

행정정보 데이터세트 기록의 관리방안

A Study on Managing Dataset Records in Government Information Systems

왕 호 성 (Ho-sung Wang)*

설 문 원 (Moon-won Seol)**

목 차

- | | |
|------------------------|--------------------|
| 1. 서 론 | 4. 데이터세트 기록의 관리 방안 |
| 2. 데이터세트 기록의 특징과 관리 현황 | 5. 결론 및 제안 |
| 3. 데이터세트 기록과 재현 | |

〈초 록〉

최근 조사에 의하면 공공기관이 사용하고 있는 전체 행정정보시스템의 수는 18,000여 종이 넘는다. 공공기록물 관리에 관한 법률은 모든 행정정보 데이터세트를 기록이라고 선언하고 있지만, 현행 전자기록관리정책으로 데이터세트 기록을 관리하기는 매우 어려운 실정이다. 이 논문은 데이터세트 기록을 관리하기 위한 정책방향을 제시하기 위한 것이다. 데이터세트 기록의 진본성은 기록의 내용, 구조, 맥락뿐 아니라 외형과 기능을 함께 보존함으로써 확보되며, 데이터세트를 전자객체가 아니라 전자기록으로 관리하기 위해서는 재현성이 전제되어야 함을 강조하였다. 또한 데이터세트 기록의 재현 및 장기보존 전략으로서 에뮬레이션의 가능성을 제시하였다. 이를 토대로 데이터세트 기록의 관리 모형과 정책 방향을 제안하였다.

주제어: 전자기록, 데이터세트, 장기보존, 에뮬레이션, 클라우드, 재현성, 외형 재현, 기능 재현, 행정정보시스템

〈ABSTRACT〉

According to a recent survey, over 18,000 government information systems have numerous different functions and characteristics. Although every dataset that is created and maintained in government information systems is declared as a collection of records according to the Public Records Management Act, current electronic records management policies cannot cover dataset records management. This study suggests the policy directions for dataset records management at the national level. It emphasizes the necessity to preserve the appearance and behavior (function) of database systems to ensure the authenticity of dataset records. In addition, this study investigates "emulation" as a representation and long-term preservation methodology for dataset-type records. It also suggests a dataset records model.

Keywords: electronic records, dataset, long-term preservation, emulation, cloud, representation of appearance, representation of behavior, government information systems

* 국가기록원 기록연구사(kinghosung@gmail.com) (제1저자)

** 부산대학교 문헌정보학과 부교수(seol@pusan.ac.kr) (교신저자)

■ 접수일: 2017년 7월 22일 ■ 초심사일: 2017년 8월 2일 ■ 게재확정일: 2017년 8월 19일

■ 한국기록관리학회지 17(3), 23-47, 2017. <<http://dx.doi.org/10.14404/JKSARM.2017.17.3.023>>

1. 서론

1.1 연구의 목적

현재 공공기관의 전자기록관리체계는 2006년 전면 개정된 「공공기록물 관리에 관한 법률」(이하 공공기록물관리법)에 토대를 두고 있다. 공공기록물관리법에서는 행정정보 데이터셋을 전자기록의 한 유형으로 구분하고 관리대상에 포함하였지만 10여 년이 지나도록 데이터셋을 기록으로 관리할 수 있는 구체적인 정책이나 절차가 결정된 바 없다. 명목과 달리 실제로는 기록으로 관리되지 못하고 있는 것이다. 현재 전자기록관리의 모든 절차와 방침은 전자문서관리시스템과 업무관리시스템에서 생산되는 전자문서 유형에 맞추어져 있어서 다양한 데이터셋 기록에는 적용할 수 없다.

조사에 의하면 공공기관의 행정정보시스템은 약 18,000여 종이 넘는다(국가기록원, 2016). 그러나 현재 전자기록관리정책은 전자문서관리시스템과 업무관리시스템이라는 2종의 시스템에서 생산되는 전자문서를 위한 정책이 전부라고 해도 과언이 아니다. 가령, 현재 전자기록은 법률에 따라 3단계 이관이 이루어져야 한다. 전자문서관리시스템과 업무관리시스템에서 생산되는 전자문서들은 기록관리시스템을 거쳐 영구기록관리시스템으로 기록관리 메타데이터와 함께 물리적으로 전송되는 3단계 이관을 거치도록 되어 있다. 이렇게 3단계 이관방식을 채택했던 이유는 전자기록관리체계가 단순히 “종이문서가 전자매체로 전환된 것”이라는 사고를 토대로 기록의 생애주기적 관점을 적용한 결과라고 볼 수 있다. 그러나 수많은 행정정보

시스템에서 생성되는 다양한 데이터셋에 이러한 방식의 물리적 이관을 적용하려면 관리대상 선별과 추출 등 수많은 문제가 해결되어야 할 뿐만 아니라 현실적으로 불가능한 경우도 있다. 현행 법령상 전자기록물 이관시 수행해야 하는 장기보존포맷 변환과 같은 보존전략이 적합하지 않은 데이터셋 기록도 많을 것이다.

본 연구는 전자기록의 단계적 관리와 물리적 보존에 집착하는 현재의 생애주기적 관리체계가 데이터셋과 같은 유형의 전자기록에 적용되어서는 안 된다는 문제의식에서 출발한다. 많은 경우에 데이터셋 기록은 행정정보시스템 내에서 구현되던 기능이 함께 재현되어야만 ‘업무 반영물로서 기록’의 의미를 유지할 수 있기 때문에 데이터셋 기록 관리를 위해서는 ‘재현성’에 특히 주목해야 한다는 것이 본 연구의 기본 입장이다. 아키비스트가 행정정보시스템의 수많은 데이터 중에서 데이터셋 기록을 선별하는 것도 에뮬레이션을 통한 재현 없이는 불가능할 것이다. 이러한 입장과 문제인식을 토대로 본 연구에서는 데이터셋 기록의 관리 요건으로서 재현성을 검토하고, 아직 확보되지 못한 행정정보 데이터셋 기록의 관리 방안을 모색하고자 한다.

1.2 선행연구

데이터셋 기록과 관련된 연구는 이미 많이 소개되었지만 아직 정책에 반영하여 실무에 적용할만한 수준의 구체적 방법론 논의는 부족한 편이다. 초기 연구들은 대체로 데이터셋 기록 관리의 필요성을 강조하고 공공기록의 누락 없

는 관리를 위한 정책 반영의 당위성을 강조하는 내용들이 주류를 이룬다. 후속 연구에서는 데이터세트 기록의 실체를 파악하기 위하여 데이터베이스의 구조를 파악하고 이를 구성하는 데이터 요소들 중 기록으로 관리되어야 하는 선별단위에 대한 논의들이 전개된다. 최근에는 빅데이터와 오픈데이터를 구분하여 데이터세트 기록을 연구한 InterPARES Trust의 산출물이 소개되었고 전자기록관리 정책 재설계를 위한 연구가 본 연구에 앞서 소개된 바 있다.

이규철(2016)은 데이터세트 기록의 관리방식을 네 가지로 구분하여 제안한다. 첫째, 데이터세트의 유형을 데이터의 구조화 여부·데이터베이스의 처리·데이터베이스의 업데이트 방식으로 구분한다. 둘째, 데이터세트 기록의 수집방식을 이관·이관 후 수집·수집의 세 가지로 제시한다. 셋째, 데이터세트 기록의 구조가 기록의 철-건 구조와 다르다는 것을 전제로 데이터세트의 단위기능을 기록관리를 위한 기준 계층으로 제시한다. 넷째, 데이터세트 기록이 기존의 보존포맷 방식을 적용할 수 없는 경우 METS (Metadata Encoding and Transmission Standard)나 스위스의 SIARD(Software Independent Archiving of Relational Databases)¹⁾ 방식 등을 고려해야 한다고 주장하였다.

이승억, 설문원(2017)은 행정정보데이터세트 기록관리의 취약성이 전자 기록관리 정책 미비에서 기인하는 것으로 보고 이에 대한 해결책을 다음과 같이 제안한다. 첫째, 데이터세트 기록관리를 위해서는 전자기록관리 패러다임의 변화가 필요하며 데이터세트 기록에 대한 논리

적 이관 개념을 수용하고 시스템 환경에 대한 기술적 이해가 필요하다고 주장한다. 둘째, 등록·생산현황보고, BRM 체계의 단위과제별 분류·보존기간 책정방식, 철-건으로 대표되는 계층적 편철방식 등 기존의 기록 관리 프로세스가 데이터세트 기록에는 맞지 않음을 지적하며 전자적·기술적 방식으로의 변경을 제안한다. 셋째, 전자기록의 신뢰체계와 논리적 통제체계 구축을 위해 데이터세트 기록의 조건표를 개발·활용하는 방식을 제시한다. 넷째, 지속가능한 전자기록관리체계 유지를 위해 영구기록관리시스템(CAMS)을 모듈 단위로 재구축할 필요가 있다고 주장한다. 또한 국가기록원의 역할이 현재와 같은 표준시스템의 공급자에서 플랫폼의 제공자로 변경될 필요가 있다는 점을 강조한다. 이승억, 설문원(2017)의 연구는 논리적 이관과 통제, 전자기록관리 전반의 프로세스 재설계, 서비스 공급자로서의 아카이브의 역할 전환 등 현행 전자기록관리체계의 전면 혁신을 제안하고 있다는 점에서 의미를 찾을 수 있다.

McDonald와 Leveille(2014)는 빅데이터와 오픈데이터 보존기간과 처분에 관해 연구하였다. 양 데이터의 상이한 목적을 구분하고 웹 환경에서의 기록관리는 무결성과 신뢰성 확보가 중요함을 제기한다. 특히 아카이빙 대상인 데이터세트 기록의 단위는 개별 데이터가 아닌 데이터세트 전체가 되어야 함을 특히 강조하고 있다. 이는 만약 데이터세트를 기록으로 보존해야 한다면 데이터 간의 복잡한 관계를 반영하는 질서와 기준을 분리하지 말고 만드시 전

1) SIARD는 Oracle, MS SQL, MySQL, DB/2, MS Access 등의 관계형 데이터베이스를 아카이빙하기 위해 zip으로 압축한 개방형 포맷임.

체 데이터를 하나의 세트로 보존해야 한다는 것이다.

박옥남, 박희진(2016)은 전자기록관리 국제 연구인 InterPARES 3와 ITrust를 분석하였다. InterPARES에서는 진본성이 기록이 표방하는 바 그대로의 정체성과 변조되지 않거나 훼손되지 않은 무결성을 유지하는 상태를 의미하는 반면, ITrust는 인터넷과 클라우드 환경에서는 '신탁(Trust)'이 중요한 화두라고 강조한다. 여기서 신탁은 특정 행동이나 혜택에 관한 가치시스템을 기반으로 성립된 상호 간의 신뢰나 상대방에 대한 신뢰라는 의미로 제시된다.

곽정(2016)은 InterPARES와 ISO15489의 진본성 개념을 검토하면서 전자기록의 진본성을 “재현에 의해 동일한 것으로 인지되는 외형적 무결성”으로 보아야 한다고 주장하였다. 여기서 전자기록의 진본성은 정체성, 무결성, 재현성이 결합된 개념이며, 생산당시의 모습 그대로 무결한지를 판단하려면 재현성이 담보되어야 한다는 점을 강조하였다(곽정, 2016).

데이터세트 기록의 장기보존 전략인 에물레이션과 관련해서는 김명훈, 오명진, 이재홍, 임진희(2013), 임진희, 최주호, 이재영(2014)의 해외 사례연구가 있다. 이들 연구에서는 CAMiiEON, KB, Planets, KEEP 등에 적용된 디바이스, 운영체제, 응용 프로그램에 대한 에물레이션 방식을 구체적으로 소개하고 있다. 이밖에 주목해야 할 해외 사례로는 독일 bwFLA (Functional Long-Term Archiving)가 있다(bwFLA, 2017). 독일의 bwFLA는 전자기록이 그 외형과 기능을 생산 당시 그대로 재현하는 것이 필수적임을 전제하면서 전자기록의 장기보존방안을 에물레이션에서 찾는 본격적인 해외사례이다. 2016년

7월에 출범한 미국의 전자기록 아카이빙에 관한 공동 프로젝트 BDAX(Born Digital Archiving & eXchange) 팀도 전자기록의 장기보존에 대한 전략으로 bwFLA의 에물레이션 도구를 적용하겠다고 발표한 바 있다. 데이터세트 기록의 기록관리 방안을 설계해야 하는 상황에서 이들의 연구결과는 우리에게 시사점을 줄 것으로 기대된다.

1.3 연구의 범위와 방법

본 논문에서는 전자기록을 논리적 실체로 인식하는 기록연속체적 관점에서 데이터세트 기록관리 방안을 연구하였다. 특별히 기록물로 표현해야 하는 경우를 제외하고는 '기록'으로 기술한 이유는 이 관점의 투영이다. 또한 전자기록은 전자적 속성과 기록으로서의 속성을 모두 갖지만 기록으로서의 속성이 간과된 관리가 이루어져서는 안 된다는 점을 강조하고자 하였다.

ISO 16175-3의 정의에 따르면 디지털객체(digital object)는 특정 시스템 혹은 애플리케이션에 의해 생성되는 파일 유형과 같이 컴퓨터에 의해 재현될 수 있는 객체(워드프로세스 문서, 스프레드시트, 이미지)이고 컴포넌트(component)는 디지털 문서와 같은 디지털 객체이거나 데이터베이스 내에 있는 하나의 데이터 요소가 될 수 있다. 전자기록은 하나 이상의 디지털객체와 컴포넌트로 구성되며, 업무 또는 행위와 의미적으로 관련될 수 있는 최소 단위는 기록물건이다. 따라서 전자기록이 하나의 전자객체로 구성될 수도 있지만 전자객체 자체가 기록물은 아닌 것이다. 따라서 본 논문에서는

아직 기록관리 대상으로 확정되지 않은 전자객체와 전자기록을 구분하여 접근하였다.

그러나 한편으로 모든 전자기록은 정형·비정형 등의 형태와 관계없이 정보시스템의 데이터세트 상태로 존재한다는 포괄적인 관점을 견지하였다. 이러한 관점에서 볼 때 전자문서시스템이나 업무관리시스템의 전자문서도 데이터세트 기록의 일종이라고 볼 수 있다. 그러나 이 연구에서는 현재 기록관리 차원에서 방안이 마련되어 있는 전자문서시스템이나 업무관리시스템의 전자문서 외의 행정정보 데이터세트를 주요 대상으로 하였다.

이 연구에서는 데이터세트 기록관리 과정 전체를 포괄하는 방법론이 아니라 데이터세트 기록관리를 위한 '접근법'을 제안하는 데에 초점을 두었으며, 이러한 차원에서 데이터세트가 전자객체에서 전자기록으로 선별되는 과정, 선별의 단위 확정이 어떻게 마련되어야 하는지, 그 과정에서 기록관리가 어떻게 이루어져야 하는지를 다루고자 하였다. 구체적인 실증 사례 분석이 이루어지지 않았다는 점에서 이 연구는 시론적 성격을 가지지만, 더 이상 미룰 수 없는 당면 과제로 떠오른 데이터세트 기록관리 정책 개발에 시사점을 줄 수 있을 것으로 기대한다.

2. 데이터세트 기록의 특징과 관리 현황

2.1 행정정보 데이터세트 기록의 범위와 특징

행정효율과 협업 촉진에 관한 규정(이하 '신

사무관리규정')에 따르면 전자문서시스템은 문서의 기안에서 활용까지 전 과정을 전자적으로 처리하는 시스템으로 정의한다. 업무관리시스템은 과제관리카드·문서관리카드 등을 이용하여 전자적으로 관리하는 시스템으로 정의한다. 이 규정에서 행정정보시스템은 행정기관이 행정정보를 생산에서 활용까지 전 과정에서 이용하는 하드웨어·소프트웨어·데이터베이스 등을 통합한 시스템이며, 전자문서시스템·업무관리시스템을 제외한 나머지 모든 시스템을 의미한다고 볼 수 있다.

공공기록물관리법에서는 전자문서, 웹기록물, 행정정보 데이터세트를 전자기록물로 보고, 행정정보 데이터세트가 행정기관에서 사용하는 행정정보시스템의 데이터라고 정의한다. 여기서 전자문서는 앞의 '신사무관리규정'에서 정의한 전자문서시스템·업무관리시스템에서 생산되는 전자문서를 의미하고 웹기록물은 공공기관의 홈페이지를 의미하므로 특정된 이 두 종을 제외한 모든 전자기록물은 행정정보 데이터세트가 된다. 또한 공공기록물관리법은 생애주기 구분에 따라 기록생산-기록관리-영구기록관리의 단계를 정한 후 전자문서시스템, 업무관리시스템, 행정정보시스템을 기록생산시스템으로 분류한다.

두 법규의 정의를 종합적으로 해석해 보면 전자문서·웹기록·데이터세트 기록을 병렬적으로 구분하지만 사실상 행정정보 데이터세트는 전자문서와 웹기록을 제외한 나머지 전자기록의 통칭으로 보아야 할 것이다. 여기서 전자문서와 웹기록만이 특칭된 이유는 이 두 종류의 전자기록에 대한 관리 방안은 확정²⁾되어 구분이 가능하기 때문일 것이다. 만약 데이터세트 기

록 중 전자문서나 웹기록과 같은 특정 유형, 예를 들어 전자메일이나 SNS의 기록관리 방안이 확보된다면 전자기록의 정의는 “전자문서, 웹기록물, 전자메일, SNS 그리고 행정정보 데이터세트 기록”으로 변경될 것이다.

정보통신 분야의 정의에 의하면 데이터세트는 이름이 붙은 실제적 레코드들의 집합을 가리키는 것으로 어떤 규칙에 따라 배열된 데이터의 집합이다(네이버 IT 용어대사전).³⁾ 미국 SAA 용어사전에서는 데이터세트를 데이터파일 및 데이터베이스와 종종 동의어로 사용되며 컴퓨터가 분석할 수 있도록 구조화된 관련 정보들의 집합으로 정의한다.⁴⁾ 기록학용어사전은 “컴퓨터가 처리하거나 분석할 수 있는 형태로 존재하는 관련 정보의 집합체”라고 데이터세트를 정의한다(한국기록학회 편, 2008). 한편 공공기록물관리법에서는 행정정보 데이터세트를 “각급 행정기관에서 업무상 사용하고 있는 행정정보시스템에서 생산되는 문자, 숫자, 도형, 이미지 및 그 밖의 데이터”라고 정의한다(공공기록물관리법 시행령).

데이터베이스는 논리적으로 연관된 하나 이상의 자료의 모음으로 그 내용을 고도로 구조화함으로써 검색과 갱신의 효율화를 목적으로 한다(한국데이터진흥원, 2017). 이러한 개념을 데이터세트에 적용하면 데이터세트는 검색과 접근이 용이하도록 구조화된 데이터들의 집합체로서 정보시스템의 RDBMS에 연결된 데이터베이스 상태로 존재한다. 공공기록물관리법

에서 데이터세트 기록은 행정정보 데이터세트를 지칭하지만 실체적으로는 정보시스템에 연결된 데이터베이스로 존재하므로 그 특징을 파악할 필요가 있다.

데이터베이스는 데이터에 대한 실시간 접근과 동시 공유 등을 목적으로 고도로 구조화되어 구축되기 때문에 검색·갱신 효율이 매우 높다는 특징을 가진다. 또한 데이터의 일관성·무결성·보안성을 유지하기 좋은 반면 데이터의 복잡성으로 인해 장애 발생시 백업과 복구가 어렵다는 점도 특징으로 들 수 있다. 데이터베이스는 응용 프로그램들이 직접 접근하여 조작하는 것이 아니라 별도의 소프트웨어인 DBMS의 기능을 통해 데이터 접근·저장·갱신·삭제·읽기 등을 제어한다. DBMS는 인터페이스를 통해 이 기능을 제공하지만 접근 인터페이스와는 별도로 데이터베이스의 검색과 관리, 데이터베이스 스키마의 생성과 수정 등을 위해 SQL⁵⁾(structured query language) 언어를 사용하여 이 기능들을 조정할 수 있다.

검색과 갱신 효율, 즉 활용을 목적으로 구축되는 데이터베이스는 데이터 중복 최소화, 최신 데이터 유지, 데이터 저장 공간 확보, 일관성 유지 등을 위해 데이터 수정, 테이블 생성 등 일련의 동적인 변경 활동을 지속적으로 수행할 수밖에 없다. 이때 변경 대상 데이터나 테이블은 이러한 활동 과정에서 규칙 없이 발견되거나 생성되는 불특정 데이터일 가능성이 높기 때문에 미리 예측하여 사전에 폐기하거나 변경하기

2) 국가기록원의 연도별 기록물 관리지침에는 이 두 유형의 기록물에 대한 관리방안이 포함되어 있음.

3) terms.naver.com

4) www2.archivists.org/glossary/terms/d/data-set

5) ISO는 SQL을 약어가 아닌 고유명사로 사용하여 표준화하고 있음(ISO 9075-1992).

가 쉽지 않을 것이다.

현재 가장 폭넓게 사용되는 RDBMS는 키(key)와 값(value) 등 데이터요소를 테이블 단위의 행과 열에 배치하고 각각의 테이블들을 일정한 규칙에 따라 관계를 맺어 놓은 것이다. 데이터베이스의 규모가 크면 클수록 테이블과 데이터요소는 증가하고 그 관계와 구조를 정의한 스키마 역시 그에 비례하여 더욱 복잡해진다. 장애로 인해 특정 데이터가 손실되더라도 이러한 복잡성이 백업이나 복구를 매우 어렵게 할 것이다.

한편 기존 RDBMS에서는 구조화하기 어려운 비정형 데이터가 존재한다. 비정형 데이터는 형태와 구조가 다양하고 복잡해 정형화 되지 않은 데이터를 말하는 것으로 문서·그림·영상·이메일·트위터 등이 대표적이다(네이버 IT용어대사전). NoSQL과 NewSQL DBMS는 이러한 비정형 데이터를 구조화하기 위하여 개발된 것으로 정형화된 구조를 갖지 않은 빅데이터 분석에서 주로 활용된다. 이처럼 현재의 IT 기술은 형태와 관계없이 데이터를 관리할 수 있도록 진화하고 있다.

그런데 정보시스템 내에 존재하는 모든 데이터를 기록으로 관리할 필요는 없다. ISO 16175-3에서는 “업무시스템 내의 모든 정보를 기록으로 관리할 필요는 없으며, 업무과정 분석 등을 통해 업무의 증거가 되는 정보를 선별하여” 기록으로 관리해야 한다고 말한다(KS X ISO 16175-3, pp. 5-7). 따라서 데이터세트 기록은 데이터베이스의 모든 데이터가 아니라 선별과정을 통해 기록으로 선언된 것만을 지칭한다고 볼 수 있다. 이때 선별의 기준은 ‘업무활동의 증거력’이라는 관점에서 마련되어야 한다. 만약 특정 정보시스

템의 데이터베이스가 데이터세트 기록으로 선별되었다면 앞에서 언급한 데이터베이스의 특징들을 신중히 고려하여 기록관리 행위를 설계해야 할 것이다.

또한 데이터세트를 기록으로 관리할 때 핵심은 업무활동과의 연계정보를 확보하고 기록의 ‘고정성’을 확보하는 것이다. “기록은 단순한 데이터의 집적물이 아니며 특정 사건(events)의 결과물 내지 산물”이며 기록이 증거로서의 가치를 가지려면 “기록의 내용이 고정화된 형식(fixed form)으로 존재”해야 하기 때문이다(KS ISO 16175-3, p. 4).

2.2 행정정보 데이터세트 기록의 관리 실태

데이터세트 기록에 대한 선별 요건과 절차, 이관 및 폐기 등의 처분 정책이 확립되지 않은 상태에서 대통령기록물을 포함한 일부 폐지기관의 데이터세트가 물리적인 방식으로 이관된 바 있다. 선별과정 없이 이관되었기 때문에 데이터세트의 무엇이 이관되었고 어떻게 처분되었는지 구체적으로 알려진 바 없지만 이미 공개된 대통령기록물 이관 관련 자료와 국가기록원의 자료를 통해 데이터세트 기록의 관리 실태를 확인할 수 있다.

대통령기록관에 이관된 17대 대통령기록물 중 행정정보 데이터세트는 약 330만 건이다. 330만 건의 산정은 데이터베이스의 테이블 수량을 기준으로 하였다고 한다. 이관은 각각의 데이터베이스를 해당 정보시스템 별로 구분해 하드디스크에 저장 후 전달하는 복제방식이었다. 이렇게 이관된 하드디스크의 수량은 수십 개 정도였지만 테이블 단위로 산정되어 330만 건의 기록

물이 이관된 것으로 발표되었다(박성진, 2013).

기록관리의 관점에서 이관 과정을 되짚어보면 몇 가지 문제들을 발견할 수 있다. 첫째, 전자기록과 전자객체가 구분되지 못한 상태에서 모든 데이터셋이 기록으로서 평가받았다는 점이다. 둘째, 각각의 데이터셋을 재현하는 행정정보시스템들의 재현도구가 확인되지 않았고 이관되지도 못했다는 점이다. 셋째, 하드디스크에 데이터베이스를 복제하는 물리적 이관방식이 채택되었다는 점이다. 재현도구가 없는 이관은 전자기록의 진본성을 확보하지 못한 것과 동일하다. 이 상황에서는 전자기록을 식별할 수 없기 때문에 전자객체가 저장된 하드디스크가 기록으로서의 지위를 부여받아 관리될 것이다. 하지만 현재 저장된 전자객체만으로는 원래의 외형과 기능을 파악하기는 어려울 것이다.

전자기록의 요건인 재현성을 통해 선별되지 않은 상태로 하드디스크에 저장된 전자객체를 데이터셋 기록으로 인정하기는 어렵다. 모든 데이터셋이 기록이라는 포괄적 정의와 전자기록을 물리적 실체로서 접근하는 인식이 극복되지 않는다면 이러한 상황은 반복될 것이다. 16대 대통령기록물과 몇몇 폐지기관의 데이터셋 또한 같은 방식으로 이관되었다. 하지만 재현도구를 통해 전자기록을 선별하는 절차가 생략되었다면 이관된 데이터셋은 전자기록으로서가 아닌 전자객체의 상태에서 하드디스크와 같은 매체로 보존될 수밖에 없을 것이다.

최근 국가기록원은 행정정보 데이터셋 기록에 대한 실질적인 관리체계가 부재하여 기록관리 사각지대에 있음을 인정하고 새로운 방안을 모색하고 있다(국가기록원, 2016). 여기에는 공공기관이 현재 사용하고 있는 다양한 행

정정보시스템을 시스템·데이터베이스의 구성에 따라 유형을 구분하여 관리방안을 확보한다는 전략이 포함되어 있다. 특히 주목할 점은 기록관리 대상에서 일부 행정정보시스템을 제외해야 한다는 선별의 개념이 등장한다는 것이다. 기록관리 단위가 메뉴를 구성하는 데이터셋으로서 시스템 전체·업무기능·데이터(테이블, 레코드, 필드 등)가 각각의 구성에 따라 선별될 수 있다는 개념은 비록 전자기록의 요건 또는 속성과 같은 기준을 전제하지는 않았지만 선별과 관련된 논의의 수준을 한층 진전시킨다는 점에서 의미를 찾을 수 있다(국가기록원, 2016).

과거에 검색도구는 계층적 관리의 원칙에 따라 제공되었고 계층적 기술 없이는 기록에 접근하거나 활용하기 어려웠다. 그런데 태생적 전자기록이 양산되는 디지털 환경에서 상황은 급변하였다. 데이터베이스에 저장된 전자객체들은 계층적으로 관리되지 않아도 고유 식별자나 자연어 검색을 통해 곧바로 찾고, 자동분류가 이루어진 기록을 쉽게 탐색할 수 있게 되었다. 과거에 비해 계층분류의 필요성이 적어도 검색이나 접근 차원에서는 줄어든 것이다. 전통적 환경에서 종이기록의 평가 선별과 처분은 업무 및 기록의 계층분류를 기반으로 이루어졌다. 현재 이루어지고 있는 단위과제 기반의 기록 평가정책도 이러한 계층적 분류에 토대를 두고 있다. 그러나 데이터셋 기록의 경우 분류되어야 할 단위가 무엇인지, 업무기반의 기록 보존기간 평가체계가 데이터셋 기록에는 어떻게 적용될 수 있는지 등 데이터셋을 기록으로 관리하는데 필요한 기본 사항들을 확정해야 한다. 이 과정에서 기록관리를 위한 전문적 판단을 하려

면 각 데이터세트가 업무에 활용될 당시의 외형과 기능을 재현할 수 있는 환경 마련이 절대적으로 필요할 것이다.

3. 데이터세트 기록과 재현

3.1 전자기록 재현의 의미

기록은 내용, 구조, 맥락의 3가지 요소로 구성된다. 이는 기록으로 성립하기 위하여 필요한 조건이기도 하다. 전자기록물이 생산되기 이전의 전통적인 환경에서는 내용, 구조, 맥락의 3가지 요소들이 결합된 종이문서를 사람이 곧바로 읽을 수 있었다. 기록으로서의 정보적·증빙적·역사적 가치는 가독 가능한 3가지 요소를 갖춘 외형을 보고 판단하였다. 종이문서의 외형은 특별한 조치가 없어도 생산당시 그대로를 육안으로 확인할 수 있었기 때문에 가독을 위한 재현성을 강조할 필요가 없었다. 그런데 전자기록이 본격적으로 생산되면서 상황은 바뀌었다. 종이기록과 달리 컴퓨터를 통해 생산되는 전자기록은 모니터 화면에 표현되는 인터페이스를 통해서만 기록을 구성하는 세 가지 요소를 확인할 수 있게 되었다. 또한 기능이 포함된 일부 전자기록을 생산 당시 그대로의 모습으로 표현하려면 그 기능까지 재현하여야 한다.

전자기록의 재현성의 개념과 중요성은 이미 다양한 연구들에서 강조되어 왔다. ISO 15489는 기록의 이용가능성을 기록의 위치·검색·재현·해석으로 정의하면서 재현의 중요성을 강조한 바 있다(ISO 15489-1, 7.2.5항). 설문원(2005)은 기록의 품질요건인 이용가능성과 관

련하여 전자기록에서 재현은 이용가능성의 필수적 요건임을 강조하였다. Dollar(2002)는 전자기록의 장기보존을 위한 요건 중의 하나로 재현성을 제시하면서 그 개념을 모니터 화면으로 제공되어 생산 당시와 동일한 논리적·물리적 구조 및 내용을 그대로 표현하는 것으로 설명한다. ICA(2005) 보고서에서는 전자기록을 볼 수 있는 하드웨어·소프트웨어 기술을 통해 전자기록을 종이에 기록하거나 스크린에 재생하여 사람이 그 내용을 전달받고 이해할 수 있는 접근성과 이해가능성에 주목하면서 전자기록의 보존은 재현성이 담보되어야 함을 강조하였다.

이상의 연구들을 요약하여 정리해보면 전자기록의 재현성은 전자기록의 내용뿐 아니라 업무와 활용될 당시의 외형과 기능을 사람이 가독하고 해석할 수 있도록 모니터 화면을 통해 인위적으로 제공되어야 하는 속성으로 이해할 수 있다. 본 논문에서는 현용단계에서 업무담당자들이 행정정보시스템을 통해 전자기록의 내용, 구조, 맥락을 볼 수 있다는 점에서 이를 전자기록의 '외형'이라고 표현하였으며, 여기에 행정정보시스템에서 지원되는 '기능'을 재현되어야 할 또 다른 측면으로 보았다.

재현성은 기록의 4대 품질 요건인 이용가능성의 필수 구성요소이며 진본 전자기록의 장기보존을 위해 반드시 함께 보존되어야 하는 속성으로서 꾸준히 강조되어온 개념이다. 재현성이 확보되지 않은 전자객체는 이용가능성 요건을 충족하지 못했으므로 진본 전자기록으로 인정받는 것도 불가능할 것이다. 그러나 실제 우리나라 전자기록관리정책에서는 재현성에 대한 심도 있는 고려가 부족하다. 현재 공공기록

물관리법규가 종이기록 시대의 연장선에서 있으며, 물리적 매체 중심의 보존정책이 중심을 이루고 있다는 것이 그 이유일 것이다.

앞에서 밝혔듯이 행정정보시스템에 존재하는 모든 데이터셋이 기록으로 관리되어야 하는 것은 아니며 선별을 거쳐야 한다. 데이터셋 기록의 선별요건 중 핵심은 “보존해야 할 증거 및 증거 요건을 도출하고, 증거를 구성하는 내용이나 데이터를 확인”(KS X ISO 16175-3, p. 7) 하는 것이다. 이러한 점에서 전자기록의 재현성 확보는 기록을 선별하는 데에 필요한 첫 번째 절차이다. 재현되지 않은 데이터집합체를 대상으로 증거로서의 가치 선별은 거의 불가능할 것이다. 또한 전자기록의 지적통제활동과 이용 가능성을 보장하는 필수 요건으로 기록관리 전 과정에서 재현성은 끊임없이 강조되어야 할 것이다.

한편 외형과 기능이 재현된 전자객체를 대상으로 기록을 선별해야 하므로 아키비스트는 전자객체의 시스템적 속성을 이해하고 그 유형을 구분할 수 있는 역량을 갖추어야 할 것이다. 현재의 시스템 환경에서 생산되는 전자객체의 유형은 대체로 포맷이 있는 물리적 파일과 데이터 베이스와 같이 포맷이 없는 논리적 데이터로 구분할 수 있다. 이러한 구분은 전자기록에 대한 기록관리 행위를 설계하는데 유용하다. 포맷이 있는 물리적 파일들의 종류는 매우 방대하며 파일들이 갖는 내용정보는 고유한 뷰어를 통해 재현된다는 특징을 가진다. 따라서 물리적 파일은 그 포맷에 따라 기록관리 행위가 달라질 수 있으므로 이를 대비하는 것이 국가기록원의

DFR(Digital Format Registry)이나 미국 NARA ERA(Electronic Records Archives)의 파일 수집기능으로 볼 수 있다. ERA의 경우 현재까지 수집된 고유한 포맷 파일의 수는 8억 개 이상으로 알려져 있으며 현재도 계속 수집되고 있다(미국 NARA ERA 홈페이지). 다양한 애플리케이션에서 생산된 수많은 고유 포맷 파일 유형의 기록을 관리할 수 있는 방안은 아직 미해결 과제로 남아있다. 우리의 경우를 돌아보면, 과거 구전자문서시스템에서 생산된 여러 문서포맷들의 관리방안을 아직도 마련하지 못하고 있다. 만약 고유한 포맷 파일들이 그 파일을 생산한 애플리케이션을 통해 재현될 수 있다면 아직 해결하지 못한 이러한 전자기록에 대한 관리 방안이 마련될 수도 있을 것이다. 전자기록의 재현성 유지는 기록관리 절차를 지원할 뿐만 아니라 기록의 활용성을 증가시키고 오픈데이터로의 전환도 용이해질 것이다.

3.2 재현 방식

재현방식에 대한 논의를 위하여 관련 프로젝트 및 연구를 살펴보고 그 시사점을 분석하였다. 전자기록의 재현성을 중시한 프로젝트들은 다양하다. 일찍이 전자기록의 유형별 장기보존 방안을 실험한 네덜란드의 DPT(Digital Preservation Testbed) 프로젝트에서는 전자기록의 장기보존을 위해 내용, 구조, 맥락과 함께 외형과 기능⁶⁾이 재현되어야 한다고 제시하였고(DPT, 2003, pp. 15-17), 결론적으로 웹에서의 재현 등을 고려한 XML(eXtensible Markup Language)을

6) DPT에서는 기능을 ‘행위(behavior)’라는 용어로 표현함(DPT, 2003, p. 17).

제시한다. 이는 장기보존이 재현을 전제로 한다는 것을 강조한 것으로 전자기록의 외형과 기능이 재현되어야만 장기보존의 목적이 달성되는 것으로 해석된다.

데이터세트 기록의 재현성을 고려한 대표적인 해외사례로 미국 NARA의 AAD와 영국 TNA의 NDAD, 스위스의 SIARD가 있다. 미국과 영국의 사례는 DBMS가 관리하는 데이터베이스의 외형과 기능을 일부 흉내 내어 사람이 가독할 수 있는 형식인 웹 인터페이스로 제공하는 방식이고 스위스는 향후 웹으로 구현할 수 있도록 XML의 모든 구성요소와 데이터베이스를 테이블 단위로 묶어 압축파일로 제공하는 방식이다. 데이터세트 기록에 대한 각국의 기록관리 방식을 보여주는 세 사례들은 모두 데이터세트 기록의 장기보존방안으로 웹이라는 플랫폼을 선택하여 데이터세트를 재현한다는 접근법을 택하고 있음을 알 수 있다.

네덜란드의 DPT 프로젝트가 소개된 이래 전자기록의 장기보존 방안으로 웹의 표현도구인 XML이 자주 거론되는 이유는 특정 기술에 종속되지 않는다는 점과 기록의 계층구조를 가장 잘 표현하는 언어라는 장점 때문일 것이다. XML은 스키마를 정의하는 XSD와 외형을 정의하는 XSL이 하나의 세트로 구성된다. 즉 XML은 XSD에서 사전 정의된 형식으로 기록의 계층과 관계를 표현하고 XSL의 정의를 통해 외부의 파일객체를 웹 인터페이스에서 어떻게 위치시킬지를 정하는 것이다.

스위스의 SIARD는 이러한 XML의 특징을 이용하여 데이터베이스 내의 테이블 단위를 XML 파일들과 함께 압축한 ZIP 파일이다. 압축파일로 변환하여 제공한다는 점에서 일견 우

리의 장기보존포맷(NEO) 변환 방식과 비슷하게 생각할 수도 있으나 SIARD의 경우 압축을 해제하면 테이블, XML, XSD, XSL 파일이 나타나 웹브라우저로 바로 재현(Marciani & Kaufmann, 2015)되지만, NEO는 한글, PDF, 메타데이터·전자서명 파일 등으로 구성된 것으로 별도의 뷰어가 있어야 재현 가능하다는 차이가 있다. 이는 재현을 표방했음에도 불구하고 우리는 보존에 초점을 맞춰진 장기보존포맷을 채택한 반면, 미국·영국·스위스의 경우 재현된 화면을 서비스하거나 재현성 자체를 장기보존하고 있다는 점에서 차이가 있다.

이상의 사례들은 데이터세트 기록의 보존을 위해서는 외형과 기능이 함께 재현될 필요가 있다는 점을 말해준다. 데이터세트 기록은 응용 프로그램들과 데이터가 결합된 집합체이며 기능을 가지고 있는 전형적인 전자기록 유형이다. 따라서 데이터세트가 생산 당시의 'Look & Feel'을 유지하고 기록으로서의 요건을 갖추기 위해서는 내용, 구조, 맥락과 더불어 기능의 재현이 반드시 필요할 것이다. 기능의 재현이 결여된 채 데이터세트의 화면을 캡처 또는 출력하거나 포맷을 변경하여 보존하는 것은 데이터세트 기록의 정체성과 무결성이 훼손된 것으로 볼 수 있다. 그렇다면 데이터세트 기록 선별의 위한 첫 번째 절차는 재현도구의 제공 여부를 확인하고, 그 외형과 기능의 재현성을 점검하는 작업이 되어야 할 것이다. 외형과 기능이 재현되지 못한 데이터세트는 전자기록으로 대접받을 수 없으며 전자기록이 아닌 전자객체 상태에서 기록관리 처분 프로세스를 확정할 수는 없을 것이다.

한편 앞서도 언급했듯이 데이터세트 기록

이 진본이고 신뢰성 있는 증거로 간주되기 위해서는 그 내용이 특정 시점에 고정화되어 변경될 수 없도록 해야 한다. 일반적으로 데이터베이스는 수시로 읽고, 쓰고, 수정되고, 삭제가 가능한 기능적 특성을 가진다. 만약 아카이빙 대상인 데이터세트가 현재 업무에 활용되고 있는 행정정보시스템이라면 이러한 유동적인 기능이 기록의 증빙적 가치를 훼손하지 않도록 해야 한다. 이와 관련하여 ISO 16175-3에서는 고정성 유지전략을 제시하고 있다. 업무시스템에 연결된 개별 테이블의 데이터들이 삭제될 수 있으므로 삭제이력을 남기거나 아예 삭제기능을 없애도록 기록관리 차원에서 통제하라는 것이다(KS X ISO 16175-3, pp. 25-26). DBMS의 수정·삭제 기능은 통제되어야 하며, 특정 시점을 기준으로 기록의 변경을 동결하고 증빙적 가치를 위해 고정된 기록을 재현할 수 있는 기능이 함께 보장되어야 한다.

아울러 데이터세트 기록을 업무활동과 연계할 수 있는 전략이 필요하다. 데이터세트 기록을 업무활동과 연계하려면 우선 데이터베이스에 존재하는 구조와 맥락, 기능을 유지하는 것이 우선적이며 필요한 맥락정보를 추가적으로 기술할 수 있 수 있도록 지원해야 한다.

4. 데이터세트 기록의 관리 방안

4.1 데이터세트 기록의 선별

모든 행정정보 데이터세트가 기록이라고 선언되었음에도 불구하고 모든 데이터세트가 전자기록의 요건을 갖추고 있는 것은 아니다. 이

미 검토한 바와 같이 전자기록은 내용, 구조, 맥락 등의 외형과 기능이 모두 재현되었을 때 비로소 기록으로서의 진본성이 성립된다. 재현성은 전자기록의 선별을 위한 전제이자 기록관리 전 과정에서 늘 수반되어야 하는 필수 요건이다.

데이터세트 기록의 재현은 컴퓨터에 연결되어 모니터 화면으로 나타나는 인터페이스를 기준으로 해야 한다. 정보시스템은 화면을 통해 의미를 파악할 수 없는 비트스트림이나 기계를 훈련받은 사람만이 해독 가능한 코드 형식의 시스템 관리자용 인터페이스를 제공할 수도 있다. 하지만 불가피하고 예외적인 상황 외에는 이러한 인터페이스는 재현도구로서 인정하지 않는 것이 권장된다. 왜냐하면 기록의 재현성은 그 내용, 구조, 맥락을 갖는 외형을 사람이 가독하고 해석할 수 있을 때 비로소 성립하기 때문이다. 따라서 재현도구의 기준은 일반적인 환경에서 보통의 사람이 가독할 수 있는 사용자 인터페이스로 제공되어 생산 당시 의도된 'Look & Feel'을 그대로 재현할 수 있어야 하는 것이다. 만약 아키비스트가 데이터세트 기록을 선별할 때 이러한 재현성을 기준으로 전자기록과 전자객체의 영역을 구분하지 않으면 전자객체가 전자기록으로 선별되는 오류를 범할 가능성이 높아지게 될 것이다.

전자기록은 시스템엔지니어의 전자 영역과 아키비스트의 기록 영역이 합성된 조어라 할 수 있다. 하지만 병렬적 합성어이기 때문에 두 단어가 동일한 가치를 가지는 것은 아니다. 수많은 데이터가 전자객체로 존재하지만 그 중 기록의 가치를 인정받은 전자기록은 전자객체에 우선한다. 다만 전자기록은 전자적인 속성과 기록

의 속성이 조화롭게 통제되었을 때 유지되는 것이므로 양자가 분리될 수는 없다. 데이터세트 기록의 선별을 위해 아키비스트와 엔지니어가 협업해야 하는 것은 이와 같은 이유에서이다. 분야가 다른 양자 간의 원활한 협업과 의사소통을 위해 서로의 역할과 책임을 구분할 필요가 있다.

재현성은 전자기록과 전자객체, 아키비스트와 시스템엔지니어의 역할을 구분하는 비교적 명확한 기준이다. 전자기록의 선별은 아키비스트의 영역이고 선별된 전자기록을 구성하는 전자객체의 관리는 시스템엔지니어의 영역이다. 아키비스트가 재현된 화면 인터페이스를 통해 기록을 선별하면 엔지니어는 재현된 화면을 구성하는 응용 프로그램과 데이터베이스의 스키마를 분석하여 연결된 테이블을 식별해낸다. 아키비스트는 정보시스템의 기능을 통해 다양한 데이터세트를 조합하여 원하는 정보를 찾고 아카이빙 여부를 판단한다. 엔지니어는 외형과 기능을 재현하기 위해 필요한 H/W와 S/W 구성, 유지비용을 산정한다. 아키비스트는 부족한 예산 등을 감안하여 물리적 이관을 하지 않기로 결정하고 관리자 권한 등을 확보하여 원격에서 논리적 통제를 시작한다. 마이그레이션 방식은 기능 재현에 한계를 보여 상용 프로그램을 이용한 에뮬레이션이 장기보존 방안으로 채택된다. 엔지니어는 시스템을 구성하는 H/W와 S/W의 수명주기를 고려하여 정기적인 마이그레이션을 실시하고 중요 데이터를 백업하여 장애를 대비한다. 만약 재현성을 기준으로 아키비스트와 시스템엔지니어의 역할이 명확히 구분된다면 이와 같은 협업상황이 가능하게 될 것이다.

재현성을 기준으로 하는 선별절차는 몇 가지로 정리될 수 있다. 첫째 재현도구의 유무 확인, 둘째 재현된 외형·기능평가를 통한 기록의 평가·선별, 셋째 데이터세트 기록의 재현을 구성하는 요소 식별과 재현비용 산정, 넷째 물리적 이관 또는 논리적 이관 결정 등이다. 데이터세트 기록의 재현기능을 제공하는 정보시스템의 구성요소는 하드웨어는 물론 OS, 응용프로그램, 데이터베이스 등이 포함된다. 이들을 물리적으로 재구성하는 이관방식은 초기 구축에 버금가는 막대한 비용이 발생될 수 있어 한계가 있다. 따라서 대부분의 경우 데이터세트 기록은 물리적인 이관보다 논리적인 이관방식이 적합할 것이다.

재현된 전자기록을 대상으로 선별하는 절차가 생략되면 기록관리와 시스템관리의 구분이 불명확해질 가능성이 크다. 특정 행정정보시스템의 전자객체 전체를 전자기록으로 간주할 경우 아키비스트의 기록관리 활동은 물리적인 매체보존 수준으로 제한될 수 있다. 전자기록과 전자객체의 식별이 불가능한 상황에서 아키비스트의 역할은 축소되고 그 전문성은 의심받을 것이며 결국 아키비스트 무용론이 대두될 가능성이 크다. 전자객체와 전자매체를 가장 잘 다루는 전문가는 시스템 엔지니어이기 때문에 그들을 고용한 기관은 아키비스트 대신 이들에게 전자기록의 관리를 맡기는 것이 더 효율적이라고 판단할지도 모른다.

아키비스트는 요건을 갖춘 기록의 정보적 가치를 인지하여 지적인 통제활동을 수행하지만 전자객체 상태에서는 지적통제 활동의 대상이 불분명해지게 된다. 엔지니어는 전자객체의 상태에서도 매체와 시스템을 통제할 수 있지만

아키비스트의 경우 시스템에 대한 접근 자체도 제한을 받을 뿐 아니라 자신들의 전문성을 발휘하기 힘들 것이다. 선별이나 지적통제 없이 전자객체 상태로 전달받은 미래의 이용자들은 그 재현수단을 고민할 것이며 힘들게 재현에 성공하더라도 증거로서의 가치가 거의 없는 데이터라는 사실만을 확인하게 될 수도 있다. 아직 관리방안이 확보되지 않은 다수의 데이터세트에 대한 재현성 확보를 통해 지적으로 통제된 상태로 보존하고 이를 미래세대에 전달하는 것이 아키비스트의 역할이자 책무일 것이다.

4.2 데이터세트 기록의 관리 단위

전통적 기록관리 환경에서 관리단위는 기록물이라는 물리적 실체를 대상으로 결정되었다. 이관·평가·보존 등 기록물을 관리하기 위한 행위의 대상은 언제나 실체가 있었다. 전통적인 방식에 익숙한 아키비스트는 정보시스템의 데이터세트에서도 기록물과 같은 물리적 실체를 찾으려고 노력하였지만 수많은 데이터들 속에서 실체를 식별하기가 쉽지 않았고 현재까지도 그 실체에 대한 의견은 분분한 상태이다.

데이터세트 기록과 관련한 국내외 연구 사례들은 데이터세트 기록의 실체를 DBMS가 관리하는 데이터베이스, 데이터베이스 내의 테이블, 테이블 내의 행·열 등 데이터베이스의 구조 내에 저장된 데이터요소들로 보는 등 다양하지만 역시 확정적이지는 않다. 이규철은 데이터세트

기록의 관리계층인 행정정보시스템-데이터베이스의 계층 사이에 단위기능-시리즈라는 계층을 포함시켜 기록관리 계층은 단위기능으로 하자는 제안을 하였다. 그의 정의에 따르면 단위기능은 데이터세트를 대상으로 하는 최소 기능 단위 중 생성·갱신·삭제의 연산이 일어나는 기능⁷⁾이고 시리즈는 이관된 데이터세트의 묶음이다(이규철, 2016). 데이터세트와 관련한 국가기록원의 자료(국가기록원, 2016)에서는 데이터세트 기록은 시스템이 제공하는 단위기능이 기준이며 해당 기능과 연결된 데이터베이스 내의 단위는 가변적일 수 있다고 보았다. 여기서의 관리단위는 보통 메뉴로 표현되는 단위기능에 따라 시스템 전체일 수도 있고 일부 업무 기능에 연결된 테이블 단위 또는 테이블의 행과 열에 저장된 데이터요소일 수도 있다는 것이다.

언급된 사례들을 통해 볼 때 데이터세트 기록의 관리단위를 보는 관점이 데이터요소에서부터 정보시스템 전체까지 다양할 수 있고, DBMS가 제공하는 기능이 그 기준이 되고 있음을 알 수 있다. 하지만 기록 처분의 단위가 이처럼 다양해지면 실제 기록관리 행위를 설계하기가 어려워진다. 그러므로 일부 예외적인 상황이 발생하더라도 그 대상을 확정하여 기록관리를 위한 행위를 논리적으로 수행할지 물리적으로 적용할지 등의 방식을 설계하여야 한다.

이와 관련하여 스위스의 SIARD를 예시해 보겠다. SIARD는 ZIP으로 압축된 스위스 전자정부의 표준 장기보존방식으로서 압축과일에 포함된 데이터세트 기록의 관리단위는 테이블

7) Read는 제외함.

이다. 또한 DBMS 기능의 사용을 위한 언어인 SQL의 사용을 공식화⁸⁾하고 있다. 이 상황을 분석해보면 스위스 국가기록원은 데이터베이스와 이를 활용할 수 있는 기능을 데이터세트 기록으로 선별하고 그 외형과 기능을 웹을 통해 재현하는 것을 장기보존방안으로 확정하고 있다. 이를 위해 기록관리 대상을 테이블 단위로 정해 각각의 테이블을 XML 구문으로 포장한다. 그런데 테이블 단위로 압축하지만 스키마를 제공하는 XSD 파일이 포함되고 SQL이 사용 가능한 점을 볼 때, 각각의 테이블이 분절적이지 않고 상호 간의 관계가 보존될 수 있도록 전체 데이터베이스의 기능을 사용하는 것으로 보아야 한다.

정보시스템에 연결된 데이터베이스는 시스템이 제공하는 개별 기능에 따라 하나 또는 복수의 테이블을 수시로 참조하고 또한 참조될 수 있다. 데이터베이스의 테이블, 필드, 레코드에 있는 데이터 요소들은 조합되어 제공되고 유동적이므로 그 중 하나를 특정하여 기록관리 단위로 확정하는 것은 데이터베이스의 유기적 기능성을 이해하지 못하는 것이다. 또한 정보시스템의 화면에 기능이나 메뉴형식으로 표현되지는 않지만 이용자의 요청이나 필요에 따라 평소 참조되지 않던 테이블의 데이터가 이용되는 경우도 종종 있음을 감안할 필요가 있다. 이처럼 데이터베이스 내의 데이터들은 복잡하게 연계되어 있을 뿐 아니라 모든 테이블의 잠재적 활용 가능성이 있으므로 특정 테이블이나 그 이하 단위를 기록으로 보고 분리하거나 삭제하는 것은 위험부담이 크다. ITrust의 일환으로 진행된 데

이터세트 관련 연구에서는 데이터베이스가 스키마에 의해 연계된 복잡한 구조이므로 테이블을 인위적으로 분리하지 말 것을 권장하면서 데이터세트를 보존해야 한다면 필수적으로 데이터베이스 전체를 보존해야 할 것을 강조한다 (McDonald & Leveille, 2014, p. 111). 데이터베이스 환경에서는 기록의 내용과 메타데이터의 구분조차 쉽지 않다는 사실 역시 데이터베이스의 복잡성과 관리대상의 모호성을 말해준다 (KS X ISO 16175-3, p. 10).

이러한 데이터베이스의 속성과 선행연구를 볼 때, 특정 테이블이나 행·열의 데이터요소를 분리하여 처분행위를 진행하는 것은 데이터베이스의 무결성을 훼손할 뿐만 아니라 전체 데이터베이스가 가지고 있는 잠재적 활용성을 훼손할 위험이 크다. 혹시 산출물을 분석하여 테이블을 추출하면 된다는 의견이 있을 수 있으나 이 작업은 엔지니어가 수행하는 전자객체의 영역이지 기록의 가치를 평가하는 아키비스트의 영역으로 보기는 어렵다. 또한 데이터베이스 스키마를 분석하여 일부 테이블들을 선별하는 일련의 작업이 데이터베이스를 구축하는 것보다 많은 비용과 노력을 필요로 할지도 모를 일이다. 보존해야 한다면 전체를 보존하라는 InterPARES의 권고는 이러한 현실적 편익이 고려된 결과일 것이다.

한편 또 다른 관점에서 데이터베이스 전체를 기록관리 단위로 보는 것이 유익한 이유를 발견할 수 있다. 데이터세트 기록은 정책의 개방성, 투명성, 경제·사회적 이익을 목적으로 이용되는 오픈데이터로서의 활용 가능성이 높다. 빅

8) SIARD 공식 쿼리용 언어는 SQL 1999로 불리는 3번째 SQL 버전임.

데이터는 목적과 이용자에 있어서 오픈 데이터와 다른 경우가 많지만 빅 데이터의 분석방법을 오픈데이터에도 적용할 수 있다(McDonald & Leveille, 2014, p. 102). 빅데이터 분석방법이란 데이터의 규모가 크면 클수록 정보의 정확도가 비례하여 높아진다는 것으로 가급적 규모 있는 데이터의 지속적인 축적이 권장된다. 방대한 규모의 데이터가 의미 있고 정확한 결과 도출에 도움이 된다는 것은 이미 여러 사례에서 증명된 바 있다(정용찬, 2012).

마지막으로 필자는 기록관리의 원칙에 입각하여 데이터베이스 전체 선별의 유익함을 설명하고자 한다. 기록을 계층적으로 관리하고 계층별로 집합적 기술을 하는 것은 각 계층에 포함된 기록에 쉽게 접근하고 검색할 수 있도록 하기 위해서이다. 이는 다양한 출처를 가진 기록물에 통합적이고, 구조적인 질서를 부여하는 전통적인 기록관리 방식이기도 하다. 그런데 데이터베이스에 포함된 데이터들은 이미 구조화된 질서를 가지고 있다. 굳이 인위적인 편철과 하위 데이터에 대한 색인 작업을 하지 않아도 테이블 내의 데이터 요소들은 검색되고 통제 가능하며 테이블 간의 관계도 얼마든지 연결 가능하다. 이미 기록으로 관리될 수 있도록 질서정연하게 구조화되어 있는 데이터베이스를 선별 단계에서 분리하는 것이 기록관리의 원칙에 부합하는지 고려해 볼 일이다. 데이터베이스는 데이터마이닝과 같은 검증된 데이터 품질관리 방식이 있고 이는 DBMS의 기능만 재현된다면 얼마든지 가능한 일이다. 적어도 데이터세트 기록에 대한 방법론이 축적되고 진본성과 무결성의 우려가 해소되는 시점까지는 데이터베이스 전체를 기능과 함께 보존하고

재현하는 것이 바람직할 것이다. 다만 데이터베이스가 활용된 업무 맥락을 체계적으로 기술하고 이러한 맥락정보를 논리적으로 연계하는 지적 통제가 병행될 수 있도록 해야 할 것이다.

설명책임성이 중요한 공공기관에서 수집하고 축적한 데이터베이스는 기본적으로 품질과 정확도가 매우 높은 데이터의 집합체이다. 당장은 행정정보시스템의 일부 데이터들만 활용되어 전체 데이터베이스로서의 쓰임이 적을지라도 드러나지 않은 정보들의 잠재적 활용가치는 결코 작지 않을 것이다. 기록으로서의 미래가와 데이터베이스의 일부를 분리하는 방식의 위험성 등을 모두 감안한다면 데이터베이스 전체를 기록관리 단위로 처분행위를 실시하는 것이 모범실무를 위한 최선의 선택이 될 수 있을 것이다.

4.3 에물레이션 전략의 적용

4.3.1 에물레이션에 대한 전환적 고찰

데이터세트 기록의 재현과 관련하여 주목할 만한 기법은 에물레이션이다. 여러 나라에서 전자기록의 재현과 장기보존을 위하여 수많은 기법과 전략들이 연구되어 왔지만 최근 에물레이션이 새로운 대안으로 부상하고 있다. 그러나 장기보존의 전략으로서 에물레이션이 우리나라에서는 유독 과도하게 배제되고 있다. 에물레이션 관련 해외 연구사례인 미국의 CAMiLEON, 네덜란드의 KB, 영국의 Planets, 프랑스의 Keep 프로젝트를 분석했던 최근 연구는 에물레이션과 관련된 기술적 이해도를 한층 진전시켰지만 역시 막대한 예산과 기술적 한계가 걸림돌이라는

다소 부정적인 전망을 결론으로 서술하고 있다 (임진희, 2014, pp. 80-81).

하지만 전자기록은 정체성이 무결하게 보존되고 생산 당시의 외형과 기능을 그대로 재현할 수 있어야 비로소 진본성이 보장된다는 사실을 상기하면 에뮬레이션 전략을 보다 적극적으로 검토할 필요가 있다. 데이터세트 기록은 기능이 재현되어야 진본성이 보장되는 대표적 전자 기록 유형이다. 마이그레이션과 인캡슐레이션 기법이 적용된 현행 보존포맷변환 방식은 동적 기능을 가진 전자기록의 진본성 보장에는 한계가 있다.⁹⁾

선행연구에서도 언급했듯이 국제적으로는 에뮬레이션에 대한 연구와 실험이 본격화되고 있다. BDAX(Born Digital Archiving & eXchange) 팀은 2016년 미국 SAA 연례회의에서 독일의 bwFLA¹⁰⁾의 에뮬레이션 기법을 실험하겠다는 계획을 밝혔다(왕호성, 2016). bwFLA에서는 전자기록이 실행되던 최초의 고유 환경을 에뮬레이션 기법으로 재현하여 원래의 특징·외형·기능을 그대로 사용할 수 있도록 한다. bwFLA 팀이 사용한 에뮬레이터는 MS사가 제공하는 원격데스크톱 프로토콜(RDP)¹¹⁾이다. 에뮬레이션 기법의 구현 방식은 서비스로서의 에뮬레이션 전략(EaaS: Emulation as a Service)을 제안하는데 이는 클라우드 컴퓨팅과 같은 서비스 모델이다.

원격데스크톱 프로토콜(RDP)은 전기통신 분야 표준화 기구인 국제전기통신연합(ITU)¹²⁾의 T.120 제품군으로 에뮬레이션 구현과 관련된 통신 분야의 국제표준이다. 사실 기록관리 분야에서 장기보존 전략의 관점으로 보는 에뮬레이션은 정보통신 분야에서 터미널서비스, 썬-클라이언트, 가상화, 클라우드 등의 이름으로 변형되면서 정보통신 기술의 발전과 함께 폭넓게 활용되어 왔던 기법이다.

에뮬레이션 기법이 적용된 다양한 사례 중 몇 가지만 언급하겠다. 1997년부터 2000년대 초반, 학술정보 웹서비스가 도입되기 전까지 국내 도서관들은 학술정보를 CD-ROM으로 제공받아 에뮬레이션 기법을 적용한 서비스를 연구자들에게 제공하였다(이영준, 1997). 2000년대 중후반 IT 관련 기업들은 PC의 보안 문제와 비용 절감을 위해 적극적으로 썬-클라이언트 환경을 구축한다(김경식, 2010). 썬-클라이언트란 업무에 필요한 다양한 응용 프로그램들과 데이터베이스를 서버에서만 실행하도록 하고 이용자는 에뮬레이션을 통해 그 화면과 기능을 사용하는 방식이다. 적용된 에뮬레이션이 기록관리 관점에서는 다른 기법으로 인식될 수도 있겠지만 데이터세트 기록의 장기보존이 외형과 기능의 재현이라는 점을 이해한다면 동일한 ICT 기법이라고 볼 수 있다.

정리하자면 에뮬레이션은 터미널서비스, 썬-

9) 더욱이 각 기록관이 보존기간 10년 이상의 모든 한시기록을 장기보존포맷으로 변환하도록 한 규정은 변환 오류와 일선 기록관의 피로감을 누적시키고 있음. 실효성을 검증하지 않은 정책이 10여 년간 유지되는 데에는 학계의 방관도 일조했다고 보며, 합리적인 대안 마련과 정책 개선이 시급함.
 10) 정식 명칭은 "bwFLA(Functional Long-Term Archiving) - Emulation as a Service"임.
 11) RDP(Remote Desktop Protocol)는 마이크로소프트사가 개발한 프로토콜. 에뮬레이션을 가능하게 하는 프로토콜로서 IT분야의 터미널서비스 또는 가상화서비스가 모두 유사한 개념임.
 12) ISO는 전기통신 분야를 제외한 국제표준화기구임.

클라이언트, 가상화, 클라우드서비스 등의 이름으로 발전해온 ICT 분야의 핵심 기법으로 꽤 오랫동안 폭넓게 활용되어 왔다. 전기통신 분야의 국제표준은 ISO가 아닌 국제전기통신연합(ITU)에서 관장하며, 애플레이션 국제 표준 규약(프로토콜)은 ITU T.120이다. 이 규약에 따라 개발된 상용 애플레이터는 씨트릭스의 ICA와 마이크로소프트의 RDP¹³⁾ 등이 대표적 제품군이며, 각종 시스템 환경을 PC에서 구현하는 방식의 과도한 비용을 절감할 목적으로 대학도서관의 해외학술정보, 기업의 ERP 분야 등에서 일찍부터 활용되어 왔다는 것이다.

본 논문에서 애플레이션과 관련된 정보통신 분야의 현황까지 소개한 이유는 전자기록관리 관련된 문제들을 정보통신 분야와의 협력을 통해 극복할 수 있기를 바라는 때문이다. 전자기록의 장기보존전략은 전자기록의 재현성을 장기적으로 보장하기 위한 것이지만 '재현'이라는 용어를 명시적으로 사용하지 않고 있어서 '보존'에 무게중심이 실리는 듯이 보인다. 현재의 장기보존전략들이 물리적 포맷 변환에 집중하는 것도 이와 무관하지 않다고 본다. 다시 강조하지만 전통적 환경의 종이기록은 본질적으로 이미 재현성을 가지고 있어 보존만으로 미래의 재현성이 보장되었지만 전자기록은 별도의 재현도구가 함께 보존되지 않으면 보존행위 자체가 무의미해진다. 재현의 중요성을 인식될 때까지 전자기록의 이용가능성이 진본성 못지않게 강조되어야 할 것이다. 또한 아키비스트

들이 재현성에 기반한 전자기록의 장기보존 철학을 정보통신 분야의 엔지니어들과 공유할 때 기록관리 기술의 발전도 함께 이루어질 수 있을 것이다.

4.3.2 데이터세트 기록의 장기보존 전략으로서 애플레이션

전자기록의 장기보존 전략은 기록에 대한 접근과 활용이 최대한 장기적으로 보장되어야 함을 의미하므로 데이터세트 기록에 대한 접근과 활용을 보장하는 수단이 우선 확정될 필요가 있다. 데이터세트 기록의 처분단위인 데이터베이스는 기록 보존서고의 서가처럼 행·열이 질서 있게 구조화된 논리적인 서고의 모습을 하고 있다. 아키비스트가 이용자가 요구에 따라 서로 다른 서가에 위치한 기록물을 제공하듯이 정보시스템은 이용자의 명령을 받아 데이터베이스 내 각각의 테이블에서 데이터를 추출하여 제공한다. 만약 기록물이 매체에 수록되어 있다면 그 매체를 가독할 수 있는 도구를 함께 제공하듯이 정보시스템 역시 데이터베이스에서 추출·조합된 데이터들을 조합하여 가독할 수 있는 재현도구를 화면 인터페이스로 제공한다.

정보시스템은 화면 인터페이스를 제공하기 위하여 백오피스의 공간에서 다양한 응용 프로그램들¹⁴⁾이 함께 구동된다. 이는 마치 기록물의 열람을 위해 사무실 앞편의 열람 담당자와 뒤편의 서고 담당자들이 공동으로 작업하는 것과 비슷하다. 이와 같이 데이터세트 기록의 재현도구

13) 씨트릭스(Citrix)사는 이 분야의 기술력이 가장 높은 미국의 IT 기업으로 평가됨. RDP는 Citrix사에서 개발한 ICA 프로토콜을 라이선스하여 개발된 것임.

14) 메인 응용프로그램, DBMS, 검색엔진, 뷰어 등.

는 눈에 보이지 않는 백오피스에 있는 여러 기능들의 공동 작업을 통해 제공된다. 만약 데이터세트 기록의 모든 기능들이 재현되어야 하고 그 재현성이 장기보존 대상이라면 현재 전자문서 유형에 적용되고 있는 개체별 포맷 변환 방식은 적합하지 않다.

현재 우리나라가 채택하고 있는 장기보존 방안은 전자문서의 포맷을 변경하는 마이그레이션 방식과 인캡슐레이션 방식이다. 문서보존포맷으로의 변환은 마이그레이션의 적용이라고 볼 수 있으며, 문서·이미지 등 포맷이 있는 파일을 Adobe사가 공개한 PDF/A-1 포맷으로 변환하는 것을 의미한다. PDF/A가 마이그레이션 포맷으로 채택된 이유는 공개표준이므로 개발사에 종속되지 않고 내용 변경이 어렵도록 이미지화한 것이기 때문에 무결성 유지에 유리하다는 판단이 반영된 것이다. 인캡슐레이션 방식의 장기보존포맷¹⁵⁾은 문서보존포맷과 원본 파일을 기록관리 메타데이터와 함께 캡슐화함으로써 생성된다. 인캡슐레이션 기법은 DPT 프로젝트에서 제시된 전자기록의 장기보존 방안인 XML 캡슐화 방식과 호주의 장기보존 프로젝트 산출물인 VEO¹⁶⁾의 영향을 받아 채택되었다. 웹의 표현도구인 XML이 장기보존 방안으로 인식된 이유는 개방성이 특징인 웹을 통해 재현되도록 조치한다면 특정사에 종속된 형식의 포맷에 비해 재현 가능성이 높을 것이라는 사상이 담겨있다. 그러나 장기보존포맷의 경우 캡슐 내의 .pdf, .hwp 파일 등을 위한 뷰어도 함께 포함시켜야 완전한 재현이 가능하다는 약점이 있다. 이러한 점 때문에 인캡슐레이션만으

로 장기보존이 어렵고 “에물레이션 전략과 결합”할 필요가 있다(임진희, 2013, p. 323).

데이터세트 기록은 데이터베이스와 응용 프로그램의 기능이 함께 재현되어야 진본기록으로서의 정체성이 완성된 것으로 보아야 한다. 외형과 기능의 재현성 보장되지 않은 상태에서 기록에 행하는 무결성 조치는 이미 기록관리 행위로 보기 어려울 것이다. 기존의 포맷변환 방식이 기능이 중요한 일부 전자기록의 장기보존을 담보할 수 없다면 해당 유형의 전자기록을 장기보존 하는 방안은 다르게 모색되어야 한다. 부정적인 견해도 불구하고 에물레이션 전략을 대한 전환적 검토를 제안하는 이유는 이를 모색하기 위함이다.

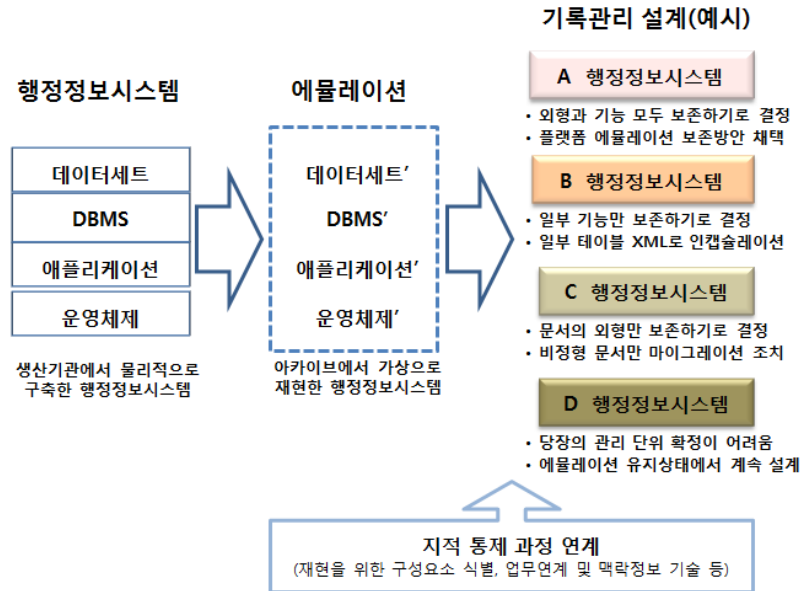
4.4 관리모형

앞의 내용을 토대로 데이터세트 기록을 관리보존하는 방안을 <그림 1>과 같이 정리하였다. 이러한 과정을 통해 전자객체가 비로소 전자기록으로 선언되고, 기록관리 절차에 편입될 수 있을 것이다.

이 모형에서 강조하는 점은 다음과 같다. 첫째, 행정정보시스템 전체를 대상으로 기록으로서의 가치를 우선 평가하고 선별절차를 실행하는 것이 바람직하다. 둘째, 기록관리 설계가 완성될 때까지 행정정보시스템의 외형과 기능을 유지해야 한다. 셋째, 논리적인 재현방안인 에물레이션 기법을 선별, 보존조치 결정, 지적통제업무의 기반으로 활용해야 한다.

15) 캡슐 내에 .pdf, .hwp 등의 파일의 뷰어도 함께 포함시켜야 완전한 재현이 가능하다는 약점이 있음.

16) VERS Encapsulation Object(VEO)의 약자로 XML로 캡슐화하는 개념임.



〈그림 1〉 데이터세트 기록의 관리 및 보존 모형(안)

5. 결론 및 제안

디지털혁명이 우리가 일하는 방식, 공부하는 방식 등 삶의 방식을 변화시켰듯 전자기록은 종이기록 시대의 기록관리 방식을 효율적으로 바꿀 수 있는 가능성이 잠재되어 있다. 하지만 이는 기록을 바라보는 기존의 물리적 패러다임에서는 쉽게 얻을 수 없는 사고의 대전환을 전제하는지도 모른다. 공공기록물관리법에 의한 현재의 기록관리는 ‘전자기록의 관리’라기보다는 ‘종이기록의 전자적 처리’라는 패러다임의 틀에서 크게 벗어나지 못하다고 볼 수 있다. 서고관리·항온항습·소독·복원 등 종이매체를 대상으로 행해지던 물리적 관리 행위들은 전자기록의 물리적 매체에 행해지는 시스템 유지·관리 행위로 대체되었지만, 아직 전자기록과 전자객체의 관리행위는 명쾌하게 구분되지 않고

있는 실정이다. 생산현황보고·선별·이관·분류·기술·평가·폐기·공개·서비스 등 기록의 지적 통제를 위해 행해지던 기존의 기록관리 행위들은 데이터세트와 같은 전자기록에 맞게 반드시 재설계 되어 아키비스트의 역할을 명확하게 해야 할 것이다.

거의 모든 공공기관이 행정정보시스템을 운영하고 있는 현실을 감안하면 향후 제도화될 데이터세트 기록관리가 미칠 파급효과는 매우 광범위할 것으로 예상된다. 만약 데이터세트 기록의 관리가 전자객체의 물리적 보존 등에만 치중한다면 기록관리 본연의 가치와 목적은 왜곡될 가능성이 높아질 것이다. 따라서 데이터세트 기록의 선별부터 활용의 전 과정에서 이를 경계해야 할 것이다. 이러한 관점으로 수행된 본 연구의 제안을 정리하면 다음과 같다.

첫째, 전자기록에 대한 생애주기적 기록관리

관념을 탈피하지는 것이다. 비현용 기록을 대상으로 아카이빙 조치를 취하는 단계적 기록관리 인식은 전자기록에 적용되기 어려우며 물리적 이관이 지적·법적인 이관을 보장하지도 않는다. 단계에 비례하여 오류들이 증가하여 전자기록의 무결성이 훼손되는 상황에서 3단계 물리적 이관방식을 고수하는 것은 근본적인 문제해결을 요원하게 할 것이다.

둘째, 재현성의 확인과 재현도구의 제공을 전자기록관리 전 과정을 위한 필수적 선행절차로서 규정해야 한다는 것이다. 화면 인터페이스로 제공된 재현 도구를 확인하여 전자기록과 전자객체가 구분되고 데이터세트의 내용이 가독·해석될 수 있도록 재현성은 지속적으로 유지되어야 할 것이다. 데이터세트 기록을 위한 지적 통제 활동은 모니터에 재현된 화면과 기능을 통해 전자기록의 진본성을 보장할 수 있는 방식으로 이루어져야 하는 것이다.

셋째, 데이터세트 기록의 관리단위는 전체 데이터베이스를 대상으로 하는 것이 가장 효과적이라는 것이다. 데이터베이스는 그 단위기능의 구현 유무에 관계없이 전체 데이터베이스가 온전히 존재한다면 새로운 정보를 조합할 수 있고 기록으로서 선별될 수 있는 2차적 가치가 잠재되어 있다. 또한 이미 구조화 되어 접근성이 보장된 데이터베이스는 접근, 검색을 위해 계층적 철·건 구조로 재구성하는 편철행위를 필요로 하지 않는다. 만약 재현성이 확인되지 않은 데이터세트의 테이블 단위에 기록의 가치를 부여하고 관리를 수행한다면 그것은 아직 기록으로 선별되지 않은 전자객체를 대상으로 시스템을 관리하는 것이지 전자기록을 관리하는 것으로 보기 어려울 것이다. 재현도구를 이용한

데이터베이스 전체의 아카이빙을 우선 수행한 후 개별 스키마를 분석하고 기록관리 단위를 정하는 것이 재현성 확인 이전에 기록관리 단위를 확정하는 것 보다는 기록의 잠재적 이익을 더 많이 제공할 것으로 판단된다.

넷째, 전자기록의 장기보존에 대한 관점을 물리적 보존에서 이용가능성과 기능의 재현으로 전환해야 한다는 점이다. 이용가능성의 첫 번째 조건은 기록으로의 접근이며, 전자기록의 진본성은 원래의 외형과 기능이 무결하게 재현되어야만 인정될 수 있다. 현재 지구상에서 가장 접근성이 강한 개방 공간은 웹이며 외형과 기능의 재현을 구현할 수 있는 가장 유력한 방안은 에뮬레이션이다. 데이터베이스를 비롯한 수많은 디지털객체들이 웹 플랫폼을 통해 재현되고 있고 웹으로의 변환을 돕는 도구들이 끊임없이 발전되고 개발되는 이유는 웹의 접근성과 개방성 때문일 것이다. 해외 국가기록원들의 데이터세트 기록에 대한 장기보존 방안이 공통적으로 웹에서의 재현을 전제로 하는 것도 같은 이유일 것이다. 웹의 표현도구인 XML은 데이터베이스 기록과 같은 구조화된 계층 관계까지 표현이 가능하며 SQL 기능을 허용하여 웹 플랫폼에서 데이터베이스를 활용할 수 있게 한다. 미국아키비스트협회의 메타데이터 인코딩 표준인 EAD가 3번째 버전까지 확장되면서도 여전히 XML을 기반으로 하고, ICA가 개발 중인 기술 표준인 Ric-CM이 기록의 맥락 확장을 위해 웹을 이용하는 것도 같은 맥락에서 이해할 수 있을 것이다.

이와 같은 연구결과를 토대로 필자는 현재 우리의 상황에서 실제 적용할만한 기록관리의 방향을 다음과 같이 제안하고자 한다. 아직 데이터세트 기록의 관리를 위한 구체적인 노하우가

확보되지 않은 상황이므로 모든 행정정보 데이터세트를 대상으로 하여 포괄적으로 접근하지 말고 영구보존 가치가 있는 가장 대표적인 행정정보시스템을 기관별로 하나씩 선별하는 현실적이고 실질적인 접근방식을 채택하자는 것이다. 가령 교육부의 교육행정시스템, 국가기록원의 중앙영구기록관리시스템, 인사혁신처의 인사관리시스템 등이 각 기관의 대표 행정정보시스템이라면 그 시스템들의 대상으로 우선 기록관리 방안을 설계하는 것이다. 각각의 시스템들은 모두 현재 활용되고 있는 해당 기관의 핵심적인 행정정보시스템이자 현용 기록물이다. 현용 기록물에 아카이빙 조치를 취하는 것은 기록의 생애주기 이론과는 맞지 않지만, 기록의 생산·활용 단계에서부터 기록관리를 목적으로 생산시스템에 기록관리 조치를 취하는 기록연속체 이론과 일치한다. 기관 당 많게는 100여 개가 넘는 전체 다종의 행정정보시스템들을 시스템적 구성에 따라 구분하여 관리방안을 찾는 것보다는, 대한민국 행정 각부 업무기능의 중요성에 따라 우선순위를 정하여 데이터세트 기록을 선별하고 지적통제 방안을 설계하는 것이 더

‘기록관리적’이면서도 현실적인 대안이 될 수 있을 것이다. 이때 중요한 점은 이미 구축되어 있는 시스템을 물리적으로 재구성하는 접근법이 아닌 논리적인 에물레이션을 활용한다는 전략으로 접근해야 생산기관과의 협상이 수월할 것이며 예산의 부담도 줄일 수 있을 것이다.

2008년 이래 ERA 프로젝트를 발족시키며 데이터세트를 포함한 전자기록관리를 적극적으로 준비해온 미국 NARA는 기금지원 프로그램을 통해 전자기록 연구를 촉진시키고 있다. 민간영역의 기록전문가협회인 SAA는 이에 호응하여 다양한 공동 프로젝트를 출범시키고 있으며 연구결과를 NARA의 성과로 돌려주는 선순환 활동을 지속하고 있다. NARA가 그 허부조직으로 데이터센터와 시스템운영센터를 구축하는 등 조직의 외연을 확장할 수 있었던 것도 이같이 전자기록관리를 위한 이론과 콘텐츠를 민간의 전문가협회가 지속적으로 제공했기 때문이라고도 볼 수 있다. 데이터세트 기록과 같은 전자기록관리 체도를 개선하기 위하여 중앙기록물관리기관은 민간에서 관련 연구가 활성화될 수 있도록 촉진하고 재정적으로 지원해야 할 것이다.

참 고 문 헌

- 곽정 (2016). 전자기록의 진본성 확보를 위한 공공기록관리 개선방안 연구. 박사학위논문. 한국외국어대학교 대학원, 정보기록관리학과.
- 국가기록원 (2016). 행정정보데이터세트 기록관리 방안. 대전: 국가기록원.
- 김명훈, 오명진, 이재홍, 임진희 (2013). 전자기록 장기보존 전략으로서의 에물레이션 사례 분석. 기록학 연구, 38, 265-309.
- 김경식, 김영곤, 김신령 (2010). 씬 클라이언트 환경에서 업무과정 정량화에 대한 연구. 한국정보과학회 2010 한국컴퓨터종합학술대회 논문집, 37(2), 114-118.

- 박성진 (2013). 제17대 대통령기록물 이관현황. 한국기록학회 3월 월례발표회(발표문).
- 박옥남, 박희진 (2016). 전자기록관리에 대한 국제 연구 동향 분석. 한국기록관리학회지, 16(1), 89-120.
- 설문원 (2005). 기록의 품질 기준 분석: 진본성, 신뢰성, 무결성, 가용성을 중심으로. 기록학연구, 11, 41-89.
- 왕호성 (2016). 미국 SAA 연례회의 참가 국외출장 결과보고서. 성남: 국가기록원.
- 이규철 (2016). 행정정보시스템 데이터세트의 이해와 기록관리 고려사항. 기록관리 표준 거버넌스 포럼 자료집, 72-78.
- 이승익, 설문원 (2017). 전자기록관리정책의 재설계에 관한 연구. 기록학연구, 52, 5-37.
- 이영준 (1997). 웹환경에서의 학술 CD-ROM 네트워킹 검색시스템 구축 기술. 한국도서관·정보학회 디지털도서관 시스템 구축기술 워크샵, 57-67.
- 임진희 (2013). 전자기록관리론. 서울: 선인.
- 임진희, 최주호, 이재영 (2014). 전자기록 애플리케이션 서비스 개발 사례 연구. 한국기록관리학회지, 14(3), 55-82.
- 정용찬 (2012). 빅데이터 혁명과 미디어 정책 이슈. KISDI Premium Report(정보통신정책연구원), 12-02, 9~12.
- 한국기록학회 편 (2008). 기록학 용어 사전. 서울: 역사비평사.
- 한국데이터진흥원 (2017). 2017 데이터산업 백서. 서울: 한국데이터진흥원.
- KS X ISO 16175-3: 2010. 문헌정보-전자사무환경에서 기록에 대한 원리 및 기능요건-제3부: 업무시스템의 기록관리지침 및 기능요건. 서울: 한국표준협회.
- Dollar, C. M. (2002). Authentic electronic records: Strategies for long-term access. Chicago: Cohasset Associates Inc.
- DPT (2006). 전자기록의 유형별 보존기법(=Digital Preservation Testbed. From digital volatility to digital permanence). (이미화, 현문수 공역). 서울: 한국국가기록연구원 (원전 발행년 2003).
- ICA (2005). Electronic records: A workbook for archivist (ICA Study 16). Paris, International Council on Archives.
- Marciani, C. & Kaufmann, R. (2015). Save your databases using SIARD. Swiss Federal Archive - Using SIARD. from <http://www.eark-project.com/resources/specificationdocs/32-specification-for-siard-format-v20>
- McDonald, J. & Leveille, V. (2014). Whither the retention schedule in the era of big data and open data?. Records Management Journal, 24(2), 99-121.

[웹사이트]

bwFLA, Retrieved July 19, 2017, from <http://bw-fla.uni-freiburg.de>

NARA, Retrieved July 6, 2017, from <http://aad.archives.gov/add>

Swiss Federal Archive, Retrieved July 25 from

<https://www.bar.admin.ch/bar/en/home/archiving/tools.html>

TNA, Retrieved July 6, 2015, from <http://www.ndad.nationalarchives.gov.uk>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

Jung, Young Chan (2012). The Revolution of big data & the policy issue of media. KISDI Premium Report, 12-02, 9~12.

Kim, Kyoung-Sik, Kim, Young-Gon, & Kim, Sin-Ryeong (2010). A study on work process quantification based thin-client environment. Academic Publications of Korean Information Science Society, 37(2), 114-118.

Kim, Myung-Hun, Oh, Myung-Jin, Lee, Jae-Hong, & Yim, Jin-Hee (2013). An analysis of cases of emulation for long term electronic records preservation strategy. The Korean Journal of Archival Studies, 38, 265-309.

Korea Data Agency (2017). 2017 Data industry white paper. Seoul: Korea Data Agency.

Korea Society of Archival Studies (2008). Dictionary of records and archival terminology. Seoul: Yuksa Bipyung Sa.

KS X ISO 16175-3: 2010. Information and documentation- Principles and functional requirements for records in electronic office environments-Part 3: Guidelines and functional requirements for records in business systems. Seoul: Korean Standard Association.

Kwak, Jung (2016). A study on improvement of public records management for the authenticity of electronic records. Unpublished doctoral thesis, Hankuk University of Foreign Studies, Seoul, Korea.

Lee, Kyu-Chul (2016). Understanding for administration information system dataset and considerations for recordkeeping. Records Management Standard Forum Resources, 72-78.

Lee, Seung-eok & Seol, Moon-won (2017). A study of redesigning electronic records management policies. The Korean Journal of Archival Studies, 52, 5-37.

Lee, Young-jun (1997). A study on the construction technology of CD-ROM networking searching system on the web. The Korean Library & Information Studies Workshop for Digital Library System Construction Technology, 57-67.

National Archives of Korea (2016). Study on the methodology for dataset records. Daejeon:

- National Archives of Korea.
- Park, Ok Nam & Park, Heejin (2016). A study on the international research trends in electronic records management: InterPARES3 and ITrust achievements. *Journal of Korean Society of Archives and Records Management*, 16(1), 89-120.
- Park, Sung Jin (2013). The 17th presidential records transferred to the Presidential Archives. *Monthly Seminar of the Korean Society of Archival Studies* (March 2013).
- Seol, Moon-won (2005). Quality Criteria for Measuring Authenticity, Reliability, Integrity and Usability of Records. *The Korean Journal of Archival Studies*, 11, 41-89.
- Wang, Ho-sung (2016). Result Report for SAA Annual Meeting attendance. Seong Nam: National Archives of Korea.
- Yim, Jin-Hee (2013). *Electronic records management*. Seoul: Seon-in.
- Yim, Jin-Hee, Choi, Joo-Ho, & Lee, Jae-Young (2014). A case study for the emulation service of electronic records. *Journal of Korean Society of Archives and Records Management*, 14(3), 55-82.

