

영구기록물관리를 위한 기록물 데이터베이스 스키마 개발 방향*

임진희**, 이대욱***, 김은실****, 김익한*****

1. 머리말
2. 영구기록물관리시스템의 데이터베이스 스키마 현황
 - 1) 영구기록물관리 시스템 검토
 - 2) 기록물 테이블 검토
 - 3) CAMS 데이터베이스의 과제
3. 영구기록물관리 데이터베이스 스키마 요건 확인
 - 1) 데이터베이스 스키마 설계 고려사항
 - 2) 데이터베이스 관리 고려사항
 - 3) 메타데이터 표준 관점에서의 고려사항
4. 영구기록물관리 데이터베이스 개념스키마 예시
 - 1) 입수단 개념스키마
 - 2) 보존단 개념스키마

* 국가기록원에서는 2015년부터 시작되는 대량의 전자기록물 이관에 대응하기 위하여 <차세대 전자기록관리 기술기반 연구>를 다년사업으로 진행하고 있다. 이 논문은 2011년 명지대 디지털아카이빙연구소가 주관연구기관이 되어 (주)엠포스, (주)세미콘네트웍스와 공동수행한 <차세대 전자기록관리 인프라 응용 기술 연구개발>(이하 “차세대연구”) 중 제1세부과제의 내용을 보완한 것이다.

** 주저자, 명지대 기록정보과학전문대학원 조교수. 제1세부과제 연구자.

*** 공동저자, 서강대 컴퓨터공학과 연구교수. 제1세부과제 연구자.

**** 공동저자, 대구광역시 달성교육지원청 기록연구사. 제1세부과제 연구자.

***** 교신저자, 명지대 기록정보과학전문대학원 원장. 총괄연구책임자.

- 3) 제공단 개념스키마
- 4) 논리스키마 개발 방향
- 5. 맺음말

[국문초록]

국가기록원 영구기록물관리시스템 CAMS(Central Archives Management System)는 2015년부터 매년 대량의 전자기록물을 이관받아 관리해야 할 중요한 시스템이다. CAMS 데이터베이스를 스키마 설계의 관점에서 진단해보고 전반적인 개선방향을 논의하는 일이 시급하다.

CAMS 데이터베이스의 중심부분인 기록물철 및 기록물건 테이블을 살펴본 결과 두 테이블 모두 정규화가 되어 있지 않으며, 용도 불명의 칼럼들이 혼재하고 있어 기록물 데이터의 품질을 신뢰하기 어려운 상황임을 알 수 있었다. 이 논문에서는 다음과 같이 기록물철 및 기록물건 테이블의 정규화방향을 제시하였다. 첫째, 두 테이블 간 중복 항목을 최소화하는 방향으로 칼럼을 재배치할 것. 둘째, 분류체계 정보 항목을 별도 테이블로 분리할 것. 셋째, 기록물의 형태 및 유형별 기술 항목을 별도 테이블로 분리할 것. 넷째, 인수 및 인계, 보존처리 등 기록관리 과정의 기술 항목을 별도 테이블로 분리할 것.

나아가 이 논문에서는 기록물의 입수, 보존, 제공 단계별로 데이터베이스 스키마 설계 및 관리 시 고려할 사항을 제시하였다. 입수단에서는 매년 대량의 이관기록물을 정해진 기간 안에 일괄 처리할 수 있어야 한다는 점, 보존단에서는 재분류, 재평가, 보존 처리와 같은 다양한 관리이력을 남길 수 있어야 한다는 점, 제공단에서는 접근도구에 필요한 데이터를 정해야한다는 점 등을 논

의하고 있다. 또한, 메타데이터 표준을 준수하는 방향에서 일부 개념 스키마를 개발하여 제시하고 있다.

주제어 : 전자기록관리, 영구기록물관리시스템, 기록물 데이터베이스 설계, 기록물 개념 스키마, 기록물 엔티티관계모형

1. 머리말

국가기록원은 우리나라 공공기록물 관리를 총괄 지도·감독하는 기관이며 기록물관리를 위해 영구기록물관리시스템 CAMS(Central Archives Management System)를 운영하고 있다. CAMS의 최초 모듈은 16비트 컴퓨터에서 구현되었으며 이후 32비트 컴퓨터를 거쳐 현재는 여러 서버로 구성된 시스템에 구현되어 있다. 2015년부터 시작될 대량의 전자기록물 이관에 대비하여 CAMS 데이터베이스의 스키마가 이관을 효과적·효율적으로 수행할 수 있도록 설계되었는지 점검해야할 시점이다.

전자기록관리시스템의 데이터베이스를 설계한 경험은 2004년 정부기관에 보급하기 시작했던 자료관시스템이 최초라 할 수 있다. 자료관시스템은 그간 기록관리의 주요대상이던 종이기록과 행정박물류를 전자적으로 관리하기 위한 기능과 2004년 본격적으로 보급되기 시작한 신전자문서시스템의 전자문서를 이관받아 관리하기 위한 기능을 갖춘 하이브리드시스템이었다. 전자정부의 추진으로 빠르게 컴퓨팅환경이 변화하던 각급 정부기관에서는 신전자문서시스템과 더불어 자료관시스템을 활발히 도입하였고, 그 결과 자료관시스템은 CAMS보다 먼저 본격적인 전자기록물 관리 기능을 탑재하게 되었다. 처분일정 상 전자기록이 정부기관에서 국가기록원으로 이관될 때까지는 몇 년의 시간 간격이 존

재하므로 CAMS는 자료관보다 뒤늦게 전자기록물 관리기능을 구현하게 되었다.

2005년 9월부터 2006년 2월까지 정부혁신지방분권위원회 산하 기록관리혁신위원회에서 작성한 기록관리혁신 로드맵에 따라 ‘기록관리시스템 혁신을 위한 정보화전략수립(ISP, Information Strategy Planning)’ 프로젝트가 진행되었다.¹⁾ 2007년에는 ISP 결과를 기반으로 자료관시스템을 업그레이드한 표준RMS(Records Management System)가 개발되었고, CAMS도 업그레이드 하게 되었다.²⁾

표준RMS의 데이터베이스 스키마는 자료관시스템에서 사용한 스키마를 보완하는 형태로 설계되었다. ISP 수행자와 표준RMS 개발자들 대부분이 자료관시스템을 개발했던 집단으로 기존의 스키마에 익숙했기 때문이다. CAMS의 경우는 통합되지 못한 여러 서버들 중 중심이 되는 서버의 스키마만 보완하게 되었는데 이 과정에서 과거 16비트 컴퓨터 시절부터 사용하던 스키마에 자료관시스템의 스키마가 추가되어 결합되는 양상이 되었다.

CAMS 데이터베이스 스키마의 변천과정에서 중요한 특징은 과거 종이기록 기반의 스키마를 새로운 전자적 관리체계에 맞춰 완전히 재편하는 방식이 아니라 기존의 스키마를 중심으로 두고 덧붙여 가는 방식을 채택한 점이다. 그 결과 현재 CAMS의 데이터베이스 스키마는 2015년 이후를 감당하기 어려운 상태가 되었으며 보완이 시급한 상황이다. 이에 이 논문에서는 CAMS의 데이터베이스 스키마 중에서 핵심이라 할 수 있는 기록물철 테이블과 기록물건 테이블의 현황과 문제점을 살펴보고, 향후 기록물 데이터베이스 스키마의 개선 방향을 제시하고자 한다.

1) 광정, 「행정기관의 기록관리시스템 개선모델 분석-2006년 기록관리시스템 혁신을 중심으로-」, 『기록학연구』 제14호, 2006, 154쪽.
2) 송병호, 「기록관리시스템의 현황과 전망」, 『기록학연구』 제21호, 2009, 390쪽.

데이터베이스를 설계하고 구축하기 위해서는 몇 가지 과정을 거치게 된다. 그 과정은 요구사항분석-개념적 설계-논리적 설계-물리적 설계이다. 먼저 사용자 요구사항을 분석하고 이를 기반으로 데이터베이스관리시스템의 종류와는 독립적으로 개념적 설계를 한다. 다음으로 특정 데이터베이스관리시스템에 적합한 논리적 설계를 거쳐 마지막으로 물리적 설계가 이루어진다. 각각의 개발 단계에서 개념스키마, 논리스키마, 물리스키마들이 만들어진다. 개발 관점의 구분법과 달리 사용자 관점에서는 데이터베이스 스키마를 외부스키마, 개념스키마, 내부스키마로 나누기도 한다. 이 논문에서는 데이터베이스 설계 과정 중 논리적 설계에 앞서 이루어지는 개념적 설계의 과정이나 그 결과를 의미하는 개념스키마, 그리고 사용자 관점에서 본 개념스키마에 해당하는 것으로서 '개념스키마'라는 용어를 사용하고자 한다.

이 논문에서는 먼저 국가기록원의 CAMS에 구축되어 운영 중인 데이터베이스를 분석하여 문제점들을 도출하였다. 이는 현재의 데이터베이스 시스템이 수정·보완되어야 하는 점에 대한 타당성을 제시한다. 또한 국가기록원의 메타데이터 표준을 참조하여 CAMS 데이터베이스를 위한 개념스키마를 제시하였다. 현재의 CAMS 데이터베이스의 기록물철 및 기록물건 테이블의 정규화는 물론 기록물관리를 위한 데이터베이스에 구현되어야 할 핵심적인 엔티티들을 제시한다.

2. 영구기록물관리시스템의 데이터베이스 스키마 현황

1) 영구기록물관리 시스템 검토

CAMS는 정보시스템의 일종이다. 모든 정보시스템은 업무활동을 자

동화하거나 의사결정과정을 지원하기 위해 입력데이터를 처리하여 필요한 정보를 산출한다. 따라서, 특정 정보시스템의 특성을 파악하기 위해서는 어떤 업무활동과 의사결정을 위한 시스템인가를 이해하는 것이 중요하다. 이 절에서는 CAMS에 구현하고 있는 업무기능 현황을 살펴보고자 한다.

CAMS에 구현된 업무프로세스³⁾는 총 14개의 대분류업무 하에 35개의 중분류 업무, 그 하위에 61개의 단위 업무프로세스로 구성되어 있다. 14개의 대분류업무는 인수, 등록, 기술·재분류, 매체수록, 서고, 보존, 기록관리기준, 통계, 검색, 전거관리, 사용자관리, 일반지원, 분류체계, 코드관리 등으로 구성되어 있다.⁴⁾

한편 국가기록원에서는 정부공공기관의 다양한 행정정보시스템에서 생산되어 관리되는 데이터세트 형태의 기록 중 보존가치가 높은 대상을 관리하기 위한 체계와 시스템을 마련하고 있다. 2007년 ‘행정정보시스템 데이터세트 기록관리연구’ 사업에서는 행정정보시스템 유형을 분류하고 기록관리 대상을 선정하는 기준을 마련한 바 있다. 2009년과 2010년에 걸쳐 ‘데이터세트 및 비표준문서 기록관리체계 시범구축’ 사업을 통해 데이터세트아카이브관리시스템(DAMS, Dataset Archives Management System)를 시범 구축하였다. DAMS의 시스템 기능은 총 11개의 대기능, 46개의 중기능, 60개의 소기능으로 구성되어 있다.⁵⁾

3) <차세대 전자기록관리 인프라 응용기술 연구개발> 사업 최종보고서, 국가기록원, 2011, 105-108쪽.

4) CAMS 매뉴얼에는 ‘대분류 업무명’, ‘중분류 업무명’, ‘단위 프로세스명’이라고 제목이 붙어 있는데 내용적으로는 시스템으로 구현되어 있는 업무들을 분류하여 기능메뉴로 제시하고 있는 것으로 볼 수 있다. 업무명칭을 정의할 때는 해당 업무의 주요 자원(resource)명과 처리내역이 조합되었을 때 명료해지며 자원명이 보편적일 때는 이를 생략할 수 있다. 현재 CAMS의 업무명칭은 명명규칙에 맞춰 재정리할 필요가 있다.

5) <차세대 전자기록관리 인프라 응용기술 연구개발> 사업 최종보고서, 국가기록원, 2011, 105-108쪽.

DAMS는 CAMS의 일부로 해석할 수 있다. CAMS 자체가 단일한 하나의 서버가 아니고 국가기록원이 장기보존 기록물을 통합적으로 관리보존하기 위해 운영하는 모든 시스템을 통칭하여 부르는 이름이기 때문이다. [그림 1]⁶⁾은 2000년부터 2011년 12월 현재까지 국가기록원이 나라장터를 통해 기록물관리시스템의 개발 및 유지보수 사업 발주를 하면서 작성한 제안요청서(Request for Proposal)를 근거로 구성해 본 CAMS 서브시스템과 모듈 현황이다. CAMS는 기록물 입수를 위해 각 생산기관의 기록관리시스템, 자료관시스템과 연계되며, 전자기록의 장기보존을 위해 행정전자서명, 정부기능분류, 행정표준코드 등의 여러 시스템과 연계된다. CAMS는 정부기록보존소 시절 16비트 컴퓨터용 목록관리 소프트웨어로부터 32비트 컴퓨터용 소프트웨어, 이후 본격적인 전자기록관리시스템으로 변화, 발전해 왔으며, 위 그림의 각 모듈에는 기록물에 대한 전자적 관리의 변천사가 반영되어 있다고 볼 수 있다. 또한, CAMS는 대부분 Oracle 데이터베이스관리시스템을 사용하고 있는 것으로 파악되고 있다.

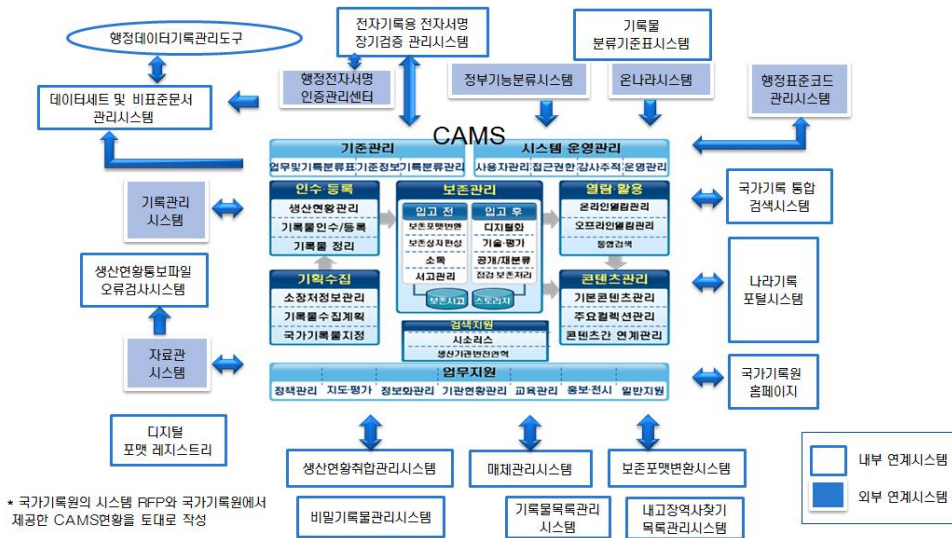
[그림 1]에 서로 연관되어 있는 서브시스템과 모듈을 영구기록물관리의 생애주기 단계별⁷⁾로 정렬해 보면 다음 [그림 2]와 같다.⁸⁾ 입수단은 ISO14721 OAIS참조모형에서 입수(Ingest)기능을 수행하는 서브시스템과 모듈의 합집합으로 구성된다. 보존단은 OAIS참조모형에서 보존계획(Preservation Planning), 운영(Administration), 데이터관리(Data Management), 저장(Storage) 기능을 수행하는 서브시스템과 모듈의 합집합으로 구성된다.

6) <차세대 전자기록관리 인프라 응용기술 연구개발> 사업 최종보고서, 국가기록원, 2011, 69쪽.

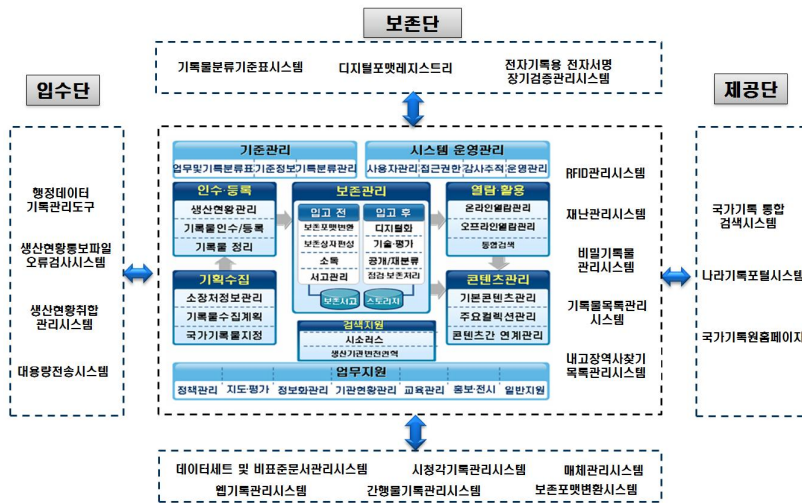
7) 영구기록물관리의 생애주기 단계 구성은 임진희·이대옥, 「대량기록물 처리를 위한 영구기록물관리시스템의 디지털저장소 배치형상 연구」, 『기록학연구』 제 32호, 2012, 181-192쪽을 참조하였다.

8) <차세대 전자기록관리 인프라 응용기술 연구개발> 사업 최종보고서, 국가기록원, 2011, 70쪽의 그림을 재구성한 것임.

[그림 1] 나라장터 조달정보에 근거하여 재구성한 CAMS의 구성현황



[그림 2] 기록물 생애주기별로 재구성한 CAMS의 구성 현황



2) 기록물 테이블 분석

(1) 스키마 분석 대상

데이터베이스 스키마는 개념, 논리, 물리스키마로 구성된다. 개념스키마는 데이터베이스를 설계하기 위한 첫 단계에서 생성하는 스키마로 데이터에 대한 요구사항 분석 결과를 엔티티관계도(ERD, Entity Relationship Diagram)와 같은 모형으로 표현하고 정의한 것이다. 추상적 수준에서 주요 관리대상 데이터가 무엇인지, 데이터 간의 연관관계 중 의미있는 것이 무엇인지에 집중하는 단계이다. 논리스키마는 데이터베이스관리 시스템과 데이터베이스의 유형을 정한 후 개념스키마를 이에 맞게 구체화하는 단계로 만약 관계형 데이터베이스를 선택한 경우라면 테이블과 칼럼을 정의하고 나아가 성능을 고려하여 파티션(Partition), 클러스터(Cluster), 뷰(View), 인덱스(Index) 등을 정의한 것이다. 데이터베이스의 유형, 데이터의 양과 사용방식에 따라 논리스키마를 산출하는 기법과 주안점이 달라진다. 물리스키마는 데이터베이스가 구현될 저장공간과 매체의 특성을 고려하여 데이터파일의 위치와 데이터 블록의 크기를 지정하는 등 데이터의 저장방식을 정의한 것이다. 이는 컴퓨터시스템에서 실행시켜 데이터베이스를 생성할 수 있는 명령문을 만들어내는 단계라고 볼 수 있다.

국가기록원 CAMS 데이터베이스 스키마는 기록물철과 기록물건 테이블에 대한 분석으로 범위를 국한하였다. 전체 데이터베이스 스키마에 대한 전반적인 현황 분석과 개선 방향의 제시가 필요함에도 불구하고 범위를 이와 같이 좁힌 이유는, 연구자가 데이터베이스 스키마 전체에 접근할 수 없고 기록물철과 기록물건 테이블에 대한 정보만 제공받을 수 있었던 현실적 제약이 직접적인 이유이다. 하지만, 기록물철과 기록물건 테이블에 대한 분석 결과 CAMS 데이터베이스 스키마의 현황 수준을 추론할 수 있었으며, 연구목적에 맞는 유의미한 시사점을 도출할 수

있었다.

(2) 기록물철 테이블과 기록물건 테이블 간의 칼럼 재조정

먼저 기록물철 테이블과 기록물건 테이블의 칼럼이 적절하게 설계되었는지를 정규화의 관점에서 분석해 보았다.⁹⁾ CAMS에서 기록물철 테이블은 ‘관리기본목록’이라는 이름의 테이블로, 기록물건 테이블은 ‘관리상세목록’이라는 이름의 테이블로 구현되어 있다. ‘관리기본목록’ 테이블은 총 144개의 칼럼으로 구성되어 있으며 ‘관리상세목록’ 테이블도 비슷한 개수의 칼럼으로 구성되어 있다. 그런데, 두 테이블에 같은 명칭을 가진 칼럼이 다수 정의되어 있고, 그 칼럼이 철 단위 고유의 정보인지 건 단위 고유의 정보인지 구분되지 않은 상태로 다중 정의되어 있는 것으로 파악되었다.

[표 1]은 ‘관리기본목록’과 ‘관리상세목록’ 테이블에 공통으로 정의된 칼럼들을 정리한 것이다. 총 39개 칼럼이 중복 정의되어 있다. 중복 칼럼 중 일부는 기록물철과 기록물건 각각에 대해 정보를 보유할 필요가 있으나, 칼럼 중 일부는 기록물철 혹은 기록물건에만 국한되어야 하는 고유 정보로서 정규화된 테이블이라면 중복 정의가 불가한 것들이다. 기록물철과 기록물건 중 어느 대상에 대한 고유 정보인지 판단하는 것은 CAMS 응용(Applications)에서 각 칼럼 항목들이 어떻게 쓰이고 있는지를 살펴보고 데이터베이스에 축적된 실제 데이터 현황을 추적함으로써 가능하다. 즉, 정보시스템 설계 시 데이터베이스의 테이블 단위, 칼럼 단위로 응용의 모듈별 CRUD 분석을 수행하게 되는데 이 분석결과를 확인하면 되는 것이다. CRUD 분석이란, 업무과정에서 응용의 어떤 메뉴 화면이나 기능을 수행하여 어떤 데이터를 생성(Create), 조회(Retrieve), 갱신(Update), 삭제>Delete)하게 되는지 살펴보는 것이다. 이는 시스템의

9) 테이블의 분석은 2011년 차세대연구 수행 시점의 CAMS 데이터베이스 현황을 반영하고 있다.

기능측면과 데이터측면의 설계가 제대로 되었는지를 상호 확인하는 방법으로 활용되기도 한다.

[표 1] 기록물철과 기록물건 테이블에 공통으로 정의되어 있는 칼럼

칼럼명	데이터타입	설명
BSID	Varchar2(12)	기본목록SID
KIKWANCODE	Varchar2(10)	처리기관코드
OLDPRODKIKWAN	Varchar2(120)	구기록물생산기관명
DOCTYPE	Varchar2(1)	기록물형태_RG001(A: 문서대장, B: 도면, C: 사진/필름, D: 녹음/동영상, E: 카드, G: 국무회의록)
ENDYEAR-	Varchar2(4)	종료년도
DAMDANG	Varchar2(60)	업무담당자
OLDGUBUN	Varchar2(1)	신구기록물구분
ISUPDATE	Varchar2(1)	수정여부
PAGE	NUMBER(6,0)	쪽수
CONTEXTMEMO	Varchar2(4000)	내용요약
MINGNO	Varchar2(10)	기록물관리번호
ORGTTYPE	Varchar2(2)	기록물유형(RG008: Z=파일(사진/필름류) 추가
KYUKYUK	Varchar2(16)	규격
WONBONGUBUN	Varchar2(1)	원본사본여부
JAEJIL	Varchar2(40)	재질
FST_REG_DT	DATE(7)	최초등록일시
FST_REGR_ID	Varchar2(20)	최초등록자ID
DOCBUNRYUNO	Varchar2(86)	기록물철분류번호
REMARK	Varchar2(500)	비고
OLDDOCNO	Varchar2(50)	구기록물문서번호
OPN_YN	Varchar2(2)	공개여부(1: 공개, 2: 부분공개, 3: 비공개)
ORI_JEMOK	Varchar2(500)	원어명
COLLECT_STATUS	Varchar2(1)	수집상태
RECORD_CENTER_FLAG	Varchar2(2)	영구보존포맷변환여부

SUB_TTL	Varchar2(500)	부제목
UP_TTL	Varchar2(500)	정정제목
SECU_CLSS_CD	Varchar2(2)	보안분류코드(RM정보->01:일반, 02:대외비, 03:3급비밀, 04:2급비밀, 05:1급비밀)
SECU_CLSS_RSN	Varchar2(200)	보안분류사유(RM정보)
READ_RANGE	Varchar2(2)	열람범위
LST_UPDR_ID	Varchar2(20)	최종수정재ID
LST_UPD_DT	DATE(7)	최종수정일시
LIST_OPEN_DIV	Varchar2(2)	목록공개
EQ_TTL	Varchar2(500)	대등제목
ETC_TTL	Varchar2(500)	기타제목
TMP_TTL	Varchar2(1000)	제목(띄어쓰기)
SERVICE_BND_TTL	Varchar2(500)	서비스용제목(기통검)
ISSERVICE	CHAR(1)	서비스제목적용여부(1:적용)
SVC_LST_UPD_DT	DATE(7)	서비스제목최종수정일자
SVC_LST_UPDR_ID	Varchar2(20)	서비스제목최종수정재ID

아쉽게도 차세대연구에서는 CAMS 응용별로 데이터베이스의 사용현황을 조사할 수는 없었다. 소량의 데이터 샘플을 통해 칼럼의 용도를 추정해 보고 CAMS 담당자의 인터뷰를 통해 일부 칼럼의 용도에 대해 확인할 수 있었을 뿐이다. 조사결과 최소한 종료년도(ENDYEAR), 신규기록물구분(OLDGUBUN), 기록물철분류번호(DOCBUNRYUNO) 칼럼들은 기록물철 단위에서만 의미있는 정보이며 기록물건 단위에서는 의미가 없는 정보라는 판단을 할 수 있었다. 또한, 처리기관코드(KIKWANCODE), 구기록물생산기관명(OLDPROD KIKWAN), 기록물형태(DOCTYPE), 업무담당자(DAMDANG), 쪽수(PAGE), 기록물유형(ORGTYP), 규격(KYUKYUK), 원본사본여부(WONBONGUBUN), 재질(JAEJIL), 구기록물문서번호(OLDDOCNO), 공개여부(OPN_YN), 영구보존포맷변환여부(RECORD_CENTER_FLAG), 보안분류코드(SECU_CLSS_CD), 보안분류사유(SECU_CLASS_RSN), 열람범위

(READ_RANGE), 목록공개(LIST_OPEN_DIV) 등 16개의 칼럼은 기록물건에만 관련된 정보를 저장하는 항목으로 기록물건 테이블에만 정의하는 것이 타당하다고 판단할 수 있었다. 수집상태(COLLECT_STATUS) 칼럼의 경우에는 기록물의 수집이 철 단위로만 가능한 것인지 건 단위로도 가능한 것인지에 따라 기록물철 테이블에 둘 것인지 기록물건 테이블에 둘 것인지 결정해야 할 것으로 판단하였다.

분석결과 기록물철 단위와 기록물건 단위 모두에서 공통으로 필요한 칼럼은 [표 2]와 같이 총 18개로 판단되었다.

[표 2] 기록물철과 기록물건 테이블에 공통으로 필요한 칼럼 목록

칼럼	데이터타입	설명
ISUPDATE	Varchar2(1)	수정여부
CONTEXTMEMO	Varchar2(4000)	내용요약
MINGNO	Varchar2(10)	기록물관리번호
FST_REG_DT	DATE(7)	최초등록일시
FST_REGR_ID	Varchar2(20)	최초등록재ID
REMARK	Varchar2(500)	비고
ORI_JEMOK	Varchar2(500)	원어명
SUB_TTL	Varchar2(500)	부제목
UP_TTL	Varchar2(500)	정정제목
LST_UPDR_ID	Varchar2(20)	최종수정재ID
LST_UPD_DT	DATE(7)	최종수정일시
EQ_TTL	Varchar2(500)	대등제목
ETC_TTL	Varchar2(500)	기타제목
TMP_TTL	Varchar2(1000)	제목(띄어쓰기)
SERVICE_BND_TTL	Varchar2(500)	서비스용제목(기통검)
ISSERVICE	CHAR(1)	서비스제목적용여부(1: 적용)
SVC_LST_UPD_DT	DATE(7)	서비스제목최종수정일자
SVC_LST_UPDR_ID	Varchar2(20)	서비스제목최종수정재ID

현재 CAMS의 데이터베이스에는 ‘생산기본목록’ 테이블과 ‘이관기본목록’ 테이블도 정의되어 있다. ‘관리기본목록’ 테이블에 정의되어 있는 대부분의 칼럼이 이 두 테이블에도 상당부분 유사하게 반복되고 있다. 누가 언제 어떤 칼럼에 어떤 값을 입력하게 되는지 세밀히 살펴보고, 혹시 동일한 데이터가 중복 저장되고 있다면 데이터의 무결성과 일관성이 보장되도록 조치가 취해지고 있는지 확인해야 한다. 조치가 미흡하다면 데이터의 품질 보장을 위해 꼭 필요한 테이블에만 관련 칼럼을 정의하도록 조정해야 한다.

(3) 기록물철 테이블 정규화

하나의 테이블에 정의된 칼럼들 중에 데이터값의 발생시점이 동일하고 이용되는 시점이 동일한 부분집합이 존재한다면 이를 별도의 테이블로 분해할 필요가 있는지 고려해야 한다. 만약, 데이터값이 지속적으로 변경되고 그 이력을 남길 필요가 있다면 해당 칼럼들은 별도의 테이블로 분리해야만 한다. 현재 기록물철 테이블에 정의된 144개의 칼럼 중에는 ‘관리상세목록’, ‘생산기본목록’, ‘이관기본목록’ 테이블의 칼럼들과 중복되는 것이 많다. 동일한 값을 저장하는 중복 칼럼을 없애는 방향으로 조정을 한다면 ‘관리기본목록’에는 최소한의 칼럼만을 남기고 칼럼들을 별도의 테이블로 독립시키되 테이블끼리의 참조 관계를 맺어 주도록 해야 한다. 기록물철 테이블을 사례로 하여 어떻게 정규화를 할 수 있는지 관점과 과정을 아래와 같이 제시하고자 한다. 이는 기록물건, 생산기록물, 이관기록물 테이블에도 유사하게 적용할 수 있을 것이다.

첫째, 중복 칼럼을 삭제한다. 기록물철 테이블에 있는 칼럼 중 기록물건 테이블에 있어야 하는 칼럼은 삭제하도록 한다. 삭제 대상 칼럼은 권호수(VOLNO), 쪽수(PAGE), 전자파일갯수(ELECFILENO), 기록물관리번호(MNGNO), 규격(KYUKYUK), 원본사본여부(WONBONGUBUN), 권수

(DOCQTY), 사용문자(USECHAR), 공개여부(1: 공개, 2: 부분공개, 3: 비공개)(OPN_YN), 전시여부(DISPLAYGUBUN), 기록매체(MEDIUMTYPE), 영구 보존포맷변환여부(RECORD_CENTER_FLAG),

보존매체관리번호(OD_MNGNO), 매체등록일련번호(MEDIA_SEQ_NO) 등이다.

둘째, 마스터 성격의 데이터집합은 분리한다. ‘관리기본목록’에는 기록물의 분류에 관련한 칼럼이 [표 3]과 같이 11개로 정의되어 있다. 기능분류와 기술분류, 목적별분류에 대해 ID나 번호, 명칭, 경로명 등이 칼럼으로 정의되어 있다. CAMS의 경우 해마다 기록물의 양이 누적적으로 증가하게 되는데 모든 기록물철 단위에서 ID 뿐만 아니라 명칭과 설명 등 분류체계에 대한 모든 정보를 보유할 것인지 성능의 관점에서 재고해야 한다. 일반적으로 분류체계와 같은 기준값들은 별도의 테이블에 따로 저장한다. 마스터 성격을 갖는 테이블의 정보는 일시에 데이터가 저장된 후 간헐적으로 수정보완이 되며, 다른 여러 테이블에서 ID를 통해 내용을 참조하는 방식으로 사용되는 것이 일반적이다. 현재 [표 3]을 통해 유추할 수 있는 것은 별도로 분류체계 테이블이 존재하더라도 분류체계의 ID만 참조하지 않고 명칭도 참조하는 방식으로 비정규화되어 있다는 점이다. 기록물 데이터베이스의 성능을 고려한 조치일 수 있으나 무결성 통제에는 취약할 수 있다는 점이 고려되어야 할 것이다.

[표 3] 분류체계 관련 칼럼 목록

칼럼	데이터타입	설명
OLDBUNRYUNO	Varchar2(35)	구기록물철분류번호
BIS_CLSS_PATH_NM	Varchar2(1000)	기능별분류체계경로명
DOCBUNRYUNO	Varchar2(86)	기록물철분류번호
DESC_CLASS_ID	Varchar2(12)	기술분류ID
DESC_LST_UPD_DT	DATE(7)	기술분류최종수정일시
DESC_LST_UPDR_ID	Varchar2(20)	기술분류최종수정자ID
PJT_CLSS_PATH_NM	Varchar2(1000)	목적별분류체계경로명
CLSS_ID	Varchar2(35)	분류체계ID
CLSS_DIV_ID	Varchar2(4)	분류체계구분ID
CLSS_NM	Varchar2(200)	분류체계명
TITLE_COL	Varchar2(120)	콜렉션명

셋째, 성격이 전혀 다른 서브타입에 해당하는 데이터집합은 분리한다. ‘관리기본목록’에는 기록물의 형태 및 유형에 관련한 칼럼들이 [표 4]와 같이 15개가 정의되어 있다. 칼럼 설명을 참조하여 이들의 용도를 구분해보면 마이크로필름류, 구술기록류, 도면류, 스캐닝파일류, 국무회의록류, 정부간행물류, 행정박물류 등에 관한 칼럼들이 포함되어 있음을 알 수 있다. 우선, 기록물의 ‘형태’와 ‘유형’이 각각 무엇을 의미하는지 명확한 구분이 선행되어야 하며, 기록물이 특정한 유형일 때 값을 지정할 수 있는 칼럼들이 무엇인지 충분히 분석하여 정의해야 한다. 예를 들어, 현재 행정박물류의 경우 물리적 특성에 대해 최대 지름을 저장하는 칼럼밖에 정의되지 않았으나 실제 물리적 관리를 위해서는 높이, 폭, 재질 등 여러 관리 값을 입력할 필요가 있을 것이다. 기록물의 유형별로 필요한 칼럼은 메타데이터 표준을 준용하고 실무적 경험에 따라 추가함으로써 상세하게 정의할 수 있다. 만약 하나의 기록물철에 하나의 형태 혹은 유형의 기록물들만이 편철되도록 관리하는 경우라면 해당하

는 하나의 형태나 유형에 관한 칼럼에만 값이 지정되고 나머지는 비게 된다. 상황에 맞게 값의 유무를 통제하기 위해서는 기록물 유형별 고유 의 칼럼들을 별도의 테이블로 분리하는 것이 유리하다.

[표 4] 기록물철의 형태 및 유형 칼럼 목록

칼럼	데이터타입	설명
MFILMNO	Varchar2(20)	M/F번호(색인대장)
MF_FSH	Varchar2(1)	M/F촬영완료여부(1:완료)
MFSNO	NUMBER(7,0)	구M/F 번호
REG_SEQ_NO	Varchar2(10)	구술기록등록일련번호
KYUKYUK	Varchar2(16)	규격
MEDIUMTYPE	Varchar2(1)	기록매체
DOMYUNCODE	Varchar2(1)	도면규격코드
DIGITAZED_FSH	Varchar2(1)	디지털화작업 완료여부(Y/N)
SCAN_FSH	Varchar2(1)	스캐닝완료여부(Y:완료)
FRCHASU	NUMBER(4,0)	시작회차(국무회의록 전용)
REGISTRY_FSH	Varchar2(1)	정부간행물 전용 등록완료여부(Y/N)
TOCHASU	NUMBER(4,0)	종료회차(국무회의록 전용)
HANGJONG_USERSERVICE	Varchar2(3)	행정박물 사용용도
HANGJONG_TYPEGUBUN	Varchar2(3)	행정박물 유형구분
MAXDIAMETER	NUMBER(5,0)	행정박물 최대지름

또한, ‘관리기본목록’에는 기록물의 보존에 관련한 칼럼이 [표 5]와 같이 19개 정의되어 있다. 기록물철을 보존하는 과정에서 종이기록물철과 전자기록물철은 상당히 다른 정보를 필요로 한다. 그러나 현재는 종이 기록물철을 보존하는데 필요한 정보 항목을 중심으로 칼럼이 정의되어 있다. 따라서, 현재의 칼럼들은 별도의 테이블로 분리하고 또한 전자기록물철의 보존관리에 필요한 칼럼들도 별도의 테이블로 분리한 후 기록물철이 종이기록물철인지 전자기록물철인지에 따라 연관관계를 맺도

록 처리해야 할 것이다.

[표 5] 종이기록물철 관련 칼럼 목록

칼럼	데이터타입	설명
POSITION	Varchar2(2)	기록물위치
BOKWANBOXNO	Varchar2(21)	보관상자번호
KIKAN	Varchar2(2)	보존기간
PRESERVE_END_YEAR	Varchar2(4)	보존기간만료년도
KIKAN_FIX_RSN	Varchar2(200)	보존기간책정사유명
OD_MINGNO	Varchar2(12)	보존매체관리번호
METHOD	Varchar2(1)	보존방법 (1: 병행보존 2: 매체보존 3: 원본보존)
ENVELOPE_STANDARD	Varchar2(2)	보존봉투규격
ENVELOPE_CNT	NUMBER(5,0)	보존봉투수량
ARCH_PLACE	Varchar2(2)	보존시설 : A=서울, B=부산, C=대전, D=성남
PLACE	Varchar2(1)	보존장소(1: 자료관, 2: 전문관리기관)
ENDSAYU	Varchar2(100)	비치사유
ENDDATE	Varchar2(10)	비치종결일자
BOXWSIZE	NUMBER(5,0)	상자가로
BOX_TYPE	Varchar2(1)	상지구분(1: 보존상자(소), 2: 보존상상(대), 3: 제작상자)
BOXHSIZE	NUMBER(5,0)	상자높이
BOXWEIGHT	NUMBER(10,3)	상자무게
BOXLSIZE	NUMBER(5,0)	상자세로
DISPLAYGUBUN	Varchar2(3)	전시여부

또한, '관리기본목록'에는 구기록물 및 생산관련 정보를 담고 있는 항목이 18개가 존재하는데 [표 6]과 같다. 이 중에는 구기록물의 경우에만 값이 있는 칼럼, 생산맥락정보를 담은 칼럼, 이관정보를 담은 칼럼 등이 포함되어 있다. 최소한 구기록물 정보는 별도의 테이블로 분리해내

는 것이 필요한데 이유는 CAMS에서 관리하는 기록물 중 구기록물은 일부이므로 이 칼럼들에 값이 필수로 지정되지 않을 것이고, 이처럼 선택적으로 값이 지정되는 칼럼들은 별도의 테이블로 관리하는 것이 무결성 통제나 성능 향상에 유리하기 때문이다.

[표 6] 구기록물철 관련 칼럼 목록

칼럼	데이터타입	설명
OLDDOCNO	Varchar2(50)	구기록물문서번호
OLDPRODKIKWAN	Varchar2(120)	구기록물생산기관명
OLDKIKWANCODE	Varchar2(10)	구기록물처리과코드
OLDBUNRYUNO	Varchar2(35)	구기록물철분류번호
FILESNO_ORI	Varchar2(100)	구문서번호(도면번호), 원문서번호
OLDSOJANGCHEO	Varchar2(40)	구소장처
NATIONCODE	Varchar2(3)	생산국가
COUNTRYNAME	Varchar2(40)	생산국명
PRODKIKAN	Varchar2(2)	생산기관
PRODKIKWANNAME	Varchar2(40)	생산기관명
PRODYEAR	Varchar2(4)	생산년도
PRODSYSDIV	Varchar2(2)	생산시스템구분(RG053)
OLDGUBUN	Varchar2(1)	신구기록물구분
INSUKIKWAN*	Varchar2(10)	이관기관
DHOUSECODE	Varchar2(10)	자료관코드
INSUYEAR	Varchar2(4)	수집년도
COLLECT_STATUS	Varchar2(1)	수집상태
SUJIPCHEO	Varchar2(40)	수집처

넷째, 다중값이 발생하는 데이터집합은 분리한다. ‘관리기분목록’에는 기록물의 인수 및 인계에 관련한 칼럼이 [표 7]과 같이 12개가 정의되어 있다. 기록물철이 영구기록물관리기관으로 입수되기 전 생산기관에서

인수 및 인계가 수행된 정보를 입력하는 칼럼들로 파악된다. 그런데, 기록물철의 인수 및 인계가 처리과 단위에서 이루어질 수도 있고, 기록관으로 이관 시 이루어질 수도 있다면 하나의 기록물철에 대해 다중의 인수 및 인계 정보가 발생할 수도 있다고 판단된다. 그런데, 현재 한 번의 인수 및 인계 내역만을 관리할 수 있도록 설계되어 있으므로 관련 칼럼들을 별도의 테이블로 분리하여 인수 및 인계 이력을 관리할 수 있도록 해야 한다. 또한, [표 7]의 목록에서 보는 바와 같이 인수일자, 입수일자, 인수인계일자 등 여러 종류의 날짜 데이터가 존재한다. 각각 어떤 날짜를 의미하는지 구분 기준을 명확히 해야 한다. 만약 인수일자와 인수인계일자가 동일한 정보를 저장하는 것으로 판단될 경우 칼럼을 하나로 통합하는 것이 필요하다.

[표 7] 기록물철의 인수 및 인계 칼럼 목록

칼럼	데이터타입	설명
INSUKIKWAN	Varchar2(10)	이관기관
TAKEOVER_NM	Varchar2(20)	인계자
INSUGUBUN	Varchar2(1)	인수인계_구분
INSUVOLNO	Varchar2(3)	인수인계_권호수
INSUDOCYEARNO	Varchar2(35)	인수인계_기록물철등록일련번호
INSUDANWICODE	Varchar2(35)	인수인계_단위업무코드
INSUPRODDATE	Varchar2(4)	인수인계_생산년도
INSUKIKWANCODE	Varchar2(10)	인수인계_처리기관코드
INSUPROD_DT	Varchar2(8)	인수인계일자
INSUDATE	Varchar2(8)	인수일자
INSUJA	Varchar2(40)	인수자
ACQDATE	Varchar2(8)	입수일자

(4) 칼럼의 데이터타입 조정

기록물철과 기록물건 테이블의 칼럼을 점검하는 과정에서 명칭과 데이터 타입을 조정해야할 필요가 있음을 파악할 수 있었다. 동일한 정보를 담는 칼럼이면서 명칭만 다른 것으로 추정되는 항목으로는 기록물철 테이블의 공개여부(OPN_YN)라는 칼럼과 기록물건 테이블의 공개구분(ISOPEN)이라는 칼럼을 예로 들 수 있다. 두 칼럼 모두 기록물의 내용을 국민에게 공개할 수 있는 범위를 지정하도록 정의되었다. 공개, 부분공개, 비공개의 3가지 종류 값을 가질 수 있으며, 공개면 1, 부분공개면 2, 비공개면 3의 값을 지정하도록 설계되어 있다. 또 다른 예로, 기록물 테이블에 NAME이라는 칼럼과 HUMANNAME이라는 칼럼이 존재한다. 같은 의미의 속성임에도 불구하고 이전 설계된 내용에 대해 파악이 안되어 새로 속성을 추가하여 이중적으로 사용하고 있는 것으로 밝혀졌다. 기타제목, 대등제목과 같은 칼럼들도 비슷한 사례이다.

다음으로 데이터타입을 조정해야 하는 사례로는 첫째, 공개여부 칼럼을 살펴볼 수 있다. 앞에서 살펴본 두 칼럼의 데이터타입은 동일하지 않다. 기록물철 테이블 칼럼은 Varchar2(2) 타입인데 기록물건 테이블 칼럼은 Varchar2(9)로 정의되어 있다. 최대 입력값의 길이가 문자 2개와 문자9개로 서로 다르게 정의되어 있는 것이다. Varchar2 타입의 경우 실제 발생할 수 있는 값을 정확히 분석하여 최대값 길이를 지정함으로써 유효한 값을 통제할 수 있다. 공개여부 칼럼의 경우 세 가지 종류의 값밖에 가질 수 없다면 Varchar2(1)로 충분하다. 2나 9와 같이 최대값을 크게 주면 오류값이 저장될 가능성이 있기 때문에 권장하지 않는다.

두 번째 사례로는, 날짜 칼럼들을 살펴볼 수 있다. 기록물철 테이블에는 종료년도(ENDYEAR)라는 칼럼이 Varchar2(4)로, 기록물건 테이블에는 종료일자(EDATE)라는 칼럼이 Varchar2(8)로 정의되어 있다. 기록물철의 경우 연도 정보만 저장하고, 기록물건의 경우 일자까지 저장하기 위해 데이터의 길이가 달라진 것으로 해석된다. 기록물 데이터베이스

전체적으로 연도나 일자를 저장하기 위한 데이터타입을 일관되게 정할 필요가 있다. 오라클 데이터베이스를 사용하는 경우라면 날짜나 시간 값을 저장하기 위해 Date, DateTime, TimeStamp 등의 데이터타입을 사용할 수 있다. Varchar2와 같은 문자열로 일자를 저장하지 않고 DateE 타입에 일자를 저장하는 경우 시간적 순서에 따라 값을 정렬하거나 시간값을 구할 때 편리하다. 전자기록물의 진본성을 중시하는 경우 생산 시점이나 이관시점 등을 정밀하게 관리하기 위해 TimeStamp 타입을 사용하려는 추세에 있다는 점도 고려할 필요가 있다.

3) CAMS 데이터베이스의 과제

차세대연구 수행과정에서 살펴본 CAMS 데이터베이스의 문제점을 해결하기 위해 필요한 향후 개선과제를 종합해 보면 다음과 같다.

첫째, 스키마 검토가 필요하다. 기록물철 테이블과 기록물건 테이블에 대한 분석을 통해 테이블들이 심각하게 비정규화되어 있음을 알 수 있었다. CAMS 데이터베이스 내에 기록물의 목록을 담고 있는 테이블이 여러 개 더 존재하는 것으로 파악되는데, 이 테이블들 또한 유사한 상태인 것으로 보여진다. 따라서, CAMS 데이터베이스 전체적으로 정규화의 관점에서 스키마를 검토해볼 필요가 있다. 정규화는 데이터의 무결성과 일관성을 확보하기 위해 필요한 것이다. 성능을 제고하기 위해 일부 반정규화를 고려할 필요도 있으나 이것은 먼저 정규화된 스키마를 완성하고 난 후에 선별적으로 검토할 사안이다. 기존의 스키마를 최소한으로 변경하면서 개선해 갈 것인지, 완전히 새로운 스키마를 개발할 것인지 전략적 선택이 필요한 시점이다.

둘째, 성능 튜닝이 필요하다. 차세대연구 과정에서 CAMS의 기록물 검색 속도가 현저히 느리다는 보고가 있었는데, 현재 전자기록물이 본격적으로 이관되어 오지도 않은 상태에서 검색 속도에 문제가 있다는

것은 시스템의 성능 튜닝이 제대로 되어 있지 않다는 것을 의미한다. 현재 CAMS에서는 검색 속도의 문제를 해결하기 위해 핵심 데이터만 선정한 각종 뷰를 생성하고 이에 대한 인덱스를 만들어 사용하고 있다. 데이터 조회 시 성능향상을 위한 적절한 인덱스 생성 전략이 부재한 상태라 볼 수 있다. 데이터베이스는 살아있는 생물체처럼 변화하므로 수시로 최적화를 해주어야 제 성능을 발휘할 수 있다. 정보시스템에 대한 기능변경이 있을 때마다 데이터베이스 스키마는 재검토되어야 하며 업무프로세스와 함께 최적화되어야 한다. 그러나, 현재 살펴본 바에 의하면 CAMS에서 사용하는 데이터베이스관리시스템에 대해서는 일반적 유지보수 계약만 되어 있을 뿐 데이터베이스 스키마에 대한 평가나 정기적인 튜닝이 실시된 바가 전혀 없었다. 향후에는 CAMS 데이터베이스 성능 튜닝을 위한 예산을 필수적으로 책정하도록 해야할 것이다.

셋째, 데이터베이스 스키마에 대한 문서를 생성하고 관리해야 한다. 현재 스키마에 대한 문서가 제대로 남아있지 않아 왜 그런 칼럼들로 구성되었으며 각 칼럼의 용도는 무엇인지 전반에 대해 제대로 파악하는 것이 불가능하다. 문서가 없고, 이력을 꿰뚫고 있는 업무담당자가 없다 보니 이런 일도 발생한다. 얼마 전 공개재분류 사업 결과를 CAMS에 입력했는데, 기록물 활용 화면에서 재분류한 결과값이 보이질 않고 이전 값이 그대로 보이는 문제가 발생했다는 것이다. 문제의 원인을 추적해보니 공개재분류 사업 결과를 입력하는 엔티티가 존재하고 있었는데 외주 사업자들이 그것을 모르고 별도의 엔티티를 추가로 만들어 재분류값을 입력했고, 기존의 활용 화면에서는 기존의 엔티티를 조회하도록 되어 있어 그리 되었다는 것이다. 외주 사업자들이 신규 기능을 추가할 때 기존의 스키마에 대해 이해를 하고 적합한 테이블과 칼럼을 이용하도록 해주어야 하는데 이를 안내해줄 최신의 문서도 없고 업무담당자도 없다는 것이다. 한편, 국가기록원이 EA(Enterprise Architecture) 프로젝트의 결과물을 잘 만들어 보유하고 있었다면 이러한 문제는 발생할

수 없을 것이다. 현재 CAMS의 문제는 복잡한 시스템 개발 과정과 연관된다고 볼 수 있다. CAMS 데이터베이스에 얽힌 이러한 문제들이 어디에서 기인하는 것인지를 밝히는 것은 문제 해결을 모색하는 방향에 실마리를 제공할 것이다. 차세대연구에서도 일부 스키마에 대한 정보를 문서에서 얻은 것이 아니라 데이터베이스에 직접 접속하여 역공학(Reverse Engineering)의 방법으로 얻어낼 수 있었다. 하루 속히 현재의 스키마에 대한 현황을 파악하여 문서화함으로써 통일된 이해가 가능하도록 해야 할 것이다.

3. 영구기록물관리 데이터베이스 스키마 요건 확인

2장에서 기록물관리의 기본 단위인 첩과 건의 테이블 설계를 검토하면서 2015년을 준비하기 위해 CAMS 데이터베이스 스키마 전반에 대해 검토를 할 필요가 있음을 인식하였다. 이에 3장에서는 CAMS 데이터베이스 스키마가 준수해야 하는 요건을 검토해 보고자 한다.

1) 데이터베이스 스키마 설계 고려사항

먼저 개념 스키마에서 고려할 사항은 다음과 같다. 첫째, 기본 속성 정보를 중심으로 하여 첩과 건 엔티티를 정의하고 나머지 속성들은 유형별 엔티티로 분리해야 한다. 둘째, 기록물 생애주기에 따른 관리업무별로 이력 엔티티를 정의해야 한다. 셋째, 감사증적, 접근제어목록(ACL, Access Control List) 등 전자기록 진본성 확인 및 유지를 위한 정보를 엔티티로 정의해야 한다. 넷째, 분류체계, 처분일정, 생산자 전거 등과 같은 기준정보나 참조정보를 위한 엔티티를 정의해야 한다.

다음으로 논리스키마에서 고려할 사항은 다음과 같다. 첫째, 기록물철과 기록물건의 양이 방대하므로 연도별 파티션 테이블로 구성하는 것이 유리하다. 연도별, 기관별 파티션 테이블 구성도 가능하다. 둘째, 기록물철과 그 하위의 기록물건은 함께 조회되고 디스플레이되는 경우가 많으므로 클러스터 테이블로 정의하는 게 유리하다.

CAMS 데이터베이스 스키마 설계 시 기본 원칙은 모든 유형의 전자기록을 통합한 데이터베이스를 설계해야 한다는 것이다. 공공기관에서 2003년 말까지 전자결재시스템을 이용하여 생산한 전자문서와 2004년 이후 신전자문서시스템을 이용하여 생산한 전자문서는 여러 측면에서 차이를 갖고 있다. 2003년까지 구전자문서시스템에서 생산된 전자기록은 기록물분류체계가 적용되어 있지 않으며 철·건 구조를 갖추고 있지 않다. 2004년부터 신전자문서시스템에서 생산된 전자기록은 표준 규격에 맞춰져 있다. 따라서 메타데이터의 종류에 상당한 차이를 보이고 있다. CAMS에는 비표준 규격 전자문서와 표준 규격 전자문서 모두가 기록으로 관리될 수 있어야 한다. 전자문서류의 전자기록 이외에도 웹문서, 데이터세트, 간행물, 시청각, 이메일 등 다양한 유형의 전자기록은 물론이고 비전자기록까지 모두 통합해 관리해야 한다. 향후 복합전자기록물이나 SNS기록물과 같이 특수한 유형의 전자기록물이 지속적으로 등장할 것이며 CAMS는 이러한 신규 기록유형을 추가하여 관리할 수 있는 융통성있는 구조라야 한다. 따라서, 향후 기록물 데이터베이스는 모든 기록유형에 공통적인 메타데이터를 중심으로 하는 마스터 엔티티와 기록의 유형별 서브엔티티를 연결하는 방식으로 구성해야 한다.

그 밖에도 CAMS 데이터베이스 설계 시 특별히 고려해야 할 사항을 정리해 보면 다음과 같다. 첫째, 메타데이터의 무결성 확인을 위해 메타데이터 값에 대한 체크섬 속성을 추가하여 관리하는 것이 필요하다. 둘째, 기록물철, 기록물건, 디지털 컴포넌트의 ID는 UCI로 처리하도록 해야 한다. 셋째, 개인정보보호 등 관련 법규를 준수하기 위해 주민번호

호와 같은 메타데이터에 대해서는 칼럼 단위 암호화 보안을 적용하도록 해야 한다. 넷째, 기록관리 업무이력 정보 중 실행일자는 시간대(Time Zone)를 표기한 타임스탬프(TimeStamp) 데이터 타입으로 지정해 주어 실행 시점을 정확히 특정해 주어야 한다. 다섯째, 기록정보 간의 관련성을 데이터베이스 제약조건(Constraints)으로 정의하여 데이터의 무결성을 강화해야 한다.

2) 데이터베이스 관리 고려사항

CAMS 데이터베이스는 관리적 측면에서 다음과 같은 특징을 갖는다. 첫째, 데이터 관리자 역할이 중요하므로 국가기록원 내에 담당자를 두어 일관된 데이터 정책을 수립하고 집행하도록 해야 한다. 예를 들어, 현재 표준RMS에서는 기록물건에 대한 공개재분류로 인해 공개 속성 값이 변경된 후 이에 대한 감사증적 정보가 발생하지 않고 있다. 또한, 보존포맷 재생성도 하지 않는다. 따라서 최신의 메타데이터와 NEO 내부의 메타데이터 값에 격차가 발생할 수 있을 뿐만 아니라 NEO 이관 시 함께 이관되어야 하는 감사증적 정보조차 없으므로 메타데이터를 따로 보낼 수밖에 없는 구조이다. 물론, 현재로는 어느 것이 최신이며 정확한 값인지에 대한 검증도 불가능하다. 이관 직전에 보존포맷 재생성을 의무화하고, 재생성 이후에는 메타데이터에 대한 변경을 시스템적으로 통제한다면 NEO만 이관해도 무방할 것이다. 그러나, 지금 현재 표준RMS와 CAMS의 운영 실태를 감안한다면 기록의 신뢰성을 보장하기 어렵다고 판단된다. 시스템의 기능을 이용하지 않고 유지보수 담당자에게 의뢰하여 데이터베이스에 직접 접근하여 데이터를 수정하는 경우도 존재하기 때문이다. 이러한 문제의 발생을 원천적으로 막기 위해서는 통제권한을 보유한 데이터 관리자를 명시적으로 임명하여 활동하도록 하는 것이 필요하다.

둘째, 국가기록원의 EA를 제대로 구축하는 것이 필요하다. 국가기록원 EA는 공공기록물 영구보존을 위한 업무 및 정보시스템의 청사진을 담고 있어야 한다. 또한, 현행화된 업무 및 정보시스템 현황을 정확히 담고 있어야 한다. EA의 데이터 아키텍처(Data Architecture)에는 CAMS의 데이터베이스 스키마가 상세히 기록되어 있어야 하며, 기술 아키텍처(Technology Architecture)에는 데이터베이스관리시스템 현황 정보가 기록되어 있어야 한다. 현재 국가기록원의 EA는 CAMS 데이터에 관한 문서를 새로 만들어 등록하는 수준에 머물러 있다. EA 구축을 통해서 데이터베이스 스키마의 통합성과 일관성을 확보할 수 있다.

셋째, DBMS 유지보수에 정기적인 데이터베이스 성능 진단 및 튜닝 서비스를 필수항목으로 추가해야 한다. CAMS에는 매년 대량의 데이터가 추가될 것이므로 입수를 마친 후 성능을 진단하고 튜닝하는 것을 정례화해야 할 것이다. 기록물에 대한 메타데이터가 대량으로 변경되는 업무가 수행된 후에는 변경 값과 연관된 인덱스를 재생성해야 한다. 클러스터, 파티션 뷰 등을 이용하여 성능을 향상시키는 것은 데이터베이스의 논리 설계에 해당하는 것이므로 현재의 CAMS 데이터베이스를 새롭게 구축하게 될 때 반영할 수 있을 것이다.

넷째, 새로운 CAMS 데이터베이스 설계 후 기존 데이터를 마이그레이션하기 위한 전략 수립이 필요하다. 새로운 데이터베이스를 구축하여 기존 데이터베이스의 데이터를 마이그레이션하는 작업은 신중하게 실행되어야 한다. 이 과정에서 기록의 진본성, 신뢰성, 무결성, 이용가능성이 훼손될 가능성이 높기 때문이다. 마이그레이션 과정에서 수정·보완해야 하는 데이터가 있을 경우 누가 어떤 권한으로 데이터를 정정할 것인지 절차와 방법을 정해야 한다. 또한, 마이그레이션 과정에서 변경된 값에 대해 어떻게 감사증적을 남길 것인지 사전에 정의해야 하며, 데이터의 정정이 발생할 경우 디지털 컴포넌트나 NEO에 영향을 줄 수 있는지 파악하고 조치를 취할 수 있어야 한다. 현재 기록물철과 기록물

건을 설명하는 속성항목들이 정규화되지 못하고 한 두 개의 테이블에
 몰려있어 마이그레이션을 하기 전에 데이터의 품질에 대해 다음과 같
 은 검토가 선행되어야 한다. 첫째, 기록의 유형별 필수 메타데이터가
 있는지 확인해야한다. 없는 경우 값을 복원하는 작업을 실시해야 한다.
 둘째, 유사한 데이터타입의 메타데이터들 간의 데이터 타입이 불일치하
 는 경우 타입 변환을 실시하여 하나로 통일시켜야 한다. 셋째, 데이터
 값이 정당한 범위 내에 속해 있는지 확인해야 한다. 값의 범위에 오류
 가 있는 경우 확인 후 수정해야 한다. 넷째, 각종 코드나 참조 ID값들은
 새로 재편된 체계에 맞춰 새로운 값을 지정해 주어야 한다.

다섯째, 새로운 데이터베이스가 CAMS 응용에 미치는 영향을 분석하
 고 그에 따른 개선 전략을 수립해야 한다. 새로운 스키마를 적용했을
 때는 CAMS 응용에 큰 변화가 예상된다. 스키마 변경 후 기존의 응용을
 그대로 사용할 수 있는 범위와 가능성을 확인해야 한다. 이를 위해 스
 키마 변경 시 응용에 대한 영향분석(Impact Analysis)을 할 수 있는 도구
 가 필요하다. 영향분석 결과 새로운 스키마에 맞춰 변경, 혹은 재개발
 해야 하는 응용과 뷰를 이용하여 재사용이 가능한 응용을 가려낸다. 응
 용개발에 투여할 수 있는 예산과 소요기간 등을 고려하여 데이터베이
 스 스키마 개선 전략을 수립해야 한다.

3) 메타데이터 표준 관점에서의 고려사항

우리나라 공공기관의 현용·준현용 기록물은 국가기록원이 제정한
 기록관리 메타데이터 표준에 따라 관리되어야 한다.¹⁰⁾ 이들 기록물이
 CAMS로 이관되어 올 것이므로 CAMS 데이터베이스 스키마도 입수되는
 기록물의 속성정보를 포괄할 수 있도록 설계되어야 한다. 즉, 현재 기

10) 차세대연구에서는 2007년 12월 제정된 “기록관리 메타데이터 표준-현용·준현
 용 기록물 용-”(NAK/S 8:2007(v1.0))을 참조하였다.

록관리 메타데이터 표준에 정의된 21개 상위요소들이 스키마에 포함되어 있어야 한다. 21개 상위요소는 1) 행위자(Agent), 2) Mandate(관련법규), 3) Identifier(고유식별자), 4) Title(표제), 5) Description(기술), 6) Type(유형), 7) Format(포맷), 8) Extent(크기), 9) Classification(분류), 10) Subject(주제어), 11) Date(일시), 12) Business History(생산이력), 13) Preservation(보존), 14) Location(위치), 15) Retention(보존기간), 16) Rights(권한), 17) Management History(관력이력), 18) Use History(이용이력), 19) Relation(관계), 20) Aggregation Level(기록계층), 21) Language(언어) 등이다. 이 요소들이 CAMS 데이터베이스 스키마에 어떻게 구현되어 있는지를 파악하기 위해서는 스키마 전반에 관한 정보와 저장된 데이터에 관한 정보가 필요하다. 이 논문에서는 연구자가 접근할 수 있었던 범위에서 21개 요소별로 현재의 스키마 현황과 향후 개선 방향을 제시하고자 한다.

먼저, 1) 행위자(Agent) 요소이다. 각 목록 테이블과 별도로 Agent 테이블을 유지하며, 목록 테이블의 키(표준 요소로 정의된 Identifier와 관련된됨)를 외래키로 갖고, 일련번호를 두어 같은 목록에 대한 여러 행위자의 행위에 순서 관계를 갖게 해야 한다. 하위요소로 정의된 Corporate Name, Section Name, Position Name는 각각 표준에 같이 정의된 Corporate Code, Section Code, Position Code로 대체해야 하며, 각각의 코드 테이블을 만들어 관리해야 한다. 국가기록원 CAMS 시스템에는 현재 Position Code는 CM_CLASS 테이블에 직급코드로, Section Code는 SM_NARS_STRUC 테이블에 조직코드로 코드화하여 관리 중이다. 사용자 테이블에서는 직급과 조직이 각각 CLS_COD, GRP_STRUC_COD 칼럼으로 코드정보를 갖고 있다. 하위요소 중 행위자의 Email, Address는 행위자에 대한 개인정보로 전체 행위자, 즉 사용자와 관리자들에 대한 메타데이터를 관리하는 별도의 테이블에 저장해야 한다. 국가기록원 CAMS 시스템에 SM_USER 테이블에 해당하며 현재 스키마는 [표 8]과 같다.

[표 8] SM_USER 테이블의 칼럼 목록

칼럼	데이터타입	설명
USR_ID	Varchar2(20)	사용자ID
USR_PASSWD	Varchar2(60)	사용자암호
MOV_PASSWD	Varchar2(40)	이전암호
USR_NM	Varchar2(60)	사용자명
POS_GRD_COD	Varchar2(3)	소속계급코드
CLS_COD	Varchar2(5)	직급코드
USR_CITNID	Varchar2(13)	사용자주민등록번호
USR_SCT	Varchar2(2)	사용자구분
ORGL_APCH_YN	Varchar2(1)	원문접근여부
NOPN_APCH_AUTH	Varchar2(2)	비공개접근권한
GRP_STRUC_COD	Varchar2(6)	그룹조직코드
ARC_EQU_COD	Varchar2(2)	보존시설코드
EXEC_BS	Varchar2(120)	수행업무
NEW_CHG_YN	Varchar2(1)	
REQ_SCT_COD	Varchar2(2)	
REQ_CONT	Varchar2(250)	
APPR_YN	Varchar2(1)	승인여부
APPRR_ID	Varchar2(20)	승인자ID
NEW_OLD_YN	Varchar2(1)	신규여부
REMARK	Varchar2(500)	비고
FST_REGR_ID	Varchar2(20)	최초등록자ID
FST_REG_DT	DATE(7)	최초등록일시
LST_UPDR_ID	Varchar2(20)	최종수정자ID
LST_UPD_DT	DATE(7)	최종수정일시
USE_YN	Varchar2(1)	사용여부(Y: 사용, N: 미사용)
APCH_AUTH	Varchar2(1)	목록접근권한(CM070)
BS_AUTH_COD	Varchar2(4)	업무권한코드
SECRET_ORGAPCH_YN	Varchar2(1)	기관자체(폐쇄기관) 전체기관명 접근권한여부

ORG_CODE	Varchar2(7)	
EMAIL	Varchar2(30)	사용자 E-Mail
PHONE	Varchar2(20)	사용자 전화번호
REQ_NO	Varchar2(60)	접근권한 접수번호(문서번호)

기록관리 표준 메타데이터와 CAMS 시스템 데이터를 비교해보면 [표 9]와 같다.

[표 9] 메타데이터 표준 요소별 CAMS 데이터 항목 비교

메타데이터 표준 요소	CAMS 데이터 항목
Corporate Code	없음
Section Code	GRP_STRUC_COD, 코드는 SM_NARS_STRUC 테이블에서 관리 중
Position Code	CLS_COD, 코드는 CM_CLASS 테이블에서 관리 중
Email	EMAIL
Address	없음

2) Mandate(관련법규) 요소의 경우에는 하위요소들을 구성요소로 갖는 별도의 테이블로 관리해야 한다. 각 목록 테이블의 키(표준 요소로 정의된 Identifier와 관련됨)를 외래키로 갖고, 일련번호를 두어 기본목록 1건당 여러 법규가 연결되도록 해야 한다. 상위요소 중 Business history, Rights, Management history, Use history에 해당하는 칼럼들과 결합하여 사용되며, 그 구분을 위한 칼럼(가칭 Related data)이 추가로 필요하다.

3) Identifier(고유식별자) 요소는 기록물의 고유한 식별자로 기록물철, 기록물건, 디지털 컴포넌트 등의 정보를 저장하는 테이블에서 키로 사용되어야 한다. 기록관리 표준의 각 메타데이터가 별도의 테이블로 관리될 경우 이를 연결할 수 있는 외래키로 사용하게 된다.

4) Title(표제) 요소는 기록물에 주어진 제목으로 기록물철, 기록물건 및 문서 등에 대한 목록테이블과 별도의 테이블로 관리해야 한다. 각

목록 테이블의 키(표준 요소로 정의된 Identifier와 관련됨)를 외래키로 갖고, 일련번호를 두어 기본목록 1건당 여러 표제가 연결되도록 해야 한다. 표제 테이블에는 표제의 종류와 표제의 내용 칼럼이 정의되어 한다. 표제 메타데이터의 3개 하위요소 Title Words, Alternative Title Type 및 Alternative Title Words 중 Title Words는 표제의 종류를 '기본'으로 하여 저장하고, 대체표제들도 종류를 정하여 저장하도록 설계할 수 있다.

5) Description(기술) 요소는 표준에 정의된 Description의 하위요소들인 Description Type과 Description Text를 구성요소로 갖는 별도의 테이블로 관리해야 한다. 각 목록 테이블의 키(표준 요소로 정의된 Identifier와 관련됨)를 외래키로 갖고, 일련번호를 두어 기본목록 1건당 여러 Description들이 연결되도록 해야 한다. 영구보존 기록의 경우 장기간의 관리과정에서 기술할 내용이 여러 종류가 발생할 수 있기 때문에 다양하고 풍성한 기술을 가능하게 하기 위해서는 기술의 종류, 기술의 내용 칼럼을 가진 테이블이 필요하다.

6) Type(유형) 요소는 기록물의 유형에 대한 정보로 기록물철과 기록물건에 대한 목록 테이블에 2개의 하위요소 Record Type과 Alternative Record Type을 가진다. 만약 기록물철과 기록물건의 유형이 다중값을 가질 수 있다면 유형을 별도의 테이블로 관리해야 한다. 또한, 기록의 유형 정보는 별도의 코드값으로 관리해야 한다.

7) Format(포맷) 요소는 기록의 매체나 데이터 포맷과 물리적인 저장 매체의 종류로 문서(본문/첨부) 목록 테이블에 하위요소인 Media Format, Data Format 및 Medium을 가진다. 만약 복합기록물이 등장하여 하나의 문서에 다중 포맷이 적용되게 된다면 포맷 테이블을 별도로 관리해야 한다. 각 하위요소는 코드화가 필요하며 별도의 테이블로 코드정보를 유지해야 한다.

8) Extent(크기) 요소는 기록물의 물리적인 크기나 용량으로 각 목록 테이블에 6개의 하위요소 Size, Count, Publication Length, Page, Speed 및

Audiovisual Length를 가진다. 포맷에 따라 크기를 재는 단위가 달라지고 크기값이 단일 값이 아닐 수도 있으므로 크기 테이블을 별도로 관리할 필요가 있다. 포맷 별로 크기의 요소를 정의하여 코드 테이블로 관리해야 한다.

9) Classification(분류) 요소는 기록을 생산한 업무기능으로 하위요소인 Classification Type, Classification ID, Classification Name, Classification Level을 구성요소로 갖는 별도의 테이블로 관리해야 한다. 분류체계 테이블이 독립되어야 한다는 것은 이미 2장의 정규화 과정에서도 언급한 바 있다. 분류체계는 자체가 계층구조를 갖는 경우가 많으므로 스스로 참조를 하는 구조로 만들어진다. 기록물철 목록 테이블에서 분류체계 계층의 최하위 Classification ID를 외래키로 참조하는 방식으로 분류값을 연결할 수 있다. 하위요소 Classification Name은 Classification ID를 이용하여 분류체계 테이블로부터 가져올 수 있다.

10) Subject(주제어) 요소는 기록에 대한 색인어로 기록물철, 기록물건 목록 테이블에 각각 존재하며, 여러 주제어를 구분자를 두어 포함할 수 있다. 주제어 자체가 시소러스로 구성되는 경우에는 주제어 테이블이 독립적으로 관리되면서 기록물에서는 관련 주제어를 참조하는 방식으로 연결시킬 수 있다. 향후 동일 주제의 기록물들을 용이하게 검색하기 위해 고려해볼 만한 선택사항이다.

11) Date(일시) 요소는 기록의 생산, 종료와 관련된 행위가 발생한 일시로 기록물철과 기록물건에 대한 목록 테이블에 4개의 하위요소 Date Time Created, Date Time Receipted, Date Time Opened, Date Time Closed를 가진다. 관리과정에 관련된 행위가 발생한 일시까지 포함한다면 별도의 테이블로 관리해야 하지만, 생산 및 종료 행위 일시로만 국한한다면 기록물철과 기록물건의 목록 테이블 칼럼으로 포함하는 것이 효과적이다.

12) Business History(생산이력) 요소는 생산부서간 인수·인계 사항을

포함하며 기록의 생산과정에서 발생하는 중요 정보이다. 하위요소들은 모두 목록 테이블이 아닌 별도의 테이블로 관리해야 한다. 해당 테이블은 목록테이블의 키를 외래키로 가지고 일련번호를 두어 목록 1건당 여러 Business history들이 연결되도록 해야 한다. Transaction Agent는 Agent Type과 같은 코드를 공유하도록 하며, Addition Number Type에 대해 코드화해야 한다. Taking Over Agent 및 Under Taking Agent는 해당 기관, 처리과, 개인에 대한 id를 갖고 있고, 표준에 명시되지 않았으나 두가지 모두 기관, 처리과, 개인에 대한 타입종류를 명시하는 별도의 칼럼(예시:Taking Over Agent Type, Under Taking Agent Type)이 필요하다. Transition Mandate는 Mandate(관련법규) 부분을 참조하도록 해야 한다.

13) Preservation(보존) 요소는 기록관리시스템으로 기록물이 인수된 이후 행해진 모든 보존처리 행위에 대한 설명하는 항목이다. 하위요소들은 기록물의 목록 테이블이 아닌 별도의 테이블로 관리해야 한다. 보존 테이블은 목록 테이블의 키를 외래키로 가지고 일련번호를 두어 목록 1건당 여러 보존행위들이 연결되도록 해야 한다. Action Agent는 행위자의 id를 저장하며 상위 표준요소인 Agent(행위자)에서 본 바와 같이 전체 행위자(사용자, 관리자)들에 대한 데이터를 관리하는 별도의 테이블에 대한 키(행위자 id)를 외래키로 참조해야 한다.

14) Location(위치) 요소는 전자와 비전자기록 모두의 현재 소장위치에 대한 정보이다. 문서 하나를 디지털화한 경우, 문서의 서고 저장 위치와 디지털 파일의 시스템 저장 위치를 동시에 가지고 있어야 하므로 하위요소인 Current Location은 위치 테이블로 별도 관리해야 한다. 혹은, 위치의 변경 이력을 관리하고자 한다면 위치 테이블에 현재의 위치뿐만 아니라 위치의 변동이력을 관리하기 위한 칼럼을 추가해야 한다. 서고의 상세정보를 의미하는 Storage Details는 서고에 ID와 명칭이 부여되어 있고 서고정보 테이블을 따로 관리하는 경우 서고의 ID를 참조하

는 것으로 변경할 수 있다. RMS ID 역시 기관들의 RMS에 관한 정보 테이블이 따로 관리되고 있다는 전제하에 ID를 참조하는 것으로 표준에 제시되어 있다.

15) Retention(보존기간) 요소는 기록이 폐기되기 전까지의 보유기간과 책정한 사유이다. 기록물철 목록 테이블과 별도의 처분지침 테이블, 처분일정 테이블이 존재해야 하며, 하위요소인 Retention Period, Retention Reason 처분 관련 테이블에 수용되어야 한다. 기록물철 목록 테이블에서 처분지침을 참조해야 하며, 처분일정 테이블에서 기록물철을 참조해야 한다.

16) Rights(권한) 요소는 기록의 이용 및 접근을 관리하고 통제하기 위한 정보로 기록물건 목록 테이블에 10개의 하위요소들 중 Security Level, Security Reason, Protection Period, Security Mandate에 해당하는 칼럼과, Internal Access Control에 해당하는 칼럼, External Access Control, External Access Authority, Limited Contents, External Access Due 및 External Access Mandate 에 해당하는 칼럼들이 존재해야 한다. 비밀이나 비공개로 분류한 정보는 시간이나 상황이 변하면서 변경된다. 따라서, 변경된 이력을 관리하고자 한다면 하위요소의 그룹별로 별도의 테이블을 만들고 특정 기간 동안 분류된 상황을 저장하도록 해야 한다. Security Mandate와 External Access Mandate는 Mandate(관련법규)를 참조하도록 해야 한다.

17) Management History(관리어력) 요소는 기록물이 이관되면서부터 폐기될 때까지 단계별 흐름에 따라 기록의 상태에 영향을 미치는 모든 관리행위를 나타낸다. 하나의 기록물철, 기록물건, 문서별로 다중의 관리행위가 행해지므로 하위요소들은 모두 목록 테이블이 아닌 별도의 테이블로 관리해야 한다. 해당 테이블은 목록테이블의 키를 외래키로 가지고 일련번호를 두어 목록 1건당 여러 관리어력들이 연결되도록 해야 한다. 하위요소인 Event Type과 Event Description은 코드테이블로 분리해야 하고, 관리행위의 정밀한 시각이 향후 진본성, 무결성 검증에

핵심이 될 수 있다면 Event Date Time은 TimeStamp 데이터타입으로 정의해야 한다. Event Agent는 상위 표준요소인 Agent(행위자)에서 본 바와 같이 전체 행위자(사용자, 관리자)들에 대한 메타데이터를 관리하는 별도의 테이블에 대한 키(행위자 id)를 외래키로 참조해야 한다. Event Mandate는 Mandate(관련법규)를 부분 참조하도록 해야 한다.

18) Use History(이용이력) 요소는 기록이 생산, 등록되면서부터 폐기, 보존되는 동안 발생한 모든 접근과 이용에 관한 정보를 나타낸다. 하위 요소들은 모두 목록 테이블이 아닌 별도의 테이블로 관리해야 한다. 해당 테이블은 목록 테이블의 키를 외래키로 하여 일련번호를 두어 목록 1건당 여러 이용이력들이 연결되도록 해야 한다. Use Agent는 행위자의 id를 저장하며 상위 표준요소인 Agent(행위자)에서 본 바와 같이 전체 행위자(사용자, 관리자)들에 대한 메타데이터를 관리하는 별도의 테이블에 대한 키(행위자 id)를 외래키로 가지게 된다. Use Type, Use Description은 이용유형을 코드화하여 별도 테이블로 관리해야 하고 이용이력 테이블에서 이를 참조해야 한다.

19) Relation(관계) 요소는 기록물 간의 수평적인 관계, 수직적인 관계 및 기록물과 다른 정보원들 상의 관계를 나타낸다. 하위요소들은 모두 목록 테이블이 아닌 별도의 테이블로 관리해야 한다. 해당 테이블은 목록테이블의 키를 외래키로 가지고 일련번호를 두어 목록 1건당 여러 관계들이 연결되도록 해야 한다. Relation Type은 코드화하여 코드테이블로 분리해야 하며, Relation Description 항목이 관계의 유형별로 상세 항목으로 분화될 필요가 있다면 관계테이블이 하나가 아니라 관계유형별 테이블로 분화될 수도 있다. 알려진 코드종류는 포함, 첨부, 대체, 참조, 관련과제, 관련 건, 기타 등이 존재하는데 포함관계의 경우는 기록계층에 관련된 것은 목록 테이블에서 계층의 목록 테이블 ID를 참조하는 칼럼을 포함하도록 하는 것도 생각해볼 수 있다. 왜냐하면, 기록물건과 기록물철의 관계처럼 기록계층 정보는 대부분의 기록물에 필수

사항이며, 기록물철, 기록물건, 문서 등의 계층적 기록정보는 함께 보는 경우가 많으므로 관계테이블을 만들어 추가로 조인하기 보다는 기록물 목록 테이블끼리만 조인하도록 하는 것이 유리하기 때문이다.

20) Aggregation Level(기록계층)은 기록의 계층구조에서 해당 기록을 검색할 수 있게 하기 위한 요소이다. 관계 요소 설명에서 확인한 바와 같이 기록물철 목록 테이블, 기록물건 목록 테이블, 문서 목록 테이블 간에 포함관계를 특화하여 기록계층을 표현할 수 있다. 각 목록 테이블에 포함되는 상위 계층의 ID를 참조하는 칼럼을 추가해야 한다.

21) Language(언어)는 기록의 내용을 작성한 언어를 의미한다. 기록물건이나 문서 등이 2개 이상의 언어로 작성되었을 경우 기록물언어테이블을 별도로 관리해야 한다. 이때, 언어의 유형을 코드테이블로 관리해야 한다.

CAMS에서는 이관받은 기록물이 표준 메타데이터 항목들을 가지고 입수된다는 가정하에 위에서 제시한 테이블별로 나누어 정보를 저장할 수 있어야 한다. 위에서 살펴본 바와 같이 메타데이터 요소들을 고려했을 때 데이터베이스 스키마 상에는 다수의 코드테이블과 마스터테이블, 이력테이블들이 설계되어야 한다. 기록물 이관시 기록물의 목록 테이블뿐만 아니라 관련된 테이블들을 적합하게 조인하여 필요한 데이터들을 함께 이관하는 것이 중요하며 이를 지원해줄 고도의 기록물 내보내기(Export) 도구가 필요하다.

4. 영구기록물관리 데이터베이스 개념스키마 예시

이 장에서는 영구기록물의 생애주기를 입수단, 보존단, 제공단의 3단계로 나누어 개념스키마의 사례를 제시하고 논리스키마의 개발 방향을

제시하고자 한다. 개념스키마는 3장에서 살펴본 정규화 방향, 표준 메타데이터 항목의 설계 방향에 기록물 생애주기별 특징을 감안하여 작성해 본 것이다. 각 생애주기 단계별로 필요한 전체의 개념스키마를 제시하는 것이 이 논문의 연구 범위는 아니며, 다만 앞에서 제시한 개념스키마의 개발 방향을 수용한 형태를 예시하여 향후 데이터베이스 설계시 참고가 되고자 하는 것이다. 엔티티와 핵심 속성 위주의 ERD (Entity-Relationship Diagram) 형태로 스키마의 일부를 제시하고자 한다.

1) 입수단 개념스키마

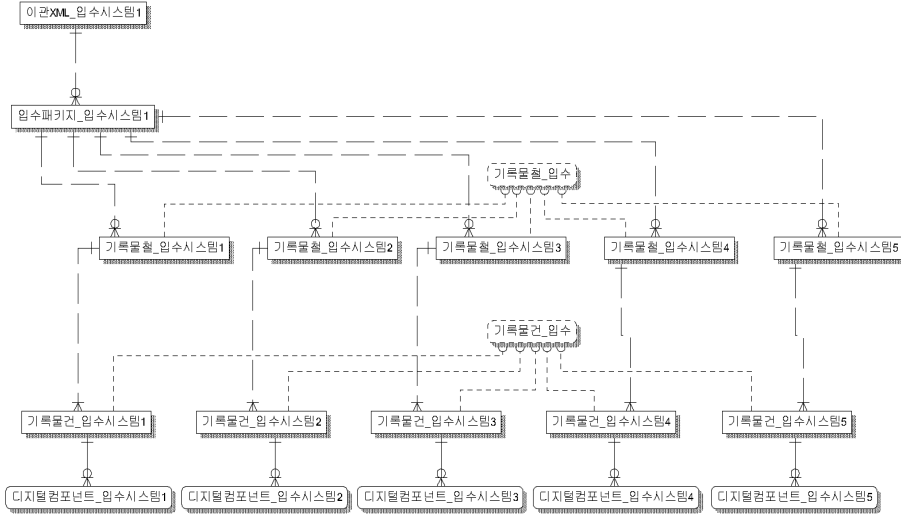
입수단 시스템은 기록물을 입수처리시스템 별로 나누어 관리하고, 전체를 다시 뷰로 통합하는 방식의 모형이 필요하다. 즉, 각 입수처리 시스템에서는 해당 시스템에서 처리할 기록물만 관리하도록 하고, 삽입, 삭제, 갱신, 검색 등 테이블 접근에 대한 성능을 향상시키며, 통합 뷰를 통한 전체 데이터에 대한 일괄 관리를 효율적으로 할 수 있도록 모델링해야 한다.

[그림 3]에서 보는 바와 같이 입수처리시스템이 5개인 경우라면, 각 시스템에서 기록철과 건, 컴포넌트를 분담하여 입수처리하고 하나의 이관입수시스템에서 그 결과를 취합할 수 있도록 설계해야 한다.¹¹⁾

11) 2015년 이후 대량의 전자기록물이 국가기록원에 입수될 때 이를 처리하는 서버를 5개 정도를 두어 부하 분산을 하도록 제안한 내용에 근거하고 있다. 임진희·이대욱, 위의 책, 206-210쪽 참조.

2) 보존단 개념스키마

[그림 3] 입수단의 주요 엔티티 관계 모형



기록물은 기본적으로 기록물철과 기록물건으로 나뉘지며, 각각 철 메타데이터를 관리하는 테이블과 그 하위에 건 메타데이터를 관리하는 테이블의 이중 구조로 관리해야 한다. 건 메타데이터는 상위 철 메타데이터의 ID를 외래키로 가지도록 한다. 기록물 유형에 따라 메타데이터를 분리 관리하여 처리 및 접근 성능 향상시키기 위해 다음과 같은 원칙을 가지고 [그림 4]와 같이 모델링해 보았다. 첫째, 목록 테이블에는 핵심적인 메타데이터만 관리하고, 둘째, 각 기록물 유형별로 상이한 메타데이터를 관리하며, 셋째, 각 기록물 유형별로 상이한 메타데이터를 분리하고, 넷째, 기록물 관리 업무적인 부분에 대한 메타데이터를 분리하도록 한다.

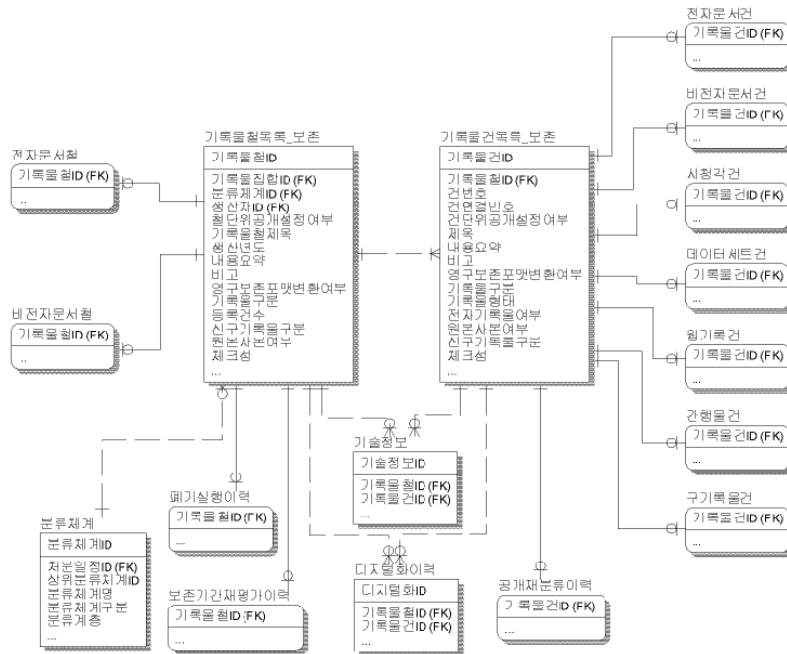
다음으로 기록물철/기록물건, 분류체계 및 기록물집합(Aggregation) 관

계는 다음과 같은 방향으로 [그림 5]와 같이 모델링하였다. 첫째, 기록물철 목록 상위 레벨로 기록물집합 메타데이터를 다단계로 관리하고, 둘째, 분류체계는 기록물집합 및 기록물철과 연계되며 처분 관련 데이터를 보유하도록 하며, 셋째, 보유기록 메타데이터는 기록물철 목록에 연계되며 보유원인사건 메타데이터에 연결하도록 했고, 넷째, 필수기록 목록은 기록물건 목록과 연계하도록 하였다.

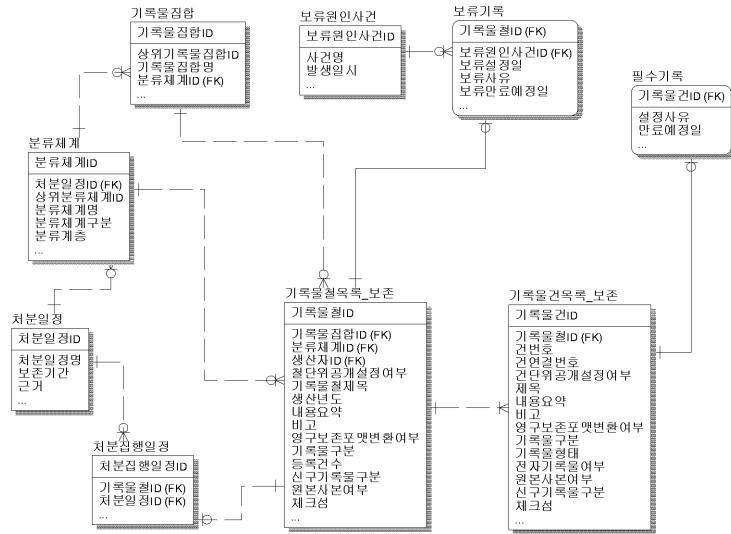
다음으로 장기보존포맷 및 디지털컴포넌트와 스토리지 관계를 [그림 6]과 같이 모델링하였는데 여기에 적용한 원칙은 다음과 같다. 첫째, 보존단의 기록물 장기보존포맷 및 디지털컴포넌트는 기록물건 목록과 연결한다. 둘째, 스토리지 시스템에 대한 정보를 스토리지 관리 데이터로 관리한다. 셋째, 각 스토리지의 저장경로들은 스토리지 관리 메타데이터의 하위로 저장경로 관리 데이터를 두어 관리한다. 넷째, 장기보존포맷 및 디지털컴포넌트는 저장경로관리 데이터와 연계한다. 다섯째, 스토리지 시스템, 스토리지 디스크에 대한 백업, 마이그레이션 등 관리의 효율성을 고려한다.

마지막으로 사용자, 그룹 및 역할과 접근제어 관계는 [그림 7]과 같이 모델링하였는데 첫째, 기록관리시스템 기능요건 표준들에 정의된 사용자, 사용자그룹, 역할관계, 접근제어, 분류체계를 반영하였으며, 둘째, 기록물철, 기록물건 혹은 기록집합체에 대한 접근 제어를 위해 접근제어목록 내 접근제어항목, 권한을 가지는 사용자와 사용자 그룹, 권한의 종류를 정해주는 역할 사이의 다중관계 설정이 가능하도록 하였다.

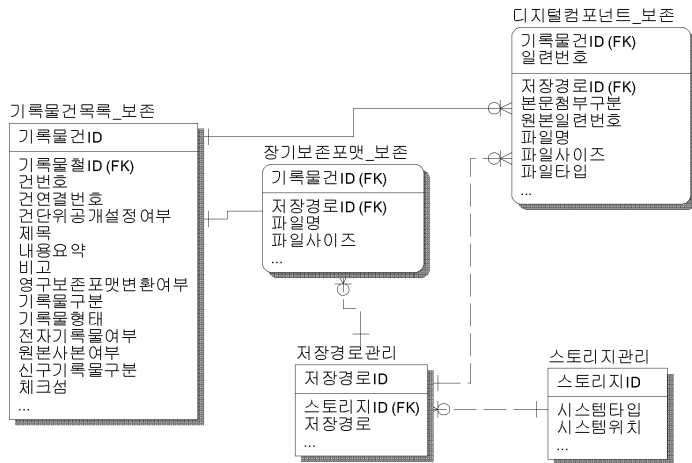
[그림 4] 보존단의 기록물 철과 건, 컴포넌트 엔티티 관계 모형



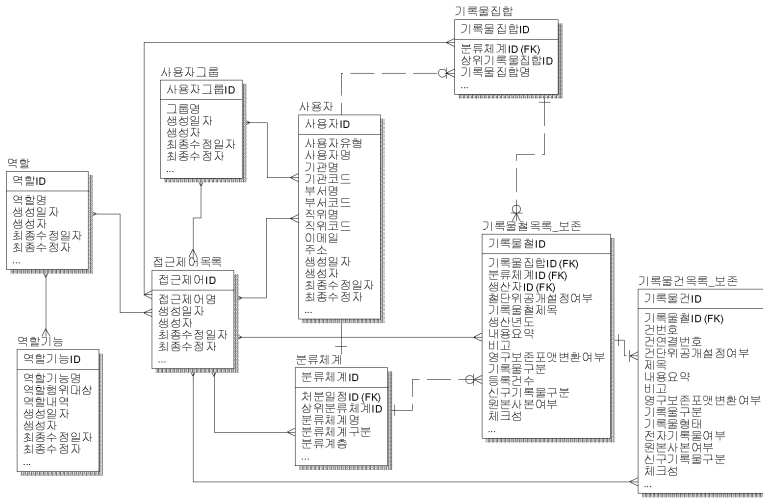
[그림 5] 보존단의 분류와 처분일정 엔티티 관계 모형



[그림 6] 보존단의 장기보존포맷, 컴포넌트, 스토리지 엔티티 관계 모형



[그림 7] 보존단의 사용자, 그룹 및 역할, 접근제어 엔티티 관계 모형



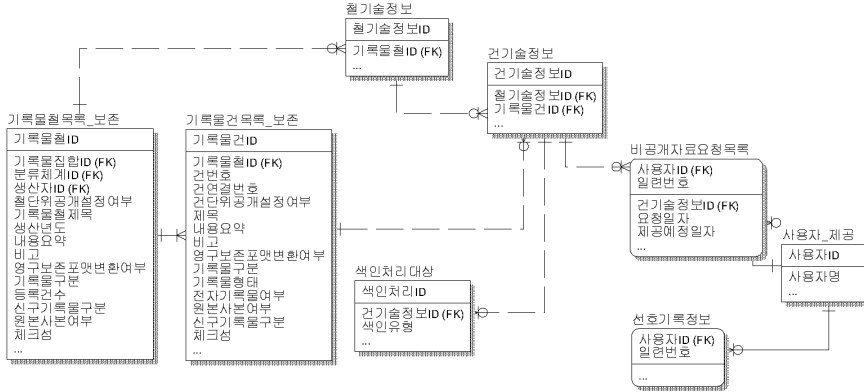
3) 제공단 개념스키마

제공단에서는 [그림 8]과 같이 기록물철과 기록물건에 대한 기술정보 데이터를 각각 분리하도록 모델링하였는데 그 이유는 서비스 사용자에 의해 접근되는 데이터를 별도로 관리함으로써 서비스 데이터 이외의 정보에 대한 보안성을 강화하기 위해서이다. 또한, 서비스 사용자 관련 데이터를 분리하도록 모델링하였는데 그 이유는 비공개자료요청목록 (발체사본 등)을 관리하고 선호기록정보 관리를 통한 사용자 개인화 서비스를 가능하도록 하기 위해서이다.

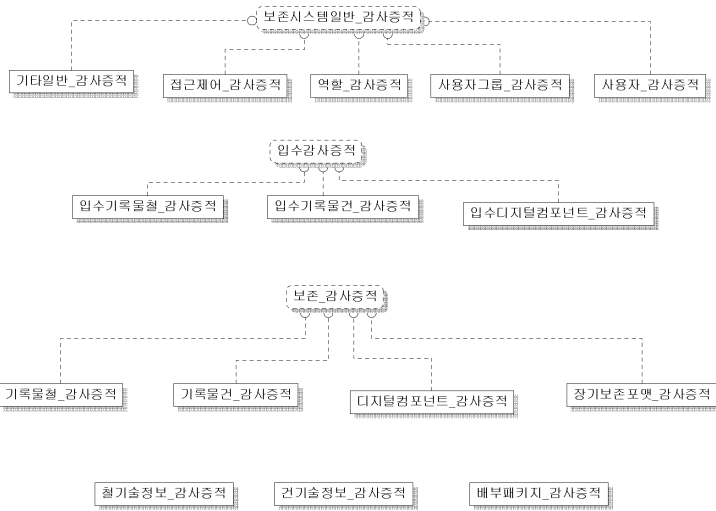
다음으로 [그림 9]와 같이 제공단에서의 감사증적(Audit Trail)을 위한 데이터베이스 스키마를 설계하여 감사증적 정보를 구조화하여 관리하도록 하였는데 이는 첫째, 기존에는 감사증적 정보가 별도로 관리되지 못하거나 다른 메타데이터와 공존해야 하는 상황을 개선하고, 둘째, 감사

증적의 타입에 따라 별도 관리하고 관련 있는 감사증적 정보를 뷰로 통합함으로써 감사증적정보의 저장 및 관리 효율성을 높이기 위해서이다.

[그림 8] 제공단의 기록물철과 기록물건 기술정보 엔티티 모형



[그림 9] 제공단의 감사증적 엔티티 모형



4) 논리스키마 개발 방향

CAMS의 경우 2015년 이후 대량의 전자기록물이 입수되는 것에 대비하여 논리스키마에 대한 정밀한 설계가 필요하다. 2015년 국가기록원으로 이관될 전자기록물의 양은 약 2.5테라바이트로 추정된다.¹²⁾ 매년 이관기록물의 양이 증가할 것이라는 점을 고려하여 기록물철과 기록물건 테이블에 대해 다음의 설계방식을 고려할 수 있다.

첫째, 테이블에 대한 파티셔닝(Partitioning)을 적용할 것인지 결정해야 한다. 매년 지속적으로 증가는 기록물의 특성으로 인해 기록물철과 기록물건 테이블의 레코드 개수는 증가하게 된다. 테이블 용량이 극도로 큰 경우에는 인덱스로 검색 성능을 높이는데 한계가 있다. 인덱스를 만들게 되면 검색 성능은 좋아지는 반면에 기록물의 삽입, 삭제 및 갱신 성능은 저하될 수 있다. 대안으로 기록물철과 기록물건 테이블에 대한 파티셔닝을 검토해 봐야 한다. 연도별 이관이 정례화된다는 점에 착안한다면 기록물철과 기록물건 테이블을 연도별 수평 파티셔닝을 하는 것이 유용할 것이다.

둘째, 테이블에 대한 파티션 뷰(Partitioned View)를 생성할 것인지 결정해야 한다. 뷰에 참여하는 각 테이블들의 공통 칼럼에 제약조건을 설정함으로써 뷰에 대한 질의 시 제약조건 칼럼에 대한 검색키 적용하게 되면 뷰에 참여하는 모든 테이블을 스캔하지 않고 제약조건이 설정된 칼럼에 해당 검색키를 갖는 테이블만 선별적으로 스캐닝하게 되어 검색 성능이 향상될 수 있다. 보존단 스키마 설계에서 제시한 기록물 유형별 테이블에 적용이 가능하다.

셋째, 테이블 간의 클러스터링(Clustering)을 적용할 것인지 결정해야 한다. 클러스터링이란 일반 질의(Query) 및 조인 질의(Join Query)에서 검색 칼럼의 값 혹은 칼럼들 값의 쌍이 동일한 레코드들을 물리적으로

12) 임진희·이대욱, 위의 책, 207쪽 참조.

같은 지역에 배치함으로써 검색 성능을 향상시키는 것이다. 제공단에서 사용자 서비스를 위해 기록물철과 기록물건에 대한 기술정보를 담은 테이블에 적용이 가능하다.

넷째, 인덱스(Index)의 활용 전략을 결정해야 한다. 인덱스는 일반적으로 가장 많이 사용되는 검색 성능 향상 기법이다. 검색 키워드로 자주 참여하는 칼럼에 대해 인덱스를 생성하는 것이 기본이다. 활용 패턴에 대한 분석이 요구된다.

5. 맺음말

이 연구에서는 현재 CAMS의 기록물철과 기록물건 테이블의 현황을 살펴보고 예상되는 문제점과 그 원인, 향후 개선방향을 논의해보았다. 또한, 입수단, 보존단, 제공단으로 나누어 CAMS 데이터베이스의 개념스키마를 예시하고 지속적으로 누적될 기록물 관리의 특성을 고려하여 성능을 담보할 수 있는 논리스키마의 개발 방향을 살펴보았다. 여기서 제시한 개념스키마와 논리스키마 개발 방향은 CAMS뿐만 아니라 표준 RMS 설계에도 상당부분 적용 가능하다.

이 논문에서 살펴본 바와 같이 CAMS의 데이터베이스 스키마를 제대로 개선하기 위해서는 대대적인 정보화전략계획수립(ISP, Information Strategy Planning)이 시급하다. 현재의 CAMS는 2006년도의 ISP에 기반하여 업그레이드된 것이었다. 당시의 ISP는 표준RMS 구축을 위한 ISP로 시작되었으나 CAMS를 범위에 추가한 것으로 CAMS를 집중적으로 다루지 못한 한계가 있었다. 전자기록의 생산 환경과 활용 환경이 급격히 변화해 간다는 점과 ISP를 수행한 지 7년 가까운 시간이 지났음을 감안한다면 기록관리시스템 전반에 대한 새로운 검토와 개선이 시급하다

할 수 있다.

새로운 ISP에서는 2015년부터 시작될 본격적인 전자기록 이관의 성공적 완수, 대량 전자기록의 안전한 보존 체계의 정립, 진본 전자기록의 효율적 이용체계 확립을 위해 기록의 생애주기 전반을 검토할 필요가 있다. 현재의 장기보존포맷 NEO 구조 적합성 검증, 전자서명장기검증 체계 적합성 검증 등 기존 전자기록 진본성 체계에 대한 검토가 필요하다. 여기에 전자기록의 종추적 무결성 검증을 위해 필요한 감사증적 정보의 생성 및 관리, 활용 체계 수립, 접근통제 설정 및 적용 체계 수립, 기록의 완전한 마이그레이션을 위한 내보내기 도구 설계 등을 통해 진본성 확인 및 유지 체계를 완성해야 한다. 영구기록물관리시스템을 위한 정보화전략계획 수립 및 시스템 개발을 위한 예산이 충분히 확보되어 안심하고 2015년을 맞이할 수 있기를 기대해 본다.

ABSTRACT

Directions for Developing Database Schema of Records in Archives Management Systems

Yim, Jin-Hee·Lee, Dae-Wook·Kim, Eun-Sil·Kim, Ik-Han

The CAMS(Central Archives Management System) of NAK(National Archives of Korea) is an important system which receives and manages large amount of electronic records annually from 2015. From the point of view in database design, this paper analyzes the database schema of CAMS and discusses the direction of overall improvement of the CAMS.

Firstly this research analyzes the tables for records and folders in the CAMS database which are core tables for the electronic records management. As a result, researchers notice that it is difficult to trust the quality of the records in the CAMS, because two core tables are entirely not normalized and have many columns whose roles are unknown.

Secondly, this study suggests directions of normalization for the tables for records and folders in the CAMS database like followings: First, redistributing the columns into proper tables to reduce the duplication. Second, separating the columns about the classification scheme into separate tables. Third, separating the columns about the records types and sorts into separate tables. Lastly, separating metadata information related to the acquisition, takeover and preservation into separate tables.

Thirdly, this paper suggests considerations to design and manage the

database schema in each phase of archival management. In the ingest phase, the system should be able to process large amount of records as batch jobs in time annually. In the preservation phase, the system should be able to keep the management histories in the CAMS as audit trails including the reclassification, revaluation, and preservation activities related to the records. In the access phase, the descriptive metadata sets for the access should be selected and confirmed in various ways.

Lastly, this research also shows the prototype of conceptual database schema for the CAMS which fulfills the metadata standards for records.

Key words : Electronic Records Management, Archives Managemetn System, Database Design for Records, Conceptual Schema for Records, Entity Relationship Model for Records