한국기록관리학회지
25(3), 331-359, 2025.8
https://doi.org/10.14404/JKSARM.2025.25.3.331
pISSN 1598-1487 eISSN 2671-7247

JKSARM
Journal of Korean Society of Archives and Records Management

개인정보 보호를 위한 아카이브 정책 설계 방향*

A Study on Policy Design Framework for Personal Information Protection in Archives*

이경남(Kyungnam Lee)

E-mail: coarchivist@gmail.com

한신대학교 대학원 기록관리학과 강사

논문접수 2025.07.21
최초심사 2025.07.24
게재확정 2025.08.22

ORCID

Kyungnam Lee
https://orcid.org/0000-0001-6357-4632

초 록

이 연구는 아카이브의 고유한 특성과 기록의 맥락 정보를 반영한 개인정보 보호 정책의 필요성을 제기하고, 이를 구현하기 위한 정책 설계 방향과 제도적 기반을 모색하였다. 개인정보 보호법과 공공기록물법 간의 제도적 충돌 지점을 분석하고, 기존 식별자 삭제 중심의 보호 조치만으로는 정보주체의 권리 보호에 한계가 있음을 지적하였다. 특히 기록은 집합체를 구성하는 특성에 따라 맥락에 따른 재식별 위험성을 고려해야 함을 강조하였다. 연구 결과로 아카이브가 보유한 맥락 정보를 활용한 개인정보 민감도 검토와 재식별 위험성 평가 체계 구축의 필요성을 제시하고, 아카이브의 특성을 고려한 윤리적 판단 기준과 다학제간 전문가 협의체의 구성, 전담조직의 운영, 관련 규정 정비와 기술 도입의 통합적 접근이 필요함을 제안하였다. 본 연구는 향후 아카이브에서 개인정보 보호 정책의 실효성 있는 체계 구축을 위한 기반을 제공하고자 한다.

ABSTRACT

This study highlights the necessity of developing a personal information protection policy that reflects the unique characteristics of archives and the contextual information embedded in records. It explores the policy design directions and institutional foundations necessary for implementing such a policy. Moreover, the study analyzes institutional conflicts between the Personal Information Protection Act and the Public Records Management Act and identifies the limitations of protection measures that rely solely on the deletion of identifiers to safeguard data subjects' rights. In particular, it emphasizes the need to consider reidentification risks that emerge from contextual relationships resulting from the aggregative character of records.

The findings of this study suggest that archives should utilize contextual information to develop a framework for sensitivity reviews and reidentification risk evaluation. Furthermore, an integrated policy framework is required—one that reflects the unique characteristics of archives and incorporates ethical decision-making, the formation of a multidisciplinary expert advisory committee, the establishment of a dedicated organizational unit, regulatory refinement, and the integration of relevant technical requirements. By outlining both institutional and technical strategies, this study aims to provide a foundational framework for implementing effective and sustainable personal information protection policies within archival environments.

■ 이 논문은 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5B5A17049873)

Keywords: 개인정보 보호, 재식별 위험, 맥락 정보, 아카이브의 개인정보 보호 정책, 윤리적 고려
Personal information protection, Re-identification risk, Contextual information, Personal information protection policy in archives, Ethical considerations

1. 서론

1.1 연구 배경 및 목적

디지털 사회에서는 정보량이 기하급수적으로 증가함과 동시에 다양한 정보들이 결합되어 새로운 정보가 만들어지기도 한다. 그러므로 개인정보의 보호 문제는 기술적 비공개 조치를 넘어 헌법으로 보장되는 기본권으로서의 개인정보 자기결정권을 보장할 수 있어야 한다. 이러한 관점에서 개인정보 보호를 수동적 특성이나 상태가 아닌 주체성을 갖는 능력이며, 개인 데이터의 유통까지 통제할 수 있는 능력으로 정의하기도 한다(Windon & Youngblood, 2024, 206). 그러나 빅데이터 환경에서 개인정보 보호는 더욱 취약해졌으며 보호를 위해 가명처리를 했더라도 재식별 가능성은 여전히 존재한다(이양복, 2020). 아카이브 역시 디지털 기록의 양적 증가와 외부 데이터와의 결합 가능성을 고려했을 때, 소장 기록의 개인정보 보호 정책을 수립하는 것은 중요한 과제로 대두된다고 할 수 있다.

공공데이터 개방 확대와 데이터기반 행정 활성화가 정부 주요 정책으로 추진됨에 따라 공공기록의 활용 가치 역시 주목받고 있으며(임진희, 2021), 공공기록과 행정 데이터의 결합 및 활용 가능성 역시 커지고 있다. 아카이브가 소장한 기록은 민감한 정보를 포함하는 경우가 많다. 의료 분야 아카이브, 범죄 관련 기관의 아카이브와 같이 특수 목적의 업무가 수행되는 기관의 아카이브에는 건강이나 유전, 인종, 범죄 경력 자료와 같은 개인정보 보호법에서 정의한 민감정보가 수집·저장되고 보존되기 때문에 이러한 기록의 공개와 활용을 처리하기 위한 추가적인 조치나 규정이 필요하다.

또한 아카이브가 소장한 기록의 공개여부를 판단하는 주체로서 아키비스트는 윤리적이고 전문적인 판단을 해야 하며, 이를 수행하기 위한 명확한 업무 지침이 있어야 한다. 기록의 생애주기에 따라 개인정보 보호의 고려사항은 다를 수 있다. 기록이 활발히 활용되는 현용단계와 일정기간 이후 아카이브로 이관된 기록을 비교했을 때, 이관 이후에는 개인정보 보호 측면에서 요구되는 긴급한 민감성은 상실한 상태의 기록이 많을 수 있다. 반면에 아키비스트가 직접 기록을 수집하는 경우 개인정보 보호의 책임은 더욱 확장되기도 한다(Windon & Youngblood, 2024, 209). 아카이브 내 기록의 공개여부를 판단하는 것은 이용자뿐 아니라 정보주체의 권한에도 직접적인 영향을 미칠 수 있으므로 현재의 법적 기준을 실제 기록에 적용하기 위한 상세 가이드라인이나 실무 지침 등의 제도적 기반이 마련될 필요가 있다.

기존의 기술적 보호조치는 대체로 개인식별정보(Personally Identifiable Information)나 키워드 기반으로 작동한다. 비전자적 형태의 기록은 광학 문자 인식(Optical character recognition) 기술을 통해 디지털화하여 기록 내 정보를 식별하고 추출할 수 있도록 변환하고, 개인정보를 식별하여 마스킹 처리 등을 통해 비식별화하는 보호 솔루션이 활용되고 있다. 그러나 이러한 보호 기술에는 기록이 생산된 맥락에 대한 고려사항은 반영되지 않는다. 기록은 생산 기관이나 시기, 유형, 다른 기록과의 관계와 같은 기록 맥락에 따라 의미가 형성되고, 개인정보의 민감도나 재식별 위험성 평가 역시 기록의 맥락에 영향을 받는다. 기록 내에 직접적인 식별 정보를 삭제하거나 가명처리했더라도 다른 공개된 정보와 결합하여 개인을 식별해내거나, 기록 내에 직접 명시되지 않았더라도 생산 맥락이나 다른 기록과 연결함으로써 민감한 정보가 도출될 수 있다. 따라서 기록의 맥락을 고려하지 않은 보호 조치는 개인정보를 보호하는 데 충분하지 않다고 할 수 있다.

본 연구는 기록의 특성을 반영한 아카이브의 개인정보 보호 정책 필요성을 고찰하고, 아카이브가 이미 보유하고 있는 맥락 정보나 메타데이터를 활용하여 기록 맥락을 고려한 개인정보 식별 체계와 민감도를 검토하는 체계 설계의 기반을 마련하는 데 목적을 둔다. 기록을 생산한 조직이나 업무적 맥락 정보, 매체의 특성 정보 등은 관리를 위한 행정적 속성이 아니라 기록을 구성하는 핵심적인 구조이다. 이러한 정보를 개인정보 보호 기술과 연계하여

https://doi.org/10.14404/JKSARM.2025.25.3.331

기존의 개인정보 재식별 위험 한계를 보완하고 민감도 분류 등의 사전적 조치를 취함으로써 실질적인 보호 체계를 강화할 수 있을 것이다.

1.2 선행연구

개인정보 보호 기술에 관한 선행연구는 정책 개선 방안과 최신 기술의 도입 과정에서 제기되는 문제점을 다룬 연구가 많다. 개인정보 보호 법제 측면의 연구들은 법 개정 과정에서 제기된 주요 쟁점과 한계를 분석하고 개선 방안을 제시하는 데 초점을 두고 있다. 권은정(2020)은 2020년 개정된 데이터 3법에 대해 보완이 필요함을 지적하고 공공데이터 개방의 리스크를 관리하고 안정성을 확보하기 위해 데이터 거버넌스 체계 구축의 필요성을 강조하였다. 이규철(2013)은 빅데이터 환경에서 법적 고려사항과 빅데이터 기술의 정착을 위한 개인정보 보호 법규 방안을 제시하였으며, 최원상 외(2019)는 4차 산업혁명 기술 환경에서 개인정보 보호를 위해 해외 법제를 분석하여 시사점을 도출하였다. 2023년 일부 개정된 개인정보 보호법에 대해 정혜영(2024)은 정보주체의 권리를 강화하는 한편, 사전 동의 제도의 경직성을 보완함으로써 변화하는 기술 환경에서 정보의 활용을 확대할 수 있는 기반을 마련하였다고 평가하였다. 이 외에 천지영과 노건태(2020)는 가명처리 후 공개된 공공데이터로부터 정보 주체를 재식별한 실증적 사례 연구를 진행하여, 이러한 데이터의 프라이버시 침해 우려를 제기하고 데이터의 신뢰성에도 문제를 야기할 수 있음을 밝혔다.

최근 정보통신기술의 발달에 따른 개인정보 보호 과제와 기술적 방안을 모색한 연구에도 주목할 필요가 있다. 손영화(2022)는 인터넷 사용과 클라우드 컴퓨팅의 확대, 빅데이터와 사물인터넷의 활용, 전자 투표 및 보안 감시 시스템 사용에 따른 개인정보 보호 과제를 제기하였다. 빅데이터 활용과 관련하여, 엄수현 외(2018)는 개인정보 비식별화 기술 동향을 정리하였으며, 김승환과 전성해(2019)는 최적 절단값을 이용한 모델을 통해 익명성을 보장하면서도 데이터 가치를 유지할 수 있는 비식별화 방안을 논의하였다. 인공지능 개발을 위한 학습 데이터의 경우 김병필(2023)의 연구에서 재식별 위험을 낮추고 활용성을 높이기 위해 차분 프라이버시 기술 등의 새로운 기술적 접근 도입을 제안하였다. 또한 블록체인 기술과 관련하여, 정진명(2019)은 블록체인 유형별로 블록체인에 저장된 정보의 개인정보 해당 여부와 개인정보 처리 책임, 정보주체의 권리 보호에 관한 문제를 분석하였다.

기록관리기관의 개인정보 보호 문제는 이용자의 열람 요청이나 정보공개 청구에 대응하거나 기록의 공개재분류 과정에서 부각된다. 재분류 제도와 관련한 연구에는 재분류 대상 기록의 범위나 재분류 주기, 재분류 절차에 대한 개선을 지적한 연구(임희연, 2016)와 기록관 단위의 기록 공개업무 개선을 위해서는 공개관리 제도의 교육 강화와 실무에 필요한 시스템이나 제도적 지원을 지적한 연구(윤연화, 이은주, 2021)가 있다.

비공개 세부기준에 대한 연구로는 비공개 세부기준 개발을 위한 사업 진행의 측면에서 고려사항을 정리한 연구(황진현 외, 2021)를 비롯하여, 비공개 대상정보 세부기준 수립의 실효성을 확보하기 위한 법제화 정비 및 세부기준의 현행화를 위한 정기적 점검, 제공방식 개선을 제안한 연구(김유승, 2023), 그리고 국회의 비공개 세부기준 정비와 함께 전담조직의 필요성과 비공개 정보 유형의 세분화를 도출한 연구가 있다(김유승, 2022). 이 외에도 정보공개 제도와 관련하여 사전에 원문정보를 공개하는 서비스상에서 개인정보가 노출된 사례 분석을 통해 제도 개선을 다룬 연구(안혜미, 2019)에서는 실무에 적용하기 위한 개인정보 보호 지침이 명확하게 재설계 될 필요가 있음을 지적하였고, 필터링 기술 역시 개선이 필요하다는 결과를 제시하였다. 권미현(2019)은 국가기록원에서의 공개재분류 업무 절차를 분석하고, 서식 개선, 업무 절차의 간소화, 시스템 지원 등 단계별 문제점에 대한 실질적인 개선방안을 제안하였다.

국가기록원은 2021년 비공개정보 필터링 및 마스킹 기술 적용 가능성을 검토하기 위한 연구를 수행하였다(국

https://doi.org/10.14404/JKSARM.2025.25.3.331

가기록원, 2021). 이 연구에서는 정형화된 개인정보 패턴을 자동으로 식별하고 마스킹하는 솔루션을 테스트하고 학습데이터 구축 방안을 모색하였다. 그러나 패턴화된 개인정보의 기술 조치에 한정되어 있어, 맥락 기반의 판단이나 제도적 측면은 충분히 다루지 못한 한계가 있다. 최근 국가기록원에서는 공개재분류 업무에 AI기반 개인정보 비식별화 서비스 모델을 적용하기 위한 연구가 진행중에 있어, 향후 발표될 최신 연구 결과를 주목할 필요가 있다(국가기록원, 2025a).

이처럼 기록관리 측면에서의 개인정보 보호 문제는 기록의 공개재분류 업무의 실무적 절차의 개선방안이나 정보공개 제도 운영을 위한 비공개 세부기준 수립의 관점에서 연구가 진행되어 왔다. 최근 비공개 대상 정보를 비식별 처리하기 위한 기술 개발이 시도되고는 있으나, 패턴화된 개인정보 유형에 한정되어 있다는 한계를 갖는다. 아직까지 기록의 생산 맥락에 따라 개인정보 보호 수준이 달라지거나 비식별 처리된 개인정보가 맥락에 따라 재식별 될 수 있는 위험성에 대한 기록관리 측면의 연구는 이루어지지 않았음을 확인할 수 있다.

1.3 연구 범위 및 방법

본 연구는 기록의 맥락정보를 고려한 개인정보 보호 기준 설정의 필요성을 고찰하고자 하며, 이를 위해 다음의 세 가지 주요 내용을 분석하였다.

첫째, 문헌검토를 통해 개인정보 보호와 관련된 법·제도적 배경을 정리하고, 특히 공공기록관리와 개인정보 보호 간의 상이한 규범적 요구를 분석하였다. 최근 공공데이터 개방 확대와 빅데이터 분석·활용이 강조되는 정책 환경에서 개인정보 보호 침해에 대한 우려가 증가하고 있으며, 이는 아카이브에서 관리되는 공공기록의 공개 및 활용에 있어 다양한 쟁점을 야기한다. 이러한 상황은 명확하고 세부적인 개인정보 보호 기준 마련의 필요성을 부각시키고, 특히 기준 적용의 과정에서 아키비스트의 전문성과 책임성을 더욱 강조한다. 본 연구는 현행 법제 체계를 분석하여 한계를 도출하고, 아카이브의 특성을 반영한 개인정보 보호 정책 수립의 필요성과 시사점을 제시하였다.

둘째, 기록이 갖는 맥락적 특성을 반영하여 아카이브의 개인정보 보호 정책 수립의 필요성을 검토하였다. 이를 위해 개인정보 보호 기술 관련 국제표준을 살펴보고, 현재의 개인정보 보호가 비식별화에 중점을 두고 있기 때문에 발생하는 한계를 분석하였다. 이와 함께 기록이 갖는 맥락적 특성이 재식별 가능성에 미치는 영향을 분석하기 위해 수행되었던 사례 연구를 고찰하였다. 구체적으로는 기록의 생산 기관, 생산 시기, 업무 특성, 관련된 다른 기록과의 연결 관계 등의 맥락 정보가 개인정보의 민감도와 재식별 가능성에 미치는 영향을 파악하였다. 그리고 이러한 맥락 정보를 현재 개발되고 있는 개인정보 보호 기술과 통합하여 개인정보 보호 기준을 수립하고 차등화된 보호 수준을 개발하는 전략의 실현 가능성을 탐색하였다.

마지막으로, 아카이브 현장에서 개인정보 보호 정책의 실효성을 제고하기 위한 정책적·기술적 요구사항을 제안하였다. 아카이브 업무 수행 과정에서 개인정보 보호 조치를 결정해야 하는 주요 지점마다 요구되는 자동화 기술 개발 및 정책 설계의 방향성을 구체화하였다. 또한 기록의 공개 여부 결정이나 기록의 유형에 따른 민감도 분류 기준 등에 요구되는 윤리적 고려 요소를 도출하고, 이를 효과적으로 운영하기 위한 제도적 기반과 실천 과제를 제시하였다.

https://doi.org/10.14404/JKSARM.2025.25.3.331

2. 아카이브의 개인정보 보호 정책 개발 필요성

2.1 개인정보 보호법 체계 현황

개인정보는 살아 있는 개인에 관한 정보로서 개인을 알아볼 수 있는 정보로 정의된다. 정보통신기술의 발전과 디지털 사회로의 전환에 따라 개인정보 보호의 중요성은 더욱 강조되고 있다. 정보의 수집과 저장, 처리과정이 전산화되고, 전산망의 연결을 통해 정보의 결합·통합 활용이 증대되고 있기 때문이다. 개인정보의 부당사용 및 무단유출 우려에 따라 개인정보 보호를 위한 입법 논의는 1990년대 들어서며 본격적으로 시작되었다. 「통신비밀보호법」(1994년 시행)과 「공공기관의개인정보보호에관한법률」(1995년 시행), 「신용정보의이용및보호에관한법률」(1995년 시행), 「정보통신망이용촉진등에관한법률」(1999년 전부개정 시행) 등이 제정 또는 전부개정되어 시행되었고, 각각의 개별법령 내에 정보보호 관련 규정이 포함되었다. 그러나 1994년 제정되어 이듬해부터 시행되었던 「공공기관의개인정보보호에관한법률」은 국가행정기관·지방자치단체·기타 공공단체 중 대통령령이 정하는 기관을 적용대상으로 한정하였고, 민간영역의 개인정보 보호에 대해서는 「신용정보의이용및보호에관한법률」 및 「정보통신망이용촉진등에관한법률」 등과 같은 개별 법령의 일부 조항에서 규정하고 있어 규제 내용과 대상 역시 각 법령과 부처별로 다르게 운영되고 있었다.

「개인정보 보호법」(이하 개인정보 보호법)의 제정 이유에서 밝히고 있듯이, 개인정보 보호의 사각지대의 발생과 개인정보의 유출·오용·남용 사례가 발생함에 따라 국가사회 전반을 규율하는 통일된 개인정보 보호 원칙과 처리기준 적용의 필요성이 대두되며, 마침내 2011년 개인정보 보호법이 제정·시행되었다. 개인정보 보호법(법률 제10465호)은 공공과 민간을 포괄하는 일반법으로서 개인정보처리자에게 일관된 규정을 적용하도록 하였고, 개인정보 보호에 관한 심의·의결 기구로서 개인정보 보호위원회를 기존 국무총리 소속에서 대통령 소속으로 격상하였다. 이로써 관리주체가 행정안전부의 개인정보보호과에서 개인정보보호위원회로 변경되었다. 이와 함께 개인정보의 수집·이용·제공 등의 단계별 보호기준을 제시하고, 정보주체의 권리를 강화하였으며, 고유식별정보의 개념을 처음으로 규정하여 처리 및 관리를 엄격히 제한하였다. 또한 영상정보처리기기의 설치 제한 근거를 마련하였으며, 개인정보 영향평가제도 도입 등을 주요 내용으로 제정하였다. 이 개인정보 보호법이 제정되며 「공공기관의개인정보보호에관한법률」은 폐지되었다.

2011년 개인정보 보호법이 제정된 이후에도 이 법은 디지털 데이터의 활용 활성화 등의 환경 변화 요구에 따라 지속적으로 개정 변경되어 왔다. 특히 개인정보 보호법을 비롯한 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」과 「신용정보의 이용 및 보호에 관한 법률」, 즉 데이터 3법의 개정은 4차 산업혁명 시대 신산업 육성을 위한 핵심 자원으로서 데이터 이용 활성화를 위한 규제 혁신과 개인정보 보호 체계 정비를 위한 입법 조치로 발의되었고 2020년 1월 국회를 통과하며 시행되었다(과학기술정보통신부, 2020). 이 데이터 3법의 개정 과정에서 불거졌던 개인정보 침해에 대한 우려와 논란은 지속되고 있다.[1] 한편 데이터 3법의 주요 틀은 유럽연합의 일반 개인정보 보호법(EU General Data Protection Regulation, 이하 GDPR)의 영향을 받아 개인정보 보호 원칙, 정보주체의 권리 강화 및 독립적인 감독기구 등의 설치에 관한 내용을 반영하고 있다. EU 집행위원회는 한국의 개인정보 보호 법체계가 EU GDPR과 동등한 수준임을 확인하는 개인정보 보호 적정성 결정을 2021년 말에 채택하였다(개인정보보호위원회, 2021).

개인정보 보호법에서 규정한 개인정보 처리 기본원칙(법 제3조)은 다음과 같다. 첫째, 개인정보처리자[2]는 구체

1) 한국소비자연맹, 경제정의실천시민연합, 소비자시민모임, 진보네트워크, 참여연대 (2023.7.23.). 정보주체 권리 외면, 개인정보 무분별한 유통 조장 마이데이터 사업 중단하라. 참여연대.

2) "개인정보처리자"란 업무를 목적으로 개인정보파일을 운용하기 위하여 스스로 또는 다른 사람을 통하여 개인정보를 처리하는 공공기관, 법인, 단체 및 개인 등을 말한다(개인정보 보호법 제2조 제5호).

https://doi.org/10.14404/JKSARM.2025.25.3.331

적이고 명확한 처리 목적에 따라 개인정보를 수집해야 하고, 그 목적에 필요한 최소한의 개인정보를 수집해야 한다. 둘째, 개인정보 처리 목적 범위 내에서만 적합하게 처리할 수 있다. 셋째, 개인정보 처리 목적에 필요한 범위 내에서 개인정보의 정확성과 최신성을 유지하고, 개인정보가 변경 및 훼손되지 않도록 완전하게 보장해야 한다. 넷째, 정보주체의 권리 침해 가능성 및 위험을 고려하여 적절한 보호조치를 취해 개인정보를 안전하게 관리해야 한다. 다섯째, 개인정보 처리방침 등 개인정보 처리에 관한 사항을 공개하고, 정보주체의 열람청구권 등의 권리를 보장하는 합리적 절차를 마련해야 한다. 여섯째, 정보주체의 사생활 침해를 최소하는 방법으로 개인정보를 처리해야 한다. 일곱째, 익명 및 가명으로 처리하여도 개인정보 수집목적을 달성하는 것이 가능하다면, 가능한 한 익명으로 처리하고, 익명처리로 목적을 달성할 수 없으면 가명으로 처리될 수 있도록 해야 한다. 이때 가명처리는 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보 없이는 특정 개인을 알아볼 수 없도록 처리하는 것을 말하며, 익명처리는 개인정보 일부 또는 전부를 삭제하거나 다른 정보로 대체하여 개인을 식별할 수 없도록하는 것으로 정의된다. 여덟째, 개인정보처리자는 개인정보 보호법에 따른 책임과 의무를 준수하고 정보주체의 신뢰를 확보하기 위해 노력해야 한다는 원칙을 제시하고 있다(개인정보보호위원회, 2024b, 5-6).

이와 더불어 개인의 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖의 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보를 민감정보로 정의하고 법에서 정하는 예외적 허용에 해당하지 않으면 원칙적으로 개인정보 처리를 금지하고 있다(법 제23조). 또한 개인을 고유하게 구별하기 위해 부여된 식별정보인 고유식별정보의 처리 역시 원칙적으로 처리를 금지하고, 예외적으로 처리를 허용하는 경우에는 안전성 확보 조치를 취하도록 하였다(법 제24조). 이 외에 14세 미만 아동의 개인정보 처리를 위해서는 법정대리인의 동의를 의무화하는(법 제22조의2) 조항과 AI 기술을 적용한 자동화된 시스템 등에서 개인정보를 처리하여 이루어지는 결정인 자동화된 결정에 대한 정보주체의 거부 또는 결정에 대한 설명을 요구하는 등의 정보주체 권리를 명시하는(법 제37조의2) 내용을 신설하여 2023년 개정하였다. 이는 EU GDPR와 동등한 수준의 개인정보 보호 법제를 마련하고 EU 개인정보 보호 적정성 결정 평가를 얻기 위한 개정 과정에서 반영된 것으로 볼 수 있다(박노형, 김효권, 2022, 367).

2.2 공공기록관리와 개인정보 보호 쟁점

공공기관에서는 「공공기록물 관리에 관한 법률」(이하 공공기록물법)에 따라 업무와 관련하여 생산하거나 접수한 기록의 공개 여부를 분류하여 관리해야 하며(법 제19조), 관할 기록물관리기관으로 이관할 때 공개 여부를 재분류하고 비공개로 재분류된 기록은 재분류된 연도의 다음 해부터 5년마다 공개 여부를 재분류해야 한다(법 제35조 제1항, 제2항). 또한 생산 후 30년이 경과한 기록은 공개하는 것을 원칙으로 한다(법 제35조 제3항). 영구기록물관리기관에서는 생산 후 30년이 경과한 기록의 공개원칙에도 불구하고 비공개로 재분류된 기록에 대한 현황을 공고하도록 되어있는데(법 제35조 제4항), 기록의 비공개 유형 중 상당 부분은 다음의 <표 1>과 같이 개인정보에 해당하는 사유로 비공개되고 있다.[3)]

3) <표 1>은 생산 후 30년이 경과한 기록 중 비공개 기간이 연장된 기록을 포함하여, 이관 후 5년마다 공개 여부를 재분류 한 기록의 수량이 합산된 결과임. 비공개 유형은 「공공기관의 정보공개에 관한 법률」 제9조 제1항의 각 호에 따른 것이며, 부분공개를 포함하고 있고, 비공개 사유가 2개 이상인 경우 중복하여 산정함.

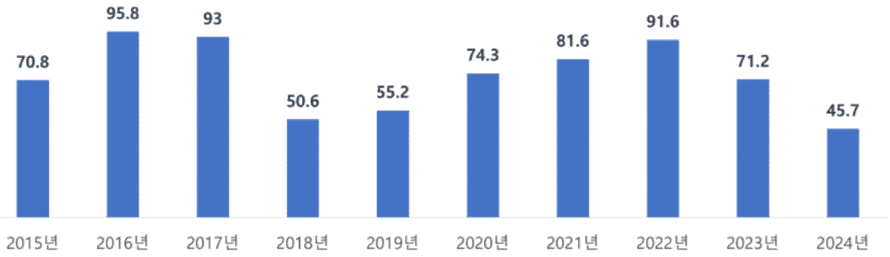
https://doi.org/10.14404/JKSARM.2025.25.3.331

<표 1> 연도별 기록물 공개재분류 결과(2015년~2024년) (단위: 건)

재분류 연도	재분류 된 비공개기록 총 수량	비공개 유형							
		법령상 비밀·비공개 (제1호)	국방 등 국익침해 (제2호)	국민의 생명 등 공익침해 (제3호)	재판 관련 정보 등 (제4호)	공정한 업무수행 지장 등 (제5호)	개인 사생활 침해 (제6호)	법인 등 영업상 비밀침해 (제7호)	특정인의 이익·불이익 (제8호)
2024년	4,379,359	0	151	160,329	2,205,545	0	2,001,863	11,471	0
2023년	51,538	0	0	0	14,868	0	36,670	0	0
2022년	32,052	132	373	1,836	322	0	29,348	40	1
2021년	19,915	387	73	811	1,739	15	16,256	632	2
2020년	339,431	2,995	17,765	64,037	529	522	252,084	1,494	5
2019년	3,096,716	2,062	40,048	651,759	493,365	73,329	1,706,806	125,806	541
2018년[4]	3,681,792	6,851	30,909	777,954	491,787	37,414	1,861,173	141,554	2,064
2017년	1,211,316	1,161	50,199	9,939	21,260	2,623	1,125,951	169	14
2016년	795,837	28	4,436	10,416	18,648	0	762,309	0	0
2015년	140,697	46	9,660	11,952	17,368	0	99,645	2,026	0

출처: 국가기록원 공고 "공개재분류 결과 비공개 기록물 유형별 현황"을 연도별 집계 재구성

재분류 대상 기록의 특성에 따라 「공공기관의 정보공개에 관한 법률」(이하 정보공개법) 제9조의 각 호에 해당하는 비공개 유형은 달라지겠으나, 전반적으로 개인정보 보호에 해당하는 제6호의 비중이 높음을 알 수 있다. 최근 10년간의 공개 재분류 결과 비공개 기록의 유형 중 제6호에 해당하는 기록은 <그림 1>과 같이 높은 비중을 보이고 있다.

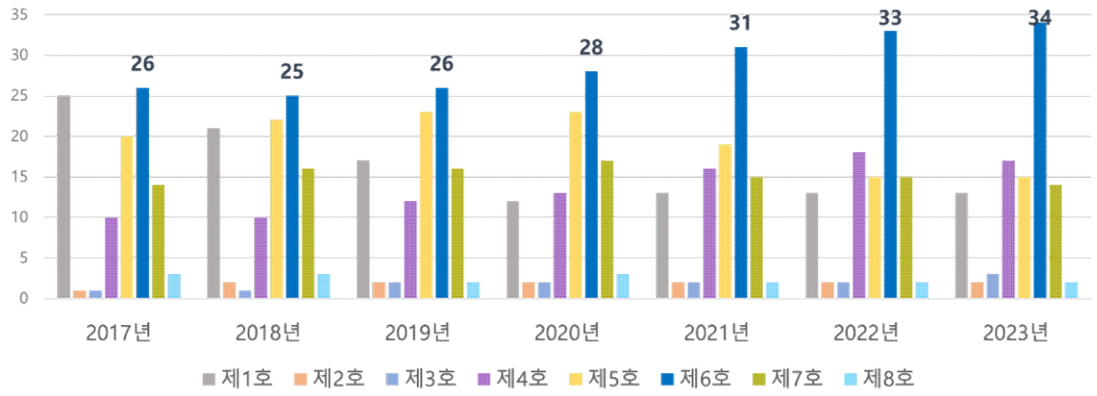


<그림 1> 최근 10년간 비공개로 재분류된 기록 중 개인정보 유형 비중
출처: 국가기록원 공고 "공개재분류 결과 비공개 기록물 유형별 현황"을 연도별 집계 재구성

공공기록의 공개 여부 재분류 외에도 정보공개법에 따라 정보공개 청구 시 공개 여부를 결정하여 청구인에게 통지하도록 되어 있다(법 제11조). 비공개 결정을 한 경우에는 비공개 사유와 함께 비공개 유형을 청구인에게 통지해야 한다(법 제13조 제5항). 2017년 기준 정보공개 청구에 대한 비공개 결정은 25,131건으로 전체 청구 건의 약 4%를 보이고 있으며, 2023년에는 비공개 결정이 61,113건으로 전체 청구 건의 6%에 해당한다(행정안전부, 2024, 25). 국민 알권리의 신장으로 정보공개 청구 자체가 증가하고 있는 동시에, 개인정보 보호를 근거로 하는 비공개 결정 역시 급증하고 있다. 2017년 비공개 사유 중 개인정보 보호에 해당하는 건은 6,482건으로 전체 비공개 결정의 약 26%를 보이다가, 이후 매년 증가하여 2023년에는 20,751건으로 34%를 차지하였다(행정안전부, 2024, 32). <그림 2>와 같이 개인정보 보호가 매년 비공개 결정 사유의 가장 높은 비중을 보이고 있음을 알 수 있다.

4) 국가기록원이 공고한 2018년 공개재분류 결과 비공개 기록물 유형별 현황 자료에서 2018년 비공개로 재분류된 기록물은 총 3,681,792로 공고되어 있으나, 비공개 유형별 기록물을 합산한 결과는 3,349,706건으로 33만 건 이상 적은 수량으로 차이가 있음. <그림 1>에서는 공고된 수량으로 비율을 산정하였음.

https://doi.org/10.14404/JKSARM.2025.25.3.331



<그림 2> 연도별 정보공개 청구에 대한 비공개 사유 현황(2017년~2023년)
출처: 2024 정보공개 연차보고서 p.32 재구성

국가기록원의 기록 공개 여부 재분류에 따른 비공개 사유와 공공기관의 정보공개 청구에 대한 비공개 사유 모두에서 개인정보 보호를 근거로 하는 비중이 가장 높음을 <그림 1>과 <그림 2>에서 확인할 수 있다. 다만, 정보공개 청구는 생산기관이 보유하는 기록을 대상으로 하는 경우가 많고, 기록 생산일로부터 보유기간이 비교적 짧은 기록을 대상으로 하는 경우가 많기 때문에 국가기밀이나 업무수행 중인 정보 등 개인정보 이외에도 다양한 유형의 비공개 사유가 유효한 상태로 존재한다고 볼 수 있다. 반면, 국가기록원으로 이관된 장기보존 대상 기록의 공개 여부 재분류 결과에서는 생산시점부터 이관시기까지 시간의 경과에 따라 비공개 사유가 상당 부분 소멸되고, 개인정보 보호 사유는 지속적으로 유지되는 경우가 많기 때문에 개인정보 보호를 사유로 한 비중이 정보공개 청구 결과에 비해 상대적으로 더욱 높은 비중을 차지하는 것으로 유추해 볼 수 있다.

공공기록의 비공개 사유 중 개인정보 비중이 높다는 것은 기록의 관리와 활용에 있어 복합적인 법적 요구사항을 충족해야 함을 의미한다. 공공기록물법은 공공기관의 투명하고 책임 있는 행정 구현과 공공기록물의 안전한 보존 및 효율적 활용을 위해 제정되었으며, 개인정보 보호법은 개인정보의 처리 및 보호에 관한 사항을 정하는 것을 목적으로 한다. 즉, 공공기록물법은 기록을 통한 행정의 투명성을 확보하고 기록의 공개와 활용을 강조하고 있으며, 개인정보 보호법은 정보 처리 과정에서의 사생활 침해를 방지하는 것이 목적임을 드러내고 있다. 이러한 법령 제정 목적의 차이는 기록 관리 실무에서의 충돌을 야기하고 있다. 개인정보처리자는 보유기간이 경과하거나 개인정보 처리 목적의 달성 및 가명정보의 처리 기간이 경과한 경우 지체없이 개인정보를 파기해야 하지만, 공공기관은 기록의 보존기간동안 보존해야 하며, 기록의 폐기 시에는 공공기록물법에서 정한 절차에 따라 폐기하도록 규정하고 있다. 국가는 과제 수행을 위해 요구되는 지식과 정보를 조사·처리·저장해야 함과 동시에 필요 이상의 정보를 수집하거나 전달·저장해서는 안되며(김일환, 2008, 353), 처리 목적을 달성한 개인정보는 즉시[5] 파기해야 한다. 그러나 업무의 투명성을 확보하고 책임있는 행정 구현을 위해서는 법정 보존기간 동안 기록의 보존이 필요하므로, 개인정보 보호법과 공공기록물법 간에는 개인정보 처리에 관한 상이한 규범적 요구가 존재한다.

이를 해결하기 위해 현행 개인정보 보호법은 다음의 두 가지 예외 규정을 포함하고 있다. 개인정보 수집 목적의 달성으로 파기 의무가 발생하였음에도 다른 법령에 의해 보존의무가 존재하는 경우에는 보존의무가 있는 정보를 물리적 또는 기술적 방법으로 분리하여 보관하고 접근권한을 최소화하여 관리하도록 하였다. 이와 함께 가명정보 처리에 관한 특례 규정을 두어, 통계작성이나 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있도록 하였다.

5) 「표준 개인정보 보호지침」 제10조에 개인정보가 불필요하게 되었을 때는 그로부터 5일내에 해당 개인정보를 파기하도록 규정함.

https://doi.org/10.14404/JKSARM.2025.25.3.331

그러나 이러한 단서 조항이나 특례 조항을 통한 부분적 해결은 아카이브 실무에서 한계를 드러낸다. 개인정보 보호법의 최소 수집 원칙이나 목적 제한 원칙은 공공기록물법의 공개원칙이나 기록의 연구, 활용, 전시 등의 기록 관리 요구사항과 구조적으로 충돌할 여지가 있다. 체계적이고 일관된 법리 적용을 위한 지침이나 통합적 해석 기준이 마련되는 제도적 보완이 필요하다.

2.3 현행 아카이브의 개인정보 보호 정책과 시사점

GDPR을 비롯한 국내의 개인정보 보호법에서도 공익적 기록보존의 목적일 경우 정보주체의 동의 없이 가명정보를 처리하거나 가명정보간의 결합을 예외적으로 허용하고 있다. 아카이브가 사회적 기억과 역사적 증거를 장기간 보존하고, 민주주의와 시민의 권리보호, 공공의 설명책임성과 문화유산으로서의 가치를 장기적으로 보존하기 위한 활동을 수행하기 때문에(TNA, 2018) 아카이브에서 공익을 위한 기록 보존 목적의 처리를 허용하고 있는 것이다. 따라서 아카이브에서의 개인정보 보호 정책에는 일반적인 개인정보 보호 정책과 접근방법에 있어 근본적인 차이가 요구된다. 일반적인 개인정보 보호는 정보주체의 자기결정권을 중심으로 수집 최소화 원칙, 처리 목적 외 활용 금지 원칙, 개인정보 침해 최소화 원칙 등을 강조한다. 그러나 아카이브에서 기록은 공개되어 활용되는 것을 전제로 관리하므로 열람과 보호 간의 균형이 필요하다. 또한 공공기록은 동일 업무기능 내의 다른 연관된 기록집합 속에 존재하므로, 개별 기록의 식별정보를 가명처리하거나 삭제하는 것으로는 개인정보 보호 조치의 완전성을 보장하기 어려울 수 있다. 아카이브에서의 기록은 역사적 가치와 연구 목적 등으로 장기간 보존되며 시간의 경과에 따라 개인정보 보호의 대상 여부가 변화될 수 있으므로 이에 대한 고려사항 역시 필요하다. 그리고 공공기록의 열람과 공개는 대중과 다수의 연구자를 대상으로 하므로 일반적인 개인정보 보호 정책과는 다른 위험 요소까지 고려해야 할 필요가 있다. 따라서 아카이브는 공공의 이익을 위한 기록의 공개 확대와 정보주체의 권리 사이의 균형을 고려한 아카이브의 특수성을 반영한 개인정보 보호 정책을 수립해야 한다.

그러나 현재 국내 공공기록물관리기관에서 시행되는 개인정보 보호 규정은 아카이브가 고려해야 할 특수성을 충분히 반영하지 못한 일반적인 개인정보 처리 방침만을 제공하고 있다. 개인정보 보호법 제32조에 따르면 공공기관은 개인정보파일의 보유 현황을 등록·공개하도록 되어 있다. 이 규정에 따라 기관별 개인정보 처리 방침을 제정하여 홈페이지 등에 게시하고 있다. 하지만 이 개인정보 처리 방침은 기관을 운영하는 과정에서 수집하는 개인정보의 유형, 처리 목적, 처리 근거, 보유 및 이용 기간 등을 명시한 것으로, 아카이브의 소장 기록을 대상으로 한 공개나 활용 측면에서의 개인정보 보호 이슈들을 다루고 있지 않다.

또한 공공기록물법에는 비공개로 재분류된 기록 중, 개인정보 보호 사유에 해당하는 정보공개법 제9조 제1항 제6호에 해당하는 기록에 대해서는 생산연도 종료 후 30년까지 공개여부 재분류 대상에서 제외할 수 있도록 하였다(법 제35조 제2항). 이 조항은 개인정보가 포함된 기록의 재분류 주기를 조정하여 실질적 보호 조치를 강화하는 근거가 될 수 있다. 그러나 한편으로는, 개인정보의 재식별 가능성 판단 없이 보호 중심의 판단을 강화하여 기록의 공개 실무를 경직화할 우려도 있다. 따라서 개인정보의 비식별화 처리 기준이나 재식별 가능성에 대한 평가 도구 등이 도입될 필요가 있다.

한편 공공기록물법은 영구기록물관리기관의 비공개 기록에 대한 제한적 열람을 규정하고 있다. 권리구제나 직무수행, 학술연구 등의 목적이 인정되는 경우 비공개로 분류된 기록에 대해 해당 목적 달성을 위한 최소한의 범위 내에서 열람을 허용하고 있다. 이 조항은 공익적 목적에 부합하는 경우 비공개 기록의 열람 제한을 완화하여 운영할 수 있도록 하는 근거가 된다. 하지만 법률 및 시행령의 조항은 추상적인 규정으로, 열람 청구의 목적을 판단하고 기록을 제공하기 위해 필요한 구체적인 실무적 기준이 필요하다. 제한적 열람을 허용할 수 있는 평가 도구나 가이

https://doi.org/10.14404/JKSARM.2025.25.3.331

드라인 등이 미비하여 실효성이 제한되는 한계가 있다. 즉 비공개 기록의 제한적 열람 기준이나 개인정보의 가명처리 기준 제정이 필요하고, 개인정보의 민감도 검토 및 재식별 위험성 평가 도구 등이 마련되어 이용자 접근 정책과 개인정보 보호간의 균형을 위한 제도적 보완이 필요하다.

해외 아카이브의 개인정보 보호 관련 규정은 개인정보 보호와 기록 접근권의 확대를 실효성있게 구현할 수 있도록 보다 더 명확한 규정을 제시하고 있다. 미국 국립기록청(NARA)의 연방규정집(CFR) 36 CFR에서 살아있는 개인에 대한 정보가 포함된 인사기록이나 의료기록, 이와 유사한 기록의 경우에는 공개되었을 때 개인 사생활 침해를 초래할 수 있는 세부적 사항에 대해 공개하지 않도록 제한하는 규정을 두고 있다(§1256.56). 하지만 생명의학 및 사회과학 연구 등 승인된 연구 목적으로의 예외적 접근은 허가하고 있다(§1256.28). 연구 목적으로 제한된 기록에 접근을 요청하는 경우, 요청을 위해 제출해야 하는 정보 목록과 이 요청에 대한 승인 과정의 고려사항, 준수해야 하는 접근 조건을 명확히 규정하였다. 접근이 제한되는 기록의 경우 일부 비공개 정보를 마스킹하거나 제거한 사본을 제공할 수 있다는 규정도 명시하였다(§1256.24). 이 외에도 NARA 직원의 윤리 규정에도 기록 시스템의 설계, 개발, 운영, 유지관리 업무를 수행하는 자는 개인정보 보호법과 5 CFR 2635.703에 명시된 비공개 정보 보호 행동 규칙을 준수해야 한다고 규정하였다(§1202.28). 그리고 NARA가 보유한 개인정보가 포함된 기록은 개인정보 주체의 요청이나 통계적 목적으로의 사용(개인식별정보는 삭제), 법적 요구(법 집행 활동이나 법원의 명령), 공공의 이익 또는 개인의 건강이나 안전에 영향을 미치는 상황 등에서 공개할 수 있으며, 이러한 경우를 제외하고는 개인의 동의 없이 다른 개인이나 기관에 공개할 수 없음을 명문화하고 있다(§1202.60).

영국 국립기록원(TNA)은 「Data Protection Act 2018」에 따라 정보위원회(ICO)가 발행한 「Guide to archiving personal data」를 준수하고 있다. 이 가이드는 아카이브 목적으로 개인정보를 처리할 때 적용할 수 있는 개인정보 처리 예외 조항과 절차를 안내하고 있다. 공익적, 역사적, 연구 목적의 장기 보존 목적에 적용되는 예외 조항을 분명히 명시하고 있으며, 동시에 이러한 예외는 처리 필요성, 적절한 안전장치, 정보주체의 피해 최소화, 투명성 확보의 조건을 충족할 때 적용할 수 있다고 설명한다(TNA, 2018).

이 외에도 미국 아키비스트 협회에서는 개인정보 보호를 “개인의 정보나 활동이 타인에 의해 무단 공개되지 않도록 보호되는 상태나 특성”으로 정의하는데(Society of American Archivists, n.d.), 이는 정보의 비공개 상태를 넘어서 보호 조치가 필요한 상태를 강조한다고 할 수 있다. SAA는 기록관의 운영 및 관리에 관한 표준화된 지침인 「Archives Policy Manual」에서 개인정보 보호 및 접근 정책에 관한 가이드를 제공한다. 개인정보는 개인 사망 후 30년 후 접근이 가능하며, 건강정보는 생산 후 120년 후 또는 사망 후 50년 경과까지 보호하는 것을 설명한다. 또한 개인정보는 본인에게만 공개하고, 사진 및 동영상 기록을 출판 및 전시하려면 본인의 동의를 받아야 하며, 어린이 및 어려움에 처한 사람을 기술한 자료는 전시나 출판을 할 수 없으며, 처리되지 않은 컬렉션에 대해서는 접근할 수 없도록 하는 등 아카이브에서의 개인정보 보호를 위한 제한 조치 13가지 기준을 구체적으로 제시한다(SAA, 2020).

위의 사례에서 검토했듯이, 아카이브는 기록의 보존이라는 공익적 목적과 정보주체의 권리 보호 간의 균형을 모색해야 한다. 포괄적인 개인정보 보호 법률을 준수하되, 아카이브는 내부 정책과 절차를 통해 이를 명확히 해석하고 적용하는 것이 필요하며, 일괄적인 비공개 조치가 아니라 구체적이고 기한이 있으며 상호 협의된 방식의 접근 제한을 제시해야 한다(Windon & Youngblood, 2024). 아카이브에서 기록의 공개를 확대하면서도 정보주체의 권리를 보호할 수 있는 구체적이고 실행 가능한 지침을 마련할 필요가 있다.

https://doi.org/10.14404/JKSARM.2025.25.3.331

3. 기록의 맥락적 특성을 반영한 개인정보 보호 전략

3.1 공공기록의 공개재분류와 개인정보 비식별화 기술 적용

공공기록물법에 따라 공공기관은 관할 기록물관리기관으로 기록을 이관할 때, 해당 기록의 공개 여부를 재분류하여 이관하여야 하며, 비공개로 재분류된 기록은 5년마다 다시 공개 여부를 재분류해야 한다. 공개 재분류 대상 기록은 매년 증가하고 있으나 현재의 업무 수행방식으로는 한계가 있다. 실제로 이관 당시 공개 여부가 분류되지 않은 미분류 기록이 존재하며, 여기에 생산 후 30년이 경과해 재분류 시기가 도래한 기록까지 포함하면, 매년 약 200만 권에 달하는 것으로 보고되고 있다(권미현, 2019, 1). 이에 따라 공개 재분류 미처리 누적 기록은 지속적으로 증가하고 있는 실정이다. 공개 재분류 결과는 기준서 서식에 맞추어 기록 건별로 작성하도록 되어 있는데,[6] 기록 유형을 구분하고, 생산기관, 제목, 생산연도, 수량, 기록의 내용, 재분류 결과, 재분류 결과가 비공개일 경우에는 근거 및 사유, 생산기관 의견조회 내용, 비공개 대상정보를 작성해야 한다. 기준서 서식 내 중복 작성되는 정보가 많고 입력한 정보를 검수하는 과정을 거친다.

이처럼 수작업으로 이루어지는 기록 공개 재분류 업무는 소요시간이 오래걸리며 인력 운용의 부담이 있어 자동화 지원이 필요한 분야이다. 효과적인 공개 재분류 업무 수행을 위해 국가기록원에서는 전자기록물 대상 비공개정보 자동 필터링 연구개발사업을 2020년부터 진행하고 있다. 물론 이전부터 개인정보 식별을 위한 상용화된 소프트웨어들이 개발되어 공공기관에서 사용중이었고, 주로 대상 데이터·기록에서 개인정보 탐지하여 마스킹처리하는 기능을 제공하여 소장 데이터·기록을 관리하고 활용하는 데 사용하고 있다. 하지만 여기에 사용되는 대부분의 사용 소프트웨어는 주민등록번호나 이메일주소와 같이 패턴화된 비공개 키워드를 검출하는 기능에 그치고 있다.

개인정보 보호법에서는 주민등록번호, 여권번호, 운전면허번호, 외국인등록번호 등 개인을 고유하게 구별하기 위해 부여된 식별정보를 “고유식별정보”로 정의하고, 이 정보들은 일반 개인정보보다 더욱 엄격한 보호 조치를 명시하고 있다. 특히 고유식별정보를 포함해 신용카드정보, 계좌정보, 생체인식정보 등은 「개인정보의 안전성 확보조치 기준」에 따라 반드시 암호화하여 저장하고 정보주체의 동의나 법령상의 처리 근거 요구, 접근권한 제한 등의 안전성 강화 조치를 의무화하고 있다. 이와 함께 개인정보 보호법에서는 개인정보 수집 목적을 달성할 수 있다면 익명정보나 가명정보 형태로 처리하도록 하여 정보주체의 권리를 보호하면서 정보활용을 할 수 있도록 하였다. 이러한 맥락에서 개인정보를 식별할 수 없도록 처리하여 활용할 수 있도록 하기 위한 비식별화 기술의 활용이 핵심적인 수단으로 쓰이고 있다. 특히 개인정보 보호법에서 규정한 통계작성, 과학적 연구, 공익적 기록보존의 목적으로 정보주체의 동의 없이 사용하기[7] 위해서는 비식별화 조치가 전제되어야 한다.

식별자(identifier)는 “특정 운영 맥락에서 데이터 주체(data principal)를 고유하게 식별할 수 있는 속성 집합”으로 정의되며(ISO/IEC 20889:2018, §3.13), 이러한 식별자는 두 가지 유형으로 구분된다. 직접식별자(direct identifier)는 “단독으로 데이터 주체를 고유하게 식별할 수 있는 속성” 정보이며, 간접식별자(indirect identifier)는 “다른 속성과 결합해 데이터 주체를 식별할 수 있는 속성”으로 정의된다(ISO/IEC 20889:2018, §3.10, §3.16). 비식별화(de-identification)는 이러한 식별자와 데이터 주체 간의 연관성을 제거하는 모든 과정을 의미하는 것으로, 개인 정보 노출 위험을 방지함과 동시에, 유의미한 통계적 활용 가능성을 확보하는 데 목적이 있다(NIST,

6) 공공기록물법 시행령 제72조 제1항에 따라 공개여부를 재분류하여 관할 기록물관리기관으로 이관하는 경우에는 기록의 건단위 또는 쪽단위로 공개여부를 구분하도록 되어 있으나, 매년 수행되는 “비공개기록물 공개재분류 결과 건별입력 위탁사업”의 제안요청서를 검토해 보면, 실제 업무는 건별로 재분류 결과를 입력하여 관리하고 있음을 알 수 있다.

7) 개인정보 보호법 제28조의2에 대해 정보주체의 자기정보 통제권 침해 및 기본권 제한의 과잉금지원칙 위배 여부로 헌법소원이 제기되었으나, 헌재는 가명정보 처리 특례 조항에 대해 합헌이라고 판단함(헌법재판소 2023.10.26. 선고 2020헌마1476). 사용 목적이 공익성에 기반하며, 가명정보는 재식별 가능성이 낮고 보호조치를 강화하고 있으므로 과도한 기본권 침해가 아니라는 판단임.

https://doi.org/10.14404/JKSARM.2025.25.3.331

2023, 12).

개인정보 보호 강화 데이터 비식별 기술 표준 ISO/IEC 20889에서는 비식별화 기술로 다음의 <표 2>와 같은 기술을 설명한다. 이 표준은 정형화된 테이블 구조의 데이터세트의 비식별화 기술에 초점을 두고 있어, 텍스트나 이미지 및 동영상 형식에는 적용이 어려울 수 있다. 비식별화 기술은 데이터의 속성이나 활용 목적에 따라 그 효과나 적합성이 달라지므로, 데이터의 구조와 데이터의 유용성 등을 종합적으로 고려하여 선택하는 것이 필요하다. 특히 텍스트 기반의 비정형 기록에 대해서는 정교한 기준을 마련하는 것이 요구된다.

<표 2> ISO/IEC 20889의 비식별화 기법

기법 유형	대표 사례	설명
통계적 기술	샘플링(Sampling)	무작위 샘플링으로 재식별 가능성을 감소
	집합(Aggregation)	속성별 총합, 평균 등 특정 속성에 대한 대표 정보를 표시하여 데이터의 정보량 감소
암호화 기술	결정론적 암호화(Deterministic encryption)	같은 입력은 같은 암호 값으로 변환되며, 재식별을 위해서는 암호키가 있어야 가능
	순서유지 암호화(Order-preserving encryption)	같은 키로 암호화된 경우 원래 순서가 유지되어 데이터 유용성 증가
	동형 암호화(Homomorphic encryption)	암호화된 상태에서 연산 가능
	형식유지 암호화(Format-preserving encryption)	입력 형식을 유지하며 암호화하는 방식
	준유사 비밀 공유(Homomorphic secret sharing)	데이터 내의 민감속성을 분할하여 여러 소유자에게 분산 저장
억제 기술	마스킹(Masking)	식별자를 제거하거나 속성값 일부를 삭제 또는 가림처리
	국지적 억제(Local suppression)	범주형 속성값 중 희소값을 제거
	기록 억제(Record suppression)	이상값 등 식별 가능한 레코드 전체 제거
가명화 기술	가명처리(Pseudonymization)	식별자 등을 가상 식별자로 대체
일반화 기술	반올림(Rounding)	속성값을 반올림하여 정밀도 감소
	상단 및 하단 코딩(Top and bottom coding)	속성값을 특정 임곗값 이상/이하로 범주화
랜덤화 기술	노이즈 추가(Noise addition)	속성에 임의 값인 노이즈를 추가하여 원래의 통계 속성은 유지하며 데이터를 수정
	순열(Permutation)	속성값의 순서만 재배열하여 통계 분포 유지
	미세 응집(Microaggregation)	근접한 값끼리 그룹화 후 평균값으로 대체
합성데이터	합성 데이터(Synthetic data)	실제 데이터의 통계적 특성을 분석한 모델을 사용하여 가상 데이터를 생성

출처: ISO/IEC 20889:2018, Clause 9 내용을 바탕으로 재구성

개인정보보호위원회의 「가명정보 처리 가이드라인」(2024)은 개인정보 보호법의 가명정보 처리에 관한 특례 규정에 근거하여 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 가명정보 처리를 위한 기준을 제공하고 있다. 특히 비정형 데이터의 개인식별 위험성 검토 사항을 예시를 제공하여 설명함으로써 실무의 활용도를 높였다. 그러나 비정형 데이터에서 식별자를 명확히 정의하거나, 자연어 처리 기반의 자동화된 식별자 탐지 및 비식별화 기술에 대한 구체적 기준이나 기술적 가이드라인은 미비한 수준이다.

3.2 기록의 맥락적 특성과 비식별화 기술의 한계

비식별화는 구별, 연결성, 추론 가능성을 기준으로 개인정보의 식별성을 제거하는 방식이며, 가명처리는 이 중에서 연결성과 추론 가능성을 차단하여 개인 식별이 불가능하도록 하는 방법이다(김정수, 2025, 88). 그러나 단일 항목으로는 개인 식별 가능성이 없지만, 다른 항목과 결합했을 때 식별 가능성이 높아질 수 있는 정보를 판단하는 절대적 구분이 어렵기 때문에(개인정보보호위원회, 2024a, 48), 맥락 관계 속에서 재식별 가능성을 판단

https://doi.org/10.14404/JKSARM.2025.25.3.331

하는 것이 중요하다. 현행 개인정보 보호법에서는 가명정보는 개인정보에 포함되는 개념이다. 그러나 다른 정보와 결합하기 어렵거나 다른 정보를 입수할 가능성이 낮은 경우 개인정보에 해당하지 않을 수 있기 때문에, 정보를 처리하는 상황과 맥락에 따라 개인정보 여부가 달라질 수 있다(개인정보보호위원회, 2024b, 2).

이러한 비식별화의 복잡성은 기록이 갖는 고유한 특성에서 더욱 뚜렷하게 부각된다. 기록은 단일 문서가 아닌 기록을 생산한 업무적 맥락과 연속성을 유지하는 정보의 유기적 집합체로 존재하게 된다. 이로 인해 개별 기록 건에는 식별정보가 존재하지 않더라도, 동일 기록 철 내의 기록 건 간의 상호 연계성이나 메타데이터 등을 통해 개인을 식별할 수 있는 가능성이 있다. 업무 기능 중심으로 분류되어 편철된 기록의 경우 맥락 관계 속에서 재식별될 가능성이 있는 것이다. 예를 들어, 인사기록 내에서 주민등록번호는 제거했더라도 발령 내역이나 직급 등이 상호 결합되면 특정 개인을 유추할 수 있다. 즉 비식별화의 기준인 구별성, 연결성을 제거해도, 추론 가능성이 남을 때 재식별 가능성이 발생하며, 외부 데이터와 결합되었을 때 재식별의 위험은 더욱 증가한다.

이러한 맥락 기반 재식별 위험성은 해외 실험에서도 확인되었다. Sweeney et al.(2017)은 환경·건강 관련하여 수집된 데이터를 대상으로, 미국 의료정보 보호법(Health Insurance Portability and Accountability Act)의 비식별화 조치(Safe Harbor)에서 요구하는 수준의 비식별 처리를 한 후 재식별 가능 여부를 연구하였다. 이 연구에서는 개인의 이름과 같은 명시적 식별자 정보를 삭제하고, 추가적인 안전조치로 생년 정보를 범주화하여 Safe Harbor 기준 이상의 비식별 처리를 했더라도, 외부에 공개된 정보 자원(부동산 정보 웹사이트, 온라인 주소록 등)을 결합하여 개인의 이름(25%)과 주소(28%)를 정확히 식별한 결과를 보여준다. 과거에는 인구통계 정보만을 기반으로 재식별 가능성을 평가했기 때문에 재식별 위험이 거의 없다고 보았으나, 공개된 부동산 정보의 주택 특성과 같은 외부 데이터를 결합하면 높은 확률로 재식별이 가능하므로 더 이상 이 Safe Harbor 기준은 개인정보 보호에 충분하지 않다는 것을 알 수 있다(Sweeney et al., 2017).

식별정보의 단순 삭제나 범주화와 같은 일률적 비식별화 조치는 속성정보나 메타데이터, 외부의 공개된 데이터베이스와의 결합을 통해 쉽게 재식별될 수 있다. 따라서 재식별 위험을 효과적으로 줄이기 위해서는 단순한 정보 제거뿐만 아니라 다른 속성정보, 메타데이터, 연계 가능한 정보와의 노출 위험성을 종합적으로 평가하고 보호조치를 설계해야 한다. 그러나 현행 법령이나 가이드라인은 대부분 식별정보의 암호화나 가명처리, 삭제 등의 방법에 머물러 있어, 기록의 생산 맥락에 의한 재식별 위험성에는 한계를 보인다. 그러므로 개인정보 보호의 실효성을 확보하고 안전한 활용을 위해서는 맥락 정보나 연계 가능성을 체계적으로 점검하고 대응하기 위한 표준화된 위험 평가 체계를 마련하는 것이 중요하며, 적절한 보호조치를 적용하기 위한 세분화된 지침이 제공되어야 한다.

더욱이 아카이브에서 개인정보 보호를 이유로 기록을 비공개 결정하거나 개인정보를 비식별처리하여 부분공개로 처리하는 기준, 그리고 개인정보를 비식별 처리하는 경우 비식별 처리의 범위나 처리 방식에 대한 명확한 지침은 더욱 필요하지만, 현재 제정된 규정은 없다. 명확한 기준이 없이 업무 작업자마다 판단이 자의적으로 이루어지기 때문에 가명처리를 통해 부분공개 가능한 기록이 비공개로 분류되거나, 반대로 맥락 정보가 유지된 상태로 공개되어 재식별 위험이 상존하기도 한다. 이는 개인정보 보호와 정보 공개의 균형을 위한 정책적 판단과 기술적 기준이 수립되지 못한 현실에 기인하고 있는 것이다.

비식별조치를 수행할 대상 정보의 유형과 활용 목적 등에 따라 제거해야 하는 식별자는 달라질 수 있다. 예를 들어 미국 의료정보 보호법(Health Insurance Portability and Accountability Act, 1996)에서는 개인을 식별할 수 없도록 삭제해야 하는 항목으로 18개의 식별자 목록을 제시한다. 이름, 지리적 정보, 날짜(생년월일, 입퇴원일, 사망일 등), 전화번호, 팩스번호, 이메일주소, 사회보장번호, 의료기록 번호, 건강보험 수혜자 번호, 계좌번호, 인증서 및 라이선스 번호, 차량번호, 기기 식별자 및 일련번호, 웹주소, IP주소, 생체식별자, 얼굴사진, 기타 개인을 고유하게 식별할 수 있는 정보가 그것이다. 이와는 다르게 학교 기록의 경우에는 학적 정보, 평가 정보 등이 지정되

https://doi.org/10.14404/JKSARM.2025.25.3.331

어야 할 것이다. 데이터의 종류와 유형, 활용 목적 등에 따라 개인 식별정보는 달라질 수 있기 때문에 사전에 대상 정보의 특성이나 분야, 범주에 따라 보다 엄격하게 식별해야 하는 정보들의 상세 항목을 정의할 필요가 있다. 특히 비정형데이터에서 개인정보가 비식별 처리되지 않으면 대규모 AI 학습데이터 구축에 이용되는 과정에서 개인정보가 유출되거나 도용될 수 있으므로 주의가 필요하다. 텍스트 형태의 문서정보에서 개인정보 식별을 위해서는 식별자 정보와 속성자와 같은 비신원확인정보 모두에 대한 기준을 만드는 것이 더욱 중요해지고 있다.

가명처리는 맥락 정보를 보존하면서도 개인정보 보호를 충족하기 위한 기술적 조치이지만, 실질적인 적용기준은 모호하다. 이로 인해 자의적 판단에 의한 정보의 은폐, 비공개 남용이 지적되기도 한다. 특히 아카이브의 기록활용 측면에서 보면, 과도한 비식별화 조치는 기록의 역사적·증거적 가치를 훼손할 수 있으며, 원본 맥락과 의미가 왜곡될 가능성을 내포한다. 공공기록이 사회적 맥락과 공무 행위의 증거적 가치를 유지할 수 있도록 기록의 특성을 고려한 정책적 조정이 요구된다.

또한 비식별 처리의 실효성을 담보하기 위해서는 비식별 처리 후 그 결과의 효과성과 재식별 위험에 대한 검증이 실행되어야 할 필요가 있다. 그러나 이러한 후속 검증 체계 역시 마련되어 있지 않다. 기록에 포함된 개인정보의 민감성을 판단하는 개인정보 민감도 검토(sensitivity review)를 비롯하여 메타데이터, 외부 정보와의 연계 가능성 등 다양한 요인을 종합적으로 고려하여 재식별 위험 평가 체계도 구축해야 하고, 이를 바탕으로 기술적 보호조치를 선택하는 체계가 필요하다. 특히 과도하거나 불완전한 비식별 처리는 오히려 재식별 위험을 증가시키거나 데이터의 활용도를 저해하므로 위험성 평가 도구를 통한 가이드라인 제정과 검증 절차의 제도화를 고려해야 한다. 또한 아카이브의 기록관리 기준과 통합하는 정책적 노력이 필요하다.

3.3 아카이브의 맥락정보 기반 개인정보 보호 전략

기록은 내용, 맥락, 구조를 갖추어 기록을 생산하게 한 활동에 대한 증거를 제공할 수 있는 정보로 정의된다(ICA, 1997, 22). 이러한 기록의 본질적 특성으로 인해 아카이브에서의 개인정보 보호는 단순 삭제나 익명화를 넘어서 복합적인 접근이 필요하다. 기록은 기록 객체의 보존뿐 아니라 기록 간 메타데이터 연결과 맥락 정보의 체계적 보존이 필요하기 때문이다. 또한 기록은 정보주체, 생산기관, 이용자 등의 다양한 이해관계자의 권리가 복합적으로 작용하기 때문에 개인정보 보호와 정보 접근권 사이의 균형점을 찾는 것이 중요하다. 이는 기록의 수정·삭제 불가 원칙과 개인정보 보호를 위한 수정·삭제 요구 사이의 충돌을 야기한다. 따라서 아카이브에서의 개인정보 보호는 기록의 보존 목적과 구조적 특성을 고려한 보호 전략 설계가 필요하다.

아카이브에서의 개인정보 보호 핵심은 기록의 구조적 특성을 반영한 맥락 정보 속에서 민감도를 판단하는 것에 있다. 기록은 업무 기능, 조직 정보, 주제 영역과 같은 정보를 기준으로 유기적으로 연결되어 있다. 따라서 개인정보의 식별 가능성이나 민감도 검토, 재식별 위험성을 판단할 때 기록의 내용뿐만 아니라 기록을 생산한 조직이나 업무 기능, 인물 정보, 관련 사건 등 구조적 정보를 종합적으로 고려해야 한다. 예를 들어, 동일한 개인정보 요소일지라도 그것이 포함된 기록이 징계나 재판과 같은 민감한 행정절차의 일부인지, 혹은 공식적인 사건이나 업무에 관련된 것인지에 따라 공개 여부와 보호 강도는 달라질 필요가 있다. 혹은 특정 질병이 언급된 기록도 의료진의 진료 차트인지, 복지 정책이나 보험료 책정에 사용되는 통계자료인지, 개인 간에 주고받은 서신인지에 따라 보호 수준은 차등화되어야 한다.

이러한 맥락 정보는 기록 생산 시점부터 관리 과정에 이르기까지 지속적으로 생성되고 축적되는 메타데이터와 전거정보로 관리된다. 이를 활용하면 개별 기록에 포함된 개인정보의 유형과 범위를 세분화하고 해당 정보의 민감도 수준과 보호 범위를 정교하게 설정할 수 있다. 따라서 기존의 단편적이고 일률적인 비식별화 조치에서 벗어나

https://doi.org/10.14404/JKSARM.2025.25.3.331

아카이브의 맥락 정보를 활용한 기준을 마련하는 것이 실질적인 개인정보 보호 전략의 핵심이 될 수 있다.

이러한 기록을 대상으로 한 개인정보 민감도 검토는 접근 제한 기록 유형의 복잡성, 비정형적이고 구조화되지 않은 기록의 특성, 그리고 맥락에 따라 달라지는 민감도 속성 등으로 판단이 쉽지 않다(Lemieux & Werner, 2024, 2). 방대한 양의 기록을 수작업 방식으로 처리하는 것은 한계가 있기에, 기술적 접근법을 도입하는 것이 필요하다. 이러한 상황에서 최신의 인공지능 기술을 이용한 자동화 솔루션이 아카이브에서 실험되고 있으며, 특히 개인정보를 찾아내 익명화하거나 삭제하는 도구가 개발되어 사용되고 있다. 더 나아가 개인정보를 직접 노출하지 않으면서 데이터를 처리하거나 분석할 수 있도록 하는 개인정보 보호 강화 기술(Privacy-Enhancing Technologies, 이하 PETs)도 주목받고 있다.

PETs는 기록의 실제 내용을 직접 열거나 파일을 복사하고 이동하는 방식이 아니라, 수학적 방법을 통해 유의미한 검색 결과나 분석 결과만을 도출하여 데이터를 활용하도록 하여 데이터 처리 과정에서 개인정보를 보호하는 것을 핵심으로 한다(Lemieux & Werner, 2024, 5). 정보기술 연구 및 자문회사인 가트너는 2022년 전략기술 중 하나로 '개인정보 보호 강화 컴퓨팅 기술(Privacy-Enhancing Computation)'을 선정하며(Gartner, 2021) 이러한 접근법의 중요성을 강조하기도 했다. PETs의 주요 기술로는 동형암호(Homomorphic Encryption), 신뢰 실행 환경(Trusted Execution Environments), 안전한 다자간 계산(Secure Multi-Party Computation), 차분 프라이버시(Differential Privacy), 개인 데이터 저장소(Personal Data Stores), 프라이버시 보호 기계 학습(Privacy-Preserving Machine Learning) 등이 있으며, 이러한 PETs의 기술범주에는 포함되지 않지만 이 기술들을 지원하는 토큰화(Tokenization), 합성 데이터(Synthetic Data), 블록체인 및 분산 원장 기술(Blockchain and Distributed Ledger Technologies), 영지식 증명(Zero-Knowledge Proofs)등도 아카이브에서의 활용이 검토되고 있다. 예를 들어, 업무 기능이나 메타데이터 등을 통해 주제 영역으로 분류된 기록의 민감도나 활용 목적을 고려하여 PETs을 선택 적용하는 전략을 사용하여, 기록 자체의 원본 데이터를 훼손하지 않고 데이터를 분석·활용할 수 있다. 이 과정에서 PETs는 기록 내용을 직접 수정하거나 이용자에게 노출하지 않고, 유의미한 통계값이나 패턴 분석 결과 등을 제공할 수 있을 것이다.

하지만 PETs의 기술 성숙도를 고려하거나, 아카이브가 보유한 상당수 기록이 텍스트 기반이라는 점에서 실질적인 적용에는 한계가 있을 것으로 판단된다. 동형암호나 안전한 다자간 계산은 연산 속도가 느리고, 차분 프라이버시나 합성데이터 기법 등은 텍스트 기반의 문서에 적용하기에는 적합하지 않은 경우가 많다. 또한 개인정보 보호를 위해 '노이즈 추가' 등을 실행해야하므로 기록의 진본성이나 무결성의 훼손이 불가피한 측면이 있다. 그럼에도 불구하고 방대한 양의 디지털 기록의 개인정보 보호를 위해서는 기술적 접근 방법에 대한 적극적 검토가 필요하며, 이러한 기술적 검토 결과 역시 관련 정책 설계에 반영될 필요가 있다.

아카이브에서 맥락 정보를 활용한 개인정보 보호 기준을 수립하기 위해서는 단순히 문서 내 특정 단어를 삭제하거나 익명화하는 방식을 넘어서야 한다. 기록이 속한 기능적·조직적 맥락과 주제별 도메인을 종합적으로 고려하여 보호 수준을 차등화하는 것이 필요하다. 이를 구현하기 위해서는 다음의 과제들이 수행되어야 한다. 먼저 기록분류체계와 맥락 정보를 기반으로 한 민감도 검토 모델 설계가 필요하다. 예를 들어 인사기록 내에서도 채용, 승진, 포상, 징계 등 업무 기능에 따른 개인정보의 민감도와 보호 수준 차등화 체계를 수립해야 한다. 또한 재식별 위험성 평가 도구 개발이 필요하다. 개별 기록에 포함된 개인정보가 다른 기록이나 외부 데이터와 결합하여 개인을 식별할 수 있는 위험성을 평가하고 보호 조치 강도를 결정하기 위한 체계가 필요하다. 이와 함께 메타데이터 정보를 활용한 자동화 도구를 구현할 필요가 있다.

이러한 맥락 정보 기반 개인정보 보호 정책이 실무에 정착되기 위해서는 다음 사항에 대한 전제조건이 충족되어야 할 것이다. 첫째, 기록의 메타데이터 정보와 전거 정보가 정확하고 신뢰할 수 있는 방식으로 생성되고 관리되어

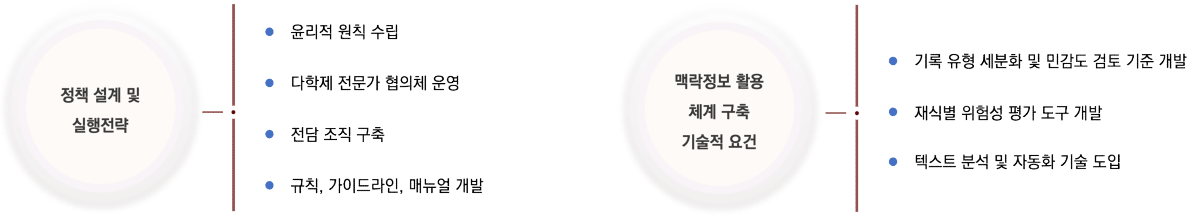
https://doi.org/10.14404/JKSARM.2025.25.3.331

야 하며 기록이 유지되는 동안 지속적으로 축적되는 체계가 마련되어야 한다. 둘째, AI 기반의 개인정보 식별 기술이 적용 가능하도록 맥락 정보가 기계가독형으로 생산되고 관리되도록 설계되어야 한다. 셋째, 기록의 진본성과 무결성 원칙이 지켜지면서도 개인정보가 보호되며 활용될 수 있도록 기술·윤리·법제를 아우를 수 있는 정책 설계가 필요하다. 넷째, 개인정보 보호와 정보 접근의 확대를 위한 균형점을 찾기 위해서는 아카이브 내의 여러 전문직군과 이해관계자의 협력 체계를 구축해야 한다. 이러한 맥락 정보 기반 접근법은 아카이브에서의 개인정보 보호를 위한 핵심 전략 역할을 할 수 있으며, 디지털 환경에서 기록관리와 개인정보 보호의 새로운 패러다임을 제시할 수 있을 것이다.

4. 아카이브의 개인정보 보호 정책 설계와 제도 정비 방안

지금까지의 분석을 통해 아카이브에서의 개인정보 보호는 기존의 개인정보 보호법 및 기록관리법 체계만으로 대응하는 것은 한계가 있으며, 아카이브의 특성을 반영한 정책 설계가 필요함을 확인하였다. 개인정보 보호법이 주로 정보주체의 권리와 개인정보 처리자의 책임에 초점을 두고 있는 반면, 아카이브에서의 기록은 장기간 보존·활용을 전제로 하며 공공의 접근 확대를 목표로 하기 때문에 아카이브에서 발생하는 복합적 개인정보 보호 문제를 조율하기 위해서는 기존 법제도의 해석과 적용의 문제를 넘어서는 별도와 정책 마련이 필요하다. 즉 장기보존을 전제로 한 기록의 특성과 함께, 디지털 기술 환경의 변화, 정보주체의 권리 보호 강화 요구를 동시에 고려한 정책적 접근이 요구되는 문제이다. 또한 해외 사례 분석을 통해 개인정보 보호를 위한 윤리적 기준과 개인정보 민감도 검토 및 재식별 위험성 평가 등이 마련되어야 한다는 시사점도 얻을 수 있었다.

이렇게 국내 법제 분석 및 해외 연구 사례 검토 등을 통해 도출된 주요 쟁점을 종합하여, 실무 적용 가능성을 중심으로 아카이브의 개인정보 보호를 위한 정책 개발의 방향과 맥락 기반 개인정보 보호 제도 구축에 반영되어야 하는 핵심 요소들을 아래의 <그림 3>과 같이 제안하고자 한다. 먼저 아카이브 환경에 적합한 개인정보 보호 정책 설계 방향으로 윤리적 기준 수립, 전문가 협의체 구성, 전담조직 및 인력 체계, 제도적 실행 기반 마련이 필수적이라고 판단하였다. 이와 함께 앞서 분석되었던 공개 여부 판단의 주관성이나 재식별 위험에 대한 미흡한 검토 체계, 보호 조치의 실효성 부족과 같은 문제점을 해소하고 정책 구현의 기술적 실행 기반을 구축하기 위해 기록 유형별 개인정보 민감도 검토 체계, 텍스트 분석 기반 자동화 기술 개발, 외부 데이터 자원의 연계를 통한 재식별 위험도 평가 체계 구축이 필요함을 제시하였다.



<그림 3> 아카이브의 개인정보 보호 정책 및 체계 구축 방향

https://doi.org/10.14404/JKSARM.2025.25.3.331

4.1 개인정보 보호 정책 설계 및 실행 전략

4.1.1 개인정보 보호 정책의 윤리적 원칙 수립

아카이브가 소장한 다양한 유형의 기록 중에서 개인정보를 포함한 기록은 광범위하지만, 보호대상에 대한 규정과 보호 처리에 대한 표준화된 기준은 마련되어 있지 않다. 기록관리 관점에서 개인정보에 대한 개념은 다른 분야와 다르지 않지만, 개인정보 보호의 접근 방식은 궁극적으로 기록의 접근성을 보장하는 것을 목적으로 하고 있다(Windon & Youngblood, 2024, 206). 이러한 목적을 보장하면서, 잊힐 권리(Right to be Forgotten)나 삭제권(Right to Erasure), 나아가 정보주체의 개인정보 자기결정권을 보호하는 방향으로 정책이 제정되어야 한다. GDPR에서는 공익을 위한 기록 보존의 목적인 경우로서 삭제권을 실행하기에 불가능하거나 삭제할 경우 해당 처리 목적 달성을 심각하게 저해할 가능성이 있는 경우에는 정보주체의 삭제 요구에 대해 거부할 수 있도록 한다(§17.3.d). 이처럼 개인정보 삭제권과 기록 보존 의무 간에는 상호 충돌되는 지점이 있으므로, 개인의 권리와 사회적 공익간의 균형있는 조정과 판단 기준을 마련하는 것이 필수적이라 할 수 있다. ICA 기록 접근 원칙(Principles of Access to Archives)에서도 개인정보 보호 요건을 수용하고 준수하면서도 기록의 접근과 이용에 대해 적극적으로 제공해야 한다고 명시하고 있다(ICA, 2012, 9). 특히 기록의 온라인 공개의 경우는 보다 적극적인 형태의 정보 제공 방식으로, 기존과는 차원이 다른 접근성과 노출의 범위를 제공하게 되므로, 이에 따른 추가적인 개인정보 보호 이슈를 고려해야 한다. 이에 대해 TNA는 온라인에 게시되는 개인정보는 검색 엔진을 통한 노출 가능성을 포함하고 있으므로 특히 신중한 검토가 필요하다고 지적하고 있다(TNA, 2018, 34).

이러한 맥락에서 기록의 공개에는 윤리적 측면을 고려하는 것이 요구된다. Ketelaar(1995)는 정보가 생산된 맥락에 따라 개인정보의 민감도와 접근 범위를 판단하는 것이 필요하며, 국가–기관–시민의 삼자간의 상호작용에 따라 개인정보의 가치 평가가 이루어져야 한다고 하였다. 예를 들어, 시민 스스로 자발적으로 기관에 제출한 정보라면 정보주체가 통제권을 갖고 있으므로 이 기록의 연구목적 사용은 연구자 서약으로 보호 목적을 달성할 수 있겠지만, 첩보 기관이 고발이나 감시 활동을 통해 수집한 민감정보와 같이 정보주체가 반론 기회를 갖지 못한 기록은 민감도가 높기 때문에 더 강력한 보호 조치가 필요하다고 주장한다(Ketelaar, 1995, 15). 즉 기록을 생산하는 과정에서 개인정보가 왜곡된 권력관계를 반영하여 불공정하게 수집된 기록은 법규에 따라 정보 자체의 민감도를 판단하는 것에 그치지 않고, 맥락까지 고려하여 보호 수준을 강화해야 하는 것으로 이해할 수 있다. 아카이브는 시민의 권리를 보호해야 하며, 아키비스트는 기록의 맥락과 권력 구조를 검토하여 공개에 대한 윤리적 책임을 다해야 한다.

이렇듯 아카이브의 기록 공개에 의한 개인정보 보호 문제를 접근할 때 기관이나 이슈 중심이 아닌 개인 중심으로 문제를 생각하고, 특정 기록 컬렉션이나 기록 집합의 공개로 인해 영향을 받을 수 있는 실제 개인이나 공동체가 존재한다는 것을 염두에 둔 돌봄 윤리(Ethics of Care)를 적용할 필요가 있다(Windon & Youngblood, 2024, 208). Windon과 Youngblood(2024)는 아키비스트가 재난이나 참사 기록을 수집하는 경우 기록 생산에도 적극적으로 참여하기 때문에 개인정보 보호 결정권이 기록 생산자뿐 아니라 아키비스트에도 위임된다는 점을 강조하였다. 또한 웹 아카이빙의 윤리적 문제로 보존된 게시물이 게시자에게 불리한 증거로 활용되거나 그들을 위험에 노출시킬 가능성이 있어, 보다 정교한 윤리적 가이드라인의 수립이 필요하다고 지적하였다. 공개된 플랫폼에 게시된 콘텐츠라 하더라도 일시적 소비나 제한된 청중을 상정한 경우가 많기 때문에, SNS상의 집회·시위 관련 기록과 같은 사회운동이나 동시대 사건의 기록화를 목적으로 한 웹 아카이빙에 신중한 접근을 요구하였다(Windon & Youngblood, 2024, 209). 특히 사회적 약자나 취약계층의 보호를 위한 기준은 보다 섬세한 주의가 필요하다. 이들은 기록의 생산 방식이나 유통에 대해 통제하는 것이 더욱 어렵기 때문에 이러한 정보의 아카이빙이 차별과

https://doi.org/10.14404/JKSARM.2025.25.3.331

감시의 수단이 될 수 있음을 인지하는 것이 필요하다. 해당 기록이 문맥을 벗어나 이용되거나 왜곡되지 않도록 맥락 정보가 함께 아카이빙 되도록 해야 하며, 필요한 경우 비식별화와 익명화, 접근 제한 등의 조치가 윤리적 관점에서 실행될 수 있는 제도적 기반이 마련되어야 한다.

더 나아가 기록의 생산자나 기증자가 아닌 제3자의 정보가 포함되어 있는 경우 제3자 개인정보의 보호 방안도 정책에 반영되어야 한다. Windon과 Youngblood(2024)의 연구에서는 제3자의 개인정보 문제, 특히 수감자나 미등록 이주민, 국가폭력이나 제노사이드의 피해자와 같은 취약계층의 기록화에 신중해야 한다고 지적한다. 미시시피 주 주권위원회의 기록이나 동독 비밀경찰 슈타지의 감찰 기록은 이러한 대표적인 사례로 언급된다. 또한 전통지식(Traditional Knowledge)와 같이 특정 개인이 아닌 공동체의 전통문화 아카이빙의 경우, 공동체의 개인정보 권리를 존중하여 신중한 접근이 필요하다는 것 역시 상기할 필요가 있다(Windon & Youngblood, 2024, 212).

또한 이러한 고려 사항을 반영하여 제정된 개인정보 보호 정책일지라도, 제정 이후 시대적 변화에 따라 주기적으로 재정비 되어야 한다. 현재의 개인정보 보호 조치가 향후에 연구자료로서의 활용 가치나 사회적 기억 형성에 미치는 영향을 고려하여 지속적으로 재점검하고 갱신되어야 한다. 최근 인공지능 기술의 급격한 발전과 빅데이터 시대의 도래는 기존 정책의 지속적인 재검토가 요구되며, 이러한 기술적 변화를 체계적으로 모니터링하고 정책에 반영하는 것이 필요하다.

4.1.2 다학제적 전문가 협의체 운영

아카이브의 개인정보 보호 정책 수립은 복합적이고 다원적인 검토가 필요하다. 기록이 장기적으로 보존되는 과정에서 다양한 사회적 맥락과 기술 환경 속에서 활용되기 때문에, 기술의 변화와 사회적 요구를 반영한 정책 설계가 요구된다. 이에 따라 기록관리 전문가뿐만 아니라 정보 보호 전문가, 법률 전문가, 정보기술 전문가, 시민사회 활동가 등이 참여하는 체계를 구축하여 전문성과 이해관계를 고려한 협력체계를 수립하여 균형 있는 정책 설계를 도모해야 한다. 각 참여 주체의 역할과 책임을 분명히하고 상호 협력과 합의 과정을 체계화하는 것은 정책의 실효성을 확보하는 핵심 요소이다.

디지털 환경에서 개인정보 보호는 기술적 대응만으로는 해결하기 어려운 윤리적, 법적, 사회적 문제들이 복합적으로 작용한다. 특히 AI 기술을 적용하기 위해서는 정보주체의 권리 침해나 알고리즘의 편향성, 책임 소재의 불분명 등의 문제가 발생한다. Jaillant와 Rees(2023)는 이해관계자 간의 신뢰 부족과 기술에 대한 불신이 이러한 위험을 증폭시킨다고 지적하며, AI의 효과적 적용과 아카이브의 공적 신뢰를 강화하기 위해서는 이해관계자 간의 상호 이해와 협력이 필수적이라고 주장한다. 이는 단순한 계산능력을 강화하는 것을 넘어서 알고리즘이 다루는 객체에 대한 도메인 지식을 통합하고, 기술의 윤리적 활용과 정책 결정의 투명성, 설명 책임성을 확보하는 구조의 중요성을 강조한 것이다(Jaillant & Rees, 2023).

의사결정 과정의 기록화와 투명한 공개를 통해 정책 결정 과정의 근거를 확보하고 설명책임성을 제고하는 것은 아카이브의 공적 신뢰를 강화하는 기반이 된다. 전문가 협의체는 이러한 투명성과 참여를 제도화하는 구조로 아카이브의 개인정보 보호 정책의 실효성과 사회적 수용성을 확보할 수 있도록 기여할 것이다.

4.1.3 전담 조직 구축 및 운영

아카이브에서 개인정보 보호의 실효성을 제고하기 위해서는 정책 설계뿐 아니라, 실행하기 위한 내부 운영 기반이 조성되어야 한다. 다양한 기록간의 정보 재조합 가능성과 외부 공개 범위가 확대되는 환경에서 규정 제정만으로는 한계가 있으므로 개인정보 보호를 전담하는 전문 조직과 인력을 갖춘 체계적인 내부 구조가 필요하다. 이 전담

https://doi.org/10.14404/JKSARM.2025.25.3.331

조직은 아카이브 전반의 개인정보 보호 조치를 기획하고 수행하며 기술 및 법률 담당부서와 조율하는 역할을 담당해야 한다.

이와 함께 실제 업무를 수행하는 공개 여부 담당자를 대상으로 한 전문 교육 체계의 정비도 이루어져야 한다. 기록 이관 시 개인정보 보호 등급을 확인하는 체크리스트의 정비, 기증 기록의 계약 내용을 확인하여 보호 기간을 명시하는 매뉴얼의 제정, 재식별 사고 사례를 축적하고 분석하여 교육 프로그램에 반영하는 등의 실질적 역량 강화 교육이 필요하다.

아울러 디지털 기술의 빠른 변화에 유연하게 대응하기 위해 최신 법제 및 기술 동향을 상시적으로 모니터링하고 환경 변화에 따른 영향 평가 체계를 제도화해야 한다. 특히 인공지능, 자동화 기술, 빅데이터 분석 기술 등의 도입은 개인정보 보호 체계에 새로운 위협요인이 될 수 있으므로, 아카이브 내 시스템 도입이나 고도화 시, 사전적·사후적 영향평가 체계를 안정적으로 운영할 필요가 있다. 개인정보 보호법 제33조에 따라 공공기관은 개인정보파일의 신규 도입·변경 시 위험요인을 분석하고 개선사항을 도출하기 위해 사전 영향평가 제도로 “개인정보 영향평가(Privacy Impact Assessment)”를 운영하고 있다. 개인정보 처리 업무 현황을 분석하고 침해 요인을 도출하고 위험도를 산정하여 개선계획을 수립하여 이행하도록 하는 것을 주요 골자로 한다. 주로 신규 시스템 구축이나 대규모 개인정보 처리의 위험 분석에 초점을 두고 있다. 그러나 아카이브의 소장 기록을 대상으로 보존 정책을 검토하고 개선하도록 하는 체계는 마련되어 있지 않다. 아카이브는 기록을 장기 보존하며 생산 당시의 목적을 넘어서는 맥락에서 지속적으로 재이용되므로, 일반적인 정보처리기관과는 다른 유형의 리스크가 존재한다. 생산 이후 수십 년이 지나 공개될 수 있는 기록이 갖는 잠재적 리스크나 과거 맥락이 단절된 상태에서 재이용 될 때 발생할 수 있는 재식별 위험 및 2차 피해 등은 통상적인 영향평가 항목에 포함되기 어렵다. 따라서 기록의 생애주기를 고려하고 검색 매커니즘과 공개 방식을 고려한 영향평가 기준 및 절차의 설계가 필요하다.

마지막으로 아카이브의 개인정보 보호 역량을 제고하기 위해 국제적 기준과 모범 실무를 수용할 필요가 있다. GDPR이나 개인정보 관리 시스템 표준(ISO/IEC 27701) 등을 준용하여 정책을 운용하고, 해외 아카이브의 사례나 정책 경험을 공유하고 참조하여 정책에 반영할 필요가 있다.

4.1.4 관련 기준 및 제도 정비

① 아카이브의 개인정보 보호 규칙 제정

아카이브에서의 개인정보 보호를 위한 규칙을 제정해야 한다. 이 규칙을 통해 아카이브에서 개인정보가 포함된 기록을 장기적으로 보존하기 위한 활동을 실행할 수 있는 근거를 명확히 제공해야 한다. 아카이브의 기록 보존 행위가 지향하는 공공의 이익을 분명히 하고, 이러한 공익적 기록 보존 목적으로 하는 정보주체의 권리 제한에 대한 면책 범위를 설정해야 한다. 개인정보 수집 목적 외 처리 및 최소 처리, 보유기간 제한의 원칙을 넘어서는 보관과 가명 처리 활용의 근거를 명확히 하여 업무 수행의 기반을 마련해야 한다. 또한 잊힐권리와 같은 정보주체의 권리에 대한 보장을 위한 지원 방안이 필요하다. 정보주체는 법에서 보장하는 개인정보의 처리정지, 정정, 삭제, 파기를 요구할 권리를 가지지만, 공익적 보관 목적과 상충하는 경우의 처리 방안에 대한 규정이 있어야 한다. 공익을 위한 보관 목적으로의 처리가 필요한 상황에서 개인정보를 삭제로 인하여 보관 목적 달성이 불가능하다면 삭제 권리를 적용할 것인가를 판단할 기준이 필요하다(TNA, 2018, 17). 혹은 정보주체의 정정 요청이 있는 경우 정보의 직접적인 수정이 아니라 메타데이터나 카탈로그 설명 등을 통한 부기 조치 등을 통해 기록의 진본성과 무결성을 보장하는 원칙 등을 제시해야 한다(TNA, 2018, 17−19). 나아가 아카이브의 개인정보 보호 책임자의 역할과 책임의 범위 또한 명시되어야 한다.

이와 함께 위에서 언급한 실행상의 윤리적 고려사항 역시 포함되어야 하는데, 여기에는 전문가의 윤리적 판단과

https://doi.org/10.14404/JKSARM.2025.25.3.331

사회적 책임성, 그리고 기록의 생산 및 이용 맥락에 대한 정보를 반영할 필요가 있다. 아카이브에서 개인정보 보호를 위한 규정 제정 시 다층적 보호 구조를 고려해 볼 수 있다. Ketelaar는 아카이브에서 개인정보 보호를 위한 다음의 다섯 가지 보호 층위(Layers of Protection)를 제안했다. 개인정보 보호법이나 기록관리법에 따른 접근 제한을 결정하는 입법(Legislation), 기록이 아카이브로 이관될 때의 접근 조건을 설정하는 이관조건(Conditions of transfer), 민감정보에 접근하려는 연구자의 비공개 준수 서약 동의(Researchers' undertaking), 보관 시설의 보안과 열람 규칙에 관한 물리적·관리적 조치(Physical·Practical regulations), 법과 규정을 넘어서는 전문적 직업 윤리(Professional ethics)의 다섯 개 층위를 제시하며 다양한 맥락과 책임 주체를 고려해야 함을 강조하였다(Ketelaar, 1995). 여기에 더하여 개인정보 보호 기관과 관련 사회 분야가 협력하여 합의한 행동강령(Codes of conduct)을 추가하는 것이 논의하기도 하였다(de Groot & van der Sloot, 2018, 305).

Ketelaar의 논의로부터 개인정보 보호를 위한 다양한 접근점이 필요함을 인지해야 한다. 아카이브는 GDPR의 개인정보 기본 처리 원칙과 개인정보 보호법의 개인정보 보호 원칙, 그리고 ICA의 윤리강령에서의 접근 제공 원칙과 프라이버시 존중 원칙에서 도출된 개인정보 보호 원칙을 전제로 해야할 필요가 있다. 이러한 원칙과 기준을 기록관리 활동 과정에서 실현하기 위해서는 기록의 생산과정부터 관리, 보존 과정의 아키비스트뿐 아니라 입법자, 아카이브 관리자와 같은 다양한 행위주체가 법과 제도와 윤리적 고려 속에서 상호작용하며 맥락적으로 판단하는 다층적 보호체계가 필요하다. 이러한 다층적 보호구조는 법적 규제가 포괄하지 못하는 실무적 한계를 보완하고 각 특성과 상황에 맞는 보호 조치를 통해 개인정보 보호의 실효성을 높이는 데 활용할 수 있다.

② 아카이브의 개인정보 보호 처리 가이드라인 제정

아카이브에서의 개인정보 보호 원칙을 실행하기 위한 실무 가이드라인을 제정해야 한다. 아카이브가 소장한 기록 분야별, 그리고 개인정보 유형별로 세분화하여 실무적 실행 수준의 업무 지침을 제공해야 한다. 개인정보의 종류에 따른 민감도에는 차이가 있으며 동일한 정보일지라도 맥락에 따라 식별이나 노출에 따른 위험도는 달라지게 되므로, 세부 유형별·상황별로 차등화된 민감도 검토 체계를 정량화한 도구로 개발해야 한다. 또한 디지털 기록의 경우 온라인 공개에 대한 더욱 엄격한 평가 체계를 사전에 적용할 필요가 있다. 검색 엔진에 노출될 경우의 피해까지 고려해야 하며, 목록 정보나 메타데이터, 검색 도구까지 사전에 위험도 평가를 적용할 수 있는 체계를 구축하는 것이 중요하다.

특히 업무 절차별로 개인정보 민감도 검토 및 위험성 평가가 필요한 시점과 업무를 실행하는 매뉴얼이 함께 제시되어야 할 것이다. 단계별 매뉴얼은 업무를 지원하는 도구로서의 역할을 할 수 있도록 스코어링 시스템이나 체크리스크 형식으로 개발될 필요가 있다. 예를 들어, 아래와 같이 TNA의 개인정보 기록 보존 가이드의 일부 질문 예시처럼(TNA, 2018, 30), 업무를 수행할 때 개인정보 보호에 영향을 미치는 내용을 확인할 수 있는 질문들을 체크리스트 형태로 목록화하여 개인정보 영향을 판단할 수 있도록 할 수 있다.

- 해당 인물이 생존해 있는가, 사망했는가?
- 해당 인물을 기술내용에 명시하는 것이 반드시 필요한가?
- 기술 내용을 공개하는 데 법적 제한이 있는가?
- 관련된 기록은 현재 공개되어 있는가, 비공개 상태인가?
- 이미 (다른 공공영역에) 공개되어 있는 정보인가?
- 설명정보에 포함된 개인정보를 공개했을 경우 정보주체에게 고통이나 피해를 줄 가능성이 있는가?

③ 아카이브의 개인정보 보호 업무 절차 매뉴얼 개발

아카이브의 개인정보 보호 전담 조직은 아카이브 전체의 개인정보가 포함된 기록의 처리 현황을 파악할 수

https://doi.org/10.14404/JKSARM.2025.25.3.331

있는 모니터링 절차를 개발해야 한다. 개인정보 유출 위험을 사전에 식별하기 위한 위험요인을 정의하고 자동 측정하는 모니터링 시스템을 구축할 필요가 있다. 아카이브 시스템 내 보안솔루션을 적용하고 개인정보를 검출하여 통지하는 시스템을 구축해야 한다. 아울러 국내외 법제 환경의 변화와 기술 동향을 주지하여 이를 반영하여 개선하는 업무 프로세스 역시 정의되어야 한다.

또한 아카이브의 개인정보 보호 정책 전반을 총괄할 책임자 지정이 필요하다. 개인정보 보호 책임자는 개인정보 보호 정책을 개발하고 현행화하며, 정보주체의 권리 침해 모니터링을 수행해야 한다. 그리고 아카이브에서 기록보존 활동 과정에서 발생할 수 있는 위험요인을 사전에 식별하기 위한 개인정보 영향 평가 제도를 운영해야 하며, 이와 더불어 아카이브 직원을 대상으로 정기적인 교육 훈련을 실시하고 외부 관련 기관 및 단체와의 소통과 협력을 수행하는 역할을 담당해야 한다.

4.2 맥락 정보 기반 개인정보 보호 체계 구축을 위한 요건

개인정보 보호를 위해 아카이브에서는 정보의 민감도나 위험성을 예측하고 해당하는 부분을 가명처리 혹은 익명처리하거나 선제적으로 비공개하는 조치를 취할 수 있다. 보호 대상 개인정보를 식별하여 분류하는 업무의 자동화 지원 도구는 인력 자원의 한계를 고려했을 때 미룰 수 없는 과제이다. 국가기록원에서는 공개재분류 사업을 통해 축적한 공개재분류 기준서를 바탕으로 비공개정보 중 개인정보 대상 유형을 세분화하고 학습데이터를 구축하여 자동화하는 연구를 진행하였다(국가기록원, 2020). 이 연구는 개인정보의 세부 유형 가운데 주로 패턴화가 가능한 정보를 필터링하여 마스킹 모듈을 적용하였다. 주로 이메일이나 주민등록번호와 같은 정규 표현식 패턴이나, 글자수 패턴, 주소DB, 사전에 정의된 단어사전 등을 활용한 개인정보의 필터링 효과성을 검증하는 것에 중점을 두고 있어, 맥락에 따른 재식별 위험성 등은 고려되지 않았다. 이 연구에서 중앙행정기관의 이관 기록에 대한 공개재분류 기준서를 분석하여 비공개대상 개인정보 유형을 도출한 부분은 향후 개인정보 보호 수준의 차등화 방안 연구에 활용할 수 있다. 업무 분류체계 및 개인정보 유형별 비공개 대상 세부정보 예시는 <표 3>과 같다. 이와 같이 기록 유형별로 보호해야 할 세부정보 유형을 분류하는 것은 유형별로 차등화된 민감도 검토 도구 설계 및 도메인별 재식별 위험 모델 등을 설계하고 가중치를 부여하기 위한 연구의 기초 자료로 활용할 수 있을 것이다.

<표 3> 비공개정보대상 6호 유형(예)

구분	항목	세부 항목 예시
업무분류체계	감사	전과, 징계사항, 자격사항, 사건번호 진술조서, 고발서, 구속영장 감찰카드, 감사결과 중 관련자의 소속, 직위, 성명, 관리기관, 담당업무 등의 정보, 연수교육 불참자명단
	인·허가	지문번호, 인적사항, 면허번호 재산증명서, 재직증명서, 취임승낙서, 범죄경력조회 등 임대차계약서, 예금잔액증명서, 자본현황, 주주별 소유주식(지분)현황
	학적관리	졸업사항, 활동경력, 학번, 학과, 학년, 학력, 특기사항, 학적부, 성적표, 입학원서, 전역증, 어학능력확인서
.....		
공통 개인정보	개인식별정보	이름, 성별, 본과, 본적, 주소, 출생지, 전화번호, 생년월일, 주민등록번호, 연령, 직업, 소속, 직위, 직업, 계좌번호, 여권번호
	개인증빙기록	주민등록표, 등기부등본, 인감증명서, 인감신고서, 주민등록초본, 주민등록등본, 인사기록카드, 신원증명서, 공무여행심사기록, 대학 졸업증명서
개인정보 유형	가족정보	가족구성원 이름, 성별, 관계, 출생지, 부양가족사실증명
	의료정보	의료기록, 채용신체검사서, 건강검진결과, 건강카드, 진단세부기록서
	통신정보	전자우편, 전화통화내용, 로그파일, 쿠키정보
	위치정보	GPS나 휴대폰에 의한 개인의 위치정
.....		

출처: 국가기록원(2020) pp.24~30 일부 발췌

https://doi.org/10.14404/JKSARM.2025.25.3.331

맥락에 따른 개인정보 민감도 검토를 위해서는 어휘를 분석하여 단어의 빈도, 주변 단어와의 조합, 중요도 등을 수치화하고, 품사 구문 시퀀스로 문장의 구조를 파악한 후, 문장 내의 의미 기반 분석, 즉 행위주체와 대상, 내용 관계 등을 파악하여 벡터화하여 민감도 검토 결과를 제시하는 기술(Technology-Assisted Review)을 적용해 볼 수 있다(McDonald, 2019). McDonald(2019)는 정보는 맥락에 따라 민감도가 달라진다는 것을 전제로 텍스트 분류를 통해 맥락을 분석하여 민감도를 판단하고 검토 우선순위를 선정하며, 실제 작업자의 피드백까지 반영하여 학습모델로 활용할 수 있는 프레임워크를 제안하였다. 그러나 아카이브의 다양한 유형의 기록을 대상으로 텍스트 분석을 위한 학습 모델을 구축하는 것은 기술의 복잡성과 높은 비용을 필요로 한다. 현실적인 방안으로 아카이브 내에 이미 구조화된 메타데이터 활용을 우선 고려해 볼 수 있다. 이는 복잡한 기술 구현보다 상대적으로 쉽고 비용을 절감할 수 있다는 장점이 있다. 맥락 정보로써 활용 가능한 현재 메타데이터 표준의 요소에는 생산기관 정보, 제목, 기술 정보, 주제 정보, 기록 유형, 매체 유형, 분류 정보 일시, 생산이력 정보, 권한 정보, 관리이력 정보, 이용이력 정보, 다른 기록과의 관계 정보 등이 포함되어 있다.

물론 이와 함께 장기적으로 개인정보 민감도 검토를 위한 텍스트 분류 기술의 도입을 준비해야 한다. 전통적인 텍스트 마이닝 기법으로 단어의 빈도를 통계적 수치로 분석하여 가중치를 부여하는 TF-IDF(Term Frequency-Inverse Document Frequency) 기술이나 머신러닝 알고리즘의 지도학습 모델 SVM(Support Vector Machine)에 의한 텍스트 분류 적용은 활발히 활용되고 있으나, 개인정보 보호 분야에 적용하려는 사례는 많지 않다. 최근 트랜스포머 계열의 BERT 모델을 적용한 텍스트 분류 기법이 정보자유법 대응 업무에 적용하기 위한 실험이 진행된 바 있다. 정보공개 요청에 대응하기 위한 분석 지원 솔루션(MITRE FOIA Assistant™)을 개발하여 연방정부 기록을 대상으로 테스트하였는데, 이름과 웹사이트, 전화번호, 사회보장번호와 같은 개인식별 정보를 탐지하는 기능을 제공한다.[8)]

여기에 더하여 전거 정보 DB나 인명 식별을 위한 외부 데이터셋 자원을 함께 활용한다면 자연어 처리 기술의 개체명 인식 정확도를 제고하여 개인정보 검출률 향상에 기여할 수 있다. 국립중앙도서관에서 운영하는 국가전거 데이터베이스를 활용하여 인명, 단체명, 지리명, 주제명 등의 식별률을 높일 수 있다. 또한 한국지능정보사회진흥원이 운영하는 개체명 태깅이 부가된 인공지능 학습용 데이터세트도 활용해 볼 수 있다.

이미 구축된 메타데이터 등의 맥락 정보를 활용하고, 텍스트 분석 결과와 외부 자원과의 연결성을 종합하여, 민감도 수준이나 재식별 위험도를 스코어로 표기하는 등의 수치화 알고리즘을 설계할 필요가 있다. 개인정보 민감도 및 재식별 위험도는 맥락과 다른 정보와의 연결성 등을 종합적으로 고려하여 판단해야 하므로 작업자마다 결과가 불일치할 수 있는 주관적 판단의 영역으로 작동되는 경우가 많다. 정량화된 수치 기준을 제시함으로써 주관적 판단을 대체하여 일관된 기준을 적용할 필요가 있다. 또한 스코어링 프레임워크에 따른 결괏값이 높을 경우 우선 처리 대상으로 분류하거나 주의해서 처리할 수 있는 식별지표로 활용할 수 있다. 아카이브의 관리 대상 기록은 더욱 증가하고 있지만 한정된 인력 및 자원을 운용해야 하는 상황에서 민감도가 높은 기록부터 우선 검토할 수 있다. 키워드 중심의 필터링이나 반복되는 패턴을 식별하는 것이 아니라 문맥에 따른 세분화된 내용과 주제 특성을 반영한 종합적인 판단을 가능하게 함으로써 동일 정보라도 상황에 따른 평가를 지원할 수 있다. 또한 평가 결과를 데이터화 함으로써 기록 집합체별, 혹은 생산 기관별, 업무 기능별 위험 분포도 등의 시각화 자료를 제공할 수 있어 기록에 대한 심층적인 이해와 분석이 가능할 것이다. 이와 더불어 평가 결과 자체를 메타데이터로 관리하여 시간에 따른 민감도 변화 추이를 모니터링하고 지속적인 재분류 업무에도 활용할 수 있을 것이다.

이러한 개인정보의 민감도 검토 체계, 재식별 위험성 평가 도구, 비식별화 조치 솔루션 등의 개발은 기존의

8) Kamb, L. (2023). Some U.S. government agencies are testing out AI to help fulfill public records requests. NBC News. Available: https://www.nbcnews.com/news/us-news/federal-agencies-testing-ai-foia-concerns-rcna97313

https://doi.org/10.14404/JKSARM.2025.25.3.331

아카이브시스템과 연계된 통합 모듈로 설계되어야 한다. 기록의 생애주기 전 과정에 이르는 일관된 개인정보 보호 체계를 구축하기 위해서는 기록관리 단계에서 획득되는 맥락 정보의 체계적인 획득과 관리가 필요하기 때문이다. 특히 민감도 검토와 재식별 위험성 평가는 시간의 경과에 따라 변화하는 환경에 영향을 받기 때문에 지속적인 환경 변화 모니터링과 개선 과정이 이루어져야 하며, 이를 통해 알고리즘의 정확도를 제고하고 정책적 타당성, 신뢰성을 확보하는 것이 중요하다.

5. 결론

본 연구는 아카이브의 특성을 고려한 개인정보 보호 정책 개발의 필요성을 검토하고 맥락 정보 기반의 개인정보 보호 정책 설계 방향과 제도적 기반을 제시하였다. 기존의 단편적인 개인정보 검출 및 삭제 기술 조치만으로는 개인정보 보호의 실효성을 확보하기 어렵다는 점에서 기록의 맥락을 고려한 민감도 검토와 위험 평가 체계가 핵심임을 강조하였다. 연구를 통해 도출한 주요 결과는 다음과 같다.

먼저 개인정보 보호법 체계와 공공기록물법 관계에서 기록의 공개와 보호를 위한 충돌지점과 제도적 공백을 확인하고, 아카이브 관점의 제도적 기반 마련의 필요성을 도출하였다. 그리고 아카이브의 개인정보 보호 정책과 실질적인 가이드라인이 부재함으로써 과도한 비공개 판단이 이루어지고, 이로 인해 공개 실무가 경직화될 우려가 있음을 지적하였다.

또한 상용화된 개인정보 보호 솔루션이 개인정보의 식별과 삭제 처리에 집중하고 있어 이러한 조치만으로는 정보주체의 권리 보호에 한계가 있음을 논증하였다. 맥락이나 다른 정보와의 연계성을 통해 재식별되는 위험성이 존재하므로, 기록이 갖는 구조와 맥락 고려가 필요하다는 것을 확인하였다.

이러한 논의를 종합하여 개인정보 보호 정책의 실효성 있는 설계 방향을 제안하였다. 아키비스트의 윤리적 고려사항을 반영한 보호 정책 수립이 필요하며, 다원화된 전문가 협의체의 구성, 전담조직 체계 운영, 정책 이행에 요구되는 관련 규정의 정비와 함께 기술적 요구사항을 반영한 통합적 접근이 필요함을 시사하였다. 이는 향후 아카이브에서의 개인정보 보호 제도의 개선과 실행 체계 구축을 위한 요건으로 실질적 기여를 할 수 있을 것이다.

그러나 본 연구에서는 비공개 사유 중 개인정보 보호로 분류된 비공개 기록에 대한 정량적 분석이나 데이터 기반 실증 연구를 수행하지 못한 한계가 있다. 향후 개인정보 보호 가이드라인이나 평가 도구의 개발 과정에서는 실제 기록을 대상으로 효과성을 검증하고 정책을 구체화하는 과정이 필요할 것이다.

아카이브에서 개인정보 보호는 정보주체의 권리를 보호하는 것과 기록 공개 확대 사명 간의 균형을 찾는 것이 중요하다. 개인정보 보호 기술의 궁극적 목적은 비공개정보 검출과 비식별 처리를 넘어 정보주체의 권리를 안전하게 보장하는 것에 있다. 이 연구가 기록의 공적 활용을 확장하는 실질적인 도구를 마련하는 데 단초를 제공할 수 있기를 바란다.

참고문헌

개인정보 보호법. 법률 제10465호.

개인정보 보호법. 법률 제19234호.

개인정보보호위원회 (2021. 12. 17.). 한국, EU 「개인정보보호 적정성 결정」 최종 통과 [보도자료]. 출처:

https://doi.org/10.14404/JKSARM.2025.25.3.331

https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=7749
개인정보보호위원회 (2024a). 가명정보 처리 가이드라인.
개인정보보호위원회 (2024b). 개인정보 처리 통합 안내서(안).
개인정보의 안전성 확보조치 기준. 개인정보보호위원회고시 제2023-6호.
공공기관의 정보공개에 관한 법률. 법률 제19408호.
공공기관의개인정보보호에관한법률. 법률 제4734호.
공공기록물 관리에 관한 법률. 법률 제20309호.
과학기술정보통신부 (2020. 1. 9.). 데이터 3법 개정을 계기로 데이터 산업 육성 지원 강화 [보도자료]. 출처: https://www.msit.go.kr/bbs/view.do?sCode=user&mId=307&mPid=208&bbsSeqNo=94&nttSeqNo=2486388
국가기록원 (2021). 전자기록물 공개재분류를 위한 비공개정보 필터링 및 마스킹 기술 적용방안 연구.
국가기록원 (2024. 1. 2.). 2023년 공개재분류 결과 비공개 기록물 유형별 현황 (국가기록원 공고 제2023-17호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=101524
국가기록원 (2025a). 2025년 국가기록관리 · 활용기술 연구개발 사업 「AI를 활용한 비공개기록물 공개재분류 및 비식별화 서비스 모델 연구」 제안요청서.
국가기록원 (2025b. 1. 10.). 2024년 공개재분류 결과 비공개 기록물 유형별 현황 (국가기록원 공고 제2025-21호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=101802
권미현 (2019. 10. 16.). 비공개기록물 공개재분류 업무절차 개선 방안. 제8차 기록관리 연구세미나, 대전.
권은정 (2020). 공공데이터 영역 리스크 관리에 관한 법적 소고: 데이터 개방에 따른 개인정보 관련 리스크를 중심으로. 행정법연구, 60, 165-190. https://doi.org/10.35979/ALJ.2020.02.60.165
김병필 (2023). 인공지능 개인정보 보호 기술과 개인정보 보호 법제의 과제. 법경제학연구, 20(1), 113-152. https://doi.org/10.46758/kjle.2023.04.20.1.113
김승환, 전성해 (2019). 데이터 비식별화를 이용한 빅데이터 통합. 한국지능시스템학회 논문지, 29(3), 235-241. https://doi.org/10.5391/JKIIS.2019.29.3.235
김유승 (2022). 국회 비공개 대상 정보 세부 기준 연구: 「국회정보공개규정」을 중심으로. 한국기록관리학회지, 22(3), 37-53. https://doi.org/10.14404/JKSARM.2022.22.3.037
김유승 (2023). 행정각부 비공개 대상정보 세부기준 개선방안 연구. 한국기록관리학회지, 23(3), 115-136. https://doi.org/10.14404/JKSARM.2023.23.3.115
김일환 (2008). 전자정부와 개인정보보호. 공법연구, 37(1-1), 339-372.
김정수 (2025). 개인정보 보호법과 가명정보 및 가명처리에 관한 헌법적 검토: 헌법재판소 2023.10.26. 선고 2020헌마1476 결정을 중심으로. 연세법학, 47, 75-116. https://doi.org/10.33606/YLA.47.3
박노형, 김효권 (2022). 자동화된 결정에 관한 개인정보보호법 정부 개정안 신설 규정의 문제점: EU GDPR과의 비교 분석. 사법, 1(62), 361-390. https://doi.org/10.22825/juris.2022.1.62.010
손영화 (2022). 정보통신기술의 발전과 개인정보 보호. 기업법연구, 36(1), 171-211. https://doi.org/10.24886/BLR.2022.03.36.1.171
안혜미 (2019). 원문정보공개 서비스에서의 개인정보 보호 실태. 한국기록관리학회지, 19(2), 147-172. https://doi.org/10.14404/JKSARM.2019.19.2.147
엄수현, 이인경, 이우기 (2018). 빅데이터 기반 개인정보 비식별화 동향. 정보화연구, 15(4), 545-552. https://doi.org/10.22865/jita.2018.15.4.545
윤연화, 이은주 (2021). 부산 지방자치단체의 기록물 공개관리에 관한 연구. 한국기록관리학회지, 21(1), 57-73. https://doi.org/10.14404/JKSARM.2021.21.1.057

이규철 (2013). 新기술(빅데이터) 등장과 개인정보의 보호. 과학기술법연구, 19(1), 3-36. https://doi.org/10.32430/ilst.2013.19.1.3

이양복 (2020). 데이터 3법의 분석과 향후과제. 비교사법, 27(2), 423-465. https://doi.org/10.22922/jcpl.27.2.202005.423

임진희 (2021). 공문서의 기계가독형(Machine Readable) 전환 방법 제언. 기록학연구, 67, 99-138. https://doi.org/10.20923/kjas.2021.67.099

임희연 (2016). 기록관에서의 공개재분류 제도 개선 방안: 서울특별시교육청 사례 중심. 기록학연구, 49, 277-297. https://doi.org/10.20923/kjas.2016.49.277

정진명 (2019). 블록체인 기술과 개인정보 보호의 법률문제. 법조, 68(2), 248-280. https://doi.org/10.17007/klaj.2019.68.2.008

정혜영 (2024). 개정 개인정보보호법의 분석과 평가: 개인정보자기결정권의 범위와 한계를 중심으로. 동아법학, 102, 1-35. https://doi.org/10.31839/DALR.2024.02.102.1

천지영, 노건태 (2020). 데이터 3법 시대의 익명화된 데이터 활용에 대한 제언. 정보보호학회논문지, 30(3), 503-512. https://doi.org/10.13089/JKIISC.2020.30.3.503

최원상, 이종용, 신진 (2019). 4차 산업혁명기 인공지능과 빅데이터 운용을 위한 개인정보 보호와 이용에 관한 연구. 융합보안 논문지, 19(5), 63-73. https://doi.org/10.33778/kcsa.2019.19.5.063

표준 개인정보 보호지침. 개인정보보호위원회 고시 제2025-4호.

한국소비자연맹, 경제정의실천시민연합, 소비자시민모임, 진보네트워크, 참여연대 (2023. 7. 23.). [공동논평] 정보주체 권리 외면, 개인정보 무분별한 유통 조장 마이데이터 사업 중단하라. 참여연대. 출처: https://www.peoplepower21.org/publiclaw/1971552

행정안전부 (2019. 4. 30.). 2018년 공개재분류 결과 비공개 기록물 유형별 현황 (행정안전부 공고 제2019-269호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100844

행정안전부 (2020. 3. 27.). 2019년 공개재분류 결과 비공개 기록물 유형별 현황 (행정안전부 공고 제2020-120호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100912

행정안전부 (2021. 3. 19.). 2020년 공개재분류 결과 비공개 기록물 유형별 현황 (행정안전부 공고 제2021-159호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100527

행정안전부 (2022. 4. 4.). 2021년 공개재분류 결과 비공개 기록물 유형별 현황 (행정안전부 공고 제2022-215호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100387

행정안전부 (2023. 4. 12.). 2022년 공개재분류 결과 비공개 기록물 유형별 현황 (행정안전부 공고 제2023-588호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=101323

행정안전부 (2024). 2024 정보공개 연차보고서.

행정자치부 (2016. 3. 28.). 2015년 공개재분류 결과 비공개 기록물 유형별 현황 (행정자치부 공고 제2016-86호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100034

행정자치부 (2017. 3. 8.). 2016년 공개재분류 결과 비공개 기록물 유형별 현황 (행정자치부 공고 제2017-86호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=99864

행정자치부 (2018. 2. 20.). 2017년 공개재분류 결과 비공개 기록물 유형별 현황 (행정자치부 공고 제2018-109호). 출처: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100855

헌법재판소 2020헌마1476 (2023. 10. 26.)

황진현, 임지민, 변우영, 임진희 (2021). 공공기관 '비공개 세부 기준' 개발 전략. 한국기록관리학회지, 21(1), 117-139. https://doi.org/10.14404/JKSARM.2021.21.1.117

de Groot, A. & van der Sloot, B. (2018). Privacy from an archival perspective. In van der Sloot, B. & de Groot A. (Eds.). The handbook of privacy studies: An interdisciplinary introduction. Amsterdam: Amsterdam University Press.

European Union (2016). General Data Protection Regulation 2016/679. Available: https://gdpr-info.eu/

Gartner (2021, October 18). Top Strategic Technology Trends for 2022. Available: https://www.gartner.com/en/documents/4006913

Health Insurance Portability and Accountability Act of 1996. Pub. L. No. 104-191.

International Council on Archives (1997). Guide for Managing Electronic Records froman Archival Perspective. Available: https://www.ica.org/app/uploads/2023/12/ICA-Study-8-guide_eng.pdf

International Council on Archives (2012). Principles of Access to Archives. Available: https://www.ica.org/resource/principles-of-access-to-archives/

International Organization for Standardization & International Electrotechnical Commission (2018). Privacy enhancing data de-identification terminology and classification of techniques (ISO/IEC 20889: 2018).

International Organization for Standardization & International Electrotechnical Commission (2019). Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines (ISO/IEC 27701: 2019).

Jaillant, L. & Rees, A. (2023). Applying AI to digital archives: Trust, collaboration and shared professional ethics. Digital Scholarship in the Humanities, 38(2), 571-585. Available: https://doi.org/10.1093/llc/fqac073

Kamb, L. (2023). Some U.S. government agencies are testing out AI to help fulfill public records requests. NBC News. Available: https://www.nbcnews.com/news/us-news/federal-agencies-testing-ai-foia-concerns-rcna97313

Ketelaar, E. (1995). The right to know, the right to forget? Personal information in public archives. Archives & Manuscripts, 23(1), 8-17.

Lemieux, V. L. & Werner, J. (2024). Protecting privacy in digital records: The potential of privacy-enhancing technologies. ACM Journal on Computing and Cultural Heritage, 16(4), 1-18. https://doi.org/10.1145/3633477

McDonald, G. (2019). A framework for technology-assisted sensitivity review: Using sensitivity classification to prioritise documents for review. Doctoral dissertation, University of Glasgow, United Kingdom.

National Archives and Records Administration (2025). Code of Federal Regulations: Title 36 - Parks, Forests, and Public Property. Available: https://www.ecfr.gov/current/title-36

National Institute of Standards and Technology (2023). De-identification techniques and best practices(NIST SP 800-188). Available: https://doi.org/10.6028/NIST.SP.800-188

Society of American Archivists (2020). Archives Policy Manual. Available: https://www2.archivists.org/sites/all/files/Archives-Policy-Manual-2020.pdf

Society of American Archivists (n.d.). Dictionary of Archives Terminology. Available: https://dictionary.archivists.org/index.html

Sweeney, L., Yoo, Ji Su, Perovich, L., Boronow, K. E., Brown, P., & Brody, J. G. (2017). Re-identification risks in HIPAA Safe Harbor data: A study of data from one environmental health study. Technological Science. Available: https://techscience.org/a/2017082801

The National Archives (2018). Guide to archiving personal data. Available: https://cdn.nationalarchives.gov.uk/documents/information-management/guide-to-archiving-personal-data.pdf

Windon, K. & Youngblood, J. (2024). Privacy considerations in archival practice and research. In Lacity, M. & Coon, L. (Eds.), Human privacy in virtual and physical worlds: Multidisciplinary perspectives. London: Palgrave Macmillan, 205-234.

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

Ahn, Hye-mi (2019). The Status of Personal Information Protection for Original Text Information Disclosure Service. Journal of Korean Society of Archives and Records Management, 19(2), 147-172. https://doi.org/10.14404/JKSARM.2019.19.2.147

Choi, Wonsang, Lee, Jong yong, & Shin, Jin (2019). A Study on the Protection and Utilization of Personal Information for the Operation of Artificial Intelligence and Big Data in the Fourth Industrial Revolution. Jouranl of Information and Security, 19(5), 65-73. https://doi.org/10.33778/kcsa.2019.19.5.063

Chun, Ji young & Noh, Geontae (2020). Suggestions for Applications of Anonymous Data under the Revised Data Privacy Acts. Journal of the Korea Institute of Information Security & Cryptology, 30(3), 503-512. https://doi.org/10.13089/JKIISC.2020.30.3.503

Chung, Jin-myung (2019). Blockchain Technology and Legal Issues of Privacy. Korean Lawyers Association Journal, 68(2), 248-280. https://doi.org/10.17007/klaj.2019.68.2.008

Constitutional Court of Korea 2020Hun-Ma1476 (2023, October 26).

Consumers Union of Korea, Citizens' Coalition for Economic Justice, Consumers Korea, Korean Progressive Network, People's Solidarity for Participatory Democracy (2023, July 23). [Joint comment] Stop the MyData Project: It Ignores Data Subject Rights and Fuels Reckless Distribution of Personal Information. People's Solidarity for Participatory Democracy Newspaper. Available: https://www.peoplepower21.org/publiclaw/1971552

Eom, Soo-hyun, Lee, In-kyung, & Lee, Wookey (2018) BigData-based Trend of Personal Information De-idendification. Journal of Information Technology and Architecture, 15(4), 545-552. https://doi.org/10.22865/jita.2018.15.4.545

Hwang, Jinhyun, Lim, Jimin, Byeon, Wooyeong, & Yim, Jinhee (2021). Strategies for the Development of "Detailed Nondisclosure Standards" for Public Institutions. Journal of Korean Society of Archives and Records Management, 21(1), 117-139. https://doi.org/10.14404/JKSARM.2021.21.1.117

Jung, Hye young (2024). Analysis and evaluation of the revised Personal Information Protection Act: Focusing on the scope and limits of the right to self-determination of personal information. DONG-A LAW REVIEW, 102, 1-35. https://doi.org/10.31839/DALR.2024.02.102.1

Kim, Byoung-pil (2023). Privacy-Enhancing Technologies for AI and the Challenges of Legal Frameworks. Korean Journal of Law and Economics, 20(1), 113-152. https://doi.org/10.46758/kjle.2023.04.20.1.113

Kim, Il hwan (2008). E-Government and Personal Information Protection. Korean Public Law Association, 37(1-1), 339-372.

Kim, Jeong-soo (2025). Constitutional Review of the Personal Information Protection Act and Pseudonymous Information and Pseudonymization: Based on the decision of the Constitutional Court on October 26, 2023, 2020 Hun-Ma1476 -. Yonsei Law Journal, 47, 75-116. https://doi.org/10.33606/yla.47.3

Kim, Seung whan & Jun, Sunghae (2019). Big Data Integration using Data De-identification.Journal of Korean Institute

of Intelligent Systems. Journal of Korean Institute of Intelligent Systems, 29(3), 235-241. https://doi.org/10.5391/jkiis.2019.29.3.235

Kim, Youseung (2022). A Study on Detailed Nondisclosure Criteria for the National Assembly: Focused on National Assembly Information Disclosure Regulations. Journal of Korean Society of Archives and Records Management, 22(3), 37-53. https://doi.org/10.14404/JKSARM.2022.22.3.037

Kim, Youseung (2023). A Study on Detailed Nondisclosure Criteria for the Administrative Departments. Journal of Korean Society of Archives and Records Management, 23(3), 115-136. https://doi.org/10.14404/JKSARM.2023.23.3.115

Kwon, Eunjeong (2020). A Legal Study on Data Risk Management in the Public Sector: Focused on Personal Information Risk Associated with Opening up Public Data. Administrative law journal, 60, 165-190. https://doi.org/10.35979/alj.2020.02.60.165

Kwon, Mi hyun (2019, October 16). Improving procedures for public reclassification of previously restricted records. Conference presentation, 8th Records Management Research Seminar, Daejeon, Korea.

Lee, Kure chel (2013). The Protect of Privacy in the new Technology(Big Data) Society. Hannam Journal of Law & Technology, 19(1), 3-36. https://doi.org/10.32430/ilst.2013.19.1.3

Lee, Yang-bok (2020). A Study on the Revision Trend of Data 3 Act. The Journal of Comparative Private Law, 27(2), 423-465. https://doi.org/10.22922/jcpl.27.2.202005.423

Lim, Heeyeon (2016). Improvement of access re-review in archives: The Seoul Metropolitan Office of Cases. The Korean Journal of Archival Studies, 49, 277-297. https://doi.org/10.20923/KJAS.2016.49.277

Ministry of Scieence and ICT (2020, January 9). Strengthening support for the data industry with the revision of the Three Data Law [Press release]. Available: https://www.msit.go.kr/bbs/view.do?sCode=user&mId=307&mPid=208&bbsSeqNo=94&nttSeqNo=2486388

Ministry of the Interior (2016, March 28). Status of Previously Restricted Records by Type Based on 2015 Reclassification Results (Ministry of the Interior Public Notice No.2016-86). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100034

Ministry of the Interior (2017, March 8). Status of Previously Restricted Records by Type Based on 2016 Reclassification Results (Ministry of the Interior Public Notice No.2017-86). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=99864

Ministry of the Interior (2018, February 20). Status of Previously Restricted Records by Type Based on 2017 Reclassification Results (Ministry of the Interior Public Notice No.2018-109). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100855

Ministry of the Interior and Safety (2019, April 30). Status of Previously Restricted Records by Type Based on 2018 Reclassification Results (Ministry of the Interior and Safety Public Notice No.2019-269). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100844

Ministry of the Interior and Safety (2020, March 27). Status of Previously Restricted Records by Type Based on 2019 Reclassification Results (Ministry of the Interior and Safety Public Notice No.2020-120). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100912

Ministry of the Interior and Safety (2021, March 19). Status of Previously Restricted Records by Type Based on 2020 Reclassification Results (Ministry of the Interior and Safety Public Notice No.2021-159). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100527

Ministry of the Interior and Safety (2022, April 4). Status of Previously Restricted Records by Type Based on 2021 Reclassification Results (Ministry of the Interior and Safety Public Notice No.2022-215). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=100387

Ministry of the Interior and Safety (2023, April 4). Status of Previously Restricted Records by Type Based on 2022 Reclassification Results (Ministry of the Interior and Safety Public Notice No.2023-588). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=101323

Ministry of the Interior and Safety (2024). 2024 Information Disclosure Annual Report.

National Archives of Korea (2021). A Study on the Application of Non-public Information Filtering and Masking Technology for Public Reclassification of Electronic Records.

National Archives of Korea (2024, January 2). Status of Previously Restricted Records by Type Based on 2023 Reclassification Results (National Archives of Korea Public Notice No.2023-17). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=101524

National Archives of Korea (2025a). Request for Proposal (RFP) for the 2025 National Research and Development Project on Records Management and Utilization Technologies: A Study on AI-based Reclassification and De-identification Service Models for Disclosing Previously Restricted Records.

National Archives of Korea (2025b, January 10). Status of Previously Restricted Records by Type Based on 2024 Reclassification Results (National Archives of Korea Public Notice No.2025-21). Available: https://www.archives.go.kr/next/newnews/announcementDetail.do?board_seq=101802

Official Information Disclosure Act. Act No. 19408.

Park, Nohyoung & Kim, Hyokwon (2022). A Critique of the Automated Decision-making Provision in the Proposed Amendment to the Personal Information Protection Act: Comparative Analysis with the EU GDPR. Juris, 1(62), 361-390. https://doi.org/10.22825/juris.2022.1.62.010

Personal Information Protection Act. Act No. 10465.

Personal Information Protection Act. Act No. 19234.

Personal Information Protection Commission (2021, December 17.). Final approval of Korea-EU 'Adequacy Decision' on personal data protection [Press release]. Available: https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=7749

Personal Information Protection Commission (2024a). Guidelines on the Processing of Pseudonymized Data.

Personal Information Protection Commission (2024b). Draft Integrated Guidelines on Personal Information Processing.

Protection of Personal Information Maintained by Public Institutions Act. Act No. 4734.

Public Records Management Act. Act No. 20309.

Son, Young-hoa (2022). Advances in Information and Communication Technology and Personal Information Protection. Business law review, 36(1), 171-211. https://doi.org/10.24886/BLR.2022.03.36.1.171

Standard Personal Information Protection Guidelines. Personal Information Protection Commission Notice No.2025-4.

Standards for Safeguarding Personal Information. Personal Information Protection Commission Notice No.2023-6.

Yim, Jin hee (2021). Suggestions on how to convert official documents to Machine Readable. The Korean Journal of Archival Studies, 67, 99-138. https://doi.org/10.20923/KJAS.2021.67.099

Yoon, Yeonhwa & Lee, Eun-ju (2021). A Study on Records Disclosure Management of Local Governments in Busan. Journal of Korean Society of Archives and Records Management, 21(1), 57-73. https://doi.org/10.14404/JKSARM.2021.21.1.057