

공공기록관리분야의 빅데이터 활용 방법과 시사점 제언

Big Data Utilization and Policy Suggestions in Public Records Management

홍덕용(Deokyong Hong)

E-mail: igre@korea.kr

부산광역시 수영구청



논문접수 2021-07-19
최초심사 2021-07-21
게재확정 2021-11-22

ORCID

Deokyong Hong
https://orcid.org/0000-0002-8052-6963

초 록

본 연구에서는 오늘날 기록관리는 정보통신 기술의 발전과 업무환경이 급변하고 정부의 규모와 여러 기능들이 확대되면서 행정업무에서 발생하는 기록과 그에 따른 데이터 생산량이 대폭 증가함에 따라 관리에 대한 중요도가 커졌다. 빅데이터의 특성을 가진 공공기록물의 개념과 빅데이터 특징을 연계하여 사례로 설명한다. 빅데이터 발생 환경에 따른 사회적, 기술적, 환경적, 경제적, 정치적 영역으로 살펴보기 위해 'STEEP' 분석을 실시하였다. 공공기록관리분야에서 빅데이터 기술 적용 적절함과 필요성을 알아보고 활용이 가능한 업무 분석을 통해 공공기록관리 업무의 최우선 적용 가능한 프레임워크를 도식하고 업무 시사점을 제시하였다. 첫째, 공공기록관리 절차와 표준에 '분석' 단계를 넣고 기록관과 기록물관리전문요원들에 의해 빅데이터 분석기술을 적용할 수 있는 신규 조직과 추가연구와 시도가 필요하다. 둘째, 많은 양의 데이터 속에 비구조화 되어있고 숨겨져 있는 패턴을 발견할 수 있도록 통합적 사고와 관련이 있는 '빅데이터 분석 자격'을 갖춘 기록물관리전문요원을 양성하여야 한다. 셋째, 공공기록분야에 빅데이터기술과 인공지능을 결합하여 자가 학습 시킨 후, 맥락을 분석하고 이를 통해 공공기관의 사회 현상과 환경을 분석하고 예측 되도록 하여야 한다.

ABSTRACT

Today, record management has become more important in management as records generated from administrative work and data production have increased significantly, and the development of information and communication technology, the working environment, and the size and various functions of the government have expanded. It is explained as an example in connection with the concept of public records with the characteristics of big data and big data characteristics. Social, Technological, Economical, Environmental and Political (STEEP) analysis was conducted to examine such areas according to the big data generation environment. The appropriateness and necessity of applying big data technology in the field of public record management were identified, and the top priority applicable framework for public record management work was schematized, and business implications were presented. First, a new organization, additional research, and attempts are needed to apply big data analysis technology to public record management procedures and standards and to record management experts. Second, it is necessary to train record management specialists with "big data analysis qualifications" related to integrated thinking so that unstructured and hidden patterns can be found in a large amount of data. Third, after self-learning by combining big data technology and artificial intelligence in the field of public records, the context should be analyzed, and the social phenomena and environment of public institutions should be analyzed and predicted.

Keywords: 빅데이터, 빅데이터 분석기술, 공공기록물, 기록물관리, 공공기록물빅데이터
big data, big data analysis technology, public records, records management,
public records big data

© 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.

<https://jksarm.koar.kr>

1. 서론

정보통신 기술의 발달과 업무환경이 급변함에 따라 행정업무에서 발생하는 기록과 그에 따른 데이터 생산량이 폭발적으로 증가하고 있다. 이에 따라 국민들의 알권리 보장과 국정운영의 투명성을 확보하기 위한 기록관리의 역할이 공공기관의 데이터 관리 영역으로 확장되면서 그 중요성이 더욱 강조되고 있다.

공공기관의 기록과 데이터는 규모가 방대하고, 생성 주기도 빠르며 형태도 수치 데이터 뿐 아니라 문자와 영상 데이터를 포함하는 대량의 데이터이기 때문에 ‘빅데이터’(Manyika & Chui, 2011)라 할 수 있다. 빅데이터는 다양한 분야에서 실시간으로 대규모의 데이터를 저장·수집·분석하여 미래에 적용할 수 있는 정보와 가치를 지니고 있으며 전 세계에서 빅데이터의 특성을 앞세워 기술개발과 활용을 공격적으로 추진하고 있다.

행정안전부는 2015년, ‘정부 최초, 빅데이터 전문 조직’을 신설하였고 2018년 공공부문의 시스템 통합이나 정보기술의 변화·발전 등으로 인한 중요 원천 데이터를 보존하기 위한 제도를 마련하고 다양한 시스템을 수용할 수 있는 통합 보존 플랫폼 구축을 마련하였다. 또한 2020년 “데이터 기반행정 활성화에 관한 법률”을 제정하여 데이터를 기반으로 한 행정을 활성화하기 위한 절차 및 방법을 정하여 행정의 책임과 대응력을 높이고 신뢰성을 높이는 등 국민의 삶과 질을 향상시키고 있다(행정안전부, 2020). 이렇게 공공기관의 데이터를 관리하기 위해 전담부서와 제도가 생겨나고 있지만 공공부문에서는 전자문서 중심의 기록관리 기준과 절차에 공공데이터를 적용시키는 것이 어렵기 때문에 빅데이터 특징과 기술 등을 적용한 연구는 상대적으로 미흡하다.

「공공기록물 관리에 관한 법률」 제19조(기록물의 관리 등), 동법 시행령 제42조(기록물 생산현황 통보)등에 따라 전국 공공기관에서 생산·접수된 기록물의 취합 결과를 살펴보면 2018년 기준 전국의 공공기록물은 약 636,480천건이 생산되었다(국가기록원, 2020). 이는 공공기록물은 신뢰성을 갖춘 시스템으로 대량으로 생산되고 있으며 각 기관의 대용량 송수신의 기능을 포함시킨 시스템을 사용하고 있어 상호간에 연계가 가능한 것이다¹⁾. 즉, 생산되는 기록과 데이터의 생산량이 많아졌기 때문에 용량도 커지게 된 것이고 이는 빅데이터 기술을 적용할 수 있는 가능성을 보여준다.

하지만 국가기록원에서 생산현황을 분석한 결과를 살펴보면 각 기관에서의 생산된 기록의 건수를 단순히 취합 후 연도와 유형별 기록물의 합계 통계 변동 추이만 파악할 뿐 정형 또는 비정형화된 방대한 메타데이터를 처리 분석하는 기술이 적용되지 않아 업무 추진에 전혀 도움이 되지 않는 문제점이 있다. 또한 정부의 규모와 기능이 확대되고 국정에 대한 국민들의 관심도가 높아짐에 따라 정보공개 청구의 양이 급격히 증가하였고 그 유형도 다양해지고 있다. 하지만 국민들은 원하는 정보만 청구 하고 이를 받아가는 수준이기 때문에 정부가 요구하는 국정운영에 대한 국민들의 참여를 유도하여 정책을 펼치는 것과도 거리가 먼 상태이다.

이러한 한계 속에서도 국가기록원에서는 공공기관의 데이터 관리에 있어 기록관리의 역할을 강화하고자 「공공기록물 관리에 관한 법률」 제20조(전자기록물의 관리) 및 동법 시행령 제34조의 제3(행정정보데이터세트의 관리)에 의거 공공기관에서 운용되고 있는 행정정보시스템의 데이터세트를 모두 기록관리 대상에 포함(국가기록원, 2020)하는 체계 구축을 진행하고 있다. 그러나 기록물관리전문요원이 1인으로 구성되어 있는 대부분의 기록관 체계에서 기록관리는 대량의 기록과 데이터를 보존기간 동안만 관리를 하는 단순 업무로 인식되고 있는 실정이다. 이러한 상황에서 처리과 직원들이 기록을 생산하고 편철할 때 해당 문서의 보존기간이나 해당 단위과제와의 매핑이 적절하지 않은 경우와 같은 문제가 발생할 경우 제재나 확인이 어렵다는 한계가 있다.

이와 같이 데이터 관리 영역으로 기록관리 업무의 범위를 확장하는 것에는 여러 가지 현실적인 어려움이 존재한다. 그럼에도 불구하고 기하급수적으로 공공기록물의 양과 생산속도가 증가하고 있고 유형이 다양해짐에 따라

1) 국가기록원 표준기록관리시스템 서비스데스크 홈페이지(<http://rms.archives.go.kr>) - 알림마당 99번 - 표준기록관리시스템 도입대상 H/W 및 상용 S/W 목록 중 대용량송수신SW(4.6G이상 데이터에 대한 대용량 전송 기능) 내용 참조.

기록으로 발생하는 데이터를 어떻게 활용할지 고민하고 기록관리업무의 새로운 가치 창출과 국가와 조직의 공공 이익을 극대화하기 위한 노력은 계속될 필요가 있다. 이러한 측면에서 공공기록을 빅데이터의 맥락에서 논의한 여러 선행연구들이 존재한다.

윤철(2013)은 빅데이터 관점에서 공공기록의 증가와 정보공개청구와 기록열람을 통해 쌓여가는 이용자 정보를 기록정보서비스의 데이터로 보았다. 수집보존중심주의 기록관리 방식에서 이용자 중심의 패러다임의 변화로 인해 빅데이터로서의 기록의 활용 가능성에 대한 관심이 높아지고 있으며, 소셜 미디어 전략과 활성화 방안과 이용자 맞춤 서비스를 통해 빅데이터에 기반한 기록정보서비스의 방향을 제안하였다.

안대진과 임진희(2017)는 제4차 산업혁명의 신기술 즉, 클라우드, 빅데이터, 인공지능, 사물인터넷 등 네 가지 기술을 활용하여 기록관리 방안과 변화된 생산기록의 특징을 분석하고 로드맵을 설계하였다. 유연한 데이터 구조를 갖추고 표준화된 형식으로 기록을 생산하면서 공공서비스의 품질을 높여야 함을 강조하였다.

임진희(2021)는 빅데이터 시대에 정형데이터 뿐만 아니라 공문서의 텍스트기반의 대형 비정형데이터도 빅데이터 분석의 대상이라고 보았다. 그러나 공문서의 상당수가 개방포맷이 아니어서 비트스트림에서 텍스트를 추출하는 전처리 과정이 요구되며 문서파일 내에 맥락 메타데이터가 충분하지 않아 별도의 메타데이터 확보가 필요함을 주장하였다.

Dootson et al.(2020)는 빅데이터의 성격을 가진 다양한 공공기록이 디지털로 생산되는 시대에 기존의 방식이 아닌 소셜미디어와 같이 신뢰성이 확보되기 어려운 공간에서도 공공기록이 배포될 수 있다는 것을 인지하고 다중 출처, 집단 기억, 메타데이터, 기관 외 기록, 광역지역 간 기록, 모바일 기록 등을 포괄하는 ‘네트워크 기록 개념’을 제안하였다. 또한 기업과 정부 의사결정에 관한 통찰과 예측을 제공하기 위해 기록관리는 보다 전문적으로 빅데이터 및 데이터 중심 도구를 활용할 수 있는 방식으로 전환되어야 한다고 제시하였다. 이를 통해 공공기록관리 전문가가 디지털 기술을 활용하여 기록관리업무의 적용 범위를 확대하고 사후 판단과 통찰력을 기반으로 새로운 가치를 창출하는 데 주도적인 역할을 할 수 있는 직중임을 제시하였다.

선행연구를 통해 빅데이터 시대의 공공기록관리의 성장 가능성을 위한 논의와 빅데이터 관리를 위한 실무적인 고려 사항에 대한 논의가 이루어졌음을 알 수 있다. 본 연구에서는 STEEP 분석(거시환경분석 방법, Social, Technological, Environmental, Economic, Political analysis)을 통해 공공기록관리 분야에서 빅데이터 기술 적용 필요성을 알아보고 빅데이터 활용이 가능한 기록관리 업무 영역을 살펴보았다. 이를 통해 공공기록관리 분야의 빅데이터 활용을 위한 프레임워크와 빅데이터 분석 기술을 업무에 적용하기 위한 시사점을 제시하고자 한다.

2. 빅데이터로서의 공공기록물

공공기관의 전자기록생산시스템은 관계형 데이터베이스시스템(RDBMS)을 사용하여 데이터를 생산·운용하고 있지만 최근 빅데이터 처리를 위해 비관계형 데이터베이스 시스템의 사용이 증가하는 추세이다(국가기록원, 2020). RDBMS에서는 대량의 정형 데이터로 데이터를 저장·수집하고 있고 NoSQL로 영상이나 문자 등 정형화되지 않은 데이터도 관리가 가능하다.

공공기록물을 생산·운영하는 시스템은 관련 메타데이터를 이관하고 생산시스템을 지정하여 현행화하여 연동한다. 이러한 메타데이터를 분석하여 해당 업무의 맥락을 이해하고 추세를 파악하여 향후 문서의 가치를 측정할 수 있다면 효율적인 기록관리가 가능할 것이다. 다양한 종류와 많은 양의 데이터를 가지고 있는 기존의 RDBMS와 NoSQL로 효율적인 기록관리업무가 수행되지 못한다면 빅데이터 분석 기술을 적용해봄으로써 해결방법을 찾아야 할 것이다.

2.1 빅데이터의 개념과 특성 · 형태

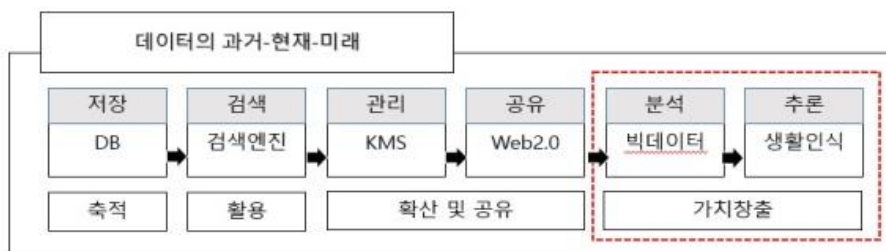
2.1.1 빅데이터의 개념

빅데이터라는 용어는 2011년 해외 블로그나 저널에서 사용하기 시작하였고 국내에서도 같은 시기에 빅데이터 관련 기사들이 등장하기 시작하였다. 맥킨지(McKinsey) 리포트(2011)는 빅데이터를 ‘기존 데이터베이스 소프트웨어를 통해 수집, 분석, 저장할 수 있는 규모를 초과하는 데이터’라 정의하고 있다. 미국의 IT 마켓리서치 회사인 IDC(International Data Corporation)는 ‘대규모의 다양한 데이터들로부터 검색, 분석, 수집을 빠르게 처리하여 효과적으로 가치를 발굴하여 수행하도록 설계된 차세대 기술 및 아키텍처’라고 정의하고 있다.

가트너의 애널리스트 더그 레이니(2010)는 데이터의 급성장에 따른 이슈와 기회를 데이터의 양(volume), 데이터 입출력의 속도(velocity), 데이터 종류의 다양성(variety)이라는 세 개의 차원으로 3V라 정의하고 있으며 빅데이터 기술 기업인 SAS(2012)에서는 가치(Value)를 추가하였다. MGI(2011)는 빅데이터를 양적인 측면에서 기존의 데이터베이스관리 도구인 DBMS의 데이터 수집 · 저장 · 분석하는 역량을 넘어서는 데이터세트(data set)이며 몇 십 테라바이트(tera byte)에서 수 페타바이트(peta byte)규모의 데이터세트라 하였다. 삼성경제연구소(2011)는 “대규모의 데이터의 집합으로 거대한 데이터와 관련한 기술과 도구를 포함하는 것”으로 정의하고 있다. 이러한 빅데이터의 개념은 외부데이터, 비정형, 실시간데이터 및 서로 상이한 형태의 정보를 결합하여 새로운 지식을 창출하는 영역(최성곤, 오진태, 장중수, 2012)으로 까지 그 의미가 확대되고 있다.

정보기술발전 이전의 데이터의 개념은 시간 순으로 수집 · 저장된 정보 즉, 변형되지 않은 순수 데이터 그 자체를 의미하였다. 하지만 정보기술발전이 되어 대규모와 실시간으로 쌓이는 데이터의 정의는 각종 IT 기기나 스마트 모빌리티들을 통해 수집 · 저장된 것 뿐만 아니라 그 안에서 활용될 수 있는 정보를 검색하여 이용자들이 이를 빠르게 이해할 수 있고 개인이나 기업 · 기관들의 아이디어와 마케팅에 활발히 사용될 수 있는 것을 말한다. 특히 ‘빅데이터 기술’이란 기존 데이터 저장 · 수집 · 분석 · 관리의 역량을 넘어서는 대량의 정형 또는 비정형 데이터세트(Manyika & Chui, 2011) 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술(Gantz & Reinsel, 2011)을 말한다.

즉, 빅데이터는 <그림 1>과 같이 기존과 달리 단순히 데이터를 수집 그리고 가공하는 것뿐만 아니라 엄청난 양의 데이터를 모아 한꺼번에 해석을 넘어서 새로운 사실 발견 및 예측을 가능하게 한다. 한마디로 우리가 다양한 데이터들을 수집, 종합하여 그 자료들을 통해 결론을 도출해내는 것과 달리 빅데이터는 과거의 데이터들을 바탕으로 미래의 사건을 예측을 할 수 있는 것이다.



<그림 1> 데이터에 대한 패러다임의 변화

(출처: 최경진 외, 빅데이터 환경에서 개인정보보호 강화를 위한 법·제도적 대책 방안 연구, 개인정보 보호위원회, 2012, 9)

2.1.2 빅데이터의 특성

가트너(Gartner, 2011)에서는 양(Volume), 속도(Velocity), 다양성(Variety)를 빅데이터의 특성이라 하였으며, 연

구자에 따라 가치(Value) 또는 복잡성(Complexity)을 추가하기도 한다. 빅데이터는 기관에서 구축된 데이터베이스 관리 시스템(DBMS: Data Base Management System)의 정형화된 데이터 뿐만 아니라 소셜미디어의 데이터와 같이 실시간 비정형화 데이터까지 포함하는 광범위한 대상이다. 즉, 빅데이터는 기존의 기관에서 구축된 시스템의 정보 단위를 넘어선 대규모의 양과 데이터의 생산과 빠르게 처리되는 데이터 속도, 동영상이나 사진 등 비구조화된 다양한 형태의 집합데이터인 것이다. 또한 빅데이터의 구조가 일반적으로 기준과 표준이 잡혀 있지 않아 데이터 관리가 복잡하고 난해하여 새로운 방법이 필요하며, 분석결과가 향후에 쓰일 수 있도록 가치가 부여되는 특징이 있다.

2.1.3 빅데이터의 형태

빅데이터의 기술을 적용하는 근본적인 이유는 기술적, 사회적, 경제적 환경에서 존재하는 방대한 데이터 속에서 흐름을 파악하고 그 안에 숨겨진 패턴을 찾아내는 것이다(Park & Leydesdorff, 2013). 기존의 데이터 분석과 빅데이터 분석의 큰 차이점은 기존에는 예전에 미리 구축된 정형데이터를 이용한 분석이 주를 이루었다면 빅데이터 분석에서는 현재 활용할 수 있는 반정형 또는 비정형데이터를 기반으로 하는 분석을 통해 가치를 창출하다는 점이다. 기록관리분야에서 활용하는 다양하고 대규모의 데이터에 포함되는 정형화 정도에 따라 정형, 반정형, 비정형 데이터로 구분하여 정리할 수 있다. <표 1>은 데이터 유형을 정리한 것이다.

<표 1> 빅데이터의 형태와 공공기록물 예시

구분	내용	종류	빅데이터 특성을 가진 기록관리분야 (연구자 정리)
정형데이터	미리 정해 놓은 형식과 구조에 따라 저장되도록 구성된 데이터	DB, 스프레드시트	기록생산·관리시스템(RDBMS)의 테이블과 같이 고정된 컬럼에 저장되는 데이터, 지정된 행과 열로 데이터의 속성이 정해져 있는 스프레드시트 데이터 등
반정형데이터	정형데이터처럼 테이블의 행과 열로 구조화 되어 있지는 않으나, 파일에 포함된 데이터 구조 정보를 바탕으로 테이블 형태의 데이터 스키마로 변환하고 데이터를 매핑하여 정형데이터로 변환가능	HTML, XML	기록관리 업무로 인해 시스템에 쌓이는 Log파일
비정형데이터	정의된 구조가 없이 정형화되지 않은 데이터	동영상파일, 사진, 보고서(문서), 메일본문 등	CLOB/BLOB ²⁾ 종이보고서(문서), 시청각기록 등의 형식으로 RDBMS내에 구조화되는 경우도 있음

2.2 공공기록물 현황

2.2.1 공공기록물의 개념과 종류

공공기록물이라 함은 국가기관·지방자치단체 그 밖에 대통령령이 정하는 기관이 업무와 관련하여 생산 또는 접수한 문서·도서·대장·카드·도면·시청각물·전자문서 등 모든 형태의 기록정보 자료와 행정박물을 말한다(공공기록물관리에 관한 법률 제3조). 이 중 전자기록물은 정보처리능력을 가진 장치에 의하여 전자적인 형태로 작성하여 송신·수신 또는 저장되는 전자문서, 웹기록물 및 행정정보 데이터 세트 등의 기록정보자료를 의미한다(공공기록물관리에관한법률 시행령 제2조). 전자기록물 유형별 정의는 <표 2>에 제시되어 있다.

2) BLOB데이터 유형은 구조화되지 않은 2진 대형 객체를 저장(문자집합 의미가 없는 비트 스트림), CLOB데이터 유형은 1바이트 및 복수바이트 문자 데이터를 저장(고정폭 문자 세트와 가변 폭 문자 세트가 모두 지원되며 둘 다 데이터베이스 문자 세트를 사용).

<표 2> 전자기록물관련 용어 정리

웹기록물	공공기관에서 운영·활용하는 웹사이트·블로그·소셜네트워크서비스(SNS) 등 웹을 기반으로 생산된 기록 정보자료와 웹사이트 운영 및 구축과 관련된 관리정보
행정정보데이터세트	「행정 효율과 협업 촉진에 관한 규정」 제3조 제12호에 따른 행정정보시스템에서 생산, 수집, 가공, 저장, 검색, 제공, 송신 및 수신 등을 위해 조합된 문자, 숫자, 도형, 이미지 및 그 밖의 데이터
기록관리 메타데이터	기록물의 내용, 맥락, 구조 및 기록물관리 이력과 관련된 사항을 기술하는 데이터

출처: 공공기록물 관리에 관한 법률 시행령 제2조(정의)

2.2.2 공공기록물을 생산·관리하는 시스템

공공기록물을 생산·관리하는 시스템은 먼저 기록관리 과정과 기록관리 프로그램을 마련하기 위한 절차의 표준인 ISO 15489을 적용한다. 이후 국제표준인 OAIIS 참조모형(Reference Model for an Open Archival Information System)을 바탕으로 정보 모형, 정보 패키지 모형, 아카이브 기능 모형 등 디지털 아카이빙과 관련된 기본적인 개념(ISO 14721, 2002)을 바탕으로 정의되었다. 또한 미국 국방부에서 개발된 기록관리시스템 표준인 DoD 5015.2 STE에서는 기록을 관리하는 시스템의 소프트웨어를 개발할 때 필수적인 메타데이터로 파일플랜, 기록물그룹명, 기록물그룹고유번호, 기록물그룹기술, 정리처분, 정리처분권한, 영구기록물표시자, 중요기록물 표시자, 중요기록물점검 및 기간, 이용자 정의 사항 등을 제시하였다.

특히 기록 메타데이터는 ISO 23081-1:2017에 따르면 시간과 공간을 초월하여 기록의 생산, 관리와 이용이 가능하도록 하는 구조화된 혹은 반구조화된 정보로 정의되고 있다. 기록 메타데이터를 통해 증거로서의 기록을 보호하고 시간이 지나더라도 기록의 접근성과 이용가능성을 보장할 수 있으며 기록의 효율적인 검색, 재이용, 기록에 대한 접근과 권한 및 개인정보 보호를 지원하는 등 다양한 혜택이 존재한다.

이러한 메타데이터를 총괄 관리할 수 있는 방법 중 하나가 관계형 데이터 모델이다. 데이터의 성격, 속성 또는 표현방법 등을 체계화 할 수 있으며, 데이터의 표준화를 통한 데이터 품질을 확보 할 수 있다. 관계형 데이터 모델은 핵심 3개의 구성요소로 구성되어 있고 모든 업무체계를 아래의 3가지로 모두 표현할 수 있다는 개념이다.

- 개체(Entity): 시스템화하고자 하는 사건, 사물
- 관계(Relationship): 개체간, 속성간의 연관성
- 속성(Attribute): 개체, 관계성의 성질을 나타내는 더 이상 쪼갤 수 없는 정보의 단위

위에서 표현된 각 구성요소들을 구현한 것이 관계형 데이터베이스 관리 시스템(Relational Database Management System)이다(Codd., 2010). 관계형 데이터베이스 관리 시스템(RDBMS)은 사용자에게 데이터를 관계로 표현하는 데 행과 열의 집합으로 구성된 테이블의 묶음 형식으로 데이터를 제공하며 테이블 형식의 데이터를 조작할 수 있는 관계 연산자로 제공한다. <표 3>은 공공기록물을 생산·관리 하는 시스템을 정리한 것인데 RDBMS는 일반적인 정보시스템에서 폭넓게 사용되고 있으나 행정정보시스템의 경우 비관계형 데이터베이스도 포괄하고 있다(양호성, 2017).

〈표 3〉 공공기록물 생산·관리 시스템

시스템명	내용	관련규정
기록관리시스템	기록관 및 특수기록관에서 기록관리를 전자적으로 수행하는 시스템(기록물법)	공공기록물 관리에 관한 법률 시행령 제2조(정의) 제8호
업무관리시스템	행정(공공)기관이 업무처리의 전 과정을 과제관리카드 및 문서관리카드 등을 이용하여 전자적으로 관리하는 시스템(규정)	행정업무의 효율적 운영에 관한 규정 제3조(정의) 제11호
전자기록생산시스템	행정업무의 효율적 운영에 관한 규정 제3조 제10호~제12호에 따른 전자문서시스템, 업무관리시스템, 행정정보시스템을 말함(규정)	
전자문서시스템	문서의 기안·검토·협조·결재·등록·시행·분류·편철·보관·보존·이관·접수·배부·공람·검색·활용 등 모든 처리절차가 전자적으로 처리되는 시스템(규정)	행정업무의 효율적 운영에 관한 규정 제3조(정의) 제10호
행정정보시스템	행정(공공)기관이 행정정보를 생산·수집·가공·저장·검색·제공·송신·수신 및 활용하기 위한 하드웨어·소프트웨어·데이터베이스와 처리절차 등을 통합한 시스템(규정)	행정업무의 효율적 운영에 관한 규정 제3조(정의) 제12호

출처: 국가기록원 전자기록생산시스템 기록관리 기능요건(v1.0), 작성자 재구성

2.2.3 빅데이터 특성을 지닌 공공기록관리 업무 사례

빅데이터는 어느 순간 갑작스럽게 만들어지지 않으며 장기간 축적을 통해 생성된다. 기록관리 관점에서 보면 과거에는 업무를 위해 단순하게 기록을 생성하였다면, 지금은 기록을 생산·관리하는 시스템에 문서기안자나 결재자의 행적을 알 수 있는 로그(Log)기록이 자동 데이터로 저장된다. 업무에 대한 어떤 의견이 있는지도 기록으로 남기며 얼마 동안의 기간에 걸쳐 기록이 생산되었는지도 알 수 있다. 또한 담당자가 민원이나 문서를 생산하는데 참고하기 위해 키워드를 검색하고 해당 문서를 몇 번 접근하여 열람하는지도 데이터로 남게 된다. 사실상 공직생활의 행동 하나하나가 빠짐없이 데이터로 저장되고 있다.

그 중 기록물 생산현황 통보 제도는 1999년 「공공기관의 기록물 관리에 관한 법률」의 제정 이후 각급 기관의 기록물 생산 및 관리 현황을 파악하기 시작하였으며, 현재 「공공기록물관리에 관한 법률」, 동법 시행령, 동법 시행규칙의 관련 조항을 근거³⁾로 운영되고 있다. 2019년 현재 생산현황 통보 기관은 중앙행정기관(50), 특별행정기관(149, 군기관포함), 지방자치단체(광역 17, 기초 227), 교육청(광역 17, 지역 176), 정부산하공공기관(39), 국·공립대학(43) 등 718개 기관이다. 각 기관은 유형별 기록물 7종(일반기록물·시청각·회의록·조사/연구/검토서·간행물·비밀기록물·행정박물)의 생산·보유현황을 국가기록원으로 통보한다.

〈표 4〉 기록물 유형별 생산현황(2018년 생산분)

(19.11.20. 기준, 단위: 권, 건, 점)

구 분	일반기록물		조사/연구/검토서(건)	회의록(건)	시청각(건)	행정박물(점)	간행물(권)
	종이(권)	전자(건)					
합 계	8,137,650	636,429,687	8,546	625,032	851,108	239,573	26,386
중앙행정기관	1,015,233	36,892,521	2,821	4,708	28,985	14,121	4,339
특별행정기관	1,618,678	89,297,173	194	191,527	175,557	40,177	1,652
광역지자체	395,269	39,307,613	382	7,494	20,584	28,406	1,217
기초지자체	2,395,043	248,755,882	4,099	50,961	458,201	75,549	2,810
광역교육청	538,825	44,573,774	134	85,049	24,369	7,645	5,449
지역교육청	1,302,303	139,369,771	40	271,032	88,886	52,432	7,474
정부산하기관	769,113	22,670,149	712	5,053	40,685	7,363	3,062
국공립대학	103,186	15,612,804	164	9,208	13,841	13,880	383

출처: 국가기록원 홈페이지(www.archives.go.kr)

3) 『공공기록물관리에 관한 법률』 제16조, 제17조, 제19조, 제22조, 제23조, 제24조, 동법 시행령 제16조, 제17조, 제18조, 제19조, 제20조, 제33조, 제42조, 동법 시행규칙 제21조, 제27조 등.

또한 공공기록물의 공개는 공공기관이 업무 수행 중 발생한 기록에 대한 정보를 국민에게 공개함으로써 국민의 알권리를 보장하고 국정운영에 참여를 유도를 위한 1996년에 「공공기관의 정보공개에 관한 법률」을 제정·공포하고, '98년 1월 1일부터 시행에 제도이다(정보공개포털, 2021). <표 5>는 5년간 전국 공공기관에서 정보목록, 원문정보, 사전공개 건수를 정리한 것이다. 정보목록은 대국민공개, 부분공개, 비공개로 생산된 모든 문서의 제목을 말하며 원문정보는 대국민공개 및 부분공개 등 부단체장 이상 생산된 문서의 원문을 제공하는 것이다.

<표 5> 5년간 공공기관 정보공개, 원문정보, 사전공개 건수

(2020.12.31. 기준, 단위: 건)

구분 \ 년도	2016년	2017년	2018년	2019년	2020년
정보목록	137,802,071	143,696,650	109,770,483	131,872,818	126,321,477
원문정보	5,804,107	5,202,881	4,585,249	6,652,825	22,177,157
사전공개	734,856				

출처: 정보공개포털(www.open.go.kr)

이와 같이 공공기록물 중 전자기록생산시스템에서 생산되는 건수가 연간 6억4천만건 정도이고 1억3천만건이 공개되어지는 것을 확인하였다면 이것이 양적인 측면에서 뿐만 아니라 내용적 측면에서도 그 규모가 상당하다는 것을 알 수 있다. 그러나 <표 4>와 <표 5>에서 제시된 공공기록의 생산과 공개 현황은 종이와 전자결재문서 등 문서등록대장의 목록 건수만 포함된 것이며, 행정정보데이터세트가 공공기록에 포함되지만 빅데이터로써 누락되는 기록들은 그 양을 추산할 수 없다. 비록 전체 공공기록의 생산 및 공개 현황을 파악할 수 없다는 한계가 있지만, 공공기관에서 누가 어떻게 왜 문서를 생산하고 활용하는지 파악하고 전자문서처럼 구조화된 형태 뿐만 아니라 행정정보데이터세트와 같은 그 외 전자기록물 유형 등을 분석하는 등 빅데이터의 관점에서 공공기록의 활용 가능성을 고려할 필요가 있다.

공공기록이 빅데이터의 3V의 특징에 부합할 수도 있지만 이러한 기록들이 폐기가 되더라도 메타데이터나 로그 파일의 형태로 전산화되어 보존된다. 시간의 경과에 따라 방대한 양의 데이터가 축적된다는 면에서 3V 중 ‘양’과 ‘다양함’의 측면에서 ‘빅데이터’라고 볼 수 있다. 이처럼 다양하고 방대한 규모의 기록은 미래 경쟁력 향상의 중요한 자원으로 활용될 것이다. 대규모 데이터를 분석하여 의미 있는 정보를 찾아내는 시도는 예전에도 존재하였지만 현재의 빅데이터 환경은 과거와 비교하여 데이터의 규모는 물론 품질과 다양성 측면에서 패러다임의 전환을 의미한다. 기존의 시스템으로 실제 업무에 적용시키는 것이 힘들다면 이를 객관화시킨 도구로 활용하여 궁극적인 가치를 창출하는 빅데이터로서의 공공기록 활용 논의가 필요하다.

3. 공공기록관리분야의 빅데이터 기술 적용 논의 및 필요성

3.1 STEEP분석을 통한 공공기록관리분야 빅데이터 기술 적절성 분석

STEEP분석은 사회(Social), 기술(Technology), 환경(Ecological), 경제(Economic), 정책/법규(Political)의 약자로 5가지 분석 틀을 활용하여 해당 산업 및 비즈니스에 영향을 주는 거시적 동향을 탐색할 수 있는 분석 방법으로 미래를 예측하거나 이해하는데 목적이 있다(배정환, 2013). 주로 국가 미래전략이나 기업 경영전략 등을 수립할

때 STEEP 분석이 활용된다(Bensoussan & Fleisher, 2013, 187). STEEP 분석 방법은 다음과 같다. 첫째, 사회 이슈나 미래변화와 관련된 보고서나 논문 등을 참고하여 사회·기술·경제·환경·정치와 관련된 부분들을 이해하고 각 영역별로 정리한다. 둘째, 각 트렌드들의 간의 연관성을 이해하여 주요 이슈들을 산출한다. 마지막으로 미래의 방향을 예측하고 시사점을 도출한다(Bensoussan & Fleisher, 2013, 194-196).

본 연구는 공공기록관리 분야의 빅데이터 활용 현황을 파악하고 미래를 예측하거나 이해하기 위해 STEEP 분석을 하였다. 전술하였듯이 공공기록관리분야의 빅데이터 활용 현황이 어느 하나에 국한된 것이 아니라 사회적, 기술적, 환경적, 경제적, 정치적인 복합 요인들에 의해 발생하기 때문에 STEEP 분석을 활용하는 것이 적절하다고 보았다.

첫째, Social(사회) 부문으로 정부 및 공공기관, 기업 등에서 빅데이터 활용이 조직에 발전이 있을 것이라는 인식을 바탕으로 전 세계의 국가 및 기업에서 관련 기술을 적용하고 있다. 최근에 스마트 모바일 기기의 출현과 SNS 사용, IoT 등으로 인해 빅데이터가 폭발적으로 증가하고 있다. 이러한 빅데이터를 통해 새로운 효용을 찾을 수 있을 것이라는 기대효과 아래 빅데이터는 IT분야의 신성장 동력과 새로운 패러다임으로 급부상 하고 있다. 공공기관의 기록관리와 관련된 법정업무, 정보공개, 국민신문고의 민원관련 그리고 공공기관의 마케팅 등 실시간으로 공공기관의 빅데이터를 활용하여 예측 가능한 공공기록관리 정책 결정의 선제적 대응이 요구되고 있다.

둘째 Technical(기술) 부문으로 기존에 이미 구축되어진 과거 정형데이터만을 활용하였다면 지금은 방대한 데이터들을 관리하기 위한 새로운 기술이 적용이 되면서 실시간으로 반정형 또는 비정형데이터를 가지고 수집, 저장, 분석이 가능하며 그 결과를 바탕으로 다양한 활용 가치를 창출하고 있다. 예를 들어 종전의 공공기록관리 분야에서는 기록물을 생산하는 공급자 위주의 정보공개서비스를 주로 하였다면 지금은 수요자 중심 즉, 국민들이 필요로 하는 정보공개서비스로 제공할 필요성이 증가하고 있다. 더불어 정보공개시스템과 기록관리시스템의 연계를 통해 빅데이터에 의한 개인화된 맞춤형 공공기록관리 분야 정보 제공 요구가 증가하고 있다. 또한 공공기관이 행정정보시스템을 통해 수행한 업무의 결과로 생성된 데이터의 기록관리도 법적 의무사항이 되면서 행정정보 데이터 세트의 관리 필요성과 실행방안에 대한 논의가 강조되고 있다.

셋째 Economic(경제) 부문과 관련하여 다양한 분야에서 빅데이터 기술이 경제·사회 발전의 원동력이자 사회 현안 해결 방안 중 하나로 인식하고 있다. 과거에는 다룰 수 없었던 문제들이 빅데이터 기술을 통해 해결되고 있으며 미국, 일본 등지의 국제 기업에서 빅데이터 기술을 적용하고 확대하여 신규 비즈니스 모델 개발을 위해 주력하고 있고 이에 따라 새로운 활용사례가 등장하고 있다.

넷째 Ecological(환경) 부문으로 행정 및 업무환경이 바뀌고 있는 시대적 현상에 주목할 필요가 있다. 저성장 구조가 지속됨에 따라 경제를 살리기 위한 신규 모멘텀이 필요하다는 인식과 함께 다양한 사회문제가 대두되고 있다. 더불어 정부-기업-국민 간 관계가 변화하면서 발생하는 문제들을 빅데이터 기술을 활용하여 해결하고자 하는 노력이 전개되고 있다. 예를 들어 공공데이터법과 정보공개법을 제정하여 사전 정보 공표를 확대하고 민간데이터 분야와의 협력을 통해 고품질의 데이터를 제공함으로써 정부서비스의 신뢰도를 향상시킬 수 있다.

다섯째 Political(정책) 부문으로 새로운 패러다임에 맞는 업무 혁신 요구가 크게 증가하고 있다. 발전된 전자정부시스템을 토대로 정부기관에서는 대용량의 데이터를 생산하고 있고 공공데이터의 개발, 소통, 공유, 협력을 신조로 빅데이터 산업 발전을 적극 추진하고 있다. 빅데이터를 통해 예측 가능한 분석을 제공함으로써 국가 경쟁력을 강화해야 한다는 것에 정부와 기업이 모두 공감하고 있고 이는 공공기록관리 분야에도 적용될 수 있을 것이다.

이상과 같이 STEEP분석을 통해 공공기록을 빅데이터로 바라보고 이를 활용함으로써 얻는 가치는 높다고 볼 수 있다. 공공기록은 그 규모와 다양성, 축적되는 속도의 측면에서 빅데이터의 특성인 3V(Volume, Variety, Velocity)와 밀접히 연관된다. 또한 IT기술의 발전과 사회, 경제, 환경, 정책적 변화에 따라 공공기록관리 분야에

서 빅데이터 기술 적용의 필요성이 증가하고 있으며 <표 6>에서 각 부문에서의 빅데이터 활용 적절성을 정리하였다.

<표 6> STEEP분석을 통한 공공기록관리분야 빅데이터 활용 적절성 분석

분야	공공기록관리분야의 빅데이터 활용 경향
S(사회)	<ul style="list-style-type: none"> • 빅데이터 분석 필요성 인식 확산(정부/공공기관, 기업, 국민) • 정보통신기술(스마트기기, SNS, IoT) 발달로 대규모 데이터 유통 및 생산 • IT분야에서의 빅데이터기술이 신 성장동력이자 새로운 패러다임으로 급부상 • 데이터를 기반한 예측가능한 공공기록관리정책의 예측 대응 요구 증대
T(기술)	<ul style="list-style-type: none"> • 지식과 기술의 합치로 새로운 기술 등장 • 행정업무와 국민을 이어주는 행정정보데이터 세트 데이터 관리 필요성 증대 • 시스템 연계·통합으로 개인맞춤형 공공기관 정보 제공 요구 증가
E(경제)	<ul style="list-style-type: none"> • 현안해결과 조직발전 등 사회·경제 발전의 원동력 • 세계 주요 국가 및 국제 기업의 빅데이터 사업 활용 및 육성 주력 • 신규 비즈니스 모델 제공 및 공공기록관리 데이터 통합을 통한 일자리 창출
E(환경)	<ul style="list-style-type: none"> • 민간데이터 융합과 공공데이터 개방 확대에 의한 통찰력 창출 • 경제 문화의 고도화된 국민 신뢰도 향상 요구 증가 • 데이터 정보 공개·공유에 따른 정보격차 및 개인정보 보호 제한
P(정책)	<ul style="list-style-type: none"> • 새로운 기술변화와 이데올로기에 맞는 혁신 요구 증대 • 미래대비 동력 및 국가경쟁력 확보라는 공감대 형성

3.2 공공기록관리와 빅데이터의 연관성 분석

빅데이터의 특성인 빠른 속도로 축적되는 대규모의 다양한 정형데이터와 비정형데이터를 포함한 공공기록을 기존의 방식대로 의미 없이 축적하는 것보다는 실시간으로 이를 수집·저장하여 과학적 알고리즘에 의해 분석, 활용할 수 있는 환경을 구축하는 것이 필요하다. 이를 통해 특정 업무에 대한 예측 가능한 결과를 제시함으로써 공공기록관리가 정책적으로 중요한 도구로 인식될 수 있다.

현재까지 기록생애주기 개념에 따라 기록에 담긴 정보의 특성을 세분화하고 그 의미를 정형화함으로써 기록물을 관리하는 개념틀을 제공하였다면 앞으로는 공공기록관리 분석을 통해 기록관리가 나아갈 수 있는 방법을 제시하는 것이 필요하다. 이는 관계형 데이터베이스(RDBMS)와 비관계형 데이터베이스(NoSQL 등)를 포괄하는 행정정보시스템에서 생산되는 행정정보데이터세트를 기록관리 대상으로 포함하여 실행방안을 논의하고 있는 현 상황과도 연결된다. 행정정보데이터세트는 업무행위의 결과로 생산된 법령상 중요 전자기록물의 하나로, 공식적인 기록관리 절차를 준수하고, 국가적으로 보존 및 공유·활용되어야 한다. 하지만 공공기관이 운영 중인 행정정보시스템은 약 1.6만개로(2020년 기준) 기관별 업무 특성에 따라 다양한 데이터세트가 운영 및 누적되어 국가적으로 관리에 부담이 되고 있으며 행정·기술적 어려움으로 기록관리 절차 적용이 다소 지연되고 있다. 이에 국가기록원에서는 다수의 행정정보시스템이 구축된 후 약 10년 이상 경과로 인해 대규모 데이터가 누적되어 관리 한계가 발생하였음을 언급하였으며 이를 통해 새로운 기록관리 방안이 요구됨을 알 수 있다. 본 연구에서는 빅데이터 분석 기술 적용을 통해 그러한 가능성을 논의하였고 공공기록의 생산, 분류, 평가, 활용 등의 업무를 효율적으로 처리할 수 있는 방안을 다음 장에서 살펴보았다.

4. 빅데이터로서의 공공기록물 활용 방안 및 프레임워크 구현

4.1 빅데이터 관점에서의 공공기록물 활용 방안

공공기록관리 분야에서 행정업무와 국민의 일상생활의 밀접한 각종 데이터를 신속하고 손쉽게 수집하여 ‘좋은 거버넌스’로 발전할 수 있는 총체적인 분석을 제공할 수 있다면 매우 긍정적인 효과를 기대할 수 있다. 공공기록관에 빅데이터 기술을 적용하는 것의 장점은 다음과 같다. 첫째, 현재보다 정확하게 향후 행정업무의 추이를 예상할 수 있다는 것이다. 공공기관의 행정업무와 관련하여 발생할 수 있는 현상에 대해 점진적으로 쌓인 출처가 다양한 데이터의 분석을 통해 업무의 인과관계와 상호작용을 파악할 수 있다. 이와 같이 업무로 인해 누적된 데이터를 통해 국가 정책, 시장의 변화와 파급효과의 인과관계를 도출하고 실시간 동향 파악을 가능하게 함으로써 좀 더 정확한 업무수행이 가능할 것이다.

둘째, 행정업무가 요청되는 현상에 대한 정확한 진단과 빠른 대응이 가능하다. 일상생활과 행정업무에서 웹 로그, SNS, 블로그, 뉴스 등에서 발생하는 데이터를 수집하여 처리하는 분석 기술의 발전으로 국민이 원하는 정보와 정책의 방향성을 파악하고 빠른 대응을 통해 해결방안을 강구할 수 있다. 따라서 공공기록물이 빅데이터로서 활용될 수 있게 하는 플랫폼 구축이 실행되어야 할 것이다.

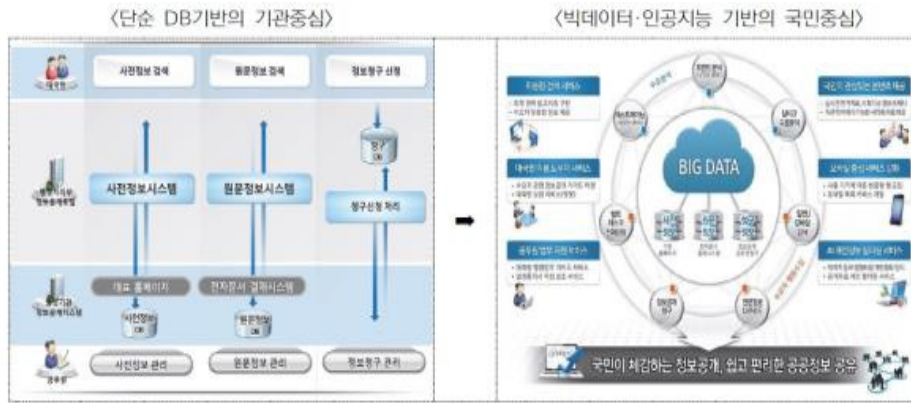
셋째, 다양한 이용자 요구를 만족시키는 서비스 제공이 가능하다. 공공기관에서 생산되어 보관되어진 대량의 정형데이터와 비정형데이터의 패턴분석 또는 형태분석을 통해 행정업무, 개인정보관리, 홍보 등 다양한 목적으로 이를 활용할 수 있다. 또한 비정형데이터를 기반으로 실시간 현상과 예측을 통해 검증하는 자료로 활용할 수 있을 것이다(김대중, 윤서연, 2013).

이러한 전제하에 공공기록관리 분야에서 빅데이터 기술을 적용할 수 있는 활용방안을 제안하고자 한다. 첫째, 공공기관에서는 보유하고 있는 정보를 국민의 정보공개 청구에 의하여 선별적·소극적으로 제공하는 ‘공급자 위주’에서 행정정보를 사전에 공표하는 등 ‘수요자 중심’으로 사전적·적극적으로 공개하고 있다. 국민의 눈높이에 부응하는 선제적·능동적 정보공개를 통해 정부의 책임성 제고 및 공공기관의 업무분담을 완화하고 국민의 알권리 증진 및 행정 투명성 확보에 중점을 두고 추진하고 있다. 정보공개포털(<http://www.open.go.kr>)에서는 2006년에 구축된 DB 기반의 시스템으로 정보량(연간 정보목록 2억건, 원문정보 660만건 등 축적)이 방대함에도, 국민이 원하는 정보를 신속하고 정확하게 찾기 어려운 문제가 있어서 근본적인 개선이 필요하다(행정안전부, 2020).

빅데이터 분석 기술을 통해 정부의 업무정책 수립과정을 정량분석을 통해 정책과 정부 신뢰성·투명성을 제고시켜 예측 가능한 국민 공감형 공공기록물 맞춤형 정보공개서비스의 실현이 가능할 것이다. 빅데이터 분석 기술은 공공기관에서 생산한 기록물을 관리하는 정책과 관련하여 국민들의 니즈(needs)를 파악하여 양방향 소통하여 업무정책에 대한 국민평가를 반영한 선제적 국민만족형 기록관리 정책 추진을 가능하게 한다. 정책의제 발굴과정에서는 공공기록관리 분야 관련 키워드와 소셜미디어의 데이터를 통해 국민이 필요한 자료를 수집 후 국민의 반응을 실시간으로 모니터링하여 현장에 어떤 영향을 미치는지 분석함으로써 공공기관 기록관리의 문제와 국민정서를 파악하여 맞춤형 정책과제를 발굴하고 실효성을 바로 모니터링 한 후 적용시킬 수 있다. 빅데이터 기술을 활용한 검색서비스, 사례정보, 실시간 시각화 서비스와 보안성 강화 등을 추진할 수 있다.

<그림 2>는 행정안전부에서 차세대 정보공개서비스 구축 방향을 제시한 것이다. 좀 더 구체적으로 살펴보면 수요자 입장에서 공공기관에서 기록과 데이터를 자발적으로 공개하는 사전정보공개제도를 국민 트렌드 및 패턴 기반의 맞춤형 정책으로 추진할 수 있다. 공공기관에서 생산되는 기록으로 어떤 이슈가 발생하고 있는지를 빠르게 파악한 후 이를 분석, 예측할 수 있는 다양한 패턴이 빅데이터 분석을 통해 축적되고 있는 현실에서 국민이 원하는 사전정보공표의 내용과 패턴을 파악하여 선제적인 정책수립을 할 수 있다. 예를 들어 행정안전부 사전정보공표의

표준기준의 사례를 살펴보면 국민이 실제 필요로 하고 원하는 정보인 일자리·민생경제 등 국민 실생활 관련 기록정보, 생활안전·건강 등 국민 관심 정보를 빅데이터로 조사·분석하여 <표 7>과 같이 집중 관리하고 제공할 필요가 있다.



<그림 2> 차세대 정보공개서비스 구축 방향
출처: 정보공개포털

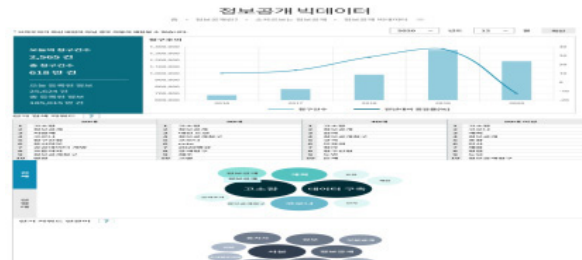
<표 7> 정보공개 빅데이터 중점 분석 분야(행정안전부 사전정보공개 표준기준 사례 정리)

1. 주요 정책 및 업무보고(계획)	6. 각종 공사·용역·물품 발주 및 용역관련 기록
2. 기관장·부기관장 업무추진비	7. 예산·결산 및 재정관련 기록
3. 대규모 예산이 투입되는 사업	8. 각종 평가·조사 등 행정감시 관련 기록
4. 국정감사·조사관련 기록	9. 국외출장 여비 기록
5. 정부업무평가 관련 기록	10. 자치단체 복지비 기록

공공기록관리 분야의 대용량 데이터에 대한 알맞은 분석과 선제적인 예측을 통해 공공기관에서 추진하는 전략이나 정책의 사회적 영향력을 파악하는 것이 필요하므로 이에 빅데이터 분석 기술을 적용할 수 있다. 그 절차는 <그림 3>과 같이 도출하였으며 이를 활용한 시각화 정보는 <그림 4>와 같다. 이와 같이 공공기록을 빅데이터 기술을 적용하여 분석함으로써 변화의 수요를 충족시키고 예측하는 정책을 추진할 수 있다.



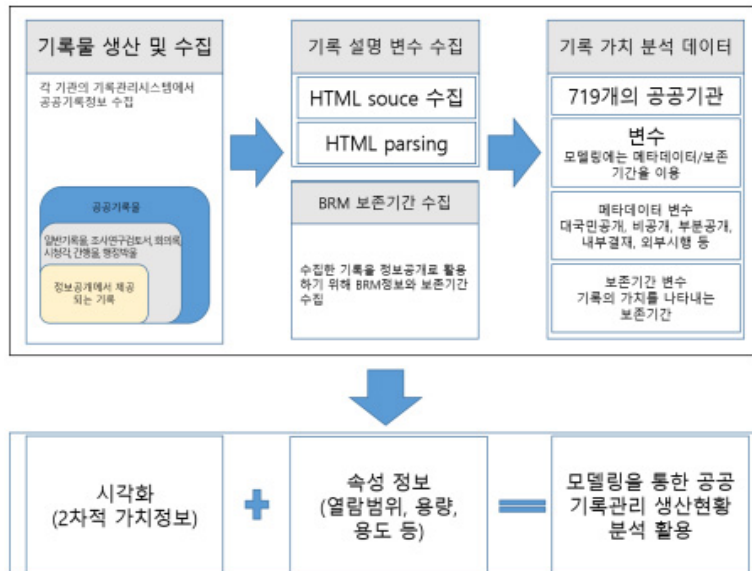
<그림 3> 사전정보공개 대상정보 빅데이터기술 적용 발굴 프로세스(예시)



<그림 4> 정보공개 빅데이터 시각화 모습(예시)

다음으로 빅데이터 분석 기술을 적용하여 「공공기록물 관리에 관한 법률」, 동법 시행령, 동법 시행규칙의 관련 조항을 근거로 운영되고 있는 기록물 생산현황보고의 효율성을 높일 수 있다. 각 공공기관에서 기록물의 생산현황을 통보하는 목적은 『공공기록물법』 제19조 제6항에 따라 ‘공공기록물의 효과적인 수집과 이관’이라고 밝히고 있으며 각 기관에서도 기록관리에 활용하는 기초자료라고 설명하고 있다. 하지만 현재 국가기록원은 각 기관에서 생산한 기록물을 단순하게 기록물 유형별로 생산 변동추이만 파악만 하고 있는 실정이다.

빅데이터 분석 기술을 활용한 공공기록물 생산현황 분석은 각 기관에서 생산되는 다양한 기록물에 대한 데이터 마이닝에 의해 정형·비정형 데이터를 수집하여 분석 후 시계열에 의한 추이와 패턴에 대한 결과를 가지고 문서의 활용도와 기록가치의 정확성과 신뢰성을 높일 수 있다. 예를 들어 ‘연관문서의 정보’, ‘결재자 의견’, ‘공개재분류 이력’, ‘문서요지’, ‘문서열람횟수’ 등 문서가 가지는 보존기간 이외의 많은 정보를 분석하여 문서의 보존기간보다 다양한 방식으로 이를 활용할 수 있을 것이다. 이와 같이 공공기록물과 데이터의 생산현황에 빅데이터 분석 기술을 활용하면 기관에서 활용도가 높은 기록물을 찾아 정보공개에 활용하고, 개별 기록이 가지고 있는 특성을 정밀하게 판단하여 보존기간의 맵핑 등 기록물평가업무와 관리가 용이해질 것이며 조직 내에서 의사 또는 정책결정 근거로 활용할 수 있을 것이다(<그림 5> 참조).



<그림 5> 공공기록의 생산현황 빅데이터 기술 분석 및 피드백 절차도(예시)

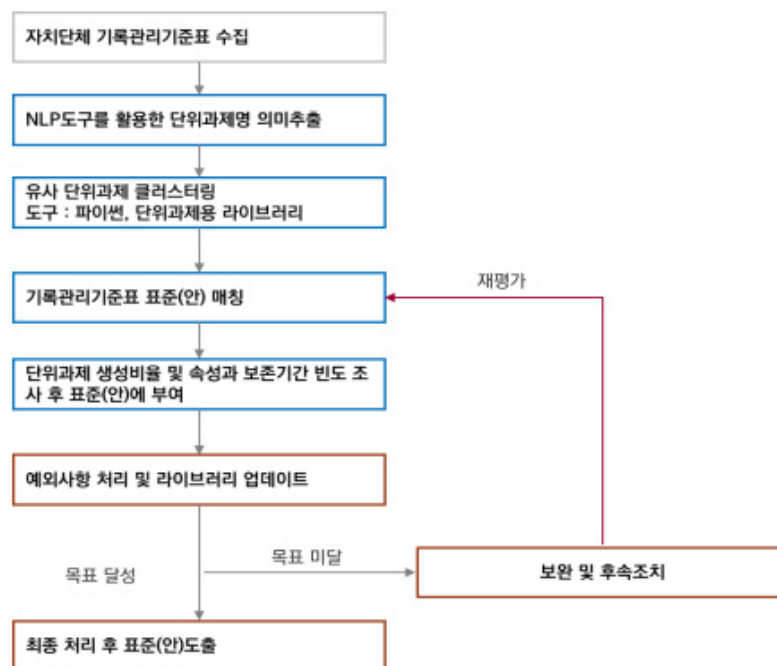
마지막으로 정부기능분류표(BRM)과 기록관리기준표 또는 기록물분류기준표의 단위과제명과 분류체계의 표준화와 보존기간 오적용을 해결하고 공개여부 설정 오류 등 현안 과제의 분석결과에 빅데이터 분석 기술을 활용할 수 있다. 현재 법령에 따라 기록관리시스템에서 관리되는 단위과제는 시도(17개) 단위과제 평균 6천 내지 7천 건, 시군구(228개) 4천 내지 5천 건 내외로 최소 100만 건 이상으로 추산되며 기록관리시스템의 엑셀 함수기능만으로는 체계적 관리의 한계가 있다. 또한 일반적인 공공기관 분류업무는 공공기관에서 수행하는 업무를 6레벨로 분류하여 상위레벨인 1레벨과 2레벨은 정부 BRM 분류체계, 3레벨은 일부 준용하고 하위레벨(4, 5, 6레벨)은 기관 특성을 반영하여 단위과제 이름과 보존기간을 책정하고 있다. 이러한 상황에서 같은 업무라고 하더라도 보존기간이 다르게 책정되는 문제점이 있고 분류 관련 데이터가 무작정 축적되기만 하는 한계가 있다.

예를 들어 전국의 지방자치단체에서 가장 큰 분류단계인 정책영역 12건을 시작하여 업무 간 유사성·독자성을 고려하여 세분화한 업무영역인 5단계 소기능의 현황은 24,988건을 사용하고 있다. 가장 하위단계인 6단계 단위과

제 영역에서 각 기관에서 사용하고 있는 보존기간을 부여하고 있는데 총 400,578건을 사용하고 있다(지방자치단체 기능분류시스템, 2021). 이렇게 대규모의 단위과제들의 명칭과 보존기간에 대한 데이터가 발생하는 환경에서 단위과제의 자연어를 클러스터링(Clustering) 혹은 기계 학습(Machine Learning)이나 딥러닝(Deep Learning)의 방식으로 수집·정리한 뒤 전자문서생산시스템과 기록관리시스템의 사용자들의 접근과 열람 빈도 등과 연계하여 분석하는 과정이 요구된다. 이러한 단위과제명과 보존기간 등을 지능화 기술을 통해 대량으로 수집하여 단위과제의 사용자 접근과 열람 빈도 변수를 활용한 모델링을 통해 단위과제명을 표준화하고 보존기간 오적용과 공개여부 설정 및 비치기록지정 등에 활용할 수 있다.

공공기록관리 분야는 아니지만 참고할 수 있는 사례로 국립중앙도서관에서는 2014년부터 도서관 내외부의 데이터를 체계적으로 수집·분석·활용하여 도서관 운영 및 이용자 서비스 제공방안을 도출하고자 “도서관 빅데이터 분석활용 체계 구축” 사업을 한국과학기술정보원(이하 KISTI)과 같이 수행하고 있다. 해당 프로젝트 수행 결과로 도서관 정보 나무, 도서관 빅데이터 분석 서비스 솔로몬, R기반 빅데이터 분석 툴 Radar, 이용자 도서추천 서비스 책콕콕 등을 도서관 이해관계자에게 제공할 수 있게 되었다. 특히 도서관 빅데이터 분석 서비스 솔로몬은 각종 데이터 분석 내용을 기반으로 도서관 특성에 적합한 수서 및 도서관 운영 정책 수립을 지원하고 있다(도서관 정보나무 홈페이지, 2021).

단위과제의 소기능(5단계)과 단위과제(6단계)와 관련된 다양한 특성을 고려한 데이터를 빅데이터로 분석함으로써 각 단계 간 서로 작용하여 영향을 주는 관계를 파악하여 보존기간을 책정하고 단위과제명을 명명하는 것을 고려해볼 수 있다. 또한 누적된 데이터를 통해 행정의 변화와 파급효과의 인과관계를 도출하여 기록관리 담당자에게 피드백을 제공할 있을 것이다(<그림 6> 참조).



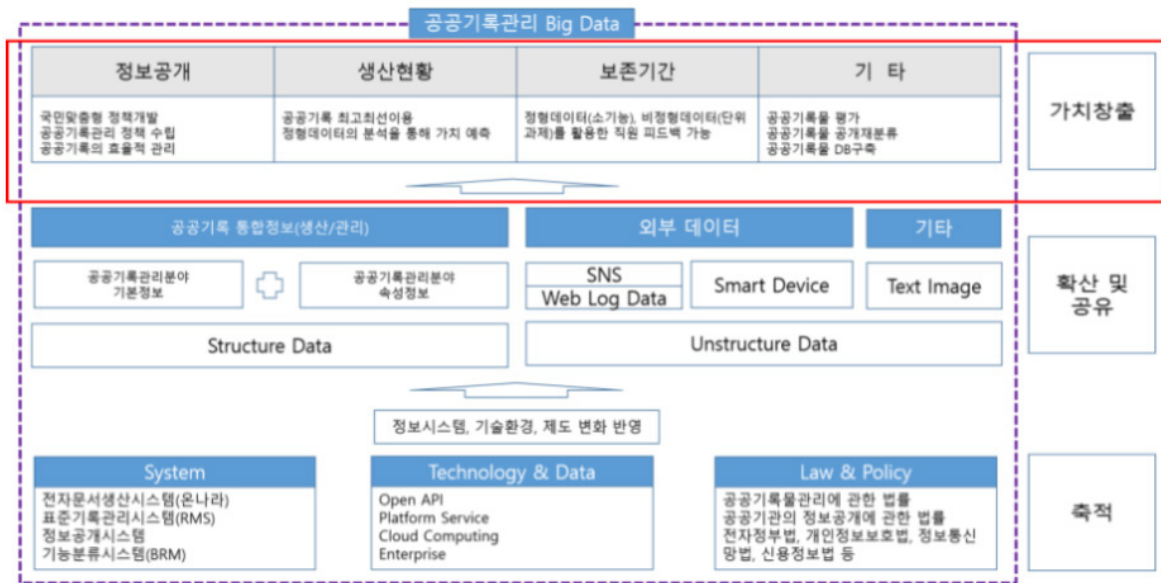
<그림 6> 기록관리기준표 자연어처리를 통한 전처리방법(예시)

4.2 공공기록관리분야의 빅데이터분석 기술을 활용한 프레임워크 구축

공공기록관리 분야에서 업무를 통해 발생하는 기록과 데이터는 시스템으로 관리·보존되며 이러한 정형·비정

형 데이터는 자연스럽게 누적되고 있다. 단순히 정형화된 기록을 수집하고 가공하는 것뿐만 아니라 빅데이터 분석 기술 적용을 통해 새로운 사실을 발견하고 예측 가능하게 하며 향후 업무를 효율적으로 지원하도록 하는 공공기록관리의 변화가 필요하다.

공공기록관리 절차인 생산과 분류, 평가, 활용단계에서 최종적으로 빅데이터 분석기술을 추가하여 업무에 적용한 프레임워크 도식화는 <그림 7>과 같다. 정보시스템과 기술환경, 법과 제도 등이 발전하며 공공기록과 데이터는 생산되며 이후 정형·비정형데이터들로 축적되어 확산 및 공유된다. 이러한 데이터를 가지고 정보공개, 생산현황, 단위과제의 보존기간 등 기록관리업무에 빅데이터 기술을 적용시켜 가치창출을 위한 프레임워크를 구축할 수 있다.



<그림 7> 공공기록분야의 빅데이터를 활용한 프레임워크 도식화

5. 결론 및 시사점

정보기술의 발전으로 이전과는 달리 많은 정보원을 통해 대량의 기록 데이터를 수집하여 관리하는 시대가 도래하면서 초창기의 기록관리업무 환경과는 다른 모습이 전개되고 있다. 공공기관의 기록관에서도 조직의 규모가 커지고 기록 생산량이 급증하면서 새로운 가치를 창출하려는 정부의 요구가 증가되었기 때문에 빅데이터 분석 기술은 공공기관에서 처한 기록관리 문제해결을 위한 최적의 기술이다. 공공기관에서 생산, 축적되는 대량의 기록 및 데이터를 빅데이터 관점에서 수집하고 분석하여 업무정책, 생산현황, 정보공개, 보존기간결정 등에서 효율성을 높일 수가 있을 것이다.

공공기록관리 분야에서 빅데이터 활용을 적용하기 위해서는 다음과 같은 몇 가지를 문제를 해결하여야 할 것이다. 첫째, 기록관리 국제표준이자 한국표준인 KS X ISO 15489는 기록관리과정을 획득, 등록, 분류, 저장, 접근, 추적, 처분 총 7단계로 정의하고 있는데 ‘추적’과 ‘처분’ 사이에 ‘분석’을 넣어 8단계로 하고 ‘처분’ 결정(평가, 폐기, 공개재분류 등)에 보다 신중을 기하기 위한 보완 조치가 가능하도록 해야 한다.

둘째, 많은 양의 기록과 그 안의 데이터 속에 비구조화 되어있고 숨겨져 있는 패턴을 발견할 수 있도록 통합적

사고와 관련이 있는 빅데이터 분석 자격을 갖춘 기록물관리전문요원을 양성하거나 기록관과 기록물전문요원들에 의해 빅데이터 분석기술을 적용할 수 있는 ‘신규 조직 또는 부서’를 신설하여 활발한 연구와 시도가 이루어지도록 해야 한다. 빅데이터를 활용하고 관리하기 위한 빅데이터 운영자 그리고 업무에 적용시켜 분석과 활용을 위한 빅데이터 기록물관리전문요원을 행정안전부 국가기록원 등에서 양성하여야 할 것이다.

셋째, 기존 기록관리단계나 절차를 고수하고 있는 현실과 빅데이터 분석 기술 적용의 기대 사이에 큰 격차가 존재한다. 공공기록에 인공지능을 결합한 자가 학습을 통해 맥락을 분석하고 이를 통해 공공기관의 사회 현상과 환경을 분석하여 국민들의 니즈를 충족시키도록 예측하는 새로운 기록관리 방안에 대한 지속적인 연구가 필요하다.

참고문헌

- 경정의 (2013. 10. 1.). 디지털타임즈 ‘부동산 빅데이터 큐레이터’ 양성하자.
출처: http://www.dt.co.kr/contents.html?article_no=2013100102012251746001
- 국가기록원 (2012). 전자기록생산시스템 기록관리 기능요건(v1.0). 대전: 국가기록원.
- 국가기록원 (2020). 기록물관리지침(공통메뉴얼). 대전: 국가기록원.
- 국가기록원 (2020). 행정정보 데이터세트 기록관리 기준 - 관리기준표 작성 및 이관 규격. 대전: 국가기록원.
- 국가기록원 (2020. 1. 15.). 기록관리 이슈페이퍼 제11호, 기록물 생산현황 분석 결과(2018년 생산분).
출처: <https://www.archives.go.kr/next/data/recordIssue.do>
- 김대중, 윤서연 (2013). 국토정책 선진화를 위한 빅데이터 활용에 관한 기초연구. 국토연구원.
- 김태영, 강주연, 김진, 오효정 (2018). 지능형 기록정보서비스를 위한 선진 기술 현황 분석 및 적용 방안. 한국기록관리학회지, 18(4), 149-182. <https://doi.org/10.14404/JKSARM.2018.18.4.149>
- 김해찬술, 안대진, 임진희, 이혜영 (2017). 기계학습을 이용한 기록 텍스트 자동분류 사례 연구. 정보관리학회지, 34(4), 321-344.
<https://doi.org/10.3743/KOSIM.2017.34.4.321>
- 도서관 정보나루 (2021). 문화빅데이터 센터 소개. 출처: <https://www.data4library.kr/utilizationCase>
- 배정환 (2013). STEEP분석을 통한 국내 테러발생 가능성에 대한 연구. 한국경찰학회보, 15(6), 31-68.
- 안대진, 임진희 (2017). 제4차 산업혁명 기술의 기록관리 적용 방안. 기록학연구, 54, 211-248.
<https://dx.doi.org/10.20923/kjas.2017.54.211>
- 왕호성, 설문원 (2017). 행정정보 데이터세트 기록의 관리방안. 한국기록관리학회지, 17(3), 23-47.
<http://dx.doi.org/10.14404/JKSARM.2017.17.3.023>
- 윤철 (2013). 빅데이터에 기반한 기록정보서비스의 방향. 석사학위논문, 한신대학교 대학원.
- 이정미 (2013). 빅데이터의 이해와 도서관 정보서비스에의 활용. 한국비블리아학회지, 24(4), 53-73.
<https://doi.org/10.14699/kbiblia.2013.24.4.053>
- 임진희 (2021). 공문서의 기계가독형(Machine Readable)전환 방법 제언. 기록학연구, 67, 99-137.
<https://doi.org/10.20923/kjas.2021.67.099>
- 정보공개포털 (2021). 정보공개포털. 출처: <http://www.open.go.kr>
- 정부3.0 추진위원회(2014). 정보3.0 발전계획. 서울: 행정안전부.
- 정부부처합동 (2013). 빅데이터 산업 발전전략. 서울: 미래창조과학부.
- 채승병 (2011). 정보홍수 속에서 금맥 찾기: ‘빅데이터(Big Data)’ 분석과 활용, 삼성경제연구소 SERI 경영 노트, 91, 1-12.
- 최성곤, 오진태, 장중수 (2012). 미래 지식 서비스를 위한 빅데이터 처리의 기술적 요구 사항. 정보처리학회지, 19(2), 5-10.
- 한국정보화진흥원 (2011). 스마트시대의 패러다임변화 전망과 ICT전략. 서울: 전국경제인연합회.
- 한형상, 이창호 (2014). 빅데이터 분석의 현황과 발전전략. KEIT PD Issue Report.

- 행정안전부 (2017). 정부3.0 추진계획. 서울: 행정안전부.
- 행정안전부 (2018). 공공 빅데이터 우수사례집. 서울: 행정안전부.
- 행정안전부 (2020). 정보공개연차보고서. 서울: 행정안전부.
- Bensoussan, Craig (2013). *Analysis Without Paralysis: 12 Tools to Make Better Strategic Decisions*. Pearson Education.
- Codd, E. F. (1983). A relational model of data for large shared data banks. *Communications of the ACM*, 26-1, 64-69.
https://doi.org/10.1007/978-3-642-59412-0_16
- Gartner (2011). Hype Cycle for Emerging Technologies. Available: www.gartner.com/en.
- James Manyika & Michael Chui (2011). Big data: The next frontier for innovation, competition and productivity. McKinsey Global Institute, 33, Available: <https://www.mckinsey.com/>
- John Gantz & David Reinsel (2011). Extracting Value from Chaos, IDC IVIEW, p 6. Available: <http://idcdocserv.com/1142>
- Park and Leydesdorff, Park, H. W., & L. Leydesdorff (2013). Decomposing Social and emantic Networks in Emerging Big Data Research. *Journal of Informetrics*, 7, 756-765.
- Paula Dootson, Mary Tate, Kevin C. Desouza, Peter Townson (2020). Transforming public records management: Six key insights, *Journal of the Association for Information Science and Technology*, 72(5), 643-648.
<https://doi.org/10.1002/asi.24429>
- SAS (2012). SAS Homepage. Available: <https://www.sas.com>
- World Economic Forum (2012). Big data, big impact: New possibilities for international development. Available: https://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

• 국문 참고자료의 영어 표기

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

- An, Dae-jin & Yim, Jin-hee (2017), Application of 4th industrial revolution technology to records managemen. *The Korean Journal of Archival Studies*, 54, 211-248. <https://doi.org/10.20923/kjas.2017.54.211>
- Bae, Jung-hwan (2013). A study on the possibility of terror in South Korea through STEEP analysis. *Korean Association of Police Science*, 15(6), 31-68.
- Chae, Seung Byung (2011). Finding gold beer in the flood of information: “big data” analysis and utilization. *Samsung Economic Research Institute SERI Management Note*, 91, 1-12.
- Choi, Sunggon, Oh, Jintae, & Jang, Jongsoo (2012). Technical requirements for processing big data for future knowledge services. *Korea Information Processing Society*, 19(2), 5-10.
- Data for Library (2021). Introduce the Cultural Big Data Center. <https://www.data4library.kr/utilizationCase>
- Government 3.0 Promotion Committee (2014). *Information 3.0 Development Plan*, Seoul: Ministry of the Interior and Safety.
- Han, Hyung-sang & Lee, Chang-ho (2014). *The Current Status and Development Strategy of Big Data Analysis*, KEIT PD Issue Report.
- Information Open Portal (2021). Information Open Portal. <http://www.open.go.kr>
- Joint Government Department (2013). *Big Data Industry Development Strategy*. Seoul: Ministry of Science and ICT.
- Kim, Daejong & Yoon, Seoyeon (2013). *Basic Research on the Utilization of Big Data for the Advancement of National Territory Policy*. Korea Research Institute for Human Settlements.
- Kim, Hae Chan Sol, An, Dae Jin, Yim, Jin Hee, & Rieh, Hae-Young (2017). A study on automatic classification of record text using machine learning. *Journal of the Korean Society for Information Management*, 34(4), 321-344.
<https://doi.org/10.3743/KOSIM.2017.34.4.321>
- Kim, Tae-Young, Gang, Ju-Yeon, Kim, Geon, & Oh, Hyo-Jung (2018). A study on the current status and application strategies for intelligent archival information services. *Journal of Korean Society of Archives and Records Management*, 18(4),

- 149-182. <https://doi.org/10.14404/JKSARM.2018.18.4.149>
- Kyung, Jung Ik (2013. 10. 1.). Let's train "Real Estate Big Data Curator". Digital Times. Available: http://www.dt.co.kr/contents.html?article_no=2013100102012251746001
- Lee, Jeong-Mee (2013). Understanding big data and utilizing its analysis into library and information services. Journal of the Korean Biblia Society for Library and Information Science, 24(4), 53-73. <https://doi.org/10.14699/kbiblia.2013.24.4.053>
- Ministry of the Interior and Safety (2017). Government 3.0 Promotion Plan, Seoul: Ministry of the Interior and Safety.
- Ministry of the Interior and Safety (2018). Best Practices for Public Big Data, Seoul: Ministry of the Interior and Safety.
- Ministry of the Interior and Safety (2020). Annual Report on Information Disclosure, Seoul: Ministry of the Interior and Safety.
- National Archives of Korea (2012). Requirements for the record management function of the electronic record production system (v1.0), Daejeon: National Archives of Korea.
- National Archives of Korea (2020). Administrative information data set record management standards - management standard table preparation and transfer specification, Daejeon: National Archives.
- National Archives of Korea (2020). Record management guidelines (common manual), Daejeon: National Archives.
- National Archives of Korea (2020. 1. 15.). Record Management Issue Paper No. 11, Results of Analysis of Record Production Status (Production in 2018). Available: <https://www.archives.go.kr/next/data/recordIssue.do>
- National Information Society Agency (2011). Paradigm change prospects and ICT strategies in the smart era. Seoul: The Federation of Korean Industries.
- Wang, Ho-sung & Seol, Moon-won (2017), A study on managing dataset records in government information systems. 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 (2021), Suggestions on how to convert official documents to machine readable. The Korean Journal of Archival Studies, 67, 99-137. <https://doi.org/10.20923/kjas.2021.67.099>
- Yoon, Cheol (2013), Archival Reference Services Based on Big Data, Master's thesis, Hanshin University.