

디지털 기록의 상호운용을 위한 지식그래프의 평가*

Evaluation of Knowledge Graph for Interoperating Digital Records*

박하람(Haram Park)¹, 김학래(Haklae Kim)²

E-mail: haram9553@gmail.com, haklaekim@cau.ac.kr

¹제1저자 중앙대학교 일반대학원 문헌정보학과 문헌정보학전공 박사과정

²교신저자 중앙대학교 사회과학대학 문헌정보학과 교수



논문접수 2023-10-17

최초심사 2023-10-24

게재확정 2023-11-20

ORCID

Haram Park

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

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 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임

초 록

디지털 아카이브는 지속적으로 보존할 가치가 있는 디지털 기록을 보존하고 활용하기 위한 온라인 플랫폼이다. 그러나 국내에서 운영되고 있는 디지털 아카이브는 기능, 메타데이터, 데이터의 기술원칙과 관련된 공통 원칙이 존재하지 않는다. 이는 분산적으로 존재하는 디지털 기록을 연계하기 힘들게 만드는 요인이 된다. 본 연구는 디지털 기록의 상호운용을 개선하기 위한 방안으로 디지털 아카이브를 위한 공통 어휘를 제안하고, 공통 어휘로 구축된 디지털 아카이브의 상호운용성을 평가한다. 1997 외환위기 아카이브의 데이터를 수집 분석하여 지식그래프를 구축하고, RiC-O로 구축된 지식그래프와 상호운용성을 비교한다. FAIR 데이터 원칙의 평가 프레임워크는 1997 외환위기 아카이브와 지식그래프를 평가하는 데 활용된다. 구축된 지식그래프는 기록의 다양한 개체가 서로 연계되고, 기록의 이해에 도움이 되는 맥락 정보를 제공한다. 검증 결과는 공통 어휘로 구축된 지식그래프가 기존 아카이브에 비해 디지털 기록의 연계와 검색, 상호운용 관점에서 향상된 결과를 보인다.

ABSTRACT

A digital archive is an online platform for preserving and utilizing digital records worthy of continued preservation. However, there are no shared standards for functionality, metadata, or data technical principles across digital archives in Korea. These issues create challenges in linking distributed digital records. This study proposes a common vocabulary for digital archives to enhance the interoperability of digital records and evaluates the interoperability of the digital archive built with the common vocabulary. We collect and analyze data from the digital archive on the Korean financial crisis of 1997 to construct a knowledge graph and compare its interoperability with the knowledge graph built with RiC-O. The archive and the knowledge graph underwent evaluation using the FAIR data principles evaluation framework. The constructed knowledge graph links various objects in the archive and provides contextual information to aid in understanding the archive. The results demonstrate that a knowledge graph built with a common vocabulary significantly improves the linkage, search, and interoperability of digital records compared to a traditional archive.

Keywords: FAIR 데이터 원칙, 디지털 아카이브, 지식그래프, DCAT, RiC-O

FAIR Data Principles, Digital Archive, Knowledge Graph, DCAT, RiC-O

1. 서론

인공지능, 빅데이터로 대표되는 정보통신기술의 발전은 기록관리를 둘러싼 환경변화를 이끌고 있다. 전통적으로 기록관리는 종이기록물을 관리하는 데에 집중되었으나, 디지털 기록을 보존하기 위한 기록관리체계의 패러다임이 변화하고 있다. 디지털 기록의 보존과 함께 공유와 활용을 위해 디지털 아카이브의 연구가 활발히 진행되고 있다. 디지털 아카이브는 디지털 기록을 보존하고 동시에 기록물을 공유하고 활용할 수 있는 기능을 제공한다. 일반적으로 아카이브는 보존 가치가 있는 자료를 관리, 기록하는 것을 의미한다. 디지털 아카이브는 데이터베이스 또는 온라인 서비스의 일종이며, 디지털 기술을 활용하여 방대한 디지털 기록을 안전하게 보관하고, 사용자들이 기록에 접근하여 기록 콘텐츠를 활용하거나 기록 자체를 공유하기 위한 기능을 제공한다. 디지털 아카이브는 운영주체에 따라 정부와 민간, 기록물의 유형에 따라 통합형과 주제형으로 구분된다(한희정, 2018). 통합형 디지털 아카이브는 주제에 관계 없이 기록물을 제공하고, 주제형 디지털 아카이브는 특정한 주제에 한정하여 서비스를 제공한다.

기술적 측면에서 보면 디지털 아카이브는 디지털 기록에 범용적으로 접근하고 탐색할 수 있으며 디지털화된 기록을 시각적으로 보여주는 공통 기능을 갖고 있다. 그러나 현재 운영되고 있는 디지털 아카이브에서 제공하는 기능은 매우 다른 것이 현실이다. 예를 들어 ‘운산 김관석 아카이브’는 비교적 단순한 게시판 형식으로 기록을 제공하고(한국기독교사회문제연구원, 2016), ‘경기도 메모리’는 경기도와 관련된 다양한 주제를 제공하는 통합형 아카이브로 패킷 기반의 검색 방식을 제공한다(신정아, 2020). 반면 국립중앙도서관의 ‘고신문 디지털 컬렉션’은 LOD(Linked Open Data) 기술을 적용하고 있고, 디지털 기록의 메타데이터를 기계가 읽을 수 있는 형식으로 제공한다(국립중앙도서관, 2023). 그러나 국내에서 운영되고 있는 디지털 아카이브는 기능, 메타데이터, 데이터의 기술원칙과 관련된 공통 원칙이 존재하지 않는다(김학래, 2021). 이와 같은 문제는 디지털 아카이브의 특성을 분석·평가하는데 제약을 만들고, 분산적으로 존재하는 디지털 기록을 연계하기 힘들게 만드는 요인으로 지적되고 있다(김학래, 2021). LOD 기술이 적용된 디지털 아카이브는 이론적으로 데이터의 의미적 연계가 가능하지만, 적절하지 않은 메타데이터의 표현으로 디지털 아카이브 사이의 데이터 연계가 쉽지 않다.

현재 국내의 디지털 아카이브에서 구축한 지식그래프(Singhal, 2012)가 충분히 상호운용이 가능하거나 재사용할 수 있는 형태로 제공된다고 보기 어렵다. 예를 들어, 남산예술센터 디지털 아카이브와 국립극단 디지털 아카이브, 아르크예술기록원의 구술채록 인물관계망 데이터는 소수의 속성을 제외한 대부분의 속성을 신규 어휘로 정의하고 있다(박미란, 2021; 정주영, 2021). 즉, 대부분의 디지털 아카이브가 기존 어휘를 재사용하지 않고, 어휘 사이의 상호운용을 고려하지 않는다. 디지털 기록이 상호 연결되기 위해서 서로 공유된 어휘의 사용이 필수적이다. 커뮤니티에서 공유된 어휘의 사용은 용어에 대한 이해를 서로 공유하고, 합의된 규칙에 따라 데이터를 표현하기 때문에, 데이터 사이의 상호운용이 가능해진다(Cox et al., 2021). 즉, 다수가 합의한 표준 어휘의 사용은 디지털 기록의 상호운용과 재사용을 위한 전제 조건이다. 이런 맥락에서 온톨로지 어휘의 재사용을 FAIR 데이터 원칙으로 검토하는 접근을 고려할 필요가 있다(Amdouni & Jonquet, 2022; Cox et al., 2021; Mazimwe, Hammouda, & Gidudu, 2021; Trojahn, 2022). FAIR 데이터 원칙은 지식그래프를 적용한 디지털 아카이브가 실제로 상호운용이 가능한지 판단하는 지표로 사용될 수 있다(김학래, 2021). 특히 FAIR 데이터 원칙은 인간의 개입 없이 기계가 자동으로 디지털 자원을 찾고 접근할 수 있는 특성을 강조한다(GO FAIR, 2022). 문헌정보학 분야에서 FAIR 데이터 원칙에 대한 논의는 초기 단계에 있다. Koster와 Woutersen-Windhower(2018)는 FAIR 데이터 원칙을 확장하여 LAM(Libraries, archives, museums) 컬렉션의 재사용을 위한 실용적인 가이드라인을 제공한다. Park과 Kim(2022)에서 제안된 지식 모델은 일본군 ‘위안부’ 기록을 표현하기 위해 다양한 어휘의 재사용을 제안하고 있다.

본 논문은 디지털 기록의 상호운용을 개선하기 위한 방안으로 디지털 아카이브를 위한 공통 어휘를 제안한다. ‘1997 외환위기 아카이브’의 데이터를 수집·분석하여 지식그래프를 구축하고, RiC-O로 구축된 지식그래프와 상호

운용성을 비교한다. FAIR 데이터 원칙의 평가 프레임워크는 1997 외환위기 아카이브와 지식그래프를 평가하는데 활용된다. 본 논문의 구성은 다음과 같다. 2장은 디지털 아카이브와 FAIR 데이터 원칙, 온톨로지 어휘 등 관련 연구를 설명한다. 3장은 디지털 기록을 표현하는 공통 어휘의 수용력을 판단하고, 1997 외환위기 디지털 아카이브를 지식그래프로 구축하는 방법을 설명한다. 4장은 공통 어휘로 구축된 디지털 기록의 상호운용이 가능한지 의미적으로 질의하고, 기존 아카이브와 검색 결과를 비교한다. 5장은 FAIR 관점에서 공통 어휘로 구축된 디지털 아카이브의 효과를 검증한다. 6장은 연구를 요약하고, 향후 연구 방향을 기술한다.

2. 선행연구

2.1 디지털 아카이브

전통적인 아카이브는 ‘기록 보관소’의 의미로, 영구적 보존의 가치를 가진 기록의 집합 또는 물리적 기록을 보존하는 장소로 인식되었다(Yadav, 2016). 그러나 디지털 환경에서 대량의 기록이 디지털화되면서 보존 측면에서 디지털 아카이빙에 대한 연구가 광범위하게 진행되고 있다. 대표적으로 OAI(Reference Model for an Open Archival Information System) 참조모델이 디지털 정보를 아카이빙하기 위한 표준으로 논의되고 있다. OAI 참조모델은 장기간에 걸쳐 디지털 자원에 접근하는 것을 유지·보존하는 기록 관리 시스템을 제안한다(Lavoie, 2000). 국내의 디지털 아카이빙 연구는 OAI 참조모델을 국내에 확장 적용하기 위한 연구가 진행되고 있다(강민정, 장우권, 2021; 김희정, 2003; 이승민, 2017; 임진희, 2006). 대표적으로 전자기록물 장기보존패키지 기술규격(NAK 31-2 2022(v1.1))은 국가기록원에서 개발한 기록관리 공공표준으로, OAI 모델의 정보패키지를 참조하여 관련 패키지에 대한 기술을 관리하기 위한 표준 방법을 정의한다.

한편, 기록이 대량으로 디지털화되고, 디지털 형태로 생성되는 기록이 기하급수적으로 증가하며, 대량의 메타데이터가 개방되면서 디지털 환경에서 무수히 많은 양의 아카이브와 관련 데이터에 접근하는 것이 가능해졌다(Hawkins, 2022). 이런 맥락에서, 거대한 양의 디지털 기록은 읽어야 하는 텍스트(texts to be read)가 아닌 발굴되어야 하는 데이터의 컬렉션(collections of data to be mined)으로 인식될 수 있다(Moss, Thomas, & Gollins, 2018). 즉, 디지털 아카이브는 디지털 기록의 보존과 함께, 디지털 기록의 공유와 활용을 위한 온라인 플랫폼으로 기능할 수 있다(김유승, 2010; 한상은, 박희진, 2022; 한희정, 2018; Ridolfo, Hart-Davidson, & McLeod, 2011). 이용자가 디지털로 직접 기록에 접근할 수 있는 디지털 아카이브는 점차 증가하고 보편화되고 있다. 기록의 활용 관점에서 다양한 주제(예: 고고학, 광주학생독립운동가 장재성, 일본고전서적)의 디지털 아카이브를 구축하기 위한 연구가 광범위하게 진행된다(김문희, 장우권, 2021; 양성윤, 2022; 이해림, 2018; 정주영, 2021; 한상은, 박희진, 2022).

그러나 디지털 아카이브는 디지털 기록의 검색과 재사용, 활용에 제한이 있다(김학래, 2021). 디지털 아카이브가 잘 활용될 수 있으려면, 디지털 기록 사이의 통합과 상호운용이 될 수 있어야 한다(Koho et al., 2021; Zeng, 2019). 이런 맥락에서 링크드 데이터 기술은 디지털 기록의 검색과 상호운용을 용이하게 한다(Hawkins, 2022). 대표적으로 WarSampo는 링크드 데이터 기술을 활용해 분산되어 존재하는 제2차 세계대전과 관련된 기록을 통합한 데이터 포털이다(Koho et al., 2021). WarSampo는 제2차 세계대전과 관련된 기록을 통합한 지식그래프를 구축했기 때문에 서로 다른 기록을 통합 검색하고 기록에 포함된 개별 개체(예: 군인, 육군부대)에 대한 정보를 제공하는 것이 용이해진다(Hyvönen et al., 2016).

본 연구는 활용 관점에서 디지털 아카이브가 개선될 수 있는 방안을 검토하고, 활용성이 높은 디지털 아카이브의 구축을 위해 지식그래프 기술을 적용한다.

2.2 FAIR 데이터 원칙

FAIR 데이터 원칙은 연구 데이터의 재사용성을 강화하기 위해 설계된 일련의 가이드 원칙이다(Wilkinson et al., 2016). FAIR 데이터는 데이터를 찾을 수 있고, 데이터에 접근할 수 있으며, 데이터의 상호운용과 재사용을 가능하게 한다(FORCE11, 2016). 특히, FAIR 데이터 원칙은 연구 데이터 관리 환경에 맞추어 기계 스스로 데이터를 찾고 사용할 수 있는 특성을 강조한다(GO FAIR, 2022). FAIR 데이터 원칙은 초기에 오픈 사이언스의 맥락에서 연구 데이터의 재사용성을 강화하기 위한 도구로 인식되었으나, 장기적으로 데이터를 보존하고 관리하기 위한 보편적인 프레임워크로 확장되고 있다(Collins et al., 2018; Corpas et al., 2018; Haux & Knaup, 2019).

FAIR 데이터 원칙은 탐색성(Findable), 접근성(Accessible), 상호운용성(Interoperable), 재사용성(Reusable) 요소로 구성된다. GO FAIR(2022)와 유럽 연합 집행위원회(Collins et al., 2018)에 따르면, 탐색성은 풍부한 메타데이터로 기술되어야 하며, 이용자들이 찾을 수 있는 자원에 데이터가 등록되거나 색인되어야 한다는 특성을 강조한다. 전역적으로 고유하고 영구적인 식별자(globally unique and persistent identifier)가 데이터와 메타데이터에 할당되면, 데이터 탐색과 재사용성을 강화할 수 있다. 접근성은 표준적이고 보편적인 프로토콜을 통해 사람과 기계가 모두 접근할 수 있는 데이터의 특성을 강조한다. 그러나 데이터의 접근성은 개방을 의미하지 않는다(Mons et al., 2017). 데이터 개방은 전적으로 데이터 소유자의 재량이며, 숨겨진 데이터에 접근하거나 재사용할 수 있는 조건을 명시적으로 제공하면 된다. 상호운용성은 데이터와 메타데이터가 지식 표현(knowledge representation)을 위해 공식적이고, 접근 가능하며, 공유되고, 광범위하게 적용할 수 있는 언어의 사용을 강조한다. 기계가 데이터를 읽고 이해할 수 있는 능력(machine-actionable)은 의미적 수준에서 상호운용을 가능하게 한다. 이를 위해, 커뮤니티에서 공통적으로 사용하는 형식, 어휘나 표준을 적용하는 것이 필요하다. 재사용성은 커뮤니티 표준을 충족하는 풍부한 메타데이터와 문서를 제공하고, 데이터의 출처에 대한 풍부한 정보를 메타데이터로 제공할 것을 강조한다. 커뮤니티에서 통용되는 표준 라이선스는 데이터의 재사용에 대한 명확한 규정을 제공할 수 있다.

FAIR 성숙도 평가는 개별 FAIR 지표보다 상세한 지표를 제공하고, FAIR 데이터 원칙을 준수하는 정도를 정량적으로 평가할 수 있다. 세부 지표는 FAIR 성숙도 평가를 개발한 기관마다 조금씩 차이가 있다. FAIR 성숙도 평가는 대표적으로 FAIR 매트릭스(Wilkinson et al., 2019), 데이터 성숙도 모델(Bahim et al., 2020), FAIRsFAIR(Devaraju & Huber, 2021)가 있다. FAIR 매트릭스는 확장성 있고, 자동화 가능한 FAIR 프레임워크를 제안한다. FAIR 매트릭스는 FAIR의 각 항목을 자동적으로 측정하는 성숙도 지표(maturity indicator)와 디지털 자원을 성숙도 지표로 평가하는 지표 준수 테스트(compliance test), 평가에 대한 구체적인 리포트를 제공하는 웹 어플리케이션을 제공한다. FAIR 매트릭스는 자동화된 방식으로 디지털 자원의 FAIRness를 평가하는 것이 큰 장점이다. FAIR 매트릭스는 15개의 성숙도 지표로 평가된다. 성숙도 지표는 웹 어플리케이션이 각 지표마다 자동으로 통과(pass)와 미흡(fail)을 평가한다. 도메인마다 중요하게 생각하는 성숙도 지표가 다를 수 있기 때문에, 도메인의 요구사항에 맞는 성숙도 지표 컬렉션을 선정할 수 있다.

RDA 워킹그룹에서 개발한 데이터 성숙도 모델은 FAIR의 모든 항목을 균등하게 평가할 수 있는 프레임워크다. 데이터 성숙도 모델은 데이터와 메타데이터를 모두 평가할 수 있도록 설계되었다. 이 원칙에서 FAIR 데이터 원칙의 지표는 여러 개의 세부 지표로 구분된다(<표 4> 참고). 예를 들어, F1 지표는 데이터를 측정하는 지표(RDA-F1-01D)와 데이터를 측정하는 지표(RDA-F1-01M)로 구분된다. 세부 지표는 우선순위에 따라 필수, 중요, 유용한 지표로 구분된다. 평가는 각 지표를 5단계로 측정하는 방법과 통과/미흡으로 측정하는 방법이 있다. 5단계 측정 방법은 0점(전혀 적용되지 않음), 1점(아직 고려되지 않음), 2점(계획 단계에서 고려 중), 3점(이행 단계), 4점(완전히 적용됨)으로 평가된다. 통과/미흡 측정은 개별 지표의 이행 여부를 판단한다.

FAIRsFAIR는 RDA의 데이터 성숙도 모델에 기초하여 실용적으로 측정할 수 있는 평가 기준을 제시한다.

FAIRsFAIR의 평가 매트릭스는 RDA의 데이터 성숙도 모델에 기초하지만, 설계 목적에 맞게 일부 지표가 수정된 17개의 세부 지표를 제공한다. 평가 방법은 자가 진단 방식(FAIR-Aware)과 자동화된 방식(F-UJI)이 있다. FAIR-Aware는 FAIR 데이터를 자가 진단할 수 있는 설문 방식의 평가다(FAIRsFAIR, 2021). 설문은 10개의 질문으로 구성되고 '예' 또는 '아니오'로 응답할 수 있다. F-UJI는 디지털 객체의 FAIRness를 자동적으로 측정하는 웹 서비스를 제공한다(Devaraju & Huber, 2021). F-UJI는 디지털 자원의 URI를 입력하면, 자동적으로 FAIRness의 단계와 각 항목별 점수를 측정하는 서비스를 제공한다.

FAIR 매트릭스와 FAIRsFAIR는 URI 또는 PID로 디지털 자원을 평가하는 자동화된 방식을 지원한다. 그러나 디지털 아카이브가 제공하는 대부분의 디지털 기록은 자동적으로 평가할 수 있는 URI 또는 PID를 제공하지 않는다. 본 연구는 디지털 기록의 상호운용과 재사용을 평가할 수 있는 방법으로 RDA의 데이터 성숙도 모델을 사용한다.

2.3 온톨로지 어휘

지식그래프는 개념과 개체 사이의 관계를 기계가 읽고 이해할 수 있는 온톨로지 어휘로 표현된다. 온톨로지는 공유된 개념의 형식적으로(formal) 명시화된(explicit) 명세(specification)로 정의된다(Borst, 1997; Gruber, 1993). Studer, Benjamins, & Fensel(1998)가 제시한 각 개념의 정의는 다음과 같다. '개념화(conceptualization)'는 실세계의 추상적인 모델을 의미하며, 집단이 '공유(share)'하고 있는 지식이 된다. '형식적(formal)'은 자연어가 아닌, 기계가 읽고 이해할 수 있는 언어로 표현된다는 것을 의미한다. '명시화(explicit)'는 사용된 개념의 유형과 사용에 대한 제약이 명확히 정의된다는 것을 의미한다. 즉, 온톨로지는 지식이나 담론을 모델링하기 위한 표현적 요소의 집합을 정의한다(Gruber, 2009). 온톨로지는 재사용이 가능한 지식 베이스를 구축하고, 이기종의 시스템과 데이터베이스를 상호운용이 가능한 형태로 구축하는 데 사용되고 있다(Gruber, 2009). 최근 추세는 데이터의 연결과 공유를 용이하게 하는 범용적인 어휘(예: DCMI, SKOS, Schema.org, DCAT)가 많이 사용되고 있다(김학래, 2017).

기록관리 분야에서 디지털 환경에서 기록의 맥락 기술과 상호운용을 위한 어휘로 RiC-O(Records in Contexts - Ontology)가 논의되고 있다. 이 어휘는 ICA EGAD(International Council on Archives Expert Group on Archival Description)에서 개발한 기록 표준 어휘로, 기록이 갖고 있는 다양한 맥락을 기술하고, 개체 사이의 연결을 위한 온톨로지 어휘로 개발되었다(Clavaud & Wildi, 2021). RiC-O는 계층 구조가 아닌 다차원적으로(multidimensional) 기록을 기술함으로써, 기록과 기록 집합 사이의 관계, 기록과 행위자, 활동, 규칙 등의 상호관계를 표현하는 것이 가능하다. 그러나, RiC-O는 어휘 차원의 연계와 상호운용에 한계가 있다(박하람, 김학래, 2021). RiC-O 어휘의 설계 원칙에서 RiC-O는 도메인 어휘이기 때문에, 어떤 요소도 기존 어휘(예: CIDOC-CRM, PROV-O)를 재사용하지 않는다고 설명한다(Clavaud & ICA EGAD, 2021). 예를 들어, 제목(dct:title)과 개념(skos:Concept)과 같이 광범위하게 사용되고 있는 기존 어휘가 존재하지만, RiC-O는 새로운 어휘(예: rico:title, rico:Concept)로 정의한다. RiC-O의 설계 원칙은 RiC-O가 기존 어휘를 재사용하지 않기 때문에 기록과 관련된 기관이나 프로젝트에서 RiC-O를 적용하기 더 쉽다고 설명한다(Clavaud & ICA EGAD, 2021). 그러나 디지털 기록이 상호 연결되기 위해 커뮤니티에서 다수가 사용하고, 서로 공유하고 있는 어휘의 사용이 필수적이다. RiC-O가 디지털 기록의 맥락 기술과 데이터 사이의 상호운용이 가능하다는 논의는 많지만 스키마 수준의 논의에 머물러 있다(박선희, 2019; 정희명, 이성숙, 2021; Mikhaylova & Metilli, 2023). 이유경과 김학래(2020) 연구와 같이 RiC-O 어휘로 디지털 기록을 표현한 지식그래프를 구축하고, 실질적으로 디지털 기록의 개체 사이에 상호운용이 될 수 있는지 검토가 필요하다.

DCAT은 분산된 웹 환경에서 공개되는 데이터 카탈로그를 기술하기 위한 RDF 어휘(Albertoni et al., 2020)로, 공공데이터 분야에서 데이터세트, 데이터 서비스의 관리와 상호운용을 위해 광범위하게 사용되고 있다. DCAT의 기본 구조는 데이터세트(dcat:Dataset) 배포 형식(dcat:Distribution)과 데이터 서비스(dcat:DataService)로 구

성된다. 데이터세트의 일반적인 특징(예: 제목, 설명, 분류체계)은 `dc:Dataset`으로 표현된다. 데이터세트의 서로 다른 표현 형식(예: JSON, CSV, XLSX)은 `dc:Distribution`으로 정의된다. API와 같은 데이터 서비스는 `dc:DataService`으로 표현된다. DCAT은 데이터세트가 갖고 있는 일반적인 특징을 일관적인 방식으로 표현하는데 강점이 있다. 특히 DCAT은 어휘 사이의 상호운용을 위해 DCMI(Dublin Core Metadata Initiative), SKOS(Simple Knowledge Organization System), FOAF(Friend of a Friend)를 포함한 표준 어휘를 광범위하게 재사용한다(Maali, Cyganiak, & Peristeras, 2010). 이는 검색 엔진에서 데이터의 탐색성을 높이고, 분산되어 존재하는 데이터의 통합 검색(federated search)을 가능하게 한다(Albertoni et al., 2023). 디지털 아카이브 관점에서 DCAT은 디지털 아카이브가 생산하거나 제공하는 데이터를 표현하는 데 활용될 수 있다.

최근 기록관리와 도서관 분야는 검색에 특화된 범용적인 어휘의 사용이 증가하고 있다(Fons, Penka, & Wallis, 2012; Gracy, 2014; Han et al., 2015; Lampron, Mixer, & Han., 2016; Matienzo, Roke, & Carlson, 2017; Mitchell, 2013). 대표적으로 Schema.org가 검색에 용이한 범용적인 어휘다. Schema.org 어휘는 구글, 마이크로소프트, 야후와 안덱스에서 공동 개발한 어휘로, 광범위한 주제(예: 사람, 장소, 사건)에 대한 표현을 포괄한다(Schema.org, 2022). 범용적인 어휘는 도메인에 관계없이 광범위한 자원을 기술하기 때문에, RiC-O나 BIBFRAME(Bibliographic Framework) 어휘에 비해 다차원적인 기록 체계를 기술하는 데 한계가 있다. 그러나 범용적인 어휘는 일반적인 검색 엔진(예: 구글)에서 디지털 아카이브의 컬렉션을 쉽게 검색할 수 있다는 장점이 있다(Han et al., 2015). 범용적인 어휘는 비교적 구조가 단순하기 때문에, 쉽게 링크드 오픈 데이터를 구축할 수 있다(Matienzo, Roke, & Carlson, 2017). 이런 맥락에서, 디지털 기록의 상호운용을 위한 어휘는 어휘 수준에서 상호운용성을 지원하고, 검색에 용이한 어휘인지 검토해야 한다. 본 논문은 RiC-O로 표현된 디지털 기록의 상호운용성을 검토하고, 범용적인 수준에서 기록의 탐색과 상호운용이 가능한 DCAT 어휘를 디지털 아카이브의 공통 어휘로 제안한다.

3. 디지털 아카이브를 위한 공통 어휘

디지털 기록의 상호운용을 위한 공통 어휘는 다음의 사항을 고려해야 한다. 첫째, 공통 어휘는 디지털 기록의 메타데이터를 충분히 표현할 수 있어야 한다. 디지털 아카이브에서 공통적으로 사용할 수 있는 어휘가 되려면, 공통 어휘가 디지털 기록이 갖고 있는 다양한 메타데이터를 수용할 수 있어야 한다. 둘째, 공통 어휘로 표현된 디지털 기록은 상호운용이 가능해야 한다. 공통 어휘로 기술된 디지털 기록이 실제로 연계되었는지 확인할 수 있어야 한다. 온톨로지 어휘로 표현된 기록이 표준 질의 언어인 SPARQL(SPARQL Protocol and RDF Query Language)로 검색될 수 있어야 하고, 서로 다른 기록 사이의 연계 검색이 가능해야 한다. 셋째, 공통 어휘가 어휘 수준에서 상호운용이 가능해야 한다. 어휘 수준의 상호운용이 가능하려면 기존 어휘의 재사용을 우선적으로 고려하는 것이 바람직하다. 기존 어휘는 합의된 규칙에 따라 데이터를 표현하고, 해당 어휘로 표현된 큰 규모의 데이터와 연계하는 것이 용이해진다.

본 논문은 RiC-O로 표현된 디지털 기록의 상호운용성을 검토하고, 디지털 기록을 기술하기 위한 새로운 공통 어휘로 DCAT 어휘를 제안한다. DCAT 어휘가 디지털 기록의 공통 어휘로 사용될 수 있는지 검토하기 위해 다음의 과정을 거친다. 첫째, DCAT 어휘가 디지털 기록을 충분히 표현할 수 있는지 검토한다. 검증 방법은 박하람과 김학래(2022)의 수용력 지표를 활용한다. 둘째, 디지털 기록이 DCAT 어휘로 표현된 지식그래프를 구축한다. DCAT으로 표현된 디지털 기록이 상호운용이 가능한지 평가하기 위해 지식그래프를 구축한다. RiC-O로 구축된 디지털 기록과 비교하기 위해, 동일한 1997 외환위기 아카이브의 디지털 기록이 DCAT 기반의 지식그래프로 구축된다. 셋째, DCAT으로 구축된 디지털 기록이 상호 연계되어 검색 가능한지 검증한다. 디지털 기록의 상호운용이 가능한지 평가하기 위해 3가지의 질의문이 SPARQL로 질의된다. 질의 결과는 기존의 1997 외환위기 아카이브의 검색 결과와 비교된다. 넷째, 디지털 기록의 상호운용이 가능한지 평가하기 위해 FAIR 성숙도 평가를 진행한다. FAIR

성숙도 평가는 기존의 1997 외환위기 아카이브와 RiC-O로 구축된 외환위기 지식그래프(이유경, 김학래, 2020), DCAT으로 구축된 외환위기 지식그래프를 비교한다.

3.1 공통 어휘의 수용력 평가

디지털 아카이브는 기록을 제공하는 웹 공간이며 동시에 데이터 관점에서 기록에 대한 데이터를 포함하는 컨테이너의 역할을 할 수 있다. DCAT 어휘는 도메인에 관계없이 데이터의 일반적인 메타데이터를 기술하는 어휘로 광범위하게 사용되고 있다. 동시에 DCAT은 데이터의 상호운용을 위해 기존 어휘(예: DCMI, SKOS)를 우선적으로 재사용한다. 디지털 아카이브에서 DCAT 어휘를 적용하면 데이터세트 관점에서 범용적인 어휘의 상호운용이 가능하고, 데이터 탐색을 향상시킬 수 있다. 그러나 DCAT 어휘는 범용적인 수준에서 데이터를 기술하기 때문에 기록물에 대한 특수성을 모두 포함하기 어렵다. 이런 맥락에서, DCAT으로 디지털 아카이브의 데이터를 표현하기 위한 메타데이터의 적용 범위를 검토하고 메타데이터의 수용력에 대한 진단이 필요하다.

데이터세트 관점에서 DCAT 어휘가 디지털 기록에 적용 가능한지 판단하기 위해 디지털 아카이브의 메타데이터 항목을 검토한다. 검증은 박하람과 김학래(2022) 연구의 수용력 지수를 활용했다. 수용력 지수는 전체 메타데이터의 개수 대비 DCAT으로 표현 가능한 메타데이터의 개수로 1에 가까울수록 DCAT 어휘가 디지털 기록을 충분히 표현할 수 있다는 의미이다. 평가는 1997 외환위기 아카이브를 포함한 대표적인 7개의 디지털 아카이브를 대상으로 수행되었다. 7개의 디지털 아카이브가 갖고 있는 디지털 기록의 메타데이터 항목을 조사한 후 아카이브의 메타데이터 항목은 DCAT의 속성으로 표현된다. 메타데이터 항목과 DCAT의 속성은 의미적 동일성을 기준으로 매핑된다. 예를 들어 제목(dct:title), 설명(dct:description)은 DCAT의 속성과 매핑되고, 기록유형과 대주제, 소주제는 DCAT의 속성과 매핑할 수 없다.

디지털 아카이브의 수용력 지수는 평균적으로 약 0.58이다. 즉, 디지털 아카이브는 절반 정도의 항목이 DCAT으로 표현될 수 있다. 예를 들어, 제목이나 설명, 라이선스, 식별자 항목과 같이 범용적인 수준에서 데이터를 기술하는 항목은 DCAT으로 충분히 표현 가능하다. 그러나 기록유형, 기록형태, 기증자 등과 같이 기록관리 도메인에 특화된 메타데이터는 DCAT으로 표현하기 어렵다. 데이터를 기술하는 데 필요한 일반적인 메타데이터는 DCAT으로 표현하고, 기록의 특성이 반영된 메타데이터 항목은 RiC-O나 DCMI, Schema.org와 같은 외부 어휘를 사용하는 것이 필요하다.

3.2 DCAT 기반의 지식그래프 구축

RiC-O는 디지털 환경에서 기록의 맥락 정보를 연계할 수 있는 대표적인 기록표준 어휘로 논의되고 있다. 그러나, 실제 RiC-O로 디지털 기록을 표현하고 상호운용을 검증한 연구는 미흡하다. 이유경과 김학래(2020)의 연구가 RiC-O 기반의 지식그래프를 구축한 대표적인 사례다. 본 연구는 RiC-O의 상호운용성을 검토하기 위해 이유경과 김학래(2020)의 연구와 동일한 1997 외환위기 아카이브의 디지털 기록을 대상으로 DCAT 어휘를 적용한다. 1997 외환위기 지식그래프는 외환위기 아카이브의 소장기록과 인물/조직 목록을 대상으로 구축된다. 수집된 기록물은 소장기록 5,982건, 인물/조직 75건으로 총 6,057건의 기록이 지식그래프 구축에 사용된다. 한편 값이 없거나 의미가 없는 메타데이터 항목(예: 주기, 권한관계)은 지식그래프의 구축 범위에서 제외한다. 수집된 데이터세트는 지식그래프 변환을 위해 데이터 정제가 수행된다. 예를 들어 하나의 항목에 여러 가지 항목의 값이 있는 경우 항목을 분리하고(예: 분량 → 쪽수, 재생시간, 용량), 데이터 값의 오류나 공백은 수정된다. 최종적으로 지식그래프 구축에 사용된 메타데이터 항목은 <표 1>과 같다.

1) 선정된 디지털 아카이브는 1997 외환위기 아카이브, 국가기록원의 국가지정기록물, 서울기록원의 컬렉션, 아카이브814, 수요시위 아카이브, 성평등 아카이브, 국회기록보존소의 임시의정원 아카이브다.

〈표 1〉 지식그래프 구축에 활용된 메타데이터 목록

목록	메타데이터 항목	추가된 메타데이터 항목	메타데이터 개수
소장기록 목록	식별번호(Identifier), 컬렉션(Collection), 주제별, 제목, 생산자, 날짜, 설명, 기록유형, 기록형태, 언어, 발행처, 출처, 참고자원, 대주제, 소주제, IMF시기분류, IMF주제분류	쪽수, 재생시간, 용량	20
인물/조직 목록	식별번호(Identifier), 이름(Title), 소속, 당시 직급, 바이오그래피, 홈페이지, 참고자원, 자원유형, 파일(File)	-	9

소장기록과 인물/조직 데이터세트는 국내외 표준 어휘를 적용하여 지식그래프로 구축된다. 데이터세트의 메타데이터 항목은 의미적 동일성을 기준으로 온톨로지 어휘의 속성과 매핑한다. 외환위기 지식그래프는 어휘 수준의 상호운용을 위해 기존 어휘의 재사용을 원칙으로 한다. 외환위기 지식그래프는 우선적으로 DCAT 어휘를 적용하고, DCAT으로 표현되지 않는 항목은 데이터맵, RiC-O, DCMI, Schema.org, SKOS와 같은 국내외 표준 어휘로 표현한다.

〈표 2〉는 동일한 메타데이터 항목에 대해 DCAT으로 매핑한 속성과 RiC-O로 매핑한 속성(이유경, 김학래, 2020)을 비교한 표다. DCAT으로 표현된 외환위기 지식그래프는 어휘의 재사용 측면에서 이유경과 김학래(2020)의 외환위기 지식그래프와 구별된다. RiC-O로 표현된 지식그래프는 외환위기와 관련된 소장기록과 인물/조직에 관련된 정보의 대부분을 RiC-O로 표현했다. 특히 제목(rico:title)과 생산자(rico:createdBy)와 같이 개체를 표현하는 일반적인 속성이 RiC-O가 갖고 있는 속성으로 표현됐다. RiC-O는 모든 어휘를 새롭게 정의하고 있기 때문에 외부 어휘와 상호운용이 제한된다. 한편 DCAT으로 표현된 외환위기 지식그래프는 어휘 수준의 상호운용을 고려하여 기존 어휘를 최대한 재사용한다. 개별 기록은 데이터세트(dcat:Dataset)로 표현되고, DCAT으로 표현이 어려운 사람과 조직은 schema:Person과 schema:Organization을 재사용한다. DCAT은 어휘 자체에서 기존 어휘의 재사용을 고려하고 있어 제목(dct:title)과 생산자(dct:publisher)와 같이 일반적인 속성은 DCMI의 어휘를 광범위하게 재사용한다. 예를 들어 rico:title로 표현되었던 기록의 제목은 DCAT 어휘의 dct:title로 표현되어 더 광범위하게 재사용되는 속성으로 변경된다. 총 29개의 메타데이터 항목 중 13개의 항목이 DCAT으로 표현되고, 이 외의 항목은 Schema.org, DCMI, RiC-O 어휘로 표현된다.

〈표 2〉 동일한 메타데이터 항목에 대해 DCAT으로 매핑한 속성과 RiC-O로 매핑한 속성(이유경, 김학래, 2020)의 비교

데이터 세트	메타데이터 항목	DCAT으로 표현된 1997 외환위기 지식그래프			RiC-O로 표현된 1997 외환위기 지식그래프		
		속성	정의역	공역	속성	정의역	공역
소장 기록, 인물/조직	식별자 (Identifier)	dct:identifier	dcat:Dataset; schema:Person; schema:Organization	xsd:string	rico:identifiedBy	rico:Record; rico:Instantiation; rico:CorporateBody; rico:Event	rico:Identifier
소장 기록	제목	rdfs:label; dct:title	dcat:Dataset	xsd:string	rico:title	rico:Record; rico:Instantiation	rdfs:Literal
소장 기록	생산자	dct:publisher	dcat:Dataset	schema:Person; schema:Organization; xsd:string	rico:createBy	rico:Record	rico:Agent
소장 기록	날짜	dct:created	dcat:Dataset	xsd:date	rico:hasBeginningDate	rico:Record	rico:Date
소장 기록	언어	dct:language	dcat:Dataset	ISO 639-1 Language Codes	rico:hasLanguage	rico:Record	rico:Language

소장 기록	대주제	rico:hasOrHasMainSubject	dcat:Dataset	skos:Concept	rico:hasSubject	rico:Record	rico:Thing
소장 기록	기록유형	rico:hasContentOfType	dcat:Dataset	rico:ContentType	rico:hasContentOfType	rico:Record	rico:ContentType
인물/조직	이름 (Title)	rdfs:label; schema:name	schema:Person; schema:Organization	xsd:string	rico:hasAgentName	rico:Person; rico:CorporateBody	rico:AgentName
인물/조직	소속	schema:worksFor	schema:Person; schema:Organization	xsd:string	rico:isMemberOf	rico:Person	rico:CorporateBody

4. SPARQL 질의를 통한 검색 평가

DCAT으로 구축된 지식그래프는 디지털 기록의 상호운용이 가능한 형태로 검색될 수 있어야 한다. 이를 평가하기 위해 기존 아카이브와 DCAT 기반으로 구축된 지식그래프를 대상으로 평가를 진행한다. <표 3>은 검색 결과를 비교하기 위한 질의문 목록이다. 질의문은 목적에 따라 FAIR와 관련된 요소로 구분된다. Q1과 Q2는 FAIR 요소 중 탐색성과 접근성, 상호운용성과 관련된 질이다. Q3은 상호운용성, 재사용성과 관련된 질이다. 동일한 질의문에 대해 기존 아카이브는 고급검색의 소장기록 제한검색을 활용한다. 지식그래프는 RDF로 구축된 데이터를 검색할 수 있는 표준 질의 언어인 SPARQL을 사용해 질의문을 검색한다. <그림 1>은 개별 질의문에 대한 1997 외환위기 아카이브와 DCAT 기반의 1997 외환위기 지식그래프의 검색 결과다.

<표 3> 1997 외환위기 지식그래프의 질의문 목록

질의문	설명	FAIR 지표			
		F	A	I	R
Q1	'임창열'이 생산한 기록물의 개수를 검색한다.	○	○	○	
Q2	인물/조직 정보가 연계된 소장기록의 개수를 검색한다.	○	○	○	
Q3	참조 가능한 출처 정보가 부여된 소장기록의 개수를 검색한다.			○	○

Q1은 '임창열'이 생산한 기록물의 개수를 검색하는 질의문이다. 기존 아카이브의 고급검색 조건은 '생산자' 항목에 '포함(contains)'된 소장기록이다. 동일한 질의문에 대해 지식그래프가 기존 아카이브의 고급검색보다 8건이 많은 검색 결과를 가져온다. 기존 아카이브는 생산자 '임창열'을 검색한 결과, 지식그래프보다 적은 10건의 기록물이 검색된다²⁾. 기존 아카이브는 기록의 생산자를 텍스트 형식으로 표현하고, 동일 인물을 '임창열', "임창렬", "부총리 겸 재정경제원 장관 임창렬", "부총리"와 같이 서로 다른 이름으로 기술한다. 동일한 개체가 상이한 이름으로 기술되면, 키워드 기반의 검색은 기존 아카이브의 Q1 결과와 같이 동일한 개체가 생산한 기록을 모두 가져오지 못한다. 예를 들어 '부총리'로 검색된 4건, 임창렬로 검색된 5건의 기록물은 찾을 수 없다. 반면 지식그래프는 인물이나 조직 정보를 텍스트로 표현하지 않고 기계가 식별할 수 있는 URI로 표현한다. 지식그래프에서 '임창열'이라는

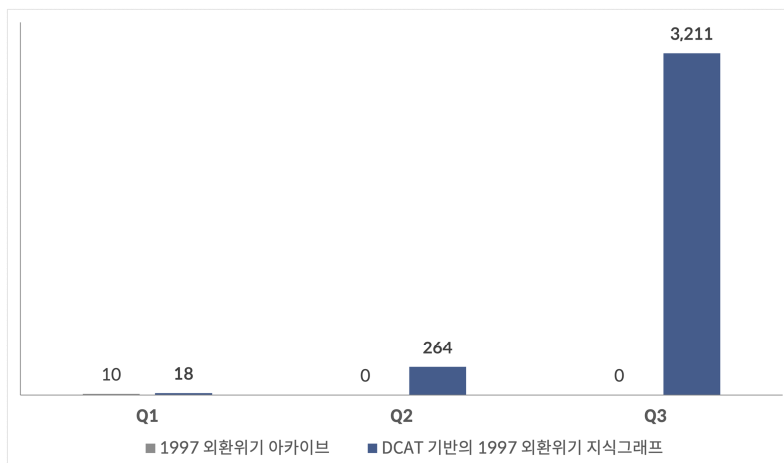
2) 검색 결과 URL:

https://97imf.kr/items/browse?search=&advanced%5B0%5D%5Belement_id%5D=39&advanced%5B0%5D%5Bterms%5D=%EC%9E%84%EC%B0%BD%EC%97%B4&advanced%5B0%5D%5Btype%5D=contains&advanced%5B0%5D%5Bjoiner%5D=and&advanced%5B1%5D%5Belement_id%5D=&advanced%5B1%5D%5Bterms%5D=&advanced%5B1%5D%5Btype%5D=contains&advanced%5B1%5D%5Bjoiner%5D=and&submit_search=%EA%B2%80%EC%83%89%ED%95%98%EA%B8%B0

개체는 URI <http://data.datahub.kr/id/1997-archive/KC-O-0001>가 부여된다. 기계는 이 URI에 대해 이름(rdfs:label)이 ‘임창열’이고, 개체의 유형(rdf:type)이 사람(schema:Person)이며, 직업(schema:jobTitle)이 통상산업부 장관이었다는 정보에 접근할 수 있다. 소장기록의 생산자 항목에 기술된 값도 ‘임창열’에 대해 동일한 URI로 표현된다. 그 결과, 지식그래프는 기존 아카이브에서 검색할 수 없었던 8건의 기록물에 대해 ‘임창열’이 생산한 기록으로 인식할 수 있게 된다.

한편, Q2는 인물/조직 정보가 연계된 소장기록을 가져오는 질의문이다. 기존 아카이브의 검색 시스템은 Q2에 대한 검색 결과를 얻지 못한다. 기존 아카이브가 제공하는 소장기록과 인물/조직 정보가 연계되지 않기 때문이다. 즉, 기존 아카이브는 소장기록의 생산자에 기술된 인물 또는 조직과 정보사전에 포함된 인물조직 정보가 연계되지 않는다. 예를 들어 “Letter from Hubert Neiss to Chang-Yuel Lim (Hand writing)”(KC-R-05404)은 Hubert Neiss가 생산한 문서이지만, 메타데이터에서 단순한 텍스트로 표현되어 있어 Hubert Neiss에 대한 추가적인 정보를 얻을 수 없다. Hubert Neiss에 대한 정보는 사용자가 정보사전에 직접 검색해야 인물 정보를 확인할 수 있다. 그러나 지식그래프는 인물/조직 정보가 연계된 264건의 소장기록의 검색이 가능하다. 지식그래프는 메타데이터 항목의 값을 단순히 텍스트로 표현하지 않고, 기계가 식별 가능한 URI로 표현하기 때문이다. 예를 들어 KC-R-05405 기록은 <http://data.datahub.kr/id/1997-archive/KC-R-05404>로 식별되고, 생산자 항목은 dct:creator로 표현된다. 생산자 항목의 레코드는 Hubert Neiss를 의미하는 <http://data.datahub.kr/id/1997-archive/KC-O-00029>로 표현된다. 즉, 기록의 URI는 생산자를 의미하는 dct:creator로 Hubert Neiss의 URI와 연결된다. 인물조직 정보가 연계된 소장기록은 총 264건으로, 의미적 수준에서 데이터를 연계하는 것이 가능해진다.

Q3은 참조 가능한 형태의 출처 정보가 담긴 기록물을 질의한다. Q2와 같이 Q3도 기존 아카이브에서 검색이 불가능하다. 기존 아카이브는 출처 정보를 단순한 텍스트로 표현하기 때문이다. FAIR 데이터 원칙은 재사용에 대한 명확한 라이선스와 출처 정보의 제공을 강조하고 있다. 기존 아카이브는 재사용이나 저작권에 대한 명확한 규정이 없고, 텍스트 형식의 출처 정보를 제공한다. 예를 들어, “재정경제부”, “IMF”와 같이 텍스트로 표현된 출처 정보는 재사용에 대한 정보를 참조하기 어렵다. 그러나 지식그래프는 참조가 가능한 URI 형태로 출처 정보를 표현하여 질의 결과로 3,211건의 기록물이 검색된다. 예를 들어 ‘재정경제부’로 기술된 출처 정보는 <http://data.datahub.kr/id/1997-archive/KC-O-00061>로 식별된다. 출처 정보는 dct:source의 속성으로 표현되므로 기계는 출처 정보를 찾기 위해 dct:source로 표현된 레코드에 접근한다. 이 레코드는 이미 ‘재정경제부’에 대한 정보(예: 이름, 유형, 설명)를 포함하고 있기 때문에 출처에 대한 풍부한 정보를 제공할 수 있다. 지식그래프로 검색된 3,211건의 기록은 기계가 출처 정보에 자동적으로 접근 가능하고 출처 정보를 참조할 수 있는 방식으로 구축된다.

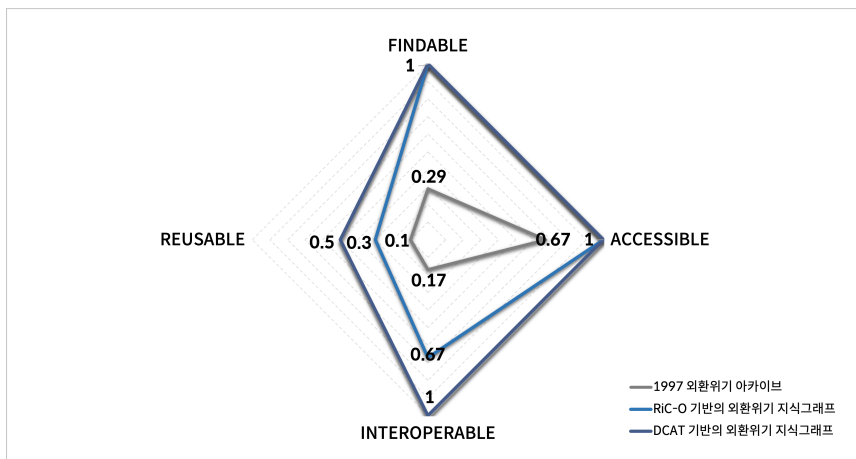


<그림 1> 1997 외환위기 아카이브와 DCAT 기반의 1997 외환위기 지식그래프의 질의문 검색 결과

5. FAIRness 평가

FAIR 데이터 원칙은 데이터 관점에서 디지털 기록이 실제로 상호운용이 가능한지 평가하는 지표로 사용될 수 있다. 특히 지식그래프에 사용된 어휘가 어휘 수준에서 상호운용을 고려하고 있는지 판단할 수 있는 적절한 기준을 제공한다. DCAT으로 구축된 외환위기 지식그래프의 상호운용성을 평가하기 위해 FAIR 데이터 원칙 기반의 평가가 수행된다. 비교를 위해 1997 외환위기 아카이브와 RiC-O로 구축된 외환위기 지식그래프(이유경, 김학래, 2020)가 함께 평가된다.

평가 지표는 RDA의 데이터 성숙도 모델(Bahim et al., 2020)을 사용한다. <표 4>와 같이 데이터 성숙도 모델은 FAIR의 요소마다 평가할 수 있는 세부 지표를 제공한다. ‘탐색성’은 7개, ‘접근성’은 12개, ‘상호운용성’은 12개, ‘재사용성’은 10개의 지표로 구성된다. 평가 방법은 개별 지표를 통과 또는 미흡으로 평가한다. F, A, I, R의 지표 점수는 전체 지표 대비 통과된 지표 개수의 비율로 계산되고, 소수점 셋째자리에서 반올림된다. 예를 들어 외환위기 아카이브의 재사용성 점수가 0.2인 결과는 10개의 재사용성 지표 중 2개의 지표만 통과했다는 의미다. <그림 2>는 1997 외환위기 아카이브와 RiC-O로 구축한 외환위기 지식그래프(이유경, 김학래, 2020), DCAT으로 구축한 외환 위기 지식그래프를 FAIR 성숙도 모델로 평가한 결과다.



<그림 2> 1997 외환위기 아카이브와 RiC-O 기반의 지식그래프, DCAT 기반의 지식그래프의 FAIR 성숙도 평가 결과

FAIR 성숙도 평가 결과는 1997 외환위기 아카이브보다 RiC-O로 구축한 외환위기 지식그래프가 더 높은 점수를 보이고, RiC-O로 구축한 지식그래프보다 DCAT으로 구축한 지식그래프가 더 높은 점수를 보인다. DCAT 기반의 지식그래프와 RiC-O 기반의 지식그래프는 기존 아카이브에 비해 탐색성과 상호운용성, 재사용성에서 뚜렷한 향상을 보인다. 한편, DCAT 기반의 외환위기 지식그래프는 RiC-O 기반의 외환위기 지식그래프에 비해 상호운용성과 재사용성에서 차이를 갖는다. 각 항목의 평가 결과에 대한 설명은 다음과 같다.

- 탐색성(Findable): DCAT 기반의 지식그래프와 RiC-O 기반의 지식그래프의 탐색성 점수는 모두 1로 기존 아카이브(0.29)보다 약 3.5배가 상승했다. 기존 아카이브와 지식그래프의 가장 큰 차이는 데이터와 메타데이터가 전역적으로 고유하고 영구적인 식별자가 부여되는지다. 기존 아카이브는 기록과 인물, 조직을 확인하는 식별자(예: KC-O-0001)를 갖고 있지만 전역적으로 고유한 식별자는 아니다. 기존 아카이브가 제공하는 메타데이터도 텍스트 형태로 제공되고 있어 개별 메타데이터 항목이 고유하게 식별되기 어렵다. 한편 지식그래프는 개별 기록과 인물, 조직에 대해 URI(예: <http://data.datahub.kr/id/1997-archive/KC-O-0001>)를 부여

하고, 해당 URI가 웹상에서 전역적으로 고유하고 영구적인 식별자로서 기능한다. RiC-O와 DCAT은 개별 어휘에 대해 URI를 부여하기 때문에 지식그래프에서 개별 어휘로 표현된 메타데이터 항목은 URI로 식별된다.

- 접근성(Accessible): 접근성은 기존 아카이브(0.67)와 지식그래프(1) 모두 비교적 높은 점수를 갖는다. 접근성 지표는 표준화된 통신 프로토콜로 디지털 기록에 접근이 가능한 것을 강조하고 있다. 기존 아카이브와 지식그래프는 HTTP 또는 HTTPS로 디지털 기록에 접근할 수 있기 때문에 상대적으로 다른 항목보다 점수가 높다. 그러나 접근성 측면에서 지식그래프는 기존 아카이브보다 더 나은 점수를 보인다. 기존 아카이브에서 텍스트 형태로 표현된 데이터와 메타데이터는 개별 이용자가 디지털 아카이브에 접근해야 확인할 수 있다. 반면 RiC-O 또는 DCAT으로 구축된 지식그래프는 모든 개체가 URI로 식별되기 때문에 사람의 개입 없이 기계가 모든 개체에 접근하는 것이 가능하다. 즉, 지식그래프는 텍스트 형식으로 표현되어 접근이 제한된 데이터와 메타데이터에 접근할 수 있게 된다.
- 상호운용성(Interoperable): 상호운용성은 3가지의 평가 대상에 대한 점수 차이가 가장 두드러진다. DCAT으로 구축된 지식그래프(1)와 RiC-O로 구축된 지식그래프(0.67)는 기존 아카이브(0.17)에 비해 상호운용성 측면에서 상승된 점수를 보인다. 기존 아카이브와 지식그래프의 큰 차이는 기계가 읽고 이해할 수 있는 표준 어휘의 사용이다. 기존 아카이브는 DCMI 중심의 메타데이터를 사용하지만, 기관이 자체적으로 정의한 항목도 많다. 기록마다 상이한 메타데이터는 서로 다른 디지털 기록의 상호운용을 제한한다. 특히 텍스트 기반의 메타데이터는 기계가 자동으로 디지털 기록을 연결하기 어렵게 만든다. 그러나 지식그래프는 RiC-O와 DCAT과 같이 기계가 읽고 이해할 수 있는 표준 어휘를 사용하여 디지털 기록을 일관적인 방식으로 기술한다. RiC-O 또는 DCAT으로 기술된 디지털 기록은 해당 어휘를 사용한 서로 다른 디지털 기록과 의미적 수준에서 상호운용이 가능해진다. 한편 DCAT으로 구축된 지식그래프가 RiC-O로 구축된 지식그래프보다 더 높은 상호운용성 점수를 보인다. RiC-O 기반의 지식그래프와 DCAT 기반의 지식그래프의 큰 차이는 어휘 수준의 상호운용이다. 상호운용성 지표는 서로 다른 어휘 사이의 참조를 강조한다. RiC-O는 기존 어휘의 재사용을 고려하지 않고 기존 어휘와 의미적인 매핑을 지원하지 않고 있다. 그러나 DCAT은 DCMI와 같은 기존 어휘를 광범위하게 재사용하고 기존 어휘와 의미적 매핑을 지원한다. DCAT으로 구축된 지식그래프는 기존의 어휘와 상호운용이 가능하기 때문에 대규모의 외부 데이터와 자동적으로 연계하는 것이 가능해진다.
- 재사용성(Reusable): 재사용성 점수도 3가지의 평가 대상에 대한 차이가 존재한다. DCAT 기반의 지식그래프(0.5)와 RiC-O 기반의 지식그래프(0.3)는 기존 아카이브(0.1)보다 다소 상승된 점수를 보인다. 지식그래프는 커뮤니티에서 다수가 합의한 표준 어휘로 디지털 기록을 표현한다. 디지털 기록이 커뮤니티에서 합의한 표준적인 방식으로 기술되면 디지털 기록의 상호 연계가 용이해지고 재사용성이 높아진다. 한편 DCAT 기반의 지식그래프가 RiC-O로 구축된 지식그래프보다 재사용성 측면에서 다소 높은 점수를 보인다. 웹 온톨로지 커뮤니티는 데이터의 연계를 극대화하기 위해 기존 어휘의 재사용을 권장한다. DCAT 기반의 지식그래프는 어휘 수준의 상호운용을 고려했기 때문에 RiC-O 기반의 지식그래프보다 더 높은 점수를 보인다. 그러나 지식그래프가 재사용성 점수를 향상시키는 데 한계가 있다. 재사용성 지표는 출처 정보와 재사용에 대한 명확한 라이선스의 제공을 강조한다. 기존 아카이브가 디지털 기록에 대한 라이선스를 제공해야 지식그래프를 통해 기계 판독이 가능한 라이선스 정보를 구축할 수 있다.

〈표 4〉 FAIR 성숙도 평가 지표와 결과

(A: 1997 외환위기 아카이브, B: RiC-O로 구축된 외환위기 지식그래프, C: DCAT으로 구축된 외환위기 지식그래프)

FAIR	원칙	RDA ID	설명	A	B	C
F	F1	RDA-F1-01M	메타데이터는 영구 식별자(persistent identifier)로 식별된다.	미흡	통과	통과
	F1	RDA-F1-01D	데이터는 영구 식별자로 식별된다.	미흡	통과	통과
	F1	RDA-F1-02M	메타데이터는 전역 고유 식별자(globally unique identifier)로 식별된다.	미흡	통과	통과
	F1	RDA-F1-02D	데이터는 전역 고유 식별자로 식별된다.	미흡	통과	통과
	F2	RDA-F2-01M	풍부한 메타데이터가 탐색을 위해 제공된다.	통과	통과	통과
	F3	RDA-F3-01M	메타데이터는 데이터를 위한 식별자를 포함한다.	통과	통과	통과
	F4	RDA-F4-01M	메타데이터는 수집과 색인이 가능한 방식으로 제공된다.	미흡	통과	통과
탐색성의 통과 지표 비율				2/7	7/7	7/7
A	A1	RDA-A1-01M	메타데이터는 사용자가 데이터에 접근할 수 있는 정보를 포함하고 있다.	통과	통과	통과
	A1	RDA-A1-02M	메타데이터는 수동으로 접근할 수 있다(사람의 개입).	통과	통과	통과
	A1	RDA-A1-02D	데이터는 수동으로 접근할 수 있다(사람의 개입).	통과	통과	통과
	A1	RDA-A1-03M	메타데이터 식별자는 메타데이터 레코드로 확인 된다.	미흡	통과	통과
	A1	RDA-A1-03D	데이터 식별자는 디지털 객체로 확인된다.	미흡	통과	통과
	A1	RDA-A1-04M	메타데이터는 표준 프로토콜을 통해 접근할 수 있다.	통과	통과	통과
	A1	RDA-A1-04D	데이터는 표준 프로토콜을 통해 접근할 수 있다.	통과	통과	통과
	A1	RDA-A1-05D	데이터는 자동으로 접근할 수 있다(컴퓨터 프로그램).	미흡	통과	통과
	A1.1	RDA-A1.1-01M	메타데이터는 무료 접근 프로토콜로 접근할 수 있다.	통과	통과	통과
	A1.1	RDA-A1.1-01D	데이터는 무료 접근 프로토콜로 접근할 수 있다.	통과	통과	통과
	A1.2	RDA-A1.2-01D	데이터는 인증과 권한 부여를 지원하는 접근 프로토콜로 접근할 수 있다.	통과	통과	통과
	A2	RDA-A2-01M	메타데이터는 데이터를 사용할 수 없는 상황에서도 사용할 수 있도록 보장한다.	미흡	통과	통과
접근성의 통과 지표 비율				8/12	12/12	12/12
I	I1	RDA-I1-01M	메타데이터는 표준 형식으로 표현된 지식 표현을 사용한다.	미흡	통과	통과
	I1	RDA-I1-01D	데이터는 표준 형식으로 표현된 지식 표현을 사용한다.	통과	통과	통과
	I1	RDA-I1-02M	메타데이터는 기계가 이해할 수 있는 지식 표현을 사용한다.	미흡	통과	통과
	I1	RDA-I1-02D	데이터는 기계가 이해할 수 있는 지식 표현을 사용한다.	미흡	통과	통과
	I2	RDA-I2-01M	메타데이터는 FAIR 데이터 원칙과 호환되는 어휘를 사용한다.	미흡	미흡	통과
	I2	RDA-I2-01D	데이터는 FAIR 데이터 원칙과 호환되는 어휘를 사용한다.	미흡	미흡	통과
	I3	RDA-I3-01M	메타데이터는 다른 메타데이터에 대한 참조를 포함한다.	미흡	미흡	통과
	I3	RDA-I3-01D	데이터는 다른 메타데이터에 대한 참조를 포함한다.	미흡	통과	통과
	I3	RDA-I3-02M	메타데이터는 다른 데이터에 대한 참조를 포함한다.	통과	통과	통과
	I3	RDA-I3-02D	데이터는 다른 데이터에 대한 정규화된 참조가 포함된다.	미흡	통과	통과
	I3	RDA-I3-03M	메타데이터는 다른 메타데이터에 대한 정규화된 참조가 포함된다.	미흡	미흡	통과
	I3	RDA-I3-04M	메타데이터는 다른 데이터에 대한 정규화된 참조가 포함된다.	미흡	통과	통과
	상호운용성의 통과 지표 비율				2/12	8/12
R	R1	RDA-R1-01M	재사용이 가능하도록 정확하고 관련이 높은 속성이 다수 제공된다.	미흡	통과	통과
	R1.1	RDA-R1.1-01M	메타데이터는 데이터를 재사용할 수 있는 라이선스에 대한 정보가 포함된다.	미흡	미흡	미흡
	R1.1	RDA-R1.1-02M	메타데이터는 표준 재사용 라이선스를 나타낸다.	미흡	미흡	미흡
	R1.1	RDA-R1.1-03M	메타데이터는 기계가 이해할 수 있는 재사용 라이선스를 의미한다.	미흡	미흡	미흡

R1.2	RDA-R1.2-01M	메타데이터는 커뮤니티에 특화된 표준에 대한 출처 정보를 포함한다.	미흡	미흡	미흡
R1.2	RDA-R1.2-02M	메타데이터는 커뮤니티 사이의 언어에 대한 출처 정보를 포함한다.	미흡	미흡	미흡
R1.3	RDA-R1.3-01M	메타데이터는 커뮤니티 표준을 따른다.	미흡	미흡	통과
R1.3	RDA-R1.3-01D	데이터는 커뮤니티 표준을 따른다.	통과	통과	통과
R1.3	RDA-R1.3-02M	메타데이터는 기계가 이해할 수 있는 커뮤니티 표준에 따라 표현된다.	미흡	미흡	통과
R1.3	RDA-R1.3-02D	데이터는 기계가 이해할 수 있는 커뮤니티 표준에 따라 표현된다.	미흡	통과	통과
재사용성의 통과 지표 비율			1/10	3/10	5/10

6. 결론

본 연구는 디지털 기록의 상호운용을 위한 공통 어휘로 DCAT을 제안하고, DCAT 어휘로 구축된 디지털 기록이 상호운용 가능한지 실증적으로 평가했다. DCAT 어휘가 디지털 기록의 특성을 표현할 수 있는지 확인하기 위해 우선적으로 DCAT 어휘의 수용력 평가를 진행했다. DCAT의 수용력 지수는 평균 0.58로 범용적인 수준에서 디지털 기록의 특성을 DCAT 어휘로 표현할 수 있다. 한편 DCAT 어휘로 구축된 디지털 기록이 실제로 상호운용 가능한지 평가하기 위해 1997 외환위기 아카이브를 대상으로 DCAT 기반의 외환위기 지식그래프를 구축했다. 외환위기 지식 그래프는 디지털 기록의 상호운용을 평가하는 SPARQL 질의를 수행한 결과, 기존 아카이브보다 향상된 검색 결과를 보인다. DCAT 기반의 외환위기 지식그래프는 FAIR 성숙도 평가의 모든 항목에서 기존 아카이브와 RiC-O 기반의 외환위기 지식그래프보다 높은 점수를 보인다. 외환위기 지식그래프는 표준 어휘로 표현되어 상호운용성과 재사용성이 높은 디지털 기록을 구축하는 것이 가능하다. 특히 DCAT 어휘는 기존 어휘를 광범위하게 재사용하고 있어 더욱 상호운용과 재사용이 용이한 디지털 기록을 구축할 수 있다.

연구 결과에 따르면, DCAT 어휘는 어휘 수준의 상호운용을 고려하고 있어 디지털 기록의 연계와 탐색에 용이하다. 그러나 DCAT 어휘는 디지털 기록의 일반적인 특성을 표현하는 것이 가능하지만, 기록관리에 특화된 특성을 표현하기에 한계가 존재한다. 향후 디지털 기록의 특성을 담을 수 있도록 DCAT 어휘의 확장이 검토되어야 한다. 한편, RiC-O 어휘는 기록관리 분야에서 디지털 기록의 상호운용을 위한 어휘로 광범위하게 논의되고 있다. 그러나 FAIR 성숙도 평가에서 알 수 있듯이 RiC-O는 어휘 수준에서 상호운용을 고려하지 않기 때문에 디지털 기록의 상호운용에 제한이 있다. RiC-O가 시맨틱 웹이 갖고 있는 장점을 극대화하려면 기존 어휘의 재사용을 검토하거나 RiC-O 어휘와 기존 어휘의 의미적 매핑을 고려하는 노력이 진행되어야 한다.

마지막으로 디지털 아카이브는 디지털 기록의 검색과 상호운용을 위해 기존 온톨로지 어휘를 적극적으로 재사용해야 한다. 구축된 온톨로지 어휘는 어휘가 목표하는 도메인의 합의된 지식 표현을 보장한다. 즉, 광범위하게 활용되는 온톨로지 어휘는 다수가 합의한 방식으로 지식을 표현하기 때문에 온톨로지 어휘로 표현된 지식의 상호운용성과 재사용성을 보장한다. 이런 맥락에서 FAIR 데이터 원칙은 데이터의 상호운용과 재사용을 위해 커뮤니티에서 합의한 표준 어휘의 사용을 권장한다. 디지털 아카이브의 맥락에서 온톨로지 어휘의 재사용은 디지털 기록이 갖고 있는 다양한 개체를 외부 도메인의 데이터와 연계할 수 있어 풍부한 맥락 정보를 제공하는 디지털 아카이브를 구축할 수 있다.

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>
- 국립중앙도서관 (2023). 고신문 디지털 컬렉션. 대한민국 신문 아카이브. 출처:
https://nl.go.kr/newspaper/oldnews_age.do
- 김문희, 장우권 (2021). 광주학생독립운동가 장재성 디지털 아카이브 구축에 관한 연구. 한국기록관리학회지, 21(4), 19-43.
<https://doi.org/10.14404/JKSARM.2021.21.4.019>
- 김유승 (2010). 아카이브 2.0 구축을 위한 이론적 고찰. 한국기록관리학회지, 10(2), 31-52.
<https://doi.org/10.14404/JKSARM.2010.10.2.031>
- 김학래 (2017). 지식그래프. 서울: 커뮤니케이션북스.
- 김학래 (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>
- 김희정 (2003). 전자저널 아카이빙을 위한 OAIS 참조모형의 적용방안에 관한 연구. 한국기록관리학회지, 3(2), 115-141.
<https://doi.org/10.14404/JKSARM.2003.3.2.115>
- 박미란 (2021). 국립극단 디지털 아카이브의 작품 관계망 개선 방안에 대한 제언. 한국극예술연구, 73, 51-86.
<https://doi.org/10.17938/tjkdat.2021..73.51>
- 박선희 (2019). 기록물 맥락정보 향상 및 통합시스템 개발에 관한 연구 - RiC-CM 및 RiC-O를 중심으로. 기록과 정보·문화 연구, 9, 55-96.
- 박하람, 김학래 (2021). 일본군 '위안부' 지식그래프: 파편화된 디지털기록의 연결. 한국기록관리학회지, 21(3), 61-78.
<https://doi.org/10.14404/JKSARM.2021.21.3.061>
- 박하람, 김학래 (2022). DCAT-AP-KR: 국내 데이터 포털의 상호운용을 위한 애플리케이션 프로파일. 디지털콘텐츠학회논문지, 23(11), 2249-2258. <https://doi.org/10.9728/dcs.2022.23.11.2249>
- 신정아 (2020). 지역 디지털 아카이브 구축: '경기도메모리' 사례. 한국기록관리학회지, 20(3), 161-166.
<https://doi.org/10.14404/JKSARM.2020.20.3.161>
- 양성윤 (2022). 일본 고전 서적의 디지털 아카이브 구축현황에 관해 - '역사적 전적 NW 사업'의 현황 및 주요 학술 데이터베이스 소개-. 고전문학과 교육, 49, 7-44. <https://doi.org/10.17319/cle.2022..49.7>
- 이승민 (2017). 디지털 아카이빙을 위한 보존 기술항목 프레임워크 구축. 한국도서관·정보학회지, 48(4), 129- 151.
<https://doi.org/10.16981/kliss.48.4.201712.129>
- 이유경, 김학래 (2020). 1997 외환위기 지식그래프 : 디지털 아카이브의 관계 중심적 접근. 한국기록관리학회지, 20(4), 1-17.
<https://doi.org/10.14404/JKSARM.2020.20.4.001>
- 이혜림 (2018). 국가 고고학 데이터 디지털 아카이브 개발을 위한 연구. 한국기록관리학회지, 18(2), 1-28.
<https://doi.org/10.14404/JKSARM.2018.18.2.001>
- 임진희 (2006). 전자기록의 장기보존을 위한 보존정보패키지(AIP) 구성과 구조. 기록학연구, 13, 41-90.
- 전자기록물 장기보존패키지 기술규격 - 제2부: 디렉터리로 구조화된 방식(NEO3)(v1.1). NAK 31-2 2020(v1.1).
- 정주영 (2021). 남산예술센터 시맨틱 디지털 아카이브 구축에 관한 연구. 한국연극학, 1(77), 211- 248.
<https://doi.org/10.18396/ktsa.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>
- 한국기독교사회문제연구원 (2016). 운산 김관석 아카이브. 출처: <http://jpic.org/archive/>
- 한상은, 박희진 (2022). 디지털 아카이브의 위키데이터 활용방안 연구. 한국기록관리학회지, 22(1), 201-217.
<https://doi.org/10.14404/JKSARM.2022.22.1.201>
- 한희정 (2018). 국내 디지털 아카이브 현황분석 및 시사점. 디지털문화아카이브지, 10(1), 43-52.

- Albertoni, R., Browning, D., Cox, S., Beltran, A. G., Perego, A., Perego, A., & Winstanley, P. (2020). Data Catalog Vocabulary (DCAT) - Version 2. W3C. Available: <https://www.w3.org/TR/vocab-dcat-2/>
- Albertoni, R., Browning, D., Cox, S., Beltran, A. G., Perego, A., Perego, A., & Winstanley, P. (2023). Data Catalog Vocabulary (DCAT) - Version 3. W3C. Available: <https://www.w3.org/TR/vocab-dcat-3/>
- Amdouni, E. & Jonquet, C. (2022). FAIR or FAIRer? An Integrated Quantitative FAIRness Assessment Grid for Semantic Resources and Ontologies. In Garoufallou, E., Ovalle-Perandones, MA., & Vlachidis, A. Eds. *Metadata and Semantic Research. MTSR 2021. Communications in Computer and Information Science*, 1537. Cham: Springer International Publishing, 67-80.
- 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, <https://doi.org/10.5334/dsj-2020-041>
- Borst, W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. Enschede: Centre for Telematics and Information Technology (CTIT).
- Clavaud, F. & ICA EGAD (2021). International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2. Available: https://www.ica.org/standards/RiC/RiC-O_v0-2.html
- Clavaud, F. & Wildi, T. (2021, September 13). ICA Records in Contexts-Ontology (RiC-O): a Semantic Framework for Describing Archival Resources. *Proceedings of Linked Archives International Workshop 2021 co-located with 25th International Conference on Theory and Practice of Digital Libraries (TPDL 2021)*.
- Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., Mietchen, D., Petrauskaitė, R., & Wittenburg, P. (2018). *Turning Fair into Reality: Final Report and Action Plan from the European Commission Expert Group on Fair Data*. Research report. European Commission; European Commission.
- Corpas, M., Kovalevskaya, N. V., McMurray, A., & Nielsen, F. (2018). A FAIR Guide for Data Providers to Maximise Sharing of Human Genomic Data. *PLOS Computational Biology*, 14(3), <https://doi.org/10.1371/journal.pcbi.1005873>
- Cox, S., Gonzalez-Beltran, A. N., Magagna, B., & Marinescu, M. C. (2021). Ten simple rules for making a vocabulary FAIR. *PLOS Computational Biology*, 17(6), <https://doi.org/10.1371/journal.pcbi.1009041>
- Devaraju, A. & Huber, R. (2021). An automated solution for measuring the progress toward FAIR research data. *Patterns*, 2(11), <https://doi.org/10.1016/j.patter.2021.100370>
- FAIRsFAIR (2021). FAIR Aware. Available: <https://fairaware.dans.knaw.nl/>
- Fons, T., Penka, J., & Wallis, R. (2012). OCLC's linked data initiative: Using schema.org to make library data relevant on the web. *Information Standards Quarterly*, 24(2-3), 29-33. <http://dx.doi.org/10.3789/isqv24n2-3.2012.05>
- FORCE11 (2016). *The Fair Data Principles*. Available: <https://force11.org/info/the-fair-data-principles/>
- GO FAIR (2022). *FAIR Principles*. GO FAIR. Available: <https://www.go-fair.org/fair-principles/>
- Gracy, K. F. (2015). Archival description and linked data: a preliminary study of opportunities and implementation challenges. *Archival Science*, 15(3), 239-294. <https://doi.org/10.1007/s10502-014-9216-2>
- Gruber, T. (2009). *Ontology*. In LIU, L. & ÖZSU, M. T. eds. *Encyclopedia of Database Systems*. Boston: Springer.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220. <https://doi.org/10.1006/knac.1993.1008>

- Han, M. K., Cole, T. W., Lampron, P., & Sarol, M. J. (2015). Exposing Library Holdings Metadata in RDF Using Schema.org Semantics. *Proceedings of International Conference on Dublin Core and Metadata Applications*, São Paulo, Brazil.
- Haux, C. & Knaup, P. (2019). Using FAIR Metadata for Secondary Use of Administrative Claims Data. *Studies in Health Technology and Informatics*, 264, 1472-1473.
<https://doi.org/https://doi.org/10.3233/SHTI190490>
- 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, 319-344.
<https://doi.org/10.1007/s10502-021-09381-0>
- Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, E., & Mäkelä, E. (2016). WarSampo Data Service and Semantic Portal for Publishing Linked Open Data About the Second World War History. In Sack, H., Blomqvist, E., d'Aquin, M., Ghidini, C., Ponzetto, S., & Lange, C. eds. *The Semantic Web. Latest Advances and New Domains. ESWC 2016. Lecture Notes in Computer Science*, 9678. Cham: Springer.
- Koho, M., Ikkala, E., Leskinen, P., Tamper, M., Tuominen, J., & Hyvönen, E. (2021). WarSampo knowledge graph: Finland in the second world war as linked open data. *Semantic Web*, 12(2), 265-278.
<https://doi.org/10.3233/sw-200392>
- Koster, L. & Woutersen-Windhower, S. (2018). FAIR Principles for Library, Archive and Museum Collections: A Proposal for Standards for Reusable Collections. *Code4Lib Journal*, 40,
- Lampron, P., Mixter, J., & Han M. K. (2016). Challenges of Mapping Digital Collections Metadata to Schema.org: Working with CONTENTdm. In Garoufallou, E., Subirats Coll, I., Stellato, A., & Greenberg, J. eds. *Metadata and Semantics Research. MTSR 2016. Communications in Computer and Information Science*, 672. Cham: Springer, 181-186.
- Lavoie, B. (2000). Meeting the challenges of digital preservation: the OAI reference model. *OCLC Newsletter*, 243, 26-30.
- Maali, F., Cyganiak, R., & Peristeras, V. (2010). Enabling interoperability of government data catalogues. In Wimmer, M. A., Chappelet, J.L., Janssen, M., & Scholl, H. J. eds. *Electronic Government. EGOV 2010. Lecture Notes in Computer Science*, 6228. Berlin, Heidelberg: Springer.
- Matienzo, M. A., Roke, E. R., & Carlson, S. (2017). Creating a Linked Data-Friendly Metadata Application Profile for Archival Description. *International Conference on Dublin Core and Metadata Applications*, 112-116.
- Mazimwe, A., Hammouda, I., & Gidudu, A. (2021). Implementation of FAIR principles for ontologies in the disaster domain: A systematic literature review. *ISPRS International Journal of Geo-Information*, 10(5),
<https://doi.org/10.3390/ijgi10050324>
- Mikhaylova, D. & Metilli, D. (2023). Extending RiC-O to model historical architectural archives: The ITDT ontology. *Journal on Computing and Cultural Heritage*, 16(4), 1-15. <https://doi.org/10.1145/3606706>
- Mitchell, E. T. (2013). Linked Data Publishing for Libraries, Archives, and Museums: What Is the Next Step?. *Journal of Web Librarianship*, 7(2), 231-236. <https://doi.org/10.1080/19322909.2013.785849>
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49-56. <https://doi.org/10.3233/ISU-170824>
- Moss, M., Thomas, D., & Gollins, T. (2018). The reconfiguration of the archive as data to be mined. *Archivaria*, 86, 118-151.

- Park, H. & Kim, H. (2023). Japanese Military 'Comfort Women' Knowledge Graph: Linking Fragmented Digital Records. *Information Technology and Libraries*, *Information Technology and Libraries*, 42(1), <https://doi.org/10.6017/ital.v42i1.15799>
- Ridolfo, J., Hart-Davidson, W., & McLeod, M. (2011). Imaging The Michigan State University Israelite Samaritan Scroll Collection as the Foundation for a Thriving Social Network. *The Journal of Community Informatics*, 7(3S1).
- Schema.org (2022). Documentation. Available: <https://schema.org/docs/documents.html>
- Singhal, A. (2012). Introducing the Knowledge Graph: things, not strings. Available: <https://blog.google/products/search/introducing-knowledge-graph-things-not/>
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge Engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1-2), 161-197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- Trojahn, C. (2022). FAIR Ontologies, FAIR Ontology Alignments. *Proceeding of the 23rd International Conference on Knowledge Engineering and Knowledge Management*, Bozen-Bolzano, Italy.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A., Groth, P., Goble, C., Grethe, J. S., Jeringa, J., Hoen, P., Hooft, R., Kuhn, R., Kok, R., Kok, J., Lusher, S. J., Marton, M. E., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., Schaik, R., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Lei, J., Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), <https://doi.org/10.1038/sdata.2016.18>
- Wilkinson, M. D., Dumontier, M., Appleton, I.J., Gabrielle, A., Axton, M., Baak, A., Blomberg, N., Boiten, J., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., Schaik, R., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., Lei, J., Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2019). Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data*, 6(1), <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.
- Zeng, M. L. (2019) Semantic enrichment for enhancing LAM data and supporting digital humanities. review article. *Profesional De La información*, 28(1), <https://doi.org/10.3145/epi.2019.ene.03>

● 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

Christian Institute for the Study of Justice and Development (2016). Unsan Kim Kwan-seok Archive. Available: <http://jpic.org/archive/>

Han, Hui-Jeong (2018). A study on the Current Status and Implications of Digital Archive in Korea. *Journal of D-Culture Archives*, 10(1), 43-52.

- Han, Sangeun & 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>
- 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/ktsa.2021.1.77.006>
- Kang, Minjeong & Chang, Wookwon (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, Haklae (2017). *Knowledge Graph*. Seoul: CommunicationBooks.
- 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, Hee-Jung (2003). A Study on e-Journal Archiving based on the OAIS Reference Model. *Journal of Korean Society of Archives and Records Management*, 3(2), 115-141.
<https://doi.org/10.14404/JKSARM.2003.3.2.115>
- Kim, Moonhee & Chang, Wookwon (2021). A Study on the Establishment of Digital Archives by Jang Jae-seong, a Gwangju Student Independence Activist. *Journal of Korean Society of Archives and Records Management*, 21(4), 19-43. <https://doi.org/10.14404/JKSARM.2021.21.4.019>
- Kim, You-Seung (2010). A Theoretical Study on Establishing Archive 2.0. *Journal of Korean Society of Archives and Records Management*, 10(2), 31-52. <https://doi.org/10.14404/JKSARM.2010.10.2.031>
- Lee, Seungmin (2017). Construction of Preservation Description Framework for Digital Archiving. *Journal of Korean Library and Information Science Society*, 48(4), 129- 151.
<https://doi.org/10.16981/kliss.48.4.201712.129>
- 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 Library of Korea (2023). Old Newspaper Digital Collection. Korea Newspaper Archive. Available: https://nl.go.kr/newspaper/oldnews_age.do
- 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, Haram & Kim, Haklae (2022). DCAT-AP-KR: Application Profile for Interoperability of Data Portals in Korea. *Journal of Digital Contents Society*, 23(11), 2249-2258.
<https://doi.org/10.9728/dcs.2022.23.11.2249>
- Park, Miran (2021). A Study on the Improvement of Work Networks of the National Theater Company of Korea Digital Archive. *The Journal of Korean drama and theatre*, 73, 51-86.
<https://doi.org/10.17938/tjkd.2021..73.51>
- 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.

Rhee, Hea Lim (2018). Developing the Korean National Archaeological Data Digital Archive: An Exploratory Study. *Journal of Korean Society of Archives and Records Management*, 18(2), 1-28.

<https://doi.org/10.14404/JKSARM.2018.18.2.001>

Shin, JeongA (2020). Building Local Digital Archives: The Case of “Gyeonggi-do Memory”. *Journal of Korean Society of Archives and Records Management*, 20(3), 161-166.

<https://doi.org/10.14404/JKSARM.2020.20.3.161>

Technical Specification for Long-Term Preservation Package Part 2: Directory Structured Format(NEO3) Version 1.1. NAK 31-2 2020(v1.1).

Yang, Seoung Yoon (2022). Status of classical Japanese digital archives Introduction to the ‘NIJL-NW project’ and the academic database of classical literature. *Journal of Korean Classical Literature and Education*, 49, 7-44. <https://doi.org/10.17319/cle.2022..49.7>

Yim, Jin Hee (2006). The composition and structure of Archival Information Packages(AIP) for a long-term preservation of electronic records. *The Korean Journal of Archival Studies*, 13, 41-90.