행정안전부
국가기록원

2019년 제2차 기록관리 현안과제 세미나

디지털 시청각기록 장기보존정책 수립 전략

2019.9.

2019년 제2차 기록관리 현안과제 세미나

디지털 시청각기록 장기보존정책 수립 전략

1. 장기보존정책 패키지 모델 시험

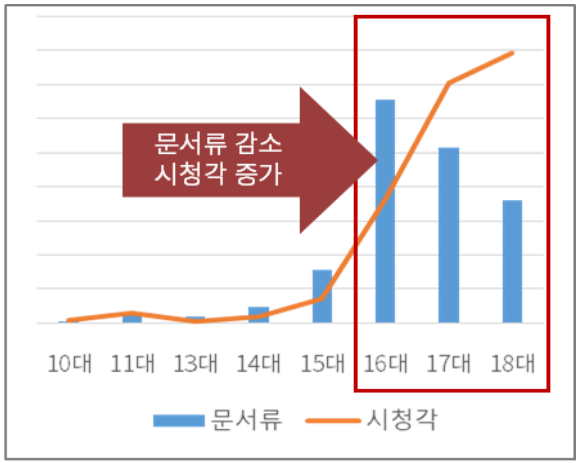
행정안전부
국가기록원

1.01 시청각기록 증가

역대 대통령기록물 증가 추세

대통령	문서류	시청각	비고
박근혜	721,220	**1,587,211**	18대
이명박	1,028,953	**1,407,352**	17대
노무현	1,307,219	736,580	16대
김대중	311,822	146,767	15대
김영삼	97,093	34,619	14대
노태우	38,695	8,715	13대
전두환	42,770	55,681	11대
최규하	8,639	19,552	10대
총계	**3,644,119**	**4,077,062**	

2800배 이상 증가



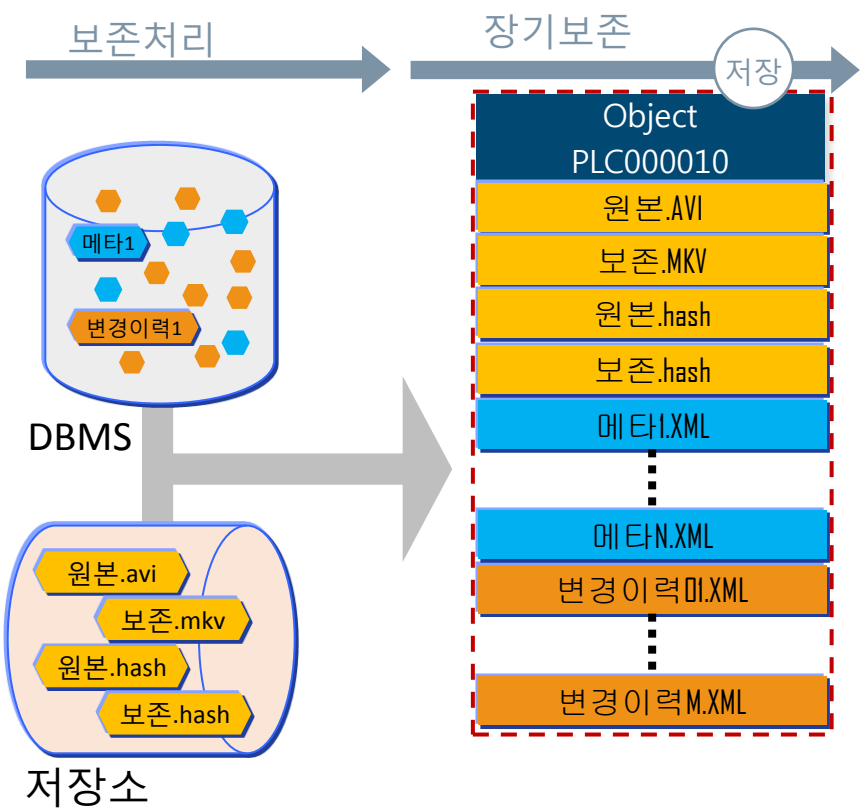
출처 : 대통령 기록관 홈페이지 (http://pa.go.kr/portal/info/report/recordReport.do)

시청각기록 파일용량 증가

아날로그 기록물 → 저용량 디지털 기록물 → 고용량 디지털 기록물 → UHD → 8K

1.02 시청각기록물 장기보존 설계

폴더형(NEO-X) 장기보존



메타데이터 변경/처리 이력관리

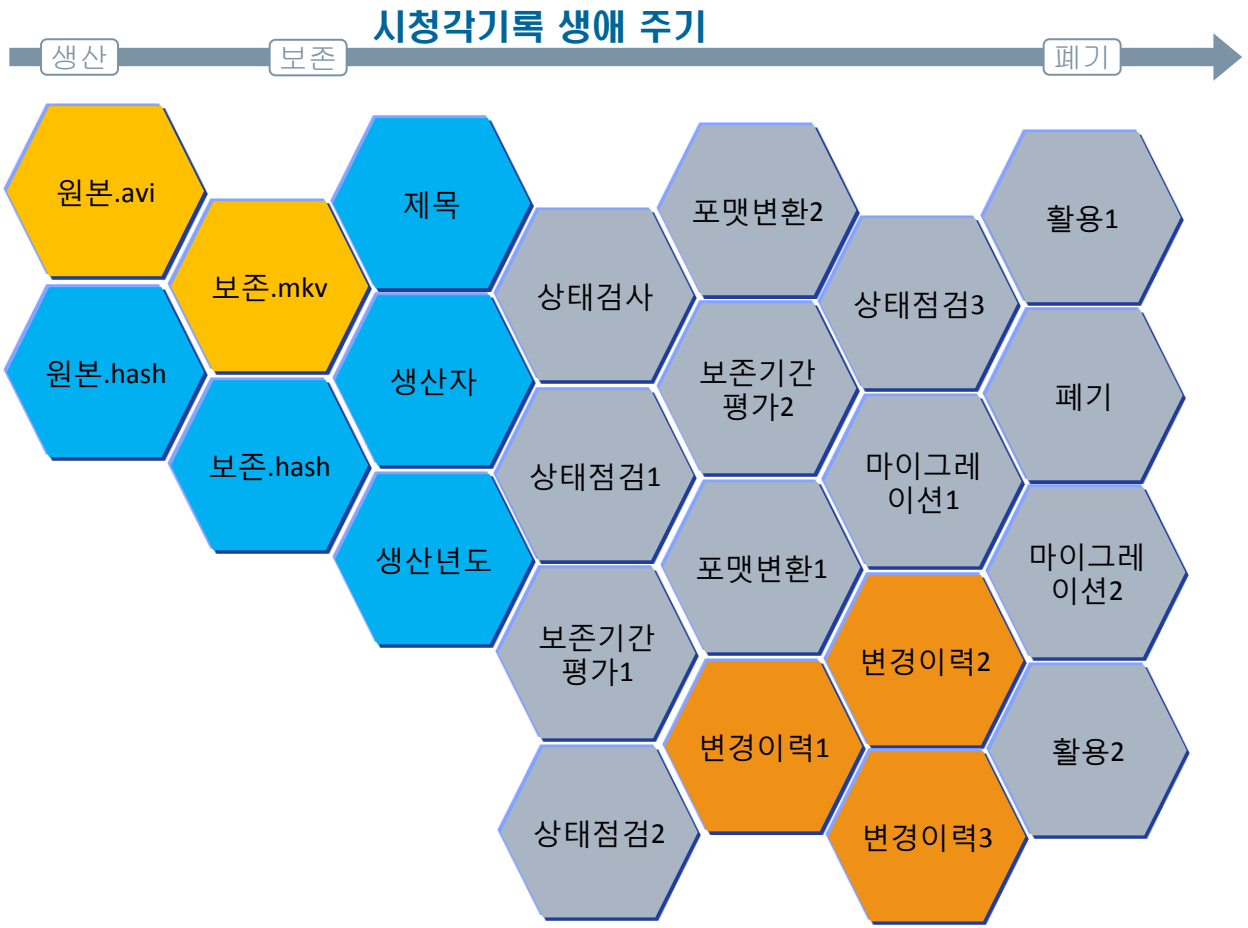
메타변경 이력

순번	수정메타	삭제메타	수정자	수정일
1	기록물 제목		홍길동	YYYY.MM.DD
2	훼손등급		홍길동	YYYY.MM.DD
3	상태결과		홍길동	YYYY.MM.DD
4		내용기술	홍길동	YYYY.MM.DD
중략				

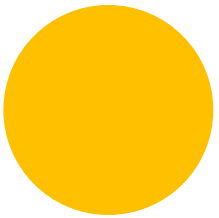
보존처리 이력

작업유형	작업상태	작업자	처리일
인제스트	완료	홍길동	YYYY.MM.DD
보존포맷 생성	완료	홍길동	YYYY.MM.DD
품질검사	완료	홍길동	YYYY.MM.DD
상태검사	완료	홍길동	YYYY.MM.DD
폐기	대기	홍길동	YYYY.MM.DD
중략			

1.03 시청각기록 장기보존 패키지 구성요소



패키지 크기(1,743MB)



- 메타데이터(10KB)
- 변경이력(10KB个)
- 관리행위(10KB个)
- 기록실물(1,742MB)

패키지의 99.99%이상

1.04 시청각기록 장기보존 패키지 프로세스

장기보존 패키징

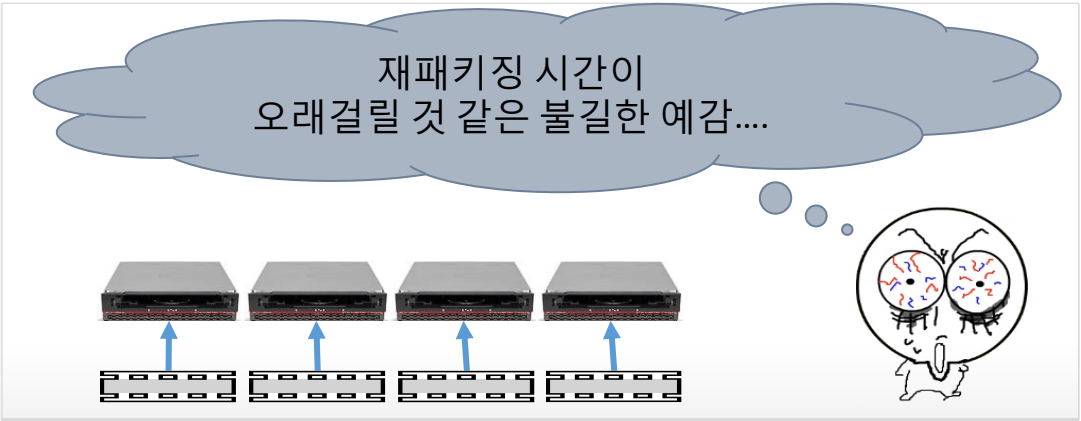
기록물 스토리지 → 장기보존 패키지 → 장기보존 저장장치

원본.avi
원본.hash
원본.xml
보존본.mkv
보존본.hash

ID: PLC00001

원본.avi	보존본.mkv
원본.hash	보존본.hash
원본.xml	

장기보존 Tape



장기보존 재패키징(메타데이터 변경)

언패키징 → 수정 → 패키징

장기보존 저장장치 → 장기보존 패키지 → 기록물 스토리지 → 메타데이터 수정 → 장기보존 재 패키지 → 장기보존 저장장치

장기보존 Tape

ID: PLC00001

원본.avi	보존본.mkv
원본.hash	보존본.hash
원본.xml	

원본.avi
원본.hash
원본.xml
보존본.mkv
보존본.hash

원본.avi
원본.hash
수정.xml
보존본.mkv
보존본.hash

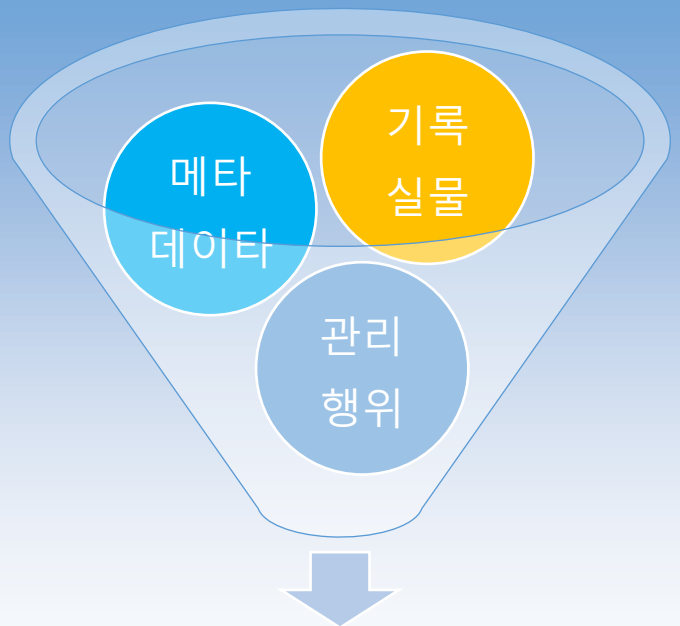
ID: PLC00001

원본.avi	보존본.mkv
원본.hash	보존본.hash
수정.xml	

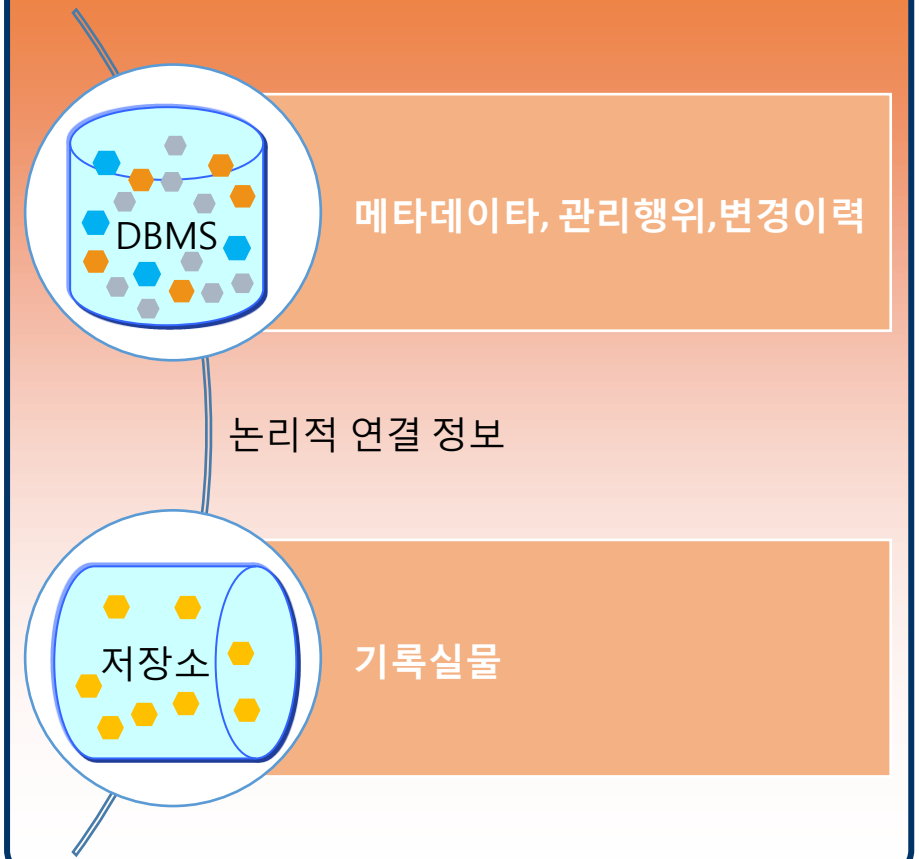
장기보존 Tape

패키징 소요시간 비교측정 시뮬레이션

(1차) NEO-X(BagIt)

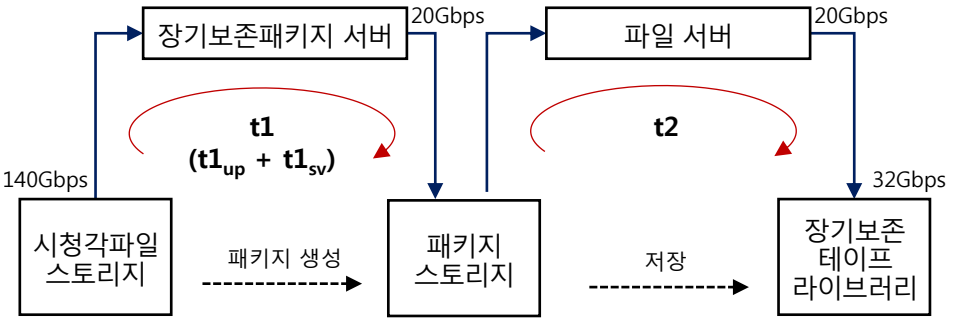


(2차) 분산관리 방식

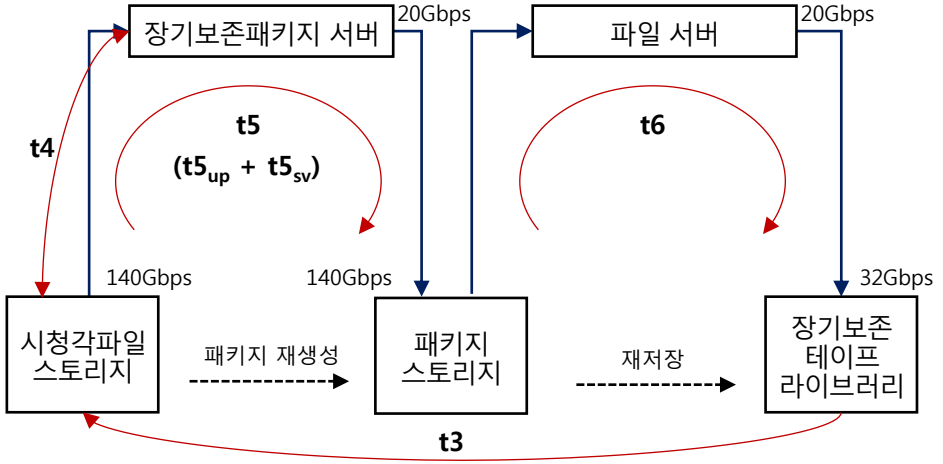


2.02 1차 시뮬레이션 시스템 구성도

시험 환경 시스템 사양	
시스템명	사양
장기보존패키지 서버	• Windows Server 2016 • 32GB 메모리 • 20Gbps 네트워크 I/O
파일서버	• Windows Server 2012 • 32GB 메모리 • 20Gbps 네트워크 I/O • 8Gbps SAN I/O
장기보존 테이프 라이브러리	• 10PB 저장 공간 • 8Gbps I/O
시청각 스토리지	• 520TB 저장 공간 • 2.9TB Cache 메모리 • 140Gbps 네트워크 I/O



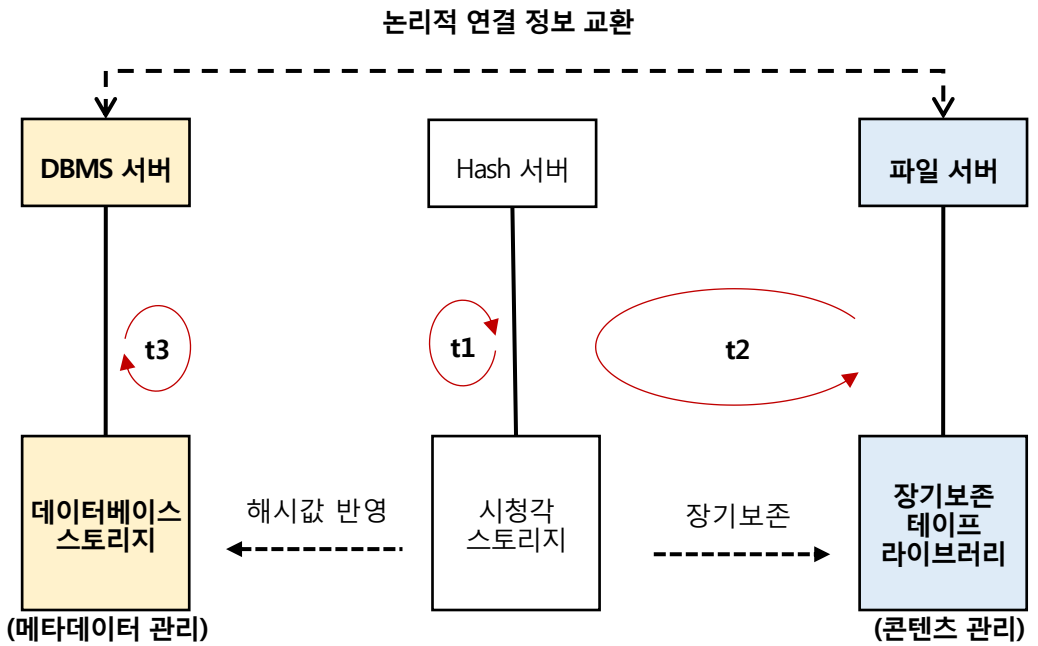
(최초 패키지 생성 시)



(메타데이터 변경 시)

2.03 2차 시뮬레이션 시스템 구성도

시험 환경 시스템 사양	
시스템명	사양
Hash 서버	• Windows Server 2016 • 32GB 메모리 • 20Gbps 네트워크 I/O
파일서버	• Windows Server 2012 • 32GB 메모리 • 20Gbps 네트워크 I/O • 8Gbps SAN I/O
장기보존 테이프 라이브러리	• 10PB 저장 공간 • 8Gbps I/O
시청각 스토리지	• 520TB 저장 공간 • 2.9TB Cache 메모리 • 140Gbps 네트워크 I/O



2.04 영상기록물 재패키징 소요 시간

1차

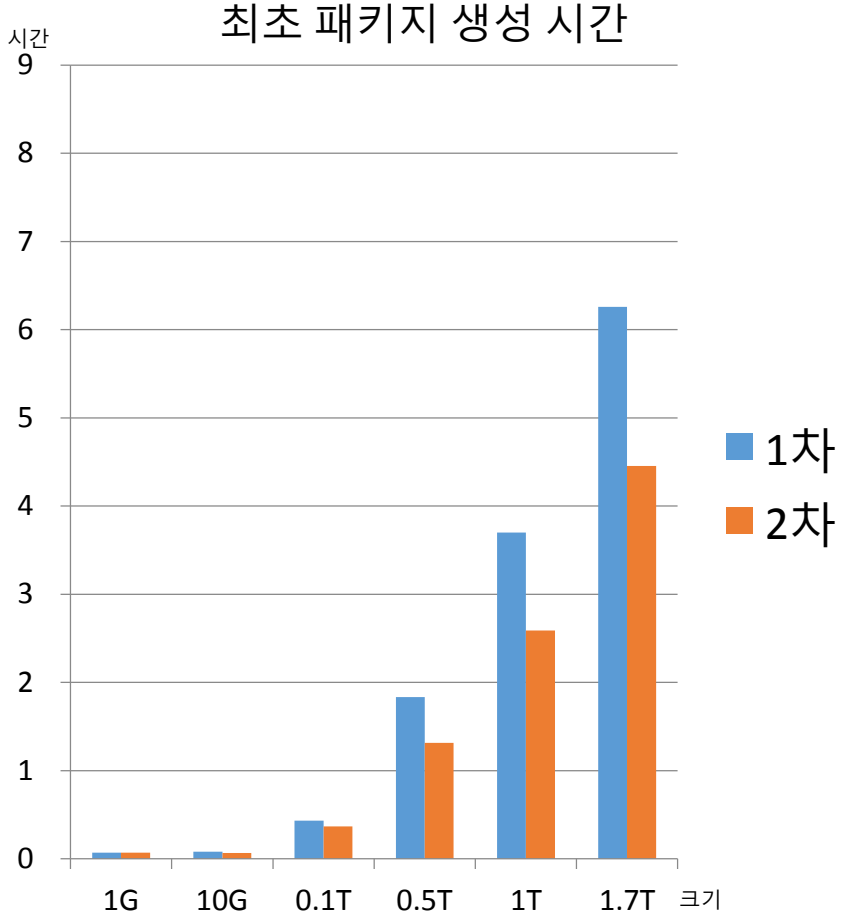
패키지 파일용량	최초 패키지 생성 시			메타데이터 변경 시				
	t1		t2	t3	t4	t5		t6
	$t1_{up}$	$t1_{sv}$				$t1_{up}$	$t1_{sv}$	
1GB	4s	6s	4m 01s	3m 31s	< 1s	5s	5s	3m 51s
10GB	35s	36s	3m 33s	3m 56s	< 1s	36s	38s	4m 01s
100GB	11m 38s	3m 10s	11m 04s	12m 14s	< 1s	9m 32s	3m 15s	10m 04s
500GB	39m 43s	31m 29s	38m 44s	39m 24s	< 1s	38m 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

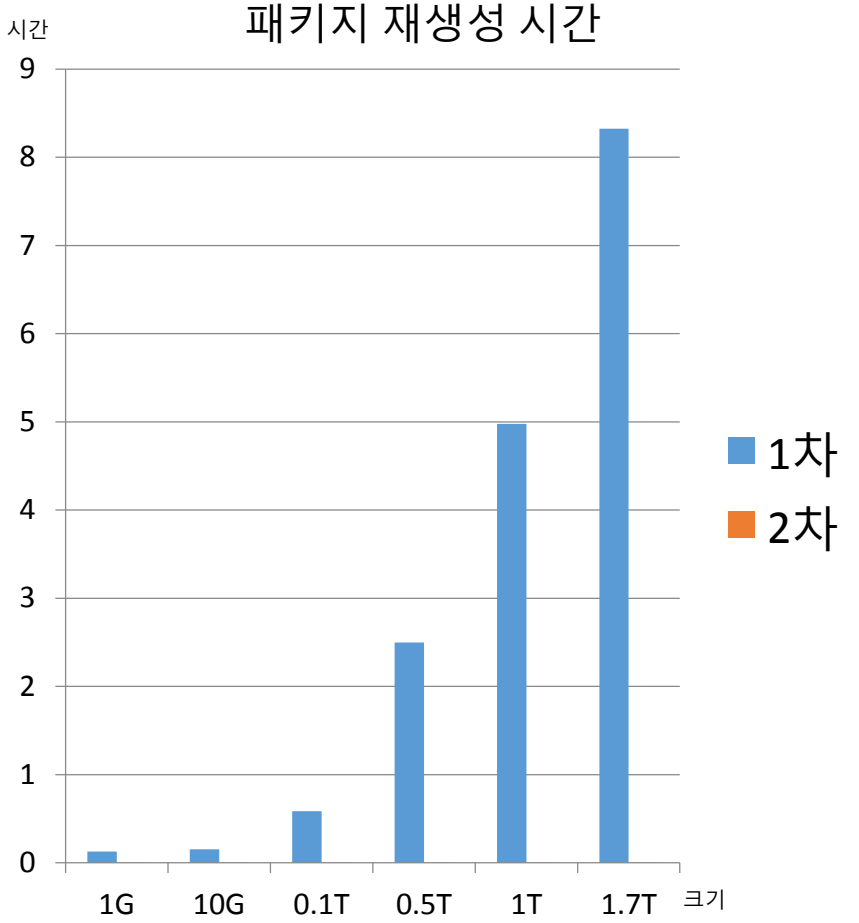
2차

패키지 파일용량	최초 장기보존 시		메타데이터 변경 시
	t1	t2	t3
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

1,2차 소요시간 그랙픽, 비율 그래픽

1, 2차 시뮬레이션 비교

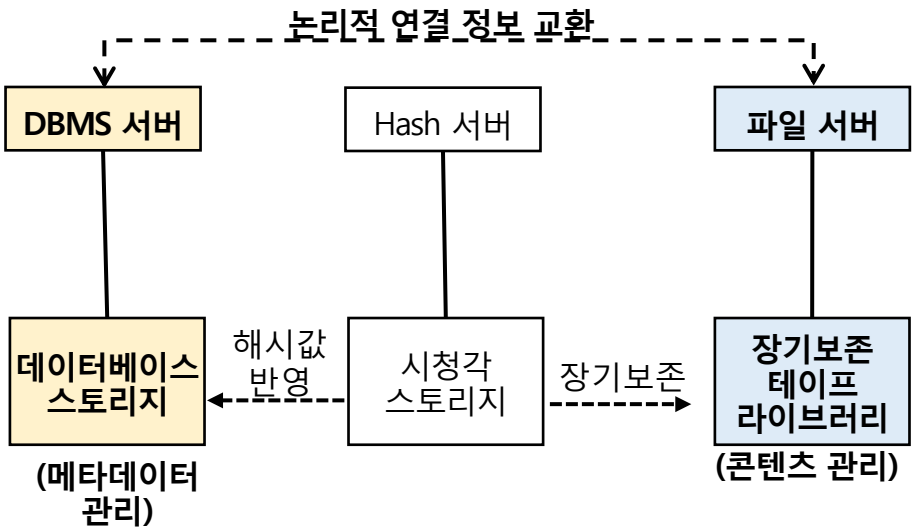

최초 패키지 생성 시간
시간
9
8
7
6
5
4
3
2
1
0
1G
10G
0.1T
0.5T
1T
1.7T
크기
1차
2차


패키지 재생성 시간
시간
9
8
7
6
5
4
3
2
1
0
1G
10G
0.1T
0.5T
1T
1.7T
크기
1차
2차

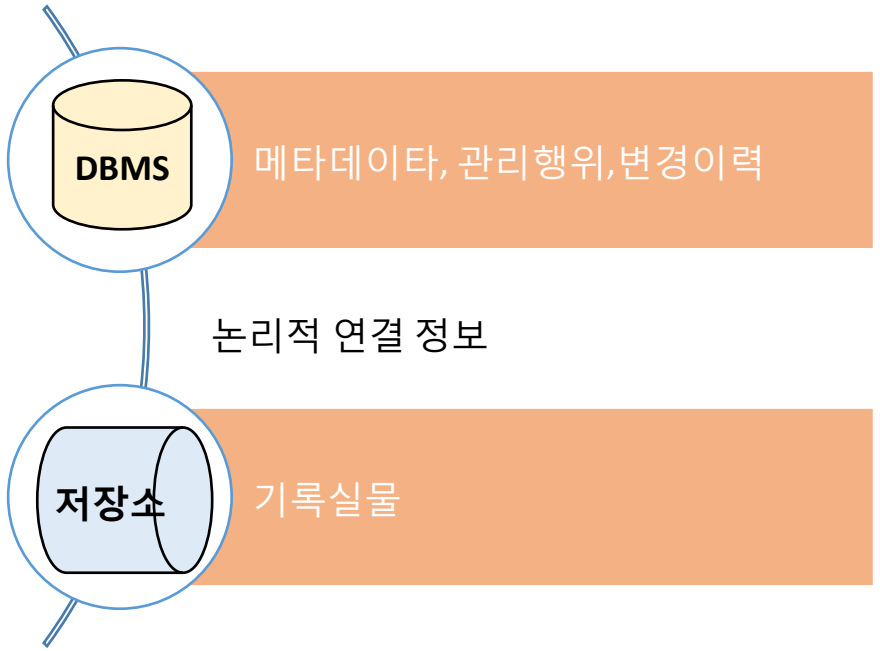
시청각기록물 장기보존 방식 제안

분산관리 방식

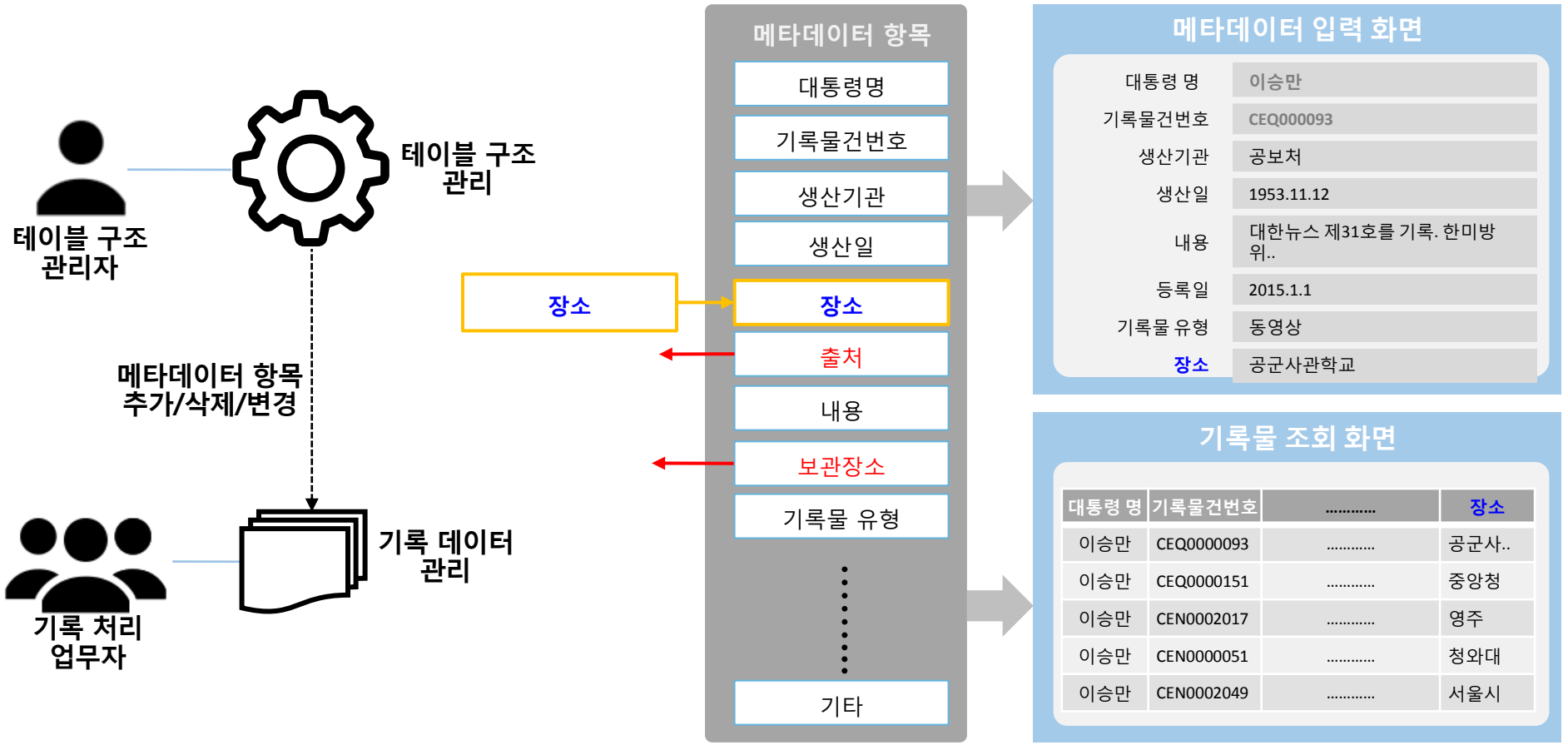
시스템 구성



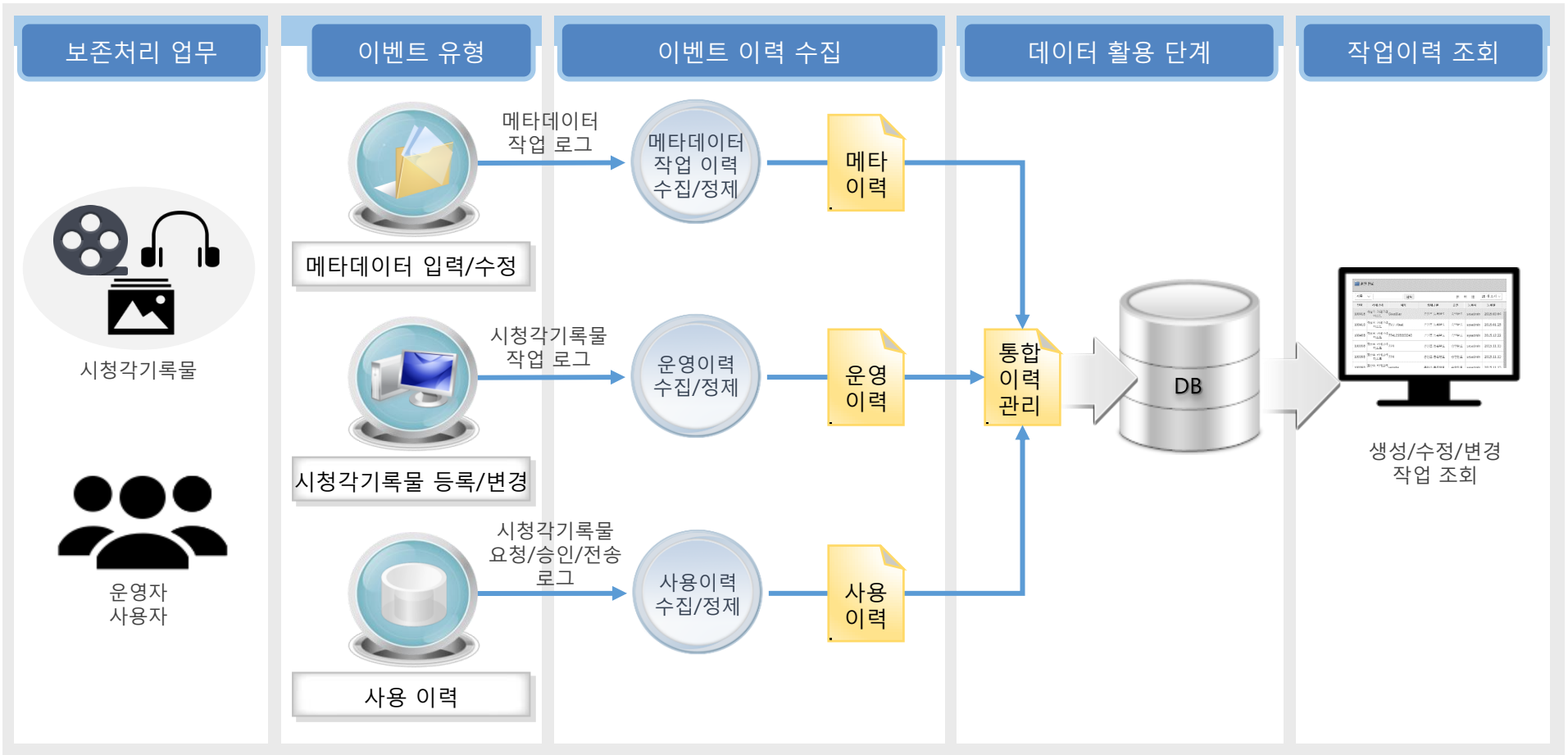
콘텐츠+메타데이터 관리 장소



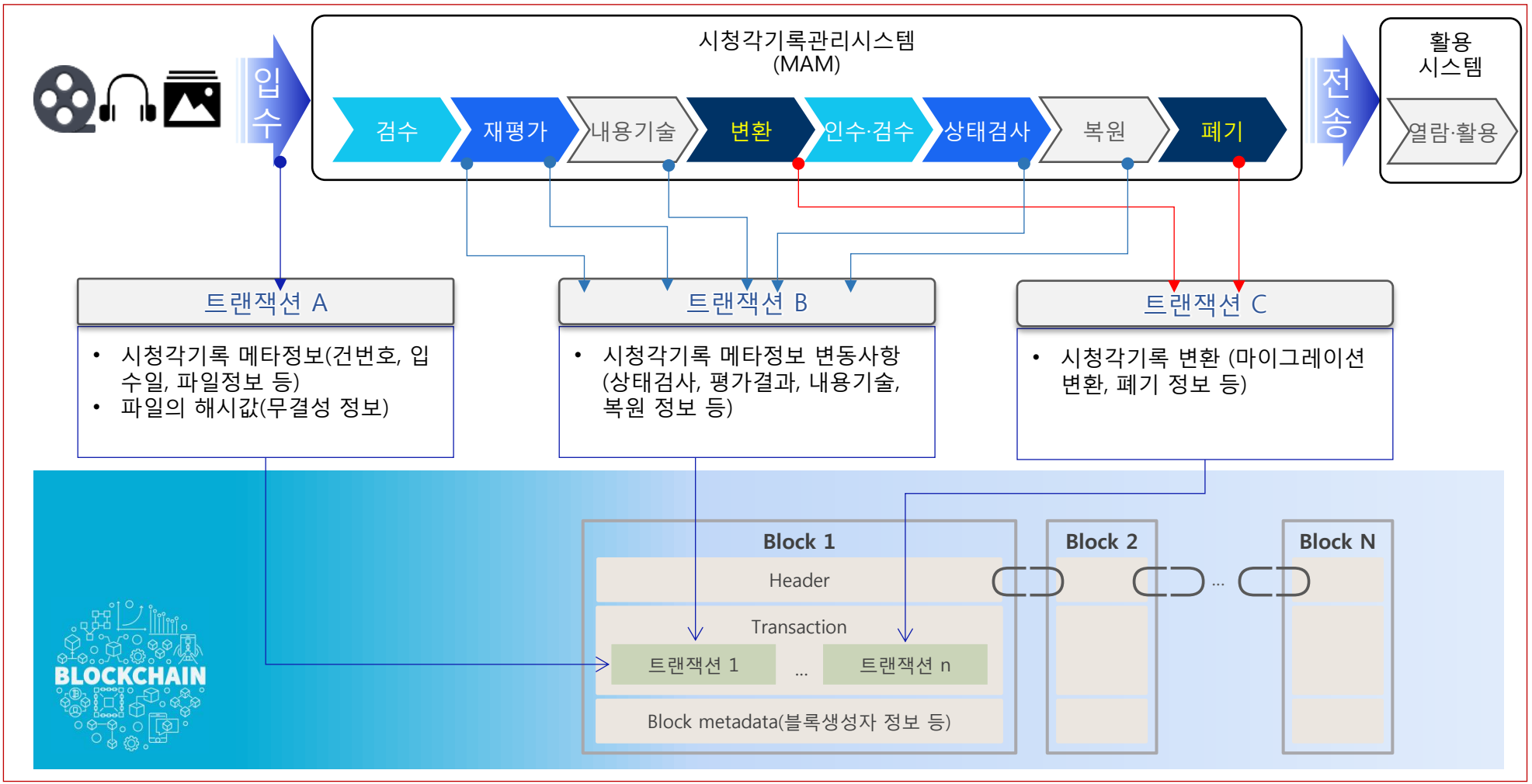
3.01 분산관리 방식 고려사항 – 가변메타



3.02 분산관리 방식 고려사항 – 이력관리



3.03 분산관리 방식 고려사항 – 신뢰성 제고



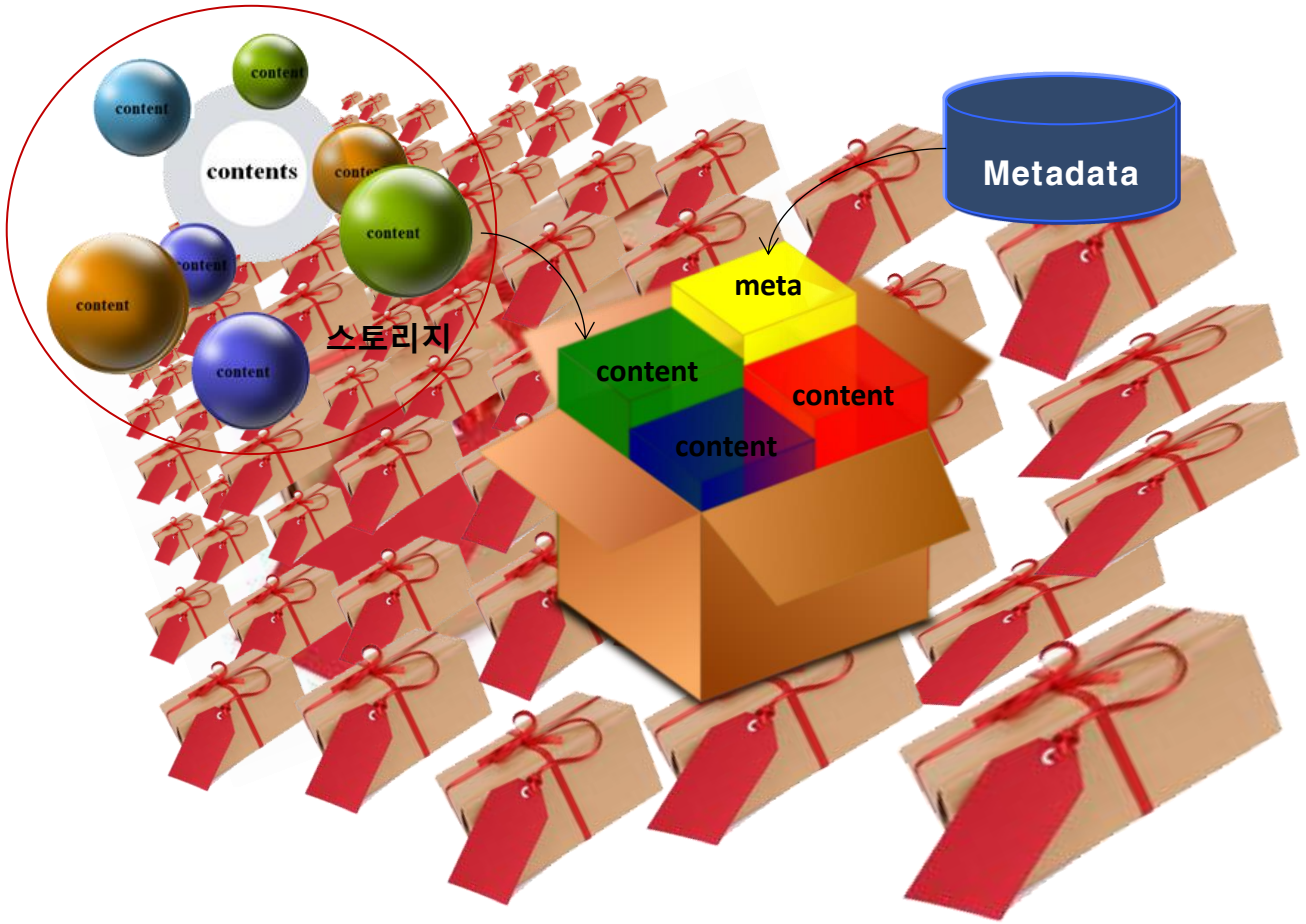
2019년 제2차 기록관리 현안과제 세미나

디지털 시청각기록 장기보존정책 수립 전략

2. 방법의 근거 및 신뢰요건

행정안전부
국가기록원

1.01 패키징 방법 → 분산관리 방법 / 블록체인 제안



공개/비공개

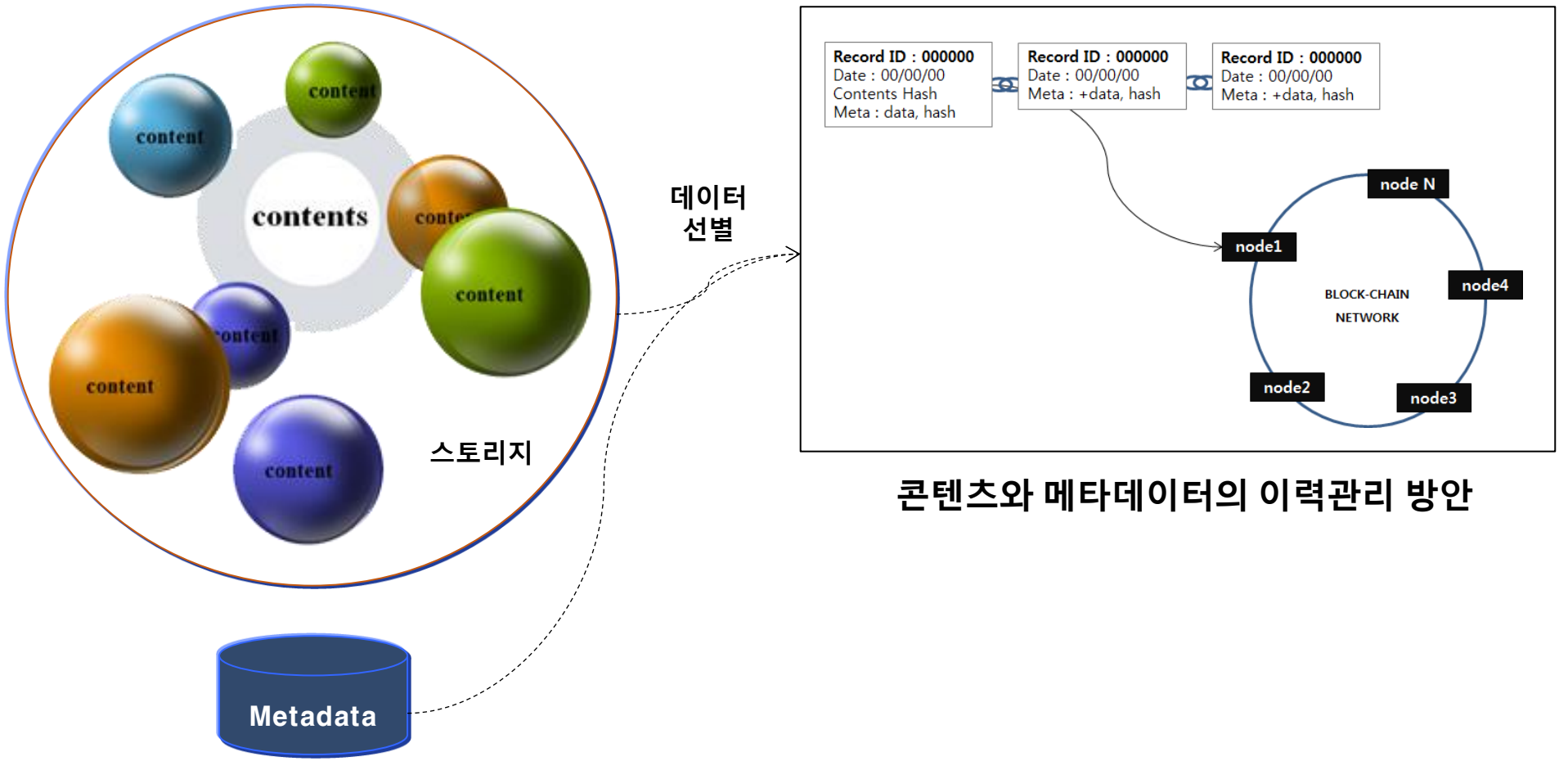
보존기간

내용기술

시스템 변경

등등

1.02 패키징 방법 → 분산관리 방법 / 블록체인 제안



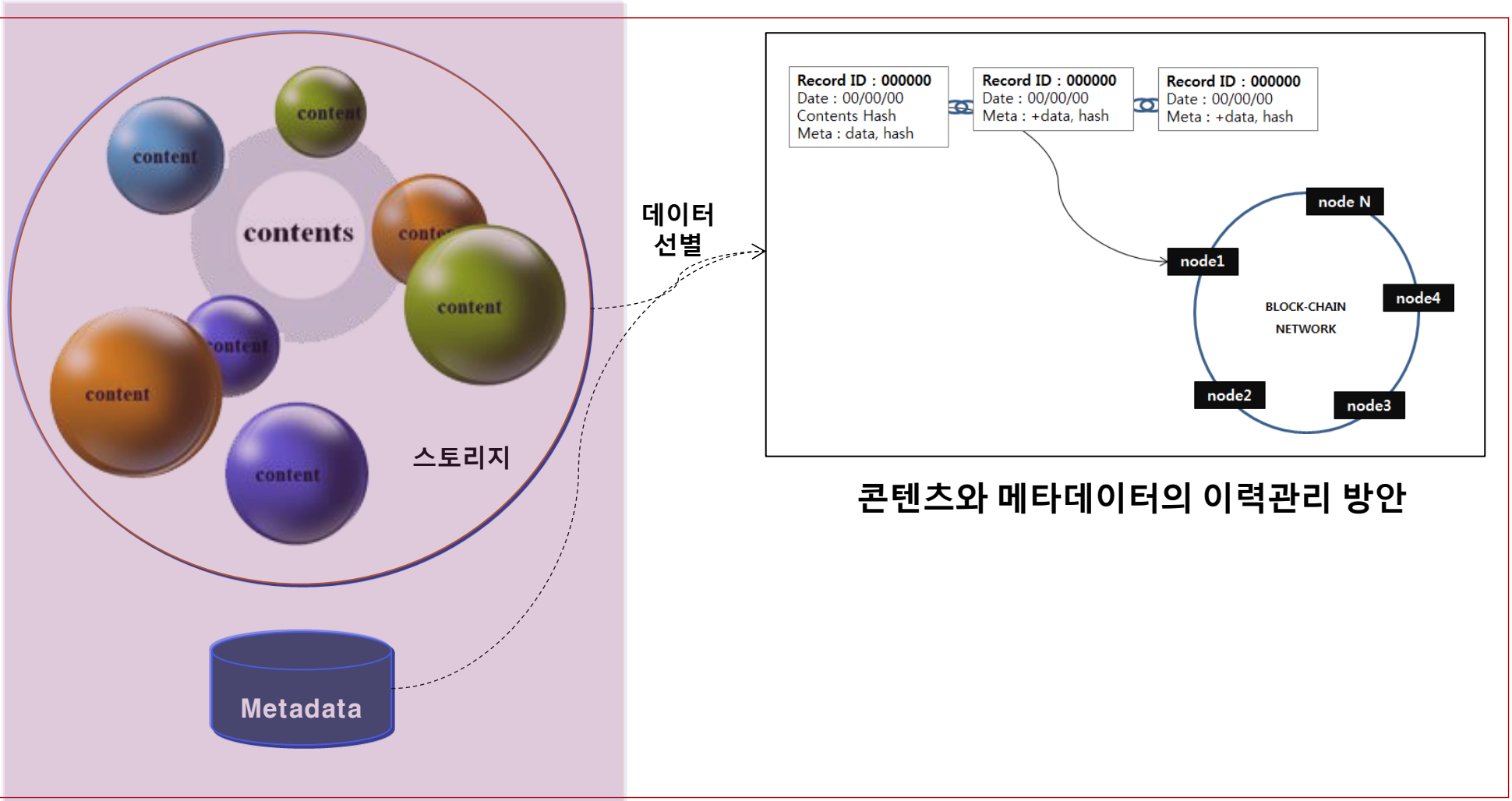
콘텐츠와 메타데이터의 이력관리 방안

1.03 분산관리, 전자기록의 신뢰요건

- ISO 15489-1 : Records management Part I, 2016.4월
- ISO 14721 : OAIS Reference model, 2012.9월
- InterPares project - I, II 결과

  - 전자기록 관리와 보존을 위한 국제협력 아젠다 개발 (2007.12월, 국가기록원)

2.01 분산관리 방법 근거



콘텐츠와 메타데이터의 이력관리 방안

2.01 분산관리 방법 근거

ISO 15489-1

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.

기록 콘텐츠와 메타데이터는 공존하는 복수의 시스템(위치) 또는 단일 시스템 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

2.02 분산관리 방법 근거

ISO 15489-1

Records systems should :

c) support the creation and maintenance of logical relationships between records content and metadata for records.

기록 시스템은 기록 콘텐츠와 메타데이터 간 논리적 연결정보의 생성 및 유지를 지원하여야 한다.

5.3 Records systems, ISO 15489-1

콘텐츠와 메타데이터는 ;
하나의 객체로 관리할 수 있거나 또는
논리적 연결정보와 함께 개별 관리 가능

2.03 분산관리 방법 근거

ISO 14721

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.**

정보패키지(information package)는 콘텐츠 정보와 보존기술정보에 대한 개념적 컨테이너이다. 콘텐츠 정보와 보존기술정보는 패키징 정보(packaging information)에 의하여 캡슐화되고 식별 가능하다.

2.2.2 Information package definition, ISO 14721

2.04 분산관리 방법 근거

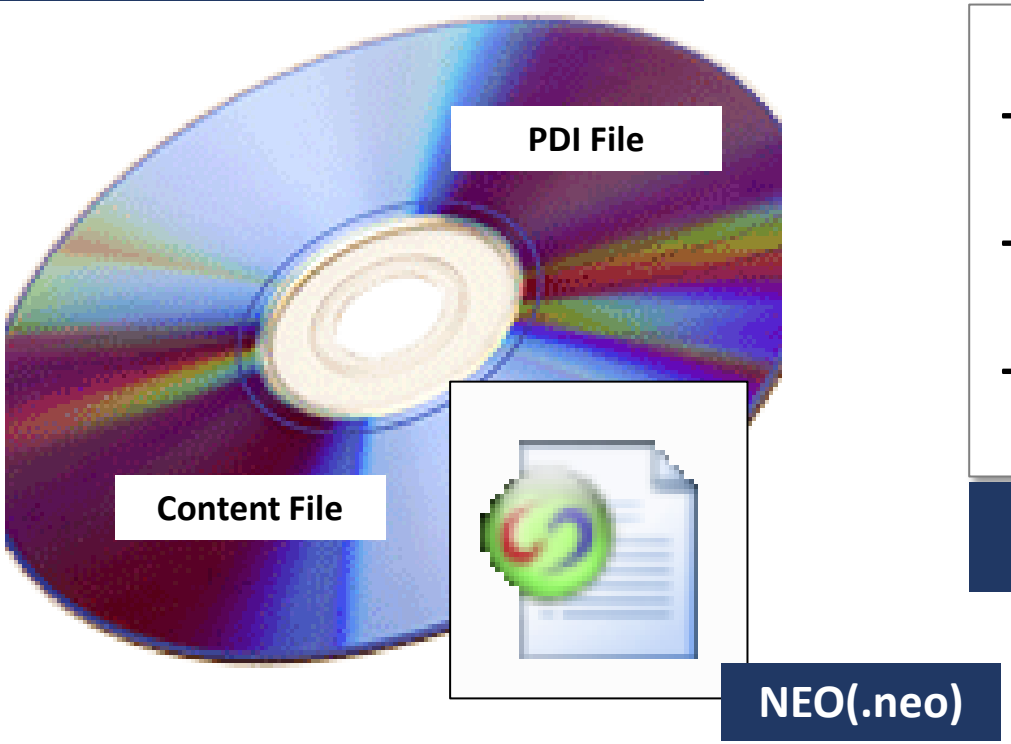
ISO 14721

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.

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-ROM내의 파일에 대한 파일명 및 디렉토리 정보를 포함할 것이다.

2.2.2 Information package definition, ISO 14721

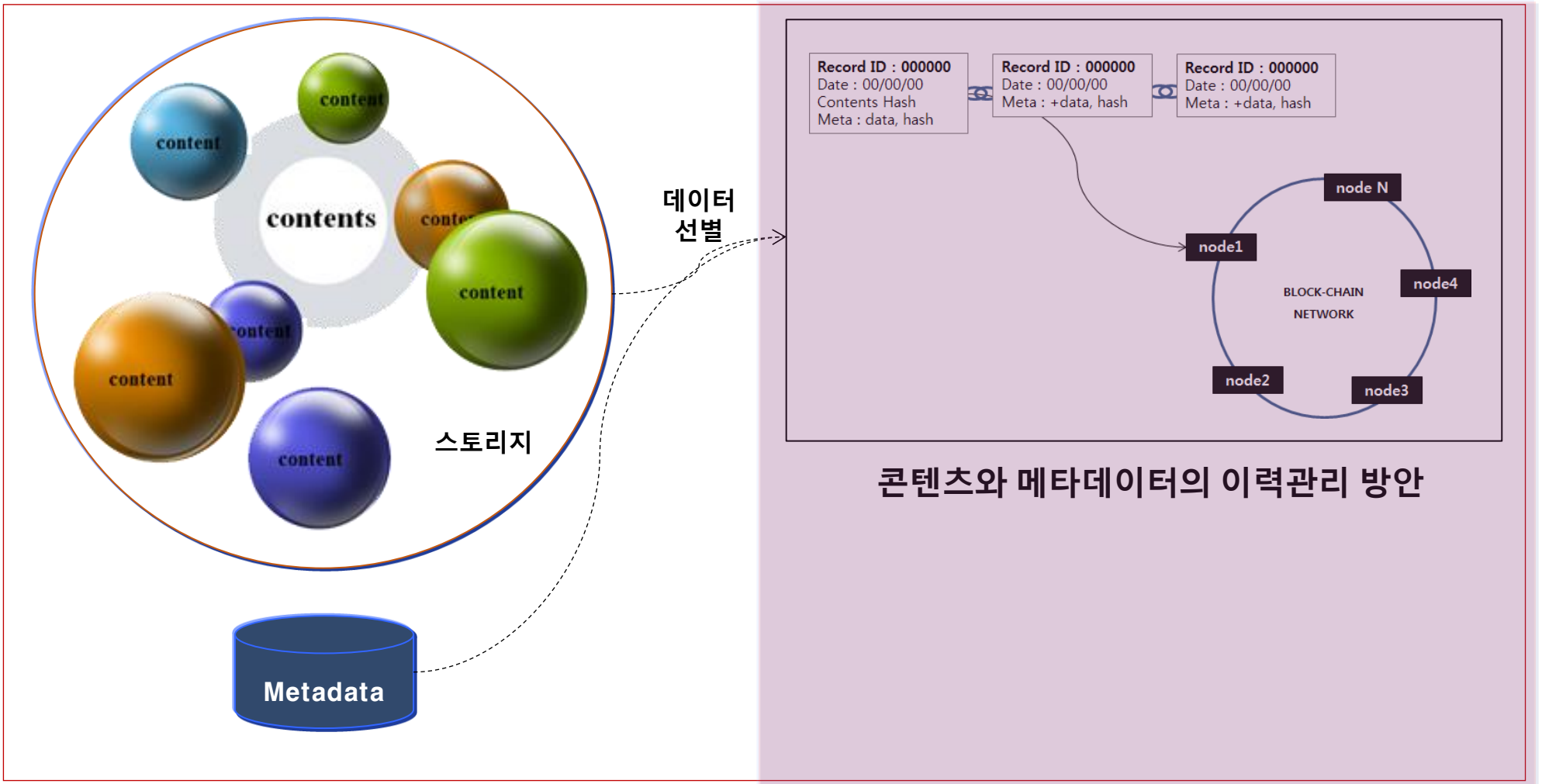
Information package (Conceptual container)



- CD의 파일시스템 : ISO-9660
- Content File : \records\AAA\
- PDI File : \metadata\CCC\

Packaging Information

3.01 전자기록관리 신뢰요건



콘텐츠와 메타데이터의 이력관리 방안

3.01 전자기록관리 신뢰요건

ISO 15489-1

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.

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. (요약)

5.3.2.1 Reliable, ISO 15489-1

3.02 전자기록관리 신뢰요건

ISO 15489-1

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.

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 발생되는 process metadata가 있다. (요약)

5.2.3 Metadata for records, ISO 15489-10

3.03 전자기록관리 신뢰요건

정책프레임워크 : 정책수립-범위

- 연속적 정책의 모든 과정에, 다시 말해서 생산, 활용과 유지 및 처분 등 기록생애주기 전반에 걸친 모든 단계에서 보존에 관련된 고려사항을 편입시켜야 한다. (생산자원칙 7 및 보존자원칙 7, InterPares project 2)
- 마이그레이션이나 에뮬레이션 등 장기보존 조치를 지속적으로 취하는데 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)

정책프레임워크 : 진본성 – 정체성

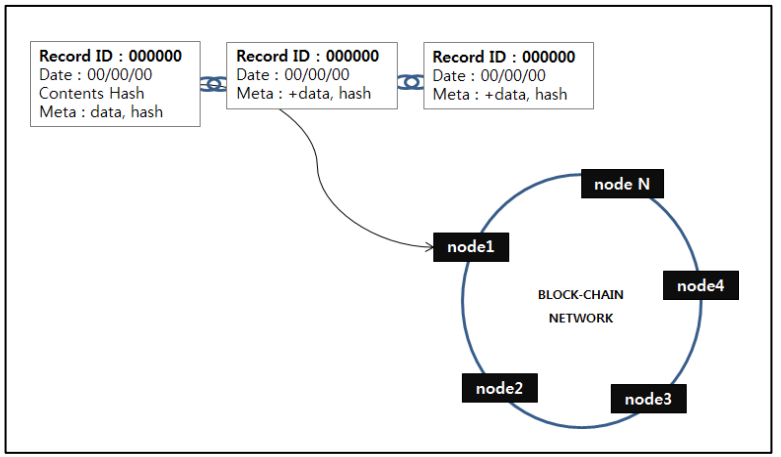
- 기록의 진본성 유지와 평가를 위한 벤치마크 요건은 기록의 속성을 표현하고 기록을 표현한 그 정보를 기록에 연계하는 것을 그 첫 번째 요건으로 들고 있다. (벤치마크 요건 1, InterPares project 1)

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, InterPares project 1)

01 전자기록관리 신뢰요건 정의

- Capture metadata와 Process metadata의 구분 관리
- 기록 생애주기 동안 기록에 영향을 미치는 모든 행위 및 결과의 기록화
- 기록 및 시스템 감사기록 존재 (감사증적)
- 믿을 수 있는 기록관리시스템의 사용



01 정리

□ 장기보존패키지(장기보존포맷) 인식 변화

- 전자기록의 장기보존은 장기보존패키지로 완성되는 것이 아님
- 장기보존패키지는 전자기록의 장기보존을 위한 여러 요소 중 하나
- 다양한 형태로 나타날 수 있는 개념적 컨테이너에 불과

□ 시스템 설계 시 고려 사항

- 최대한 추가 예상되는 메타데이터 수용 → 감사증적(Audit trail)
  - 데이터베이스 테이블 구조의 동적 설계
- 기록과 기록관리시스템의 신뢰 확보
  - 예. 블록체인시스템

3.04 정리

□ 전자기록의 장기보존 원칙

- 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.

(전자기록 관리와 보존을 위한 국제협력 아젠다 개발, 2007. 12월, 국가기록원)

토 론