

데이터형 전자기록을 위한 출처 개념 모델 개발 방향*

Toward Developing a Provenance Conceptual Model for
Data-driven Electronic Records

현문수(Hyun, Moonsoo)**

1. 들어가며
2. 출처와 맥락의 개념과 관계
 - 1) 출처와 맥락의 개념
 - 2) 출처와 맥락의 역할
3. 데이터 출처의 개념과 적용 가능성 검토
4. 데이터형 전자기록의 출처 모델 개발을 위한 방향
 - 1) 소급형 출처와 전망형 출처 개념의 도입
 - 2) 기본 모델 설계의 방향 제안
5. 출처 모델의 활용 가능성 및 맺음말

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A5A8045436).

** 이화여자대학교 문헌정보학과 강사, 충남대학교 문헌정보학과 강사(moonsoo925@gmail.com).

■ 투고일: 2023년 12월 31일 ■ 최초심사일: 2024년 01월 08일 ■ 최종확정일: 2024년 01월 16일.

■ 기록학연구 79, 305-341, 2024, <https://doi.org/10.20923/kjas.2024.79.305>

〈초록〉

이 연구는 디지털 환경에서 데이터가 중심이 되는 전자기록의 출처에 새롭게 접근하기 위해, 데이터 출처 개념과 출처 모델을 검토하고 수용하여, 어떻게 전자기록을 대상으로 새롭게 출처 개념을 적용할 수 있을지의 가능성을 살펴보았다. 이어서 데이터 중심의 전자기록을 대상으로 한 출처 표현 모델을 개발하기 위해 기초 연구를 진행하였다. 특히 소급형 출처와 전망형 출처 개념으로 전환할 것과, 기록관리 메타데이터와는 별개의 모델을 통해 출처를 표현하고 기록과 연결할 수 있는 모델을 개발할 것을 제안하였다. 기록과 동적 관계를 맺으면서도 독립적으로 출처를 표현할 수 있는 모델을 개발할 수 있다면, 오히려 기록의 유동성을 보장할 수 있으면서도, 기록의 속성과 이를 지원할 출처의 역할을 더 충실히 수행할 수 있을 것이다. 결국, 이 연구가 제안한 기본적인 모델 개발 방향을 수용하는 출처 모델은 기록의 고정성과 활동의 재현성, 재현의 신빙성을 뒷받침할 수 있을 것이며, 디지털 환경에서 적합한 출처 모델로서 역할을 할 수 있을 것이다.

주제어 : 출처, 소급형 출처, 전망형 출처, 출처 표현 모델, 데이터 출처, 데이터형 전자기록, 데이터세트

〈Abstract〉

This study explored the possibilities of a new approach to developing the provenance concept to electronic records in the data-driven digital environments by reviewing and adopting data provenance concepts and models. It then conducted basic literature review to develop a ground for a model representing the provenance of data-driven electronic records. In particular, it proposed to embrace to the concepts of retrospective and prospective provenance, and to develop a different model for representing provenance from records management metadata. If the model can be developed that can represent provenance

independently while maintaining a dynamic relationship with records, it can be ensure the fluidity of records and even support to secure the record's attributes and play the roles of provenance. Eventually, it proposed the direction to develop the provenance model which can support the fixity of records, the reproducibility of activities, and the trustworthiness of representations. It is expected to be a fit provenance model in the data-driven digital environment.

Keywords : provenance, retrospective provenance, prospective provenance, provenance model, data provenance, data-driven electronic records, dataset

1. 들어가며

출처(provenance)의 개념은 기록과 다른 형태의 정보를 구분하여 관리하도록 하는 중요한 개념이다. 기록은 출처에서 그 의미와 가치가 도출되므로, 어떤 문서의 출처를 모른다면 그 문서는 생산의 기원과 맥락이 제거된 정보 소스, 즉 더 넓은 의미와 증거적 가치가 거의 없는 정보 객체에 지나지 않을 수 있다(Cunningham, 2016).

SAA 기록학 용어 사전은 출처를 “어떠한 것의 기원이나 원천, 그리고 어떤 아이템이나 컬렉션의 기원과 관리 및 소유권에 대한 정보”이자 기록에 대한 핵심 원칙으로 정의한다(SAA, 2005-2023, “provenance”). InterPARES 용어 데이터베이스와 ICA도 출처를 “개인이나 기관이 활동을 수행하는 과정에 기록을 생성, 축적 및/또는 유지 관리하고 사용한 조직이나 개인과 기록 간의 관계”로 정의한다(ICA, 2000, “provenance”; InterPARES2, n.d., “provenance”). 이처럼 출처는 애초에 “기록관이나 보존기록관, 매뉴스크립트 보존소로 이관되기 전, 기록을 생산·축적·유지·활용한

조직이나 개인”으로 정의되었으나, 현재는 기록을 생산하거나 유지하고 관리한 실체가 출처의 핵심이 아니라 이들과 기록 사이의 관계를 출처의 본질이며 이는 “기능 출처 개념과 맥을 같이하는 것”이라는 입장(한국기록학회, 2008, “출처”)이 공식화되고 있다.

여러 정의에서 드러나듯이, 업무 활동 또는 맥락과의 연관성은 기록학 담론의 초점이자 기록학 원칙의 기초인 출처 개념과 밀접하게 연결된다. 디지털 환경에서 기록의 개념을 탐구한 Yeo(2007)와 설문원(2019)도 “활동의 지속적” 또는 “활동의 고정적 재현물”의 정의가 기록을 잘 표현한다고 보았다. 또한 기능이나 활동과 기록과의 관계로서 출처는 기록의 조직화(organization)와 평가(appraisal)를 포함하여 광범위한 기록관리 과정에서 기록학 분야의 연구 초점을 드러내는 중요한 개념이다.

물론, 타 학문 영역에서도, 출처의 개념을 다루며 이를 활용하기 위한 연구를 진행한다. 출처는 기록을 포함하는 개체 생산과 관리 연속성의 증거를 제공하는 정보를 포착하여 개체의 진본성을 보장할(Sweeney, 2008) 수 있게 한다. 그러므로 특히 디지털 환경에서 출처 정보는 연구자의 학문 배경과 관계 없이 중요한 연구 대상(Corrado & Sandy, 2017)이다. Lemieux(2016)는, 다학제적 측면에서 출처에 대한 이해를 도모하기 위한 연구 및 실무자 워크숍을 통해, 다른 학문 영역에서도 출처에 관심을 보이고 있으며 출처가 무엇이고 어떻게 효과적으로 표현될 수 있는지에 대한 연구를 늘려나가고 있다고 보았다. 나아가, 정보통신기술을 점점 더 많이 사용하고 있어서 출처에 관한 관심도 더 높아진다고 조심스럽게 추정한다(Lemieux & imProvenance Group, 2016). 특히 기록학 영역에서는 출처 원칙의 적용 문제를 다시 들여다보게 되었으며, 컴퓨터 공학 영역에서는 분산 네트워크로 연결된 컴퓨팅 환경에서 ‘데이터 출처(data provenance)’를 추적하고 분석하기 위한 요구가 증가하고 있는데, 이러한 변화는 새롭게 디지털 형식으로 기록화되는 커뮤니케이션을 관리 및 보존하고 이에 대한 접근을 제공할 필요 때문(Lemieux

& imProvenance Group, 2016)이다.

이 연구는 특히 데이터형 전자기록의 출처에 주목할 것인데, 일상의 활동과 업무 과정에서 이미 분산 운영되는 다종의 정보시스템을 다양하게 활용하고 있으며, 이 과정에서 다양한 유형의 데이터형 전자기록이 생산되고 있기 때문이다. 우리도 이러한 환경에서 출처 개념의 적용과 확장을 위한 과제가 제기되기도 하였는데, 출처 정보의 표현을 위해 기록관리 메타데이터나 기술 모델을 어떻게 활용할 것인지의 문제를 비롯해, 디지털 환경에서의 원질서 존중 개념에 대한 재해석, 데이터형 기록에 대한 출처·원질서 원칙 적용과 생산자를 벗어난 출처 개념의 확장(설문원, 2017) 등이 그것이다. 특히 정부 업무 과정에서 대량의 다양한 전자기록, 특히 데이터세트와 같은 데이터형의 기록이 “유동적이며 복잡한 연계망” 안에서 대량으로 생산되고 있으며, 이렇게 생산된 무수한 “빵부스러기”와 같은 데이터형 기록 간 연결 관계를 측정하여 기록의 가치와 의미를 드러낼 수 있도록 하는 기록관리 정책의 변화가 필요하다(이승억, 설문원, 2017). 아직은, 데이터형 전자기록에 대해서 근대성을 탈피한 관리의 패러다임부터 새롭게 정립할 필요성(이승억, 설문원, 2017)을 제시하거나, 관리기준표 개발(김수영, 2022; 이정은 외, 2022), 보존속성(이정은, 이동민, 2023), 이관 방법(김수영, 2023; 변우영, 임진희, 2022; 양동민 외, 2023) 등에 초점을 두어 연구가 진행되고 있다. 본격적으로 분산 환경을 전제로 생산·관리되는 데이터형 전자기록을 위해 출처 개념을 재해석하고 적용하기 위한 원칙과 방법론을 모색하는 작업은 초기 단계라 할 수 있다.

데이터형 전자기록의 출처에 관한 연구를 진행할 때, 컴퓨터 공학이나 자연과학 분야처럼 데이터를 다루는 타 학문 분야에서 진행한 데이터 출처 연구의 초점, 특히 데이터의 기원과 데이터베이스로 도착한 과정인 ‘혈통(pedigree)’ 또는 ‘계보(lineage)’로서의 데이터 출처(Buneman, Khanna & Tanet, 2001)와 컴퓨터 시스템에서 “데이터로 이어지는 프로

세스”로서 출처(Moreau, 2010)를 주의 깊게 살펴볼 필요가 있다. 데이터 출처에서는 실행된 프로그램, 입출력되는 데이터, 설정 사항, 사용자 등을 포함해 실행과 관련한 많은 사항을 프로세스에 따라 획득할 수 있다는 점을 주목한다. 그 결과 2013년 데이터의 생산과 활용의 출처 정보를 표현하기 위한 W3C 표준으로 PROV 모델이 개발되었고, 자연과학 및 사회과학을 포함하여 다양한 학문 분야에서 이를 적용하거나 확장하여 사용하기 위해 노력하고 있다. PROV 모델에서 출처는 하나의 데이터나 사물을 산출하는 데 관여한 개체와 활동 및 사람에 대한 정보로 보며, 그 품질이나 신뢰성(reliability), 신빙성(trustworthiness)을 평가하는데 출처 정보를 사용할 수 있다(Moreau et al., 2013)고 제안한다. ‘데이터의 생산과 관리에 누가, 어떤 행위를 하면서 참여하였나’가 데이터의 품질과 데이터를 산출한 활동에 대한 신뢰성을 평가하는 기준으로 사용될 수 있음을 중요하게 인식하는 것이다.

최근 기록학 연구에서도 이러한 데이터 출처 개념을 도입해야 한다는 주장을 제기한다. Van Bussel(2017)은 출처와 맥락 원칙을 구분하면서, 출처는 기록의 생산 및 이동 이력과 연관되는 개념이며, 디지털 환경에서 출처주의는 데이터 계보를 포착해야 적용할 수 있다고 주장한다. 또한 그는 출처와 구분되는 맥락에 대해서도, 기록을 생산한 업무 처리나 정책 등의 상황과 환경에 관한 것(Van Bussel, 2017)으로 보았다. 설문원(2019)도 그의 주장에 주목하면서, 개념과 실무적 측면에서 출처와 맥락을 구분해야 한다고 강조하였으며, 디지털 데이터 환경에서는 데이터 계보를 확보할 수 있어야 출처주의를 적용할 수 있다는 반 뷰셀의 주장을 인용하였다. 나아가 기록을 의미 있게 해석하기 위해서는 맥락 역시 중요하게 획득되어야 하며, 여기에는 디지털 데이터를 처리하기 위해 사용된 알고리즘도 포함되어야 한다는 주장을 지지(설문원, 2019)하고 있었다.

따라서, 기존의 출처 원칙을 근대 기록관리 환경을 전제로 한 패러다

임으로 인식하고, 계속해서 변화하는 디지털 환경에 맞게 데이터형 전자기록 관리 환경을 형성하기 위한 근본적인 논의부터 진행될 필요가 있다(이승억, 설문원, 2017). 또, 디지털 환경에서 기록의 출처에 새롭게 접근하기 위해서는, 출처를 다루는 개별 연구 영역에서의 접근방식을 수용하며 포괄함으로써 상호운용할 수 있는 프레임워크를 추구하는 작업이 중요하다(Lemieux & imProvenance Group, 2016). 이 연구는 이를 위한 초기 작업의 수준에서 진행되었다. 즉, 데이터 출처의 개념과 출처 모델을 검토하고 수용하여, 특히 데이터형 전자기록을 대상으로 어떻게 출처 개념을 적용할 수 있을지의 가능성을 살펴보았다. 범위상으로도, 사후적인 기술(description)보다, 시스템 환경에서 기록이 산출될 때 함께 출처 기록이 생산되고 기록과 연결될 수 있는 출처 개념과 표현 모델의 개발을 제안할 목적으로 진행되었다.

2. 출처와 맥락의 개념과 관계

1) 출처와 맥락의 개념

19세기에 정립된 출처 원칙을 구성하는 측면은 일반적으로 크게 두 가지로, 품 존중(respect des fonds) 원칙과 원질서(original order) 원칙이다(ICA EGAD, 2023b; Van Bussel, 2017). 첫 번째, 품 존중 원칙은, 개인이나 조직이 일상생활과 업무 과정에서 생성하거나 축적 및 사용한 기록을 다른 출처의 기록과 섞지 않고 함께 보관해야 한다는 것이다. 즉, 한 개인이나 조직이 존재하는 동안 축적하는 기록에 대해서는 그 존재를 반영하고 문서화하며, 이 기록이 상호 연관된 전체, 즉 일관된 증거를 구성한다는 것을 인정함으로써, 기록 축적의 무결성을 보장하기 위한 것이다. 두 번째, 원질서 원칙을 통해서도 축적과 사용 맥락에서 형

성된 기록의 지적 그룹화와 순서가 기록 간 상호 관계를 이해하는 데 필수적이며 기록이 어떻게 사용되었는지에 대한 증거가 된다는 점을 인식하게 한다. 출처 원칙의 두 측면을 적용하려면 기록이 저장되는 방식, 즉 함께 보관되고 순서가 부여되고 유지되는 방식과 더불어, 지적 설명, 즉 전체와 부분에 대한 설명을 모두 포함해야 한다(ICA EGAD, 2023b).

전통적인 출처 개념을 디지털 환경에 맞게 해석하고 적용하기 위한 연구가 늘어남에 따라, 출처의 개념도 확대되고 있다. ICA의 RiC-CM(Records in Context: Concept Model) 개발에서, 그리고 디지털 환경에서 드러난 출처주의의 쟁점을 지적한 설문원(2017)의 연구에서 살펴본 것처럼, 출처의 개념은 점점 더 확대되어 ‘사회적 출처’까지도 수용해야 한다는 주장이 있고, 출처로 수용해야 할 맥락 요소의 범위도 넓어지는 방향성은 두드러진다. 하지만, 기록이 산출되고 이용되는 직접적인 과정을 출처 정보로서 구조적으로 획득하되, 이를 맥락과 구분할 필요(설문원, 2019; Van Bussel, 2017)도 함께 드러난다. Duranti(1997)나 Van Bussel(2017), 설문원(2019)이 구분한 것과 같이, 출처 원칙이 대상으로 하는 범위를 “생산 조직 전체의 기록 집합과 집합체 내의 관계 구조”로 보는 것이 타당하다. 그러므로 기록을 통해 그 생산자와 생산 시점까지 역추적할 수 있고 기록의 결합구조(archival bond)를 재구성할 수 있는지 확인할 수 있게 하는 것(Duranti, 1997; Van Bussel, 2017), 이것을 기본적인 출처의 기능으로 보아야 한다. 물론, 기록은 법적·행정적 맥락, 출처적 맥락, 절차적 맥락, 문서 맥락 등 다양한 맥락과 연결(Duranti, 1997)되어야 한다. 하지만, 맥락(context)이 가지는 사전적인 의미를 고려한다면¹⁾ 이는 어떤 행위자가 활동을 수행하면서 생산하는 기록과의 직접적인 관계를

1) 케임브리지(Cambridge) 사전에 의하면 ‘맥락(context)’은 “어떤 것이 존재하거나 일어난 상황, 그리고 이를 설명하는 데 도움이 되는 것”을 의미한다. (출처: Cambridge Dictionary. Available: <https://dictionary.cambridge.org/dictionary/english/context>).

벗어나 존재하는 간접적인 것이다.

즉, 맥락은 기록에 의미를 부여하고 기록을 해석하게 하는 환경적 상황을 대상으로 기록화된 정보인데, 기록의 생산을 이해하도록 하는 조직의 정책과 업무 프로세스, 법·규제 환경은 물론이고 사회문화적 영향까지 포함할 수 있다(Van Bussel, 2017). 분명히 맥락은 기록에 중요한 의미를 부여하는 요소이지만 출처의 대상과 맥락의 대상은 분명하게 구분하여 획득하고 기록과 연결해야 할 것이다. 다만, 이를 기록관리 메타데이터의 범위 내에서 출처와 맥락을 기술하는 방법을 벗어나, 별도의 출처 정보를 획득할 수 있는 모델을 수용하는 방안을 고민할 필요가 있다. 예를 들어 정보처리 행위에 사용되는 소프트웨어나 알고리즘은 기록관리 메타데이터 이외의 도구로 획득해야 하는 맥락이다. 디지털 환경에서는 기록의 생산도, 관리 과정에서 발생하는 기록의 의미·구문·물리적 변화도, 대부분 소프트웨어를 통해 수행(Fu, 2015)된다. 따라서 이러한 경우에 기록을 이해하고 그 품질을 평가하려면, 적어도 두 가지 범주의 소프트웨어 정보에 주목할 필요가 있다. 첫 번째는 소스 코드와 도큐멘테이션과 같이 기록을 산출한 소프트웨어를 이해하고 재사용하는 데 필요한 정적 정보이며, 두 번째 범주는 특정 실행 환경과 실행 시 통신 기록이나 변수와 명령어처럼 소프트웨어를 사용하여 수행한 실행 환경을 되풀이할 수 있도록 해주는 동적 정보(Fu, 2015)이다. 특히, AI와 같은 인지 기술의 발달 속도가 빨라지고 활용 범위가 넓어지는 상황에서 기록의 신뢰 여부를 판단하고 활동 과정의 투명성도 확보(Goudarouli, Sexton & Sheridan, 2019)하려면, 기록의 생산과 이용에 활용되는 이러한 소프트웨어 알고리즘을 업무 프로세스 맥락과 유사하게 포착할 수 있어야만 할 것이다.

2) 출처와 맥락의 역할

서로 연결된 전체를 구성하고 기록의 지적 순서를 보여줌으로써, 출처는 다양한 목적으로 사용된다. 무엇보다 출처의 주요한 목적 중 첫 번째는, 기록을 신뢰할 수 있는지 결정할 수 있게 하는 것(Lemieux & imProvenance Group, 2016)이다. 기록이 진본이고 이를 신뢰할 수 있으려면, 우선 기록이 어디에서 왔으며, 어떻게 관리되었는지를 확인할 수 있어야 한다. 디지털 시대 이전에는, 기록을 신뢰할 수 있는 저장소, 즉 아카이브에 배치하고, 신뢰할 수 있는 환경이나 프로세스에 따라 보존하면서 접근 및 이용을 제공했기 때문에 아카이브에서 관리하는 기록을 신뢰할 수 있었다(Michetti, 2016). 연속적 관리의 무결성에 근거하여 간접적으로 기록에 대한 신뢰를 결정하는 방법을 사용해 왔다고 해석할 수 있다. 하지만 디지털 환경에서도 이러한 믿음을 지속할 수 있다고 의심 없이 주장하기 어렵고, 신뢰 확보를 위해 사용해 온 전통적인 방법을 점점 더 적용하기 어려워지고 있다. 특히, 분산 및 네트워크 컴퓨팅 환경의 맥락에서 기록이 생산되고 관리되면서, 기록의 신빙성(trustworthiness)을 확립하고 평가하는 전통적인 접근방식에 문제가 발생하고 있다(이승억, 설문원, 2017; Lemieux & imProvenance Group, 2016). 이를 해결하려면, 기존의 간접적 접근을 넘어서서, 명시적으로 출처에 대한 통제와 관리를 드러낼 수 있어야(Michetti, 2016) 한다. 따라서 기록관리 전문가들은 새로운 모델과 메커니즘 및 도구를 찾고 사용해야 하며, 그 과정에 다양한 전문영역과 협력하면서 전통적인 기록학적 모델을 새로운 모델과 비교하고 통합해야 하는 과제도 해결해야 한다(Michetti, 2016).

두 번째, 기록의 출처를 이해할 때 그 기록의 의미와 가치를 드러낼 수 있다는 점에서, 출처는 기록의 의미를 해석하게 해준다(Lemieux & imProvenance Group, 2016). 기록의 출처는 기록의 내용에 국한하여서는 찾을 수 없는 광범위한 의미와 증거가치를 제공하여, 과거의 의미를

현재 이용자가 해석할 수 있게 한다. 반대로, 누가 기록을 생산했거나 받았는지와 어떤 목적을 위해 작성했는지 알지 못하면 기록은 맥락이 제거되어 탈출처화되고, 광범위한 의미와 증거가치가 거의 없는 일반적인 정보 객체가 되고 만다(Lemieux & imProvenance Group, 2016).

세 번째, 출처는 기록의 조직화 기준을 제시하는 역할을 한다. 기록을 정리하고 기술할 때, 출처는 기록의 기원을 추적하고, 서로 다른 기록을 식별하며, 그룹화하기 위한 첫 번째 단서가 된다(Michetti, 2016). 또한 출처를 기준으로 기록을 유의미한 클래스로 분류하게 되며 기술을 통해 출처를 표현한다. 디지털 환경에서 출처 개념이 확장되고 분산 환경으로 인한 복잡성이 강화되면서 기록관리 메타데이터나 기술을 통한 출처 표현도 영향을 받고 있으며 이를 기록을 조직화하는 과정에 어떻게 활용할 수 있을지(Michetti, 2016) 고민해야 한다. ICA는 최근 수십 년 동안 진행된 출처 원칙에 대한 도전을 인정하면서, 출처를 존중하는 원칙에 대한 전통적인 이해를 수용하면서도 복잡적이며 다양한 행위자를 수용하면서 확장되는 출처를 수용할 수 있도록 하는 기술 모델로 RiC-CM의 버전 1.0을 개발(ICA EGAD, 2023a)하기도 하였다. 한편, 조직화 과정에 출처는 주요 접근점을 제공하여, 기록에 대한 검색을 지원한다. Bearman과 Lytle(1985)은 출처를 기록의 생성 맥락과 기록 간 관계에 대한 정보를 통해 기록의 관련성을 드러내는 검색 도구로 인식할 것을 제안했다. 기록의 검색 수단으로서 출처 정보의 효과는 그 규모(extent)와 표준화에 따라 증가하는데, 기록관리 전문가가 기록의 생산 및 기록체 보존과 관련된 행위자와 행위에 대해 더 많이 알고 전달할수록, 이용자는 이용 목적에 맞게 그 기록물을 더 쉽게 활용할 수 있다(Bearman & Lytle, 1985)는 것이다. 그러므로 출처가 점점 더 복잡한 관계망으로 얽힐수록 사용자가 대량의 정보로 압도당하지 않고 이러한 복잡성을 이해할 수 있도록 출처를 충분하고 철저하게 제공하는 것이 중요해졌다(Michetti, 2016).

마지막으로, 출처는 투명성과 책임성을 지원할 수 있다. 거버넌스의 맥락에서 볼 때, 어떤 정부 정보 객체의 출처 정보에 접근함으로써 정부 기록과 데이터를 신뢰할 수 있고, 이를 근거로 공무원에게 책임을 물을 수 있게 된다(Lemieux & imProvenance Group, 2016). 특히, 지능형 기술을 사용하는 환경에서는 거시적 맥락과 함께 세밀한 프로세스 출처 표현이 더 중요해질 수 있을 것이다.

기록을 정의하고 관리하며 기록에 의미를 부여하는 데에 출처가 중요한 역할을 한다면, 출처 정보는 우선 기록의 속성을 지원해야 할 것이다. 또, 출처와 맥락을 구분한다면, 이러한 역할은 어느 측면에서 지원해야 하는지도 구분해 보아야 할 것이다. 이를 위해, 설문원(2019)이 제안한 기록의 속성을 중심으로 출처의 역할을 분해하고, 다시 출처와 맥락의 측면에서 어떻게 지원할 수 있는지 <표 1>과 같이 분석하였다.

<표 1> 출처의 역할과 기록 속성의 관계

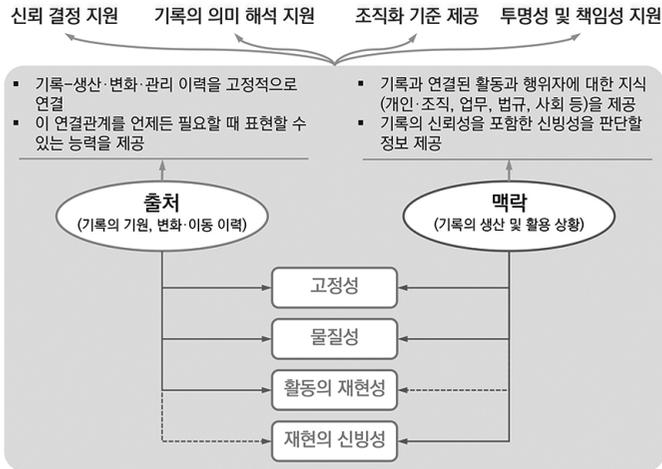
역할	기록 속성과의 관계		구분
신뢰 결정 지원	물질성 (이하 공통)	발생한 활동과 행위자 및 기록 사이의 관계, 기록의 진화 단계를 출처 기록으로 세밀하게 획득하여, 물질성 증거를 확보함(이하 물질성 공통)	<ul style="list-style-type: none"> ■ 출처: 기록과 그 생산·변화·관리 이력을 고정적으로 연결하여, 언제든지 필요할 때 이를 추적하고 표현할 능력을 제공 ■ 맥락: 기록과 연결된 활동과 행위자에 대한 지식을 제공해 출처 및 기록의 신빙성을 평가할 증거를 제공
	고정성	발생했던 활동과 행위자, 기록 사이의 관계를 출처 기록으로 산출하고 당시의 맥락을 기록화하여, 기록의 신뢰 여부를 판단할 기초 자료로 활용하게 함	
	활동의 재현성	획득된 출처 기록과 맥락 정보를 추적·활용하여, 필요한 시점에 발생했던 활동을 다시 표현할 수 있게 함	
	재현의 신빙성	기록과 연결된 활동과 행위자에 대한 지식을 제공하여, 기록과 기록을 통해 재현된 활동의 품질을 평가할 수 있도록 함	

역할	기록 속성과의 관계		구분
기록의 의미 해석 지원	고정성	발생했던 활동과 행위자, 기록 사이의 관계를 출처 기록으로 산출하고 당시의 맥락을 기록화하여, 필요한 시점에 언제든지 기록의 생산 및 관리의 의미를 이해하게 함	<ul style="list-style-type: none"> ■ 출처: 기록과 그 생산·변화관리 이력을 고정적으로 연결하여, 기록 산출·변화의 인과와 흐름을 이해할 능력을 제공 ■ 맥락: 기록과 연결된 활동과 행위자에 대한 지식을 제공해 기록 생산·관리의 의미를 이해하고 해석할 정보를 제공
	활동의 재현성	획득된 출처 기록과 맥락 정보를 추적·활용하여, 기록을 생산하고 관리한 활동과 기록을 이해할 수 있게 함	
	재현의 신빙성	기록과 연결된 활동과 행위자에 대한 지식을 제공하여, 기록의 의미를 폭넓게 이해하고 해석할 수 있게 함	
조직화 기준 및 접근점 제공	고정성	발생했던 활동과 행위자, 기록 사이의 관계를 출처 기록으로 산출하고 당시의 맥락을 기록화하여, 기록과 출처 및 맥락을 고정적으로 연결해 줌	<ul style="list-style-type: none"> ■ 출처: 기록과 그 생산·변화관리 이력을 고정적으로 연결하여, 기록의 분류 지점과 접근점을 제공 ■ 맥락: 기록과 연결된 활동과 행위자에 대한 지식을 제공해, 기록의 분류·기술 및 접근점 활동을 위한 정보 제공
	활동의 재현성	획득된 출처 기록과 맥락 정보를 추적·활용하여, 발생했던 활동을 기준으로 기록을 집합화하고 발생했던 일련의 활동을 동일하게 재구성할 수 있도록 분류·기술하여 미래 접근을 도움	
	재현의 신빙성	기록과 연결된 활동과 행위자에 대한 지식을 제공하여, 기록의 분류·기술의 원자료를 제공	
투명성 및 책임성 지원	고정성	발생했던 활동과 행위자, 기록 사이의 관계를 출처 기록으로 산출하여, 필요한 시점에 전달하고 규정되거나 합의된 업무 과정과 비교할 수 있도록 함	<ul style="list-style-type: none"> ■ 출처: 기록과 그 생산·변화관리 이력을 고정적으로 연결하여, 필요할 때 전달할 능력을 제공 ■ 맥락: 기록과 연결된 활동과 행위자에 대한 지식을 제공해, 미리 정의된 법규제 환경과 업무 과정에 따라 업무를 수행하였음을 공개하고, 설명책임의 의무 이행을 지원
	활동의 재현성	획득된 출처 기록과 맥락 정보를 추적·활용하여 발생했던 활동을 재현할 수 있어, 설명책임의 의무를 지원함	
	재현의 신빙성	기록과 연결된 활동과 행위자에 대한 지식을 제공하여, 활동에 대한 설명책임과 기록의 투명한 공개를 도움	

이처럼, 출처 정보는 기록의 ‘물질성’, ‘고정성’, ‘활동의 재현성’, ‘재현의 신빙성’의 속성과 밀접하게 연결되어야 그 역할을 다할 수 있다. 또, 일반적으로 제시되는 출처의 역할 범위에는, 기록과 그 생산·변화·관

리 이력을 고정적으로 연결하는 ‘출처’의 역할은 물론이고, 기록과 연결된 활동 및 다양한 행위자에 대한 지식을 제공하는 ‘맥락’의 역할까지 포함하고 있음을 분석할 수 있었다. 정리하면, 기록을 관리할 때 출처가 의미 있는 역할을 하려면, <표 1>과 <그림 1>로 정리한 것처럼, 기록의 기원을 추적하는 능력과 기록과 연결된 활동 및 행위자에 대한 지식을 제공하는 능력이 함께 필요하며, 이 능력을 활용해서 기록의 속성을 확보하는 데에도 각각 중요한 역할을 하게 됨을 알 수 있었다. 다만, ‘활동의 재현성’ 측면에서는 맥락 정보보다 출처 정보가 더 직접적인 역할을 할 것이며, ‘재현의 신빙성’ 측면에서는 맥락 정보가 출처 정보보다 더 직접적인 역할을 할 것이기에, <그림 2>에서는 출처와 ‘재현의 신빙성’ 속성, 그리고 맥락과 ‘활동의 재현성’ 속성을 점선으로 연결하였다.

<그림 1> 출처와 맥락의 역할 구분



이렇게 출처와 맥락의 능력을 구분하면, 기록의 출처에 관한 개념을 더 분명하게 드러낼 수 있으며, 기록의 속성을 지원할 가능성도 함께

탐구할 수 있을 것이다. 그러려면, 활동 과정에서 기록과 독립적으로 출처 기록을 산출하고 이를 관리함으로써 좀 더 자유롭게 출처의 역할과 기록의 속성을 지원할 수 있는 출처 표현 모델이 필요하다. 이를 통해 “선택된 구조물”(Van Bussel, 2017)인 기록보다 더 풍부하고 조밀하게 기록의 기원을 보여주어서, 활동 당시에 발생했던 사건을 더 구체적으로 재현할 수도 있을 것이다. 따라서, ‘기록이 어디에서 왔는가?’를 다루었던 전통적 개념을 벗어나지 않으면서도, 시스템 기반의 활동 과정에서 산출되고, 유동적이며 분산·연계적인 특징을 갖는 데이터형 전자기록을 지원할 수 있는 확장된 출처 모델의 개발 가능성을 탐구할 필요가 있다.

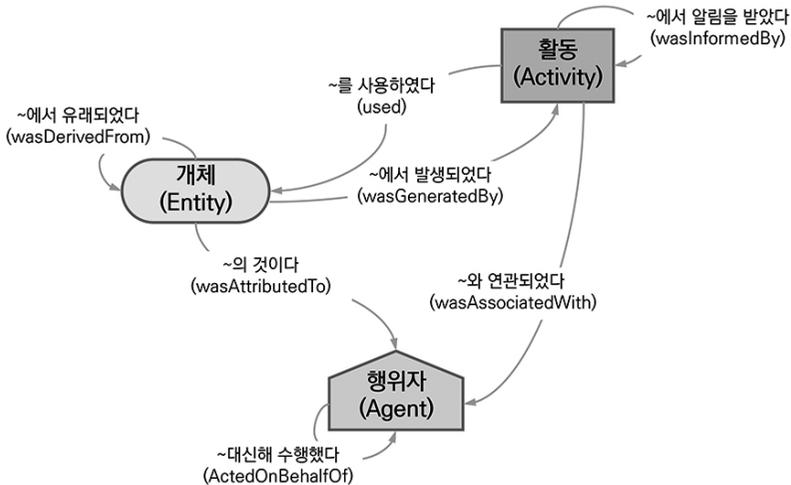
3. 데이터 출처의 개념과 적용 가능성 검토

과학 및 컴퓨터 공학에서 다루는 데이터 출처는 기록이 어디에서 왔는지의 출처 개념을 적용하는 데 도움이 될 수 있다. 이 전문영역에서는 학술 출판물을 뒷받침하는 실험 결과의 증거를 제공하기 위해 출처 정보를 중요하게 다루므로 컴퓨팅 환경을 사용하는 실험 환경에서 어떻게 표준적으로 출처 정보를 획득할 수 있을지(Cuevas-Vicentín et al., 2016) 고민을 해왔다. 2000년대 중반부터 데이터의 출처를 기술하고 이를 교환하기 위해 실험하고 공통 모델을 개발하기 위한 연구가 이어졌으며, 이를 바탕으로 2009년 9월, W3C 출처 인큐베이터 그룹(Provenance Incubator Group)이 창설된 후, 결국 2013년 6월에 W3C의 표준 출처 모델로 PROV 최종 권고안을 발표했다(Missier, 2016). 여러 정보 및 메타데이터 표준에서도 일부 정보 요소로 출처를 표현하도록 하지만, 완전하게 데이터 개체의 계보를 기록화하기에는 부족하다는 판단에 따라, 데이터를 포함하는 개체의 출처를 충분히 서술할 수 있는 표준(Car, 2017.3.15.)

으로 PROV 모델이 개발되었다.

출처에 대한 범용 데이터 모델인 PROV 규격에서는 데이터를 포함하는 ‘사물을 생산하거나 전달하고, 그것에 영향을 주는 데 참여한 사람, 기관, 개체 및 활동을 기술하는 기록’(Moreau et al., 2013)을 출처로 정의한다. 따라서 PROV 모델의 핵심 출처 구조는 <그림 2>와 같이, 발생한 활동과 이 활동에서 사용하거나 생산한 개체, 이 과정에 영향을 미친 행위자를 중심으로 구성된다. PROV에 따라 출처 정보를 생산하면, 개별 데이터나 데이터세트의 생산을 설명하는 기원에 관한 주장을 모을 수 있다(Missier, 2016). 따라서 이렇게 표현되는 출처 기록을 통해 개체인 정보를 신뢰할 수 있는지, 개체에 대한 소유권을 포함하여 다양한 권리 및 책임이 누구에게 있는지를 판단할 수 있으며, 표준화된 출처 정보를 서로 다른 시스템 간에도 상호 교환할 수 있어 출처를 확인하기 위한 요구를 지원(Moreau et al., 2013)할 수 있게 한다.

<그림 2> PROV 개념 모델의 핵심 구조(Moreau et al., 2013)



PROV 규격을 구성하는 핵심 개념은 ‘개체’와 ‘활동’ 및 ‘행위자’로, 각 각을 간략히 설명하면 다음과 같다(Moreau et al., 2013). 개체(entity)는 문서나 파일, 자동차, 아이디어처럼, 물리적이거나 디지털이며 고정된 어떤 사물, 또는 개념적인 것으로서 실재하거나 가상의 존재일 수 있다. 활동은 일정 기간 발생하며 개체에 대해 행위를 했거나 개체와 함께 행위를 수행한 것으로, 디지털 이미지를 사용하여 웹에 문서를 생산해 게시하기, 서울에서 부산까지 자동차 몰기, 데이터세트를 활용해 통계 분석하기 등 넓은 범위를 포함하는 개념이다. 활동은 개체를 사용하고 개체를 생산하며, 후속 활동으로 이어질 수도 있다. 또한, 활동을 통해 개체가 다른 개체로 변형되거나 갱신되어 새로운 개체가 유래될 수 있다. 행위자는 수행되는 활동, 개체의 존재 또는 다른 행위자의 활동에 대해 어떤 형태로든 책임을 지는 것이며 활동을 수행할 권한과 책임을 다른 행위자에게 위임할 수 있다. 서울에서 부산까지 자동차를 운전하는 사람은 물론이고 디지털 문서의 검증에 사용한 소프트웨어, 도서를 판매하는 웹사이트, 호스팅 업체 등도 모두 행위자다.

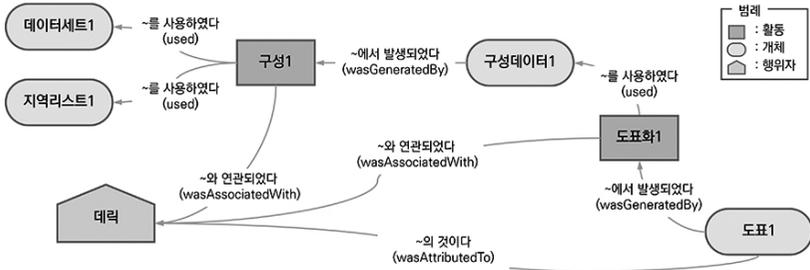
개체와 활동, 행위자는 다음과 같은 이진 관계로 연결된다(Moreau et al., 2013). 우선 개체와 활동은 두 가지 방식으로 상호 연결되는데, 활동은 개체를 사용했고(used), 활동에 따라 새로운 개체가 산출되었다(wasGeneratedBy)는 관계가 형성된다. 예를 들어, 어떤 프로그램이 파일을 읽기 시작했고(사용), 새로운 버전의 파일을 생성했음(산출)을 표현할 수 있다. 행위자도 개체와 활동과 관련된다. 개체는 어떤 행위자와 책임(wasAttributedTo) 관계가 있으며, 행위자는 활동에 어떤 역할을 하면서 연관(wasAssociatedWith)된다. 게시된 웹 페이지 개발 과정에 어떤 웹디자이너가 참여했다면(연관) 이 웹 페이지에 대해서는 그 디자이너가 개발 책임을 갖는다. 서로 다른 구성 개념 간의 연결 이외에, 같은 구성 개념 사이를 연결할 수도 있다. 한 개체가 다른 개체에서 파생(wasDerivedFrom)되고, 어떤 활동이 개체를 생성하고 뒤따르는 활동에 넘겨 사용하도록 하

는 커뮤니케이션을 통한 연결(wasInformedBy)이 발생하며, 대리인이나 대표자로서 특정 활동을 위임받아 수행(actedOnBehalfOf)하는 관계도 표현할 수 있다.

한편, PROV는 용도에 맞추어 출처 기록을 달리 구성할 수 있도록 하므로, 출처에 대한 관점에 따라 결과적으로 다른 유형의 정보가 출처 기록으로 획득될 수 있다(Gil et al., 2013). 대략 세 가지 관점에서 출처 기록을 포착할 수 있을 것인데, 첫 번째는 데이터세트와 같은 개체 관점에서 출처 기록을 구성하는 것이다. 즉, 특정 개체의 기원을 다른 객체로부터 추적할 수 있도록 하는 것인데, 어떤 웹 페이지를 구성한 뉴스 기사 내용과 인터뷰 인용문, 통계 데이터 등을 추적할 수 있도록 출처 기록을 생산하는 경우가 여기에 해당한다. 두 번째는, 행위자 관점으로 출처를 표현하는 것으로, 어떤 개체를 생산하거나 조작하는데 관련한 개인이나 기관 등에 초점을 맞춘다. 예를 들어, 뉴스 기사에 실린 사진 출처로, 처음 이 사진을 찍은 사진작가, 이 사진을 편집한 사람, 이 뉴스 기사를 발행한 신문 등에 대한 정보를 획득할 수 있도록 할 수 있다. 세 번째는, 프로세스에 초점을 두어 출처 기록을 산출하는 것으로, 어떤 개체를 생성하기 위해 취한 조치와 단계를 획득하게 된다. 어떤 서비스를 호출해 데이터베이스에서 데이터를 검색한 후, 통계 패키지로 통계 데이터를 추출한 후 그래프 도구를 사용해 도표화하여 최종 그래프를 산출했다면, 이를 일련의 프로세스 중심의 출처 기록으로 산출할 수 있다.

PROV로 표현한 간략한 출처 사례를 살펴보면 <그림 3>과 같다. 데릭 [행위자]은 '데이터세트1'[개체]과 '지역리스트1'[개체]을 활용해 구성[활동]하여 '구성데이터1'[개체]를 제작하였고, 이를 사용하여 도표화[활동]한 결과로 '도표1'을 산출하였다. 따라서 데릭은 구성 활동[구성1]과 도표화 활동[도표화1]과 연관되며, 최종 결과물인 '도표1'에 대해서도 책임이 있다.

〈그림 3〉 PROV의 출처 표현 사례(Gil et al., 2013; 재구성)



출처를 표현하기 위한 표준으로 W3C에서 PROV 규격을 채택하였음에도, 이 표준은 이미 생성된 개체, 발생한 활동, 이 활동 수행에 참여한 행위자 등 과거로 그 기원을 소급해 추적할 수 있는 출처 표현에 초점을 둔다. 하지만 일부 전문영역에서는 발생한 프로세스의 이력을 추적하도록 하는 소급형(retrospective) 출처 표현은 물론, 사전적으로(proactive) 정의된 프로세스에 대한 정보도 함께 획득할 필요가 있어서 확장형 PROV를 제안하기도 하였다. 특히 데이터 출처를 다루는 연구 영역에서 표준 출처 모델 PROV를 확장해 전망형(prospective) 출처까지 지원할 수 있도록 적극적으로 시도하고 있다.

Freire et al.(2008)과 Lim et al.(2010), Missier et al.(2013)은 과학 연구용 워크플로우 시스템 환경에서 특정 프로세스를 실행하며 산출되는 소급형 출처와 함께 이들 프로세스 자체에 대한 표현인 전망형 출처까지 기록화하기 위한 모델을 제안하기도 하였다. 특히 ProvONE 모델은 기존에 제안된 과학 분야 워크플로우 출처 모델을 통합하면서 PROV를 확장한 것으로, 소급형 출처와 전망형 출처를 동시에 획득(Cuevas-Vicentín et al., 2016)하기 위해 제안되었다. PROV는 전망형 출처를 일반적인 것으로 보아, 의도적으로 모델에서 제외하였다(Missier, 2016). 하지만 ProvONE에서는, 데이터 객체를 생산하기 위해 수행해야 하는 단계를 설명하는 클

래스가 있어야 프로그램 실행과 관련한 단계를 이해할 수 있을 뿐 아니라, 프로그램이 예상대로 실행 때 데이터가 생성되는 과정을 이해할 수 있다(Cuevas-Vicenttín et al., 2016)고 보았기 때문이다.

또한, 이 모델은 지구과학 공동체에 서비스를 제공할 목적으로 구축된 대규모 데이터 인프라인, DataONE 프로젝트 맥락에서 개발되어 ProvONE으로 명명되었으나, 이에 제한하지 않고 폭넓은 과학 공동체에서 워크플로우를 기반으로 한 계산 실험을 이해하고 분석하는 데 필요한 출처를 표준화하여 표현할 것을 목표로 한다(Cuevas-Vicenttín et al., 2016). ProvONE에서 다루는 대상을 정리하면 다음의 표와 같다.

〈표 2〉 ProvONE 모델의 표현 대상(Cuevas-Vicenttín et al., 2016; 재구성)

ProvONE의 표현 대상		표현 내용	활용
출처	전망형 출처 (prospective provenance)	<ul style="list-style-type: none"> ■ 데이터를 생성하는데 필요한 절차 또는 방법 ■ 과학 실험용 워크플로우의 경우, 데이터를 생성하기 위해 수행해야 하는 단계를 자세히 설명하는 워크플로우 규격 ■ 일정 정도 추상화한 표현 가능 	프로그램이 예상대로 실행되면 데이터가 어떻게 생성되는지 이해할 수 있음
	소급형 출처 (retrospective provenance)	<ul style="list-style-type: none"> ■ 데이터를 생성할 때 실행된 단계, 관련된 기본 입력물과 중간 데이터에 대한 기록, 관련 환경에 대한 정보 ■ 일련의 입출력 데이터를 사용하고 생성하는 소프트웨어 에이전트의 계산 실행 정보 	실제로 데이터가 어떻게 생성되었는지 이해할 수 있음
	프로세스 출처 (process provenance)	<ul style="list-style-type: none"> ■ 데이터를 생성하는 데 사용된 프로그램 자체의 변화, 워크플로우의 진화 정보 	원하는 데이터를 생성하기 위해 워크플로우를 다양하게 변경한 사항을 이해할 수 있음
데이터	데이터 구조	<ul style="list-style-type: none"> ■ 실험 및 계산 프로세스에서 사용되거나 생성된 데이터가 어떻게 구성되고 표현되는지의 정보 ■ 과학 실험 워크플로우의 일부를 구성하는 다양한 작업의 입력물과 출력물 	실행 과정에서 사용되고 생성된 데이터 자체에 대해 이해할 수 있음

표와 같이, ProvONE이 지원하는 출처 표현은 크게 3개 유형으로 나뉜다. PROV 규격에서 다루는 소급형 출처를 포함하며, 이 밖에 전망형 출처와 프로세스 출처도 표현할 수 있도록 설계되었다. ProvONE의 구성요소로서 전망형 출처는 데이터 생성에 사용되는 일반 절차나 방법을 표현하기 위한 것으로, 일련의 절차와 방법 내에서 하나 이상의 프로그램을 사용하여 데이터를 처리할 수 있고, 프로그램에서 작동하는 입출력 방식에 관해서도 설명할 수 있다(Cuevas-Vicentín et al., 2016). 따라서 실제 데이터의 생성 이전에, 또는 프로그램이 예상대로 실행되면, 데이터가 어떻게 생성되는지 이해하거나 데이터 생성 이후 미리 정의된 절차나 방법에 따라 데이터가 생성되었는지 확인하는데 전망형 출처를 사용할 수 있게 한다. 프로세스 출처는 데이터 생성에 사용된 워크플로우 또는 프로그램 자체가 어떻게 진화해 나갔는지를 표현한다(Cuevas-Vicentín et al., 2016). 예를 들어, 실제 데이터를 생성하는 과정에 프로그램 실행의 시행착오가 있거나 오류가 발생해 보정이 필요했다면, 워크플로우나 프로그램을 변경해 나가며 데이터를 최종적으로 산출할 수 있을 것이다. 이러한 변경 사항을 추적하고 전달할 필요가 있다면 프로세스 출처로 기록하여 활용할 수 있다. 이 밖에 출처 표현은 아니지만, 워크플로우나 프로그램 내에서 사용하고 산출하는 데이터 자체에 대한 정보도 데이터 구조 측면에서 ProvONE을 적용해 표현할 수 있도록 설계되었다.

한편, 기록학의 영역에서는 출처 정보를 기록화하기 위해 기록관리 메타데이터 모델(ISO 23081 시리즈)이나 보존기록 기술 모델(RiC-CM)을 사용한다. 기록과 행위자, 업무, 규정 및 이들의 관계를 표현하면서 기록의 출처와 맥락을 드러내도록 이들 모델을 사용할 수 있다. 하지만, 앞서 기술한 출처와 맥락의 구분을 염두에 둔다면, 이들의 초점은 다르다. PROV 모델은 개체의 산출 과정을 보여주는 소급형 출처의 궤적을 표현하기 위한 모델이며, 특정한 데이터세트에 어떤 행위가 가해졌는

지를 미시적인 수준에서 표현해 주는 정보이다. 그런 이유로, 출처 표현을 과거형으로 읽도록 관계 화살표의 방향을 시간의 역순으로 연결한다. 이와 달리 기록관리 메타데이터 모델과 보존기록 기술 모델에서는 출처와 맥락의 개념이 명확하게 구분되지 않는다. 소급형 출처의 개념을 일부 메타데이터 그룹(예. 사건 이력 메타데이터 그룹)에서 찾아볼 수는 있지만, 이는 특정 개체 클래스에 부여되는 속성 수준이며, 보편적으로 적용하기 어렵다. 또, 기록의 산출과 변화, 이동 등의 궤적을 충실하게 기록화하기도 어렵고 기록의 의미를 전달하고 신빙성을 평가하기 위한 충분한 정보를 전달하기도 어렵다.

특히, 이승억과 설문원(2017)이 지적한 것처럼, 데이터세트로 불리는 데이터형 전자기록은 개별 시스템이 무수히 연결되는 환경에서 생산되며, 물리적인 관리 단위로 격리되기 어렵다. 즉, 물리적·법적·행정적 관리와 장기보존의 범위를 넘나드는 논리적 연계 구조를 갖는다. 나아가 사후적으로 출처에 따라 기록을 재구성하며 그 의미와 가치를 부여하기도, 증거력을 구성하기도 어렵다. 그러므로 실효성 있게 데이터형 전자기록을 관리 및 보존하고 그에 대한 신뢰를 보장하기 위해서는 기존의 원칙과 기준을 넘어서는 접근이 필요하다(이승억, 설문원, 2017).

일련의 업무 과정을 지원하는 시스템에서 산출되는 기록의 출처를 명확하게 유형화하여 표현하기 위해서 W3C PROV의 범위인 소급형 출처와 함께 ProvONE에서 수용한 전망형 출처 개념을 적용할 가능성을 탐구해 볼 만하다. 이 연구에서는 이후의 장에서 데이터형 전자기록의 소급형 출처와 전망형 출처를 함께 표현할 수 있는 모델을 설계할 방안을 모색할 것을 제안한다. 다만, ProvONE의 프로세스 출처의 경우는 현재 단계에서는 제외할 것이다. 과학 실험에 사용되는 프로그램과는 달리, 일반 업무 과정에 기록을 산출하는 시스템의 경우는 실제 실행 과정에 진화되기보다 시스템 설계에 따라 시스템 개발이 완료된 후 실행된다고 보았다. 또, 변화가 필요한 경우라 하더라도 업무 절차의 변화

에 따른 시스템 갱신 등의 재설계 과정을 거치므로, 전망형 출처 표현을 통해 그 변화를 추적할 수 있다고 판단하였다. 결국, 기록의 출처와 맥락 정보를 어떻게 소급형 출처와 전망형 출처로 범주화할 수 있는지 검토하고, 이를 표현하기 위한 모델을 설계할 방향을 제안한다.

4. 데이터형 전자기록의 출처 모델 개발을 위한 방향

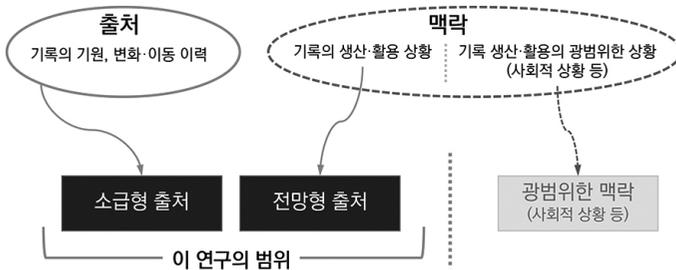
1) 소급형 출처와 전망형 출처 개념의 도입

이 연구는 기록과 기록을 생산한 행위자 및 활동의 직접적인 관계를 벗어나는 확장된 출처 개념을 지지한다. 기록을 생산하는 활동과의 직접적인 관계만으로 출처를 좁힌다면, 기록학에서 출처를 존중하는 의미와 출처의 역할을 충분히 드러내기 어려울 것으로 판단했기 때문이다. 또한, 기록의 생산과 변화 및 이동 이력을 추적할 수 있는 계보적 성격의 궤적으로서 출처 정보는, 기록에 의미를 부여하고 해석하도록 하는 맥락 정보와 구분해야 한다는 주장 역시 수용한다. 다만, 예를 들어 사회적 출처로 언급되는 광범위한 맥락 정보의 경우, 이 연구에서 다루는 출처 개념에서는 제외하였다. 너무 광범위한 맥락까지 출처의 범위에서 다룬다면, 여전히 분명하게 출처의 범주를 드러내고 기록과 함께 출처 기록을 생산하기 어려워질 것으로 판단했기 때문이다. 유사한 이유에서, 보존기록을 대상으로 하여 새롭게 기술(description)되는 출처 정보보다, 기록과 함께 시스템 기능성과 시스템 설계 과정을 통해 산출될 수 있는 출처 표현에 초점을 맞추었다.

종합하면, <그림 4>와 같이 기록의 기원과 궤적으로서의 출처 정보와 함께, 기록의 생산과 활용 상황을 전달할 수 있는 맥락 정보의 범위를 일부 포함해야만 시스템을 통해 산출되는 데이터형 전자기록에 대

한 신뢰를 더 잘 판단할 수 있고, 따라서 업무 활동의 증거로 활용할 가능성도 커질 것으로 보였다. 따라서 이 연구는 일부 맥락 정보의 범주를 출처, 즉 전망형 출처의 개념으로 포함하면서, 출처 개념을 다음과 같이 두 개의 유형으로 범주화할 것을 제안한다.

〈그림 4〉 출처 개념 확장과 모델의 수용 범위



첫째, 발생한 활동과 기록과의 직접적인 관계, 즉, 기록이 어디에서 왔는지를 보여주는 정보 범주를 소급형(retrospective) 출처로 유형화한다. 둘째, 맥락 정보 중에서 기록을 산출하는 활동에 대한 일반 정보, 업무 과정 또는 워크플로우 자체에 대한 정보, 조직 환경이나 구조를 포함해 활동 행위자를 이해할 수 있는 정보, 계획된 데이터의 입출력 방식 등의 정보 범주를 전망형(prospective) 출처로 유형화한다. 이미 반뷰셀은 ‘출처’와 ‘맥락’의 경계를 분명히 할 것을 제안하였지만, 또한 기록과 활동을 해석하는 데 활용할 수 있다면 ‘맥락’의 범위를 계속해서 확장할 수 있게 허용했다고 판단할 수 있다. 맥락의 범위를 확장해야 할 요구는 수용할 필요가 있다. 그럼에도 이 연구에서는, 기록의 생산과 활용 상황을 비교적 직접적으로 이해하기에 필요한 유형의 맥락 정보와 사후적인 기술(description) 과정에서 작성되어야 하는 사회적 상황과 같은 정보 유형을 모두 하나의 맥락 개념 범주로 묶는다면, 출처와

맥락에 대한 이해를 어렵게 할 것으로 판단하였다. 따라서, 맥락 정보 중에서 기록의 생산과 변화, 활용 등을 이해하고 기록의 신빙성을 지원하기 위해 직접적으로 필요한 정보를 오히려 출처 개념으로 수용하되, 기록의 기원을 표현하는 소급형 출처와 구분하여 전망형 출처의 범위로 분명하게 구분할 것을 제안한다.

우선, 소급형(retrospective) 출처는 과거에 활동이 실행되면서 데이터가 산출된 정보, 즉 어떤 행위자가 어떤 작업을 수행하였고, 그 과정에 어떻게 데이터 객체가 생산되거나 파생되었는지를 보여준다. 이러한 정보는 시스템에서 특정 행위가 수행되는 과정에 획득될 수 있게 구성될 수 있다. 이는 컴퓨터 및 데이터 공학에서 수용하고 있는 출처의 개념. 즉, '객체, 즉 데이터나 어떤 것을 생산하는 것에 관여하거나, 해당 객체에 영향을 주거나, 이를 전달하는 데 관여한 사람, 기관, 활동, 개체를 기술하는 기록'(Moreau et al., 2013)으로 설명할 수 있다. W3C의 PROV 모델이 수용한 출처 개념이며, Van Bussel(2017)과 설문원(2019)이 맥락과 구분하여 제안한 좁은 의미의 출처 개념과도 합치한다. 정리하면, 소급형 출처는 실행된 활동의 각 단계와 특정 데이터 산출물을 도출하는 과정, 이 과정에 행위자와 활동 및 개체는 어떤 관계로 상호작용하였는지 추적할 수 있는 정보를 표현한다. 컴퓨터 시스템의 작업 실행에 대한 일종의 상세 로그 정보일 수 있으며, 소급형 출처를 표현하기 위한 대표적인 출처 모델은 앞서 살펴본 PROV 규격이다.

다음으로, 전망형(prospective) 출처는 데이터를 산출하기 위해 정해진 "레시피"(Cuevas-Vicentín, 2016)로서 추상적인 워크플로우 정보를 표현한다. 행위자와 활동, 개체가 어떻게 특정 시스템 내·외부에서 상호작용해야 하는지에 대한 사전적 정보이므로, 이는 시스템 설계 과정에 획득되어야 하고, 기록과 관련된 행위의 맥락 정보를 표현할 수 있도록 한다. 구체적인 데이터 산출물 간의 상세한 의존성 및 계보 정보는 프로세스를 실행 후에만 이용할 수 있지만, 일반적인 계보에 대한 정보에

대해서는 프로세스를 실행하기 이전에도 확인할 수 있기(Luäscher, 2016) 때문이다.

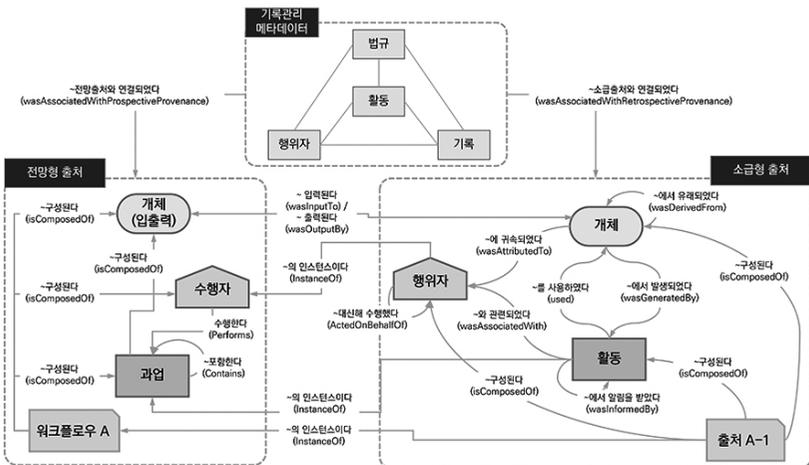
데이터의 기원에 대해 알려주는 소급형 출처가 ‘이 데이터세트가 어떤 과정을 거쳤으며 누가 생산에 관여하였나’의 질문의 답을 확인할 수 있게 해준다면, ‘왜 생산되었는가?’, ‘정해진 절차와 규칙, 지정된 사람 또는 대리인이 생산하였는가?’와 같은 질문에는 답하기는 어렵다(Missier et al., 2013; 재구성). 어떻게 데이터세트가 흘러왔는지에 대한 정보와 별개로, 미리 정의된 절차와 행위의 순서에 따라 데이터세트가 산출되었는지를 보여주는 정보는 데이터세트가 업무 활동을 얼마나 신뢰성 있게 재현해 줄 수 있는가의 문제(설문원, 2019)를 다루기 위해서도 필수적인 정보이다. 데이터 유형으로 생산되는 전자기록을 관리하려면, 행위가 발생한 이후의 궤적으로서 소급형 출처는 물론이고, 특정한 업무 과정 내에서 어떻게 기록을 산출되는지, 즉 업무 과정을 설명해 줄 수 있는 정보와 규정된 절차에 따라 기록이 산출되었는지를 판단할 수 있는 정보도 필요하다는 것이다. 기록의 개념을 탐구한 Duranti와 Thibodeau(2006)도 활동의 재현성을 지원하는 기록 유형에는 과거 행위를 기억할 수 있게 소급하여 보여주는 유형도 있지만, 미래 지향적인 기능을 수행하는 기록 유형이 존재함을 드러낸 바 있다. 이들은 이런 기록의 유형을 전망형 기록(prospective records)으로 정의하였으며, 이를 다시 행위나 프로세스 수행에 대한 지시사항을 포함하는 지시기록(예. 절차 매뉴얼, 법규, 연주 악보 등)과, 행위나 프로세스 수행에 개입하는 기반기록(예. 전자악기와 컴퓨터의 상호작용을 지원하는 소프트웨어 패치, 입력 데이터를 해석해 다음 출력물 표현을 결정하는 소프트웨어 등)으로 다시 구분하였다(Duranti & Thibodeau, 2006). 특히, 물리적으로 분산되어 생산·저장, 관리되는 디지털 기록의 특성을 고려한다면, 디지털 기록을 생산한 기반 규칙을 포함하여 기록의 구조와 의미를 이해하기 위한 정보는 필수적이다.

따라서 전망형 출처는 소스 코드나 스크립트, 정의된 워크플로우 정보까지 컴퓨터가 수행하는 작업의 명세(specification), 데이터 산출물이나 데이터 산출물 클래스를 생성하기 위해 따라야 하는 단계 등을 표현한다. 이미 Missier et al.(2013)도, 특히 정보시스템에서 워크플로우 실행과 관련하여 생산되는 데이터세트의 경우는 소급형과 전망형 출처 정보 모두를 표현할 수 있어야 한다고 주장하여, ProvONE의 기초가 된 모델을 제시하기도 하였다.

2) 기본 모델 설계의 방향 제안

이 연구에서는 데이터형 전자기록의 출처 모델을 지원하기 위해, 소급형 출처와 전망형 출처의 범주를 표현할 수 있도록 다음의 그림과 같이 기본 모델을 설계할 것을 제안한다.

〈그림 5〉 데이터형 전자기록을 위한 확장형 출처 모델의 기초 설계(안)



전제로 한 것은, 기본적으로 출처 모델은 출처 정보를 특정 기록관리 환경에서 정의한 기록과 기록관리 메타데이터와는 별도의 기록으로서 획득하고 관리하는 모델로 설계할 필요가 있다는 것이다. 출처 정보를 기록관리 과정에 활용하려면 출처 정보와 연결된 산출물로서의 데이터 형 전자기록과는 다른 데이터 계층이 필요(Missier, 2016)하다고 판단했기 때문이다. 또한, 기록은 개별 기록관리 환경의 범위에서 공식적으로 정의된 개체이므로, 시스템에서 이루어지는 프로세스를 따라 모든 개체나 활동의 궤적을 추적하기 어려운 형태로 기록화될 수 있다는 점도 고려해야 하기 때문이다.

출처 모델의 기초로 활용한 모델은 앞서 살펴본 PROV 및 ProvONE 모델이다. 다양한 영역과 표준적으로 출처 기록을 상호교환할 가능성을 고려한다면, 별개의 특수 모델보다 표준을 수용한 확장 모델을 개발하는 것이 효과적이라고 판단하였다.

우선 <그림 5>의 오른쪽 영역은 소급형 출처 영역으로, PROV 규격의 기본 모델을 수용하여 개발할 것을 제안한다. 즉, 데이터와 활동의 궤적, 행위자가 어떤 개체에 대해 어떤 활동을 수행했는지 등, 수행된 활동과 연관된 이력을 표현할 수 있도록 개체, 활동, 행위자와 이들의 관계를 설정하는 방식이다. 이를 위해서 PROV의 구성요소인 개체와 행위자, 활동 및 이들의 관계를 기초로 하되, 이 관계 네트워크를 출처 기록 단위의 논리적 개체로 구성하기 위해 ‘출처’ 구성요소를 추가할 것도 제안한다. ‘출처’는 특정 인스턴스를 기준으로 생성된 소급 출처 기록을 하나의 단위로 묶어주는 한편, 전망형 출처로 표현된 워크플로우가 실행된 인스턴스로서 이와 연결되고, 기록관리 메타데이터와 연결되는 단위로도 활용할 수 있을 것으로 보았다.

다음으로, <그림 5>의 왼쪽 영역은 전망형 출처 영역인데, ProvONE의 전망형 출처 표현을 통한 목표를 수용하되, 그 구성요소는 Lim et al.(2010)의 모델에서 구성한 요소를 참조해서 개발할 것을 제안한다.

ProvONE의 경우, 과학 실험을 위한 계산 프로그램을 중심으로 구성요소를 도출하였다. 하지만 이 연구에서는 개별 프로그램보다 과업 단위로 특정 워크플로우를 구성하는 것이 적절하다고 보았다. 이미 기록관리를 위한 업무분석 표준이 개발되었으므로, 표준에 근거한 업무분석 결과를 활용하여 워크플로우와 과업, 과업 수행자와 입출력 개체를 정의함으로써 전망형 출처를 구성할 수 있다고 판단하였다. 즉, 기록 생산과 활용의 기반이 되는 조직의 정책과 업무 프로세스, 법제 정보에서부터 기록의 생산 및 활용에 쓰이는 소프트웨어 알고리즘, 나아가 이 과정을 이해하기에 필요한 조직 환경이나 구조를 포함해 기록 산출 과정을 이해할 수 있게 하는 일반 정보 등을 전망형 출처 영역의 구성요소를 통해 표현하도록 하였다. 이런 유형의 정보는 대부분 정보시스템을 개발하는 과정에 함께 정의될 수 있을 것이다. 따라서 정보시스템을 통해 활동이 수행되기 이전에 산출되고, 특정 워크플로우가 실행될 때 산출되는 소급형 출처 기록과 인스턴스화 관계로 연결될 수 있을 것이다.

이렇게 산출된 소급형 출처 기록과 전망형 출처 기록은 특정 기록관리 환경에서 정의된 기록 및 기록관리 메타데이터와 연결되어야 할 것이다. 따라서 출처 기록은 다시 <그림 5>의 상단 영역과 같이 특정 기록관리 환경에서 정책적으로 정의된 기록 및 기록관리 메타데이터와 연결되어, 그 출처를 분명하게 표현할 수 있도록 설정할 것을 제안한다.

이러한 연결 관계를 통해서 각 출처 기록의 기능에 따라 분명하게 출처의 유형을 구분해 줄 수 있을 것이다. 또 기록관리 메타데이터로 수용하기 어려운 수준까지 소급형 출처와 전망형 출처를 기록화함으로써 조밀하고 풍부하게 특정 기록의 출처를 표현하는 효과도 얻을 수 있을 것으로 기대한다.

5. 출처 모델의 활용 가능성 및 맺음말

이 연구는 데이터형 전자기록의 출처 표현 모델을 개발하기 위한 초기 단계로 진행되었다. 특히 이를 위해 출처와 맥락의 개념을 소급형 출처와 전망형 출처 개념으로 전환할 것과, 기록관리 메타데이터와는 별개로 출처를 표현하고 기록과 연결할 수 있는 모델을 개발할 것을 제안하였다. 단일 생산자를 기준으로 기록 집합에 “구속복”(Cook, 1997)을 입히듯 출처를 연결하는 방법은 이미 과거 전통으로 인식된다. 1960년대 중반, 피터 스콧(Peter Scott)의 제안으로 기록물군 대신 시리즈를 기록 통제 단위로 채택한 호주 국가기록관리기관의 결정은 기록과 생산 맥락을 다중으로 연결할 수 있는 유동적 관계로 바꾸었다(Cunningham, 2016). ICA도 Records in Context의 이름으로 디지털 환경에서 기록과 생산 맥락의 다중 연결을 지원할 통합 기술(description) 도구를 개발했다.

이 연구는 여기에 그치지 않고, 기록 표현 모델과 구분되는 출처 표현 모델을 개발할 것을 제안하였다. 기록과 연관되면서 독립적으로 출처를 표현할 수 있는 모델을 개발하여, 오히려 기록의 유동성을 보장할 수 있고, 기록의 속성과 이를 지원할 출처의 역할을 더 충실히 수행할 수 있을 것이라 보았다. 이 연구가 제안하는 기본 모델을 활용하여 개발될 출처 모델은 기록의 고정성과 활동의 재현성, 재현의 신빙성을 지원함으로써 디지털 환경에서 필요한 출처의 역할을 할 수 있도록 개발되어야 한다.

첫째, 이후 개발될 출처 모델은 기록의 고정성을 확보하기 위한 과정을 지원할 수 있을 것이다. 동적 디지털 환경에서도 활동을 지속해서 재현하기 위해서는 기록이 갖는 고정성의 개념을 배제하기는 아직 어렵다. 반 뷰셀과 존커(Jonker)를 비롯한 많은 기록학 연구자가 여전히 이를 지지하거나, 적어도 케텔라(Ketelaar)와 같이 기록 획득의 중요성을 인정한다(설문원, 2019). 그만큼 고정성은 기록과 연결된 활동을 재

현하는데 기본 전제가 되기 때문이다. 그런데, 데이터의 도처편재성(ubiquity)이 특징인 디지털 환경에서, 고정성은 정책적 결정에 따른 것(Glaudemans, Jonker & Smit, 2017)일 수밖에 없다. 따라서 '무엇을 고정할 것인가, 왜 고정하는가, 어떻게 고정할 것인가'(Jonker, 2017)를 결정할 수 있어야만, 유동적 디지털 환경에서 생산되는 데이터형 기록에 대해서도 고정성을 확보할 수 있게 될 가능성이 높다.

이 연구가 제안하는 방향에 맞추어 출처 모델을 개발하여 활용하면, 이와 같은 고정의 대상과 근거, 정보기술을 활용하여 고정할 방법 등을 결정할 정책 지원 도구로서 역할을 할 것이라고 기대한다. 소급형 및 전망형 출처 기록을 함께 사용함으로써, 정보시스템의 업무 과정 및 행위자와 연계하여 그 실행에서 생산·전달된 데이터의 궤적을 미시적으로 드러내 줄 수 있다. 이를 근거로 '무엇을 기록으로 고정·획득할 것인가'의 과제를 상향식 관점에서 업무와 데이터 연계 속에 해결할 수 있다고 판단한다. 또한, 출처 기록 자체가 기록의 고정성을 대신할 가능성도 있을 것인데, 데이터의 유동성을 수용하는 한편, 데이터형 전자 기록의 궤적에 따라 기록화된 소급형 출처 기록을 통해 기록의 고정성에 대한 요구를 상당 부분 해소해 줄 수 있을 것으로 판단한다.

둘째, 기록의 전체 생애주기에 걸쳐 활동을 재현하도록 할 수 있을 것이다. 동적 환경에서 활동의 재현성(reproducibility)과 출처 사이의 관계성은 기록관리나 과학 연구 및 실험을 포함하여 정보시스템을 사용하는 다양한 영역이 공통으로 직면하고 있는 과제 영역(Lemieux & imProvenance Group, 2016)이다. 더욱이, 이질적인 분산 시스템 환경에서 생산되는 데이터를 증거로 활용하려면, 기본적으로, 데이터를 산출한 프로세스가 종료된 이후에도 지속해서 이를 재현할 수 있어야 한다. 활동의 재현성은 기록의 내용과 맥락, 구조를 필요할 때 재구축할 수 있는 능력(Van Bussel, 2017)을 의미하기 때문이다. 또, 이 과정에는 기록의 기원을 드러내는 소급형 출처와 기록 산출의 규칙이나 워크플로우 정보와 같은

전망형 출처를 모두 활용해야 한다.

예를 들어, 일련의 차량 충돌 실험을 수행하였고(활동), 그에 따라 실험 데이터세트(개체)가 산출되었음을 표현하는 것은 소급형 출처 기록이다. 하지만 이것만으로는 해당 실험 과정을 지속해서 재현하는데 한계가 있다. 이를 동영상으로 촬영하는 것만으로 부족하다. 필요할 때 과거의 활동을 재현할 수 있어야 한다는 의미에서 본다면, 충돌 실험 과정에 사용된 소프트웨어, 실험을 위해 설정된 복합적인 환경 등에 해당하는 정보를 담은 출처 기록이 함께 포착되어야 한다. 이 연구가 제안하는 출처 모델을 기반으로 출처를 기록화하면 소급형 출처와 전망형 출처 기록을 연계하여 활동을 기록화할 수 있게 한다. 또한 필요한 시기에 산출된 데이터와 출처 기록을 함께 연계함으로써 당시의 활동을 지속해서 재현할 가능성을 지닐 것이다. 이를 통해, Zhang(2012)과 설문원(2017)이 해석했던 것처럼, “생산, 관리, 이용 과정에서 이루어지는 모든 행위와 그 관계”를 확보할 수 있는 디지털 환경의 원질서 개념도 의미 있게 다룰 수 있을 것이다.

셋째, 기록에 대한 신빙성, 나아가 기록을 통해 재현된 활동의 신빙성을 판단하도록 지원할 수 있을 것이다. 기록화된 사실의 증거로서 기록의 신빙성은 기록이 진본이고 신뢰할 수 있는 것이라는 믿음과 관련되는데, 이는 기록의 특성 분석과 평가(assessment), 특히 출처의 평가 결과인 경우가 많다(Yeo, 2013). 그런데 앞서 지적한 것처럼, 특히 분산·네트워크 컴퓨팅 환경의 맥락에서는, 객체의 신빙성을 확립하고 평가하기 위한 전통적인 방법론으로서 신뢰할 수 있는 보관소(Duranti, 1989) 개념을 넘어서 더 직접적인 신빙성 평가 방법론을 적용해야 하는 도전이 계속되고 있다(Lemieux & imProvenance Group, 2016).

이 연구에서 제안한 출처 모델을 개발하여 소급형 출처와 전망형 출처 기록을 산출한다면, 활동의 재현물인 기록을 신뢰할 수 있는지, 그리고 종합적으로 기록이 신빙성(trustworthiness)을 갖추었는지 평가할

도구로 역할을 할 수 있다. 기록관리 국가표준(KS X ISO 15489-1)은 기록의 신뢰성을 “기록이 입증하고자 하는 처리행위, 활동 또는 사실의 완전하고 정확한 표현물로서 해당 기록의 내용을 신뢰할 수 있고, 이후의 처리행위 또는 활동 과정에서 의존할 수 있는 것”으로 설명한다. 기록관리는 “맥락 및 구조와의 고정적 결합성을 유지하는 전략에 근거하여 ‘신빙성’을 추구”(설문원, 2019)하며, 이는 기록의 신뢰성을 포함하는 품질 요건을 종합적으로 평가하여 판단할 수 있는 것(현문수, 2022; Makhoulf Shabou, 2015)으로 받아들여지고 있다. 이 연구가 제안하는 출처 모델이 개발된다면, 소급형 출처 기록을 통해서도 컴퓨팅 환경 내에서 이루어진 활동과 기록의 흐름을 구체적으로 표현할 수 있으며, 전망형 출처 기록을 활용하여서는 산출된 기록의 구조와 활동 맥락, 업무 기반 규칙과 함께 기반 규칙에 따라 활동이 수행되고 기록이 산출되었는지 등을 설명할 수 있을 것이다. 지속적인 활동의 재현성을 확보한 기반 위에, 이후 처리행위와 활동 과정에 의존할 수 있도록 해주어, 기록으로서 데이터세트의 신뢰성과 신빙성을 보장하는 역할을 할 것이다.

이에 후속 연구를 통해서, 이 연구에서 제안하는 기초 모델을 구체화한 출처 표현 모델을 개발할 것이며, 그 활용 가능성 역시 함께 탐구하고자 한다. 출처 모델의 구성요소와 이들의 관계 구조를 구체화하여 설계할 것이며, 기록 및 기록관리 메타데이터와의 연결 관계도 더 세밀하게 구성하여, 디지털 환경에 맞게 출처로서 역할을 할 수 있는지 검토할 것이다. 또한, 실상 대부분의 전자기록이 데이터형 전자기록으로 생산되고 있다고 볼 수 있으므로, 전자기록의 일반 출처 모델로 사용될 가능성도 함께 고려하여 향후 모델 개발이 진행될 것을 기대한다.

〈참고문헌〉

- 김수영 (2022). 행정정보시스템 분류방식을 적용한 행정정보 데이터세트 기록관리의 실제적 실행연구. 기록과 정보·문화 연구, 14, 55-88.
- 김수영 (2023). 행정정보 데이터세트 이관 방안 적용 사례 연구. 기록과 정보·문화 연구, 16, 7-50.
- 변우영, 임진희 (2022). 행정정보 데이터세트 이관도구 SIARD_KR의 개선방안. 정보관리학회지, 39(1), 195-217.
- 설문원 (2017). 기록관리 원칙의 해석과 적용을 위한 담론 분석: 출처주의를 중심으로. 기록학연구, 52, 59-117.
- 설문원 (2019). 기록이란 무엇인가: 활동의 고정적 재현물로서의 개념 탐구. 기록학연구, 59, 5-46.
- 양동민, 최광훈, 김지혜, 유남희 (2023). 행정정보 데이터세트의 이관규격의 다양화 및 재현 방안에 관한 연구. 정보관리학회지, 40(4), 167-200.
- 이승억, 설문원 (2017). 전자기록관리정책의 재설계에 관한 연구. 기록학연구, 52, 5-37.
- 이정은, 양동민 (2023). 데이터세트 유형 전자기록의 필수보존속성 연구. 한국비블리아학회지, 34(4), 259-283.
- 이정은, 김지혜, 왕호성, 양동민 (2022). 행정정보 데이터세트의 관리기준표 개선방안 연구. 한국기록관리학회지, 22(1), 177-200.
- 한국기록학회 (2008). 기록학 용어 사전. 서울: 역사비평사.
- 현문수 (2022). 전자기록의 신뢰가치 확립을 위한 증거능력 구현체계 연구: 우리나라 공공 전자기록의 신뢰가치 모델 개발을 중심으로. 기록학연구, 73, 5-46.
- Bearman, D. & Lytle, R. (1985). The power of the principle of provenance. *Archivaria*, 21, 14-27.
- Buneman, P., Khanna, S. & Tan, W. (2001). Why and where: A characterization of data provenance. In Van den Bussche, J. & Vianu, V. eds, *Database Theory — ICDT 2001*. ICDT 2001. Lecture Notes in Computer Science, vol 1973. Berlin: Springer. Available: <http://db.cis.upenn.edu/DL/whywhere.pdf>
- Car, N. (2017.3.15). Intro to PROV. Provenance and Social Science Data Webinar. Available: <https://www.youtube.com/watch?v=elPcKqWoOPg>
- Cook, T. (1997). What is past is prologue: A history of archival ideas since 1898, and the future paradigm shift. *Archivaria*, 43, 17-63.
- Corrado, E. & Sandy, H. (2017). *Digital Preservation for Libraries, Archives, and Museums* (2nd ed.). Lanham: Rowman & Littlefield. [e-Book].

- Cuevas-Vicentín, V., Ludäscher, B., Missier, P., Belhajjame, K., Chirigati, F., Wei, Y., Dey, S., Kianmajd, P., Koop, D., Bowers, S., Altintas, I., Jones, C., Jones, M., Walker, L., Slaughter, P., Leinfelder, B. & Cao, Y. (2016). ProvONE: A PROV Extension Data Model for Scientific Workflow Provenance. Available: <https://jenkins-1.dataone.org/jenkins/view/Documentation%20Projects/job/ProvONE-Documentation-trunk/ws/provenance/ProvONE/v1/provone.html#>
- Cunningham, A. (2016). Describing Archives in Context: Peter J Scott and the Australian 'Series' System. In Lemieux, V. ed. *Building Trust in Information: Perspectives on the frontiers of provenance*. Cham: Springer International Publishing, 49-57.
- Duranti, L. (1989). The odyssey of records manager: Part I: From the dawn of civilization to the fall of the Roman Empire. *ARMA Quarterly*, 23(3), 3-11.
- Duranti, L. (1997). The Archival Bond. *Archives & Museum Informatics*, 11(3-4), 213-218.
- Duranti, L. & Thibodeau, K. (2006). The concept of records in interactive, experiential and dynamic environments: The view of InterPARES. *Archival Science*, 6, 13-68.
- Freire, J., Koop, D., Santos, E. & Silav, C. T. (2008). Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 10(3), 11-21. DOI: 10.1109/MCSE.2008.79
- Fu, L. (2015). Automatic provenance capturing for research publications. Doctoral dissertation, Rensselaer Polytechnic Institute, United States.
- Gil, Y., Miles, S., Belhajjame, K., Deus, H., Garijo, D., Klyne, G., Missier, P., Soiland-Reyes, S. & Zednik, S. (2013). PROV model primer. Available: <https://www.w3.org/TR/prov-primer/>
- Glaudemans, A., Jonker, R. & Smith, F. (2017). Beyond the traditional boundaries of archival theory: An interview with Eric Ketelaar. In Smit, F., Glaudemans, A. & Jonker, R. eds. *Archives in Liquid Times*. s'-Gravenhage: Stichting Archiefpublicaties, 73-91.
- Goudarouli, E., Sexton, A. and Sheridan, J. (2019). The Challenge of the Digital and the Future Archive: Through the Lens of The National Archives UK. *Philosophy & Technology*, 32, 173-183. <https://doi.org/10.1007/s13347-018-0333-3>
- ICA (2000). ISAD(G): General International Standard Archival Description (2nd ed.). Available: https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf

- ICA EGAD (2023a). Records in Context: Conceptual Model, Version 1.0. Available: https://www.ica.org/sites/default/files/ric-cm-1.0_0.pdf
- ICA EGAD (2023b) Records in Contexts: Foundations of archival description, Version 1.0. Available: https://www.ica.org/sites/default/files/ric-fad-1.0_1.pdf
- InterPARES2 (n.d.). InterPARES 2 Terminology Database. Available: http://www.interpares.org/ip2/ip2_terminology_db.cfm
- Jonker, R. (2017). A perfect match? Connecting partners in the labyrinth of information. In Smit, F., Glaudemans, A., & Jonker, R. eds. *Archives in Liquid Times*. s'-Gravenhage: Stichting Archiefpublicaties, 297-305.
- Lemieux, V. (2016). Preface. In Lemieux, V. ed. *Building Trust in Information: Perspectives on the Frontiers of Provenance*. Cham: Springer International Publishing, vii-viii.
- Lemieux, V. & the imProvenance Group (2016). Provenance: Past, present and future in Interdisciplinary and multidisciplinary perspective. In Lemieux, V. ed. *Building Trust in Information: Perspectives on the Frontiers of Provenance*. Cham: Springer International Publishing, 3-45.
- Lim, C., Lu, S., Chebotkot, A. & Fotouhi, F. (2010). Prospective and retrospective provenance collection in scientific workflow environments. 2010 IEEE International Conference on Services Computing, 449-456. Miami, FL.
- Luäscher, B. (2016). A brief tour through provenance in scientific workflows and databases. In Lemieux, V. ed. *Building Trust in Information: Perspectives on the Frontiers of Provenance*. Cham: Springer International Publishing, 103-126.
- Makhlouf Shabou, B. (2015). Digital diplomatics and measurement of electronic public data qualities: What lessons should be learned? *Records Management Journal*, 25(1), 56-77. <https://doi.org/10.1108/RMJ-01-2015-0006>
- Michetti, G. (2016). Provenance: An archival perspective. In Lemieux, V. ed. *Building Trust in Information: Perspectives on the Frontiers of Provenance*. Cham: Springer International Publishing, 59-68.
- Missier, P. (2016). The lifecycle of provenance metadata and its associated challenges and opportunities. In Lemieux, V. ed. *Building Trust in Information: Perspectives on the Frontiers of Provenance*. Cham: Springer International Publishing, 127-137.
- Missier, P., Dey, S., Belhajjame, K., Cuevas-Vicenttin, V. & Ludäscher, B. (2013). D-PROV: Extending the PROV provenance model with Workflow structure. *Proceedings of the 5th USENIX Workshop on the Theory and Practice of Provenance (TaPP 13)*, USENIX Association, Berkeley. Available: <https://www.usenix.org/system/files/conference/tapp13/tapp13-final3.pdf>

- Moreau, L. (2010). The foundations for provenance on the web. *Foundations and Trends in Web Science*, 2(2-3), 99-241.
- Moreau, L., Missier, P., Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., Miles, S., Myers, J., Sahoo, S. & Tilmes, C. (2013). PROV-DM: The PROV data model. Available: <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>
- SAA (2005-2023). Dictionary of archives terminology. Available: <https://dictionary.archivists.org/entry/provenance.html>
- Sweeney, S. (2008). The ambiguous origins of the archival principle of "provenance". *Libraries & the Cultural Record*, 43(2), 193-213.
- Van Bussel, G. J. (2017). The theoretical framework for the 'archive-as-is'. An organization oriented view on archives: Part II. An exploration of the 'archive-as-is' framework. In Smit, F., Gludemans, A. & Jonker, R. eds. *Archives in Liquid Times*. s'-Gravenhage: Stichting Archiefpublicaties, 43-71.
- Yeo, G. (2007). Concepts of Records(1): Evidence, information, and persistent representations. *American Archivist*, 70(2), 315-343.
- Yeo, G. (2013). Trust and context in cyberspace. *Archives and Records*, 34(3), 214-234. DOI: 10.1080/23257962.2013.825207
- Zhang J. (2012). Original order in digital archives. *Archivaria*, 35, 167-193.

〈표준〉

문헌정보-기록관리-제1부: 개념과 원칙. KS X ISO 15489-1:2016.