

전자기록의 장기보존을 위한 보존정보패키지(AIP) 구성과 구조

임진희*

1. 머리말
 - 1.1 연구 목적과 범위
 - 1.2 연구 방법
2. OAIS 정보모델과 AIP
 - 2.1 OAIS 정보모델의 구성 요소
 - 2.2 정보패키지의 유형 - SIP, AIP, DIP
 - 2.3 정보의 기본단위와 컬렉션 - AIU와 AIC
 - 2.4 OAIS 정보모델의 기록관리 적용
3. 기록AIP 구성과 구조
 - 3.1 개념 레이어 - 기록AIP의 구성
 - 3.2 논리 레이어 - 기록AIP의 구조
4. 맺음말

* 한국국가기록연구원 책임연구원

[국문초록]

전자기록 시대를 맞아 여러 유형의 생산시스템에서 만들어지는 다양한 전자기록을 일관된 방식으로 입수하여 보존관리하면서 이용자들이 기록의 존재를 쉽게 검색할 수 있게 하고 원하는 내용을 배부해줄 수 있는 유연하면서도 안정적인 전자기록관리 시스템을 설계하는 것이 필요하게 되었다. 전자기록관리시스템 설계의 가장 기본이 되는 것이 바로 보존관리되는 과정에서의 기록의 구성과 구조라고 할 수 있다. 이 논문은 3-레이어 접근방식에 따라 전자기록의 개념적 구성과 논리, 물리적 구조를 설계하는 과정을 설명하고 있으며, 이 과정에서 3-레이어 접근방식의 유용성을 보여주고 있다.

이 논문은 디지털 정보객체의 장기보존을 위한 참조모델인 OAIS의 정보 모델을 벤치마킹하여 보존패키지(AIP) 개념을 기록에 적용한 ‘기록AIP’의 구성과 구조에 대해 논의한다. 개념 레이어에서는 기록AIP를 구성하는 메타데이터의 그룹을 식별해내는 작업이 핵심이며 특히 OAIS 정보모델의 벤치마킹을 통해 기록정보의 서비스를 위해 필요한 메타데이터 그룹을 기록AIP의 구성에 필수적으로 포함해야 한다는 것을 강조하고 있다. 논리 레이어에서는 보존전략(마이그레이션과 에뮬레이션)이 기록AIP의 구조에 미치는 영향과 기록AIP를 인캡슐레이션할 것인지 여부에 따른 구조에 대해 논의한다. 이 논문은 현재의 수준에서 사용할 수 있는 정보통신 기술과 각종 표준을 기준선으로 하여 기록AIP를 어떻게 구성하고 구조화할 수 있는지 논의를 정리하고 있다.

주제어 : 보존정보패키지, AIP, OAIS 모형, 정보모델, 3-레이어 접근방법, 장기보존, ISO 14721

1. 머리말

1.1 연구 목적과 범위

이 논문의 목적은 전자기록의 장기보존(Long-term Preservation)을 위한 포맷 연구에 있다. 여기서 장기보존이란 미래에도 기록의 재구성이 가능하고 활용이 가능하도록 보존하는 하는 것을 의미하며, 보존포맷이란 컴퓨터로 접근과 활용이 가능한 매체에 기록정보가 저장되는 형식을 의미한다. 여기서 ‘보존포맷’이라는 말은 여러 의미를 내포한다. 먼저, 전자기록이 어떤 메타데이터들과 함께 저장될 것이냐는 구성의 측면과, 전자기록과 메타데이터를 인캡슐레이션(Encapsulation)할 것인지와 어떤 파일포맷으로 저장할 것인지에 대한 물리적 구조의 측면이다.

본 논문은 최근 대통령비서실과 행정자치부에서 구축하는 전자기록관리시스템에 구현되어야 할 전자기록의 저장과 보존형태에 대한 고민에서 시작되었다. 여러 유형의 생산시스템에서 만들어지는 다양한 전자기록을 일관된 방식으로 입수하여 보존관리하면서 이용자들이 기록의 존재를 쉽게 검색할 수 있게 하고 원하는 내용을 배부해줄 수 있는 유연하면서도 안정적인 전자기록관리시스템을 설계하려할 때, 가장 기본이 되는 것이 바로 보존관리되는 과정에서의 기록의 구성과 구조, 그리고 존재형태와 형식이다. 전자기록의 디지털객체로서의 특성과 기록으로서의 특성을 아우르는 보존 포맷에 대한 본 논문은 앞으로 공공영역에서 활발하게 진행될 전자기록관리시스템의 설계에 기초적인 아이디어를 제공하고자 작성되었다.

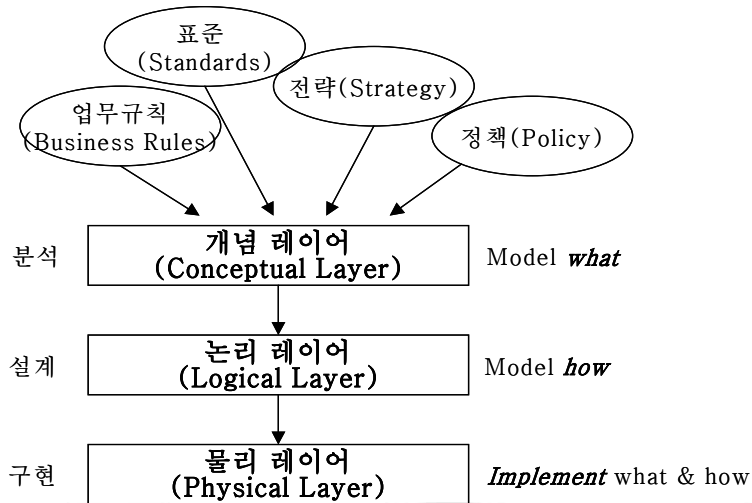


그림1-1. 정보시스템의 개발 단계

*개념 레이어(Conceptual Layer)*는 개념 모델링 단계로서 정보시스템에 갖추어져야 하는 내용을(What) 찾아내는 과정이다. 전자기록관리시스템의 경우, 시스템이 수행할 수 있는 관리 기능이 무엇이며, 관리해야 하는 기록정보가 무엇인지 찾아내어 이름지어주고 그 특성을 파악하여 기능과 정보의 모델을 생성하게 된다. 기록관리의 관점에서 본다면 관리의 대상이 되는 기록을 찾아내고, 그 기록을 구성하는 요소들을 하나하나 정의해주고 세부 속성 값들을 정의해 주는 과정이 될 것이다. 개념 모델링 단계에서는 정보시스템이 어떤 소프트웨어와 하드웨어로 구축이 될 것이냐와는 무관하게 독립적으로 진행이 되는 것으로, 하나의 개념 모델이 완성되면 동일한 모델을 여러 종류의 기술 기반에 적용하는 것이 가능하다.

*논리 레이어(Logical Layer)*는 논리 모델링 단계로서 정보시스템을 구성하는 소프트웨어와 하드웨어, 기술 기반에 대한 선택을 전제로 하여 개념 레

이어에서 찾은 대상을 어떻게(How) 구현할 지를 설계하는 과정이다. 채택한 플랫폼의 특성을 반영하여 설계가 이루어지는데 예를 들어, 기록정보의 저장과 관리를 위해 관계형 데이터베이스를 선택할 것인지, 혹은 계층형 데이터베이스를 선택할 것인지, 혹은 XML 데이터베이스를 선택할 것인지에 따라 저장하여 관리하는 기록의 구조가 결정되게 된다. 또한, 기록정보를 인캡슐레이션할 때 XML 스키마를 활용할 것인지, 아니면 RDF를 채용할 것인지에 따라 구조와 저장포맷이 다르게 된다. 또한, 시스템이 최대한의 성능을 보장할 수 있도록 정보모델에 대한 반정규화 등의 고도화된 설계 기법을 구사하는 단계이기도 하다.

*물리 레이어(Physical Layer)*는 물리적 구현 단계로서 논리 레이어에서 설계한 내용을 주어진 예산에 맞춰 전개하여 구현하는 과정이다. 예를 들어, 이용자들이 자주 접근하여 활용하는 기록들은 비싸지만 접근 속도가 빠른 매체에 두어 서비스 수준을 향상시키도록 한다든지, 필요하다면 기록을 복제하여 한 벌 더 덩으로써 더욱 효율적인 서비스를 할 수 있게 하는 등의 조치를 포함한다. 물리 레이어는 기록정보의 구성과 구조와는 별개로 매체 선택과 분산, 복제 등 물리적 기술(Technology)에 대한 고려가 중심이므로 이 논문의 논의에서는 제외하기로 한다.

3-레이어 접근법의 장점

3-레이어 접근법의 가장 큰 장점은 레이어간의 독립성에 있다. 정보시스템의 기반이 되는 IT 기술이 빠르게 발전해 나감에 따라 그에 상응하여 기록정보의 물리적인 구현도 적절한 시점에 새로운 기술을 적용하며 따라가 주어야 하는데, 그 때마다 정보시스템을 처음부터 다시 만들어 가는 것이 아니라 개념모델은 그대로 두고 논리적 설계부터 또는 물리적 구현만을 바

뛰주면 IT의 발전과정을 빠르게 시스템에 적용할 수 있다. 예를 들어, 전자 기록의 구성요소는 개념적인 수준에서 하나로 모델링이 되지만, 전자 기록을 저장하고 관리하기 위한 전문 데이터베이스관리시스템을 현재 전 세계적으로 가장 많이 사용되고 있는 안정적인 관계형 데이터베이스관리시스템을 사용할 것인지 또는 XML 문서를 분해하지 않고 그대로 저장하여 관리할 수 있는 XML nature 데이터베이스관리시스템을 사용할 것인지에 따라서 전자 기록의 논리적인 설계는 달라진다. 관계형 데이터베이스의 경우에는 전자 기록을 구성하는 개념적인 요소들이 각각 테이블로 쪼개져서 저장되는 것으로 설계될 것이며, 인캡슐레이션은 개념적으로만 존재하고 설계 시에는 인캡슐레이션의 개념이 충실하게 보장될 수 있도록 하는 설계 방식을 취하게 된다. 하지만, XML 데이터베이스를 선택하는 경우에는 전자 기록이 하나의 XML 문서로 인캡슐레이션되어 물리적인 하나의 덩어리로 저장되는 것으로 설계될 수 있어서, 개념 레이어의 인캡슐레이션이 논리층에 변형없이 그대로 적용될 수 있다. 또한, 관계형 데이터베이스에서는 테이블, 칼럼, 기본키, 외래키 등의 개념으로 정보를 정의하고 연결하여 하나의 의미 실체를 담게 되지만 XML 데이터베이스는 전혀 다른 용어와 개념으로 정보를 정의하게 된다.

즉, 개념 레이어에서 정의한 내용은 동일할 수 있지만 논리 레이어에서는 어떤 기술기반의 시스템으로 갈 것이냐를 전제로 하여 매우 용어를 사용하여 다른 설계가 진행될 수 있는 것이다. 만약 개념 모델은 그대로 유지하면서 기술 기반만 새로운 플랫폼으로 변경하기를 원한다면 이 때에는 분석단계를 반복하지 않고 논리 모델을 새로 설계하는 것부터 진행할 수 있다는 것이 3-레이어를 분리하여 시스템 구축을 하는 최대의 장점인 것이다.

OAIS 모델은 3-레이어 접근방식에서의 최상위인 개념 레이어에서 기술

된 것으로 이러한 OAIS 모델을 기록관리에 적용한다는 것은 바로 개념 레이어에서 기록관리에 필요한 부분을 섭취한다는 것을 의미한다. 결국 논리 레이어와 물리 레이어는 기록관리의 특수한 환경과 조건에 따라 다양하게 설계하고 구축하는 것이 가능하게 된다. 그런 면에서 OAIS 모델은 기록뿐만 아니라 모든 유형의 전자객체의 장기보존을 위한 모델로서 그 적용 범위가 매우 넓은 모델이라 할 수 있으며, 이러한 적용성은 모델의 기술 수준에 개념 레이어에서 이루어졌기 때문이다.

2. OAIS 정보모델과 AIP

OAIS(Open Archival Information System) 모델은 디지털 객체의 *장기보존(Long-term Preservation)*을 위해 만들어진 참조모델(Reference Model)로서 이미 국내에는 ISO14721 표준과 함께 논문“디지털 아카이빙의 표준화와 OAIS 참조모형”(이소연, 2002)을 통해 소개가 시작되었으며, OAIS 모델을 적용한 외국 사례들도 국내 연구자들에 의해 소개가 된 바 있다. 이 논문에서는 정보모델을 중심으로 OAIS를 소개하고, 소개된 개념 모델들을 전자기록의 시스템 구현에 실용적으로 적용해 보고자 한다.

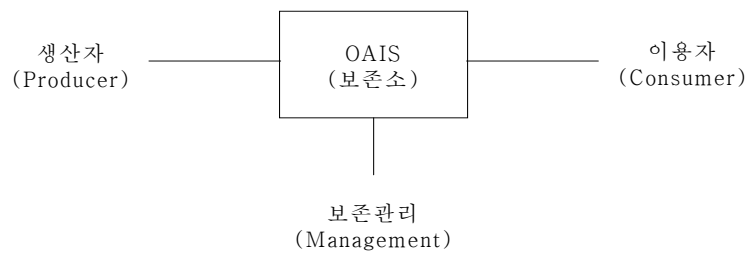


그림2-1. OAIS의 외부환경 모델

OAIS 참조모델은 정보를 생산하여 보급하는 생산자(Producer)와 그 정보를 접근 활용하려는 이용자(Consumer)를 위해 전략과 계획에 의거하여 정보를 보존, 관리하는(Management) 주체들을 위한 것이다. OAIS 참조 모델은 크게 기능 모델과 정보 모델로 구성되어 있다. 기능 모델(Function Model)은 디지털 정보의 장기보존을 위해 필요한 기능을 ①정보의 입수, ②저장, ③데이터유지관리, ④운영관리, ⑤보존계획, ⑥접근 및 이용 등의 6개의 대기능으로 나누어 설명하고 있다. 정보모델(Information Model)은 객체지향 모델링 언어인 UML²⁾을 사용하여 정보객체(Information Object), 정보패키지(Information Package), 정보컬렉션(Information Collection)이라는 개념을 통해 정보의 계층적인 구조를 설명하고 있다. 이러한 OAIS 모델은 Open이라는 말이 의미하듯이 모든 유형의 디지털 객체를 보존하기 위한 개방적이고 확장가능한 모델로서 디지털 정보의 장기보존 모델 표준 ISO 14721 로 채택될 만큼 매우 유용한 시사점을 주고 있다.

이 논문에서는 OAIS의 정보모델을 중심으로 전자기록의 장기보존을 위한 개념적 구성과 논리적 구조를 도출해 보고자 한다. 먼저 OAIS의 정보모델을 요약하면 다음과 같다.

2.1 OAIS 정보모델의 구성 요소

OAIS 정보모델의 목적은 정보를 효과적으로 장기보존하기 위해 필요한 요소들을 모형화하는 것이다. 정보모델을 구성하는 각 요소들은

2) UML은 Unified Modeling Language의 약자로, 객체지향 방법론에서 기능과 정보를 모델링할 때 사용하는 언어이다.

OAIS의 기능에 의해 교환되거나 관리되면서 구성의 변화가 이루어질 수 있으며, 정보모델은 기능에 의한 모든 변화 가능성을 포괄하는 요소들로 구성되어야 한다.

정보객체 모델

OAIS 모델에서 정보객체는 데이터객체(Data Object)와 표현정보(Representation Information)가 조합된 개념이다.

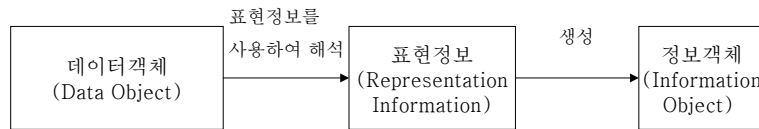


그림2-2. 데이터로부터 정보를 구성하는 과정

위의 그림은 전자적인 형태로 생산된 데이터객체가 OAIS에서 보존, 관리할 수 있는 대상으로서의 정보객체로 구성되는 과정을 보여준다. 여기서 데이터객체란 컴퓨터 상에 입력되어 있는 바이너리 데이터 자체를 말한다. 표현정보는 데이터객체를 수신하여 활용하고자 하는 사람 또는 컴퓨터가 해당 데이터를 이해할 수 있도록 해주는 정보를 말한다. 정보객체는 표현정보를 사용하여 해석된 데이터로서 컴퓨터의 화면을 통해 보여지는 데이터를 말한다. 일반적으로 모든 데이터들은 표현정보를 사용해서 해석되고 정보를 생성하게 된다.

표현정보의 예

예를 들어, 우리가 메모장에 'ABC123' 이라는 문자를 입력하고 텍스트로 저장을 하면 .txt 파일이 생기게 된다. 이때, 이 텍스트 파일은

‘ABC123’ 라는 문자들을 컴퓨터가 알아볼 수 있는 바이너리 값의 조합, 즉 0과 1의 조합인 비트스트림으로 만들어지게 된다. 이 때, 우리가 메모장에 쓴 ‘ABC123’ 이라는 문자 정보를 컴퓨터가 바이너리 데이터로 변환하여 저장할 때는 여러가지 방법이 있을 수 있으며, 가장 많이 사용되는 방법이 문자를 각각의 ASCII 코드 값으로 변환하여 저장하는 방법이다. ‘ABC123’ 이라는 문자가 ASCII 코드를 사용하여 데이터 객체화 되면 그 바이너리 비트스트림은 ‘01000001 01000010 01000011 00110001 00110010 00110011’ 가 될 것이며 총 6바이트로 구성될 것이다.

그렇다면, 만약 컴퓨터에 저장되어 있는 비트스트림이 ‘01000001 01000010 01000011 00110001 00110010 00110011’ 라고 했을 때, 아무런 해석 정보가 없는 상태라면 이것을 어떤 의미를 가진 정보로 볼 것이냐를 생각해 보면 답이 없음을 알게 된다. 이 비트스트림은 ASCII 코드에 의한 문자 값으로 저장된 것임을 알아야 해석이 가능하게 되는 것이다. 비트스트림으로 저장된 바이너리 데이터객체 자체는 그것이 어떤 정보인지를 해석하기 위한 기본적인 표현정보가 함께 있어야만 정보로서의 가치를 갖게 된다.



그림2-3. 비트스트림이 ASCII 코드라는 표현정보를 통해 해석되어 보여지는 과정

위의 그림에서 “ASCII 코드 문자열로 해석” 이라는 정보가 표현정보로서, “01000...”라는 비트스트림을 해석하여 ‘ABC123’이라는 의미있는 정보로 재구성하도록 해주는 것이다.

또 다른 예로, 한글이나 마이크로소프트 워드와 같은 문서편집기를 이용하여 문자들을 입력하고 저장하게 되면, 우리가 입력한 데이터들이 한글이나 마이크로소프트 워드의 자체방식에 의해 일정한 규칙을 가지고 비트스트림으로 변환되어 컴퓨터에 저장되는데 이때는 단순한 문자 정보 외에도 문자의 폰트와 문서의 형식 등에 대한 부가적인 정보들이 함께 저장되게 된다. 이처럼, 한글이나 MS 워드 등은 각자의 방식으로 비트스트림을 생성하여 저장하므로 이런 전자파일에 대해서는 어떤 소프트웨어를 사용하여 작성된 파일인지를 주요 표현정보를 작성해 주어야 하며, 가능하다면 해당 소프트웨어가 어떤 형식으로 비트스트림을 구성하는 지에 대한 정보까지를 표현정보에 포함해 준다.

정보패키지 모델

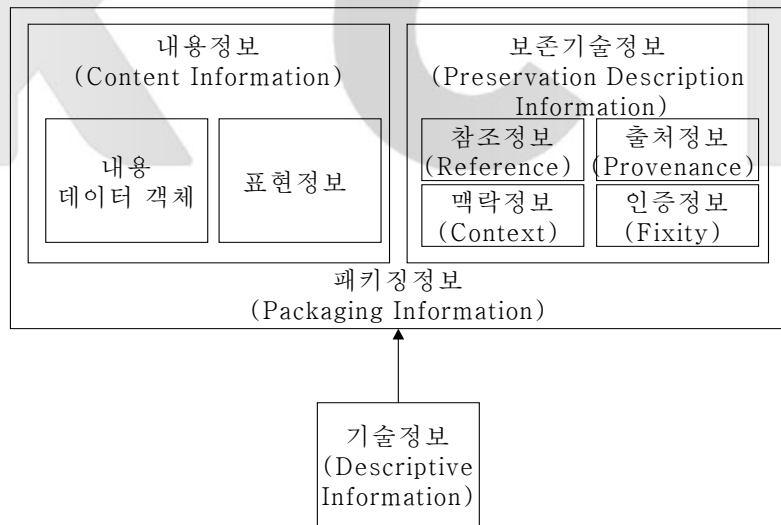


그림2-4. 정보패키지의 구성 개념도

OAIS 정보모델에서는 정보객체 개념을 사용하여 입수와 보존, 관리, 접근, 활용의 기본 단위를 형성하는 정보패키지(Information Package)를 정의하고 있다.

여기서 정보패키지는 내용정보(Content Information)과 보존기술정보(Preservation Description Information), 패키징정보(Packaging Information)가 합쳐진 개념적 구조체이다. 그림2-5에서 표현하고 있듯이 내용정보, 보존기술정보, 패키징정보, 기술정보 모두 앞에서 살펴본 정보객체 종류로 각각 데이터객체와 표현정보로 구성되며, OAIS에 저장되는 모든 정보들은 최소의 의미있는 기본단위로서 정보객체의 유형을 갖게된다.

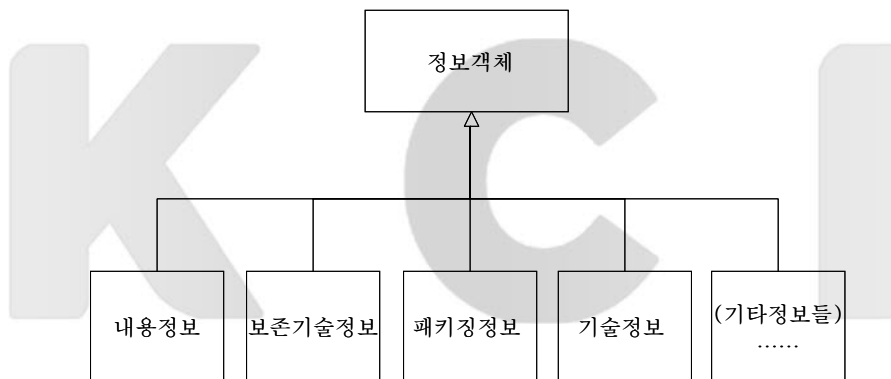


그림2-5. 정보객체의 종류

그림2-4에서, 내용정보와 보존기술정보는 하나의 캡슐로 싸인 것으로 보며³⁾ 패키징정보를 통해서 캡슐 안의 구성을 확인할 수 있다.⁴⁾ 이렇게 패키징된 전체가 OAIS에 저장되고 보존되며, 한편으로는 OAIS내

3) 여기서 말하는 캡슐은 개념적인 것으로, 내용정보와 보존기술정보는 떨어져서는 안되는 통합된 일체로서 하나의 패키지를 구성한다는 것을 의미한다.

4) 패키징정보는 패키지에 대한 메타데이터로서의 성격을 갖는다.

에 존재하는 정보패키지를 찾아내기 위해 필요한 기술정보(Descriptive Information)⁵⁾를 정의하게 되는데, 이 기술정보는 정보패키지의 보존기술정보와 패키징정보에서 추출된다.

표2-1. 정보패키지의 구성 객체

정보객체	내 용	
내용정보	<ul style="list-style-type: none"> ● 보존해야할 정보 집합으로 표현 정보를 동반하는 내용 데이터 객체 ● 생산자와의 협상을 통해 내용정보를 결정하게 됨 	
보존기술정보	출처정보	내용정보의 이력 정보로 과거와 현재 상태를 기술 (기원, 원천, 관리변화 등)
	맥락정보	내용정보와 외부환경과의 관계 정보 (생산배경, 다른 정보와의 관계 등)
	참조정보	내용정보에 대한 식별자 (식별자 지정 매커니즘은 분류시스템 등)
	인증정보	문서화되지 않은 방법으로 내용정보가 변형되지 않았음을 보장해주는 정보
패키징정보	<ul style="list-style-type: none"> ● 내용정보와 보존기술정보 검색에 필요한 정보내용정보와 보존기술정보 검색에 필요한 정보 	
기술정보	<ul style="list-style-type: none"> ● 원하는 내용정보가 어떤 패키지에 담겨 있는지 검색하는데 이용되는 정보 ● 단순히 표제 정보일 수도 있고, 목록서비스를 위한 속성값 전체일 수도 있음 	

내용정보는 보존의 목표 대상인 정보객체로서, 앞에서 설명한 대로 내용의 데이터객체와 데이터 객체를 이해할 수 있게 해주는 표현정보가 연결

5) 전자객체의 경우 정보시스템을 떠나서는 존재할 수 없으므로, 전자객체를 활용하고자 할 때 원하는 전자객체를 검색할 수 있도록 목록이나 인덱스, 키워드 등을 제공하는 것은 기본이라 할 수 있다.

되어 구성된다. 전자기록의 경우, 본문이나 첨부 파일 자체(또는 비트스트림)가 내용데이터객체가 되고, 각 파일(또는 비트스트림)이 어떤 형식으로 데이터를 표현하고 있는가에 대한 정보, 즉 한글 파일이라거나 엑셀 파일이라는 등의 생산 소프트웨어 정보로부터 한글과 엑셀이 화면상에 보여지는 데이터를 내부적인 비트스트림으로 변형 저장할 때의 방식에 대한 정보 모두가 표현정보가 된다.

*보존기술정보*는 내용정보를 보존하기 위해 필요하고, 내용정보를 명확하게 식별해 내고, 내용정보가 생산된 배경을 이해하기 위한 정보이며 출처정보(Provenance), 맥락정보(Context), 참조정보(Reference), 인증정보(Fixity)의 네 가지 유형으로 구분된다. 출처정보는 내용정보의 소스 또는 원생산자를 나타내며 원본 생산에서부터 관리권의 변화와 이력(처리 이력 포함) 내역을 포함한다. 맥락정보는 내용정보가 정보패키지 외부에 있는 다른 정보와 어떤 관계를 가지고 있는지를 나타내는 것으로 예를 들면, 내용정보가 생산된 이유에 관한 기술이나 다른 내용정보와 어떤 관련이 있는지에 관한 정보 등이다. 참조정보는 내용정보를 식별할 수 있는 이상의 역할을 하는 참조코드들로서 시스템 id, 서비스 id, ISBN, URL 정보, 핸들 등이 여기에 속한다. 인증정보는 내용정보를 허가되지 않은 변경으로부터 보호하기 위한 정보로서 체크섬이나 CRC와 같은 값들로 구성된다.⁶⁾

6) 체크섬은 수신자가 같은 수의 비트가 도착했는지를 확인 할 수 있도록 전송단 위 내의 비트 수를 세는 것이다. 만약 계산이 맞으면, 오류없이 원만하게 수신된 것으로 간주된다. CRC(Cyclic Redundancy Checksum 또는 Cyclic Redundancy Check)는 데이터의 무결성을 체크하는데 사용되는 대표적인 “디지털 인증”방법으로, 예를 들어, 전송이 이루어졌을 때 전송전의 파일과 전송 후의 파일이 서로 동일한 내용인지 확인하고자 할 때 원래 파일에 대한 CRC값을 구하고 전송된 파일의 CRC값을 구해서 두 개의 값이 같으면 두 파일의 내용이 동일하다고 추정할 수 있는 인증방법이다.

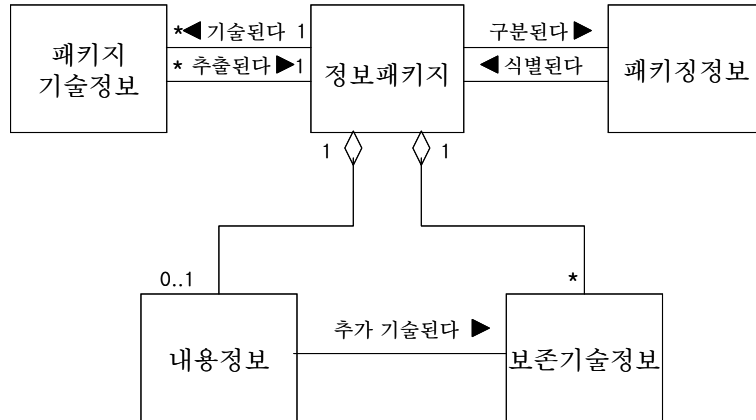


그림2-6. 정보패키지와 연결 정보

패키징정보는 내용정보와 보존기술정보의 실제적인 혹은 논리적인 결합관계를 나타내는 정보이다. 예를 들어, 어떤 패키지의 내용정보과 보존기술정보가 CD-ROM에 각각 파일로 저장되어 있을 때 패키징정보는 “ISO 9660 volume/파일구조(=파일디렉토리와 파일명)”⁷⁾의 형태로 CD-ROM 안에 내용정보와 보존기술정보 파일 저장규격과 위치 및 이름 정보로 표시될 것이다.

기술정보는 원하는 내용정보를 담고 있는 패키지가 어떤 것인지 찾아내는 데 사용하는 정보로 정보패키지의 표제 정보만 가지고 있을 수도 있고 목록서비스에서 검색할 수 있도록 전체 메타데이터의 집합일 수도 있다. 기술정보는 일반적으로 보존기술정보의 일부와 내용정보의 핵심어를 추출하여 작성하며, 사용자가 원하는 패키지를 찾는데 유용한 색인정보 역

7) ISO 9660은 1988년도에 국제표준화기구가 제정한 CD-ROM 을 위한 파일 시스템을 정의한 표준규격으로 마이크로소프트의 윈도우즈, 매킨토시의 OS, 유닉스 등 서로 다른 시스템에서 데이터를 교환할 수 있도록 하기 위해 제정되었다.

할을 하게 된다.

2.2 정보패키지의 유형 - SIP, AIP, DIP

OAIS 참조모델에서는 정보패키지의 유형을 제출정보패키지(SIP: Submission Information Package), 보존정보패키지(AIP: Archival Information Package), 배부정보패키지(DIP: Dissemination Information Package)의 세 가지로 구분하고 있다.

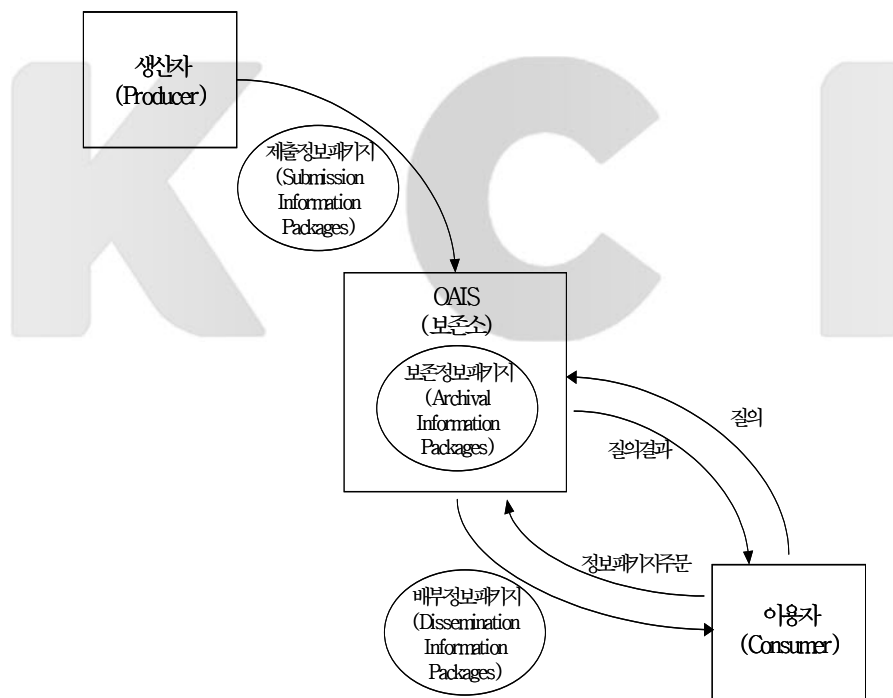


그림2-7. 정보패키지의 용도별 유형

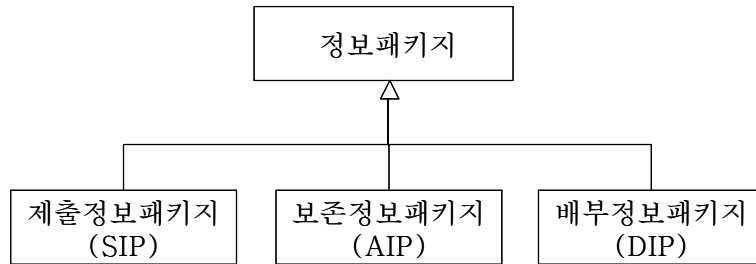


그림2-8. 정보패키지의 유형

SIP는 생산자(Producer)가 OAIS로 보내는 정보패키지로 생산자와 OAIS 간의 협약을 통해서 내용과 형태가 결정된다. 하나의 SIP가 하나의 AIP로 구성되는 경우가 대부분이나 여러 개의 SIP가 하나의 완전한 내용정보를 구성하고 보존기술정보를 연결하여 AIP 하나로 구성될 수도 있다. 또한, 하나의 SIP안에 여러 AIP의 정보가 들어있을 수도 있다. 어떠한 경우에도, SIP는 일정한 형태, 즉 개념적으로 동일한 구조를 가지고 있다.

AIP는 내용정보와 이에 대한 보존기술정보의 완전한 집합으로 만들어지며, 패키징정보는 OAIS가 정한 내부 표준을 준수한다. AIP는 다른 여러 개의 AIP의 집합으로 구성되기도 하며, OAIS로 입수된 하나 이상의 SIP가 하나 이상의 AIP로 변형되어 저장된다.

DIP는 AIP의 일부 또는 전부를 이용자(Consumer)에게 제공하기 위해 만들어진 정보패키지로, AIP의 집합을 포함하여 만들어질 수도 있고, 때로는 보존기술정보를 포함하지 않고 만들어질 수도 있다. 패키징정보는 일정한 형식으로 표현되어야 이용자(Consumer)가 요청한 정보를 구분하여 볼 수가 있으며, 이용자의 요구사항이나 DIP를 전달하는 매체 종류에 따라 다양한 형태를 갖게 된다.

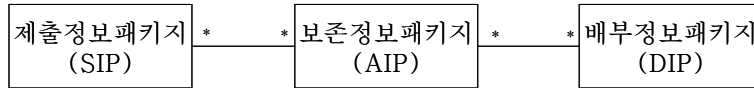


그림2-9. SIP, AIP, DIP의 대응 관계

표2-2. SIP, AIP, DIP의 내용과 특징

정보 패키지	내 용	비 고
SIP	<ul style="list-style-type: none"> ● 생산자가 OAIS로 보내는 정보패키지 ● 형태와 내용은 OAIS와 생산자 간의 협약에 따름 ● 내용정보와 보존기술정보로 구성되며 표현정보와 연결됨 	<ul style="list-style-type: none"> ● 다수 SIP가 동일한 표현정보를 이용하면 표현정보는 한번만 제공함 ● 다수 SIP에 동일한 보존기술정보가 적용되면 내용정보 없는 SIP 형태로 제공함 ● 기술정보는 SIP 제출시점에 관계없이 제공됨
AIP	<ul style="list-style-type: none"> ● SIP를 기반으로 OAIS에 보존하기 위해 변형된 정보패키지 ● 내용정보와 보존기술정보의 묶음, 혹은 AIP들의 집합 	<ul style="list-style-type: none"> ● 패키징 정보는 내부의 표준을 따름 ● AIP 기술정보를 이용해 이용자가 원하는 내용정보를 찾고 주문 가능하도록 함
DIP	<ul style="list-style-type: none"> ● 이용자의 주문에 대해 AIP의 전체 또는 일부를 제공하는 정보패키지 ● AIP들의 집합이거나 보존기술정보 없는 AIP 등 	<ul style="list-style-type: none"> ● 배포매체와 이용자요구에 따라 패키징정보가 다양함 ● DIP 기술정보는 DIP 전송 이전, 이후, 전송 시 제공 가능

AIP의 정보요소

정보패키지로서의 AIP는 보존시스템에 실재하면서 보존되는 대상으로서 장기적인 안정성을 확보할 수 있는 구성을 갖추어야 하며, 한편으로는 이용자들의 이용을 용이하게 해줄 수 있는 구성을 갖추어야 한다. 후자의

측면을 고려하여 앞에서 살펴본 일반적인 정보패키지의 구성요소에 연결 요소를 추가하여 AIP의 구성을 정의해 보면 다음과 같다.

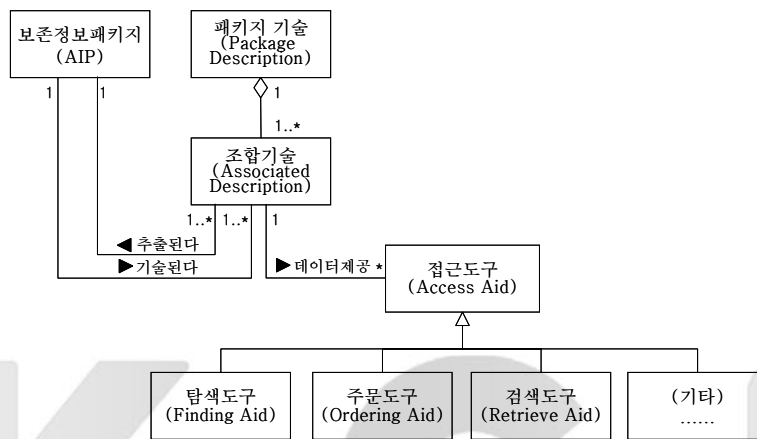


그림2-10. AIP의 연결요소 정보 구성

표2-3. AIP의 정보 요소

구성 요소	내용정보	● 보존해야할 내용 데이터 객체와 표현정보
연결 요소	보존기술정보	● 출처정보, 맥락정보, 참조정보, 인증정보로 구성
	패키징정보	● AIP를 구성하는 하위요소를 한정하고 식별해주는 정보
연결 요소	기술정보 ⁸⁾	● 이용자가 AIP를 탐색, 분석, 주문을 하는데 필요한 정보
		● 접근도구(Access Aid) ⁹⁾ 를 위해 필요한 수만큼 조합기술 (Associated Description) 정보를 구성
		● 각 조합기술(Associated Description)은 하나 이상의 접근도구를 지원하는 데이터를 포함할 수 있도록 구성

8) 그림2-10의 패키지기술에 해당

9) 접근도구(Access Aid)에는 탐색도구(Finding Aid), 검색도구(Retrieve Aid), 주문도구(Ordering Aid) 등이 있다.

구성요소의 내용정보는 SIP와 DIP 모두 동일하게 필요로 하는 요소이나, 구성요소 중 보존기술정보와 연결요소는 정보패키지 유형마다 필요한 요소가 다르게 된다. 예를 들어, 연결요소 중 기술정보의 경우는 SIP와 DIP에는 필요없는 요소이며, 패키징정보의 경우는 SIP와 DIP에도 필요한 요소이나 정보를 기술하는 방식이나 수준은 매우 다를 것이다.

2.3 정보의 기본단위와 컬렉션 - AIU와 AIC

OAIS에 저장 보관되는 AIP는 기본단위인 AIU(Archival Information Unit)와 기본단위들의 집합체인 AIC(Archival Information Collection)라는 두 가지 유형으로 분화된다.

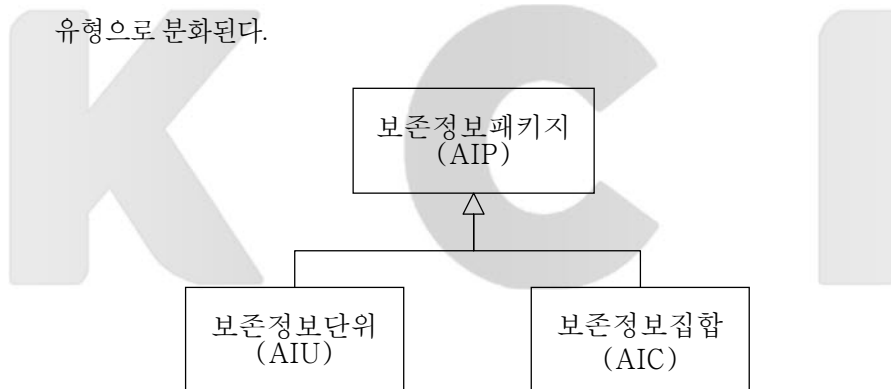


그림2-11. AIP의 분화 유형 - AIU와 AIC

AIU란 보존소에서 저장하는 정보의 기본단위로 하나의 내용정보와 보존기술정보로 구성된다.

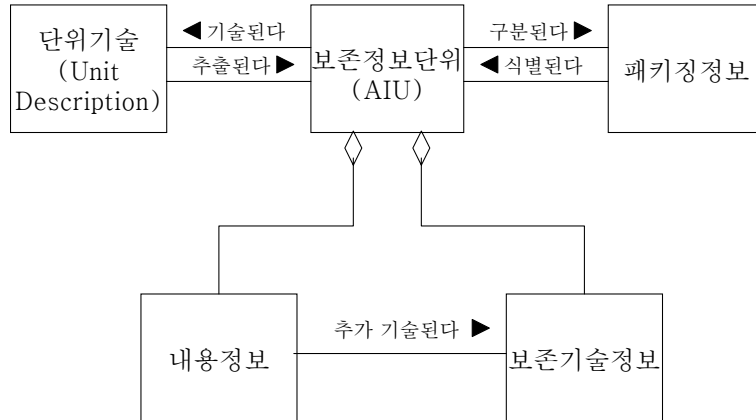


그림2-12. AIU의 구성요소 모형

이 때, AIU를 구성하는 단위기술(Unit Description)은 정보객체가 보존소로 입수될 때 내용정보와 보존기술정보로부터 추출되고 OAIS에 의해 유일 식별자 등과 같은 정보가 부가되어 생성된다. 이 단위기술 정보는 인가된 이용자가 저장되어 있는 AIU를 검색할 수 있도록 제공되어야 한다.

AIC란 보존관리자가 정한 범주에 따라 모인 AIP들의 집합을 말한다. OAIS는 전체 시스템 안에 최소한 하나 이상의 AIC를 갖게 되며, AIC는 그 자체가 내용정보와 보존기술정보를 가진 완전한 AIP이다. OAIS에는 수많은 정보객체가 존재하므로 이용자가 원하는 정보를 찾기 위해 AIU와 해당 단위기술정보를 검색하는 일은 매우 힘든 일이다. AIC는 동일한 범주의 AIU들을 묶어서 하나의 집합을 정의하고 그 집합의 출처, 맥락, 참조, 인증 정보를 기술하도록 하여 집합적 기술과 활용을 가능하게 한다.

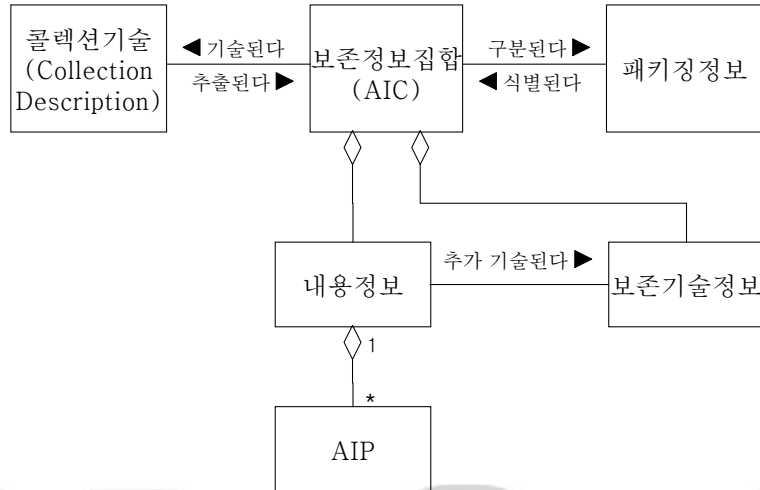


그림2-13. AIC의 구성 모형

AIC의 기술정보, 즉 콜렉션기술은 전체 콜렉션을 기술해주는 내용과 콜렉션의 멤버 각각을 기술해주는 0개 이상의 멤버기술을 포함한다. 즉, 해당 AIC에 속하는 AIP들에 대한 기술 내용이 포함될 수 있다.

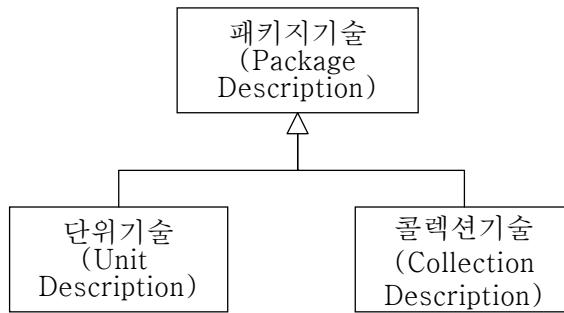


그림2-14. 기술정보의 분화 유형

2.4 OAIS 정보모델의 기록관리 적용

OAIS의 정보 모델은 어떤 유형의 정보이든 보존을 위한 시스템으로 입수될 때, 보존이 될 때, 배부가 될 때 등 어느 시점에서든지 동일한 정보패키지의 개념으로 구조화하여 취급하는 단순하고도 융통성 있는 모델이다. 전자기록 역시 OAIS의 정보 모델을 이용하여 구조화가 가능하며 특히 영구기록 보존 단계의 기록관리시스템의 기록 정보 모델에 참조될 만하다. 그러나, 기록관리가 전자기록 중심인 시점이 되면서부터는 전자기록이 영구보존 단계의 시스템에 이관되기 전부터 일관되게 영구 보존 포맷으로 저장, 보관되는 것이 전자기록의 안정성과 진본성 보장에 도움이 될 수 있다는 점에서 OAIS의 정보모델은 영구보존 단계 이전의 기록관리시스템에도 적용이 가능한 참조모델이라 할 수 있겠다.¹⁰⁾

OAIS 모델의 정보패키지는 물리적인 구조체가 아니라 개념적인 구조체로 이해된다. 개념 레이어에서 정의한 정보패키지를 논리, 물리 레이어에서 어떻게 구조화하고 설계할 것인지는 구체적인 상황에 따라 다양한 적용이 가능한 것이다. 즉, OAIS 정보모델에서 정의한 하나의 정보패키지는 물리적으로 하나의 파일로 저장될 수도 있고 여러 개의 파일로 나뉘어 저장될 수도 있는 것이다. 3장에서는 OAIS 보존정보패키지 개념을 참조하여 먼저 기록의 AIP의 개념적 구성을 논의하고, 주요한 보존전략에 따라 기록 AIP의 논리적 구조가 어떻게 달라질 수 있

10) 행정자치부에서 수행한 기록관리시스템 혁신을 위한 ISP 프로젝트 결과보고에 따르면, 기록관 단계에서 사용할 기록관리시스템에서 5년 이상의 보존기간을 가진 전자기록물을 입수했을 때 문서보존포맷 PDF/A를 적용하여 문서단위 포맷전환을 하고, 기록 건 단위에서 하나의 XML 문서로 인캡슐레이션하여 저장하는 것을 제안하고 있다.

는지에 대해 논의할 것이다.

현재의 OAIS 정보 모델이 기록관리에 곧바로 적용되는 데 한 가지 아쉬운 점이 있다면 그것은 OAIS 참조모델에서는 기록관리의 기본 틀인 분류 체계에 대응하는 명시적인 개념이 없다는 것이다. 기록의 경우 해당 기록이 어떤 조직에서 또는 어떤 업무기능의 맥락에서 생산, 접수되었는지를 분류체계에 반영하여 이에 따라 기록을 분류, 등록, 기술하는 것이 관리의 첫 걸음인데, OAIS에서는 기록의 분류체계가 갖고 있는 계층성과 계층별 기술을 가능하게 하는 AIU와 AIC에 대한 개념만 제시하고 있을 뿐이다. 그러므로, 기록관리에서 구체적으로 필요한 기록철(File), 시리즈(Series), 레코드 그룹 등의 개념은 OAIS 참조모델의 AIC를 창조적으로 적용 구현해 나가야 할 부분이다. 물론, OAIS의 정보모델이 기록관리의 분류체계를 구현하는 데 용이하지 않다면 이 모델이 전자기록의 장기보존 측면에서 주는 아이디어만을 기록관리에 차용하고 분류체계 등의 구성에 대해서는 다른 모델을 만들어 조합해 나갈 수도 있겠다.

이 논문에서는 기록철, 시리즈 등의 상위개념은 향후의 과제로 남겨두고 전자기록 건 하나에 대응하는 개념으로서의 기록AIP에 대한 구성과 구조를 장기보존의 관점에서 어떻게 만들어갈 것인지에 대해 OAIS의 정보모델을 적용해 보고자 한다.

3. 기록AIP 구성과 구조

OAIS가 보존, 관리 대상으로 하는 정보객체는 일반 전자문서, 전자저널,

과학기술 데이터세트, 웹페이지, 시청각 디지털 저작 등 디지털화된 모든 유형이다. 이 논문에서는 OAIS가 포괄하는 정보객체의 유형 중에서 업무 수행과정에서 생산 접수된 전자기록이라는 정보객체에 초점을 맞춰 논의를 진행하고자 한다. 또한, 전자기록을 안정적으로 장기간 보존하기 위해서는 기록의 AIP를 어떤 구성과 구조로 보존 관리해야 하는가에 초점을 맞춰 논의를 진행하고자 한다.

논의에 앞서 한 가지 전제를 확인하면, 전자기록의 장기보존을 위한 포맷은 전자기록을 보존하는 주체에 의한 *보존전략 및 계획(Preservation Strategy & Planning)* 수립이 선행되어야 논의가 가능하며 내용이 정해질 수 있다는 것이다. 보존전략 및 계획 수립 활동은 기록의 생산자와 이용자 양쪽에 대해 현재 사용하고 있는 컴퓨터 플랫폼과 선호하는 소프트웨어 등에 대해 모니터링하고, 신규 디지털 기술에 대한 동향을 살펴면서 현재 보존 중인 기록의 마이그레이션 계획을 수립하고 새로운 포맷으로 생산되어 등록되는 기록에 대한 보존 전략을 수립하며, 표준을 개발하고 신규 AIP를 설계하는 등, 기록보존시스템에 저장된 기록은 장기간에 걸쳐 안정적으로 이용될 수 있다는 것을 보증하기 위한 활동으로 이루어진다. 이러한 전략과 계획에 의해 전자문서로 생산 또는 접수된 기록을 등록하여 관리할 때 본문 파일의 포맷을 그대로 보존할 것인지, 아니면 표준으로 정한 보존 포맷으로 전환하여 보존할 것인지가 정해지며, 원래의 본문파일과 보존포맷으로 전환된 파일을 둘 다 보존할 것인지의 여부도 정해진다. 최초의 전자 기록 AIP는 이러한 전략과 계획에 의해 그 구성과 구조가 결정되게 되며 또한 전략과 계획에 의해 보존 관리되는 과정에서 그 구성과 구조가 변형될 수 있다.

이러한 보존전략 및 계획의 수립은, 기록의 이용자들이 누구이며, 어떤

형태로 기록정보를 서비스 받고자 하는가, 이러한 이용자의 요구에 대해 기록보존관리자는 어느 수준에서 서비스를 제공하기로 결정했는가 등에 따라 정해지게 된다. 기록관리 조직도 일반적인 조직들과 마찬가지로 한정된 자원과 예산으로 조직과 시스템을 구축하고 운영해야 하므로 비용 대비 효과를 고려하여 서비스 수준을 결정하고 시스템의 성능을 보완해 가야 한다. 이러한 현실적인 한계들이 AIP의 구현 양상에 중요한 영향을 미치게 된다.

OAIS는 기록을 포함한 다양한 종류의 디지털 정보를 장기보존하기 위한 모델로 제시된 것이므로 기록정보에 대한 특별한 고려사항을 담고 있지는 않다. 이 논문에서는 기록에 적용된 AIP를 ‘기록AIP’라고 부르기로 하고 기록AIP의 개념적 구성과 논리적 구조에 대해 3-레이어 접근법에 따라 논의해 보기로 한다. 이때의 기록AIP는 기록 자체의 구성요소 뿐만 아니라 그 기록이 보존되는 이유가 이용자의 접근활용을 위한 것이라면 연결요소를 유지해야만 의미가 있다는 점에서 연결요소까지를 포함한 확장된 의미로서의 기록AIP를 다루고자 한다.

3.1 개념 레이어 - 기록AIP의 구성

“기록은 내용(Content), 맥락(Context), 구조(Structure)로 구성되며, 업무활동의 사실(Content)을 일정한 서식(Structure I)에 담아 다른 기록들과 함께 업무 기능과 활동의 증거 및 맥락(Content & Context)을 전달해 주며, 파일링시스템 등의 물리적 구조체(Structure II)로 존재하게 된다”¹¹⁾고 하는 기록의 구성요소에 대한 고전적인 정의에서 출발해 보자.

11) Shepherd & Yeo, 2003

표3-1. 기록의 구성요소

구성요소	표현내용
내용(Content)	<ul style="list-style-type: none"> ● 활동의 사실 ● 업무 지식과 정보
맥락(Context)	<ul style="list-style-type: none"> ● 업무기능과 활동의 증거
형식(Structure I)	<ul style="list-style-type: none"> ● 문서 서식 ● 기록하는 방식
구조(Structure II)	<ul style="list-style-type: none"> ● 논리적 관련성 ● 물리적 구성

전자기록의 경우, 위 표의 “내용”에 해당하는 것이 전자문서 파일 안에 들어있는 내용이 될 것이고, “맥락”은 전자기록이 생산되게 된 배경에 해당하는 업무기능의 분류 정보와 조직 정보가 될 것이며, “형식”은 전자문서에 적용된 약속된 포맷구조가 될 것이며, “구조”는 기록이 몇 개의 전자문서 파일로 구성되고 각 전자문서 파일은 어느 것이 본문이고 어느 것이 첨부인지 등의 논리적인 구성 정보와 각 전자문서가 어떤 소프트웨어에 의해 생산되어서 어느 매체에 어떤 이름으로 저장되었는가 등에 대한 물리적 정보가 될 것이다.

“전자기록 진본성 유지를 위한 기록관리시스템 기능 및 정보모형”에 따르면 전자기록이 진본성을 유지하면서 중장기적으로 안정적으로 보존되기 위해서는 기록의 메타데이터가 다음과 같은 8가지의 범주에서 상세히 기술되어야 한다.

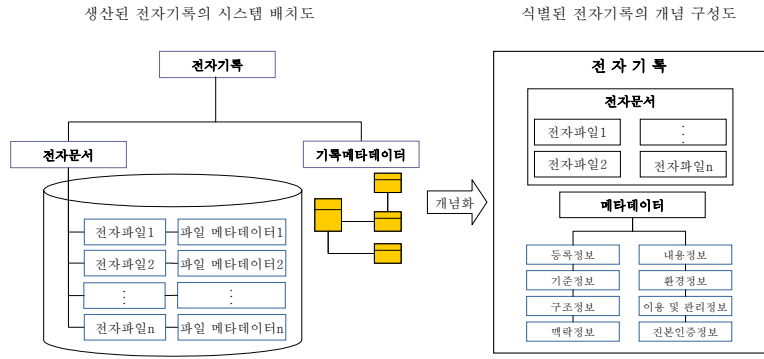


그림3-1. 전자기록의 개념적 구성도¹²⁾

표3-2. 기록의 메타데이터 그룹¹³⁾

메타데이터 그룹	내용	요소
등록정보	● 기록의 유일한 식별기호, 날짜 등과 같이 등록 및 확인에 필요한 정보	식별자 날짜 위치값
기준정보	● 기록의 권한관리와 처리기준 정보 ● 기록의 이용과 관리 조건에 대한 정보	권한관리 처리일정
구조정보	● 기록의 물리적·논리적 구조에 관한 정보 ● 보존관리를 포함하여 기록에 행해질 관리조치를 통제하기 위한 정보	기록유형 기록계층 보존이력
맥락정보	● 기록의 생산배경이나 기록을 생산한 업무기능 정보 ● 다른 기록과의 관계정보	기관 기능 법규 관계
내용정보	● 기록자체의 내용을 설명하는 정보	제목 주제 파일내용
환경정보	● 전자기록의 형태나 생산 및 이용을 위해 필요한 기술적(technological) 환경에 대한 정보	포맷 소프트웨어 하드웨어
이용 및 관리정보	● 기록이 등록된 시점부터 폐기될 때까지 기록에 취해진 관리행위나 이용에 대해 감사추적하기 위한 정보	관리이력 이용이력
진본인증정보	● 이관대상 기록이나 기록관리시스템에서 관리하는 기록에 불법적인 변화가 없음을 보증하는 정보	기록인증 전자파일인증

12) 김익한, forthcoming

13) 김익한, forthcoming

전자기록의 구성요소를 그림으로 나타내보면 크게 전자문서 파일들과 메타데이터 집합으로 볼 수 있다. 그림3-1에서 왼쪽은 전자기록이 생산되었을 때 기록을 구성하는 요소들이 시스템에 어떻게 배치되어 있는지를 나타낸 것으로 기록을 구성하는 전자파일들은 파일시스템에 저장되고 각 파일단위의 메타데이터를 갖게 되며, 기록자체의 메타데이터들은 주로 별도의 데이터베이스에 저장되어 있음을 표현하였다. 오른쪽 그림은 식별된 하나의 전자기록의 개념 구성도로 전자파일들과 파일단위와 기록단위의 메타데이터를 합한 전체 메타데이터를 유형별로 재정의한 구성을 보여주고 있다.

기록AIP의 메타데이터에는 기록의 메타데이터, 보존을 위한 메타데이터, 기술적인(technological) 메타데이터 3가지 종류가 있다. 기록의 메타데이터는 기록의 내용이나 맥락과 관련된 정보로서 DIP에도 포함되는 주요 내용이 되며, 보존을 위한 메타데이터는 기록관리시스템에서 보존관리를 위해 필요한 내용으로 DIP에는 거의 포함되지 않는다. 기술적인 메타데이터는 DIP에도 포함되어야 하는 필수적인 정보로 이 정보가 있어야 이용자들이 DIP의 전자파일을 해석하여 활용할 수 있다. 그런데, 전자기록관리시대에 중요하게 취급해야 할 메타데이터가 또 있다. 바로 기록AIP에 접근하고 활용하기 위해 필요한 메타데이터에 대한 것이다. 이 시점에서 메타데이터에 대한 개념의 확장이 필요하다. 메타데이터를 만드는 가장 궁극적 목표는 원 자원(Original Resource) 즉, 기록자체를 찾기 위한 안내자의 역할을 하기 위해서이므로 더블린 코어 메타데이터에서도 메타데이터는 자원기술(Resource Description)이 아닌 자원 발견(Resource Discovery)의 개념으로 사용되고 있다. 기록의 메타데이터도 보존과 관리를 위한 메타데이터 이외에도 접근 및 활용을 위한 메타데이터(색인어 등)의 구분이 필요하며 이러한 메타데이터는 기록의 내

용정보는 아니지만 기록과 함께 패키징되어 이관될 필요가 있으며 메타데이터 유형별로 관리 방식과 활용 방식이 달라져야 한다. 앞에서 살펴본 OAIS 정보모델에서의 연결정보가 이런 메타데이터에 해당한다.

전자기록의 메타데이터를 잘 구비하여 유지하는 것의 목적은 기록이 자체만으로도 내용을 충분히 이해할 수 있도록 하기 위해서이다. 메타데이터는 전자기록의 관리와 보존을 위해 없어서는 안 되는 것으로 지적 통제, 관리이력 기술, 보존기술, 진본성 표현, 장기적인 가독성, 보존, 검색지원 등의 다양한 기능을 지원한다. 전자기록을 위한 메타데이터 표준은 현재 작성 중이며, 보존기술(Archival description)을 위한 ISAD(G) 표준은 전자기록의 디지털 특성과 속성에 대한 몇 개의 필드만을 제공하고 있다.

이러한 기록의 구성요소가 고전적인 기록의 구성요소 및 OAIS의 AIP 구성요소들과 어떻게 매핑되는지를 도해해보았다.

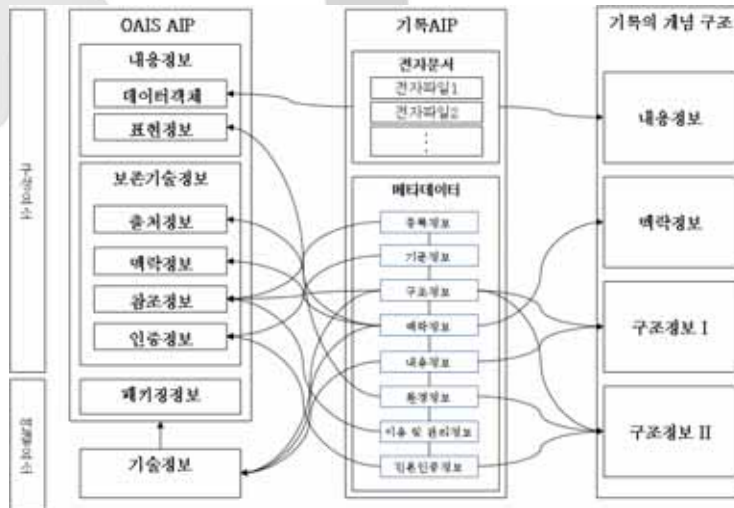


그림3-2. OAIS, 기록AIP, 고전적 기록개념구성 간의 매핑

그럼에서 이해되는 것처럼 고전적인 기록의 개념구조에서는 강조되지 못했던 기록AIP의 연결요소가 OAIS 모델을 적용하면서 분명하게 드러나게 된다. 전자기록은 정보시스템을 떠나서는 존재할 수 없으며 기본적으로 정보시스템의 서비스를 통해 이용자에게 제공된다는 환경적 제약을 생각하면 이용자의 탐색, 검색, 주문 등을 통한 기록의 접근과 이용을 위해 기록AIP에 대한 일정 부분의 정보를 연결정보로서 제공하는 것이 필요하다. 이러한 연결정보를 기록AIP의 구성요소로 반영하는 것이 전자기록관리 시대의 특성을 반영하는 핵심 변화 내용이 될 것이다.

보존전략에 따른 구성변화

생산된 기록은 SIP 형태로 전자기록관리시스템에 제출되고, 기록관리자가 수립한 보존 전략에 따라 AIP로 재구성되어 보존 관리되며, 이용자의 요구에 의해 DIP가 생성되어 배포된다. 기록AIP는 관리 과정에 따라 또는 이용자의 요구에 따라 그 상세한 내부 구성이 달라질 수 있다.

기록AIP의 내부 구성에 중요한 영향을 끼치는 관리행위는 앞서도 언급했듯이 기록에 대한 보존 전략이다. 조직에서의 기록에 대한 보존 전략은 어떤 기록정보가 보존되고 이 정보를 디지털 리퍼지토리안에서 어떻게 관리할 것인지를 결정하는 것으로, 전자기록이 향후에도 생산시점과 동일하게 표현가능하게 남아 있도록 하면서 동시에 진본성과 의미 해석이 가능하도록 보장하는 방법을 정하는 것이다. 현재로는 디지털 정보로서의 전자기록에 대해 마이그레이션(Migration)과 에뮬레이션(Emulation)이 주요한 보존 전략으로 제시되고 있으며, 양자의 장단점에 대해서는 많은 논의가 이루어지고 있다. 물론, 양자의 접근 방법이 서로 배타적이지 않으며 상호 보완적인 전략으로서 채택이 가능하므로 기록관리조직에서는 서로 다른

전자문서의 유형에 대해 두 가지 전략을 따로 적용할 수도 있다.

어떤 전략을 선택하든 기록의 생명주기 동안 마이그레이션과 에플레이션은 반복되어 적용될 것이며, 특정 기록 건에 대한 전략이 어느 시점에 바뀔 수도 있다고 본다. 기록에 대한 가독성을 최대한 보장하기 위해, 그리고 미래의 어느 시점에서 해당 기록에 대한 마이그레이션과 에플레이션 중 전략 전환이 가능하기 위해서는 기록이 디지털 리파지토리에 원래의 파일 포맷과 보존 포맷 둘 다 함께 통합되어 저장되어 있어야 한다. 미래의 기술적 진보에 따라서 지금 당장은 활용하지 못하는 비트스트림도 나중에는 보존 조치에 유용하게 사용될 수 있기 때문이다.

하나의 기록에 속한 전자파일의 포맷이 여러 가지 존재하는 경우 이용하는 그 기록에 대해 여러 포맷이 존재하고 있으며 그 시점에 접근이 가능한 포맷이 어떤 것들이 있는지를 확인할 수 있어야 하며, 권한이 있는 경우 원하는 특정 포맷의 기록에 접근할 수 있어야 한다. 이러한 전략의 다양성과 적용변화 가능성을 고려한 기록AIP구성은 다음과 같이 도해될 수 있다.

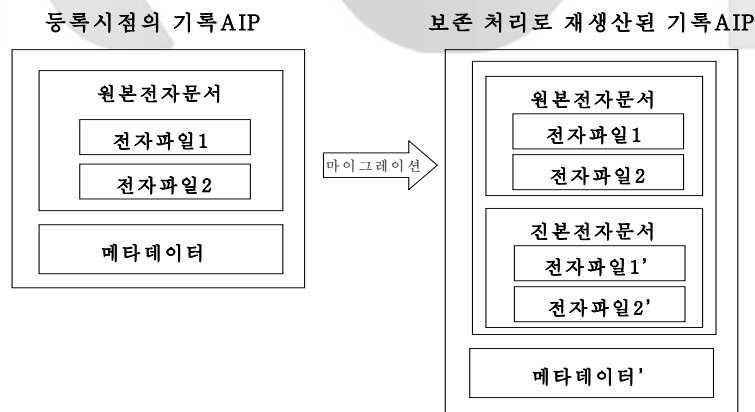


그림3-3. 기록AIP의 변화과정

위의 그림에서 등록시점의 기록AIP는 원본 전자파일들과 메타데이터로 구성되지만 이후 보존 처리의 필요에 의해 각 전자파일에 대한 마이그레이션 파일이 생성될 수 있다. 마이그레이션 파일의 포맷을 원본과 유사한 형태로 가져갈 것인지 아니면 향후의 추가 마이그레이션 필요성을 최소한으로 하는 안정적인 보존 포맷으로 가져갈 것인지는 보존전략에 따르게 된다. 또한, 원본을 마이그레이션한 사본 전자파일을 진본으로 인증하는 절차 또한 조직의 전자기록 진본성 확보 요건에 맞춰 진행되며 보존처리 및 진본인증 과정에 대한 관리이력 정보가 메타데이터에 반영되게 된다. 그림3-3에서 마이그레이션된 결과로서의 “진본전자문서”는 이러한 진본인증과정을 반영하여 명명한 것이다. 이러한 마이그레이션의 결과로 기록AIP는 전체적인 볼륨이 커지게 되므로 원본과 진본사본을 둘 다 보존할 것인지 여부 또한 보존 전략에 따라 결정하게 된다.

기록AIP의 저장공간

이렇게 등록, 보존관리되는 기록AIP는 다음과 같은 3 종류의 개념 영역에 나누어 저장된다.

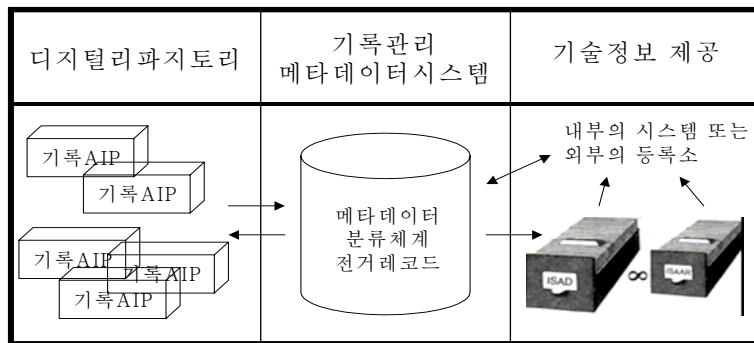


그림3-4. 기록AIP의 저장 공간 개념도¹⁴⁾

디지털 리파지토리는 기록AIP의 전자파일을 중심으로 하는 내용정보가 저장되는 공간이며, 기록관리메타데이터 시스템영역은 기록AIP의 메타데이터와 분류체계 정보가 저장되는 공간이고, 기술정보 영역은 이용자들이 기록AIP를 탐색, 검색, 주문하기 위해 필요로 하는 정보를 저장하는 공간이다.

이 3 종류의 공간은 개념 레이어에서 정의한 것이므로 모두 다 정보시스템으로 구현될 수도 있고 일부만 구현될 수도 있다. 예를 들어, 기술정보 영역의 경우, 기술정보를 목록정보시스템으로 구축할 수도 있고 종이로 만든 목록카탈로그를 작성하여 제공할 수도 있고, 또는 다른 정보시스템에 등록정보로 올려 더 넓은 범위의 이용자가 접근 가능하도록 조치할 수도 있다. 어떤 영역을 정보시스템화할 것인지, 어떤 기술로 구현할 것인지는 개념 레이어에서 논의할 문제가 아니다. 개념 레이어에서는 관리 범위에 들어가는 기록AIP가 어떤 것들이며, 그 기록AIP들을 이용하고자 하는 고객들이 어떤 사람들인가를 정의하고, 기록AIP의 정보들이 어떤 저장 공간에 저장되어야 고객들이 편리하게 기록AIP에 접근할 수 있을 것인지를 개념화하여 빠짐없이 정확하게 모델링하는 것이 주안점이다.

3.2 논리 레이어 - 기록AIP의 구조

논리 레이어에서는 전자기록관리시스템 구축을 전제로 하여 기록AIP의 설계를 하게 된다. 이때는 시스템을 구성하는 소프트웨어와 하드웨어에 대한 선택이 전제되고, 선택한 플랫폼의 장점을 확대하고 단점을 보완하는 측면과 시스템의 전반적인 성능 향상을 고려하는 측면에서의 설계가 진

14) Filip Boudrez, 2005

행된다. 물론, 설계의 과정에도 조직의 기록에 대한 보존 전략이 관철되게 되는데 어떠한 전략을 채택하느냐에 따라 시스템의 소프트웨어와 하드웨어를 선택하는 폭이 좁아지기도 한다. 기록보존관리자는 전자기록을 어떻게 재현해보일 것인지, 메타데이터를 어떻게 저장할 것인지에 대한 여러 가지의 선택을 할 수가 있다. 또한, 디지털 리퍼지토리를 구성하면서 저장의 기본단위를 무엇으로 할 것인지, 기록에 대한 식별을 어떻게 할 것인지 등을 선택할 수 있다.

전자기록의 저장 방법으로서의 인캡슐레이션

기록AIP 설계에 큰 영향을 미치는 보존 전략은 기록을 인캡슐레이션할 것이냐 여부이다. 기록AIP의 저장 방법을 선택할 때 가장 중요한 점은 전자기록의 식별과 다양한 구성요소들 간의 관계에 있다.

기록의 구성요소는 보존되는 동안 하나라도 결코 잃어버려서는 안 되는 단일한 논리적 객체를 형성한다. 그런데, 대부분의 기록 저장 방법은 전자기록의 다양한 구성요소들을 단일한 물리적 객체로 형성하지 않고 데이터베이스, 파일시스템 등의 별도 장소에 별개의 디지털 객체로 저장하고 상호간의 관계는 링크, 데이터베이스 릴레이션, 포인터와 파일이름 등으로 표시한다. IT의 빠른 진보로 인해 플랫폼이 자주 변하고 그 때마다 디지털 객체들이 마이그레이션을 반복하게 되는 경우, 이러한 참조정보는 장기보존 과정에서 유실될 가능성이 있다. 만약 상호 참조관계 정보가 깨져 기록을 재구성할 수 없게 되면 그 기록은 잃어버린 것으로 간주된다. 기록보존관리자는 전자파일 안에 메타데이터를 포함해 넣음으로써 이러한 위험을 피할 수 있다. 즉, 하나의 물리적 객체 안에 두 개의 컴포넌트를 결합함으로써 기록과 메타데이터의

관계가 유실되는 것을 막을 수 있는 것이다.

이렇게 디지털 객체에 메타데이터를 추가하는 것을 인캡슐레이션 또는 임베딩(Embedding)이라고 하며, 인캡슐레이션은 실제로 하나의 디지털 객체에 메타데이터를 추가하거나, 하나의 디지털 객체 안에 여러 전자파일이 그룹으로 존재하는 저장 기술을 말하는 것일 뿐 전자기록이 미래에도 안전하게 재구성되고 활용 가능하도록 보장해 주는 방법은 아니다.

디지털 객체에 메타데이터를 임베딩하는 아이디어는 새로운 것이 아니다. 인캡슐레이션은 객체지향 프로그래밍의 기본 원칙의 하나이다. 객체지향에서 사용되는 인캡슐레이션은 하나의 객체를 구성하는 기능과 정보 중에서 정보를 은닉하기 위한 방법으로 인캡슐레이션을 하게 된다. 그러나, 기록AIP 설계에서의 인캡슐레이션은 정보 은닉의 관점에서가 아니라 정보객체의 독립성과 내부 구성요소 간의 결합성이라는 장점을 얻기 위해서 하게 된다¹⁵⁾.

기록AIP를 인캡슐레이션한다는 것은 앞에서 살펴본 기록AIP의 구성요소들을 묶어 하나의 객체, 예를 들면 XML 파일로 만들어 저장한다는 것이다. 이렇게 했을 때의 장점은 기록 내부의 구성요소들끼리 링크나 참조정보를 가지고 있을 필요가 없이 간단하게 저장·관리할 수 있다는 것이며, 단점은 여러 요소를 하나로 묶게 되므로 저장 파일의 크기가 커져서 기록을 검색하여 조회할 때나 DIP를 생성하고자 할 때 큰 파일을 열어 작업해야 하므로 전체적인 성능이 저하될 수 있다는 것이다.

15) 전자기록을 유지하는 데 있어서 인캡슐레이션의 중요성은 David Bearman에 의해 제기되어 왔으며, NARA의 Persistent Object Preservation, 빅토리아주 Public Record Office의 VERS 아카이빙 전략, 호주 National Archives의 AtoR 프로젝트 등에서 적용되고 있다.

인캡슐레이션의 적용

기록보존관리자는 기록AIP의 저장방법으로 인캡슐레이션 기법을 다양하게 적용할 수 있다. 예를 들어, TIFF 파일의 헤더에 메타데이터 태그를 채우는 식으로 기록의 특정 전자파일 안에 메타데이터를 포함하는 방법이 있다. 그러나, 이 방법은 기록의 여러 전자파일에 대해 동일한 메타데이터가 여러 번 저장되고 이 파일들을 서로 참조관계로 연관지어줘야 하기 때문에 유용성이 떨어진다. 또한, 기술적인 부분에서도 여러 단점이 있다. 보통의 텍스트 파일에서는 문제가 없지만 바이너리 파일 포맷에서는 메타데이터를 포함하기 위해 몇 가지 표준 필드를 제공하긴 하지만 보존기술에 필요한 모든 필드를 제공하지는 않는다는 것이다. 게다가 제공되는 메타데이터 필드는 공간이 너무 작거나 필드를 확장했을 때 가독성(Readability)에 문제를 일으킬 수 있다. 또한, 바이너리 파일에 메타데이터를 추가하는 것은 현재의 컴퓨터 프로그램이 기능을 제공하지 않는 경우가 일반적이므로 각 바이너리 포맷마다 메타데이터 추가를 위한 소프트웨어 툴을 개발해야 한다.

이러한 문제점들을 피하기 위해 하나의 기록에 속하는 모든 전자파일들을 한꺼번에 인캡슐레이션할 수 있다. 그러면, 인캡슐레이션되는 전자파일들은 하나의 파일에 묶여있으므로 서로에 대한 링크정보를 따로 가질 필요가 없게 되고 기록에 대한 기술정보도 한번만 포함하면 된다. 이 때, 인캡슐레이션된 결과 파일은 OAIS 모델의 AIU에 해당하는 디지털 리파지토리의 기본 저장단위가 된다. 인캡슐레이션이 저장방법으로 채택하면 하나 또는 여러 개로 나뉘어 입수되는 기록SIP를 하나의 기록AIP로 팩킹(packaging)하는 과정을 거쳐야 하고, 이 상태로 저장 보존되는 기록AIP들은 팩킹을 풀어야만 보거나 활용할 수가 있게

된다.

기록AIP의 구조

기록AIP를 하나의 파일로 인캡슐레이션하는 경우에 기록AIP의 내부를 식별자(identifier), 기록의 모든 전자파일과 필수 메타데이터, 체크섬 등의 3부분으로 구조화할 수 있다. 식별자는 정보시스템 내에서 기록AIP파일의 유일한 식별자 역할을 하는 값으로 장기보존 과정에서 가급적이면 변하지 않고 하나의 값을 유지할 수 있도록 할당되는 것이 좋다. 체크섬은 OAIS의 인증정보에 대응되는 것으로 기록AIP에 체크섬으로 삽입된 해쉬값은 기록AIP가 전송되거나 마이그레이션 등의 조치 이후에 다시 계산한 해쉬값과 비교함으로써 기록AIP가 변조 또는 훼손되지 않았음을 확인하기 위한 정보이다. 체크섬을 이용하여 기록AIP가 잘 보존되고 있는지를 확인하는 것은 기록보존관리자의 주요 임무 중의 하나이며, 이러한 확인 작업은 전자기록관리시스템이 주기적으로 자동으로 수행하도록 할 수도 있고 기록보존관리자가 필요에 의해 임의로 수행할 수도 있다. 삽입된 해쉬값이 다시 계산한 해쉬값과 일치하지 않으면 시스템이 기록보존관리자에게 그 사실을 알려줘서 관리자가 기록AIP의 훼손된 상태를 파악하고 백업해 놓은 기록AIP를 가져다가 디지털 리퍼지토리에 저장할 수 있도록 한다. 체크섬에는 해쉬값뿐만 아니라 해쉬알고리즘이 함께 보존되어야 한다.

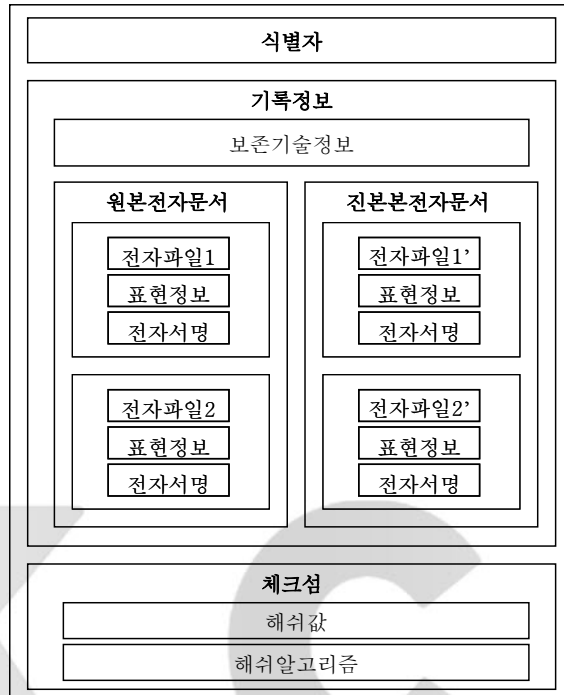


그림3-5. 기록AIP 파일의 구조

기록AIP의 기록정보에는 보존과 관리에 필요한 메타데이터들이 보존기술정보라는 하위요소로 포함된다. 이 메타데이터는 기록AIP 전체와 관련된 것으로 한 번만 저장되면 된다. 개념 레이어에서 찾아낸 기록의 메타데이터 중 표현정보와 인증정보 일부를 제외하고는 거의 모든 메타데이터가 보존기술정보에 함께 기술된다. 기록정보의 하위 요소로 표시된 원본전자문서에는 전자기록이 가졌던 모든 전자파일들의 내용정보와 전자서명을 포함한다. 또한, 진본전자문서에는 마이그레이션된 파일 포맷을 저장하고 표현정보와 전자서명을 포함한다. 마이그

레이션을 누가 언제 어떻게 왜 수행하게 되었는지 등의 이력정보가 보존기술정보에 포함되며, 원본이나 진본전자문서의 모든 전자파일이 동일한 표현정보를 갖는 경우 기록AIP에 표현정보는 한 번만 기술할 수 있다.

전자파일을 마이그레이션할 때 어떤 파일포맷으로 할 것인가는 보존전략에서 중요한 선택사항이다. 예를 들어, 한글 워드프로세서로 98버전으로 작성된 전자파일을 언제쯤 어떤 포맷으로 마이그레이션할 것인지 기록보존관리자는 결정해야 한다. 2004버전이 출시된 시점까지 마이그레이션을 하지 않았을 때, 현재는 98버전 이상의 한글 워드프로세서로 호환이 되므로 98버전의 파일을 읽는 데에 문제가 없을 수 있다. 그러나, 어느 시점에서 워드프로세서가 더 이상 98버전의 파일을 지원하지 않는 제품으로 출시되기 시작한다면 반드시 마이그레이션을 검토해야 할 것이다. 해당 워드프로세서의 최신 버전 파일로 마이그레이션하던가 아니면 향후 마이그레이션을 덜 할 수 있는 안정적인 보존포맷인 PDF/A¹⁶⁾파일로 마이그레이션 할 수 있을 것이다. 마이그레이션이 반복되는 경우 기록정보에 진본전자문서를 모두 보존할 것인지 역시 보존전략에 따른다. 기록보존관리자는 진본전자문서가 반복하여 생산되는 경우 기록AIP의 파일사이즈가 커지므로 저장 공간이 더 필요해 진다는 것과 이용자의 요구에 의해 기록AIP를 접근할 때마다 큰 파일을 조작해야 하는 시스템의 부담이 증가한다는 것을 고려해야 한다.

16) ISO 19005-1:2005로 채택된 전자문서의 장기보존포맷인 PDF/A는 Adobe사가 PDF1.4포맷을 로열티없이 공개함으로써 공개용표준과 편재성의 특성을 모두 만족하는 세계표준 전자문서 보존포맷이 되었다.

기록AIP 보존포맷으로의 XML

기록AIP를 디지털 리파지토리에 하나의 디지털 객체로 저장할 때 여러 가지 방법이 고려될 수 있다. 기록AIP의 저장 파일포맷으로는 XML 포맷이 tar, zip, jar 등과 같은 압축된 포맷보다 장점이 많다. 왜냐하면, 압축을 사용하게 되면 압축알고리즘에 의해 원래의 비트스트림이 인코딩되어 변형된 상태로 보존되므로, 나중에 완벽하게 원래의 비트스트림으로 재구성할 수 있다고 장담할 수 없으며, 원래의 비트스트림 안에 워터마크가 삽입되어 있는 경우에는 압축과 해제 과정에서 데이터가 손실될 수도 있기 때문이다. 또한, 압축파일 안에 어떤 전자파일들이 들어있는지에 대한 참조정보를 생성하여 관리해야 하며, 압축된 포맷은 메타데이터를 포함하는 것에 한계가 있으므로 기록AIP의 저장포맷으로는 적합하지 않다.

XML포맷은 전자파일을 메타데이터와 함께 직접 XML문서 안에 포함할 수가 있는데, XML에 포함되기 위해서 바이너리 파일들은 먼저 Base64로 인코딩되어 텍스트 문자로 변환되어야 한다. Base64를 거쳐 인코딩했다가 다시 디코딩하기 위해서는 재구성을 위한 참조정보가 추가되어야 하지만, Base64는 이미 잘 알려져 있으며 다양한 플랫폼에 광범위하게 사용하는 아주 단순한 방식이다. 대신, Base64를 사용하여 바이너리 파일을 변환하게 되면 파일 사이즈가 3분의 1만큼 증가하게 된다. XML 문서에 메타데이터를 포함하면 함께 저장되는 전자파일과 메타데이터 사이에 참조정보를 추가하지 않아도 된다는 장점이 있다. XML 문서에 메타데이터를 삽입하는 것은 메타데이터와 기록정보를 하나로 묶어준다는 장점뿐만 아니라, 메타데이터가 구조화되고 안정적인 방식으로 보존됨으로써 향후의 활용가능성을 보증할 수 있다는 것

이다. 또한, XML은 필요에 의해 구조를 가변적으로 변화시킬 수 있어서 XML 문서 안에 메타데이터 등의 요소들을 언제라도 확장하거나 변형하는 것이 가능하다. 이러한 XML의 구조적 융통성이 기록에 대한 추가적인 기술(description)과 마이그레이션에 따른 진본전자문서와 메타데이터를 용이하게 포괄할 수 있는 최대의 장점이 된다.

기록AIP의 파일포맷으로서 XML은 기술적인(technological) 이유와 장기간에 걸친 활용가능성의 확보를 위해서 뿐만이 아니라 AIP자체에 대해 문서화를 할 때 편리하다. 기록AIP의 구성과 내용에 대한 정보는 장시간이 지난 후에도 이해될 수 있어야 하며, 미래의 기록보존관리자는 기록AIP 파일에 패키징되어 있는 전자기록과 메타데이터가 어떻게 재구성되어야 하는지를 알 수 있어야만 한다. 그러므로, 기록AIP의 각 부분의 의미와 상호간의 관계는 분명하게 문서화되어야만 한다. XML로 만들어진 기록AIP는 구성 정보를 보여주지 않고도 잘 구조화된 방식으로 외부와 소통될 수 있다. 기록AIP를 위한 XML스키마는 OAIS 모델에서 패키징정보로 기능하여 내용정보와 보존기술정보의 요소들이 무엇이 있는지를 식별하고 서로 연관시켜준다.

결합레벨(Aggregation Level)

이제까지 기록AIP를 기록 한 건과 대응하는 개념으로 설명해왔지만 사실 기록보존관리자는 OAIS 모델의 정보패키지 하나가 기록 건 하나와 일치될 필요는 없다는 점을 고려하여야 한다. OAIS는 디지털 정보의 일반적인 장기보존을 위해 설계된 것으로 하나의 AIP는 하나의 기록과 대응할 수도 있고 기록 건의 전자파일 하나, 또는 기록 여러 건과 대응할 수도 있다. OAIS 모델에서는 다양한 종류의 AIP들이 결합되어

컬렉션을 형성할 수 있으므로 기록보존관리자는 기록관리의 입장에서 기본 저장단위를 설정할 때 다음과 같은 다양한 결합레벨(Aggregation level) 중 하나를 선택할 수 있다.

- (1) 기록 한 건의 하나의 컴포넌트 : 예를 들어, 워드문서 파일에 포함되는 이미지 하나.
- (2) 기록 한 건의 하나의 전자파일 : 예를 들어, 워드문서 파일 하나
- (3) 기록 한 건의 모든 전자파일 : 예를 들어, 본문과 첨부 파일들 모두와 PDF/A 포맷으로 마이그레이션된 파일 모두
- (4) 기록 건 여러 개 또는 레코드그룹 : 예를 들어, 기록철(File)을 구성하는 건들의 일부, 또는 기록철 여러 개¹⁷⁾

전자기록을 장기간 안정적으로 보존하는 것이 논의의 출발점이었던 기록보존관리자는 기록 한 건을 하나의 기록AIP로, 즉 기록 한 건의 모든 전자파일을 결합레벨로 선택하는 것이 일반적일 것이다. 특히 기록보존관리자가 인캡슐레이션의 장점을 살리고 싶어한다면 기록AIP를 하나의 물리적 파일로 저장하는 것이 좋다.

기록AIP의 메타데이터 저장

인캡슐레이션을 하여 기록AIP를 저장하고자 할 때 결정적인 것은 기록AIP파일 안에 어떤 메타데이터를 포함할 것이냐를 결정하는 것이다. 기술적으로는 모든 관련 메타데이터를 삽입하는 것이 가능하지만 효율적이지는 않다. 삽입되는 메타데이터는 자주 변경되지 않는 필수적인

17) 한 파일 내의 기록 건 묶음이 물리적으로 하나의 파일에 포함된다. 이러한 접근방식은 저장되는 물리적 파일의 크기가 매우 커져서(100 MB 이상) 처리가 용이하지 않게 된다.

메타데이터만 포함하는 것이 좋다. 그래야 기록AIP를 재생산하는 횟수가 줄어들게 된다. 또한 모든 메타데이터를 인캡슐레이션하는 것은 기록AIP 파일이 커지게 하고 디지털 리파지토리 내에 중복된 정보가 넘치게 만든다. 각각의 기록AIP파일에 특정 전자파일 포맷의 명세와 뷰어를 포함한다던가, 전거레코드에 해당하는 기록생성자에 대한 다큐멘테이션을 포함한다던가하는 해서는 안 된다. 또한, 개념적으로는 기록과 메타데이터가 캡슐화되지만, 기록의 모든 메타데이터들은 실제로는 디지털 리파지토리와 별도의 기록관리메타데이터시스템에 저장되거나 안에 또는 외부에 저장될 수도 있는 것이다.

그림3-4에서 기록관리메타데이터시스템은 다양한 메타데이터를 보유할 수 있다. 예를 들어, 기록생성자를 포함한 전거레코드, 파일포맷에 대한 표현정보, 저장 매체에 대한 관리 정보, 기록의 상위 레벨 즉, 기록철, 시리즈, 레코드그룹 등에서의 보존기술 메타데이터 등이다. 이러한 메타데이터의 일부는 시스템 외부에 존재하는 곳, 예를 들면 파일포맷에 대한 레지스트리 등에서 추출되거나 연결되어 구성될 수 있다. 이용자들의 기록에 대한 검색속도 등 접근 성능을 높이기 위해 기록AIP들에 삽입된 메타데이터들의 일부를 기록관리메타데이터시스템에 중복할 수도 있다. 이렇게 되면, 기록AIP파일 안에 삽입된 메타데이터와 기록관리메타데이터시스템에서 관리되는 메타데이터 사이에 값을 일치시키는 작업이 필요할 수도 있다. 한편으로 이러한 메타데이터의 중복 저장은 백업으로서 보안기능을 하기도 한다. 기록보존관리자가 기록AIP파일에 삽입될 메타데이터를 선별해야 하는데, 기록AIP파일이 자체만으로도 완전한 기술이 되어야 한다는 측면과 다른 한편으로는 기록AIP파일을 쉽게 관리할 수 있어야 한다는 측면을 두고 균형점

을 찾아야 할 것이다.

4. 맺음말

이 논문에서는 OAS 정보모델과 시스템 구축의 3-레이어 접근법을 이용하여 전자기록관리시스템에서 장기간에 걸쳐 안정적으로 보존관리하기 위한 기록AIP의 구성과 구조, 보존포맷에 대한 논의를 진행하였다. OAS는 개념 레이어에서 기록AIP의 구성을 제시해주었고, 그러한 구성을 바탕으로 기록AIP를 어떤 보존포맷으로 구조화할 것인지에 대해서는 구체적인 보존전략에 따라서, 이용자의 접근 활용 요구에 따라서, 인캡슐레이션의 채택할 것인지의 여부에 따라서, 메타데이터를 선택하는 범위에 따라서 다양한 선택이 있을 수 있다는 것을 논리 레이어의 논의를 통해 확인했다.

이 논문에서 사용한 3-레이어 접근법은 정보공학 등의 정보시스템 개발 방법론에서 보편적으로 적용하는 접근법으로, 기록AIP의 구성과 구조에 대한 논의 수준을 차별화하기 위해 활용되었다. 3-레이어 접근법은 시스템 개발의 초기 단계에서는 개념적 수준에서의 분석을 하고, 다음 단계에서는 분석결과를 현실적인 제약조건에 맞춰 논리적으로 설계를 하고, 마지막 단계에서 설계된 내용을 토대로 물리적인 구현을 한다는 것이 기본 틀이다. 기록AIP의 구성과 구조에 3-레이어 접근법을 적용하여 먼저 개념적인 수준에서 패키지가 어떤 내용들로 구성되어야 하는지를 정리해보고, 다음 단계에서 어떤 전략과 기술을 채용할 것

인가를 중심으로 패키지의 구성요소들을 구조화하는 여러 방법들에 대해 논의해 보았다. 이러한 3-레이어 접근법은 기록AIP가 장기보존이 가능하기 위해서 어떤 구성을 유지해야 하는지, 어떤 구조로 저장되어야 하는지 등에 대해 원칙적인 부분과 상황에 맞춰 유연하게 변형 적용해야 하는 부분을 나누어 생각할 수 있게 함으로써 기록의 정보모델과 보존포맷 사이의 간격을 입체감 있게 바라볼 수 있게 한다. 개념 레이어의 모델은 상대적으로 장기적으로 공유할 수 있는 지식이며, 논리 레이어의 설계는 시스템을 구축하는 주체의 주관적, 객관적 조건에 따라 다양한 변이가 가능하다. 물리 레이어는 실제적인 구현으로 비용을 고려한 전개가 필요한 부분이므로 기록관리 자체의 요구사항보다는 조직을 둘러싼 기본적인 제약, 즉 자원의 배분을 둘러싼 기본 제약의 영향을 받는 영역이라 하겠다.

대부분의 표준이 개념 레이어 수준의 서술 중심이며, 실제로 기록관리 주체들은 현실적인 요건 즉 기록관리시스템의 구축 목적, 예산과 요구되는 성능 등의 제약조건 하에서 개념 레이어에서 정의된 내용을 그대로 유지하면서 현실적인 설계를 하고 구현을 하게 되는 것이다. 이러한 레이어의 분리를 통해 개념에 담긴 원칙을 그대로 고수하면서 현실적인 여건에 따라 시스템에 대한 제각기 다양한 설계와 구현이 가능하며, 개념은 상대적으로 안정적이며 불변이지만 기술의 발전은 빠르므로 개념은 그대로 두고 논리적인 설계만을 바꾸어 시스템을 고도화를 해나가는 것이 가능해진다. 3-레이어 접근법의 관점에서 보자면 OAIS 모델은 개념 레이어에서 정의된 표준모델이며 일부 전략적 측면에서 논리 레이어에 대한 권고를 포함하고 있다. 국내에서도 OAIS와 같은 개념 레이어 수준의 기록관리 기능 및 정보모델의 표준화가 조속

히 작성되어 전자기록관리시스템을 구축하려는 공공기관과 민간기업들에게 가이드라인으로 제공되기를 기대해 본다.

이 논문에서는 보존정보패키지의 구성 측면보다는 구조의 측면에 논의를 더 집중하였다. 장기보존을 위한 메타데이터 구성요소에 대한 식별과 정의는 OAIS 보존기술정보에 대한 최근의 논의와 OCLC/RLG의 보존메타데이터 연구그룹 PREMIS의 최근의 연구 결과 등을 반영하여 좀 더 상세히 다뤄질 필요가 있다. 또한 이 논문의 후속으로 OAIS 모델에 따른 기록DIP의 구성과 구조 설계, OAIS 보존관리기능이 AIP와 DIP에 주는 영향과 그에 따른 AIP와 DIP의 구성과 구조, AIP와 DIP의 구성과 구조에 진본성 유지 요건 적용, OAIS 모델의 컬렉션 개념을 기록의 분류체계에 적용 등의 여러 연구가 필요하다고 본다.

참 고 문 헌

- CCSDS 650.0-B-1:Reference Model for an Open Archival Information System, 2002
- 이소연, 디지털 아카이빙의 표준화와 OAIS 참조모형, 2002
- 김희정, 디지털 아카이빙 최근 연구동향 및 OAIS 참조모형에 관한 연구, 2003
- Moore R., A. Merzky , Persistent Archive Concepts, 2003
- Shepherd, Elizabeth and Geoffrey Yeo. Managing Records: A Handbook of Principle and Practice, Facet Publishing, 2003
- 이승휘 외, 대통령비서실 기록관리체계 재설계 연구방안 보고서, 2004
- 헨디소프트 컨소시엄, 대통령비서실 기록관리혁신을 위한 정보화전략 계획수립 결과보고서, 2005
- Filip Boudrez, Digital containers for shipment into the future, 2005
- Filip Boudrez, Digital signatures and electronic records, 2005
- LG CNS 컨소시엄, 국가기록원 기록관리시스템혁신 정보화전략계획수립 결과보고서, 2006
- 김익한, 전자기록 진본성 유지를 위한 기록관리시스템 기능 및 정보모형, forthcoming

Abstracts

**The composition and structure of Archival Information
Packages(AIP) for a long-term preservation of electronic records**

YIM, JIN HEE

It's needed for the archivists to design a flexible and stable ERMS(Electronic Records Management System) which can ingest and store records through a consistent way and let users search and use records easily what they want. The basis of the design for ERMS are the conceptual composition and the logical and physical structure of the records when they are stored and managed in the ERMS. This paper explains the process of defining components and designing structure of electronic records using 3-layered approaches which consist of conceptual, logical and physical layer and shows advantages of this approaches.

After benchmarking the information models of OAIS which is a reference model for the long-term preservation of digital information objects, this paper applies the model of AIP to a record as a 'Record AIP' and discusses the composition and structure of it. It's a critical task to identify mandatory or optional metadata groups which consists of the 'Record AIP's in the conceptual layer. This paper emphases that the metadata group related to services for the record information to users is required as a result of benchmarking OAIS information models. Various issues about the structure of 'Record AIP's are discussed according to the kind of preservation strategy such as migration or emulation and whether the encapsulation of records is required or not in the logical layer.

**Key words : Archival Information Package, AIP, OAIS, Information Model,
3-Layered approach, long-term preservation, ISO 14721**

전자기록의 장기보존을 위한 보존정보패키지(AIP) 구성과 구조

임진희*

1. 머리말
 - 1.1 연구 목적과 범위
 - 1.2 연구 방법
2. OAIS 정보모델과 AIP
 - 2.1 OAIS 정보모델의 구성 요소
 - 2.2 정보패키지의 유형 - SIP, AIP, DIP
 - 2.3 정보의 기본단위와 컬렉션 - AIU와 AIC
 - 2.4 OAIS 정보모델의 기록관리 적용
3. 기록AIP 구성과 구조
 - 3.1 개념 레이어 - 기록AIP의 구성
 - 3.2 논리 레이어 - 기록AIP의 구조
4. 맺음말

* 한국국가기록연구원 책임연구원

[국문초록]

전자기록 시대를 맞아 여러 유형의 생산시스템에서 만들어지는 다양한 전자기록을 일관된 방식으로 입수하여 보존관리하면서 이용자들이 기록의 존재를 쉽게 검색할 수 있게 하고 원하는 내용을 배부해줄 수 있는 유연하면서도 안정적인 전자기록관리 시스템을 설계하는 것이 필요하게 되었다. 전자기록관리시스템 설계의 가장 기본이 되는 것이 바로 보존관리되는 과정에서의 기록의 구성과 구조라고 할 수 있다. 이 논문은 3-레이어 접근방식에 따라 전자기록의 개념적 구성과 논리, 물리적 구조를 설계하는 과정을 설명하고 있으며, 이 과정에서 3-레이어 접근방식의 유용성을 보여주고 있다.

이 논문은 디지털 정보객체의 장기보존을 위한 참조모델인 OAIS의 정보 모델을 벤치마킹하여 보존패키지(AIP) 개념을 기록에 적용한 ‘기록AIP’의 구성과 구조에 대해 논의한다. 개념 레이어에서는 기록AIP를 구성하는 메타데이터의 그룹을 식별해내는 작업이 핵심이며 특히 OAIS 정보모델의 벤치마킹을 통해 기록정보의 서비스를 위해 필요한 메타데이터 그룹을 기록AIP의 구성에 필수적으로 포함해야 한다는 것을 강조하고 있다. 논리 레이어에서는 보존전략(마이그레이션과 에뮬레이션)이 기록AIP의 구조에 미치는 영향과 기록AIP를 인캡슐레이션할 것인지 여부에 따른 구조에 대해 논의한다. 이 논문은 현재의 수준에서 사용할 수 있는 정보통신 기술과 각종 표준을 기준선으로 하여 기록AIP를 어떻게 구성하고 구조화할 수 있는지 논의를 정리하고 있다.

주제어 : 보존정보패키지, AIP, OAIS 모형, 정보모델, 3-레이어 접근방법, 장기보존, ISO 14721

1. 머리말

1.1 연구 목적과 범위

이 논문의 목적은 전자기록의 장기보존(Long-term Preservation)을 위한 포맷 연구에 있다. 여기서 장기보존이란 미래에도 기록의 재구성이 가능하고 활용이 가능하도록 보존하는 하는 것을 의미하며, 보존포맷이란 컴퓨터로 접근과 활용이 가능한 매체에 기록정보가 저장되는 형식을 의미한다. 여기서 ‘보존포맷’이라는 말은 여러 의미를 내포한다. 먼저, 전자기록이 어떤 메타데이터들과 함께 저장될 것이냐는 구성의 측면과, 전자기록과 메타데이터를 인캡슐레이션(Encapsulation)할 것인지를 어떤 파일포맷으로 저장할 것인지에 대한 물리적 구조의 측면이다.

본 논문은 최근 대통령비서실과 행정자치부에서 구축하는 전자기록관리시스템에 구현되어야 할 전자기록의 저장과 보존형태에 대한 고민에서 시작되었다. 여러 유형의 생산시스템에서 만들어지는 다양한 전자기록을 일관된 방식으로 입수하여 보존관리하면서 이용자들이 기록의 존재를 쉽게 검색할 수 있게 하고 원하는 내용을 배부해줄 수 있는 유연하면서도 안정적인 전자기록관리시스템을 설계하려할 때, 가장 기본이 되는 것이 바로 보존관리되는 과정에서의 기록의 구성과 구조, 그리고 존재형태와 형식이다. 전자기록의 디지털객체로서의 특성과 기록으로서의 특성을 아우르는 보존 포맷에 대한 본 논문은 앞으로 공공영역에서 활발하게 진행될 전자기록관리시스템의 설계에 기초적인 아이디어를 제공하고자 작성되었다.

1.2 연구 방법

이 논문은 디지털 정보객체의 장기보존을 위한 참조모델인 OAIS의 정보 모델을 벤치마킹하는 것으로 시작하고 있다. OAIS는 기능모델과 정보모델로 구성되어 있고, 두 모델은 밀접한 연관성을 지니며 통합되어 있으나 본 논문에서는 기록의 보존 포맷에 대한 현실적인 여러 문제제기와 이에 대한 검토를 중심으로 논의가 진행되므로 기능을 중심에 두지 않고 정보모델을 중심으로 살펴보고, 이 정보모델을 적용하여 기록의 보존패키지 구성과 구조 및 보존포맷에 대해 논의를 진행한다. 단, 정보모델에 결정적인 영향을 주는 장기보존 전략과 접근이용 서비스 등의 몇 가지 기능에 대해서는 이슈에 대한 논의 과정에서 조금씩 언급할 것이다. 이 논문에서는 기록 한 건에 대응하는 개념으로서의 기록 보존패키지에 대한 구성과 구조를 살펴보기 위해 OAIS 모델을 적용하고 있으며, OAIS의 컬렉션 개념과 컬렉션 기술(Description) 모델을 기록의 분류체계와 연관지어 해석하고 적용하려는 시도는 제외되었다.

3-레이어 접근법

또한, 이 논문에서는 보존전략에 따라 전자기록의 구성과 구조를 설계하고 전자기록관리시스템에 구현하는 과정을 설명하기 위해 3-레이어 접근관점을 제안한다. 일반적으로 하나의 정보시스템을 구축하는 과정은 그림 1.1.1.1과 같이 3 단계의 레이어(Layer)로 나누어 볼 수 있다.¹⁾

1) 소프트웨어 공학의 주요 방법론인 구조적 개발방법론, 정보공학방법론 등에서 소프트웨어의 개발 단계를 일반적으로 크게 분석, 설계, 구현 등 3단계로 나누어 설명하고 있으며, 각 단계는 개념레이어, 논리레이어, 물리레이어에 조응한다.

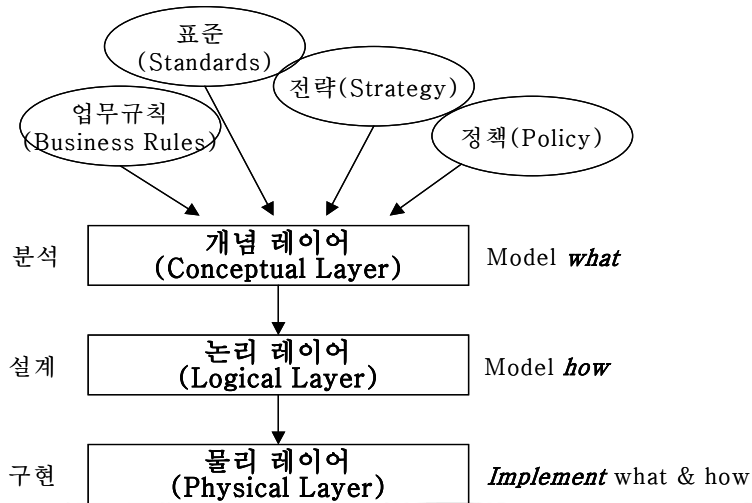


그림1-1. 정보시스템의 개발 단계

*개념 레이어(Conceptual Layer)*는 개념 모델링 단계로서 정보시스템에 갖추어져야 하는 내용을(What) 찾아내는 과정이다. 전자기록관리시스템의 경우, 시스템이 수행할 수 있는 관리 기능이 무엇이며, 관리해야 하는 기록정보가 무엇인지 찾아내어 이름지어주고 그 특성을 파악하여 기능과 정보의 모델을 생성하게 된다. 기록관리의 관점에서 본다면 관리의 대상이 되는 기록을 찾아내고, 그 기록을 구성하는 요소들을 하나하나 정의해주고 세부 속성 값들을 정의해 주는 과정이 될 것이다. 개념 모델링 단계에서는 정보시스템이 어떤 소프트웨어와 하드웨어로 구축이 될 것이냐와는 무관하게 독립적으로 진행이 되는 것으로, 하나의 개념 모델이 완성되면 동일한 모델을 여러 종류의 기술 기반에 적용하는 것이 가능하다.

*논리 레이어(Logical Layer)*는 논리 모델링 단계로서 정보시스템을 구성하는 소프트웨어와 하드웨어, 기술 기반에 대한 선택을 전제로 하여 개념 레

이어에서 찾은 대상을 어떻게(How) 구현할 지를 설계하는 과정이다. 채택한 플랫폼의 특성을 반영하여 설계가 이루어지는데 예를 들어, 기록정보의 저장과 관리를 위해 관계형 데이터베이스를 선택할 것인지, 혹은 계층형 데이터베이스를 선택할 것인지, 혹은 XML 데이터베이스를 선택할 것인지에 따라 저장하여 관리하는 기록의 구조가 결정되게 된다. 또한, 기록정보를 인캡슐레이션할 때 XML 스키마를 활용할 것인지, 아니면 RDF를 채용할 것인지에 따라 구조와 저장포맷이 다르게 된다. 또한, 시스템이 최대한의 성능을 보장할 수 있도록 정보모델에 대한 반정규화 등의 고도화된 설계 기법을 구사하는 단계이기도 하다.

*물리 레이어(Physical Layer)*는 물리적 구현 단계로서 논리 레이어에서 설계한 내용을 주어진 예산에 맞춰 전개하여 구현하는 과정이다. 예를 들어, 이용자들이 자주 접근하여 활용하는 기록들은 비싸지만 접근 속도가 빠른 매체에 두어 서비스 수준을 향상시키도록 한다든지, 필요하다면 기록을 복제하여 한 벌 더 덩으로써 더욱 효율적인 서비스를 할 수 있게 하는 등의 조치를 포함한다. 물리 레이어는 기록정보의 구성과 구조와는 별개로 매체 선택과 분산, 복제 등 물리적 기술(Technology)에 대한 고려가 중심이므로 이 논문의 논의에서는 제외하기로 한다.

3-레이어 접근법의 장점

3-레이어 접근법의 가장 큰 장점은 레이어간의 독립성에 있다. 정보시스템의 기반이 되는 IT 기술이 빠르게 발전해 나감에 따라 그에 상응하여 기록정보의 물리적인 구현도 적절한 시점에 새로운 기술을 적용하며 따라가 주어야 하는데, 그 때마다 정보시스템을 처음부터 다시 만들어 가는 것이 아니라 개념모델은 그대로 두고 논리적 설계부터 또는 물리적 구현만을 바

뛰주면 IT의 발전과정을 빠르게 시스템에 적용할 수 있다. 예를 들어, 전자 기록의 구성요소는 개념적인 수준에서 하나로 모델링이 되지만, 전자 기록을 저장하고 관리하기 위한 전문 데이터베이스관리시스템을 현재 전 세계적으로 가장 많이 사용되고 있는 안정적인 관계형 데이터베이스관리시스템을 사용할 것인지 또는 XML 문서를 분해하지 않고 그대로 저장하여 관리할 수 있는 XML nature 데이터베이스관리시스템을 사용할 것인지에 따라서 전자 기록의 논리적인 설계는 달라진다. 관계형 데이터베이스의 경우에는 전자 기록을 구성하는 개념적인 요소들이 각각 테이블로 쪼개져서 저장되는 것으로 설계될 것이며, 인캡슐레이션은 개념적으로만 존재하고 설계 시에는 인캡슐레이션의 개념이 충실하게 보장될 수 있도록 하는 설계 방식을 취하게 된다. 하지만, XML 데이터베이스를 선택하는 경우에는 전자 기록이 하나의 XML 문서로 인캡슐레이션되어 물리적인 하나의 덩어리로 저장되는 것으로 설계될 수 있어서, 개념 레이어의 인캡슐레이션이 논리층에 변형없이 그대로 적용될 수 있다. 또한, 관계형 데이터베이스에서는 테이블, 칼럼, 기본키, 외래키 등의 개념으로 정보를 정의하고 연결하여 하나의 의미 실체를 담게 되지만 XML 데이터베이스는 전혀 다른 용어와 개념으로 정보를 정의하게 된다.

즉, 개념 레이어에서 정의한 내용은 동일할 수 있지만 논리 레이어에서는 어떤 기술기반의 시스템으로 갈 것이냐를 전제로 하여 매우 용어를 사용하여 다른 설계가 진행될 수 있는 것이다. 만약 개념 모델은 그대로 유지하면서 기술 기반만 새로운 플랫폼으로 변경하기를 원한다면 이 때에는 분석단계를 반복하지 않고 논리 모델을 새로 설계하는 것부터 진행할 수 있다는 것이 3-레이어를 분리하여 시스템 구축을 하는 최대의 장점인 것이다.

OAIS 모델은 3-레이어 접근방식에서의 최상위인 개념 레이어에서 기술

된 것으로 이러한 OAIS 모델을 기록관리에 적용한다는 것은 바로 개념 레이어에서 기록관리에 필요한 부분을 섭취한다는 것을 의미한다. 결국 논리 레이어와 물리 레이어는 기록관리의 특수한 환경과 조건에 따라 다양하게 설계하고 구축하는 것이 가능하게 된다. 그런 면에서 OAIS 모델은 기록뿐만 아니라 모든 유형의 전자객체의 장기보존을 위한 모델로서 그 적용 범위가 매우 넓은 모델이라 할 수 있으며, 이러한 적용성은 모델의 기술 수준에 개념 레이어에서 이루어졌기 때문이다.

2. OAIS 정보모델과 AIP

OAIS(Open Archival Information System) 모델은 디지털 객체의 *장기보존(Long-term Preservation)*을 위해 만들어진 참조모델(Reference Model)로서 이미 국내에는 ISO14721 표준과 함께 논문“디지털 아카이빙의 표준화와 OAIS 참조모형”(이소연, 2002)을 통해 소개가 시작되었으며, OAIS 모델을 적용한 외국 사례들도 국내 연구자들에 의해 소개가 된 바 있다. 이 논문에서는 정보모델을 중심으로 OAIS를 소개하고, 소개된 개념 모델들을 전자기록의 시스템 구현에 실용적으로 적용해 보고자 한다.

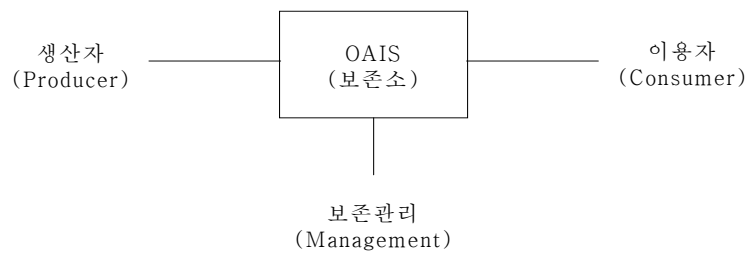


그림2-1. OAIS의 외부환경 모델

OAIS 참조모델은 정보를 생산하여 보급하는 생산자(Producer)와 그 정보를 접근 활용하려는 이용자(Consumer)를 위해 전략과 계획에 의거하여 정보를 보존, 관리하는(Management) 주체들을 위한 것이다. OAIS 참조 모델은 크게 기능 모델과 정보 모델로 구성되어 있다. 기능 모델(Function Model)은 디지털 정보의 장기보존을 위해 필요한 기능을 ①정보의 입수, ②저장, ③데이터유지관리, ④운영관리, ⑤보존계획, ⑥접근 및 이용 등의 6개의 대기능으로 나누어 설명하고 있다. 정보모델(Information Model)은 객체지향 모델링 언어인 UML²⁾을 사용하여 정보객체(Information Object), 정보패키지(Information Package), 정보컬렉션(Information Collection)이라는 개념을 통해 정보의 계층적인 구조를 설명하고 있다. 이러한 OAIS 모델은 Open이라는 말이 의미하듯이 모든 유형의 디지털 객체를 보존하기 위한 개방적이고 확장가능한 모델로서 디지털 정보의 장기보존 모델 표준 ISO 14721 로 채택될 만큼 매우 유용한 시사점을 주고 있다.

이 논문에서는 OAIS의 정보모델을 중심으로 전자기록의 장기보존을 위한 개념적 구성과 논리적 구조를 도출해 보고자 한다. 먼저 OAIS의 정보모델을 요약하면 다음과 같다.

2.1 OAIS 정보모델의 구성 요소

OAIS 정보모델의 목적은 정보를 효과적으로 장기보존하기 위해 필요한 요소들을 모형화하는 것이다. 정보모델을 구성하는 각 요소들은

2) UML은 Unified Modeling Language의 약자로, 객체지향 방법론에서 기능과 정보를 모델링할 때 사용하는 언어이다.

OAIS의 기능에 의해 교환되거나 관리되면서 구성의 변화가 이루어질 수 있으며, 정보모델은 기능에 의한 모든 변화 가능성을 포괄하는 요소들로 구성되어야 한다.

정보객체 모델

OAIS 모델에서 정보객체는 데이터객체(Data Object)와 표현정보(Representation Information)가 조합된 개념이다.

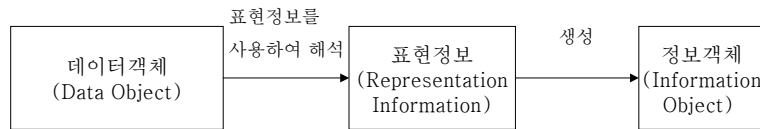


그림2-2. 데이터로부터 정보를 구성하는 과정

위의 그림은 전자적인 형태로 생산된 데이터객체가 OAIS에서 보존, 관리할 수 있는 대상으로서의 정보객체로 구성되는 과정을 보여준다. 여기서 데이터객체란 컴퓨터 상에 입력되어 있는 바이너리 데이터 자체를 말한다. 표현정보는 데이터객체를 수신하여 활용하고자 하는 사람 또는 컴퓨터가 해당 데이터를 이해할 수 있도록 해주는 정보를 말한다. 정보객체는 표현정보를 사용하여 해석된 데이터로서 컴퓨터의 화면을 통해 보여지는 데이터를 말한다. 일반적으로 모든 데이터들은 표현정보를 사용해서 해석되고 정보를 생성하게 된다.

표현정보의 예

예를 들어, 우리가 메모장에 'ABC123' 이라는 문자를 입력하고 텍스트로 저장을 하면 .txt 파일이 생기게 된다. 이때, 이 텍스트 파일은

‘ABC123’ 라는 문자들을 컴퓨터가 알아볼 수 있는 바이너리 값의 조합, 즉 0과 1의 조합인 비트스트림으로 만들어지게 된다. 이 때, 우리가 메모장에 쓴 ‘ABC123’ 이라는 문자 정보를 컴퓨터가 바이너리 데이터로 변환하여 저장할 때는 여러가지 방법이 있을 수 있으며, 가장 많이 사용되는 방법이 문자를 각각의 ASCII 코드 값으로 변환하여 저장하는 방법이다. ‘ABC123’ 이라는 문자가 ASCII 코드를 사용하여 데이터 객체화 되면 그 바이너리 비트스트림은 ‘01000001 01000010 01000011 00110001 00110010 00110011’ 가 될 것이며 총 6바이트로 구성될 것이다.

그렇다면, 만약 컴퓨터에 저장되어 있는 비트스트림이 ‘01000001 01000010 01000011 00110001 00110010 00110011’ 라고 했을 때, 아무런 해석 정보가 없는 상태라면 이것을 어떤 의미를 가진 정보로 볼 것이냐를 생각해 보면 답이 없음을 알게 된다. 이 비트스트림은 ASCII 코드에 의한 문자 값으로 저장된 것임을 알아야 해석이 가능하게 되는 것이다. 비트스트림으로 저장된 바이너리 데이터객체 자체는 그것이 어떤 정보인지를 해석하기 위한 기본적인 표현정보가 함께 있어야만 정보로서의 가치를 갖게 된다.



그림2-3. 비트스트림이 ASCII 코드라는 표현정보를 통해 해석되어 보여지는 과정

위의 그림에서 “ASCII 코드 문자열로 해석” 이라는 정보가 표현정보로서, “01000...”라는 비트스트림을 해석하여 ‘ABC123’이라는 의미있는 정보로 재구성하도록 해주는 것이다.

또 다른 예로, 한글이나 마이크로소프트 워드와 같은 문서편집기를 이용하여 문자들을 입력하고 저장하게 되면, 우리가 입력한 데이터들이 한글이나 마이크로소프트 워드의 자체방식에 의해 일정한 규칙을 가지고 비트스트림으로 변환되어 컴퓨터에 저장되는데 이때는 단순한 문자 정보 외에도 문자의 폰트와 문서의 형식 등에 대한 부가적인 정보들이 함께 저장되게 된다. 이처럼, 한글이나 MS 워드 등은 각자의 방식으로 비트스트림을 생성하여 저장하므로 이런 전자파일에 대해서는 어떤 소프트웨어를 사용하여 작성된 파일인지를 주요 표현정보를 작성해 주어야 하며, 가능하다면 해당 소프트웨어가 어떤 형식으로 비트스트림을 구성하는 지에 대한 정보까지를 표현정보에 포함해 준다.

정보패키지 모델

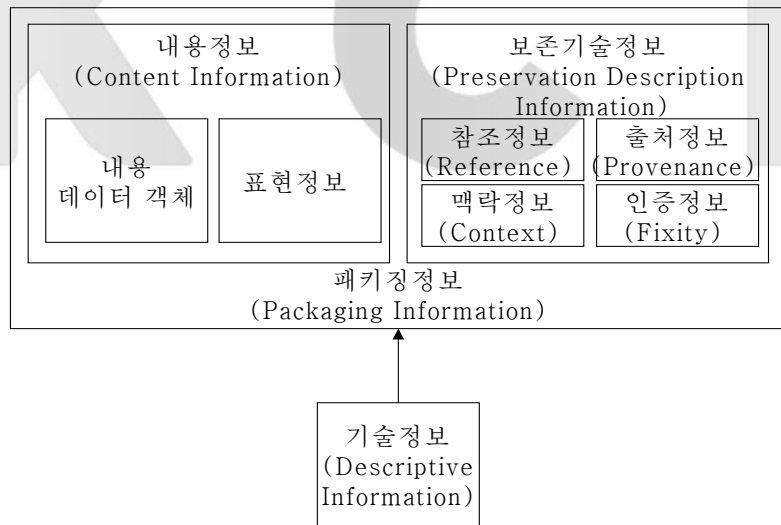


그림2-4. 정보패키지의 구성 개념도

OAIS 정보모델에서는 정보객체 개념을 사용하여 입수와 보존, 관리, 접근, 활용의 기본 단위를 형성하는 정보패키지(Information Package)를 정의하고 있다.

여기서 정보패키지는 내용정보(Content Information)과 보존기술정보(Preservation Description Information), 패키징정보(Packaging Information)가 합쳐진 개념적 구조체이다. 그림2-5에서 표현하고 있듯이 내용정보, 보존기술정보, 패키징정보, 기술정보 모두 앞에서 살펴본 정보객체 종류로 각각 데이터객체와 표현정보로 구성되며, OAIS에 저장되는 모든 정보들은 최소의 의미있는 기본단위로서 정보객체의 유형을 갖게된다.

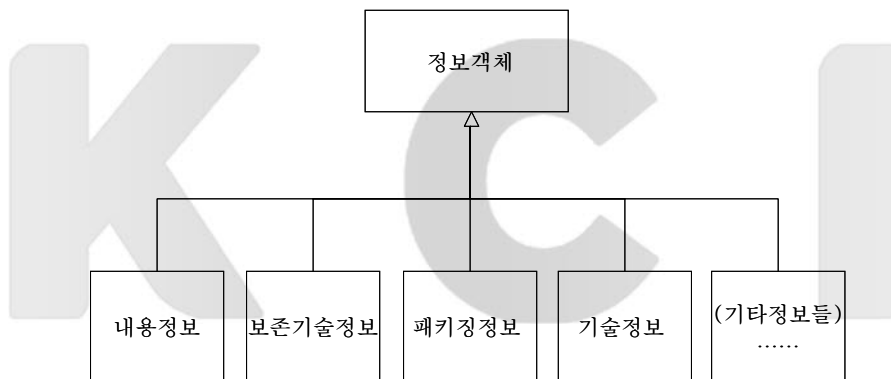


그림2-5. 정보객체의 종류

그림2-4에서, 내용정보와 보존기술정보는 하나의 캡슐로 싸인 것으로 보며³⁾ 패키징정보를 통해서 캡슐 안의 구성을 확인할 수 있다.⁴⁾ 이렇게 패키징된 전체가 OAIS에 저장되고 보존되며, 한편으로는 OAIS내

3) 여기서 말하는 캡슐은 개념적인 것으로, 내용정보와 보존기술정보는 떨어져서는 안되는 통합된 일체로서 하나의 패키지를 구성한다는 것을 의미한다.

4) 패키징정보는 패키지에 대한 메타데이터로서의 성격을 갖는다.

에 존재하는 정보패키지를 찾아내기 위해 필요한 기술정보(Descriptive Information)⁵⁾를 정의하게 되는데, 이 기술정보는 정보패키지의 보존기술정보와 패키징정보에서 추출된다.

표2-1. 정보패키지의 구성 객체

정보객체	내 용	
내용정보	<ul style="list-style-type: none"> ● 보존해야할 정보 집합으로 표현 정보를 동반하는 내용 데이터 객체 ● 생산자와의 협상을 통해 내용정보를 결정하게 됨 	
보존기술정보	출처정보	내용정보의 이력 정보로 과거와 현재 상태를 기술 (기원, 원천, 관리변화 등)
	맥락정보	내용정보와 외부환경과의 관계 정보 (생산배경, 다른 정보와의 관계 등)
	참조정보	내용정보에 대한 식별자 (식별자 지정 매커니즘은 분류시스템 등)
	인증정보	문서화되지 않은 방법으로 내용정보가 변형되지 않았음을 보장해주는 정보
패키징정보	<ul style="list-style-type: none"> ● 내용정보와 보존기술정보 검색에 필요한 정보내용정보와 보존기술정보 검색에 필요한 정보 	
기술정보	<ul style="list-style-type: none"> ● 원하는 내용정보가 어떤 패키지에 담겨 있는지 검색하는데 이용되는 정보 ● 단순히 표제 정보일 수도 있고, 목록서비스를 위한 속성값 전체일 수도 있음 	

내용정보는 보존의 목표 대상인 정보객체로서, 앞에서 설명한 대로 내용의 데이터객체와 데이터 객체를 이해할 수 있게 해주는 표현정보가 연결

5) 전자객체의 경우 정보시스템을 떠나서는 존재할 수 없으므로, 전자객체를 활용하고자 할 때 원하는 전자객체를 검색할 수 있도록 목록이나 인덱스, 키워드 등을 제공하는 것은 기본이라 할 수 있다.

되어 구성된다. 전자기록의 경우, 본문이나 첨부 파일 자체(또는 비트스트림)가 내용데이터객체가 되고, 각 파일(또는 비트스트림)이 어떤 형식으로 데이터를 표현하고 있는가에 대한 정보, 즉 한글 파일이라거나 엑셀 파일이라는 등의 생산 소프트웨어 정보로부터 한글과 엑셀이 화면상에 보여지는 데이터를 내부적인 비트스트림으로 변형 저장할 때의 방식에 대한 정보 모두가 표현정보가 된다.

*보존기술정보*는 내용정보를 보존하기 위해 필요하고, 내용정보를 명확하게 식별해 내고, 내용정보가 생산된 배경을 이해하기 위한 정보이며 출처정보(Provenance), 맥락정보(Context), 참조정보(Reference), 인증정보(Fixity)의 네 가지 유형으로 구분된다. 출처정보는 내용정보의 소스 또는 원생산자를 나타내며 원본 생산에서부터 관리권의 변화와 이력(처리 이력 포함) 내역을 포함한다. 맥락정보는 내용정보가 정보패키지 외부에 있는 다른 정보와 어떤 관계를 가지고 있는지를 나타내는 것으로 예를 들면, 내용정보가 생산된 이유에 관한 기술이나 다른 내용정보와 어떤 관련이 있는지에 관한 정보 등이다. 참조정보는 내용정보를 식별할 수 있는 이상의 역할을 하는 참조코드들로서 시스템 id, 서비스 id, ISBN, URL 정보, 핸들 등이 여기에 속한다. 인증정보는 내용정보를 허가되지 않은 변경으로부터 보호하기 위한 정보로서 체크섬이나 CRC와 같은 값들로 구성된다.⁶⁾

6) 체크섬은 수신자가 같은 수의 비트가 도착했는지를 확인 할 수 있도록 전송단 위 내의 비트 수를 세는 것이다. 만약 계산이 맞으면, 오류없이 원만하게 수신된 것으로 간주된다. CRC(Cyclic Redundancy Checksum 또는 Cyclic Redundancy Check)는 데이터의 무결성을 체크하는데 사용되는 대표적인 “디지털 인증”방법으로, 예를 들어, 전송이 이루어졌을 때 전송전의 파일과 전송 후의 파일이 서로 동일한 내용인지 확인하고자 할 때 원래 파일에 대한 CRC값을 구하고 전송된 파일의 CRC값을 구해서 두 개의 값이 같으면 두 파일의 내용이 동일하다고 추정할 수 있는 인증방법이다.

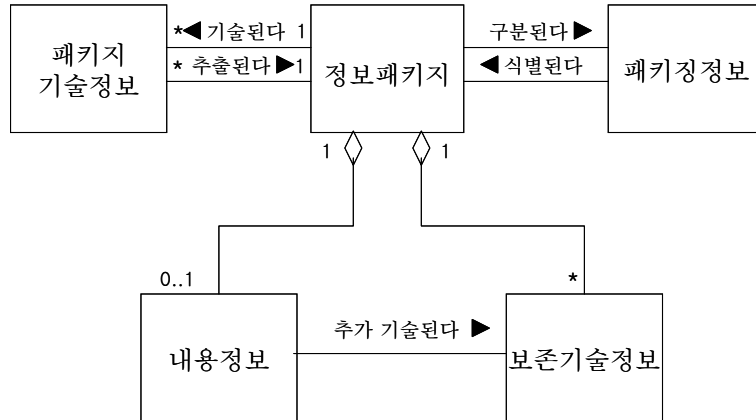


그림2-6. 정보패키지와 연결 정보

패키징 정보는 내용정보와 보존기술정보의 실제적인 혹은 논리적인 결합관계를 나타내는 정보이다. 예를 들어, 어떤 패키지의 내용정보과 보존 기술정보가 CD-ROM에 각각 파일로 저장되어 있을 때 패키징 정보는 “ISO 9660 volume/파일구조(=파일디렉토리와 파일명)”⁷⁾의 형태로 CD-ROM 안에 내용정보와 보존기술정보 파일 저장규격과 위치 및 이름 정보로 표시될 것이다.

기술정보는 원하는 내용정보를 담고 있는 패키지가 어떤 것인지 찾아내는 데 사용하는 정보로 정보패키지의 표제 정보만 가지고 있을 수도 있고 목록서비스에서 검색할 수 있도록 전체 메타데이터의 집합일 수도 있다. 기술정보는 일반적으로 보존기술정보의 일부와 내용정보의 핵심어를 추출하여 작성하며, 사용자가 원하는 패키지를 찾는데 유용한 색인정보 역

7) ISO 9660은 1988년도에 국제표준화기구가 제정한 CD-ROM 을 위한 파일 시스템을 정의한 표준규격으로 마이크로소프트의 윈도우즈, 매킨토시의 OS, 유닉스 등 서로 다른 시스템에서 데이터를 교환할 수 있도록 하기 위해 제정되었다.

할을 하게 된다.

2.2 정보패키지의 유형 - SIP, AIP, DIP

OAIS 참조모델에서는 정보패키지의 유형을 제출정보패키지(SIP: Submission Information Package), 보존정보패키지(AIP: Archival Information Package), 배부정보패키지(DIP: Dissemination Information Package)의 세 가지로 구분하고 있다.

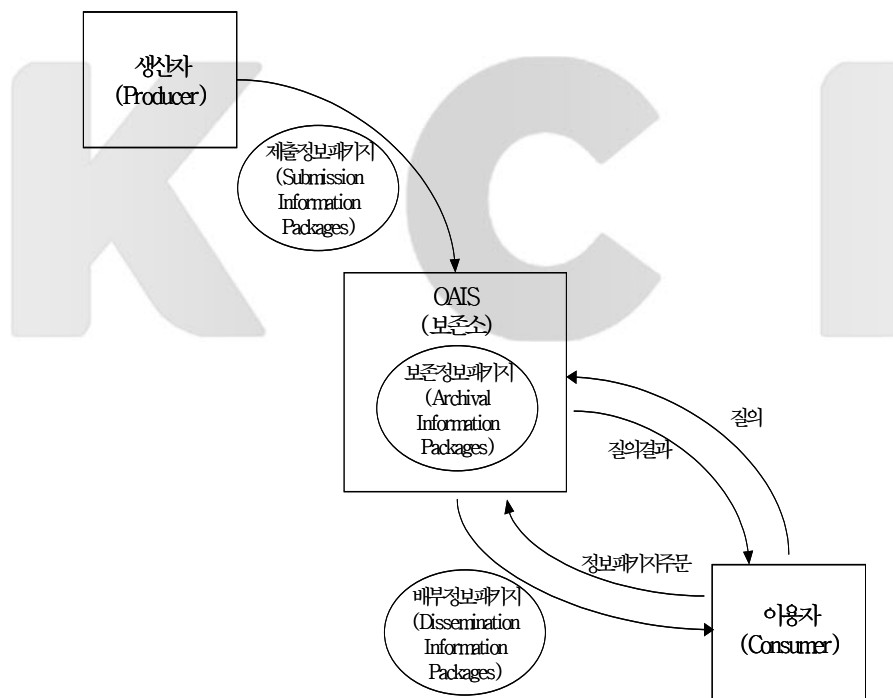


그림2-7. 정보패키지의 용도별 유형

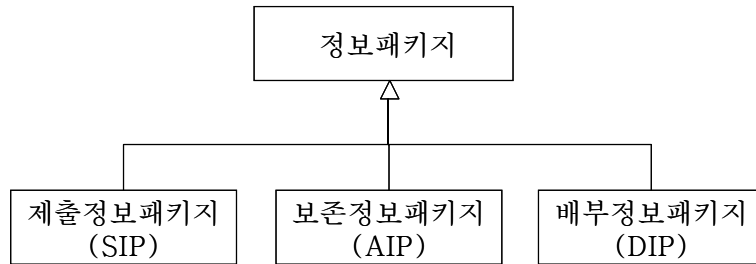


그림2-8. 정보패키지의 유형

SIP는 생산자(Producer)가 OAIS로 보내는 정보패키지로 생산자와 OAIS 간의 협약을 통해서 내용과 형태가 결정된다. 하나의 SIP가 하나의 AIP로 구성되는 경우가 대부분이나 여러 개의 SIP가 하나의 완전한 내용정보를 구성하고 보존기술정보를 연결하여 AIP 하나로 구성될 수도 있다. 또한, 하나의 SIP안에 여러 AIP의 정보가 들어있을 수도 있다. 어떠한 경우에도, SIP는 일정한 형태, 즉 개념적으로 동일한 구조를 가지고 있다.

AIP는 내용정보와 이에 대한 보존기술정보의 완전한 집합으로 만들어지며, 패키징정보는 OAIS가 정한 내부 표준을 준수한다. AIP는 다른 여러 개의 AIP의 집합으로 구성되기도 하며, OAIS로 입수된 하나 이상의 SIP가 하나 이상의 AIP로 변형되어 저장된다.

DIP는 AIP의 일부 또는 전부를 이용자(Consumer)에게 제공하기 위해 만들어진 정보패키지로, AIP의 집합을 포함하여 만들어질 수도 있고, 때로는 보존기술정보를 포함하지 않고 만들어질 수도 있다. 패키징정보는 일정한 형식으로 표현되어야 이용자(Consumer)가 요청한 정보를 구분하여 볼 수가 있으며, 이용자의 요구사항이나 DIP를 전달하는 매체 종류에 따라 다양한 형태를 갖게 된다.

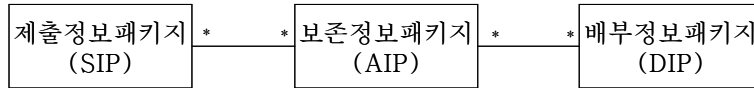


그림2-9. SIP, AIP, DIP의 대응 관계

표2-2. SIP, AIP, DIP의 내용과 특징

정보 패키지	내용	비고
SIP	<ul style="list-style-type: none"> ● 생산자가 OAIS로 보내는 정보패키지 ● 형태와 내용은 OAIS와 생산자 간의 협약에 따름 ● 내용정보와 보존기술정보로 구성되며 표현정보와 연결됨 	<ul style="list-style-type: none"> ● 다수 SIP가 동일한 표현정보를 이용하면 표현정보는 한번만 제공함 ● 다수 SIP에 동일한 보존기술정보가 적용되면 내용정보 없는 SIP 형태로 제공함 ● 기술정보는 SIP 제출시점에 관계없이 제공됨
AIP	<ul style="list-style-type: none"> ● SIP를 기반으로 OAIS에 보존하기 위해 변형된 정보패키지 ● 내용정보와 보존기술정보의 묶음, 혹은 AIP들의 집합 	<ul style="list-style-type: none"> ● 패키징 정보는 내부의 표준을 따름 ● AIP 기술정보를 이용해 이용자가 원하는 내용정보를 찾고 주문 가능하도록 함
DIP	<ul style="list-style-type: none"> ● 이용자의 주문에 대해 AIP의 전체 또는 일부를 제공하는 정보패키지 ● AIP들의 집합이거나 보존기술정보 없는 AIP 등 	<ul style="list-style-type: none"> ● 배포매체와 이용자요구에 따라 패키징정보가 다양함 ● DIP 기술정보는 DIP 전송 이전, 이후, 전송 시 제공 가능

AIP의 정보요소

정보패키지로서의 AIP는 보존시스템에 실재하면서 보존되는 대상으로서 장기적인 안정성을 확보할 수 있는 구성을 갖추어야 하며, 한편으로는 이용자들의 이용을 용이하게 해줄 수 있는 구성을 갖추어야 한다. 후자의

측면을 고려하여 앞에서 살펴본 일반적인 정보패키지의 구성요소에 연결 요소를 추가하여 AIP의 구성을 정의해 보면 다음과 같다.

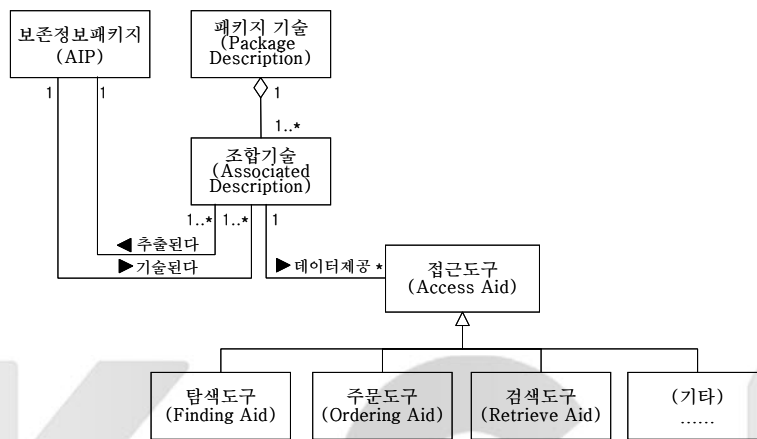


그림2-10. AIP의 연결요소 정보 구성

표2-3. AIP의 정보 요소

구성 요소	내용정보	● 보존해야할 내용 데이터 객체와 표현정보
연결 요소	보존기술정보	● 출처정보, 맥락정보, 참조정보, 인증정보로 구성
	패키징정보	● AIP를 구성하는 하위요소를 한정하고 식별해주는 정보
연결 요소	기술정보 ⁸⁾	● 이용자가 AIP를 탐색, 분석, 주문을 하는데 필요한 정보
		● 접근도구(Access Aid) ⁹⁾ 를 위해 필요한 수만큼 조합기술 (Associated Description) 정보를 구성
		● 각 조합기술(Associated Description)은 하나 이상의 접근도구를 지원하는 데이터를 포함할 수 있도록 구성

8) 그림2-10의 패키지기술에 해당

9) 접근도구(Access Aid)에는 탐색도구(Finding Aid), 검색도구(Retrieve Aid), 주문도구(Ordering Aid) 등이 있다.

구성요소의 내용정보는 SIP와 DIP 모두 동일하게 필요로 하는 요소이나, 구성요소 중 보존기술정보와 연결요소는 정보패키지 유형마다 필요한 요소가 다르게 된다. 예를 들어, 연결요소 중 기술정보의 경우는 SIP와 DIP에는 필요없는 요소이며, 패키징정보의 경우는 SIP와 DIP에도 필요한 요소이나 정보를 기술하는 방식이나 수준은 매우 다를 것이다.

2.3 정보의 기본단위와 컬렉션 - AIU와 AIC

OAIS에 저장 보관되는 AIP는 기본단위인 AIU(Archival Information Unit)와 기본단위들의 집합체인 AIC(Archival Information Collection)라는 두 가지 유형으로 분화된다.

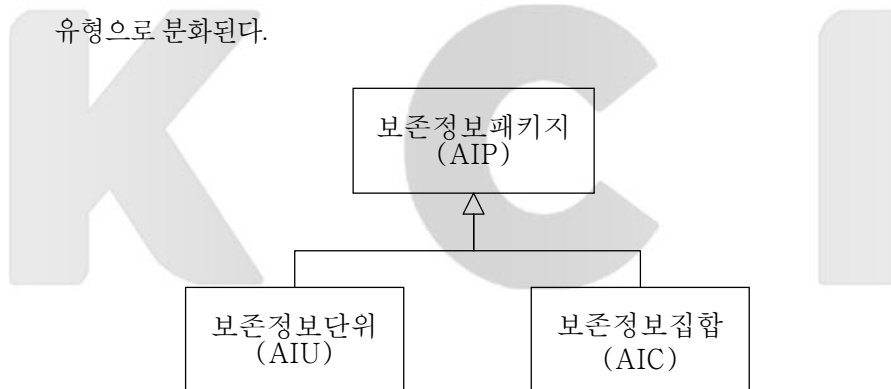


그림2-11. AIP의 분화 유형 - AIU와 AIC

AIU란 보존소에서 저장하는 정보의 기본단위로 하나의 내용정보와 보존기술정보로 구성된다.

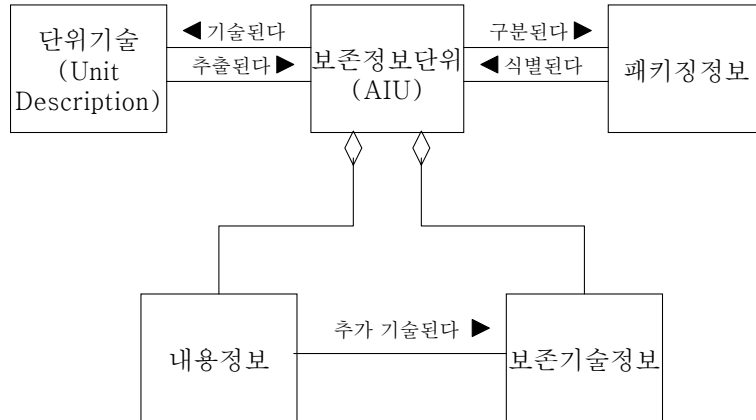


그림2-12. AIU의 구성요소 모형

이 때, AIU를 구성하는 단위기술(Unit Description)은 정보객체가 보존소로 입수될 때 내용정보와 보존기술정보로부터 추출되고 OAIS에 의해 유일 식별자 등과 같은 정보가 부가되어 생성된다. 이 단위기술 정보는 인가된 이용자가 저장되어 있는 AIU를 검색할 수 있도록 제공되어야 한다.

AIC란 보존관리자가 정한 범주에 따라 모인 AIP들의 집합을 말한다. OAIS는 전체 시스템 안에 최소한 하나 이상의 AIC를 갖게 되며, AIC는 그 자체가 내용정보와 보존기술정보를 가진 완전한 AIP이다. OAIS에는 수많은 정보객체가 존재하므로 이용자가 원하는 정보를 찾기 위해 AIU와 해당 단위기술정보를 검색하는 일은 매우 힘든 일이다. AIC는 동일한 범주의 AIU들을 묶어서 하나의 집합을 정의하고 그 집합의 출처, 맥락, 참조, 인증 정보를 기술하도록 하여 집합적 기술과 활용을 가능하게 한다.

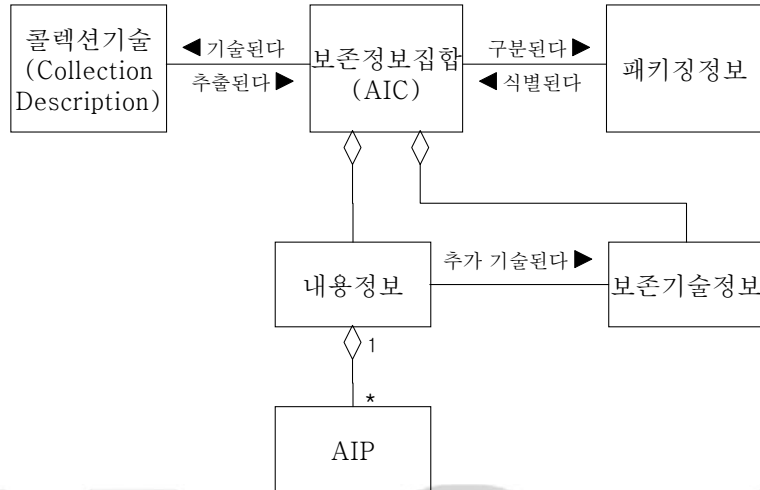


그림2-13. AIC의 구성 모형

AIC의 기술정보, 즉 콜렉션기술은 전체 콜렉션을 기술해주는 내용과 콜렉션의 멤버 각각을 기술해주는 0개 이상의 멤버기술을 포함한다. 즉, 해당 AIC에 속하는 AIP들에 대한 기술 내용이 포함될 수 있다.

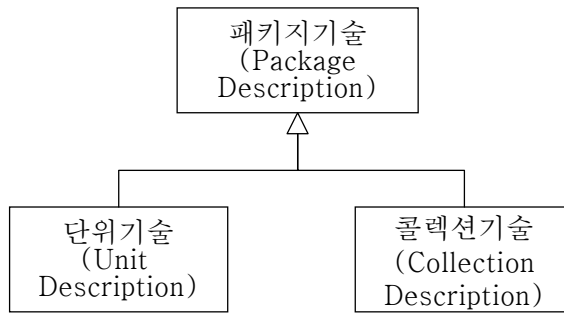


그림2-14. 기술정보의 분화 유형

2.4 OAIS 정보모델의 기록관리 적용

OAIS의 정보 모델은 어떤 유형의 정보이든 보존을 위한 시스템으로 입수될 때, 보존이 될 때, 배부가 될 때 등 어느 시점에서든지 동일한 정보패키지의 개념으로 구조화하여 취급하는 단순하고도 융통성 있는 모델이다. 전자기록 역시 OAIS의 정보 모델을 이용하여 구조화가 가능하며 특히 영구기록 보존 단계의 기록관리시스템의 기록 정보 모델에 참조될 만하다. 그러나, 기록관리가 전자기록 중심인 시점이 되면서부터는 전자기록이 영구보존 단계의 시스템에 이관되기 전부터 일관되게 영구 보존 포맷으로 저장, 보관되는 것이 전자기록의 안정성과 진본성 보장에 도움이 될 수 있다는 점에서 OAIS의 정보모델은 영구보존 단계 이전의 기록관리시스템에도 적용이 가능한 참조모델이라 할 수 있겠다.¹⁰⁾

OAIS 모델의 정보패키지는 물리적인 구조체가 아니라 개념적인 구조체로 이해된다. 개념 레이어에서 정의한 정보패키지를 논리, 물리 레이어에서 어떻게 구조화하고 설계할 것인지는 구체적인 상황에 따라 다양한 적용이 가능한 것이다. 즉, OAIS 정보모델에서 정의한 하나의 정보패키지는 물리적으로 하나의 파일로 저장될 수도 있고 여러 개의 파일로 나뉘어 저장될 수도 있는 것이다. 3장에서는 OAIS 보존정보패키지 개념을 참조하여 먼저 기록의 AIP의 개념적 구성을 논의하고, 주요한 보존전략에 따라 기록 AIP의 논리적 구조가 어떻게 달라질 수 있

10) 행정자치부에서 수행한 기록관리시스템 혁신을 위한 ISP 프로젝트 결과보고에 따르면, 기록관 단계에서 사용할 기록관리시스템에서 5년 이상의 보존기간을 가진 전자기록물을 입수했을 때 문서보존포맷 PDF/A를 적용하여 문서단위 포맷전환을 하고, 기록 건 단위에서 하나의 XML 문서로 인캡슐레이션하여 저장하는 것을 제안하고 있다.

는지에 대해 논의할 것이다.

현재의 OAIS 정보 모델이 기록관리에 곧바로 적용되는 데 한 가지 아쉬운 점이 있다면 그것은 OAIS 참조모델에서는 기록관리의 기본 틀인 분류 체계에 대응하는 명시적인 개념이 없다는 것이다. 기록의 경우 해당 기록이 어떤 조직에서 또는 어떤 업무기능의 맥락에서 생산, 접수되었는지를 분류체계에 반영하여 이에 따라 기록을 분류, 등록, 기술하는 것이 관리의 첫 걸음인데, OAIS에서는 기록의 분류체계가 갖고 있는 계층성과 계층별 기술을 가능하게 하는 AIU와 AIC에 대한 개념만 제시하고 있을 뿐이다. 그러므로, 기록관리에서 구체적으로 필요한 기록철(File), 시리즈(Series), 레코드 그룹 등의 개념은 OAIS 참조모델의 AIC를 창조적으로 적용 구현해 나가야 할 부분이다. 물론, OAIS의 정보모델이 기록관리의 분류체계를 구현하는 데 용이하지 않다면 이 모델이 전자기록의 장기보존 측면에서 주는 아이디어만을 기록관리에 차용하고 분류체계 등의 구성에 대해서는 다른 모델을 만들어 조합해 나갈 수도 있겠다.

이 논문에서는 기록철, 시리즈 등의 상위개념은 향후의 과제로 남겨두고 전자기록 건 하나에 대응하는 개념으로서의 기록AIP에 대한 구성과 구조를 장기보존의 관점에서 어떻게 만들어갈 것인지에 대해 OAIS의 정보모델을 적용해 보고자 한다.

3. 기록AIP 구성과 구조

OAIS가 보존, 관리 대상으로 하는 정보객체는 일반 전자문서, 전자저널,

과학기술 데이터세트, 웹페이지, 시청각 디지털 저작 등 디지털화된 모든 유형이다. 이 논문에서는 OAIS가 포괄하는 정보객체의 유형 중에서 업무 수행과정에서 생산 접수된 전자기록이라는 정보객체에 초점을 맞춰 논의를 진행하고자 한다. 또한, 전자기록을 안정적으로 장기간 보존하기 위해서는 기록의 AIP를 어떤 구성과 구조로 보존 관리해야 하는가에 초점을 맞춰 논의를 진행하고자 한다.

논의에 앞서 한 가지 전제를 확인하면, 전자기록의 장기보존을 위한 포맷은 전자기록을 보존하는 주체에 의한 *보존전략 및 계획(Preservation Strategy & Planning)* 수립이 선행되어야 논의가 가능하며 내용이 정해질 수 있다는 것이다. 보존전략 및 계획 수립 활동은 기록의 생산자와 이용자 양쪽에 대해 현재 사용하고 있는 컴퓨터 플랫폼과 선호하는 소프트웨어 등에 대해 모니터링하고, 신규 디지털 기술에 대한 동향을 살펴면서 현재 보존 중인 기록의 마이그레이션 계획을 수립하고 새로운 포맷으로 생산되어 등록되는 기록에 대한 보존 전략을 수립하며, 표준을 개발하고 신규 AIP를 설계하는 등, 기록보존시스템에 저장된 기록은 장기간에 걸쳐 안정적으로 이용될 수 있다는 것을 보증하기 위한 활동으로 이루어진다. 이러한 전략과 계획에 의해 전자문서로 생산 또는 접수된 기록을 등록하여 관리할 때 본문 파일의 포맷을 그대로 보존할 것인지, 아니면 표준으로 정한 보존 포맷으로 전환하여 보존할 것인지가 정해지며, 원래의 본문파일과 보존포맷으로 전환된 파일을 둘 다 보존할 것인지의 여부도 정해진다. 최초의 전자 기록 AIP는 이러한 전략과 계획에 의해 그 구성과 구조가 결정되게 되며 또한 전략과 계획에 의해 보존 관리되는 과정에서 그 구성과 구조가 변형될 수 있다.

이러한 보존전략 및 계획의 수립은, 기록의 이용자들이 누구이며, 어떤

형태로 기록정보를 서비스 받고자 하는가, 이러한 이용자의 요구에 대해 기록보존관리자는 어느 수준에서 서비스를 제공하기로 결정했는가 등에 따라 정해지게 된다. 기록관리 조직도 일반적인 조직들과 마찬가지로 한정된 자원과 예산으로 조직과 시스템을 구축하고 운영해야 하므로 비용 대비 효과를 고려하여 서비스 수준을 결정하고 시스템의 성능을 보완해 가야 한다. 이러한 현실적인 한계들이 AIP의 구현 양상에 중요한 영향을 미치게 된다.

OAIS는 기록을 포함한 다양한 종류의 디지털 정보를 장기보존하기 위한 모델로 제시된 것이므로 기록정보에 대한 특별한 고려사항을 담고 있지는 않다. 이 논문에서는 기록에 적용된 AIP를 ‘기록AIP’라고 부르기로 하고 기록AIP의 개념적 구성과 논리적 구조에 대해 3-레이어 접근법에 따라 논의해 보기로 한다. 이때의 기록AIP는 기록 자체의 구성요소 뿐만 아니라 그 기록이 보존되는 이유가 이용자의 접근활용을 위한 것이라면 연결요소를 유지해야만 의미가 있다는 점에서 연결요소까지를 포함한 확장된 의미로서의 기록AIP를 다루고자 한다.

3.1 개념 레이어 - 기록AIP의 구성

“기록은 내용(Content), 맥락(Context), 구조(Structure)로 구성되며, 업무활동의 사실(Content)을 일정한 서식(Structure I)에 담아 다른 기록들과 함께 업무 기능과 활동의 증거 및 맥락(Content & Context)을 전달해 주며, 파일링시스템 등의 물리적 구조체(Structure II)로 존재하게 된다”¹¹⁾고 하는 기록의 구성요소에 대한 고전적인 정의에서 출발해 보자.

11) Shepherd & Yeo, 2003

표3-1. 기록의 구성요소

구성요소	표현내용
내용(Content)	<ul style="list-style-type: none"> ● 활동의 사실 ● 업무 지식과 정보
맥락(Context)	<ul style="list-style-type: none"> ● 업무기능과 활동의 증거
형식(Structure I)	<ul style="list-style-type: none"> ● 문서 서식 ● 기록하는 방식
구조(Structure II)	<ul style="list-style-type: none"> ● 논리적 관련성 ● 물리적 구성

전자기록의 경우, 위 표의 “내용”에 해당하는 것이 전자문서 파일 안에 들어있는 내용이 될 것이고, “맥락”은 전자기록이 생산되게 된 배경에 해당하는 업무기능의 분류 정보와 조직 정보가 될 것이며, “형식”은 전자문서에 적용된 약속된 포맷구조가 될 것이며, “구조”는 기록이 몇 개의 전자문서 파일로 구성되고 각 전자문서 파일은 어느 것이 본문이고 어느 것이 첨부인지 등의 논리적인 구성 정보와 각 전자문서가 어떤 소프트웨어에 의해 생산되어서 어느 매체에 어떤 이름으로 저장되었는가 등에 대한 물리적 정보가 될 것이다.

“전자기록 진본성 유지를 위한 기록관리시스템 기능 및 정보모형”에 따르면 전자기록이 진본성을 유지하면서 중장기적으로 안정적으로 보존되기 위해서는 기록의 메타데이터가 다음과 같은 8가지의 범주에서 상세히 기술되어야 한다.

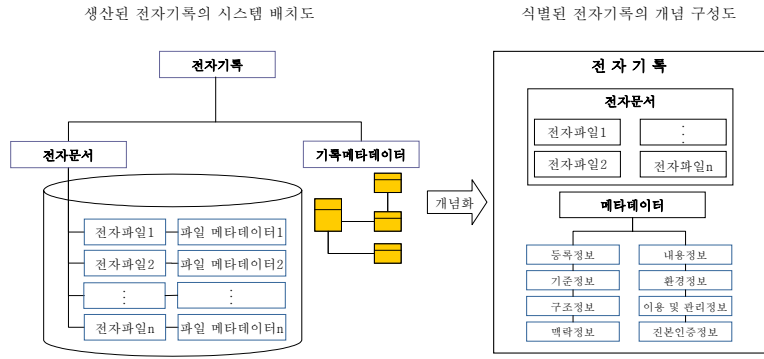


그림3-1. 전자기록의 개념적 구성도¹²⁾

표3-2. 기록의 메타데이터 그룹¹³⁾

메타데이터 그룹	내용	요소
등록정보	● 기록의 유일한 식별기호, 날짜 등과 같이 등록 및 확인에 필요한 정보	식별자 날짜 위치값
기준정보	● 기록의 권한관리와 처리기준 정보 ● 기록의 이용과 관리 조건에 대한 정보	권한관리 처리일정
구조정보	● 기록의 물리적·논리적 구조에 관한 정보 ● 보존관리를 포함하여 기록에 행해질 관리조치를 통제하기 위한 정보	기록유형 기록계층 보존이력
맥락정보	● 기록의 생산배경이나 기록을 생산한 업무기능 정보 ● 다른 기록과의 관계정보	기관 기능 법규 관계
내용정보	● 기록자체의 내용을 설명하는 정보	제목 주제 파일내용
환경정보	● 전자기록의 형태나 생산 및 이용을 위해 필요한 기술적(technological) 환경에 대한 정보	포맷 소프트웨어 하드웨어
이용 및 관리정보	● 기록이 등록된 시점부터 폐기될 때까지 기록에 취해진 관리행위나 이용에 대해 감사추적하기 위한 정보	관리이력 이용이력
진본인증정보	● 이관대상 기록이나 기록관리시스템에서 관리하는 기록에 불법적인 변화가 없음을 보증하는 정보	기록인증 전자파일인증

12) 김익한, forthcoming

13) 김익한, forthcoming

전자기록의 구성요소를 그림으로 나타내보면 크게 전자문서 파일들과 메타데이터 집합으로 볼 수 있다. 그림3-1에서 왼쪽은 전자기록이 생산되었을 때 기록을 구성하는 요소들이 시스템에 어떻게 배치되어 있는지를 나타낸 것으로 기록을 구성하는 전자파일들은 파일시스템에 저장되고 각 파일단위의 메타데이터를 갖게 되며, 기록자체의 메타데이터들은 주로 별도의 데이터베이스에 저장되어 있음을 표현하였다. 오른쪽 그림은 식별된 하나의 전자기록의 개념 구성도로 전자파일들과 파일단위와 기록단위의 메타데이터를 합한 전체 메타데이터를 유형별로 재정의한 구성을 보여주고 있다.

기록AIP의 메타데이터에는 기록의 메타데이터, 보존을 위한 메타데이터, 기술적인(technological) 메타데이터 3가지 종류가 있다. 기록의 메타데이터는 기록의 내용이나 맥락과 관련된 정보로서 DIP에도 포함되는 주요 내용이 되며, 보존을 위한 메타데이터는 기록관리시스템에서 보존관리를 위해 필요한 내용으로 DIP에는 거의 포함되지 않는다. 기술적인 메타데이터는 DIP에도 포함되어야 하는 필수적인 정보로 이 정보가 있어야 이용자들이 DIP의 전자파일을 해석하여 활용할 수 있다. 그런데, 전자기록관리시대에 중요하게 취급해야 할 메타데이터가 또 있다. 바로 기록AIP에 접근하고 활용하기 위해 필요한 메타데이터에 대한 것이다. 이 시점에서 메타데이터에 대한 개념의 확장이 필요하다. 메타데이터를 만드는 가장 궁극적 목표는 원 자원(Original Resource) 즉, 기록자체를 찾기 위한 안내자의 역할을 하기 위해서이므로 더블린 코어 메타데이터에서도 메타데이터는 자원기술(Resource Description)이 아닌 자원 발견(Resource Discovery)의 개념으로 사용되고 있다. 기록의 메타데이터도 보존과 관리를 위한 메타데이터 이외에도 접근 및 활용을 위한 메타데이터(색인어 등)의 구분이 필요하며 이러한 메타데이터는 기록의 내

용정보는 아니지만 기록과 함께 패키징되어 이관될 필요가 있으며 메타데이터 유형별로 관리 방식과 활용 방식이 달라져야 한다. 앞에서 살펴본 OAIS 정보모델에서의 연결정보가 이런 메타데이터에 해당한다.

전자기록의 메타데이터를 잘 구비하여 유지하는 것의 목적은 기록이 자체만으로도 내용을 충분히 이해할 수 있도록 하기 위해서이다. 메타데이터는 전자기록의 관리와 보존을 위해 없어서는 안 되는 것으로 지적 통제, 관리이력 기술, 보존기술, 진본성 표현, 장기적인 가독성, 보존, 검색지원 등의 다양한 기능을 지원한다. 전자기록을 위한 메타데이터 표준은 현재 작성 중이며, 보존기술(Archival description)을 위한 ISAD(G) 표준은 전자기록의 디지털 특성과 속성에 대한 몇 개의 필드만을 제공하고 있다.

이러한 기록의 구성요소가 고전적인 기록의 구성요소 및 OAIS의 AIP 구성요소들과 어떻게 매핑되는지를 도해해보았다.

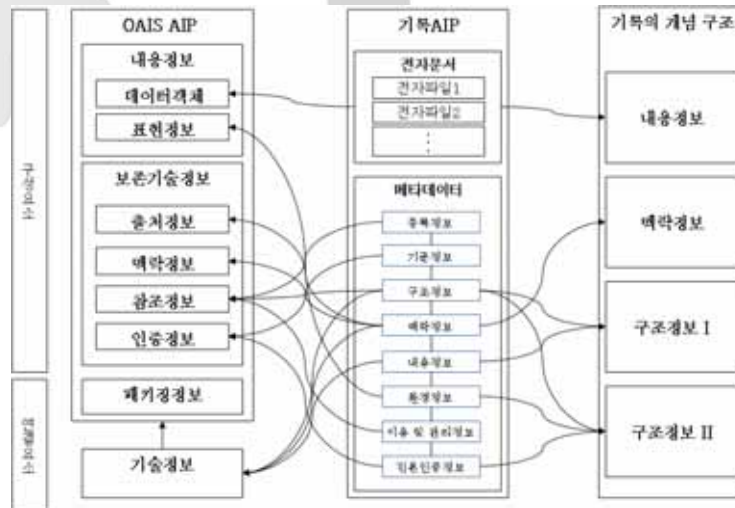


그림3-2. OAIS, 기록AIP, 고전적 기록개념구성 간의 매핑

그럼에서 이해되는 것처럼 고전적인 기록의 개념구조에서는 강조되지 못했던 기록AIP의 연결요소가 OAIS 모델을 적용하면서 분명하게 드러나게 된다. 전자기록은 정보시스템을 떠나서는 존재할 수 없으며 기본적으로 정보시스템의 서비스를 통해 이용자에게 제공된다는 환경적 제약을 생각하면 이용자의 탐색, 검색, 주문 등을 통한 기록의 접근과 이용을 위해 기록AIP에 대한 일정 부분의 정보를 연결정보로서 제공하는 것이 필요하다. 이러한 연결정보를 기록AIP의 구성요소로 반영하는 것이 전자기록관리 시대의 특성을 반영하는 핵심 변화 내용이 될 것이다.

보존전략에 따른 구성변화

생산된 기록은 SIP 형태로 전자기록관리시스템에 제출되고, 기록관리자가 수립한 보존 전략에 따라 AIP로 재구성되어 보존 관리되며, 이용자의 요구에 의해 DIP가 생성되어 배포된다. 기록AIP는 관리 과정에 따라 또는 이용자의 요구에 따라 그 상세한 내부 구성이 달라질 수 있다.

기록AIP의 내부 구성에 중요한 영향을 끼치는 관리행위는 앞서도 언급했듯이 기록에 대한 보존 전략이다. 조직에서의 기록에 대한 보존 전략은 어떤 기록정보가 보존되고 이 정보를 디지털 리퍼지토리안에서 어떻게 관리할 것인지를 결정하는 것으로, 전자기록이 향후에도 생산시점과 동일하게 표현가능하게 남아 있도록 하면서 동시에 진본성과 의미 해석이 가능하도록 보장하는 방법을 정하는 것이다. 현재로는 디지털 정보로서의 전자기록에 대해 마이그레이션(Migration)과 에뮬레이션(Emulation)이 주요한 보존 전략으로 제시되고 있으며, 양자의 장단점에 대해서는 많은 논의가 이루어지고 있다. 물론, 양자의 접근 방법이 서로 배타적이지 않으며 상호 보완적인 전략으로서 채택이 가능하므로 기록관리조직에서는 서로 다른

전자문서의 유형에 대해 두 가지 전략을 따로 적용할 수도 있다.

어떤 전략을 선택하든 기록의 생명주기 동안 마이그레이션과 에플레이션은 반복되어 적용될 것이며, 특정 기록 건에 대한 전략이 어느 시점에 바뀔 수도 있다고 본다. 기록에 대한 가독성을 최대한 보장하기 위해, 그리고 미래의 어느 시점에서 해당 기록에 대한 마이그레이션과 에플레이션 중 전략 전환이 가능하기 위해서는 기록이 디지털 리파지토리에 원래의 파일 포맷과 보존 포맷 둘 다 함께 통합되어 저장되어 있어야 한다. 미래의 기술적 진보에 따라서 지금 당장은 활용하지 못하는 비트스트림도 나중에는 보존 조치에 유용하게 사용될 수 있기 때문이다.

하나의 기록에 속한 전자파일의 포맷이 여러 가지 존재하는 경우 이용하는 그 기록에 대해 여러 포맷이 존재하고 있으며 그 시점에 접근이 가능한 포맷이 어떤 것들이 있는지를 확인할 수 있어야 하며, 권한이 있는 경우 원하는 특정 포맷의 기록에 접근할 수 있어야 한다. 이러한 전략의 다양성과 적용변화 가능성을 고려한 기록AIP구성은 다음과 같이 도해될 수 있다.

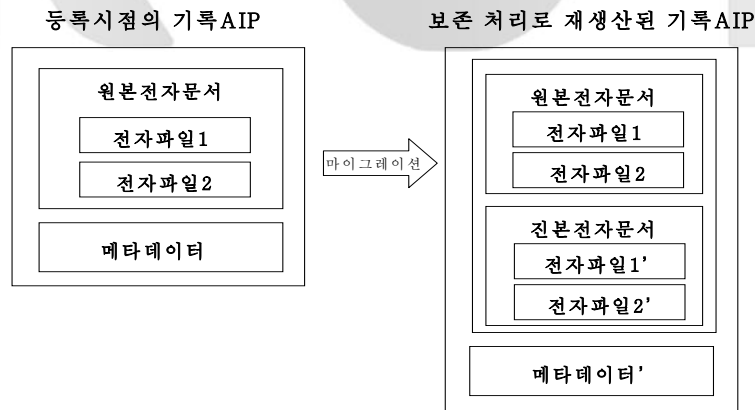


그림3-3. 기록AIP의 변화과정

위의 그림에서 등록시점의 기록AIP는 원본 전자파일들과 메타데이터로 구성되지만 이후 보존 처리의 필요에 의해 각 전자파일에 대한 마이그레이션 파일이 생성될 수 있다. 마이그레이션 파일의 포맷을 원본과 유사한 형태로 가져갈 것인지 아니면 향후의 추가 마이그레이션 필요성을 최소한으로 하는 안정적인 보존 포맷으로 가져갈 것인지는 보존전략에 따르게 된다. 또한, 원본을 마이그레이션한 사본 전자파일을 진본으로 인증하는 절차 또한 조직의 전자기록 진본성 확보 요건에 맞춰 진행되며 보존처리 및 진본인증 과정에 대한 관리이력 정보가 메타데이터에 반영되게 된다. 그림3-3에서 마이그레이션된 결과로서의 “진본전자문서”는 이러한 진본인증과정을 반영하여 명명한 것이다. 이러한 마이그레이션의 결과로 기록AIP는 전체적인 볼륨이 커지게 되므로 원본과 진본사본을 둘 다 보존할 것인지 여부 또한 보존 전략에 따라 결정하게 된다.

기록AIP의 저장공간

이렇게 등록, 보존관리되는 기록AIP는 다음과 같은 3 종류의 개념 영역에 나뉘어 저장된다.

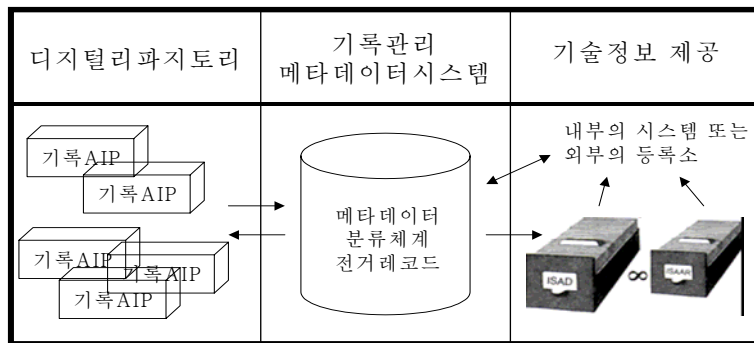


그림3-4. 기록AIP의 저장 공간 개념도¹⁴⁾

디지털 리파지토리는 기록AIP의 전자파일을 중심으로 하는 내용정보가 저장되는 공간이며, 기록관리메타데이터 시스템영역은 기록AIP의 메타데이터와 분류체계 정보가 저장되는 공간이고, 기술정보 영역은 이용자들이 기록AIP를 탐색, 검색, 주문하기 위해 필요로 하는 정보를 저장하는 공간이다.

이 3 종류의 공간은 개념 레이어에서 정의한 것이므로 모두 다 정보시스템으로 구현될 수도 있고 일부만 구현될 수도 있다. 예를 들어, 기술정보 영역의 경우, 기술정보를 목록정보시스템으로 구축할 수도 있고 종이로 만든 목록카탈로그를 작성하여 제공할 수도 있고, 또는 다른 정보시스템에 등록정보로 올려 더 넓은 범위의 이용자가 접근 가능하도록 조치할 수도 있다. 어떤 영역을 정보시스템화할 것인지, 어떤 기술로 구현할 것인지는 개념 레이어에서 논의할 문제가 아니다. 개념 레이어에서는 관리 범위에 들어가는 기록AIP가 어떤 것들이며, 그 기록AIP들을 이용하고자 하는 고객들이 어떤 사람들인가를 정의하고, 기록AIP의 정보들이 어떤 저장 공간에 저장되어야 고객들이 편리하게 기록AIP에 접근할 수 있을 것인지를 개념화하여 빠짐없이 정확하게 모델링하는 것이 주안점이다.

3.2 논리 레이어 - 기록AIP의 구조

논리 레이어에서는 전자기록관리시스템 구축을 전제로 하여 기록AIP의 설계를 하게 된다. 이때는 시스템을 구성하는 소프트웨어와 하드웨어에 대한 선택이 전제되고, 선택한 플랫폼의 장점을 확대하고 단점을 보완하는 측면과 시스템의 전반적인 성능 향상을 고려하는 측면에서의 설계가 진

14) Filip Boudrez, 2005

행된다. 물론, 설계의 과정에도 조직의 기록에 대한 보존 전략이 관철되게 되는데 어떠한 전략을 채택하느냐에 따라 시스템의 소프트웨어와 하드웨어를 선택하는 폭이 좁아지기도 한다. 기록보존관리자는 전자기록을 어떻게 재현해보일 것인지, 메타데이터를 어떻게 저장할 것인지에 대한 여러 가지의 선택을 할 수가 있다. 또한, 디지털 리퍼지토리를 구성하면서 저장의 기본단위를 무엇으로 할 것인지, 기록에 대한 식별을 어떻게 할 것인지 등을 선택할 수 있다.

전자기록의 저장 방법으로서의 인캡슐레이션

기록AIP 설계에 큰 영향을 미치는 보존 전략은 기록을 인캡슐레이션할 것이냐 여부이다. 기록AIP의 저장 방법을 선택할 때 가장 중요한 점은 전자기록의 식별과 다양한 구성요소들 간의 관계에 있다.

기록의 구성요소는 보존되는 동안 하나라도 결코 잃어버려서는 안 되는 단일한 논리적 객체를 형성한다. 그런데, 대부분의 기록 저장 방법은 전자기록의 다양한 구성요소들을 단일한 물리적 객체로 형성하지 않고 데이터베이스, 파일시스템 등의 별도 장소에 별개의 디지털 객체로 저장하고 상호간의 관계는 링크, 데이터베이스 릴레이션, 포인터와 파일이름 등으로 표시한다. IT의 빠른 진보로 인해 플랫폼이 자주 변하고 그 때마다 디지털 객체들이 마이그레이션을 반복하게 되는 경우, 이러한 참조정보는 장기보존 과정에서 유실될 가능성이 있다. 만약 상호 참조관계 정보가 깨져 기록을 재구성할 수 없게 되면 그 기록은 잃어버린 것으로 간주된다. 기록보존관리자는 전자파일 안에 메타데이터를 포함해 넣음으로써 이러한 위험을 피할 수 있다. 즉, 하나의 물리적 객체 안에 두 개의 컴포넌트를 결합함으로써 기록과 메타데이터의

관계가 유실되는 것을 막을 수 있는 것이다.

이렇게 디지털 객체에 메타데이터를 추가하는 것을 인캡슐레이션 또는 임베딩(Embedding)이라고 하며, 인캡슐레이션은 실제로 하나의 디지털 객체에 메타데이터를 추가하거나, 하나의 디지털 객체 안에 여러 전자파일이 그룹으로 존재하는 저장 기술을 말하는 것일 뿐 전자기록이 미래에도 안전하게 재구성되고 활용 가능하도록 보장해 주는 방법은 아니다.

디지털 객체에 메타데이터를 임베딩하는 아이디어는 새로운 것이 아니다. 인캡슐레이션은 객체지향 프로그래밍의 기본 원칙의 하나이다. 객체지향에서 사용되는 인캡슐레이션은 하나의 객체를 구성하는 기능과 정보 중에서 정보를 은닉하기 위한 방법으로 인캡슐레이션을 하게 된다. 그러나, 기록AIP 설계에서의 인캡슐레이션은 정보 은닉의 관점에서가 아니라 정보객체의 독립성과 내부 구성요소 간의 결합성이라는 장점을 얻기 위해서 하게 된다¹⁵⁾.

기록AIP를 인캡슐레이션한다는 것은 앞에서 살펴본 기록AIP의 구성요소들을 묶어 하나의 객체, 예를 들면 XML 파일로 만들어 저장한다는 것이다. 이렇게 했을 때의 장점은 기록 내부의 구성요소들끼리 링크나 참조정보를 가지고 있을 필요가 없이 간단하게 저장·관리할 수 있다는 것이며, 단점은 여러 요소를 하나로 묶게 되므로 저장 파일의 크기가 커져서 기록을 검색하여 조회할 때나 DIP를 생성하고자 할 때 큰 파일을 열어 작업해야 하므로 전체적인 성능이 저하될 수 있다는 것이다.

15) 전자기록을 유지하는 데 있어서 인캡슐레이션의 중요성은 David Bearman에 의해 제기되어 왔으며, NARA의 Persistent Object Preservation, 빅토리아주 Public Record Office의 VERS 아카이빙 전략, 호주 National Archives의 AtoR 프로젝트 등에서 적용되고 있다.

인캡슐레이션의 적용

기록보존관리자는 기록AIP의 저장방법으로 인캡슐레이션 기법을 다양하게 적용할 수 있다. 예를 들어, TIFF 파일의 헤더에 메타데이터 태그를 채우는 식으로 기록의 특정 전자파일 안에 메타데이터를 포함하는 방법이 있다. 그러나, 이 방법은 기록의 여러 전자파일에 대해 동일한 메타데이터가 여러 번 저장되고 이 파일들을 서로 참조관계로 연관지어줘야 하기 때문에 유용성이 떨어진다. 또한, 기술적인 부분에서도 여러 단점이 있다. 보통의 텍스트 파일에서는 문제가 없지만 바이너리 파일 포맷에서는 메타데이터를 포함하기 위해 몇 가지 표준 필드를 제공하긴 하지만 보존기술에 필요한 모든 필드를 제공하지는 않는다는 것이다. 게다가 제공되는 메타데이터 필드는 공간이 너무 작거나 필드를 확장했을 때 가독성(Readability)에 문제를 일으킬 수 있다. 또한, 바이너리 파일에 메타데이터를 추가하는 것은 현재의 컴퓨터 프로그램이 기능을 제공하지 않는 경우가 일반적이므로 각 바이너리 포맷마다 메타데이터 추가를 위한 소프트웨어 툴을 개발해야 한다.

이러한 문제점들을 피하기 위해 하나의 기록에 속하는 모든 전자파일들을 한꺼번에 인캡슐레이션할 수 있다. 그러면, 인캡슐레이션되는 전자파일들은 하나의 파일에 묶여있으므로 서로에 대한 링크정보를 따로 가질 필요가 없게 되고 기록에 대한 기술정보도 한번만 포함하면 된다. 이 때, 인캡슐레이션된 결과 파일은 OAIS 모델의 AIU에 해당하는 디지털 리파지토리의 기본 저장단위가 된다. 인캡슐레이션이 저장방법으로 채택하면 하나 또는 여러 개로 나뉘어 입수되는 기록SIP를 하나의 기록AIP로 팩킹(packaging)하는 과정을 거쳐야 하고, 이 상태로 저장 보존되는 기록AIP들은 팩킹을 풀어야만 보거나 활용할 수가 있게

된다.

기록AIP의 구조

기록AIP를 하나의 파일로 인캡슐레이션하는 경우에 기록AIP의 내부를 식별자(identifier), 기록의 모든 전자파일과 필수 메타데이터, 체크섬 등의 3부분으로 구조화할 수 있다. 식별자는 정보시스템 내에서 기록AIP파일의 유일한 식별자 역할을 하는 값으로 장기보존 과정에서 가급적이면 변하지 않고 하나의 값을 유지할 수 있도록 할당되는 것이 좋다. 체크섬은 OAIS의 인증정보에 대응되는 것으로 기록AIP에 체크섬으로 삽입된 해쉬값은 기록AIP가 전송되거나 마이그레이션 등의 조치 이후에 다시 계산한 해쉬값과 비교함으로써 기록AIP가 변조 또는 훼손되지 않았음을 확인하기 위한 정보이다. 체크섬을 이용하여 기록AIP가 잘 보존되고 있는지를 확인하는 것은 기록보존관리자의 주요 임무 중의 하나이며, 이러한 확인 작업은 전자기록관리시스템이 주기적으로 자동으로 수행하도록 할 수도 있고 기록보존관리자가 필요에 의해 임의로 수행할 수도 있다. 삽입된 해쉬값이 다시 계산한 해쉬값과 일치하지 않으면 시스템이 기록보존관리자에게 그 사실을 알려줘서 관리자가 기록AIP의 훼손된 상태를 파악하고 백업해 놓은 기록AIP를 가져다가 디지털 리퍼지토리에 저장할 수 있도록 한다. 체크섬에는 해쉬값 뿐만이 아니라 해쉬알고리즘이 함께 보존되어야 한다.

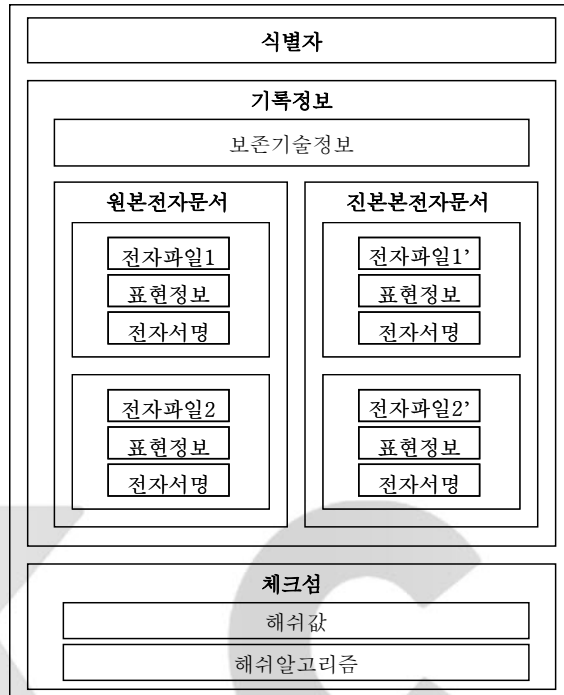


그림3-5. 기록AIP 파일의 구조

기록AIP의 기록정보에는 보존과 관리에 필요한 메타데이터들이 보존기술정보라는 하위요소로 포함된다. 이 메타데이터는 기록AIP 전체와 관련된 것으로 한 번만 저장되면 된다. 개념 레이어에서 찾아낸 기록의 메타데이터 중 표현정보와 인증정보 일부를 제외하고는 거의 모든 메타데이터가 보존기술정보에 함께 기술된다. 기록정보의 하위 요소로 표시된 원본전자문서에는 전자기록이 가졌던 모든 전자파일들의 내용정보와 전자서명을 포함한다. 또한, 진본전자문서에는 마이그레이션된 파일 포맷을 저장하고 표현정보와 전자서명을 포함한다. 마이그

레이션을 누가 언제 어떻게 왜 수행하게 되었는지 등의 이력정보가 보존기술정보에 포함되며, 원본이나 진본전자문서의 모든 전자파일이 동일한 표현정보를 갖는 경우 기록AIP에 표현정보는 한 번만 기술할 수 있다.

전자파일을 마이그레이션할 때 어떤 파일포맷으로 할 것인가는 보존전략에서 중요한 선택사항이다. 예를 들어, 한글 워드프로세서로 98 버전으로 작성된 전자파일을 언제쯤 어떤 포맷으로 마이그레이션할 것인지 기록보존관리자는 결정해야 한다. 2004버전이 출시된 시점까지 마이그레이션을 하지 않았을 때, 현재는 98버전 이상의 한글 워드프로세서로 호환이 되므로 98버전의 파일을 읽는 데에 문제가 없을 수 있다. 그러나, 어느 시점에서 워드프로세서가 더 이상 98버전의 파일을 지원하지 않는 제품으로 출시되기 시작한다면 반드시 마이그레이션을 검토해야 할 것이다. 해당 워드프로세서의 최신 버전 파일로 마이그레이션하던가 아니면 향후 마이그레이션을 덜 할 수 있는 안정적인 보존 포맷인 PDF/A¹⁶⁾파일로 마이그레이션 할 수 있을 것이다. 마이그레이션이 반복되는 경우 기록정보에 진본전자문서를 모두 보존할 것인지 역시 보존전략에 따른다. 기록보존관리자는 진본전자문서가 반복하여 생산되는 경우 기록AIP의 파일사이즈가 커지므로 저장 공간이 더 필요해 진다는 것과 이용자의 요구에 의해 기록AIP를 접근할 때마다 큰 파일을 조작해야 하는 시스템의 부담이 증가한다는 것을 고려해야 한다.

16) ISO 19005-1:2005로 채택된 전자문서의 장기보존포맷인 PDF/A는 Adobe사가 PDF1.4포맷을 로열티없이 공개함으로써 공개용표준과 편재성의 특성을 모두 만족하는 세계표준 전자문서 보존포맷이 되었다.

기록AIP 보존포맷으로의 XML

기록AIP를 디지털 리파지토리에 하나의 디지털 객체로 저장할 때 여러 가지 방법이 고려될 수 있다. 기록AIP의 저장 파일포맷으로는 XML 포맷이 tar, zip, jar 등과 같은 압축된 포맷보다 장점이 많다. 왜냐하면, 압축을 사용하게 되면 압축알고리즘에 의해 원래의 비트스트림이 인코딩되어 변형된 상태로 보존되므로, 나중에 완벽하게 원래의 비트스트림으로 재구성할 수 있다고 장담할 수 없으며, 원래의 비트스트림 안에 워터마크가 삽입되어 있는 경우에는 압축과 해제 과정에서 데이터가 손실될 수도 있기 때문이다. 또한, 압축파일 안에 어떤 전자파일들이 들어있는지에 대한 참조정보를 생성하여 관리해야 하며, 압축된 포맷은 메타데이터를 포함하는 것에 한계가 있으므로 기록AIP의 저장포맷으로는 적합하지 않다.

XML포맷은 전자파일을 메타데이터와 함께 직접 XML문서 안에 포함할 수가 있는데, XML에 포함되기 위해서 바이너리 파일들은 먼저 Base64로 인코딩되어 텍스트 문자로 변환되어야 한다. Base64를 거쳐 인코딩했다가 다시 디코딩하기 위해서는 재구성을 위한 참조정보가 추가되어야 하지만, Base64는 이미 잘 알려져 있으며 다양한 플랫폼에 광범위하게 사용하는 아주 단순한 방식이다. 대신, Base64를 사용하여 바이너리 파일을 변환하게 되면 파일 사이즈가 3분의 1만큼 증가하게 된다. XML 문서에 메타데이터를 포함하면 함께 저장되는 전자파일과 메타데이터 사이에 참조정보를 추가하지 않아도 된다는 장점이 있다. XML 문서에 메타데이터를 삽입하는 것은 메타데이터와 기록정보를 하나로 묶어준다는 장점뿐만 아니라, 메타데이터가 구조화되고 안정적인 방식으로 보존됨으로써 향후의 활용가능성을 보증할 수 있다는 것

이다. 또한, XML은 필요에 의해 구조를 가변적으로 변화시킬 수 있어서 XML 문서 안에 메타데이터 등의 요소들을 언제라도 확장하거나 변형하는 것이 가능하다. 이러한 XML의 구조적 융통성이 기록에 대한 추가적인 기술(description)과 마이그레이션에 따른 진본전자문서와 메타데이터를 용이하게 포괄할 수 있는 최대의 장점이 된다.

기록AIP의 파일포맷으로서 XML은 기술적인(technological) 이유와 장기간에 걸친 활용가능성의 확보를 위해서 뿐만이 아니라 AIP자체에 대해 문서화를 할 때 편리하다. 기록AIP의 구성과 내용에 대한 정보는 장시간이 지난 후에도 이해될 수 있어야 하며, 미래의 기록보존관리자는 기록AIP 파일에 패키징되어 있는 전자기록과 메타데이터가 어떻게 재구성되어야 하는지를 알 수 있어야만 한다. 그러므로, 기록AIP의 각 부분의 의미와 상호간의 관계는 분명하게 문서화되어야만 한다. XML로 만들어진 기록AIP는 구성 정보를 보여주지 않고도 잘 구조화된 방식으로 외부와 소통될 수 있다. 기록AIP를 위한 XML스키마는 OAIS 모델에서 패키징정보로 기능하여 내용정보와 보존기술정보의 요소들이 무엇이 있는지를 식별하고 서로 연관시켜준다.

결합레벨(Aggregation Level)

이제까지 기록AIP를 기록 한 건과 대응하는 개념으로 설명해왔지만 사실 기록보존관리자는 OAIS 모델의 정보패키지 하나가 기록 건 하나와 일치될 필요는 없다는 점을 고려하여야 한다. OAIS는 디지털 정보의 일반적인 장기보존을 위해 설계된 것으로 하나의 AIP는 하나의 기록과 대응할 수도 있고 기록 건의 전자파일 하나, 또는 기록 여러 건과 대응할 수도 있다. OAIS 모델에서는 다양한 종류의 AIP들이 결합되어

컬렉션을 형성할 수 있으므로 기록보존관리자는 기록관리의 입장에서 기본 저장단위를 설정할 때 다음과 같은 다양한 결합레벨(Aggregation level) 중 하나를 선택할 수 있다.

- (1) 기록 한 건의 하나의 컴포넌트 : 예를 들어, 워드문서 파일에 포함되는 이미지 하나.
- (2) 기록 한 건의 하나의 전자파일 : 예를 들어, 워드문서 파일 하나
- (3) 기록 한 건의 모든 전자파일 : 예를 들어, 본문과 첨부 파일들 모두와 PDF/A 포맷으로 마이그레이션된 파일 모두
- (4) 기록 건 여러 개 또는 레코드그룹 : 예를 들어, 기록철(File)을 구성하는 건들의 일부, 또는 기록철 여러 개¹⁷⁾

전자기록을 장기간 안정적으로 보존하는 것이 논의의 출발점이었던 기록보존관리자는 기록 한 건을 하나의 기록AIP로, 즉 기록 한 건의 모든 전자파일을 결합레벨로 선택하는 것이 일반적일 것이다. 특히 기록보존관리자가 인캡슐레이션의 장점을 살리고 싶어한다면 기록AIP를 하나의 물리적 파일로 저장하는 것이 좋다.

기록AIP의 메타데이터 저장

인캡슐레이션을 하여 기록AIP를 저장하고자 할 때 결정적인 것은 기록AIP파일 안에 어떤 메타데이터를 포함할 것이냐를 결정하는 것이다. 기술적으로는 모든 관련 메타데이터를 삽입하는 것이 가능하지만 효율적이지는 않다. 삽입되는 메타데이터는 자주 변경되지 않는 필수적인

17) 한 파일 내의 기록 건 묶음이 물리적으로 하나의 파일에 포함된다. 이러한 접근방식은 저장되는 물리적 파일의 크기가 매우 커져서(100 MB 이상) 처리가 용이하지 않게 된다.

메타데이터만 포함하는 것이 좋다. 그래야 기록AIP를 재생산하는 횟수가 줄어들게 된다. 또한 모든 메타데이터를 인캡슐레이션하는 것은 기록AIP 파일이 커지게 하고 디지털 리파지토리 내에 중복된 정보가 넘치게 만든다. 각각의 기록AIP파일에 특정 전자파일 포맷의 명세와 뷰어를 포함한다던가, 전거레코드에 해당하는 기록생성자에 대한 다큐멘테이션을 포함한다던가하는 해서는 안 된다. 또한, 개념적으로는 기록과 메타데이터가 캡슐화되지만, 기록의 모든 메타데이터들은 실제로는 디지털 리파지토리와 별도의 기록관리메타데이터시스템에 저장되거나 안에 또는 외부에 저장될 수도 있는 것이다.

그림3-4에서 기록관리메타데이터시스템은 다양한 메타데이터를 보유할 수 있다. 예를 들어, 기록생성자를 포함한 전거레코드, 파일포맷에 대한 표현정보, 저장 매체에 대한 관리 정보, 기록의 상위 레벨 즉, 기록철, 시리즈, 레코드그룹 등에서의 보존기술 메타데이터 등이다. 이러한 메타데이터의 일부는 시스템 외부에 존재하는 곳, 예를 들면 파일포맷에 대한 레지스트리 등에서 추출되거나 연결되어 구성될 수 있다. 이용자들의 기록에 대한 검색속도 등 접근 성능을 높이기 위해 기록AIP들에 삽입된 메타데이터들의 일부를 기록관리메타데이터시스템에 중복할 수도 있다. 이렇게 되면, 기록AIP파일 안에 삽입된 메타데이터와 기록관리메타데이터시스템에서 관리되는 메타데이터 사이에 값을 일치시키는 작업이 필요할 수도 있다. 한편으로 이러한 메타데이터의 중복 저장은 백업으로서 보안기능을 하기도 한다. 기록보존관리자가 기록AIP파일에 삽입될 메타데이터를 선별해야 하는데, 기록AIP파일이 자체만으로도 완전한 기술이 되어야 한다는 측면과 다른 한편으로는 기록AIP파일을 쉽게 관리할 수 있어야 한다는 측면을 두고 균형점

을 찾아야 할 것이다.

4. 맺음말

이 논문에서는 OAS 정보모델과 시스템 구축의 3-레이어 접근법을 이용하여 전자기록관리시스템에서 장기간에 걸쳐 안정적으로 보존관리하기 위한 기록AIP의 구성과 구조, 보존포맷에 대한 논의를 진행하였다. OAS는 개념 레이어에서 기록AIP의 구성을 제시해주었고, 그러한 구성을 바탕으로 기록AIP를 어떤 보존포맷으로 구조화할 것인지에 대해서는 구체적인 보존전략에 따라서, 이용자의 접근 활용 요구에 따라서, 인캡슐레이션의 채택할 것인지의 여부에 따라서, 메타데이터를 선택하는 범위에 따라서 다양한 선택이 있을 수 있다는 것을 논리 레이어의 논의를 통해 확인했다.

이 논문에서 사용한 3-레이어 접근법은 정보공학 등의 정보시스템 개발 방법론에서 보편적으로 적용하는 접근법으로, 기록AIP의 구성과 구조에 대한 논의 수준을 차별화하기 위해 활용되었다. 3-레이어 접근법은 시스템 개발의 초기 단계에서는 개념적 수준에서의 분석을 하고, 다음 단계에서는 분석결과를 현실적인 제약조건에 맞춰 논리적으로 설계를 하고, 마지막 단계에서 설계된 내용을 토대로 물리적인 구현을 한다는 것이 기본 틀이다. 기록AIP의 구성과 구조에 3-레이어 접근법을 적용하여 먼저 개념적인 수준에서 패키지가 어떤 내용들로 구성되어야 하는지를 정리해보고, 다음 단계에서 어떤 전략과 기술을 채용할 것

인가를 중심으로 패키지의 구성요소들을 구조화하는 여러 방법들에 대해 논의해 보았다. 이러한 3-레이어 접근법은 기록AIP가 장기보존이 가능하기 위해서 어떤 구성을 유지해야 하는지, 어떤 구조로 저장되어야 하는지 등에 대해 원칙적인 부분과 상황에 맞춰 유연하게 변형 적용해야 하는 부분을 나누어 생각할 수 있게 함으로써 기록의 정보모델과 보존포맷 사이의 간격을 입체감 있게 바라볼 수 있게 한다. 개념 레이어의 모델은 상대적으로 장기적으로 공유할 수 있는 지식이며, 논리 레이어의 설계는 시스템을 구축하는 주체의 주관적, 객관적 조건에 따라 다양한 변이가 가능하다. 물리 레이어는 실제적인 구현으로 비용을 고려한 전개가 필요한 부분이므로 기록관리 자체의 요구사항보다는 조직을 둘러싼 기본적인 제약, 즉 자원의 배분을 둘러싼 기본 제약의 영향을 받는 영역이라 하겠다.

대부분의 표준이 개념 레이어 수준의 서술 중심이며, 실제로 기록관리 주체들은 현실적인 요건 즉 기록관리시스템의 구축 목적, 예산과 요구되는 성능 등의 제약조건 하에서 개념 레이어에서 정의된 내용을 그대로 유지하면서 현실적인 설계를 하고 구현을 하게 되는 것이다. 이러한 레이어의 분리를 통해 개념에 담긴 원칙을 그대로 고수하면서 현실적인 여건에 따라 시스템에 대한 제각기 다양한 설계와 구현이 가능하며, 개념은 상대적으로 안정적이며 불변이지만 기술의 발전은 빠르므로 개념은 그대로 두고 논리적인 설계만을 바꾸어 시스템을 고도화를 해나가는 것이 가능해진다. 3-레이어 접근법의 관점에서 보자면 OAIS 모델은 개념 레이어에서 정의된 표준모델이며 일부 전략적 측면에서 논리 레이어에 대한 권고를 포함하고 있다. 국내에서도 OAIS와 같은 개념 레이어 수준의 기록관리 기능 및 정보모델의 표준화가 조속

히 작성되어 전자기록관리시스템을 구축하려는 공공기관과 민간기업들에게 가이드라인으로 제공되기를 기대해 본다.

이 논문에서는 보존정보패키지의 구성 측면보다는 구조의 측면에 논의를 더 집중하였다. 장기보존을 위한 메타데이터 구성요소에 대한 식별과 정의는 OAIS 보존기술정보에 대한 최근의 논의와 OCLC/RLG의 보존메타데이터 연구그룹 PREMIS의 최근의 연구 결과 등을 반영하여 좀 더 상세히 다뤄질 필요가 있다. 또한 이 논문의 후속으로 OAIS 모델에 따른 기록DIP의 구성과 구조 설계, OAIS 보존관리기능이 AIP와 DIP에 주는 영향과 그에 따른 AIP와 DIP의 구성과 구조, AIP와 DIP의 구성과 구조에 진본성 유지 요건 적용, OAIS 모델의 컬렉션 개념을 기록의 분류체계에 적용 등의 여러 연구가 필요하다고 본다.

참 고 문 헌

- CCSDS 650.0-B-1:Reference Model for an Open Archival Information System, 2002
- 이소연, 디지털 아카이빙의 표준화와 OAIS 참조모형, 2002
- 김희정, 디지털 아카이빙 최근 연구동향 및 OAIS 참조모형에 관한 연구, 2003
- Moore R., A. Merzky , Persistent Archive Concepts, 2003
- Shepherd, Elizabeth and Geoffrey Yeo. Managing Records: A Handbook of Principle and Practice, Facet Publishing, 2003
- 이승휘 외, 대통령비서실 기록관리체계 재설계 연구방안 보고서, 2004
- 헨디소프트 컨소시엄, 대통령비서실 기록관리혁신을 위한 정보화전략 계획수립 결과보고서, 2005
- Filip Boudrez, Digital containers for shipment into the future, 2005
- Filip Boudrez, Digital signatures and electronic records, 2005
- LG CNS 컨소시엄, 국가기록원 기록관리시스템혁신 정보화전략계획수립 결과보고서, 2006
- 김익한, 전자기록 진본성 유지를 위한 기록관리시스템 기능 및 정보모형, forthcoming

Abstracts

**The composition and structure of Archival Information
Packages(AIP) for a long-term preservation of electronic records**

YIM, JIN HEE

It's needed for the archivists to design a flexible and stable ERMS(Electronic Records Management System) which can ingest and store records through a consistent way and let users search and use records easily what they want. The basis of the design for ERMS are the conceptual composition and the logical and physical structure of the records when they are stored and managed in the ERMS. This paper explains the process of defining components and designing structure of electronic records using 3-layered approaches which consist of conceptual, logical and physical layer and shows advantages of this approaches.

After benchmarking the information models of OAIS which is a reference model for the long-term preservation of digital information objects, this paper applies the model of AIP to a record as a 'Record AIP' and discusses the composition and structure of it. It's a critical task to identify mandatory or optional metadata groups which consists of the 'Record AIP's in the conceptual layer. This paper emphases that the metadata group related to services for the record information to users is required as a result of benchmarking OAIS information models. Various issues about the structure of 'Record AIP's are discussed according to the kind of preservation strategy such as migration or emulation and whether the encapsulation of records is required or not in the logical layer.

**Key words : Archival Information Package, AIP, OAIS, Information Model,
3-Layered approach, long-term preservation, ISO 14721**