

행정정보데이터세트 기록관리 방안

2019.5.



행정안전부

국가기록원
전자기록관리과



행정안전부

국가기록원

2019년 제4차 기록관리 연구세미나

목 차

- I 추진배경 및 경과**
- II 데이터세트 분석**
- III 관리원칙 및 관리체계**
- IV 기록관리 프로세스**
- V 데이터세트 이관·보존 전략**

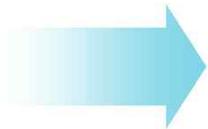
※ 참고 데이터세트 관리항목표, 그간 고려 사항 등

I 추진배경

정의	행정정보시스템에서 생산된 데이터
현황	공공기관 기준 : 16,400여개(중앙행정 기관 : 2,200여개)
운영	업무 수행을 목적으로 구축한 부서에서 운영
기록관리	법령상 기록관리대상



구체적 기록관리 방법이 명시되지 않은 경우 기존 방식*(지침) 준용
*기존 방식 : 전자문서 중심의 정형화된 객체 단위 관리(온나라-RMS)



**특정 정보시스템에 구체화
다양한 유형의 데이터세트에 일괄 적용 불가
범용적으로 적용하여, 실행이 가능한 방안 필요**



추진경과

'17.4~11.	차세대 전자기록관리 모델 재설계 R&D 사업 추진
'17.6~7.	유형별 대표 행정정보시스템 6개 기관 방문 현장조사(11회)
'17.2~10.	행정정보데이터세트 기록관리지침 작업반 논의(3회)
'17.11.	2017년 제2회 전자기록관리전문위원회 안건 상정·논의
'17.12.	행정정보 데이터세트 기록관리체계(안) 마련
'18.2~10.	행정정보데이터세트 기록관리 기관 설명회(1차, 2차, 3차)
'18.12.	행정정보 데이터세트 기록관리 매뉴얼 개발(안) (BPR/ISP)
'19.1~3.	행정정보 데이터세트 기록관리 추진방향 현안연구 워크숍(5회)

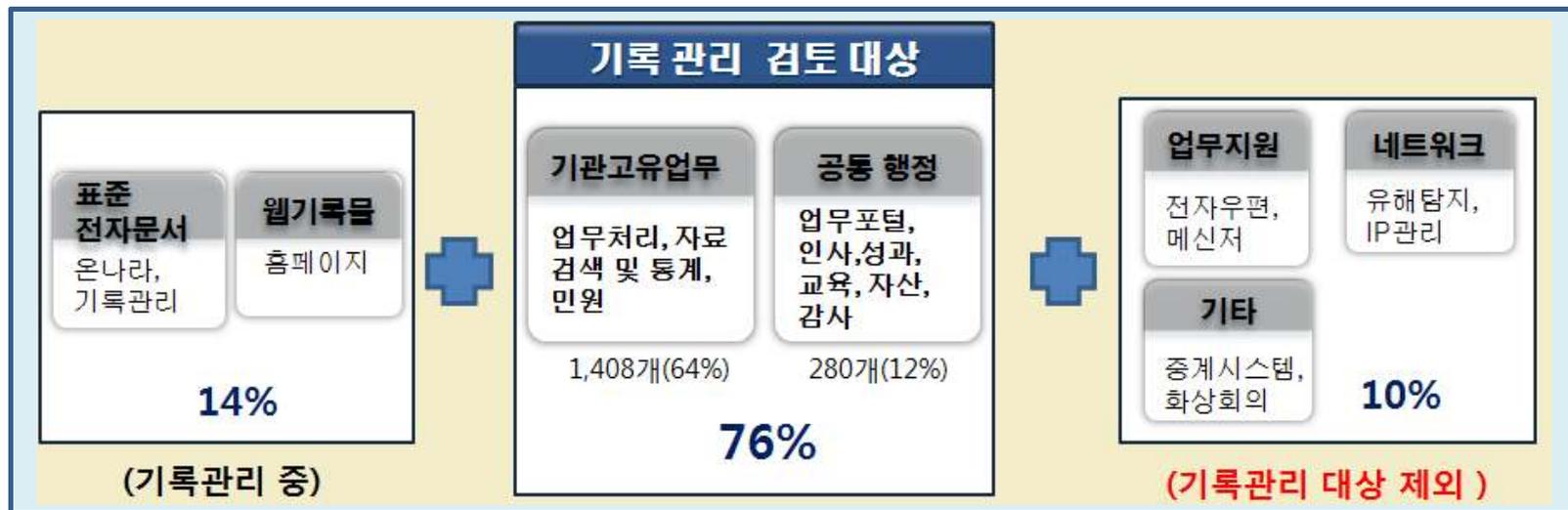


데이터세트 기초분석1(구분)

□ 데이터세트 기초 분석(중앙부처 2,200여개)

기관 고유업무(64%)	국세통합시스템, 건축행정, 관세청FTA, 보호관찰, 민원24 등
공통 행정업무(12%)	내부행정포털, 인사관리, 자산관리, 그룹웨어 등
단순 업무지원(4%)	내부메일, 문자전송, 웹하드 등
보안·네트워크(3%)	방화벽, 유해탐지, IP관리시스템 등
기타(3%)	자료전송, 화상회의 시스템 등

□ 기록관리 대상 식별



II 데이터세트 기초분석2(유형)

□ 데이터세트 유형

구분	개념도	설명 및 예시
단일기관 단일시스템		<ul style="list-style-type: none"> • 업무 또는 서비스를 위해 단독 시스템으로 구성되어 있으며, 단일기관에서 운영, 관리 • 타기관 및 타시스템과 데이터 연계가 없는 단순 구조 • 단순한 구조의 유형으로 극소수 <p>EX) KAIST 전자연구노트 시스템</p>
단일기관 연계시스템		<ul style="list-style-type: none"> • 한 개 기관에서 업무별 다수 시스템이 통합DB를 통해 상호간 데이터를 공유·참조·생성 • 대형 시스템의 경우 이와같은 구성으로 구축된 사례가 많음 <p>EX) 특허청 특허넷 시스템</p>
중앙집중형		<ul style="list-style-type: none"> • 동일 부처 내 다수의 기관이 중앙시스템에 접속하여 데이터를 생성·처리 • 중앙시스템에 다수 기관에서 생성된 데이터는 업무처리를 위해 각 기관에서 참조 <p>EX) 미래부 주파수분석 시스템</p>

구분	개념도	설명 및 예시
다수기관 데이터 연계형		<ul style="list-style-type: none"> • 하나의 기관 시스템에서 생산된 데이터를 다수 기관에서 연계 또는 참조하여 업무 처리하는 유형 • 각 부처간 업무처리 효율성 등을 위해 이와 같이 구축된 유형의 시스템 사례가 매우 많음 <p>EX) 식약청식약약품 정보시스템</p>
중앙-지방 연계형		<ul style="list-style-type: none"> • 국가위임사무의 경우, 업무처리를 위한 시스템은 각 지자체에서 운영·관리되며, 지자체별 생성된 데이터는 중앙에서 수집하여 통계·분석 등 업무 참조 • 중앙에서는 각 지자체 데이터를 참조만 하고 중앙시스템에 데이터를 저장하지 않음. * 업무참고를 위해 지자체 데이터 열람 <p>EX) 국토부 국토정보시스템</p>
중앙-지자체 간 복합 연계형		<ul style="list-style-type: none"> • 시도·시군구별 데이터를 생성하며, 지자체간 중계서버를 통해 상호 데이터를 공유 및 저장 • 중앙 중계서버를 통해 시도·시군구에서 생성된 데이터를 중앙 부처에서 활용을 위해 연계 • 중계서버에는 데이터 저장 않음 <p>EX) 시도행정정보시스템, 서울행정정보시스템</p>



데이터세트 기초분석3 (현장조사개요)

□ 유형별 현장조사

- 조사기간 : '17.6 ~ 8월(2개월, 11회 현장방문)
- 대상시스템(기관) : 6개
 - (다수기관 데이터연계) 산림자원통합관리시스템(산림청), 국민신문고시스템(권익위)
 - (단일기관 단일) 전자연구노트시스템(한국과학기술원)
 - (단일기관 연계) 특허넷시스템(특허청)
 - (중앙-지방 연계형) 국토정보시스템(국토부)
 - (중앙집중형) 화학물질종합정보시스템(환경부)
- 조사방법 : 담당자 인터뷰 및 조사표 작성, 시스템 개발 산출물 분석 등
- 조사내용 : 시스템 운영·관리 주체, 데이터베이스(DB) 구조, 비정형데이터 종류, 데이터 생산·연계 현황, 업무기능, 사용자(이용자) 현황, 기록관리 대상·주기 등



데이터세트 기초분석4(조사절차)

□ 현장조사 수행단계 및 수행내용

① 조사대상 선정	• 사전 수요조사를 통해 기록관리 시급성, 중요도, 우선순위 등을 고려하여 대상시스템 선정
② 사전자료 입수	• 행정시스템 구축 관련 사업계획, 개발산출물 등 기술자료 사전 입수하여 예비 분석
③ 현장조사(1차)	• 사전 세부조사계획에 따라 기록관리 및 업무, 시스템 운영, 관리 담당자 인터뷰를 통한 사전조사 내용 확인·보완
④ 항목표작성(1차)	• 사전조사 및 1차 기관 현장조사 결과를 취합, 1차 항목표 작성
⑤ 현장조사(2차)	• 정보시스템 실사, 개발산출물 분석을 통해 DB구조, 시스템 구조, 데이터 종류 등 구체적인 데이터 유형 조사
⑥ 항목표작성(2차)	• 시스템 현황, 업무기능, DB구조, 기록관리 요건 등 데이터 세트 기록관리를 위한 항목표 세부내용 작성
⑦ 검토 및 확정	• 작성된 항목표 기록관리담당자 검토 및 협의체 최종 확정
* 참조모델 구축	• 항목표 + 기록관리 절차 시범 적용(처분 스크립트 등) 재정사업으로 추진

**데이터세트
관리항목표
작성**



데이터세트 기초분석6(조사결론)

□ 현장조사 결과 시사점

관리단위
식별 필요

- 업무상 여러 기능의 데이터가 동일한 데이터베이스에 저장·운영
- ← 동일한 물리적 데이터베이스(DB)구조를 기반으로 단위기능 결정

동일
단위기능 내
보존기간 상이

- 물리적 DB구조가 동일(단위기능 1개)하나, 유형별 보존기간 상이한 경우 존재
- ← 항목표는 단위기능별 작성, 단위기능 아래 처분단위 및 방법 명시

생산과 관리
불일치

- 데이터 생산기관과 관리기관(시스템 운영기관)이 불일치하여 처분에 어려움
- ← 데이터 관리기관(시스템 운영기관)에 처분권한 부여

보존기간
기산일 정의

- 지속 운영상태로 확정되지 않는 데이터 유형 존재, 보존기간 기산일 산정 어려움
- ← 데이터 생성 다음연도 1월1일을 기산일로 산정, 특수한 경우 협의 후 산정

운영시스템
데이터 폐기

- 운영시스템에서 보존기간이 지난 데이터세트의 폐기(물리적 완전 삭제 문제)
- ← 기록물평가 심의위원회를 거쳐 데이터 삭제(스크립 등), 삭제 관련 증적자료 보관



행정정보 데이터세트 기록관리 원칙

- **범용적인 관리모델 & 효율적 실행 절차**
.....
- **행정정보 데이터세트 생산시스템부터 협업관리 시작** (최적의 대국민 기록서비스)
.....
- **행정정보데이터세트 관리 기준항목(표) 개발·운영** (협업으로 기준 제시 및 관리)
.....
- **국가기록원 기술지원 및 협업** (기술지원센터)
.....
- **데이터세트 보존 및 서비스 체계 마련** (최소화, 국가기록원 이관)
.....



행정정보 데이터세트 관리체계



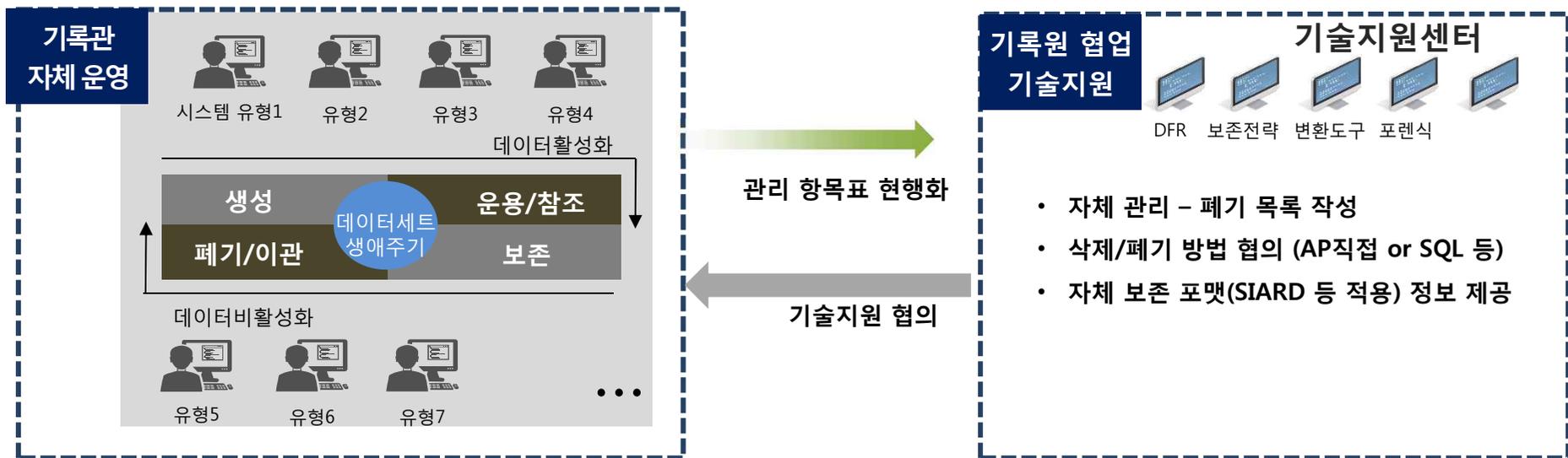


행정정보 데이터세트 관리체계 - 협의체



항목표(항목) 분석-협의

- 기록관, 관리운영부서, 기록원 참여
- 관리 단위(시스템 > 업무 > 기능 > 테이블)
- 관리 방법(자체관리 - 연속체 모델 적용)
- 보존기간 책정 / 기록관리처분(폐기, 보존 등) 방식

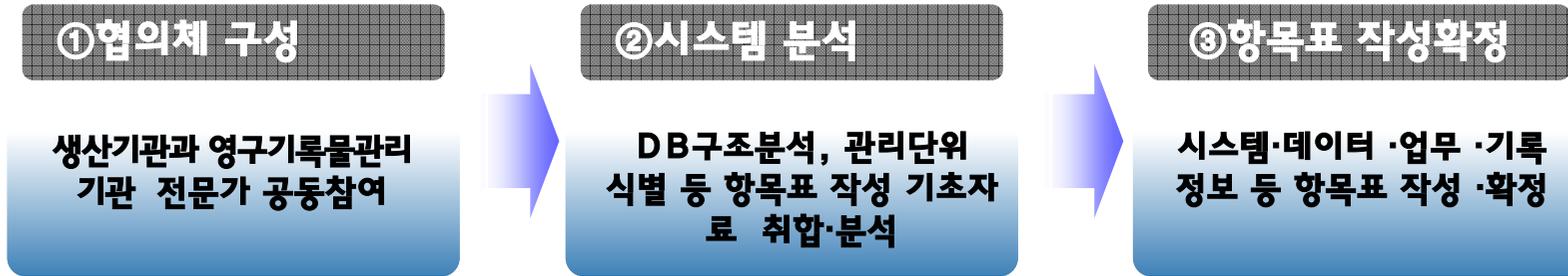


□ 데이터세트 기록관리 절차(전체)



IV 기록관리 프로세스

□ 기록관리절차 - ①준비등록



- 협의체 구성**

 - (생산기관) 기록연구사, 업무담당자, 시스템 담당자, 유지관리·개발업체
 - (영구기록물관리기관) 법·제도, 시스템, 보존 전문가
- 시스템 분석**

 - (사전조사) 홈페이지, 나라장터 등을 통해 시스템 구축·운영 관련 현황 파악
 - (담당자인터뷰) 시스템, 기록관리 등 관련 업무담당자 인터뷰를 통한 조사
 - (현장실사) 시스템 실사, 개발산출물 분석 등 세부항목 확인·분석하여 항목표 작성
- 항목표 작성 확정**

 - (생산기관) 항목표 확인·확정, 식별번호 부여 후 관리시스템 등록
 - (영구기록물관리기관) 항목표 확인/확정

□ 기록관리절차 - ②기관기록관리

항목표 관리



변경이력
관리

- 시스템 변경사항 발생 시 기록관리 영향을 평가하여 항목표에 반영
- (주요 변경사항) SW/HW와 업무기능의 추가·삭제, 데이터(DB) 구조 변경 등

항목표
관리

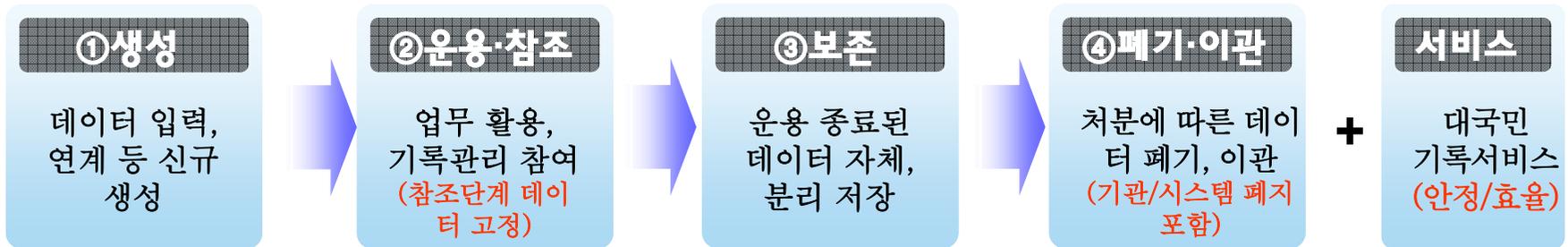
- 데이터세트관리항목표를 최신 정보로 유지관리
- 항목표 등록정보의 변경사유 발생시 항목표 반영 후 영구기록물관리기관 통보

절차는 준수(등록관리, 폐기심의 절차, 폐기목록 작성 등), 단위 행위는 시스템현실 반영 (범위적용, 지연 삭제 등)

IV 기록관리 프로세스

□ 기록관리절차 - ②기관기록관리(계속)

데이터세트 관리



생성

- 시스템에 데이터 직접 입력, 또는 연계에 의해 신규 데이터 생성
- 생성과 동시에 관리단위별 항목표에 따라 자동 분류·편철된 것으로 간주

운용·참조

- 생성된 데이터에 추가, 수정, 조회 등 업무 수행, 변동사항 관리
- 더 이상 변경사항이 없는 경우, 해당 데이터세트 고정 방안 권장 (최소한 log)

보존

- 업무에 운용 참조되지 않는 데이터세트 자체 또는 분리(시스템, 매체 등) 저장
- 진본성, 무결성 및 이용가능성(재현성)을 유지하는 상태 권장 / (ex. 뷰어+고정DB)

폐기·이관

- 시스템 폐지, 업무 종결, 보존기간 등에 따른 폐기 수행(기록물평가심의회 심의)
- 이관 대상의 데이터세트는 사전에 상호 협의하여 방식, 시기, 범위 등 정하고 실시

□ 기록관리절차 - ③영구기록관리(계속)

기록관리 모니터링



모니터링

- 생산기관 데이터세트 관리활동 전반에 대한 모니터링 수행
- 점검 및 확인, 피드백을 통한 생산기관 관리활동 지원

관리시스템 운영

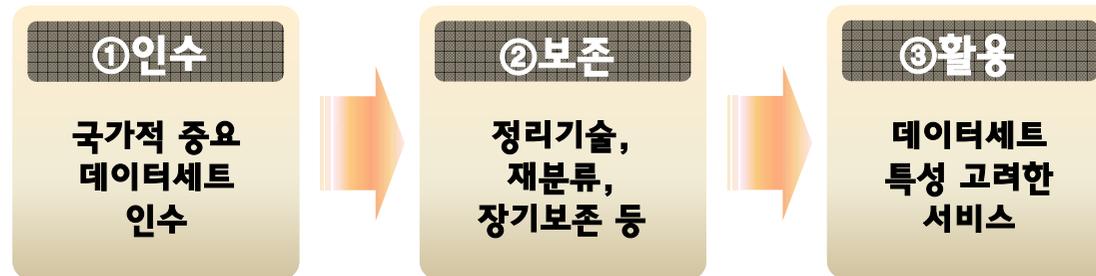
- 항목포에 대한 입력, 확인, 변경, 접근제어 등의 기능 관리
- 생산기관의 법·제도, 업무, 시스템 변경에 따른 항목포 현행화 여부 분석 등

절차는 준수(등록관리, 폐기심의 절차, 폐기목록 작성 등), 단위 행위는 시스템 현실 반영 (범위적용, 지연 삭제 등)

IV 기록관리 프로세스

□ 기록관리절차 - ③영구기록관리

보존·활용



인수

- 국가적 중요 데이터, 폐지기관, 기관요청 등에 의한 데이터세트 인수
- 이관대상 확정 후 규모·규격·방법·시기 등 협의하여 인수(마이그레이션, 에뮬레이션 등)

보존

- 인수한 데이터세트 정리기술·재분류·장기보존 등 업무처리
- 업무처리는 기존 처리절차를 준용하되 관련부서와 충분히 검토하여 마련

활용

- 생산기관의 법령, 정책, 업무 및 데이터 구조 등을 고려하여 서비스 개발(예정)

관리시스템 = 관리항목표(ERD, 설계서 등) 등록관리 + 데이터세트 실체 관리(저장소 개념)

데이터세트 실체 = 이관파일(SIARD, DB_dump 등) + 직접 이관 시스템 목록 + ...



데이터세트 이관 및 보존 전략(구분)

□ 마이그레이션



□ 에뮬레이션





데이터세트 이관 및 보존 전략(SIARD)

SIARD : Software Independent Archiving of Relational Databases

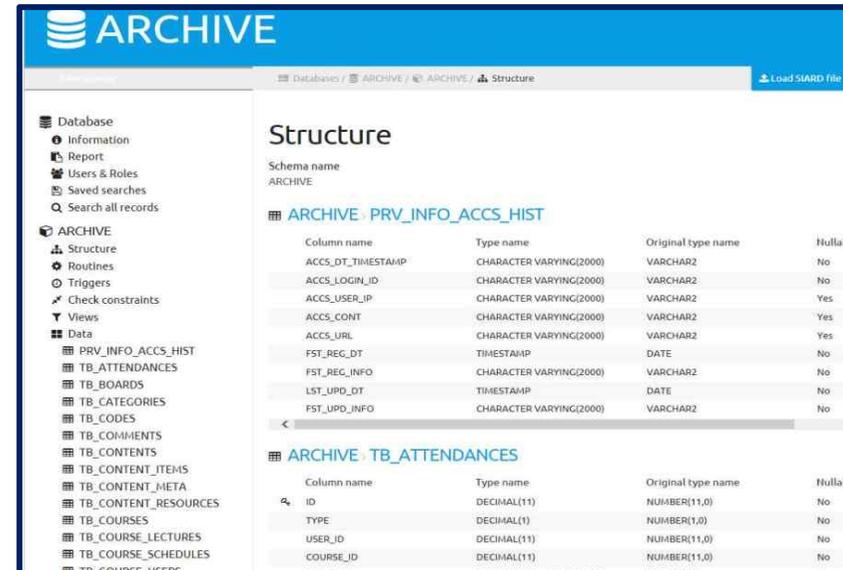
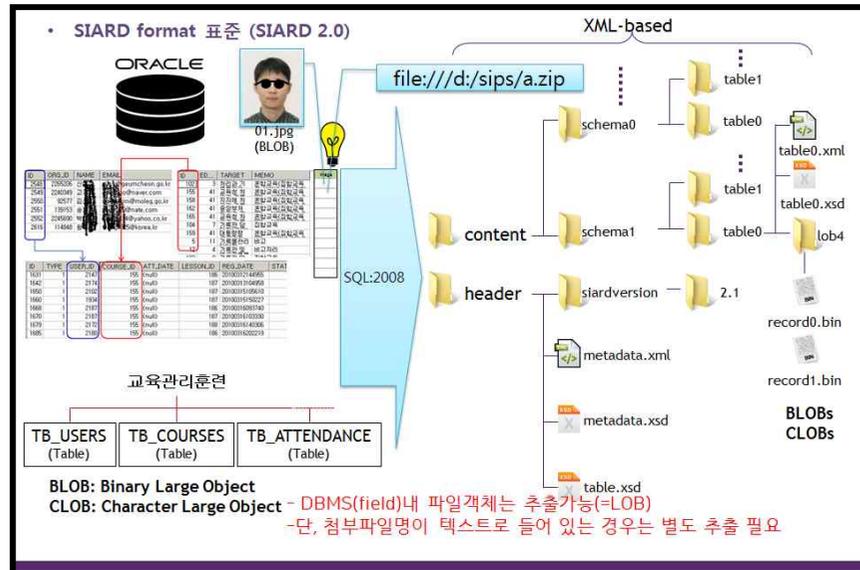
스위스 아카이브 2008년 개발(Enter AG), EU Planets 프로젝트 공식 포맷,
ISO 표준 유니코드, XML, SQL/인터넷표준 URI/산업표준 ZIP 기반으로 개발

* 국제적으로 인정받는 표준들을 기반으로 해서 장기보존과 액세스(접근열람)을 보장
독점 'dump' 포맷에 독립적으로 관계형 데이터베이스 추출 + 데이터세트간 콘텐츠 활용

RODA프로젝트(E-ARK 주관, KEEP Sol.(c) 개발) → 범용성, 사용가능성

SIARD 구성 : SIARD 포맷(75장), DB 추출 toolkit, SIARD파일 뷰어, SIARD Suite(관리용 tool)

SIARD 리뷰(SIARD toolkit으로 '기록교육훈련시스템' 90여개 DB테이블 추출, 뷰어로 재현)



→ 이관, 보존에 활용가능, 보완(추가개발)은 필요





□ '19년 이후 추진일정

'19.3.~	행정정보데이터세트 기록관리 적용 의견 수렴(매뉴얼 검토)
'19.3~4.	시범사업 대상 수요조사 및 사업계획 수립(협업팀 구성)
'19.4.~6.	협업팀 운영(14개 시스템) 및 중기 예산 확보(100개, 8.7억)
'19.7.~12.	행정정보데이터세트 기록관리 실행 및 시범 적용(14개)
'20.3.	행정정보데이터세트 기록관리 매뉴얼 보완 및 1차 공지
'20.7.~	행정정보데이터세트 등록관리시스템 설계 및 구축
'21.3.~	행정정보시스템 기록관리 기능 요건(안) 개발 - 사전협의

※참고1 데이터세트 기록관리항목표 / 매뉴얼에 포함

구분	구상항목	설명	필수여부	
관리기관 정보	기관명	정보시스템 관리기관 기재	필수	
	부서명	정보시스템 관리 부서 기재	필수	
시스템 정보	시스템명	정보시스템 공식 명칭 기재	필수	
	구축연도	정보시스템 최초 구축연도 기재	필수	
	HW내역	HW명	HW명 기재(예. 서버, 스토리지 등)	필수
		용도	해당 HW 용도 기재	필수
		수량	해당 HW 수량 기재	필수
		사양	해당 HW 사양 기재	옵션
	SW내역	비고	해당 HW 특이사항 기재	옵션
		용도	구동되는 SW의 용도를 기재	필수
		제품명	SW제품명 기재	필수
		제조사	SW 제조사 기재	필수
		수량	SW 수량 기재	필수
		버전	SW버전 기재	옵션
		라이선스	해당 SW 라이선스 정책 기재	옵션
		설치HW	해당 SW가 설치된 HW명 기재	옵션
	DBMS	정형데이터가 저장되는 DBMS정보 기재	필수	
	백업정책	백업대상	백업 대상 기재	필수
		백업주기	백업 주기 기재	필수
		백업방법 및 매체	백업방법이나 백업된 정보가 저장된 매체 기재	옵션
	사용자 현황	정보시스템의 주 사용자 기재. 해당 사용자수를 추정할 수 있으면 추정치를 기재	필수	
	이력관리	기준표 변경의 원인이 되는 시스템 정보 변경 사항 기재	필수	
개발산출물	ERD, Table Layout, 프로그램연관도, 개발 산출물 등 시스템 구축 및 운영과 관련된 산출물만 표기	옵션		
업무 정보	업무 목적	시스템의 업무적 목적 기재	필수	
	업무 내용	업무명	해당 시스템에서 제공하는 주요 서비스 업무명 기재	필수
		세부내용	업무의 상세내용 기재	필수
대상	서비스의 대상 기재	필수		
법규 정보	법령, 고시, 규정, 표준	정보시스템과 관련된 법령, 고시, 규정, 표준 등 기재	필수	

※ 시스템 정보, 업무 정보 등은 EA 등록 정보 활용 가능, 향후 구축예정인 데이터세트 항목표 관리시스템은 EA 시스템과 데이터 연계·활용방안 모색
www.geap.go.kr 계정신청(기록관 필수)

구분	구상항목	설명	필수여부	
데이터 정보	대표 데이터	대표 데이터의 내용과 컬럼명 작성	필수	
	비정형데이터	종류	비정형데이터 종류나 확장자 기재	필수
		데이터	해당 비정형 데이터의 구분명 기재	필수
		구동가능 SW	해당 비정형 데이터를 구동할 수 있는 SW를 판별할 수 있는 경우 해당 SW명 기재	옵션
	데이터 해석의 존성	저장방법	비정형데이터의 저장 방법을 기재	옵션
		데이터	데이터만으로는 해석이 불가능한 데이터의 현상과 해석방법 기재	옵션
		비정문 기술	데이터를 인코딩한 기술 기재	옵션
	정보공개	해석방법	데이터를 해석할 수 있는 기술적 방법 기재	옵션
		사유	데이터가 비정문으로 저장되지 않는 사유 기재	옵션
		구분	정보공개 구분명을 기재	필수
		방법	정보를 공개하는 방법 기재	필수
	연계 데이터 현황	공개자료	공개자료의 대상(또는 범위) 기재	필수
		데이터 내용	타 시스템 연계로 생성 또는 활용되는 데이터 기재	필수
		연계 시스템	연계가 되는 대상 시스템명 기재	필수
		연계 방법	연계를 하는 기술적 방법 기재	필수
		활용 방법	해당 데이터의 활용 방법 기재	필수
		연계 방향	입력, 출력 등 연계 방향 기재	필수
		이력관리	기준표 변경의 원인이 되는 데이터 정보 변경사항 기재	필수
		기록관리 정보 (단위기능 별기록)	단위기능	도출된 단위기능 기재
	업무활용 목적		관리단위의 업무활용상 목적을 기재	필수
주제어	단위기능과 관련된 주제어(keyword) 기재		필수	
데이터 소유권	데이터를 생산 및 처리하는 기관 기재		필수	
데이터 관리권	데이터 관리 기관 기재		필수	
접근권한	단위기능 시스템 및 데이터에 접근할 수 있는 권한 기록기재		옵션	
비공개 대상	대국민 비공개 정보 포함 여부와 종류 기재		옵션	
비밀 여부	비밀 기록 여부 기재		옵션	
보존기간 및 처분 (처분 단위법)	처분단위		처분단위의 명칭 기재	필수
	보존기간		1년부터 영구 사이 년도 단위로 기간 기재	필수
	기산일		보존기간 시작일 기준 설정	필수
	적용범위		단위기능 전체 또는 일부 추출 조건 명시	필수
	최정사유	보존기간을 최정한 법률적 사유 또는 정책 기재	필수	
처분방법	보존기간 도래 시 처분 방법 기재	필수		



※참고2 그간 논의 사항1

1. 데이터세트 기록의 관리(안)의 이론적 근거와 고려사항

- 생산기관 자체 관리 · 국가기록원 선별 이관 모두 허용 필요(단계적 관리 탈피)
- 생애주기 관리모형과 기록연속체 모형의 병행 적용 필요
- 물리적 이관 · 관리 · 보존 패러다임의 전환(보관우선주의 극복) : → 법령 뒷받침 필요
(시행령에 데이터세트 관리 별도 정함 명기, 시행규칙 등에 프로세스 정의, 구체적 실행 가이드 개별 적용 등)
- 현용 여부와 보존기간 도래 여부에 관계없는 처분행위 가능(조기·지연처분 허용)
- 현장관리 중심, 일부 중요 데이터세트 국가기록원 이관 방안도 고려(비용,효율)
- 행정정보시스템의 유형에 따라 별도 관리방식 적용 고려 ex. NEIS vs. 통계시스템
- 가독 · 이해 · 해석이 가능한 재현성의 확보(이용가능성, 기본원칙)
→ 보존방안, 포맷(SIARD 포함) 등은 재현성 포함

2. BRM 체계의 기록관리기준과 데이터세트 기록관리기준 비교

- 시스템 계층 또는 단위기능 모두 기록관리 기준계층으로 유연하게 적용
(BRM 체계와 참조하되, “기관-시스템-단위기능”의 3계층을 기본으로 확대·축소 가능)
- 복수의 시스템이 연계(통합)된 경우 개별 시스템 또는 전체가 관리단위 가능(특허넷)
- 행정정보시스템의 데이터세트 기록관리주체(협의체 결정, 소유권한/관리권한 개념 분리 가능)
- 데이터세트는 이미 구조화되고, 자동화된 관리질서가 존재
(특정 분류체계로 재구조화하는 것은 손실이 많음 → 구조화된 DB를 유지한 분류 적용)
- 기록관리항목표(안)은 6개 영역 60여개 항목, 단일 관리시스템 구축(RM, AM 이외)
- 기록물 산정은 업무수/시스템 수 등으로 산정(테이블 단위 등은 과다 산정 문제 발생, 과거사례)

3. 데이터세트 기록의 장기보존방안, 처분과 시범적용 사업

- 다종의 행정정보 데이터세트 장기보존은 알려진 모든 방식을 적용할 수 있도록
마이그레이션 · 인캡슐레이션 · 에뮬레이션 방안 적용 가능(절차표준)
- 유럽의 E-ARK에서 채택하고 있는 SIARD는 DBMS의 테이블과 비정형데이터에
대한 보존방안으로 사용가능, 향후 기술적 보완 필요('19년 R&D 예정)
- 데이터세트의 처분은 프로세스 준수, 방식은 현장 적용(다수 기관 요청사항)
 - 처분 프로세스 참조모델, 현장 응용이 가능한 참조모형 필요
- 데이터세트 기록관리체계 구축은 '19년 시범사업 후 단계적 확산 예정(예산)
 - 추진체계 : 시범기관(기록관, 시스템부서) + 국가기록원(부처담당, 데이터세트팀)
 - 목 표 : 관리항목표조사 + 관리매뉴얼 적용 및 시범 실행(폐기심의, 목록, 처분도구 등)

4. 데이터세트 주요 유형 관리방안(메가시스템 or 다중복합시스템)

- 교육행정정보시스템(NEIS) 등 주요 데이터세트의 법정서식 관리방안 필요
 - NEIS는 학교(1만)·교육지원청(177)·교육청(17) 및 교육부의 대형 행정정보시스템
 - 생활기록부 등 교육 제증명 발급, 성적관리, 교무업무, 인사·급여 등 행정업무 수행

※ NEIS는 전자문서(서식), 정형·비정형데이터 모두 존재하는 전형적인 데이터세트

- 데이터세트 자체 관리도 필요하나, 중요 법정 서식형 DB는 UI 재현도 필요(viewer)
- 생활기록부, 성적표, 병적기록부, 수형자신분증, 출입국관리카드, 공무원인사카드 등은 출력관리(PDF), 별도 뷰어 제공 등의 방식도 고려 가능