

Format)이라 한다.

3.18 컴포넌트(Component)

기록물건을 구성하는 기록물의 최소단위. 일반문서 유형인 경우 종이문서, 전자문서의 본문, 첨부 데이터파일 등이 기록물건을 구성하는 컴포넌트가 되고, 다른 기록 유형인 경우 각 기록물건을 구성하는 고유의 컴포넌트 형식을 가진다.

3.19 해시함수(Hash function)

임의의 길이의 문자열을 고정된 길이의 이진 문자열로 매핑하여 주는 함수. 데이터를 자르고, 치환하거나 위치를 바꾸는 방법들로 결과를 만들어 내며, 이 결과를 해시 값(hash value)이라 한다. 해시함수는 데이터의 무결성, 인증, 부인 방지 등에서 응용되는 중요한 함수 가운데 하나이다.

[TTA 정보통신용어사전]

3.20 행정전자서명(GPKI Digital Signature)

전자문서를 작성한 행정기관 또는 행정기관과 전자문서·행정정보를 유통하는 법인·단체에서 직접 업무를 담당하는 자의 신원과 전자문서의 변경 여부를 확인할 수 있는 정보

3.21 확장성 스타일시트 언어 변환(eXtensible Stylesheet Language Transformations, XSLT)

확장성 마크업 언어(XML) 문서를 다른 스타일의 XML 문서로 변경하거나 다른 문서 형식으로 변환하기 위해 개발된 언어. XSLT 기반으로 문서 변환 시 원본 문서는 변경되지 않고 변환 결과로 새로운 문서가 생성된다. HTML, 텍스트(txt), PDF, 포스트스크립트(PostScript), 이미지(PNG 등), 한글(HWP) 등의 형식으로 변환할 수 있다.

[TTA 정보통신용어사전]

3.22 Base64 인코딩(Base64 Encoding)

바이너리 데이터를 아스키 텍스트로 변환하거나, 그의 반대로 변환하는 인코딩 방법. Base64는 2진 데이터의 각 3바이트씩을 4개의 6비트 단위로 나누어 하나의 64진수 문자로 변환한다. 이것은 파일 크기를 대체로 원래보다 약 1/3 정도 증가시킨다.

[텀즈 컴퓨터용어사전 참조하여 개작]

4 장기보존패키지 개요

4.1 OAIS의 정보 패키지와 장기보존패키지의 관계

이 표준에서 규정하고 있는 장기보존패키지는 ISO 14721 OAIS 참조모형의 정보 패키지(Information Package)를 참고한다. 정보 패키지는 **그림 1**과 같이 콘텐츠 정보(Content Information), 보존기술정보(Preservation Description Information), 패키징 정보(Packaging Information), 기술 정보(Descriptive Information)로 구성되어 있다. 콘텐츠 정보는 실제 보존의 대상인 기록물을 의미하며, 원문과 보존포맷으로 구성될 수 있다. 보존기술정보는 콘텐츠 정보의 보존과 설명에 필요한 메타데이터로서 본 표준에서 규정하는 있는 장기보존패키지에서는 장기보존 메타데이터로 표시한다. 패키징 정보는 콘텐츠 정보와 보존기술정보를 실제적으로 아니면 논리적으로 묶고, 식별하고, 관련 짓는 정보이다. 기술정보는 이용자가 해당 정보 패키지를 검색할 때 활용될 수 있는 정보로 전자기록물 목록 정보에 해당된다.

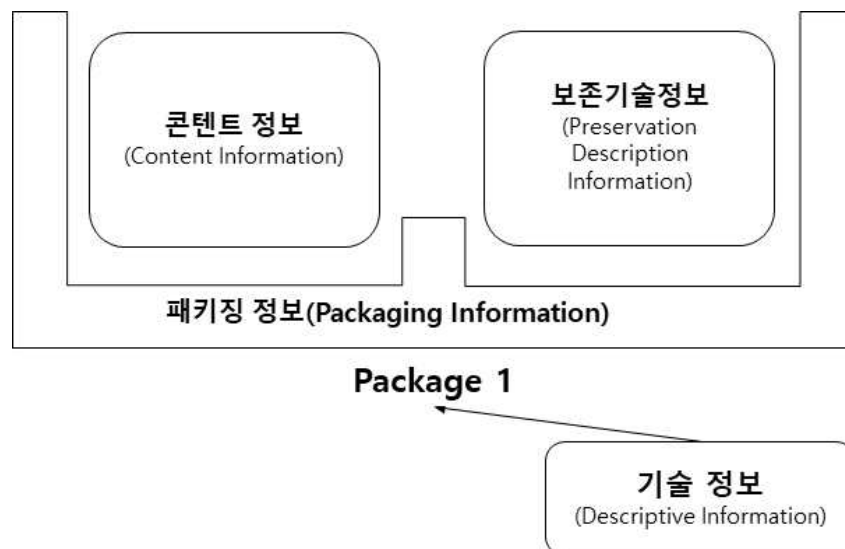


그림 1 - ISO 14721: OAIS 정보 패키지 개념 및 관계

장기보존패키지는 전자기록물의 진본성·무결성을 유지하고 장기간 안전하게 보존할 수 있도록 장기보존 메타데이터와 원문 및 보존포맷을 실제적 아

다면 논리적으로 캡슐화하여 관리하는 개념적인 컨테이너이다. 즉, 장기보존 패키지는 전자기록물 자체와 장기보존을 위한 보존포맷 및 메타데이터를 포함할 수 있도록 설계되었으므로 ISO 14721 OAIS 참조모델에서 제시하고 있는 정보 패키지(SIP, DIP, AIP) 중에서 AIP에 해당한다.

4.2 장기보존패키지 필요성

전자기록물은 그 특성상 변형, 훼손, 유실되기 쉬우므로 보다 안전하게 보존할 수 있도록 하는 기술적 기반을 필요로 한다.

즉, 오랜 기간이 경과해도 기록물이 생산된 당시에 가지고 있던 내용을 그대로 재현하여 접근할 수 있도록 보존되어야 하며, 업무 활동에 대한 증거와 업무에 대한 책임 소재를 분명하게 밝혀주는 법적 증거를 확보할 수 있도록 하여야 한다.

기록물이 가진 법적 증거는 기록물의 진본성과 무결성이 유지되어야만 확보될 수 있으므로 전자기록물에 대해 이러한 진본성과 무결성을 유지하면서 장기간 보존할 수 있게 하도록 하는 장기보존패키지의 적용이 필요하다.

4.3 장기보존패키지 고려사항

4.3.1 자체 충족성

기록물이 생산된 당시의 내용을 그대로 재현하여 읽고 이해할 수 있도록 하기 위하여 전자기록물은 시스템, 외부 데이터 등에 독립적이어야 한다.

시스템 의존적일 경우 시스템의 손실은 전자기록물의 손실이 되고, 외부데이터의 손실 역시 전자기록물의 손실이 되기 때문에 전자기록물은 자체적으로 기록물이 생산된 당시의 내용을 그대로 재현할 수 있는 충분한 능력을 가져야 한다.

4.3.2 자체 문서화

전자기록물은 자체적으로 기록물과 관련된 기술(記述)과 맥락에 대한 이해를

줄 수 있는 정보를 포함하여, 미래의 이용자들이 장기 보존된 기록물의 내용을 이해할 수 있도록 하여야 한다.

즉 기록물을 기술(記述)함과 동시에, 다른 기록물이나 기관 또는 조직과의 관계를 기술하고, 기록물의 지속적인 관리에 관련된 메타데이터를 포함하여야 한다. 메타데이터가 포함됨으로써 장기보존패키지가 독립적인 객체로서 기능할 수 있도록 한다.

4.3.3 진본성 및 무결성 유지

이관, 수집 등을 통해 영구기록물관리기관 등으로 제출(submission)된 전자기록물이 장기보존패키지로 변환되거나 저장·관리되는 동안 진본성을 유지할 수 있어야 한다. 이를 위해 전자서명 등의 기술을 활용한 진본확인 절차를 통해 검증할 수 있어야 한다.

전자기록물이 위조 또는 변조되지 않았음을 검증하여 기록물의 법적 증거를 확보하여야 한다. 장기보존패키지를 저장·관리하는 동안 무결성을 검증할 수 있어야 한다.

4.4 장기보존패키지 구성요소

장기보존패키지는 원문, 보존포맷, 장기보존 메타데이터, 진본확인 정보 등으로 구성된다. 이 구성 요소들은 장기보존패키지로 캡슐화되어야 하며, 캡슐화 방법은 국제표준 ISO 14721 OAIS 참조모형에 따르면, 논리적(logically) 아니면 실제적(actually) 방식으로 크게 구분한다. 캡슐화 방식에 대한 상세 유형은 **4.5 장기보존패키지 생성방식**을 참조한다.

장기보존패키지의 구성요소는 다음과 같다.

- **원문** : 생산자가 생산 또는 접수한 전자기록물 원본(진본)으로 진본성을 보장하기 위해 포함한다.
- **보존포맷** : 전자기록물 생산 당시의 애플리케이션이 없이도 해당문서의 내용과 외형을 그대로 재현한 포맷 또는 원본의 디지털객체 특성정보(내용과

문맥정보, 유형별 특성정보 등)를 원본과 동일한 수준으로 보존 가능한 포맷으로 시간과 기술변화에 상관없이 이용자가 기록물 내용에 접근할 수 있게 한다. 다만, 변환 가능한 보존포맷이 정해지지 않은 원문의 경우는 보존 대책을 마련할 때까지 원문만 포함할 수 있다.

- **장기보존 메타데이터** : 기록물의 생산부터 관리·보존에 이르는 전 과정을 기술(記述)한 정보로, 기록물 생애주기 전 기간에 걸쳐 진본성, 신뢰성, 무결성, 이용가능성을 유지하며, 기록물을 관리하고 보존·이해할 수 있도록 지원한다. 또한 패키지를 설명하는 패키징 정보 등 추가적인 메타데이터를 포함한다.
- **장기보존 메타데이터 스키마 정보** : 장기보존 메타데이터는 XML 문서로 저장된다. 메타데이터의 XML 구조는 시대가 흐르고 생산 시스템이 변화함에 따라 지속적으로 변화된다. XML 스키마는 해당 패키지의 장기보존 메타데이터 유효성을 특정 시스템에 종속되지 않고 독립적으로 검증하기 위해 포함된다.
- **진본확인 정보** : 전자기록물의 진본성 및 무결성 유지를 위해 장기보존패키지에 포함되는 정보이다. 이 표준에서는 진본확인 기술 중 전자서명을 사용하며 관련 정보로 GPKI 인증서, XML전자서명 파일 등을 포함할 수 있다. 영구기록물관리기관 등이 전자서명 외의 다른 진본확인 기술을 적용하는 경우, 해당 기술에 따른 진본확인 정보를 관리할 수 있어야 한다.

비고 2 장기보존패키지 구성요소별 자세한 내용은 '5 XML로 포맷화된 방식의 장기보존패키지(NEO2)'를 참조한다.

4.5 장기보존패키지 생성방식

4.5.1 일반사항

장기보존 메타데이터와 전자기록물의 콘텐츠 정보에 해당하는 원문 및 보존포맷은 국제표준 ISO 14721 OAIS 참조모형의 패키징 정보에 의해 실제적으로(actually) 아니면 논리적으로(logically) 묶고, 식별하고 관련지을 수 있다.

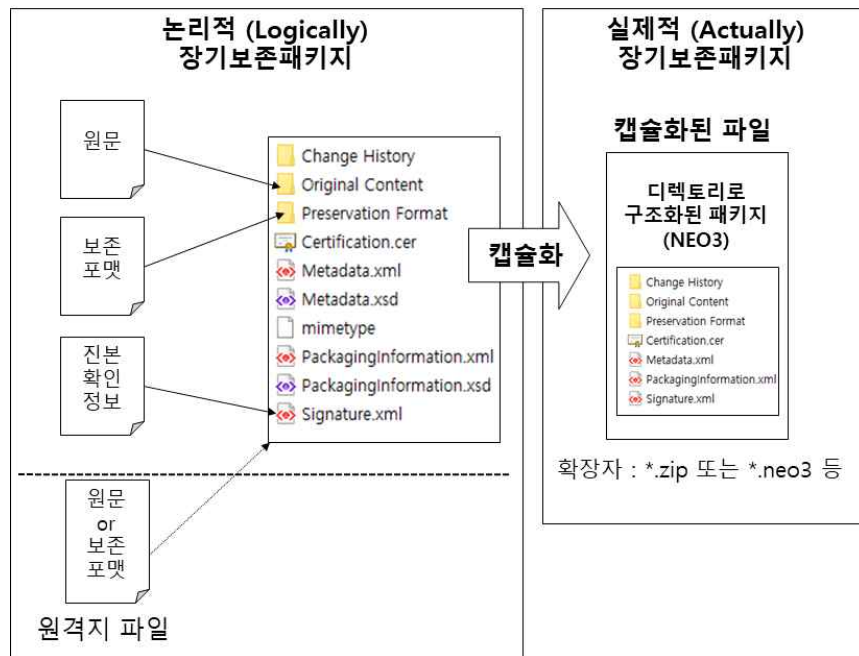


그림 2 - 논리적 · 실제적 장기보존패키지 사례

장기보존패키지는 **그림 2**와 같이 장기보존 메타데이터와 원문 및 보존포맷을 단일한 디지털 객체로 생성하는 방식의 실제적 장기보존패키지 또는 여러 디지털 객체들로 나뉘어 있는 것을 서로 연결하는 방식인 논리적 장기보존패키지가 모두 가능하다.

논리적 장기보존패키지는 구성요소인 장기보존 메타데이터, 원문 및 보존포맷 등을 패키징 정보에 포함된 식별자를 통해 서로 연결, 식별이 가능하며, 패키징 정보는 이를 위해 URL(Uniform Resource Locator), URN(Uniform Resource Name), URI(Uniform Resource Identifier) 등을 포함한다. 원문 파일이나 보존포맷이 동일한 물리적 스토리지에 존재하지 않거나, 진본확인 기술의 보안성 강화를 위해 전자서명 등 진본확인 정보가 저장된 파일만을 보안이 강화된 별도의 스토리지에 보존하는 등의 경우에 논리적 장기보존패키지 방식을 적용할 수 있다. 다만, 이 장기보존패키지 방식은 장기보존 메타데이터 또는 원문 및 보존포맷의 물리적인 위치가 변경되더라도 패키징 정보의 논리적인 연결 관계가 단절되지 않도록 장기간 지속적으로 확인할 수 있는 별도의 식별자 검증 체계가 필요하다.

논리적 장기보존패키지는 실제로 하나의 객체로 캡슐화하지 않지만 장기보존 메타데이터, 패키징 정보, 원문, 보존포맷, 진본확인 정보를 논리적으

로 캡슐화함으로써 장기보존패키지의 구성요소가 서로 물리적으로 떨어져 있는 것을 허용한다. 앞서 언급한 바와 같이 이 패키지 방식을 적용하기 위해서는 원격지에 존재하는 디지털 객체의 위치정보 및 구분 정보를 관리하기 위한 디지털 객체의 식별자 관리 시스템이 마련되어야 한다. 이를 위해 디지털 객체 식별자(DOI) 시스템과 같은 방식이 사용될 수 있으며 이를 통해 각각의 요소를 영구적으로 연계하고 관리할 수 있다.

이 유형의 캡슐화 방식은 정보시스템, 매체 등 다양한 형태로 분산 저장되어 유실된 기록물에 대한 복구가 더 어려울 수 있다. 그러나 대용량의 전자기록물을 장기 보존하거나 비전자기록물과 전자기록물을 연계하는 경우, 특히 영구기록물관리기관 등에 이관된 대용량 멀티미디어 기록물의 캡슐화를 위해 대용량으로 이동, 복사 등 장기보존 처리 절차가 힘든 경우에는 불가피하게 논리적으로 캡슐화하는 방식으로 장기보존패키지를 생성할 수 있다.

실제적 장기보존패키지 방식은 현행 장기보존패키지인 NEO2(NAK Encapsulated Object 2)가 해당되며, 공공표준 'NAK 31-1:2022(v2.3) 전자기록물 장기보존패키지 기술규격-제1부: XML로 포맷화된 방식(NEO2)'에서 그 내용을 상세하게 규정하였다. 논리적 장기보존패키지 방식은 NEO3(NAK Encapsulated Object 3)가 해당되며, 공공표준 'NAK 31-2:2022(v1.1) 전자기록물 장기보존패키지 기술규격-제2부: 디렉토리로 구조화된 방식(NEO3)'에서 그 내용을 규정하였다. NEO3는 논리적 장기보존패키지로 설계되었지만 ZIP64 등으로 압축 또는 압축 없이 하나의 디지털 객체로 묶여서 보관되면 실제적 장기보존패키지가 된다.

NEO3에서는 기록물 생산·보존 기관에서 장기보존패키지를 생성할 때, 장기보존 메타데이터가 포함된 파일(Metadata.xml) 및 패키징 정보 파일(PackagingInformation.xml)이 **그림 2**의 사례처럼 최상위 디렉토리에 위치해야 하는 규칙을 제외하고는 기관별 상황에 따라 적용할 수 있도록 장기보존 패키지 내 물리적으로 구성되는 디렉토리의 구조와 명칭은 별도 정의하지 않고 장기보존 메타데이터(Metadata.xml)의 패키징 정보(PackagingInformation.xml)의 내부 경로 정보로 확인할 수 있도록 한다. 디렉토리로 구조화된 방식의 장기보존패키지(NEO3)를 논리적 장기보존패키지로 구현한 사례는 **그림 3**에서 볼 수 있다.

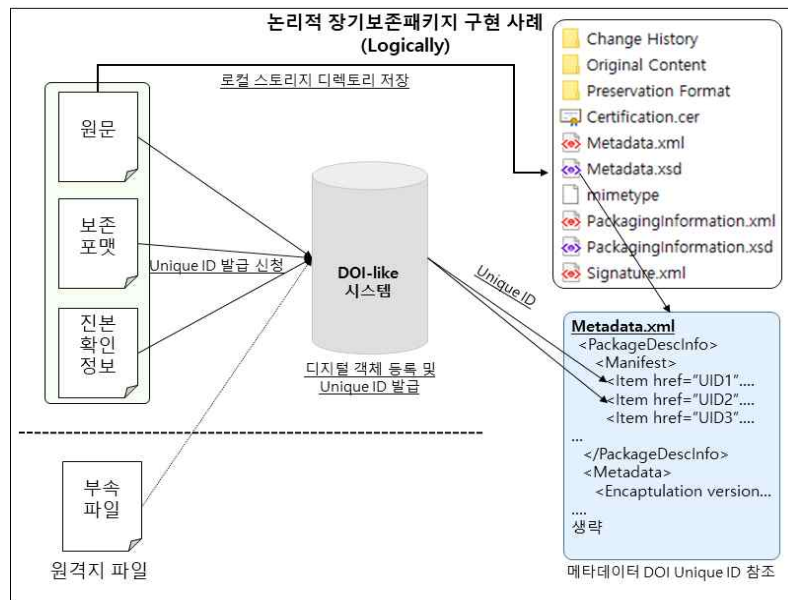


그림 3 - 논리적 장기보존패키지 구현 사례(NEO3)

XML로 포맷화된 방식의 장기보존패키지 또는 아카이브 포맷으로 압축된 디렉토리로 구조화된 방식의 장기보존패키지에 포함되는 컴포넌트 중 일부 항목의 위치를 다른 서버에 있는 저장소 등으로 지정하거나 비전자 문서를 연결하는 경우 실제적 장기보존패키지와 논리적 장기보존패키지의 혼합 방식이 된다.

장기보존패키지를 생성하는 방식은 실제적 장기보존패키지로 설계된 **4.5.2 XML로 포맷화된 방식**과 논리적 장기보존패키지로 설계된 **4.5.3 디렉토리로 구조화된 방식**으로 구분한다. 다만, 디렉토리로 구조화된 방식의 논리적 장기보존패키지가 ZIP64 등으로 압축 또는 압축 없이 하나의 디지털 객체로 묶여서 관리되는 경우에는 실제적 장기보존패키지라고 할 수 있다. 이때, 사용되는 압축 기술은 압축하는 파일의 크기의 제한, 파일의 수, 파일명에 대한 제한 등이 64비트 주소 체계가 지원되는 ZIP64 등 시장표준(De Facto) 기술을 활용하도록 한다.

4.5.2 XML로 포맷화된 방식

이 방식은 전자기록물 및 관련 메타데이터 정보를 단일 XML 파일에 모두 포함한다. 즉, 보존대상인 전자기록물을 하나의 객체처럼 캡슐화하는 방식이다. 다르게 말하면 XML로 포맷화된 방식이라고 할 수 있다. 원문과 보존포

맷이 바이너리(binary) 데이터라 하더라도 base64방식으로 인코딩함으로써 장기보존 메타데이터, 원문 및 보존포맷, 진본확인 정보 등을 텍스트 형식인 단일 XML 문서 내에 객체로 캡슐화한다. 이 방식을 채택한 해외 사례로는 호주 빅토리아 주립 기록보존소의 VEO2, 호주 국가기록원의 XENA 등이 있다.

이 방식은 공공표준 'NAK 31-1:2022(v2.3) 전자기록물 장기보존패키지 기술규격-제1부: XML로 포맷화된 방식(NEO2)'에서 기술규격을 상세하게 정하였다. 영구기록물관리기관 등은 XML로 포맷화된 방식을 채택하는 경우라 하더라도 기관의 환경 등에 따라 NEO2가 아닌 다른 장기보존패키지를 정의하고 적용할 수 있다.

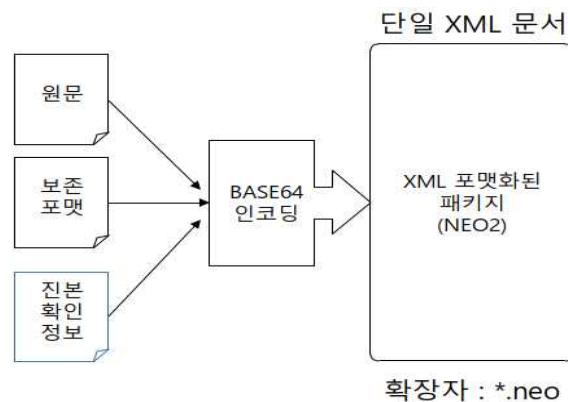


그림 4 - XML로 포맷화된 장기보존패키지

4.5.3 디렉토리로 구조화된 방식

이 유형은 보존대상인 전자기록물을 디렉토리로 구조화하여 캡슐화하는 방식으로 NEO2 패키지 방식과 달리 장기보존 메타데이터, 원문 및 보존포맷, 진본확인 정보를 각각 디렉토리와 하위 디렉토리 등에 구분하여 보관하는 캡슐화 방식이다. 디렉토리로 구조화된 패키지는 ZIP64 등으로 압축 또는 압축 없이 하나의 디지털 객체로 묶여서 보관될 수 있다. 이와 유사한 방식을 채택한 해외 사례로는 미국 의회 도서관의 BagIt, 호주 빅토리아 주립 기록보존소의 VEO3 등이 있다.

이 방식은 본 표준에 기술규격을 상세하게 정하였다. 영구기록물관리기관 등은 디렉토리로 구조화된 방식을 채택하는 경우라 하더라도 기관의 환경 등에 따라 NEO3가 아닌 다른 캡슐화 방식을 정의하고 적용할 수 있다.

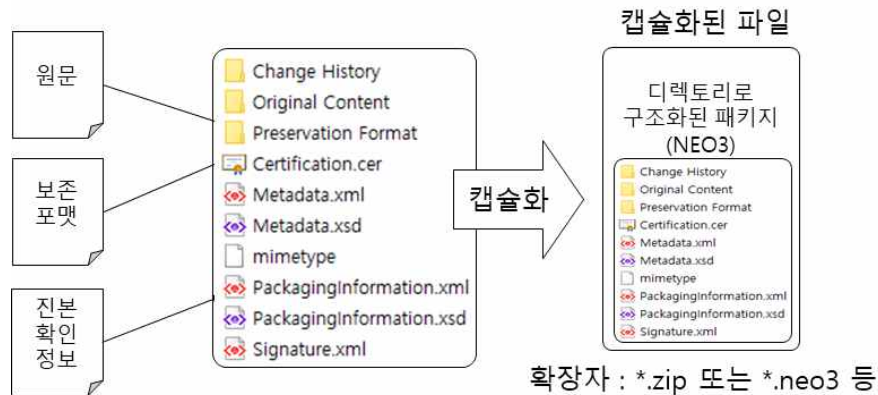


그림 5 - 디렉토리로 구조화된 장기보존패키지

5 XML로 포맷화된 방식의 장기보존패키지(NEO2)

5.1 원문

원문은 생산자가 생산 또는 접수한 기록물 원본(진본)을 뜻하며 업무활동의 증거로서 법적가치를 가진다. 전자기록물의 경우 특성상 다수의 생산자에 의해 시간 및 공간에 제한 없이 여러 개의 복본이 생산 가능하다.

그러므로 전자기록물은 위변조가 쉽게 이루어질 수 있으며, 계속 변화한다. 따라서 진본성 확인을 위해, 적법한 절차에 따라 관리되어 온 것인지에 대한 과정 등이 메타데이터를 통해 관리되어야 한다.

또한, 기록물 생산 당시의 모습이 변하지 않고 그대로 존재한다고 확인할 수 있어야 진본성을 가지고 있다고 할 수 있다. 그러므로 원문을 장기보존패키지에 포함하여 진본성을 보장한다.

5.2 보존포맷

보존포맷은 원문이 생성된 당시의 애플리케이션이 없어도 해당 원문의 내용과 외형을 그대로 재현하여 내용보기를 가능하게 하거나, 원문의 내용과 맥락 정보 등 유형별 중요 정보를 추가적인 타 기술에 의존하지 않거나, 표준

화되고 공개된 방식으로 기술적 복원이 가능하게 하는 포맷이다. 즉, 보존포맷은 기록이 생산되었을 때의 기록물(원문)의 의미, 구조 등 기록의 특성을 유지하며, 하드웨어나 소프트웨어의 노후화나 매체의 퇴화 및 기술발전으로 인한 데이터 포맷의 변화에 상관없이 전자기록물을 보존하여 원본이 손상되어 접근이 불가하거나 어려운 경우, 이용자가 원문과 유사한 전자기록물을 접근하여 활용하기 위한 변환된 포맷이다.

원문을 변환할 적합한 보존포맷이 지정되지 않았거나 에뮬레이션 등 다른 보존방식에 의해 재현이 가능한 경우 장기보존패키지에 보존포맷을 포함하지 않을 수 있다. 원문이 장기보존에 취약한 포맷이며 적합한 보존방식이 없는 경우 영구기록물관리기관 등은 가급적 빠른 시일 내에 보존대안을 마련할 수 있도록 노력하여야 한다.

장기보존패키지에 이미 포함되어 관리되고 있는 보존포맷을 다른 보존포맷으로 대체하거나 병행 관리하고자 하는 경우 행위자, 일시, 사유 등을 장기보존 메타데이터에 추가하여 관리할 수 있어야 한다. 이와 관련한 세부 사항은 관할 영구기록물관리기관이 정한 방식을 따른다.

5.3 장기보존 메타데이터

5.3.1 장기보존 메타데이터 개요

전자기록물의 장기보존을 위해서는 기록물 원문과 함께 기록물의 생산부터 관리, 보존에 이르는 전 과정을 기술(記述)한 정보를 보존해야한다. 그러므로 장기보존 메타데이터의 확보는 필수적이다.

장기보존패키지는 독립된 객체로서 기능할 수 있도록 별도의 메타데이터를 정의하며, 이를 위해 “NAK 8:2022(v2.3) 기록관리 메타데이터 표준”을 반영하여 기록관리 메타데이터와 호환성을 유지한다.

비고 1 장기보존패키지 메타데이터 요소별 자세한 설명은 **부속서 A (규정) NE02 장기보존 메타데이터 스키마**를 참조한다.

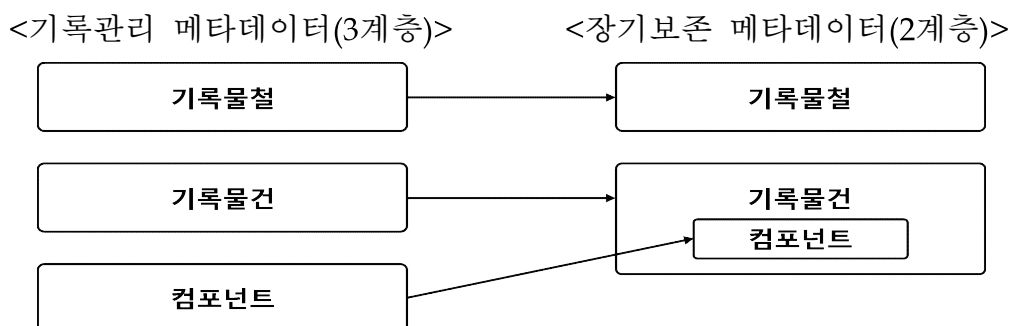
또한, 장기보존 메타데이터는 국가표준 KS X ISO 15489-1과 KS X ISO

23081-1에서 정의한 기록관리 원칙과 메타데이터 요소를 반영하여 기록물 생애주기 전 기간에 걸쳐 진본성, 신뢰성, 무결성을 보장하고, 장기간 기록물에 접근할 수 있도록 지원한다.

장기보존패키지는 기록계층별로 각각의 메타데이터 요소를 정의하여 장기보존과 접근을 보장하는 모든 정보를 제공한다.

비고 2 스키마 설계의 용이성과 안정성을 위하여 사용되지 않는 메타데이터 요소는 실제 삭제하는 대신 더 이상 사용되지 않는 항목으로 변경하여 관리한다. 이러한 항목은 편의상 “사용중지(deprecated)”로 명명하며 해당항목은 **그림6**에서 **그림10**까지 취소선을 추가하여 표시하였다. 다만, 메타데이터 항목명, 선택값 등에서 ‘장기보존패키지’, ‘보존포맷’은 이미 ‘XML로 포맷화된 방식의 장기보존패키지’를 적용하고 있는 기관간 호환을 위해 기존에 사용하던 명칭을 그대로 사용하며, 이 유형의 장기보존패키지를 새로 적용하고자 하는 영구기록물관리기관은 메타데이터 항목명 변경과 관련하여 중앙기록물관리기관과 협의하도록 한다.

기록관리 메타데이터와 장기보존 메타데이터는 적용되는 기록계층에 대해 각각 다음과 같이 매핑된다.



장기보존 메타데이터는 기록계층에 따라 기록물철 장기보존 메타데이터와 기록물건 장기보존 메타데이터로 나뉘고, 장기보존패키지가 수정되는 경우 수정된 장기보존패키지로 구분된다. 각 유형별 메타데이터는 **5.3.2 장기보존 메타데이터 구성**을 따른다.