

행정정보 데이터세트의 장기보존을 위한 장기보존패키지(NEO3) 확대 방안 연구

A Study on Enhancing the NEO3 Long-Term Preservation Package for Administrative Information Dataset

김지혜(Jihye Kim)¹, 이지영(Jiyoung Lee)², 현경빈(Gyeongbin Hyun)³,
양동민(Dongmin Yang)⁴

E-mail: kimjh129056@gmail.com, athezoo@korea.kr, hkb3216@gmail.com, dmyang@jbnu.ac.kr

¹ 제 1저자 전북대학교 기록관리학과 박사과정

² 국가기록원 공업연구관

³ 전북대학교 기록관리학과 석사과정

⁴ 교신저자 전북대학교 기록관리학과 교수, 문화융복합아카이빙연구소/고려인연구센터 공동연구원



논문접수 2026-07-15

최초심사 2026-07-27

게재확정 2026-08-10

ORCID

Jihye Kim
<https://orcid.org/0000-0002-0045-3670>

Jiyoung Lee
<https://orcid.org/0009-0007-1881-4802>

Gyeongbin Hyun
<https://orcid.org/0009-0004-6554-6837>

Dongmin Yang
<https://orcid.org/0000-0002-4029-9372>

© 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.

- 본 연구는 한국연구재단의 '2024년도 인문 사회분야 학술지원사업 학문후속세대 지원사업(박사과정생 연구장려금지원)'의 지원을 받아 수행된 연구임.
(과제번호: NRF-2024S1A5B5A19044086).
- 본 연구는 '2025년 행정안전부 국가기록원 기록관리 연구개발사업'의 연구비를 지원받아 수행되었음.

초 록

국가기록원은 보존가치가 높은 전자기록물을 장기간 보존·활용하기 위해 보존포맷과 장기보존패키지를 중심으로 장기보존 정책을 운영하고 있다. 보존기간이 30년 이상인 전자기록물은 법령에 따라 장기보존패키지로 변환하여 관리하여야 한다. 행정정보 데이터세트 역시 장기보존패키지의 적용 대상이나, 현행 장기보존패키지 기술규격(NAK 31-2:2022)은 일부 구성요소가 문서유형 전자기록물을 중심으로 설계되어 있어 행정정보 데이터세트에 그대로 적용하기 어렵다. 따라서 본 연구는 장기보존패키지(NEO3)를 행정정보 데이터세트에 적용하기 위한 확대 방안을 제시하는 것을 목적으로 한다. 이를 위해 장기보존패키지의 구조와 각 구성요소의 개념 및 범위를 분석하고, 장기보존패키지에 포함해야 할 행정정보 데이터세트 기록의 보존 범위를 도출하였다. 나아가 행정정보 데이터세트의 보존 범위를 기준으로 현행 장기보존패키지 구성요소를 분석하여 데이터세트 유형에 적합한 개선(안)을 제시하였다. 본 연구는 행정정보 데이터세트의 장기보존을 위한 장기보존패키지의 활용 방향을 제시하였다는 점에서 의미가 있다.

ABSTRACT

The National Archives of Korea (NAK) operates long-term preservation policies centered on preservation formats and long-term preservation packages to ensure the long-term preservation and utilization of electronic records. Electronic records with retention periods of 30 years or longer must, under relevant legislation, be converted into long-term preservation packages and managed accordingly. Administrative information datasets are likewise subject to the long-term preservation package framework; however, the current long-term preservation package technical specification (NAK 31-2:2022) includes components designed specifically for document-type records, which limits its applicability to administrative information datasets. This study, therefore, aims to propose ways to expand the long-term preservation package (NEO3) to accommodate administrative information datasets. To this end, the structure of the long-term preservation package and the concept and scope of each component were analyzed, and the scope of the administrative information dataset records to be included in the long-term preservation package was identified. Based on the identified preservation scope, the components of the current long-term preservation package were analyzed, and an improvement plan suitable for administrative information datasets was proposed. This study is significant in that it provides guidance on the use of long-term preservation packages for the long-term preservation of administrative information datasets.

Keywords: 행정정보 데이터세트, 장기보존패키지, 장기보존, NEO3, 전자기록관리

Administrative information dataset, Long-term preservation package, Long-term preservation, NEO3, Electronic records management

<http://jksarm.koar.kr>

www.kci.go.kr

1. 서 론

1.1 연구배경 및 목적

우리나라의 공공기관은 「전자정부법」에 따라 행정정보시스템을 구축·운영하여 행정업무를 전자적으로 수행하고 있다. 「공공기록물 관리에 관한 법률 시행령」(이하 공공기록물법 시행령)은 행정정보시스템에서 생산되는 데이터의 집합을 행정정보 데이터세트(이하 데이터세트)로 정의한다. 또한 데이터세트를 전자기록물의 한 유형으로 분류하여 기록관리 대상으로 규정한다. 데이터세트는 법적으로 새로운 유형의 전자기록물이지만, 그 내용은 기존의 기록관리 대상이 데이터 형태로 데이터 형태로 생산·관리되는 것으로 볼 수 있다. 즉, 기록의 표현 방식이 문서에서 데이터로 변화하였을 뿐, 업무 수행 과정에서 생산되는 기록이라는 본질과 기록관리 필요성은 기존 기록과 다르지 않다. 이에 따라 데이터세트 역시 다른 전자기록물과 마찬가지로 생산부터 보존 및 활용에 이르기까지 전자기록관리 체계에 따라 관리되어야 한다. 그러나 데이터세트는 시스템이 운영되는 동안 언제든지 변경될 수 있어 기록을 확정하는 고정시점을 특정하기 어렵다. 또한 데이터세트는 다수의 레코드로 구성된 집합적 구조이므로 어느 범위를 하나의 기록으로 볼 것인지 명확하지 않으며, 기록물건과 같은 문서 중심의 기록관리 단위에도 대응되기 어렵다. 이로 인해 기존 전자기록관리 체계를 그대로 적용하기보다 데이터세트의 특성을 반영한 기록관리 방안이 요구되었다.

이러한 필요성에 따라 국가기록원은 데이터세트 기록관리 제도를 마련하고, 「공공기록물법 시행령」에 관련 조항을 신설하였다. 법령에서는 행정정보시스템을 기본 단위로 데이터세트를 관리하고, 그 실행방안으로 데이터세트 관리기준표를 작성·운영하도록 규정하고 있다(제25조제6항, 제34조의3). 또한 데이터세트에 대해서는 물리적 이관뿐 아니라 처리과와 기록관 간 관리권한을 이전하는 방식도 이관의 한 형태로 인정하고 있다(제31조의2). 국가기록원은 세부 기준으로 기록물관리 공공표준 「행정정보 데이터세트 기록관리 기준」(국가기록원, 2020, NAK 35:2020, v1.0)을 제정하여 관리기준표의 작성 방법과 관리항목을 구체화하였다. 그러나 현행 제도와 표준은 데이터세트의 관리 대상 식별과 보존가치 평가에 관한 사항을 중심으로 규정하고 있으며, 이관 및 장기보존에 관한 구체적인 기준은 별도로 제시되어 있지 않다. 데이터세트는 대부분 행정정보시스템 내에서 관리·보존되고 있으며, 행정정보시스템은 업무 수행을 목적으로 구축·운영되기 때문에 기록관리를 위한 기능이 충분히 갖추어져 있지 않다. 이로 인해 시스템이 교체·폐기되거나 운영 환경이 변화할 경우, 해당 시스템에 의존하여 관리되던 데이터세트의 지속적인 접근과 재현이 어려워질 수 있다. 따라서 데이터세트를 장기간 보존하기 위해서는 특정 시스템 환경에 종속되지 않는 형태로 데이터세트를 분리하여 보존·관리할 수 있는 방안이 필요하다.

국가기록원은 전자기록물의 장기보존을 위해 보존포맷과 장기보존패키지를 중심으로 관련 정책을 운영하고 있다. 「공공기록물법 시행령」에서는 보존기간이 10년 이상인 전자기록물에 대해서는 보존포맷으로 변환하여 관리할 수 있도록 하고, 보존기간이 30년 이상인 전자기록물은 장기보존패키지로 변환하여 관리하도록 의무화하고 있다. 국가기록원은 이를 위한 세부 기준으로 보존포맷 선정기준과 장기보존패키지 기술규격을 기록물관리 공공표준으로 제정하였다. 두 표준은 초기에는 문서 유형의 전자기록물을 중심으로 마련되었으나, 전자기록물의 유형이 다양해짐에 따라 이를 수용하기 위한 개정과 개선이 이루어졌다. 보존포맷 선정기준의 경우 개정을 통해 시청각기록물과 데이터세트까지 적용할 수 있도록 대상 유형을 확장하였다. 장기보존패키지 또한 전자기록물의 대용량화와 유형 다양화에 대응하여 패키징 방식이 개선되었다. 초기 장기보존패키지(NEO)는 모든 구성요소를 하나의 XML 문서에 Base64 방식으로 인코딩하여 캡슐화하는 구조로 개발되었다. 이후 대용량 전자기록물을 보다 효율적으로 처리하고 다양한 유형의 전자기록물을 수용하기 위해 디렉토리 기반의 논리적 구조와 ZIP64 패키징 방식을 적용한 장기보존패키지(NEO3)가 개발되었으며, 이와 구분하여 기존의 Base64 기반 장기보존패키지는 NEO2로 명명되었다. NEO3는 문서뿐만 아니라 시청각 기록물과 데이터세트 등 다양한 유형의 전자기록물을 수용할 수 있도록 설계되었으나, 장기보존 메타데이터를 비롯한 일부 구성요소의 정의와 범위에는 여전히 문서

유형을 전제로 한 특성이 남아 있다. 따라서 데이터세트에 NEO3를 적용하기 위해서는 데이터세트의 특성과 보존 범위를 고려하여 해당 구성요소의 개념과 범위를 검토할 필요가 있다.

데이터세트는 전자기록물의 한 유형으로, 장기적으로 보존할 가치가 있는 경우 장기보존패키지로 변환하여 관리하여야 한다. 그러나 현행 장기보존패키지 기술규격은 데이터세트의 무엇을 보존 범위에 포함하고 이를 각 구성요소에 어떻게 수용할 것인지 등 데이터세트를 장기보존패키지로 변환·관리하기 위한 구체적인 기준을 제시하지 않고 있다. 따라서 데이터세트에 장기보존패키지를 적용하기 위해서는 먼저 데이터세트의 특성을 고려하여 보존 범위를 설정하고, 이를 기준으로 현행 장기보존패키지 구성요소의 적용 가능성을 검토할 필요가 있다. 이에 본 연구에서는 장기보존패키지에 포함되어야 할 데이터세트의 보존 범위를 도출하고, 이를 기준으로 현행 장기보존패키지 구성요소의 적용 가능성을 분석하여 데이터세트 유형에 적합한 장기보존패키지의 구조와 구성요소를 제안하고자 한다.

1.2 선행연구

데이터세트에 관한 연구는 그 특성에 적합한 기록관리 방안을 마련하기 위해 다양한 측면에서 이루어져 왔다. 초기에는 데이터세트 가운데 기록관리 대상을 식별·선별하는 방안에 관한 연구가 주로 이루어졌으며, 이후 관리 기준표의 작성과 적용 등 실무적 차원의 기록관리 방안에 관한 논의로 확장되었다. 이와 함께 데이터세트를 지속적으로 보존·관리하기 위한 방안에 관한 연구도 이루어져 왔다. 그중 데이터세트의 장기보존에 관한 연구는 데이터세트의 재현과 SIARD를 중심으로 한 보존포맷 및 이관규격에 관한 연구를 중심으로 이루어졌다. 왕호성과 설문원(2017)은 데이터세트를 전자기록으로 관리하기 위해서는 내용, 구조, 맥락뿐만 아니라 외형과 기능까지 함께 보존하여 재현성을 확보해야 한다고 제시하였다. 특히 데이터세트는 생산시스템과 분리되는 과정에서 사용자 인터페이스와 기능이 소실될 수 있으므로, 이를 고려한 장기보존 방안이 필요함을 강조하였다. 또한 데이터세트의 최소 관리단위 설정과 재현 전략에 대한 논의를 통해 기존 문서 중심 기록관리와는 차별화된 접근이 필요함을 제시하였다. 한희정 외(2020)는 데이터세트의 특성을 반영한 보존포맷 선정기준과 평가체계를 마련하고, 해당 기준을 바탕으로 SIARD의 적용 가능성을 검토하였다. 이어 윤성호 외(2021)는 다양한 DBMS 환경에서 SIARD를 적용한 실험을 수행하여 데이터 타입과 구조 요소의 보존 가능 범위를 검증하고, 일부 데이터 유형에서 정보 손실과 호환성 문제가 발생할 수 있음을 확인하였다. 두 연구는 데이터세트 보존포맷으로서 SIARD의 적용 가능성을 검토하는 한편, 지속적인 기능 개선과 기술적 검증의 필요성을 제시하였다. 변우영, 임진희(2022)는 SIARD_KR이 데이터세트의 특성을 충분히 반영하는지를 분석하고 구조적 한계와 개선방안을 제시하였다. 현행 SIARD_KR은 데이터베이스 전체를 추출하는 방식으로 설계되어 의미 있는 데이터세트 단위의 식별과 사용자 인터페이스, 첨부 파일 등의 관련 정보 보존에 한계가 있음을 지적하였다. 연구에서는 데이터세트의 의미 단위와 재현 특성을 고려한 이관도구의 기능 개선과 구조 보완의 필요성을 제안하였다. 김수영(2023)은 실제 데이터세트 이관 사례를 중심으로 SIARD_KR 적용 과정에서 발생하는 오류와 문제점을 분석하였다. DB 덤프 방식과 오픈포맷 방식 등 다양한 이관 방식을 비교하여 데이터세트의 특성과 시스템 환경을 고려한 이관 절차의 개선 방향을 제시하였으며, 실제 이관 과정에서 발생하는 기술적 문제를 분석하여 데이터세트 장기보존을 위한 실무적 개선방안을 제안하였다. 박라미와 양동민(2026)은 기존 데이터 중심의 장기보존 방식이 사용자 인터페이스를 포함하지 못하는 한계를 지적하고, HTML 기반 사용자 인터페이스를 XSLT로 재현하는 방안을 제안하였다. 다양한 웹 기반 행정정보시스템의 HTML 구조를 유형화하고 유형별 XSLT를 설계하여 재현 가능성과 한계를 실증적으로 검증하였으며, 데이터뿐 아니라 사용자 인터페이스를 함께 보존함으로써 데이터세트의 구조와 맥락을 보다 충실하게 재현할 수 있음을 제시하였다.

한편, 전자기록물 장기보존패키지에 관한 연구는 다른 기록관리 분야에 비해 활발하게 이루어지지 않았다. 임진희(2006)는 OAIS 참조모형의 정보패키지 개념을 기록관리에 적용하여 기록 보존정보패키지(AIP)의 구성과

구조를 제안하였다. 연구에서는 기록 보존정보패키지(AIP)의 구조를 개념, 논리, 물리의 세 가지 레이어로 구분하고, 보존전략과 이용 요구에 따라 다양한 구조를 설계할 수 있음을 제시하였다. 박병주 외(2010)는 당시 전자기록물 장기보존포맷(NEO)이 웹기록물의 특성을 충분히 반영하지 못한다는 한계를 지적하고, 표면 웹기록물과 심층 웹기록물을 장기보존포맷에 수용할 수 있도록 메타데이터와 콘텐츠 구조를 확장하였다. 또한 웹기록물의 보존과 복원에 필요한 기술 메타데이터를 추가하고 기존 장기보존포맷의 구조를 확장하여 웹기록물을 전자기록물과 함께 장기적으로 보존할 수 있는 방안을 제시하였다. 이 연구는 특정 기록 유형을 수용하기 위해 기존 장기보존포맷의 구조를 확장한 초기 연구라는 점에서 의의를 가진다. 한능우(2023)는 전자기록 장기보존정책의 변천 과정을 종합적으로 검토하면서 장기보존패키지 정책의 발전 과정과 향후 과제를 분석하였다. 연구에서는 XML 기반의 NEO2와 디렉토리 기반의 NEO3의 특징을 비교하고, 논리적 장기보존패키지의 도입 배경과 의의를 설명하였다. 나아가 기록 유형별 장기보존패키지의 확대 적용과 SIP 및 DIP에 대한 추가 설계의 필요성을 제시하며 장기보존정책의 발전 방향을 논의하였다.

선행연구를 검토한 결과, 데이터세트의 장기보존에 관한 연구는 재현, 보존포맷, 이관 등 다양한 측면에서 이루어져 왔으나 장기보존패키지에 관한 논의로는 이어지지 않았다. 한편 장기보존패키지에 관한 연구는 개념과 구조, 특정 기록 유형의 적용 및 정책적 발전 방향 등을 다루었으나, 기록 유형별 특성을 반영한 구체적인 적용 방안에 관한 연구는 제한적으로 이루어졌다. 이에 본 연구에서는 데이터세트의 보존 범위를 도출하고, 이를 바탕으로 데이터세트 유형에 적합한 장기보존패키지의 구조와 구성요소를 제안하고자 한다.

1.3 연구방법

본 연구는 행정정보 데이터세트를 장기보존패키지로 관리할 수 있도록 장기보존패키지(NEO3)의 확대 방안을 제안하는 것을 목적으로 한다. 이를 위해 다음과 같은 절차로 연구를 수행한다. 첫째, 기록물관리 공공표준인 「장기보존패키지 기술규격」(국가기록원, 2022, NAK 31-2:2022, v1.1)을 검토한다. 장기보존패키지 구성요소의 개념, 범위 및 역할과 장기보존패키지의 기본 구조 및 구성 방식을 검토함으로써, 현행 장기보존패키지가 무엇을 어떠한 방식으로 관리하도록 설계되어 있는지를 확인한다. 둘째, 데이터세트의 보존 범위를 도출한다. 데이터세트를 장기보존패키지로 관리하기 위해서는 먼저 어떠한 대상을 보존하여야 하는지를 설정할 필요가 있다. 이를 위해 데이터세트를 시스템으로부터 분리하여 별도로 보존하는 상황을 전제로 기록의 내용과 구조, 맥락을 유지하는 데 필요한 보존 범위를 도출한다. 셋째, 데이터세트의 보존 범위를 기준으로, 현행 장기보존패키지의 구성요소가 해당 범위를 수용할 수 있는지를 비교·분석한다. 이 과정에서 데이터세트 유형에 그대로 적용 가능한 요소, 개념 또는 범위의 재설계가 필요한 요소, 그리고 새롭게 추가되어야 하는 요소로 구분한다. 마지막으로, 앞선 분석 결과를 종합하여 데이터세트 유형에 적합한 장기보존패키지(NEO3) 확대 방안을 제안한다. 이를 위해 데이터세트의 특성을 반영한 장기보존패키지의 구조와 구성요소를 설계하고, 생성 및 관리방안을 제시한다.

2. 이론적 배경

2.1 장기보존패키지

전자기록물은 디지털 환경에서 생성·관리되기 때문에 이를 생성하거나 실행하는 하드웨어와 소프트웨어, 파일포맷 등의 기술적 환경에 영향을 받는다. 시간이 경과함에 따라 이러한 기술적 환경이 변화하거나 저장매체가 노후화될 경우 기록물의 훼손·유실뿐만 아니라 접근성과 재현성이 저하될 수 있다. 따라서 전자기록물을 장기간

보존하기 위해서는 기록의 내용과 구조, 맥락을 유지하면서 진본성과 무결성, 이용가능성을 보장할 수 있는 장기 보존 기술이 필요하다. 장기보존패키지는 이러한 요구를 충족하기 위해 기록물 자체와 이를 이해하는 데 필요한 정보를 하나의 패키지로 캡슐화하여 보존하는 장기보존전략이다.

장기보존패키지가 갖추어야 할 주요 요건으로는 자체 충족성, 자체 문서화, 진본성 및 무결성 유지가 있다. 자체 충족성은 전자기록물이 특정 시스템이나 외부 데이터에 의존하지 않고 생산 당시의 내용을 재현할 수 있도록 필요한 정보를 자체적으로 갖추는 것을 의미한다. 자체 문서화는 기록물의 내용과 맥락을 이해하는 데 필요한 정보를 포함하여 미래의 이용자가 장기보존된 기록물을 이해할 수 있도록 하는 것이다. 이를 위해 기록물에 대한 기술정보와 다른 기록물이나 기관·조직과의 관계, 지속적인 관리와 관련된 메타데이터 등을 포함할 필요가 있다. 진본성과 무결성 유지는 전자기록물이 장기보존패키지로 변환되고 저장·관리되는 과정에서도 위조되거나 변조되지 않았음을 확인할 수 있도록 하는 것이다. 이를 위해 전자서명 등의 기술을 활용하여 기록물의 진본성을 확인하고, 장기보존 과정에서 무결성을 지속적으로 검증할 수 있어야 한다.

국가기록원은 2008년 전자기록물의 진본성을 보장하고 장기적으로 보존하기 위해 ISO 14721 OAIS 참조모형에서 제시한 정보패키지 구조를 기반으로 NEO를 개발하였다. 이 과정에서 호주 PROV(Public Record Office Victoria)의 VEO를 벤치마킹하고, 원문 등을 Base64 방식으로 인코딩하여 하나의 XML 문서에 포함하는 정보패키지 구조를 적용하였다. NEO에는 OAIS에서 제시하는 내용정보(Content Information), 보존기술정보(Preservation Description Information), 패키징 정보(Packaging Information), 기술정보(Descriptive Information) 등의 개념이 반영되었다. 전자기록물의 용량이 증가하고 유형이 다양해지면서 XML과 Base64 인코딩을 기반으로 하는 기존 패키징 방식의 한계가 나타났다. 기록물의 용량이 커질수록 XML 문서의 크기와 처리 시간이 함께 증가했기 때문이다. 이를 개선하기 위해 기존 패키징 방식은 NEO2로 구분하고, 디렉토리나 ZIP을 기반으로 하는 새로운 방식인 NEO3가 마련되었다. NEO3는 각 구성요소를 파일과 디렉토리 형태로 구조화한 후 하나의 패키지로 구성하는 방식을 채택하였다. 원문은 별도의 Base64 인코딩 없이 파일 형태로 유지되며, 각 구성요소 역시 파일과 디렉토리 단위로 구분하여 관리할 수 있다. 최종적으로는 전체 구성요소를 ZIP64 규약에 따라 하나의 장기보존패키지로 구성한다. 이러한 구조는 대용량 전자기록물과 문서 이외의 다양한 유형의 전자기록물을 수용하기 위해 마련되었다.

2.2 행정정보 데이터세트

데이터세트는 공공기관의 행정업무 수행 과정에서 행정정보시스템을 통해 생산·관리되는 전자기록물이다. 「공공기록물법 시행령」에서는 이를 “공공기관의 행정정보시스템에서 생산, 수집, 가공, 저장, 검색, 제공, 송신 및 수신 등을 위하여 조합된 문자, 숫자, 도형, 이미지 및 그 밖의 데이터”로 정의하고 있다. 데이터세트는 일반적인 문서 유형의 전자기록물과 달리 데이터베이스를 기반으로 생산·관리된다는 특징을 가진다. 현재 대부분의 행정정보시스템은 관계형 데이터베이스를 기반으로 데이터를 저장·관리하고 있으며, 최근에는 빅데이터와 비정형 데이터의 활용이 증가함에 따라 비관계형 데이터베이스(NoSQL)를 적용하는 사례도 나타나고 있다. 그러나 데이터베이스의 구현 방식과 관계없이 데이터세트는 데이터뿐 아니라 데이터와 연계된 첨부파일, 데이터베이스 구조, 이를 처리하는 응용프로그램 등이 함께 구성되는 복합적인 기록 객체이며(김지혜 외, 2022), 생산 당시의 행정정보시스템 환경에 기술적으로 종속된다는 공통적인 특성을 가진다.

데이터세트는 「행정정보 데이터세트 기록관리 기준」(국가기록원, 2020, NAK 35:2020, v1.0)에 따라 관리된다. 해당 표준은 먼저 관리 대상 행정정보시스템을 선정할 후 관리기준표를 작성하도록 규정하고 있으며, 관리기준표를 중심으로 데이터세트를 관리하도록 하고 있다. 관리기준표에는 행정정보시스템과 데이터세트의 기본 정보, 단위기능, 보존기간, 관리부서 등 데이터세트 기록관리에 필요한 정보를 작성하도록 규정하고 있다. 또한 단위기능은 행정정보시스템 전체를 대상으로 설정할 수도 있으며, 필요한 경우 시스템 내 업무기능을 기준으로 설정할 수도 있다. 관리기준표는 데이터세트 기록관리의 기준으로 활용되며, 관리 대상 행정정보시스템과 데이터세트의

현황을 지속적으로 관리하도록 하고 있다. 관리기준표를 통해 데이터세트의 관리 범위와 관리 단위를 일관성 있게 유지하며, 시스템 운영 과정에서 발생하는 변경 사항도 지속적으로 반영하도록 규정하고 있다. 이와 같이 현재 데이터세트는 관리 대상 행정정보시스템의 선정과 관리기준표를 중심으로 관리되는 기록관리 체계를 갖추고 있다.

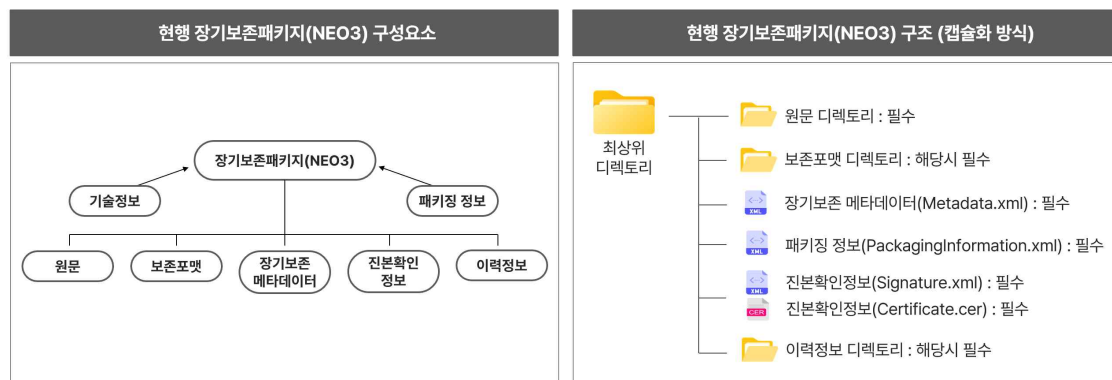
3. 현행 장기보존패키지(NEO3) 기술규격 분석

3.1 장기보존패키지(NEO3)의 구성요소 및 구조

현행 장기보존패키지(NEO3)의 구성요소는 <표 1>과 같이 원문, 보존포맷, 장기보존 메타데이터, 패키징 정보, 기술정보, 진본확인 정보, 이력정보로 구성되며, 각 구성요소는 <그림 1>과 같이 최상위 디렉토리 아래에 디렉토리 또는 파일 형태로 계층적으로 배치된다. 장기보존 메타데이터, 패키징 정보 및 진본확인 정보는 최상위 디렉토리에 파일 형태로 저장되며, 원문, 보존포맷 및 이력정보는 각각의 하위 디렉토리에 해당 파일을 저장하는 방식으로 구성된다.

<표 1> 장기보존패키지(NEO3)의 개념적 구성요소

구분	내용	필수 여부
원문	생산자가 생산(접수)한 기록물 원본 파일	필수
보존포맷	원문의 의미와 구조 등 기록의 특성을 유지하면서 기술환경 변화에도 원문의 내용과 외형을 재현하거나 기술적으로 복원하여 지속적인 접근과 활용을 가능하게 하는 포맷	해당시 필수
장기보존 메타데이터	기록물의 생산부터 관리·보존에 이르는 전 과정을 기술한 정보로, 기록물의 진본성, 신뢰성, 무결성 및 이용가능성을 유지하고 관리·보존을 지원하는 정보	필수
패키징 정보	장기보존패키지의 구성요소를 실제적 또는 논리적으로 묶고, 식별하며, 상호 연결하기 위한 정보	필수
기술정보	이용자 또는 시스템이 장기보존패키지를 검색, 식별 및 추출하는 데 활용하기 위한 정보	선택
진본확인 정보	전자서명, 인증서, 해시값 등을 이용하여 장기보존패키지와 그 구성요소의 진본성 및 무결성을 검증하기 위한 정보	필수
이력정보	장기보존패키지의 변경 이력과 버전 정보를 관리하고, 이전 상태의 추적 및 진본성·무결성 검증을 지원하기 위한 정보	해당시 필수



<그림 1> 장기보존패키지(NEO3) 구성요소 및 구조

원문은 업무 수행 과정에서 생산하거나 접수한 기록의 원본(진본)에 해당하는 구성요소이다. 전자기록물은 기록관리 과정에서 이관, 보존포맷 변환 등 다양한 관리행위를 거치면서 형태가 변환될 수 있으며, 복제가 용이하고 위·변조가 가능하다. 장기보존 과정에서는 기록 생산 당시의 내용이 변경되지 않았음을, 혹은 적법한 절차에 따라 변경되었음을 지속적으로 검증하여 기록의 무결성을 확보하고 진본성을 보장할 필요가 있다. 이에 기록 생산 당시의 원본을 보존하기 위하여 최상위 디렉토리 하위에 원문에 대한 디렉토리를 생성하고 원문에 해당하는 전자기록물을 배치하여 관리한다. 다만 표준에는 기록 유형별로 원문의 구성 범위를 어디까지 설정해야 하는지에 대한 구체적인 기준은 제시하지 않는다. 보존포맷은 하드웨어·소프트웨어의 노후화, 저장매체의 퇴화, 데이터 포맷의 변화 또는 원본의 손상 등으로 원문에 지속적으로 접근하기 어려워지는 상황에 대비하여 전자기록물을 장기간 접근·활용할 수 있도록 하는 구성요소이다. 현행 표준에서는 원문의 의미와 구조 등 기록의 특성을 유지하면서, 생산 당시의 애플리케이션이 없어도 원문의 내용과 외형을 재현하여 내용을 확인하거나 원문의 내용과 맥락정보 등 기록 유형별 중요 정보를 표준화되고 공개된 방식으로 기술하여 복원할 수 있는 형태로 변환하도록 규정하고 있다. 다만 원문을 변환할 적합한 보존포맷이 지정되지 않았거나 에뮬레이션 등 다른 보존방식으로 재현할 수 있는 경우에는 보존포맷을 포함하지 않을 수 있다. 생성된 보존포맷은 장기보존패키지의 최상위 디렉토리 하위에 보존포맷 관련 디렉토리를 생성하여 배치한다.

장기보존 메타데이터는 전자기록물의 생산부터 관리·보존에 이르는 전 과정을 기술하여 기록물이 적법한 절차에 따라 관리되어 왔음을 입증하고 진본성과 무결성을 유지하기 위한 구성요소이다. 현행 표준에서는 패키지명과 버전 등 장기보존패키지 자체에 관한 정보, 기록물의 관리 전 과정에 관한 기록관리 메타데이터, 진본성 검증 정보, 장기보존패키지가 재변환된 경우 재변환 전후의 관계에 관한 정보 등을 장기보존 메타데이터로 관리하도록 하고 있다. 기록관리 메타데이터는 「기록관리 메타데이터 표준」(NAK 8)을 참조하며, 다양한 기록물 유형의 특성에 따른 정보를 수용할 수 있도록 기록물체와 기록물건 메타데이터 하위에 확장 가능한 상위 요소인 ExtraMetadata를 두고 있다. 장기보존 메타데이터는 Metadata.xml 파일로 생성하여 장기보존패키지의 최상위 디렉토리에 배치한다. 패키징 정보는 장기보존패키지에 포함된 정보 구성요소를 하나로 묶고 목록화하기 위한 정보이다. 원문과 보존포맷으로 구성되는 콘텐츠 정보와 장기보존 메타데이터를 논리적으로 묶고, 장기보존패키지에 추가로 포함되는 부속파일에 관한 정보도 함께 목록화한다. 또한 장기보존패키지가 수정된 경우에는 이력정보에 대한 경로나 식별자를 포함하며, 이는 패키지 내부 파일의 무결성 확인 및 진본확인을 위한 정보로 활용된다. 패키징 정보는 PackagingInformation.xml 파일로 생성하여 장기보존패키지의 최상위 디렉토리에 배치한다. 기술정보는 장기보존패키지에 보존된 전자기록물에 대한 검색 및 목록 제공을 지원하기 위한 메타데이터이다. 이는 장기보존패키지 자체의 필수 구성요소가 아니라 기록물관리시스템 등에서 장기보존패키지의 검색·추출 기능을 지원하기 위해 선택적으로 생성·관리할 수 있는 정보이다. 기술정보는 새로운 메타데이터 요소를 별도로 정의하는 것이 아니라, 장기보존 메타데이터에 포함된 요소 가운데 검색과 목록 제공에 필요한 항목을 선별하여 구성한다.

진본확인 정보는 장기보존패키지에 포함된 기록물이 특정 시점 이후 변경되지 않았음을 검증하기 위한 구성요소로, 기록물의 진본성과 무결성 유지를 목적으로 한다. 기술규격(NAK 31-2)에서는 진본확인 기술로 전자서명을 채택하고 있으며, 이는 공개키 기반구조(Public Key Infrastructure, PKI)를 기반으로 X.509 전자서명용 인증서를 사용한다. 생성된 전자서명 정보와 인증서는 장기보존패키지에 함께 보존되어, 서명 주체를 확인하고 서명 시점의 진본성과 무결성을 검증하는 근거가 된다. 전자서명의 대상은 장기보존 메타데이터 파일이며, 원문·보존포맷 등 콘텐츠 정보의 해시값이 장기보존 메타데이터 또는 패키징 정보에 포함되어 있어 장기보존패키지 전체에 대한 연쇄적인 무결성 검증이 가능하다.

이력정보는 장기보존패키지의 변경 이력을 관리하기 위한 구성요소이다. 장기보존패키지는 생성 이후에도 관리이력의 추가, 바이러스 검사 및 오류 검증에 따른 메타데이터 수정, 전자서명 생성, 새로운 보존포맷의 추가 등의 사유로 구성요소가 변경될 수 있다. 이 경우 변경된 최신 파일은 기존 위치에 반영하고, 수정 전 파일은 별도의 버전을 부여하여 Change History 디렉토리에 함께 보존한다. 이를 통해 변경 내용을 버전별로 관리함으로써

씩 버전 간 비교와 각 버전에 대한 진본확인 기술의 적용을 효율적으로 수행할 수 있다.

3.2 데이터세트의 보존 범위

장기보존패키지의 기본 원칙 중 하나인 자체 충족성은 전자기록물이 시스템이나 외부 데이터에 의존하지 않고 생산 당시의 내용을 재현하여 읽고 이해할 수 있어야 한다는 원칙이다. 데이터세트는 행정정보시스템을 통해 동적으로 구현되는 기록이므로 시스템에서 분리될 경우 화면 구성, 데이터의 배치와 표현 방식 등 생산 당시의 이용 환경을 그대로 재현하기 어렵다. 즉, 기록의 룩앤필(Look & Feel)을 유지하는 데 한계가 있다. 따라서 데이터 세트를 장기보존하기 위해서는 데이터 자체뿐만 아니라 그 내용과 구조를 이해하고 재현하는 데 필요한 정보도 함께 보존할 필요가 있다. 본 연구에서는 이러한 특성을 고려하여 <표 2>와 같이 데이터세트의 보존 범위를 설정하였다. 데이터, 첨부파일, DB 구조 정보 및 재현정보는 데이터세트 기록의 원문을 구성하는 요소로 보존 범위에 포함하였다. 주요 산출물은 데이터세트 기록을 직접 구성하는 요소는 아니지만 데이터세트가 생산된 행정정보시스템의 구조와 운영을 이해하는 데 활용할 수 있는 참고 자료로서 보존 범위에 포함하였다.

<표 2> 장기보존패키지 적용을 위한 데이터세트 보존 범위

구분	보존 범위	내용	필수 여부
원문	데이터	행정정보시스템의 데이터베이스에 저장된 값	필수
	첨부파일	생산자가 데이터를 입력하거나 등록하는 과정에서 함께 첨부한 파일	해당시 필수
	DB 구조 정보	데이터세트의 논리적 구조를 구성하는 요소	해당시 필수
	재현정보	생산 당시의 사용자 인터페이스와 실행환경을 재현하기 위한 정보	해당시 필수
참고자료	주요산출물	시스템의 구축·개발 및 운영 과정에서 생산된 문서	선택

데이터는 행정정보시스템의 데이터베이스에 저장된 값이다. 데이터를 보존하지 않으면 데이터세트의 내용을 확인하거나 재현할 수 없으므로 장기보존패키지에 필수적으로 포함하여야 한다. 첨부파일은 생산자가 데이터를 입력 및 등록하는 과정에서 함께 첨부한 파일이다. 시스템의 환경에 따라 첨부파일은 BLOB(Binary Large Object) 또는 CLOB(Character Large Object) 형태로 데이터베이스에 저장되거나, 외부 저장소에 저장된 후 데이터베이스에는 첨부파일의 저장경로나 식별정보만 저장될 수 있다. 저장 방식과 관계없이 첨부파일은 데이터와 함께 데이터 세트의 내용을 구성하는 요소이므로 필수적으로 보존하여야 한다.

DB 구조 정보는 데이터베이스의 스키마, 테이블과 컬럼, 테이블 간 관계, 제약조건 등 데이터가 조직된 구조와 트리거, 프로시저, 뷰 등 데이터베이스의 처리·동작에 관한 정보를 의미한다. 데이터세트를 데이터베이스로부터 분리하여 보존할 경우 데이터 값만으로는 데이터가 어떠한 구조로 조직되어 있으며 서로 어떠한 관계를 가지는지 파악하기 어려울 수 있다. 특히 테이블과 컬럼, 키와 관계 등은 데이터세트의 논리적 구조를 구성하므로, 데이터의 내용과 관계를 지속적으로 이해하기 위해 함께 보존할 필요가 있다. 따라서 DB 구조 정보는 데이터의 구조적 의미와 재현에 미치는 영향을 고려하여 필요한 요소를 보존 범위에 포함할 수 있다.

재현정보는 데이터세트의 내용을 이해하고 이용하거나 필요한 경우 생산 당시의 표현 방식과 이용 환경을 재현하는 데 필요한 정보를 의미한다. OAIS 참조모형의 Representation Information은 데이터 객체를 의미 있는 정보로 해석하고 이해하는 데 필요한 정보를 포괄하는 개념으로, 데이터의 구조와 의미에 관한 정보뿐만 아니라 소프트웨어나 알고리즘 등도 이에 해당한다. 데이터세트는 데이터베이스에 구조화된 형태로 저장되어 관리되지만, 이용자는 이를 직접 이용하기보다 행정정보시스템을 통해 제공되는 결과를 확인하고 활용한다. 행정정보시스템은 업무 기능에 따라 저장된 데이터를 조회·결합·가공하여 특정 화면과 형식으로 제공한다. 그러나 데이터와 DB 구조

정보만으로 이러한 처리 방식과 이용 환경을 충분히 파악하기 어려운 경우가 있다. 특히 데이터가 특정 기능이나 화면 구동을 전제로 생산되었거나, 기능적 맥락 없이는 데이터의 의미를 충분히 이해하기 어려운 경우에는 재현정보가 필요하다. 재현정보의 예로는 HTML, CSS, 화면 구동에 필요한 프로그램, 화면 캡처 등이 있다. HTML, CSS 및 프로그램은 데이터세트의 표현이나 실행을 지원하며, 화면 캡처는 기능 자체를 재현하지는 못하지만 생산 당시의 화면 구성과 데이터의 표현 및 이용 방식을 확인하는 데 활용될 수 있다.

주요 산출물은 행정정보시스템의 구축·개발 및 운영 과정에서 생산된 문서로서 제안요청서, 설계서, 시스템 구성도, ERD, 테이블·컬럼 명세서, 사용자 및 운영자 매뉴얼 등을 의미한다. 주요 산출물은 데이터세트 자체를 구성하는 요소가 아니므로 원문의 범위에 포함되는 필수 보존 대상은 아니다. 그러나 데이터세트가 어떠한 업무 목적과 시스템 환경에서 생산·관리되었는지, 데이터세트를 지원하는 구조와 기능이 어떠한 방식으로 구현되었는지를 이해하는 데 필요한 참고자료가 될 수 있다. 이에 주요 산출물은 데이터세트의 생산 맥락과 시스템 환경을 이해하는 데 유용한 자료를 중심으로 보존 범위에 선택적으로 포함할 수 있다.

3.3 데이터세트 보존 범위에 따른 장기보존패키지 구성요소 분석

본 절에서는 현행 장기보존패키지의 구성요소가 데이터세트의 보존 범위를 포괄할 수 있는지 검토하였다. 그 결과, <표 3>과 같이 현행 기술규격을 그대로 적용할 수 있는 요소, 데이터세트의 특성을 반영하여 재설계가 필요한 요소, 현행 기술규격에 대응하는 구성요소가 없어 새롭게 추가하여야 하는 요소로 구분하였다.

<표 3> 현행 장기보존패키지(NEO3) 구성요소의 데이터세트 적용 가능성 분석

구분	구성요소	내용
적용 가능 요소	보존포맷	원문을 보존에 적합한 파일포맷으로 변환한다는 개념은 기록 유형과 무관하므로, 데이터세트 유형에도 그대로 적용 가능
	패키징 정보	구성요소를 식별자로 연결하여 하나의 패키지로 묶는 개념은 기록 유형과 무관하므로, 데이터세트 유형에도 그대로 적용 가능
	진본확인 정보	전자서명을 기록물이 아닌 장기보존 메타데이터 파일을 대상으로 적용하므로, 서명·검증 방식이 기록 유형과 무관하게 적용 가능
재설계 요소	원문	기록 원문을 보존한다는 원문의 개념은 그대로 적용되나, 데이터세트의 원본 범위(데이터, 첨부파일, DB 구조 정보, 재현정보 등)가 현행 표준에 규정되어 있지 않아 범위 설정이 필요함
	장기보존 메타데이터	기록물철, 기록물건 구조를 전제로 항목이 정의되어 데이터세트에 그대로 적용하기 어려우므로 항목 재설계 필요
	기술정보	표준 규격에는 포함 가능성만 규정될 뿐 어떤 요소로 구성하는지가 정의되어 있지 않아, 기록 유형별 특성을 반영하되 외부 접근에 필요한 최소 요소 제시 필요
	이력정보	변경 시 수정 전 파일을 버전별로 별도 보존하도록 규정되어 있으나, 이 방식은 대량·대용량 데이터세트에 적용할 경우 저장 부담이 커, 변경분 중심의 효율적 관리 방식으로 보완 필요
신규 요소	재현정보	재현정보는 원문의 범위에 포함되지만 다른 원문 요소와 보존 목적과 역할이 구분되므로 원문 디렉토리 내에서 이를 별도로 관리할 수 있는 디렉토리 신설 필요
	주요 산출물	데이터세트의 구조·맥락 이해를 보조하는 시스템 산출물을 담을 구성요소가 현행 규격에 없어 신설 필요

먼저 현행 정의를 그대로 적용할 수 있는 요소는 보존포맷, 패키징 정보, 진본확인 정보이다. 이들은 기록물을 보존에 적합한 형태로 변환하거나, 구성요소를 하나의 패키지로 조직하고 진본성과 무결성을 검증하기 위한 기능을 수행하므로, 기록물의 내용이나 유형과 관계없이 동일하게 적용할 수 있다. 보존포맷은 원문을 보존에 적합한 파일포맷으로 변환한다는 개념이 기록 유형과 무관하다. 더욱이 「전자기록물 보존포맷 선정기준(NAK 37:2025)」

이 개정을 통해 데이터세트 유형에 대한 선정기준을 추가하였으므로, 별도의 기준을 새로 마련할 필요 없이 해당 기준에 따라 데이터세트에 적합한 보존포맷을 선정하여 적용할 수 있다. 패키징 정보는 버전 정보, 파일 목록, 개별 파일의 경로 및 해시값 등을 기록하여 구성요소를 목록화하고 식별자로 연결하는 기능을 수행한다. 이는 연결 대상이 데이터세트인지 전자문서인지와 무관하게 동일하게 적용된다. 진본확인 정보 역시 기록 유형과 무관하게 동일하게 적용할 수 있다. 현행 표준은 전자서명을 기록물 원문이 아니라 장기보존 메타데이터 파일에 적용하고, 원문과 보존포맷 등의 해시값을 장기보존 메타데이터에 포함하여 함께 검증하는 방식을 사용한다. 따라서 전자서명의 적용 대상이 장기보존 메타데이터 파일로 한정되어 있으므로, 패키지에 포함된 기록물의 유형과 관계 없이 동일한 방식으로 적용할 수 있다.

개념은 유지되나 사실상 · 보완이 필요한 요소는 원문, 장기보존 메타데이터, 기술정보, 이력정보이다. 이들은 개념 자체는 데이터세트에 유효하나, 현행 정의나 관리 방식을 그대로 적용하기 어려워 데이터세트의 특성에 맞게 조정할 필요가 있다. 원문은 생산된 기록 원본을 보존한다는 개념 자체는 데이터세트에도 동일하게 적용된다. 그러나 전자문서는 원본이 파일 형태로 고정되어 있어 원문의 범위가 비교적 명확한 반면, 데이터세트는 원본이 데이터베이스에 존재하므로 무엇을 원문으로 볼 것인지가 자명하지 않다. 현행 표준은 이에 대한 범위를 별도로 규정하고 있지 않으므로, 데이터 · 첨부파일 · DB 구조 정보 · 재현정보 가운데 어디까지를 원문에 포함할 것인지에 대한 기준을 새롭게 설정할 필요가 있다. 특히 데이터세트는 데이터 자체만으로는 생산 당시의 의미와 이용 형태를 충분히 이해하거나 재현하기 어려우므로, 재현정보 역시 원문의 범위에 포함하여 관리할 필요가 있다. 즉, 원문은 기존 개념을 변경하는 것이 아니라 데이터세트의 특성에 맞게 원문의 범위를 구체화하는 보완이 요구된다. 장기보존 메타데이터는 보존 대상 기록물을 관리하고 설명하기 위한 정보로, 데이터세트 유형 전자기록물의 장기보존에도 필수적으로 요구된다. 그러나 현행 장기보존 메타데이터는 「기록관리 메타데이터 표준(NAK 8)」을 참조하여 기록물철과 기록물건을 기본 관리 단위로 구성되고, 세부 항목 역시 전자문서의 기록관리를 전제로 정의되어 있다. 반면 데이터세트는 「행정정보 데이터세트 기록관리 기준(NAK 35)」에 따라 행정정보시스템 또는 단위기능을 관리 단위로 하므로 철 · 건 구조를 그대로 적용하기 어렵다. 또한 표준은 다양한 기록물 유형에 대응하기 위해 확장 요소(ExtraMetadata)를 제공하지만, 이는 기존 철 · 건 메타데이터 하위에 항목을 추가하는 방식이므로 관리 단위 자체가 다른 데이터세트를 수용하기에는 한계가 있다. 따라서 데이터세트의 관리 단위와 특성을 반영할 수 있도록 메타데이터의 계층 구조와 항목을 새롭게 설계할 필요가 있다. 기술정보는 이용자가 장기보존패키지를 검색 · 추출할 때 활용하는 접근점에 해당한다. 현행 기술규격에서 기술정보는 장기보존패키지의 필수 구성요소가 아니라 기록물관리시스템 등에서 검색과 활용을 지원하기 위해 선택적으로 생성 · 관리하는 정보이며, 장기보존 메타데이터에 포함된 요소 가운데 검색 및 목록 제공에 필요한 항목을 활용하여 구성한다. 따라서 기술정보의 개념은 데이터세트에도 그대로 적용할 수 있다. 다만 검색 · 추출에 필요한 접근점은 기록 유형에 따라 달라질 수 있으므로, 데이터세트의 특성을 반영한 검색 요소를 장기보존 메타데이터로부터 선별하여 구성할 필요가 있다. 이력정보는 장기보존패키지가 수정될 때 변경 전후의 이력을 보존하기 위한 요소이다. 현행 표준은 수정이 발생하면 최신 정보는 현재 파일에 반영하고 수정 전 파일은 버전을 부여하여 별도 디렉토리에 보존하며, 그 전후 관계는 패키징 정보의 생성 시점 정보로 확인하도록 규정한다. 다만 버전마다 파일 전체를 보존하는 방식은, 대량 · 대용량으로 구성되는 데이터세트에 적용할 경우 변경되지 않은 부분까지 반복 보존하게 되어 저장 공간을 과도하게 점유한다. 데이터세트에는 저장 부담을 줄이면서도 변경 이력을 추적할 수 있는 관리 방식이 별도로 요구된다.

마지막으로 현행 규격에 대응하는 구성요소가 없어 새롭게 추가해야 하는 요소는 재현정보와 주요 산출물이다. 두 요소 모두 데이터세트의 보존 범위에는 포함되지만, 현행 장기보존패키지에는 이를 담을 구성요소가 존재하지 않는다. 재현정보는 데이터가 생산 당시 어떠한 화면과 기능을 통해 이용자에게 제공되었는지를 보여주는 정보로, 데이터세트의 내용을 이해하고 재현하기 위해 필요한 원문의 일부에 해당한다. 그러나 HTML, CSS, 화면 캡처 이미지, 구동 애플리케이션 등 성격이 서로 다른 자료를 데이터 · 첨부파일과 함께 하나의 원문 디렉토리에서

관리할 경우 식별과 관리가 어렵다. 따라서 원문의 범위에는 포함하되, 관리의 명확성과 활용성을 높이기 위해 별도의 디렉토리로 구분하여 관리할 필요가 있다. 주요 산출물은 ERD, 테이블 명세서, 운영 매뉴얼 등 시스템의 구조와 기능을 설명하는 문서이다. 이는 데이터세트 기록 자체를 구성하는 정보는 아니지만, DB 구조 정보와 재현정보를 해석하고 시스템의 구축 배경과 업무 목적을 이해하는 데 중요한 참고자료가 된다. 현행 장기보존패키지에는 이러한 참고자료를 관리하기 위한 구성요소가 마련되어 있지 않으므로, 해당 자료가 존재하는 경우 선택적으로 포함할 수 있는 별도의 구성요소를 마련할 필요가 있다.

4. 데이터세트 적용을 위한 장기보존패키지(NEO3) 개선(안)

본 장에서는 데이터세트의 특성을 반영한 장기보존패키지의 확장 방안을 제시한다. 3장의 분석 결과를 바탕으로 도출한 설계 방향과 재설계 결과는 <표 4>와 같다. 이를 토대로 장기보존패키지의 구조와 구성요소, 장기보존 메타데이터, 생성 및 변경이력 관리방안을 설명하며, 시범 구현을 통해 제안한 장기보존패키지의 실현 가능성을 검토한다.

<표 4> 데이터세트 수용을 위한 장기보존패키지(NEO3) 재설계 개요

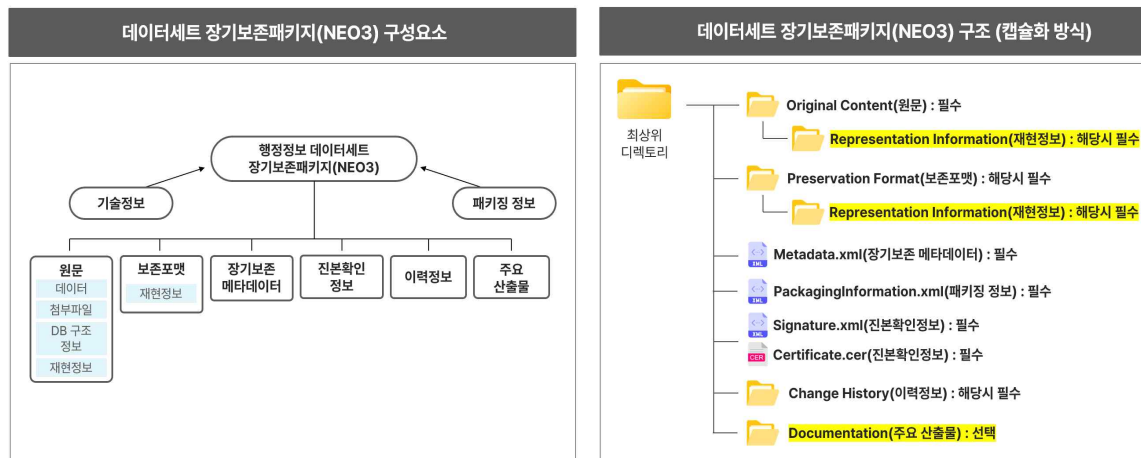
구분	기존 한계	재설계 방향	재설계 결과
구조 및 구성요소	장기보존 시 고려되어야 하는 데이터 세트의 보존 범위를 기존 장기보존패키지(NEO3)의 구성요소와 구조만으로는 충분히 수용하기 어려움	기존 장기보존패키지(NEO3) 체계와의 호환성을 유지하면서 데이터세트의 보존 범위를 수용할 수 있도록 구성요소를 확장	기존 구조에 재현정보 및 주요 산출물 디렉토리 추가
장기보존 메타데이터	기록물철·건 계층으로 구분된 메타데이터 구조로 요소가 정의되어 있어 데이터세트에 적용 어려움	기존 메타데이터 체계와의 호환성을 유지하면서 데이터세트의 관리단위와 특성을 반영한 메타데이터 구조 및 요소 설계	차상위요소, 하위요소, 최하위요소, 세부요소 등 4계층으로 구성된 데이터세트 장기보존 메타데이터 설계
변경이력 관리	전체 패키지 단위의 버전 관리는 대용량 데이터세트에서 저장공간 부담과 관리 비효율을 초래함	대용량 데이터세트의 특성을 고려하여 변경분 중심의 버전 관리 체계를 적용하고, 데이터 무결성과 복원 가능성을 확보할 수 있는 변경이력 관리방안 마련	최신본 유지, 변경분 중심 버전 관리, 버전별 진본확인정보·메타데이터·패키징 정보 관리, 역방향 병합 기반 복원 방안 제시

4.1 데이터세트 유형 장기보존패키지 구성

4.1.1 구조 및 구성요소

본 연구에서 제안하는 데이터세트 유형 장기보존패키지는 <그림 2>와 같이 현행 장기보존패키지(NEO3)의 기본 구조를 유지하면서 데이터세트의 장기보존에 필요한 디렉토리를 신설하는 방향으로 설계하였다. 차상위 디렉토리 바로 하위에 배치되는 디렉토리와 파일의 명칭 및 배치를 규정하였으며, 각 디렉토리 내부의 세부 구조는 별도로 규정하지 않아 필요에 따라 유연하게 구성할 수 있도록 하였다. 다만 보존포맷은 국가기록원에서 정한 데이터세트 보존포맷 규격에 따라 구성한다.

추가한 디렉토리는 Representation Information(재현정보) 디렉토리와 Documentation(주요 산출물) 디렉토리이다. 데이터세트 원문에는 데이터, 첨부파일, DB 구조 정보와 필요한 경우 재현정보가 포함되며, 이들은 Original Content 디렉토리에서 관리한다. 재현정보는 생산 당시와 동일하거나 유사한 방식으로 이용·재현하는 데 필요한 기술정



〈그림 2〉 데이터세트 장기보존패키지(NEO3) 구조 및 구성요소

보로, 내용을 구성하는 다른 원문 요소와는 보존 목적과 역할이 구분된다. 이에 재현정보를 함께 보존하는 경우에는 Original Content 디렉토리 하위에 Representation Information 디렉토리를 두어 별도로 관리하도록 하였다. 또한 보존포맷에 재현정보가 필요한 경우에도 Preservation Format 디렉토리 하위에 Representation Information 디렉토리를 두어 관리할 수 있도록 하였다. 재현정보는 원문과 보존포맷에 일률적으로 포함하지 않고, 각각 필요한 경우 유연하게 구성할 수 있다. 시스템의 주요 산출물은 데이터세트의 보존 범위에는 포함되지만 데이터세트 기록 자체를 구성하는 요소는 아니므로 필수는 아니다. 다만 데이터세트의 생성·구조·운영 환경을 이해하는데 활용할 수 있는 참고자료이므로, 필요한 경우 별도의 Documentation 디렉토리를 두어 주요 산출물을 포함하도록 하였다.

현행 장기보존패키지(NEO3)의 구성요소 가운데 장기보존 메타데이터와 이력정보는 기본 개념과 역할을 유지하면서 데이터세트에 적용할 수 있도록 개선하였다. 장기보존 메타데이터는 데이터세트에 필요한 정보를 기술할 수 있도록 메타데이터 계층과 요소를 새롭게 설계하였으며, 세부 내용은 4.2절에서 설명한다. 이력정보는 장기보존패키지 생성 이후 발생하는 변경 사항을 관리하고 특정 시점의 패키지 상태를 복원할 수 있도록 관리방안을 마련하였다. 특히 대용량 데이터세트의 특성을 고려하여 변경된 구성요소를 중심으로 이력을 관리하는 방식을 제안하였으며, 세부 내용은 4.3절에서 설명한다.

4.1.2 장기보존패키지 생성 시나리오

데이터세트 유형 장기보존패키지의 구조와 구성요소를 실제 장기보존패키지로 구성하는 절차는 <표 5>와 같다. 이는 각 구성요소가 어떠한 순서로 생성되고 장기보존패키지에 배치되는지를 개념적으로 나타낸 것이다.

첫 번째 단계에서는 장기보존 대상 데이터세트를 식별하고 보존 범위와 추출 기준 시점을 확인한 후, 장기보존 패키지 구성에 필요한 자료를 수집한다. 먼저 원문을 구성하기 위해 데이터베이스에서 보존 대상 데이터를 추출한다. 데이터는 데이터베이스의 특성과 보존 방식 등을 고려하여 SQL dump, SIARD 등 적절한 형식으로 추출할 수 있다. 첨부파일은 데이터베이스 내부 또는 외부 저장소 등 저장 위치와 관리 방식에 따라 적절한 방법으로 확보한다. 재현정보는 원문의 내용을 이해하고 이용하는 데 필요한 경우 함께 수집한다. 원문의 형식이 장기보존에 적합한 경우에는 별도의 보존포맷을 생성하지 않으며, 보존포맷이 필요한 경우에는 국가기록원에서 정한 데이터세트 보존포맷 규격에 따라 생성한다. 보존포맷은 추출한 원문을 변환하여 생성하거나, 데이터베이스에서 필요

〈표 5〉 데이터세트 유형 장기보존패키지 생성 시나리오

구분	내용
1단계: 데이터세트 식별 및 수집	① (필수) 보존 대상 데이터세트의 범위와 추출 기준 시점 확정 ② (필수) 원문을 형식에 맞게 추출 및 수집(데이터, 첨부파일, DB 구조 정보, 재현정보 등) ③ (해당시 필수: 원문의 파일포맷이 장기보존에 적합하지 않은 경우) 국가기록원에서 정한 데이터세트 보존포맷 규격에 따라 보존포맷 생성 ④ (선택) 행정정보시스템에 관한 산출물 수집 ⑤ (필수) 데이터세트 관리기준표, 정보자원관리시스템(IRM) 등 장기보존 메타데이터 작성을 위한 정보 수집
2단계: 디렉토리 배치	① (필수) Original Content 디렉토리를 생성하여 원문에 해당하는 데이터, 첨부파일, DB 구조 정보 등 배치, 필요시 하위에 Representation Information 디렉토리를 생성하여 재현정보 배치 ② (해당시 필수: 원문을 보존포맷으로 변환한 경우) Preservation Format 디렉토리를 생성하여 보존포맷을 배치하고, 필요시 하위에 Representation Information 디렉토리를 생성하여 재현정보 배치 ③ (해당시 필수: 장기보존패키지 생성 후, 변경이 발생한 경우) Change History 디렉토리를 생성하여 향후 변경이력 정보 배치 ④ (선택) Documentation 디렉토리를 생성하여 행정정보시스템의 주요 산출물 배치
3단계: 장기보존 메타데이터 작성	① (필수) 신규 정의한 데이터세트 유형 장기보존 메타데이터에 따른 Metadata.xml 생성 ② (필수) 행정정보시스템, 데이터세트 등에 관한 메타데이터 기술 ③ (필수) Original Content 디렉토리, Preservation Format 디렉토리, Representation Information 디렉토리, Documentation 디렉토리 내 모든 파일의 경로값, 해시값 등 기술 ④ (선택) 장기보존 메타데이터 요소 중 DescriptiveMetadata 요소 설정
4단계: 전자서명	① (필수) Metadata.xml을 대상으로 GPKI 방식의 행정전자서명을 수행하여 행정전자서명 값 생성 ② (필수) 행정전자서명 인증서로 Certificate.cer 파일 생성하여 최상위 디렉토리에 배치 ③ (필수) 행정전자서명 값, 공개키, 유효기간 등을 담은 Signature.xml 파일을 생성하여 최상위 디렉토리에 배치
5단계: 패키징	① (필수) 장기보존패키지 전체 구성요소의 생성일, 식별자, 해시값, 경로정보 등을 목록화한 PackagingInformation.xml 파일 생성 및 배치 ② ZIP으로 패키징 수행

한 데이터를 별도로 추출하여 생성할 수 있다. 행정정보시스템의 주요 산출물은 보존 필요성이 인정되는 경우 함께 수집하며, 이는 시스템 관리자 또는 정보자원관리시스템(IRM) 등을 통해 확보할 수 있다. 마지막으로 장기보존 메타데이터 작성에 필요한 정보를 수집한다. 이때 데이터세트 관리기준표를 우선 활용하며, 행정정보시스템, 주요 산출물 및 정보자원관리시스템(IRM) 등의 정보를 통해 보완한다.

두 번째 단계에서는 장기보존패키지의 구조에 따라 각 구성요소를 해당 디렉토리에 배치한다. 원문은 Original Content 디렉토리에 위치시킨다. 원문에 대한 재현정보를 함께 보존하는 경우에는 그 하위에 Representation Information 디렉토리를 생성하여 배치한다. 보존포맷을 생성한 경우에는 Preservation Format 디렉토리를 생성하여 보존포맷을 배치하고, 필요에 따라 하위에 Representation Information 디렉토리를 생성하여 보존포맷에 대한 재현정보를 포함한다. 장기보존패키지 생성 이후 변경이 발생하는 경우를 대비하여 Change History 디렉토리를 생성하며, 필요에 따라 Documentation 디렉토리를 생성하여 행정정보시스템의 주요 산출물을 배치한다.

세 번째 단계에서는 데이터세트 유형에 맞게 설계한 장기보존 메타데이터를 작성한다. 먼저 Metadata.xml을 생성한 후 행정정보시스템과 데이터세트에 관한 정보, 장기보존패키지를 구성하는 각 파일의 경로, 식별자 및 해시값 등 메타데이터 요소별 값을 입력한다. 또한 기술정보(Descriptive Information)는 장기보존패키지의 검색·조회 및 활용을 지원하기 위해 패키지 외부에서 활용되는 정보이므로, 장기보존 메타데이터 중 기술정보로 활용할 요소도 함께 설정한다.

네 번째 단계에서는 행정전자서명을 적용하여 장기보존패키지의 무결성을 검증할 수 있도록 한다. 앞선 3단계에서 작성한 Metadata.xml에는 장기보존패키지를 구성하는 파일의 경로와 해시값이 기록되며, 이를 기준으로 패

키지의 변경 여부를 확인할 수 있다. Metadata.xml을 대상으로 GPKI 방식의 행정전자서명을 수행하고, 생성된 전자서명값은 공개키, 전자서명 유효기간 등의 정보와 함께 Signature.xml에 기록한다. 인증서는 Certificate.cer 파일로 저장하여 장기보존패키지의 최상위 디렉토리에 배치한다.

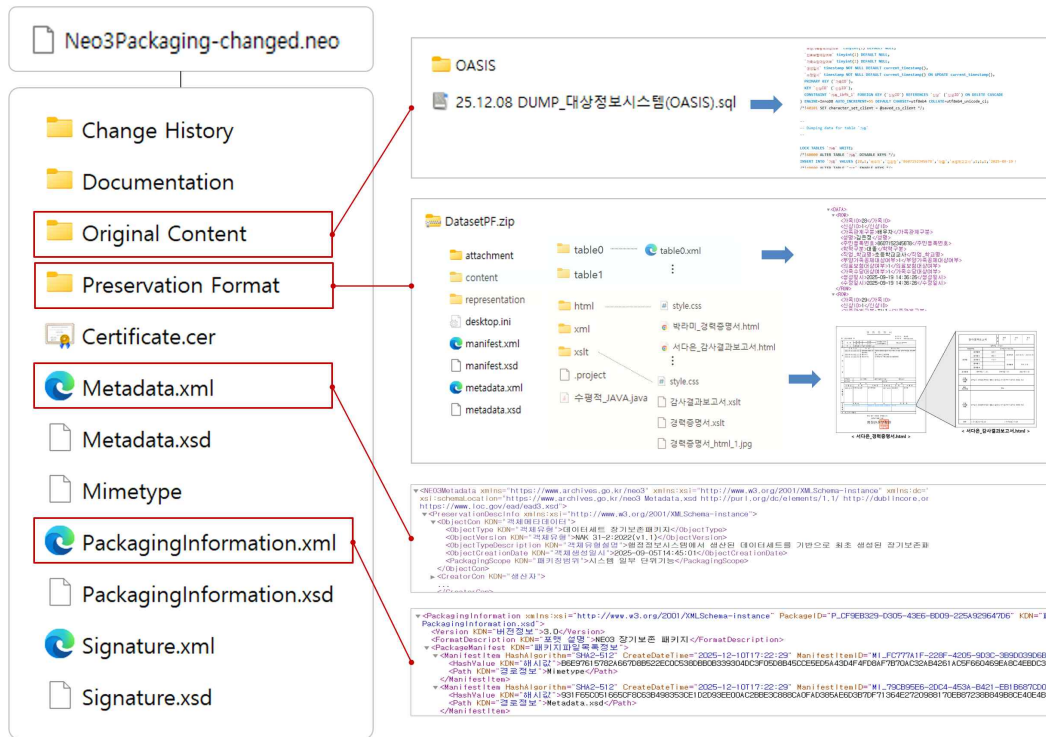
마지막 단계에서는 장기보존패키지에 포함된 구성요소의 생성일시, 식별자, 해시값, 경로 등의 정보를 목록화한 PackagingInformation.xml을 생성한다. 이후 최상위 디렉토리를 기준으로 전체 파일을 하나의 ZIP 파일로 압축하여 장기보존패키지 생성을 완료한다.

4.1.3 장기보존패키지 구현 및 검증

앞서 제안한 데이터세트 장기보존패키지의 구현 가능성을 확인하기 위하여 관계형 데이터베이스 기반의 검증 환경을 구축하고 장기보존패키지를 시범 구현하였다. 본 검증에서는 실제 데이터베이스를 구성하는 데이터와 관련 정보를 추출하여 원문, 보존포맷, 재현정보, 장기보존 메타데이터 등을 생성하고, 이를 제안한 장기보존패키지의 구조에 따라 구성하는 데 중점을 두었다. 검증환경은 MySQL 기반의 관계형 데이터베이스로 구성하였으며, 총 16개의 테이블을 생성하고 여기에 493건의 데이터(첨부파일 포함)를 입력하였다. 각 테이블에는 데이터 타입과 기본키(PK), 외래키(FK), NULL 허용 여부 등의 제약조건을 설정하고 테이블 간 참조관계를 구성하였다. 또한 데이터베이스의 구조와 기능을 정의하는 View, Trigger, Procedure, Function, Event, Index와 사용자 및 권한정보를 포함하여 관계형 데이터베이스의 구조적·기능적 특성이 반영되도록 하였다.

원문은 데이터세트의 추출 범위에 해당하는 데이터와 DB 구조 정보를 SQL문 형태로 추출하여 저장하였다. SQL 파일은 별도의 전용 소프트웨어 없이 일반적인 텍스트 편집기를 통해 내용을 확인할 수 있으므로, 원문을 위한 별도의 재현정보는 구성하지 않았다. 첨부파일은 데이터와 별도로 추출하여 원문 디렉토리 하위에 저장하였다. 보존포맷은 국가기록원(2025)의 보존포맷 규격에 따라 생성하였다. Python을 이용하여 MySQL 서버에 접속한 후 보존 대상 데이터와 테이블 구조를 조회하고 SQL-XML 변환 코드를 통해 테이블별 XML 파일로 추출 및 변환하였다. 보존포맷에 대한 재현정보로는 데이터(XML) 파일에 스타일(XSLT, CSS) 파일을 적용하여 화면 레이아웃, 색상, 글꼴, 표 구조 등을 반영한 HTML 파일을 생성하고, 관련 HTML 파일 간에 링크를 설정하여 연계된 정보를 탐색할 수 있도록 하였다. 시범 구현 대상에 대한 장기보존 메타데이터를 작성한 후, 기록관에서 발급받은 인증서를 이용하여 Metadata.xml의 전자서명값을 생성하고 이를 Signature.xml 파일에 기록하였다. 이후 장기보존패키지의 구성요소 목록을 기술한 패키징 정보를 생성하였다. 마지막으로 ZIP 규약에 따라 디렉토리 내 전체 파일을 압축·패키징하여 .NEO3 확장자로 저장하였다. 이렇게 만들어진 장기보존패키지는 <그림 3>과 같다.

검증 결과, 데이터, 첨부파일, DB 구조 정보, 재현정보 등 원문에 해당하는 데이터세트의 보존 범위가 장기보존패키지의 구성체계 내에서 함께 보존·관리될 수 있음을 확인하였다. 다만 본 검증은 관계형 데이터베이스 기반의 제한된 검증환경을 대상으로 수행되었다는 점에서, 실제 행정정보시스템의 규모와 다양한 구성요소, 구조 및 운영환경을 충분히 반영하지 못하였다. 또한 장기보존패키지에 포함된 파일에 대한 장기보존 메타데이터 작성, 패키징 정보 작성, 재현정보 구성 등의 과정이 수작업으로 수행되었는데, 대용량·대규모 데이터세트를 장기보존패키지로 변환할 경우 상당한 시간과 자원이 소요될 것으로 예상된다. 이를 해결하기 위해 자동화가 가능한 작업을 식별하고, 이를 자동화하는 방안을 마련할 필요가 있다.



〈그림 3〉 데이터세트 장기보존패키지(NEO3) 샘플

4.2 데이터세트 장기보존 메타데이터 요소 재설계

데이터세트 장기보존 메타데이터는 <그림 4>와 같이 차상위요소 6개, 하위요소 29개, 최하위요소 52개, 세부요소 36개로 구성된 4계층 구조로 재설계하였다. 기존 장기보존 메타데이터 요소는 호환성을 유지하기 위해 가능한 범위에서 유지하였으며, 데이터세트의 관리단위와 특성을 반영하기 위해 필요한 요소를 보완하거나 신규로 추가하였다. 각 메타데이터 요소의 정의와 기술규칙은 국가기록원(2025)에서, 장기보존 메타데이터의 스키마는 GitHub (<https://github.com/likeba/NEO3-for-Dataset/blob/main/Metadata.xsd>)에서 확인할 수 있다.

객체메타데이터 차상위요소는 기존 장기보존 메타데이터를 기본적으로 유지하되, 데이터세트의 관리단위를 반영하기 위해 패키징범위를 신규 요소로 추가하였다. 「행정정보 데이터세트 기록관리 기준」(국가기록원, 2020, NAK 35:2020, v1.0)에 따르면 데이터세트는 시스템 전체 또는 시스템 내 일부 단위기능을 관리단위로 설정할 수도 있다. 이에 하나의 장기보존패키지가 시스템 전체를 대상으로 하는지, 특정 단위기능만을 대상으로 하는지를 식별할 수 있도록 패키징범위 요소를 신설하였다.

생산자 차상위요소는 기존 장기보존 메타데이터의 요소를 유지하되, 데이터세트의 생산 특성에 맞게 의미를 재정의하였다. 문서와 달리 데이터세트는 다수의 사용자가 입력한 데이터가 축적된 결과물이므로 개인 단위의 생산자를 특정하기 어렵다. 이에 데이터세트의 생산 또는 시스템 관리에 관련한 기관을 생산자로 기록하는 것을 원칙으로 하고, 부서명·개인명은 필요한 경우에만 선택적으로 기술하도록 하였다.

기록관리식별자 차상위요소는 기존 장기보존 메타데이터에서 기록물철ID 또는 기록물건ID를 식별하던 방식을 데이터세트 유형에 맞게 재설계하였다. 기존에는 장기보존패키지가 단일 기록물철 또는 기록물건을 대상으로 생성되었으나, 데이터세트는 철·건 구조를 따르지 않는다. 이에 식별 대상을 개별 기록물이 아닌 장기보존패키지 자체로 전환하고, 데이터세트 장기보존패키지의 식별코드를 기술하도록 단일 요소로 재구성하였다.



〈그림 4〉 데이터세트 장기보존메타데이터 개요

기록물명 차상위요소는 기록관리식별자와 동일한 원칙에 따라 기록물철·기록물건 중심의 기술 방식을 장기보존패키지 중심으로 재설계하였다. 기존에는 기록물철 또는 기록물건의 제목을 기술하기 위해 제목, 기타제목 등의 하위요소를 두었으나, 데이터세트 유형에서는 장기보존패키지 자체를 식별 대상으로 하므로 하위요소를 제거하고 단일 명칭으로 기술하도록 단순화하였다.

시스템정보 차상위요소는 데이터세트가 생산·관리된 시스템 환경을 장기적으로 이해하고 재현하는 데 필요한 정보를 기술하기 위해 설계하였다. 데이터세트는 행정정보시스템과 데이터베이스 환경에서 생산·관리되는 기록으로, 시스템으로부터 분리될 경우 생산 환경과 기술적 맥락을 함께 보존하지 않으면 데이터의 구조와 의미를 정확하게 해석하기 어렵다. 이에 데이터세트 관리기준표의 시스템 관련 정보를 기반으로 장기보존 이후에도 데이터세트의 생산 환경과 시스템 구성을 확인할 수 있도록 시스템정보를 재구성하였다. 시스템명, 시스템개요, 시스템구축연도 및 DBMS정보는 데이터세트 관리기준표의 해당 항목을 그대로 반영하였다. 반면 시스템폐기연도는 시스템의 운영 종료 시점을 관리하기 위해 새롭게 추가하였다. 또한 연계시스템은 관리기준표보다 구체적인 정보를 기술할 수 있도록 보완하여, 연계 시스템명, 식별자, 보유기관뿐 아니라 연계 방식과 연계 방향까지 기록하도록 하였다. 산출물은 장기보존패키지에 시스템 산출물이 포함되는 경우에 작성하는 요소로, 해당 산출물 파일의 파일명, 파일식별자, 저장경로 및 해시값 등을 기술하여 장기보존패키지 내 산출물 파일을 식별하고 관리할 수 있도록 구성하였다.

데이터세트정보 차상위요소는 데이터세트의 내용, 이용조건 및 생애주기에 관한 정보를 기술하기 위해 설계하였으며, 단위기능정보, 단위기능범위, 권한, 생산기간, 보존기간정보, 처분, 매체, 관리이력, 보존이력 및 컴포넌트의 10개 하위요소로 구성하였다. 이 가운데 권한, 생산기간, 보존기간정보 및 처분은 데이터세트 관리기준표에서 관리하는 정보를 장기보존 이후에도 연속적으로 활용할 수 있도록 그대로 반영하였다. 생산기간은 장기보존패키지에

포함된 데이터세트가 언제부터 언제까지 생산된 기록인지를 나타내는 요소로, 데이터세트의 시간적 범위를 명확하게 기술하기 위해 포함하였다. 단위기능정보와 단위기능범위는 장기보존패키지와 데이터세트 관리기준표를 연계하기 위해 새롭게 구성한 요소이다. 데이터세트 관리기준표에서는 단위기능을 기준으로 기록관리 단위를 설정하고, 이를 실제 행정정보시스템의 데이터 범위에 적용하여 관리한다. 따라서 장기보존패키지에 포함된 데이터가 어떠한 단위기능에 해당하는지와 해당 단위기능이 실제로 어떠한 테이블과 컬럼을 대상으로 하는지를 함께 기술할 필요가 있다. 이러한 정보는 관리기준표의 단위기능과 장기보존패키지에 포함된 실제 데이터 간의 대응관계를 명확히 하는 데 활용된다. 특히 단위기능범위의 조화용쿼리는 장기보존패키지에 포함된 데이터를 어떠한 기준으로 추출하였는지를 기록하기 위해 신설한 요소이다. 데이터 추출 기준을 함께 보존함으로써 동일한 범위의 데이터를 재구성하고 재현할 수 있도록 하였다. 권한 요소는 데이터세트의 부분공개 특성을 반영하여 보완하였다. 문서형 기록물은 기록물 단위로 공개 여부를 판단하면 되지만, 데이터세트는 하나의 단위기능 안에서도 공개 가능한 테이블 또는 컬럼과 비공개 대상인 테이블 또는 컬럼이 함께 존재할 수 있다. 이러한 특성을 표현하지 못하면 부분공개가 가능한 데이터를 일괄적으로 비공개 처리하거나, 반대로 민감정보를 포함한 채 공개하는 문제가 발생할 수 있다. 이에 공개 여부가 적용되는 계층(단위기능·테이블·컬럼)과 그 적용 범위를 기술할 수 있도록 공개대상계층과 공개대상범위를 신규 요소로 추가하였다. 매체, 관리이력 및 보존이력은 장기보존패키지 생성 이후에도 지속되는 관리 과정을 기록하기 위한 요소이다. 데이터세트는 보존기간 재책정, 공개 재분류, 보존포맷 변환 등의 사유로 장기보존 이후에도 메타데이터가 변경될 수 있으므로, 이러한 변경 사항을 기록하지 않으면 특정 시점의 메타데이터 상태를 추적하기 어렵다. 이에 관리이력에 변경요소를 추가하여 변경된 요소명과 이전 값을 함께 기록하도록 하였다. 컴포넌트는 장기보존패키지를 구성하는 개별 파일의 관리 정보를 기술하기 위해 신설한 요소이다. 데이터세트 장기보존패키지는 데이터, 첨부파일, DB 구조 정보, 재현정보, 주요 산출물 등 다양한 파일로 구성되므로, 각 파일을 개별적으로 식별하고 관리할 수 있는 정보가 필요하다. 이에 장기보존패키지에 포함되는 모든 파일을 대상으로 파일명, 저장경로, 해시값, 식별자 등을 기술할 수 있도록 컴포넌트 요소를 구성하였다.

```

<DatasetInfoCon KDN="데이터세트 정보">
  <UnitFunctionCon KDN="단위기능정보">
    <UnitFunctionID KDN="단위기능식별자">UF-83427519-947</UnitFunctionID>
    <UnitFunctionName KDN="단위기능명">기본인적정보(OAS_01)</UnitFunctionName>
    <UnitFunctionOverview KDN="단위기능개요">성명, 주민등록번호, 주소, 연락처 등 인적사항을 포함하여 인사행정의 기초자료로 활용되는 데이터</UnitFunctionOverview>
    <SubjectWords KDN="주제어">인적정보, 개인기본정보, 인사정보</SubjectWords>
  </UnitFunctionCon>
  <UnitFunctionScopeCon KDN="단위기능범위">
    <RetrievalQuery KDN="DB질의문">SELECT * FROM oasis.기본인적</RetrievalQuery>
    <ApplicationScope KDN="적용범위">단위기능 전체</ApplicationScope>
    <ApplicationScopeInfo KDN="적용범위관련정보">개인인사기본 > 기본인적(탐)</ApplicationScopeInfo>
  </UnitFunctionScopeCon>
  <RightsCon KDN="권한">
    <DataOwningCorporateCon KDN="데이터 보유권한">
      <CorporateName KDN="기관명">전북대학교</CorporateName>
      <CorporateCode KDN="기관코드">7000903</CorporateCode>
      <SectionName KDN="부서명">정보혁신부</SectionName>
      <SectionCode KDN="부서코드">C135810</SectionCode>
    </DataOwningCorporateCon>
    <SystemManagementCorporateCon KDN="시스템 관리권한">
      <CorporateName KDN="기관명">전북대학교</CorporateName>
      <CorporateCode KDN="기관코드">7000903</CorporateCode>
      <SectionName KDN="부서명">정보혁신부</SectionName>
      <SectionCode KDN="부서코드">C135810</SectionCode>
    </SystemManagementCorporateCon>
    <InternalAccessScope KDN="접근권한">인사담당자 및 권한 승인 사용자</InternalAccessScope>
  </RightsCon>
</DatasetInfoCon>

```

〈그림 5〉 데이터세트 장기보존 메타데이터(안) 작성 예시

4.3 변경이력 관리 및 복원 방안

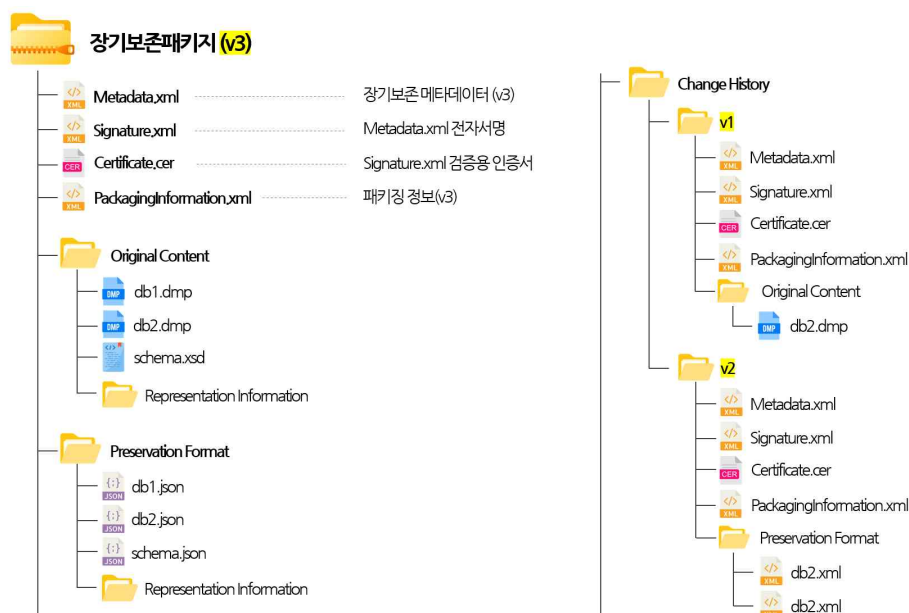
본 연구에서는 데이터세트의 특성을 고려하여 데이터 무결성(Data Integrity)과 중복 최소화(Redundancy Minimization)를 변경이력 관리방안의 설계 원칙으로 설정하였다. 데이터 무결성은 장기보존패키지의 생성 이후 발생하는 모든 변경 과정과 버전을 누락 없이 보존하여, 변경이력의 검증과 특정 시점으로의 복원이 가능해야 한다는 원칙이다. 중복 최소화는 장기보존패키지의 변경 사항만을 선별적으로 저장함으로써 불필요한 중복을

줄이고 저장 및 관리 효율성을 높이기 위한 원칙이다. 데이터세트 장기보존패키지(NEO3)는 대량·대용량의 기록 구성요소로 구성되므로, 변경이 발생할 때마다 패키지 전체를 새로 저장한다면 저장 용량이 급격히 증가하고 관리 효율성 또한 저하될 수밖에 없다.

이와 같은 원칙을 반영하여 다음과 같은 변경이력 관리방안을 설계하였다. 첫째, 변경 사항을 반영한 장기보존 패키지의 구성요소는 항상 최상위 디렉토리에 유지하여, 관리자가 별도의 탐색이나 복원 과정 없이 최신 상태의 구성과 내용을 확인하고 열람·점검·검증 요청에 쉽게 대응할 수 있도록 한다. 둘째, 변경이 발생한 구성요소만 Change History 디렉토리 하위에 버전별 디렉토리를 생성하여 선택적으로 관리한다. 버전 번호는 변경이 발생할 때마다 순차적으로 부여하고, 각 버전 디렉토리에는 변경 이전 구성요소만 원래의 디렉토리 구조에 따라 배치하여 보존한다. 이를 통해 변경 시점별 범위를 명확히 구분하고 이전 버전 복원에 필요한 구성요소를 체계적으로 관리할 수 있으며, 변경되지 않은 구성요소는 중복 저장하지 않으므로 저장공간 사용도 최소화된다. 셋째, 변경이력이 발생할 때마다 해당 버전의 진본확인 정보, 장기보존 메타데이터, 패키징 정보를 함께 보존한다. 이 정보는 버전별 장기보존패키지의 진본성 및 무결성을 검증하고 구성요소의 완전성을 확인하는 근거가 된다. 넷째, 이전 버전의 장기보존패키지는 역방향 병합(Retrieve Merge) 방식을 통해 복원할 수 있도록 한다. 최신 상태의 장기보존패키지를 기준으로 변경이력을 역순으로 적용한 후 패키징 정보를 이용하여 해당 시점의 구성요소를 재구성함으로써, 저장 효율성을 유지하면서도 특정 시점의 장기보존패키지를 복원할 수 있다. 이상의 방안을 통해 장기보존패키지는 현재 상태에 대한 접근성을 유지하면서도, 변경이력을 체계적으로 관리하고 저장 효율성과 데이터 무결성을 동시에 확보할 수 있다.

4.3.1 변경이력 관리 절차

본 연구에서는 장기보존패키지의 변경 전후 상태를 서로 다른 버전으로 구분하여 관리하도록 변경이력 관리방안을 설계하였다. 구성요소에 변경이 발생하면 변경 이전의 상태를 $v(n)$, 변경이 반영된 이후의 상태를 $v(n+1)$ 로 정의하였다. 이에 따라 <그림 6>과 같이 장기보존패키지는 최상위 디렉토리에 최신 버전을 유지하고, Change History 디렉토리에는 이전 버전의 변경이력을 관리하는 구조를 갖는다.



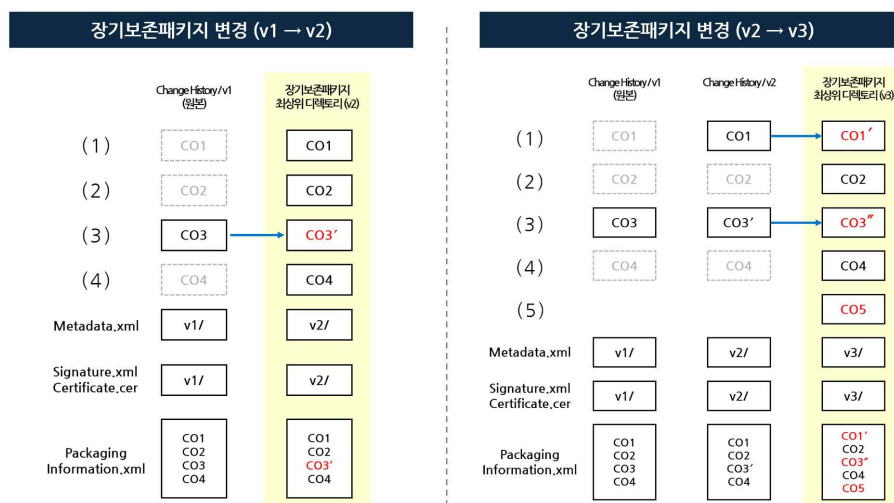
<그림 6> 장기보존패키지 Change History 디렉토리 예시

본 변경이력 관리방안을 장기보존패키지에 적용하면, 구성요소의 변경은 <표 6>과 같은 4단계 절차에 따라 수행된다.

<표 6> 변경이력 관리 절차

구분		내용
1단계	변경될 구성요소 식별	Metadata.xml, PackagingInformation.xml, Certificate.cer, Signature.xml, Original Content 디렉토리, Preservation Format 디렉토리, Documentation 디렉토리 내 파일 등에서 변경 요소 식별
2단계	Change History/v(n) 디렉토리 생성	Change History 디렉토리 하위에 변경 이전 버전을 저장하기 위한 버전 디렉토리 생성 (최초 변경 발생 시 Change History/v(n) 디렉토리 생성, 이후 변경마다 버전 번호를 1씩 증가시켜 디렉토리 생성)
3단계	변경 이전 구성요소 보존	변경 이전 구성요소와 Metadata.xml, PackagingInformation.xml, Certificate.cer, Signature.xml을 Change History/v(n)/ 디렉토리에 원래의 구조를 유지한 상태로 보존
4단계	변경된 구성요소 반영	기존 디렉토리 구조의 해당 위치에 변경된 최신 구성요소 반영

<그림 7>은 장기보존패키지 원본(v1)에 두 차례의 변경이 발생하는 상황을 가정하여 변경이력 관리 과정을 도식화한 것이다. 1차 변경(v1 → v2)에서는 구성요소 CO3이 CO3'으로 변경되고, 2차 변경(v2 → v3)에서는 CO1이 CO1'으로, CO3'이 CO3''으로 변경되며 CO5가 새롭게 추가된다. 먼저 1차 변경에서는 Change History 디렉토리 하위에 v1 디렉토리를 생성한 후 변경 이전의 CO3을 복사하여 배치한다. 이와 함께 해당 버전의 Metadata.xml, Signature.xml, Certificate.cer, PackagingInformation.xml도 함께 보존한다. 이는 구성요소의 변경이 장기보존 메타데이터, 진본확인 정보 및 패키징 정보에도 연쇄적인 변경을 발생시키기 때문이다. 장기보존 메타데이터에는 변경된 구성요소 정보와 관리이력이 반영되고, 진본확인 정보는 변경된 장기보존패키지에 대해 새롭게 행정전자서명을 수행하여 갱신되며, 패키징 정보 역시 변경된 구성요소의 경로, 해시값 및 식별자 등을 반영하도록 갱신된다. 이후 장기보존패키지 최상위 디렉토리에는 변경된 구성요소(CO3')와 갱신된 메타데이터, 진본확인 정보 및 패키징 정보가 배치되어 최신 버전(v2)을 구성한다. 2차 변경도 동일한 방식으로 수행된다. Change History 디렉토리 하위에 v2 디렉토리를 생성하여 변경 이전의 CO1, CO3' 및 해당 버전의 장기보존 메타데이터, 진본확인 정보, 패키징 정보를 함께 보존한다. 이후 최상위 디렉토리에는 변경된 CO1', CO3''과 새롭게 추가된 CO5를 반영하고, 이에 맞추어 장기보존 메타데이터, 진본확인 정보 및 패키징 정보를 갱신하여 최신 버전(v3)을 구성한다.



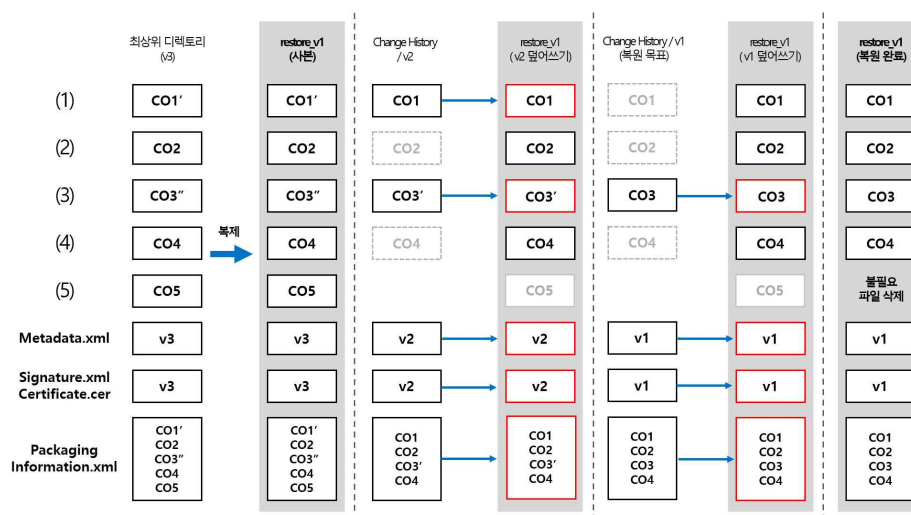
<그림 7> 변경이력 관리방안 예시

4.3.2 변경이력 복원 절차

본 연구에서는 최신 버전의 장기보존패키지를 기준으로 복원 대상 버전에 이를 때까지 변경이력을 역순으로 적용하는 방식으로 변경이력 복원 방안을 설계하였다. 복원은 최상위 디렉토리에 유지되는 최신 버전(v(n))에서 시작하여 v(n-1), ..., v(p)의 순서로 변경이력을 역순으로 탐색하면서 수행한다($1 \leq p < n$). 각 단계에서는 해당 버전에 보존된 변경 이전의 구성요소를 현재 장기보존패키지의 구성요소에 덮어쓰기(Overwrite) 방식으로 반영하며, 이 과정을 복원 대상 버전인 v(p)에 이를 때까지 반복한다. 이와 같이 변경된 구성요소는 순차적으로 이전 상태로 복원되고, 변경되지 않은 구성요소는 최신 버전의 상태를 그대로 유지함으로써 복원 대상 시점의 장기보존패키지를 재구성할 수 있다. 이러한 복원 방안을 장기보존패키지에 적용하면 <표 7>과 같은 5단계 절차에 따라 특정 버전의 장기보존패키지를 복원할 수 있다.

<표 7> 변경이력 복원 절차

구분	내용
1단계	복원 대상 버전 지정 v(1), v(2), ..., v(n) 중 복원하려는 장기보존패키지의 버전(v(p), $1 \leq p < n$)을 선택
2단계	최신 버전 복제 현재 최상위 디렉토리에 위치한 전체 구성요소를 복제한 별도의 장기보존패키지(restore_vP) 생성
3단계	패키징 정보 기반 역순 병합 Change History/v(n-1)부터 Change History/v(p)까지 역순으로 탐색하면서 해당 버전에 저장된 변경 이전의 구성요소를 복제본(restore_vP)에 덮어쓰기(Change History/v(p)/PackagingInformation.xml에 정의된 구성요소만을 최종 복원 대상으로 함)
4단계	구성요소 정합성 확인 Change History/v(p)/PackagingInformation.xml에 정의된 파일 목록을 기준으로 restore_vP의 구성요소가 복원 대상 버전과 일치하는지 확인하고, 목록에 없는 파일을 제거하여 구성요소의 정합성을 확보
5단계	무결성 검증 Change History/v(p)/PackagingInformation.xml에 정의된 해시값과 restore_vP의 각 파일에 대한 실제 해시값을 비교하여 복원 결과의 무결성 검증



<그림 8> 변경이력 복원 방안 예시

<표 7>의 복원 절차를 도식화하면 <그림 8>과 같다. <그림 8>은 현재 장기보존패키지의 최신 버전이 v3인 상태에서 v1으로 복원하는 과정을 예시로 나타낸 것이다. 먼저 복원 대상인 v1에 해당하는 별도의 장기보존패키지(restore_v1)를 생성하고, 최상위 디렉토리에 유지되는 최신 버전(v3)의 모든 구성요소를 복제한다. 이후 Change History/v2 디렉토리에 저장된 변경 이전의 구성요소를 restore_v1에 덮어써 v2 상태로 복원한다. 이 과정에서

Metadata.xml, Signature.xml, Certificate.cer 및 PackagingInformation.xml도 함께 덮어써져 v2 버전에 맞게 갱신된다. 다음으로 Change History/v1 디렉토리에 저장된 변경 이전의 구성요소를 restore_v1에 다시 덮어써 v1 상태로 복원한다. 이와 함께 앞선 방식과 같이 장기보존 메타데이터, 진본확인 정보 및 패키징 정보도 v1 버전에 맞게 갱신한다. 이후 패키징 정보(PackagingInformation.xml)에 정의된 구성요소 목록을 기준으로 restore_v1의 파일을 확인하고, 목록에 포함되지 않은 파일을 제거하여 v1 버전의 구성요소만 남도록 한다. 마지막으로 복원된 각 파일의 실제 해시값과 패키징 정보에 기록된 해시값을 비교하여 무결성을 검증하고, 모든 값이 일치하면 v1 버전 장기보존패키지의 복원이 완료된다.

5. 결 론

「공공기록물법 시행령」에는 보존기간 30년 이상인 전자기록물을 장기보존패키지로 변환하여 이관·보존하도록 규정하고 있으며, 국가기록원은 이를 뒷받침하는 표준으로 장기보존패키지 기술규격을 제정하여 운영 중에 있다. 데이터세트 역시 전자기록물의 한 유형인 만큼, 보존가치가 높은 경우에는 장기보존패키지로 변환하여 관리할 필요가 있다. 그러나 현행 장기보존패키지(NEO3)는 다양한 전자기록물 유형을 수용할 수 있음을 전제로 설계되었음에도 일부 구성요소의 개념과 범위가 여전히 문서유형을 중심으로 정의되어 있어 데이터세트 유형에 그대로 적용하기 어렵다는 한계가 있다. 이에 본 연구는 이러한 한계를 보완하여 데이터세트 유형에도 적용할 수 있도록 장기보존패키지의 확장 방안을 제안하였다.

이를 위해 데이터세트의 특성을 분석하여 장기보존패키지에 포함되어야 하는 보존 범위를 도출하였다. 그 결과 데이터, 첨부파일, DB 구조 정보, 재현정보는 데이터세트의 원문으로 보존되어야 하며, 주요 시스템 산출물은 데이터세트의 구조와 생산 맥락을 이해하기 위한 참고자료로 함께 보존될 수 있음을 확인하였다. 이러한 보존 범위를 기준으로 현행 장기보존패키지(NEO3)의 구조와 구성요소를 분석하여 그대로 적용 가능한 요소와 재설계 또는 신설이 필요한 요소를 구분하였다. 이를 바탕으로 데이터세트 유형에 적합한 장기보존패키지(안)를 제안하였다. 구체적으로는 현행 장기보존패키지의 기본 구조를 유지하면서도, 재현정보와 주요 시스템 산출물을 체계적으로 관리할 수 있도록 Representation Information과 Documentation 디렉토리를 추가하고, 데이터세트의 관리 특성을 반영한 장기보존 메타데이터를 설계하였다. 또한 장기보존패키지의 생성 절차, 변경이력 관리 및 복원 방안을 함께 제시함으로써 데이터세트 유형의 장기보존패키지가 생성 이후에도 지속적으로 관리될 수 있는 체계를 마련하였다.

본 연구는 문서유형 전자기록물을 중심으로 설계된 현행 장기보존패키지(NEO3)를 데이터세트 유형에도 적용할 수 있도록 보존 범위를 설정하고 구조를 재설계하였다는 점에서 의의가 있다. 다만 다음과 같은 한계를 지니며, 이를 해결하기 위한 후속 연구가 필요하다. 먼저 본 연구에서 제안한 장기보존패키지(안)의 실효성을 실제 운영 중인 행정정보시스템 수준에서 검증하지 못하였다. 이에 향후 연구에서는 보다 복잡하고 대용량·대규모 데이터세트를 대상으로 적용 가능성을 검증할 필요가 있다. 또한 데이터 유형이 통합·공유되는 환경으로 변화하면서 기록물의 원문이 DB와 응용 내부에 분산되어 저장되는 사례가 늘고 있어, 최초 생성 당시의 형태와 내용을 담은 원문을 특정하기 어려운 경우가 발생하고 있다. 이를 대비하기 위한 보존 전략을 마련할 필요가 있다. 한편, 장기보존패키지로 변환된 데이터세트가 영구기록물관리기관으로 이관된 이후에도 원본 데이터세트가 행정정보시스템 내에 남아 계속 활용되는 등 원본과 사본의 관계가 아직 명확히 정립되지 않은 상태이므로, 행정정보시스템 내 데이터세트와 장기보존패키지 내 데이터세트 간 원본·사본 관계를 규명하는 연구가 이루어져야 한다. 마지막으로 원문 추출, 보존포맷 변환, 장기보존 메타데이터 생성, 패키징 정보 생성 등 현재 수작업으로 이루어지는 업무를 대상으로 자동화 가능성을 검토하여 업무 효율성을 높일 수 있는 기술적 방안을 마련할 필요가 있다.

참고문헌

- 공공기록물 관리에 관한 법률, 법률 제21066호.
공공기록물 관리에 관한 법률 시행령, 대통령령 제35506호.
국가기록원 (2020). 행정정보 데이터세트 기록관리 기준 - 관리기준표 작성 및 이관규격 (NAK 35:2020, v1.0).
국가기록원 (2022). 전자기록물 장기보존패키지 기술규격 - 제2부: 디렉토리로 구조화된 방식(NEO3) (NAK 31-2:2022, v1.1).
국가기록원 (2025). 오픈 포맷을 활용한 행정정보 데이터세트 보존 방안 연구 (11-1741050-100022-01).
김수영 (2023). 디렉토리 이관 방안 적용 사례 연구. 기록과 정보·문화 연구, 16, 7-50.
<https://www.doi.org/10.23035/KAICS.2023.1.16.007>
김지혜, 윤성호, 양동민 (2022). 행정정보 데이터세트 종합관리시스템의 서비스 방안 연구. 한국기록관리학회지, 22(2), 27-49.
<https://www.doi.org/10.14404/JKSARM.2022.22.2.027>
박라미, 양동민 (2026). 웹 기반 행정정보 데이터세트의 XSLT 재현 방안 연구: HTML 구조 유형화를 중심으로. 한국기록관리학회지, 26(2), 131-156. <https://www.doi.org/10.14404/JKSARM.2026.26.2.131>
박병주, 차승준, 이규철 (2010). 웹기록물 보존을 위한 전자기록물 장기보존포맷 확장 설계. 한국디지털산업학회지, 15(4), 33-47.
변우영, 임진희 (2022). 행정정보 데이터세트 이관도구 SIARD_KR의 개선방안. 정보관리학회지, 39(1), 195-217.
<https://www.doi.org/10.3743/KOSIM.2022.39.1.195>
왕호성, 설문원 (2017). 행정정보 데이터세트 기록의 관리방안. 한국기록관리학회지, 17(3), 23-47.
<https://www.doi.org/10.14404/JKSARM.2017.17.3.023>
윤성호, 이정은, 양동민 (2021). 행정정보 데이터세트 보존포맷으로서 SIARD 검증에 관한 연구. 한국기록관리학회지, 21(3), 99-118. <https://www.doi.org/10.14404/JKSARM.2021.21.3.099>
임진희 (2006). 전자기록의 장기보존을 위한 보존정보패키지(AIP) 구성과 구조. 기록학연구, 13, 41-90.
한능우 (2023). 전자기록 장기보존정책의 현황과 과제. 기록학연구, 78, 79-151. <https://www.doi.org/10.20923/kjas.2023.78.079>
한희정, 오효정, 양동민 (2020). 전자기록물의 장기보존을 위한 보존포맷 선정 방안에 관한 연구. 한국기록관리학회지, 20(1), 69-87. <https://www.doi.org/10.14404/JKSARM.2020.20.1.069>

• 국문 참고자료의 영어 표기

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

- Byeon, Woo-yeong & Yim, Jin-hee (2022). Improvement of administration information dataset transfer tools 'SIARD_KR'. Journal of the Korean Society for Information Management, 39(1), 195-217.
<https://www.doi.org/10.3743/KOSIM.2022.39.1.195>
Enforcement Decree of the Public Records Management Act. Presidential Decree No.35506.
Han, Hui-Jeong, Oh, Hyo-Jung, & Yang, Dongmin (2020). A study on the selection of preservation format for long-term preservation of electronic records. Journal of Korean Society of Archives and Records Management, 20(1), 69-87.
<https://www.doi.org/10.14404/JKSARM.2020.20.1.069>
Han, Nungwoo (2023). Long-Term preservation strategy of digital records. The Korean Journal of Archival Studies, 78, 79-151.
<https://www.doi.org/10.20923/kjas.2023.78.079>
Kim, Ji-Hye, Yoon, Sung-Ho, & Yang, Dongmin (2022). A study on the service of the integrated administrative information dataset management system. Journal of Korean Society of Archives and Records Management, 22(2), 27-49.
<https://www.doi.org/10.14404/JKSARM.2022.22.2.027>
Kim, Su Young (2023). A study on the application of administrative information dataset transfer methods. The Korean Journal of Archival, Information and Cultural Studies, 16, 7-50. <https://www.doi.org/10.23035/KAICS.2023.1.16.007>
National Archives of Korea (2020). Record Keeping Criteria for Dataset - Composition of Dataset Management Reference

- Table & Exchange of Dataset (NAK 35:2020, v1.0)
- National Archives of Korea (2022). Technical Specification for Long-Term Preservation Package Part 2: Directory Structured Format(NEO3) (NAK 31-2:2022, v1.1).
- National Archives of Korea (2025). A Study on Preservation Methods for Administrative Information Datasets Using Open Formats (11-1741050-100022-01).
- Park, Boung-Joo, Cha, Seung-Jun, & Lee, Kyuchul (2010). Extension of the long-term archival information package for electronic records to accommodate web records. *Korean Journal of Digital Industry*, 15(4), 33-47.
- Park, Lami & Yang, Dongmin (2026). A study on XSLT-based reproduction of web-based administrative information datasets: Focusing on HTML structure typology. *Journal of Korean Society of Archives and Records Management*, 26(2), 131-156. <https://www.doi.org/10.14404/JKSARM.2026.26.2.131>
- Public Records Management Act. Act No.21066.
- Wang, Hosung & Seol, Moon-won (2017). A study on managing dataset records in government information systems. *Journal of Korean Society of Archives and Records Management*, 17(3), 23-47. <https://www.doi.org/10.14404/JKSARM.2017.17.3.023>
- 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.
- Yoon, Sung-Ho, Lee, Jung-eun, & Yang, Dongmin (2021). A study on SIARD verification as a preservation format for data set records. *Journal of Korean Society of Archives and Records Management*, 21(3), 99-118. <https://www.doi.org/10.14404/JKSARM.2021.21.3.099>