

기록관리 이슈페이퍼 vol.09

2019

전자기록 장기보존패키지 모델 시험과 새로운 모델 제안



전자기록 장기보존패키지 모델 시험과 새로운 모델 제안

신동혁 공업연구소(대통령기록관 보존복원과)

김상국 전산사무관(기록보존서비스부 복원관리과)

나미선 학예연구관(대통령기록관 보존복원과)

목 차

- I. 시작하며
- II. 추진배경
- III. 장기보존 패키지 모델 시뮬레이션
- IV. 장기보존 패키지 모델 제안
- V. 끝맺으며

요약

기록물 장기보존의 목적은 장기간에 걸쳐 독립적이고 해석 가능하며 안전하게 보존하고 지속적 접근이 가능하도록 하는 것이다. 이와 함께 기록을 확인하는 시점에서 해당 기록이 적법하고 적정하다는 것이 객관적으로 입증되어야 한다. 전자적으로 생산·관리되어야 하는 모든 전자기록(본 디지털 또는 디지털화 기록을 모두 포함)은 기록물의 특성상 전자적 형태와 방법으로 장기보존 전략을 수행하여야 한다. 또한 전자기록의 수적 양적 증가를 고려할 때 인적·물적·시간적 자원에 대한 비용 역시 고려하지 않을 수 없다.

본 보고서에서는 OAIS(Open Archival Information System)모델의 개념을 적용하여 하나의 파일 또는 폴더 객체의 형태를 갖는 모델(NEO-X)과 콘텐츠, 메타데이터가 개념적 패키징에 포함되어 있는 분산관리모델에 대해 시험하고 그 결과를 제시하였다. 모델에 대한 시뮬레이션은 대용량 기록물인 디지털 시청각기록물을 대상으로 하였다.

또한, 전자기록 장기보존패키지가 전자기록을 무한한 시간동안 안전하게 보존할 수 있도록 하는 완성체가 아닌 하나의 요소임을 확인하였다. 장기보존패키지는 전자기록의 장기보존을 위한 많은 방법 중의 하나이다. OAIS(ISO-14721)에서도 장기보존패키지의 특정한 형식이나 구성 방법에 대해서는 제시하고 있지 않다. 오히려 기록보존소의 상황과 여건에 맞게 장기보존에 필요한 요건들을 갖추도록 제안하고 있다.

모델 시뮬레이션 결과를 토대로 전자기록의 장기보존을 위한 패키지의 적합성 여부를 판단하였고, 장기보존의 신뢰성을 높일 수 있는 프로세스 메타데이터를 제안하였다. 프로세스 메타데이터에 해당하는 기록물에 대한 관리 이력과 속성 정보의 변경사항을 수집 관리하는 방법을 제안하였다. 이를 위한 시스템 설계 사항으로 가변메타데이터 및 이력 수집 기능을 제안하였다.

기술은 지속적으로 변하며, 전자기록의 장기보존을 위한 기술적 수단은 기록의 생애주기 동안 몇 번이 될지는 예상할 수 없지만 계속 변화할 것이다. 따라서 주변 환경과 여건의 변화에 따른 장기보존방법에 대한 연구는 지속되어야 한다. 전통적 장기보존 전략을 토대로 새로운 전략에 대한 제안이 계속될 수 있기를 바란다.

I. 시작하며

전자기록관리에서는 콘텐츠의 생산과 동시에 정체성을 확보하고 시간적으로 기록에 행해지는 모든 행위와 그 결과(PDI : Preservation Description Information) ISO-14721(OAIS Reference Model)¹⁾를 남기면서 오랜 기간에 걸쳐 콘텐츠와 행위 이력에 대한 접근을 허용하여야 한다. 또한, 이 모든 것들을 안전하게 보존하는 업무를 수행하면서 동시에 장기간 보존하고 그 결과가 객관적으로 신뢰받을 수 있어야 하며 적용된 모든 절차와 방법이 검증 가능하도록 하여야 한다. 국가기록원에서는 그동안 전자기록의 엄정한 관리를 위하여 많은 노력을 해 왔으며, 그 결과의 실체가 기록관리 시스템의 기능으로 구현되었다. 이 중 장기보존포맷은 전자기록의 장기보존을 위한 움직임 수 없는 정론으로 인식되어 왔고, 2018년 혁신과제의 일환으로 추진 중인 「전자기록물 장기보존 정책(안)」에서도 여전히 주요 전략의 하나로 자리매김하고 있다. 본 보고서에서는 디지털 시청각기록관리시스템 설계 및 운용 경험을 토대로 시청각기록의 장기보존패키지 모델 시험을 통해 전자기록 장기보존패키지에 대한 적합성 판단과 함께 대안 및 시스템 설계 시 고려사항을 제시하였다.

II. 추진배경

장기보존의 목적은 보존할 가치가 있는 전자기록을 장기간에 걸쳐 안전하게 보존하고 지속적 접근이 가능하도록 하는 것이다. 기록을 둘러싼 환경과 시간 경과에 관계 없이 보존되고 있는 기록에 대한 객관적, 과학적 그리고 합리적 신뢰성이 유지되어야 하며, 신뢰성을 유지하기 위한 방법이 검증 가능할 때 기록이 안전하게 보존되고 있다고 할 수 있을 것이다.

그러나 여기에는 몇 가지 고려사항이 있다. 첫 번째는 보존하고자 하는 전자기록의 수와 양이 전통적 기록물과는 비교할 수 없을 만큼 폭발적으로 증가하고 있다는 사실이며, 두 번째는 전자기록의 수적·양적 증가와 전자적 관리를 위한 시스템과 네트워크의 성능은 비례관계에 있으며, 세 번째는 물리적·시각적 실체를 갖고 있는 전통적 기록물과 달리 실체가 없는 전자기록과 이를 관리하는 시스템에 대한 막연한 의구심을 해소할 수 있는 객관적 신뢰수단이 확보되어야 한다. 마지막으로 전자기록의 생애주기

1) ISO-14721(OAIS Reference Model) 규정하고 있는 기록물의 내용을 이해할 수 있도록 하는 정보 집합체

동안 아키비스트가 수행하는 다양한 업무행위가 시스템으로 이루어지는 만큼 사용자 입장에서 편의성과 대량의 전자기록 처리 부담을 완화시킬 수 있는 방안이 마련되어야 한다.

보고서에서는 전자기록의 장기보존을 위하여 2018년 기록관리 혁신과제의 일환으로 추진 중인 「전자기록물 장기보존 정책(안)」에서 제시하고 있는 장기보존포맷(이하 ‘장기보존패키지’라고 함)의 적합성 판단을 위하여 디지털 시청각기록을 예로 들어 시뮬레이션을 수행·분석한 후, 결과의 도출과 함께 기술적 대안 및 기록관리시스템 설계 시 고려사항을 제시하였다. 또한 기술적 대안의 정당성을 판단하기 위하여 기록 관련 국제표준 및 국제포럼의 결과로부터 그 근거를 제시하였다.

III. 장기보존 패키지 모델 시뮬레이션

1. 전자기록물 장기보존정책(안)²⁾ 주요 내용

2018년 혁신과제로 추진 중인 「전자기록물 장기보존 정책(안)」의 주요 내용을 요약하면 다음과 같다.

- 보존정책의 방향
 - 환경변화에 관계없이 전자기록물을 장기간 보존할 수 있는 지속가능성
 - 이론과 실무의 토대로서 인정받은 국내·외 표준의 준수
 - 새로운 기술변화에 쉽게 대응할 수 있는 유연성
 - 정보자원 및 기능의 확대·축소가 용이한 확장성
- 정책의 적용 대상
 - (기관) 기록물법에 따라 기록물을 장기간 보존하는 기관
 - * 기관 특성에 따라 장기보존정책에 대해 재량권 보장
 - (대상) 전자적으로 생산된 기록물, 원본이 폐기된 전자화기록물
- 전자기록물 장기보존 원칙
 - 전자기록물의 진본성·무결성 보장 및 유지
 - 장기간 접근 및 이용가능성

2) 전자기록 장기보존정책(안), 2019. 전자기록관리과 - 논의 중에 있는 초안임.

- 장기보존을 위한 OAIS 참조모델 준수
- 메타데이터의 생성·관리
- 장기보존 전략
 - 마이그레이션, 에뮬레이션, XML의 개별 또는 혼합 방식
 - 보존포맷 다양화
 - 장기보존패키지(기록물, 기록관리 메타데이터 포함)
 - 매체 재수록 및 복제

2. 시뮬레이션 (1차)

시뮬레이션은 시험대상인 디지털 시청각기록에 「전자기록물 장기보존 정책(안)」 중, 전자기록물 장기보존 원칙에 부합하는 장기보존패키지를 적용하여 패키지 생성, 저장, 메타데이터 변경을 위한 업로드, 변경 및 재패키징, 재저장 등에 소요되는 시간을 측정하였다. 시스템 운용 경험을 바탕으로 측정된 소요시간이 시스템의 성능과 사용자 편의성 및 대량의 기록처리에 어느 정도 영향을 미치는지를 정성적으로 분석하였다. 단, 시뮬레이션에 사용된 장기보존패키지에 대한 기술적 적정성, 메타데이터의 충실도 등은 보고서의 주제 범위에서 벗어난 관계로 분석대상에서 제외하였다.

1) 사용된 장기보존패키지

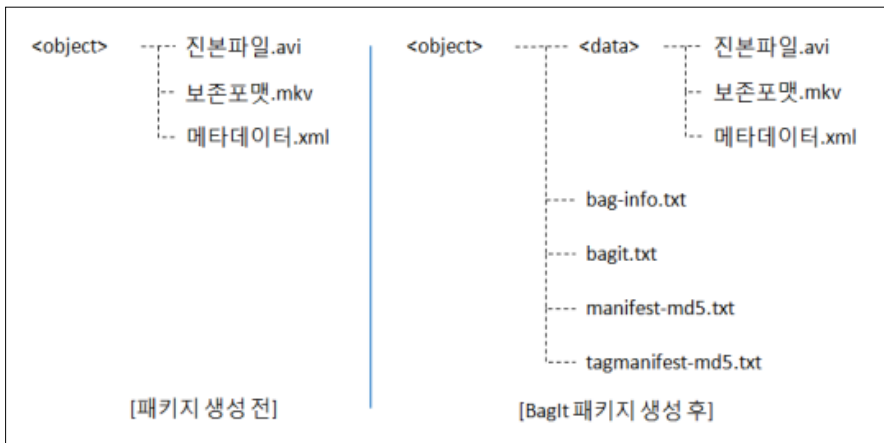
OAIS에서 제시하는 참조모델인 AIP(Archival Information Package)를 준수하였다. AIP는 단일객체 파일형식으로 시뮬레이션에서는 2018년 전자기록관리 고도화를 위한 BPR/ISP에서 제시된 NEO-X의 형태를 갖는 BagIt 패키지를 선택하였으며, 패키지 생성 및 검증에 사용되는 도구는 Bagger를 이용하였다.

- 장기보존패키지 : BagIt(버전 0.97)
- 장기보존패키지 생성 및 검증 도구 : Bagger(버전 2.8.1)
- 도구출처 : <https://github.com/LibraryOfCongress/bagger/releases>

BagIt 패키지는 흔히 알고 있는 ZIP과 유사한 방식이나, 차이점은 패키지 대상인 파일 각각에 대하여 해쉬함수(MD5)를 적용한 검증 목록을 생성하고, 검증 목록 파일과 BagIt정보를 담고 있는 파일들에 대한 해쉬 검증 목록을 생성하여 교차

검증이 가능하도록 되어 있다. 패키징 방법이 직관적이어서 단순해 보일 수 있으나 구성 파일에 대한 임의 수정, 삭제 등의 위·변조에 대한 강점을 갖고 있다. 3개 파일(진본 파일, 보존포맷 파일, 메타데이터 파일)을 갖는 object 폴더에 대한 BagIt 패키지 생성 전·후의 비교 모습을 <그림 1>에 제시하였다.

BagIt 패키지 생성 시, 가장 많은 시간이 소요되는 것은 3개의 대상파일에 대한 해쉬값을 갖는 manifest-md5.txt 파일의 생성이다. 3개의 파일을 <data> 폴더로 이동시키고, bagIt 정보를 나타내는 bag-info.txt와 bagit.txt의 생성 및 tagmanifest-md5.txt (manifest-md5.txt, bag-info.txt, bagit.txt 각각의 해쉬값을 담고 있음)를 생성하는 데 소요되는 시간은 manifest-md5.txt 파일의 생성 소요시간에 비해 무시할 만한 수준이다.



<그림 1> BagIt 패키지 생성 전·후 폴더 구조

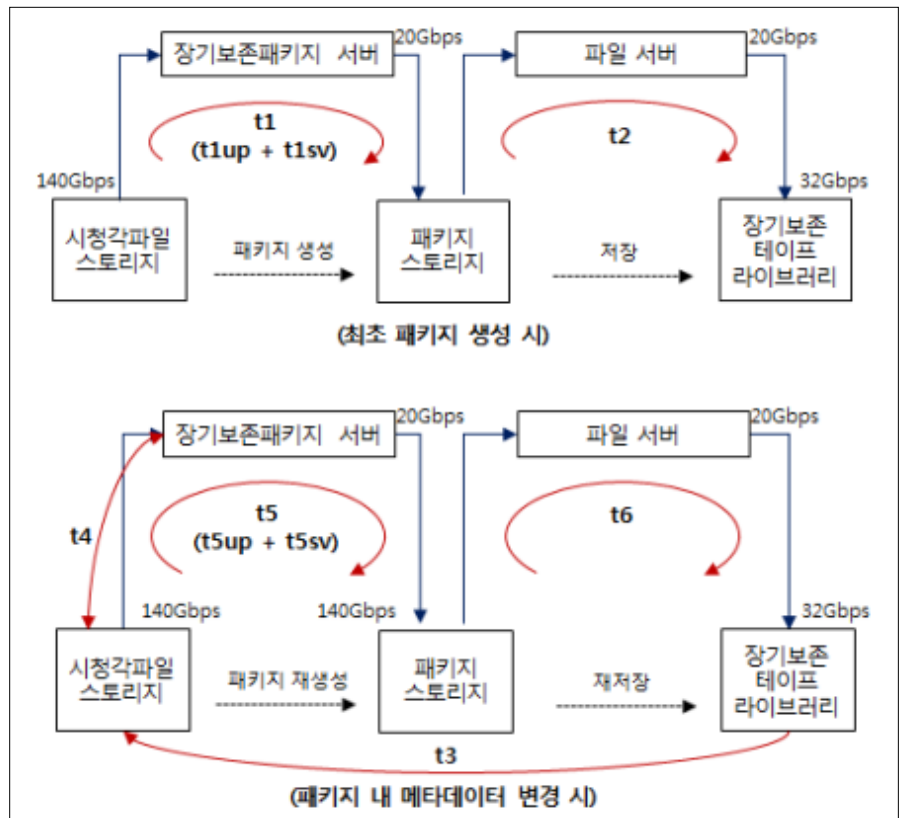
2) 분석항목

- 보존객체 용량별 장기보존패키지 생성에 소요되는 시간(t1)
- 장기보존패키지를 아카이빙 저장장치에 저장 시 소요 시간(t2)
- 메타데이터 변경을 위하여 아카이빙 저장장치로부터 패키지를 가져오는 시간(t3)
- 메타데이터 변경 소요 시간(t4)
- 메타데이터 변경 후 재패키징 소요시간(t5)
- 재패키징 후 아카이빙 저장장치에 저장 시 소요 시간(t6)

3) 시험 환경

시뮬레이션은 현재 운용 중인 대통령기록관의 시청각기록관리시스템(MAM, Media Asset Management)에서 실시하였다. 국가기록원 본원의 MAM 시스템도 동일한 구성이나, 대통령기록관 MAM 시스템의 서버 성능 및 네트워크 속도가 훨씬 우수한 관계로 최대한 좋은 환경에서 결과를 얻고자 하였다. 시뮬레이션에 사용된 장비의 주요 사양 및 구성은 <그림 2>와 같다.

- 장기보존패키지 서버 : Windows server 2016, 32GB DDR, 20Gbps I/O
- 파일 서버 : Windows server 2012, 32GB DDR, 20Gbps I/O, 8Gbps SAN I/O
- 시청각/패키지 스토리지 : 520TB volume, 2.9TB cache memory, 140Gbps I/O
- 장기보존 테이프 라이브러리 : 10PB volume, 32Gbps I/O



<그림 2> 1차 시뮬레이션 시스템 구성도

4) 측정 결과 및 분석

〈표 1〉은 장기보존패키지 생성과 장기보존패키지 내의 메타데이터 변경 시 소요 되는 시간을 보여준다.

〈표 1〉 장기보존패키지 생성 및 메타데이터 변경 소요시간

패키지 파일 용량	최초 패키지 생성 시			패키지 내 메타데이터 변경 시				
	t1		t2	t3	t4	t5		t6
	t1up	t1sv				t5up	t5sv	
1GB	4s	6s	4m 01s	3m 31s	< 1s	5s	5s	3m 51s
10GB	35s	36s	3m 33s	3m 56s	< 1s	36s	38s	4m 01s
100GB	11m 38s	3m 10s	11m 04s	12m 14s	< 1s	9m 32s	3m 15s	10m 04s
500GB	39m 43s	31m 29s	38m 44s	39m 24s	< 1s	38m 33s	32m 49s	39m 09s
1TB	78m 06s	68m 15s	75m 34s	77m 14s	< 1s	77m 26s	67m 04s	76m 48s
1.7TB	139m 48s	109m 24s	126m 12s	125m 45s	< 1s	137m 18s	110m 14s	126m 12s

t1은 세부적으로 다음과 같이 구분할 수 있다.

- 시청각파일스토리지에서 서버로 파일을 업로드와 동시에 패키지를 생성하는 시간(tup)
- 장기보존패키지서버에서 패키지스토리지로 저장하는 시간(tsv)

시뮬레이션 결과, 1TB 파일의 최초 패키지 생성 후, 장기보존 테이프라이브러리에 저장하기까지 약 3시간 40분이 소요됨을 확인할 수 있다. 이 중 시청각파일을 장기보존패키지 서버에 업로드하고 패키지를 생성하는 시간은 1시간 18분인데 비해 나머지 2시간 22분은 생성된 패키지 저장과 이를 다시 장기보존 테이프라이브러리로 이동하는 데 소요되는 시간이다.

1TB 패키지 내의 메타데이터 변경을 위해 소요되는 시간은 약 5시간으로 실제로 메타데이터 변경은 1초 내외에 불과하며 모든 시간은 메타데이터 변경을 위하여 패키지를 업로드, 해체, 재패키징 및 테이프 라이브러리에 저장하는 데 소요되는 시간이다.

시스템 측면에서 특정 데이터의 변경을 위하여 이처럼 많은 시간이 소요된다는 것은

대단히 비정상적이며 시스템의 효율성을 떨어뜨리는 구조라고 밖에 할 수 없다. 많은 예산이 투입되는 전산시스템(서버, 스토리지, 네트워크 등)이 1초 내외의 처리를 위하여 부가적으로 5시간을 소비한다는 것은 자원의 효율성을 떠나서 비용의 낭비라고 판단된다. InterPares project 및 ISO-15489 표준의 곳곳에서 언급하는 비용(cost)의 범주에는 시스템 자원관련 비용도 포함되어 있다고 이해하였다. 전산시스템 설계 시, 이러한 비용의 낭비는 허용되지 않아야 하며 반드시 필요한 기능이라면 우회방법을 마련하여 시스템 자원이 효율적으로 사용되도록 하여야 한다.

3. 시뮬레이션 (2차)

1차 시뮬레이션은 패키지 생성시간보다 패키지 생성을 위하여 서버와 스토리지 간 파일을 전송하기 위한 시간이 훨씬 더 소요된다는 것을 보여준다. 특히 기록관리 행위의 결과로서 단순히 메타데이터만을 변경하고자 할 경우, 데이터 변경에 비해 패키지 파일을 불러오고, 재패키징 및 재저장을 위한 시간이 비정상적이라고 할 만큼 많이 소요된다. 이는 시스템 자원을 비효율적으로 사용하게 되는 결과를 발생시켜 시스템을 통한 전자적 관리의 장점보다는 오히려 기록관리의 불편을 가중시키는 현상을 초래하게 되리라는 것을 쉽게 예상할 수 있다.

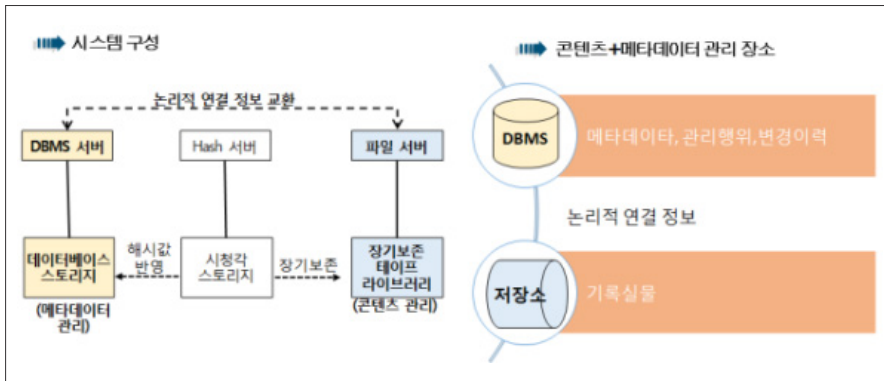
소요시간을 단축시키기 위한 가장 쉬운 방법은 동시 처리를 위하여 서버 수를 늘리면서 네트워크 대역폭을 상향 조정하는 방안이 있으나, 그럼에도 불구하고 스토리지와 장기보존 테이프라이브러리의 네트워크 인터페이스 속도 한계에 봉착하여 시스템 비용 대비 그다지 효과를 기대하기 어려울 것으로 예측된다.

1) 새로운 방법의 제안

2차 시뮬레이션에서는 소요시간의 획기적 단축과 시스템을 통한 기록관리의 효율성 극대화를 위하여 전산시스템의 본래 기능과 성능을 최대한 이용할 수 있도록 다음과 같은 개념을 도입하였다.

– 제안 개념 : 콘텐츠와 메타데이터의 분산관리³⁾

3) NEO 패키지를 구성하는 파일을 예로 들면, 전자문서 파일과 보존포맷 파일을 콘텐츠로 분류할 수 있으며, 전자 문서의 생산과 함께 취득된 정보 및 기록관리 과정에서 얻어진 정보를 메타데이터로 분류할 수 있다.



〈그림 3〉 분산관리 방식 개념도

전자기록관리에서 장기보존의 주요 전략으로 채택한 콘텐츠(컴퓨터에서는 이를 파일로 인식)와 메타데이터(컴퓨터에서는 데이터베이스 파일의 레코드⁴⁾로 인식)를 또 다른 한 개의 파일형식으로 객체화하고자 한 전략은 전통적으로 물리적 실체를 갖고 있는 기록물에 대한 보존방법을 전자기록에도 동일하게 적용하고자 한 고정관념에 그 원인이 있다고 보았으며, 이는 여러 기능을 시스템적으로 조합하여 목표기능을 수행하는 전산시스템의 장점을 충분히 이용하지 못한 이해 부족도 원인의 하나라고 판단하였다. 〈그림 3〉은 분산관리 방식의 개념을 설명하고 있다.

일정 규모 이상의 전산시스템에서 파일(콘텐츠에 해당)은 파일 시스템이 탑재된 컴퓨터(서버)에 의하여 스토리지라고 하는 별도의 저장장치에서 관리되는 것이 최적이다. 또한 데이터베이스 파일은 데이터베이스관리시스템(DBMS, DataBase Management System)이 탑재된 컴퓨터(서버)에 의하여 파일이 저장된 저장장치와 다른 독립적인 스토리지에서 관리하는 것이 데이터의 안정적 운용을 위한 일반적 방법이다. 데이터베이스의 특정 레코드(record)가 콘텐츠와 관련되어 있다면 이들 간의 논리적 연결정보를 데이터베이스의 해당 레코드의 특정 필드(field)에 반영하여 관리하도록 한다.

따라서 2차 시뮬레이션에서는 콘텐츠(파일)와 메타데이터를 분산하여 관리할 수 있도록 〈그림 4〉와 같이 시스템을 구성하였으며, 시뮬레이션에 사용된 주요 장비의 사양은 1차 시뮬레이션과 동일하다.

4) 데이터베이스에서 레코드(record)는 필드(field)로 이루어진 정보를 의미하는 것으로, 전자기록에서 하나의 철 또는 건 정보로 이해할 수 있다.



〈그림 4〉 2차 시뮬레이션 시스템 구성도

2) 분석항목

- 콘텐츠 용량별 해쉬값 계산에 소요되는 시간(t1)
- 콘텐츠를 아카이빙 저장장치에 저장 시 소요 시간(t2)
- 메타데이터 변경 소요 시간(t3)

3) 측정 결과 및 분석

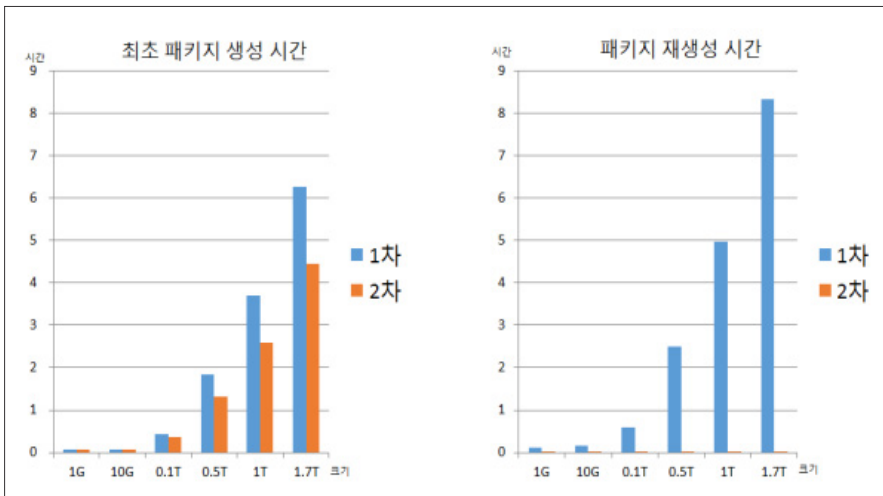
〈표 2〉는 콘텐츠와 메타데이터를 분산관리하는 구조에서 소요되는 시간을 보여준다.

〈표 2〉 장기보존 저장 및 메타데이터 변경 소요시간

패키지 파일 용량	최초 장기보존 시		메타데이터 변경 시(sec)
	t1up	t1sv	t5up
1GB	3s	4m 05s	< 1s
10GB	38s	3m 19s	< 1s
100GB	11m 11s	10m 52s	< 1s
500GB	39m 31s	39m 24s	< 1s
1TB	78m 49s	76m 29s	< 1s
1.7TB	139m 32s	127m 37s	< 1s

최초 장기보존 시, 해쉬값을 생성하고 장기보존 테이프라이브러리로 저장하는 시간은 1차 시뮬레이션에 비해 2/3정도의 시간이 소요된다. 장기보존 테이프라이

브러리로 저장하는 시간은 1차 시뮬레이션과 동일하다고 볼 수 있으나, 기록물에 대한 패키징 과정을 생략하고 해쉬값을 생성하는 소요시간은 1차 시뮬레이션의 절반 수준에 불과하다. 또한 메타데이터 변경은 1초 내외이기 때문에 동시에 대량의 전자기록에 대한 기록관리 행위 및 결과를 반영한다는 측면에서는 매우 유리한 구조라 할 수 있다. <그림 5>에 나타난 것처럼 1차 시뮬레이션과 2차 시뮬레이션의 결과는 확연한 차이를 나타낸다.



<그림 5> 1차~2차 시뮬레이션 패키지 생성시간 비교

IV. 장기보존 패키지 모델 제안

1. 분산관리 방법의 근거 및 전자기록관리의 신뢰요건

콘텐츠와 메타데이터가 분산관리되더라도, 각각의 데이터와 이를 관리하는 시스템이 신뢰할 수 있고 객관적 입증 가능성이 가능하다면 결과적으로 전자기록의 장기보존이 신뢰할 수 있는 수준에서 수행되고 있다고 할 수 있을 것이다. 이 절에서는 분산관리 방법에 대한 타당성과 전자기록 관리의 신뢰요건을 기록관리 국제표준 및 연구결과로부터 그 근거를 제시하였다.

1) 분산관리 방법의 근거

- 기록 콘텐츠와 메타데이터는 공존하는 복수의 시스템(위치) 또는 단일 시스템(위치) 내에서 관리 될 수 있으며, 콘텐츠와 메타데이터 간의 논리적 연결정보(관계)는 자동 또는 수동절차에 의해 생성, 관리되어야 한다.(5.2.3 Metadata for records, ISO 15489-1)⁵⁾
- 기록 시스템은 기록 콘텐츠와 메타데이터 간 논리적 관계의 생성 및 유지를 지원하여야 한다.(5.3 Records systems, ISO 15489-1)⁶⁾

ISO 15489-1(Information and documentation – Records management, part 1)에서는 기록콘텐츠와 메타데이터의 관리를 위한 구체적 구현 방법을 제시하지 않고 있으나,⁷⁾ 콘텐츠와 메타데이터를 하나의 객체로 관리할 수 있거나 또는 시스템 내에서 개별 서브시스템에 의하여 별도 관리할 수 있으며, 기록시스템은 이 모든 경우를 지원할 수 있도록 설계되어야 한다고 설명하고 있다.

- 정보패키지(information package)는 콘텐츠 정보와 보존기술정보에 대한 개념적 컨테이너이다. 콘텐츠 정보와 보존기술정보는 패키징 정보(packaging information)에 의하여 캡슐화되고 식별 가능하다고 할 수 있다.(2.2.2 Information package definition, ISO 14721)⁸⁾
- 패키징 정보(packaging information)는 콘텐츠 정보와 보존기술정보를 실질적 또는 논리적으로 묶고, 식별하고 연결하는 정보이다. 예를 들어, 콘텐츠 정보와 보존기술정보가 CD-ROM에 저장되어 있는 파일이라고 확인되는 경우, 패키징 정보는 CD-ROM의 ISO-9660 볼륨 및 파일구조(파일시스템이라고 함)와 CD-ROM 내의 파일에 대한 파일명 및 디렉토리 정보를 포함할 것이다.(2.2.2 Information

5) A record's content and its associated metadata may be managed either in multiple coexistent locations and systems, or in a single location and system. Logical relationships, or linkages, between a record's content and its associated metadata should be created and maintained using automated or manual processes.

6) Records systems should : c) support the creation and maintenance of logical relationships between records content and metadata for records.

7) This part of ISO 15489 is not intended to provide detailed implementation advice for specific environments in which records are created, captured and managed.

8) An Information Package is a conceptual container of two types of information called Content Information and Preservation Description Information (PDI). The Content Information and PDI are viewed as being encapsulated and identifiable by the Packaging Information.

package definition, ISO 14721)⁹⁾

- 패키징 정보(packaging information)는 콘텐츠 정보와 보존기술정보를 실질적 또는 논리적으로 묶고, 식별하고 연결하는 정보이다. 예를 들어, 콘텐츠 정보와 보존기술정보가 CD-ROM에 저장되어 있는 파일이라고 확인되는 경우, 패키징 정보는 CD-ROM의 ISO-9660 볼륨 및 파일구조(파일시스템이라고 함)와 CD-ROM 내의 파일에 대한 파일명 및 디렉토리 정보를 포함할 것이다.(2.2.2 Information package definition, ISO 14721)

ISO 14721(OAIS-reference model)에서는 명확하게 콘텐츠와 메타데이터를 패키징 정보(packaging information)에 의해 개념적으로 패키징 할 수 있다고 설명하고 있다. 즉, 콘텐츠와 메타데이터가 단일 객체의 파일형식으로 캡슐화될 수도 있는 반면, 콘텐츠와 메타데이터가 위치를 달리하고 패키징 정보에 의해 논리적 연결이 가능한 방식도 정보 패키지(information package)라고 설명하고 있는 것이다.

이러한 사실로부터 그동안 장기보존패키지(또는 장기보존포맷)가 객체화된 한 개의 파일형식이라는 공식을 벗어나 콘텐츠와 메타데이터의 연결방법에 따라 패키지의 형태가 달라질 수 있으며, 오히려 논리적 연결정보의 유지·관리가 중요하다는 사실을 인식하는 계기가 되었으면 한다.

2) 전자기록관리의 신뢰 요건

하나의 파일형식으로 이루어진 단일 객체 또는 논리적 연결정보를 갖는 개념적 패키지에 관계없이 보존대상인 전자기록의 신뢰를 어떻게 확보하고 유지할 수 있는냐가 가장 중요하다고 할 수 있다.

- 기록관리시스템의 운용·절차 및 기술적 사항에 관한 내용들이 기록으로 관리될 때, 시스템의 신뢰성이 있다고 할 수 있다. 기록에 대한 process metadata가 관리될 때, 기록 시스템의 신뢰성이 있다고 할 수 있다.(5.3.2.1 Reliable, ISO 15489-1, 의역)¹⁰⁾

9) The Packaging Information is that information which, either actually or logically, binds, identifies and relates the Content Information and PDI. For example, if the Content Information and PDI are identified as being the content of specific files on a CD-ROM, then the Packaging Information would include the ISO 9660 volume/file structure on the CDROM, as well as the names and directory information of the files on CD-ROM disk.

10) The reliability of records systems should be documented by creating and maintaining records of their operational, procedural and, where applicable, technological routines. Process metadata for records should also demonstrate the reliability of system(s) in which they have been managed.

기록 메타데이터는 두 가지로 구분할 수 있다. 기록이 생산 또는 획득 당시 확보되는 메타데이터로 변하지 않는 특징을 갖고 있는데, 이를 capture metadata라고 한다. 또 하나는 기록의 생애주기 동안 기록관리 행위의 결과로 발생하는 process metadata가 있다.(5.2.3 Metadata for records, ISO 15489-1, 요약)¹¹⁾

권위 있는 기록의 특징(Characteristics of authoritative records)인 4대 속성 중 진본성은 capture metadata에 가까우며, 신뢰성 및 무결성은 process metadata로, 이용가능성은 capture metadata와 process metadata로 설명할 수 있을 것으로 판단된다. 또한 신뢰할 수 있고, 안전하고, 규정을 준수하고, 포괄적이며 체계적인 기록 시스템을 기록 권위에 대한 필요조건으로 설명¹²⁾하고 있는데, 이는 기록 시스템이 capture metadata와 process metadata를 어떻게 다룰 것인가를 묻는 중요한 기준이라고 판단한다.

InterPares project 1, 2의 성과물을 토대로 진본 전자기록의 장기보존을 위한 정책 프레임워크를 제시한 연구보고서¹³⁾에서 신뢰요건의 기준이 될 만한 내용을 참조하였다.

■ 정책프레임워크 : 정책수립 - 범위

- 연속적 정책의 모든 과정에, 다시 말해서 생산, 활용과 유지 및 처분 등 기록 생애주기 전반에 걸친 모든 단계에서 보존에 관련된 고려사항을 편입시켜야 한다.(생산자원칙 7 및 보존자원칙 7, InterPares project 2)
- 마이그레이션이나 에뮬레이션 등 장기보존 조치를 지속적으로 취하는 데 꼭 필요한 메타데이터를 충분하고도 정확하게 확보할 수 있는 메타데이터는 가능한 한 인간이 개입하지 않는 자동화 과정을 통해 생산 시점부터 확보할 수 있어야 한다.

■ 정책프레임워크 : 정책수립 - 믿을 수 있는 아카이브 시스템

- 믿을 수 있는 기록관리시스템을 사용하여 정확하고 진본성을 유지하는 기록을 관리하라는 원칙이다.(생산자원칙 6, InterPares project 2)

11) Some of a record's metadata is derived or attributed at the time the record is created or captured and does not change. This is point of capture metadata for records. Metadata about actions on the record and other events in the record's existence, including the participating agents, continue to accrue over time as the record is used and managed. This is process metadata for records.

12) The authoritativeness of records is supported by their being managed by records systems that are reliable, secure, compliant, comprehensive and systematic.

13) 국가기록원, 『전자기록의 관리와 보존을 위한 국제협력 아젠다 개발』, 2007. 12.

- 정책프레임워크 : 정책수립 – 믿을 수 있는 아카이브 시스템
 - 믿을 수 있는 기록관리시스템을 사용하여 정확하고 진본성을 유지하는 기록을 관리하라는 원칙이다.(생산자원칙 6, InterPares project 2)
- 정책프레임워크 : 진본성 – 정체성
 - 기록의 진본성 유지와 평가를 위한 벤치마크 요건은 기록의 속성을 표현하고 기록을 표현한 그 정보를 기록에 연계하는 것을 그 첫 번째 요건으로 들고 있다.(벤치마크 요건 1, InterPares project 1)
- 정책프레임워크 : 진본성 – 무결성
 - 시스템에 대한 접근 권한의 관리 및 사용을 통제하기 위해, 자료에 대한 접근 기록을 확인할 수 있는 감사기록을 관리해야 한다.(벤치마크 요건 2, Inter-Pares project 1)

위에서 열거한 사실들로부터 전자기록관리를 위한 신뢰요건을 다음과 같이 정리하였다.

- 첫째, capture metadata와 process metadata의 구분 관리
- 둘째, 기록 생애주기 동안 기록에 영향을 미치는 모든 행위 및 결과의 기록화
- 셋째, 기록 및 시스템 감사기록 관리
- 넷째, 믿을 수 있는 기록관리시스템의 사용

2. 기록관리시스템 설계 시 고려사항

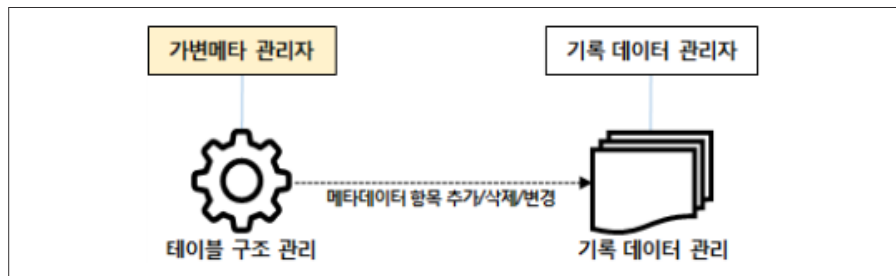
이 절에서는 전자기록관리를 위한 신뢰요건을 중심으로 시스템 설계 시 고려하여야 할 사항을 제시하였다. 다만, 제시내용은 충분한 검토와 검증을 거친 것이 아닌 개념 수준의 내용임을 밝혀둔다.

1) 데이터베이스의 가변 구조

capture metadata가 변하지 않는 특징을 가진 반면, process metadata는 기록 관리 행위 및 그 결과를 반영하여야 한다고 설명하였다. 이를 위해서는 capture metadata의 테이블(table)¹⁴⁾이 일반적인 데이터베이스처럼 고정된 형식으로

14) 데이터베이스에서 세로줄과 가로줄의 모델을 이용하여 정렬된 한 개 이상의 데이터 집합(값)의 모임

설계할 수 있는 반면, 기록 생애주기 동안 횟수를 예측할 수 없는 기록관리 행위와 결과를 반영하는 process metadata를 데이터베이스화하기 위해서는 테이블 구조가 가변적이어야 한다. 즉 어떤 레코드는 생애주기 동안 10개의 필드만으로 충분한 반면, 어떤 레코드는 20개 또는 그 이상의 필드가 필요할 수도 있다. 우리에게 익숙한 기록물의 건 개념을 예로 들어 설명하면 어떤 기록물 건은 <제목, 생산자, 공개평가 결과>의 단지 3개의 항목만 필요한 반면, 어떤 기록물 건은 <제목, 생산자, 공개평가 결과, 재평가 결과, 포맷변환, 마이그레이션> 항목이 필요하다. 더욱이 이 기록물 건은 앞으로 항목이 추가될 수도 있다. 이렇듯 예측할 수 없는 항목들을 데이터베이스로 관리하기 위해서는 레코드의 필드 수가 가변성을 가져야 할 것이다.



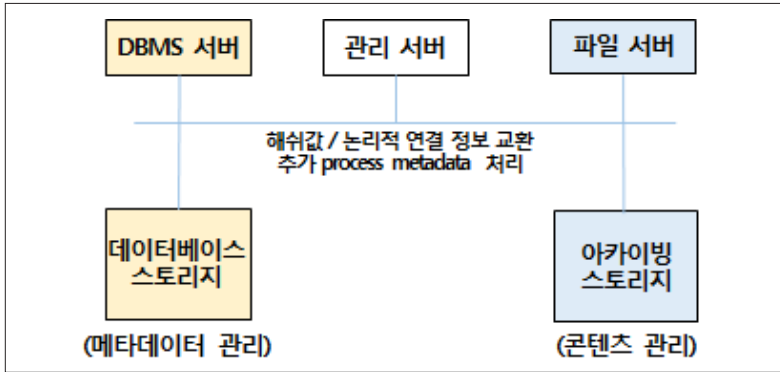
〈그림 6〉 동적 메타데이터 필드 관리

2) 객관적 신뢰를 위한 블록체인 개념 적용

2차 시뮬레이션에서 장기보존패키지를 대체하기 위한 방안으로 콘텐츠와 메타데이터의 분산관리 개념을 제안하였다. 모든 전산시스템의 서버와 스토리지는 그 자체로 사용자의 접근권한을 제한함과 동시에 보안 관리를 엄격히 수행하고 있다. 또한, 서버와 스토리지에 저장된 데이터베이스와 파일시스템에 접근하기 위해서는 별도의 접근권한을 가져야만 하는 이 중 구조의 권한 정책을 적용하는 것이 일반적이다. 현실적으로 일반 사용자가 서버와 스토리지에 저장된 메타데이터와 콘텐츠에 직접 접속하는 경우는 거의 없다. 사용자가 인지하는 콘텐츠와 메타데이터는 기록관리시스템이 제공하는 사용자 인터페이스(UI)를 통해서만 가능할 뿐이다.

전통적으로 전산시스템은 자체의 보안 솔루션 등으로 데이터의 안전과 신뢰를 주장해 왔다. 제3자가 문제제기를 하더라도 입증할 수 없다면 시스템 운용주체의 주장을 수용할 수밖에 없는 것이 일반적이다. 그러나 기록관리 분야에서 만큼은 반대현상이 나타나고 있고 이러한 이유로 인해 기록물관리기관 및 기록 관계자의

어려움이 가중되고 있는 것이 현실이다. 이를 위해서는 전자기록의 엄정함과 객관적 신뢰성을 오랫동안 유지하기 위한 추가적 방법을 고려할 필요가 있다. 보고서에서는 그 대안으로 블록체인 개념을 제안한다.



〈그림 7〉 분산관리 방식의 시스템 구성도

〈그림 7〉은 분산관리 방식의 시스템 구성도이다. 콘텐츠와 메타데이터는 별개의 서버와 스토리지에서 각각 관리되고 있다. 관리서버는 콘텐츠 파일의 해쉬값과 논리적 연결정보를 DBMS 서버로 전달하여 메타데이터로 관리되도록 한다. 이후 process metadata가 추가적으로 발생하게 되면 동일한 방법으로 메타데이터 파일에 기록을 하는 방식이다. 〈그림 5〉는 식별번호(ID) 9999인 전자기록의 메타데이터로 보존단계에서 생성된 논리적 연결정보와 해쉬값까지를 capture metadata로, 이후 발생되는 행위 1~N까지를 process metadata로 관리하는 모습을 보여주고 있다.



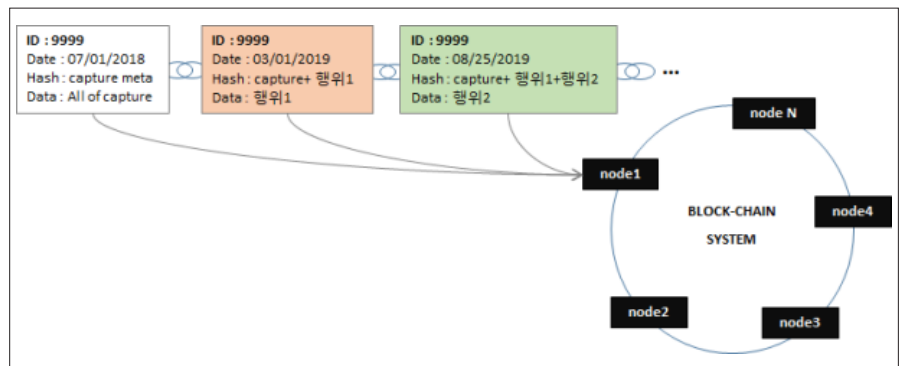
〈그림 8〉 ID 9999 전자기록의 메타데이터

〈그림 8〉은 메타데이터를 블록으로 생성하고 이를 블록체인시스템의 노드를 통해 관리하는 개념을 보여주고 있다. ID 9999인 전자기록의 최초 capture metadata 전체와 해쉬값을 블록으로 만든 후, 이를 node1을 통하여 블록체인 시스템 전체로 전파되도록 한다. 이후 행위1이 발생하게 되면 행위1의 데이터

그리고 capture metadata와 행위1의 데이터를 추가한 결과의 해시값을 새로운 블록으로 생성하고 node1을 통하여 앞서 등록된 블록에 체인으로 연결되도록 한다. 이후 행위2~N까지도 동일한 방식으로 블록체인시스템을 통하여 관리된다.

전자기록관리에 블록체인 개념 적용 시, 예상되는 기대효과는 다음과 같다.

- 메타데이터의 시계열적 변화 관리(감사관리)
- 제3자에 의한 객관적 관리 및 검증



〈그림 9〉 ID 9999의 메타데이터를 블록체인시스템으로 관리하는 개념도

다만, 모든 기록관리 행위와 그 결과가 메타데이터로 기록되어야 한다는 것을 전제로 하여야 하며, 사용자 입장에서 메타데이터의 기록화가 쉽게 이루어지기 위해서는 데이터베이스의 구조를 어떻게 설계하고 시스템에 의한 자동화를 어떻게 구현할지와 밀접한 관계에 있다.

V. 끝맺으며

전자기록의 장기보존을 위한 패키지의 적합성 여부를 판단하고, 그 결과 전자기록 관리를 위한 신뢰요건을 도출하고 기록관리시스템 설계 시 고려사항까지를 제시하였다. 다만, 시뮬레이션에서는 디지털 시청각기록을 예로 들고 이후 모든 과정에서도 시청각 기록을 대상으로 검토하였다.

그러나 전자기록의 수적·양적 증가를 고려할 때, 전자기록을 다루는 아카이브 기관에서는 디지털 시청각기록 외에도 모든 전자기록에 보고서에서 검토한 개념을

적용하는 것이 타당하다고 판단한다.

첫 번째 이유는 파일의 용량에 따라 차이가 있을 수 있으나, 컴퓨터가 처리해야 하는 파일의 수가 증가할수록 파일 I/O를 신중하게 고려해야 하며 이에 따라 업무프로세스 및 하드웨어 구성, 소프트웨어가 설계되어야 하기 때문이다.

두 번째 이유는 전자기록의 장기보존이 장기보존패키지로 완성되는 것이 아니기 때문이다. 장기보존패키지(장기보존포맷)는 전자기록의 장기보존을 위한 많은 요소 중의 하나에 불과하며, 그 이상도 이하도 아님을 이해하는 계기가 되었으면 한다. 기술은 지속적으로 변하며, 전자기록의 장기보존을 위한 기술적 수단은 기록의 생애주기 동안 몇 번이 될지는 예상할 수 없지만 변경될 것이다.

세 번째 이유는 아래의 표현으로 인용¹⁵⁾하며 동시에 전자기록의 장기보존을 위한 원칙이라고 표현하며 결론으로 마무리한다.

모든 기록을 장기적으로 훼손 없이 보존하려면, 기록 주기의 각 단계와 관련된 모든 활동에서 보존에 대한 고려를 하여야 한다.

연속적 정책의 모든 과정에, 다시 말해서 생산, 활용과 유지 및 처분 등 기록생애주기 전반에 걸친 모든 단계에서 보존에 관련된 고려사항을 편입시켜야 한다.

믿을 수 있는 기록관리시스템을 사용하여 정확하고 진본성을 유지하는 기록을 관리하여야 한다.

15) 국가기록원, 『전자기록의 관리와 보존을 위한 국제협력 아젠다 개발』, 2007. 12.

〈참고 문헌〉

1. 서은경, 「디지털 아카이브의 영구적 보존을 위한 개념적 모형 설계에 관한 연구」, 『한국문헌정보학회지』 38, 2004.
2. 국가기록원, 『전자기록 관리와 보존을 위한 국제협력 아젠다 개발』, 2007. 12.
3. 소정의 · 한희정 · 양동민, 「국외 전자기록물의 장기보존 정책 비교 분석」, 『한국기록관리학회지』 제18권 제4호, 2018.
4. 이윤주 · 이소연, 「진본 전자기록의 장기보존을 위한 정책프레임워크 : InterPARES성과물에 기초하여」, 『기록학연구』 제19호, 2009.
5. ISO 14721 Space data and information transfer systems – Open archival information system(OAIS)–Reference model, September 1, 2012.
6. ISO 15489-1 Information and documentation – Records management – Part 1: Concepts and principles, April 15, 2016.
7. The BagIt File Packaging Format(v1.0) draft-kunze-bagit-17, March 21, 2019.

|| 기록관리 이슈페이퍼 발간 목록 ||

발간호	제목	작성자	발간일
vol. 1	기관 심층인터뷰를 통한 BRM 단위과제 운영 개선 방안 수립	황정원 기록연구사	2019. 10. 8.
vol. 2	「공공기록물법」상의 기록의 개념 검토 ① 기록의 개념과 성립요건 - 정보와 증거로서의 기록의 함의를 기록물법에 적용하기 - ② 기록이란 무엇인가? - 「공공기록물법」에 따른 기록관리 대상의 범위와 관련하여-	이점마 서기관 임신영 기록연구사	2019. 10. 22.
vol. 3	대통령기록물 평가체계 개선 방안	윤정훈 행정사무관	2019. 10. 31.
vol. 4	“도전! 기록관리 명감사되기” 기록관리 감사양성제도 도입	김명옥 사서사무관	2019. 11. 15.
vol. 5	국가기록원 블록체인 기록관리 플랫폼 구축사업의 의미와 전망	왕호성 기록연구사	2019. 11. 22.
vol. 6	전자기록 장기보존정책의 방향	이지영 공업연구사	2019. 12. 5.
vol. 7	디지털기반 대통령기록관리체계 모델 재설계	김현숙 공업연구사	2019. 12. 12.
vol. 8	건축아카이브의 해외 동향 및 향후 과제 - ICAA BRAGA 2019 참가기 -	김수연 전문임기제 허인영 전문임기제	2019. 12. 13.
vol. 9	전자기록 장기보존패키지 모델 시험과 새로운 모델 제안	신동혁 공업연구사 김상국 전산사무관 나미선 학예연구관	2019. 12. 17.
vol. 10	기록물 매체수록 해외 동향 및 향후 과제	박지혜 공업연구관	2019. 12. 24 발간 예정

발간 예정 목록

- 2019년 기록물 생산현황 분석과 통보제도 개선 방안
- 기록물관리 전문요원 양성 및 자격제도 개선 방안
- 기록관리 자원 향상을 위한 컨설팅 및 심층인터뷰 개선 방안
- 행정정보데이터세트 기록관리 방안
- 특수지 기록물과 보존
- 국가 중요기록물 보존성 향상을 위한 (맞춤형 복원·복제 지원사업), 10년을 돌아보며
- 비밀기록물 생산현황 통보서식 개선방안
- 국외 소재 한국 병풍 <오륜행실도> 복원처리 지원을 위한 영국박물관 방문기
- 속기록 생산의무 회의 지정제도 운영 현황 분석 및 개선 방안
- 공공기록물법과 다른 법률과의 관계

※ 향후 이슈페이퍼의 주제 및 발간 일정은 원내 사정에 의해 일부 변경될 수 있습니다.

「기록관리 이슈페이퍼」는 기록관리 현장의 다양한 현안 논의와 기록인 여러분의 귀중한 연구성과 공유를 기다립니다.

국가기록원 연구협력과 ☎ (042) 481-6353 ✉ issuepaper@korea.kr

'신뢰받는 기록관리로 정부는 투명하게, 국민은 행복하게'



행정안전부
국가기록원

35208 대전광역시 서구 청사로 189 정부대전청사 2동
Tel 042-481-6353 Fax 042-481-6234