



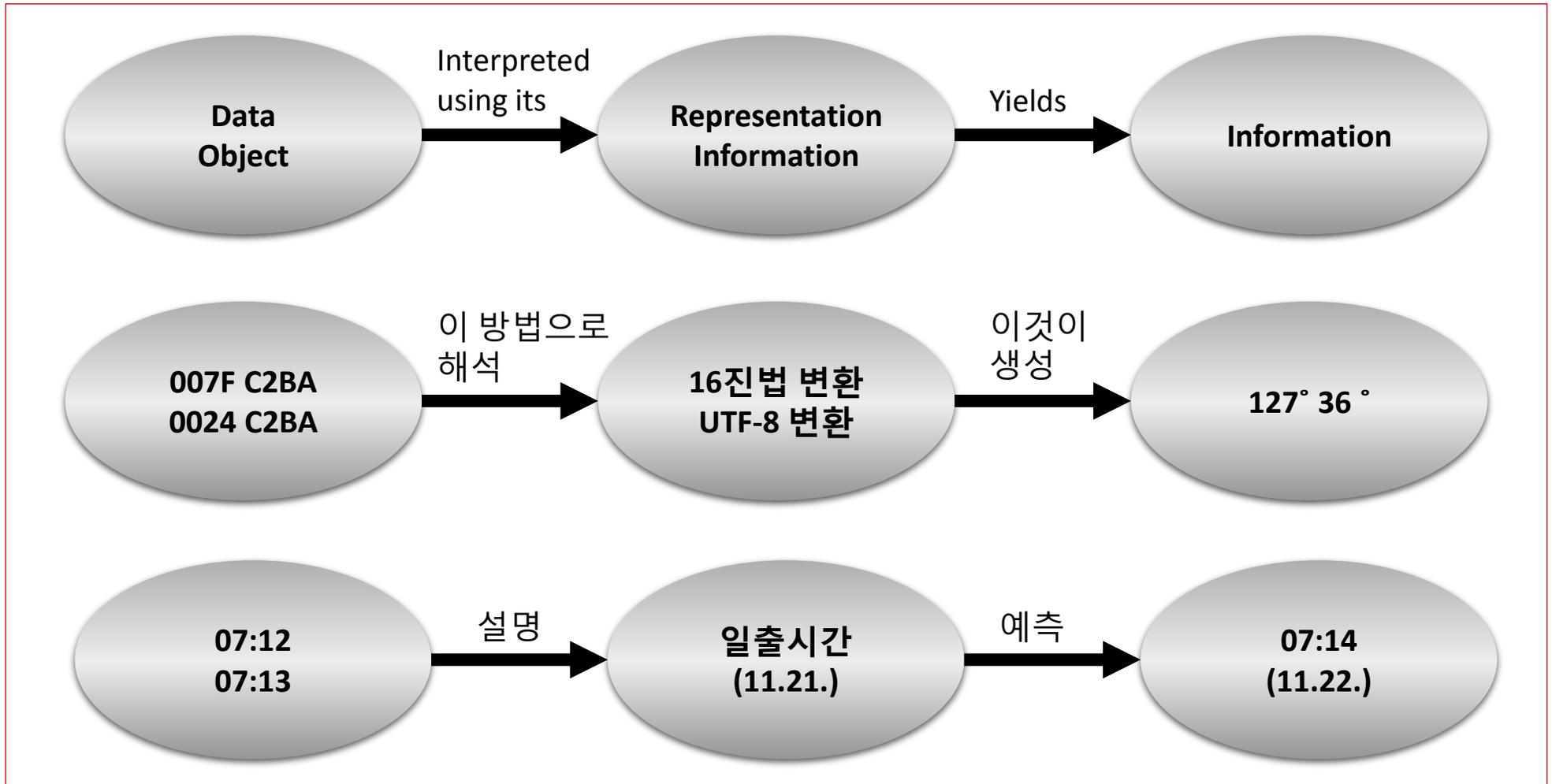
행정안전부
국가기록원

OAIS에서 말하는 I(Information) P(Package)의 이해

2019년 제6차 전자기록관리 연구포럼

대통령기록관 신동혁
2019.11.21.





$$I = \log \frac{1}{P(x)}$$

P는 사건이 발생할 확률, 발생 확률이 작은 것을 맞힐수록 정보량은 높다고 할 수 있음.
정보량이 많을수록 많은 경우의 수를 표현할 수 있다.

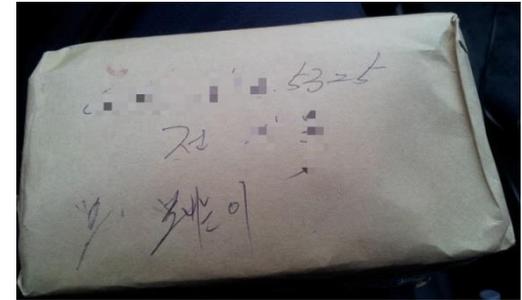
패키지(생활)



상자



소포



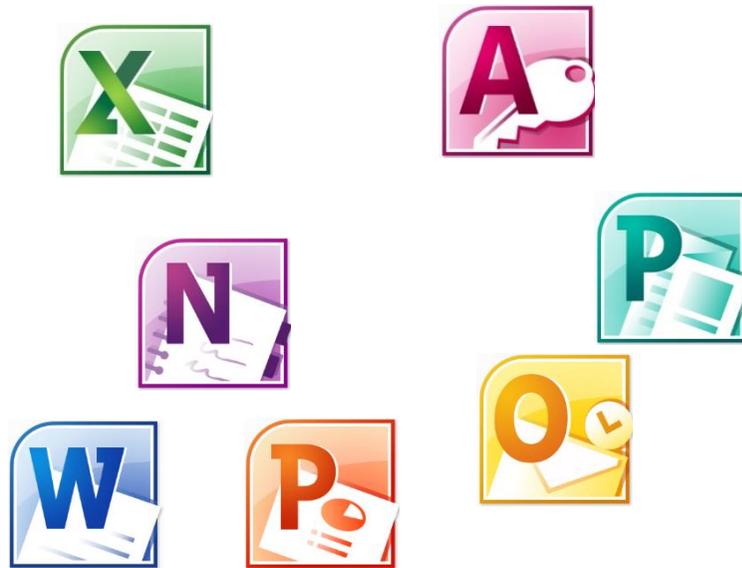
소화물



포장물

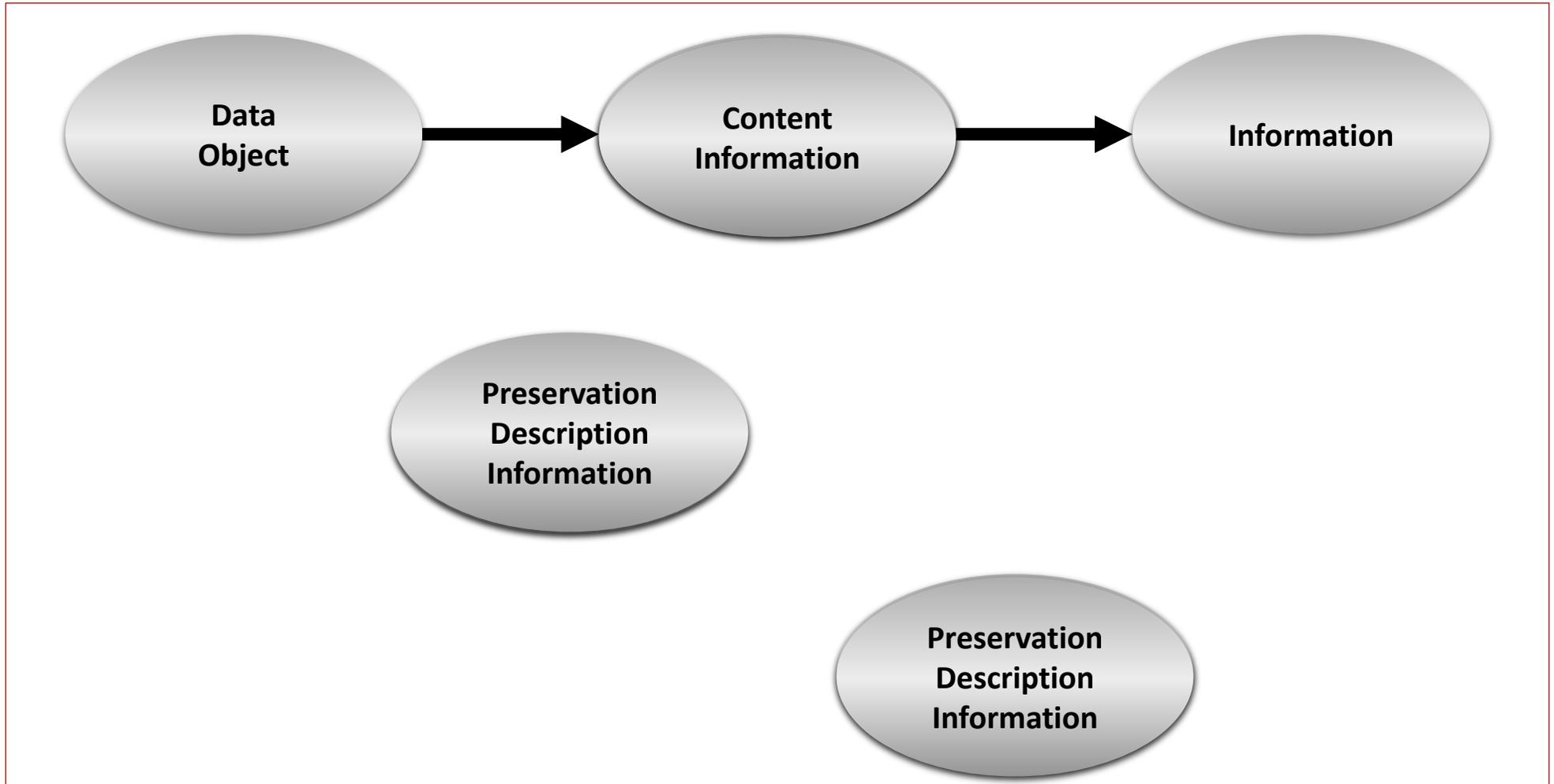


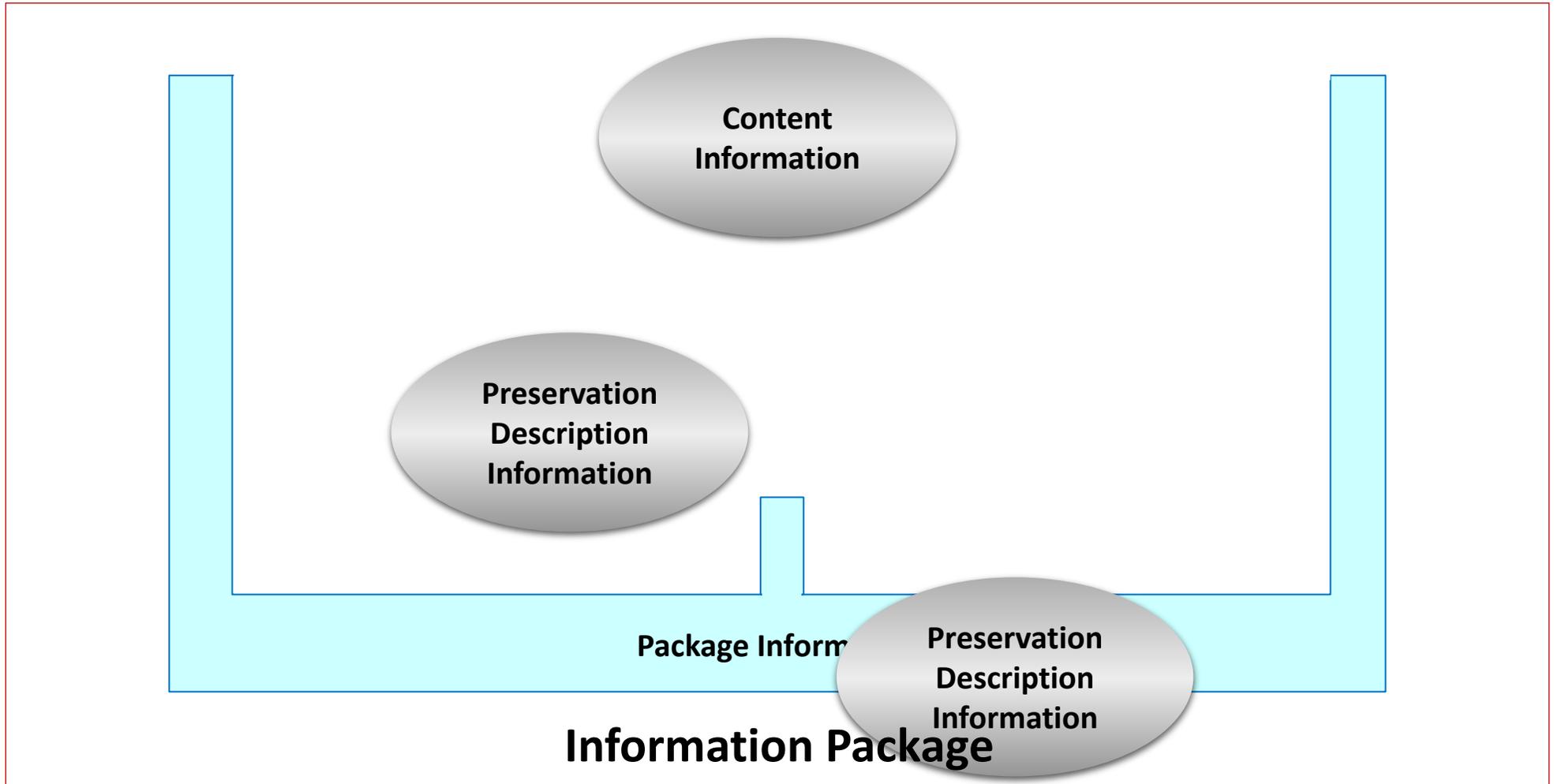
꾸러미



컴퓨터 이용자가 일상 업무에서 사용하는 응용 프로그램의 한가지이며, 시판되고 있는 완성품 프로그램 또는 프로그램 **집합체**이다.

데이터 → 정보 → 패키지





UML의 정의

Unified Modeling Language 시스템설계, 시스템 구현 등의 시스템 개발 과정에서, 개발자간의 의사소통을 원활하게 이루어지게 하기 위하여 표준화한 Modeling Language임

객체지향 설계의 3단계

1. 실세계 : 객체가 존재하는 실세계
2. 모델링 : 실세계에 존재하는 객체에 대해 객체가 어떻게 구성되었는지 분석하는 단계
3. 컴퓨터세계 : 추상화하여 생성한 상태에 대하여 구체화 과정을 통하여 소프트웨어로 변환



결합(Association)

- 개념적으로 서로 연결되어 있음
- 각클래스는 해당 결합 관계 내에서 역할 가짐
- 클래스 간을 이어주는 실선(연관선)으로 표기, 결합관계 명칭 표기



결합의 다중성(Multiplicity of Association)

Many(* 또는 n)

클래스 B의 인스턴스 하나에 A의 인스턴스 여러 개가 관계



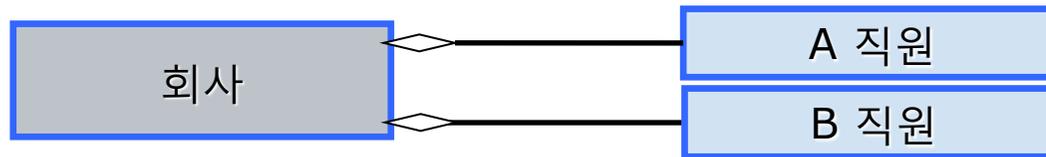
Zero to More(0..n)

클래스 B의 인스턴스 하나에 A의 인스턴스 0~n개가 관계



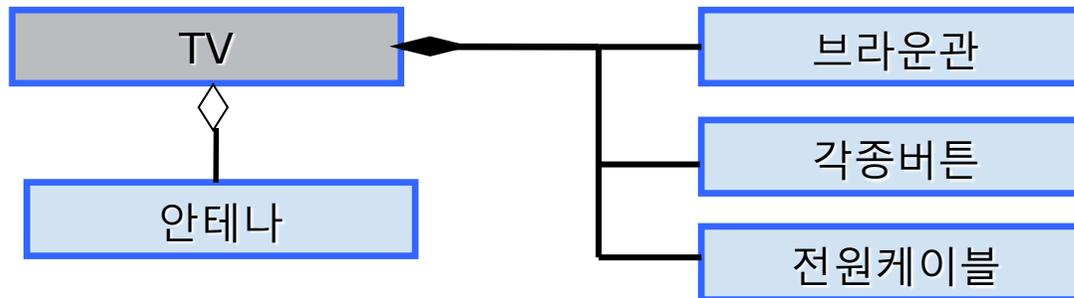
집합(Aggregation)

- 클래스가 전체-부분 관계, 약한 연결.
- 전체 인스턴스가 소멸 \times 부분 인스턴스가 소멸
- 표기법 : 클래스 쪽에 흰색 마름모 표기



합성(Composition)

- 클래스가 전체-부분 관계, 강하연결.
- 전체 인스턴스가 소멸 \times 부분 인스턴스가 소멸
- 표기법 : 전체가 되는 클래스 쪽에 검은색 마름모 표기

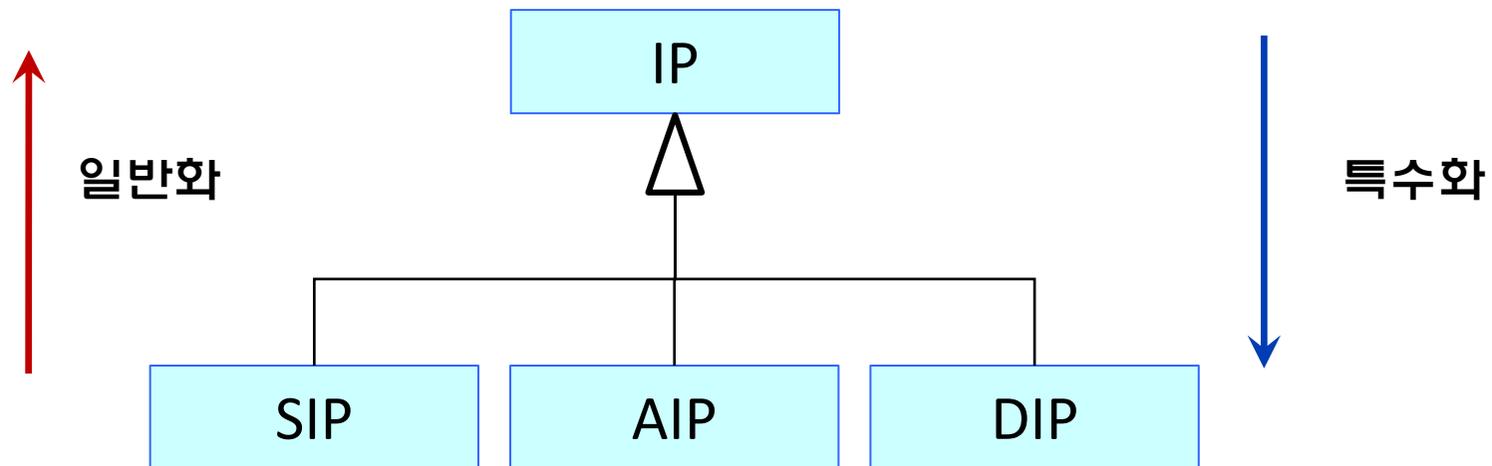


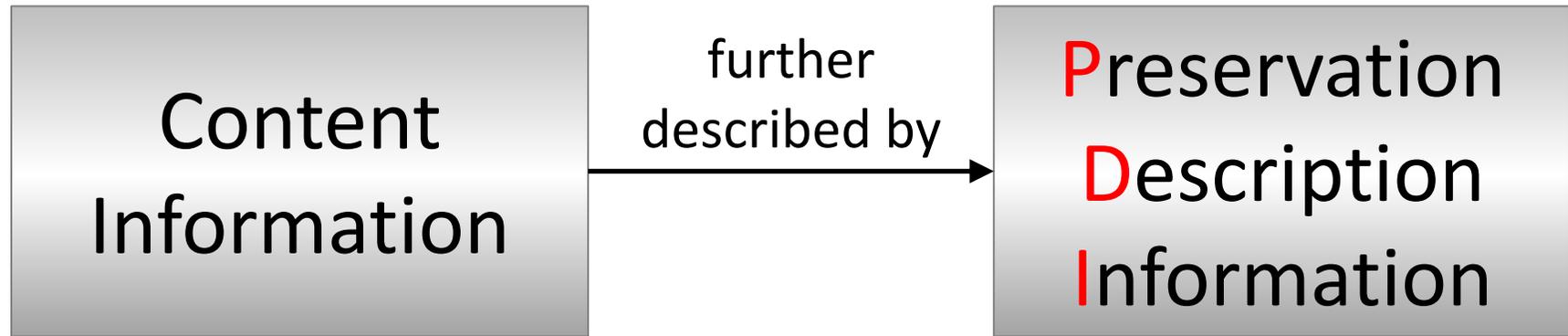
일반화/특수화(Generalization/Specialization)

클래스들간의 상속관계 지칭

일반화는 하위클래스들에 있어 공통적인 성질들을 상위클래스로 정의

특수화는 상위클래스에서 하위클래스를 보는 관점으로, 하위개념으로 내려 갈수록 상세화





PDI : 내용정보를 이해할 수 있도록 하는 정보
보존, 설명, 확인, 무결

Preservation Description Information

▶ PDI 예

열람정보 (Reference)	출처(기원.유래.)정보 (Provenance)	정황(맥락)정보 (Context)	고정(수리.수정)정보 (Fixity)
내용정보를 위한 확인자	생성 후 발생한 모든 사건(행위)의 이력정보	연관정보	비정상정 수정.접근 차단
<ul style="list-style-type: none"> 제목 인덱스 	<ul style="list-style-type: none"> 처리과정 이력 디지털화 과정 메타데이터 이전버전의 포인터 마이그레이션 정보 변환정보 	<ul style="list-style-type: none"> 첨부파일 포인터 사용법 뷰어 	<ul style="list-style-type: none"> 해쉬 증명서 암호화 CRC
DIP	AIP	SIP,DIP	SIP

ISO 14721:2003(참조 : *Preserving Digital Information*)

Reference	Provenance	Context	Fixity
-----------	------------	---------	--------

ISO 14721:2012(참조 : *Preserving Digital Information*)

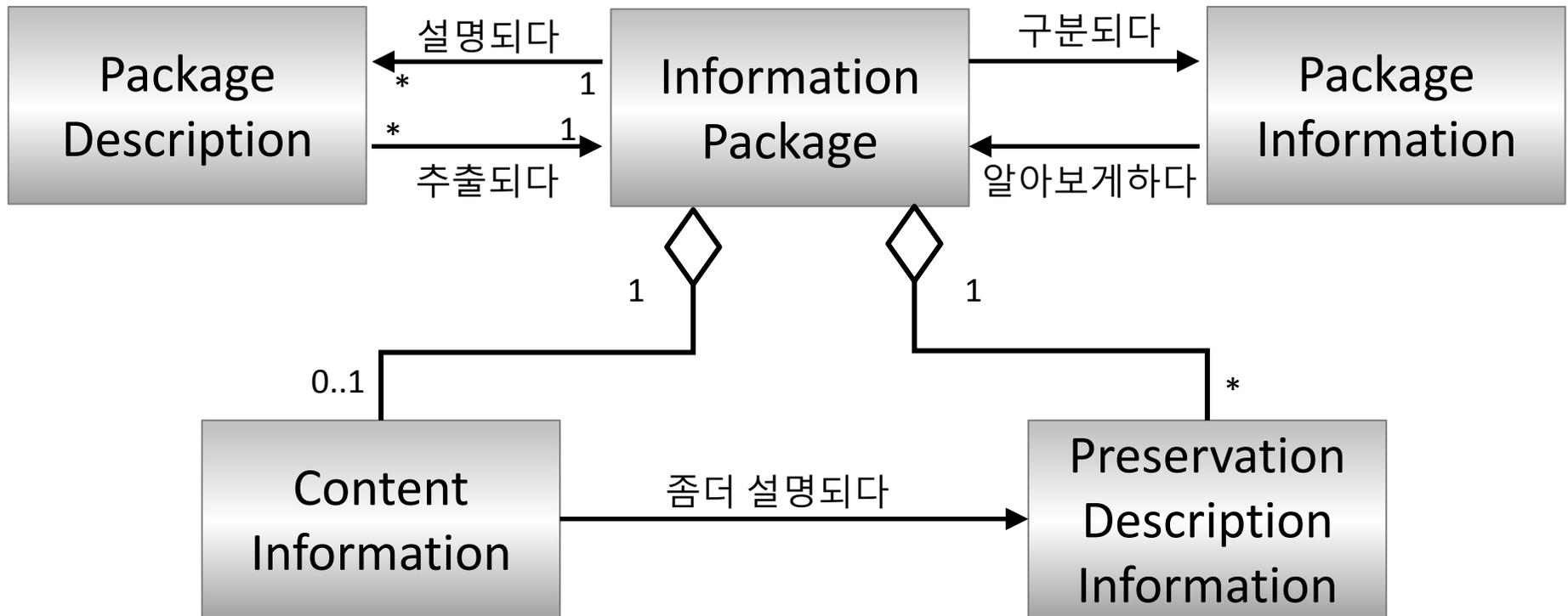
Reference	Provenance	Context	Fixity	Access Rights
-----------	------------	---------	--------	---------------

MPDI(*Multimedia Preservation Description Information requirements document : MPEG w12785*)

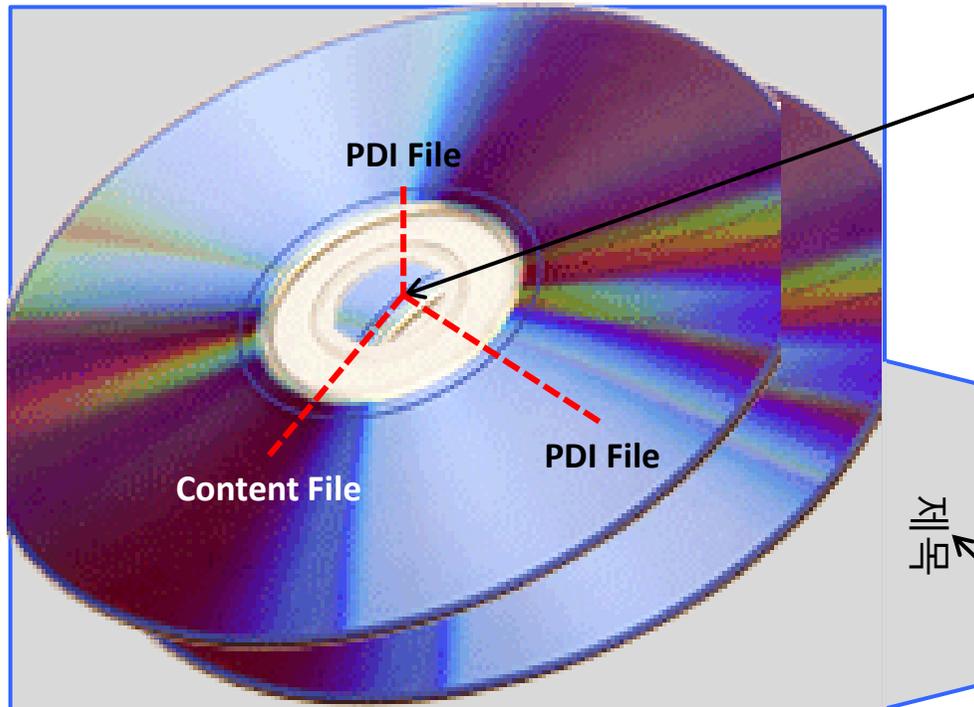
Reference	Provenance	Context	Fixity	Rights	Quality	Integrity	Authentication
-----------	------------	---------	--------	--------	---------	-----------	----------------

☞ 디지털 기록보존소 상황에 따라
항목과 분류는 조정·확대

Information Package



Information package (Conceptual container)



- CD의 파일시스템(ISO-9660)
- Content File 위치, 이름
- PDI File 위치, 이름
- Content, PDI의 연결정보

Packaging Information

- 제목, 키워드, 주제

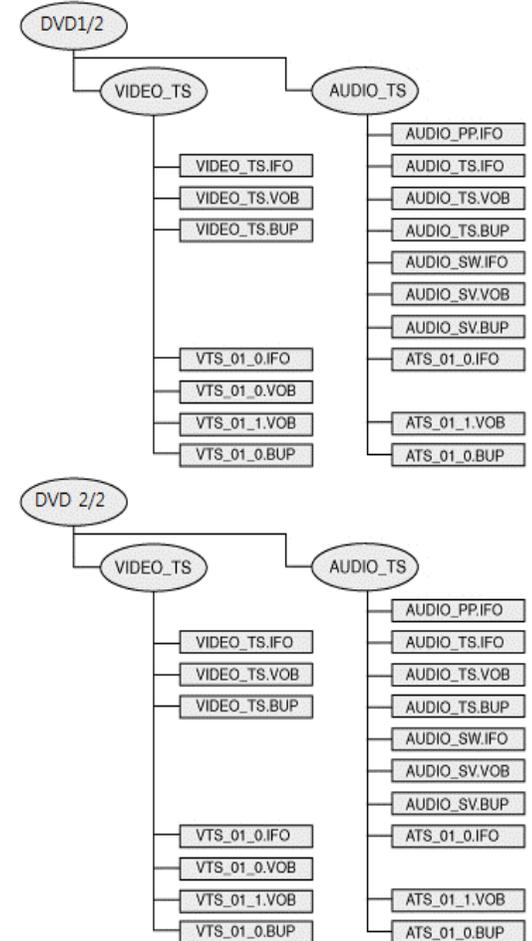
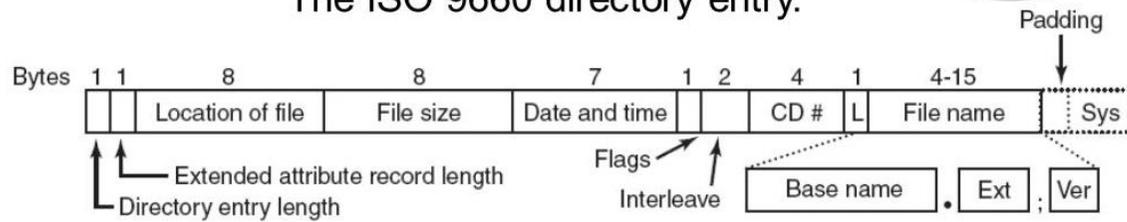
Package Description

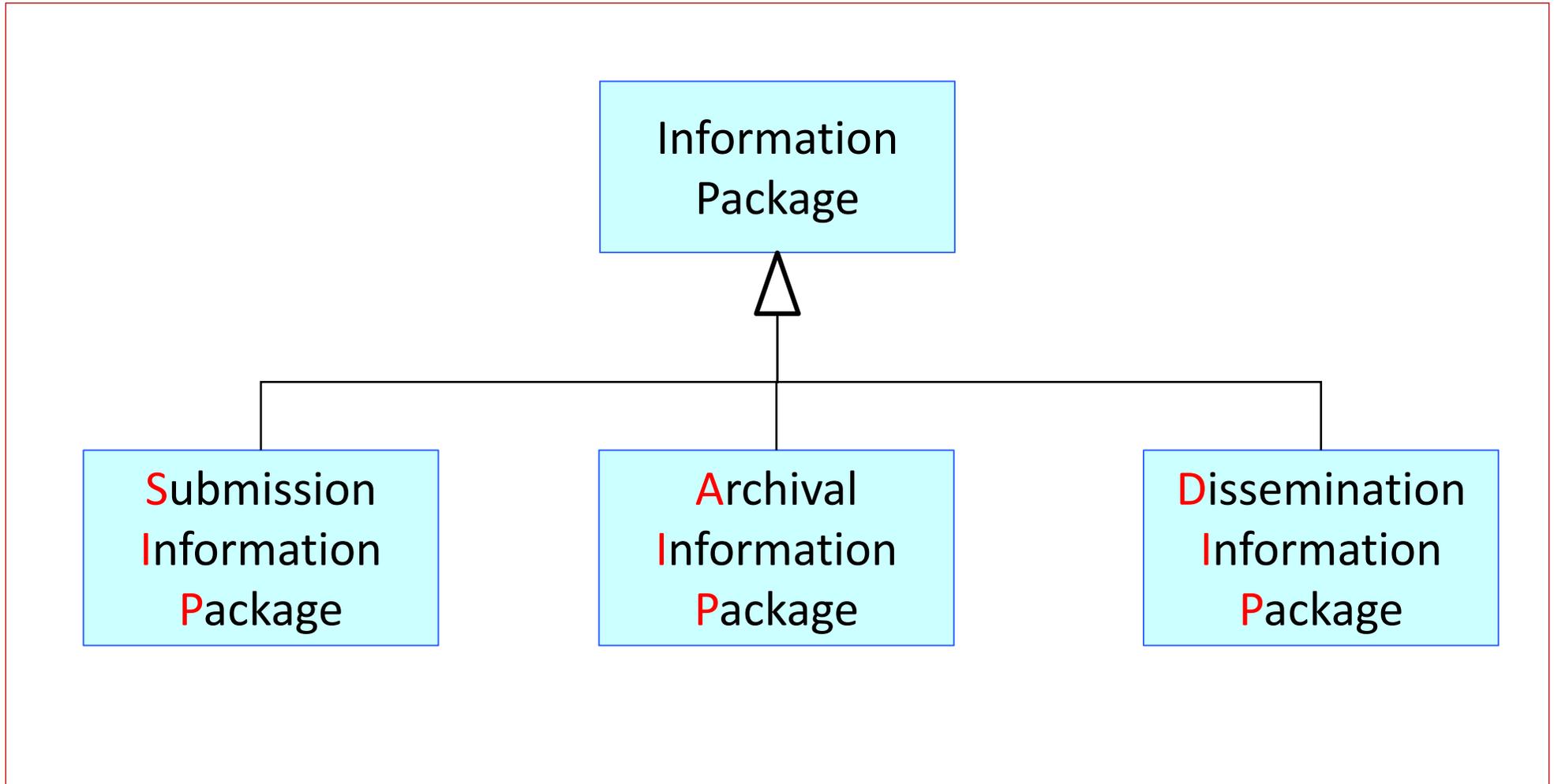
ISO에서 제정한 CD-ROM 매체를 위한 파일시스템 표준

The ISO 9660 File System

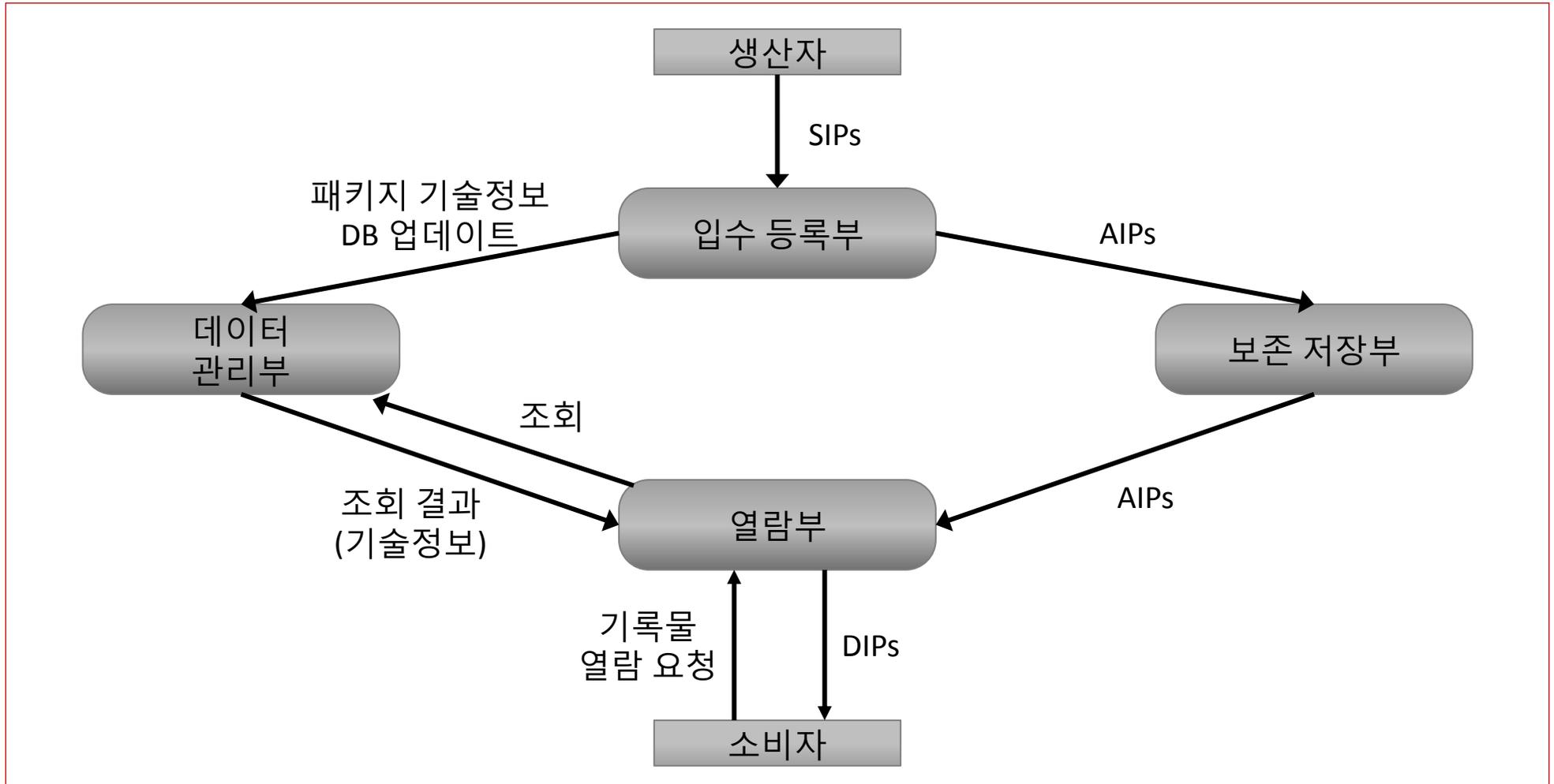


The ISO 9660 directory entry.





Information Package 흐름도



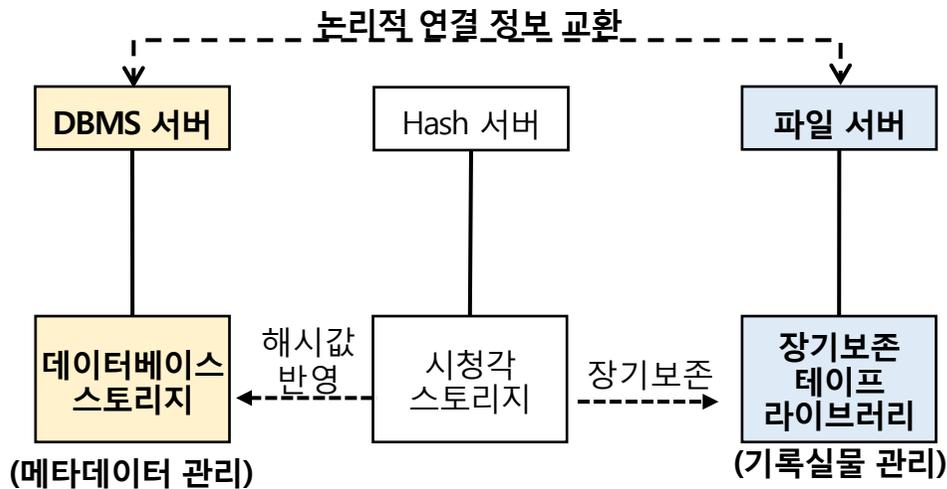
Information Package 3형제 관계



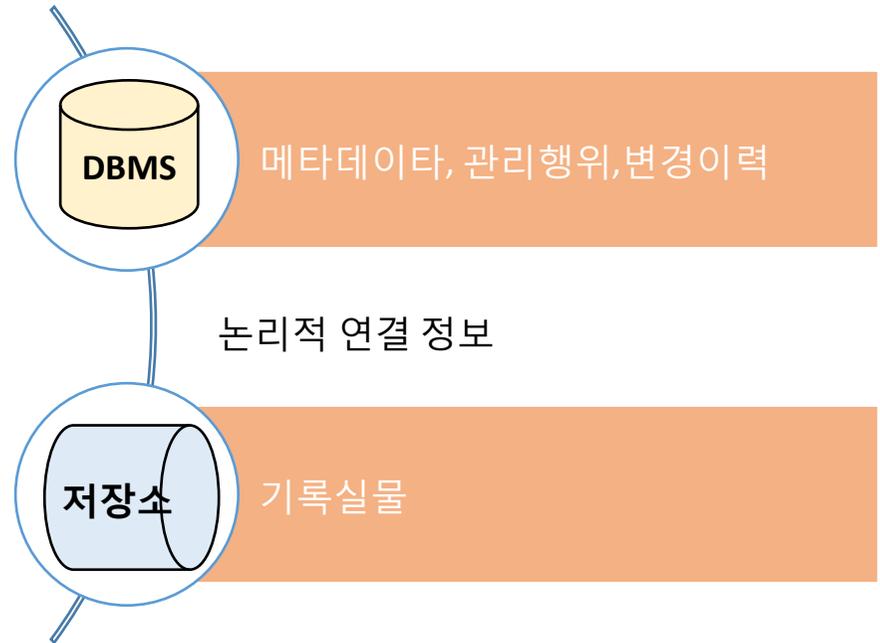
IP	중점 PDI
SIP	생산 및 전송에 대한 설명 정보 ▷ 생산정보(장비,사람,시간,장소...) 전송정보(방식,시간...)
AIP	보존에 대한 설명 정보 ▷ 보존이력, 변환(Migration), 변경이력, 버전,
DIP	해석 및 이해에 대한 설명 정보 ▷ 설명서, 해설서, 뷰어, 포맷정보

장기보존 시스템 예 - 대통령기록관 MAM

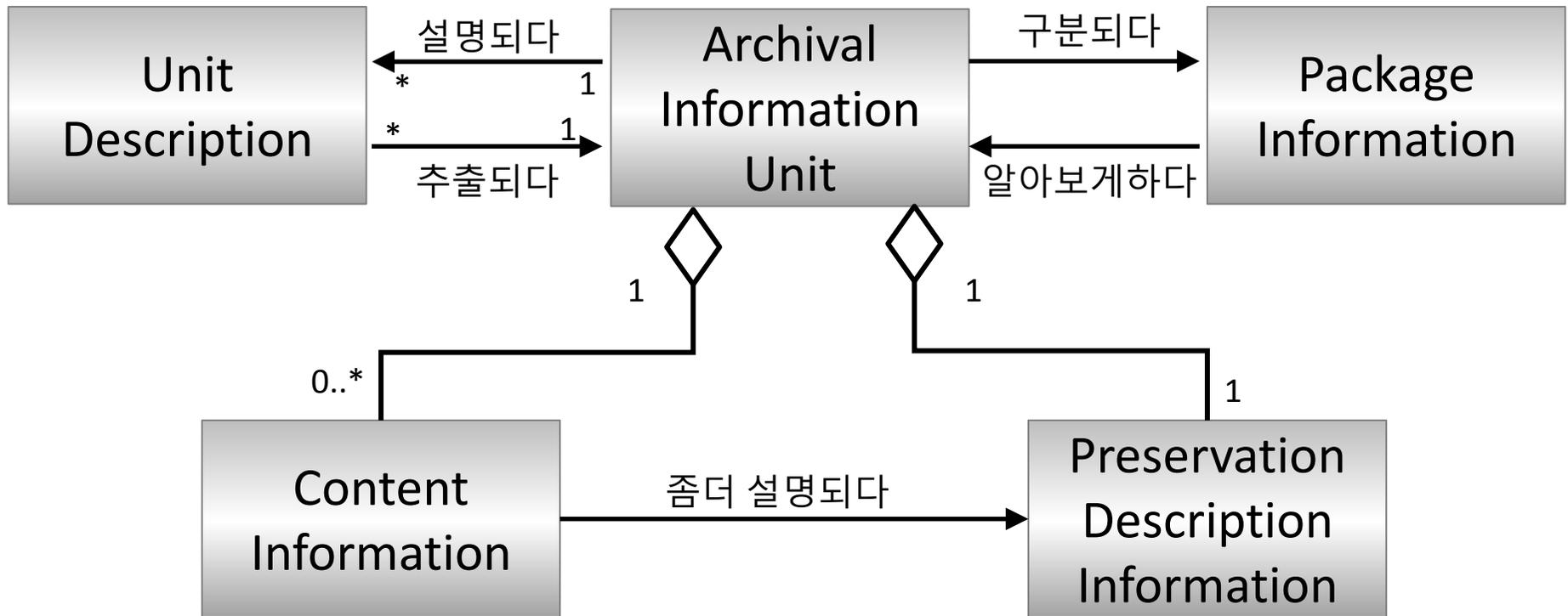
시스템 구성



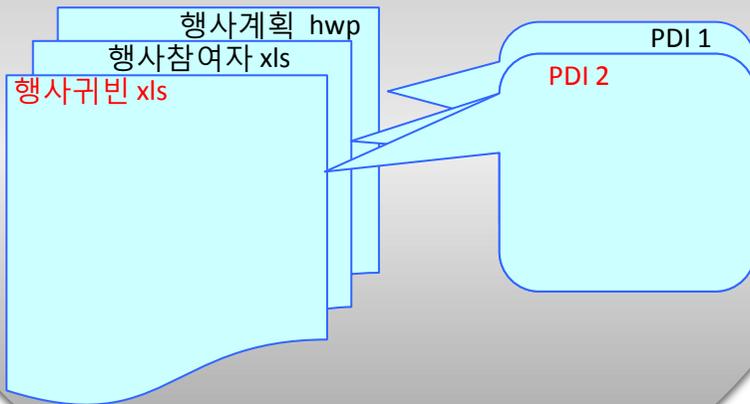
시스템 구성 콘텐츠+메타데이터 관리 장소



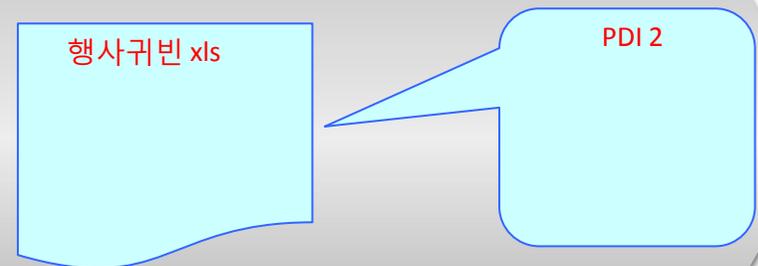
보존정보 최소단위(AIU)



Submission Information Package



Archival Information Unit



보존정보의 최소 단위로 하나의 PDI세트, 단일 내용정보를 갖는다.
서로 다른 PDI정보(문서,수치 등)를 갖는 단위의 AIP

☞ **다양한 관리(처리) 단위 정의의 기반이 됨**

최소단위 정의 예 - MAM의 시청각정보관리

Born-Digital 영상에 대한 최소 관리 단위

WMV
(WMV)

입수본ID
생산자
기록관리번호
입수(Ingest)일
포맷정보
파일위치
해상도
화질
Hash

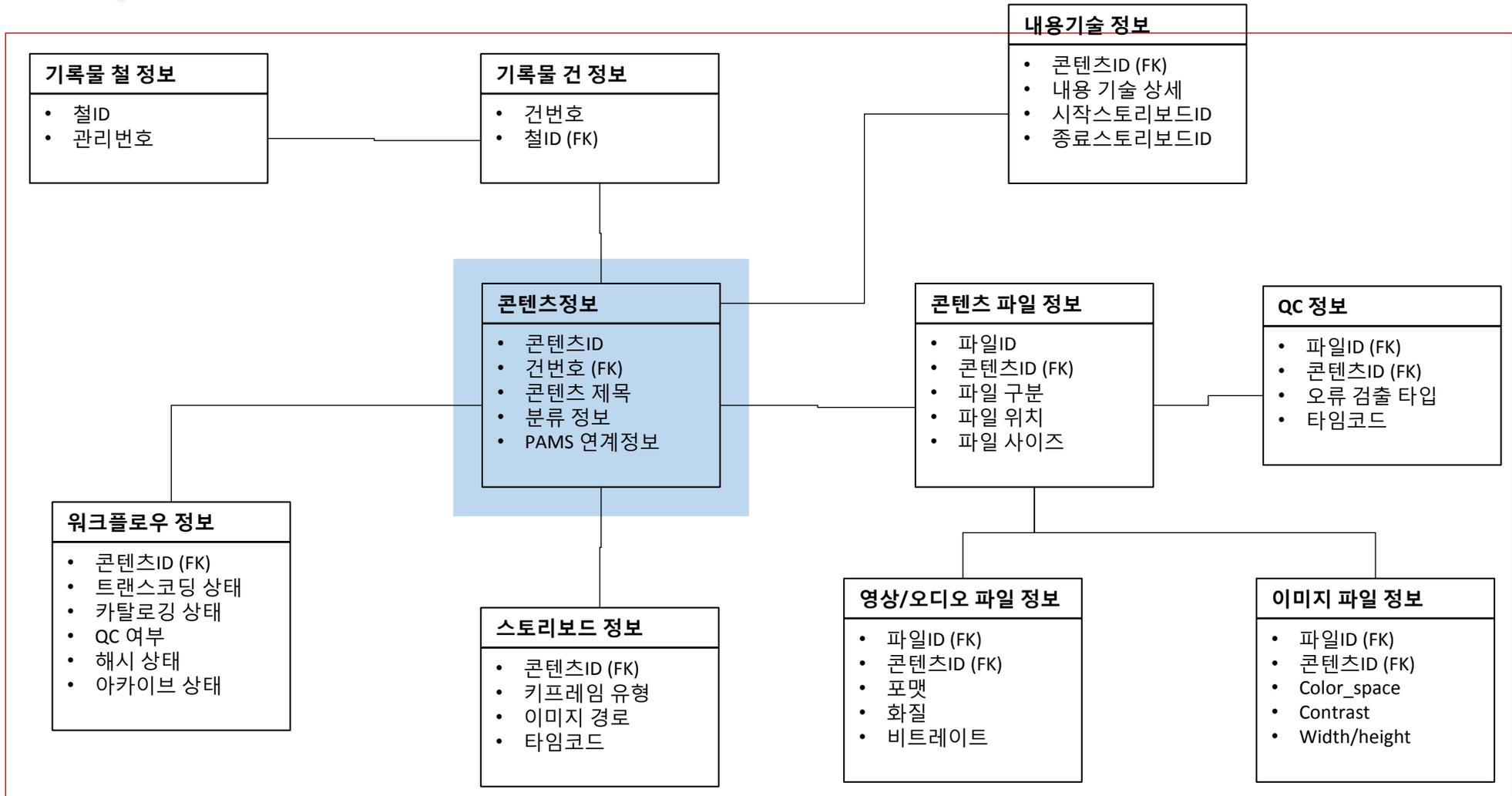
MKV
(FFV1,FLAC)

보존본ID
입수본ID
생성자
포맷정보
파일위치
해상도
화질
Hash

MP4
(H.264)

활용본ID
입수본ID
생성자
포맷정보
파일위치
해상도
장면분석정보
미리보기URL

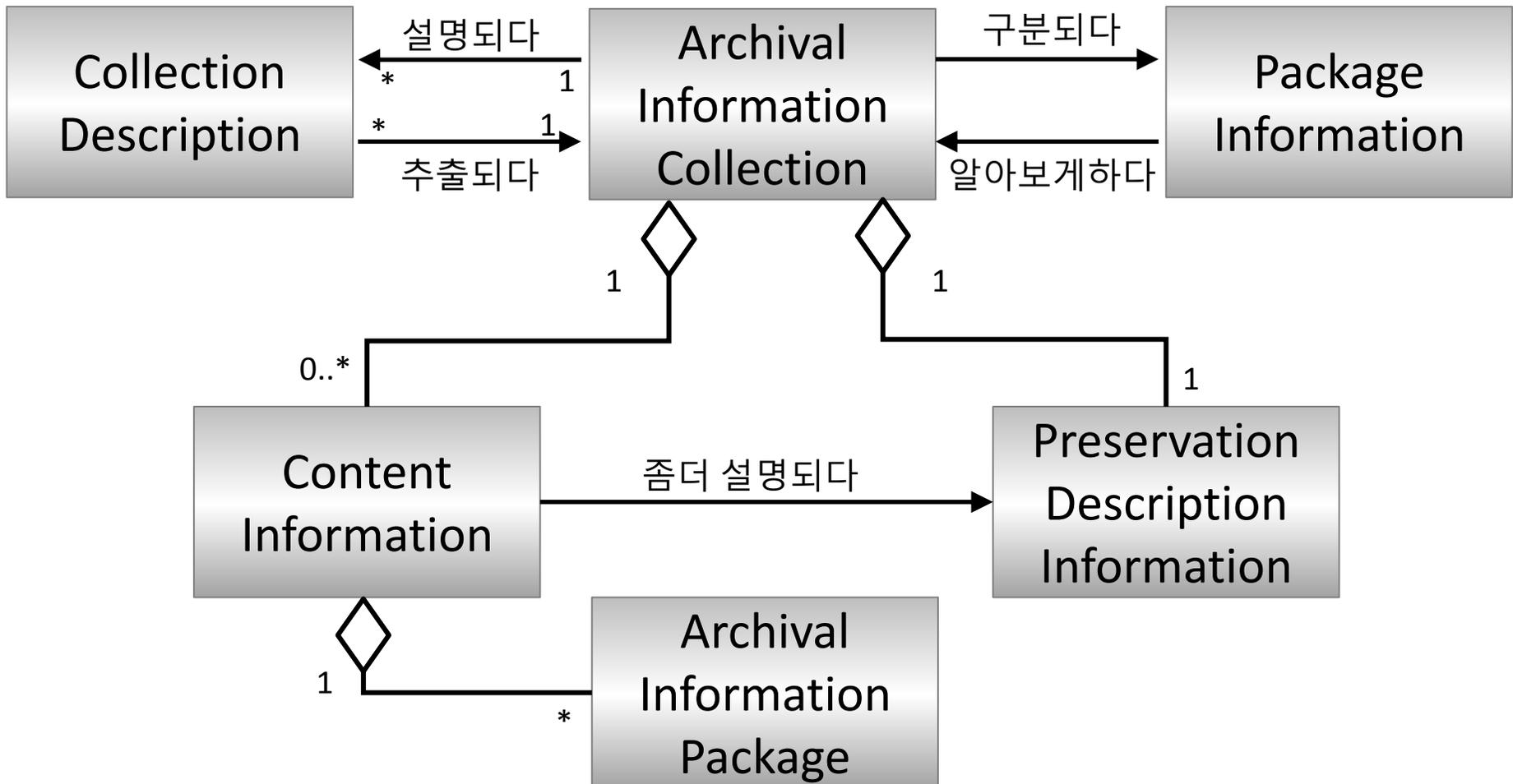
최소단위 DBMS 테이블 구성 예

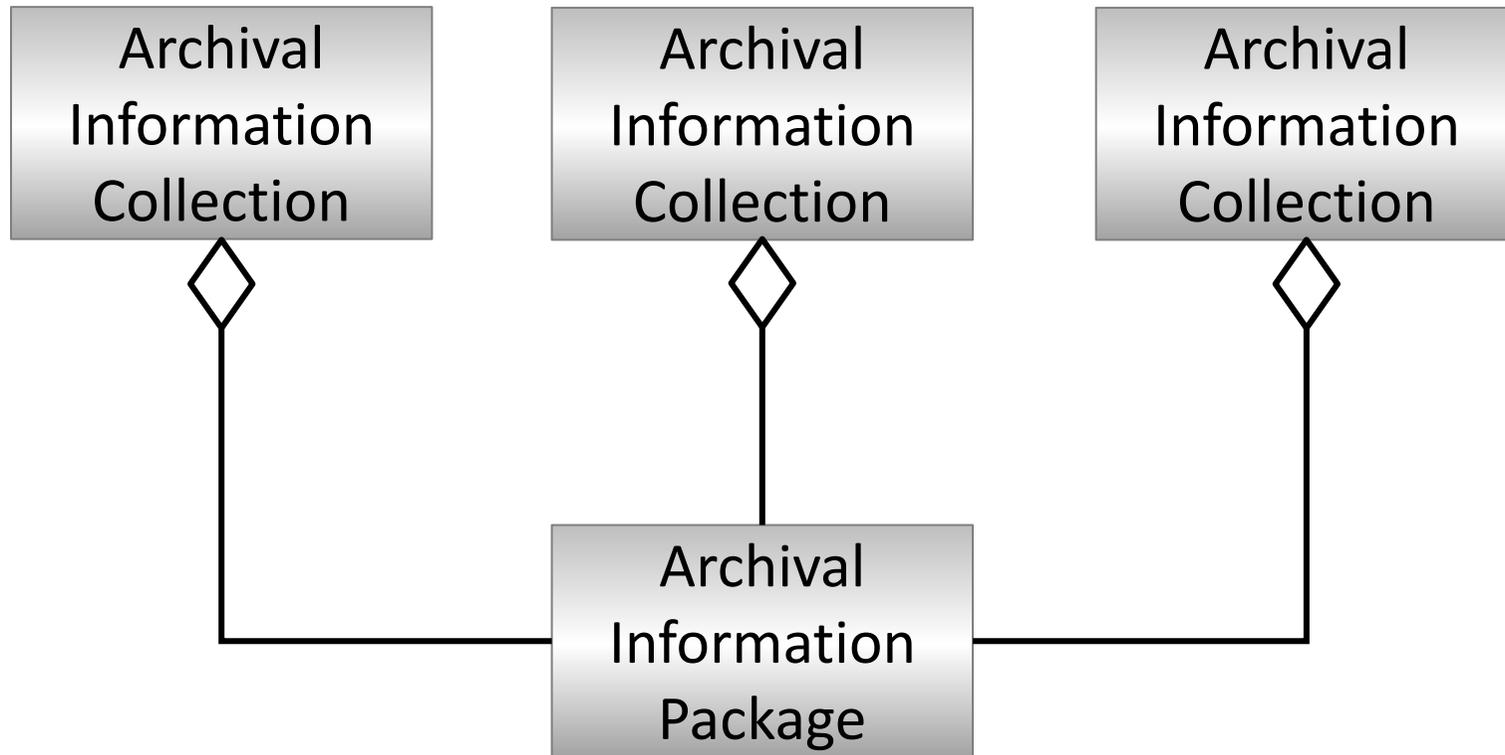


기록물 처리(관리) 단위 모델



보존정보 컬렉션(AIC)





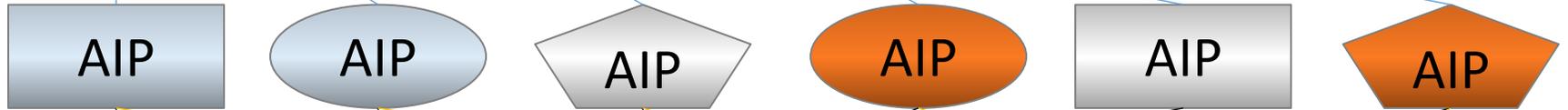
하나의 AIP는 다수의 AICs에 속할 수 있다.

☞ 정보에 대한 **다중 분류의 기초**이며, **활용성·관리편의성 증대**, 새로운 연관기술의 발견을 가져온다.

색 분류



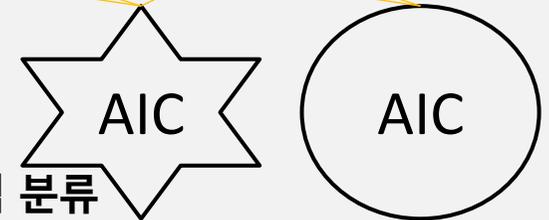
AIP는 요소, 속성등에 따라 다양한 AIC 구성



형태 분류



꼭지점 분류



다중분류 설계 - 대통령기록관 MAM

다중분류에 의한 기록물 관리



하나의 기록물은 1개 이상의 상위 분류에 속할 수 있음

분류그룹 관리 화면

대통령기록관 | 기록물 검색 | 기록물 관리 | 포탈관리 | 통계 | 마이페이지

그룹 관리 | 분류 현황

기록물 분류 관리

등록일 ▼ [] ~ []

대표이미지	기록물 ID [등록상태]	상태	진행	정
<input checked="" type="checkbox"/>	상대	공개	EN	
<input checked="" type="checkbox"/>	상대	공개	EN	
<input type="checkbox"/>	상대	공개	EN	
<input type="checkbox"/>	상대	공개	EN	

중략

다중그룹 연결

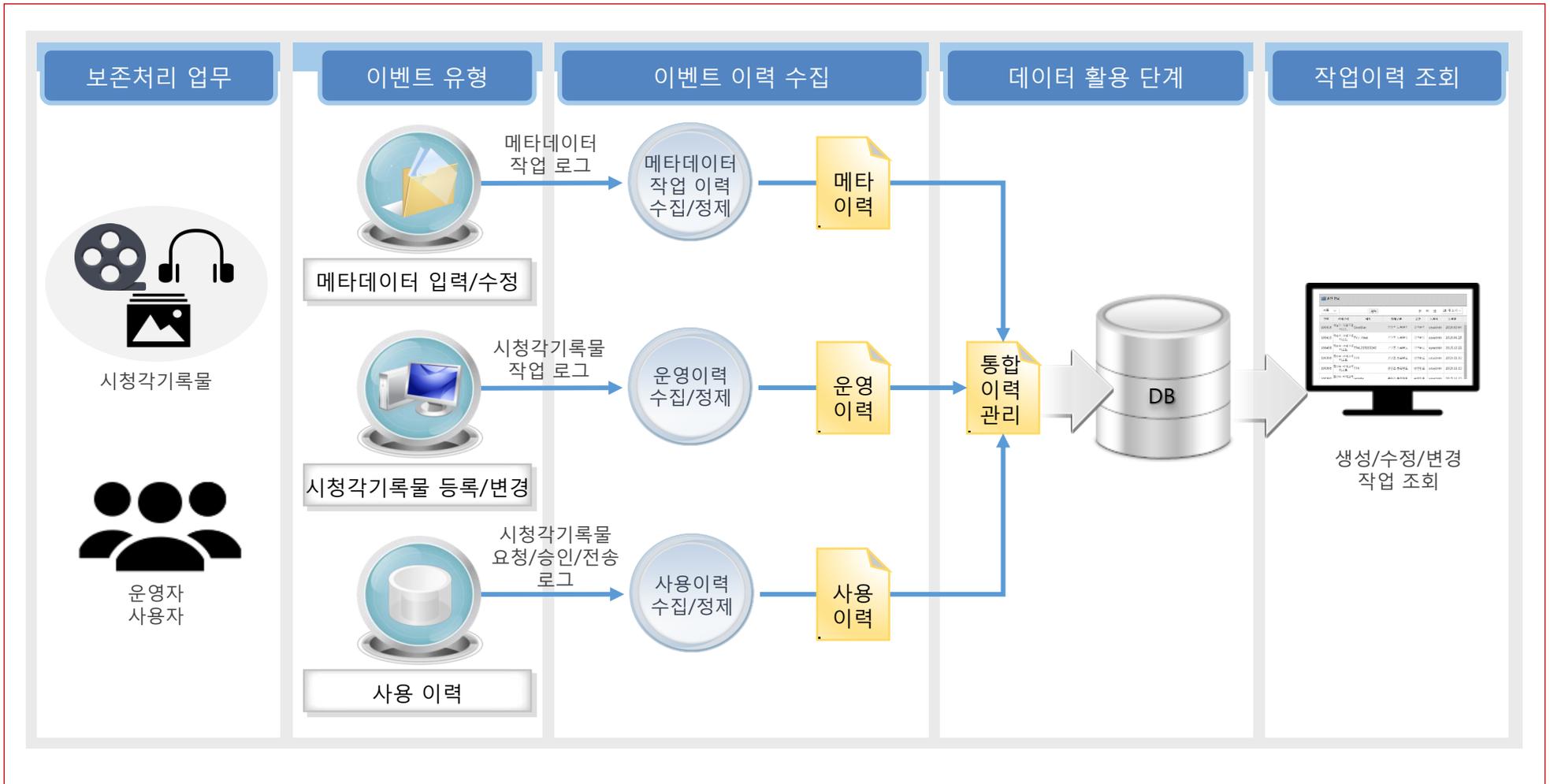
[다중그룹]

- 18대 대통령
 - > 18대-취임식1
 - > 18대-취임식2
- 취임식
 - > 18대-취임식1
 - > 18대-취임식2
 - > 17대-취임식
 - > 16대-취임식

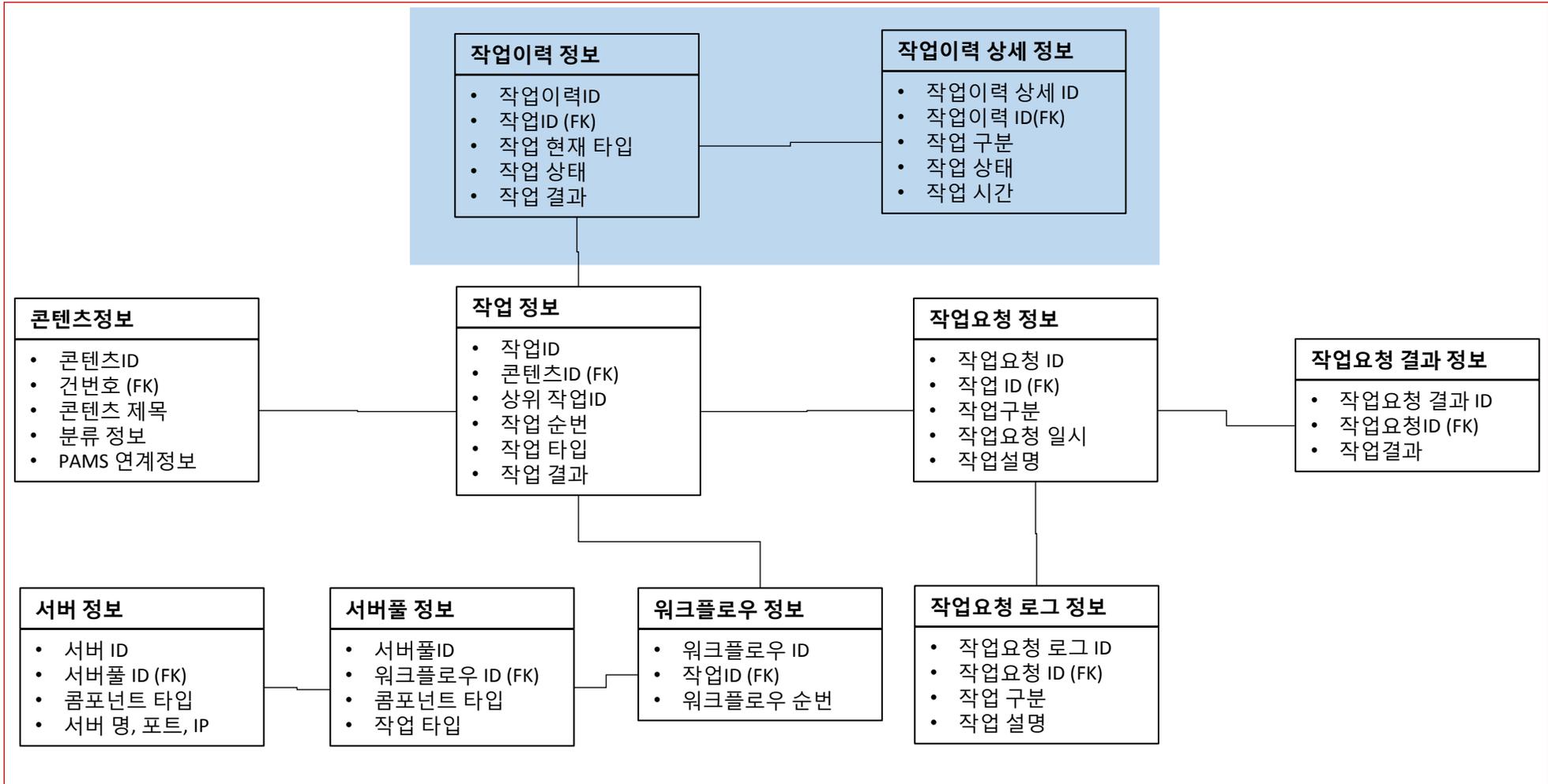
저장 | 닫기

그룹 연결

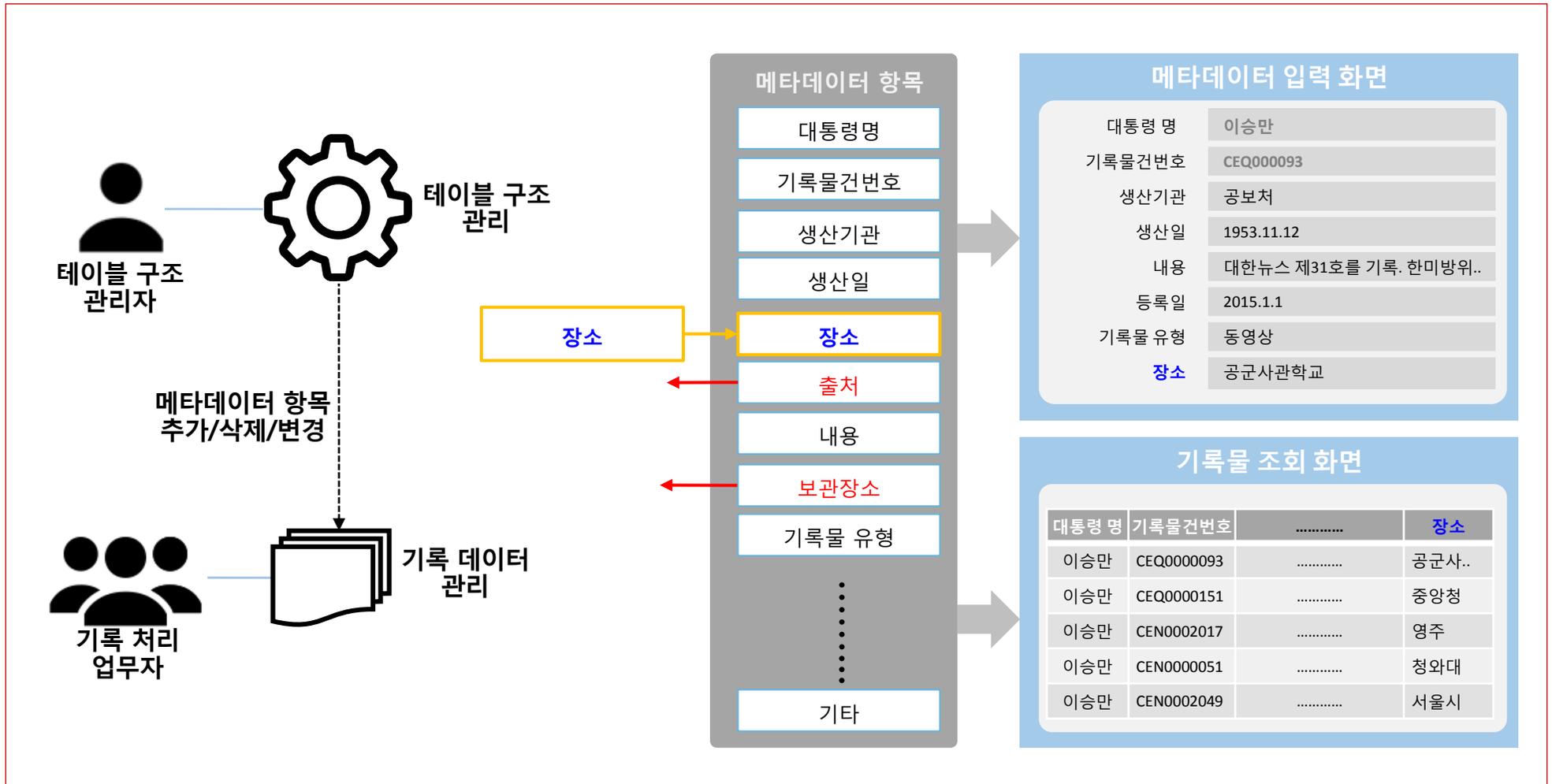
이력관리 설계 예 - 대통령기록관 MAM



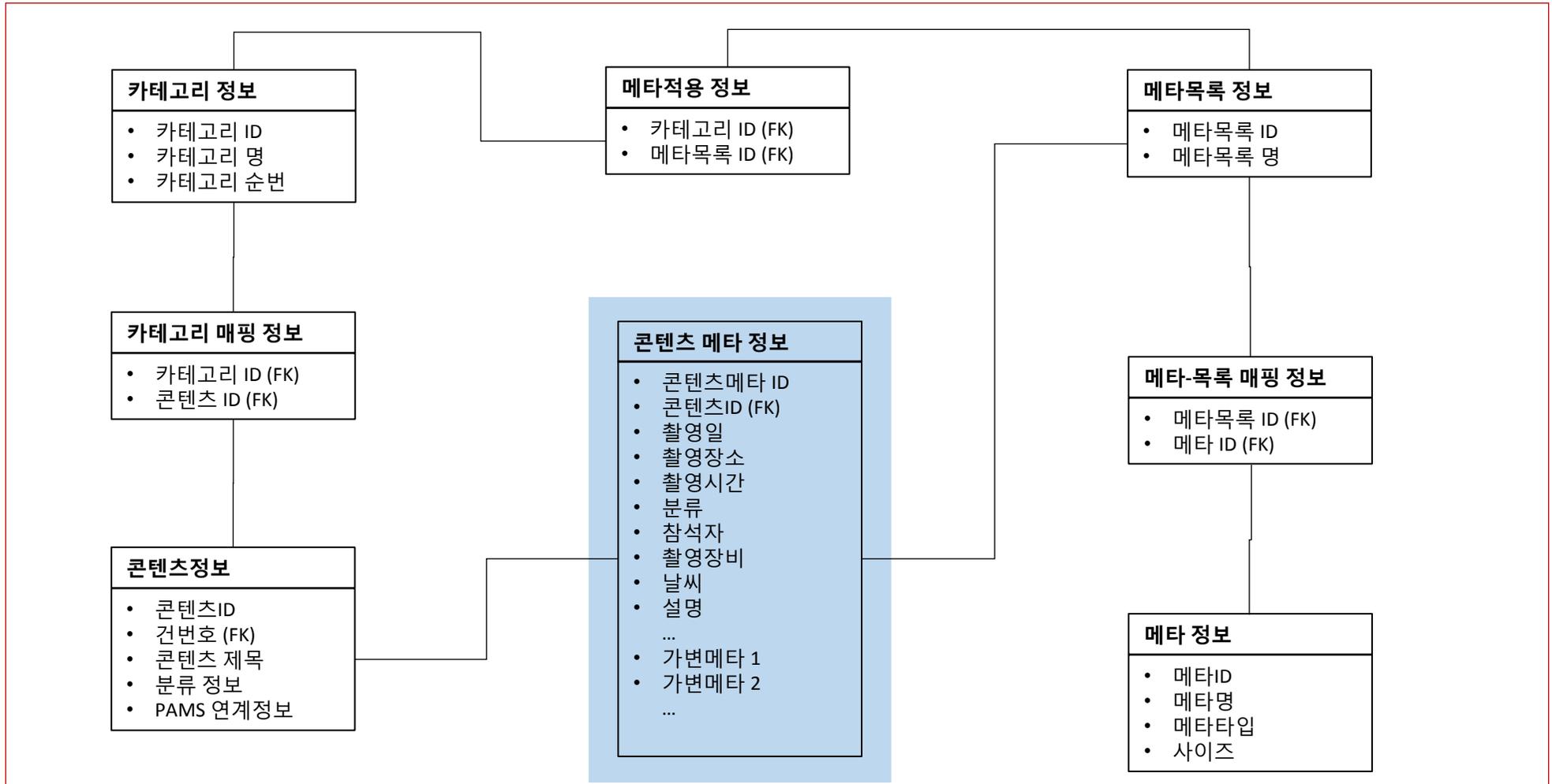
이력관리를 위한 DBMS 테이블구조



유연한 테이블 구조 설계 예 - 가변메타



가변메타 운용을 위한 DBMS 테이블 구조



무결성과 일관성을 유지하여 사용자(운영자)가 이해하기 쉽고 접근하기 편하게함.

종 류	설 명
표준용어정의서	엔터티, 테이블, 컬럼 등의 이름 명명규칙, 약식규칙
도메인정의서	데이터의 범위, 형식, 크기
논리데이터모델 다이어그램	논리데이터 모델을 도식화하여 표현한 것(ERD)
엔터티(개체)정의서	하나이상의 속성으로 이루어진 정보 단위, ROW와 유사
애트리뷰트(속성)정의서	개체를 설명해주는 속성, Column와 유사
데이터베이스정의서	전체적인 정의서, 논리·물리 DB명, 구축일자, DBMS정보 등.
물리데이터모델 다이어그램	DBMS의 테이블 모델을 도식화하여 표현한것
테이블정의서	테이블에 대한 설명, 테이블명, 관련엔터티, 볼륨
컬럼정의서	테이블의 컬럼에 대한 설명, 컬럼명, 데이터타입·길이·형식

공공기관의 데이터베이스 표준화 지침 [행정안전부고시 제2019-20호, 2019. 3. 20.]

데이터베이스 설계 문서 예 - 용어사전

단어사전 (표준용어)	
논리명	물리명
콘텐츠	CONTENT
대표	MAIN
ID	ID
이미지	IMG
제목	TITLE
내용	DESC
순번	SEQ
구분	GUBN
원본	ORG
파일	FILE
경로	PATH
명	NM
여부	TF
요청	RQST
상태	STATUS
디스크	DISK
DTL	DTL

용어사전						
논리명	물리명	도메인그룹	도메인 유형	도메인명	논리데이터타입	길이
등록시스템구분	RGST_SYTM_GUBN	코드명	일반	코드명VC20	VARCHAR	20
대표이미지	MAIN_IMG	VC일반	일반	일반VC200	VARCHAR	200
제목	TITLE	제목	일반	제목VC500	VARCHAR	500
콘텐츠내용	CONTENT_DESC	내용	일반	내용VC4000	VARCHAR	4000
원본파일경로	ORG_FILE_PATH	VC일반	일반	일반VC500	VARCHAR	500
원본파일명	ORG_FILE_NM	명	일반	명VC200	VARCHAR	200
원본파일사이즈	ORG_FILE_SIZE	크기	일반	파일사이즈DEC20	NUMERIC	20
디스크존재여부	DISK_EXIST_TF	여부	일반	여부CH1	CHAR	1
업로드상태	UPLD_STATUS	상태	코드	상태VC6	VARCHAR	6
삭제요청상태	DEL_RQST_STATUS	상태	코드	상태VC6	VARCHAR	6

데이터베이스 설계 문서 예 - 도메인정의서

제6차 전자기록관리 연구포럼

도메인그룹(분류)	설명	도메인명	데이터타입	데이터길이
순번	발생순서대로 부여된 일련번호	순번DEC8	NUMERIC	8
		순번DEC10	NUMERIC	10
시간	일자와 시간으로 표시, YYYY-MM-DD HH24:MI:SS 로 표기	일시TM	DATE	0
	하루를 의미, YYYYMMDD 로 표기	일VC8	VARCHAR	8
ID	특정한 목적의 식별자로 사용하기 위해 설계된 문자의 나열	컨텐츠ID	VARCHAR	20
		오브젝트ID	VARCHAR	20
사이즈	특정한 사물의 길이, 크기 등을 표현한 값	사이즈DEC20	NUMERIC	20
		사이즈DEC10	NUMERIC	10
수량	셀수 있는 사물을 세어서 나타낸 값	수량DEC5	NUMERIC	5
		수량DEC10	NUMERIC	10
코드	공통의 성질, 특징이 있는 것들을 묶은 하나의 틀 및 그 틀에 속하는 것	코드VC20	VARCHAR	20
		코드VC4	VARCHAR	4
여부	둘로 구분되는 속성을 가지는 사물 또는 개념을 구분하는 값	여부CH1	CHAR	1
		여부VC4	VARCHAR	4
제목	글의 내용을 보이거나 대표하는 이름	제목VC1000	VARCHAR	1000
		제목VC100	VARCHAR	100
		제목VC500	VARCHAR	500
내용	글의 내용	내용VC1000	VARCHAR	1000
		내용CL	CLOB	0
		내용NVC500	NVARCHAR	500

데이터베이스 설계 문서 예 - 컬럼정의서

테이블ID	TB_CONTENT	테이블명		콘텐츠정보						
테이블정의	시청각기록 파일단위의 콘텐츠정보 관리									
No	컬럼명	한글명	Type	Len	Null	PK	FK	설명		
1	CONTENT_SEQ	콘텐츠순번	NUMBER	10		Y		시스템에서 인식하는 ID로 10자리 숫자		
2	CONTENT_ID	콘텐츠ID	VARCHAR2	20		Y		사용자가 인지하는 시청각기록 ID, 비디오 V, 오디오 A, 이미지 I로 시작하며, 연월일+일련번호로 구성		
9	CONTENT_DESC	콘텐츠내용	VARCHAR2	4000	Y			시청각기록물 전체에 대한 설명, 각 장면 또는 구간에 대한 설명은 내용기술화면의 구간정보에 기술		
10	DOC_NO	건번호	CHAR	16	Y			PAMS에서 받아오는 건번호, 문자와 숫자 결합된 12자리.		
22	DISPRDION_POSITION	소산위치	VARCHAR2	100	Y			LTO 테이프의 소산 위치, 지역 이름 입력, 예, 성남, 세종, 대전, 부산		
23	TAPE_NO	테이프번호	VARCHAR2	20	Y			DTL장치가 인식하는 Main 저장영역 그룹의 테이프 번호, Main + 일련번호 형식으로 되어 있음.		
37	MAM_LINK_RGST_NM	MAM연계등록자명	VARCHAR2	40	Y			MAM과 PAMS연계 목록(건과 시청각파일명 연결)을 등록한 PAMS 사용자 이름. 예) 홍길동		
38	INGEST_RGST_USERID	인제스트등록자ID	VARCHAR2	20	Y			MAM에 시청각파일을 등록한 사용자의 ID, 예, hyeok		
39	INGEST_RGST_NM	인제스트등록자명	VARCHAR2	20	Y			MAM에 시청각파일을 등록한 사용자의 이름, 예, 신동혁		

대통령기록관 시청각기록물관리시스템

제6차 전자기록관리 연구포럼

MAM의 기능구성별 메타데이터 및 주요 모듈

기능	입수 Digitize&Ingest	데이터관리 Data Management	스토리지 Archival Storage	보존 Preservation	접근 Access	운영 Administration
메타 데이터 구성	제목, 주제, 언어 등	포맷정보	저장정보(위치)	보관정책정보	사용자정보	관리자정보
	파일정보	상태이력정보 (포맷/매체 변경)	인증정보 (ID)	포맷별 저장 환경정보	권한정보	배포정보
		인덱싱				워크플로우
		분류정보(카탈로그)				시스템정보
주요 모듈 & CORE 서비스	인코더	카탈로거	아카이브 관리		스트리밍	전송관리
	인제스트 관리	품질보증			권한관리	승인관리
					콘텐츠 조회	
	워크플로우 & 콘텐츠 관리					