

행정정보 데이터세트 종합관리시스템의 서비스 방안 연구

A Study on the Service of the Integrated Administrative Information Dataset Management System

김지혜(Ji-Hye Kim)¹, 윤성호(Sung-Ho Yoon)², 양동민(Dongmin Yang)³

E-mail: kimjh129056@gmail.com, efmarchive@gmail.com, dmyang@jbu.ac.kr



¹ 제 1저자 전북대학교 기록관리학과 석사과정
² 은평구시설관리공단 기록물관리 전문요원
³ 교신저자 전북대학교 기록관리학과 부교수, 문화융복합아카이빙 연구소 공동연구원

논문접수 2022.04.18
최초심사 2022.04.21
게재확정 2022.05.10

ORCID

Ji-Hye Kim
https://orcid.org/0000-0002-0045-3670
Sung-Ho Yoon
https://orcid.org/0000-0001-6197-5336
Dongmin Yang
https://orcid.org/0000-0002-4029-9372

© 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.

- 본 연구는 '2021년 행정안전부 국가기록원 기록관리 연구개발사업'의 연구비를 지원 받아 수행되었음.
- 본 연구는 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (과제번호: NRF-2019S1A5B8099507).
- 이 논문은 2021년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음.

초 록

2020년 「공공기록물에 관한 법률 시행령」 개정에 따라 행정정보 데이터세트 기록관리 방안이 법제화되며, 국가기록원은 행정정보 데이터세트 기록관리 업무를 지원하기 위해 행정정보 데이터세트 종합관리시스템을 구축할 계획을 밝혔다. 하지만 데이터세트와 관리 기준표의 특성을 고려한 구체적인 서비스 방안은 부재한 작금이다. 이에 본 논문은 국내·외 공공데이터 포털 및 기록관 웹사이트 14곳을 대상으로 데이터세트 서비스 현황을 비교 분석하고 시사점을 도출하여 행정정보 데이터세트 종합관리시스템에 적용 가능한 서비스 방안 6가지를 제안했다. 본 연구의 결과가 행정정보 데이터세트 활용 및 서비스 활성화로 이어지기를 기대한다.

ABSTRACT

According to the amendment of the Enforcement Decree of the Public Records Management Act in 2020, an administrative information dataset record management plan will be enacted, and the National Archives of Korea plans to establish an integrated administrative information dataset management system to support it. However, there is no specific service plan that considers the characteristics of the datasets and the Management Reference Table. Therefore, this paper compared and analyzed the current status of dataset services at 14 domestic and foreign public data portals and archives websites, derived implications, and proposed 6 service plans applicable to the integrated administrative information dataset management system. This study's results will lead to utilizing the administrative datasets and the activation of services.

Keywords: 행정정보 데이터세트, 행정정보 데이터세트 관리기준표, 기록정보서비스, 공공데이터, 전자기록관리
Administrative Information Dataset, Management Reference Table for Administrative Information Dataset, Archival Information Service, Open Data, Electronic Records Management

https://jksarm.koar.kr

1. 서론

1.1 연구배경 및 필요성

행정정보 데이터세트는 두 번에 걸친 공공기록물법 시행령 개정을 통해 전자기록의 한 유형으로 공인되어 기록관리의 영역에 편입되었다. 2007년 시행령 개정으로 행정정보시스템이 전자기록생산시스템의 유형에 포함되었다. 이어 2010년 시행령 개정에서 행정정보 데이터세트의 정의를 신설하여 전자기록물의 범위를 확장하는 법제화가 이루어졌다. 하지만 행정정보 데이터세트 기록관리를 위한 체계의 부재로 데이터세트 상실, 행정정보 데이터세트 특성에 따른 기존 전자기록관리방안 적용의 어려움 등의 문제점이 발생하여 데이터세트의 구체적인 기록관리 방안 마련의 필요성과 시급성이 역설되었다(오세라, 이해영, 2019, 52). 국가기록원은 행정정보 데이터세트 기록관리 방안 설계를 위한 연구용역(2007)을 착수하였으며 여러 차례에 걸친 연구 끝에 2017년 12월에 행정정보 데이터세트 기록관리체계(안)을 발표하였다. 이를 계기로 행정정보 데이터세트 공공표준과 법령개정을 준비하기 시작했으며, 2020년 3월에는 행정정보 데이터세트 기록관리방안을 반영한 공공기록물법 시행령을 공포하였고, 나아가 공공표준 행정정보 데이터세트 기록관리 기준(NAK 35:2020(v1.0))을 제정하였다.

법령과 공공표준을 기반으로 행정정보 데이터세트의 기록관리가 본격적으로 시작되었고, 각급 공공기관은 행정정보 데이터세트 관리기준표를 작성 및 운영하는 기록관리업무를 이행해야 하는 법적 의무를 지니게 되었다. 국가기록원은 데이터세트 관리기준표를 등록, 검색, 열람 등의 관리와 이관대상 행정정보 데이터세트를 SIARD (Software Independent Archiving of Relational Database, 이하 SIARD) 포맷 형태로 인수하여 보존, 열람, 재현 및 관리하는 종합관리시스템의 체계 설계 및 구축 방안을 제시하였다(국가기록원, 2020b). 이는 법을 준수하기 위한 데이터세트 관리기준표 작성에 초점을 맞췄음을 알 수 있다.

데이터에 기반한 빅데이터나 인공지능과 같은 기술이 발전하고 실생활에 스며들며 따라 데이터세트의 가치가 높아지고 데이터세트 분석이 각광받고 있다. 특히 정부와 공공기관이 업무활동을 수행하면서 생산된 데이터세트는 일반 데이터세트보다 공신력 있고 신뢰성 있는 기록으로 여겨진다. 여기서 장기적으로 보존할 가치가 있는 데이터세트의 경우 증거적·행정적·역사적·연구적으로 중요한 의미를 지니므로 영구기록물관리기관으로 이관되어 보존·관리되도록 한다. 이러한 데이터세트를 국민이 자유롭게 이용할 수 있도록 다양한 형태로 개방·제공되어 새로운 데이터 가치를 창출할 수 있는 서비스를 제공해야 할 필요가 있다.

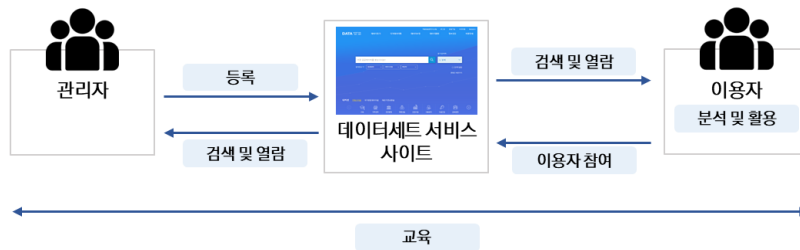
현재 행정정보 데이터세트 기록관리 업무는 시행 초기 단계이므로 데이터세트 관리기준표 작성하는 업무를 수행하면서 발생하는 시행착오나 혼선을 개선하는 데 방점을 두고 있다. 향후 행정정보 데이터세트의 기록관리가 안정화된 후 행정정보 데이터세트 종합관리시스템은 데이터세트의 가치와 특징을 극대화할 수 있는 시스템으로 확장되어야 한다. 이에 본 연구는 각급 기관의 기록물관리 담당자들이 수행해야 할 행정정보 데이터세트 기록관리 업무를 지원하고 대국민을 대상으로 행정정보 데이터세트에 접근하고 이를 활용할 수 있도록 행정정보 데이터세트 종합관리시스템에 적용 가능한 서비스 방안을 제안하고자 한다.

1.2 연구방법

본 연구는 향후 구축 예정인 행정정보 데이터세트 종합관리시스템이 행정정보 데이터세트의 기록관리 업무를 지원함과 동시에 아카이브된 행정정보 데이터세트의 서비스 방안을 제안하는 것을 목적으로 한다. 이를 위한 연구방법으로 첫째, 웹사이트에서 제공하는 데이터세트의 이동 흐름에 따른 서비스 유형 분석과 웹사이트에서 제공하는 데이터세트 서비스 현황 조사하여 분석 기준을 도출하였다(<표 1>, <그림 1> 참고). 둘째, 데이터세트 서비스 제공을 목적으로 구축 및 운영되는 웹사이트를 연구대상으로 선정하여 <표 1>을 기준으로 서비스 현황을

분석하였다. 마지막으로, 분석결과를 토대로 시사점을 도출하고, 이를 반영하여 종합관리시스템에 적용 가능한 서비스 방안을 제안하였다.

데이터세트 서비스 현황을 분석에 앞서 공유·개방에서 이용·활용까지 데이터세트의 이동 흐름을 살펴보았다. 그 결과 <그림 1>과 같이 데이터세트 서비스 제공 사이트의 서비스 대상을 관리자와 이용자로 구분하여 각 서비스 대상 그룹에 필요한 서비스 유형을 도출하였다. 관리자는 각급 기관의 데이터세트 관리 담당자로 데이터세트를 웹사이트에 등록하여 공유하는 업무를 수행한다. 본 논문에서 관리자를 행정정보 데이터세트 관리기준표를 작성하는 기관의 기록물관리 담당자로 보았다. 이들은 검색을 통해 타 기관에서 작성한 데이터세트 관리기준표를 열람 및 참조할 수 있는 서비스 기능이 필요하다. 반면 이용자는 데이터세트를 개방하고 있는 사이트에서 제공하는 검색 서비스를 통해 데이터세트에 접근하고 이를 활용한다. 관리자와 이용자 사이에 위치한 데이터세트 서비스 사이트는 데이터세트를 등록하고, 등록된 데이터세트를 검색 및 열람하고 이를 분석 및 활용할 수 있도록 지원하는 서비스가 필요함을 확인하였다. 더불어 앞선 세 가지 서비스 유형에 긍정적인 영향을 끼치는 서비스로 이용자 참여 서비스와 교육 서비스를 제시하여 데이터세트 서비스 사이트에서 제공하는 모든 서비스를 등록, 검색 및 열람, 분석 및 활용, 이용자 참여, 교육으로 총 5가지 유형으로 정리하였다.



<그림 1> 데이터세트 이동 흐름에 따른 서비스 유형

<표 2>에 제시된 웹사이트에서 제공하는 데이터세트 서비스를 분석하는데 필요한 항목을 다음 <표 1>과 같이 정리하였다. 각 유형에 해당하는 세부 항목은 웹사이트의 데이터세트 서비스 현황을 조사하며 도출하였다.

등록 서비스는 관리자가 데이터세트와 맥락정보를 함께 등록할 수 있도록 지원하는 서비스이다. 데이터세트 현행화를 진행할 때 기존에 등록한 데이터세트와 관련 정보를 편집 및 삭제할 수 있어야 하며, 유효성을 검증할 수 있는 서비스를 제공해 데이터세트의 신뢰성을 높일 수 있어야 한다.

검색 및 열람 서비스는 이용자가 찾고자 하는 데이터세트에 접근하여 열람할 수 있도록 지원하는 서비스이다. 검색의 효율성을 위해 여러 유형의 검색 방식과 검색 도구 제공과 열람을 위해 웹 브라우저에서 데이터세트 미리 보기나 다운로드가 가능해야 한다.

분석 및 활용 서비스는 데이터를 분석하고 활용할 수 있도록 필요한 정보와 관련 도구를 지원하는 서비스이다. 이용자가 데이터를 분석하는데 사용하는 소프트웨어, 분석을 위한 정보 등을 제공할 수 있는지 확인한다.

이용자 참여 서비스는 이용자의 참여를 통해 데이터세트를 재발견하고 활용가치를 촉진할 수 있도록 지원하는 서비스이다. 이용자들은 데이터세트를 분석하고 활용한 결과를 다시 웹사이트에 업로드하여 다른 이용자와 공유하며 정보를 교류하는 것을 확인할 수 있었다. 이외에도 정보 공유 목적으로 이용자들 간의 의사소통을 위한 커뮤니티와 데이터에 관련된 문의, 데이터 신청, 데이터를 평가하는 등의 이용자들이 활동하는 것을 알 수 있다.

교육 서비스는 관리자와 이용자 모두 웹사이트에서 제공하는 데이터세트 서비스를 원활히 이용하기 위해 필요한 이용 가이드를 제공하고, 관리자의 경우 데이터세트 등록 시 참고할 수 있는 등록 가이드 등이 필요하다. 또한, 데이터 분석 및 활용 능력을 향상하기 위해 교육 프로그램을 제공하는 것이 필요하다.

<표 1> 데이터세트 서비스 현황 분석 기준

서비스 유형		서비스 항목	내용	
등록	데이터세트를 등록할 수 있도록 지원하는 서비스	계정 로그인	기관의 데이터세트 등록 권한을 가진 사람만 접근할 수 있도록 승인받은 계정으로만 데이터세트 등록 기능을 제공하는지	
		데이터세트 등록	개방하려는 데이터세트 업로드가 가능한지	
		메타데이터 등록	등록하는 데이터세트를 설명할 수 있는 메타데이터를 함께 등록할 수 있는지	
		편집/삭제	데이터세트와 메타데이터의 편집과 삭제가 가능한지	
		유효성 검증	데이터세트의 형식, 구조, 품질 등 신뢰성 확보를 위한 유효성 검사를 제공하는지	
검색 및 열람	개방하고 있는 데이터세트를 검색하고 열람할 수 있도록 지원하는 서비스	검색	데이터세트 검색	통합 키워드 검색, 상세 검색 또는 카테고리 검색 등을 제공하는지
			검색 결과 필터	질의어에 따라 도출되는 검색 결과를 이용자가 원하는 특정 조건에 맞춰 재검색할 수 있는 필터를 제공하는지
			검색어 서비스	관련 검색어, 실시간 이슈 검색어 등의 서비스를 제공하는지
		열람	데이터 미리보기	웹 브라우저에서 데이터세트를 테이블 형태로 재현이 가능한지
			계정 로그인	데이터세트를 다운받기 위해 로그인할 필요가 있는지
			다운로드 프로그램 설치	데이터세트를 다운받기 위해 추가로 프로그램을 설치해야 하는지
다운로드	데이터세트 전체 파일 혹은 필요한 일부 파일을 다운받아 열람할 수 있는지			
분석 및 활용	데이터세트를 분석하고 활용할 수 있도록 정보를 제공하고 관련 도구를 지원하는 서비스	분석	분석 정보 제공	데이터세트를 분석하는데 참고할 수 있는 정보를 제공하는지
			분석 도구 제공	데이터세트를 분석하는데 사용할 수 있는 도구를 지원하는지
		시각화	시각화 정보 제공	데이터세트를 시각화하는데 참고할 수 있는 정보를 제공하는지
			시각화 도구 제공	데이터세트를 시각화하는데 사용할 수 있는 도구를 지원하는지
이용자 참여	이용자의 참여를 통해 데이터세트의 가치를 재발견하고 추가적 공개를 촉진하는 서비스	데이터세트 활용 사례 공유	이용자의 데이터세트 활용 사례 공유가 가능한지	
		데이터세트 재배포	이용자가 데이터세트를 재배포하여 새롭게 만든 데이터세트 재배포를 지원하는지	
		이용자 커뮤니티	이용자들 간 데이터세트에 관한 정보 교류(문제 해결을 위한 질문, 데이터세트 공유, 개발자 노하우 공유 등)를 할 수 있는 공간이 있는지	
		데이터세트 문의/신청	이용자가 데이터세트에 관한 문의와 개방하지 않는 데이터세트를 신청할 수 있는지	
		평가/피드백/리뷰/설문조사	개방하고 있는 데이터세트에 대해 평가하고, 리뷰를 남기고, 서비스에 대한 피드백과 설문조사에 참여할 수 있는지	
교육	이용자와 관리자의 데이터 리터러시 역량을 기르기 위해 필요한 교육 프로그램을 제공하는 서비스	데이터세트 등록 가이드	데이터세트 등록 방법과 과정에 관한 설명을 제공하는지	
		데이터세트 검색 가이드	데이터세트 검색 방법과 과정에 관한 설명을 제공하는지	
		활용 교육 영상	활용 교육을 동영상으로 제공하는지	
		활용 교육 문서	활용 교육용 문서 자료로 제공하는지	
		활용 실습 프로그램	활용 교육을 통해 얻은 기술적인 지식을 적용해 간단한 문제를 해결해 볼 수 있도록 실습 프로그램을 제공하는지	

1.3 분석대상

본 연구에서는 데이터세트 서비스 제공 현황을 분석하기 위해 다음 <표 2>와 같이 국외 국립기록원의 데이터세트를 서비스하는 포털 3개, 국외 공공데이터 포털 6개, 국내 공공데이터 포털 4개와 Internet Archive까지 총 14개의 웹사이트를 분석대상으로 선정하였다.

〈표 2〉 연구대상 웹사이트 일반사항

구분	웹사이트(웹사이트 URL)	운영주체	데이터세트 건수	개시 연도	
국외	국립 기록원 Access to Archival Database(AAD) (https://aad.archives.gov)	정부(NARA)	247,000,000	2003	
	National Digital Archive of Datasets(NDAD) (https://www.nationalarchives.gov.uk)	정부(TNA)	-	2005 (~2010)	
	-	정부(NAA)	-	-	
	공공 데이터포털	미국 공공데이터 포털 (https://data.gov)	정부	342,973	2009
		영국 공공데이터 포털 (https://data.gov.uk)	정부	51,938	2010
		EU 공공데이터 포털 (https://data.europa.eu)	정부(EU)	1,305,969	2013
		호주 공공데이터 포털 (https://data.gov.au)	정부	106,380	2013
		캐나다 공공데이터 포털 (https://open.canada.ca)	정부	28,127	2013
		스위스 공공데이터 포털 (https://opendata.swiss)	정부	6,617	2014
	Internet Archive (https://archive.org)	비영리 단체	-	1996	
	국내	공공 데이터 포털	한국 공공데이터 포털 (https://www.data.go.kr)	정부/공공기관 (행정안전부, NIA)	51,114
서울 열린데이터 광장 (https://data.seoul.go.kr)			지자체 (서울시)	7,063	2012
경기데이터드림 (https://data.seoul.go.kr)			지자체 (경기도)	1,507	2015
AI Hub (https://aihub.or.kr)			공공기관 (NIA)	191	2017

국외 국립기록원은 기록관리 선진사례로 자주 언급되는 미국 국가기록청(National Archives and Records Administration, 이하 NARA), 영국 국가기록원(The National Archives, 이하 TNA), 호주 국가기록원(National Archives of Australia, 이하 NAA)을 연구대상으로 선정하였다. 이상 세 국가의 국립기록원을 대상으로 중앙기록물관리기관으로 이관된 데이터세트를 이용자들이 접근하여 활용할 수 있는 서비스를 제공하고 있는지 확인하였다. NARA와 TNA는 각각 AAD(Access to Archival Database, 이하 AAD)와 NDAD(National Digital Archive of Datasets, 이하 NDAD)라는 데이터세트 아카이브를 운영하고 있으며, 호주 국가기록원의 경우 별도의 데이터세트 서비스를 제공하고 있지 않음을 확인하였다.

국외 공공데이터 포털은 미국, 영국, EU, 호주, 캐나다, 스위스에서 운영하는 국외 공공데이터 포털을 분석대상으로 선정하였다. 각국은 정부기관에서 생산한 데이터를 국민이 쉽게 접근 및 활용할 수 있도록 오픈 데이터 정책을 시행하였고, 이 정책의 일환으로 공공데이터 포털을 구축하여 운영하고 있다. 현재 공공데이터 포털은 시민과 기업, 정부기관에 이르기까지 모두가 데이터를 검색, 접근, 재사용할 수 있도록 장려하고 있다.

Internet Archive는 디지털 라이브러리로 웹, 도서, 비디오, 오디오, 이미지, 소프트웨어 등의 자료를 디지털화하여 보존하고 있다. 데이터세트와 관련된 서비스를 제공하고 있지 않으나 특징적으로 소프트웨어에 한정하여 애플리케이션을 제공한다. 또한 수많은 웹페이지를 지속적으로 스냅샷하여 보존하고 있어 데이터세트와 관련된 웹페이지가 아카이빙되어 이용자들이 해당 페이지에 접근 가능하다.

국내 공공데이터 포털로 정부에서 운영하는 우리나라 공공데이터포털과 지자체에서 운영하는 서울 열린데이터 광장, 경기데이터드림, NIA(한국지능정보사회진흥원)에서 운영하는 AI Hub를 분석 대상으로 선정하였다. 지자체에서 운영하는 공공데이터 포털은 시정활동 과정에서 수집된 데이터를 개방하고 있다. AI Hub는 AI 학습용 데이터 구축·확산 사업의 일환으로 AI 기술·서비스 개발에 필수적인 AI 데이터, 소프트웨어, 컴퓨팅자원 등을 제공한다.

본 연구는 웹사이트에서 제공하는 데이터세트 서비스 현황을 파악하기 위해 총 14개의 웹사이트를 조사하였다. 조사 결과를 토대로, 웹사이트에서 데이터세트 서비스를 제공하지 않거나 다른 웹사이트와 비교해 봤을 때 특정 서비스 없이 공통적인 서비스를 제공하는 웹사이트인 국외 웹사이트 3곳(호주 국가기록원, 영국 공공데이터 포털, 스위스 공공데이터 포털), 국내 웹사이트 1곳(서울 열린데이터 광장) 총 4곳의 웹사이트를 생략하였다. 호주 국가기록원은 현재 데이터세트 서비스를 제공하지 않으며 공공데이터 포털을 통해 데이터세트를 개방하고 있어 데이터세트 서비스 현황을 파악할 수 없었다. 영국 공공데이터 포털은 검색 및 열람 서비스를 제외한 다른 특정 서비스를 제공하지 않으며, 스위스 공공데이터 포털도 이와 마찬가지로 검색 및 열람, 활용 사례 공유 서비스만 제공하고 있어 다른 공공데이터 포털과 비교될만한 특정 서비스를 제공하지 않고 있다. 특히 서울 열린데이터 광장과 경기데이터드림에서 제공하는 서비스를 비교했을 때 데이터세트 검색, 통계 데이터 제공, AI학습데이터 제공, 활용사례 공유 등 공통적인 서비스를 제공하고 있다. 그러나 경기데이터드림이 분석과 활용에 관한 서비스를 추가적으로 제공하고 있어 서비스가 유사한 서울 열린데이터 광장을 분석대상에서 제외하였다. 이 4곳의 사이트를 제외하더라도 웹사이트를 비교 분석 대상으로 삼아 2장에서 분석을 진행하였으며 2.6절에서 생략한 4곳을 포함한 데이터세트 서비스 현황을 함께 제시하였다.

1.4 선행연구

행정정보 데이터세트 기록관리의 필요성을 필두로 기록관리체계 수립을 위한 선행연구가 수행되었다. 국가기록원은 2007년 ‘행정정보시스템 데이터세트 기록관리 방안 연구용역’을 시작으로 데이터세트 기록관리체계를 위한 R&D를 추진하였다. 2015년 ‘데이터세트 구조분석 및 진본성 보장 기록관리 기능모델 연구’를 진행하여 행정정보시스템 유형 분석, 데이터세트 관리단위를 단위기능으로 정의하고, SIARD를 이관포맷으로 채택하는 등의 연구결과를 도출하였다. 2017년에는 ‘차세대 전자기록관리 모델 재설계’ 연구용역 사업을 추진하여 6개의 행정정보시스템의 데이터세트 현황을 조사 및 분석하였고 분석결과를 바탕으로 관리기준표 요소를 제안하였다. 앞선 연구결과를 기반으로 국가기록원은 행정정보 데이터세트 기록관리체계(안)을 확정지었다. 특히 2020년 행정정보 데이터세트 기록관리 체계구축 사업을 진행하면서 61개 행정정보시스템 분석 및 행정정보 데이터세트 관리기준표를 작성하는 시범사업과 함께 향후 개발할 종합관리시스템 절차 및 일부 상세 기능을 설계하였다.

윤성호 외(2020)는 행정정보시스템과 관련된 단위과제를 분석하고 유형별 보존기간 책정 현황을 파악하여 유사한 단위과제 사이 상이한 보존기간 책정, 데이터세트 관리 인식 미비, 관리 방안 부재 등의 문제점과 이에 대한 개선방안은 제안하였다. 서지인(2020)은 행정정보 데이터세트를 공공데이터의 한 유형이라는 전제하에 기록관리 표준과 공공데이터법을 비교 분석하였다. 그 결과 관리기준표 작성이 작성주체의 역량에 의존적이며 설명형인 관리기준표 변경사항을 반영하기 취약한 구조임을 확인하였다. 신정엽(2021)은 관리기준표 작성 절차, 항목별 작성 사례와 관리기준표 작성 과정 중에 생기는 업무의 혼선과 고충을 해결할 수 있는 별도의 선별 정책 마련 필요성을 제안하며, 전체 데이터세트 보존 방식으로 운영, 비정형데이터에 대한 선별, 관리기준 추가 등 관리기준표 운영 개선방안을 제시하였다. 행정정보 데이터세트 기록관리 표준(NAK 35:2020(v1.0))의 제정으로 행정정보시스템의 관리대상 선정, 관리기준표 구성항목과 작성방법, 이관규격 등에 데이터세트 기록관리에 관한 전반적인

프로세스와 방안이 마련되었다. 그러나 데이터세트 기록관리 방안이 확립됐음에도 구체적인 실행방안은 정의되지 않고 관리기준표 양식만 표준화되어 있는 상황이라 기록관리 관련 전반 업무를 수행하기 어려운 실정임을 확인하였다.

초기 선행연구는 데이터세트의 특성을 반영할 수 있는 기록관리체계 구축을 위한 연구가 주를 이루었다. 기록관리 방안이 확립된 후에는 실제 데이터세트 기록관리 업무를 수행하면서 발생하는 문제점의 개선방안을 마련하는 연구가 수행되었다. 2020년 국가기록관리 중장기 발전계획에 따르면 국가기록원은 행정정보시스템 기록관리체계 적용 확산과 함께 행정정보 데이터세트 종합관리시스템 구축을 목표로 하고 있다(국가기록원, 2020c). 종합관리시스템에서 접근할 수 있는 데이터세트의 높은 보존가치를 고려했을 때 대국민을 대상으로 개방된 데이터세트를 적극적으로 활용을 장려하는 서비스가 제공될 필요가 있다. 이에 본 연구는 종합관리시스템에서 기록물관리 담당자뿐만 아니라 대국민을 대상으로 제공할 수 있는 서비스 방안을 제안하고자 한다.

2. 국내외 데이터세트 서비스 현황 분석

본 장에서는 1.2절에서 정리한 <표 1>을 기준으로 웹사이트에서 제공하는 데이터세트 서비스 현황을 분석하고자 한다. 분석대상은 국외 국가기록원(미국 NARA, 영국 TNA)에서 제공하는 데이터세트 아카이브와 국외 공공데이터 포털(미국, EU, 호주, 캐나다), Internet Archive, 국내 공공데이터 포털(우리나라, 경기데이터드림, AI Hub)로, 조사한 14개의 웹사이트 중 10개의 웹사이트에서 제공하는 데이터세트 서비스를 비교 분석하였다. 데이터세트 등록, 데이터세트 검색 및 열람, 데이터세트 분석 및 활용, 이용자 참여, 교육이라는 총 5개의 서비스 유형으로 구분하여 각 유형의 세부 항목에 따른 서비스를 제공 여부를 확인하였다.

2.1 등록

데이터세트 등록 서비스 제공 현황 분석결과는 다음 <표 3>과 같다. 등록 서비스는 등록 권한을 가진 계정으로 로그인해야지만 이용할 수 있기에 웹사이트마다 어떤 등록 서비스를 제공하는지 직접 확인할 수 없었다. 그래서 등록 가이드에서 확인할 수 있는 정보의 수준에서 조사를 진행하였고, 이마저 확인할 수 없는 웹사이트의 경우 ‘-’(해당 서비스 제공 여부 확인 불가)로 표시하였다. 서비스 현황 분석결과, 일부 웹사이트에서 제공하는 데이터세트 등록 서비스는 전반적으로 로그인-데이터파일 등록-메타데이터 등록순의 단계를 거쳐 데이터파일을 업로드할 수 있도록 한다. 등록 방식은 크게 수동 등록과 자동 등록으로 구분된다. 수동 등록은 데이터세트 파일 업로드 시 메타데이터를 직접 작성하도록 하며, 자동 등록은 소프트웨어나 자동화 도구 등을 사용하여 등록할 수 있다. 국외 국립기록원은 직접적인 등록 서비스는 제공하지 않으나 이관된 데이터세트 중 공개하고자 하는 데이터세트를 웹사이트에 업로드한다고 보아 ‘△’로 표시하였다. 우리나라 공공데이터 포털의 경우 공공데이터목록 등록관리시스템(all.data.go.kr)을 지원하며, 행정안전부에서 표준데이터셋 점검 서비스와 공공데이터 품질관리 수준 자체평가 서비스를 제공하여 각 기관의 담당자가 데이터세트 구성 시 올바른 파일명 생성, 항목구성, 데이터 등록을 안내하고 있다. 등록 서비스는 관리자를 대상으로 데이터세트를 등록하고 이를 유지관리 업무를 수행할 수 있도록 지원한다. 행정정보 데이터세트 기록관리를 위해 기록물관리 담당자는 데이터세트 관리기준표를 작성하고 변경사항이 있을 때마다 데이터세트 관리기준표를 수정해야 한다. 종합관리시스템에서 데이터세트 관리기준표를 작성하여 등록하여 관리할 수 있도록 서비스를 제공하면 데이터세트 기록관리 업무 수행을 지원할 수 있을 것이다.

<표 3> 데이터세트 등록 서비스 현황 비교표

	국외							국내		
	국립기록원		공공데이터 포털				Internet Archive	공공데이터 포털		
	NARA	TNA	미국	EU	호주	캐나다		한국	경기도	AI Hub
계정 로그인	△	△	-	○	○	-	○	○	X	X
데이터세트 등록	△	△	-	○	○	-	○	○	X	X
메타데이터 등록	△	△	-	○	○	-	○	○	X	X
편집/삭제	△	△	-	○	○	-	○	○	X	X
유효성 검증	△	△	-	X	X	-	X	○	X	X

(○: 해당 서비스 제공, △: 해당 서비스 일부 제공, X: 해당 서비스 제공하지 않음, -: 해당 서비스 제공 여부 확인 불가)

2.2 검색 및 열람

데이터세트 검색 및 열람 서비스 제공 현황 분석결과는 다음 <표 4>와 같다. 대부분의 웹사이트는 메인화면에 키워드 기반의 검색창을 위치시킨 검색형 포털로 운영되고 있다. 질의한 키워드에 따라 도출되는 검색 결과 내에서 이용자가 원하는 특정 조건에 맞춰 재검색할 수 있도록 검색 결과 필터를 제공하고 있다. 입력한 키워드와 연관 있는 주제를 검색창 하단에 추천해주는 검색어 서비스는 일부 웹사이트에서 제공하고 있다. 모든 웹사이트에서는 이용자가 데이터세트를 열람할 수 있도록 다운로드 서비스를 제공하며, 일부 사이트에서는 데이터파일을 다운로드하기 전, 일부 데이터를 확인할 수 있도록 미리보기 서비스를 제공하고 있다.

<표 4> 데이터세트 검색 및 열람 서비스 현황 비교표

	국외							국내		
	국립기록원		공공데이터 포털				Internet Archive	공공데이터 포털		
	NARA	TNA	미국	EU	호주	캐나다		한국	경기도	AI Hub
검색	데이터 검색	○	○	○	○	○	○	○	○	○
	검색 결과 필터	X	X	○	○	○	○	○	○	X
	검색어 서비스	○	X	X	X	X	X	○	○	X
열람	데이터 미리보기	○	○	△	○	○	○	○	○	△
	계정 로그인	X	X	X	X	X	X	X	X	○
	다운로드 프로그램 설치	X	X	X	X	X	X	X	X	○
	다운로드	○	○	○	○	○	○	○	○	○

(○: 해당 서비스 제공, △: 해당 서비스 일부 제공, X: 해당 서비스 제공하지 않음)

NARA의 AAD는 필드 검색을 통해 이용자가 원하는 레코드 일부만 검색하고 다운로드 받을 수 있도록 서비스 하고 있다. 호주 공공데이터 포털은 CSV 파일에 한정하여 테이블 형태와 그래프(막대형, 꺾은선, 버블형 등)로 시각화할 수 있는 미리보기를 제공하여 데이터세트의 이해를 돕는다. AI Hub에서 데이터세트를 다운로드 받아 열람하기 위해서는 계정 로그인 후 데이터세트 사용 신청 절차를 거쳐 승인을 받아야 한다. 데이터세트 다운로드를 위해 별도의 프로그램을 설치해야 하지만 데이터세트를 구성하는 여러 데이터파일 중 이용자가 필요한 파일만 선택하여 다운로드할 수 있다. 범정부EA에 따르면 2020년 기준 16,199개의 행정정보시스템이 공공기관에서 운영되고 있다. 행정정보시스템이 생산하는 데이터세트는 종합관리시스템을 통해 데이터세트 관리기준표를 작성하고 등록 시 다양한 데이터세트 관리기준표가 기록물관리 담당자와 대국민에게 공개될 것이다. 이에 이용자 의도에

적합한 검색 결과를 얻을 수 있도록 다양한 검색 접근점을 제공하고, 효과적으로 데이터세트 및 관리기준표에 접근할 수 있는 기능이 필요하다.

2.3 분석 및 활용

데이터세트 분석 및 활용에 관한 서비스 현황 분석결과는 다음 <표 5>와 같다. 일부 웹사이트에서만 데이터세트를 분석 및 활용할 수 있도록 관련 정보와 도구를 지원한다. 분석 서비스를 제공하는 웹사이트의 경우 분석 정보는 분석 리포트와 같이 미리 분석된 정보를 제공하는 수준에 그쳤고, 미국 공공데이터 포털과 AI Hub만이 분석 도구를 지원한다. 미국 공공데이터 포털은 resources.data.gov을 새롭게 편성하여 데이터 관리, 교환 및 사용을 지원하기 위한 정책, 도구, 사례 연구 등을 제공한다. AI Hub는 AI API·DATA를 통해 인공지능 기술 및 데이터를 체험하고 연구목적으로 사용할 수 있도록 제공하고 있다. 시각화 서비스를 제공하는 사이트가 분석 서비스보다 상대적으로 많았다. 시각화 정보의 경우 미국과 EU는 기관별 데이터 공개 정책 수행 현황이나 나라별 메타데이터 품질 현황과 같은 정보를 대시보드로 공개하고 있으며 그 외 캐나다, 한국, 경기도는 인포그래픽, 그래프 등 다양한 시각화 자료를 공개하고 있다. 호주는 nationalmap을 통해 데이터세트를 지도와 매쉬업해주는 서비스를, 우리나라 공공데이터 포털은 단순 그래프를 나타낼 수 있도록 시각화 도구를 웹 브라우저에서 제공하고 있다.

데이터세트 개방의 목적이 활용인 만큼 앞서 살펴본 다양한 웹사이트에서 데이터세트 분석을 위한 서비스를 제공하려는 움직임을 보이고 있다. 종합관리시스템 역시 데이터세트를 분석하여 활용할 수 있도록 필요한 도구와 환경을 마련할 필요가 있다. 데이터세트가 가지는 높은 활용도는 기존의 기록물과는 차별화된 특징이므로 종합관리시스템 이용도를 제고하기 위해선 이를 충족시키는 서비스 제공이 필수적일 것이다.

<표 5> 데이터세트 분석 및 활용 서비스 현황 비교표

		국외						Internet Archive	국내		
		국립기록원		공공데이터 포털					공공데이터 포털		
		NARA	TNA	미국	EU	호주	캐나다		한국	경기도	AI Hub
분석	분석 정보 제공	X	X	△	X	X	○	X	X	○	○
	분석 도구 제공	X	X	○	X	X	△	X	X	X	○
시각화	시각화 정보 제공	X	X	△	△	X	○	X	○	○	X
	시각화 도구 제공	X	X	○	X	○	X	X	○	X	X

(○: 해당 서비스 제공, △: 해당 서비스 일부 제공, X: 해당 서비스 제공하지 않음)

2.4 이용자 참여

이용자 참여와 관련된 서비스 현황 분석결과는 다음 <표 6>과 같다. 전반적으로 이용자가 참여할 수 있는 활동은 주로 개방하고 있는 데이터세트를 평가하고, 데이터세트의 활용 사례를 공유, 이용자-이용자, 이용자-웹사이트 사이에서 이루어지는 정보 교류 등에 한정되어 있다. 이용자들이 많이 참여하는 서비스는 활용 사례 공유와 정보와 경험, 데이터세트 활용 시 발생하는 문제점 등을 공유할 수 있는 커뮤니티 서비스이다. 우리나라 공공데이터 포털은 데이터 공유하기라는 커뮤니티를 통해 이용자가 만든 데이터세트 재배포가 이루어지고 있다. 미국은 Open Data Stack Exchange란 질의응답 사이트와 연계하여 개발자와 연구자 누구나 질문하고 질문에 답변할 수 있도록 서비스를 제공하고 있으며, 호주도 마찬가지로 누구나 질문을 올리고 답변할 수 있는 커뮤니티를 제공하고 있다. 그러나 포털 내 메뉴에 커뮤니티 서비스가 명시되어 있지 않아 접근하는데 불편함이 있다.

종합관리시스템의 이용대상은 크게 기록물관리 담당자와 대국민으로 구분할 수 있다. 기록물관리 담당자는 데이터세트 기록관리 업무 수행을 위해, 대국민은 필요한 데이터세트에 접근하고 이를 분석 및 활용하고자 종합관리시스템을 이용할 것이다. 이때 관련 업무를 수행하거나 데이터세트를 분석하는 과정에서 도출되는 문제점을 공유하고 다른 이용자로부터 솔루션을 얻을 수 있는 정보 교류 공간을 마련할 필요가 있으며, 다양한 이용자가 쉽게 접근할 수 있도록 홍보, 배너 노출 등을 고려해야 한다.

<표 6> 이용자 참여 서비스 현황 비교표

	국외							국내		
	국립기록원		공공데이터 포털				Internet Archive	공공데이터 포털		
	NARA	TNA	미국	EU	호주	캐나다		한국	경기도	AI Hub
데이터 활용 사례 공유	X	X	X	○	X	○	X	○	○	X
데이터세트 재배포	X	X	X	X	X	X	X	○	X	X
이용자 커뮤니티	X	X	○	X	○	X	○	○	X	X
데이터세트 문의/신청	△	X	○	X	○	○	X	○	○	○
평가/피드백/리뷰 /설문조사	X	X	X	X	X	○	○	○	○	X

(○: 해당 서비스 제공, △: 해당 서비스 일부 제공, X: 해당 서비스 제공하지 않음)

2.5 교육

교육 관련 서비스 제공 현황 분석결과는 <표 7>과 같다. 검색 가이드와 등록 가이드는 웹사이트의 전체 서비스 이용 가이드에 포함되어 제공되며, 등록 가이드의 경우 등록 서비스를 제공하고 있는 웹사이트에서 확인할 수 있다. EU 공공데이터 포털과 AI Hub는 데이터에 대한 지식과 활용 능력 향상에 이바지하고자 활용 교육 서비스를 제공하고 있음을 확인하였다. EU 공공데이터 포털은 data.europa academy를 운영하여 데이터 이용자와 데이터 품질관리를 수행하는 관리자를 위한 교육을 제공한다. AI Hub는 유튜브 채널을 운영하여 포털에서 제공하는 데이터세트의 이해를 돕고 활용 교육 영상과 함께 학습 자료를 제공하고 있다. 더불어 인공지능 놀이터라는 프로그램을 운영하여 수준별로 나누어진 문제를 데이터를 가지고 이용자가 집적 코딩을 해보고 결과를 도출해내는 프로그램을 경험할 수 있다.

<표 7> 교육 서비스 현황 비교표

	국외							국내		
	국립기록원		공공데이터 포털				Internet Archive	공공데이터 포털		
	NARA	TNA	미국	EU	호주	캐나다		한국	경기도	AI Hub
데이터 등록 가이드	X	X	-	○	○	-	○	△	X	X
데이터 검색 가이드	○	○	○	○	X	○	○	○	X	X
활용 교육 영상	X	X	X	○	X	X	X	X	X	○
활용 교육 문서	X	X	X	○	X	X	X	X	X	○
활용 실습 프로그램	X	X	X	X	X	X	X	X	X	○

(○: 해당 서비스 제공, △: 해당 서비스 일부 제공, X: 해당 서비스 제공하지 않음, -: 해당 서비스 제공 여부 확인 불가)

이용 가이드와 교육 서비스는 웹사이트에서 제공하는 서비스를 더욱 유용하게 활용할 수 있도록 지원하는

서비스이므로 데이터 포털과 데이터세트를 활용하는 이용자들의 만족도 제고에 도움이 될 수 있는 유형의 서비스이다. 종합관리시스템의 이용자를 위한 상세 이용가이드와 관리기준표 작성이나 데이터세트 활용을 제고하기 위한 교육 서비스를 제공하여 데이터세트 분석에 대한 진입장벽을 낮춰 누구나 데이터세트를 분석하고 활용할 수 있어야 한다.

2.6 국내외 데이터세트 서비스 현황 분석결과 비교표

본 절에서는 2.1절부터 2.5절까지의 서비스 유형에 따른 분석결과를 <표 8>과 같이 정리하였다. 등록 서비스는 국내외 공공데이터 포털을 중심으로 제공되는 서비스이며, 검색과 열람 서비스는 모든 웹사이트에서 제공되고 있었다. 그러나 분석 및 활용에 관한 서비스와 교육 서비스는 소수 웹사이트에서만 서비스되고 있었으며, 이용자 참여에 관한 서비스는 활용 사례 공유와 데이터세트 신청에 국한되어 있음을 확인했다.

<표 8> 국내외 웹사이트 데이터세트 서비스 현황 비교표

		국의 국립기록원			공공데이터 포털										Internet Archive
		N A R A	T N A	N A A	국외					국내					
					미국	영국	EU	호주	캐나다	스위스	한국	서울시	경기도	AI Hub	
등록	계정 로그인	△	△	X	-	○	○	○	-	○	○	X	X	X	○
	데이터세트 등록	△	△	X	-	○	○	○	-	○	○	X	X	X	○
	메타데이터 등록	△	△	X	-	○	○	○	-	○	○	X	X	X	○
	편집/삭제	△	△	X	-	○	○	○	-	○	○	X	X	X	○
	유효성 검증	△	△	X	-	X	X	X	-	X	○	X	X	X	X
검색 및 열람	검색	데이터세트 검색	○	○	X	○	○	○	○	○	○	○	○	○	○
		검색 결과 필터	X	X	X	○	○	○	○	○	○	○	○	X	○
		검색어 서비스	○	X	X	X	X	X	X	X	X	○	X	○	X
	열람	데이터 미리보기	○	○	X	△	○	○	○	○	○	○	○	△	○
		계정 로그인	X	X	X	X	X	X	X	X	X	X	X	X	○
		다운로드 프로그램 설치	X	X	X	X	X	X	X	X	X	X	X	○	X
		다운로드	○	○	X	○	○	○	○	○	○	○	○	○	○
분석 및 활용	분석	분석 정보 제공	X	X	X	△	X	X	X	○	X	X	X	○	X
		분석 도구 제공	X	X	X	○	X	X	X	△	X	X	X	○	X
	시각화	시각화 정보 제공	X	X	X	△	X	△	X	○	X	○	X	○	X
		시각화 도구 제공	X	X	X	○	X	X	○	X	X	○	X	X	X
이용자 참여	데이터세트 활용 사례 공유	X	X	X	X	X	○	X	○	○	○	○	○	X	
	데이터세트 재배포	X	X	X	X	X	X	X	X	X	○	X	X	X	
	이용자 커뮤니티	X	X	X	○	X	X	○	X	X	○	X	X	○	
	데이터세트 문의/신청	△	X	X	○	○	X	○	○	X	○	○	○	X	
	평가/피드백/리뷰/설문조사	X	X	X	X	X	X	X	○	X	○	○	○	X	
교육	데이터세트 등록 가이드	X	X	X	-	○	○	○	-	○	△	X	X	X	
	데이터세트 검색 가이드	○	○	X	○	X	○	X	○	X	○	X	X	○	
	활용 교육 영상	X	X	X	X	X	○	X	X	X	X	X	X	○	
	활용 교육 문서	X	X	X	X	X	○	X	X	X	X	X	X	○	
	활용 실습 프로그램	X	X	X	X	X	X	X	X	X	X	X	X	○	

(○: 해당 서비스 제공, △: 해당 서비스 일부 제공, X: 해당 서비스 제공하지 않음, -: 해당 서비스 제공 여부 확인 불가)

3. 시사점

앞서 데이터세트 활용을 취지로 구축된 국내외 웹사이트에서 제공하는 데이터세트 서비스 현황을 분석하였다. 그 결과 데이터세트 개방과 함께 이용자들의 데이터세트 활용 편의를 증진하기 위한 서비스를 함께 고려하고 있음을 확인하였다. 이에 본 장에서는 선행 장에서 분석한 결과를 통해 행정정보 데이터세트 종합관리시스템에 적용 가능한 서비스 기능 제안을 위한 시사점을 도출하였다.

첫째, 행정정보 데이터세트 관리기준표 등록을 지원하는 서비스가 필요하다. 데이터세트 등록 서비스 현황 분석결과, 데이터 활용을 위해서는 고품질의 데이터 등록이 선행되어야 한다. 이에 국내외 공공데이터 포털은 등록 서비스가 활성화되어 있으며, 각급 기관은 데이터세트 개방에 방점을 두고 있다. 이처럼 행정정보 데이터세트 관리기준표 작성을 독려하고 기관 간 상호 참조를 위해선 관리기준표 등록 서비스가 필수적으로 제공되어야 한다.

둘째, 행정정보 데이터세트 관리기준표와 행정정보 데이터세트를 검색하기 위한 다양한 접근점이 필요하다. 기록물관리 담당자는 데이터세트 관리기준표 작성 사례를 참고하고, 이용자가 원하는 데이터세트를 찾으려면 종합관리시스템에서 다양한 검색 기능을 제공해야 한다. 검색 및 열람 서비스 현황 분석결과, 우리나라 공공데이터 포털을 제외한 웹사이트에서는 데이터세트 접근을 위해 검색 기능을 제공하나 키워드에 기반한 통합검색과 카테고리별 데이터세트 목록을 제공하는 정도에서 그쳤다. 찾고자 하는 자료를 빠르고 효율적으로 검색할 수 있도록 통합검색뿐만 아니라 주제별 검색, 기관별 검색, 상세검색 등 여러 검색 방식을 제공하고, 검색 결과 내에서 필요한 자료만 필터링해서 볼 수 있도록 검색 서비스를 제공할 필요가 있다.

셋째, 데이터세트 분석 및 활용을 위한 서비스 방안 마련이 필요하다. 분석 및 활용 서비스 현황 분석결과, 미국, 우리나라, 미국 공공데이터 포털 등 국내외 데이터 포털에서 데이터세트 활용 활성화를 위한 서비스 제공 필요성을 인지하여 데이터세트를 분석하여 활용할 수 있는 도구를 직접 혹은 간접적으로 제공하려는 움직임을 보인다. 그러나 단순 분석결과 제공이나 분석에 필요한 도구 소개에 그치는 경우가 대다수이다. 하지만 고품질 기록정보인 행정정보 데이터세트의 특성을 효과적으로 활용하려면 종합관리시스템에서 데이터 분석이나 시각화를 수행할 수 있는 도구와 분석환경을 마련할 필요가 있다.

넷째, 정보 교류를 위한 커뮤니티 서비스가 필요하다. 미국, 호주, 우리나라 공공데이터 포털과 Internet Archive는 이용자 간 정보를 교류할 수 있는 이용자 커뮤니티 서비스를 제공하고 있다. 이를 통해 개발자와 일반 이용자들이 서비스를 이용하고 데이터세트를 활용하면서 겪는 어려운 점을 공유할 수 있다. 이처럼 이용자-종합관리시스템 관리 담당자 간 소통할 수 있는 창구 마련 시 종합관리시스템 서비스나 이용에 관한 피드백을 얻을 수 있으며, 이를 통해 기록관리 업무 수행이나 데이터세트를 활용할 때 나타나는 문제점을 공유함으로써 솔루션을 얻고, 경험과 정보를 교류할 수 있어야 한다.

다섯째, 서비스 이용 및 데이터 활용 교육 서비스 지원 방안이 필요하다. EU 공공데이터 포털과 AI Hub만이 교육과 관련된 서비스를 제공하고 있었다. 교육 서비스에는 다양한 프로그램이 포함될 수 있다. EU 포털의 경우 데이터세트 관리자와 이용자 모두 참여할 수 있는 웨비나와 필요한 교육 등, AI Hub는 데이터세트 활용에 집중한 교육 프로그램과 실습 프로그램을 제공하고 있다. 이처럼 종합관리시스템에서 기록물관리 담당자를 위한 데이터세트 관리기준표 작성에 관한 교육이나 모범 사례 공유 등의 세미나, 데이터세트를 이해하고 분석하는 데 필요한 지식을 학습할 수 있는 분석 교육 등 여러 프로그램을 기획하여 이용자의 역량을 강화할 수 있다.

지금까지 국내외 데이터세트 서비스 현황 분석을 통한 참고할만한 시사점을 도출하였다. 이와 함께 본 연구에서는 종합관리시스템이 행정정보 데이터세트를 위해 추가적으로 제공할 특정 서비스 방안을 다음과 같이 제안한다.

먼저, 웹브라우저에서 데이터세트의 원문을 확인할 수 있도록 뷰어(Viewer) 제공이 필요하다. 국외 국립기록관과 국내외 공공데이터 포털에서 제공하는 데이터 미리보기 기능은 테이블 형태의 데이터세트를 확인할 수 있으나

극히 일부의 레코드만 확인할 수 있다. 원문을 보기 위해서는 웹사이트에서 데이터세트 파일을 다운로드하고 이를 실행할 수 있는 상용 애플리케이션 설치가 선행되어야 한다. 이러한 과정을 생략하고 이용자들이 웹상에서 바로 데이터세트 원문을 확인할 수 있는 서비스 방안이 마련되어야 한다.

데이터세트 서비스 현황 조사한 결과, 데이터세트 개방과 더불어 데이터세트 활용 증진을 위해 데이터 분석 및 시각화 서비스를 제공하고 있다. 그러나 대부분이 웹브라우저에서 애플리케이션 실행을 지원하기보다는 분석과 시각화를 위해 필요한 애플리케이션을 이용자가 개인 컴퓨터에 직접 설치하여 사용할 수 있도록 지원하는 것에 그쳤다. 데이터 분석에 대한 전문적 지식수준이 상대적으로 낮은 이용자들에게 데이터 분석 및 활용을 위해 쓰이는 애플리케이션을 설치하고 환경에 맞게 설정을 세팅하는 것은 복잡하고 많은 시간이 소요될 것이다.

이러한 문제점을 개선하기 위해서 새로운 신기술을 도입하여 이용자의 편의성을 향상할 수 있는 서비스를 제공하여야 한다. 이에 본 연구는 종합관리시스템에 클라우드 컴퓨팅 기술에 기반한 재현 및 분석 서비스 방안 마련을 제안한다. 종합관리시스템에서 데이터를 재현하고 분석, 활용할 수 있는 응용 프로그램을 사용할 수 있도록 서비스할 시 데이터 활용의 효율성이 높아질 것으로 기대할 수 있다.

4. 행정정보 데이터세트 종합관리시스템 서비스 방안

본 장에서는 국내외 데이터세트 서비스 현황 분석결과를 토대로 도출한 시사점을 바탕으로 행정정보 데이터세트 종합관리시스템에 적용 가능한 서비스 방안을 제안한다(<표 9> 참고). 더불어 종합관리시스템에 적용한 서비스 방안 모형을 함께 제공하여 이해를 돕고자 한다.

<표 9> 행정정보 데이터세트 종합관리시스템 기능요건

기능요건	내용
등록	1) 각급 기관에서 운영하는 전체 행정정보시스템 등록 2) 데이터세트 기록관리 적용 시스템을 대상으로 행정정보 데이터세트 관리기준표 등록
검색 및 열람	행정정보 데이터세트 관리기준표와 행정정보 데이터세트 검색을 위한 여러 검색 방식 제공 예) 통합검색, 기관별 검색, 주제별 검색, SIARD 파일 검색, 상세 검색 등
분석 및 활용	클라우드 컴퓨팅 기반 행정정보 데이터세트 분석을 지원하는 다양한 애플리케이션 제공 예) 데이터 분석 도구(Python, R, Numpy 등), 데이터 시각화 도구 등
이용자 참여	각급 기관의 기록물관리 담당자-영구기록물관리기관, 기록물관리 담당자-기록물관리 담당자, 이용자-이용자 등 서로의 해결하기 어려운 문제점을 공유해 해답을 얻고, 정보와 경험을 교류하는 커뮤니티 공간 제공
교육	종합관리시스템에서 제공하는 서비스 이용 가이드 및 데이터세트 활용 교육 제공
재현	클라우드 컴퓨팅 기반 행정정보 데이터세트 재현을 지원하는 다양한 애플리케이션 제공 1) 뷰어 소프트웨어를 통한 외형 재현 ① SIARD 포맷 데이터세트를 SIARD 뷰어(SIARD GUI)를 통해 외형 재현 ② DBMS에서 SIARD 파일 업로드하여 생산 당시 외형과 일부 기능 재현 2) 에뮬레이션을 통해 데이터세트의 외형과 기능 재현

4.1 행정정보시스템과 행정정보 데이터세트 관리기준표 등록

현행 행정정보 데이터세트 기록관리 기준에 의거, 기록물관리 담당자는 행정정보시스템의 유형을 구분하여 기록관리 적용 여부를 판단한다. 적용 제외대상 시스템은 그 사유와 근거를 기록으로 남기며, 기록관리 적용대상

시스템은 행정정보 데이터세트 관리기준표를 작성하여 관리기준표에 명시된 바와 같이 기록관리를 적용한다(이주광, 2020). 이러한 업무를 지원하기 위해 행정정보 데이터세트 종합관리시스템은 각급 기관마다 운영하는 행정정보시스템 중 기록관리 대상 시스템에 한정하여 데이터세트 관리기준표를 등록할 수 있는 기능을 제공할 필요가 있다. 행정정보 데이터세트는 행정정보시스템 내에 동적인 상태로 존재한다는 특성이 있다. 행정정보시스템은 기관의 업무를 수행하기 위해 지속적으로 운영된다. 이에 따라 행정정보 데이터세트는 업무가 종결되거나 기관 폐지, 행정정보시스템이 사용 종료되지 않는 한 수정, 삭제와 같은 변경이 수시로 일어난다. 또한 업무가 종결되어도 보존기간이 남아있는 데이터세트는 다른 업무에 활용될 수 있어 시스템 내에 데이터세트를 그대로 두고 각급 기관에서 자체관리할 수 있도록 한다. 더는 데이터세트에 변경이 일어나지 않는 상태이고 국가적으로 중요한 자료이기 때문에 영구기록물관리기관으로 이관이 결정 난 후에 종합관리시스템에 등록할 수 있다. 이관 시 국가기록원에서 제공하는 이관도구 SIARD_KR(SIARD_KR, 2020)을 사용하여 이관할 데이터세트를 추출하고 이를 종합관리시스템으로 등록한다.

이에 본 연구에서는 행정정보 데이터세트를 종합관리시스템에 등록하는 기능을 제외한 행정정보시스템 등록-행정정보 데이터세트 관리기준표 등록 두 단계를 거쳐 데이터세트 관리기준표 등록 서비스 방안을 제시하고자 한다. 행정정보시스템과 데이터세트 관리기준표를 등록할 수 있도록 웹상에서 주어진 항목의 내용을 직접 작성하여 등록하는 수동 등록 방식과 서식에 맞춰 미리 작성한 파일을 업로드하여 등록하는 자동 등록 방식, 총 두 가지의 등록 방식을 채택하였다. 기록물관리 담당자는 본인이 선호하는 방식을 선택하여 행정정보시스템과 데이터세트 관리기준표를 등록할 수 있다.

4.1.1 행정정보시스템 등록

데이터세트 관리기준표를 등록하기에 앞서 각급 기관에서 운영하는 전체 행정정보시스템 등록이 선행되어야 한다. <그림 2>와 <그림 3>은 본 연구에서 제안하는 종합관리시스템의 행정정보시스템 등록 모형이다. 수동 등록 시 작성해야 하는 항목은 데이터세트 관리지침의 제외유형 구분 양식 항목을 참고하여 기관명(부서명), 담당자 이름과 연락처, 시스템명과 시스템 개요, 생성되는 데이터세트 내용, 선정유형과 선정 제외 사유로 구성하였다. 제외대상 시스템(C, D 유형)인 경우 선정 제외 사유를 반드시 작성해야 하지만, 기록관리 적용대상 시스템(A, B 유형)은 해당 사항이 없으므로 생략한다. 자동 등록 시 데이터세트 관리지침 제외유형 구분 양식에 맞춰 작성한 파일을 업로드한다. 이때 파일 포맷은 XML 기반 개방형 문서 포맷인 hwpX 또는 xlsx로 한정하여 기계가독성을 확보할 수 있도록 한다.

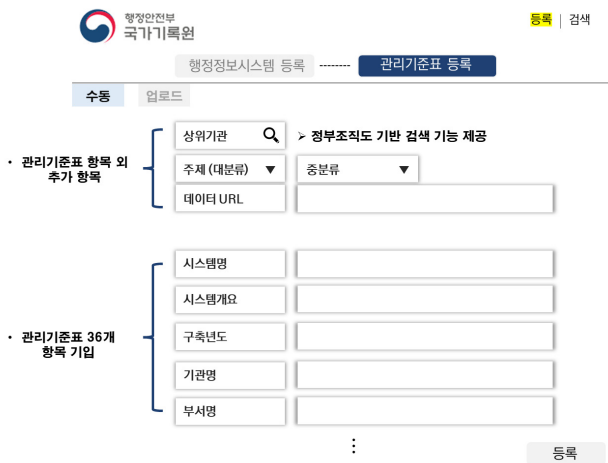
<그림 2> 행정정보시스템 수동 등록 서비스 예시모형

<그림 3> 행정정보시스템 자동 등록 서비스 예시모형

4.1.2 행정정보 데이터세트 관리기준표 등록

기록관리 제외대상 시스템의 경우 행정정보시스템 등록만 마치면 되지만, 기록관리 적용대상 시스템의 경우 데이터세트 관리기준표 등록 단계로 넘어간다. 다음 <그림 4>, <그림 5>는 종합관리시스템에서 제공하는 데이터세트 관리기준표 등록 예시모형이다. 수동 등록 시 데이터세트 관리기준표 6개 영역 36개 항목에 해당하는 내용을 직접 작성하고, 자동 등록 시 미리 작성한 데이터세트 관리기준표 파일을 업로드한다. 행정정보시스템 등록과 마찬가지로 파일 포맷은 hwp와 xls로 한정한다. 데이터세트 관리기준표의 기관명, 시스템명, 주제어 등은 검색 키워드가 될 수 있는 항목으로 검색 색인어로 사용될 수 있다.

데이터세트 관리기준표 항목 외 추가적인 항목으로 상위기관, 주제(대분류), 데이터세트 URL 등을 입력하도록 한다. 이는 데이터세트 관리기준표를 등록한 기관들을 상위기관으로 범주화한 정부조직도 기반 검색과 주제별 검색을 위한 기반이 되며, 외부 데이터 URL 기입하여 연계되는 외부 웹사이트 페이지와 연결될 수 있도록 한다. 이러한 추가 항목은 수동 등록과 자동 등록에 상관없이 직접 입력하도록 한다.



<그림 4> 관리기준표 수동 등록 서비스 예시모형



<그림 5> 관리기준표 자동 등록 서비스 예시모형

4.2 행정정보 데이터세트 관리기준표와 행정정보 데이터세트 검색

검색은 행정정보 데이터세트 종합관리시스템에서 제공하는 행정정보 데이터세트 관리기준표와 행정정보 데이터세트에 접근하기 위한 필수 기능이다. 종합관리시스템 이용자인 기록물관리 담당자와 대국민의 다각적인 검색 요구를 수용할 수 있도록 다양한 검색 방식을 제공할 필요가 있다. 이를 위해 본 연구에서는 통합검색을 기반으로 기관별 검색, 주제별 검색, SIARD 파일 검색, 상세검색 총 다섯 가지의 검색 방식과 함께 검색의 효율을 높이는 부가기능을 제안하고자 한다.

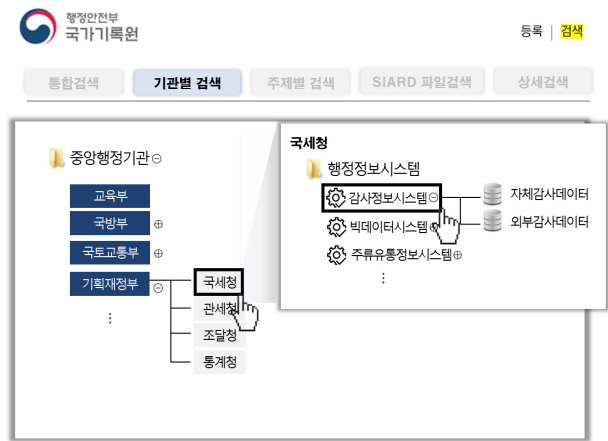
통합검색은 가장 보편적인 검색 방식으로 이용자가 검색창에 키워드 입력해 찾고자 하는 데이터세트 관리기준표 혹은 데이터세트를 결과로 제공한다. 앞서 기록물관리 담당자가 등록한 데이터세트 관리기준표의 시스템명, 업무명, 업무내용 및 목적, 단위기능명, 주제어 등과 같은 항목은 검색 키워드로 사용될 수 있어 검색 색인어로도 사용된다. 이를 통해 기록물관리 담당자는 데이터세트 기록관리 단위인 단위기능 검색이 가능해 데이터세트 관리기준표 작성 시 관련 내용을 참고할 수 있다. <그림 6>과 같이 검색창 주위에 인기 데이터세트나 최신 관리기준표 등과 같은 정보 볼 수 있도록 배치해 이용자가 쉽고 빠르게 관리기준표나 데이터세트에 접근할 수 있도록 한다.

기관별 검색은 <그림 7>과 같이 중앙행정기관 기준 부-처-청의 순서로 구성된 계층형 조직도를 제공하는 디렉

터리형 검색 방식이다. 각 조직에 속하는 기관-행정정보시스템-행정정보 데이터세트 관리기준표/데이터세트 순으로 계층 부여하여 상위 카테고리부터 하위 카테고리까지 차례로 들어가 정보를 찾을 수 있다. 이를 통해 각급 기관의 전체 행정정보시스템 목록 확인이 가능하며, 각 시스템에 속한 데이터세트 관리기준표에 접근할 수 있다. 데이터세트 관리기준표 작성자는 본인이 속해 있는 기관과 유사한 업무를 수행하는 유관기관이 등록한 데이터세트 관리기준표에 쉽게 접근할 수 있어 데이터세트 관리기준표 작성 시 참고할 수 있다.

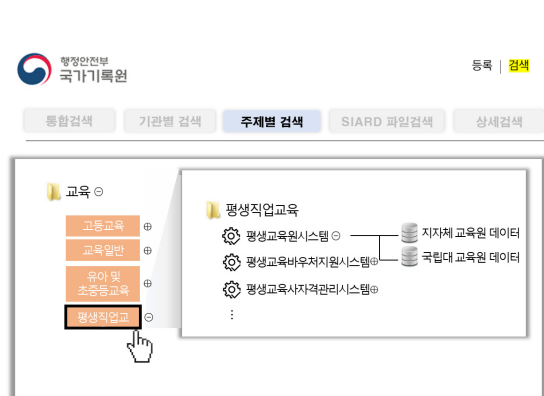


<그림 6> 통합검색 예시모형

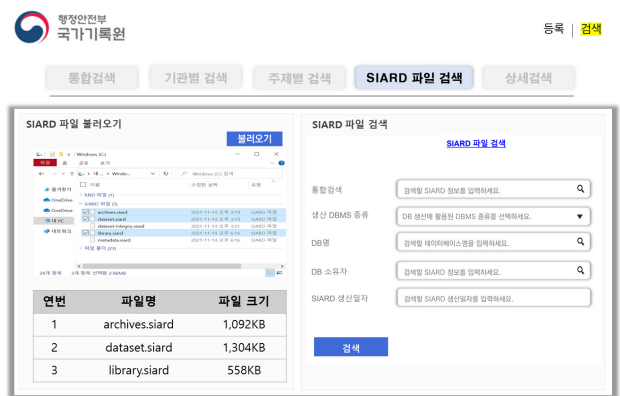


<그림 7> 기관별 검색 예시모형

주제별 검색은 사전에 설정한 주제별 범주에 따라 분류된 행정정보시스템 목록을 통해 각 시스템에 속한 행정 정보 데이터세트 관리기준표와 데이터세트에 접근할 수 있도록 한 검색 방식이다. 이용자가 찾고자 하는 정보에 대한 주제어나 키워드를 알기 어려워도 계층적으로 분류된 주제를 따라 쉽게 검색할 수 있다. 본 연구는 이를 위해 19개의 대분류(<표 10> 참조)와 75개의 중분류를 제안하여 대분류-중분류-행정정보시스템-데이터세트 계층 구조로 검색할 수 있도록 한다(<그림 8> 참조). 향후 공공데이터 포털과 종합관리시스템의 연계성 강화를 위해 공공데이터 포털과 같은 주제 분류체계를 적용하거나 기존 종합관리시스템의 주제 분류체계의 대분류와 공공데이터포털의 주제 분류체계와 대응되도록 매핑(예. 농림 + 해양수산 → 농림해양수산)하는 방안을 적용하여 상호운용성을 확보하고 공공데이터를 비롯한 행정정보 데이터세트 활용도 제고 가능성을 열어둔다.



<그림 8> 주제별 검색 예시모형



<그림 9> SIARD 파일 검색 예시모형

<표 10> 종합관리체계와 공공데이터포털 주제 분류체계 대분류

행정정보 데이터세트 주제 분류체계(안) 대분류(19개)				공공데이터포털 주제 분류체계 대분류(16개)			
공공질서 및 안전	공동행정 지원	과학기술	교통 및 물류	공공질서 및 안전	과학기술	교육	국방
국방	국회	농림	문화체육관광	국회	농림해양수산	문화체육관광	보건
보건	사회복지	산업·통상·중소기업	일반공공행정(고유)	사회복지	산업·통상·중소기업	교통 및 물류	일반공공행정
재정·세제·금융	지역개발	통신	통일·외교	지역개발	통신	통일·외교	환경보호
해양수산	환경						

SIARD 파일 검색은 행정정보 데이터세트 특성에 부합하는 SIARD 파일의 주요 메타데이터를 검색할 수 있도록 한 검색 방식이다. <그림 9>와 같이 이용자의 로컬에 저장된 SIARD 파일 중 불러올 파일을 선택할 경우 파일의 항목을 확인할 수 있다. 불러온 SIARD 파일에 대한 검색은 생산 DBMS 종류, DB명, DB 소유자, SIARD 생산 일자 등 주요 메타데이터를 대상으로 수행된다. 주요 메타데이터로 색인어로서 활용도가 높은 항목을 선정했으며, SFA(스위스 연방 기록원)에서 발표한 SIARD 표준을 참고하였다.

상세검색은 기본 검색에 여러 검색조건을 추가하여 더 정확한 검색을 결과를 얻을 수 있는 검색 방식이다. 관리기준표의 항목을 기준으로 종합관리시스템의 검색조건 설정하였다. <그림 10>과 같이 기관명, 시스템명, 데이터세트명, 주제어, 기관 유형으로 검색조건을 구성하였고, 각 검색 질의어마다 불리언 연산자(AND, OR)를 선택하여 상세 검색식을 만들어 검색 범위를 제한할 수 있다. 검색조건 중 데이터세트 관리유형은 자체관리과 이관되어 관리되는 것으로 구분하여 필요시 체크박스를 선택하도록 한다.



<그림 10> 상세검색 예시모형

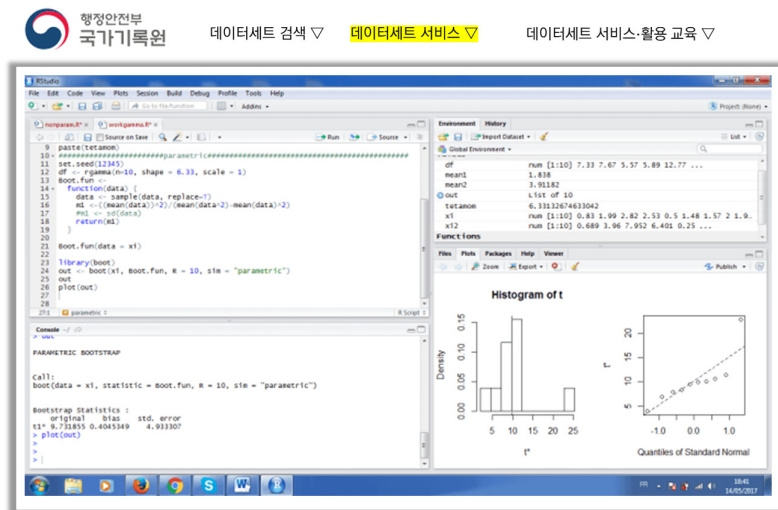


<그림 11> 검색 결과 화면 예시모형

다각도로 검색할 수 있는 검색 방식과 함께 검색의 효율성을 높일 수 있도록 다음과 같은 부가적인 기능을 고려할 수 있다(<그림 11> 참고). 질의어 검색 시 한/영 키가 바뀌어 질의어를 잘못 검색한 경우, 한글로 자동 변환 및 인식하여 (예. wosks → 재난) 올바른 검색 결과가 나올 수 있도록 한다. 검색창 하단에 검색 질의어와 상관도가 높은 키워드(주제어)를 추출하여 추천 키워드를 제공한다. 검색 결과는 질의어와 연관도가 높은 순으로 정렬되도록 하며, 관리기준표 등록 시 입력되는 항목을 메타데이터로 사용하여 기본적인 정보를 검색 결과 목록에서 볼 수 있도록 한다. 또한 검색 결과의 범위를 좁힐 수 있도록 필터링할 수 있는 기능을 제공한다.

4.3 행정정보 데이터세트 분석 서비스

기록을 보존하는 궁극적인 목표는 활용에 있는 만큼 종합관리시스템은 보존가치가 높은 행정정보 데이터세트를 적극적으로 활용할 수 있는 서비스를 제공할 필요가 있다. 이를 위해 본 연구는 클라우드 컴퓨팅에 기반한 분석 서비스를 제공할 것을 제안한다. 클라우드 환경을 구축한 종합관리시스템은 이용자들에게 데이터세트 분석에 필요한 인프라를 지원할 수 있다. <그림 12>는 종합관리시스템의 클라우드 서버 내 분석 도구 일종인 R을 탑재한 예시모형이다. 데이터를 수집하고 정제하는 가공 단계부터 분석 기법을 활용한 분석에 이르기까지 사용되는 분석 도구의 종류는 다양하며, 분석결과를 시각적으로 표현하는 시각화 도구 역시 종류가 다양하다. 이러한 도구들을 클라우드 서버 내에 탑재해 제공함으로써 이용자는 별도의 분석 소프트웨어를 설치할 필요 없이 종합관리시스템의 웹 브라우저에서 분석 및 시각화를 수행할 수 있는 편의를 지원받을 수 있다. 더불어 여러 분석 플랫폼을 지원하여 이용자의 선호하는, 실력 수준에 맞는 플랫폼을 선택하여 사용할 수 있도록 한다.



<그림 12> R Programming 활용 예시화면

4.4 행정정보 데이터세트 이용자 참여 서비스

향후 구축 예정인 행정정보 데이터세트 종합관리시스템에는 데이터세트 기록관리 업무를 수행하고자 하는, 행정정보 데이터세트를 얻어 활용하고자 하는 목적을 가진 자들이 모일 것이다. 데이터세트 기록관리 초기 단계인 지금 기록물관리 담당자들은 익숙지 않은 데이터세트 기록관리를 하며 수반되는 어려움을 마주할 것이다. 데이터세트를 분석하여 활용하고자 하는 이용자들 역시 분석과정에서 발생하는 문제점을 해결하지 못하는 어려움을 겪을 수도 있다. <그림 13>은 관리기준표 작성이나 데이터세트 활용에 관한 게시글을 올려 의견을 나눌 수 있는 이용자 커뮤니티 예시모형이다. 이처럼 종합관리시스템을 통해 소통할 수 있는 각각의 커뮤니티를 구축하여 비슷한 문제를 겪고 있는, 이미 문제를 해결한 이들이 서로의 정보와 경험, 의견을 공유할 수 있는 소통창구로 활용할 수 있도록 한다.



<그림 13> 종합관리시스템 이용자 커뮤니티 예시모형

4.5 데이터세트 활용 교육 서비스

다양한 서비스를 제공하는 만큼 이용자들이 서비스를 온전히 이용할 수 있어야 한다. 종합관리시스템은 등록, 검색, 분석, 교육, 재현 서비스 기능과 이용 방법을 모두 포함한 이용가이드를 배포하여 도움을 주어야 한다. 데이터세트를 분석하는 데 있어 하드웨어·소프트웨어 인프라, 양질의 데이터, 분석 전문성 등이 요구된다. 그러나 종합관리시스템의 이용자 저변확대에 따라 이용자마다 분석에 대한 전문지식 수준이 다양하다. <그림 14>는 데이터세트 활용 교육 예시 커리큘럼이다. 기초-초급-고급-심화와 같이 단계별로 구성된 교육을 제공하여 기초부터 심화까지 학습해가며 기술력을 향상할 수 있도록 한다. 종합관리시스템에서 데이터세트 분석과 활용을 위한 서비스 제공과 함께 이용자의 데이터 활용을 위한 교육 프로그램을 제공하여 이용자가 데이터세트를 이해하고 활용할 수 있는 폭을 넓힐 수 있도록 한다.

The screenshot shows the '국가기록원' website with a navigation menu including '데이터세트 검색', '데이터세트 서비스', and '데이터세트 서비스: 활용 교육'. Below the menu is a table titled '데이터세트 활용 교육 상세 프로세스 예시모형'.

구분	시간	상세내용	
Day 1	데이터세트 이해	1H	행정정보 데이터세트 정의와 유형을 살펴보고 대표적인 활용 사례 소개
	SIARD GUI 활용	1H	SIARD 포맷을 지원하는 전용 어플리케이션인 SIARD GUI 활용법 소개
Day 2	데이터세트 검색	2H	국가기록원 종합관리체계가 제공하는 유형별 검색 방법 소개
	Open API 활용	2H	다양한 Open API 활용 기법 소개
	Crawling 기법 활용	2H	웹사이트, SNS 등에 적용 가능한 Crawling 기법 소개
Day 3	Python 기반 데이터세트 분석	2H	Python을 활용한 데이터세트 가공, 전처리, 분석 방법 소개
	Python 기반 데이터세트 분석 심화	2H	Python이 제공하는 다양한 라이브러리를 활용한 데이터세트 분석 기법 소개
	R 기반 데이터세트 분석	2H	R Programming을 이용한 데이터세트 분석 방법 소개
Day 4	인공지능 이해	2H	인공지능의 정의와 인공지능 설계를 위한 전반적인 프로세스 소개
	머신러닝, 딥러닝 기법 활용	5H	머신러닝과 딥러닝의 대표적인 기법을 소개하고 데이터세트 특성에 부합하는 기계학습 기법 적용

<그림 14> 데이터세트 활용 교육 상세 프로세스 예시모형

4.6 행정정보 데이터세트 재현 서비스

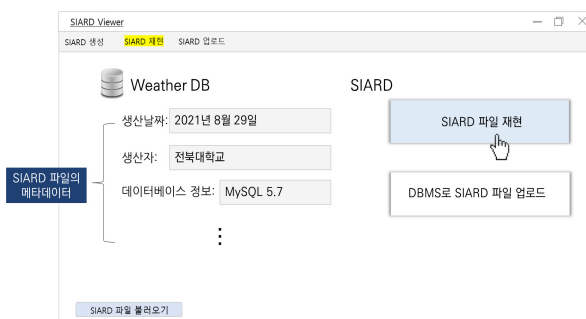
행정정보 데이터세트 종합관리시스템에서 개방하고 있는 데이터세트의 외형과 기능에 접근할 수 있어야 한다. 이에 본 연구는 클라우드 컴퓨팅을 활용한 재현 서비스를 제안한다. 재현 서비스는 크게 두 단계로 구분한다. 첫 번째 단계는 종합관리시스템 내에서 SIARD 파일을 재현하기 위한 플랫폼 제공이다. 더 나아가 두 번째 단계는 애플레이션을 통한 재현 서비스로 데이터세트 생산 당시의 컴퓨팅 환경까지 재현하는 것이다.

4.6.1 SIARD 파일의 재현을 위한 플랫폼 제공

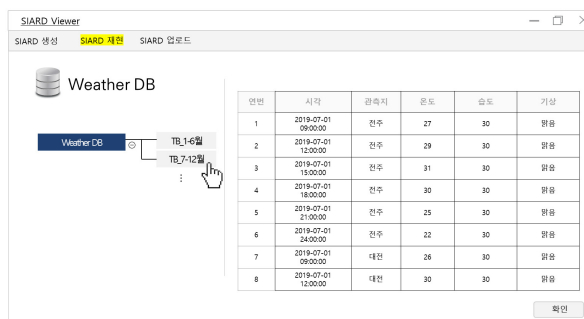
SIARD 파일 형태로 보관·관리되고 있는 데이터세트 다운로드할 수 있으면서 웹상에서 데이터파일을 열람할 수 있어야 한다. 별도의 프로그램 설치 없이 클라우드 서버 내에 SIARD 파일 형태로 이관된 데이터세트를 열람하고 분석할 수 있도록 웹기반 SIARD GUI 애플리케이션 및 DBMS를 탑재해 SIARD 파일의 내용을 확인할 수 있도록 재현하는 방안을 제안하였다.

1) SIARD GUI를 통한 SIARD 파일 재현

클라우드 서버 내에 웹기반 SIARD GUI 애플리케이션을 탑재해 이용자는 별도의 설치 없이 SIARD_KR(SIARD_KR, 2020) 프로그램을 통해 추출된 SIARD 파일의 내용을 확인할 수 있다. <그림 15>는 행정정보 데이터세트 생산 당시 활용된 DBMS 종류 및 접근정보를 설정하여 클라우드 서버에 저장된 데이터세트를 재현하는 예시모형이다. 접속 정보를 설정하면 SIARD 파일에 저장된 테이블 정보와 메타데이터, 데이터 등을 시각화하여 이용자에게 제공한다(<그림 16> 참고).



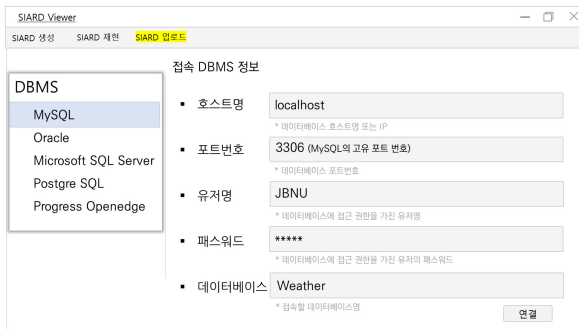
<그림 15> SIARD 파일 재현 서비스 예시모형 1



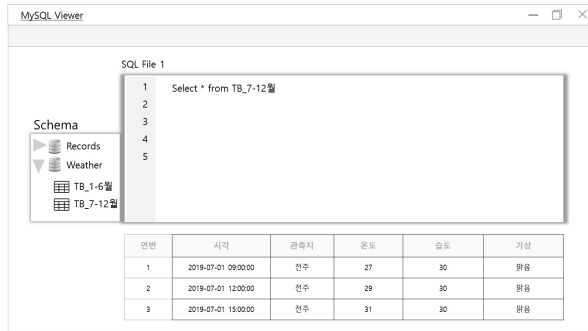
<그림 16> SIARD 파일 재현 서비스 예시모형 2

2) DBMS로 SIARD 파일 업로드하여 재현

클라우드 서버 내에 여러 DBMS를 탑재하여 웹 브라우저에서 SIARD 파일을 DBMS에 업로드해 행정정보 데이터세트의 생산 당시 모습을 재현할 수 있도록 한다. <그림 17>과 같이 SIARD 파일을 탑재할 DBMS에 접속할 수 있는 정보를 입력하면 SIARD 파일의 계층구조(Schema-Table-Row-Column)를 확인할 수 있고(<그림 18> 참조) SW 설치 없이 바로 작업이 가능하다.



<그림 17> DBMS로 SIARD 파일 업로드 예시모형 1



<그림 18> DBMS로 SIARD 파일 업로드 예시모형 2

4.6.2 에뮬레이션을 통한 데이터세트 재현

SIARD 파일의 데이터세트를 SIARD 뷰어를 통해 열람하고, DBMS에 업로드하여 열람할 수 있도록 하는 데이터세트 재현 방식은 생산 당시의 테이블 형태의 모습을 재현할 수 있다. 그러나 생산 당시의 하드웨어, 소프트웨어 등의 컴퓨팅 환경까지 완전하게 재현하지 못한다. 이를 재현하기 위해서는 기록관리 대상으로 규정되지 않은 WAS, Web Server 등의 보존이 선행되어야 한다. 생산 당시 시스템마다 설치된 프로그램, 시스템 설정 등의 정보가 필요하다. 클라우드 서버 내부에 이러한 정보가 담긴 시스템 프로파일을 탑재하거나, 시스템 이미지가 저장된 위치를 찾아가 시스템 이미지 자체를 클라우드 내에 설치하여 에뮬레이션하는 방안도 고려할 수 있다. 시스템 프로파일 명세나 저장된 시스템 이미지를 통해 에뮬레이션하는 방식은 서로 다르나 이용자에게 보이는 모습은 같다.

5. 결론

2020년 공공기록물법 시행령 개정에 따라 행정정보 데이터세트 기록관리 방안이 법제화되어 기록물관리 담당자들은 데이터세트 기록관리를 위해 관리기준표를 작성할 의무를 지니게 되었다. 국가기록원은 이들이 수행해야 할 데이터세트 기록관리 업무를 지원하기 위해 행정정보 데이터세트 종합관리시스템을 구축할 계획을 밝혔다. 그러나 데이터세트 관리기준표와 데이터세트의 특성을 고려한 구체적인 서비스 방안이 부재한 상황이다. 이에 본 연구는 데이터세트 기록관리 업무를 지원하면서 데이터세트의 활용을 제고할 수 있는 종합관리시스템의 서비스 방안을 제안하고자 하였다. 이를 위해 데이터세트 등록, 검색 및 열람, 분석 및 활용, 이용자 참여, 교육 5가지 서비스 유형을 기준으로 국외 국립기록원과 국내외 공공데이터 포털 14곳의 데이터세트 서비스 현황을 비교 분석하였다.

국내외 웹사이트의 데이터세트 서비스 현황을 종합 분석한 결과, 등록, 검색 및 열람 서비스보다 분석 및 활용, 이용자 참여, 교육 서비스 제공이 상대적으로 부족한 것으로 파악되었다. 데이터세트 개방과 이용자들의 데이터세트 활용 편의 증진을 위한 서비스를 함께 고려하려는 움직임을 보이거나 여전히 데이터세트 공개와 제공에 중점을 두고 있음을 확인하였다. 본 연구는 기록물관리 담당자의 행정정보 데이터세트 관리기준표 작성 업무 지원과 이용자에게 데이터세트를 제공하는데 필요한 서비스 기능 제안을 일차적 목적으로 두었고 나아가 이용자의 데이터세트 활용 활성화를 위한 서비스 방안을 제안하였다. 그 결과 행정정보 데이터세트 관리기준표 등록, 관리기준표와 데이터세트 검색, 데이터세트 분석과 재현을 위한 도구 제공, 이용자 커뮤니티 구축, 서비스 이용 및 활용

교육 제공 등의 서비스 방안을 제안하였다. 현재 국가기록원의 종합관리시스템은 기록관리업무 지원 측면에 방점을 두고 구축되고 있다. 이는 데이터세트 기록관리의 시작 단계로 데이터세트 관리기준표 작성을 선제적으로 추진하기 위함이다. 행정정보 데이터세트 기록관리가 안정적으로 도입되면 데이터세트의 구체적인 평가 체계, 이관, 처분 등에 관한 연구가 적극적으로 수행될 것이라 기대한다. 향후 연구가 행정정보 데이터세트 관리기준표 작성, 서비스, 평가로 이어지는 기록관리 프로세스 고도화로 이어지기를 기대한다.

참고문헌

- 경기데이터드림 (2015). 경기데이터드림 홈페이지. 출처: <https://data.gg.go.kr/portal/mainPage.do>
- 공공기록물 관리에 관한 법률 시행령. 대통령령 제32223호(2021.12.16., 타법개정).
- 공공기록물 관리에 관한 법률. 법률 제16661호(2019.12.03., 일부개정).
- 공공데이터포털 (2013). 공공데이터포털 홈페이지. 출처: <https://www.data.go.kr/>
- 국가기록원 (2007). 행정정보시스템 데이터세트 기록관리 방안 연구보고서.
- 국가기록원 (2015). 2015년 기록보존기술 연구개발사업 연구결과 보고서: 데이터세트 구조분석 및 진본성 보장 기록관리 기능모델 연구.
- 국가기록원 (2017). 차세대 기록관리 모델 재설계 연구 개발 연구보고서.
- 국가기록원 (2020a). 행정정보 데이터세트 기록관리 기준: 관리기준표 작성 및 이관 규격. 대전: 국가기록원.
- 국가기록원 (2020b). 행정정보 데이터세트 기록관리 실행 매뉴얼(기록관리 지침).
- 국가기록원 (2020c). 신뢰받는 기록관리를 위한 국가기록관리 중장기 발전계획(안).
- 국가기록원 (2021a). 2020년 행정정보 데이터세트 기록관리 체계구축 사업 결과보고서.
- 국가기록원 (2021b). 행정정보 데이터세트 기록정보 서비스 및 활용 모형 연구.
- 서울 열린데이터 광장 (2012). 서울 열린데이터 광장 홈페이지. 출처: <https://data.seoul.go.kr/>
- 서지인 (2020). 행정정보 데이터세트 관리 개선방안 연구: 공공데이터와의 비교를 중심으로. 한국기록학회지, 20(4), 41-58. <https://doi.org/10.14404/JKSARM.2020.20.4.041>
- 신정엽 (2021). 행정정보 데이터세트 기록관리 적용 사례 분석: 전자인사관리시스템 데이터세트 관리기준표 작성을 중심으로. 한국기록학회지, 21(3), 227-246. <http://dx.doi.org/10.14404/JKSARM.2021.21.3.227>
- 오세라, 이해영 (2019). 행정정보 데이터세트의 기록관리 방안. 한국기록학회지, 19(2), 51-76. <https://doi.org/10.14404/JKSARM.2019.19.2.051>
- 윤성호, 유신성, 최기쁨, 오효정 (2020). 공공기관 행정정보시스템 관련 단위과제 및 보존기간 책정 현황분석. 한국비블리아학회지, 31(2), 139-160. <http://dx.doi.org/10.14699/kbiblia.2020.31.2.139>
- 이주광 (2020). 행정정보 데이터세트의 기록관리 방안: 데이터세트 형태의 전자기록에 대한 기록관리 실행 방안. 국가기록원 기록관리 이슈페이퍼, 23, 3-26.
- Access to Archival Databases(AAD) (2003). AAD homepage. Available: <https://aad.archives.gov/aad/index.jsp>
- AI Hub (2017). AI Hub 홈페이지. 출처: <https://aihub.or.kr/>
- Australian Government Open Data Portal (2013). Australian Government Open Data Portal homepage. Available: <https://data.gov.au/>
- Canada Government Open Data Portal (2013). Canada Government Open Data Portal homepage. Available: <https://open.canada.ca/en/open-data>
- EU Open Data Portal (2013). EU Open Data Portal homepage. Available: <https://data.europa.eu/en>
- National Archives of Australia(NAA) (2013). NAA homepage. Available: <https://www.naa.gov.au/>
- SIARD_KR (2020). SIARD_KR open Source. Available: https://github.com/nakdataset/SIARD_KR

Swiss Government Open Data Portal (2014). Swiss Government Open Data Portal homepage. Available:

<https://opendata.swiss/en/>

U.K. Government Open Data Portal (2010). U.K. Government Open Data Portal homepage. Available: <https://data.gov.uk/>

U.S. Government Open Data Portal (2009). U.S. Government Open Data Portal homepage. Available: <https://data.gov/>

UK Government Web Archive (2010). National Digital Archive of Datasets(NDAD). Available:

https://webarchive.nationalarchives.gov.uk/ukgwa/*/http://www.ndad.nationalarchives.gov.uk/

• 국문 참고자료의 영어 표기

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

AI Hub (2017). AI Hub homepage. Available: <https://aihub.or.kr/>

Enforcement Decree Of Public Records Management Act. Presidential Decree No. 32223, December 16, 2021.

Gyeong-gi Datadream (2015). Gyeong-gi Data dream homepage. Available: <https://data.gg.go.kr/portal/mainPage.do>

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.

National Archives of Korea (2007). Reports on Records Management of Administrative Datasets.

National Archives of Korea (2015). A Research on Dataset Structure Analysis and Record Management Functional Model for Assurance Record Authority.

National Archives of Korea (2017). A Study on Conceptual Redesign of the Next Generation Record Management.

National Archives of Korea (2020a). NAK 35:2020(v1.0) Records Keeping Criteria for Dataset – Composition of Dataset Management Reference Table & Exchange of Dataset.

National Archives of Korea (2020b). Record Keeping Manual for Dataset.

National Archives of Korea (2020c). Mid-and Long-term Development Plan for National Records Management for Reliable Records Management.

National Archives of Korea (2021a). Report on the Project Results for the Construction of the Record Management System for Administrative Information Data Sets in 2020.

National Archives of Korea (2021b). Study on Modeling for Information Service & Application of Administrative Information Dataset.

Oh, Seh-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. <https://doi.org/10.14404/JKSARM.2019.19.2.051>

Open Government Data Portal (2013). Open Government Data Portal homepage. Available: <https://www.data.go.kr/>

Public Records Management Act. Act No. 16661, December 3, 2019.

Seo, Ji-In (2020). A study on the improvement of data set management in government information systems: a comparison with public data. Journal of Korean Society of Archives and Records Management, 20(4), 41-58. <https://doi.org/10.14404/JKSARM.2020.20.4.041>

Seoul Open Data Plaza (2012). Seoul Open Data Plaza homepage. Available: <https://data.seoul.go.kr/>

Shin, Jeong-Yeop (2021). Case study on managing dataset records in government information system: focusing on establishing records management reference table for electronic human resource management system. Journal of Korean Society of Archives and Records Management, 21(3), 227-246.

<http://dx.doi.org/10.14404/JKSARM.2021.21.3.227>

Yoon, Sung-Ho, Yu, Sin-Seong, Choi, Ki-ppeum, & 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. <http://dx.doi.org/10.14699/kbiblia.2020.31.2.139>

