

공공기관 데이터세트 식별과 평가 절차 연구*

국가철도공단 전자조달시스템 사례를 중심으로

Study on Public Institution Dataset Identification and Evaluation Process :
Focusing on the Case of KR Electronic Procurement System

황진현(Hwang, jin hyun)**

백영미(Baek, young mi)*** · 임진희(Yim, jin hee)****

1. 서론
1) 연구의 배경 및 목적
2) 연구의 범위 및 방법
2. 데이터세트 기록의 식별 및 관리기준표 작성
1) 데이터세트 기록 식별의 단위와 범위
2) 데이터세트 기록 식별 절차
3) 공통데이터세트 기록 식별
4) 데이터세트 기록 관리기준표 작성
3. 데이터세트 기록 평가를 위한 도구 개발
1) 가치평가 지표
2) 데이터 활용성 평가 지표
3) 데이터 품질 평가 지표
4) 평가 결과에 따른 보존기간 책정
4. 데이터세트 기록의 이관 및 수집
1) 장기보존을 위한 이관
2) 기관 내·외부 연계 · 결합을 위한 수집
5. 결론

* 본 연구는 2021년 국가철도공단 'KR 기록정보자원 장기보존 체계 구축 사업' 중 컨설팅 결과를 바탕으로 작성되었음.

** 강릉원주대학교 기록관리협동과정 강사(제1저자)(jinhyunii@gmail.com).

*** 국가철도공단 기록물관리 전문요원(제2저자)(praha@kr.or.kr).

**** 명지대학교 기록정보과학전문대학원 조교수(교신저자)(yimjihkr@mju.ac.kr).

■ 투고일: 2021년 09월 30일 ■ 최초심사일: 2021년 10월 09일 ■ 최종확정일: 2021년 10월 15일.

■ 기록학연구 70, 41-83, 2021, <https://doi.org/10.20923/kjas.2021.70.041>

〈초록〉

공공기록물법 시행령 개정 이후 기록관에서는 데이터세트 기록 관리를 위한 관리기준표를 작성하고 이에 대한 관리와 통제를 수행하게 되었다. 이에 본 연구에서는 기록관의 체계적 데이터세트 기록 관리를 위한 데이터세트 기록 식별 절차와 평가 지표를 개발하였다. 이를 적용하여 국가철도공단의 KR전자조달시스템의 8개 데이터세트 기록 식별 후 관리기준표를 작성하였고, 평가지표에 따라 평가를 진행, 보존기간과 이관 및 수집 여부를 책정하였다. 데이터세트 기록의 관리 절차에 대한 구체적 사례가 부족한 시점에, 이 사례연구가 기록관에 실무적 도움이 되기를 바란다.

주제어 : 데이터세트, 행정정보시스템, 식별, 데이터세트 기록 식별, 관리기준표, 평가, 사전평가

〈Abstract〉

After the revision of the Enforcement Decree of the Public Records Act, the archives created a management standard table for data set records management and performed management and control. Therefore, in this study, the data set record identification procedure and evaluation index were developed for systematic data set record management of archives. By applying this, a management standard table was prepared after identifying the records of 8 datasets in kr's electronic procurement system, and the evaluation was carried out according to the evaluation index, and the retention period, transfer, and collection were determined. It is hoped that this case study will be of practical use to the archives at a time when concrete examples of procedures for the management of dataset records are lacking.

Keywords : dataset, public administrative information system, identification, dataset records identification, records management standard table, appraisal, pre-appraisal

1. 서론

1) 연구의 배경 및 목적

행정정보 데이터세트 기록(이하, 데이터세트 기록)의 관리는 오랜 시간 기록관리계의 화두였다. 2005년 기록관리혁신 로드맵에서 행정정보 데이터세트가 기록관리 대상으로 언급된 이후 2010년 개정된 공공기록물법에 행정정보 데이터세트가 법적 관리 대상에 포함되었고, 기록의 유형 중 하나로 정의되었다. 그러나 여전히 관리범위의 사각지대에 놓인 것이 현실이며, 데이터세트 기록관리에 대한 기록관의 부담과 혼란은 해결되지 못한 상태이다.

이러한 문제들을 해결하고자 전국기록인대회, 학회 등에서도 논의가 여러 차례 이루어졌다. 그리고 기록관리계 여러 의견들을 수렴하여, 2020년 공공기록물법 시행령 개정을 시작으로 데이터세트 기록관리가 본격적으로 시작되었다. 국가기록원은 행정정보 데이터세트 기록관리 기준-관리기준표 작성 및 이관규격(NAK 35:2020(v1.0) 표준을 개발하고, 관련 실행매뉴얼을 배포하였으며 ‘행정정보 데이터세트 기록관리 실행방안 설명회’를 개최하였다. 기록관리계가 그동안 깊게 고민하던 과제인 행정정보 데이터세트의 기록관리에 대한 가장 기초적인 접근 방안과 방법을 제시(한국기록학회 외, 2020, 78)하였지만 여전히 보완점이 필요해 보이며, 이를 기록관에서 직접 실행하기에는 다양한 부분에서 어려움이 있을 수 있다. 국가기록원의 관계자는 데이터세트 관리를

위한 방법이 어렵지 않다고 설명한 바 있으나 실무 현장에서 기록관리 기준표 작성 등 데이터세트 기록을 관리하는 기록관에서는 간단하거나 쉬운 일이 아닐 수 있기 때문이다(한국기록학회 외, 2020, 57).

모든 공공기관이 체계적인 데이터세트 기록 관리를 위해 고심하고 있지만 기타 공공기관(이하 공공기관)의 경우는 관리 방안에 대한 접근이 다소 다를 수 있다. 국가기록원 직접관리기관을 제외한 대다수의 공공기관은 기록관의 역할과 함께 영구기록물관리기관으로서의 책임을 다해야 하기 때문이다. 관리기준표 작성은 물론 평가, 이관, 장기보존 등도 함께 고민해야 한다.

이에 본 연구에는 기록의 생산시점부터 관리, 보존까지 기록의 생애주기 전반을 통제해야 하는 기타 공공기관 기록관에서 데이터세트 기록을 관리하기 위한 방안을 소개하고자 한다. 구체적 사례로 국가철도공단(Korea National Railway, KR)에서 운영 중인 'KR전자조달시스템'을 대상으로 하여 데이터세트 기록 식별절차와 관리기준표 작성, 데이터세트 기록의 평가를 위한 평가지표 개발, 평가지표에 따른 데이터세트 기록의 보존기간 및 이관여부 책정의 과정을 설명한다.

데이터세트 기록 관리를 위해 국가기록원에서는 실행매뉴얼(국가기록원, 2020b)을 배포하고 설명회를 개최하였음에도 데이터세트 기록의 관리 절차에 대한 구체적 사례는 여전히 부족하다. 각급 기관에서 수행하는 업무와 행정정보시스템 운영 상황이 다르므로, 다양한 행정정보시스템 데이터세트 기록 관리의 사례들이 공유되고 축적될 필요가 있다. 이 사례연구가 데이터세트 기록관리를 시작하는 기록관에 실무적 도움이 되기를 희망한다.

2) 연구의 범위 및 방법

데이터세트 기록은 행정정보시스템의 구조적인 복잡성과 데이터베

이스 관리시스템에 종속적인 데이터의 IT기술적 특성 및 그 종류와 수량의 방대함으로 기록관리 절차를 만드는데 어려움(국가기록원, 2020a)을 겪었고, 해결방안을 도출하기 위해 여러 연구들이 수행되었다.

데이터세트 기록에 관한 첫 연구용역은 국가기록원에서 발주한 행정정보시스템 데이터세트 기록관리 연구(국가기록원, 2007)이다. 데이터세트 기록 관리를 위한 개념과 프로세스 등을 설계하였는데 이를 시작으로 후속 연구들이 이어졌다. 국가기록원은 행정정보데이터세트 기록관리 시범 시스템을 구축¹⁾하였으며 2016년에는 데이터세트 기록관리 방안을 마련²⁾하기도 하였다. 이러한 노력들은 현재의 체계를 마련하는데 중요한 밑거름이 되었다.

데이터세트 구조분석 및 진본성 보장 기록관리 기능모델 연구(국가기록원, 2015)에서는 균일한 데이터세트를 대상으로 하는 최소 기능단위 중 READ를 제외한 CREATE, DELETE, UPDATE 연산이 일어나는 기능으로 ‘단위기능’을 정의하고, 단위기능을 단위로 행정정보 데이터세트 기록 관리기준표를 작성해야 함을 설명하였다. 단위기능의 등장은 현 데이터세트 기록 관리기준표의 기반이 되는 내용이었다고 볼 수 있다.

2017년 차세대 기록관리 모델 재설계 연구에서는 중앙행정기관 등의 행정정보시스템에 대한 현장조사 및 시스템 분석을 통해 데이터세트의 생산 및 운영 현황을 파악하고 결과를 분석하였다. 시스템 현황 및 운영정보, 데이터베이스 스키마의 분석을 기반으로 한 관리단위 식별, 데이터세트 관리기준표 설계 및 작성은 데이터세트 기록 관리 프로세스와 관리기준표에 대한 구체적 사례를 도출하였다는 데 큰 의미가 있다.

지난 해 60여개의 행정정보시스템을 대상으로 한 2020년 행정정보 데이터세트 기록관리 체계 구축 사업에서는 중앙행정기관 외 각급 기관의 61개 행정정보시스템을 대상으로 시스템 분석 및 데이터세트 기

1) 국가기록원 기록정보화팀에서 2009년 수행함

2) 국가기록원 전자기록관리과에서 수행함

록 관리기준표를 작성하였으며 기록관리 실행 매뉴얼을 구축하였다. 이는 향후 관리기준표를 직접 작성 및 관리해야 하는 기록관들에게 일부이지만 데이터세트 관리기준표의 실질적인 사례를 보여주었다는 데에 의미가 있다. 그러나 대다수의 관리기준표가 1~2개 등 소수의 단위 기능을 도출하는 등 규모가 큰 편이었고 큰 편이었고 관리기준표 항목에 기입한 내용이 간략하여, 이러한 결과물로 데이터세트 기록의 체계적 관리가 가능할지에 대해 의문점을 갖게 하기도 하였다.

학계에서도 연구는 꾸준히 이루어졌다. 조은희, 임진희(2009)는 데이터세트 기록관리를 위해 식별의 중요성을 언급하였으며, 데이터세트를 기록 건 단위로 식별하고 관리단위로 삼아야 함을 주장하였다. 차세대 기록관리 모델 재설계 연구의 결과물을 기반으로 한 오세라, 박승훈, 임진희(2018), 오세라, 이해영(2019)의 연구에서는 행정정보시스템에서의 데이터세트 생산 및 관리 환경사례를 조사하여 관리 방안 개발의 기초 자료 또는 유사 연구에서 활용가능한 조사방법론을 제시, 데이터세트의 관리 기준을 설계하고, 실무 현장에서 적용 가능한 관리 절차를 제안하였다.

최근에는 공공기록물법에서 지칭하는 데이터세트가 공공데이터의 제공 및 이용 활성화에 관한 법률(이하 공공데이터법), 데이터 기반 행정 활성화에 관한 법률(이하 데이터행정기반법기반법) 등에도 영향을 받으면서 데이터세트의 활용적 측면을 고려하지 않을 수 없게 되었다. 이에 서지인(2020)은 공공데이터 사업과의 비교를 통해 행정정보 데이터세트관리의 문제점을 확인하고, 개선방안을 도출하였다. 단위기능 책정 시 정부기능분류표(BRM) 참조, 행정정보 데이터세트 관리기준표의 재구성, 유관 시스템과의 데이터 연계를 행정정보 데이터세트 관리 개선방안으로 제안하였다. 다만 공공데이터에서 의미하는 데이터는 데이터세트 기록 중 공개적으로 활용 가치가 있는 일부를 다루는 것이기에 관리 및 보존의 측면에서의 데이터세트 기록 관리에 대한 제안이 다

소 아쉬웠다.

다양한 연구들이 선행되었으나 기록관에서 데이터세트 기록의 관리를 위한 구체적 방법에 대해 설명한 연구는 부재하였다. 이에 본 연구에서는 행정정보시스템 내 데이터세트 기록의 식별 및 관리기준표 작성, 평가, 보존기간 책정 및 (영구)기록관리시스템으로의 이관 여부 결정까지의 과정 중 필요한 절차, 도구 등을 개발하고 이를 설명하였다. 국가철도공단에서 수행한 KR 기록정보자원 장기보존 체계 구축 사업의 컨설팅 과업에서 KR전자조달시스템을 대상으로 데이터세트 기록의 식별 및 평가를 연구자가 직접 수행하였기에 이를 사례로 하여 구체성을 더하였다. 전자조달 업무는 대다수의 공공기관에서 수행하는 업무이자 조달시스템을 직접 개발 또는 나라장터를 사용하고 있기 때문에 해당 업무와 데이터세트 기록에 대한 이해가 수월할 것이라 판단하였기 때문이다.

데이터세트 기록 관리에서 가장 먼저 수행해야 하는 것은 행정정보시스템 내에서 데이터세트 기록을 식별하는 것이다. 데이터세트 기록의 식별을 위해 행정정보시스템의 기능과 목적을 식별하는 1)시스템 구조 식별 단계와 데이터세트 기록 식별을 위한 2)시스템 세부 식별 단계로 구분하여 절차를 설계하였다.

해당 절차는 국가기록원의 행정정보 데이터세트 기록관리 실행매뉴얼(2020), 행정정보데이터세트 기록관리 체계 구축 연구보고서(2020) 결과를 참고하였다. 실행매뉴얼에서는 관리기준표 작성 절차, 연구보고서에서는 전체적인 사업 절차와 61개 시스템에 대한 시스템 분석서 및 관리기준표 예시들을 중점적으로 살펴보았고, 이를 기반으로 식별절차를 1차로 작성하였다. 이후 KR전자조달시스템에서 데이터세트 기록을 식별하는 과정을 실제 수행해보면서 필요한 절차들을 추가, 보완하여 최종 절차를 완성하였다. 데이터세트 기록의 식별은 기록관의 기록물관리전문요원은 물론 시스템 운영담당자의 협조를 반드시 필요로 하기

에 각 단계 별 주된 수행주체를 함께 제시하였다.

데이터세트 기록이 식별되면 기록관에서는 이에 대한 보존기간을 책정하고, 장기보존이 필요하다고 판단한 경우 (영구)기록관리시스템으로의 이관여부를 결정해야 한다. 기타 공공기관의 경우 국가기록원 이관 대상이 아니므로, 기관 내 (영구)기록관리시스템으로의 이관·수집 여부를 결정하고 데이터세트 기록의 안전한 장기보존의 책무가 있기 때문이다. 이를 위해서는 먼저 식별한 데이터세트 기록에 대한 평가가 선행되어야 하므로 평가를 위한 지표를 개발하였고, 업무 및 내용가치, 데이터 활용가능성, 데이터 품질 등 3가지 영역으로 구성하였다.

업무적, 내용적 가치는 국가철도공단의 기록물분류기준표, 보존기간 준칙을 참고하였고, 기록관에서 기록물 평가심의 시 기록의 가치를 판단하기 위한 업무 참고 빈도, 역사적 가치, 학술적 가치 등에 대한 질문이 포함된다. 데이터 활용 지표는 공공기관 데이터 자원 보존 개선방안 연구(행정안전부, 2019)의 중요 정보자원 선정지표 파트를 참고하였다. 국가적 보존가치가 있는 중요 디지털 행정정보 선정을 목적으로 개발한 이 선정지표에는 통계적 활용성 및 활용 요구의 다양성을 평가하기 위한 요소가 제시되어 있어 이를 참고하여 5개의 지표를 개발하였다. 데이터 품질평가를 위한 지표는 한국지능정보사회진흥원의 공공데이터 품질관리 매뉴얼(2018)의 품질관리 수준평가 자가진단 체크리스트의 지표 및 공공데이터 예방적 품질관리 진단가이드(2021)를 참고하여 개발하였다.

개발한 영역 별 평가지표를 바탕으로 보존기간과 (영구)기록관리시스템으로의 이관·수집 여부 결정을 하도록 돕는 문항도 함께 작성하였다. 데이터세트 기록의 경우 기록의 생산종결시점과 보존기간 기산시점 등이 다양하게 적용될 수 있기 때문에 조건에 따라 기산시점을 적용하는 트리거 기반의 유연한 보존기간 책정이 가능하도록 구성하였다. 이관·수집 관련 문항 또한 데이터세트 기록의 이관·수집여부는

물론 이관시점, 이관 주기, 이관 후 조치, 이관 방법, 알고리즘 및 화면 인터페이스 포착 필요 여부 등을 체크할 수 있도록 하였다. 특히 데이터세트 기록의 경우 시스템 내에서 운영되던 프로그래밍 수식 등을 함께 이관하여야 활용이 가능하기 때문에 이에 대한 필요 여부를 체크하도록 하는 등 유형에 맞는 이관·수집에 대한 정보를 파악하도록 하였다.

이관과 수집에 대한 정보를 함께 파악하도록 한 것은 본 연구에서 기록의 ‘평가’ 개념을 보다 넓게 적용하였기 때문이다. 보존기간이 만료된 시점에 이루어지는 사후평가라는 관점이 일반적이었던 기존의 평가 개념이 ISO 15489의 개정 등으로 넓게 확장되었고, 국가기록원에서 제시하고 있는 데이터세트 기록 관리기준표 양식에도 처분의 제약 발생 사항, 처분방법을 포함하고 있다. 이에 보존기간의 책정은 물론 기록의 처분 단계의 기록의 이관 및 수집 여부의 결정, 이관 및 수집 시 고려해야 할 사항까지를 판단하는 것을 평가로 보았다. 특히 데이터세트 기록은 시스템을 운영하려면 데이터를 계속 가지고 있어야 하는 경우가 있기에 사전평가가 필수적이라 보고, 평가지표에 따라 데이터세트 기록의 가치 판단 후 보존기간 책정과 이관 및 수집에 대한 정보까지를 판단하는 것을 평가 영역으로 포함하였다.

이 연구는 기타 공공기관 기록관에서 데이터세트 기록 관리를 하기 위한 구체적 절차와 도구, 그리고 이를 적용한 결과를 설명하는데 목적이 있다. 개발한 데이터세트 기록 식별절차와 평가지표의 적합성, 식별과 평가 결과의 타당성을 확인하기 위해 국가철도공단의 데이터세트 기록 평가심의회에서 2차례에 걸쳐 검토 받았다. 기록관에서 데이터세트 기록을 관리하는데 필요한 절차와 도구라는 의견을 들을 수 있었다. 타 공공기관에서도 본 절차와 지표 등을 적용할 수 있으나 기관 별 업무의 특성, 행정정보시스템의 규모와 운영방식 등에 따라 변경·보완이 필요할 수 있다.

2. 데이터세트 기록의 식별 및 관리기준표 작성

2장에서는 데이터세트 기록 식별 기준과 절차, 관리기준표 작성 과정을 KR전자조달시스템을 대상으로 하여 설명한다. KR전자조달시스템은 국가철도공단에서 보유 중인 행정정보시스템 중 하나로 조달업무를 전자적으로 처리하기 위해 운용하는 정보시스템이다. 2004년 ERP(Enterprise Resource Planning) 기반의 통합정보시스템 구축 사업 중 하나의 응용프로그램으로 구축되어 몇 차례의 고도화를 거쳐 현재 버전까지 이어져왔다. 그러나 다가오는 2025년 조달청에서 구축하는 (신)조달시스템의 기능을 국가철도공단에서도 다수 사용할 예정으로 현 조달시스템의 기능들이 축소되고 사용량이 줄어들 것으로 전망하고 있다. KR전자조달시스템은 현재 대부분의 업무담당자가 사용할 뿐만 아니라 공단 내 주요 업무 수행의 과정과 결과가 남아있는 주요 시스템이자, 향후 시스템의 기능 축소 및 변경 가능성이 있어 데이터세트 기록에 대한 식별은 물론 장기보존 대상 및 이관 대상의 선별을 위한 평가가 중요하다.

구체적인 데이터세트 기록 식별 기준과 절차, 관리기준표 작성 과정은 다음과 같다.

1) 데이터세트 기록 식별의 단위와 범위

데이터세트를 기록으로 관리하기 위해서는 가치 평가 이전에 데이터세트의 어느 범주를 하나의 기록으로 볼 것인지 식별(identification)하고, 확정하여야 한다. 보존기간을 책정하거나 이관 여부를 결정하는 등 통제와 관리를 수행하기 위해서는 이를 위한 관리 단위를 확정하는 것이 필요하기 때문이다.(조은희, 임진희, 2009, 265)

데이터세트 기록 식별 시에는 그 단위를 어떻게 구분하고 정할 것인가

지에 대해 기준이 마련되어야 한다. 데이터세트 기록 별로 ①업무 대상 또는 유형이 상이한가, ②업무 프로세스가 동일하여 CRUD(Create, Read, Update, Delete)되는 데이터의 집합인가(데이터세트의 생애주기 동일), ③데이터세트 기록을 생산하는 처리과가 구분되는가, ④데이터세트 기록에 대한 업무의 근거법령이 다른가, ⑤데이터세트 기록의 원천 시스템이 어디인가 등 여러 요소들이 데이터세트 식별 단위 구분의 기준이 될 수 있다.

데이터세트 기록 식별을 위해 먼저 KR전자조달시스템을 살펴보았다. 국가철도공단은 다수의 사업을 발주 및 관리하는 기관으로 업무담당자들이 해당 시스템을 자주 사용하고 있었다. KR전자조달시스템에서는 크게 발주계획 및 요청, 입찰, 계약, 공개견적, 청구관리 등의 업무를 수행할 수 있다.

〈표 1〉 KR전자조달시스템 내 주요 업무

분야	주요 내용
발주계획 및 요청	발주계획서 작성, 발주 요청
입찰	공고등록, 게시(공동수급협정서, PQ심사, 현장설명, 투찰), 투찰마감(예기작성), 개찰(적격심사), 낙찰/유찰
계약	계약서 작성, 업체 서명, 전자결재, 공단 서명, 단순데이터 변경 요청
공개견적	견적서 요청, 견적서 작성, 견적서 확인, 견적서 개봉, 견적서 결과 확인
청구	기성(준공), 선금청구 / 청구서 작성 - 확인 - 업체서명 - 전자결재
현황정보	입찰현황, 계약현황, 낙찰현황, 청구현황, 업체정보, 계약자정보

시스템 내 주요 업무를 살펴본 후 기준과 절차에 따라 데이터세트 기록을 식별할 수 있다. 기준은 데이터세트 기록을 식별하는데 유용하나, 시스템마다 업무의 규모나 프로세스의 복잡도가 다르므로 경우에 따라 기준을 어떻게 적용할 것인지에 대한 논의가 활성화될 필요가 있다.

조달업무의 경우 하나의 사업이 발주계획 요청단계부터 청구 단계까지 하나의 세트로 이루어지는 케이스파일(case-file) 성격을 가지고 있다. 물론 업무를 구분하지 않고 KR전자조달시스템 데이터베이스 전체를 하나의 데이터세트 기록으로 식별하는 것도 가능하다. 다만 국가철도공단의 경우 공사기간이 10년 이상을 필요로 하는 사업들이 다수이고, 이는 하나의 데이터세트 기록이 완결되기 까지 오랜 시간이 소요될 수 있다는 것을 의미한다. 또한 발주계획은 하였으나 입찰이 되지 않거나, 입찰은 하였으나 계약이 완료되지 않는 등 모든 사업이 동일한 프로세스로 종결되지 않는 경우도 있다. 매년 대응해야 하는 국회, 감사원의 자료 제출 요구에서 사업 하나의 세트가 아닌 입찰현황, 계약현황과 같이 구분된 현황정보를 요구하는 등 업무 프로세스 별로 데이터세트 기록이 활용되기도 한다. 이러한 경우 케이스파일 성격의 업무라 할지라도 프로세스 별로 데이터세트 기록을 식별하는 것이 유용할 수 있다.

결국 시스템 마다 상황이 다를 수 있기에 업무와 생산되는 데이터의 특성에 따라 기준을 참고하여 데이터세트 기록으로 식별하는 것이 중요하다. 이 때에는 데이터세트 기록의 보존 외에도 활용성이 반드시 고려되어야 할 것이다. 시스템에서 수행하는 업무와 데이터가 동종으로 생산될 경우 하나의 데이터세트 기록으로 식별이 가능할 수 있지만, 다수의 업무와 그에 따른 트랜잭션이 발생하는 행정정보시스템의 경우 데이터세트 기록을 구분하여 식별하는 것이 체계적 관리 및 보존은 물론 활용에 도움이 될 수 있다. 무엇보다 시스템 내 '모든' 데이터들을 데이터세트 기록으로 식별하는 것이 중요하다. 보존가치나 활용가치가 낮거나 이관 대상으로 선별하지 않을 데이터세트 기록이라 하더라도 시스템 내 모든 데이터세트 기록을 식별한 후 관리 방안을 마련하여야 철저한 관리와 통제가 가능하다.

기준에 따라 데이터세트 기록을 식별할 때, 쿼리(query)와 같이 대상 데이터를 구체적으로 확정할 수 있는 명령문을 기술하는 등 논리적인

정의가 식별의 과정이 된다. 데이터세트 기록은 데이터베이스 내 저장되고 관리되기 때문에 데이터세트 기록 식별을 위해서는 해당 행정정보시스템의 데이터베이스 구조, 데이터 간 연관관계 등을 파악해야 한다. 이러한 분석을 통해 데이터세트 기록 별 쿼리문을 추출, 보정하여 하나의 데이터세트 기록을 식별할 수 있다.

2) 데이터세트 기록 식별 절차

데이터세트 기록을 식별하기 위하여 1)시스템 구조 식별과 2)시스템 세부 식별로 절차를 구분하였다. 먼저 시스템 구조 식별은 시스템 개발과 관련된 다양한 정보, 즉 EA, 테이블 명세서, ERD, 매뉴얼 등을 토대로 시스템의 목적과 기능 등에 대한 대략의 정보를 파악하여 기록성격을 고려한 상세기능 구분과 단위기능을 도출하는 단계다.

〈표 2〉 데이터세트 기록 식별 : 시스템 구조 식별 단계

단계	구분	내용	(주요)담당	비고
1	관련자료 확보 및 기본 내용 파악	EA, DB설명, DBMS 형태, 테이블명세, 기능명세, 적용업무, 매뉴얼 등을 확인하여 전반적인 시스템구조를 파악하고 심도있는 시스템 분석 인터뷰를 위한 자료를 생산함	시스템 운영 담당자	
2	시스템 분석	전반적인 시스템의 특징과 흐름, 외부 연계 및 시스템 사용자의 업무양상과 양태 등을 파악하고 시스템에 접근하여 데이터의 형태 및 수량 등 필요한 정보를 확인함	시스템 운영 담당자	
3	시스템 유형 및 유사(기능) 시스템 존재여부 파악	상용SW, 공동이용 시스템 및 유사시스템의 존재여부 및 각 시스템별 차이점, 시스템의 존재 목적과 관련된 근거 법령 등의 유무를 확인함	시스템 운영 담당자, 기록관	유사 시스템 통합 처리 여부 및 이해관계자 파악

4	타 시스템과 연계 여부 파악	타 시스템과 연계 여부에 따라 전송, 수신, 동기화 여부 및 주기, 데이터 사본여부 확인과 현재 데이터의 정확성에 문제가 있는지를 확인함	시스템 운영 담당자	데이터 공유 여부 및 주요 내용 확인
5	데이터 개방 여부 파악	데이터 개방 여부에 따라 개방 방식 및 시기, 공개 주기, 개방하는 데이터의 단위와 명칭을 확인하여 '데이터세트' 단위로 삼을 수 있을지 검토함	기록관	개방 시 개방단위 파악
6	시스템이 단일업무 수행용인지 파악	시스템 전체사용자와 권한 부여 정보 등을 확인하여 단일 사용자 혹은 부서 이상이 사용하는 시스템인지를 파악하고 시스템별 이해관계자를 식별하며 기록분류체계의 특정 계층과 시스템 전체를 연결가능한지 검토함	시스템 운영 담당자	BRM과 매핑 가능 여부 및 부서와 매핑 가능 여부 확인
7	시스템의 주요기능 도출 및 흐름 파악	각각의 기능을 대-중-소-최종기능 수준으로 계층을 구성하여 파악하고 해당 기능의 사용기간, 개별 법령이나 규정, 법적 서식의 존재여부 및 종결되고 더 이상 수정이 없는 시점을 식별함	시스템 운영 담당자, 기록관	한시 사용 혹은 특정 기간 사용 파악
8	시스템 단위기능(안) 도출	데이터에 대한 원래 활용목적, 공개여부, 생산단위 등을 고려하여 처분에 대한 의사결정이 비교적 단순하게 일어날 수 있도록 주제어와 함께 도출하며 해당 단위기능에 포함된 데이터 접근권한자도 식별함	기록관	공통과 고유 기능을 구분하여 도출

먼저 관련자료 확보 및 기본 내용 파악단계에서는 시스템에 대한 기본적인 내용 파악의 결과물과 함께 인터뷰를 위한 질문지를 설계하고, 시스템 기능을 기반으로 CRUD권한과 업무담당자를 파악해야 한다. KR 전자조달시스템의 경우도 매뉴얼을 우선적으로 검토하였고 이후 시스템 운영 담당자와 인터뷰를 진행하였다.

2단계인 시스템 분석 인터뷰는 시스템 운영 담당자 또는 유지보수를 담당하는 협력사 직원을 대상으로 진행하며, 1단계에서 파악하지 못한 내용 및 그에 대한 확인이 주된 목적이다. 시스템의 특징, 주요 데이터의 형태·수량·용량 등에 대한 내용 확보가 중요하다. KR전자조달시

시스템의 경우 시스템 유지보수를 수행하는 협력사 직원이 약 15년간 조달시스템을 담당하고 있어 시스템에 대한 이해도는 물론 그간의 고도화 이력, 업무담당자들의 요구사항 등을 구체적으로 파악하고 있었다. 이는 KR전자조달시스템을 분석하는데 큰 도움이 되었다.

3단계인 시스템 유형 및 유사(기능) 시스템 존재 여부 파악 단계에서는 시스템의 유형을 정의한 후 유사시스템을 파악하고, 이와 연계하여 근거법령 등을 파악한다.

4단계에서는 타 시스템 연계여부를 파악하는 것으로 인터뷰에서 확인한 정보를 토대로 타 시스템과 연계여부를 실제 확인하고 원출처(originality)와 동기화 자체의 적절성, 동기화 실행이 적절히 일어나는지 여부 등을 확인한다. KR전자조달시스템은 조달청의 나라장터, 국가철도공단의 ERP SAP(Systems, Applications, and Products in Data Processing) 등과 연계되고, 동기화로 인한 중복데이터가 발생할 수 있어 데이터의 원출처 여부를 파악하였다.

다음 5단계인 데이터 개방 여부 파악은 기관의 개방정책을 확인하고 개방을 총괄하는 부서와 접촉하여 개방항목과 범위를 확인하는 것이다. 주로 해당 업무 담당자의 업데이트 행위를 통해 데이터가 개방된다. 시스템 내에는 개방대상항목이 존재하나, 데이터세트 기록 단위보다 더 작은 단위인 특정 칼럼 단위로 목록정보를 개방할 수도 있다. 이는 데이터세트 기록의 활용적 측면 고려 시 근거가 될 수 있다.

6단계에서는 시스템이 단일 업무 수행용인지 파악한다. 단일업무시스템 여부는 하나의 부서에서 특정 단위업무 혹은 소기능과 매핑될 수 있는 수준이라 할 수 있다. KR전자조달시스템은 입찰, 계약, 청구 관리 등 복합적인 업무를 수행하는 시스템으로, 단일업무시스템에 해당되지 않았다.

7단계인 시스템의 주요기능 도출은 전체적으로 파악한 내용을 정리하여 시스템기능을 토대로 규정, 서식, 단위업무(또는 단위과제) 등을

매핑하고 정리하는 단계이다. 시스템기능은 메뉴기능과 동일하지 않을 수 있고 데이터세트 기록을 식별하기 위한 전 단계이므로 관련규정, 활용단위 등을 고려하여 도출할 필요가 있다. 마지막 단계인 시스템 단위 기능(안) 도출은 시스템 데이터세트 기록을 도출하여 레코드스케줄을 적용하기 위한 단계이다. KR전자조달시스템의 경우 발주계획, 입찰, 계약, 공개견적, 청구 총 5개의 시스템 고유 단위기능을 식별하였다.

다음은 시스템 세부 식별 단계이다. 시스템의 테이블구조와 엔티티 모델, 칼럼 정보 등을 토대로 단위기능과 단위기능을 구성하는 테이블 및 칼럼과의 연결을 목적으로 진행한다. 이 단계에서 데이터세트를 식별하고, 데이터세트 기록 별 쿼리문을 확보하게 된다. 논리 및 물리 ERD, DB 테이블 목록 및 명세서, 워크플로우 등을 확인하며 진행하는 것이 좋다.

〈표 3〉 데이터세트 기록 식별 : 시스템 세부 식별 단계

단계	구분	내용	담당	비고
1	총 테이블 목록에서 마스터 테이블 구분	시스템 구조식별단계에서 확보된 테이블 명세서를 통해 식별된 각각의 마스터테이블과 전체 테이블의 관계 및 내용을 파악함	시스템 운영 담당자	마스터테이블 기준으로 업무기능별 데이터 구분 가능 여부 파악
2	업무단위로 메뉴 및 테이블 그룹화(데이터 세트화)	ERD 혹은 파악되는 엔티티구조를 통해 의미단위로 기능을 구분하고 이를 테이블과 매핑하며 데이터세트 기록 단위를 식별함	시스템 운영 담당자	공동사용 및 구분불가 테이블을 파악하여 데이터 세트 기록 식별
3	테이블별 적용되는 법령, 고시, 규정 존재 파악	업무기능별 특정 테이블에 적용되는 상세 법령, 규정 등이 있는지 추가로 파악함	시스템 운영 담당자	테이블별로 구분되는 경우에 한정
4	업무담당자 및 기록관담당자 인터뷰	업무담당자에게 단위기능과 데이터세트 기록을 식별한 결과를 확인하고 이에 대한 검토 및 추가 정보를 확보함	기록관	업무수행 및 활용 차원에서 접근

5	테이블별 접근 권한 및 암호화 여부 파악	암호와 여부 및 내용과 이용자 계정(조직)별 접근권한을 파악하여 테이블별로 구분하고 테이블 내 수준이 다른 경우를 파악함	시스템 운영 담당자	암호화 여부 확인
6	타 시스템과 연계 시 연계 정보 파악	타 시스템과 연계가 확인된 경우 연계 테이블 내 정보의 읽기, 업데이트, 추가전용 여부를 칼럼과 로우단위로 확인함	시스템 운영 담당자	연계로 인해 생성, 변형되는 정보를 전체적 파악
7	특정 테이블의 칼럼의 정보 파악(설명) 및 PK와 FK 파악	테이블 내 칼럼의 의미를 확인하고, 칼럼간의 관계와 필드값 형태를 식별하며, FK와 PK정보를 확보함	시스템 운영 담당자	칼럼간의 연결관계, 타입 확인 및 테이블 분리 시 발생하는 문제 파악
8	쿼리문 확보(도출) 및 첨부파일 여부 파악	시스템 내 특정 테이블과 첨부파일이 연결되는지 확인하고, 그 결과 첨부파일의 관리 필요성을 검토 후 데이터세트 기록 단위 쿼리문에 반영하여 쿼리문을 도출함	시스템 운영 담당자	첨부파일의 별도 관리, 중복 여부 파악
9	단위기능과 데이터세트 기록, 테이블 매핑 확정	최종적으로 단위기능과 데이터세트 기록, 테이블 매핑을 확정하고 중복으로 연결된 테이블의 내용을 정리함	시스템 운영 담당자, 기록관	필요 시 묶음 재정의

가장 먼저 수행해야 하는 단계는 총 테이블 목록에서 마스터 테이블을 구분하는 것으로, 앞서 진행된 구조식별단계에서 확보된 테이블 명세와 단위기능을 토대로 마스터테이블을 식별하는 단계이다. 마스터테이블은 데이터세트 기록 식별에 주요한 기준점이 될 수 있으며, 이를 토대로 테이블 관계가 확보된다면 향후 쿼리문의 도출이 수월해질 수 있다.

다음 2단계는 업무단위로 메뉴 및 테이블 그룹화하는 것으로, 최종적으로 식별되는 데이터세트 기록을 확인하는 단계이다. 의미단위 및 활용단위 등을 고려하여 식별해야 한다. 이어 3단계는 테이블별 적용되는 법령, 고시, 규정 존재 파악하는 단계로 업무의 근거가 되는 주요한 법적 근거와 함께 데이터 운영에 적용되는 별도의 규정들을 검토해야 한다. 특히 보안 및 개인정보보호, 데이터개방 등에 대한 규정은 업

무상 활용과 무관하게 테이블별도 적용되거나 칼럼 단위로 적용될 수 있다.

4단계에서는 업무담당자 및 기록관담당자 인터뷰를 통해 단위기능 및 데이터세트 기록을 파악 및 적합성을 검토하고, 업무와 기록관리차원에서 추가정보를 확보하여야 한다. 데이터세트 기록 식별이 적절하였는지, 누락 및 중복 데이터세트 기록 식별이 발생하지 않았는지 등에 대한 검토가 반드시 실행되어야 한다. 업무담당자의 실제 업무경험 및 활용단위를 토대로 데이터세트 기록을 분리하거나 통합이 필요한지를 확인하는 것이 중요하다.

5단계인 테이블별 접근권한 및 암호화 여부 확인 단계는 특정 테이블이나 특정 칼럼에 접근권한이 분리되어 부여되거나 암호화되어 있는지 여부를 파악하는 것이다. 사용자계정 권한을 토대로 파악을 진행하며 쿼리문 도출 시 정상적으로 데이터를 확보할 수 있는지를 함께 검토해야 한다.

다음 6단계는 타 시스템과 연계 시 연계정보 파악하는 것으로 앞서 시스템 구조식별 단계에서 파악된 연계여부에 따라 실제 연계되는 데이터의 테이블과 칼럼 단위를 확인하는 것이다. 근본적으로 데이터의 원출처와 CRUD 상황을 고려한 종결단위 식별을 확인하기 위함이므로 연계규격이나 작동방식에 대한 상세한 정보를 확인하는 과정까지 수행할 필요는 없다.

7단계는 특정 테이블의 칼럼의 정보파악(설명) 및 PK(Primary key, 기본키)와 FK(Foreign key, 외래키)파악하는 것으로 앞서 식별된 데이터세트 기록 단위의 쿼리문 도출 등에 따라 필요한 부분에 한하여 진행하여야 한다. 만약 특정 테이블이 시스템에서 벗어날 경우 발생할 수 있는 오류나 PK와 FK가 데이터세트 기록 단위에서 join하여 작동할 수 있는지에 대한 검토 또한 필요하다.

8단계 쿼리문 확보(도출) 및 첨부파일여부 파악 단계는 향후 데이터

세트 기록의 이관을 위해 필요한 쿼리문과 (필요한) 첨부파일을 가져올 수 있는 방식을 식별하는 단계다. 쿼리문은 시스템에 접근권한을 얻어 엔지니어가 생산하며, 가능할 경우 기존에 가지고 있던 쿼리문을 참고하여 수정하는 형태로 도출하도록 한다. 이 때 첨부파일의 확보를 위한 쿼리도 별도로 식별하는 것이 좋다. 쿼리문의 테이블명, 칼럼명 등이 영문으로 된 물리명은 사용자 및 기록관에서 이해할 수 있는 의미 정보로 바꿔주도록 한다. KR전자조달시스템의 경우에는 시스템 유지 보수 협력사의 직원의 협조를 얻어 쿼리문을 도출하였다. 화면 메뉴구성과 테이블 구조가 유사하여 데이터세트 기록 별 쿼리문을 추출하는데 큰 어려움은 없었다. 하지만 데이터세트 기록이 여러 메뉴 혹은 테이블 및 칼럼값을 조합하여 구성되는 경우 쿼리문을 추출하는데 많은 시간이 소요될 수 있다. 무엇보다 시스템 운영 담당자의 협조가 필수적이기 때문에 데이터세트 기록 식별 시 쿼리문 필요성에 대한 설득이 요구된다.

마지막 9단계는 데이터세트 기록과 테이블 매핑을 확정하는 것으로, 식별의 최종단계다. 전체적인 오류를 확인하고 데이터세트 기록, 쿼리문 등에 대한 검증을 진행한다. 이 때 식별과정에서 도출된 모든 결과물을 문서화하고 정리하는 작업이 병행되어야 한다.

데이터세트 기록 식별 과정에서 가장 유념해야 할 것은 시스템 내 모든 데이터세트를 식별하는 것이다. 기록으로 관리할 것과 아닌 것을 구분하지 않고 모든 데이터세트를 식별한 후 관리기준표를 작성하고 보존기간 책정, 이관 및 수집대상 선별 등의 과정을 거치도록 한다.

위와 같은 방법에 따라 KR전자조달시스템의 데이터세트 기록을 식별하였다. KR전자조달시스템의 매뉴얼 분석, 시스템 관련 자료 등을 사전에 검토하고 시스템 운영 담당자 인터뷰를 통해 시스템을 직접 확인하며 확인 구성과 데이터 생산, 관리에 대해 설명 들었다. 이후 전자조달 시스템에서 수행하는 고유 업무에 대한 데이터세트 5개를 식별하였다.

〈표 4〉 KR전자조달시스템 데이터세트 기록 식별

구분	NO.	세트명	설명	주요 항목 예시	데이터분포(개) *최근 3년
고유	PS01	발주계획 요청 정보	- 발주계획서를 작성하고, 발주 요청한 정보	- 입찰공고명 - 예산액 - 발주년도/발주월 - 담당자 - 공고구분명	1,565 1,769 1,635
고유	PS02	입찰 정보	- 일반 및 수의입찰 건에 대하여 공고서를 작성 및 등록하고 투찰, 개찰, 낙 찰 등 입찰 과정의 완료 정보	- 계약의뢰번호 - 공고게시일자 - 입찰공고명 - 개찰일시 - 공고년도 - 공고구분명 - 공고차수 - 일련번호 - 투찰업체수 - 낙찰금액 - 공고상태	930 1,183 1,037
고유	PS03	계약 완료 정보	- 입찰이 완료된 건에 대하 여 계약서를 작성하고 업 체 및 공단의 서명에 의 해 계약이 완료된 정보	- 계약일자 - 계약번호 - 계약명 - PO 번호 - WBS 코드 - 공사구분 - 총체금액/ 차수금액 - 계약담당자/ 주관부서 - 준공여부	6,685 6,026 4,946
고유	PS04	공개견적서 정보	- 공개견적서에 대한 요청 과 작성, 개봉 결과 확인 이 완료된 정보	- 견적명 - 상태명 - 담당자명 - 견적미감일 - 견적미감분	28 49 48
고유	PS05	청구관리 정보	- 계약 건에 대해 선금 및 기성 청구 처리 완료 정 보	- 계약번호 - 계약명 - 총체금액 - 차수금액 - 청구서 작성 - 청구서일련번호 - 기성준공구분 - 선금금지급률	5,934 5,952 5,222

국가철도공단의 다른 시스템의 경우에도 데이터세트 기록 상위에 단위기능을 두지 않았다. 처음에는 단위기능과 데이터세트 기록 모두를 식별하였는데, 단위기능과 데이터세트가 1:1로 동일하였으며 명칭도 유사하였다. 데이터세트 기록의 명칭과 단위기능의 명칭을 동일하게 사용하여 두 가지 모두를 식별하고 운영하는 것이 가능하나, 큰 유용성을 발견하지 못했다. 다른 시스템의 경우 단위기능과 데이터세트 기록이 1:N으로 식별된다 하더라도 단위기능의 역할이 모호하고 데이터세트 기록 별 관리와 통제가 이루어질 것이라 보았기 때문에, 별도로 식별하지 않아도 무방하다고 판단하였다.

최종 데이터세트 기록 식별 시 쿼리문에 포함된 SELECT절의 칼럼값을 주요 항목으로 파악하여 데이터세트 기록에 대한 설명을 돕도록 하였으며, 분포도를 함께 추출하여 연간 생산량 추이를 확인할 수 있도록 하였다. 데이터 분포도는 데이터세트 기록의 보존기간을 책정하거나 이관 및 수집 여부 결정 시 필요한 근거가 될 수 있다.

3) 공통데이터세트 기록 식별

위와 같이 시스템 매뉴얼 및 관련 정보 검토, 시스템 운영 담당자 인터뷰 등으로 해당 시스템에 대한 정보를 파악하고 데이터세트 기록을 식별한다. 특히 기록관 입장에서는 시스템 화면에서 보여지는 메뉴가 데이터세트 기록 식별에 주요한 정보가 될 수 있다. 매뉴얼과 시스템 화면에서 직접 확인할 수 있는 데이터세트 기록은 해당 시스템에서 수행 중인 고유 업무 데이터세트 기록이다. 그러나 고유 데이터세트 이외에 시스템 공통 데이터세트 기록 식별도 필요하다.

공통 데이터세트는 모든 행정정보시스템에서 공통적으로 식별 가능한 데이터세트 기록을 의미한다. 각 행정정보시스템에서 처리되는 특정 업무에 대한 데이터세트 기록을 고유 데이터세트 기록이라 한다면,

모든 행정정보시스템에서 공통적으로 생산되는 데이터세트 기록을 ‘공통데이터세트 기록’으로 명명하였다. 행정정보시스템에서 고유 데이터세트 기록을 식별하는 것처럼 DB 내 데이터세트 기록뿐만 아니라 운영체제나 소프트웨어 실행 중 발생하는 이벤트가 기록된 로그파일 등도 식별하여 기록관의 관리대상으로 삼아야 한다. 그동안 기록관리계는 행정정보시스템에서 생산될 수 있는 데이터세트 기록 유형을 파악하는데 집중하였지만(국가기록원, 2017; 오세라 외, 2018) 업무와 관련된 데이터세트 기록 외 시스템 작동 전반에 대한 데이터세트 기록을 관리하여야 한다. 특히 로그파일은 국내에서도 인정보보호법, 개인정보의 안전성 확보조치 기준, 정보보호 및 개인정보보호 관리체계(ISMS-P) 인증 기준, 개인정보의 기술적·관리적 보호조치 기준 등에 따라 이미 주요한 관리대상이기에 기록관에서는 로그 파일에 대한 보존기간을 사전에 정의하고 관리 체계를 갖추어야 할 필요가 있다. 개인정보의 기술적·관리적 보호조치 기준의 제5조(접속기록의 위·변조방지)의 ‘접속한 기록을 월 1회 이상 정기적으로 확인·감독하여야 하며, 시스템 이상 유무의 확인 등을 위해 최소 1년 이상 접속기록을 보존·관리해야 한다’는 조항은 주의 깊게 볼 필요가 있다.

해외에서도 이미 로그 파일의 중요성에 대해 인식하고 이를 관리하기 위한 정책을 마련하고 있다. 시스템에는 업무 수행의 과정과 결과가 남아있으므로 특정 업무상의 이유로 또는 법적 요구 사항을 충족하기 위해(UB, 2010), 일어날 수 있는 문제에 대한 해결을 위해(AUC, 2018) 적절한 보유기간(retention schedule)을 책정하도록 하고 접근권한에 대한 가이드도 적극적으로 제시하고 있다. 시스템의 양호한 상태를 유지하고 보유하고 있는 데이터를 보호하기 위해 로그파일의 보존 및 관리, 통제가 필수적이라고 판단하고 시스템 및 디지털 서비스에서 생성된 로그 파일을 모두 저장, 보관할 뿐만 아니라 로그 파일 보존 지침에 따라 보존기간을 책정하고 폐기 방식을 따르도록 하고 있다(AUC, 2018;

UB, 2010).

KR전자조달시스템에서도 로그기록 등을 포함하여 시스템 작동 중 생성되는 공통 데이터세트 3개를 식별하였다.

〈표 5〉 KR전자조달시스템 공통 데이터세트 기록 식별

구분	NO.	세트명	설명	주요 항목	데이터분포(개) *최근 3년
공통	PS06	이용자 권한 정보	- 시스템 이용 정보	- 사번 - 이름 - 유저권한타입	3,022
공통	PS07	웹 서버 로그	- 인터넷망을 통해 접근한 접근정보 및 요청 자원 정보	- 클라이언트 아이피 - 접근일시 - 요청리소스 - 응답결과	연 당 365
공통	PS08	WAS 서버 로그	- 애플리케이션이 요청 받은 업무처리 수행 시 관리적 목적으로 기록하는 정보	- 요청처리 일 시간 - 로그레벨 - 애플리케이션 모듈 명 - 로그 본문	365일 x 24

이용자 권한 정보는 시스템 내 이용자 권한을 부여하는 데이터세트 기록으로, 인사관리시스템에서 보유하고 있는 국가철도공단의 전체 직원정보 중 KR전자조달시스템을 이용하는 직원들에게 시스템 내 어떠한 권한이 부여되었는지를 확인할 수 있는 데이터다. 시스템에 접속한 이용자들이 언제, 어디서, 어떤 페이지를 방문했는지 등에 대한 정보가 기록된 ‘웹 서버 로그’와 ‘WAS 서버 로그’도 데이터 항목을 확보하여 식별하였다. 공통 데이터세트는 시스템 화면에서 확인하기 어렵지만 시스템을 관리하면서 반드시 생성되는 데이터로 주로 시스템 운영 담당자만이 확인할 수 있다. KR전자조달시스템에서는 3개의 공통 데이터세트 기록을 식별하였지만 시스템 환경에 따라 추가적으로 공통 데이터세트 기록을 발견할 수 있다.

4) 데이터세트 기록 관리기준표 작성

데이터세트 기록 식별을 마친 후에는 관리기준표를 작성한다. 관리기준표는 데이터세트 기록의 관리를 위한 관리정보, 법규정보, 시스템정보, 데이터정보, 업무정보, 기록관리정보 등 6개 영역 36개 항목으로 구성(국가기록원, 2020, 12)된 서식을 그대로 사용하였다.

관리정보부터 업무정보까지의 5개 정보영역에 대해서는 시스템 운영부서의 도움이 절대적으로 필요하다. 특히 시스템 정보 영역의 EA정보나 DBMS 정보 항목, 데이터정보 영역은 시스템 운영 담당자 외에는 정확히 파악하기 어려운 항목들이 다수였다. 법규정보, 업무정보 등은 실질적으로 업무를 수행하는 담당자들의 역할이 중요하다. 만약 행정정보시스템의 운영 부서와 시스템을 통해 업무를 수행하는 부서가 다르거나 시스템 운영 담당자가 업무 프로세스에 대해 구체적으로 알지 못하는 경우 두 영역의 항목을 기입하기는 어려울 것이다. 본 연구에서는 시스템 운영 담당자는 물론 데이터세트 기록 별 업무 수행담당자를 모두 인터뷰하여 해당 정보를 파악하고 관리기준표를 작성하였다. 향후 기록관에서 데이터세트 관리기준표 초안 작성 시 시스템 운영 부서에만 맡긴다면 법규 및 업무정보 영역은 내용이 부족하게 기입될 우려가 있다.

KR전자조달시스템의 경우 시스템 내 고유 데이터세트 기록 5개, 공통 데이터세트 기록 3개 총 8개의 데이터세트 기록을 식별하였다. KR 전자조달시스템은 공통업무 성격으로 대다수의 업무담당자가 비슷한 업무이해도를 가지고 있었지만, 실제 사업을 발주하는 부서와 조달업무 전체를 관장하는 계약업무 부서가 시스템 이해관계자로 관리기준표 작성 시 두 부서의 의견이 모두 필요했다.

데이터세트 기록 식별 및 관리기준표 작성 과정에서 파악한 몇 가지 이슈는 다음과 같다. 먼저 관리기준표의 작성 단위의 모호함이다. 시행

령 제34조의 3(행정정보 데이터세트의 관리) 제1항에 따르면 ‘시스템을 기본단위로 하여 데이터세트를 관리’하며 관리기준표를 작성하도록 한다. 그러나 관리기준표에는 단위기능 별 보존기간, 처분관련 정보를 기입하도록 하고 있으며, 단위기능 내 데이터세트 기록이 복수로 식별될 경우 보존기간 및 처분 관련정보가 모두 달라지게 된다. 관리정보, 법규정보 등을 비롯한 5개 정보들은 시스템 단위의 작성이 적절할 수 있지만, 기록관리 정보는 처분 등 통제와 관련이 있는 만큼 관리기준표 작성 단위를 명확히 할 필요가 있다.

다음은 단위기능의 모호함이다. 관리기준표 작성 단위에서도 보듯이 단위기능 내에는 복수의 데이터세트 기록이 식별 될 수 있다. 관리기준표에서 지칭하는 단위업무는 데이터세트가 생산되어 포함되는 업무단위를 지칭하는 것으로 데이터세트와 동일하지 않다. 단위업무는 데이터세트와 1:1 또는 1:N 관계일 수 있다. 단위기능이 있음으로 하여 데이터세트가 어떠한 업무에서 생산된 것인지 유추할 수 있다는 장점이 있지만 실질적 관리단위로서 기능하지 못한다. 단위기능 대신 데이터세트를 관리 단위로 하여 통제할 수 있도록 하는 방향을 고려해볼 필요가 있다.

세 번째는 ‘기록관리 대상’ 만을 관리기준표에 기입한다는 것이다. 실행매뉴얼에 따르면 관리기준표의 기록관리정보-단위기능명에는 데이터세트 기록관리 대상이 되는 단위기능명을 기입하도록 되어 있다. 대상이 되는 것을 파악하여 적는다는 것은 단위기능과 데이터세트를 식별하는 과정에서 평가가 진행된다는 것을 의미한다. 시스템 내 데이터세트 중 장기적으로 보존가치가 있거나, 업무담당자들의 활용이 많거나, 역사적으로 높은 가치를 가지기 때문에 관리대상으로 삼아 관리기준표에 기입한다는 것 자체는 데이터세트에 대한 평가이다. 그러나 이 평가가 기준과 지침에 따라 체계적으로 이루어지는 것이 아니라 관리기준표 작성을 위해 시스템을 분석하는 과정 중 임의적으로 진행 될 확

률이 높다. 이에 관리기준표 작성 단계에서는 시스템 내 모든 데이터세트를 식별하고, 식별된 데이터세트들에 대하여 평가 진행 후 보존기간 책정 및 이관 대상을 선별하는 작업을 수행해야 한다.

마지막으로 공통 데이터세트에 대한 식별 문제이다. 본 연구에서는 KR전자조달시스템을 대상으로 이용자권한정보, 웹 서버 로그, WAS 서버 로그 등 공통데이터세트 3종을 식별하였다. 이 중 웹 및 WAS 서버 로그의 경우 시스템 운영 부서에서 1년을 주기로 하여 폐기하고 있었다. 시스템의 접속기록은 개인정보보호법 및 보호조치 등에 따라 1년 이상의 보존을 규정하고 있기 때문에 이에 따라 시스템 내에서 보존하다가 폐기하는 것이다. 이들 데이터세트는 타 고유 데이터세트와 달리 매년 평가를 진행하거나 평가심의회 안건으로 상정할 필요 없이 데이터세트 기록 별 명확한 보존기간을 책정해 주고, 보존기간이 만료되면 주기적으로 폐기하는 방법을 고려할 수 있다. 현 관리기준표에서는 시스템에서 수행하는 고유업무에 한하여 예시를 주고 있기 때문에 이러한 공통 데이터세트에 대한 식별, 보존기간 책정 및 처분 관련 정보 등의 기입 필요성에 대해 고민하게 된다. 공통 데이터세트는 매년 평가심의회에서 다루어야 할 필요는 없으나 식별하여 처분일정에 따라 관리해야 하는 데이터세트임이 분명하기 때문에 관리기준표를 통해 관리하여야 할 것이다.

3. 데이터세트 기록 평가를 위한 도구 개발

기록관 및 영구기록물관리기관에서 기록에 대한 평가란 ‘보존기간이 만료된 기록물을 행정적·사회적·역사적 가치를 검토하여 보존여부를 판단하는 업무절차’(국가기록원, 2012, 2014)로 폐기여부를 결정하려는 시점이 도래했을 때 가치를 검토하는 것이다. 기록관리기준표(또는 기

기록물분류기준표)에 따라 기록 생산시점에 단위과제(또는 단위업무)를 설정하면서 보존기간을 부여하는 것이 사전적 평가 개념이라고도 볼 수 있지만, 실질적인 평가는 보존기간이 만료된 시점에 이루어진다고 보는 것이 일반적일 것이다. 반면 ISO 15489-1:2016에서는 생산되고 획득되어야 할 기록의 결정, 관리 방식에 대한 결정, 기록시스템 설계와 실행, 방침과 절차의 개발, 메타데이터 요구사항의 정의, 기록 통제도구(처분지침, 접근과 허용 규칙 등) 개발 까지를 평가(Appraisal) 활동으로 정의하면서 평가의 범위를 더 넓게 바라보고 있다.

문서형 기록이 생산시점에 보존기간이 책정되고 처리과-기록관-영구기록물관리기관으로 이관 절차를 따르는 것과 달리, 데이터세트 기록은 현재 행정정보시스템 내에서 저장·관리되고 있다가 시스템 운영 종료, 데이터 품질 저하 등의 이슈가 있을 시 영구기록물관리시스템으로 이관을 고려하고, 이관을 고려하는 시점이 되어야 장기적으로 보존할 가치가 있는지를 판단하게 된다. 그러나 사전에 데이터세트 기록에 대한 식별과 평가가 진행되지 않으면 장기적으로 보존할 가치가 있는 데이터세트 기록의 체계적 관리가 어려워진다. 이에 데이터세트 기록 관리기준표를 만들고, 단위기능 또는 데이터세트 기록 별 보존기간 및 처분 관련 정보를 파악하는 등 사전적 평가가 반드시 필요하다.

데이터세트 기록도 다른 기록물 유형과 마찬가지로 보존기간 준칙 및 각 기관의 기록관리기준표(또는 기록물분류기준표)에 따라 보존기간을 책정하는 등 업무기능 및 내용 가치에 따라 평가를 진행하나, 데이터세트 기록이라는 특성에 맞게 데이터 활용 및 품질에 대한 평가도 함께 진행되어야 한다. 최근 공공데이터법, 데이터행정기반법 등에 따른 활용적 측면이 강조되기 때문에 데이터의 신뢰도와 이용가능성, 활용적 측면을 평가에 반영하여야 한다. 이러한 측면이 평가에 반영되었을 때, 데이터세트 기록이 생산되고 시스템 내 저장되어 있는 상태에서도 품질을 유지하고 신뢰성을 보장받을 수 있다.

이에 본 연구에서는 데이터세트 기록 특성을 반영한 일관적이고 명확한 데이터세트 기록 평가기준을 수립하고자 평가 도구를 개발하고, 이를 KR전자조달시스템의 5개 고유 데이터세트에 대한 평가에 적용해 보았다.

국가기록원의 실행매뉴얼 등에 따르면 아래의 양식을 응용하여 평가 심의회에 데이터의 영역별(볼륨, 범위지정 및 데이터 건 수 등으로 표기)로 상정(국가기록원, 2020b, 16)하도록 하며, 기록물철 분류번호는 행정정보 데이터세트의 명칭으로, 기록물 철 제목은 데이터세트 명칭이나 단위기능 명칭으로 대응하여 사용 보존하도록 한다. 기존의 기록물 평가 심의 양식을 응용한 형태로 기록관에서 익숙하게 운영할 수 있음(국가기록원, 2020a, 20)을 주장 하였다.

〈표 6〉 행정정보 데이터세트 평가심의 양식(국가기록원, 2020b, 16)

기록물철분류번호 (시스템명칭/데이터세트명칭)	생산연도 (데이터 생성 기간, 범위)	기록물철제목 (기록관리 단위기능, 대상 명칭)	보존기간만료일 (처분 제약기간 + 보존기간 종료일)	처리과 의견		기록물관리전문위원		심의회 의견
				처리 의견 (볼륨단위 처분 내역)	사유 (처분 사유)	평가의견	사유	
교육훈련 시스템	2001~2008	교육훈련시스템 2001~2008년 고객지원 자료실 1,480라인 (건, 등)	2019. 1. 1. (통계 목적상 5년 비치 후 보존기간 5년 경과)	보존기간 만료된 2008년 이전 자료실 1,480건에 대한 데이터세트 처분 필요. 시스템 영향도 없음	보존기간 경과	단순 공지 및 자료 정보로 처분제약 (비치)기간 및 보존기간 만료 기록으로 연차별 처분 필요	보존기간 경과	시스템내 타 데이터 연관성 확인 후 시스템 운영부서 최종 폐기

그러나 이는 데이터세트 기록의 유형적 특성과 평가영역 등을 고려

하지 않은 채 기존의 심의 양식을 그대로 사용하는 것에 지나지 않는다. 처리과의 의견과 기록관의 의견도 보존기간에만 초점을 두고 진행할 수 밖에 없다. 이를 해결하기 위해 데이터세트 기록의 유형적 특성을 고려하여 업무기능 및 내용가치 평가, 데이터 활용 및 품질평가 총 3개 영역 23문항으로 구성된 평가 지표를 개발하였다.

업무기능 및 내용가치 평가지표는 기록관리기준표의 단위업무와 보존기간 책정사유를 살펴보고 현행 평가기준(보존기간 준칙, 보존기간 별 책정기준) 및 활용적·증빙적·역사적 가치를 판단하도록 지표를 개발하였다. 데이터 활용 영역은 행정안전부 공공기관 데이터 자원 보존 개선방안 사업(2019)에서 중앙행정기관이 운영하고 있는 정보자원 중 보존대상 여부를 결정하는데 사용되는 지표를 참고하였다. 데이터세트 기록 평가 지표와 목적이 유사하고 시범 적용을 통한 실효성도 검증되었기에 적절하다고 판단하였다. 데이터 품질 영역은 한국지능정보사회진흥원의 공공데이터 품질관리 매뉴얼(2018) 및 공공데이터 예방적 품질관리 가이드(2021)를 기반으로 하였다. 공공데이터의 공개와 활용의 관점에서 기존의 데이터 품질관리 지침 등을 통합하고 구체적인 내용을 정리한 자료로 데이터 품질을 평가하는데 필요한 지표들이 다수 포함되어 있었다. 이와 같은 선행연구, 기준, 매뉴얼 분석을 기반으로 23개의 평가 지표를 직접 개발하였다.

1) 가치평가 지표

기록에 대한 가치평가는 업무기능과 내용가치 두 분야로 볼 수 있다. 먼저 업무기능 평가는 식별된 데이터세트 기록과 관련된 단위업무 또는 단위과제의 보존기간을 확인하는 것을 우선으로 한다.

〈표 7〉 업무기능 평가 지표

NO.	평가지표	평가 주체	답변	구체적 설명(예시)
1	데이터세트 기록과 관련된 단위업무(또는 단위과제)가 존재합니까?	기록관	O/X	단위업무명
2	단위업무(또는 단위과제)가 존재한다면 보존기간을 얼마이고, 책정사유는 무엇입니까?		OO년	보존기간 책정사유를 적어주세요.
3	위의 단위업무(또는 단위과제) 및 보존기간은 해당 데이터세트 기록에 적용하는 것이 적절합니까?	기록관 처리과	O/X	X의 경우, 내용가치 평가 결과를 따름

식별된 데이터세트 기록과 관련된 단위과제 또는 단위업무는 부재하거나 1:N으로 매핑될 수 있다. 국가철도공단은 기록물분류기준표를 사용하고 있어 단위업무와의 매핑을 살펴보았다. KR전자조달시스템 내 5개 데이터세트의 경우 각 데이터세트 기록 별로 정확히 매핑되는 단위업무는 없었고, 계약 관련 주무부서에서 사용하는 계약 관련 단위업무만이 존재했다. 이에 계약 관련 단위업무의 보존기간을 고유업무 데이터세트 기록 5개에 적용할 수 있었다.

다음은 내용가치에 대한 평가이다. 단위과제 또는 단위업무의 보존기간 외 업무 활용·증빙·역사적·학술적 가치를 확인하여 업무기능 영역 결과와 비교 후 긴 보존기간을 1차 보존기간으로 책정한다.

〈표 8〉 내용가치 평가 지표

NO.	평가지표	평가 주체	답변	구체적 설명(예시)
4	업무 수행과정에서 데이터세트 기록을 참고하는 기간은 몇 년 정도입니까?	처리과	OO년	주로 어디에 활용하는지 적어주세요. 예) 통계정보 작성
5	데이터세트 기록을 국회, 감사원 등에 근거자료로 제출할 경우, 그 사유와 제출을 위해 보관하는 기간은 몇 년입니까?		OO년	예) 국회 요구자료 / 최근 10년 간 데이터세트
6	데이터세트 기록에 적용되는 법령, 고시, 규정 등이 있다면 그 근거와 해당 보존기간을 기입해 주세요.		OO년	예) 시설물안전법 시행규칙 / 안전점검 및 정밀안전진단 결과보고서 10년

7	데이터세트 기록의 철도사업목적, 철도역사기록 등 역사적/학술적 가치는 몇 년 정도 입니까?	기록관	00년	예) 철도기술사업원 사업자료로 활용 등
---	--	-----	-----	-----------------------

내용가치 평가 지표에는 국회 및 감사원 등에 근거자료로 데이터세트 기록을 제출하는지 여부를 포함하였다. 다수의 기관에서 평가심의 시 보존기간이 만료되었음에도 국회 및 상급기관, 감사원 등에 증빙할 목적으로 폐기를 보류하거나 보존기간 상향을 요청하는 경우가 있어 KR전자조달시스템의 데이터세트 기록도 이러한 요구사항이 있는지를 파악하고자 하였다. 역사적 또는 학술적 가치를 묻는 평가 지표는 데이터세트 기록이 현재는 물론 앞으로의 필요성을 예측하는 것으로 주관적 판단일 수 있다. 업무담당자들의 의견보다 기록관 입장에서의 의견, 평가심의회 개최 시 외부 평가심의회원들의 주요 역할인 지표라 할 수 있다. 특히 입찰, 계약 데이터세트 기록의 경우 업무 참고, 증빙적 목적이외에도 철도 관련 사건, 사고의 원인을 추적하는데 중요한 정보가 될 수 있다는 의견을 평가심의회에서 확인할 수 있었다. 전자조달 업무에서 생산된 데이터세트 기록의 또 다른 활용성을 확인할 수 있는 중요한 발견이었다.

2) 데이터 활용성 평가 지표

최근 데이터의 활용에 대한 관심이 높아지고 있다. 공공데이터법, 데이터행정기반법 등에 따라 데이터의 활용적 측면이 강조되고 있다. KR 전자조달시스템의 데이터세트도 공공데이터로서 가능성이 있는지, 내·외부 데이터와의 연계·결합 가능성은 어떠한지에 대해 평가하는 지표를 개발하였다.

〈표 9〉 데이터 활용 평가 지표

NO.	평가지표	평가 주체	답변	구체적 설명(예시)
8	기관 내부 통계작성에 데이터가 자주 활용되니까?	처리과 기록관 시스템담당자	O/X	보고서나 결재문서에 활용하는 데이터예시를 적어주세요.
9	데이터가 공공데이터로 개방할 대상입니까?		O/X	공공데이터 개방기준을 확인하여 결과를 적어주세요.
10	내부의 정보자원과 연계·결합되어 활용할 만한 가치가 있습니까?		상/중/하	연계·결합 가능한 데이터를 적어주세요. (동일한 시스템 내 데이터 연계·결합도 포함)
11	해안(범정부빅데이터플랫폼), 공공데이터 포털 등 외부 기관 데이터 세트 기록들과 연계·결합되어 활용할 만한 가치가 있습니까?		O/X	연계·결합 가능한 데이터를 적어주세요.
12	해당 데이터세트 기록은 기록관에서 보유하고 있는 결재보고서 등과 연결, 제공합직한 데이터세트 기록입니까?		O/X	

통계적, 내·외부 자원과의 연계 활용성을 확인하여 평가 결과에 따라 활용가능성이 높다면 가치평가를 통해 1차 책정된 보존기간을 상향할 수 있다. 또한 데이터세트 기록이 그룹웨어, 업무관리시스템 등과 같은 전자기록생산시스템에서 생산되는 결재보고서 등과 연계되는 내용인지를 확인하여 연계가능성 여부를 파악하고자 하였다.

3) 데이터 품질 평가 지표

데이터세트 기록의 경우 품질에 대한 평가가 필요하다. 시스템 내에서 적절한 관리조치가 부재할 경우 데이터의 품질이 저하되거나 데이터 관리의 이력정보가 부재할 수 있고, 이는 데이터의 신뢰도 저하로 이어질 수 있다. 데이터세트 기록 또한 증빙에 활용될 수 있기 때문에

품질 관리는 필수적이다. 품질관리와 관련한 평가 지표는 다음과 같이 개발하였다.

〈표 10〉 데이터 품질 평가 지표

NO.	평가지표	평가 주체	답변	구체적 설명(예시)
13	데이터세트 품질관리 지침이 있습니까? (있다면 지침명을 적어주세요)	처리과	O/X	지침명을 적어주세요.
14	데이터세트 기록 품질관리 활동이 있습니까? (있다면 구체적 활동을 적어주세요)	시스템담당자	O/X	예) 0 / 모니터링 시행
15	분석에 필요한 raw 데이터가 충분히 확보되어 있습니까?	시스템담당자	상/중/하	집계에 자주 활용되는 raw 데이터 예시를 적어주세요.
16	raw 데이터는 신뢰할만 합니까?		O/X	
17	오류 데이터 발견 시 데이터 보정작업을 수행한 적이 있습니까? (있다면 수행 경험을 적어주세요)		O/X	연계-결합 가능한 데이터를 적어주세요.
18	데이터값 변경 시 이력정보를 생성하고 있습니까?		O/X	
19	시스템 운영 중 데이터세트 기록 항목의 추가, 데이터값 구조 변경 등의 변화가 있을 때, 데이터베이스 값들이 일관되고 무결한 상태를 유지하도록 반영하고 있습니까?	처리과 시스템담당자	O/X	변경 및 조치 사례를 적어주세요
20	테이블 간 연결정보(PK, FK)가 데이터베이스에 정의되어 있습니까?		O/X	
21	테이블명과 칼럼명의 의미정보가 데이터베이스에 정의되어 있습니까?		O/X	
22	데이터값 유효성 규칙이 데이터베이스에 정의되어 있습니까?		O/X	
23	사용되는 코드의 의미정보가 데이터베이스에 정의되어 있습니까?		O/X	

시스템 내 데이터세트의 품질관리 지침 또는 활동여부를 확인하고, 향후 활용 및 분석에 필요한 raw데이터의 신뢰성 여부를 평가지표에

포함하였다. 국가철도공단의 경우 수많은 철도공사에 대한 정보를 보유하고 있는 기관이므로 향후 국가철도공단 내부는 물론 외부의 데이터와 연계하여 분석할 만한 데이터를 시스템 내에서 충분히 보유하고 신뢰성 있게 관리하고 있는지를 파악하는 것이 중요하다. 또한 데이터의 보정, 변경, 추가 시 이력정보가 잘 관리되고 있는지를 확인하도록 하였다.

데이터 품질 평가 결과가 좋지 않을 경우 시스템의 데이터 관리 활동의 기능개선을 제안할 수 있다. 증빙적·활용적 목적으로 데이터세트 기록을 이관해야 할 때 데이터의 품질을 재고하고 신뢰성을 확보한 다음 이관할 수 있도록 해야 한다. 또한 행정정보시스템의 운영환경이 변화함에 따라 데이터세트 기록 품질은 주기적으로 확인하도록 한다.

4) 평가 결과에 따른 보존기간 책정

기록관에서는 행정정보 데이터세트 영역별 평가결과를 고려하여 최종 보존기간을 책정한다. 기존의 보존기간은 7종으로 구분되나 데이터세트 기록의 경우 7종에 국한되지 않고 데이터세트 쓰임에 따라 자유로운 보존기간을 책정할 수 있도록 제안하였다. 특히 기존의 단위업무 및 단위과제 보존기간이 적용되지 않을 수 있기 때문에 업무에 활용되거나 내·외부 데이터와의 연계·결합 가능성을 바탕으로 보존기간을 자유롭게 책정할 수 있도록 한다.

〈표 11〉 보존기간 책정 및 보유기간 산정

NO.	결정사항	평가주체	결정 내용	구체적 설명(예시)
1	보존기간(최종) 책정	기록관	00년	

2	생산 종결 시점	기록관	① 0000년 00월 00일 ② 직접입력	예) 0000년 00월 00일 계약 진행상태값 '완료' 공개건적서 상태값 '완료'
3	데이터세트 기록 보존 기간 기산시점	기록관	① 생산년도 다음 해 1월 1일 ② 직접입력(트리거)	예) 2022년 1월 1일 00계약 완료 후 00사업 종료 후 00사업 하자보증 종료 후
4	보유기간 종결시점	기록관	0000년 00월 00일	데이터 보존기간 기산시점 + 보존기간(최종)

보존기간의 기산일은 기록의 생산년도 다음 해 1월 1일로 하거나 특별한 경우 트리거를 입력할 수 있다. 예를 들어 '00공사 종결 후' 등으로 설정하여 세밀한 보존기간을 유지할 수 있도록 하는 것이다. 또한 공통 데이터세트의 경우 '해당 시스템 운영종료일' 등으로 책정할 수 있다. 데이터세트 기록의 보유기간은 데이터 보존기간 기산시점과 최종 보존기간을 합한 결과로 이를 위하여 생산 종결시점 및 데이터세트 기록 보존기간 기산시점도 결정하도록 한다.

마지막으로 (영구) 기록관리시스템의로의 이관 및 수집 여부를 결정하도록 하였다. 국가철도공단은 기록관과 영구기록물관리기관의 기능을 모두 수행하는 기록관리시스템을 직접 개발하여 운영 중이다. 식별한 데이터세트 기록 중 증빙 또는 활용의 목적에 의해 준영구 이상의 보존기간을 책정한 경우 기록관리시스템의로의 이관을 결정할 수 있다. 더하여 장기보존의 가치는 없으나 현 시스템 내에서는 이용자 권한 제한 등의 이유로 활용이 제한될 경우 기록관리시스템로 수집하여 국가철도공단 직원들의 자유로운 데이터세트 기록 활용에 도움이 될 수 있다. 기록관리시스템의로의 수집은 기록관의 증빙적 책임은 없으나 자유로운 활용을 보장하는 역할로, 수집한 데이터세트 기록의 품질이 다소 낮아도 무방하다. 다만 이 경우 원천시스템에서

데이터를 장기보존하는 조치들을 취하거나, 데이터세트 기록이 종결된 후 평가 결과에 따라 이관을 결정하는 등 추가적인 조치들을 고려할 수 있다.

위의 평가 지표를 기반으로 KR전자조달시스템의 8개 데이터세트 기록의 보존기간을 책정하였다. 입찰 정보(PS02), 계약 완료 정보(PS03), 공개 견적서 정보(PS04), 청구 관리 정보(PS05) 데이터세트는 계약 관련 단위업무 보존기간을 참고하여 모두 '영구'로 책정하였다. 발주계획 요청 정보(PS01)의 경우 매핑되는 단위업무는 부재하나 업무 시 참고하는 기간을 고려하여 보존기간 '10년'으로 책정하려 하였으나, 타 데이터세트 기록과의 연계 보존이 필요하다고 판단하여 최종 '영구'로 책정하도록 하였다. 공통 데이터세트인 이용자 권한정보(PS06)는 '시스템 운영기간 중', 웹 및 WAS 서버 로그(PS07, PS08)는 1년의 보존기간을 책정하였다. 웹 및 WAS 서버 로그의 보존기간 책정을 위해서는 개인정보보호법, 개인정보의 안전성 확보조치 기준, 정보보호 및 개인정보보호 관리체계(ISMS-P) 인증기준, 개인정보의 기술적·관리적 보호조치 기준을 참고하였다. 특히 개인정보의 안전성 확보조치 기준의 제8조((접속기록의 보관 및 점검)의 '시스템 내 개인정보가 있을 경우 각각 기록을 1년 이상 보관·관리'해야 한다는 조항이나 개인정보의 기술적·관리적 보호조치 기준의 제5조(접속기록의 위·변조방지)의 '접속한 기록을 월 1회 이상 정기적으로 확인·감독하여야 하며, 시스템 이상 유무의 확인 등을 위해 최소 1년 이상 접속기록을 보존·관리해야 한다'고 명시된 조항을 근거로 하였다.

보존기간 책정이 완료된 데이터세트 기록은 (영구)기록관리시스템으로 이관 또는 수집 여부를 결정해야 한다. 다음 장에서는 이관 및 수집의 시점 및 방법을 설명하도록 한다.

4. 데이터세트 기록의 이관 및 수집

1) 장기보존을 위한 이관

기록관은 장기보존 및 기록관의 증빙적 책임이 있는 데이터세트 기록에 대한 안전한 관리 및 보존의 의무가 있다. 영역 별 평가 결과를 고려하여 이관 또는 수집 여부를 결정한 후에는 (영구) 기록관리시스템의로의 이관 시점 및 주기, 이관 후 조치, 이관 방법 등을 마련해야 한다.

〈표 12〉 이관 및 수집의 주기와 방법

단계	결정사항	평가주체	결정 내용	구체적 설명(예시)
1	이관 및 수집 목적	기록관	① 원본 이관 ② 사본 수집	종료/업무종료가 완료 업무중이라도 가능/ 활용용으로만 사용 완전무결 아님 예) EPMS
2	이관 시점	기록관	① 업무 트랜잭션 종료 후 ② 시스템 고도화 시 ③ 시스템 폐지 시 ④ 직접 입력 (2023년12월31일)	
3	이관 주기	기록관	① 매일 ② 6개월 ③ 매년 ④ 직접 입력	
4	이관 후 조치	기록관	① 생산시스템 내 삭제 ② 생산시스템 내 보존 (원본) ③ 생산시스템 내 보존 (사본)	

5	알고리즘 및 화면 인터페이스 포착 필요 여부	기록관	O/X	데이터가 계산되어 보여지는 경우 수식/로직(알고리즘 포착) / 데이터가 보여지는 화면이 특징적이거나 데이터 배치가 의미있는 경우 화면 인터페이스 포착 필요
6	이관 방법	기록관	① 데이터베이스 덤프 ② 엑셀파일 형식으로 추출 ③ AIP 호출 ④ SIP Creator ⑤ 기타	

이관 시점 및 주기는 업무활동, 데이터 운용, 이관 비용 등을 고려하여 결정하도록 한다. 공공기록물법 시행령 제32조(기록물의 이관)에 따라 처리과에서 보존기간의 기산일부터 2년의 범위 내에 기록관으로 이관하여야 하며, 업무관리시스템으로 생산된 기록물은 매 1년 단위로 전년도 생산기록물을 기록관 하도록 하고 있다. 그러나 행정정보시스템의 경우 시스템 내에서 데이터가 지속적으로 쌓이고 연계되어 운영될 수 있어 일반적인 주기적 이관은 현실적이지 않다. 데이터세트 기록이 종결된 시점, 즉 데이터가 더 이상 추가 또는 수정되지 않는 시점이 되어 '종결' 되면 관리 및 활용 여부를 고려하여 이관 또는 수집을 결정하도록 한다.

알고리즘 및 화면 인터페이스 포착 필요여부는 시스템 내 화면이 특징적이거나 데이터의 프로그래밍 수식이 필수적인 경우 이관 시 포착 여부를 파악하는 것으로 데이터세트 이관 후 이용가능성에 대한 보장을 확인하기 위한 것이다. 이관방법으로는 데이터베이스 덤프, 엑셀파일 형식으로 추출, AIP 호출, SIP Creator 등의 방법을 선택지로 두었다. 표준(NAK 35:2020(v1.0))에서는 SIARD 형식규격을 권고하고 있으며, 국가기록원으로 이관하는 기관의 경우 이를 준수하여야 한다. 그러나 국가철도공단과 같은 기타 공공기관의 경우 자체관리기관이며 기관에서

독자적으로 사용하는 (영구)기록관리시스템으로 이관할 수 있기 때문에 SIARD를 제외한 이관 방법을 제시하도록 하였다. 시스템 간 데이터 이관의 방법론으로 자주 사용되는 데이터베이스 덤프, 엑셀파일 형식의 추출(export)/업로드(import), 네트워크를 통한 이관 등을 참고하여 선택지를 작성하였으며, 국가철도공단의 경우 기록관리시스템으로의 이관을 위해 직접 개발한 SIP Creator를 선택지에 포함하였다.

KR전자조달시스템의 경우 고유 데이터세트 기록 4개를 이관하기로 결정하였으며, 이관 시점을 업무 트랜잭션 종료 후로 설정하였다. 입찰 정보(PS02), 계약 완료 정보(PS03), 공개견적서 정보(PS04)는 ‘계약 완료 (체결) 후’ 라는 업무 트랜잭션의 종료를 시점으로 하였으며, 청구 관리 정보(PS05)의 경우 데이터세트 기록의 진행상태 칼럼값이 ‘완료’로 설정되었을 때를 이관 시점으로 결정하였다. 국가철도공단의 (영구)기록관리시스템의 경우 전자문서시스템의 생산 기록을 매일 이관 받는 체계를 갖추고 있어 데이터세트 기록의 경우도 이관 주기를 ‘매일’로 설정하였다. 알고리즘 및 화면 인터페이스 포착은 필요하지 않은 것으로 확인되었으며, 이관 방법으로는 국가철도공단이 직접 개발한 별도의 SIP Creator를 사용하여 이관하기로 하였다. 이외에도 기타 공공기관에서 구축하거나 운영하고 있는 (영구)기록관리시스템의 상황에 따라 이관의 시점, 방식 등을 조정할 수 있을 것이다.

2) 기관 내·외부 연계·결합을 위한 수집

(영구)기록관리시스템으로의 이관이 장기보존 및 증빙을 위한 목적이라면 수집은 활용적 측면을 강조한 행위다. 업무 트랜잭션이 종결되지 않았거나 데이터세트 기록에 대한 무결성이 다소 떨어진다고 하더라도 데이터의 연계·결합을 위한 목적이라면 (영구)기록관리시스템로 ‘사본’을 수집하는 것이다. 보존기간이 낮고 향후 이관 대상이 아니라

하더라도 수집이 가능하고, 향후 이관 대상이라 할지라도 활용의 요구가 많다면 (영구)기록관리시스템로 사본을 수집하여 활용하다가 업무 트랜잭션 종료 등 데이터세트 기록이 종결되면 원본을 이관할 수 있다.

이러한 수집을 위해서는 기관 내 또는 외부의 데이터와의 연계·결합 가능성을 살펴보는 것이 중요하다. 특히 기관 외부 데이터와의 연계·결합은 학술연구 등에 적극 이용될 수 있기 때문이다. KR전자조달시스템의 경우 철도 공사 과정 혹은 완공 후 사고가 발생했을 때 사고의 원인을 찾아나가는데 입찰 조건과 계약 등의 정보가 필수적이라는 평가심의회 외부위원의 의견이 있었다. KR전자조달시스템의 데이터세트 기록의 경우 대부분 이관되고, 해당 시스템에 대한 업무담당자들의 접근이 용이한 편이나, 시스템 내 기능 간 접근 권한이 제한되는 경우 (영구)기록관리시스템의로의 수집이 기관 내 활용을 위한 중요한 시작점이 될 수 있다.

5. 결론

2019년을 기준으로 대한민국 GEAP(범정부 EA(Enterprise Architecture) 시스템)에 등록된 공공기관의 정보시스템은 약 1만 6천여 개에 달하고, 중앙행정기관에는 약 1,700여 개의 시스템이 구축(국가기록원, 2020a)되어 있다고 한다. 실제 기관에서 사용하고 있는 행정정보시스템은 이보다 많을 것이며, 지방자치단체, 교육행정기관, 공공기관을 포함하면 셀 수 없이 많은 행정정보시스템이 존재할 것이다. 각 기관의 업무방식이 업무관리시스템, 전자문서시스템에서 생산하는 기안이 전부가 아닐 것이기에, 수많은 행정정보시스템 내 업무 과정의 설명책임성을 보여줄 수 있는 중요한 데이터세트 기록들이 보관되고 있다.

데이터세트 기록을 관리하기 위해 기록관리계의 노력은 오랜 시간

지속되었으나, 명쾌한 해결방안이 나오지 못했다. 2020년 공공기록물법 시행령 개정 이후 본격적인 데이터세트 기록 관리가 시작되었으나 여전히 모호한 부분과 보완해야 할 점들이 있다. 특히 데이터세트 기록관리라는 과업을 마주한 기록관은 데이터세트 기록이라는 유형적 부담감과 함께 참고할 만한 사례가 없어 어려움을 겪었다. 지난 해 60여개의 행정정보시스템을 대상으로 관리기준표 작성과 기록관리 실행방안을 적용하는 시범사업(국가기록원, 2020)이 이루어진 바 있고 올해도 이 대상을 확대하여 적용사례를 구현한다고 하나 관리기준표와 시스템 분석 결과가 간략하여 구체적 사례로 기능할 수 있을지 의문이 들기도 했다. 특히 공공기관의 경우 데이터세트 기록 관리 전반을 책임져야 하는 기관으로서 전체 프로세스에 대한 사례를 보여주는 것이 필요할 것이다.

본 연구는 KR전자조달시스템을 대상으로 하여 데이터세트 기록의 식별과정부터 관리기준표 작성, 평가 지표를 활용한 평가 결과를 설명하였다. 특히 행정정보시스템 내 관리대상 데이터세트 기록만을 식별하는 것이 아니라 '모든'데이터세트 기록을 식별하고, 이들을 업무기능, 내용가치, 데이터의 활용 및 품질 가치 영역에 따라 평가하고 그 결과를 도출해 내는 것이 이 연구의 핵심이다. 기록관이 데이터세트 기록을 관리하는 데 있어 시스템 전체를 파악하고 이에 대한 관리 및 통제권을 가지며, 기록의 유형적 특성에 따른 평가를 수행하는 등 전문성 있는 업무 수행이 가능토록 하는 것이 중요하다고 보았기 때문이다. 평가 지표 또한 업무기능의 가치를 절대적으로 두는 것을 넘어 데이터세트 기록의 활용성, 품질 평가 지표를 포함하였다. KR전자조달시스템을 대상으로 품질 평가 진행 시 부족한 점이 여럿 발견되었으며, 이를 통해 시스템 기능 개선 사항을 도출하는데 도움이 되었다.

본문에서 제시한 식별 기준 및 절차와 평가 지표는 필요한 모든 요건을 포함한 것으로 다소 복잡하고 상세하다. 이를 모두 적용할 수 있

다면 좋겠지만 모든 공공기관 기록관의 여건이 다르고 행정정보시스템마다 업무의 크기나 상세도 차이가 클 수 있다. 식별 절차와 평가 지표를 간소화하거나 경중을 다르게 하여 기록관 상황에 맞게 수정 및 보완하여 사용할 수 있다.

기록관이 직면한 데이터세트 기록 관리는 우리의 책무이기도 하지만 부담감의 대상이 되기도 한다. 비록 하나의 행정정보시스템을 대상으로 한 사례이기는 하나, 세부적인 과정과 결과를 공유하였기에 기록관에 실무적 도움이 되기를 바란다.

또한 연구를 진행하며 데이터세트 기록관리를 체계적으로 하기 위해서는 기록전문가들의 디지털 역량이 필요함을 느끼게 되었다. 시스템 운영 담당부서 및 담당자와의 협업 과정에서 서로 간의 소통을 원활히 하기 위해서는 IT에 대한 기본적 지식과 데이터베이스 등에 대한 이해가 요구되었다. 다양한 유형의 디지털 기록이 생겨나고 이에 대한 관리적 책임을 다해야 하는 만큼 많은 기록전문가들의 디지털 역량이 강화되어야 할 것이다.

〈참고문헌〉

- 서지인 (2020). 행정정보 데이터세트 관리 개선방안 연구: 공공데이터와의 비교를 중심으로. 한국기록관리학회지, 20(4), 41-58. DOI : 10.14404/JKSARM,2020.20.4.041
- 오세라, 박승훈, 임진희 (2018). 행정정보 데이터세트 사례 조사 연구. 한국기록관리학회지, 18(2), 109-133. DOI : 10.14404/JKSARM,2018.18.2.109
- 오세라, 이해영 (2019). 행정정보 데이터세트의 기록관리 방안. 한국기록관리학회지, 19(2), 51-76. DOI : 10.14404/JKSARM,2019.19.2.051
- 왕호성, 설문원 (2017). 행정정보 데이터세트 기록의 관리방안. 한국기록관리학회지, 17(3), 23-47. DOI : 10.14404/JKSARM,2017.17.3.023
- 조은희, 임진희 (2009). 행정정보 데이터세트 기록의 선별 기준 및 절차 연구, 기록학연구, 19, 251-291. UCI : G704-001548,2009.19.004

〈연구보고서〉

국가기록원 (2007). 행정정보시스템 데이터세트 기록관리 방안 연구보고서.
국가기록원 (2015). 데이터세트 구조분석 및 진본성 보장 기록관리 기능 모델 연구.
국가기록원 (2017). 차세대 기록관리 모델 재설계 연구 개발.
국가기록원 (2019). 데이터세트 유형 전자기록의 장기보존기술 연구.
국가기록원 (2020). 2020년 행정정보데이터세트 기록관리 체계 구축.
국가철도공단 (2022 발간예정). KR 기록정보자원 장기보존 체계 구축 최종보고서.
행정안전부 (2019). 공공기관 데이터 자원 보존 개선방안 연구.

〈표준 및 법령〉

개인정보 보호법. 법률 제16930호.
개인정보의 안전성 확보조치 기준. 행정안전부고시 제2019-47호.
개인정보의 기술적·관리적 보호조치 기준. 개인정보보호위원회고시 제2020-5호.
공공기록물관리에 관한 법률 시행령. 대통령령 제30584호.
국가기록원. 2020. 행정정보 데이터세트 기록관리 기준-관리기준표 작성 및 이관
규격(NAK 35:2020(v1.0))
한국인터넷진흥원. 정보보호 및 개인정보보호 관리체계(ISMS-P) 인증기준. [https://isms.kisa.or.kr/main/ispims/notice/?boardId=bbs_000000000000014&mode=vi
ew&cntId=9](https://isms.kisa.or.kr/main/ispims/notice/?boardId=bbs_000000000000014&mode=view&cntId=9)
한국지능정보사회진흥원 (2018). 공공데이터 품질관리 매뉴얼.
한국지능정보사회진흥원 (2021). 공공데이터 예방적 품질관리 진단가이드.
ISO 15489-1:2016,
Information and documentation — Records management — Part 1: Concepts and
principles

〈문서류〉

국가기록원 (2016). 행정정보 데이터세트 기록관리 방안
국가기록원 (2020a). 기록관리 이슈페이퍼 vol.23
국가기록원 (2020b). 행정정보 데이터세트 기록관리 실행 매뉴얼(기록관리 지침)
국가기록원 (2020c). 행정정보 데이터세트 시행령 개정에 따른 준비사항 안내(요
약).
한국기록학회, 한국기록관리학회, 한국기록전문가협회 (2020). 제12회 전국기록인
대회 자료집.