

행정정보 데이터세트 기록의 선별 기준 및 절차 연구

조은희* · 임진희**

1. 머리말
 - 1) 연구의 배경과 목적
 - 2) 선행연구
2. 행정정보 데이터세트의 개념과 유형
 - 1) 행정정보 데이터세트의 개념
 - 2) 데이터세트의 유형과 서비스의 특성
3. 데이터세트 기록화를 위한 선별의 기준
 - 1) 데이터세트 건 단위 식별의 개념
 - 2) 데이터세트 선별의 시점
 - 3) 데이터세트 기록화 대상
4. 행정정보시스템의 데이터세트 선별 절차
 - 1) 전사적 아키텍처의 DRM에 기반한 데이터세트 식별
 - 2) 데이터세트 기록화를 위한 선별 절차
5. 맺음말

* (주)딤스 책임컨설턴트

** 교신저자. (사)한국국가기록연구원 책임연구원

[국문초록]

행정정보시스템은 계속 숫자가 증가하고 있음에도 불구하고 기록관리의 사각지대에 있어서 시스템이 노후화되거나 새로 개발되는 경우 기록이 손실될 위험을 내포하고 있다. 또한 기록관리를 염두에 두지 않고 설계되었기 때문에 기록의 특성과 요건을 만족하지 않은 상태로 관리되고 있다. 기록관리 선진국에서는 정부차원에서 데이터세트 기록의 중요성을 인지하여 데이터세트 아카이브를 운영하는 사례도 있고, 보존을 위한 관리체계나 보존포맷에 대한 프로젝트를 수행한 사례도 있다. 우리나라도 데이터세트에 대한 기록관리 체제와 개별 행정정보시스템에 대한 연구가 조금씩 수행되고 있으나 공식적인 관리 체계가 수립되지는 않은 상태이다.

본 논문에서는 이러한 방향성에 기본 하여 전자정부 추진 전략의 틀에서 데이터세트 식별체계의 수립과 기록관리기준의 설정 전략을 제안하였다. 행정정보시스템 기록화를 위한 데이터세트 선별 절차를 (1)기록화 대상 데이터세트 결정, (2)데이터세트 기록 건 식별 단계, (3)데이터세트 기록의 관리계층 구성 단계로 나누어 정리해보았다. 이 논문은 데이터세트에 대한 보존포맷이나 데이터세트 아카이브 단계에서의 관리절차 등 아카이빙 단계의 문제는 연구내용에 포함하지 않았다.

주제어: 데이터세트, 행정정보시스템, 전자정부

1. 머리말

1) 연구의 배경과 목적

행정정보시스템¹⁾에 축적되는 데이터세트는 업무활동의 증거이자 대체할 수 없는 기록정보자원으로서 기록화 및 장기적인 기록관리가 필요하다. 우리나라에서도 전자정부를 추진함에 따라 주로 종이로 생산·관리되던 문서들이 행정정보시스템에 전자문서나 업무트랜잭션(Business Transaction)²⁾ 정보로 생산·관리되는 추세이다. 인사기록카드, 주민등록등본, 학적부 등이 대표적 사례이다³⁾. 이처럼 과거 문서 형태를 띠던 기록이 대부분 데이터세트 형태의 기록으로 행정정보시스템에 의해 생산됨에도 불구하고 이를 기록으로 획득하여 관리할 절차와 방법이 부재한 상태로 데이터세트의 양은 지속적으로 누적되고 있다. 업무트랜잭션 정보를 포함한 데이터세트는 업무 전반의 투명성과 설명책임성을 보여줄 뿐만 아니라 정보적 가치 측면에서도 유용한 역할을 수행하며, 사회적 통계 데이터나 과학적 관측 데이터와 같은 데이터세트는 시대상의 표현 혹은 사실 정보로서 향후 연구자들에게 활용가치가 높다.

이와 같은 데이터세트 기록의 중요성을 인지하여 외국에서는 정부차원의 데이터세트 아카이브를 운영하고 있다. 미국은 AA

-
- 1) 사무관리규정 제3조 13항 “행정정보시스템”이란 행정기관이 행정정보를 생산·수집·가공·저장·검색·제공·송신·수신 및 활용하기 위한 하드웨어·소프트웨어·데이터베이스와 처리절차 등을 통합한 시스템을 말한다.
 - 2) 여기서 업무트랜잭션 개념은 David Bearman의 정의에 따른다. 「Virtual Archives」, ICA Meeting in Beijing, 1999
 - 3) 기존의 인사관리카드는 e-사람시스템에, 주민등록카드는 주민등록전산정보시스템에, 학적부는 교육행정정보시스템(NEIS)에 각각 데이터세트 형태로 변환 저장되어 관리되고 있다.

D⁴⁾ 사이트를 구축하고 NARA에서 영구 보존되고 있는 역사적 데이터베이스의 일부를 온라인을 통해 서비스하고 있다⁵⁾. 영국에서는 각 부처의 데이터베이스에 축적되는 통계 데이터들이 영국 정부의 의사결정 과정에 대한 기초 자료로서 정책과 법률에 많은 영향을 미치고 있는 것으로 보고 데이터세트를 보존대상에 포함시키고, TNA에서 ULCC⁶⁾에 운영 외주를 주는 방식으로 NDAD⁷⁾를 운영하고 있다.

우리나라도 기록관리혁신 추진기본계획⁸⁾(2005)에서 모든 공적행위의 철저한 기록화를 위해 행정정보시스템 데이터세트의 기록화를 언급한 바 있다. 그러나 현재 행정정보시스템에서 데이터세트 형태로 생산된 기록은 국가 전반적인 기록관리 범주에 포함되어 있지 않은 상태이다. 이러한 상황으로 인해 데이터세트를 담고 있는 행정정보시스템이 노후하여 폐기되거나 새롭게 진화하는 시점에 기존의 데이터세트 기록이 손실될 위험에 처해있다. 늦기 전에 우리나라 행정정보시스템의 데이터세트를 기록화하고 보존하기 위한 다각적인 대책이 마련되어야 할 것이다. 본 연구에서는 공공기록물관리법령에 명시된 기록의 생산시스템으로서 행정정보시스템에서 생산·관리되는 데이터세트가 갖는 특성을 살펴보고 그에 따른 기록화 방안을 제시하고

4) The Access to Archival Database, [<http://aad.archives.gov/aad/>]

5) NARA에서는 거의 200,000 데이터 파일을 보유하고 있는데 그들 중에 AAD를 통해 약 475 아이템 정도를 개인, 지역, 기관, 날짜별로 선별하여 제공하고 있다. [<http://aad.archives.gov/aad/>, 2008.06 AAD 홈페이지]

6) University of London Computer Center(런던대학컴퓨터센터) <http://www.ulcc.ac.uk>

7) NDAD(The National Digital Archive of Datasets) 영국의 데이터세트 아카이브로서, 장기보존 데이터세트를 각 부처로부터 획득하여 온라인을 통해서 서비스를 제공하고 있다. 데이터세트의 특징을 살려 온라인을 통해 데이터세트에 직접 테이블 질의 기능을 통해 효율적이고 정확한 검색 서비스를 이용자에게 제공한다. (<http://www.ndad.nationalarchives.gov.uk/>)

8) 중앙기록물관리기관, 「기록관리 혁신 추진기본계획」, 2005.8, p.14

자 하며, 특히 기록화 과정 중 데이터세트를 기록으로 선별하는 구체적인 기준과 절차를 제시하고자 한다.

2) 선행연구

데이터세트에 대한 국내의 선행연구는 세 영역으로 구분할 수 있다.

첫 번째는 데이터세트의 기록관리에 관한 전반적인 이론 연구이다. 이소연⁹⁾은 매체별 아카이빙의 한 가지 유형인 데이터세트 아카이빙 분야의 필요성을 주장하였다. 데이터세트 아카이빙에서 가장 활발한 움직임을 보이고 있는 분야는 과학 및 사회과학 분야라고 진단하고, 인구통계, 항공우주과학, 천문기상, 유전자공학 등 이 분야에서 구축해 온 다양한 관찰 및 실험 데이터 사례를 제시하고 있다. 현문수¹⁰⁾는 기록으로서 데이터세트 관리의 필요성을 언급하고, 영국과 미국의 사례 분석을 통해 데이터세트의 조직 및 기술을 중심으로 데이터세트 관리방안을 제안하고 있다. 윤수영¹¹⁾은 데이터세트에 대한 기술(description) 방법을 중점적으로 제안하고 있다.

두 번째는 국가기록관리 체계 하에서 데이터세트 기록관리의 필요성을 느끼고 데이터세트 기록관리에 대한 정책, 법규적인 부분 및 데이터세트 아카이브와 관련된 사항을 대상으로 수행한 연구이다. 「국가 디지털 아카이빙 체계 구축에 관한 연구」¹²⁾에서는 과학기술 정보자원 유형별 아카이빙 전략을 제시

9) 이소연, 「디지털유산의 장기적 보존 : 국가정책 수립을 위한 제안」, 『기록학연구』, 10, 2004, p.42

10) 현문수, 「데이터세트 기록의 관리 방안」, 『한국기록학회지』, 2005

11) 윤수영, 「데이터세트 기록물의 기술요소에 관한 연구」, 전북대학교 정보과학대학원, 2007

12) 한국국가기록연구원, 「국가 디지털 아카이빙 체계 구축에 관한 연구」, 2005

하면서 데이터베이스에 대한 관리전략을 소개하고 있다. 대통령 비서실 기록관리 ISP(정보화 전략계획수립) 사업에서는 '개별업무시스템 데이터세트 획득방안'이라는 개선과제를 통해 데이터 세트에 대한 기록화 방안을 제시하고 있다¹³⁾. 국가기록원 기록관리시스템 혁신 ISP 사업에서는 행정정보 데이터세트 기록의 이관방안을 ① 행정정보시스템에 탑재된 연계모듈을 통해 기록관리시스템으로 이관하여 관리하는 방안과 ② 행정정보시스템에 탑재된 연계모듈을 통해 직접 국가기록원의 데이터세트 아카이브로 이관하는 방안 등 두 가지로 제시하고 있다¹⁴⁾. 「행정정보시스템 데이터세트 기록관리 연구용역 보고서」¹⁵⁾에서는 환경 분석과 사례분석을 통해 행정정보시스템의 유형을 분류하고, 기록관리 대상 데이터세트의 선정 기준을 마련하고, 행정정보시스템 데이터세트 아카이브 단계에서의 관리방안을 제시하고 있다.

세 번째는 행정정보시스템별로 접근한 연구사례이다. 개별 행정정보시스템 중에서 디지털예산회계시스템¹⁶⁾, 조달업무시스템¹⁷⁾, e-사람¹⁸⁾, 교육행정정보시스템(NEIS)¹⁹⁾²⁰⁾등을 대상으로 기

-
- 13) 국가기록원, 「행정정보시스템 데이터세트 기록관리 연구용역 보고서」, 2007, p.17
 - 14) 국가기록원 「기록관리시스템 혁신 ISP사업」, 2006, p.37, <국가기록원, 행정정보시스템 데이터세트 기록관리 연구용역 보고서, 2007, p.17 재인용>
 - 15) 국가기록원, 「행정정보시스템 데이터세트 기록관리 연구용역 보고서」, 2007
 - 16) 이은별, 「국가 재정정보의 기록학적 관리방안 : 디지털예산회계시스템을 중심으로」, 명지대학교 기록과학대학원, 2008, 재정기록이 전자적으로 관리될 수 있는 환경을 조성하고 현행 기록 관리체계에서 관리되지 않고 있는 정보데이터를 기록화 하여 재정기록 정보서비스를 최대화할 수 있는 방안을 연구
 - 17) 이순한, 「조달업무의 설명책임성 확보를 위한 조달정보 기록관리 요건 연구」, 한국외국어대학교, 2008
 - 18) 진채환, 「공공기관의 인사기록관리에 관한 연구」, 한국외국어대학교, 2007
 - 19) 임미숙, 「교육행정정보시스템의 기록관리 기능분석-학교생활기록부를 중심으로」, 한국외국어대학교, 2007

록관리를 위한 연구가 수행되었다. 이러한 흐름은 개별 행정정보시스템별로 생산되는 중요한 데이터세트에 관심을 두고 기록으로 보존할 필요성이 점점 부각되고 있음을 보여준다.

한편 해외에는 데이터세트 기록관리에 관련한 여러 프로젝트와 워크숍을 중심으로 연구결과가 발표되어 있다. ERPANET²¹⁾을 중심으로 데이터베이스 및 데이터세트의 장기보존 및 관리에 관한 논의가 있었다. ERPANET은 2003년 4월 스위스에서 “Long-term Preservation of Databases”라는 제목으로 워크숍을 개최하였으며, 스위스, 네덜란드, 영국, 미국 등지의 연구자들이 참석한 이 워크숍에서는 데이터베이스 보존에 대한 실험과 시도에 대해 소개하고 토론하였다²²⁾. 네덜란드의 Digital Preservation Testbed 프

- 20) 한철희, 「NEIS 교무업무시스템 데이터의 기록화 방안 연구(학교생활기록부를 중심으로)」, 명지대학교 기록과학대학원, 2007, 어떤 데이터를 기록으로 선정할 것인지, 획득 시기는 언제인지 등 데이터세트의 기록화 방법과 절차를 도출하였다. 다양한 행정정보시스템의 기록화를 대상으로 하진 않지만, 중요한 기록을 생산해 내는 행정정보시스템의 사례연구로서 해당 시스템의 기록화를 위한 방법과 절차를 제안.
- 21) <http://www.erpanet.org/>, 문화유산과 과학자원의 디지털 보존 영역의 연구를 수행하기 위해 유럽지역에서 조직된 연구 그룹으로서 사무소는 영국의 Glasgow 대학에 위치.
- 22) 워크숍 Long-term Preservation of Databases“의 홈페이지를 참고하여 각국의 발표 내용을 정리해 보면 다음과 같음
(<http://www.erpanet.org/events/2003/bern/index.php>)
 - 프랑스는 1979년 CONSTANCE 센터 설립. 역사적 가치를 갖는 데이터베이스를 선별하여 국가적 차원에서 기록으로 관리하고 있음.
 - 벨기에는 DAVID 프로젝트를 통해 각종 전자기록을 관리하기 위한 솔루션을 개발함.
 - 미국 NARA에서 소장하고 있던 소프트웨어와 독립적인 연구 데이터 20,000 파일을 전달하여 AERIC 시스템에서 아카이빙한 사례가 있음.
 - 네덜란드는 Digital Preservation Testbed 프로젝트를 수행. 데이터베이스와 이용자 응용프로그램을 대상으로 보존전략을 제시함.
 - 영국은 NDAD를 통해 보존된 자료의 온라인 공공접근을 위한 국가 아카이브 서비스를 제공함.
 - 스위스는 SIARD라는 관계형 데이터베이스 보존 솔루션을 개발하여 기존의

로젝트²³⁾에서는 결과물 중의 하나인 『From Digital Volatility to Digital Permanence』라는 보존시리즈²⁴⁾에서 프로젝트 수행의 경험과 사례를 기록의 유형별로 나누어 수록하고 있다. 또한 프로젝트의 결과물에는 데이터베이스 유형에 대한 연구를 포함하고 있는데 데이터베이스의 기록유형의 특징, 보존전략, 실행방법과 실질적 행동계획 등을 제시하고 있다.

이상에서 살펴본 바와 같이 데이터세트의 기록관리 및 보존과 관련하여 외국의 경우 아카이브를 운영하는 사례나 데이터베이스 보존을 위한 실험과 시도가 있었으나 국내 연구의 경우 필요성의 주장이나 기술, 혹은 개별 행정정보시스템에 대한 연구에 한정되고 있다. 본 연구에서는 이러한 한계를 극복하기 위해 논의를 좀 더 구체화하여 국내 행정정보시스템에서 관리되는 데이터세트를 기록화하기 위해 필요한 선별의 기준과 절차를 제시하고자 한다.

2. 행정정보 데이터세트의 개념과 유형

1) 행정정보 데이터세트의 개념

데이터세트라는 용어는 많은 곳에서 유사한 성격으로 사용되

XML 컨버전 소프트웨어의 단점을 보완하고자 함.

23) 네덜란드의 Ministry of the Interior and Kingdom Relations(BZK)와 네덜란드 국립 기록관이 공동 구축

24) DPT, 『From Digital Volatility to Digital Permanence : Preserving Databases』, Digital Preservation Testbed, 2003 (역서 : 이미화, 현문수 역, 『전자기록의 유형별 보존 기법』, 한국국가기록연구원, 2006)

<<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility=permanence-databases-en.pdf>>

고 있으나 경우에 따라 조금씩 다르게 정의되어 사용된다. 데이터세트는 데이터파일, 데이터베이스, 데이터집합 등의 다양한 용어로 표현되기도 한다.

IT 용어사전	어떤 규칙에 따라 배열된 데이터의 집합으로, 어떤 운영 체제에서는 데이터 파일과 같은 뜻으로 사용된다고 정의하고 있다. ²⁵⁾
SAA 용어사전	컴퓨터가 분석 할 수 있도록 형식을 갖춘 연관성 있는 정보들의 집합으로 데이터파일이나 데이터베이스와 유사한 개념으로 사용되기도 한다고 정의하고 있다. ²⁶⁾
기록학 용어사전	컴퓨터가 처리하거나 분석할 수 있는 형태로 존재하는 관련정보의 집합체로서 데이터세트가 정의되고 있다. 반복적으로 발생하는 사건이나 이에 대한 정보가 관계를 가지고 누적되어 컴퓨터로 처리할 수 있게 구조화되어 누적된 것을 의미한다. 습도나 온도를 지속적으로 관측한 데이터가 누적된 경우와, 학생기록과 같이 기록의 대상이 되는 객체를 공유하여 관계를 가지고 누적된 것을 사례로 들 수 있다. ²⁷⁾

〈표 1〉 데이터세트 용어의 사전적정의

위와 같은 여러 가지 정의를 참고하여 개념을 정리하면, 데이터세트는 “컴퓨터가 처리할 수 있도록 구조화된 데이터의 집합”이다. 다만, 단순히 구조화된 데이터의 집합이라기보다는 기록학 용어사전에서 정의한 바와 같이 “데이터가 서로 관계나 맥락을 가지고 하나의 대상을 중심으로 반복적으로 누적”된 경우로 한정된다.

행정정보 데이터세트는 행정정보시스템을 통해 생산·축적되

25) 네이버 IT 용어사전 [<http://terms.naver.com/>]

26) SAA 홈페이지 용어사전

http://www.archivists.org/glossary/term_details.asp?DefinitionKey=332

27) 한국기록학회 엮음, 『기록학 용어사전』, 역사비평사, 2008, pp.82-83

는 데이터세트이다. 공공기록물 관리에 관한 법률 시행령에 의하면 전자문서시스템²⁸⁾, 행정정보시스템, 업무관리시스템²⁹⁾이 전자기록 생산시스템으로 지정되어 있다³⁰⁾. 즉, 전자문서시스템, 업무관리시스템을 제외한 모든 시스템을 행정정보시스템으로 간주할 수 있다. 사례로는 디지털예산회계시스템, 교육행정정보시스템, 조달정보시스템, e-사람, 통계지식관리시스템 등 다양한 시스템이 있다.

공공부문에서는 1967년 경제기획원 통계국에 인구조사를 목적으로 도입된 컴퓨터가 최초의 행정정보시스템으로 알려져 있다. 전산화 초창기에는 단위업무를 전산화하거나 망사업에 관심을 가졌기 때문에 주목할 만한 행정정보시스템은 많지 않을 것으로 짐작된다. 전자정부가 추진되는 국민의 정부 후반기(2001~2002)부터 범정부적인 관점에서 구축되는 행정정보시스템이 나타났고, 행정정보시스템을 통해 생산, 축적되는 데이터세트의 양이 급격히 증가 하였다. 4대 사회보험 정보시스템, 정부통합전자조달 시스템, 재정정보시스템, 교육행정정보시스템, 표준인사관리시스템(현재 e-사람으로 명칭 변경)등이 추진되기 시작하였고, 현재 업무관리시스템과 전자문서시스템으로 발전한 전자결재 및 전자문서 유통 정착을 위한 조치들도 이 시기에 취해졌다³¹⁾.

현재도 많은 행정정보시스템이 구축 중이고, IT 발전의 경향

28) 사무관리규정 제2조 12항 "전자문서시스템"이라 함은 문서의 기안·검토·협조·결재·등록·시행·분류·편철·보관·보존·이관·접수·배부·공람·검색·활용 등 문서의 모든 처리절차가 전자적으로 처리되는 시스템을 말한다.

29) 사무관리규정 제2조 14항 "업무관리시스템"이라 함은 행정기관이 업무처리의 전 과정을 과제관리카드 및 문서관리카드 등을 이용하여 전자적으로 관리하는 시스템을 말한다.

30) 공공기록물관리에 관한 법률 시행령 제2조 7항의 정의

31) 정충식, 『전자정부론』, 서울경제경영, 2007에서 발췌 정리

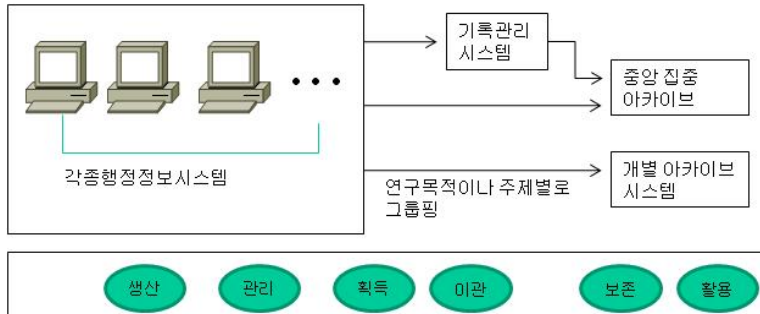
을 보면 앞으로도 행정정보시스템의 수는 꾸준히 증가할 것이다. 따라서 행정정보시스템을 통해 생산되는 데이터세트들에 대한 장기적인 보존 및 기록관리를 위한 대책 수립도 시급한 상황이다.

2) 데이터세트의 유형과 서비스의 특성

데이터세트를 어떻게 기록화하여 보존하는게 좋을지는 향후 어떤 사용자가 어떤 목적으로 데이터세트를 필요로 할 것인지를 예상하는 것에서 답을 얻을 수 있다. 행정정보 데이터세트의 경우 증거적 가치뿐만 아니라 정보적 가치도 높은 기록임을 염두에 두면서, 아직 존재하지는 않지만 데이터세트 아카이브가 향후 어떻게 서비스하는 것이 좋을 것인지를 고려하여 요구사항을 기록화 단계로 환류하는 것이 요건 도출을 위한 최선의 방안이 될 것이다.

데이터세트를 아카이빙하는 모형은 중앙집중형과 개별형으로 나누어 볼 수 있다. 첫 번째의 중앙집중형 아카이브 모형은 다시 두 가지 방법으로 나누어 볼 수 있다. 데이터세트를 기관의 기록관리시스템(RMS)을 거쳐서 중앙집중형 아카이브로 이관하는 방식과, 중앙의 아카이브로 직접 이관하는 방식이 가능하다. 전자의 경우 일반 전자기록과 동일한 방식이므로 기록전체에 대해 유형과 관계없이 단일한 통제를 할 수 있다는 장점이 있으나, 기관의 RMS에 데이터세트 유형의 기록을 이관 받아 관리하는 기능을 추가 개발해야 하는 부담이 있다. 두 번째는 데이터세트 주제나 활용목적별로 개별 아카이브 시스템을 구축하는 것이다. 예를 들어 인사기록 데이터세트 아카이브, 주민등록 아카이브 등을 각각 만들어 관리하며 서비스하는 것이다. 단일

기관 혹은 관련 기관들이 모여 구축가능한 방식이고, 증거적 가치가 중요시되는 데이터세트의 경우 엄격한 진본성 요건을 만족하는 아카이브 시스템을 구성해야 한다.



〈그림 1〉 데이터세트의 아카이브 구축방안

아카이빙된 데이터세트의 서비스 특성을 고려하여 행정정보 시스템 데이터세트 유형을 구분해 보면 다음과 같다.

첫째 각종 통계 및 설문을 수행한 원자료(raw data)이다³²⁾. 데이터에 대한 취합 결과나 분석 결과는 전자결재나 업무관리시스템 등을 통해 보고되어 기록으로 관리되지만, 원자료는 따로 관리하지 않는 경우가 많다. 공공부문에서 시행하는 통계 및 설문 자료는 시대상을 연구하는 미래의 학자나 이용자들이 추후 용이하게 검색, 가공, 활용할 수 있도록 데이터세트 형태로 보존하여 제공할 필요가 있다.

두 번째는 주민등록카드, 인사기록카드, 학적부 등의 카드 및 대장류이다. 과거 종이카드로 존재하던 것이 전자정부 추진으

32) 여기서 언급하는 통계 및 설문 자료는 통계청과 같이 그 자체를 주업무로 하는 기관을 제외한 다른 기관들이 업무수행과정에서 수집하거나 생성한 자료를 대상으로 한다.

로 행정정보시스템에 전자 대장으로 변형하여 관리하는 경우이다. 대장류의 공통점은 사람, 건물, 땅 등의 하나의 객체에 대한 이력사항을 지속적으로 기입해가는 방식으로 작성되며, 각종 증명서 발급이나 사실 확인에 수시로 사용된다는 특징이 있다. 이 때문에 이러한 대장류는 처리과에 장기간 비치하는 경우가 많다. 예를 들어 이 유형에 해당하는 주민등록카드, 인사기록카드 등은 영구보존 대상으로 책정되어 있지만 전산화 이후에 명확한 기록 관리방안은 수립되지 않은 상태이다.

세 번째는 전자문서와 업무트랜잭션 데이터가 복합적으로 누적되는 유형이다. 디지털예산회계시스템이나 조달정보시스템 등이 그것이다. 과거 오프라인 상에서 진행하던 업무를 전자적으로 추진하면서 시간이 지날수록 대량의 데이터가 축적되며, 추후 사안별 투명성과 설명책임의 추적에 활용된다는 특징이 있다. 전자정부의 추진, 그리고 민간부문의 e-비즈니스 확대 추세로 업무의 전자적 처리를 위한 이러한 정보시스템은 계속적으로 증가할 것이다.

네 번째는 관측 데이터 유형이다. 관측 데이터는 활용분야가 매우 넓고 다양하며, 데이터를 누적하여 판단을 내리거나 활용하는 경우 장기간 보존할 대상이 된다. 관측 데이터는 환경문제나 국토개발 등의 기초자료로서 중요한 역할을 수행하게 되는데, 장기간에 걸친 관측 데이터들을 활용하여 관련 정책 수립시 정교하고 올바른 의사결정을 할 수 있다. 이러한 데이터를 활용하는 시스템에는 기상정보시스템, 교통정보시스템, 토지정보시스템, 지리정보시스템 등 여러 종류가 있다. 관측 데이터는 이를 활용하기 위한 전문소프트웨어가 필요하며 기록화를 할 때 응용프로그램을 관리 대상으로 포함³³⁾하는 등 추후 데이터

33) Meijer, A. J., 「Transparent government: Parliamentary and legal accountability in an

세트의 활용·재현을 위한 조치가 중요하다.

	데이터세트 유형	서비스 특성	유형 사례
유형 I	통계 및 설문 등의 원자료	관계형 데이터베이스에 데이터를 보존하여 분석·활용 서비스 가능	여론조사 시스템
유형 II	카드 및 대장류	관계형 데이터베이스에 데이터를 보존하되 카드 및 대장 서식의 재현과 증명서 발급 서비스 가능	인사기록 카드
유형 III	전자문서와 업무트랜잭션 데이터	사안별로 진행 과정을 추적할 수 있도록 전자문서와 업무트랜잭션 데이터를 기록화	조달정보 시스템
유형 IV	관측 데이터	데이터의 구조와 연관관계를 이해하고 재현할 로직과 어플리케이션을 함께 기록화하여 활용	지리정보 시스템

〈표 2〉 데이터세트 유형별 서비스 특성

3. 데이터세트 기록화를 위한 선별의 기준

1) 데이터세트 건 단위 식별의 개념

기록의 선별(selection)이란 업무과정 중 생산되는 문서와 정보 중에서 기록으로 남겨야 할 것을 결정하는 것을 말한다. 데이터 세트 기록화를 위한 선별도 동일한 맥락에서 이해할 수 있다. 그런데, 데이터베이스에 저장된 데이터세트를 기록으로 관리하는 것은 문서를 기록으로 관리하는 것과는 차이가 있는데 특히 기록으로서 통제를 적용할 단위나 대상이 명확하지 않다는 점

information age」, Information Polity, 2003

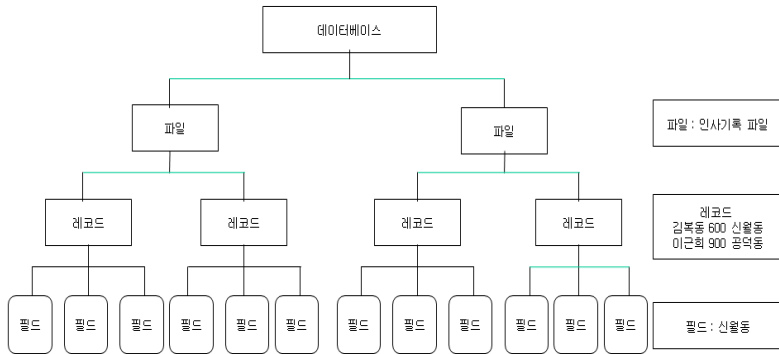
에서 그렇다. 데이터세트를 기록으로 관리하기 위해서는 가치 선별 이전에 먼저 데이터세트의 어느 범주를 하나의 기록 건으로 볼 것인지 식별(identification)하여 확정하는 것이 필요하다. 기록 건이 분명해야 건들의 집합을 정의할 수 있고 덩어리 기록에 대한 관리나 통제 기준을 도출할 수 있기 때문이다.

데이터세트의 건단위 식별이란 기록관리 단계상에서 적절한 조치를 취하기 위한 통제 및 관리 단위를 결정하는 것이다. 종이 기록의 경우 건-철-시리즈의 계층구조를 가지며 이 계층구조를 기준으로 기록관리기준값을 적용하게 된다. 데이터세트도 기록으로 관리되기 위해서는 동일한 구성이 필요하다. 이 논문에서는 데이터세트를 기록으로 관리하기 위해 통제와 관리 단위를 확정하고 계층구조를 구성하는 것을 ‘데이터세트의 식별’이라고 정의하고자 한다. 기록건(records)은 공식적인 기록으로 선언된 논리적 개체로 내용과 메타데이터로 구성된다. 기록은 하나의 워드 문서이거나, 웹페이지, 멀티미디어 문서와 같이 단일한 객체로 취급되는 단단히 묶인 객체 집합이 될 수도 있다³⁴⁾. 이러한 논리적 선언의 과정이 식별의 과정이라고 할 수 있다.

데이터세트는 주로 데이터베이스 시스템에 의해 저장되고 관리된다. 데이터세트의 식별을 위해서는 데이터베이스의 구조를 이해하고 데이터의 연관관계를 파악해야 한다. 데이터베이스의 논리적 계층은 [데이터베이스-파일-레코드-필드]로 구성된다. 예를 들어 인사데이터베이스는 개인 신상사항, 개인 임용시험, 병역사항, 학력사항, 경력사항, 자격사항 등 다수의 파일들로 구성된다. 인사기록파일에는 “김복동, 600, 신월동”, “이근희, 900, 공덕동” 등이 각각 레코드로 저장되며, “김복동”, “600” 등이 한

34) 설문원, 「전자기록철의 구조와 관리방안」, 한국기록학회지 제5권 제2호, 2005, p.53

레코드의 필드를 구성한다.



〈그림 2〉 데이터베이스의 계층

이름	주민등록번호	우편번호	전화번호	행
홍길동	123456-1234567	123-456	123-1234	
		123-123	456-4567	

열

↑ 테이블은 여러 개의 필드로 구성

〈그림 3〉 인사데이터베이스 중에서 개인신상사항 테이블 예제

현재 행정정보시스템에 가장 많이 사용되고 있는 관계형 데이터베이스의 경우 데이터의 물리적 저장 계층은 [데이터베이스-테이블-행-필드]로 구성된다. 파일은 테이블로 사상(mapping)되며, 레코드는 행으로 사상된다. 테이블은 행(row)과 열(column)

로 구성되는데 행은 하나의 개체에 대한 값을 모아서 포함하고 있으며, 열은 각 메타데이터들에 대한 값을 담고 있다. 관계형 데이터베이스의 구조를 이해하기 위해 그림3의 인사데이터베이스의 개인신상사항 테이블을 살펴보자. 개인신상사항 테이블은 이름, 주민등록번호, 우편번호, 전화번호의 열로 구성된다. 예를 들어 홍길동이라는 사람에 대한 주민등록번호와 우편번호, 전화번호의 구체적인 값이 모여 하나의 행을 이룬다. 즉, “홍길동, 123456-1234567, 123-456, 123-1234”은 하나의 행이 된다. 이 행의 “홍길동”, “123456-1234567” 각각은 필드를 구성한다.

이와 같은 구조의 데이터베이스에서 기록으로 획득할 데이터 세트의 식별단위를 어떻게 정할 것인지 기준을 정리해 보면 첫째, 기록관리 생애 주기 및 관리단계에 따라 통제 및 관리가 가능한 단위이어야 하며 둘째, 기술·분류·처분 등을 수행할 때 계층별로 논리적 연관성을 가져서 하위 계층에서 상위계층의 속성을 상속받거나, 물리적 매체관리와의 관련성을 가지는 등 설계 및 수행에 현실적인 편의성을 제공할 수 있어야 하며 셋째, 업무적·논리적·물리적으로 연관성을 가진 데이터세트가 범주화 될 수 있어야 한다.

DPT 프로젝트에서는 기록으로서의 데이터베이스를 식별하는 것에 대한 개념을 다음과 같은 경우로 나누어 가능성으로 제시하였다³⁵⁾.

- 완전한 데이터베이스시스템 전체가 하나의 전자기록(데이터베이스, RDBMS, 응용프로그램 모두 포함)
- 데이터베이스가 하나의 전자기록
- 하나의 데이터베이스 테이블에 저장된 하나의 열이 전자기록

35) 이미화, 현문수 옮김, 『전자기록의 유형별 보존기법』, 2006, p.197

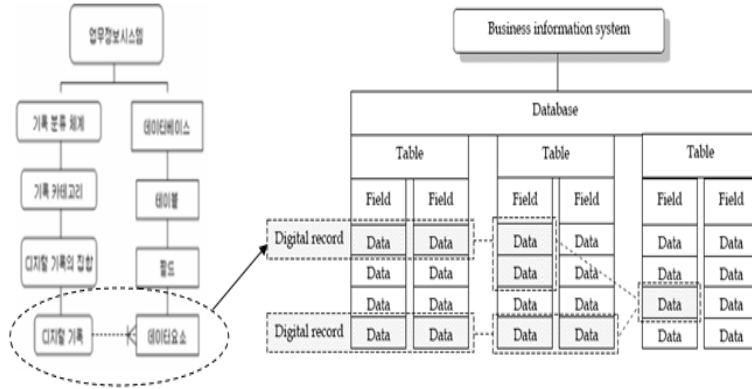
· 응용프로그램에 의해 스크린에 표현된 데이터베이스 내의 정보가 전자기록

위에서 제시된 선택 안에 내포된 문제점 혹은 가능성을 살펴보면 다음과 같다. 첫째, 데이터베이스 단위로 기록건을 식별한다면 처분, 분류, 기술 등 기록 통제의 기본 단위로는 대상 기록의 크기가 커서 세밀한 통제가 어려운 경우가 많다. 다만, 데이터베이스가 담고 있는 기록의 내용과 종류가 단일한 경우에는 이 방식이 유용할 수도 있다. 예를 들어 지속적으로 여론조사 결과만을 담아놓은 데이터베이스가 이에 해당된다. 둘째, 테이블 단위는 작은 정보 단위이기 때문에 업무트랜잭션을 충분히 반영하지 못할 가능성이 있다. 특히 정규화된 테이블의 경우엔 하나의 업무트랜잭션에 여러 테이블이 연관되는 경우가 많다. 이런 상황에서 테이블 단위로 기록건을 식별한다면 업무트랜잭션을 반영하기 위해 기록건 간의 관계 링크 정보를 세심하게 관리해주어야 하는 부담을 안게 된다. 즉, 기록의 맥락정보와 참조정보를 잘 구성해주어야 한다.

데이터베이스는 논리적 구조에 따른 계층관계를 가지고 있으나 그 자체가 업무적 연관성이나 맥락을 반영하지는 않는다. 따라서 데이터베이스 논리적 구조에 따른 계층 단위를 기준으로 데이터세트를 식별하는 것은 불가능하다. 데이터베이스 단위보다는 작고 테이블 단위보다는 큰 범주에서 업무적 연관성과 맥락을 반영해 줄 수 있는 기록건 식별이 이루어져야 한다.

호주의 '업무정보시스템을 위한 기록관리 기능요건 가이드라인'에서 기록건 식별 개념을 잘 보여주고 있다. 이 지침에 따르면, 디지털기록은 데이터베이스의 다양한 필드에 있는 관련된 데이터요소가 그룹화된 것이다³⁶⁾. 구체적으로 살펴보면, 디지털

기록³⁷⁾은 데이터베이스의 물리적 계층에서 봤을 때 하나의 테이블이나 열이 아닌 복수의 관련 테이블에서 복수의 관련된 열을 취해 기록을 구성하게 된다.



〈그림 4〉 데이터베이스에서 기록건의 식별

※ 출처 : NAA, Functional Specifications for Record-keeping in Business Information System Software ,Guidelines, 2007, p.59

- 36) NAA, 「Guidelines for Implementing the *Functional Specifications for Recordkeeping in Business Information System Software*」, 2007 : 기관들이 적합한 기록관리 기능을 갖춘 행정정보시스템을 선택할 수 있도록 NAA(호주 국립기록보존소 National Archive of Australia)에서 개발한 별도의 지침으로서 NAA, 「*Functional Specifications for Recordkeeping in Business Information System Software*」, 2007과 NAA, 「*Functional Specifications for Recordkeeping in Business Information System Software Guidelines*」, 2007로 구성되어 있음. 이 지침에서는 시스템을 업무활동을 지원하고 용이하게 하기 위한 기관의 시스템으로 정의하고 있으며 우리나라의 행정정보 시스템과 대응이 되는 개념이다.
- 37) 호주에서 BIS의 계층은 디지털폴더-디지털기록-디지털 객체-컴포넌트로 구성됨. 디지털 객체는 하나의 디지털 문서, 즉 워드 프로세스나 해당 시스템에 커스터마이징 된 특수한 데이터 포맷으로 존재할 수 있다. 예를 들어서 사안(case) 관리 시스템, 전자상거래 시스템, 콘텐츠 관리 시스템 등은 디지털 기록으로서 디지털 객체를 생산, 획득, 관리할 수 있음

2) 데이터세트 선별의 시점

데이터세트의 기록건 식별이 이루어진 다음에는 일반 기록관리의 과정에서 처럼 가치평가를 통해 해당 기록건의 기록화 여부를 결정하게 된다. 이 절에서는 데이터세트의 기록화를 위한 식별과 선별은 어느 시점에 이루어지는 것이 이상적인지를 검토해 보고자 한다.

Greg O'Shea는 기록관리연속체 모델을 기반으로, 전자기록의 생산 이전단계부터 아키비스트가 적극적으로 개입하여 평가를 수행해야 한다고 강조한다. 평가결정은 기록이 생산되기 전에 정보시스템에 미리 반영되어야 하며, 이를 위해 아키비스트는 어떠한 기록이 선별되어 유지되어야 하는지 미리 결정해야 한다고 주장한다. 전자기록의 평가는 생산 이전에 업무기능 및 활동 등에 대한 고차원적인 분석을 통해 시스템 설계단계부터 보유일정(retention scheduling)을 통해 시스템에 반영되도록 해야 한다³⁸⁾.

Charles Dollar 역시 전자기록의 평가를 위해서는 기록의 계속적 가치를 확인하고 보존하기 위해 아키비스트가 정보시스템의 설계단계부터 적극적으로 참여해야 함을 강조한다. 전자기록관리에서 아키비스트가 공헌할 수 있는 가장 핵심적인 영역은 정보시스템의 설계단계 시 전자기록의 생애주기 관리개념을 시스템에 반영시키는 것이라는 주장이다³⁹⁾. 이러한 작업을 위해서는 아키비스트와 정보기술 전문가, 정보시스템 관리자 간의 긴밀

38) 『전자기록의 이해』, p.74

39) Charles Dollar, 『Archival Theory and Information Technologies : The Impact of Information Technologies on Archival Principles and Methods』, Univ. of Macerata : Macerata, 1992, p.58, 『전자기록의 이해』 p.75에서 재인용

한 협력관계가 필요하다⁴⁰⁾. Hans Hofman은 전자기록관리의 통합적 접근 필요성을 역설하였고, 특히 전자기록 관리를 위한 기초 인프라를 상호 연관성을 가진 3가지 요소로 설정하였다. 첫째, 지적 통제 및 평가를 위한 지적 인프라, 둘째, 생산, 보존, 열람, 활용 등을 위한 기술 인프라, 셋째, 위의 두 요소를 수행하는데 필요한 조직 인프라가 그것이다. 전자기록관리를 위한 이와 같은 인프라를 기반으로, 전자기록의 평가는 가능한 생애주기의 초기단계에서부터 시행되어야 하며, 이러한 인식을 바탕으로 평가를 위한 세부 방법론 및 절차를 개발해야 한다는 것이다⁴¹⁾.

전자기록의 경우 시스템에서 데이터를 생산하기 시작하면 대량으로 생산되면서 새로운 통제를 가하기 어렵기 때문에, 행정정보시스템을 설계할 당시부터 생애주기에 따른 기록관리 요건을 반영하는 것이 필요하다. 이와 같이 데이터세트의 식별과 선별은 행정정보시스템이 설계될 당시에 수행하는 것이 이상적이다.

3) 데이터세트 기록화 대상

앞 절에서는 기록의 기본 통제 단위로서의 데이터세트 기록 식별 개념을 정의하였다. 일반적으로 정보시스템은 소프트웨어, 하드웨어, 데이터베이스, 네트워크, 사용자 등의 항목으로

40) O'Shea Greg, 「The Medium is not the Message: Appraisal of Electronic Records by Australian Archives」, Archives and Manuscripts 22(1), 1994, 『전자기록의 이해』 p.74에서 재인용

41) Hans Hofman, 「Off the Beaten Track : The Archivist Exploring the Outback of Electronic Records」, Proceedings of an Electronic Records Management Conference, 1994, pp.5-6<<http://www.naa.gov.au/govserv/techpub/keeps/hofman.htm>>, 『전자기록의 이해』 p.5에서 재인용

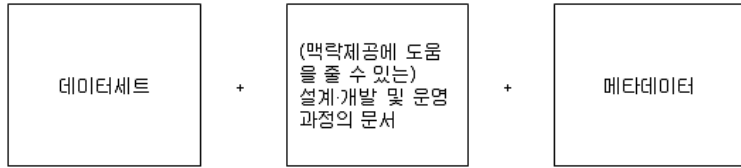
구성되는데⁴²⁾, 데이터세트를 기록화할 때 데이터베이스 저장소에 있는 업무트랜잭션 정보만을 고려해서는 기록화가 불충분하다. 이 절에서는 행정정보시스템의 데이터세트를 기록화할 때 함께 포착해야 할 기록화 대상에 대해 논의해 보고자 한다.

먼저, 영국 스쿨 센서스 사례⁴³⁾를 통해 기록화 대상을 추정해 보고자 한다. 이 사례는 오래전에 수행되었던 영국과 웨일즈 지방 학교에 대한 설문조사 사례로 컴퓨터 매체에 기록된 것이 1975년으로 추정된다. 이 설문조사 데이터세트가 NDAD로 이관되었으나 몇 가지 문제점으로 인해 활용이 어려웠다. 첫째, 마이그레이션으로 인해 열(column)에 대한 설명이 잘려나가서 손실된 부분이 있었다. 둘째, 데이터사전의 설명 부족으로 약어를 사용한 경우나 열의 의미를 해석이 불가능한 경우가 발생하였다. 이러한 문제를 해결하기 위해 당시에 작성된 설문지 양식을 참고하여 각 열들의 의미를 최대한 유추하고, 당시의 교육통계 연간보고서 등을 참고하여 데이터세트의 의미를 복구할 수 있었다.

사례에서 보듯이 데이터세트는 데이터 내용만 보존되어서는 기록으로서의 해석이 불가능한 것을 알 수 있다. 각 열이나 코드, 약어들의 의미가 충분히 해석될 수 있는 자료를 확보하여 기술하는 절차가 필요하다. 또한 데이터세트는 시스템 의존적이므로 당시에 사용되었던 시스템에 관련된 문서 등을 함께 수집하여 정보를 제공하는 것이 필요하다. 이러한 경험을 바탕으로 영국의 NDAD는 데이터세트의 맥락을 알려주고 해석하는데 도움을 줄 수 있는 설계·개발 및 운영과정의 문서 및 메타데이터를 보존 대상으로 보고 있다.

42) 강재정 외, 『정보시스템의 이해』, 한경사, 2005

43) Darlington Jeffrey, 2004, 「A National Archive of Datasets」, 『Ariadne』 39



〈그림 5〉 NDAD의 데이터세트 기록화를 위한 보존 대상

DPT 프로젝트에서는 데이터베이스의 진본성을 보장하기 위한 최소 집합을 구체화하여 함께 보존할 대상으로 제시하고 있다. 데이터베이스와 사용자 프로그램으로 나누어 제시하고 있는데 그중 데이터베이스에서의 보존대상은 표와 같다⁴⁴⁾.

데이터베이스	
맥락	조직의 맥락정보 명세 - 조직명, 업무과정, 다른 파일간의 관계 로그 유지 - 원래 파일 형식과 현재 파일 형식에 대한 정보 - 현재 파일 형식의 해석에 필요한 정보 - 날짜와 시간과 같이 취해진 보존 행위에 대한 정보
내용	테이블의 실제 내용
구조	물리적인 구조 : 테이블명, 유형, 필드명, 테이블간의 관계, 뷰와 필드 속성(이름, 데이터 유형, 필드 길이 등) 데이터베이스의 논리적 구조 (예)ERD ⁴⁵⁾ 나 XML스키마로 표현
외형·행위	없음

〈표 3〉 DPT 프로젝트에서 제시하는 데이터베이스에서의 보존대상

행정정보시스템 기록관리의 대상에서 가장 기본적인 것은 데이터베이스에 저장되는 데이터 그 자체이지만 기록의 기본요소인 내용, 구조, 맥락을 포괄하기 위해서는 데이터 자체 외의 항목들이 함께 보존되어야 한다.

44) 이미화, 현문수 옮김, 『전자기록의 유형별 보존기법』, 한국국가기록연구원, 2006, pp.203~206

45) Entity-Relationship Diagram

활동의 증거가 되는 데이터를 함께 보존해야 한다. 시간이 지난 후에도 당시의 구조와 맥락을 살려 데이터세트의 내용을 재현하기 위해서는 데이터베이스에 저장되는 데이터는 보존되어야 하지만 데이터베이스관리시스템(DBMS) 자체는 보존대상에서 제외된다. DBMS는 상용제품인 경우가 많고 제품 의존적이므로 보존하는데 있어서 기능호환이나 라이선스 비용이 문제가 된다. 그러나 DBMS에서 제공하는 일반적인 기능은 보존환경에서도 활용과 관리를 위해 필요하다. 보존단계에서 어떤 도구 및 환경을 선택하느냐에 따라 영구기록관리 단계로 데이터세트 기록을 이관할 때의 이관포맷과 보존포맷이 영향을 받게된다. 이 때, 특정 DBMS 제품에 의존적인 포맷이나 특성이 이관대상 데이터에 반영되지 않도록 유의해야 한다. 데이터세트 기록의 맥락정보 확보를 위해 필요한 메타데이터와 데이터베이스 시스템의 설계, 개발, 운영과정 문서를 보존대상으로 포함해야 한다. 데이터세트는 대량의 데이터일 가능성이 많고, 다양한 방식과 목적으로 생산·활용되었을 가능성이 크다. 데이터베이스 시스템에서의 업무 절차 및 업무의 배경을 담은 문서, 시스템이 사용되던 방식을 담은 문서, 데이터세트의 활용방식을 담은 문서들을 통해 맥락정보를 도출할 수 있다. 데이터베이스 응용프로그램은 보존의 대상은 아니지만, 업무의 절차나 방식을 보여줌으로써 맥락적 정보로 활용하거나 추후 기록을 해석하거나 이해할 때 도움이 될 수 있도록 사용자지침서 등을 확보하는 것이 필요할 수 있다⁴⁶⁾. 또한, 사용자가 조회·활용을 위해 쓰던 화면의 형태가 아카이브 단계에서도 서비스를 위해 필요할 수 있다. 이 경우 화면 양식을 보존대상에 포함하거나 추후 서비스 방안과 연계하여 관리

46) 데이터세트 유형IV에서는 아카이브 단계에서 재현을 위한 방법으로서 응용프로그램이 보존의 대상이 될 수 있다.

방안에 포함시킬 필요가 있다. 따라서 데이터세트의 기록화 대상은 선별된 데이터세트, 업무와 시스템관련 문서류와 각각에 대한 메타데이터, 필요한 경우 응용프로그램에서 조회·활용을 위해 사용되던 화면 양식 등이 포함된다⁴⁷⁾.

이 장에서는 데이터세트 관리를 위해 먼저 생각해 보아야 할 데이터세트 식별의 개념, 시기, 기록화 대상 등 선별의 기준을 살펴보았다. 다음 장에서는 이러한 기준을 전제로 데이터세트의 선별 전략과 절차를 제시하고자 한다.

4. 행정정보시스템의 데이터세트 선별 절차

1) 전사적 아키텍처의 DRM에 기반한 데이터세트 식별

데이터세트 기록관리 체계 구축을 모색하기 위한 전제는 다음의 두 가지로 정리할 수 있다. 첫째, 데이터세트는 전자기록의 일종으로서 전자기록이 생산되기 이전의 시스템 설계단계부터 이러한 체계가 반영되어야 한다는 것이다. 둘째, 조직 전체의 정보관리계획 등의 전략적 계획에 기록관리 전략이 통합되어 수립해야 한다는 것이다. ISO 15489(8:1)는 정보관리 전략적 계획에 기록 관리전략을 통합해야 하며 조직 전체의 계획문서에 이 전략 계획을 통합해야만 한다고 권고하고 있다. 이러한 맥락에서 보면 우리나라 공공기관에서 실행하고 있는 전사적 아키텍처⁴⁸⁾

47) 특히 데이터세트 유형 II의 경우 이러한 필요성이 제기될 수 있다.

48) 전사적 아키텍처는 기업이나 기관에서 사용되는 업무 및 관리 프로세스와 정보기술 간의 관계를 표현한 것이다. 즉, 전사적 아키텍처란 조직의 전략적인

(EA, Enterprise Architecture)가 기관 전체의 정보전략 청사진에 해당하므로 데이터세트의 기록화 전략은 EA에 통합적으로 추진되어야 한다.

우리나라는 ‘정보시스템의 효율적 도입 및 운영 등에 관한 법률안’을 제정하여 정부 기관 및 공공기관에 의무적으로 EA를 구축하도록 하고 있다. 위 법률은 EA 방법론을 규정한 것으로 각 기관은 EA⁴⁹⁾ 도입계획을 수립하여야 하고 정부는 공공기관에서 공동으로 활용할 수 있는 EA 참조모형⁵⁰⁾을 개발하여 보급할 수 있도록 규정하고 있다⁵¹⁾. 그 결과 전자정부 추진으로 구축되는 행정정보시스템이 기관 내에 시스템을 도입할 때는 5가지의 EA 참조모형⁵²⁾을 참조하여 계획을 수립하고, 표준화, 상호참조, 재

방향, 업무절차, 정보의 흐름, 리소스 활용 등에 대한 행동가능하고 조화된 관점을 제공하는 관리프로그램이며 문서화 하는 방법을 의미한다. 선진국에서는 1990년대 초반부터 정보 기술에 의한 관리 개혁을 추진하면서 이 개념을 도입하여 사용하고 있다. 미국의 경우 정보화 분야의 투자 관리와 성과 관리, 그리고 정보시스템의 통합 및 상호 운영성 등을 관리하는 주요 수단으로 활용하고 있다

- 49) 정보시스템의 효율적도입및운영등에관한법률 제2조(정의) 2항 정보기술아키텍처“라 함은 일정한 기준과 절차에 따라 업무, 응용, 데이터, 기술, 보안 등 조직 전체의 정보화 구성요소들을 통합적으로 분석한 뒤 이들 간의 관계를 구조적으로 정리한 체계 및 이를 바탕으로 정보시스템을 효율적으로 구성하기 위한 방법을 말한다.
- 50) 정보시스템의 효율적도입및운영등에관한법률 제2조(정의) 5항 “참조모형”이라 함은 정보기술아키텍처의 일관성, 재사용성, 상호운용성 등을 확보하기 위하여 정보기술아키텍처의 구성에 필요한 정보화 구성요소의 표준화된 분류체계와 형식을 정의한 것을 말한다.
- 51) 정보시스템의 효율적도입및운영등에관한법률제5조
- 52) 조직 또는 기관이 전체의 정보시스템(IT) 또는 정보화를 효율적으로 구현하거나 재구성하기 위한 체제로서 조직의 성과를 최대화 할 수 있도록 지원하는 도구임. 한국전산원, 「법정부 데이터 참조모델 1.0(안)」, 2005, p.1~2 - 다시 말하자면 정보기술 아키텍처를 가지기 위해 범정부적인 차원에서 일관된 형식을 가지고 설계도를 가지는 것인데, 시스템 설계에 대한 방식으로서 5가지 측면에서의 일관된 설계도를 가지도록 제시하는 셈이다.

사용성을 고려하도록 되어 있다. EA의 참조모형 중에서 데이터 참조모형(DRM, Data Reference Model)이 범정부 데이터 설계 체계와 관련한 것이며 행정정보시스템의 데이터베이스 설계 및 구축과 직접적인 연관을 가진다.

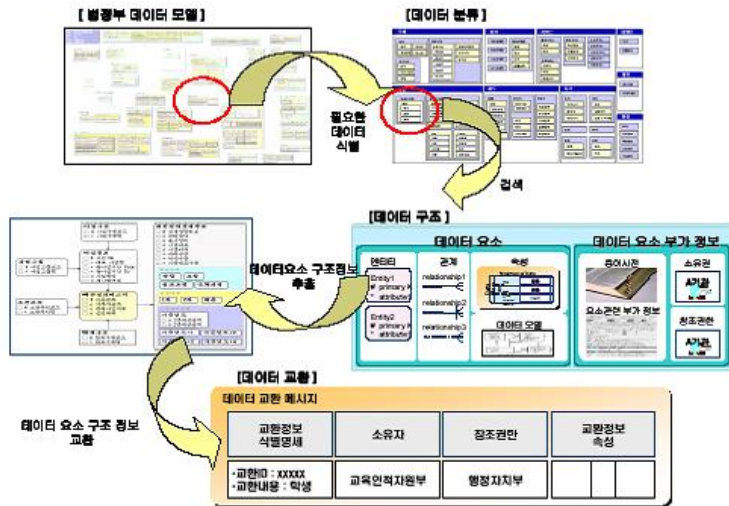
범정부 DRM은 데이터 아키텍처의 수립 및 관리를 지원하기 위한 참조모형으로서 범정부 데이터 체계의 표준화, 상호참조, 재사용을 목적으로 한다. 범정부 DRM에서는 데이터 아키텍처 구축 시 필요한 데이터 관련 정보 및 지원도구 지원도구를 분석하여 프레임워크를 제시하고 있다. 범정부 DRM 프레임워크는 데이터 모델, 데이터 분류, 데이터 구조, 데이터 교환, 데이터 관리의 5가지 요소⁵³⁾로 구성되어 있다. 범정부 차원의 표준화된 데이터 요

정보기술아키텍처를 위한 참조모형 종류 5가지

성과참조모형 (PRM, Performance Reference Model)	정보화성과의 측정을 위한 항목과 지표 및 방법을 정의
업무참조모형 (BRM, Business Reference Model)	업무아키텍처의 기준. 아키텍처 대상 기관의 사업 또는 업무 등을 전체적으로 분류하고 정의한 것
서비스참조모형 (SRM, Service Component Reference Model)	응용 아키텍처의 기준이 되며, 응용 서비스의 기능을 분류하고 정의
데이터참조모형 (DRM, Data Reference Model)	데이터아키텍처 기준. 기관 간에 교환되는 주요 데이터요소를 분석하여 이를 정의하고 표준화
기술참조모형 (TRM, Technical Reference Model)	기술 아키텍처의 기준이 되며, 정보기술을 분류 및 식별

53) ① 데이터 모델 : 범정부에서 구축 및 활용하고 있는 데이터를 식별하고 데이터 영역간의 관계를 도식화한 것으로서, 활동영역을 중심으로 한 큰 분류로서 주제, 환경, 제도, 지식자원, 물리적 자원 등으로 정의되어 있다. ② 데이터 분류 : 범정부 데이터에 대한 분류 체계이다. 내용적으로 성격이 유사한 항목을 묶어서 주제, 자원, 활동, 제도, 환경의 총 5개 데이터 주제 영역으로 분류하였고 하위에 데이터 그룹 영역으로 세분화되어 있다. 또한 다른 참조모형인 BRM과의 연계를 위해 데이터 분류체계의 활동-서비스 항목에서 범정부 BRM의 업무관점의 분류체계를 수용하였다. ③ 데이터 구조 : 범정부에서 구축 및

그러므로 EA를 참조하여 데이터 모델을 설계할 때 데이터세트의 식별을 함께 수행한다면 효과적·효율적인 작업이 될 것이다. 또한 데이터세트를 식별하는 과정이 행정정보시스템의 설계와 맞물려 진행되기 때문에 기록이 생산되기 전에 전체 정보화 전략과 맞물려 기록화 계획을 수립할 수 있게 되어 안정적이고 견고한 기록관리가 가능해진다.



〈그림 7〉 범정부 DRM 프레임워크 구성 요소간 기능관계

※출처 : 한국전산원, 「범정부 데이터 참조모델 1.0(안)」, 2005

한편 DRM에 기반한 데이터세트 식별에는 환경적 한계점과 내용적 한계점이 존재한다. 먼저, 내용적 한계점이란 DRM 자체가 갖고 있는 한계에서 기인한 것이다. DRM은 정보 모형만을 표현하고 있다. 즉, 기록이 가져야 하는 내용, 구조, 맥락 중에 내용에 대한 정보만을 표현하고 있는 것이다. 또한 DRM에서 개

체의 개념은 상호운용성이나 참조를 목적으로 만들어지는 것이기 때문에 개념상으로는 데이터세트와 많이 유사한 면이 있으나 규모나 범위에서 데이터세트의 단위와 일치하지 않을 수 있다. 따라서 DRM의 개체 개념을 참고하되 논리적, 물리적 대상은 따로 사상작업을 수행해야 한다. DRM의 개체가 데이터베이스의 테이블과 일치하는 경우 DRM 수립과 데이터세트의 식별을 함께 추진하는 것은 더욱 효율적이다. 결론적으로 조직에서 EA를 수립하는 과정에 기록화 대상인 데이터세트 식별의 과정이 추가되거나 함께 진행되는 것이 효율적이며, 데이터세트를 기록으로 선별하는 과정에서 기록연구사와 정보화담당자가 협력하여 기록에 대한 증거적 역할과 해석을 도와줄 수 있는 구조 정보와 맥락 정보를 추가적으로 확보하도록 해야 한다. 한편으로는 DRM이 상호운용, 재활용 등을 목표로 하고 있으므로 기록 관리 측면에서 모형의 한계가 있을 수 있으므로 중앙기록물관리기관을 중심으로 행정정보시스템 데이터세트 식별 및 관리에 대한 지침을 만들면서 EA의 DRM과 공유할 수 있는 부분을 정의해 나가는 것이 필요하다고 본다. DRM의 환경적 한계점은 현재 EA가 중앙부처를 중심으로 진행되고 있고 모든 기관으로 확대되어 구축된 것은 아니라는 점이다. 또한 EA를 구축한 기관에서 변화되는 내용을 현행화하는 작업을 충실히 수행했을 때 기록 관리에 활용이 가능하다는 것이다.

2) 데이터세트 기록화를 위한 선별 절차

이제까지 살펴본 내용을 토대로 행정정보시스템 기록화를 위한 데이터세트 선별 절차를 정리해 보면, (1)기록화 대상 데이터세트 결정, (2)데이터세트 기록 건 식별 단계, (3)데이터세트 기록

의 관리계층 구성 단계로 나누어 볼 수 있다. 각 단계별로 수행해야 하는 내용을 EA의 DRM과 관련하여 살펴보면 다음과 같다.

(1) 기록화 대상 데이터세트 결정

이 단계는 행정정보시스템별로 기록화할 데이터세트가 있는지를 판단하여 결정하는 과정이다. 기관의 업무를 파악하고 기록관리에 필요한 기반 사항을 파악하기 위해 업무나 정보시스템, 기록관리 관련된 각종 문서와 지침 수집하여 분석한다. 정보수집을 위해 분석단계에서 사용할 수 있는 참고정보원의 사례는 다음과 같다.

- 문서류 : 현업장표, 문서양식, 보고서
- 인터뷰/현장조사
- 시스템 자료 : 시스템관련 각종 매뉴얼, 데이터 구조/모형, 데이터와 관련한 내부문서시스템에 대한 보고서, 시스템 관련 출간물 등
- 관련 지침

기존 업무에서 사용하는 문서양식이나 인사기록카드와 같은 대장류 등은 데이터세트 식별에 중요한 자료가 되므로 중점적으로 수집하여 분석한다. 해당 업무가 전산화 될 때 이를 바탕으로 입출력 화면이나 보고서 출력 양식을 설계하는 경우가 많기 때문이다. 행정정보시스템의 메뉴 구성, 업무의 범위와 처리 방식 등도 데이터세트의 기록화 여부를 판단하는데 도움이 되는 자료로 함께 수집되어야 한다.

이 단계에서는 행정정보시스템이 기관의 업무 전체에서 차지하는 비중과 중요도를 평가하게 되며, EA의 DRM과 해당 시스템

의 데이터베이스에 대한 ERD를 참조하여 주요 데이터세트 후보 그룹을 식별한다⁵⁴⁾. 이때는 각 데이터세트를 구성하는 기본 개체들만 식별되는 상태이다. 이들 후보 데이터세트 중 현재의 체계에서 문서나 보고서로 가공되어 기록화 되고 있는 것과 그렇지 않은 것을 구분한다. 기록화되지 않고 있는 후보 데이터세트 중에서 업무 증거적 가치, 활용적 가치, 역사적 가치 등 기관의 평가 기준에 맞춰 기록화대상을 결정한다. 이미 기록화되고 있는 후보 데이터세트 중에서도 향후 데이터세트 형태로 아카이빙함으로써 연구자들이 용이하게 재가공하거나 활용할 수 있다고 판단되는 데이터세트는 기록화대상으로 결정한다. 예를 들어 인사기록카드의 ERD에서 개체들의 관계에서 개인신상, 개인임용, 교육훈련 등의 중심 개체들이 성별코드, 부처코드, 직급코드, 임용코드, 직위코드 등과 연관을 맺고 있음을 확인할 수 있으며 이 전체 덩어리가 후보 데이터세트가 될 수 있다.

성명		성별		주민등록번호	
한자		영문		이메일	
소속		직위		직급	
주소				전화번호	

근무경력

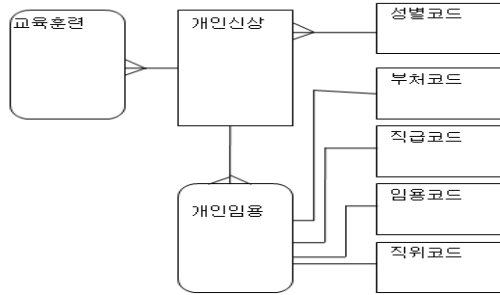
임용일	직급	임용구분	발령기관	임용종료일

교육훈련

기간	과정명	교육훈련기관

〈그림 8〉 인사기록카드 양식 샘플

54) ERD에서 개체 간에 Master-Detail 연관을 갖는 대상들을 식별하도록 한다. 이들은 외래키를 통해 연관된 정보를 구성하며, 하나의 주제영역(Subject Area)을 형성한다.



〈그림 9〉 인사기록카드 ERD 샘플

(2) 데이터세트 기록 건 식별 단계

이 단계는 기록화 대상으로 결정된 데이터세트 별로 기록 건을 식별해 내는 과정이다. 앞에서 살펴본 바와 같이 기록 건을 무엇으로 볼 것인지, 정해주는 과정으로, 관계형 데이터베이스 경우라면 어느 테이블의 어느 열로 구성되는 것인지를 결정하는 것이다.

하나의 데이터세트를 구성하는 테이블은 다수이며 각 테이블에는 다수의 열들이 있다. 데이터베이스를 설계하고 구축할 때 필요해서 구현한 테이블과 열이지만 실제 업무가 진행되면서 어떤 테이블과 열에는 정보의 입출력이 이루어지지 않는 경우도 발생한다. 그러므로 실제 각 열에 데이터의 입출력이 이루어지고 있는지를 확인해야 하며, 데이터의 정확성과 무결성, 신뢰도 등을 평가하여 기록 건에 포함시킬 지 여부를 결정해야 한다. 또한, 향후 이 데이터세트에 대한 이용과 활용 측면에서 필요한 데이터 항목인지를 평가해야 한다.

데이터세트 기록 건을 식별한 결과로 데이터베이스의 논리모델이 산출되어야 한다. 논리모델은 데이터세트의 기록 건을 구

성할 데이터 항목들의 실제 테이블의 컬럼명, 데이터타입 등 구조정보를 정리한 것이다. 이 때, 구조 정보의 정확성을 기하기 위해 실제 데이터베이스에 접속하여 테이블명과 컬럼명, 데이터타입과 길이 등의 정보를 데이터사전에서 찾아 확보하도록 한다. 열에 저장되는 데이터가 코드화된 값인 경우 해석정보를 찾아 함께 확보하도록 해야 한다.

컬럼명	속성명	데이터타입	데이터길이	키유형
reg_num	주민등록번호	VARCHAR	13	PK
name_hanja	한자설명	VARCHAR	30	
name_eng	영문성명	VARCHAR	30	
sex	성별코드	VARCHAR	2	
address	거주지	VARCHAR	100	
name	이름	VARCHAR	30	
phone_home	집전화번호	VARCHAR	15	

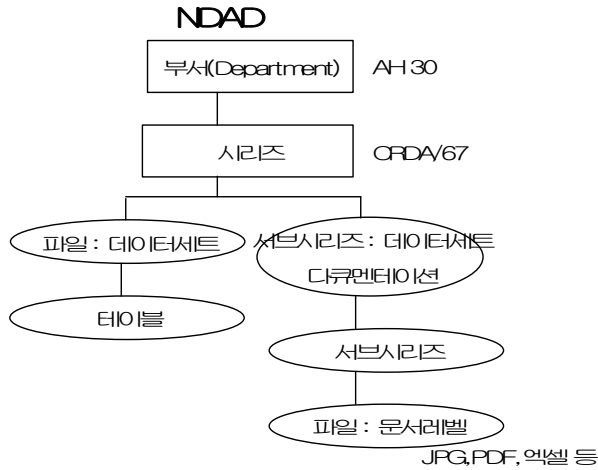
〈표 4〉 인사기록카드 데이터세트 논리 모델 샘플

(3) 데이터세트 기록의 관리계층 구성 단계

데이터세트의 기록 건 식별이 끝나면 해당 데이터세트를 관리할 계층 구조를 결정해야 한다. 데이터세트의 관리 계층을 설정하는 이유는 업무적 연관성이나 생애주기에 따라 통제가 효율적으로 가능하고, 이용자가 기록을 이해하는데 도움을 주기 위해서이다.

NDAD 홈페이지와 정책문서를 분석하여 데이터세트의 계층 구조를 구성해보면 다음 그림과 같다. 영국에서는 데이터세트, 데이터세트의 맥락정보를 제공할 수 있는 문서를 기록관리 대상으로 지정하고, 이에 대한 메타데이터와 함께 획득하여 아카이빙으로 이관 받는 체제를 가지고 있다. NDAD 데이터세트 계층구조는 논리적 연관성을 가진 데이터세트가 하나의 단위로서

계층을 이루게 되고, 하위에는 데이터베이스의 체계인 테이블-필드 등의 항목으로 구성된다. 해당 데이터세트와 관련된 문서들이 있다면 데이터세트와 함께 연관 지어서 관리되게 된다. 관련문서도 웹에서 전자파일로 제공되는데, 전자파일로 입수한 것은 그 형태대로 서비스 되고, 종이로 입수된 문서들은 스캔하여 JPG, PDF 등의 포맷으로 낱장별로 온라인 열람이 가능하다.



〈그림 10〉 NDAD 데이터세트 계층구조

여론조사 데이터의 경우 데이터세트 기록 건 식별 단계는 상대적으로 간단하다고 할 수 있다. 통계조사를 위한 항목들이 데이터베이스의 구조에 맞도록 사전에 메타데이터들이 정해지고, 그 항목들을 채워나가는 방식으로 통계조사가 수행되기 때문에 기록화 대상 항목이 쉽게 식별되기 때문이다. 또한 개체간의 관계가 복잡하게 얽혀져 있지 않고 상대적으로 단순하다. 하지만 여론조사가 매년 반복되거나 비슷한 유형의 조사가 변수를 변

경하여 반복적으로 수행되는 경우 조사 결과들을 건으로 하는 하나의 시리즈로 구성할 지, 혹은 별도의 기록 건과 시리즈로 분리하여 구성할 지를 결정하는 일은 단순하지 않다. 건과 시리즈의 계층구조가 제대로 정해져야 추후에 서비스 시점에 유용하게 사용될 수 있을 것이다. 또한 여론조사 데이터에서는 어떤 사안에 대한 조사였는지, 모집단은 무엇이었는지, 조사대행기관은 어디였는지 등 조사데이터의 맥락이 충분히 될 수 있도록 자료를 확보하는 것이 필요하다. 여론조사와 같은 통계자료는 아카이브의 서비스 단계에서 연구자의 필요에 따라 원하는 자료를 다양한 방식으로 추출가능하다는 장점을 가진다. 따라서 연구자가 자신의 필요성을 구체적으로 표현하고 인지할 수 있도록 이에 대한 관련 자료를 확보해 주는 것이 필요하다. 예를 들면 여론조사에 대한 배경, 수행시의 여러 가지 변수들, 여론조사가 수행되고 난 후의 보고서 및 분석결과, 관련된 출판물 등을 충분히 확보하고 메타데이터와 각 필드들에 대한 충분한 설명 자료를 확보해야 한다.

5. 맺음말

전자정부가 지속적으로 추진됨에 따라 행정정보시스템의 도입과 개선은 지속적으로 이루어질 것이다. 이와 더불어 행정정보시스템에는 각종 업무트랜잭션의 데이터셋이 축적될 것이다. 기록화해야 할 데이터셋을 선별하여 체계적으로 관리하지 않았을 때 관련 기관이나 국가 전반에 어떤 손실과 위험이 있을 지를 세밀히 따져볼 시점이다. 이제부터라도 중앙기록물

관리기관이 주체가 되어 데이터세트의 기록화 절차와 방법을 수립하여 각 공공기관에 지침으로 전달하고 데이터세트 아카이브의 선진적 모형을 만들어가야 할 것이다.

이 논문에서는 행정정보시스템의 데이터세트 개념을 명확히 하고, 데이터세트를 기록화 하고자 할 때 선행되어야 할 기록건 식별의 개념을 정의했으며, EA의 DRM을 활용한 데이터세트 식별의 유용성을 제안하였다. 또한, 데이터세트의 기록화를 위한 선별과정을 1)기록화 대상 데이터세트 결정, 2)데이터세트 기록건 식별 단계, 3)데이터세트 기록의 관리계층 구성 단계로 나누어 제시하였다. 이 과정에서 데이터세트 선별의 과정을 효과적 효율적으로 진행할 수 있기 위해 필요한 부가적 조치를 제안하면 다음과 같다.

첫째, 전자정부 추진 계획에 행정정보시스템의 데이터세트 기록관리에 대한 고려사항을 포함시켜야 한다. 특히, 행정정보시스템의 생애주기 관리와 관련하여 데이터세트의 기록화 방안 및 관리 절차를 마련하고 이를 보장하기 위한 관련 법제 마련이 시급하다.

둘째, 기관별 EA구축 시 각 행정정보시스템의 기록화 대상 데이터세트 식별작업을 포함하도록 해야 한다. 또한, 데이터세트 기록화에 유용하게 활용할 수 있도록 각 기관의 EA정보가 현행화 꾸준히 현행화될 수 있도록 조치가 필요하다.

셋째, 중앙기록물관리기관은 기관에서 참고할 수 있는 행정정보시스템 데이터세트 기록관리 지침을 개발·보급해야 한다. 데이터세트의 기록 관리에 대한 정책수립과 반영은 시스템을 설계하면서 함께 이루어져야 한다. 현재로서는 각 부처에서 행정정보시스템을 개발하면서 기록관리 요건을 반영하려고 해도 마땅한 지침이 없는 상황이다. 현업의 혼선을 줄이고 전체적인

체계의 확립을 위해 행정정보시스템 데이터세트 기록화를 위한 지침의 작성이 시급하다.

넷째, 기관 내의 부서 간 역할 체계를 정립이 필요하다. 데이터세트 선별 및 기록화 절차를 수행할 때 현업부서와 기록관리부서, 정보화부서 간의 협력이 중요하다. 데이터세트를 식별하는 과정은 논리적으로는 기록의 범위를 선별하는 행위이며 기관의 업무분석과 행정정보시스템의 기능분석을 통해 이루어져야 한다. 업무분석을 수행하는 기록연구사가 정보화부서와의 연계와 협조를 통해 시스템의 기능분석과 정보 분석에 접근할 수 있어야 한다. 기관의 기록관과 정보화부서 간의 적극적 협업을 통해서만이 데이터세트의 기록관리가 가능하다는 것을 인식이 필요하다.

이상에서 언급한 내용을 역할체계로 정리하면 다음과 같다⁵⁵⁾.

- 중앙기록물 관리기관 : 지도감독, 지침작성, 이관, 아카이빙
- 기록관 : 데이터세트의 논리적 식별, 생산현황보고, 이관
- 정보화부서 : 데이터세트의 논리적/물리적 식별, 기록관리 단계에서 기술적인 결정과 집행

중앙기록물관리기관의 역할은 지도감독과 이관에 대한 책임을 지는 것이다. 이를 위해 행정정보시스템 기록관리를 위한 지침을 만들어 각 기관에 배포하고, 각 기관에서는 이 내용을 기본으로 하여 행정정보시스템을 구축할 때 초기에 이러한 내용을 반영해서 효과적, 효율적으로 기록관리를 수행하도록 해야 한다. 이미 많은 숫자의 행정정보시스템이 도입된 상황이기 때문에 이들에 대한 이관 및 기록관리 대책을 수립하는 한편, 계

55) 이 논문의 범위에는 포함되지 않았던 기능도 역할체계에 일부 포함하였다.

속적으로 도입되고 있는 대규모의 행정정보시스템을 고려할 때 우선적으로 지침을 만들어서 향후 개발될 행정정보시스템부터 이 지침을 따를 수 있도록 하는 조치가 필요하다. 시스템의 특성상 나중에 이관을 받기 위한 기능이나 계획을 수립하려면 더 많은 비용이 들게 되며 현실적으로 시스템을 변경하는 것이 용이하지 않음을 고려해야 할 것이다.

기록관의 기록연구사가 업무분석 및 데이터세트의 논리적 식별을 주로 담당하고 기관의 정보화 담당자가 데이터세트의 물리적 식별 및 기술적인 집행을 하는 업무분장이 가능하다. 데이터세트의 논리적 식별이라는 것은 업무적으로 연관성을 가지는 기록의 범위를 설정해 주는 것으로 기록의 관점에서 필요한 항목들의 제안, 연계되어 관리되어야 할 기록들의 판별 등의 활동을 통해 식별의 논리적인 범위를 결정하는 것이다. 이것을 시스템에 구현할 때 생길 수 있는 여러 문제점을 정보화담당자로부터 환류받아 다시 반영하는 등의 협의과정이 필수적이라고 할 수 있다.

행정정보시스템 데이터세트의 기록관리를 위해 이 논문에서 제시한 데이터세트의 선별 이후의 단계에 대해 심도 깊은 연구가 필요하다. 데이터세트의 기록관리기준을 도출하고, 이관체계를 수립하고, 아카이브 모형을 구축하여 서비스하는 테스트베드의 구축이 필요하다. 초기 전자정부 추진과정에서 구축된 행정정보시스템의 데이터들이 시스템의 노후화나 관리부실로 망실되기 전에 국가차원의 데이터세트 기록관리 체계가 정립되어 많은 유용한 데이터세트들이 수요자와 만나게 되길 기대한다.

ABSTRACT

A Study on Record Selection Strategy and Procedure in Dataset for Administrative Information

Cho, Eun-Hee · Yim, Jin-Hee

Due to the trend toward computerization of business services in public sector and the push for e-government, the volume of records that are produced in electronic system and the types of records vary as well. Of those types, dataset is attracting everyone's attention because it is rapidly being supplied. Even though the administrative information system stipulated as an electronic record production system is increasing in number, as it is in blind spot for records management, the system can be superannuated or the records can be lost in case new system is developed. In addition, the system was designed not considering records management, it is managed in an unsatisfactory state because of not meeting the features and quality requirements as records management system.

In the advanced countries, they recognized the importance of dataset and then managed the archives for dataset and carried out the project on management systems and a preservation formats for keeping data. Korea also is carrying out the researches on an dataset and individual administrative information systems, but the official scheme has not been established yet.

In this study the items for managing archives which should be reflected when the administrative information system is designed was offered in two respects - an identification method and a quality requirement. The major directions for this system are as follows.

First, as the dataset is a kind of an electronic record, it is necessary to reflect this factor from the design step prior to production. Second, the system should be established integrating the strategy for records management to the information strategy for the whole organization.

In this study, based on such two directions the strategies to establish the identification for dataset in a frame to push e-government were suggested. The problem on the archiving steps including preservation format and the management procedures in dataset archive does not included in the research contents. In line with this, more researches on those contents as well as a variety of researches on dataset are expected to be more actively conducted.

Key words: dataset, BIS(Business Information System), Enterprise Architecture, e-governance