

데이터세트 기록물의 기술요소에 관한 연구

A Study on the Description of Archival Datasets

김 포 옥(Po-Ok Kim)*

윤 수 영(Soo-Young Yun)**

초 록

데이터베이스시스템을 활용하여 데이터를 수집·처리하는 분야가 급속도로 확대됨에 따라 데이터세트에 대해서도 일반기록물과 같이 수집·평가·보존·활용해야 할 필요성이 증대되고 있다. 그럼에도 불구하고 국내 기록관리 분야에서의 데이터세트에 대한 관심은 매우 미흡한 수준이다. 이에 본고에서는 데이터세트를 기록물로 인식하고 체계적인 관리를 하기 위한 기본항목을 제시하고자 한다. 국제표준인 ISAD(G)를 준용하여 RAD, MAD와 데이터세트를 기록물로 인정하여 서비스를 제공하고 있는 NDAD의 기술요소를 세밀히 조사 분석하여, ISAD(G)의 기술영역을 기준으로 국내 데이터세트 기술에 필요한 각 기술영역과 영역별 내 주요 기술요소안을 제시하였다.

ABSTRACT

With the rapid spread of the practice of collecting and treating data by using a data base system, it's increasingly more critical to approach data sets in the same manner as general records in the collection, evaluation, preservation, and utilization process. Despite the importance, however, the interest level in data sets in Korea's records management is very low. In order to suggest basic items to regard data sets as records and manage them systematically, this study examined the descriptive elements of data set records. descriptive elements of data set records were suggested by comparing and analyzing those ones adopted by the agencies that regarded data sets as records and provided the concerned service as well as the descriptive rules of electronic records set by the advanced nations in records management based on the descriptive areas of ISAD(G).

키워드: 데이터세트, 데이터세트 기술, 데이터세트 기록물, 기록물 데이터세트, ISAD(G)

Datasets, Datasets Description, Datasets Records, Archival Datasets, ISAD(G)

* 전북대학교 문헌정보학과 교수(pook-kim@hanmail.net)

** 전북대학교 기록관리학 석사(ysy8038@hotmail.com)

논문접수일자 : 2007년 11월 20일 논문심사일자 : 2007년 11월 24일 게재확정일자 : 2007년 12월 4일

1. 서론

1.1 연구목적

인류는 사고의 시작과 함께 다양한 매체를 통해 기록을 남겨 왔으며, 문자와 종이의 발명으로 기록의 형태가 오랫동안 획일화 되어져 왔다. 그러나 현대에 이르러 특수형태의 시청각기록물을 비롯하여 급변하는 전산화 기술로 전자기록물이 다량 발생 되는 등 다양한 형태의 기록매체들이 생산됨에 따라 기록관리 업무상에서도 새로운 국면을 맞게 되었다. 특히 방대한 데이터의 생성 및 관리에 있어 데이터베이스시스템을 활용하는 분야가 급속도로 확대됨에 따라 각 분야에서 생산된 데이터세트는 다양한 목적을 위해 재활용됨으로써 부가가치를 창출할 수 있는 주요한 원데이터 집합으로 인식되어지고 있다. 이는 데이터세트 역시 여타의 기록물과 같이 평가 및 수집을 통해 가치척도에 따라 장기간 보존되어야 하고, 후대의 이용에도 활용될 수 있도록 기록으로서 보존 관리되어야 함을 의미하는 것이다.

외국에서는 2003년 스위스 연방 아카이브에서 개최된 데이터베이스 장기 보존을 위한 ERPANET 워크숍을 중심으로 선진국들의 데이터베이스의 보존 정책과 전략, 보존과정, 데이터베이스 평가, 도큐멘테이션, 보존포맷, 접근, 보존기술, 효과적인 보존을 위한 협력 등 다양한 분야의 논의가 이루어졌다(ERPANET 2003).

반면 국내에서는 전자정부(e-Korea) 구현 사

업 이후 정부의 각급기관은 전자적 형태로 공공 기록물을 생성·유통·보존·활용하고 있으며, 2006년 전면개정된 “공공기록물관리법”¹⁾에서 전자기록물의 생산 및 관리에 대해 명확히 규정함으로써 기록물의 전자적 생산·관리 및 비전자기록물의 전자적 관리에 대한 원칙을 명시할 정도로 데이터베이스시스템을 통해 상당량의 데이터가 축적되고 있다. 그럼에도 불구하고, 이를 기록으로서 취급하고 장기보존하기 위한 관리방법에 관해서는 논의조차 매우 미흡한 수준이다.

기록물을 효과적으로 관리하기 위해서는 먼저 정확한 기술(description)작업이 선행되어야만 한다. 이와 관련하여 국내에서는 “국가기록관리 혁신 로드맵”에서 국제수준에 부합하는 국가표준 제정을 기록관리혁신 종합실천계획의 4대 정책목표로 설정하고, 2006년 국가기록원에서는 ISAD(G)를 기반으로 하여 일반문서류와 전자기록물 기술을 위한 국가기록원 기록물 기술규칙(안)을 제시하였다. 그러나 ISAD(G)는 기록물 매체와는 관계없이 모든 기록물을 관리하는데 필요한 최소한의 요소만을 제시하고 있을 뿐이다. 때문에 전자기록물의 데이터세트와 같은 특수형태의 기록물을 관리함에 있어 별도의 심도 있는 연구가 필요하다고 사료된다. 이에 본 연구자는 먼저 선행 자료를 조사하였으나, 데이터세트의 기술적 요소에 관한 연구를 찾지 못하였다. 그리하여 외국의 실태를 상세히 조사 비교한 후, 이를 기초로 국내에서 데이터세트를 관리함에 있어 필요한 주요

1) 『공공기록물관리법』 제1장(총칙) 제6조(기록물의 전자적 생산·관리) 공공기관 및 기록물관리기관의 장은 기록물이 전자적으로 생산·관리되도록 필요한 조치를 강구하여야 하며, 전자적 형태로 생산되지 아니한 기록물에 대하여도 전자적으로 관리되도록 노력하여야 한다.

요소들을 면밀히 분석하여 구체적인 기술안을 제시하고자 한다.

1.2 연구방법

본 연구목적을 성취하기 위하여 필자는 첫째, 전자기록물의 개념을 비롯하여 데이터세트의 의미와 특성을 정리하고, 데이터세트의 유형을 구분하여, 유형별로 데이터 추출방법을 살펴보았다.

둘째, 기록물의 여러 기술규칙을 비교하기 위해, ISAD(G)의 기술영역을 기준으로 하여 문헌조사 및 웹사이트를 통해 수집한 각국의 기술규칙에서 제시하고 있는 기술요소들을 분석 검토하였다.

분석대상으로는 대체유형에 관계없이 기술요소를 제시하고 있는 ISAD(G)를 비롯하여, 전자기록물에 관한 기술요소를 독립적 내용으로 제시하고 있는 캐나다의 RAD와 영국의 MAD를 비교하였다. 그리고 ERPANET의 데이터베이스 보존 프로젝트에 참여했던 영국의 NDAD 프로젝트에서 제시하고 있는 기술요소의 내용도 비교분석 대상에 포함시켜 분석하였다.

셋째, 미국, 캐나다, 영국 등 각국의 기술요소를 비교 분석하여 기술상 중요성이 제기된 주요항목 부분을 ISAD(G)의 기술영역 구분에 기초하여 재정리함으로써 국내 데이터세트 기록물 기술에 필요한 각 부분의 주요 기술요소들을 분석하였다.

2. 기록물로서의 데이터세트

2.1 데이터베이스와 데이터세트

2.1.1 데이터베이스와 데이터세트의 개념과 구성

ICA와 IRMT의 전자기록관리에 관한 보고서에서는 전자기록물의 유형을 데이터세트, 텍스트문서, 다차원문서, 멀티미디어 문서로 구분짓고 있다(ICA & IRMT 1999).

그 중 데이터세트(Datasets)는 전자기록의 일부 그룹으로 편성, 취급되어져 데이터베이스 내에서 생산·관리·활용되어지는 것으로, 데이터베이스 전자기록의 사무관리 시스템 내에서 별개의 기록으로 역할을 담당하거나, 기록의 중요한 구성부분으로서 출현되는 형식으로 나타나는 기록물로 정의되고 있다.

SAA 용어집에서는 데이터세트란 데이터베이스를 보다 조직화하고 구조화된 형태로 저장시켜 관리시스템에 의해 접근 또는 업데이트가 가능한 정보로서 다양한 목적에 부합되도록 조작되고 추출될 수 있다고 정의하고 있다.

데이터세트의 개념과 연관된 데이터베이스의 구성은 내용부분과 데이터베이스 관리 시스템(DBMS), 애플리케이션으로 되어 있다(DPT 2003).

내용부분은 필드와 레코드 및 테이블로 구성된 컴퓨터 기반의 정보를 의미하는 것으로 통상적으로 데이터베이스 자체를 의미한다. DBMS는 물리적 데이터베이스와 이용자 사이의 소프트웨어 계층으로, 파일 위치에 대한 물리적인 상세한 사항과 파일 포맷에 관한 정보를 관리함으로써 데이터의 무결성과 보안성을 집중적으로 통제할 수 있다. 또한 데이터베이스에 관련한 이

용자 쿼리를 처리함으로써 이용자는 DBMS를 통해 데이터베이스를 정의하고 생산하고 유지할 수 있다. 애플리케이션은 데이터 입출력과 관련하여 시스템을 지원하는 기능을 비롯하여 그래픽 사용자 인터페이스와 데이터베이스 내용에 대한 처리 및 검색시 이용자가 사용할 수 있는 기능을 통합하고 있다.

이상에서 살펴본 바와 같이 데이터베이스에서의 기록은 다각적으로 정의될 수 있다. 데이터베이스 시스템은 데이터베이스의 내용, DBMS, 애플리케이션이 연합하여 전자기록을 구성한다. 또한 데이터베이스 내용은 데이터베이스 자체로서 전자기록이며, 데이터베이스 테이블에 저장되어 있는 하나의 행이나 테이블에 포함되어 있는 데이터 모두가 전자기록을 구성한다. 이러한 구성상의 복잡성으로 인해 데이터베이스에서 기록을 명확히 규정하기란 어려운 일이다.

데이터베이스에 의해 생산된 데이터세트는 컴퓨터 하드웨어와 소프트웨어의 개발로 기존 형태의 데이터를 새로운 환경에서 가독할 수 있도록 데이터베이스에서 데이터만을 추출하여 보관하는 원래의 데이터 집합이다.

SAA 용어집에서는 데이터세트를 컴퓨터에 의해 형성된 상호관련성 있는 정보의 묶음이라고 정의하고 있다. 또한 데이터 파일이나 데이터베이스와 유사하게 사용된다고 언급하고 있는데 이는 데이터베이스 시스템의 내용적 측면이 강조되었을 경우를 의미하는 것이다. 이러한 데이터세트는 동일한 업무활동의 결과를 축적하여 데이터세트 기록을 구성하게 된다.

데이터는 모든 정보형태의 기본단위로 이러한 데이터가 저장되고 관리되는 원 시스템이

데이터베이스이며, 이러한 데이터베이스의 데이터를 스냅샷 등의 추출 및 보존활동을 통해 수집한 결과물이 데이터세트이다.

2.1.2 데이터세트의 특성

데이터세트의 생산시스템인 데이터베이스는 컴퓨터로 접근할 수 있는 저장매체에 저장되어 데이터의 조작이나 검색요구에 대응하여 실시간 처리와 응답이 가능토록 이루어지게 되며, 동시 데이터의 삽입이나 삭제 또는 갱신 등 지속적인 데이터의 변화를 통하여 최신성과 정확성을 유지할 수 있게 된다. 또한 하나의 데이터베이스에 포함되어 있는 다양한 내용정보에 대하여 여러 이용자의 동시접근을 허용함으로써 데이터베이스를 다양한 목적으로 활용할 수 있게 되며, 데이터베이스 내의 데이터 레코드는 주소나 위치가 아닌 데이터가 가지고 있는 값에 의해 참조되어진다.

특수한 목적을 위해 생산된 데이터베이스는 각각 유일성으로서 원래의 고유 환경 속에서 널리 이용되지 못하는 특수성을 지니고 있다. 또한 고유성을 지닌 데이터베이스 시스템에서 일정기간이나 순간적 획득과정을 통해 형성된 데이터베이스의 내용은 잦은 갱신이 이루어지므로 데이터베이스의 보존을 용이한 포맷으로 변환시키는 데에는 고도의 기술력이 필요하다.

이러한 특성을 지닌 데이터베이스에서 생산된 데이터세트는 간혹 연관 자료나 파일링시스템을 확인할 수 없는 구조화되지 않은 환경에서 존재하기도 한다. 또한 물리적 특성을 구분짓기 어렵고 범위가 명확하지 않기 때문에 보안 측면에서도 취약하다. 구조적 측면에서는 정보의 특정유형을 정의하는 열과 개별데이터

를 의미하는 횡으로 구성된 테이블 묶음으로서, 데이터세트는 개별대상이 아니라 여러 횡 데이터에 적용할 수 있는 집합체적 의미를 띠고 있다. 또한 끊임없이 갱신되는 동적인 데이터베이스 시스템에 의해 생산된 데이터세트에 대해서는 기록물 기술에 대한 전통적인 접근을 재고해야 할 필요성이 있다.

2.1.3 데이터세트의 유형

데이터세트는 데이터 축적과 데이터베이스의 활동에 의해 개방형과 폐쇄형, 정적과 동적인 것으로 구분되며 이러한 특성에 따라 데이터베이스 시스템의 보존 전략도 달라져야 할 것이다(International Council on Archives 2000).

일회성 데이터세트는 보통 특정 프로젝트를 수행하면서 생성되는 것으로 데이터베이스와 프로젝트의 범위가 명확하며, 특정기간 동안에 데이터를 수집하고 이를 데이터베이스 시스템에 입력한다. 이러한 작업이 완료되면 데이터베이스를 폐쇄하여 더 이상의 데이터 축적은 이루어지지 않는다. 일회성 데이터세트는 독립된 단위로서 단일 데이터세트 시리즈로 관리되어야만 한다.

동적 데이터세트는 일정한 간격으로 개방되어 운영되는 데이터베이스에 의해 생성되는 것

으로 기존 데이터에 대한 업데이트가 이루어진다. 정기적으로 데이터 수정이 발생하거나 지속적으로 작업이 이루어짐으로써 데이터베이스의 범주가 명확하지 않으며, 전통적인 방법으로 ISAD(G)의 파일에 해당하는 계층의 자료를 구분하는 것도 불가능하다. 즉, 시리즈를 형성하기 위해서는 인위적으로 파일계층을 만들어야만 한다.

폐쇄형 데이터세트는 데이터베이스가 정적이거나 동적인 것에 관계없이 지정된 일정기간 동안 데이터를 수집하여 더 이상의 데이터 추가는 이루어지지 않는 폐쇄적인 데이터베이스 파일의 데이터의 집합체를 의미한다.

개방형 데이터세트는 동적 데이터베이스 시스템에 의해 생산되지만 새로운 데이터가 기존의 데이터를 대체하는 것이 아니라 기존의 데이터에 새로운 데이터가 지속적으로 추가되는 형태를 취한다. 이 역시 데이터베이스의 범주를 명확히 구분하기 어려우며 이러한 경우 전통적인 방법으로 독립적인 자료의 단위를 규정하는 것은 불가능하다.

전통적으로 시스템의 기록은 정기적인 스냅샷을 통해 보존되어 왔는데, 이는 거대한 시리즈에서 파일을 수집할 수 있는 유일한 표본추출 방법이다. 스냅샷은 단지 레코드의 일부만을 캡

(표 1) 데이터세트 유형 및 특징

일회성 프로젝트 DB	폐쇄적 DB	종료시점이 비정기적이며 동적인 DB	데이터 추가가 이루어지는 동적인 DB
<ul style="list-style-type: none"> • 일정기간동안 데이터 수집 후 종결 • 단일 데이터세트가 시리즈로 인식 	<ul style="list-style-type: none"> • 특정 데이터 수집기간과 관련한 파일 • 종료된 파일에 대한 데이터 추가 불가 	<ul style="list-style-type: none"> • 기존 데이터의 갱신, 수정 • 정기적·지속적 데이터 수정 • 스냅샷에 의한 인위적 파일 형성 • 로그파일 	<ul style="list-style-type: none"> • 신규 데이터 추가 • 데이터 수집기간이 정해져 있지 않음 • 데이터베이스 전체에 대한 정기적인 스냅샷을 통해 인위적 파일 형성

* 출처: Shepherd, Elizabeth. The Application of ISAD(G) to the Description of Archival Datasets(2003).

쳐하는 것으로 일부 동적인 데이터베이스 시스템은 정기적으로 레코드의 변화 추이를 알 수 있는 로그파일을 내포하고 있으나 전체 레코드를 포함한다는 것이 현실적으로 불가능하다.

개방형 데이터세트의 경우 스냅샷은 정기적으로 데이터베이스의 전체에 대한 사진을 찍는다고 할 수 있다. 이러한 과정으로 인해 가장 오래된 데이터에 대하여 매번 반복적인 기술이 이루어진다고 할지라도, 최신의 데이터의 경우는 한 번의 기술이 이루어지게 되므로, 기술에 있어 인위적인 파일계층을 형성하게 된다.

정적 데이터베이스 시스템의 보존 문제는 동적 시스템에 비해 비교적 간단하다. 폐쇄형이면서 정적인 시스템이라면 데이터 입력이 완료된 이후 데이터베이스 시스템을 보존하고, 동적 개방형이면서 정적인 시스템은 일정 시점마다 스냅샷을 통해 데이터를 보존할 수 있다. 이와 달리 동적이며 개방적인 데이터베이스 시스템은 보존상의 어려움이 있다. 계속 이용되고 있는 데이터베이스 중 특히 정기적·비정기적인 갱신과 삭제가 이루어지는 데이터베이스의 경우, 이를 그대로 보존소에 전송한다면 전송된 데이터세트와 원본데이터세트가 달라지는 경우가 발생할 수도 있다. 데이터 보존을 위해 가장 많이 이용되는 방법은 정적 데이터베이스 시스템 보존에도 이용되는 스냅샷인데, 동적 데이터베이스 시스템의 경우 시스템의 변화 주기에 따라 스냅샷의 실행 주기도 바뀌게 된다. 이와 관련하여 ERPNET 워크샵에서는 변화주기가 불규칙적인 경우 정기적인 스냅샷의 실행은 보존 데이터의 왜곡을 낳을 수 있으며, 스냅샷 주기가 짧아질수록 비용이 증가하고, 데이터 중복의 정도도 증가하게 된다고 지적하고 있다.

2.2 데이터세트 기록에 대한 기술

2.2.1 데이터세트에 대한 기술사항

데이터베이스에 담긴 정보의 진본성을 유지하면서 장기적 보존을 위해서는 기본적으로 무결성(integrity), 진본성(authenticity), 접근성(accessibility), 명료성(clearness), 원본성(original)을 지속적으로 관리하며 데이터를 보호해야 한다. 전자기록물의 구성요소와 관련하여 데이터세트는 우선 진본성을 갖추기 위하여, 배경정보와 내용 및 구조부분을 비롯 외형과 행위의 조건들도 만족시켜야 하며, 조직의 업무처리과정에 관한 부가적인 조건까지도 포함하여 기술해야 한다(DPT 2003, 15-18).

배경정보에서는 기관명, 업무, 업무처리 등 기관의 맥락정보를 기술해야 하며, 데이터베이스를 효율적으로 보존하기 위해서는 데이터베이스에 대한 기술적(technical) 배경정보도 함께 포함되어야만 한다. 또한 배경정보에서 중요한 정보의 기술은 다른 레코드와의 관련과, 보존 활동 및 결과에 관한 것인데, 이는 기록의 진본성 및 후대의 열람을 보장해 주는 방법이 기 때문이다.

내용부분은 데이터베이스의 가장 중요한 요소로서 다양한 방법을 통해 수집·축적된 데이터이며 수문자 텍스트, 유니코드 텍스트 등 다양한 종류의 내용을 포함하고 있다.

구조부분은 데이터베이스 요소의 논리적 계층성과 구성을 의미한다. 데이터베이스는 한 개 이상의 테이블이 상호 관련성을 맺고 있으며, 각각의 테이블은 수많은 행으로, 행은 수많은 열로 구성되어 있다. 이러한 구조가 없는 데이터베이스는 개별적인 데이터에 지나지 않는

다고 말할 수 있다.

외형이란 데이터베이스의 내용이 화면상에 표현되는 방식에 관한 것으로 데이터베이스 책임자가 부여한 의미를 보다 정확하게 표현하기도 한다. 레코드의 값은 숫자이지만 날짜나 화폐형식을 적용하여 다른 의미를 표시할 수도 있으며, 글자 색깔에 변화를 주어 의미를 강조할 수도 있다. 이처럼 특정한 외형은 내용과 구조에 의해 표현될 수 없는 부가적인 사항을 제공할 수 있음이다.

종이 기록물은 동적인 행위를 표현할 수 없는 반면 전자기록물은 애플리케이션을 통해 기록을 생산하고 조작하는 행위를 포함할 수 있다. 나아가서는 데이터베이스를 보존하기 위하여 데이터입력과 질적인 제공 및 데이터의 다양한 조작을 위해 설계되고 개발된 이용자 애플리케이션이 필요하게 된다.

이와 같은 데이터베이스의 진본성을 유지하기 위한 필수적 주요 요소들을 기록물 기술원칙에 준하여 기술함으로써 기록물에 대한 효과적인 관리가 가능해지며 정확한 정보를 제공해야만 이용자의 이해를 도울 수 있다.

2.2.2 기술원칙의 적용

다계층적 기술은 기록물 전반에 관한 지적통제를 통하여 기록물의 배열을 반영시키는 중요한 원칙이다. 데이터세트의 기술에서는 계층관리의 원칙에 따라 상위계층은 기록물 생산에 관여한 행정조직의 구조 및 시스템에 기반하고 하위계층은 데이터세트의 내부구조에 기반하여 형성될 수 있다. 하지만 중간계층인 시리즈 기술에 관해서는 데이터세트의 물리적 객체가 존재하지 않으므로 임의로 지정해야 한다.

상위계층에 해당하는 품이나 서브품의 계층은 앞서 언급한 바와 같이 일반기록물과 같이 생산기관의 행정적 조직구조 및 기능에 기반하여 기술될 수 있다. 품 계층은 생산기관으로 지정하고, 서브품 계층은 데이터세트를 생산한 부서나 조직으로 기술된다. 그러나 생산기관이 안정된 조직구조를 가지고 있거나 특정 목적을 위해 단기간 동안 형성된 데이터세트는 이러한 적용이 용이하지만, 생산기관의 조직구조가 빠르게 변화하는 경우 품의 개념을 정의하기가 매우 어렵고 다출처에 의해 형성된 데이터세트는 복수의 품이 존재하게 된다. 반면 상위계층의 기술은 기술대상의 상세사항이 아닌 총괄적 기술을 원칙으로 하며 맥락을 포함해야 하므로 ISAD(G)의 기술규칙을 적용하여 기술해야 한다.

자료특성에 관한 기술이 이루어지기 시작하는 단계인 시리즈 계층과 관련하여 ISAD(G)에서는 다음과 같이 정의하고 있다. “데이터세트는 동일한 축적과정이나 활동에 의한 결과물로서 파일링 시스템에 의해 정리되거나 하나의 단위로서 유지된다. 시리즈를 데이터세트에 적용할 경우에는 데이터세트의 규칙적인 증가와 시스템에 대한 정기적인 스냅샷 등과 관련되어 있다”(Shepherd 2003). 즉, 데이터세트의 경우 데이터 축적과 데이터베이스의 활동에 대한 고찰을 통해 시리즈 계층을 정의할 수 있다.

UCL은 ISAD(G)의 정의에 따라 데이터세트를 정적 데이터세트와 동적 데이터세트로 구분하였으며, 정적 데이터세트는 시리즈를 규정짓기 용이하나 동적 데이터세트는 반드시 인위적으로 시리즈 계층을 생산해야 한다고 언급하였다.

하위계층은 데이터세트의 테이블·필드·관

련성 등의 내부적 관계에 의해 정의되고 있다. 테이블은 기술의 최하위단계로 각 필드에 대한 설명을 포함해야만 하는데, 필드가 별도의 개체로 존재하지 않는 이상 기술에 있어 필드를 독립된 기술단위로 다루는 것은 부적절한 것으로 분석된다.

원질서는 기술적 단위의 생산배경인 활동사항의 증거적 측면에서 생산기관이 부여하게 되는 기록의 질서 및 파일링 구조를 준수함이 원칙이지만, 데이터세트는 수정이 용이하고 잦은 갱신이 부가됨으로써 이러한 원질서를 유지하기가 매우 어려운 실정이다.

그러나 기록물의 내부 구조 및 기록물간의 관련성을 정확하게 기술함으로써 진본성을 유지할 수는 있다. 또한 데이터세트는 여러 필드로 구성된 개별적 레코드가 모여 테이블을 이루고, 테이블간의 관련성으로 데이터세트가 존재하게 된다. 이처럼 기술단위간의 연계를 통해 집합적으로 기술되어야만 온전한 기록물로 이해될 수 있다.

3. 전자 및 데이터세트 기록물 기술규칙의 비교 분석

3.1 ISAD(G)2d의 기술요소 분석

국제적 차원의 기록물 기술표준을 제정하기 위해 ICA는 1990년 표준특별위원회를 설치하였고, 미국의 APPM, 캐나다의 RAD, 영국의 MAD 등의 주요 기술표준을 비교분석하여 공통적으로 제시된 요소들을 표준요소로 채택하여, ISAD(G)를 개발하였으며, 각국 표준의 상이한

요소들은 배제하였다. 다시 말하면 ISAD(G)는 모든 유형에 걸쳐 광범위하게 적용할 수 있는 일반 기술규칙의 성격으로서, 이에 그 내용을 먼저 분석하고자 한다.

ISAD(G)에서는 26개 요소를 7개 영역으로 구분하여 제시하고 있으며, 국제적 차원에서 기술정보의 교환을 위하여 6개 요소를 필수적 요소로 지정하였다.

먼저 식별영역은 기술단위를 식별하기 위한 본질적인 정보를 제공하는 영역으로 참조코드, 표제, 일자, 기술계층, 기술단위의 규모와 매체를 포함시키었는데, 이는 모두 이용자 검색에 있어 접근점으로 제공될 수 있다. 이런 측면에서 ISAD(G)는 필수적 기술요소로서 제일 먼저 식별영역을 지적하고 있다.

배경영역은 기술단위의 출처와 보관에 대한 정보를 기술하는 영역으로 기록단위의 생산적 배경을 구성하고 있다. 즉, 생산자명과 행정연혁 및 개인이력 의 요소를 전거파일과 연계하여 기술하도록 규정하고 있다. 내용과 구조영역은 기술단위의 주제와 배열에 관한 정보를 기술하는 영역으로 기술단위의 범위와 내용 및 평가, 폐기, 정리체계 등과 같이 기록물의 내용 및 구조에 관한 전반적인 사항을 다루고 있다. 열람과 이용조건의 영역은 기술단위의 이용가능성에 대한 정보를 제공하는 것으로 기술단위의 이용과 관련된 법적사항 및 시스템적 제한요소, 검색보조도구, 복사에 관한 제한사항들을 기술하도록 규정하고 있다. 연관자료 영역은 기술단위와 밀접한 관련성을 가지고 있는 자료에 관한 것으로 기록물에 대한 확장된 정보를 제공해 주는 영역이다. 주기영역은 중요하지만 다른 영역에서 기술될 수 없는 기술단위에 관

한 정보를 기술하기 위한 영역이다. 기술통제 영역이란 자료에 대한 기술의 책임자나 기술시기 및 기술방법에 관하여 기술하는 영역으로서, 기록물 기술에 관한 추가적인 정보를 제시토록 하고 있다.

3.2 RAD: Rules for Archival Description Revision

캐나다의 RAD는 2부로 구성되어 있으며, Part

I에서는 기록물의 유형별 기술규칙을, Part II에서는 표목과 참조에 관한 내용을 포함하고 있다. 먼저Part I은 1장 기술의 일반규칙, 2장 복합매체 용에 관한 설명, 3장에서 12장까지는 텍스트, 그래픽, 지도, 건축과 기술도안, 동영상, 소리, 전자형태, 마이크로폼, 실물, 우편등과 같은 유형별 기술규칙을 각기 하나의 장으로서 구성하고 있다. Part II는 접근점의 선정, 인명, 지명, 단체명, 참조로 나누어져 있다(Canadian Committee on Archival Description 2007).

(표 2) RAD 전자형태 레코드의 기술요소

영역	기술요소	하위기술요소
표제 및 책임영역	본표제	공식본표제 보충본표제
	일반자료명시	
	대등표제	
	기타표제정보	
판영역	책임표시	
	판표시	
판에 대한 책임표시	판에 대한 책임표시	
	출판일	
일자영역	생산일자	
	출판지, 배포지	
	출판자, 배포자	
	출판자, 배포자의 역할에 대한 표시	
	출판일, 배포일	
물리적 기술영역	제작지, 제작자, 제작일	
	(특정자료명시를 포함한)기술단위 수량	
	기타 물리적 상세사항	
	크기	
출판사의 총서영역	딸립자료	
	출판사의 총서표시	
기록물기술영역	행정연혁	
	개인이력	
	보관이력	
	범위와 내용	
	시스템기술	시스템명 및 개발자 하드웨어 운영시스템 네트워크 또는 멀티유저형태

영역	기술요소	하위기술요소
		시스템보안 및 접근 프로그래밍어 소프트웨어/애플리케이션 역량 시스템문헌 시스템 이정표 시스템 소프트웨어 위치
주기영역	주기사항	접근을 위한 시스템 조건 이형표제 본표제 정보원 대등표제, 기타표제정보 표제연장 책임사항 기여 및 추론 판 출판·배포 등을 포함한 생산일자(축적일자) 물리적 기술(물리적 상태, 보존, 탈립자료) 출판사의 총서 알파벳-숫자 표시 직접적 출처 정리 언어 및 스크립트 원본 및 복제본(원본에 관한 결합주기, 기타 포맷의 이용가능성) 열람 제한 검색도구 관련자료 추가 다른 품 내의 관련 레코드그룹 일반주기
표준번호영역	표준번호	

또한 기술요소의 제시에서는 계층별로 주요한 정보원들을 구분하고 있다. 품과 시리즈 계층에서는 각각의 계층에 포함되어 있는 모든 자료를 대상으로 하며, 파일과 아이템 계층에서는 내부요소(영화제목, 주요 메뉴, 내부 라벨, 기타 식별정보), 생산, 편집, 제작, 편집에 관한 주요 기관(개인)이 작성한 시스템 도큐멘테이션, 기타요소(물리적 매체나 용기에 표시된 정보) 등이 주요 정보원으로 활용되고 있다(Canadian Committee on Archival Description 2007).

RAD는 전자형태의 레코드에 관해서, 컴퓨터 처리에 의해 코딩되고, 데이터(문자, 이미지)와 프로그램, 시스템 도큐멘테이션으로 이

루어진 원 디지털 레코드나 디지털화된 기록물이라고 정의하고 있다. 동시 기술적 주요요소는 표제 및 책임영역, 판영역, 일자영역, 물리적 기술영역, 출판사 총서영역, 기록물 기술영역, 주기영역, 표준번호영역 등 8개의 영역으로 나누어 제시하고 있다.

RAD의 기술영역은 ISBD를 준용하고 있으며 기록과 관련한 기술요소는 주기사항에서 제시하고 있다. 특히 기술요소에서는 전자자료의 특성을 고려하여 시스템 기술요소와 시스템명 및 개발자, 하드웨어, 운영시스템, 네트워크 또는 멀티유저형태, 시스템보안 및 접근, 프로그래밍어, 소프트웨어/애플리케이션 역량, 시스

(표 3) MAD 전자레코드 기술요소

영역		기술요소	하위기술요소
기록물 기술부문	식별표시영역	참조코드	
		표제	형식어/유형어 이름요소 단순일자
		규모	
	생산영역	목적 맥락 책임표시 복합일자 회수 또는 폐쇄 보유·평가·샘플기준	
	열람과 이용 영역	일반열람조건 법적열람조건 네트워크 열람조건 요금	
	내용과 구조영역	범위와 내용/초록	주제 및 내용 유형
		구조	링크 관련자료 역동/폐쇄
물리적·기술적(technical) 기술영역	물리적 용기 기타 특성 고유 애플리케이션/소프트웨어 의존성 고유 하드웨어 의존성		
관리정보영역	처리와 검증이력 아카이빙 업데이팅		

템문헌, 시스템 이정표, 시스템 소프트웨어 위치 등의 하위요소를 기술해 줌으로써 전자형태 레코드의 관리시스템에 관한 상세정보를 포함하고 있다.

3.3 MAD: Manual of Archival Description 3rd edition

영국의 MAD는 기록관리자협회의 재정지원을 받아 Michael Cook이 주도한 리버풀대학 프로젝트에 의해 1986년 제정되었다. MAD의 초판(1986)은 기록물기술과 서지기술에 있어

분리적 입장을 띄고 있으며, 처음에는 특수매체를 포함하지 않았다. 이후 개정이 이루어지면서 기록물의 기술과 서지기술과의 절충안이 제시되었고, 특수매체의 기술안도 포함시키면서 ISAD(G)의 원칙을 수용 하는 작업이 이루어졌다.

MAD 역시 기록물 기술에 있어 다계층의 기술원칙을 존중하고 있으며, 이에 따라 기술계층을 0~5계층으로 나누어 제시하고 있다. MAD는 기록물 기술의 본질성과 데이터 구조 및 기술모형, 기록물 기술의 유형, 특수매체 등의 다섯 부분으로 구성되어 있다.

MAD에서는 전자기록물을 전자시스템에 의해 생성되고, 가독 될 수 있는 파일이라고 정의하고, 제3·4·5계층에서만 기술이 이루어진다고 언급하였다. 3계층에서는 모든 관련 파일 집합을 기술하고 4계층에서는 가장 기초적인 배열 및 기술단위를 형성하고, 식별 가능한 파일 단위를 기술한다. 5계층에서는 레코드내의 개인파일이나 가상 폴더 내의 가상문헌 등의 아이템 부분을 기술하도록 규정하고 있다.

MAD는 앞에서 설명한 기술표준과는 다르게 기술요소를 일반기록물의 경우 일차적으로 중요한 정보를 포함하는 기록물 기술부분과 보존소 내에서 행해지는 관리 과정상의 필요정보를 포함하는 관리정보부분의 2개로 나누어 기술요소를 제시하고 있다.

반면, 전자레코드는 기록물 유형의 특성에 맞는 기록물 기술부분의 기술영역만을 제시하고 있으며, 관리정보부분의 요소를 기술부분의 관리정보영역으로 통합하여 제시하고 있다.

전자레코드는 그 정의에서 설명되듯이, 하드웨어와 소프트웨어에 대한 기술과 수정이 용이한 특성으로 인하여 기록물의 검증 및 업데이트에 관한 기술사항도 필수적이다. 이를 위하여 MAD에서는 물리적·기술적 기술요소를 별도의 영역으로 설정 제시하고 있는 것으로 분석된다.

3.4 NDAD: National Digital Archive of Dataset

NDAD는 1998년부터 영국 The National Archives(TNA)와 University of London Computer Centre(ULCC)가 협력하여 영국정

부의 데이터세트를 수집·구축하여 온라인 검색서비스를 제공하는 데이터세트 아카이브를 말한다. NDAD는 디지털 형태의 영국 중앙 정부 부처를 비롯한 다양한 기관이 생산하는 데이터 세트와 문서 중 TNA가 장기적으로 보존할 가치가 있는 기록에 대하여 보존하고 공공의 접근에 대한 서비스를 제공한다. 현재 영국 중앙정부가 컴퓨터를 이용하기 시작한 1960년대부터의 데이터세트를 포함하여 40년간의 데이터를 포괄하고 있다(The National Digital Archive of Dataset 2007).

데이터베이스는 여러 정부부처의 기록을 보존하는 수단으로서, 데이터베이스관리 시스템은 법률 및 각종 정책을 제정함에 있어 다량의 통계정보를 수집, 보관, 분석하는데 사용되고 있다.

이처럼 데이터세트는 일찍부터 영국 정부부처에서 생산되는 디지털 기록의 하나로 인식되어 왔으며, 여타 공공기록처럼 후대의 접근을 제공하기 위하여 Operational Selection Procedures를 통해 수집정책을 이행하고 있다. 일부 비밀성을 띠고 있는 기록에 대해선 열람을 제한하기도 하지만, 그 외 공공의 적극적 활용을 위하여 웹을 통해 열람서비스를 제공하기도 한다(Darlington 2004).

NDAD에서 축적하고 보존한 데이터베이스는 다양한 분야에서 영국정부의 의사결정에 따라 중요한 정보를 제공하게 되고, 각 정부부처에서 생산된 데이터세트 자료는 정책 형성에 막대한 영향을 미칠 뿐만 아니라, 정부의 정책 결정 과정을 투명하게 보여주는 중요한 증거로서의 기능을 수행하기도 한다. 통계데이터 역시 재활용될 가치가 매우 높은 내용들로, NDAD의

(표 4) NDAD 계층별 기술요소

기술영역	계층별 기술요소		
	시리즈 계층	데이터세트 계층	테이블 계층
배경	상위계층	상위계층	상위계층
요약정보			제목 NDAD 참조 범위와 내용 이용조건
식별정보	제목 NDAD 참조 생산일자 내용일자 데이터세트 규모 도큐멘테이션 생산일자 도큐멘테이션 규모 최종 입력일 최종 이용일 ISAD(G) 기술계층	제목 NDAD 참조 데이터세트 생산일자 데이터세트 내용일자 최종 입력일 최종 접근일 데이터세트 규모 ISAD(G) 기술계층	테이블명 레코드 수 파일크기 최대 레코드 크기 ISAD(G) 기술계층
행정배경	목적 및 용도 생산 책임 소재 관리 이력	목적 및 용도 책임 소재	
입수정보원		입수정보원	
특징과 내용	범위와 내용 처분일정 정보 추가 입수 계획 이전 참조	범위와 내용 디지털 처리와 변환	
접근·이용조건	법적 지위 접근조건 저작권 요건 데이터 보호법 요건 언어	접근조건	
관련자료	관련 기술단위 연합자료 생산부서 제작 출판물 연구자 제작 출판물	관련 기술단위 연합자료 생산부서 제작 출판물 연구자 제작 출판물	
원 시스템 특징	하드웨어 운영시스템 응용프로그램 이용자 인터페이스		
구조	논리구조와 스키마 동적 혹은 패쇄 여부 획득 및 검증 방식 데이터 신뢰성에 대한 제한	논리구조와 스키마 획득 및 검증 방식 데이터 신뢰성에 대한 제한	
검증	이관 이후 수행된 검증	내용 검증 변환 검증	
연결 데이터세트	데이터세트 목록 연결	관련 데이터세트	
주기	기술에 이용된 정보원	기술에 이용된 정보원	
기술			테이블 정보
필드정보			필드정보

데이터세트를 통해 각 부처로부터 수집된 다양한 분야의 통계정보를 분석할 수 있게 된다. 이처럼 NDAD는 다양한 분야에 걸쳐 중요한 정보를 제공하고, 오랫동안 데이터세트 기록에 대한 온라인 서비스에 있어 독자적인 입지를 구축해 왔다.

NDAD는 데이터세트의 구조적 특성을 고려하여 표 4와 같은 기술요소들을 제시하고 있다.

NADA는 시리즈 계층, 데이터세트계층, 테이블 계층에서 각 주요한 기술요소를 제시하고 있으며, 상위계층의 기술에서는 행정 구조를 반영하고 있어 실질적인 데이터세트 기술은 시리즈 계층에서부터 이루어진다.

시리즈계층은 11개 영역, 37개 기술요소를, 데이터세트계층은 11개 영역, 25개의 기술요소를 제시하고 있다. 실제적인 데이터세트계층에서는 시리즈계층에서는 제시되지 않은 입수정보원을 기술해야 하며, 내용 및 변환에 관한 검증요소를 상세히 기술하도록 제시하고 있다. 반면에 원시스템의 특징은 시리즈계층에서만 기술하게 되어 있다.

그러면 상기 분석된 내용을 중심으로 국내 기술규칙 중 상세히 명시되어 있지 않은 전자 기록물의 데이터세트 기술요소에 관한 주요 내용을 정리해 보고자 한다.

4. 국내 데이터세트 기록물의 기술요소(안) 분석

본 장에서는 ISAD(G)의 기술영역을 기준으로 국내 데이터세트 기술에 필요한 항목들을 각 영역별로 나누어 설정하였다. ISAD(G)는 기록

물 기술에 관한 국제표준으로서, 국제적 차원의 기록정보 교류를 위해서는 국내의 기록물 기술 규칙 역시 ISAD(G)의 기술영역을 준수하여야 할 것이다.

특히, 데이터세트의 원 데이터는 포맷에 관계없이 읽혀질 수 있는 전자적 관리시스템에 의해서 생성되거나, 또 이런 시스템에서 읽을 수 있는 파일로서 정의되어 진다. 그러므로 추가 또는 삭제과정도 용이하기 때문에 파일의 업데이트에 관한 기술력도 지속해서 이루어져야 한다.

이에 본 장에서는 ISAD(G)의 기술영역을 준용하여 식별영역, 배경영역, 내용영역, 구조영역, 열람 및 이용조건영역, 연관자료영역, 원시스템영역, 관리정보영역, 기술통제영역, 주기영역과 각 영역별로 나누어, 데이터세트 기록물에 적합한 기술요소들을 분석 제시하였다.

4.1 식별영역

데이터세트의 기록물은 생산자가 작성한 표제를 지니고 있는 사례가 드물며, 대부분의 명칭이 간략하게 코드화 되어 있어 기술단위의 내용을 표현하기가 극히 어렵다. 또한 스냅샷에 의해 수집된 데이터세트는 임의의 시리즈계층을 설정해야 하므로 하위의 기술단위를 총괄하는 적절한 표제와 기술단위의 내용적 특성을 고려한 표제를 기술해야 한다. 아이템 계층에 해당하는 각 테이블명은 데이터세트 내에서 유일한 값으로 테이블명만으로도 테이블 식별이 가능하며, 대부분의 테이블이 상호참조가 되고 있으므로 기술담당자가 인위적으로 표제를 작성할 필요성은 비교적 낮다. 이와 같은 복잡성

(표 5) 식별영역 요소

기술영역	기술요소	하위요소
식별영역	참조코드	국가코드, 보존소코드, 생산자코드, 서브폰코드, 시리즈코드, 파일코드, 아이템코드
	표제	본표제, 이형표제
	일자	생산일자, 축적일자, 내용일자
	기술계층	기술계층
	기술단위의 규모	규모, 매체유형, 매체의 규모, 기타 물리적 상세사항: 밀도, 트랙 수, 디스크의 면수 등

으로 데이터세트는 본 표제 이외의 이형표제 요소를 가능한 기술해야 한다. 특히 스냅샷에 의해 수집된 데이터세트의 표제는 시간적 범주를 포함해야 하며, 중요도에 따라 주요 생산자나 내용상의 주요인물(기관)도 표기해야 할 것이다.

일자요소는 기술단위를 구별하는데 중요한 식별요소인 동시에 기술단위의 증거적 특성을 위해서도 중요한 항목이다. ISAD(G)는 단일요소로 생산일자와 축적일자를 구분하여 기술하도록 되어 있으나, 데이터세트에서의 일자요소는 보다 다양한 측면에서 살펴볼 수 있다.

일반 종이기록물과는 달리 데이터세트의 일자는 생산측면과 내용측면에서 구분할 수 있다. 생산측면의 일자구분은 다시 데이터를 기록할 파일이나 테이블이 생산된 구조적 생산일자와 데이터세트의 내용의 입력, 수집, 연구기간 등과 같은 데이터의 축적일자로 구분할 수 있다. 데이터세트가 개방형이면서 동적인 경우 데이터베이스 설계는 지속적으로 재정비되어야 하고, 각각의 새로운 구조에 대한 일자를 기록해야 하므로, 하위계층의 일자요소를 기술하는데 어려움이 발생한다. 또한 데이터세트의 특성에 따라 내용일자도 각기 다른 의미를 지니게 된다.

이외에도 정보원이 될 만한 일자요소들은 이용자들의 혼란을 감소시키고, 기술상의 효율성

을 높이기 위하여, 제각기 관련성이 깊은 영역에서 기술되도록 할 뿐 아니라, 해당영역의 각 기술요소에 대한 이용자들의 이해력을 높일 수 있도록 참조가 이루어져야 할 것이다.

기술단위의 규모를 기술함에 있어 데이터세트는 전자기록물로서 내용과 보존매체가 각기 분리됨으로, 보존매체에 대한 적절한 기술이 제공되어야 할 것이다. 각 계층별로 데이터세트의 개수, 해당 데이터세트의 크기 및 테이블 수, 테이블별 레코드 수를 기술하여 크기를 표시하며, 특정 매체에 의해 보존되는 경우 매체유형 및 규모도 함께 기술되어야 할 것이다.

4.2 배경영역

배경영역에서 생산자명은 이용자가 검색시 가장 쉽게 이용할 수 있는 접근점으로 기술단위의 지적인 내용에 책임을 지는 개인이나 기관을 표시하는 항목이다. 데이터세트에서의 생산자는 일자요소와 마찬가지로 구조적·내용적 측면으로 구분하여 정리할 수 있다. 구조적 측면의 생산자 요소는 데이터베이스 구조의 생산·설계·관리에 관한 책임자를 기술해야 하고, 동적 데이터베이스 시스템인 경우에는 지속적으로 구조적 변화와 관련된 다양한 유형의 책임자가 존재하게 된다. 또한 내용적 측면의

(표 6) 배경영역 요소

기술영역	기술요소	하위요소
배경영역	생산자명	생산자명, 역할표시, 이형생산자명, 이형생산자의 역할표시
	행정연혁/개인이력	행정연혁/개인이력
	목적 및 용도	목적 및 용도
	기록물이력	소유기관/보존소, 소유기간/보존기간, 소유권 변경 사유
	데이터 획득 및 검증	획득방법, 검증방법
	수집/이전의 직접적 출처	수집처, 수집일자, 수집방법, 이관번호

생산자는 내용의 생산이나 편집, 분석에 대한 책임자들을 각기 상세히 기술해 주어야 한다.

이외에도 기증자나 후원자 등이 존재하지만 이러한 요소들은 일자요소의 경우처럼 각각 관련 기술영역에서 기술되어야 할 것이다.

다양한 생산자와 관리책임자가 존재하는 데이터세트의 행정연혁이나 개인이력을 기술함에 있어 기타 주요한 생산자에 대한 기원, 발전 과정, 주요 과업 등을 포함함으로써 기술단위에 대한 이용자의 이해도를 높여야 할 것이다.

목적 및 용도의 기술요소는 데이터세트의 내용 및 시스템 설계에 주요한 영향을 미치는 요소이지만, 종이기록물과 달리 데이터세트는 수집이나 축적된 원데이터를 통하여 해당 기록물의 생산목적이나 용도를 이해하기란 쉽지가 않다. 그러나 본 기술요소는 복잡하고 혼란스러운 기록물의 행정연혁이나 개인이력의 요소를 구조화시키기 위해서는 반드시 배경영역에서 필요한 기술요소이다.

데이터의 획득 및 검증에서는 원 생산시스템에서 보존소로 이관되기 전까지의 기록물 이력의 주요 변화에 따라 수행된 획득 및 검증방법을 기술하여 데이터세트에 대한 무결성 및 신뢰성을 높이는 항목으로 지적된다.

4.3 내용영역

ISAD(G)에서는 내용과 구조영역에서 범위와 내용, 평가·폐기·처리일정의 정보, 추가, 정리체계 요소등을 제시하고 있으나, 데이터세트의 구조적 복잡성으로 인해 내용영역과 구조영역을 분리하여야 할 것이다. 내용영역에서는 구조에 해당하는 정리체계를 제외한 요소들을 기술한다.

범위항목에서는 주제나 내용상에 나타난 주요한 시간, 지역, 연령, 직업, 교육수준 등의 범위를 기술하고, 내용에는 주제어나 간략한 내용요약 및 설문, 실험, 통계 등의 형식까지도 상세히 기술하여 기술단위에 대한 이용자의 이해를 돕도록 해야 한다.

4.4 구조영역

데이터세트와 관련하여 기술단위의 내부구조상의 기술은 시리즈 계층과 파일 및 아이тем 계층으로 구분하여 살펴볼 필요가 있다. 파일 및 아이тем 계층에 해당하는 테이블과 필드의 내부구조는 데이터베이스 설계자에 의해 결정되지만, 시리즈 계층은 임의로 생산될 수도 있어 배열순서가 항상 데이터베이스 설계자에 의

(표 7) 내용영역 및 구조영역 요소

기술영역	기술요소	하위요소
내용영역	범위와 내용	범위, 내용(주제어, 내용요약, 내용형식), 내용참조
	평가·폐기·처리일정	처리기관(개인), 처리방법, 처리근거, 처리일정, 보존기간, 공개수준
	추가	추가일정, 추가분량, 추가 기록물 유형, 추가목적, 추가방법
구조영역	논리적 구조 및 스키마	하위 기술단위 정보, 관련성
	데이터베이스 특성	데이터베이스 특성, 데이터 추가 및 갱신, 스냅샷의 방법 및 일정

해 결정되지는 않는다. 특히 동적인 데이터세트의 경우 아키비스트에 의해 인위적으로 설정 기술되어 지기도 한다.

논리적 구조 및 스키마는 데이터세트의 내부 구조에 대한 상세사항을 기술하고 있다. 포괄하는 기술단위의 명칭과 규모를 기술한 후, 데이터세트 내의 테이블과 파일관계에 관한 정보 및 시스템 설계와 구조에 대한 정보 등을 세밀히 기술한다. 또한 구조적 변화에 대한 사항도 함께 기술할 수 있다.

데이터베이스의 특성은 데이터세트에 관한 원 데이터베이스 시스템의 동적인 특성에 관한 상세정보를 제공해 주는 기술항목이라고 할 수 있다.

4.5 열람과 이용조건영역

데이터세트는 포맷 등에 영향을 받지 않는 원 데이터형식이기 때문에 시스템적 제한사항이 존재하지 않으므로 이용자에게는 데이터의 저장형식에 대한 정보보다 이를 활용할 수 있는 방법에 대한 정보가 더 중요하다.

열람영역은 데이터세트의 열람에 영향을 미치는 법규를 비롯하여 이와 유사한 영향력을 갖는 국제적 협약이나 규정 및 생산자 또는 기증자의 요청에 의한 열람 규제정보를 기술하는

항목으로, 본 요소를 기술할 때는 가장 기본적인 열람 조건인 기술단위의 비공개기간이나 공개일자 등을 기술하며, 부분공개인 경우 그 공개의 범위를 상세히 기술한다. 또한 저작권과 관련하여 열람 이후에 기술단위의 재생산 활동(복사, 출판)에 관한 제한조건을 기술하기 위한 재생산조건의 기술요소가 요구된다. 기술단위에서는 주요 언어와 기호체계 및 약어를 기술하기 위한 언어와 소장사항에 대한 물리적·지적통제를 표현해야 하며, 동시에 내부 이용자에게는 관리상의 편의를 주고, 외부 이용자에게는 검색상의 편의를 제공 해주는 검색도구에 관한 정보도 기술해야 한다(이소연 2002).

4.6 연관자료영역

기록물 보존소로 이관된 데이터세트는 대부분 생산부서에서 소장하고 있는 데이터의 사본이며, 경우에 따라 원 데이터세트 역시 수많은 버전이 존재하거나 지속적으로 다른 컴퓨터 시스템으로의 마이그레이션이 이루어진다. 이러한 환경에서 원본과 사본의 개념은 무의미하므로 원본 및 사본에 대한 정보보다 데이터세트의 생산에서부터 이관 및 보존에 이르기까지의 무결성 및 진본성을 증명할 검증과정에 대한 기술이 더욱 중요하다. 또한 상호 관련성이 많

은 데이터세트의 특성을 고려하여 기술단위의 내용을 파악하는데 직접적 관련성을 맺고 있는 기술단위에 관한 정보를 기술하여야 한다. 이러한 이유로 본 영역에서는 이용자들의 연구의 깊이를 심도있게 할 만한 정보를 제공하기 위해 관련기술단위와 출판자료에 관한 항목만을 기술요소로 분석 제시하였다.

4.7 원 시스템 영역

원 시스템 영역은 데이터베이스의 원 운영환경에 대한 세부사항을 기술하는 영역으로 배경 정보와 함께 생산 환경을 보다 명확하게 설명해주는 중요한 항목이다. 본 영역에서는 원래의 시스템으로 운영되었던 컴퓨터 하드웨어, 운영시스템, 애플리케이션, 사용자 인터페이스에 관한 정보를 기술함으로써 전산환경의 급진적 변화에 따라 이용자들이 원 컴퓨터 환경을 재구성할 수 있도록 세부적인 사항의 정보를 제공해야 할 것이다. 데이터세트는 포맷에 영향을 받지 않는 원 데이터 형식이기 때문에 데

이터세트가 본래 어떤 환경에서 운영되고 검색되었는지에 관한 정보를 미리 제공함으로써 이용자가 데이터세트를 보다 정확하게 이해할 수 있도록 함이다.

4.8 관리정보영역

관리정보 영역은 비독점 포맷에 저장되어 있는 데이터에 관한 신뢰성을 높이고, 데이터세트 내용에 대한 세밀한 평가를 할 수 있는 요소로서, 데이터세트가 생산된 원 시스템 환경에서 보존소로 이관될 때 이루어지는 디지털 처리과정 및 변환에 관한 정보를 기술한다. 또한, 디지털 처리 및 변환요소와 함께 보존소로 이관된 데이터세트 내용의 검증방법을 기술하는 검증요소들을 다루고 있다. 동시 데이터세트를 연구기록물로 변환할 때 업무과정상에서 전후의 데이터가 일치함을 확인하기 위하여 보존소에서 수행된 변환방법에 관한 검증요소 등도 기술함으로써 진본성·무결성 입증과 관련하여 이용자에게 중요한 정보를 제공해야 한다.

(표 8) 원 시스템 영역 요소

기술영역	기술요소	하위요소
원시스템영역	하드웨어	하드웨어 환경, 제조업체
	운영시스템	운영시스템 및 특징
	애플리케이션	애플리케이션 명칭, 제조업체, 프로그래밍 언어
	이용자인터페이스	

(표 9) 관리정보영역 요소

기술영역	기술요소	하위요소
관리정보영역	디지털 처리 및 변환	이관자료형식, 매체, 분량, 압축 및 변환
	내용 검증	검증이유, 검증일시, 검증방법
	변환 검증	변환이유, 변환일시, 변환방법

또한 기록물 기술에 관한 책임성을 부여하는 영역으로 기술담당자의 표시와 기술일자를 제공하고, 일면 기술작업에 영향을 미치는 규정 및 참조자료를 기술하기 위한 기술통제영역과 기타 다른 영역에서 기술될 수 없는 주요 정보를 제공하기 위한 주기영역을 기술영역의 일 분야로 설정하였다.

5. 결 론

데이터세트는 전통적 기록물과 달리 데이터 세트 내의 각종 정보를 재조직하여 새로운 정보를 생산할 수는 있으나 체계적인 관리활동이 이루어지지 않으면 아무리 활용성이 높은 데이터 내용이라도 본래의 생산 활동이 완료되면 원 내용과 생산 환경은 사라지는 것이 일반적이다. 이러한 현실에서 데이터세트의 보존의 필요성을 인식하고 생산단계에서부터 어떠한 활동들을 통해 어떤 유형의 데이터베이스가 설계되고, 데이터가 수집·축적되어 어떻게 활용되는지에 관한 일련의 과정을 이해하는 노력이 필요하다.

더욱이 내용 및 구조적인 면에서 복잡성을 띄고 있는 데이터세트의 경우 이용자의 이해를 돕고, 활용을 유도하기 위한 통일성 있는 기술의 필요성이 더욱 절실하다고 판단되어, 본 연구에서는 데이터세트의 기술적 요소에 초점을 맞추어 연구를 수행하였으며, 분석된 결과 내용은 다음과 같이 요약할 수 있다.

첫째, 기록으로서의 데이터세트는 개념상의 범위에서 거의 데이터베이스와 동일시되는 개념으로 데이터베이스 시스템의 활동성에 따라

정적·동적, 개방형·폐쇄형으로 구분되며, 전통적으로 스냅샷을 통해 데이터 추출이 이루어진다. 이렇게 획득된 데이터세트는 경우에 따라 인위적인 기술계층이 요구되지만 기본적인 기록물의 기술원칙은 일부 적용하여 기술업무가 가능할 수 있다.

둘째, 기록물의 기본적인 기술요소를 제시하고 있는 ISAD(G)에서도 데이터세트는 그 내용이나 구조상에서 복잡성을 지니며, 기술단위 간 밀접한 관련성을 가지고 있음에 관하여 상세한 기술이 이루어지고 있지 않다. 그러나 국제적 차원의 일관성 있는 기록물 기술을 위해 ISAD(G)를 준용하여 기술할 필요성이 따른다. 각국에서 제시하고 있는 기술영역과 기술요소들 역시 명칭은 상이하지만 많은 부분에서 ISAD(G)를 수용하고 있었다. ISAD(G)를 기준으로 공통적으로 제시되고 있는 영역들 이외에도 관리정보영역, 원 시스템특징영역은 전자기록으로서 진본성을 입증하기 위한 중요한 영역으로 데이터세트 기술에 있어 반영되어야 한다.

셋째, 특히 동적인 성향이 강한 데이터세트의 식별을 위해 본 연구에서는 정확한 일자요소 기술을 강조하였으며, 내용적·구조적으로 복잡한 관련성을 지니고 있기 때문에 생산목적 을 명확히 기술하고 내용과 구조영역의 분리를 통해 체계적으로 기술요소를 제시하고자 하였다. 또한 진본성과 무결성을 입증하기 위해 원 시스템 영역과 변환 및 검증에 관한 영역의 중요성을 강조하였다.

국내의 일부 연구를 통해 데이터세트를 기록으로서 수집·관리해야 한다는 주장이 제기되고 있으나, 이와 관련하여 국가적 차원의 실질적인 노력은 매우 미비한 수준이다. 그리하여

데이터세트와 관련해서는 기술업무만이 아니라 기록관리의 전반적인 측면에서 전산전문가와의 협력 하에 수많은 연구가 이루어져야 할

것이며, 이러한 활동에서 획득된 정보들이 보다 객관적이고 효과적인 기술규칙을 제정하는데 필히 반영되어야 할 것으로 사료된다.

참 고 문 헌

- 김명훈 외. 2004. 『전자기록관리의 이해』. 서울: 한국국가기록연구원.
국가기록원. 국가기록원 일반기록물 기술규칙. 2007. <<http://www.archives.go.kr>>
- 박진희. 2005. 『기록물용 KORMARC 데이터 필드 개발을 위한 메타데이터 요소에 관한 연구』. 박사학위논문. 중앙대학교 대학원, 문헌정보학과.
- 방효순. 2001. 기록물 기술을 위한 등록정보의 활용에 관한 연구. 『한국문헌정보학회지』, 35(4): 25-50.
- 서혜란. 2000. 기록물 기술의 표준화. 『기록학연구』, 1: 7-22.
- 설문원. 2005. 과학기술 분야 디지털 아카이빙을 위한 정책 연구. 『情報管理研究』, 36(4):23-49.
- 설문원. 2005. 국가기록관리 표준정비의 방향. 『한국기록관리학회지』, 5(1): 137-169.
- 윤대현 외. 2005. 전자기록물 관리체계에 대한 제언: 전자기록물 개념과 국제표준 사례 등을 중심으로. 『記錄管理保存』, 10: 1-32.
- 윤주범. 1999. 외국의 기록물 기술제도. 『記錄保存』, 12: 189-208.
- 윤주범. 2000. 기록물의 정리와 記述. 『記錄保存』, 13: 159-204.
- 이소연. ISAD(G)를 적용한 한국기록물기술규칙 개발에 관한 연구』. 서울: 한국국가기록원 연구보고서2.
- 이소연. 2004. 디지털유산의 장기적 보존: 국가정책 수립을 위한 제안. 『기록학연구』, 10: 27-64.
- 이영숙. 2003. 기록물 기술(Archival Description)에 대한 고찰. 『기록보존』, 16: 265-280.
- 현문수. 2005. 데이터세트 기록의 관리방안. 『한국기록관리학회지』, 5(2): 103-124.
- Canadian Committee on Archival Description. Rules for Archival Description. <<http://www.cdncouncilarchives.ca/archdesrules.html>>
- Darlington, Jeffrey. A National Archive of Datasets. Issue 39(2004). <<http://www.ariadne.ac.uk/issue39/ndad/>>
- DPT. From Digital Volatility to Digital Performance: Preserving Databases. Hague: Digital Preservation Testbed. 2003. <<http://www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence>>

- nce-databases-en.pdf>
Electronic Resource Preservation and Access
Network <ww.erpanet.org/>
ERPANET. 2003. The long-term preservation
of databases. ERPANET Workshop.
Bern.
ICA & IRMT. Managing Electronic Records.
IRMT: London.
<http://www.irmt.org/downloadlist/
education.html>
International Council on Archives. ISAD(G)2:
General International Standard Ar-
chival Description, second edition. Ot-
tawa: International Council on Ar-
chives Committee on Descriptive Stan-
dards. 2000.
Jamieson, Jim. The National Digital Archive
of Datasets. University of London
Computer Centre
<http://www.ulcc.ac.uk/fileadmin/d
ocuments/events/2005-11-02-JJ.pdf>
Miller, Ferdic. M. 2002. 『아카이브와 매뉴스
크립트의 정리와 기술』. 조경구 역. 서
울: 진리탐구.
The National Digital Archive of Dataset
<http://www.ndad.nationalarchives.
gov.uk/>
Procter, Margaret and Cook, Michael. 2000.
Manual of Archival Description. 3rd
ed. Brookfield. Vt: Gower.
Shepherd, Elizabeth. The Application of
ISAD(G) to the Description of Ar-
chival Datasets. ERPANET seminar.
2003.
Shepherd, Elizabeth & Smith, Charlotte.
2000. The application of ISAD(G) to
the description of archival datasets.
Journal of the Society of Archivists,
21(1): 55-86.
Verdegem, Remco. Databases preservation
issues. 2003.
<http://www.digitaleduurzaamheid.n
l/bibliotheek/docs/longterm__preserv
ation_of__databases.pdf>
The Society of American Archivists
<http://www.archivists.org>
University of London Computer Centre
<http://www.ulcc.ac.uk>