

행정정보 데이터세트 관리 개선방안 연구: 공공데이터와의 비교를 중심으로

A Study on the Improvement of Data Set Management in Government Information Systems: A Comparison with Public Data

서지인(Jiin Seo)

E-mail: archivething@naver.com

부산대학교대학원 문헌정보학과 기록관리학전공 박사과정



논문접수 2020-10-19

최초심사 2020-10-20

게재확정 2020-11-08

ORCID

Jiin Seo
<https://orcid.org/0000-0002-3448-8927>

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

초 록

많은 연구에서 행정정보 데이터세트의 중요성을 언급해왔으나 지금까지 실질적인 관리는 전무했다. 이런 상황에서 2020년 국가기록원은 행정정보 데이터세트 관리를 중점 추진사업으로 선정하며 본격적인 관리 작업에 착수하였다. 국가기록원의 노력에도 불구하고 업무를 수행하여야 할 기록관은 새로운 사업에 대한 우려를 보였다. 이에 이 연구에서는 공공기관에서 이미 실시하고 있는 공공데이터 사업과의 비교를 통해 행정정보 데이터세트 관리의 문제점을 확인하고, 개선방안을 찾고자 하였다. 법률, 고시, 지침, 소관 부처에서 발행한 간행물의 내용을 중심으로 대상을 분석, 비교하였다. 분석 결과를 토대로 단위기능 책정시 정부기능분류 참조, 행정정보 데이터세트 관리기준표 재구성, 유관 시스템과의 데이터 연계를 행정정보 데이터세트 관리의 개선방안으로 제안하였다.

ABSTRACT

Although numerous studies have noted the importance of data sets in government information systems, the practical management of data sets has yet to be developed. Under these circumstances, the National Archives of Korea designated data set management as a major project in 2020, initiating full-scale management work. Despite these efforts, the records center, which will conduct management, expressed great concern for the new project. As such, this study identifies problems in managing data sets and searches for possible improvements through a comparison with existing public data projects by public institutions. In particular, the following materials were analyzed: laws, notices, guidelines, and publications issued by the ministries. Based on the results, several measures were proposed as part of an improvement plan for data set management: (1) the utilization of government functional classification as a reference, (2) the reorganization of the table, and (3) data linkage with related systems.

Keywords: 데이터세트, 행정정보 데이터세트, 공공데이터, 오픈데이터, 행정정보시스템
Data set, Data sets in government information systems, Public data, Open government data, Public administrative information system

1. 서론

이때까지 많은 연구에서 공공기록물로서 행정정보 데이터세트 관리의 필요성을 강조해왔다. 또한 그 중요성에 비해 실질적인 관리체계는 빈약한 상황임을 인정하였다. 왕호성, 설문원(2017)은 공공기록물 관리에 관한 법률(이하 공공기록물법)에서 행정정보 데이터세트를 전자기록의 범주에 포함시키면서도 지난 10여 년간 구체적인 정책이나 절차가 나오지 않은 현실을 지적하였다.

행정정보 데이터세트 관리를 위해 국가기록원은 행정정보시스템 데이터세트 기록관리 연구용역(2007)을 시작으로 2016년 행정정보데이터세트 기록관리 방안을 발행하였다. 2017년에는 국가기록원 연구개발사업인 「차세대 기록관리 모델 재설계 연구」의 일환으로 행정정보 데이터세트의 실행방안을 마련하였으며, 2020년 기록관리 중점 추진사항을 통해 대외적으로 공개하였다. 이 자료에 따르면 국가기록원은 ‘행정정보데이터세트 기록관리 실행방안(안)’(이하 실행방안)은 법령 개정, 기관평가 계획, 국가기록관리 중장기 발전계획과 함께 중점 추진사항 중 하나로 정하고 공공기록물관리법 개정과 맞추어 2020년 하반기부터 해당 사업을 추진할 예정임을 밝혔다.

하지만 훨씬 이전부터 공공기관에서는 행정정보 데이터세트를 관리, 대국민에게 공개하고 있다. 바로 공공데이터 개방 정책으로, 이 정책은 “중앙정부·지방자치단체 및 공공기관이 보유·관리하는 공공데이터를 일반 국민이 자유롭게 이용할 수 있도록 다양한 형태로 개방·제공하고 새로운 데이터 가치를 창출하도록 지원”을 목적으로 하며, 2013년 10월 31일 공공데이터법 제정 및 시행을 시작으로 2020년 정책추진 8년 차를 맞이하였다. 행정안전부 홈페이지에 의하면 공공데이터 개방 관련 사업은 2013년 말과 비교하여 데이터 개방이 6.3배(730개 기관, 5,272개 → 33,600개), 민간 이용이 792배(13,973개 → 11,029,321개), 민간 웹/앱 개발사례 58.2배(42개 → 2,448개) 증가라는 성과를 거두었다. 2019년 기준 OECD 평가 3회 연속 세계 1위를 달성한 매우 성공적인 국가사업 중 하나이다.

공공데이터를 대상으로 한 연구는 행정 혹은 정보과학 분야에서 많이 다루어졌으며, 행정정보 데이터세트와 달리 활용, 공유를 위한 개방을 초점으로 진행되었다. 기록으로서의 공공데이터에 주목한 학술논문은 김유승의 연구(2014)가 유일하다고 볼 수 있다. 김유승(2014)은 각종 법령에서 공공데이터 외에도 행정정보, 정보, 전자기록물 등 여러 용어를 혼용하여 사용하며, 대다수가 이를 기록으로 간주하지 않음을 지적하였다.

이 연구는 선행사업이라고 할 수 있는 공공데이터 사업과 비교·대조를 통해 행정정보 데이터세트 관리의 개선방안을 도출하는 것에 목적이 있다. 이 연구에서는 공공기록물법과 공공데이터의 제공 및 이용 활성화에 관한 법률(이하 공공데이터법)을 비롯한 소관 부처에서 발행한 간행물 등을 분석하여 두 대상을 비교한다.

먼저 행정정보 데이터세트 관리와 관련한 사항은 행정정보데이터세트 기록관리 실행방안과 국가기록원 이슈페이퍼(이주광, 2020)를 중심으로 국가기록원에서 발간한 자료를 검토한다. 특히 실행방안은 행정정보 데이터세트의 기록관리 방향과 실행절차부터 관리기준표 작성, 처분지원 도구에 대한 사항 등이 구체적으로 포함되어 있으므로 해당 자료를 중점적으로 확인하였다.

공공데이터 관리에 관한 사항은 행정안전부에서 제공하는 공공데이터 관리지침 및 개방표준, 목록 등록가이드 등을 통해 확인한다. 공공데이터 관리지침은 공공데이터 관리 및 제공과 관련하여 공공기관이 준수해야 할 원칙과 기준을 제시한다. 이 지침은 공공데이터의 주요 용어개념 및 관리원칙에서부터 제공 단계(생성·수집, 처리·운영, 등록·관리, 제공, 사후관리)별 기준, 개방표준, 품질 관리 단계(계획, 구축, 운영, 활용)별 기준, 중복·유사 서비스 금지에 관한 사항으로 구성되어 있다.

공공데이터 운영을 위한 관리지침의 가장 최근 개정은 2019년 9월 3일(행정안전부 고시 제2019-71호)이며, 공공데이터 개방표준은 2020년 2월 4일 일부 개정되어, 2020년 2월 10일부터 시행(행정안전부 고시 제2020-5호)되었다. 관리지침 내 개방표준이 포함되어 있으나 해당 부분은 가장 최근 사항을 반영하기 위하여 행정안전부 고시 제2020-5호의 내용으로 확인하였다.

이 연구의 구성은 다음과 같다. 2장에서는 행정정보 데이터세트와 공공데이터의 범주와 관리단계를 확인하고, 3장에서 관리단계별 비교 분석을 통해 행정정보 데이터세트 실행방안의 문제점을 도출한다. 4장에서는 이를 토대로 한 행정정보 데이터세트 실행방안의 개선방안을 제안한다.

2. 행정정보 데이터세트와 공공데이터

2.1 대상

공공기록물법상 행정정보 데이터세트는 “「행정 효율과 협업 촉진에 관한 규정」 제3조제12호에 따른 행정정보 시스템에서 생산, 수집, 가공, 저장, 검색, 제공, 송신 및 수신 등을 위해 조합된 문자, 숫자, 도형, 이미지 및 그 밖의 데이터”를 말한다. 또한 데이터세트란 컴퓨터가 처리하거나 분석할 수 있는 형태로 존재하는 관련 정보의 집합체를 지칭한다. 즉 행정정보 데이터세트란 공공 기록관리의 대상으로서 행정정보시스템을 통해 처리되는 구조화된 데이터 집합체를 의미한다.

공공데이터법에 따르면 공공데이터란 “데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보”를 의미한다. 전자정부법에 따른 행정정보와 국가정보화 기본법에 따른 정보 중 공공기관이 생산한 정보, 공공기록물관리법에서의 전자기록물 중 웹기록물 및 행정정보 데이터세트 등이 해당한다. 세 유형 외에도 타 법령으로 정하는 자료 또는 정보도 공공데이터에 포함될 수 있다.

공공데이터법의 대상은 국가기관과 지방자치단체, 기타 공공기관으로 모든 공공기관이며, “개별 공공기관이 일상적 업무수행의 결과물로 생성 또는 수집취득한 다양한 형태(텍스트, 수치, 이미지, 동영상, 오디오 등)의 모든 자료 또는 정보”이다. 그리고 공공데이터 관리지침에서는 기관에서 관리하는 원천데이터 내 유사데이터의 묶음 및 집합인 데이터세트를 공공데이터 목록이라 표현하며 관리 단위로 설정하였다.

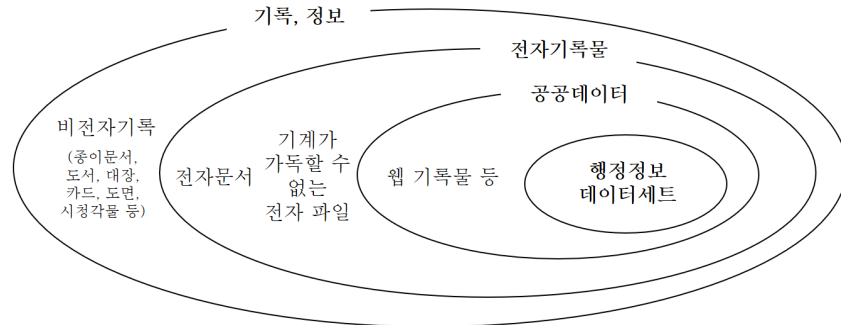
공공데이터의 기본 조건으로 기계 판독이 가능한 형태로 정하여 최소 “소프트웨어로 데이터의 개별내용 또는 내부구조를 확인하거나 수정, 변환, 추출 등 가공할 수 있는 상태”의 대상만을 선정한다. 현재 온-나라 문서시스템의 공문서 파일형식인 PDF는 대표적인 기계 판독이 불가능한 형태의 포맷으로, 전자문서는 공공데이터에 해당하지 않는다.

그리고 공공데이터 제공은 기존 정보공개와 구분하여 제공(개방)이란 용어를 사용한다(행정안전부 고시 제 2019-71호, 2019. 9. 3.). 공공기관의 정보공개에 관한 법률(이하 정보공개법)에 따른 공개 대상 중 기계판독이 가능한 형태 등 제공요건을 충족할 경우에만 공공데이터로써 개방이 가능하다고 명시한다. 비전자적 자료도 해당되며 특정한 공개 형태를 정해놓지 않은 정보공개 대상에 비해 공공데이터는 다소 한정적인 개념이라 할 수 있다.

<그림 1>은 법령을 바탕으로 공공데이터, 행정정보 데이터세트와 관련 개념의 범주를 정리한 것으로, 현행 법률에 따른 행정정보 데이터세트와 공공데이터의 범주를 비교하면 다음과 같다. 먼저 정보공개법에서의 정보는 전자형태를 포함하여 “공공기관이 직무상 작성 또는 취득하여 관리하고 있는 문서 등 이에 준하는 매체에 기록된 사항”을 의미한다. 정보공개법에서의 정보와 공공기록물법에서의 기록물의 개념이 흡사하다는 점을 미루어 보아, 법률상 정보와 기록은 동일한 개념이라 볼 수 있다. 기록과 정보는 유형에 따라 전자와 비전자로 구분할 수 있으며 행정정보 데이터세트는 전자문서, 웹 기록물과 같이 전자기록물에 해당한다.

그리고 공공데이터는 법률상 모든 자료 또는 정보라 규정되었으나 실제 기계가 가독할 수 있는 전자형태의 파일만을 대상으로 한다. 행정정보 데이터세트는 공공기관에서 활용하는 행정정보시스템에서 생산된 데이터로

웹 기록물 등과 함께 공공데이터의 한 유형으로 정해져 있다. 다시 말해 법률상 행정정보 데이터세트는 전자기록물의 대표적인 유형인 동시에 공공데이터에 해당한다.



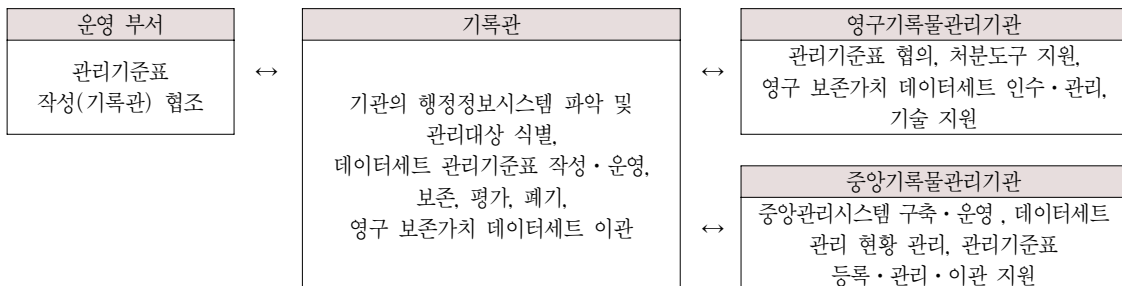
<그림 1> 행정정보 데이터세트와 공공데이터의 범주

2.2 관리 원칙 및 절차

국가기록원에서 발표한 실행방안에서 제시한 행정정보 데이터세트의 관리 방향은 총 네 가지이다. 첫째, 시스템을 기본단위로 생산시스템 자체관리 허용이다. 이는 기존 온-나라 문서시스템과 같은 업무관리시스템에서 채택한 ‘기록과 관련 메타데이터를 함께 이관’하는 방식과는 다른 방식이다. 전자기록생산시스템 기록관리 기능요건 (NAK 19-1:2012 v.1.0)에 의하면 데이터세트를 비치성 전자기록물로 보아 생산시스템과 지정 기록관리시스템을 연계하여 변동 내용을 현행화하는 방식이 적합함을 밝히고 있다. 하지만 실행방안에서는 위 표준에 제안된 바와 다르게 영구적으로 보존할 가치가 있는 일부 데이터세트에 대해서만 소관 영구기록물관리기관으로 이관토록 설계되었다는 점이 다르다.

둘째, 행정정보 데이터세트 관리기준표 작성이다. 이번 실행방안에서 가장 핵심적인 사항이라고 할 수 있으며, 실행방안에 따르면 관리기준표란 식별한 관리대상에 대한 기록관리 요소를 정리한 표이자 메타데이터라 밝혔다. 현재 전자문서 중심인 기록관리기준표의 구성요소인 보존기간, 보존기간 책정사유 등의 항목을 차용하였다.

셋째는 기록관-운영 부서-국가기록원의 유관부서와 협업체제 구성이며, 마지막으로 영구기록관의 기술지원 및 종합관리 방안 마련이다. 두 방향은 행정정보 데이터세트를 관리하는 데 있어 각 기관의 역할을 규정해둔 것이라 할 수 있다. 본문 내 기관별 역할을 확실히 명시하지 않았으나 정리해보면 <그림 2>와 같다. 운영 부서(처리부서)-기록관-영구기록물관리기관·중앙기록물관리기관으로 구성된 3단계의 구조는 기존 전자문서 관리체계에서 크게 벗어나지 않음을 알 수 있다(국가기록원 전자기록관리과, 2019).



<그림 2> 행정정보 데이터세트 단위기관별 역할

행정정보 데이터세트의 관리단계는 대상확인, 식별, 관리기준표 작성, 운영, 평가(폐기, 보류, 이관)로 구성된다. 마지막 평가단계에서는 관리기준표 작성시 기입한 처분제약사항과 보존기간에 따라 평가를 실시하며, 평가심의 결과에 따라 폐기 혹은 영구기록물관리기관으로의 이관을 진행한다. 이는 현재 기록관리체계와 유사한 형태이지만 기록 생산 전부터 설계하는 기록관리기준표와 달리 데이터세트 관리기준표는 기록 생산 후 관리대상을 식별 후 설계된다는 점, 생산 후 1년 이내 이관을 실시하도록 정해진 업무관리시스템과 달리 주기적인 이관일정이 없다는 점에서 차이가 있다.

<표 1> 행정정보 데이터세트 관리단계(국가기록원, 2020)

| 단계 | 설명 |
|----------|--|
| 식별 | 행정정보시스템(데이터세트)의 유형을 확인, 기록관리 적용 대상여부를 결정하는 단계 * 비대상 데이터세트는 제외사유 명시 후 관련 규정에 준수하여 관리 |
| 관리기준표 작성 | 행정정보데이터세트의 기록관리 요소를 정리한 표(메타데이터)를 작성하는 단계 |
| 평가 | 처분제약 기간 후 보존기간이 경과된 행정데이터(영역)에 대해 평가하는 단계 |
| 폐기 | 평가심의 결과에 따라 행정데이터(영역)를 삭제하는 단계 |
| 이관 | 중요 행정정보데이터세트를 이관 패키지 형태로 영구기록관으로 이관하는 단계 |

공공데이터의 관리는 관리주체의 지정에서부터 시작한다. 공공데이터의 관리주체는 해당 공공데이터의 소유권(ownership)을 가진 개별 공공기관이며, 해당 관리지침의 적용·시행에 대한 책임을 가진다. 각 기관은 공공데이터제공책임관 및 실무담당관을 정해 내부 규정을 정비하는 것을 시작으로 공공데이터의 생성·관리에서부터 등록·관리까지의 모든 단계를 수행한다. 또한 일부 위탁은 허용하나 산하기관 등을 포함한 총괄적인 관리는 금지한다. 공공데이터 제공 관리단계는 <표 2>와 같이 생성·수집, 처리·운영, 등록·관리, 제공, 사후관리의 5단계로 구분된다.

<표 2> 공공데이터 제공 관리단계(행정안전부 고시 제2019-71호, 2019. 9. 3.)

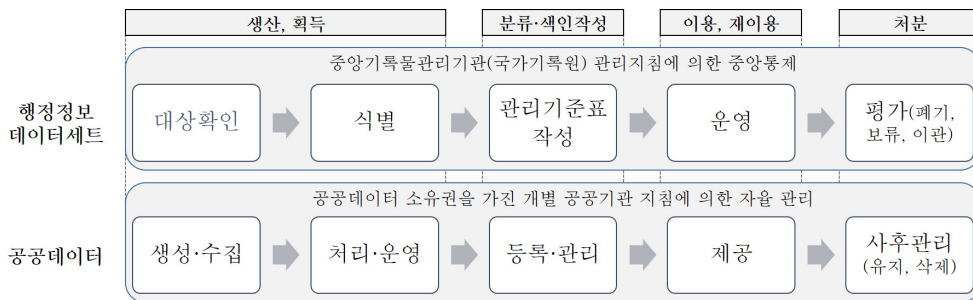
| 단계 | 설명 |
|-------|--|
| 생성·수집 | 각 기관이 생성하거나 또는 외부에서 수집·취득하는 공공데이터의 미제공대상 정보 포함여부 확인, 미제공대상 정보의 분리 가능 여부 확인 및 저작권 등 제3자 권리의 확보 등을 시행하는 단계 |
| 처리·운영 | 각 기관이 생성하거나 또는 외부에서 수집·취득한 공공데이터를 기계 판독이 가능한 형태로 생성·정비하고 데이터베이스, 파일 등으로 구축·운영하는 단계 |
| 등록·관리 | 각 기관이 생성 또는 수집·취득한 공공데이터의 목록 및 제공대상 공공데이터를 공공데이터포털(data.go.kr)에 등록하고, 변경사항 반영 등 유지·관리하는 단계 |
| 제공 | 각 기관이 공표된 소관 제공대상 공공데이터 또는 공표 제공대상 외의 공공데이터 신청에 따라 국민, 기업 등에 데이터를 제공하는 단계 |
| 사후관리 | 각 기관이 국민, 기업 등에 공공데이터 제공 후 갱신주기 및 변경사항을 반영한 현행화, 특정 사유 발생에 따라 공공데이터의 목록 제외 및 제공중단, 이용불편사항 접수·처리 등을 시행하는 단계 |

공공데이터의 획득은 생성·수집 단계에서 공공데이터 대상을 확인하고 선별하여 ‘처리·운영’ 단계에서 앞서 선별한 공공데이터를 기계 판독이 가능한 데이터 형태로 포맷을 구축하는 것부터 시작한다. 이 단계에서는 제공대상이 아닌 정보를 분리하여 데이터의 형태를 변경하고 제공방식을 정한다. 다음 등록·관리단계에서 목록을 작성하여 공공데이터포털에 등록하면 해당 데이터는 이용자에게 제공된다. 차후 데이터 갱신, 제공 중단, 관련 민원처리 등 주기적인 사후관리 단계를 통해 관리된다. 행정정보 데이터세트와 달리 공공데이터의 주요한 목적은 개방과

이용에 있으므로 이관, 보존 등의 절차가 없는 것을 확인할 수 있다.

<그림 3>은 행정정보 데이터세트와 공공데이터 관리단계를 ISO15489-1:2016의 기록 생산, 획득 및 관리 프로세스와 대조한 것이다. 각 단계가 ISO15489에서의 프로세스와 확실하게 상응하는 것은 아니겠으나 어느 정도 유사성을 고려하여 구분하였다. 행정정보 데이터세트와 공공데이터 관리단계를 생산, 획득, 분류·색인작성, 이용·재이용, 처분으로 구분하였으며 접근 통제, 저장, 마이그레이션 또는 변환은 공개된 지침에서는 명확히 확인되지 않아 제외하였다.

ISO15489에 따르면, 획득이란 특정 기록이 유지되고 관리될 필요가 있다고 평가되는 경우 시스템으로 기록을 입수하는 것을 의미한다. 생산된 기록을 확인하고, 선별하여 시스템의 영역으로 들이는 과정을 획득이라 표현할 수 있다. ISO15489의 기록 분류는 업무분류체계에 기록을 연결하는 작업을 의미한다. 보다 넓은 의미로 분류란 업무, 기능, 주제 등을 기반으로 한 정해진 질서에서 대상을 합류시켜 유기적인 관계를 형성하는 행위라고 할 수 있다. 이와 같은 사유로 행정정보 데이터세트의 관리기준표 작성, 공공데이터 목록을 작성하는 등록·관리 단계를 분류, 색인작성 단계로 선정하였다. 다음 장에서는 이용·재이용, 처분을 제외한 생산과 획득, 분류·색인작성 단계를 중점적으로 비교, 분석하고자 한다.



<그림 3> 행정정보 데이터세트와 공공데이터 관리단계 비교

3. 단계별 비교·분석 결과

3.1 생산·획득

행정정보 데이터세트의 대상 선정은 행정정보시스템의 유형을 확인하는 것에서 시작한다. 시스템 유형은 고유 업무용 시스템, 공통 행정업무 시스템, 기록관리 적용 시스템, 단순 지원 시스템과 보안·NW 관리 시스템으로 구분한다. 이 중 고유 업무용 시스템과 공통 행정업무 시스템에서 생산하는 데이터세트를 기록관리 대상으로 선정한다. 이미 기록관리가 적용된 시스템과 단순 지원 시스템, 보안·NW 관리 시스템의 데이터세트는 기록적 가치가 낮아 대상에서 제외함을 밝혔다. 관리대상이 되는 시스템으로는 전자관보, 환경분쟁조정시스템, 전자 인사 관리, 자산관리시스템 등을 예시로 들었다(국가기록원, 2020). 대상으로 결정한 행정정보시스템은 관리기준표를 작성하여 제시된 기준을 적용하며, 제외된 비대상 시스템은 그 이유를 명시하여 타 연관 규정에 따라 운용하도록 한다.

행정정보 데이터세트는 업무를 수행하는 과정에서 행정정보시스템에 자연스럽게 생산·축적된다. 이와 달리 공공데이터는 각 기관이 생성하거나 외부에서 수집·취득하는 데이터 중에서도 개인정보 등 미제공대상 정보 포함 및 대상 정보 분리 가능 여부, 저작권과 같은 제3자 권리 여부 등을 고려하여 선별된다.

먼저 공공데이터법에 따른 공공데이터 해당 여부를 판단하는 과정에서 대상이 정보공개법에 따른 정보에 해당 하더라도 개방요건에 부합하는지에 대한 여부를 확인한다. 다음으로 개인정보 등 미제공 대상정보 포함 여부 확인 후 기술적 분리 혹은 삭제, 비식별 조치를 통해 공개 가능한 상태로 제작한다. 만약 특정인이 공공데이터의 저작권을 소유하고 있을 경우에는 저작권자로부터 반드시 이용허락 동의서를 작성받아 이용 가능 범위를 확정하여야 한다. 마지막으로 공공데이터의 기본 요건인 기계 판독이 가능한 데이터 포맷으로 데이터베이스 파일을 구축하도록 하며, 활용성 증대를 위해 오픈포맷 활용을 권장한다(행정안전부 고시 제2019-71호, 2019. 9. 3.).

생산, 획득단계에서는 두 대상의 선별기준과 획득에서 차이가 있음을 알 수 있다. 선별에 있어 행정정보 데이터 세트는 데이터의 유형과 상관없이 해당 시스템의 기능과 목적을 기준으로 하는 반면 공공데이터는 데이터 유형, 포함된 정보의 내용을 고려하여 대상을 결정하였다. 획득방법에 있어서도 행정정보 데이터세트는 변조나 가공없이 시스템 단위로 관리를 시작하지만 공공데이터는 원활한 관리와 활용을 위해 기계가독한 형태로 변형하거나 데이터를 분리하는 등의 조치를 취한다. 관리단위 책정 시에도 공공데이터는 각 기관의 고유업무(단위과제, 주요 사업 등)에서 등록대상 공공데이터를 선정하여 데이터세트 단위로 관리하는 반면 행정정보 데이터세트는 기능, 업무보다는 보존과 재현을 위해 시스템을 관리단위로 선정한다.

유사한 대상을 취급함에도 이러한 차이점을 보이는 가장 큰 이유는 행정정보 데이터세트와 공공데이터의 주요한 목적은 각각 보존, 개방으로 상이하기 때문이다. 공공데이터는 정보자원으로써의 활용을 위해 최신성, 정확성, 상호연계성 등을 고려하여 지속적인 변형을 통해 품질을 확보한다. 또한 정보시스템 또는 업무 폐지 등의 사유가 발생하면 해당 데이터를 삭제하거나 목록에서 제외할 수 있다.

하지만 행정정보 데이터세트는 활용 여부, 시스템 교체 등과 상관없이 장기적인 보존 및 이용가능성 확보를 목적으로 한다. 공공데이터와 달리 데이터세트를 변형할 수 없으며, 시스템이 폐지되었다고 하더라도 적절한 절차를 거치지 않고서 폐기할 수 없다. 이러한 근본적인 차이는 대상 및 관리단위를 설정하고 획득하는 생산·획득단계뿐만 아니라 관리단계 전반에 영향을 미친다(<표 3> 참조).

<표 3> 행정정보 데이터세트와 공공데이터 생산·획득단계 비교

| 구분 | 행정정보 데이터세트 | 공공데이터 |
|--------|-------------------------------------|--|
| 목적 | 행정정보시스템 내 데이터세트의 보존과 장기 이용가능성 확보 | 민간 활용을 통한 신규 비즈니스와 일자리 창출, 국민편익 향상 |
| 대상, 형태 | 행정정보시스템 내 데이터 (데이터세트) | 전자적 자료 또는 정보 (반드시 기계판독이 가능한 형태) |
| 선정기준 | 생산 시스템 유형 (고유 업무용 시스템, 공통 행정업무 시스템) | 데이터 유형, 포함된 정보의 내용(비공개 정보 포함 여부), 저작권 등 제3자 권리 |
| 획득 | 변조나 가공없이 획득 | 기계가독한 형태로의 변형, 데이터 분리 후 획득 |
| 관리단위 | 행정정보시스템 | 데이터세트 |

3.2 분류·색인작성

분류, 색인작성은 보통 목록 혹은 메타데이터의 생성과 작성을 통해 이루어진다. 국가기록원의 실행방안에서는 행정정보 데이터세트의 기록관리 요소를 정한 표인 관리기준표를 제시하며, 관리기준표가 곧 기록관리 메타데이터 역할을 수행한다고 설명한다. 이주광(2020)은 데이터 집합으로부터 공통적인 메타요소를 도출, 시스템별로 데이터세트 기록관리를 위한 사항으로 구성된 관리용 항목표인 관리기준표를 만들고, 이를 데이터세트에 대한 간접적 관리를 위한 수단으로 활용할 것이라 언급하였다.

관리기준표를 살피기에 앞서 오세라, 이해영의 연구(2019)를 볼 필요가 있다. 이 연구는 2017년 국가기록원 연구개발사업 「차세대 기록관리 모델 재설계 연구」의 일환으로 수행된 산출물이며, 해당 연구의 산출물 중 하나가 현재의 관리기준표로 제시되었다. 오세라, 이해영(2019)은 기록 유형의 특성상 기존의 단위과제 중심의 정부 기능 분류체계를 적용하기 어렵고, 향후 다양한 분류체계의 개발을 고려하여 설명형 기준표를 채택하였으며, 기준표를 구성하는 영역과 세부 항목은 선행연구와 조사 결과를 토대로 정했다고 언급하였다. 관리기준표는 관리기관 정보, 시스템 정보, 업무 정보, 법규 정보, 데이터 정보, 기록관리 정보의 총 6개의 영역과 세부항목으로 구성되었다.

다음의 <표 4>는 오세라, 이해영의 연구(2019)와 국가기록원 실행방안을 토대로 관리기준표 항목을 비교한 표이다. 먼저 국가기록원이 공개한 기준표에는 다른 영역에 비해 오세라, 이해영(2019)의 기준표의 시스템 정보, 데이터 정보 영역 항목이 다소 제외되었음을 알 수 있다. 제외된 항목은 HW내역, SW내역, 백업정책, 이력관리(시스템), 데이터 해석 의존성, 정보공개, 연계데이터 현황(데이터 내용, 연계 방법, 활용 방법, 연계 방향), 이력 관리(데이터)로 총 8개의 상위항목이다. 특히 정보공개는 공공데이터 포털 등 대국민 공개 데이터 구분명이나 공개방법 등을 표기하여 정보공개 현황을 파악할 수 있도록 한 기술요소였다(오세라, 이해영, 2019, p. 65). 해당 영역별 제외항목 대신 시스템 개요, 업무데이터베이스명, EA정보, 암호화데이터 사용유무, 연계내역 첨부 등의 항목이 추가되었다.

<표 4> 행정정보 데이터세트 관리기준표 영역별 설명 및 항목

| 영역 | 설명 | 세부 항목(기술요소) |
|---------------|---|--|
| 관리기관 정보 | 행정정보시스템을 운영 및 관리하는 기관, 부서, 담당자 | 기관명, 부서명, 담당자 |
| 시스템 정보 | 시스템 명칭, 개요, 구성 등 | 시스템명, 시스템 개요, 구축연도, HW 내역, SW 내역, DBMS(정보, 업무DB명), 백업정책, 사용자 현황, 이력관리, 개발 산출물, EA정보 |
| 업무 정보 | 해당 시스템으로 수행하여 저장되는 데이터에 관한 업무 | 업무목적, 업무내용(업무명, 세부내용, 대상) |
| 법규 정보 | 시스템 구축 근거나 진행하는 행정업무의 근거 규정 등 | 법령, 고시, 규정, 표준 |
| 데이터 정보 | 실제 업무행위 결과로 저장되는 데이터 결과, 암호화 여부 등 | 대표 데이터, 비정형 데이터, 데이터 해석 의존성, 정보공개, 연계데이터 현황(데이터 내용, 연계 시스템, 연계 방법, 활용 방법, 연계 방향), 이력 관리, 암호화데이터 사용유무, 연계내역 첨부 |
| 기록관리 정보 | 데이터정보 및 업무정보를 근거로 기록관리를 위한 단위 선정, 보존기간, 공개여부 등 지정 | 단위기능, 업무활용 목적, 주제어, 데이터 소유권(데이터의 보유 권한), 데이터 관리권(시스템의 관리권한), 접근권한, 비공개 대상(정보공개 구분), 비밀 여부, 보존기간 및 처분(처분단위, 보존기간, 기산일, 적용 범위, 책정 사유, 처분 방법), 적용범위 관련 정보, 처분의 제약 발생사항(처분 지연의 사유) |
| 부가 정보(이관시 작성) | 어느 영역에도 해당하지 않는 정보 | 주요 서비스 내용, 이관대상(기관명, 부서명, 담당자명) |

* 기술임: 오세라, 이해영(2019), 밑줄: 국가기록원(2020)에만 있는 항목 표시

다음 기록관리 정보 영역에는 적용범위 관련 정보와 함께 처분의 제약발생사항(처분 지연의 사유)이 추가된 것을 확인할 수 있다. 이 항목은 책정된 보존기간과 별개로 대상 데이터가 다른 기능과 연계되어 활용되는 기간, 다른 업무에 참고되는 기능 등에 사용되는 기간 등을 고려하게 되면서 발생하는 처분 제약사항을 기재한다. 처분이 보류되거나 지연되는 기간이나 사유, 시스템 및 업무 운영상 필요한 데이터의 지연처분 요소 등이 이에 해당한다(이주광, 2020, p. 16). 그 외 주요 서비스 내용, 화면 캡처, 표현쿼리 등 주요 서비스 내용, 이관대상 등 이관시 작성하는 부가정보 영역이 추가되었다.

마지막으로 두 기준표에서는 공공기록물 관리체계에서 볼 수 있는 단위업무, 단위과제 대신 단위기능을 포함하였다. 단위기능은 이규철(2016)의 연구에서 처음 언급된 개념으로 전통적인 철, 건 구조를 적용할 수 없는 행정정보 데이터세트의 특성을 고려하여 데이터세트의 이관 및 보존, 관리 단위로 제시되었다. 오세라, 이해영(2019)은 기능 분석 - 데이터 모델 분석의 절차를 통해 단위기능을 도출할 수 있다고 보았으며 이를 관리기준표 중 기록관리 영역에 반영하였다.

단위기능이란 균일한 데이터세트를 대상으로 하는 최소 기능 단위 중 읽기를 제외한 생성, 수정, 삭제, 연산이 일어나는 기능을 의미한다. 단위기능은 시스템 전체일 수도 있고 일부 업무기능에 연결된 테이블 단위 또는 테이블의 행과 열에 저장된 데이터요소일 수도 있다. 왕호성, 설문원(2017)은 이렇게 기록 처분의 단위가 다양해지면 기록관리 행위를 설계하기 어려워지며, 질서정연한 상태의 데이터베이스를 기록선별 단위에서 분리하는 것은 기록의 무결성을 해칠 우려가 있음을 밝혔다. 오세라, 이해영(2019)은 이러한 위험요소도 고려하여 데이터세트의 관리 단위는 확실히 정의할 수 없고 각 시스템과 데이터세트의 특성에 따라 정해야 한다고 주장하였다.

관리기준표는 시스템을 중심으로 하여 그에 포함된 기록관리 영역의 단위기능을 추가적으로 기술하는 형태로 설계되었다. 예를 들어 인사관리시스템을 대상으로 기술한 관리기준표에는 인사, 교육, 복무, 근태관리, 행정지원 인력관리 등 여러 개의 업무정보와 단위기능을 기술하는 구조이다. 시스템의 단위기능이 여러 개일 경우, 하나의 기준표를 작성하거나 단위기능별로 작성도 가능하다(이주광, 2020, p. 19).

공공데이터 관리는 2013년부터 지금까지 각급 기관에서 공공데이터 목록을 작성하여 등록하는 식으로 진행된다. 해당 데이터의 등록 및 관리는 데이터 소유권을 가진 해당 기관에서 직접 수행한다. 공공데이터를 생성·수집·취득하는 기관은 모든 공공데이터에 대한 목록 등록서(이하 등록서)를 작성하고 15일 이내에 공공데이터포털의 공공데이터목록등록관리시스템에 등록한다. 이 중 제공대상이 될 공공데이터는 제공대상 공공데이터 등록서(이하 제공대상 등록서)를 따로 작성하여야 하며, 제공대상으로 결정된 공공데이터는 등록서와 제공대상 등록서를 모두 작성한다. <표 5>는 두 등록서의 근거, 대상과 기술요소를 비교한 것이다.

<표 5> 공공데이터 목록 등록서, 제공대상 공공데이터 등록서 기술요소

| 구분 | 공공데이터 목록 등록서 | 제공대상 공공데이터 등록서 |
|--------------|---|---|
| 근거 | 공공데이터법 시행규칙 제2조 | 공공데이터법 시행규칙 제4조 |
| 대상 | 기관에서 생산, 수집, 취득하는 모든 공공데이터 | 제공대상 공공데이터 |
| 항목 (기술요소) | 공공데이터 목록 명칭(국문/영문), 정부기능 분류 체계(BRM), 보유 근거(법령), 키워드(3개), 공공데이터 설명, 제공대상 여부, 향후 제공, 제공신청에 의한 등록 여부 | 명칭, 설명, 공공저작물 여부, 제3자 권리 포함 유무, 이용허락범위, 업데이트 주기, 차기 등록 예정일, 비용부과 유무, 비용 부과 기준 및 단위, 제공 형태, 위치(URL), 파일명, 매체 유형, 매체 유형별 건수, 언어, 제공신청에 의한 등록, 공공데이터 개방 표준, 기타 이용 유의사항 |

여기서 확인할 수 있는 사항은 데이터 이용과 제공에 관련한 항목을 중심으로 구성되었다는 점과 정부기능분류 체계(BRM)를 표기하도록 한 점이다. 우선 모든 공공데이터에 대해서 작성하는 등록서는 9개의 항목으로 구성되어 있으며 그 중 제공대상 여부, 향후 제공, 제공신청에 의한 등록 여부 등은 기입형이 아닌 선택지를 제공하였다. 제공대상 등록서는 등록서에 비해 항목이 많으나 이 역시 대체로 택일형 문항을 배치하였다.

공공데이터 분류는 정부기능분류체계를 따르며, 정확하게는 정부기능분류의 단위과제 수준에서 선정토록 한다. 별도의 분류체계가 없는 기관은 소관 부처의 분류체계에 기관 직제 규정 등을 참조하여 소기능 단위로 단위주제를 선정할 수 있다. 공공데이터법 시행령에서 정한 등록서 서식에서는 공공데이터 분류체계의 중분류표를 붙임으로 제공하여 해당하는 분류를 선택하도록 한다. 등록서에 입력한 분류체계는 공공데이터포털에 그대로 반영되

어 검색점으로 기능하며, 2020년부터 제공하는 국가데이터맵 서비스를 구축하는 기반으로 활용한다.

공공데이터를 생산, 관리하는 시스템에 대해서는 공공기관의 데이터베이스 표준화 지침(행정안전부 고시 제 2019-20호, 2019. 3. 20.)에 따라 공공데이터베이스 메타데이터를 별도로 관리한다. 위 지침에서의 메타데이터란 “공공데이터베이스내 데이터의 체계적인 관리와 편리한 검색 및 활용을 위하여 데이터의 구조, 속성, 특성, 이력 및 용어 등이 표현된 자료”를 의미한다. <표 6>은 데이터베이스 메타데이터 영역별 표준 관리항목을 정리한 것으로, 항목에 따라 작성한 공공데이터베이스 메타데이터는 표준 관리항목에 따라 기관메타관리시스템을 통해 관리하며 중앙메타데이터관리시스템과 연계된다.

<표 6> 공공데이터베이스 메타데이터 표준 관리항목

| 영역 | 세부 항목 |
|-----------|--|
| 정보시스템 정보 | 기관명, 정보시스템명, 관련 법령(보유목적), 구축년도, 운영부서명, 담당자명, 전화번호, 이메일 |
| 데이터베이스 정보 | 논리DB명, 물리DB명, DB설명, 적용 업무, DBMS 정보, 운영체제 정보, 구축일자, 테이블수, 데이터 용량, 수집제의 사유 |
| 테이블 정보 | 테이블 소유자, 테이블 영문명, 테이블 한글명, 테이블 유형, 테이블 설명, 업무분류체계, 품질진단여부, 보존기간, 테이블 볼륨, 발생주기, 공개/비공개 여부, 개방데이터 목록 |
| 컬럼 정보 | 컬럼 영문명, 컬럼 한글명, 컬럼 설명, 데이터 타입, 데이터 길이, 데이터 포맷, Not Null 여부, PK 정보, FK 정보, 제약조건, 개인정보 여부, 암호화 여부, 공개/비공개 여부 |

3.3 분석 결과

앞서 행정정보 데이터세트와 공공데이터의 생산·획득 단계와 분류·색인작성 단계를 비교해보았다. 행정정보 데이터세트와 공공데이터 관리를 비교 분석한 결과는 <표 7>과 같다. 두 대상 모두 공공기관에서 사용하는 시스템의 데이터를 대상으로 하며 관리기준표나 등록표, 메타데이터 등 관리도구가 사람에 의해 작성·관리되는 구조임을 알 수 있었다. 또한 대상의 특성상 기록의 내용, 정보에 대한 사항 외에도 데이터, 시스템 등에 대한 기술요소가 포함되어 있었다.

<표 7> 생산·획득, 분류·색인작성 단계별 비교 분석 결과

| 구분 | 행정정보 데이터세트 | 공공데이터 |
|-----|---|--|
| 목적 | 데이터세트 보존, 장기 이용가능성 확보 | 공공데이터의 개방, 활용 증대 |
| 공통점 | <ul style="list-style-type: none"> 공공기관에서 업무상 활용하는 시스템 내 데이터(베이스)를 대상으로 함 대상에 대한 목록(관리기준표, 등록표 등) 작성 및 갱신 등을 위한 관리인력이 필요함 대상의 특성상 데이터, 시스템 등을 설명하는 항목이 포함됨 | |
| 차이점 | <ul style="list-style-type: none"> 변조나 가공없이 원상태의 데이터를 획득함 시스템의 기능과 목적(고유 업무용, 공통 행정업무)을 기준으로 관리대상을 선별함 시스템 단위로 관리, 기술함 향후 다양한 분류체계 개발 등을 위해 설명형 구조인 관리기준표를 작성함 시스템, 업무정보를 통해 단위기능을 선정함 대부분 기술요소가 자유 기술형임 데이터의 장기보존을 위한 기술요소(기록관리 정보, 부가 정보) 등이 포함됨 | <ul style="list-style-type: none"> 관리와 활용을 위해 기계가독한 형태로 변형, 민감정보 삭제 등 가공 작업 후 획득함 개인정보 등 미제공대상 정보 포함 및 대상 정보 분리 가능 여부, 저작권 등 제3자 권리 여부 등을 고려하여 선별함 데이터세트 단위로 관리, 기술함 모든 공공데이터에 대한 등록서와 제공대상 등록서를 별도로 작성함 데이터에 대한 목록(등록서)과 시스템에 대한 목록(공공데이터베이스 메타데이터 표준)을 분리하여 운영함 정부기능분류를 기반으로 한 분류체계를 사용함 기술요소 중 선택형도 다수 포함됨 데이터 이용과 관련된 기술요소가 대부분이며, 그 외 관리항목이 포함됨 |

하지만 두 대상의 관리에 있어 훨씬 많은 차이점을 확인할 수 있었다. 관리단위부터 행정정보 데이터세트는 시스템, 공공데이터는 데이터세트로 선정하였으며 이 차이는 생산·획득을 포함한 이후의 모든 단계에 영향을 미쳤다. 그리고 관리도구에 관련하여 행정정보 데이터세트 관리기준표는 설명형 구조로 시스템을 기준으로 기술하기 때문에 시스템 기능이 추가되거나 삭제된다면 매번 관리기준표 전체를 갱신하여야 한다. 또한 공공데이터 등록표나 공공데이터베이스 메타데이터는 없는 장기보존을 위한 기술요소가 포함된다. 그에 비해 공공데이터는 데이터를 기준으로 제공대상에 대한 등록표를 따로 작성하며, 기술요소의 대부분이 데이터 활용과 관련된 항목이다. 시스템에 대한 정보는 공공데이터베이스 메타데이터 표준에 따라 별도로 작성하여 관리한다.

비교표를 통해 얻은 분석결과는 다음과 같이 정리할 수 있다. 첫째 행정정보 데이터세트와 공공데이터 모두 수집, 관리도구(관리기준표, 등록서 등) 작성 및 운영이 대부분 사람에게 의해 이뤄졌다. 인력에 의존적인 방식은 작성주체의 능력과 전문성에 따라 대상의 품질이 크게 좌우될 가능성이 있다. 또한 잦은 조직개편과 개칭, 업무 신설 및 삭제 등으로 인한 변경사항을 반영하기 위해 관리도구 담당자에 의한 주기적인 확인 절차가 필요하다.

특히 행정정보 데이터세트 관리기준표의 경우, 관리기준표는 행정정보 데이터세트의 메타데이터로 활용될 것이므로 작성된 관리기준표의 기계가독 여부가 중요할 것이다. 하지만 실행방안에서 예시로 든 기록관리교육훈련 시스템의 관리기준표를 보면 표기가 다소 정형화되어 있지 않음을 알 수 있다. 예를 들어 관리기관 정보의 담당자명은 성명+직위명을 병기하였으며, 시스템 정보의 개발산출물은 구분없이 한 칸에 나열식으로 기재되었다. 비록 실제 행정정보 데이터세트 관리시스템을 공개하지 않은 시점에서 구현 형태가 어떨지 알 수 없으나 제시된 예시만으로는 기계가 이를 판독하지 못하거나 또는 작성자가 아닌 다른 사람이 보았을 때 제대로 의미가 전달되지 않는 등 문제가 발생할 수 있다.

그에 비해 공공데이터 등록서는 자유기술형 뿐만 아니라 선택형 문항도 반영하여 이와 같은 문제를 방지하고 있었다. 특히 제공대상 공공데이터 등록서는 항목 18개 중 10개가 선택형 문항이 포함되었으며, 공공저작물 여부(예, 아니오 중 택 1)나 제공대상 공공데이터 제공 형태(공공데이터 포털에서 다운로드, 기관자체에서 다운로드, 전자기록매체 저장 제공, 링크드 오픈데이터(LOD) 연계파일등록, 링크드 오픈데이터(LOD) 연계URL등록 중 택 1) 등이 해당한다.

둘째 두 대상을 관리하기 위한 목록의 구조가 달랐다. 앞서 언급했듯이 행정정보 데이터세트는 시스템 단위 관리를 구현하고 향후 분류체계 개발을 위해 설명형 기준표를 채택하였다. 일체형인 관리기준표는 하나의 시스템을 기준으로 관리기관, 업무, 법규, 데이터, 기록관리 정보 등을 기재하도록 설계되었다. 한 시스템에 여러 단위기능이 있을 경우, 단위기능별로 작성할 수 있는 재량이 있으나 시스템에 대한 변경사항에는 유연하게 대처할 수 없는 구조이다. 또한 국가기록원에서 제시한 관리기준표에는 해당 시스템에 대한 백업정책, 이력관리 등 시스템에 대한 변동사항을 설명할 수 있는 항목이 누락되었다. 시스템의 기능이 추가되거나 삭제, 일부 기능이 새로운 시스템으로 대체하는 등 공공기관 행정정보시스템에 대한 변경은 매우 흔한 상황이므로 이에 대한 보완이 필요할 것이다.

반면 공공데이터 관리는 행정정보 데이터세트와 달리 데이터(등록서)와 시스템(공공데이터베이스 메타데이터)을 구분한다. 공공데이터 목록 등록서와 제공대상 공공데이터 등록서로 공공데이터 내용 및 이용조건 등을, 공공데이터베이스 메타데이터를 통해 정보시스템 구축, 운영에 대한 사항을 관리한다. 이렇듯 행정정보 데이터세트 관리도 대상을 영역별로 분리하여 이를 식별시스템을 통해 연계할 수 있다면 빈번한 정보시스템의 변화에 유연하게 대응할 수 있을 것이다.

셋째 정부기능분류라는 기존의 분류체계를 적용한 공공데이터와 달리 행정정보 데이터세트의 분류에는 일반 문서의 철, 건 구조를 따르지 않고 단위기능이라는 업무기반 단위를 도입하였다. 공공데이터 등록서는 작성시 정부기능분류체계의 중분류표를 붙임으로 제공하면서 해당 분야를 선택하도록 유도한다. 이와 같은 맥락에서 대부분 행정기관은 정부기능분류를 기반으로 한 기록관리기준표를 사용하고 있다. 이주광(2020)은 행정정보 데이

터세트 관리에 있어 행정정보시스템의 단위업무를 고려하였을 때, 현재 공공업무 영역에서의 효율성을 보아 기능 분류체계를 응용하는 것이 바람직하나 하나의 시스템에 여러 데이터베이스가 연결되었거나 여러 단위과제에 해당하는 데이터세트가 있는 등의 사례를 고려해보아야 함을 언급하였다. 오세라, 이해영(2019)의 연구에서도 관리 단위로서의 단위기능이 분석되었으나 비획일성, 다양성 등의 특성으로 관리기준표는 시스템을 기준으로 구성하였음을 밝혔다.

행정정보 데이터세트의 단위기능은 시스템의 설계구조 등에 따라 하나 혹은 그 이상을 적용할 수 있다. 또한 선행연구에서 밝힌 바와 같이 단위기능은 데이터베이스 자체 혹은 데이터 테이블이 될 수 있다. 실행방안에서 예시로 든 '기록관리교육훈련시스템 행정정보데이터세트 관리기준표'에는 교육운영 관리, 사이버교육, 성적처리, 설문평가, 강사관리, 홈페이지 관리, 교육참고 데이터베이스 7개의 업무정보와 기록관리 교육이라는 1개의 단위기능으로 구성되었다. 7개의 업무가 모두 기록관리 교육이라는 업무를 구성하는 트랜잭션으로 간주할 수도 있지만 작성자 재량에 따라 각 업무정보별로 단위기능을 세분화할 수 있다는 이야기이다.

가변적이고 유연한 단위기능의 개념적 특성은 다양한 시스템 및 행정정보 데이터세트를 수용할 수 있겠으나 현재까지 명확한 기준이 없어 제대로 분류기준으로 기능할지 알 수 없다. 이 연구에서 살펴본 단위기능이란 개념은 실질적으로 활용, 도입이 어려운 관리 단위임을 알 수 있다. 정해진 단위기능에 따라 보존기간이 책정되기 때문에 단위기능 설정은 향후 데이터세트 평가를 통한 폐기, 이관 결정 등에 중추적인 요소이다. 그렇기 때문에 행정정보시스템이 공공기관의 업무수행을 위한 도구이며, 그 기능 역시 정부기능의 범위에서 크게 벗어나지 않는다면 공공데이터와 같이 정부기능분류체계를 참조하는 방법도 고려해볼 수 있다.

마지막으로 기술요소와 관련하여 행정정보 데이터세트 관리기준표와 공공데이터 등록기준표, 공공데이터베이스 메타데이터 표준간 동일하거나 유사한 기술요소가 있었다. 관리기준표의 시스템정보에 해당하는 시스템명, 구축연도 등은 공공데이터베이스 메타데이터 정보시스템 정보의 항목과 매핑될 수 있다. 또한 실행방안의 관리기준표에서 제외되었으나 HW관리, SW관리, 데이터 해석 의존성과 같이 전문적인 지식을 요하는 기술요소는 전문 지식이 없다면 작성이 어렵다. 이러한 상황을 최대한 고려하여 기존 작성된 데이터와의 상호 연계 체계 기반을 마련한다면 전문성을 확보하면서 동시에 노동력을 절감할 뿐만 아니라 표기방식의 통일로 상호호환성도 확보할 수 있을 것이다.

이 장에서는 행정정보 데이터세트 및 공공데이터 관리 중 생산·획득, 분류·색인작성 단계에 해당하는 절차를 비교해보았다. 분석 결과 행정정보 데이터세트와 공공데이터는 모두 데이터세트를 대상으로 하지만 대상을 선정하고 획득하는 과정부터 속성을 기술하는 과정까지 차이가 있음을 알 수 있었다. 다음 장에서는 분석 결과를 바탕으로 하여 향후 행정정보 데이터세트 관리에 대한 개선방안을 제시하고자 한다.

4. 행정정보 데이터세트 관리 개선방안

4.1 단위기능 책정시 정부기능분류 참조

먼저 행정정보 데이터세트의 단위기능에 대한 객관적인 기준이 필요하다. 실제 기록관에서 행정정보 데이터세트 관리를 진행하면서 대상 선별부터 작성주체의 주관적인 판단이 많이 반영될 여지가 있다. 특히 데이터세트의 관리 단위로 제안한 단위기능은 데이터베이스 그 자체가 될 수도 있고 특정 데이터의 테이블이 될 수도 있을 정도로 가변적이므로, 단위기능에 대한 기준은 기관별로 차이가 클 것으로 예상된다.

단위기능 설정 및 운영에 대한 명확한 지침이 나오지 않은 시점에서 업무와 시스템 분석을 통해 도출된 단위기

능의 적절성은 확인하기 어렵다. 예를 들어 전자인사관리시스템에는 인사, 복무, 급여, 행정지원인력관리 등 다양한 기능이 탑재되어 많은 업무를 지원한다. 만약 이 시스템에 대한 관리기준표를 작성한다면 각 업무를 한 단위기능으로 지정할 수도 있고 단순히 인사관리라는 하나의 단위기능으로 묶어서 지정할 수도 있다. 이로 인해 유사한 대상에 대한 서로 다른 단위기능을 기재하는 상황이 발생할 수도 있다. 따라서 단위기능은 데이터세트의 보존기간 측정, 평가, 처분 등 기록관리 행위를 위한 핵심적인 정보이기 때문에 더 세분화되고 정밀한 기준이 제시되어야 할 것이다.

이와 달리 공공데이터 사업에서는 정부기능분류를 활용하여 공공데이터를 위한 분류체계를 고정시켰다. 마치 도서관에서 사용하는 십진분류법과 같이 고정적인 체계에 대상을 끼워 맞추는 식으로 운영한다. 그리고 공공기관의 기록관에서는 정부기능분류를 기반으로 한 기록관리기준표를 기록관리의 기준으로 삼고 있다. 기록관리기준표가 기록 분류에 적합하지 않은 체계라는 등 그로 인한 문제점이 많이 제기되었으나, 이를 해결하기 위한 국가기록원의 업무동향 등을 살펴보면 현행 기록관리기준표 체계를 유지하는 식으로 진행됨을 알 수 있었다(국가기록원, 2019).

기록관리기준표 운영 등으로 드러난 문제점을 감안하여 공공데이터와 같이 대기능-중기능 단위를 참조하여 기능, 업무 등에 대한 최소한의 정보를 제공한다면 단위기능 측정의 일정한 기준점으로 기능할 수 있을 것이다. 또한 대상 선별 및 관리기준표 작성 등을 지원할 수 있는 갈피를 제공하는 역할도 기대할 수 있다.

또한 관공서에서 활용하는 분류체계나 용어가 어렵고 이용자친화적이지 않다는 지적은 예전부터 있어왔으나, 공공데이터포털을 비롯하여 국가기록원 사이트에서 국경주제별 검색을 지원하는 등 이 분류체계에서 벗어나지 않는 선에서 이뤄지고 있다. 향후 행정정보 데이터세트의 보존과 함께 이용도 고려한다면 현 분류체계 반영에 대한 고려는 반드시 필요하다.

4.2 행정정보 데이터세트 관리기준표 재구성

분석 결과를 통해 국가기록원(2020)에서 제시한 행정정보 데이터세트 관리기준표에 대한 몇가지 문제점을 알 수 있었다. 우선 행정정보 데이터세트 대상 선정과 관리기준표 작성은 수작업으로 진행된다는 사실이었다. 하지만 사람에게 의한 수작업은 시간과 비용이 많이 들며, 실제로 많은 작업량을 소화할 수 없으며 그 결과물 역시 작성주체의 숙련도에 따라 품질이 고르지 않다는 것은 무시할 수 없는 단점이다.

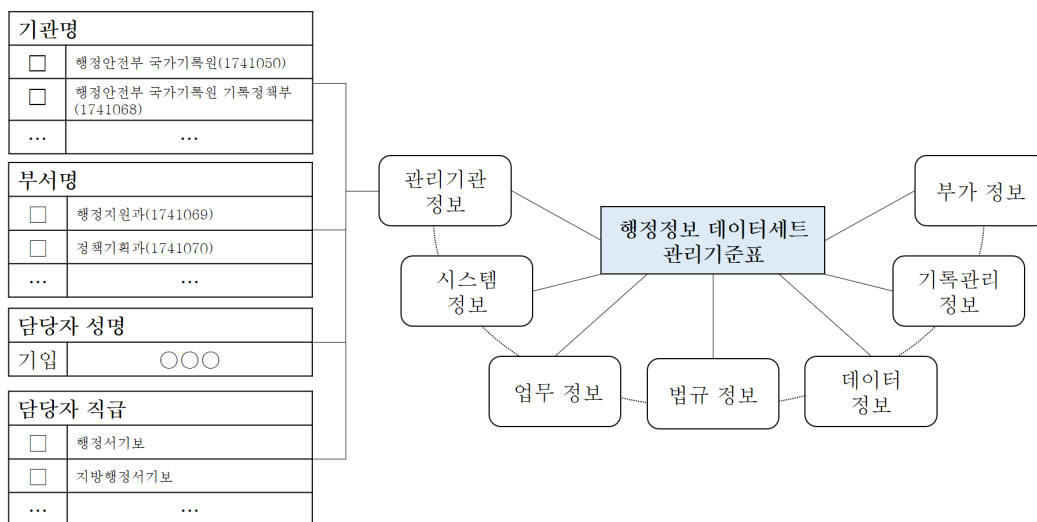
다음으로 모든 기술요소가 서술형이며 기술방식을 규정한 별도의 스키마는 없다는 점이였다. 단순한 담당자명 작성에 있어서도 성명, 성명과 직위명 혹은 직급명을 함께 기재하는지 등에 대한 지침이 없는 상태에서 현장에서 관리기준표를 작성할 시 많은 혼란을 야기시킬 것으로 예상된다.

마지막으로 시스템 단위를 중심으로 하는 설명형 구조의 관리기준표는 시스템의 변경사항을 반영하기에는 무리가 있다는 점이다. 공공기관에서 사용하는 시스템은 단순한 오류해결, 기능개선으로부터 데이터베이스 교체, 대체 시스템 개발 등 크고 작은 변화가 빈번하다. 시스템뿐만 아니라 담당자나 운영부서 변경, 유관 법령 개정 등 수시로 변경될 수 있는 요인은 매우 많다. 이런 상황을 고려해볼 때 시스템 정보와 관리기관, 업무, 법규, 데이터, 기록관리 정보 등이 한 틀에 구성된 관리기준표는 매우 고정적이며 유연하지 못하다.

위 문제점을 보완하기 위해 우선 공공데이터 사업과 같이 관리기준표 작성시 미리 정해진 값 중에서 작성주체가 선택하여 입력할 수 있는 항목을 최대한 반영할 필요가 있다. 직접 값을 입력하는 대신 시스템에서 미리 정한 적합한 입력값을 선택하는 방식은 관리기준표 작성시 혼란을 줄이고 데이터 품질을 유지할 수 있는 방안이 될 것이다.

또한 공공데이터 관리체계가 데이터와 시스템으로 구분된 점을 착안하여 행정정보 데이터세트 관리도 시스

템, 관리기관, 업무 등 각 영역을 분리하여 관리하되 특정 식별자로 연계하는 개체형 구조를 도입하여야 한다 (<그림 4> 참조). 시스템을 비롯한 관리기관, 업무, 법규 등 각 영역을 따로 관리함으로써 관리기준표에 반영해야 할 여러 상황에 보다 유연하게 대처할 수 있을 것이다.



<그림 4> 행정정보 데이터세트 관리기준표 수정(안)

4.3 유관 시스템과의 데이터 연계

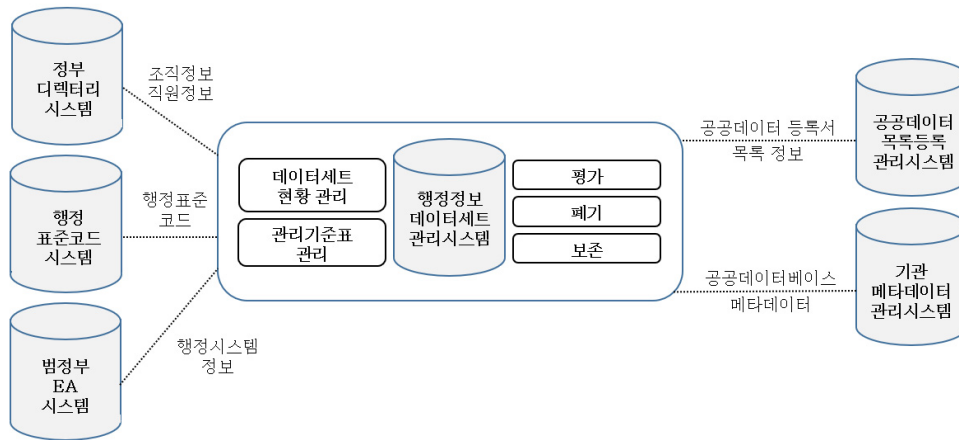
현행 법률상 행정정보 데이터세트는 공공데이터의 한 유형이며, 따라서 모든 행정정보 데이터세트는 공공데이터 등록서 작성대상이라 할 수 있다. 각 대상의 관리도구(관리기준표, 등록서, 메타데이터 등)를 살펴보면 기술요소 일부가 중복된다는 사실도 확인하였다. 업무를 수행하는 기관의 입장에서는 행정정보 데이터세트와 공공데이터 관리가 동일한 사업으로 비춰질 수 있으며, 실제로 발생할 중복작업은 행정력 낭비를 초래한다.

그리고 현재 기록관은 기록관리 외 시스템, 전산 등 특정 분야의 전문인력이 매우 부족한 상황이다. 통합기록관 체제 도입 등의 노력에도 이러한 상황은 단기간에 해결하기 어려운 사안이다. 전자기록이 주류를 이루면서 기록관 현장에서도 기록관리인력의 전산 역량에 대한 필요성이 높아지고 있으나 녹록치 않은 상황이라 할 수 있다. 이러한 상황에서 행정정보 데이터세트에 대한 전문적인 관리는 다소 기대하기 어려울 것이다.

이러한 상황을 보완하기 위해 이 연구에서는 행정정보 데이터세트 관리기준표 작성시 공공데이터의 정보(등록표, 공공데이터베이스 메타데이터)를 연계하는 방안을 제안한다. 공공데이터 사업은 2013년부터 추진되어 현재까지 많은 성과를 얻고 성공적인 행보를 걷는 사업이다. 이제 발을 떼기 시작한 행정정보 데이터세트 관리와 비교하였을 때, 이미 기반이 마련되어 높은 단계의 개발이 진행된 선행사업이라 할 수 있다. 공공데이터 목록과 공공데이터베이스 메타데이터의 품질도 역시 데이터 활용과 직결되는 사항이므로 매년 주관 부처에 의해 철저히 관리되고 있다.

더 나아가 공공데이터를 비롯한 빅데이터 등 유관 사업에서 사용하는 시스템과 정부데이터시스템, 행정표준코드시스템, 범정부EA시스템 등 기존 시스템을 활용한 데이터 연계모델 마련을 제안한다. 아래의 <그림 5>는 데이터세트 관리를 위한 데이터 연계를 표현한 것이다. 공공데이터를 비롯하여 각 시스템의 기능과 목적이 상이하므로 보유한 데이터 유형과 그 값 역시 차이가 있다는 점을 감안할 때, <그림 5>와 같은 연계 구상도를 실제로 구현한다면 식별자 부여, 메타데이터 매핑 테이블 작성 등의 작업이 필요하다. 하지만 유관 시스템과의 데이터

연계는 결과적으로 행정정보 데이터세트 관리로 소모되는 인력 낭비를 줄이고, 데이터 품질도 확보할 수 있다.



〈그림 5〉 행정정보 데이터세트 관리를 위한 데이터 연계

5. 결론

지금까지 행정정보 데이터세트에 대한 연구는 꾸준히 있어 왔지만 실무에 도입하는 시점은 요원했다. 중앙기록물관리기관인 국가기록원도 행정정보 데이터세트를 관리하기 위하여 연구사업을 진행하였지만 실제 현장에서의 적용까지 이어지지 못했다. 그러던 중 2020년에 발표한 행정정보 데이터세트 실행방안은 실제 기록관 업무에도 많은 변화를 불러일으킬 수 있는 사업 중 하나였다.

하지만 국가기록원이 제시한 행정정보 데이터세트 방안은 이제껏 몇 차례의 발표를 통해 공개되었으나 수립 과정에서 현직 기록관리 담당자의 견해가 반영되지 않았다는 의견도 있었다. 여러 해를 거쳐 심사숙고한 끝에 실행방안이 고안되었으나 기록관 현장의 입장은 다소 부정적이었다. 특히 국가기록원에서 2020년 9월 실시간으로 진행한 2020 행정정보 데이터세트 기록관리 실행방안 설명회 후 국가기록원 홈페이지 내 기록인광장에서의 반응을 통해 그 사실을 충분히 확인할 수 있었다. 이러한 상황에서 현재 공공기록물 관리와의 괴리감을 줄이고, 데이터세트에 대한 기록관리 체계를 정착시키는 것은 매우 중요한 숙제일 것이다.

여러 갈등을 차치해둔 상태로, 2020년 3월 31일자로 행정정보 데이터세트 관리에 대한 공공기록물법이 개정되었고, 2020년 10월 16일자로 행정정보 데이터세트 기록관리 기준·관리기준표 작성 및 이관규격(NAK 35: 2020 (v.1.0)) 표준이 제정되었다. 모든 문제를 해결하고 진행하기에는 이미 행정정보 데이터세트 관리의 필요성이 높아졌고, 법적 기반은 마련되어 가는 상태이므로 더 이상 늦출 수 없는 것도 사실이다. 이 연구에서는 이러한 상황에서 오래전부터 진행되어 왔던 공공데이터 사업에 관심을 두게 되었다. 2013년부터 시작된 공공데이터 개방 사업은 통상적인 업무의 하나로 자리매김하여 점점 더 그 영역을 확대해가고 있다.

이 연구에서는 공공데이터 사업과의 비교를 통해 행정정보 데이터세트 관리에 대한 개선방안을 제안하고자 하였다. 비교한 결과 현행 법률상 행정정보 데이터세트는 공공데이터의 한 유형으로 공공데이터법의 관리대상을 알 수 있었다. 이 점을 염두에 두고 관리단계별로 비교·분석한 결과 행정정보 데이터세트 관리대상 선정 및 관리기준표 작성이 오로지 작성주체의 역량에 의존적인 형태이며 설명형인 관리기준표는 변경사항을 반영하기에 취약한 구조임을 확인하였다. 또한 정부기능분류를 적용한 공공데이터와 달리 행정정보 데이터세트는 가변적일 수도 있는 단위기능이라는 개념을 도입하였으나 두 대상에 대한 기술요소는 일부 중복된다는 사실을 알

수 있었다.

분석결과를 토대로 단위기능 선정시 공공데이터와 같이 정부기능분류를 참조하며, 관리기준표 작성시 선택형 문항을 늘리고 시스템 변경 등에 보다 유연하게 대응하기 위해 영역별로 구조를 분리할 것과 공공데이터목록등록 시스템과 같은 유관 시스템과의 데이터 연계를 제안하였다. 이 연구에서 제시된 방안은 관리기준표의 데이터 품질 유지를 비롯하여 행정정보 데이터세트 관리단계에서 나타난 문제점을 개선할 수 있을 것으로 기대된다.

이 연구는 행정정보 데이터세트의 유관 법령 중에서도 공공데이터법이라는 특정 법만을 비교대상으로 삼아 개선방안을 제시하였으며, 관리단계 중 생산·획득, 분류·색인작성 단계만 다루었다는 한계가 있다. 앞으로 실제 각 기록관에서 행정정보 데이터세트 관리에 착수하면서 유관 법령과의 비교, 각 단계별 혹은 특정 대상에 대한 세분화된 연구의 필요성은 더 증가할 것이다. 그런 의미에서 이 연구는 행정정보 데이터세트 관리에 대한 기초자료로서 기록관에서의 행정정보 데이터세트 관리 제도 정착에 기여할 것이다. 이제 우리는 오랜 시간 관리하지 못했던 행정정보 데이터세트 관리를 시작하며, 앞으로 실리적인 연구를 통해 보다 나은 기록관리 방법을 찾아야 할 것이다.

참고문헌

- 강재희 (2014). 데이터세트 아카이빙 기반 공공데이터 개방체계 연구. 석사학위논문. 명지대학교 기록과학대학원, 기록관리학과.
- 국가기록원 (2019). 공공기록물 분류 및 평가제도의 재설계. 제11회 전국기록인대회 발표자료집, 73-76.
- 국가기록원 (2020). 2020년도 기록관리 중점 추진사항. 대전: 국가기록원.
- 국가기록원 전자기록관리과 (2019). 행정정보데이터세트 기록관리 방안. 2019년 제4차 기록관리 연구세미나 발표자료.
- 김남경 (2012). 공공정보 데이터세트의 아카이빙 방안 연구. 석사학위논문. 명지대학교 기록과학대학원, 기록관리학과.
- 김대기, 주원균, 김은진, 이용호 (2014) 공공데이터 유형화를 위한 분류체계 설계에 관한 사례 연구. 디지털융복합연구, 12(4), 51-68. <http://dx.doi.org/10.14400/JDC.2014.12.4.51>
- 김유승 (2014). 기록으로의 공공데이터 관리를 위한 제도적 고찰. 한국기록관리학회지, 14(1), 53-73. <http://dx.doi.org/10.14404/JKSARM.2014.14.1.053>
- 김포옥, 윤수영 (2007) 데이터세트 기록물의 기술요소에 관한 연구. 한국비블리아학회지, 18(2), 39-59.
- 설문원 (2017). 기록관리 원칙의 해석과 적용에 관한 담론 분석: 출처주의를 중심으로. 기록학연구, (52), 59-117.
- 안대진, 이해영 (2013). 지방자치단체 데이터세트의 서비스 방안 연구. 서울 열린 데이터 광장 서비스를 중심으로. 한국기록관리학회지, 13(2), 149-178.
- 오세라, 박승훈, 임진희 (2018). 행정정보 데이터세트 사례 조사 연구. 한국기록관리학회지, 18(2), 109-133. <http://dx.doi.org/10.14404/JKSARM.2018.18.2.109>
- 오세라, 이해영 (2019). 행정정보 데이터세트의 기록관리 방안. 한국기록관리학회지, 19(2), 51-76. <http://dx.doi.org/10.14404/JKSARM.2019.19.2.051>
- 왕호성, 설문원 (2017). 행정정보 데이터세트 기록의 실행방안. 한국기록관리학회지, 17(3), 23-47. <http://dx.doi.org/10.14404/JKSARM.2017.17.3.023>
- 윤성호, 유신성, 최기쁨, 오효정 (2020). 공공기관 행정정보시스템 관련 단위과제 및 보존기간 책정 현황분석. 한국비블리아학회지, 31(2), 139-160.
- 이규철 (2016). 행정정보시스템 데이터세트의 이해와 기록관리 고려사항. 기록인, 37, 56-61.
- 이주광 (2020). 행정정보 데이터세트의 기록관리 방안: 데이터세트 형태의 전자기록에 대한 기록관리 실행 방안. 국가기록원 이슈페이퍼, 23, 3-26.
- 임진희, 조은희 (2010). 행정정보 데이터세트 기록 이관 시 데이터 보정 및 품질 개선 방법 연구. 기록학연구, (25), 91-129.

조은희, 임진희 (2009). 행정정보 데이터세트 기록의 선별 기준 및 절차 연구. 기록학연구, (19), 251-291.

행정안전부 (2019). 공공기관 데이터 자원 보존 개선방안 연구, 서울: 행정안전부.

현문수 (2005). 데이터세트 기록의 실행방안. 한국기록관리학회지, 5(2), 103-124.

Jinfang Niu (2016). Organisation and description of datasets, Archives and Manuscripts, 44(2), 73-85.

<https://doi.org/10.1080/01576895.2016.1179585>

공공데이터의 제공 및 이용 활성화에 관한 법률(2017. 7. 26.)

공공데이터 관리지침(행정안전부 고시 제2019-71호, 2019. 9. 3.)

공공데이터 개방 표준(행정안전부 고시 제2020-5호, 2020. 2. 4.)

공공기관의 데이터베이스 표준화 지침(행정안전부 고시 제2019-20호, 2019. 3. 20.)

[웹사이트]

행정안전부 (2020). 공공데이터포털. 검색일자: 2020. 4. 30. www.data.go.kr

행정안전부 (2020). 행정안전부 홈페이지. 검색일자: 2020. 4. 30. <https://www.mois.go.kr/frt/sub/a06/b02/openData/screen.do>

• 국문 참고자료의 영어 표기

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

An, Dae-Jin & Rieh, Hae-young (2013). A study on the services of data-sets in the local government: Based on the cases of seoul open data portal services. Journal of Korean Society of Archives and Records Management, 13(2), 149-178.

Cho, Eun Hee & Yim, Jin-hee (2009). A study on record selection strategy and procedure in dataset for administrative information. The Korean Journal of Archival Studies, (19), 251-291.

Hyun, Moonsoo (2005). A Study on the Management of Dataset as Records. Journal of the Korean Association of Records Management, 5(2), 103-124.

Kang, Sae Hee (2014). A Study on the public data open system based on dataset archiving. Unpublished master's thesis, The Graduate School of Records, Archives and Information Science, Myongji University, Seoul, Korea.

Kim, Dae-Gi, Joo, Won Kyun, Kim, Eunjin & Lee, Young Ho (2014). A Case Study on Classification System Design for Public Sector Information Typology. Journal of Digital Convergence, 12(4), 51-68.

<http://dx.doi.org/10.14400/JDC.2014.12.4.51>

Kim, Nam Kyung (2012). A study on archiving plans for dataset of the public sector information. Unpublished master's thesis, The Graduate School of Records, Archives and Information Science, Myongji University, Seoul, Korea.

Kim, Po-Ok, Yun, Soo-Young (2007). A Study on the Description of Archival Datasets. Journal of the Korean Biblia Society for Library and Information Science, 18(2), 39-59.

Kim, You seung (2014). A Study on Legal Issues of Public Data Management as Records: Focused on Analysis of the Act on Provision and Use of Public Data. Journal of Korean Society of Archives and Records Management, 14(1), 53-73.

<http://dx.doi.org/10.14404/JKSARM.2014.14.1.053>

Lee, Ju-Kwang (2020). A study on the Records Management of Dataset in Government Information Systems: Execution of management for electronic records in the form of datasets. The National Archives of Korea Issuepaper, 23, 3-26.

Lee, Kyu-Chul (2016). Understanding dataset records in information system and suggestions for the management. GirokIN, 37, 56-61.

Ministry of the Interior and Safety (2019). A Study on the Improvement of Data Resource Conservation in Public Institutions, Seoul: Ministry of the Interior and Safety.

National Archives of Korea (2019). Redesigning the Classification and Evaluation System of Public Records. Proceeding of

- 11th National Archivist Conference, 73-76.
- National Archives of Korea (2019). Reports on records management of administrative datasets.. Presentation materials for the 4th Annual Research Seminar on Records Management in 2019.
- National Archives of Korea (2020). 2020 Records Management Majority Proposal. Daejeon: National Archives of Korea
- Oh, She-La & Rieh, Hae-young (2019). Managing Data Set in Administrative Information Systems as Records. Journal of Korean Society of Archives and Records Management, 19(2), 51-76. <http://dx.doi.org/10.14404/JKSARM.2019.19.2.051>
- Oh, She-La, Park, Seunghoon, & Yim, Jin-Hee (2018). A case study of dataset records in information management system. Journal of Korean Society of Archives and Records Management, 18(2), 109-133. <http://dx.doi.org/10.14404/JKSARM.2018.18.2.109>
- Seol, Moon-Won (2017). An analysis of discourses on interpreting and applying the principle of provenance in archival organization. The Korean Journal of Archival Studies, 52, 59-117.
- Wang, Ho-sung & Seol, Moon-Won (2017). A study on managing dataset records in government information system. Journal of Korean Society of Archives and Records Management, 17(3), 23-47. <http://dx.doi.org/10.14404/JKSARM.2017.17.3.023>
- Yim, Jin-hee & Cho, Eun Hee (2010). A study on data adjustment and quality enhancement method for public administrative dataset records in the transfer process-based on the experiences of datawarehouses' ETT. The Korean Journal of Archival Studies, (25), 91-129.
- Yoon, Sung-Ho, Yu, Sin-Seong, Choi, Kippeum, & Oh, Hyo-Jung (2020). Current Status Analysis of Business Units and Retention Period Estimation related to Administrative Information Systems of Public Institutions. Journal of the Korean Biblia Society for Library and Information Science, 31(2), 139-160.

「ACT ON PROMOTION OF THE PROVISION AND USE OF PUBLIC DATA」.

「GUIDELINE FOR MANAGEMENT OF PUBLIC DATA」.

「STANDARDS FOR OPENING PUBLIC DATA」.

「GUIDELINE FOR STANDARDIZATION OF DATABASE OF PUBLIC INSTITUTIONS」.