 품질도 함께 평가될 수 있도록 검토해야 한다.
- ◆ 재사용성: R2는 출처정보가 충분히 기술되었는지 평가하지만, ‘충분하다’는 기준이 명확하지 않다. 본 연구는 GO FAIR의 R1을 참조하여 4가지 기준을 제시하고, 이를 기준으로 판단했다. 그러나, 기록관리 측면에서 출처 정보는 기록물의 이해에 중요한 맥락정보라는 것을 고려할 때 기록관리의 특성을 충분히 반영할 수 있는 평가 기준이 마련되어야 한다. 특히, 기록의 특성(예: 아날로그 기록의 디지털화, 디지털로 만들어진 기록)에 따라 출처 정보를 평가하는 방식에 차이가 존재할 수 있다는 것을 고려해야 한다.

5.4 국내의 디지털 아카이브 개선을 위한 고려사항

국내 디지털 아카이브의 발전을 위해 기록물의 특성, 데이터 관리, 시스템 구축 등 다양한 관점에서 체계적인 개선이 요구된다. 이를 위해 세 가지 고려사항을 검토해야 한다.

첫째, 디지털 아카이브의 구체적인 가이드라인 정립과 공식적인 평가 방법론 설계가 필요하다. 디지털 아카이브는 기록, 데이터, 시스템 관점에서 기록물의 보존과 관리, 데이터의 구조화, 시스템적 연계성을 강화하는 방향으로 가이드라인을 수립하는 것이 중요하다. 동시에 디지털 아카이브를 객관적으로 평가할 수 있는 공식적인 평가 방법론을 설계하고 적용해야 한다. 이런 맥락에서 디지털 아카이브에 대한 학계와 실무 차원의 공론화와 개선 방향성을 논의해야 한다.

둘째, 기록관리, 디지털 아카이브를 담당하는 기관의 제한된 자원과 실행의 어려움을 고려한 개선 방안이 요구된다. 현재 국내 기록관리, 디지털 아카이브 관련 기관은 한정된 인력과 예산으로 운영되며, 개별 기관이 독립적으로 디지털 아카이브를 효과적으로 실행하는 데 한계가 있다. 이를 극복하기 위해 기관과 유관 전문가의 협력 체계 구성이 중요하며, 기록의 대상과 주제별 메타데이터 표준을 정립하는 노력이 필요하다. 국가 차원의 표준을 처음부

터 완벽하게 적용하는 대신, 유관 전문가들이 현장에서 실질적으로 적용할 수 있는 실용적 방안을 모색하는 접근이 현실적이다. 이러한 접근은 제한된 자원을 가진 기관들이 보다 효과적으로 디지털 아카이브를 구현하고 운영할 수 있도록 지원할 수 있다.

셋째, FAIR 데이터 원칙의 이해와 디지털 아카이브 적용 방안 논의가 필요하다. 디지털 아카이브가 단순한 기록 보존을 넘어 데이터 관리와 상호운용성을 고려한 시스템으로 발전하기 위해 FAIR 데이터 원칙에 대한 이해가 필수적이다. 디지털 아카이브와 기록관리 관점에서 FAIR 데이터 원칙의 적용하기 위해 평가 대상과 방법, 평가지표에 대한 명확한 논의와 일관성 있는 평가 지표의 수립이 요구된다.

6. 결론

본 연구는 국내 디지털 아카이브의 FAIR 데이터 원칙 준수 여부를 평가하고, 개선 방향을 모색하는 데 목적이 있었다. 69개를 대상으로 평가를 수행한 결과, 국내 디지털 아카이브는 전반적으로 FAIR 원칙을 충분히 준수하지 못하고 있다. FAIR의 평균 점수는 50.43점으로, 디지털 아카이브의 구축 과정에서 FAIR 데이터 원칙에 포함된 요소들이 고려되지 않고 있다. 조사대상의 디지털 아카이브는 형태와 기능에 있어 통일성이 부족하며, 데이터 관점에서 디지털 기록을 제공하지 않고 있다. 디지털 아카이브에 대한 구조적이고 체계적인 접근이 미흡하고, 기록관리의 이론적 틀, 기록 메타데이터의 적용은 극소수에 한정되어 있다. 종합적으로 보면, 다수의 아카이브에서 탐색과 접근이 어려워지거나 운영되지 않는 문제가 있었고, 상호운용성과 재사용성을 보장하기 위한 기술적·관리적 요소가 부족했다.

본 연구의 공헌점은 세 가지로 요약할 수 있다. 첫째, 국내 디지털 아카이브의 실질적인 운영 실태를 객관적으로 분석하고, 이를 바탕으로 FAIR 데이터 원칙을 적용하여 문제점을 진단했다. 둘째, 디지털 아카이브는 단순히 기록의 보존을 넘어서 데이터 상호운용성과 재사용성을 보장해야 함을 강조하고, 구체적인 개선 방향을 제시했다. 셋째, FAIR 원칙을 기반으로 디지털 아카이브의 각 요소를 정량적으로 평가하는 새로운 방법론을 제시하여, 향후 아카이브 평가 연구에 중요한 기초 자료를 제공하였다.

그러나 본 연구는 몇 가지 한계를 갖고 있다. 첫째, 평가 대상은 보다 객관적인 접근이 필요하다. 디지털 아카이브에 대한 개념적 모호성과 더불어 공공과 민간 디지털 아카이브의 목록은 존재하지 않는다. 평가대상으로 선정된 261개 디지털 아카이브는 상향식으로 정리되어 누락이 존재할 수 있다. 따라서 본 연구가 국내 디지털 아카이브의 실태를 완벽히 반영하는 데 부족할 수 있다. 둘째, FAIR 데이터 원칙은 디지털 자원에 적용하기 위한 도구이지만, 디지털 기록물의 특성을 완벽히 반영하지 못하는 측면이 있다. 디지털 아카이브는 기록관리, 웹 공학, 컴퓨터공학의 학제간 협력의 결과일 수 있다. 따라서 FAIR 데이터 평가 프레임워크는 평가의 목적과 더불어 진단과 개선을 안내할 수 있는 방향으로 보완될 필요가 있다.

향후 연구는 기록관리 중심의 디지털 아카이브 개념 정립에 초점을 맞출 필요가 있다. 디지털 아카이브는 기록물 보존과 열람을 넘어 데이터의 상호운용성과 재사용성을 강화해야 하며, 이를 위해 기록관리 이론을 바탕으로 체계적인 개념 정립이 필요하다. 한편, 디지털 아카이브와 FAIR 데이터 원칙에 대한 심도 있는 고찰이 필요하며, 기록관리의 이론을 기반으로 FAIR 데이터 평가지표를 세분화하는 연구가 요구된다. 이를 통해 디지털 아카이브의 다양한 요소를 보다 구체적으로 평가하고 개선할 수 있을 것이다. 더불어, 국내 디지털 아카이브의 특징, 기능, 시스템에 대한 종합적인 목록을 구축하는 작업이 필수적이다. 국내 디지털 아카이브의 현황을 명확히 파악하고,

아카이브 간의 차이점을 분석함으로써 디지털 기록물 관리의 일관성과 통일성을 강화할 수 있는 방안을 도출할 수 있을 것이다. 이러한 연구는 디지털 아카이브의 상호운용성 강화와 재사용성 향상을 위한 기반을 마련하고, 국내 디지털 아카이브가 국제적 기준에 부합하는 방향으로 발전하는 데 기여할 것이다.

참고문헌

- 강민정, 장우권 (2021). OAIS 참조모형 기반 민중가요 메타데이터 설계에 관한 연구. 한국기록관리학회지, 21(1), 211-230. <https://doi.org/10.14404/JKSARM.2021.21.1.211>
- 김명훈 (2024). 디지털 시대 기록의 폐기에 대한 재조명: 정보관리 기능의 발전 및 확대와 관련하여. 한국기록관리학회지, 24(1), 73-88. <https://doi.org/10.14404/JKSARM.2024.24.1.073>
- 김유승 (2019). 디지털시대의 공공기록평가에 관한 정책적 고찰 : 영국 TNA 사례를 중심으로. 기록학연구, 62, 5-39. <https://doi.org/10.20923/KJAS.2019.62.005>
- 대한불교조계종 (2020). 한국전통문화 사진 디지털 아카이브. 출처: <http://www.ktda.kr/main.do?language=ko>
- 도슬기, 박희진 (2023). 국채보상운동 디지털 아카이브 기록물의 활용을 위한 위키데이터 연계 방안에 대한 연구. 한국기록관리학회지, 23(2), 95-115. <https://doi.org/10.14404/JKSARM.2023.23.2.095>
- 박순철, 함한희, 이정승, 정성미 (2018). (인문학자를 위한) 디지털 아카이브즈. 서울: 민속원.
- 박태연, 신동희 (2016). 참여형 아카이브 구축을 위한 오픈소스 소프트웨어 평가 - 노근리디지털아카이브 구축을 위한 예비분석 -. 한국기록관리학회지, 16(1), 121-150. <https://doi.org/10.14404/JKSARM.2016.16.1.121>
- 박희진 (2020). 디지털 보존을 위한 PREMIS 3.0 기반 메타데이터에 관한 연구. 한국기록관리학회지, 20(4), 185-203. <https://doi.org/10.14404/JKSARM.2020.20.4.185>
- 서울시 도시건축위원회 (발행년불명). 서울시 도시건축위원회 아카이브. 출처: <https://commission.eseoul.go.kr/subMenu.do?menuId=fnArchiveDoc¶m=ADOC-0005>
- 설문원 (2005). 과학기술 분야 디지털 아카이빙을 위한 정책연구. 정보관리연구, 36(4), 23-49. <https://doi.org/10.1633/JIM.2005.36.4.023>
- 아카이브센터 (발행년불명). Digital Archiving for Digital Archive. 출처: <https://archivecenter.net/dada>
- 이상호 (2023). MOW 디지털 관련 권고의 한계와 '디지털 기록유산' 정책 방향. 국학연구, 52, 467-508.
- 이소연 (2002). 디지털 아카이빙의 표준화 OAIS 참조모형. 정보관리연구, 33(3), 45-68. <https://doi.org/10.1633/JIM.2002.33.3.045>
- 이소연 (2008). 믿을 수 있는 디지털 아카이브 인증기준: OASIS 적용사례. 정보관리학회지, 25(3), 5-25. <https://doi.org/10.3743/KOSIM.2008.25.3.005>
- 이승역, 설문원 (2021). 디지털 정보기술 환경에서 보존기록 평가론의 전환. 기록학연구, 67, 57-97. <https://doi.org/10.20923/KJAS.2021.67.057>
- 전경선, 장윤금 (2018). 디지털 아카이브 구축을 위한 공개 소프트웨어 사용성 평가. 한국문헌정보학회지, 52(3), 247-271. <https://doi.org/10.4275/KSLIS.2018.52.3.247>
- 정성미 (2023). 무형유산의 다층적 활용을 위한 디지털 아카이브 발전방안. 문화와융합, 45(12), 133-147. <https://doi.org/10.33645/cnc.2023.12.45.12.133>
- 정주영 (2021). 남산예술센터 시맨틱 디지털 아카이브 구축에 관한 연구. 한국연극학, 77, 211-248. <https://doi.org/10.18396/KTSA.2021.1.77.006>
- 정희선, 김희순, 송현숙, 이명희 (2016). 개화기 조선 채류 서양인 기록물의 디지털 아카이브 시스템 구축. 한국비블리아학

- 회지, 27(4), 229-249. <https://doi.org/10.14699/kbiblia.2016.27.4.229>
- 하은미, 설보연, 김현진 (2022). 디지털 아카이브의 온라인 서비스 개선방안 연구: 지역 아카이브의 검색서비스 평가요소 개발을 중심으로. 한국기록관리학회지, 22(3), 83-102. <https://doi.org/10.14404/JKSARM.2022.22.3.083>
- 한국기록공보 (발행년불명). 한국기록공보 디지털 아카이브. 출처: <https://www.archives.or.kr/>
- 한국산업표준 (2020). 기록의 디지털화 이행 지침 표준 (KS X ISO TR 13028:2020).
- 한상은, 박희진 (2022). 디지털 아카이브의 위키데이터 활용방안 연구. 한국기록관리학회지, 22(1), 201-217. <https://doi.org/10.14404/JKSARM.2022.22.1.201>
- HIKE (2024). FAIR-evaluation. 출처: <https://github.com/hike-lab/FAIR-evaluation>
- HIKE (2024). 부록_FAIR 원칙 세부 평가 기준. 출처: https://github.com/hike-lab/FAIR-evaluation/blob/main/%EB%B6%80%EB%A1%9D_FAIR%20%EC%9B%90%EC%B9%99%20%EC%84%B8%EB%B6%80%20%ED%8F%89%EA%B0%80%20%EA%B8%B0%EC%A4%80.pdf
- Abu, K. S. (2021). Bibliometric Review of Digital Archive Research: Contemporary Status, Research Hotspots and Future Trends. *Library Philosophy and Practice*, 1-10.
- Ammar, A., Bonaretti, S., Winkers, L., Quik, J., Bakker, M., Maier, D., Lynch, I., Rijn, J., & Willighagen, E. (2020). A semi-automated workflow for FAIR maturity indicators in the life sciences. *Nanomaterials*, 10(10), 1-14. <https://doi.org/10.3390/nano1010206>
- ARDC (2020). FAIR Data Self-Assessment Tool. Available: <https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>
- Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K., & Stall, S. (2020). The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal*, 19, 1-7. <https://doi.org/10.5334/dsj-2020-041>
- Barbuti, N. (2021). Thinking digital libraries for preservation as digital cultural heritage: By R to R4 facet of FAIR principles. *International Journal on Digital Libraries*, 22, 309-318. <https://doi.org/10.1007/s00799-020-00291-7>
- Basuki, A., Churiyah, M., Dharma, B. A., & Sakdiyyah, D. A. (2022). Archive Management System as Complementary Learning Media in the Digital Era. *International Journal Of Humanities Education and Social Sciences*, 2(2), 493-505. <https://doi.org/10.55227/ijhess.v2i2.260>
- Calamai, S. & Frontini, F. (2018). FAIR data principles and their application to speech and oral archives. *Journal of new music research*, 47(4), 339-354. <https://doi.org/10.1080/09298215.2018.1473449>
- Caplan, P. (2009). *Understanding PREMIS*. Washington, D.C.: Library of Congress.
- Caplan, P. (2010). The Florida Digital Archive and DAITSS: a model for digital preservation. *Library Hi Tech*, 28(2), 224-234. <https://doi.org/10.1108/07378831011047631>
- Carbajal, I. A. & Caswell, M. (2021). Critical Digital Archives: A Review from Archival Studies. *The American Historical Review*, 126(3), 1102-1120. <https://doi.org/10.1093/ahr/rhab359>
- Corpas, M., Kovalevskaya, N. V., McMurray, A., & Nielsen, F. G. (2018). A FAIR guide for data providers to maximise sharing of human genomic data. *PLOS Computational Biology*, 14(3), 1-10. <https://doi.org/10.1371/journal.pcbi.1005873>
- DANS (n.d.). FAIRdat: FAIR data assessment tool. Available: <https://www.surveymonkey.com/r/fairdat>
- David, R., Mabile, L., & Cambon-Thomsen, A. (2019). RDA-SHARC fairness assessment tools for

- crediting/rewarding scientists data sharing activities. Zenodo, 1–3. <https://doi.org/10.5281/zenodo.2575734>
- Devaraju, A., Mokrane, M., Cepinskas, L., Huber, R., Herterich, P., de Vries, J., Akerman, V., L'Hours, H., Davidson, J., & Diepenbroek, M. (2021). From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. *Data Science Journal*, 20, 1–14. <https://doi.org/10.5334/dsj-2021-004>
- Dobreski, B., Moulaison-Sandy, H., & Bishop, B. W. (2022). How FAIR is MARC?: FAIR Data Principles applied to a bibliographic data standard. *School of Information Sciences — Faculty Publications and Other Works*, 1–6.
- FAIRplus (2024). FAIR DataSet Maturity (FAIR-DSM) Assessment Tool. Available: <https://fairdsm.biospeak.solutions/assess>
- Force, D. C. & Smith, R. (2021). Context Lost: Digital Surrogates, Their Physical Counterparts, and the Metadata that Is Keeping Them Apart. *The American Archivist*, 84(1), 91–118. <https://doi.org/10.17723/0360-9081-84.1.91>
- FORCE11 (n.d.a). R1.2: (Meta)data are associated with detailed provenance. Available: <https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/>
- FORCE11 (n.d.b). The Fair Data Principles. GO FAIR. Available: <https://www.go-fair.org/fair-principles/>
- Gagnard, A., Rosnet, T., De Lamotte, F., Lefort, V., & Devignes, M. D. (2023). FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards. *Journal of Biomedical Semantics*, 14, 1–14. <https://doi.org/10.1186/s13326-023-00289-5>
- Hawkins, A. (2022). Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. *Archival Science*, 22(3), 319–344. <https://doi.org/10.1007/s10502-021-09381-0>
- Hedstrom, M. (1997). Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, 31, 189–202. <https://doi.org/10.1023/A:1000676723815>
- International Council on Archives. (2000). ISAD(G): General International Standard Archival Description (2nd ed.) https://www.ica.org/app/uploads/2024/01/CBPS_2000_Guidelines_ISADG_Second-edition_EN.pdf
- International Organization for Standardization. (2016). Information and documentation — Records management Part 1: Concepts and principles (2nd ed.) (ISO 15489-1:2016). <https://www.iso.org/obp/ui/en/#iso:std:iso:15489:-1:ed-2:v1:en>
- Jensen, H. S. (2021). Digital archival literacy for (all) historians. *Media History*, 27(2), 251–265. <https://doi.org/10.1080/13688804.2020.1779047>
- Kilchenmann, A., Laurens, F., & Rosenthaler, L. (2019). Digitizing, archiving... and then? Ideas about the usability of a digital archive. *Archiving Conference*, 16, 146–150. <https://doi.org/10.2352/issn.2168-3204.2019.1.0.34>
- Kim, B., Nakamura, S., & Watanave, H. (2022). Using Archivematica and Omeka S for Long-Term Preservation and Access of Digitized Archive Materials. From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries, 241–250. http://doi.org/10.1007/978-3-031-21756-2_20
- Kim, H. (2024). A Knowledge Graph of Interlinking Digital Records: the Case of the 1997 Korean Financial Crisis. *The Electronic Library*, 42(1), 60–77. <https://doi.org/10.1108/EL-05-2023-0131>
- Klareld, A. S. & Gidlund, K. L. (2017). Rethinking archives as digital: The consequences of "paper minds" in illustrations and definitions of E-archives. *Archivaria*, 83, 81–108.
- Krans, N. A., Ammar, A., Nymark, P., Willighagen, E. L., Bakker, M. I., & Quik, J. T. K. (2022). FAIR assessment tools: evaluating use and performance. *NanoImpact*, 27, 1–10. <https://doi.org/10.1016/j.impact.2022.100402>

- Laermans, R. & Gielen, P. (2007). The archive of the digital an-archive. *Image & narrative*, 17, 3-14.
- Lee, C. A. (2009). Open archival information system (OAIS) reference model. *Encyclopedia of library and information Sciences*, 3, 4020-4030. <https://doi.org/10.1081/E-ELIS3-120044377>
- Lia, M. & Colella, D. D. (2023). CkanFAIR: A digital tool for assessing the FAIR principles. 2023 IEEE International Conference on Big Data, 3980-3984. <https://doi.org/10.1109/BigData59044.2023.10386850>
- Matusiak, K. K. (2022). Evaluating a digital community archive from the user perspective: the case of formative multifaceted evaluation. *Library & Information Science Research*, 44(3), 1-8. <https://doi.org/10.1016/j.lisr.2022.101159>
- Nicholson, C., Kansa, S., Gupta, N., & Fernandez, R. (2023). Will It Ever Be FAIR?: Making Archaeological Data Findable, Accessible, Interoperable, and Reusable. *Advances in Archaeological Practice*, 11(1), 63-75. <https://doi.org/10.1017/aap.2022.40>
- Pacheco, A. P., Freitas, M. C. V., & Silva, C. G. (2022). The FAIR Accessor and the authenticity of digital archival information. *Transinformação*, 34, 1-12. <https://doi.org/10.1590/2318-0889202234e210006>
- Pacheco, A., Silva, C. G., & Freitas, M. C. V. (2023). A metadata model for authenticity in digital archival descriptions. *Archival Science*, 23, 629-673. <https://doi.org/10.1007/s10502-023-09422-w>
- Padia, K., AlNoamany, Y., & Weigle, M. C. (2012). Visualizing digital collections at archive-it. *JCDL '12: Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, 15-18. <https://doi.org/10.1145/2232817.2232821>
- Park, Haram & Kim, Haklae (2023). Japanese Military "Comfort Women" Knowledge Graph: Linking Fragmented Digital Records. *Information Technology and Libraries*, 42(1), 1-16. <https://doi.org/10.6017/ital.v42i1.15799>
- Power, C., Lewis, A., Petrie, H., Green, K., Richards, J., Eramian, M., Chan, B., Walia, E., Sijaranamual, I., & Rijke, M. D. (2017). Improving archaeologists' online archive experiences through user-centred design. *Journal on Computing and Cultural Heritage*, 10(1), 1-20. <https://doi.org/10.1145/2983917>
- Riasetiawan, M., & Mahmood, A. K. (2010). DALA Project: Digital archive system for long term access. *IEEE*, 1-5.
- Sundaram, S. S. & Musen, M. A. (2023). Making Metadata More FAIR Using Large Language Models. *DaMaLOS 2023*, 1-14. <https://doi.org/10.4126/FRL01-006444995>
- Villanueva, E. & Shiri, A. (2021). Methodological Diversity in the Evaluation of Cultural Heritage Digital Libraries and Archives: An Analysis of Frameworks and Methods. *Canadian Journal of Information and Library Science*, 43(3), 316-342.
- Welter, D., Juty, N., Rocca-Serra, P., Xu, F., Henderson, D., Gu, W., Strubel, J., Giessmann, R. T., Emam, I., Gadiya, Y., Abbassi-Daloi, T., Alharbi, E., Gray, A. J. G., Courtot, M., Gribbon, P., Ioannidis, V., Reilly, D. S., Lynch, N., Boiten, J. W., Satagopam, V., Goble, C., Sansone, S. A., & Burdett, T. (2023). FAIR in action—a flexible framework to guide FAIRification. *Scientific data*, 10, 1-9. <https://doi.org/10.1038/s41597-023-02167-2>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Santos, L. B. S., Bourne, P., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C., Finkers, R., Gonzalez-Beltran, A., Gray, A., Groth, P., Goble, C., Grethe, J., Heringa, J., Hoen, P., Hooft, R., Khun, T., Kok, R., Kok, J., Lusher, S., Martone, M. E., Mons, A., Packer, A., Persson, B., Rocca-Serra, P., Roos, M., Schaik, R., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M., Thompson, M., Lei, J., Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhou, J., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3,

1-9. <https://doi.org/10.1038/sdata.2016.18>

Wilkinson, M. D., Dumontier, M., Sansone, S.-A., Bonino da Silva Santos, L. O., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M., & Schultes, E. (2019). Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data*, 6, 1-12.

<https://doi.org/10.1038/s41597-019-0184-5>

Yadav, D. (2016). Opportunities and challenges in creating digital archive and preservation: an overview. *International Journal of Digital Library Services*, 6(2), 63-73.

Zhang, J., & Mauney, D. (2013). When archival description meets digital object metadata: a typological study of digital archival representation. *The American Archivist*, 76(1), 174-195.

<https://doi.org/10.17723/aarc.76.1.121u85342062w155>

• 국문 참고자료의 영어 표기

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

Archive Center (n.d.). Digital Archiving for Digital Archive. Available: <https://archivecenter.net/dada>

Chung, Heesun, Kim, Heesoon, Song, Hyun-Sook, & Lee, Myeong-Hee (2016). Construction of the Digital Archive System from the Records of Westerners Who Stayed in Korea during the Enlightenment Period of Chosun. *Journal of the Korean Biblia Society for Library and Information Science*, 27(4), 229-249.

<https://doi.org/10.14699/kbiblia.2016.27.4.229>

Do, Seulki & Park, Heejin (2023). A study on Wikidata linkage methods for utilization of digital archive records of the National Debt Redemption Movement. *Journal of Korean Society of Archives and Records Management*, 23(2), 95-115. <https://doi.org/10.14404/JKSARM.2023.23.2.095>

Ha, Eunmi, Seol, Boyeon, & Kim, Hyeonjin (2022). A Study on Improvements for Digital Archives Online Services: Focusing on the Development of Search Service Evaluation Factors in Local Digital Archives. *Journal of Korean Society of Archives and Records Management*, 22(3), 83-102.

<https://doi.org/10.14404/JKSARM.2022.22.3.083>

Han, Sang Eun & 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>

HIKE (2024). FAIR-evaluation. Available: <https://github.com/hike-lab/FAIR-evaluation>

HIKE (2024). Supplement_FAIR Principles Detailed Evaluation Criteria. Available:

https://github.com/hike-lab/FAIR-evaluation/blob/main/%EB%B6%80%EB%A1%9D_FAIR%20%EC%9B%90%EC%B9%99%20%EC%84%B8%EB%B6%80%20%ED%8F%89%EA%B0%80%20%EA%B8%B0%EC%A4%80.pdf

Jeon, Kyungsun & Chang, Yunkeum (2018). Usability Testing of Open Source Software for Digital Archiving. *Journal of the Korean Society for Library and Information Science*, 52(3), 247-271.

<https://doi.org/10.4275/KSLIS.2018.52.3.247>

Jeong, Seongmi (2023). For Advancing Digital Archives in Multifaceted Utilization of Intangible Cultural Heritage. *Culture and Convergence*, 45(12), 133-147. <https://doi.org/10.33645/cnc.2023.12.45.12.133>

Jung, Joo Young (2021). A Study on Establishment of Semantic Digital Archive of the Namsan Arts Center. *Journal of Korean Theatre Studies Association*, 77, 211-248. <https://doi.org/10.18396/ktsa.2021.1.77.006>

- Kang, Minjeong & Chang, Woo Kwon (2021). Study on Design of Protest Song Metadata based on OAIS Reference Model. *Journal of Korean Society of Archives and Records Management*, 21(1), 211-230.
<https://doi.org/10.14404/JKSARM.2021.21.1.211>
- Kim, Myung-hun (2024). Reconsideration on the Disposal of Records in Digital Era: Regarding the Development and Expansion of Information Management Functions. *Journal of Korean Society of Archives and Records Management*, 24(1), 73-88. <https://doi.org/10.14404/JKSARM.2024.24.1.073>
- Kim, You-seung (2019). A Study on Public Records Appraisal Policies in the Digital Age : Based on the Case of The National Archives in the UK. *The Korean Journal of Archival Studies*, 62, 5-39.
<https://doi.org/10.20923/KJAS.2019.62.005>
- Korean Industrial Standards (2020). Information and documentation – Implementation guidelines for digitisation of records (KS X ISO TR 13028:2020).
- Lee, Sangho (2023). Limitations of MOW digital-related recommendations and ‘digital documentary heritage’ policy direction. *Korean Studies*, 52, 467-508.
- Lee, Seung-eok & Seol, Moon-won (2021). Transforming Archival Appraisal in Digital Environments. *The Korean Journal of Archival Studies*, 67, 57-97. <https://doi.org/10.20923/KJAS.2021.67.057>
- Lee, Soyeon (2002). Standardization of Digital Archiving and OAIS Reference Model. *Journal of information management*, 33(3), 45-68. <https://doi.org/10.1633/JIM.2002.33.3.045>
- Lee, Soyeon (2008). Trustworthy Repositories Audit Criteria: Self-Assessment of OASIS. *Journal of the Korean Society for Information Management*, 25(3), 5-25. <https://doi.org/10.3743/KOSIM.2008.25.3.005>
- Park, HeeJin (2020) A Study of Metadata for Digital Preservation Based on PREMIS 3.0. *Journal of Korean Society of Archives and Records Management*, 20(4), 185-203. <https://doi.org/10.14404/JKSARM.2020.20.4.185>
- Park, Sunchul, Ham, Hanhee, Lee, Jeongsong, & Jung, Seongmi (2018). *Digital Archives*. Seoul: Minsokwon.
- Park, Tae-yeon & Shin, Donghee (2016). Evaluation of Open-source Software for Participatory Digital Archives: Understanding System Requirements for No Gun Ri Digital Archives. *Journal of Korean Society of Archives and Records Management*, 16(1), 121-150. <https://doi.org/10.14404/JKSARM.2016.16.1.121>
- Seol, Moon-Won (2005). A Study of Digital Archiving Policies for National S&T Information Resources. *Journal of Information Management*, 36(4), 23-49. <https://doi.org/10.1633/JIM.2005.36.4.023>
- Seoul Metropolitan City Construction Committee (n.d.). Archive of Seoul Metropolitan City Construction Committee. Available: <https://commission.eseoul.go.kr/subMenu.do?menuId=fnArchiveDoc¶m=ADOC-0005>
- The Gazette of the Presbyterian Church of Korea (n.d.). Digital Archive of The Gazette of Presbyterian Church of Korea. Available: <https://www.archives.or.kr/>
- The Jogye Order of Korean Buddhism (2020). Digital Archive of Korean Traditional Culture Photography. Available: <http://www.ktda.kr/main.do?language=ko>