

# PROV의 확장에 기초한 데이터형 전자기록의 출처 모델 연구\*

A Study on Developing a Provenance Conceptual Model for Data-driven  
Electronic Records Based on Extending W3C PROV

현문수(Hyun, Moonsoo)\*\*

1. 들어가며
2. 모델의 개발 방향
3. 개발을 위한 기초 모델 분석
  - 1) PROV: 소급형 출처 표현의 기초 모델
  - 2) P-Plan 및 ProvONE: 전망형 출처 표현의 기초 모델
4. 기초 모델의 확장을 위한 분석
  - 1) 전망형 출처 표현 영역
  - 2) 소급형 출처 표현 영역
5. 데이터형 전자기록을 위한 출처 모델 개발(안)
  - 1) 모델 구성
  - 2) 적용 예시
6. 맺음말

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

\*\* (전)부산대학교 사회과학연구원 전임연구원, (전)부산대학교·이화여자대학교·충남대학교 문헌정보학과 강사(moonsoo925@gmail.com).

■ 투고일: 2024년 03월 31일 ■ 최초심사일: 2024년 04월 01일 ■ 최종확정일: 2024년 04월 11일.

■ 기록학연구 80, 5-41, 2024, <https://doi.org/10.20923/kjas.2024.80.005>



## 〈초록〉

이 연구는 데이터형 전자기록을 대상으로 한 출처 표현 모델의 개발 방향에 맞추어 모델을 개발할 목적으로 진행되었다. 데이터형 전자기록의 생산·관리를 위해 출처와 맥락의 개념 구분을 지지하며, 이를 구분하여 표현할 수 있는 확장형 출처 모델을 제시할 것을 목표로 하였다. 이를 위해 W3C PROV를 기초 모델로 활용하며, P-Plan과 ProvONE도 일부 참고하였다. 이후, 기록관리 요건을 드러내고, 이를 바탕으로 기초 모델을 일부 확장하였다. 이 연구가 제안한 출처 모델은 데이터형 전자기록의 소급형 출처와 전망형 출처를 각각 표현하고 연결할 수 있도록 설계되었다. 향후 기록학 영역에서 출처 개념을 논의하고 모델을 확장해 나갈 수 있기를 기대한다.

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

## 〈Abstract〉

This study was conducted to develop a provenance representation model for data-type electronic records. It supports the distinction between provenance and context for the creation and management of data-type electronic records. To express both, it aims to design an extensible provenance model. For this purpose, W3C PROV is utilized as a basic model, with P-Plan and ProvONE for designing prospective provenance area. Afterward, the provenance model was extended by mapping the record management requirements. The provenance model proposed in this study is designed to represent and connect both retrospective and prospective provenance of data-type electronic records. Based on this study, it is expected to discussing the concept of provenance in the records management and archival studies area and to extending the model in the future.

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

## 1. 들어가며

일상생활과 업무 환경에서 정보기술을 활발하게 사용하면서 다양한 형태로 데이터가 생산되고 있다. 2022년 12월을 기준으로 302개 행정기관(중앙행정기관, 입법·사법·헌법·독립기관, 지방자치단체)이 보유하고 있는 정보시스템은 11,479개이며, 512개 공공기관까지 포함하면 전체 17,090개 시스템이 운영 중이다(한국지능정보사회진흥원, 2023, 8). 모바일 앱 사용 통계는 제외한 수치다. 이처럼 데이터의 생산과 활용이 일상화됨에 따라, 기록학 분야는 물론이고 정보학, 컴퓨터공학, 데이터공학 등 다양한 분야에서 데이터 관리와 활용의 문제를 다룬다. 그중에서도 특히 과학연구 영역에서 최근 집중해 논의되는 영역은 데이터의 출처 표현의 문제이다. 연구 결과의 원자료로서 데이터의 중요성이 커짐에 따라 ‘데이터 출처(data provenance)’의 이력을 확인하고 연구의 신뢰성을 확보하기 위해 이를 활용할 수 있는 방안을 모색(Lemieux & imProvenance Group, 2016)하게 된 것이다. 이미 2000년대 중반부터 시작된 데이터 출처에 관한 논의는 2013년 데이터 출처를 표준적으로 표현할 수 있는 W3C의 PROV 모델 개발로 이어졌고, 이 외에도 다양한 연구 영역에서 분야별 요구를 반영하여 모델 확장을 진행하고 있다.

반면, 우리나라 기록학 연구 영역에서 데이터형 전자기록의 출처에 대한 논의(이승억, 설문원, 2017; 설문원, 2019; 현문수, 2024a)는 최근에 일부 진행되었으며, 이 중 현문수(2024a)가 전자기록의 출처 모델을 개발할 필요성과 방향을 제시했을 만큼 구체적으로 모델 개발을 다룬 연구도 소수이다. 이 외에 박미영(2019)은 국방 인사 정보체계에서 행정정보데이터세트

획득을 위해 PROV 기본 모델과 Dublin Core 메타데이터 모델의 적용 가능성을 추상적 수준에서 검토하였다. Li와 Sugimoto(2014)는 디지털 보존에 있어 PREMIS 데이터 사전과 PROV 기본 모델의 적용 가능성을 출처 기술(description) 메타데이터의 관점에서 다루었다. Prom(2016. 6. 8.)은 클라우드 환경에서 보존기록의 조직, 즉 정리와 기술의 측면에서 PROV 모델의 가능성을 간략히 검토하였다. 그런데, 이들은 여전히 전통적 접근, 즉 맥락의 측면에 더 큰 의미를 두고 있었다. 또한 이들 연구는 추상적 수준에서 출처 모델을 다루어 구체적인 업무 흐름을 충분히 반영하지 못한 만큼, 실질적인 모형화나 기록관리 방법론으로서 적용하기 어렵다고 판단된다. 최근, ICA EGAD는 보존기록 기술 모델인 Records in Context(RiC)를 개발하는 과정에 PROV에 대한 언급이 없다는 지적(InterPARES Trust, 2019)을 받았지만, 최종 모델에서도 여전히 보존기록관리 현장에서 RiC를 쉽게 구현하기 위해 PROV의 요소나 그와의 연결 관계를 고려하지 않는다(ICA EGAD, 2023)고 밝힌다. 하지만, 이 연구에서는 현문수(2024a)가 주장한 것처럼, 기록학 영역에서 데이터 출처의 개념을 수용하면서도 이력으로서의 출처와 생산 환경으로서의 맥락을 구분하여 표현할 수 있는 출처 모델을 개발할 필요가 있다고 판단한다. 이에 데이터형 기록이 생산되는 시스템 환경에 확장된 출처 개념을 도입하고 기록화할 수 있는 새로운 출처 모델을 개발하기 위한 연구가 필요하다고 보았다.

이 연구는 데이터형 전자기록을 대상으로 한 출처 표현 모델의 개발 방향(현문수, 2024a)에 맞추어 모델을 개발할 목적으로 진행되었다. 데이터형 전자기록의 생산·관리를 위해 출처와 맥락의 개념 구분을 지지하며, 이를 구분하여 표현할 수 있는 확장형 출처 모델을 제시할 것을 목표로 하였다. 이를 위해 데이터의 이력을 보여주는 소급형 출처 모델로 개발된 W3C PROV 모델을 활용하며, 생산과 관리 및 보존 맥락을 표현할 수 있는 전망형 출처 표현을 위해 P-Plan과 ProvONE도 참고하였다. 이들 모델을 기초로 기록으로서의 데이터를 다루기 위해 기록관리 요건을 드러낸 바탕 위에

모델을 확장하였다. 이 연구는 기록의 출처를 포괄적으로 표현하기 위한 모델 제안을 목적으로 하지 않았다. 또 모든 현상이나 시스템에 수정 없이 적용할 것을 목적으로 하지도 않았다. 대신, 이 연구는 데이터형 전자기록의 출처와 맥락을 구분(설문원, 2019; 현문수, 2024a; Van Bussel, 2017) 하면서도 이를 연결하는 시도를 함으로써, 향후 기록학 영역에서 출처 개념을 논의하고 모델을 확장해 나갈 수 있도록 일반적이고 시험적 수준에서 모델을 개발하였다.

## 2. 모델의 개발 방향

이미 출처를 기록관리 메타데이터와 구분하여 기록화할 필요성을 드러내고 이를 위해 어떤 원칙에 따라 데이터형 전자기록의 출처 모델을 개발해야 하는지의 방향이 제안된(현문수, 2024a) 바 있다. 핵심 방향 중 하나로, 기록의 이력으로서 협의로 정의되는 출처와 일부 환경적 맥락을 포함하는 출처 개념을 함께 수용하는 모델을 개발할 필요성을 기본적으로 드러냈다. 즉, 기록을 생산·축적한 개인이나 조직, 그리고 기록을 산출한 업무 활동과 기록 사이에 존재하는 직접적인 관계는 물론이고 일정 정도로 확장된 관계까지 출처로 표현할 모델의 개발을 목표로 할 것을 밝힌 바 있다.

따라서 이 연구에서도 “소급형(retrospective) 출처”와 “전망형(prospective) 출처”의 두 영역을 연결하여 표현할 수 있는 출처 모델의 개발을 목표로 한다. 출처를 통해 기록을 이해하고 품질을 확보하려면 기록의 기원과 이력을 드러내는 한편 기록의 생산 및 활용 상황을 해석하게 도와주는 정보가 필요(현문수, 2024a)하므로, 두 출처 영역을 포괄하는 것이 적절하다고 판단하였다. 먼저, 소급형 출처는 기록의 궤적을 시간의 역순으로 표현하기 위한 것으로, 소급형 출처 모델을 수용함으로써 어떤 업무시스템에서 이행되는 과업의 결과로 산출되는 기록에 대해 어떤 객체와 행위자가 영향을 주면

서 기능이 수행되어 대상 기록이 생산되었는지(Moreau et al., 2013) 기술할 수 있도록 지원한다. 이 연구에서는 소급형 출처 모델 영역의 개발을 위해 W3C의 PROV 출처 모델을 기초로 사용하며, 데이터형 전자기록의 생산과 관리 맥락을 반영하도록 이를 확장할 것이다. 3장에서 분석한 것처럼 PROV는 표준적이고 범용적 성격을 지녀, 우리의 요구를 수용하면서도 인접 영역과 상호운용할 수 있을 뿐 아니라 출처와 맥락 간 개념을 구분하는 최근의 기록학적 연구(설문원, 2019; 현문수, 2024a; Van Bussel, 2017)를 반영할 수 있기 때문이다. 다음으로 전망형 출처는 기록을 산출하는 일종의 “레시파”(Cuevas-Vicenttín et al., 2016)를 표현하는데, 과업을 수행하도록 정의된 행위자와 개체는 물론이고 특정 정보시스템에서 정해진 과업을 수행하면서 어떤 규제나 통제를 따라야 하는지도 사전적으로 표현할 수 있는 출처 영역(현문수, 2024a)이다. 이 연구에서는 P-Plan과 ProvONE을 기초로 하여 전망형 출처 모델 영역을 개발하되, 규제와 절차 등을 정의할 때 데이터형 전자기록의 생산과 관리 맥락을 반영할 수 있도록 확장할 것이다.

또, 현문수(2024a)는 설문원(2019)이 정의한 기록의 속성을 기준으로 기록의 출처가 이와 어떻게 관련될 수 있는지 분석한 후, 데이터형 전자기록의 출처 모델이 기록의 고정성과 기록을 통한 활동의 재현성 및 그 신빙성 확보라는 측면에서 출처의 역할을 지원할 것을 요구하였다. 특히, 유동성을 특징으로 하는 디지털 환경에서도 여전히 고정적 기록의 속성을 고수하는 이유는 활동을 신빙성 있게 재현할 수 있는 능력이야말로 기록이 갖는 핵심적인 속성(설문원 2019)이기 때문이다. 기록이 고정성을 확보할 때 과거 발생했던 활동을 신뢰할 수 있는 수준으로 재현할 수 있다(현문수, 2024a)는 의미이다. 이러한 방향을 수용하여 이 연구가 개발할 모델은 소급형 출처와 전망형 출처를 연계함으로써 데이터형 전자기록의 발생과 진화, 관리 이력을 업무맥락과 함께 기록화할 수 있게 지원할 것이다. 기록화된 출처 정보를 근거로 첫째, 특정 업무 환경에서 기록으로 선언 및 획득할 대상을 구체적으로 결정하거나, 둘째, 유동적 기록을 대신해 기록의 고정성을 확보하는 대안

으로 기록화된 출처 정보를 활용할 가능성을 검토해 볼 수 있을 것이다(현문수, 2024a). 활동의 재현성은 기록을 통해 과거에 실행된 특정 프로세스를 추적할 수 있는 능력과 연결되므로, 소급형 출처의 기능과 연결된다. 그에 대한 신빙성은 사전적으로 정의된 업무 규범이나 절차 및 다양한 기반 규칙에 따라 기록이 생산되었는지 확인할 수 있는 능력이므로, 전망형 출처의 기능과 연결된다. 특히, 기록의 속성을 지원할 수 있도록 전자기록에 대한 신빙성 평가 기준을 활용할 수 있다고 판단하여, 기초 모델을 확장하는 과정에 “공공 전자기록의 신뢰가치 모델(안)”(현문수, 2022)을 참고하기로 하였다.

### 3. 개발을 위한 기초 모델 분석

#### 1) PROV: 소급형 출처 표현의 기초 모델

소급형 출처 표현을 위해 이 연구에서 활용한 기초 모델은 W3C 표준으로 채택된 PROV이다. PROV는 실험 과정에 데이터를 다루는 과학 분야를 중심으로 개발된 데이터 출처 모델이다. 다양한 정보 모델과 메타데이터 표준이 존재하지만, 데이터의 궤적으로서 계보(lineage)를 충분히 기록화하고 추적하려면 별도의 표현 모델이 필요하다는 요구(Car, 2017. 3. 15.)가 2000년대 중반부터 있었고, OPM(Open Provenance Model) 개발도 2006년부터 진행(OPM, n.d.)되고 있었다. 이러한 노력을 이어받아 2009년 9월부터 W3C를 중심으로 출처 모델을 본격적으로 개발하기 시작해, 2013년 PROV 모델 최종안을 발표(Missier, 2016)했다. PROV는 학술 연구의 근거가 되는 데이터의 출처 정보가 중요하게 인식됨에 따라 출처 정보를 표준적으로 표현하고 획득하는 방안(Cuevas-Vicentín et al., 2016)으로 개발되었다. 그 목적에 맞게 PROV는 어떤 개체가 발생한 이력을 소급하여 표현할 수 있게 설계되었다. 즉, 어떤 활동이 발생할 때 사용했거나 산출

한 개체는 무엇이며, 어떤 행위자가 이와 연결되었는지 과거로 계속해서 추적하여 그 이력과 기원에 대한 정보를 모아준다(Missier, 2016). 이를 위해 PROV가 채택한 구성 개념은 다음과 같이 유형화할 수 있다.

〈표 1〉 PROV의 구성(Moreau et al., 2013, 재구성)

영역	명칭	구분	설명
영역 1. 개체와 활동	개체 (Entity)	클래스	물리적, 디지털, 개념적 또는 여타 종류인 것. 고정된 측면을 지니고 있음
	활동 (Activity)	클래스	일정 기간 발생하는 것으로, 개체와 연결되어 이루어짐
	발생함 (wasGeneratedBy)	관계	활동에 따라 새로운 개체가 제작되었음
	사용함 (used)	관계	활동이 개체를 활용하기 시작했음
	전달함 (wasInformedBy)	관계	한 활동이 제작한 일부 불특정 개체를 다른 활동이 사용하도록 전달함
	시작됨 (wasStartedBy)	관계	트리거에 해당하는 개체에 의해 활동이 시작됨 (활동과 관련한 사용, 발생, 무효화는 시작 이후에 수행됨)
	종료됨 (wasEndedBy)	관계	트리거에 해당하는 개체에 의해 활동이 종료됨 (활동과 관련한 사용, 발생, 무효화는 종료 이전에 수행됨)
영역 2. 파생	무효가됨 (wasInvalidatedBy)	관계	활동에 따라 기존 개체가 파파되거나 (효력이) 중단 또는 만료됨
	파생됨 (wasDerivedFrom)	관계	한 개체가 다른 개체로 변환되거나, 새로운 개체로 갱신되거나, 기존 개체를 기초로 새로운 개체가 만들어졌음
	개정 (Revision)	종류	원본을 개정한 버전으로 결과 개체가 파생됨
영역 3. 행위자, 책임 및 영향	인용 (Quotation)	종류	원본 개체의 일부나 전체를 복사 또는 인용해 결과 개체가 파생됨
	1차정보원 (PrimarySource)	종류	결과 개체가 1차 정보원을 활용하여 파생됨
	행위자 (Agent)	클래스	발생한 활동이나 개체의 존재, 또는 또 다른 행위자의 활동에 대해 일정 책임의 형태를 지니는 것
	귀속됨 (wasAttributedTo)	관계	개체를 행위자에게 속하는 것으로 봄
	결부됨 (wasAssociatedWith)	관계	활동에 대한 책임을 행위자에게 할당함. 행위자가 활동에서 어떤 역할을 했음을 지시하며, 활동의 목표 달성을 위한 계획을 지시할 수 있음
	위임됨 (actedOnBehalfOf)	관계	대리인 또는 대표자로서 특정 활동을 수행할 권한과 책임을 행위자에게 위임하며, 행위자는 위임된 작업의 결과에 대해 일부 책임을 짐
	계획 (Plan)	종류	어떤 목표를 성취하기 위해 하나 이상의 행위자가 의도한 일련의 활동이나 단계를 표현하는 개체
	개인 (Person)	종류	행위자가 개인임
	조직 (Organization)	종류	행위자가 사회적이거나 법정 기관임(기업, 학회, 법인 등)
	소프트웨어행위자 (SoftwareAgent)	종류	행위자가 구동되는 소프트웨어임
영향받음 (wasInfluencedBy)	관계	사용, 시작, 종료, 생성, 무효화, 전달, 파생, 귀속, 연결 또는 위임의 방법으로 다른 것의 특성이나 발전 및 행위에 영향을 주기 위한 개체, 활동 또는 행위자의 능력 (일 변화를 위한 표현)	



영역	명칭	구분	설명
영역 4. 번들	번들 생성기 (bundle id ... endBundle)	컨테이너	출처에 대한 복수의 기술(description)을 하나의 출처 집합으로 구성하기 위한 것. 번들도 개체에 속하며, 번들 단위 출처 기술도 지원함
	번들 (Bundle)	종류	개체가 번들임을 지시함
영역 5. 대체 개체	대체형임 (alternateOf)	관계	같은 것에 대해 동일하거나 다른 측면을 제시하는 대안적 개체를 표현함
	특수형임 (specializationOf)	관계	일반형 개체와 특수형 개체를 연결하여, 구체적인 개체의 측면을 나타냄
영역 6. 컬렉션	컬렉션 (Collection)	종류	구성원을 포함하면서 구조를 제공하는 개체
	빈 컬렉션 (EmptyCollection)	종류	구성원이 없는 컬렉션
	구성원있음 (hadMember)	관계	컬렉션에 개체가 속함을 표현하여 컬렉션의 구성원을 언급함
추가 요소	식별자 (id)	한정어	객체를 고유하게 식별하기 위한 한정어
	속성	한정어	객체를 특수하게 표현하기 위한 한정어 레이블(prov:label): 문자열로 대상을 표현함 위치(prov:location): 지리적 위치를 포함해 다양한 방법으로 위치를 표현함 역할(prov:role): 활동과 관련하여 개체나 행위자의 기능을 표현함(사용, 발생, 무효화, 결부, 시작, 종료 표시) 종류(prov:type): 추가 입력 정보를 제공함 값(prov:value): 개체가 직접 갖는 값을 표현함

PROV의 핵심적인 출처 개념은 데이터로 대표되는 ‘개체(Entity)’, 수행된 ‘활동(Activity)’, 개체와 활동에 영향을 미치는 ‘행위자(Agent)’이다. 이를 중심으로 영역 1부터 영역 3까지 전개되는데, 영역 1과 2는 개체와 활동을, 영역 3은 개체 및 활동에 영향을 미치는 행위자와 행위자가 근거하는 계획을 표현할 수 있게 한다. 이들 영역의 개념을 사용하면 실행된 활동, 활동에 사용된 개체나 활동에서 발생한 개체, 이 과정에 관련된 행위자의 관계를 통해 출처를 표현할 수 있다(Moreau et al., 2013). 특히 영역 2의 경우는 개체의 파생 관계만으로 특정 개체의 출처를 표현할 가능성을 보여준다. 특정 개체에서 시작하여 이전에 발생한 일련의 개체를 추적하는 개체 관점의 출처(Gil et al., 2013)를 기술할 수 있도록 한다. 한편 영역 4는 복수의 출처 기술(description)을 다루기 위한 ‘번들(Bundle)’ 개념을 드러

낸다. 즉, 필요에 따라 일련의 출처 기술정보를 하나의 번들로 묶고 개별 개체로 선언할 수 있는데, 그러면 번들 개체 단위로 다시 출처의 출처를 표현(Moreau et al., 2013)할 수 있게 된다. 영역 5는 개체와 개체가 상호 대안형의 관계를 형성하거나 일반형과 특수형의 관계를 형성함을 표현하게 해준다. 영역 6은 개체의 집합인 컬렉션을 통해 개체와 집합, 집합과 집합 간 관계를 표현할 수 있게 한다.

PROV는 W3C 표준이며 범용 모델의 성격을 지니므로, 이 연구에서는 모델에 구성된 클래스와 관계, 속성 등을 이후 데이터형 전자기록을 위한 출처 모델의 기초로 활용할 수 있을 것으로 판단하였다. 이들 요소를 기초로 하여, 이후의 장에서는 데이터형 전자기록의 출처를 소급해 표현하기 위한 요소를 추가하여 확장할 것이다. 예를 들어, PROV의 관계유형 중 ‘무효가됨’의 개념은 전자기록의 폐기와 관련하여 유효하게 확장할 수 있을 것으로 본다. 특히 클라우드 컴퓨팅이나 블록체인 플랫폼과 같은 분산 디지털 환경에서는 전통적인 처분 요건에 부합하게 데이터를 완전히 폐기하기 어려운데(현문수, 2024b), 기록의 효력이나 존재성을 무효로 하는 선언을 통해 전자기록의 폐기를 지원할 가능성이 있기 때문이다. 또, PROV에 존재하지 않으나, ‘무효가됨’과 반대의 영향을 미칠 수 있도록 ‘효력발생됨’의 관계 표현도 추가를 검토할 것이다. 나아가, 소급형 출처 기술정보는 필요한 시점에 기록관리 메타데이터로 포착되거나 기록관리 메타데이터를 통해 획득된 기록의 출처 정보로 연결될 가능성도 지닐 것이다.

## 2) P-Plan 및 ProvONE: 전망형 출처 표현의 기초 모델

PROV를 기초로 하여, 이 연구에서 전망형 출처를 표현할 수 있도록 모델을 확장하는 과정에 기본적으로 참고한 모델은 P-Plan과 ProvONE이다. PROV는 개체의 일종인 ‘계획(Plan)’을 통해 행위자가 수행하는 활동에 근

거를 연결할 수 있게 하지만, 구체적으로 무엇을 표현할 것인지 명시하지는 않는다. 이를 구체적으로 표현할 수 있도록 기본적인 수준에서 확장된 기본 모델이 P-Plan이며, 나아가 PROV에 전망형 출처 영역을 더하며 다양하게 개발된 워크플로우 출처 표현 모델을 종합해 반영한 ProvONE도 개발되었다.

먼저 P-Plan은 PROV의 ‘계획(Plan)’ 클래스를 확장하여 사전에 정의된 과학 실험의 워크플로우를 일종의 프로토크로 표현하고, 실제 이행 과정에 대한 표현과 연결할 어휘를 제안하기 위해 OPM(Open Provenance Model) 작업을 기반으로 개발되었다(Garijo & Gil, 2013). P-Plan에서 출처를 표현하기 위해 채택한 대상을 정리하면 다음의 표와 같다.

〈표 2〉 P-Plan의 구성(Garijo & Gil, 2013, 재구성)

대상	명칭	구분	설명
P-Plan	계획 (Plan)	클래스	PROV의 계획에 해당하며, 입출력 변수를 사용 및 생성하는 일련의 단계로 구성됨
	단계 (Step)	클래스	계획을 구성하는 구체적인 활동에 해당하며, 계획 자체일 수 있음
	입출력 변수 (Variable)	클래스	계획된 활동, 즉 단계의 입력물이나 출력물 개체에 관한 서술임
	계획의 단계임 (isStepOfPlan)	관계	단계를 이과 조응하는 계획과 연결함
	계획의 변수임 (isVariableOfPlan) (뒤이어) 진행됨 (isPrecededBy)	관계	입출력 변수를 이과 조응하는 계획과 연결함
	입력변수있음 (hasInputVar)	관계	단계를 선행 단계와 연결함
	입력변수없음 (isInputVarOf)	관계	계획된 실행의 단계에 입력되는 개체가 있음을 나타냄. ‘입력변수있음’과 역관계
	출력변수있음 (hasOutputVar)	관계	계획된 실행의 단계에 입력되는 개체임을 나타냄. ‘입력변수없음’과 역관계
	출력변수없음 (isOutputVarOf)	관계	계획된 실행의 단계에서 산출되는 개체가 있음을 나타냄. ‘출력변수있음’과 역관계
	출력변수없음 (isOutputVarOf)	관계	계획된 실행의 단계에서 산출되는 개체임을 나타냄. ‘출력변수있음’과 역관계

대상	명칭	구분	설명
PROV (실행 출처의 요약)	개체 (Entity)	클래스	PROV의 개체로, 활동과 연결되며, P-Plan의 계획과 조응함
	활동 (Activity)	클래스	PROV의 활동으로, P-Plan의 단계에 계획된 프로세스의 이행임
	번들 (Bundle)	클래스	PROV 번들의 한 유형이며, P-Plan의 특정한 실행에 대한 출처 서술 집합임
	사용함 (used)	관계	PROV의 사용함과 동일
	발생함 (wasGeneratedBy)	관계	PROV의 발생함과 동일
	파생됨 (wasDerivedFrom)	관계	PROV의 파생됨과 같으며, 특정 실행에 대한 번들이 P-Plan의 계획에서 파생된 것임을 연결하여 표현할 수 있음
	변수와조응함 (correspondsToVariable)	관계	활동 실행에서 사용-발생한 개체를 P-Plan의 계획에 표현된 입출력 변수와 연결함
	단계와조응함 (correspondsToStep)	관계	실행된 활동을 P-Plan의 계획에 표현된 단계와 연결함

기본적으로 P-Plan은 실행 추적을 위한 모델인 PROV의 구성요소를 수용하면서 계획 클래스를 확장하기 위한 것이므로, 표의 실행 출처 영역인 PROV 부분은 P-Plan에 언급된 요소만 포함하여 기술하였다. P-Plan의 주요 초점은 확장된 계획 영역, 그리고 계획 영역과 실행 영역 간 연결 관계 형성이다. 실행에 대해 사전에 정해진 계획을 표현하는 영역은 ‘계획’과 ‘단계’, ‘입출력 변수’를 중심으로 관계를 형성할 수 있도록 확장되었다. ‘단계’는 미리 계획된 일련의 활동이나 절차를 표현하기 위한 것이며, 이 과정에 입력되어야 하거나 산출되는 개체를 ‘입출력 변수’로 표현할 수 있다. 계획에 따른 실행은 PROV의 구성요소를 통해 표현할 수 있으며, 계획 영역과 실행 영역을 연결하기 위해 조응(correspondsTo) 관계를 추가하여 구성한다. 또한 특정한 실행과 연결하여 생성된 복수의 출처 기술정보를 모은 번들을 계획 영역과도 연결할 수 있는데, 이를 위해 PROV의 ‘파생됨’ 관계 표현을 사용한다.

P-Plan은 PROV의 개체인 ‘계획(Plan)’을 확장하여 전망형 출처 영역을 구성할 가능성을 기초적인 수준에서 보여준다. 또, 전망형 영역을 구성할

때 소급형 출처 영역과 조응할 수 있도록 개체와 활동의 관계를 정의할 수 있다는 점도 드러났다. 나아가 전망형 영역과 실행 영역의 구성요소가 조응하는 관계를 형성할 수 있으며, 특정 실행 인스턴스 단위로 기록화된 출처 기술정보를 묶어 전망형 출처와 연결할 수 있게 한 점도 추후 이 연구에서 수용할 가능성이 있다고 판단하였다.

다음으로 ProvONE은 PROV의 소급형 출처 개념을 수용하는 한편, 과학 실험 프로그램이나 워크플로우에 미리 정의된 절차와 개체 사용·생산 규칙 등에 대한 정보를 사전적(proactive)으로 획득하여 소급형 출처 정보와 연결할 수 있게 DataONE 프로젝트의 맥락에서 개발(Cuevas-Vicenttín et al., 2016)되었다. DataONE(Data Observation Network for Earth)은 여러 리포지토리에 있는 대규모 데이터를 지구과학 연구와 교육 분야 및 일반에까지 제공할 목적으로 인프라를 구축하기 위해 진행된 프로젝트이며 2009년 미국 국가과학재단(NSF: National Science Foundation)의 재정 지원으로 시작되었다(DataONE, n.d.). ProvONE의 명칭도 이 맥락에 따른 것이지만 과학 분야에서 다양하게 제안된 모델을 통합하고 과학 공동체에서 이뤄지는 실험에 대해 표준화된 출처를 표현할 수 있을 만큼 폭넓게 사용되도록 개발되었다(Cuevas-Vicenttín et al., 2016). ProvONE을 사용하면 활동과 개체, 행위자의 관계 속에서 개체의 이력을 추적하는 능력을 제공하면서도, 미리 정의된 방식대로 활동과 개체, 행위자의 관계가 맺어졌는지를(Cuevas-Vicenttín et al., 2016) 검증할 수 있게 한다. ProvONE에서 출처를 표현하기 위해 채택한 대상을 정리하면 다음의 표와 같다.

〈표 3〉 ProvONE의 구성(Cuevas-Vicenttín et al., 2016, 재구성)

대상	명칭	구분	설명
워크플로우	프로그램 (Program)	클래스	입력 포트와 출력 포트를 통해 각각 데이터를 소비하고 생성하는 계산 과업이며, 복합 프로그램일 수 있음
	포트 (Port)	클래스	프로그램에서 개체를 보내거나 받을 수 있게 함
	채널 (Channel)	클래스	프로그램에 정의된 입출력 포트 간의 연결을 제공함

대상	명칭	구분	설명
추적	통제 프로그램 (Controller)	클래스	특정 계산 모델에 따라 다른 프로그램을 통제하는 프로그램
	워크플로우 (Workflow)	클래스	어떤 계산 실험 전체를 다루는 고유한 개별 프로그램
	하위프로그램있음 (hasSubProgram)	관계	프로그램에 하위 프로그램이 있음
	통제됨 (controlledBy)	관계	프로그램을 통제 프로그램과 연결함. 프로그램이 통제 프로그램에 의해 통제됨
	통제함 (controls)	관계	통제 프로그램을 프로그램과 연결함. 통제 프로그램이 프로그램을 통제함
	입력포트있음 (hasInPort)	관계	입력 포트 사용되는 특정 프로그램의 포트를 지정함
	출력포트있음 (hasOutPort)	관계	출력 포트 사용되는 특정 프로그램의 포트를 지정함
	기본전달개체있음 (hasDefaultParam)	관계	입력 포트에 기본적으로 전달할 개체가 있음을 지시함
	연결함 (connectsTo)	관계	지정된 포트가 연결되는 채널을 지정하며, 일반적으로 입력 포트에 연결된 출력 포트가 있음
	파생됨 (wasDerivedFrom)	관계	프로그램과 워크플로우의 진화를 표현함
	실행 (Execution)	클래스	프로그램 또는 워크플로우의 실행
	결부 (Association)	클래스	활동에 대해 행위자에게 책임을 할당함. 행위자가 활동에서 어떤 역할을 했음을 지시하며, 활동의 목표 달성을 위한 계획을 지시할 수 있음
	사용 (Usage)	클래스	활동에서 개체를 활용하는 시작점
	발생 (Generation)	클래스	활동에 따른 새로운 개체 생성의 완료
	이용자 (User)	클래스	실행에 책임이 있는 사람 [행위자]
	사용함 (used)	관계	실행에서 특정 개체를 입력물로 사용함
	발생함 (wasGeneratedBy)	관계	실행에서 특정 개체를 출력물로 제작함
	결부됨 (wasAssociatedWith)	관계	이용자와 특정 실행을 연관시킴
	전달됨 (wasInformedBy)	관계	두 실행 사이에 실행의 트리거로서 출력-입력을 주고받음
	부분임 (wasPartOf)	관계	상위 실행과 하위 실행의 실행 인스턴스 구조를 표현함
한정결부됨 (qualifiedAssociation)	관계	이용자가 계획으로서 특정 프로그램이나 워크플로우와 연결된 실행에 책임이 있음	
행위자임 (agent)	관계	실행에 영향을 미친 이용자임	
계획있음 (hadPlan)	관계	실행과 결부된 이용자가 채택한 계획으로서 프로그램이나 워크플로우가 있음	
한정사용됨 (qualifiedUsage)	관계	실행에서 프로세스 입력물로 어떤 개체(데이터, 시각개체, 문서)를 사용했음	
입력포트있음 (hadInPort)	관계	특정 실행에서 사용된 입력 포트(Port)를 지시하며, 이 입력 포트는 사용(Usage)에 설명됨	
출력포트있음 (hadOutPort)	관계	특정 실행에서 사용된 출력 포트(Port)를 지시하며, 이 출력 포트는 발생(Generation)에 설명됨	

대상	명칭	구분	설명
	개체있음 (hadEntity)	관계	특정 실행에서 입력포트로 사용된 개체를 지시하며, 사용(Usage)에 설명됨
	한정발생됨 (qualifiedGeneration)	관계	실행에서 프로세스 출력물로 어떤 개체(데이터, 시각개체, 문서)를 생성했음을 나타내며, 실행에서 개체를 어떻게 생성했는지 한정하여 표현함
데이터	개체 (Entity)	클래스	데이터나 시각 개체, 문서 종류의 인스턴스이거나 PROV 개체 클래스의 하위 클래스일 수 있음
	컬렉션 (Collection)	클래스	구성원을 포함하면서 구조를 제공하는 개체
	데이터(Data)	클래스	프로그램에서 소비하거나 생성하는 정보의 기본 단위로, 여러 데이터 항목을 하나의 컬렉션으로 그룹화할 수 있음
	시각개체 (Visualization)	클래스	프로그램에서 소비하거나 생성하는 정보의 기본 단위로, 디지털 시각적 표현의 형태임. 여러 시각 개체 항목을 하나의 컬렉션으로 그룹화할 수 있음
	문서 (Document)	클래스	학술지 논문이나 보고서와 같이, 실행(Execution)의 결과로 생성된 정보의 본문을 커뮤니케이션 매체의 형태로 나타낸 것임. 여러 문서 항목을 하나의 컬렉션으로 그룹화할 수 있음
	파생됨 (wasDerivedFrom)	관계	워크플로우 실행 중 제작된 데이터 간의 종속성을 설명함
	구성원있음 (hadMember)	관계	컬렉션을 구성하는 데이터를 지시함

우선 ProvONE에서는 PROV에서 제외된 전망형 출처를 표현할 수 있도록 ‘워크플로우(Workflow)’ 영역을 추가한다. 이를 통해 미리 지정된 워크플로우나 프로그램을 ‘추적(Trace)’ 영역의 ‘실행(Execution)’과 연결하여, 워크플로우에 기반해 수행된 실험을 이해할 수 있도록(Cuevas-Vicenttín et al., 2016) 해준다. 또, 워크플로우나 프로그램에서 사용하는 입출력을 지시할 수 있게 ‘포트(Port)’와 ‘채널(Channel)’ 클래스를 두고, 입출력 포트가 있다는 관계도 표현할 수 있도록 한다. 워크플로우나 프로그램도 개체의 일종임을 고려할 때, 하위프로그램이나 워크플로우가 있음을 나타내기 위한 관계 표현도 가능하도록 구성된 것으로 판단한다.

‘추적(Trace)’ 영역은 PROV를 기초로 과학 실험 맥락을 수용하기 위해 일부 확장된 소급형 출처 표현 영역이다. PROV의 ‘활동’과 ‘행위자’를 각각 프로그램이나 워크플로우 ‘실행(Execution)’ 클래스와 ‘이용자’ 클래스로 대체하였다. 워크플로우 영역과 관계를 표현할 수 있도록 ‘사용(Usage)’,

‘발생(Generation)’, ‘결부(Association)’를 클래스로 두고, 프로그램이나 포트와 연결할 수 있게 ‘한정사용함’, ‘한정발생됨’, ‘한정결부됨’이나 입출력 포트와 연결할 수 있는 관계 표현도 추가한다. 한편, ProvONE은 PROV의 ‘개체’를 데이터 영역으로 구분하여 구성한다. 개체의 집합으로서 컬렉션 및 구성원 관계, 개체 간 파생 관계를 수용하는 점은 PROV와 같다. 반면, 개체를 ‘데이터’와 ‘시각개체’, ‘문서’로 표현할 수 있도록 한 점은 PROV와 구분된다.

ProvONE은 과학 실험의 프로세스 구조[워크플로우 영역]와 실제 이행[추적 영역]을 연결하여 기록화하고 데이터세트 산출의 재현성을 지원할 수 있도록 PROV 모델을 확장한 것이다(Cao et al., 2016). 특히, 전망형 출처까지 표현하기 위해 PROV를 확장하였다는 점, 활동으로서의 과학연구 과정에 산출되는 데이터를 대상으로 출처를 표현하여 재현성을 확보하기 위해 개발되었다는 점에서 데이터형 전자기록의 출처 모델 개발과 유사한 맥락을 가진다고 판단하였다. 이미 ProvONE은 그 개발 과정에 이력으로서의 소급형 출처와 실행의 근거가 되는 규제나 통제 등을 포괄하는 전망형 출처를 표현할 수 있도록 확장되었다. 따라서, 이 연구에서도 전망형 출처 영역에 대해서는 ProvONE의 구성 개념을 중심으로 하되, 요소의 명칭은 좀 더 일반적인 모델인 P-Plan을 중심으로 하여 조정하고, 데이터형 전자기록의 생산과 관리·보존의 맥락을 수용할 수 있도록 수정 및 재확장할 수 있다고 보았다. 이어지는 장을 통해 P-Plan의 계획 영역과 ProvONE의 워크플로우 영역을 이 연구 범위에 맞추어 데이터형 전자기록의 전망형 출처 영역으로 수정할 것이며, 그 과정에는 과학분야의 실험 워크플로우를 전제로 한 ProvONE의 구조도 수정 범위에 포함할 것이다.



#### 4. 기초 모델의 확장을 위한 분석

앞 장에서 살펴본 PROV 표준 모델은 특정 영역의 적용을 목적으로 하지 않는 범용성을 지닌다. 그러므로 개별 전문 영역은 각자의 출처 개념을 원활하게 지원할 수 있도록 이 모델을 확장하여 적용할 수 있다. 그중 하나가 이 연구에서 PROV와 함께 사용할 ProvONE이며, 이 이외에도 다양한 영역에서 개별 전문 영역의 요구를 반영하여 PROV 확장을 시도하고 있는데, 개인정보보호 데이터의 수집과 활용을 지원하기 위해 Trinity College Dublin에서 개발한 GDPProv 출처 온톨로지(Pandit, 2019), 웹 기반 협업 플랫폼인 신경의학 시멘틱 웹 애플리케이션 개발 프로젝트의 맥락에서 웹 자원의 저작과 버전 및 출처에 대한 정보를 기록화하기 위한 PAV(Provenance, Authoring & Versioning) 온톨로지(Ciccarese et al., 2013) 등이 있다. 따라서 데이터형 전자기록에 적용할 수 있도록 PROV를 확장하려면, 기록 관리 영역의 요구를 반영하기 위한 분석 작업이 선행되어야 한다. 2장에서 밝힌 것처럼, 이 연구는 소급형 출처 영역과 전망형 출처 영역을 함께 구성하고 연결할 수 있는 모델 개발을 목표로 한다. 이를 위해, 3장에서 살펴본 PROV와 P-Plan, ProvONE을 기초 모델로 사용하여 기본 요소를 구성하였다. 단, ProvONE의 구성요소 명칭의 경우 프로그램 실행의 범위를 넘어 일반화될 수 있도록 일부 수정하여 사용하였다. 이를 바탕으로 전자기록의 출처를 표현하기 위한 추가 확장 과정을 진행하였다.

특히 기록관리 영역에서 다루는 개체나 수행되는 활동, 이에 관여하는 행위자, 그리고 이들의 관계를 표현할 수 있도록 보완되거나 수정되어야 할 구성요소가 무엇인지 확인하는 과정을 진행할 필요가 있었다. 출처 정보로 기록되어야 하는 구체적인 요소를 드러내기 위해서 우선 기록관리 메타데이터 표준인 KS X ISO 23081의 1부(이하 23081-1)와 2부(이하 23081-2) 및 OAI 참조 모델(이하 OAI)의 정보 모델을 참고하여 출처와 맥락 표현의 측면을 확인하였다. 전자기록관리 전반에 걸친 활동과 행위자

를 드러내기 위해서는 기록관리 일반 표준인 KS X ISO 15489의 1부(이하 15489)에 구성된 기록관리 과정을 참조하는 한편, InterPARES 2단계 프로젝트에서 개발한 “CoP(Chain of Preservation) 모델”(이하 CoP)도 함께 참고하였다. CoP를 대상에 포함한 이유는 신뢰할 수 있고 진본인 디지털 기록을 생산·관리·장기보존하기에 필요한 모든 활동과 입출력, 통제 기제를 포괄적으로 묘사(Eastwood, Hofman & Preston, 2008)하고 있어, 전자기록의 출처 모델에 필요한 일반 수준의 구성요소를 드러낼 수 있다고 판단했기 때문이다. 마지막으로 출처 기술정보를 통해 기록이 갖는 활동의 고정적 재현성과 신빙성의 속성을 지원할 수 있도록, 전자기록의 진본성과 신뢰성 및 이용가능성을 평가하기 위한 도구로 제안된 신뢰가치 모델(안)(이하 신뢰모델(현)(현문수, 2022)의 품질측정 기준도 함께 확인하여 구성요소를 보강하였다. 확장을 위해 사용한 문헌에 드러나는 메타데이터 요소나 요건은 특정 업무 환경에서 기록을 표현하는 데 충분히 활용되어야 하겠으나, 이 연구는 기록의 출처 표현 가능성에 집중하였다. 따라서 포괄적으로 요소를 추출하지 않고, 데이터형 전자기록의 출처를 표현하고 이후 출처의 역할을 할 수 있도록 지원하는 데 초점을 두어 기초 모델을 확장하였다.

## 1) 전망형 출처 표현 영역

첫 번째로 데이터형 전자기록에 적용할 전망형 출처 영역을 구성하기 위해 수행한 내용은 다음과 같다. 먼저, P-Plan의 계획 영역과 ProvONE의 워크플로우 영역의 구성요소를 바탕으로 기본적인 요소를 구성하였다. 다만, ProvONE의 경우 일반성을 표방하지만, 일부 개념은 과학 실험에서 산출된 데이터 표현에 필요한 요소를 세밀하게 표현하고 있다고 판단하여 다른 개념에 포함하여 배치하였다. 예를 들어, 채널이나 포트로 표현하기보다 P-Plan의 입출력 개체의 개념을 사용하여 미리 정의된 입력 개체와 출력 개체의 존재나 활용을 통제할 수 있을 것으로 보았다.

다음으로는 기본 구성요소를 기준으로 이 연구의 목적에 맞게 확장할 수 있도록 앞서 제시한 문헌에 드러난 전망형 출처 표현 요건을 매핑하였다. 23081-1에서는 모범적인 기록관리 지원에 필요한 메타데이터 유형을 제시한 9장의 내용을 중심으로 전망형 출처 표현 요소의 가능성이 있는 요건을 추출하여 매핑하였다. 23081-2의 경우 9장에서 기록관리를 위한 일반메타데이터 모델에 따른 메타데이터 그룹별 요소를 제시하고 있는데, 23081-1에 제시한 다양한 메타데이터 유형에 일반적으로 적용할 수 있도록 구성하고 있으므로, 개별 요소와 매핑하기에는 적합하지 않다고 판단하였다. 다만, ‘사건 계획’과 ‘이용’ 메타데이터 그룹의 경우는 미리 규범이나 절차적으로 활동을 정의하기 위해 활용할 수 있다고 판단하여 제한적으로 활용하였다. OAIS에서도 정보 모델을 구성하는 보존기술정보(PDI: Preservation Description Information)의 영역 내에서 출처와 맥락정보 요소를 배치하고 있어, 이를 활용하여 기본 구성요소와 매핑하였다. 또, 15489와 CoP에 지시된 기록관리과정은 기록 생산에서부터 관리와 보존에 이르는 기록 활동을 규범적으로 제안한다고 판단하여, 워크플로우의 하위 단계(또는 활동)를 확장하는데 적용할 수 있다고 보아 전망형 출처 영역에서도 활용하였다.

전망형 출처 영역을 확장하는 과정에 기록관리 영역의 규범과 통제를 수용하기 위한 분석도 진행되었다. 15489의 8장에서 제시하는 기록 통제에 필요한 도구 유형과 CoP에서 기록 활동에 대한 통제로 제시하는 요소는 전망형 출처로 표현되어야 하는 통제 요소를 확장하기 위해 사용될 수 있다고 보았다. 신뢰모델(현)은 전자기록의 신빙성을 평가하기 위한 요건을 제시하므로, 그 기준 중 전망형 출처 영역에서 표현되어야 충족될 수 있는 기준을 확인하여 관련 요소와 매핑하였다. 예를 들어, ‘파일명이 기관 내부의 명명규칙을 준수하는가’와 같은 기준은 소급형 출처 정보만으로 확인할 수 없는 것이며, 통제 요소를 통해 표현되어야 한다고 판단하여 매핑하였다.

이처럼 기본 구성요소를 기준으로 각 문헌에서 매핑할 수 있는 항목을 아래의 예시 표와 같이 정리하고 확장 가능성을 분석한 내용도 함께 구성하

여 이후 모델 개발에 활용하였다. 표에는, 요소를 매핑한 결과와 함께, 기본 구성요소의 확장을 검토하는 과정에 확인한 사항과 이후 최종 모델을 개발할 때 검토하거나 결정해야 할 사항도 종합하여 서술하였다. 예를 들어, ‘통제’ 요소에는 규범과 도구가 모두 포함되어 있으므로 이를 분해할 가능성을 검토할 수 있었고, ProvONE을 기준으로 한 ‘프로그램’ 요소에 대해 행위자로 명칭을 수정할 것인지 규제로서의 도구로 편입할 것인지를 검토할 필요도 확인하였다.

〈표 4〉 기초 모델 확장을 위한 검토 사례: 전만형 출처 영역

출처 영역	기본 구성요소	기초 모델	확장 검토					검토 종합
			23081	OAIS	15489	CoP 모델	신뢰모델(현)	
전만형 출처	계획/ 워크플로우	P-Plan ProvONE	(1부)(9.5) 업무과정 (2부) 사건 계획	(PDI, 맥락) s/w 도용할 파일 이용자 가이드	업무프로세스, 절차	기록, 생산·관리·보존시스템 내 가능 흐름	기록시스템-기록 간 연결 정보, 기록과 조직의 공식 업무프로세스 간 연결	업무 수행을 위해 이루어지는 단계나 절차, 입출력되는 객체, 책임 행위자를 사전 정의하는 기술 정보로 적용 개발 활동의 집합체로 적용 검토
	통제	ProvONE	(1부)(9.3) 법규, 업무규정, 절차규범, 기록 활동규범 (1부)(9.6) 접근권한, 업무 분류체계, 처분 지침, 메타데이터 스키마, 기타 통제도구 (2부) 이용	(PDI, 출처) 디커뮤테이션 및 s/w 인터페이스 규격 (PDI, 맥락) 업무, 출판 당시 원본 환경에 대한 문서	업무규칙, 방침, 법규환경, 기록 관리요구사항, 기록메타데이터 스키마, 업무분류체계, 접근 허용 규칙, 처분지침	업무, 표준, 기록관리원칙·요건, 업무-기록법규 체계, 접근규칙, 기록서식, 기록메타데이터스키마, 업무-기록화 절차규칙, 품질확보절차규칙, ...	파일명명규칙, 처분보유-요건, 접근·허용 규칙, 반출규칙, 기록생산규칙, ...	법규, 표준 등 기록 활동을 직접 통제하는 규제환경을 기술하는 정보로 적용 기록 활동을 통제하는 기제/도구를 기술하는 정보로 적용 규제환경과 통제도구를 구분할 가능성 검토
	프로그램	ProvONE	(1부)(9.4) 행위주체	(PDI, 출처) 측정도구설명	기록시스템, 행위주체	기록시스템(생산, 관리, 보존), 생산자, 보존자, 기록관리자, 도구, 설비 기록시스템	기록생산s/w정보	계획/워크플로우와 연계하여 기록 활동을 수행할 권한을 부여받은 수행자 정보로 수정 검토 또는 통제도구로 편입할 가능성 검토

## 2) 소급형 출처 표현 영역

두 번째로 데이터형 전자기록에 적용할 소급형 출처 영역을 구성하기

위해서 PROV의 구성요소를 바탕으로 기본 요소를 구성하였다. P-Plan의 실행 영역은 PROV를 따르므로 이에 대해서는 추가 검토 없이 제외하고 ‘계획’ 영역과 연결을 위한 요소만 사용하였다. ProvONE의 경우 일부 명칭을 조정하고 워크플로우 영역과 연결하기 위한 추가 요소를 구성하고 있으나, 일단 동일하거나 유사한 개념을 다루는 요소의 경우는 PROV를 기준으로 기본 요소를 구성하고, 연결을 위한 요소는 우선은 PROV에서처럼 관계 표현으로 다룬 후 추후 검토 과정을 거쳤다.

다음으로는 전망형 출처 표현 영역에 대해 진행했던 것처럼, 분석 대상 문헌에 드러난 소급형 출처 표현 요건을 매핑하였다. 23081-1에서는 9장의 내용을 중심으로 소급형 출처 표현 요소의 가능성이 있는 요건을 추출하여 매핑하였다. 23081-2의 경우 대부분 기록관리 메타데이터로 획득될 것이나, ‘사건 이력’ 메타데이터 그룹의 요소는 실행된 활동을 표현하기 위해 활용될 수 있다고 판단하였으며, 일부 ‘식별’ 및 ‘기술’ 메타데이터 그룹의 요소(예: 식별자, 제목 등)는 속성으로 활용될 수 있다고 판단하여 포함하였다. OAIS에서도 앞 절에서와 같이 보존기술정보(PDI) 중 출처와 맥락정보 요소를 활용하여 기본 구성요소와 연결될 수 있는 요소를 매핑하였다. 또한 OAIS의 정보 모델 기본 구조는 개체 요소를 확장할 때 활용할 수 있을 것으로 보아 개체와 연결하여 배치하였는데, 예를 들어 ‘디지털 데이터 객체’와 같은 단위를 개체의 하위 범주로 확장할 수 있을 것으로 판단하였기 때문이다.

15489에서는 9장에서 제시하는 기록의 생산과 획득 및 관리 프로세스 영역을, CoP에서는 기록의 생산과 관리, 보존 시스템에서의 기록 활동을 중심으로 기본 구성요소의 활동에 배치하였다. 이 과정을 통해 확인된 기록 활동을 바탕으로 이후 활동 요소를 확장하여 전개할 가능성을 검토할 수 있었다. 신뢰모델(현)에서도 소급형 출처 영역에서 표현되어야 충족될 수 있는 기준을 확인하여 관련 요소와 매핑하였다. 예를 들어, ‘허가받지 않은 기록 내용의 변경을 확인할 정보가 있는가’와 같은 기준에 대해서는 대상

기록의 출처 기록을 종합적으로 분석한 후 판단해야 하므로 특정 요소와 연결하지 않았다. 대신, ‘적절한 처분(마이그레이션, 이관/반출 등) 이후 복구 불가능하도록 (이전의) 내용객체와 메타데이터를 폐기하였음을 확인할 수 있는가’와 같은 기준의 경우는, 출처 기록으로 산출되어야 하는 필수 활동 요소를 드러내고, 이와 연결하여 일련의 출처로 기록화되어야 하는 범위를 구체화하기 위해 사용할 수 있었다.

소급형 출처 표현 영역에 대해서도 기본 구성요소를 기준으로 각 문헌에서 매핑한 항목을 아래의 예시 표와 같이 정리하고, 이 연구에서 분석한 확장 가능성과 이후 최종 모델을 개발할 때 검토하거나 결정해야 할 사항도 종합하여 서술하여 이후 모델 개발에 참고하였다. 검토 과정에 주요하게 드러난 사항은, 이후 최종 모델을 개발하는 과정에 주로 ‘개체’와 ‘활동’, 개체와 개체 간 파생 관계를 확장할 필요가 있다는 점이었다. 우선 개체 측면에서, 미시적 수준에서 개체의 출처를 표현하기 위해서는 논리적 단위인 ‘기록’보다 디지털 컴포넌트 또는 디지털 데이터 객체를 최소 개체 단위로 검토할 필요가 있었다. ‘활동’의 전개와 관련하여서는 기록관리 과정을 참조하여 확장을 검토할 것인데, 이 과정에 어떤 기록 활동을 기준으로 출처 기록과 조직 내에서 선언된 기록 및 메타데이터가 연결될 수 있는지도 함께 검토할 것이다. 관계 표현 중에서는 ‘(~에서) 파생되었다’의 확장을 요구하는 사항이 드러나, 최종 모델을 개발하는 과정에 개체와 개체 간 파생 관계의 각 유형을 구분하여 표현할 수 있도록 구성할 필요를 확인하였다. 기록의 물리적·논리적 구조가 변화하거나 마이그레이션의 세부 유형 또는 이전 등의 기록화가 기록의 무결성에 영향을 미치므로 여러 요건에서 이를 세분화하여 다루고 있었기 때문이다.

〈표 5〉 기초 모델 확장을 위한 검토 사례: 소급형 출처 영역

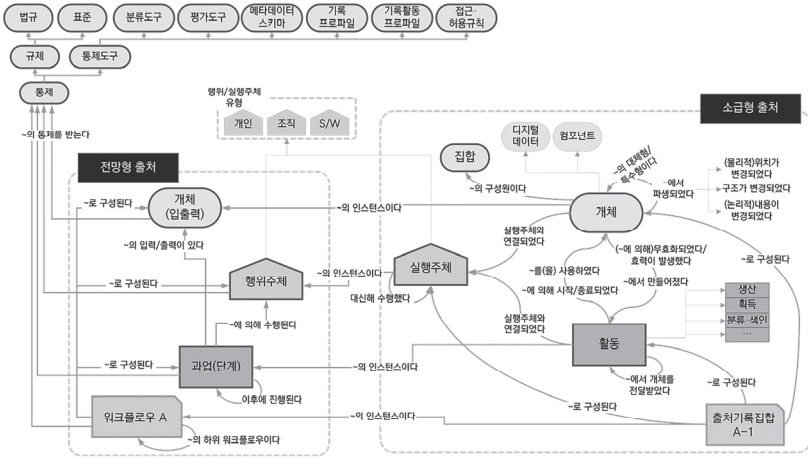
출처 영역	기본 구성요소	기초 모델	확장 검토					검토 종합
			23081	OAIS	15489	CoP 모델	신뢰모델(현)	
소급형 출처	개체(데이터, 시각개체, 문서)	PROV ProvONE	(1부)(9.2) 기록	(정보모델) 디지털 데이터 객체, 표현정보, 정보객체	기록, 메타데이터	문서, 기록, 디지털 컴포넌트, 메타데이터	데이터객체, 표현정보, 기술정보	기록의 출처 표현에 적절하게 개체를 확장할 수 있는 단위 검토
	활동	PROV	(1부)(9.5) 업무과정 (1부)(9.6) 기록관리과정 (2부) 사건 이력	(PDI, 출처) 데이터처리이력, 저장·조작 이력, 디지털화 프로세스	생산, 획득, 분류 · 색인작성, 접근 통제, 저장, 이용·재이용, 마이그레이션·변환, 처분	(기록생산시스템) 문서작성, 문서회독, 기록선언, ... (기록관리시스템) 색인, 저장, ... (기록보존시스템) 기록평가, 기록일수, ...	기록 생산을 유발한 기능 확인, 기록과 관련된 행위-이벤트 확인(획득, 접근, 이용, 사본제작, 이관, 폐기 등) ...	기록 활동을 표현할 수 있도록 확장하여 적용  ‘획득’ 활동의 대상 기록은 공식 기록으로 선언된 개체이므로, 획득 활동 종료 후 기록에 대해서는 기록관리 메타데이터 영역과 연결될 수 있음 ...
	파생되었다	PROV	(1부)(9.2) 논리적 구조 변화, 물리적 구조 변화, 기술적 속성 변화 기록 간 새로운 관계 위치 변경	(PDI, 출처) 이전 개체 포인터(마스터버전 등), 변화이력, ... (마이그레이션 유형) 리프레시, 구조변경, 내용변경	마이그레이션·변환	저장위치변경, 매체변경(리프레시), 갱신	완전한 구조-구성요소를 포함한 마이그레이션·변출 확인 정보	개체 간 파생 관계로 변화 이력을 표현할 때, 변화 유형을 표현할 수 있도록 확장하여 적용
...								

## 5. 데이터형 전자기록을 위한 출처 모델 개발(안)

### 1) 모델 구성

기초 모델을 바탕으로 출처 기록화 요건을 분석하는 과정에 드러난 검토 사항을 확인하면서 이 연구가 제안하는 데이터형 전자기록을 위한 출처 모델을 개발하였다. 이 연구가 목표로 하는 모델을 활용하여, 활동을 이행한 결과로 기록을 구성하는 디지털 데이터의 궤적을 표현하는 한편, 활동 실행의 기반이 되는 규제나 통제 요건 및 도구를 미리 정의된 메커니즘에 따라 연결할 수 있도록, 아래의 그림에서와 같이 각각 소급형 출처와 전망형 출처 영역으로 구성하고 연계하도록 하였다.

〈그림 1〉 데이터형 전자기록의 출처 모델(안)(개요)



※ 현문수(2024a)의 기초 설계를 바탕으로 재구성 및 확장함

〈그림 1〉에서 왼쪽 영역이 전망형 출처 영역으로, 개별 워크플로우를 표현하기 위한 요소와 실행을 규제하거나 통제하는 맥락을 표현할 요소를 아래의 〈표 6〉과 같이 구성하였다. 전망형 출처 영역의 기본 구성요소와 명칭은 P-Plan을 기준으로 하였다. ProvONE의 구성요소보다 일반적으로 적용할 수 있다고 판단하였으며, 선행연구(현문수, 2024a)에서 밝힌 것처럼 기록관리 실무환경에서도 기록관리를 위한 업무분석 표준에 따라 워크플로우와 활동의 단계들, 행위주체, 입·출력물을 정의할 수 있다는 점을 고려하였다. 또, 기초 모델에서는 소급형 출처 표현 요소로만 구성되었던 행위자를 전망형 출처 영역에서도 표현할 수 있도록 추가하여, 활동 실행과 개체 사용 및 발생의 근거와 함께 활동을 실행한 주체에 대해서도 미리 정의된 규제나 통제를 따랐는지 확인할 수 있도록 하였다. 앞 장에서 서술한 것처럼, 통제 요소는 ‘법규’와 ‘표준’을 표현하기 위한 ‘규제’와 좀 더 직접적으로 기록 활동을 통제하는 ‘통제도구’로 구분하였고, ‘통제도구’를 세분하여



앞서 확인한 분류나 평가도구를 포함하여 기록에 적용되는 메타데이터 스키마, 기록 및 기록활동 프로파일, 접근과 허용규칙을 표현할 수 있도록 구성하였다.

〈표 6〉 데이터형 전자기록의 출처 모델 구성(안): 전방향 출처 영역

구분	구성요소	하위 구성요소	하위-하위 구성요소	설명	관계 방향*	
개체	워크플로우			일련의 과업 수행 구조를 표현할 수 있으며, 그 실행에 해당하는 (소급형) 출처기록집합과 연결됨	-	
	입력 개체			과업 수행을 위해 입력되는 개체	-	
	출력 개체			과업 수행에서 출력되는 개체	-	
	통계	규제			기록 생산, 보관, 전달, 사용, 유지 및 보존 등 기록 활동을 제약하거나 이에 영향을 미치는 규제 환경	-
			법규		기록 활동을 수행할 때 준수해야 하는 법령 및 규정	-
			표준		모범적인 기록 활동 실무를 제안하는 국제 및 국가표준	-
			통제도구		기록요구사항을 충족하기 위해 기록 활동에 근거해야 하는 도구	-
		분류체계			기록과 생산 맥락을 연결하는 도구	-
			평가도구		기록의 평가 결정 및 처분 행위를 정해 놓은 도구	-
			메타데이터 스키마		기록 식별, 기술 및 관리에 사용되는 메타데이터 스키마	-
			기록 프로파일		기록 작성에 사용되는 기록 속성이나 특징의 집합을 나타내는 도구	-
			기록활동 프로파일		수행되어야 하는 기록 생산·관리·보존 활동을 나타내는 도구	-
			접근·허용 규칙			일련의 접근권한 식별 규칙 및 허용과 제한에 대한 체계
활동	과업(단계)			워크플로우의 일부로 수행되는 과업의 단계	-	

구분	구성요소	하위 구성요소	하위-하위 구성요소	설명	관계 방향*
행위자	행위주체			과업에 대한 책임을 부여받아 과업을 수행해야 하는 주체. 개인, 조직, 소프트웨어행위자로 구분될 수 있음	-
관계	~(으)로 구성된다			워크플로우에 포함된 구성요소를 연결함	워크플로우→과업·행위주체·개체
	~의 하위 워크플로우이다			(하위)워크플로우가 (상위)워크플로우의 하위에 있음	(하위)워크플로우→(상위)워크플로우
	~의 통제를 받는다			규제 및 통제도구의 통제를 받음	워크플로우·과업·행위주체·개체→통제
	~의 이후에 진행된다			(후행) 과업이 선행 과업 이후에 진행됨	(후행)과업(단계)→(선행)과업(단계)
	~에 의해 수행된다			과업이 행위자에 의해 수행됨	과업→행위주체
	~의 입력이 있다			과업 수행에 입력되는 개체가 있음	과업→개체(입력)
	~의 출력이 있다			과업 수행에서 산출되는 개체가 있음	과업→개체(출력)

\* 관계의 방향은 '관계'에 해당하는 요소에만 서술하였음

다음으로 소급형 출처 영역은 <그림 1>의 오른쪽 영역이다. 소급형 출처 영역에서는 미리 정의된 대로 실행이 이루어지는 과정과 결과의 이력을 표현하며, 이를 위한 요소를 아래의 <표 7>과 같이 구성하였다. 소급형 출처 영역의 기본 구성요소와 명칭은 PROV를 기준으로 하였다. 기본적으로 개체와 활동, 행위자 간의 관계를 표현하는 구조 및 PROV의 추가요소인 속성 표현 요소도 함께 수용하였다. 다만, 행위자에 해당하는 요소의 명칭은 전망형 출처의 행위자 요소와 구분할 수 있도록 '실행주체'로 변경하였다.

앞 장에서 검토 사항으로 언급했던 것처럼, 소급형 출처 영역에 대해서는 개체와 활동 및 파생 관계에 초점을 두고 모델을 확장하였다. 개체의 기본 단위는 '디지털 데이터객체'와 '컴포넌트' 수준에서 표현할 것으로 결정했는데, 논리적 단위로 여겨지는 기록을 단위로 하는 경우, 오히려 특정 과업 단위를 실행한 이력을 불명확하게 표현할 것으로 판단했기 때문이다. 디지털 데이터 객체나 컴포넌트의 집합을 표현할 수 있도록 한 PROV의 '컬렉

선' 개념을 '집합' 요소로 구성하였으며, PROV의 '번들'을 '출처기록집합'으로 구성하여 특정 전망형 출처 단위와 연결할 수 있는 구조를 그렸다.

'활동' 요소의 확장을 위해서는, 기록관리 프로세스를 기록 활동으로 보고, 이를 중심으로 세분하여 전개하였다. 앞 장에서 살펴본 결과 일부 문헌에서는 기록 생산의 범위와 관리 및 장기보존의 범위나 시스템을 엄밀히 구분하여 단계별 기록 활동을 중복 서술하고 있었다. 예를 들어, 분류·색인이나 저장, 이용 등이 대표적으로 중복된다. 하지만 이 연구의 범위에서는 이러한 기록관리 단계를 기준으로 활동을 구분하지 않고 활동을 기준으로 세분하여 요소를 구성하였다. 엄밀한 단계 구분이나 기록시스템의 역할 구분보다 어떤 시스템에서든 일정한 기록 활동을 수행할 수 있는 기능성을 갖추도록 하는 방향으로 바뀌고 있으므로, 시스템이나 기록관리 단계별 구분이 유효하지 않을 것으로 판단하였다.

관계 표현을 위한 구성요소는 PROV의 관계 개념을 수용하였다. 주요하게 확장한 대상은 파생 관계와 효력의 발생과 무효화 관계이다. 파생 관계의 경우 역시 앞서 검토한 것처럼 내용이나 구조를 변경하지 않고 물리적 위치나 매체만 변경되어 발생하는 개체 파생을, 구조나 논리적 내용이 변경되는 개체 파생과 구분할 필요가 있다고 판단하여 세분하였다. 또, 2장의 1절에서 서술한 것처럼 '무효화되었다'의 관계와 조응하여 '효력이 발생했다'의 관계를 추가하였는데, 예를 들어 특정 개체의 마이그레이션 이후에 발생하는 2개의 개체에 대해 마이그레이션 활동을 기준으로 그 이전에 존재했던 개체의 효력이 없음을 선언하고 이후 발생한 개체의 효력이 발생했다는 관계 표현을 할 수 있을 것이다. 또, 효력이 발생하였다는 관계 표현은 공식 기록으로 선언되거나, 특정한 권한의 실행 근거가 유효하게 되는 기점을 제공하는 데 도움이 될 것으로 판단했다.

〈표 7〉 데이터형 전자기록의 출처 모델 구성(안): 소급형 출처 영역

구분	구성요소	하위 구성요소	하위-하위 구성요소	설명	관계 방향*
개체	개체			물리적, 디지털, 개념적 또는 여타 종류인 것. 고정된 측면을 지니고 있음	-
		디지털 데이터 객체		비트열로 구성된 객체 커뮤니케이션, 해석 또는 처리에 적합한 형식화된 방식으로 재해석 가능한 정보 표현(비트열, 숫자 표, 페이지의 문자, 소리 녹음 등)	-
		컴포넌트		기록을 구성하는 인스턴스로서의 객체	-
		집합		일정한 디지털 데이터 객체나 컴포넌트의 집합	-
		출처기록집합		출처 기록에 대한 집합. 일련의 출처 기록을 하나의 개체로 집합화	-
활동	활동			일정 기간 발생하며 개체를 사용하거나 산출하는 행위	-
		생산		활동을 수행하면서 개체를 산출함	-
		획득		보유·관리할 필요가 있는 개체를 기록으로 선언하고 기록시스템으로 포착함	-
		분류·색인		정해진 범주(예. 업무맥락)에 해당하는 개체를 체계적으로 확인하고 정리하며, 검색에 필요한 색인을 부여함	-
		평가		평가도구에 정해진 범주에 해당하는 개체를 체계적으로 확인하고 평가 결정을 부여함	-
		접근통제		개체에 대해 접근규칙을 적용해 접근을 제공하거나 제한함	-
		저장		저장 환경과 매체에 개체를 배치함	-
		이용		개체를 탐색·활용할 수 있도록 제공함	-
		마이그레이션·변환		하나의 하드웨어나 소프트웨어 구성에서 다른 것으로 개체를 이전함. 포맷의 변화가 발생하거나 발생하지 않을 수 있음	-
		처분		처분지침이나 다른 도구에서 정한 방식으로 개체를 보유, 파기 또는 이관함	-
			이관	처분지침이나 다른 도구에서 정한 방식으로 개체를 다른 관리권으로 이전함	-
			파기	처분지침이나 다른 도구에서 정한 방식으로 개체를 복구할 수 없게 파기함	-
			보유	처분지침이나 다른 도구에서 정한 방식으로 개체를 보유함	-
	입수		특정 관리권이나 시스템으로 개체를 가져옴	-	

구분	구성요소	하위 구성요소	하위-하위 구성요소	설명	관계 방향*
행위자	실행주체			발생한 활동이나 개체의 존재, 또는 또 다른 실행주체의 활동에 대해 일정 책임의 형태를 지니는 것	-
		개인		실행주체가 개인임	-
		조직		실행주체가 사회적이거나 법정 기관임	-
		소프트웨어		실행주체가 구동되는 소프트웨어임	-
관계	~를(을) 사용하였다			활동에서 개체를 사용했음	활동→개체
	만들어졌다			활동에서 개체가 산출되었음	개체→활동
	실행주체와 연결되었다			활동이나 개체가 실행주체와 연결되었음	활동·개체→실행주체
	~의 인스턴스이다			전망형 출처 영역 요소의 인스턴스임	소급 출처→전망 출처
	~에서 개체를 전달받았다			앞선 활동이 생성한 불특정 일부 개체를 (후행) 활동이 사용함	(후행)활동→(선행)활동
	~에서 파생되었다			개체의 이력(변형, 갱신, 일부 활용 등)을 표현함	개체(후행)→개체(선행)
		물리적 위치가 변경되었다		데이터의 내용과 구조 변화 없이 위치나 접근권한이 변경되었음	개체(후행)→개체(선행)
		(구문·동사적) 구조가 변경되었다		데이터의 논리적 내용은 유지하지만 구조나 인코딩 방식이 변경되었음	개체(후행)→개체(선행)
		논리적 내용이 변경되었다		데이터의 논리적 내용이 변경되었음	개체(후행)→개체(선행)
	~를(을) 대신해 수행하였다			대리인 또는 대표자로서 특정 활동을 수행할 권한과 책임을 위임받아 수행했음	실행주체→실행주체
	시작되었다			트리거(개체)에 따라 어떤 활동이 시작되었음. 활동이 시작되어야 사용, 생성, 무효화 등이 진행될 수 있음	활동→개체
	종료되었다			트리거(개체)에 따라 어떤 활동이 종료되었음	활동→개체
	무효화되었다			활동 수행에 따라 개체가 파기되거나 효력·영향이 중단 또는 만료되었음	개체→활동
	효력이 발생하였다			개체의 효력·영향이 활동 수행에 따라 시작되었음	개체→활동
	~의 대체형이다			유호한 동일 사물임을 제시하는 개체 간 관계임(예. 웹버전 - 모바일버전)	개체→개체
	~의 특수형이다			일반형-특수형 개체를 연결하는 관계임	개체→개체
	~의 구성원이 있다			집합에 특정 개체가 있음	집합→개체
	~의 구성원이다			개체가 특정 집합에 구성되어 있음	개체→집합
	~(으)로 구성된다			출처기록집합에 포함된 구성요소를 표현함	출처기록집합→활동·실행주체·개체

구분	구성요소	하위 구성요소	하위-하위 구성요소	설명	관계 방향*
속성 (일반)	식별자			객체를 고유하게 식별하는 기호	
	제목			대상을 설명하는 문자어	
	역할			개체나 행위자가 활동에서 수행한 기능이나 임무	
	위치			지리적 위치를 포함해 대상의 위치를 다양하게 표현함	
	유형			추가 설명의 종류를 제공함	
	값			직접 갖는 값을 표현함	

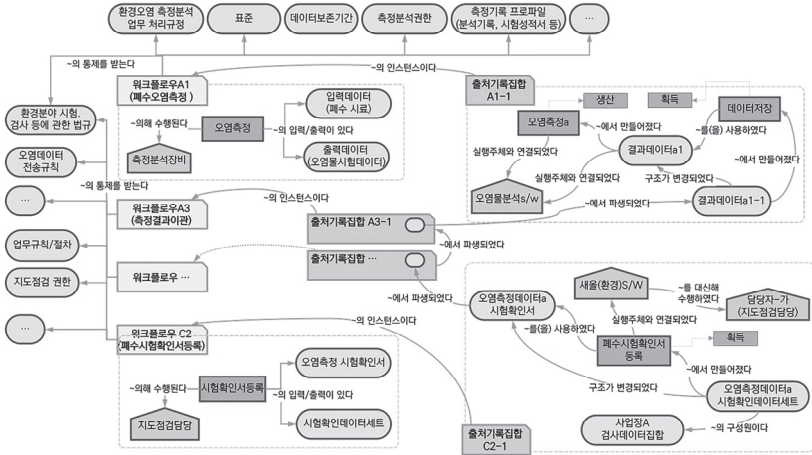
\* 관계의 방향은 '관계'에 해당하는 요소에만 서술하였음

## 2) 적용 예시

앞 절에서 구성한 모델을 활용하여 행정정보시스템 중 '서울행정시스템'의 환경업무 모듈 중 인허가 사업장에 대한 폐수오염 지도점검 기능의 일부를 <그림 2>와 같이 표현해 보았다. 사례 표현을 위해서는 공개된 '서울행정시스템<sup>1)</sup> 사용자 매뉴얼'(이하 서울매뉴얼)(한국지역정보개발원, 2017. 5. 19.)을 기초로 하였으며, 출처 표현의 확장을 위해 환경부 훈령인 '환경오염 측정분석업무 처리규정'(이하 오염측정규정)을 참고하여 지도점검에 사용되는 폐수오염시험 절차를 함께 구성하였다. 단, 이 절의 예시는 전체 흐름을 포괄적으로 표현하기보다 모델의 특징을 드러내는데 좀 더 초점을 두어, 각 영역과 구성요소를 일부 축약하여 제시하였다. 또, 데이터가 발생하는 활동을 기록 생산으로, 공식적으로 식별기호를 부여받아 최종 저장 또는 등록되는 활동을 기록 획득으로 설정하고 표기하였다. 이는 이후 실행 환경에서 달리 설정할 수 있다.

1) 서울행정시스템은 전국적으로 수행하는 환경은 물론 기획, 의회, 공보 상하수도, 농업, 산림 등 22개 행정업무를 지원하기 위해 다양한 데이터를 처리하는 정보시스템이다(행정안전부, 발행년불명).

〈그림 2〉 표현 예시: 폐수오염 측정 및 시험확인서 등록



폐수오염을 측정하고 시험확인서를 등록하는 이유는, 인허가를 신청하는 사업장의 민원을 처리하거나, 기존 인허가 사업장에 대한 지도점검을 수행하는 과정에 폐수오염도 검사를 진행하고 그 과정과 결과를 관리하여 민원에 대해 공개(한국지역정보개발원, 2017. 5. 19.)하기 위해서이다. 사업장에서 배출된 오염물질 시료를 검사하기 위해서는 법규에 따른 시험기관에서 환경오염도를 측정·분석하게 되는데, 이는 새울매뉴얼의 범위 밖에 있다. 따라서 새울매뉴얼에 안내된 시험확인서 등록 이전의 절차 구성은 오염 측정규정을 참고하여 구성하였다.

우선 <그림 2>의 오른쪽 위, ‘출처기록집합 A1-1’은 특정 폐수 시료를 대상으로 오염도를 측정하는 활동을 중심으로 구성된다. 오염분석에는 측정을 위한 도구나 소프트웨어가 사용되므로, 이를 실행주체로 표현하고 측정을 통해 ‘결과데이터a1’이 산출되어 ‘결과데이터a1-1’로 최종 저장되는 흐름을 표현하였다. 측정 후 산출된 ‘결과데이터a1’과 저장되는 ‘결과데이터a1-1’은 내용은 같지만, 저장 체계에 따라 구조가 변경되어 최종적으로

저장될 것으로 표현하였다. 이러한 측정 활동은 미리 정의된 ‘폐수오염측정 워크플로우A1’의 인스턴스 중 하나로 연결될 것이다.

〈그림 2〉의 왼쪽 위, ‘워크플로우A1’에서는 오염물질 측정을 규제하고 통제하는 법규와 표준, 측정 권한, 측정 결과로 산출될 데이터나 기록의 프로파일 등을 반영해 측정 활동이 따라야 하는 과업의 단계와 입출력 데이터를 보여준다. 따라서, 이후 오염측정 결과데이터가 어떤 과정에 거쳐 산출되었는지 확인할 수 있도록 하는 한편, 관련 규제와 통제 범위에서 신뢰할 수 있게 산출된 것인지 검증하는 과정을 지원할 수 있을 것으로 판단한다. 시험기관은 오염측정을 통해 저장된 데이터를 활용하여 시험확인서를 생산하고, 새울행정시스템에 등록할 수 있도록 전달하는데, 이 과정은 〈그림 2〉의 중앙에 축약적으로 표현하였다.

새울매뉴얼에 지시된 등록 과정은 〈그림 2〉의 오른쪽 아래에 구성되어 있다. ‘폐수시험확인서 등록’ 활동은 새울행정시스템에서 실행되는 기능 중 하나로, 전달받은 ‘오염측정데이터a 시험확인서’를 사용해 특정 인허가 사업장에 대한 오염측정데이터세트에 데이터를 추가하는 과정을 거친다. 이 활동은 새울행정시스템에서 ‘지도점검’ 기능에 구성되어 있어, 지도점검을 수행할 역할을 부여받은 ‘담당자-가(지도점검담당)’가 ‘새울(환경)S/W’를 사용해 수행하는 관계로 표현하였다. 등록 활동에 투입된 ‘오염측정데이터 a 시험확인서’의 데이터는 ‘오염측정데이터a 시험확인데이터세트’의 일부를 구성하면서, 시험확인서에 포함된 데이터 구조와 달리 저장될 수 있으므로, 구조변경의 관계도 함께 표현하였다. 또, 이 데이터세트는 폐수 시료를 채취한 사업장의 전체 검사 데이터집합의 일부일 것이므로, ‘사업장A 검사 데이터집합’의 구성원 관계도 표현하였다. 이렇게 표현된 소급형 ‘출처기록 집합C2-1’은 다시 이에 대한 전망형 출처 ‘워크플로우C2(폐수시험확인서 등록)’와 인스턴스 관계로 연결될 것이다.



## 6. 맺음말

이 연구는 일반 출처 모델로 개발된 PROV를 기초로 하여 데이터형 전자기록의 출처를 표현할 수 있는 모델을 시험적으로 개발하였다. 선행 연구의 모델 설계 방향에 따라 소급형 출처 영역과 전망형 출처 영역을 각각 표현하고 연결할 수 있도록 모델을 설계하였으며, 전망형 출처 영역을 위해 P-Plan과 ProvONE도 함께 참고하여 확장하였다. 영역별 목적에 맞게 출처를 표현할 수 있도록 기초 모델의 구성요소를 조정하여 편성하고, 기록관리 환경에 적용할 수 있도록 요건을 확인한 결과를 반영하여 요소를 세분해 전개하였다.

이미 일상 업무 환경에서 다양한 정보시스템을 사용하고 있고, 그 과정에 생산되는 데이터 형태의 전자기록을 식별하고 관리할 요구가 늘어나고 있다. 이 연구가 시범적으로 개발한 데이터형 전자기록의 출처 모델을 활용하면, 데이터객체 단위의 이력을 조밀하고 연속적으로 추적할 수 있는 한편, 이를 미리 정의된 규제나 절차, 기준에 부합하는지 확인할 수 있도록 지원할 기록화된 정보를 축적할 수 있을 것이다. 이렇게 축적된 정보는 특정 정보시스템이나 네트워크 환경, 심지어 인지 기술 기반의 시스템 환경에서 생산되는 기록의 관리와 보존 활동을 지원할 수 있는 기초자료로 활용될 가능성을 지니며, 나아가 특정 업무 환경에서 기록관리 메타데이터로도 재사용될 가능성도 높다.

이러한 가능성을 실현하기 위해서는, 실제 데이터형 전자기록을 다루는 환경에 부합하도록 이 연구의 모델을 추가 확장할 필요가 있다. 시스템별로 활동이나 행위자, 역할, 규범이나 통제 도구의 범위가 다를 것이기 때문이다. 이 연구에서는 활동이나 행위자, 통제 체계 등을 그룹화할 수 있는 수준까지 세분화하여 일반적인 적용이 가능한 범위로 기초 모델을 확장하였다. 따라서 실제 정보시스템이나 실무에 적용할 때, 개별 시스템에 구현된 기능이나 행위자 및 개체를 비롯해 시스템에서 다루는 대상을 고유하게 식별하고, 이를 활용하여 이 연구에서 제안한 모델을 한 단계 더 확장하여 대상을

표현할 수 있을 것이다. 이 연구에서 시범적으로 개발한 모델을 바탕으로, 데이터형 전자기록의 출처에 대한 논의가 이루어지고, 진전된 모델을 개발하는 후속 연구가 이어지길 기대한다. 또한 기록관리 메타데이터와 함께 전자기록의 관리를 지원하고 그의 종합적인 품질을 측정할 수 있는 도구로서 기록 출처 모델의 역할을 탐구하는 연구도 진행되길 기대한다.

### 〈참고문헌〉

- 박미영 (2019). 행정정보데이터세트 획득 방안 연구: 국방인사정보체계 개인 인사기록을 중심으로. 석사학위논문, 서울대학교 대학원 협동과정 기록학과.
- 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.
- 이승역, 설문원 (2017). 전자기록관리정책의 재설계에 관한 연구. 기록학연구, 52, 5-37.
- 한국지능정보사회진흥원 (2023). 2023년도 범정부EA기반 공공부문 정보자원 현황 통계보고서. 출처: [https://www.irm.go.kr/cmm/login/getBBSDataListDetail.do?ntt\\_id=63622#none](https://www.irm.go.kr/cmm/login/getBBSDataListDetail.do?ntt_id=63622#none)
- 한국지역정보개발원 (2017. 5. 19.). 새울행정시스템 사용자 매뉴얼: 환경업무(배출관리)(버전 1.84). 출처: <https://www.jeonnam.go.kr/D0303/boardView.do?seq=1930107&infoReturn=&menuId=east0301000000&displayHeader=&searchType=0&searchText=%ED%99%98%EA%B2%BD&pageIndex=3&boardId=D0303&displayHeader=>
- 현문수 (2022). 전자기록의 신뢰가치 확립을 위한 증거능력 구현체계 연구: 우리나라 공공 전자기록의 신뢰가치 모델 개발을 중심으로. 기록학연구, 73, 5-46.
- 현문수 (2024a). 데이터형 전자기록을 위한 출처 개념 모델 개발 방향. 기록학연구, 79, 305-341.
- 현문수 (2024b). 전자기록관리: 디지털 환경에서 전자기록을 어떻게 관리할 것인가? 한국기록관리학회 역음. 기록관리의 세계: 이론에서 실천까지. 파주: 한울 아카데미, 313-349.
- 행정안전부 [발행년불명]. 지방행정정보화. 출처: <https://www.mois.go.kr/frt/sub/a06/b04/localGovernmentIt/screen.do>

- Cao, Y., Jones, C., Cuevas-Vicentín, V., Jones, M., Ludäscher, B., McPhillips, T., Missier, P., Schwalm, C., Slaughter, P., Vieglaiss, D., Walker, L. & Wei, Y. (2016). ProvONE: extending PROV to support the DataONE scientific community. Available: <http://homepages.cs.ncl.ac.uk/paolo.missier/doc/dataone-prov-3-years-later.pdf>
- Car, N. (2017. 3. 15.). Intro to PROV. Provenance and Social Science Data Webinar. Available: <https://www.youtube.com/watch?v=elPcKqWoOPg>
- Ciccarese, P., Soiland-Reyes, S., Belhajjame, K., Gray, A., Goble, C. & Clark, T. (2013). PAW ontology: provenance, authoring and versioning. *Journal of Biomedical Semantics*, 4(37). Available: <https://doi.org/10.1186/2041-1480-4-37>
- CCSDS (2012). Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data System Practices. CCSDS 650.8-M-2. Available: <https://public.ccsds.org/pubs/650x0m2.pdf>
- 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#>
- DataONE [n.d.]. About DataONE. Available: <https://www.dataone.org/about/>
- Eastwood, T., Ballaux, B., Mills, R. & Preston, R. (2008). Appendix 14: Chain of Preservation model - Diagrams and definitions. In Duranti, L. & Preston, R. eds. *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Available: [http://www.interpares.org/display\\_file.cfm?doc=ip2\\_book\\_appendix\\_14.pdf](http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_14.pdf)
- Eastwood, T., Hofman, H. & Preston, R. (2008). Part five-Modeling digital records creation, maintenance and preservation: Modeling cross-domain task force report. In Duranti, L. & Preston, R. eds. *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*. Available: [http://www.interpares.org/display\\_file.cfm?doc=ip2\\_book\\_part\\_5\\_modeling\\_task\\_force.pdf](http://www.interpares.org/display_file.cfm?doc=ip2_book_part_5_modeling_task_force.pdf)

- Garijo, D. & Gil, Y. (2013). The P-Plan ontology. Available: <https://www.opmw.org/model/p-plan/>
- 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/>
- ICA EGAD (2023). International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 1.0. Available: <https://github.com/ICA-EGAD/RiC-O/tree/master/ontology/current-version>
- InterPARES Trust (2019). InterPARES Trust responds to EGAD-RiC. Available: <https://interparestrustblog.wordpress.com/2016/12/11/inter pares-trust-responds-to-egad-ric/>
- 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.
- Li, C. & Sugimoto, S. (2014). Provenance description of metadata using PROV with PREMIS for long-term use of metadata. International Conference on Dublin Core and Metadata Applications, 147-156.
- 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.
- 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/>
- OPM [n.d.]. The OPM Provenance Model (OPM). Available: <https://openprovenance.org/opm/>
- Pandit, H. (2019). GDPRov - The GDPR provenance ontology. Available: <https://openscience.adaptcentre.ie/ontologies/GDPRov/docs/ontology>
- Prom, C. (2016. 6. 8.). Archival arrangement and description in the cloud: A preliminary analysis. 21th Archival Science Colloquium/International Symposium, In Cooperation with InterPARES Trust(ITRUST). Available: [https://www.archivschule.de/uploads/Kolloquium/2016/Panel2\\_3\\_Prom\\_Arrangement\\_Description\\_in\\_Cloud\\_Marburg.pdf](https://www.archivschule.de/uploads/Kolloquium/2016/Panel2_3_Prom_Arrangement_Description_in_Cloud_Marburg.pdf)

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.

〈표준, 법령〉

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

문헌정보-기록관리과정-기록메타데이터-제1부: 원칙. KS X ISO 23081-1:2017

문헌정보-기록관리과정-기록메타데이터-제2부: 개념과 실행 고려사항. KS X ISO TS 23081-2:2008(2020 확인)

환경오염 측정분석업무 처리규정 [환경부훈령 제859호, 2009. 8. 18., 타법개정][시행 2009. 8. 18.]