

국내외 근대신문 디지털 서비스의 구조적 특성에 대한 비교분석 연구*

A Comparative Analysis of the Structural Characteristics of Modern-Era Newspapers Digital Services in Korea and Abroad

이혜은 (Hye Eun Lee)**

조상은 (Sang Eun Cho)***

초록

본 연구는 국내외 근대신문 디지털 서비스의 구조적 특성을 비교·분석하여 공공성과 접근성 확보를 위한 논의의 기초를 제공하는 것을 목적으로 하였다. 이를 위하여 국내 4개(대한민국 신문 아카이브, 고신문 디지털 아카이브, 근대한국학 DB, 네이버 뉴스 라이브러리)와 국외 3개(Chronicling America, Newspapers.com, British Newspaper Archive) 등 7개 사례를 대상으로 운영 주체 및 협력 구조, 콘텐츠 형태 및 디지털화 수준, 서비스 및 접근 방식, 메타데이터 구조, 저작권 및 이용 제한의 다섯 가지 항목을 분석하였다. 분석 결과, 운영 주체는 공공기관 단독 운영형, 공공-민간 파트너십형, 민간 기업 단독형, 대학 연구소형으로 나타났으며, 유형에 따라 서비스 지속성·개방성·메타데이터 표준화 수준이 상이하게 나타났다. 연구 결과를 토대로 지속 가능한 협력 운영 모델 설계, 기사 단위 메타데이터 표준화, 개방형 데이터 채널 마련, 저작권 고지 체계 구축의 시사점을 제시하였다.

ABSTRACT

This study aims to establish a strategic foundation for enhancing the public value and accessibility of modern newspaper archives through a comparative analysis of their structural characteristics in Korea and abroad. Seven representative cases were examined: four domestic platforms (Korea Newspaper Archive, Old Newspaper Digital Archive, Database of Modern Korean Studies, and Naver News Library) and three international platforms (Chronicling America, Newspapers.com, and British Newspaper Archive). The analysis focused on five key dimensions: governance and collaborative structures, content formats and levels of digitization, service delivery and access models, metadata structures, and copyright policies. The findings categorize these services into four operational models: public institution-led, public-private partnership, private enterprise-led, and university research institute-based. These models demonstrate notable differences in terms of service sustainability, openness, and metadata standardization. Based on these results, the study proposes practical strategies for developing sustainable collaborative frameworks, standardizing article-level metadata, enhancing data accessibility, and establishing systematic mechanisms for copyright disclosure.

키워드: 근대신문, 디지털 아카이브, 사례 비교분석, 메타데이터 표준화, 오픈 액세스

Modern-era Newspapers, Digital Archives, Comparative Case Analysis, Metadata Standardization, Open Access

* 본 연구는 2025년 국립중앙도서관 '한국 근대잡지 소장현황 및 활용방안 연구'의 일부를 수정·보완한 것임.

** 숙명여자대학교 문헌정보학과 교수(helee@sookmyung.ac.kr / ISNI 0000 0004 6814 4289) (제1저자)

*** 대림대학교 도서관미디어정보과 조교수(secho@daelim.ac.kr / ISNI 0000 0005 1358 5527) (교신저자)

논문접수일자 : 2026년 2월 15일 논문심사일자 : 2026년 2월 27일 게재확정일자 : 2026년 3월 5일

한국비블리아학회지, 37(1): 187-213, 2026. <http://dx.doi.org/10.14699/kbiblia.2026.37.1.187>

* Copyright © 2026 Korean Biblia Society for Library and Information Science

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.

1. 서론

1.1 연구의 배경 및 필요성

근대신문은 근대 사회에서 형성된 대표적인 연속간행물로서, 일정한 주기로 발행되며 정치·사회·경제·문화 전반의 정보를 동시에 전달하는 매체이다. 특정 주제나 분야를 체계적으로 정리한 단행본과 달리, 신문은 일상적 사건과 사회적 담론을 시계열적으로 축적한다는 점에서 근대 사회의 변화 과정을 연속적으로 추적할 수 있는 기록물로 기능해 왔다(윤정원, 1997). 본 연구에서 ‘근대신문’은 개항기부터 일제강점기를 거쳐 해방 전후까지, 대략 1880년대부터 1960년대까지의 시기에 발행된 정기 간행 신문을 지칭하는 것으로, 현재 국내 대표 디지털 아카이브인 대한민국 신문 아카이브가 수록 대상으로 설정한 1883년부터 1960년대까지의 발행 범위를 실질적 준거로 삼는다(국립중앙도서관, 발행년불명). 이 시기의 신문은 시의성(時宜性)과 연속성을 핵심 속성으로 지니며, 동일 매체 내에서 기사·논설·광고·공지 등 다양한 정보 유형이 병렬적으로 수록된다는 점에서, 근대 사회의 공적 담론과 여론 형성 과정을 반영하는 복합적 매체였음이 강조된다(유춘동, 오영식, 2021; 윤정원, 1997). 이러한 특성으로 인해 근대신문은 단행본이나 관찬 기록과는 구별되는 기록사적·매체사적 가치를 지닌 핵심적 1차 사료로 평가된다. 그러나 근대신문은 물리적 매체의 특성으로 인해 훼손과 멸실의 위험이 높아, 장기 보존과 안정적인 접근을 위한 제도적 관리가 지속적으로 요구되어 왔다. 이에 따라 국내외에서는 근대신문을 대상으로 한 디

지탈화와 아카이브 구축이 본격적으로 추진되었으며, 온라인 기반의 디지털 서비스 형태로 제공 범위와 이용 방식이 확대되고 있다(송주형, 2016). 이러한 변화는 근대신문을 단순한 보존 대상에서 벗어나, 학술 연구와 공공적 활용이 가능한 디지털 자원으로 전환시키는 계기가 되었다.

한편 근대신문 디지털 서비스는 공공기관, 학술기관, 민간 기업, 포털 서비스 등 다양한 운영 주체에 의해 구축·운영되고 있으며, 콘텐츠 구성과 디지털화 수준, 서비스 방식과 접근 구조에서 상이한 특성을 보이고 있다. 그럼에도 불구하고 기존 연구는 근대신문의 디지털화 필요성이나 개별 아카이브 구축 사례를 중심으로 논의되어 왔으며(곽승진 외, 2018; 송주형, 2016; 유춘동, 오영식, 2021; 이고은 외, 2025), 다양한 운영 주체와 서비스 유형을 동일한 분석 기준에서 구조적으로 비교한 연구는 제한적으로 이루어져 왔다.

이에 본 연구는 국내외 근대신문 디지털 서비스 사례를 대상으로, 콘텐츠와 서비스 구조의 특성을 비교·분석함으로써 근대신문 디지털 서비스의 유형과 구조적 특징을 규명하고자 한다. 이는 향후 근대신문 아카이브의 구축 및 운영 방향을 모색하고, 공공성과 접근성을 균형 있게 확보하기 위한 논의의 기초를 제공한다는 점에서 연구의 필요성을 지닌다.

1.2 연구의 내용 및 방법

본 연구는 국내외에서 구축·운영되고 있는 근대신문 디지털 서비스 사례를 대상으로, 디지털 아카이브 구축 방식과 서비스 제공 구조의 특

성을 비교·분석하는 것을 주요 내용으로 한다.

사례 선정을 위해 국내외 근대신문 디지털 서비스의 대표성과 활용 가능성을 고려한 선정 기준을 적용하였다. 사례 선정 기준은 관련 연구를 참조하여, 대표성, 디지털 아카이브 및 서비스체계, 희귀성 및 특화성, 콘텐츠의 다양성 및 연속성, 국제협력 및 오픈엑세스 활용성, 정책 활용성 및 확장 가능성의 여섯 가지 기준으로 구성하였다(〈표 1〉 참조). 이러한 기준에 따라 분석을 위한 국내 사례는 국가 차원의 근대신문 보존과 제공을 담당하는 「대한민국 신문 아카이브」를 포함하여, 대학 및 연구기관 중심의 「고신문 디지털 아카이브」, 「근대 한국학 DB」, 포털 기반의 근대신문 접근 서비스인 「네이버 뉴스 라이브러리」를 선정하였다.

국외 사례로는 공공 아카이브 기반의 「(미국)Chronicling America」와 민간 주도의 「(미국)Newspapers.com」, 「(영국)British Newspaper Archive」, 총 7개를 선정하였다(〈표 2〉 참조).

선정된 사례에 대해서는 운영 주체 및 협력 구조, 콘텐츠 형태와 디지털화 수준, 서비스 방식과 접근 방식, 메타데이터 구조, 저작권 및 이용 제한의 다섯 가지 분석 항목을 중심으로 비교·분석을 수행하였다. 이 분석 항목들은 근대신문 콘텐츠의 보존, 서비스, 활용 전반을 구조적으로 진단하고, 향후 근대신문 디지털 서비스의 발전 방향을 논의하기 위한 분석 틀로 설정되었다(〈표 3〉 참조).

연구 방법으로는 문헌분석과 사례 기반 질적 비교분석을 병행하였다. 문헌분석을 통해 근대신문의 매체적 특성과 디지털 아카이브 구축, 서비스 제공 방식에 관한 선행연구를 검토하였고, 사례 분석에서는 각 서비스의 콘텐츠 구성, 메타데이터 체계, 접근 및 이용 정책을 중심으로 구조적 특성을 분석하였다. 이 과정에서는 개별 서비스의 기능을 나열하기보다 어떠한 공통점과 차별성을 보이는지를 도출하는데 분석의 초점을 두었다.

〈표 1〉 사례 선정 기준 요약

기준	주요 고려 요소	참고문헌
대표성	공공성 높은 국가기관 중심, 안정적 운영 기반 및 정책 활용성	국립중앙도서관(2024)
디지털 아카이브 및 서비스 체계	메타데이터 구조화, 검색기능, UI/UX 등 정보서비스 구현 수준	국립중앙도서관(2024)
희귀성 및 특화성	희소한 간행물, 특정 주제·계층 특화 콘텐츠	송주형(2016), 유춘동, 오영식(2021), 홍종욱 외(2025),
콘텐츠의 다양성 및 연속성	정치, 사회, 문학 등 다분야 주제, 연대별 지속 간행물 여부	곽승진 외(2018), 국립중앙도서관(2024), 송주형(2016)
국제협력 및 오픈 액세스 활용성	국제 표준 연계 가능성	국립중앙도서관(2024), 송주형(2016), 이고은 외(2025)
정책 활용성 및 확장 가능성	메타데이터 표준화, 기관 연계, 콘텐츠 큐레이션 등 정책 수립·적용 가능성	국립중앙도서관(2024), 이고은 외(2025)

〈표 2〉 국내외 근대신문 디지털 서비스 사례 선정 목록

구분	기관명	서비스명	수록 범위	선정 기준
국내	국립중앙도서관	대한민국 신문 아카이브	1883년~1960년대 (107종, 864만 건 기사)	국가 주도, 근대신문 보존, 장기 아카이브 체계
	고려대학교 중앙도서관	고신문 디지털 아카이브	1898~1910년 (황성신문)	희귀본 디지털화, 학술 큐레이션
	연세대학교 근대한국학연구소	근대한국학 DB	1860년대~1910년대 (독립신문·황성신문·대한매일신보 외)	OCR, 원문 서비스, 학술 활용성
	네이버	네이버 뉴스 라이브러리	1920~1999년 (주요 5개 일간지)	포털 기반 민간 서비스, 근대신문 대중적 접근성
국외	(미국)Library of Congress	Chronicling America	1770년대~1963년 (수백만 페이지)	공공 접근성, 메타데이터 구조
	(미국)Ancestry Inc.	Newspapers.com	1700년대~2000년대 (10억 페이지 이상)	민관협력, 장기 디지털화
	(영국)British Library/Findmypast	British Newspaper Archive	1700년대~2000년대 초 (약 9천만 페이지, 영국·아일랜드)	민간 플랫폼, 대중적 접근

〈표 3〉 분석 항목 구성 및 선정 근거

항목	주요 내용	참고문헌
① 운영 주체 및 협력 구조	단독 혹은 공동 운영, 협력 기관 유무, 공공-민간 협업 여부	국립중앙도서관(2024)
② 콘텐츠 형태 및 디지털화 수준	스캔본, OCR 여부, 연대·주제별 분포 등	국립중앙도서관(2024), 유춘동, 오영식(2021)
③ 서비스 및 접근 방식	열람, 다운로드, 전시 및 교육 콘텐츠, 사용자 참여형 기능 포함 여부/온라인 플랫폼, 전용 사이트, 연계 포털, 모바일 최적화 등 기술적 접근 경로 구분	국립중앙도서관(2024), 송주형(2016), 이고은 외(2025)
④ 메타데이터 구조	메타데이터 체계의 구조화 수준, 표준화 여부, 교차검색 가능성, 개방성	홍종욱 외(2025)
⑤ 저작권 및 이용 제한	저작권 정보 제공, 오픈엑세스 가능, 이용범위 제한 등	유춘동, 오영식(2021)

2. 이론적 배경

2.1 근대신문의 매체적 특성과 기록학적 의미

근대신문은 근대 사회에서 형성된 대표적인 연속간행물로서, 일정한 주기로 발행되며 정치·사회·경제·문화 전반의 정보를 동시적으로 전달하는 매체이다. 신문은 특정 주제나

분야를 체계적으로 정리한 단행본과 달리, 일상적 사건과 사회적 담론을 시계열적으로 축적한다는 점에서 근대 사회의 변화 과정을 연속적으로 추적할 수 있는 기록물로 기능해 왔다. 이러한 특성으로 인해 근대신문은 근대 사회의 정보 환경과 지식 유통 구조를 분석하는 데 핵심적인 1차 자료로 평가된다(윤정원, 1997).

기록학적 관점에서 근대신문은 단순한 정보

전달 매체가 아니라, 사회적 권력과 집단 기억의 형성에 능동적으로 참여한 기록물로 이해된다. Schwartz와 Cook(2002)은 기록이 역사적 학문의 방향, 집단 기억, 국가 정체성을 형성하는 권력으로 작동하며, 기록의 생산·보존·활용 전 과정에 사회적 가치 판단이 내재한다고 논증하였다. 이러한 관점에서 근대신문은 당대 사회의 공적 담론과 정치적 담론을 반영하는 동시에, 그것을 선택·배치·유통한 발행 주체의 관점을 내포한 복합적 기록물이라 할 수 있다. Cook(2013)은 기록학 패러다임이 증거(evidence)에서 기억(memory), 정체성(identity), 공동체(community)로 이동해 왔음을 논증하면서, 기록을 단순한 행정적 증거물로 보는 관점을 넘어 사회적 기억과 정체성 구성의 매개로 재정립하였다. 근대신문은 이러한 패러다임 전환의 맥락에서, 과거 사회의 증거인 동시에 특정 집단의 기억과 정체성이 구성·경합된 현장으로 볼 수 있다.

또한, 매체적 관점에서 근대신문은 시의성과 연속성을 핵심 속성으로 지닌다. 신문은 정기적인 발행주기를 통해 사회적 사건을 즉각적으로 반영하며, 동일 매체 내에서 기사·논설·광고·공지 등 다양한 정보 유형이 병렬적으로 수록된다. 이는 근대신문이 단순한 정보 전달 매체를 넘어, 근대 사회의 공적 담론과 여론 형성 과정을 반영하는 복합적 매체였음을 보여준다. 유춘동과 오영식(2021)은 근대문헌의 범주를 논의하면서, 신문이 시대 변화에 따른 정보 생산과 유통 양상이 가장 집약적으로 드러나는 문헌 유형임을 강조하였다.

한편 근대신문은 종이 매체라는 물리적 한계로 인해 보존과 접근에 구조적 제약을 지녀 왔다. 방대한 분량과 연속성, 재질의 취약성은 신

문을 장기적으로 관리하고 활용하는 데 어려움을 초래하였다. 이러한 한계는 근대신문을 기록유산으로서 체계적으로 보존하고 활용하기 위한 디지털 아카이브 구축의 필요성을 제기하는 배경이 되었다. 송주형(2016)은 국내외 온라인 신문 아카이브 구축 사례를 분석하여, 디지털 환경이 신문 자료의 접근성과 활용 가능성을 근본적으로 확장시킨다는 점을 실증적으로 보여주었다.

디지털 환경에서 근대신문은 이미지 기반 원문 제공을 넘어, 텍스트 데이터와 메타데이터가 결합된 정보 자원으로 재구성되고 있다. 이 과정에서 Huvila(2008)가 제안한 참여형 아카이브(participatory archive) 개념, 즉 이용자의 참여와 자료에 대한 폭넓은 맥락화를 통해 아카이브의 접근성과 해석 가능성을 확장하는 방향은 근대신문 디지털 서비스의 발전 방향과 긴밀하게 연결된다. 미국 의회도서관의 Chronicling America는 이러한 방향에서 근대신문을 국가적 기록유산으로 인식하고, 표준화된 메타데이터와 장기 보존 전략을 통해 신문 자료를 공개하고 있다(Library of Congress, n.d.). 또한 대규모 신문 디지털 데이터셋을 활용한 연구에서는 신문이 단순한 열람 자료를 넘어, 데이터 분석과 정보 추출의 대상으로 확장되고 있음을 보여준다(Lee et al., 2020).

2.2 디지털 신문 아카이브의 개념과 유형

디지털 신문 아카이브는 인쇄 매체로 생산된 신문을 디지털 형태로 변환하여 장기적으로 보존하고, 온라인 환경에서 검색·열람·활용할 수 있도록 구축된 정보 관리 및 서비스 체계를

의미한다. 신문은 정부 비판·여론 형성·로컬리티 기록 등 다층적 기록학적 가치를 지니며, 기자의 취재 활동 지원에서 학문 연구 및 교육 활용에 이르기까지 다양한 목적으로 이용된다(송주형, 2016). 이에 따라 신문 아카이브는 원문 이미지의 디지털화를 넘어 체계적인 메타데이터 구축과 검색기능, 접근 및 이용 정책을 포괄하는 종합적 정보 시스템으로 발전해 왔으며, 이용자 요구 역시 단순 열람에서 디지털 원문 접근성 향상, 다양한 검색 기능, 기관 간 자료 연계, AI 기반 기능 도입 등으로 고도화되고 있다(이고은 외, 2025). 또한 근대문헌의 맥락에서는 단순한 이미지 아카이브를 넘어 온톨로지 기반의 시맨틱 데이터 아카이브 구축이 새로운 과제로 부상하고 있다(홍종욱 외, 2025).

기록학 분야의 아카이브 유형론은 전통적으로 운영 주체를 기준으로 공공기록관(public archives), 민간기록관(private archives), 특수기록관(special archives)으로 구분해 왔다. 그러나 Schwartz와 Cook(2002)은 이러한 전통적 유형론이 국가 주도의 공공기록관 중심으로 편향되어 있음을 비판하며, 아카이브의 유형은 운영 주체뿐 아니라 기록의 사회적 목적과 이용자 집단을 함께 고려하여 분류되어야 한다고 논증하였다. 나아가 Cook(2013)은 디지털 환경에서 공동체 아카이브(community archives)와 민간 주도의 접근 서비스가 기록학적으로 정당한 유형으로 확장되고 있음을 제시하였다. 이러한 논의는 디지털 신문 아카이브를 운영 주체와 목적에 따라 유형화하여 이해할 필요성을 뒷받침한다.

이에 본 연구는 선행연구를 종합하여 디지털 신문 아카이브를 네 가지 유형으로 구분한다.

첫째, 공공 아카이브형 서비스는 국가 또는 공공기관이 주도하여 신문을 기록유산으로 보존·공개하는 유형으로, 기록학의 전통적 공공기록관 유형에 해당하며 표준화된 메타데이터와 안정적인 관리 체계를 핵심으로 한다. 미국의 Chronicling America(Library of Congress, n.d.)와 국내의 대한민국 신문 아카이브(국립중앙도서관, 발행년불명)가 대표적 사례이며, Chronicling America에서는 딥러닝 기반 시각 콘텐츠 인식 모델을 적용하여 1,630만 페이지에서 헤드라인·사진·삽화·지도 등 7종의 시각 콘텐츠를 추출하는 등 데이터 고도화 시도가 이루어지고 있다(Lee et al., 2020). 둘째, 학술 지원형 데이터베이스는 대학 등 학술기관이 연구·교육 활용을 목적으로 구축한 유형으로, 기록학의 특수기록관 유형에 상응하며 고급 검색 기능과 텍스트 데이터 활용을 강조한다. 연세대학교 근대한국학연구소 DB와 고려대학교 고신문 디지털 아카이브 등 대학 기관 주도의 사례가 이에 해당한다. 셋째, 민간 아카이브형 서비스는 민간 기업이 주도하여 대규모 자료 축적과 사용자 편의성을 바탕으로 연구자와 일반 이용자 모두에게 접근 기회를 제공하는 유형으로, 기록학의 민간기록관 유형에 대응한다. Newspapers.com과 The British Newspaper Archive가 대표적이다. 넷째, 포털 기반 접근·이용형 서비스는 포털 플랫폼을 통해 디지털화된 신문에 대한 접근성과 가시성을 강화한 유형이다. 이 유형은 기록의 물리적 보존과 관리 주체의 분리를 특징으로 하는 탈관리주의적(post-custodial) 접근과 맥락이 닿아 있으며(Huvila, 2008), 기록학적 의미의 아카이브 기능—즉 장기 보존, 출처 관리, 메타데이터 표준

화—을 직접 수행하기보다는 이미 구축된 디지털 신문 자원에 대한 이용자 접근 경로를 제공하는 서비스 레이어로 기능한다. 네이버 뉴스 라이브러리가 이에 해당하며, 이 유형은 대중적 이용범위를 확대하는 데 기여하나, 메타데이터 표준화와 기록 관리 체계성 측면에서 상대적으로 제한적인 구조를 보일 수 있다(송주형, 2016).

이와 같이 디지털 신문 아카이브는 운영 주체와 서비스 목적, 제공 방식에 따라 서로 다른 유형적 특성을 지니며, 각 유형은 보존·연구·활용의 측면에서 상이한 강점과 한계를 갖는다.

2.3 선행연구 검토 및 본 연구의 차별성

근대신문과 디지털 신문 아카이브를 다룬 선행연구는 크게 세 가지 연구 흐름으로 구분된다. 첫째, 근대신문의 매체적 특성과 기록사적 위상을 규명한 연구들이다. 윤정원(1997)은 근대 정기간행물의 서지학적 분석을 통해 신문이 시의성과 연속성을 기반으로 사회적 사건과 담론을 동시에 기록한 자료임을 논증하였다. 유춘동과 오영식(2021)은 시대 변화에 따른 근대문헌 범주의 재설정 논의에서, 신문이 정보의 생산·유통 양상이 가장 집약적으로 드러나는 문헌 유형임을 강조하였다.

둘째, 기록학 이론의 관점에서 아카이브의 사회적 의미와 유형적 특성을 논의한 연구들이다. Schwartz와 Cook(2002)은 기록이 역사적 학문의 방향, 집단 기억, 국가 정체성을 형성하는 권력으로 작동하며, 운영 주체와 사회적 목적에 따른 아카이브 유형의 다양성을 인식해야

함을 논증하였다. Cook(2013)은 기록학 패러다임이 증거 중심에서 기억·정체성·공동체로 이동해 왔음을 제시하며, 디지털 환경에서 다양한 형태의 아카이브가 기록학적 정당성을 갖는다고 논증하였다. Huvila(2008)는 이용자의 참여와 자료의 맥락화를 강조한 참여형 아카이브 개념을 제시하여, 디지털 아카이브 서비스의 이용자 지향적 발전 방향을 이론화하였다. 또한 Ringel(2023)은 뉴욕타임스의 신문 디지털화 사례를 기록학적 관점에서 분석하여, 디지털 신문 아카이브가 단순한 디지털 복제물이 아니라 문화적 기억과 저널리즘 실천을 적극적으로 형성하는 기록 주체임을 실증하였다.

셋째, 디지털 환경에서의 신문 아카이브 구축 및 서비스 체계를 분석한 연구들이다. 송주형(2016)은 국내외 온라인 신문 아카이브 사례의 체계적 비교를 통해, 디지털화 수준·검색 기능·서비스 구조의 차이가 자료의 접근성과 활용 가능성을 결정짓는 핵심 변수임을 실증적으로 규명하였다. 국립중앙도서관(2024)은 정책 연구 차원에서 근대 문헌 디지털 서비스가 보존 중심의 기록 관리 단계를 넘어 이용과 재활용을 전제로 한 정보서비스 체계로 전환되어야 함을 제시하였다.

넷째, 이용자 요구와 서비스 활용을 분석한 연구들이다. 박승진 외(2018)는 이용자 인식조사를 통해 접근성·검색 편의성·맥락 정보 제공이 디지털 문헌 서비스의 핵심 품질 요소로 기능함을 실증하였으며, 이고은 외(2025)는 근대 문헌 통합 온라인 서비스 연구에서 이용자 요구가 단순 열람을 넘어 자료의 분석·재활용을 지향하는 방향으로 심화되고 있음을 확인하였다.

그러나 이러한 선행연구들은 공통적으로 개별 사례 분석이나 특정 서비스 유형에 한정된 논의에 머물러 있으며, 기록학 이론에 기반한 유형 분류 체계를 적용하면서 동시에 운영 주체·콘텐츠 구성·서비스 방식·메타데이터 구조·이용 정책을 아우르는 종합적 분석 틀을 적용하여 국내외 사례를 동일한 기준에서 구조적으로 비교한 연구는 충분히 축적되어 있지 않다. 본 연구는 이러한 공백을 바탕으로, 기록학 유형론과의 이론적 적합성을 확보하면서 근대신문 디지털 서비스를 운영 주체와 서비스 구조를 중심으로 체계적으로 비교·분석함으로써 유형적 특성과 구조적 차이를 규명하고자 한다.

3. 국내외 근대신문 디지털 서비스 사례 분석

3.1 국내 사례 분석

3.1.1 대한민국 신문 아카이브¹⁾

대한민국 신문 아카이브는 국립중앙도서관이 운영하는 근대신문 디지털 서비스로서, 1883년부터 1960년대까지 발행된 신문을 대상으로 체계적인 디지털화와 온라인 서비스를 제공하고 있다. 이는 국가 차원의 근대신문 보존과 공개를 목적으로 구축된 대표적인 공공 아카이브로 평가된다.

① 운영 주체 및 협력 구조

대한민국 신문 아카이브는 2013년부터 2015

년까지 미래창조과학부와 한국정보화진흥원이 추진한 국가 DB 사업의 일환으로 구축이 시작되었다. 이 시기에는 국립중앙도서관 소장 귀중본인 제국신문, 한성일보, 국민신보 등을 중심으로 디지털화 작업이 진행되었으며, 2016년 정식 서비스가 개시되었다. 이후 한국언론진흥재단, 국사편찬위원회 등 유관 기관으로부터 기증받은 고신문 DB를 단계적으로 통합하였으며(2017-2021년), 2022년부터 2024년까지는 조선신문, 경성일보 등 주요 신문의 디지털화와 메타데이터 표준화 작업을 수행하였다. 2026년 2월 기준으로 107종의 신문, 864만 건의 기사, 1,647만 건의 색인을 서비스하고 있으며, 국립중앙도서관의 자체 사업인 '지식정보자원 공유 기반 구축 사업'을 통해 고신문 디지털화를 지속적으로 추진하고 있다. 이러한 운영 구조는 국가 주도의 디지털화 사업과 유관 기관 간 협력 체계를 기반으로 하고 있으며, 공공 아카이브로서 장기적이고 안정적인 서비스 제공을 지향한다는 특징을 지닌다.

② 콘텐츠 형태 및 디지털화 수준

대한민국 신문 아카이브는 기사 단위와 원문면 단위의 이중 구조로 콘텐츠를 제공한다. 기사 단위로는 개별 기사의 서지정보와 썸네일 이미지를 제공하며, 원문면 단위로는 신문 전체 면의 고해상도 이미지를 제공한다. 원문 이미지는 책갈피, 자동 넘김, 화면 확대, 조판 보기 등의 뷰어 기능을 지원하나, 저작권 보호를 위해 화면 캡처 기능은 제한되어 있다.

디지털화 수준 측면에서는 이미지 기반 서비

1) <https://www.nl.go.kr/newspaper/>

스가 주를 이루며, 일부 자료에 한해 '텍스트 보기' 메뉴를 통해 OCR 처리된 텍스트를 제공하고 있다. 그러나 전체 자료에 대한 전문(全文) 텍스트 데이터 제공은 아직 제한적이며, 이는 향후 확대가 필요한 영역으로 평가된다. 원소장처 정보는 국립중앙도서관, 국사편찬위원회, 한국언론진흥재단, 관훈클럽, 경남신문사 등으로 명시하여 자료의 출처를 제시하고 있다.

③ 서비스 및 접근 방식

대한민국 신문 아카이브는 키워드 검색, 발행일 검색, 브라우징 방식의 다층적 검색 체계를 제공한다. 검색 결과는 기사 단위로 제시되며, 각 기사에는 해시태그 기능과 SNS 공유 기능(페이스북, X, URL 복사)을 결합하여 이용자 간 자료 공유를 지원한다([그림 1] 참조).

서비스 구조는 크게 세 가지로 구성된다. 첫째, 고신문 디지털 컬렉션은 시대별(대한제국 시기, 일제강점기, 군정기, 대한민국)과 주제별(경제, 문화, 사회, 스포츠, 정치, 학술)로 분류되어 있으며, 부동산, 전쟁, 3·1운동, 감염병, 독도·울릉도 등 특정 주제에 대한 큐레이션 서비스를 제공한다. 둘째, 통계 서비스는 신문 통계, 발행연도, 기사 주제, 선거인명, 선거단체, 선거 지명, 선거 사건 등의 데이터를 제공하며 엑셀 파일로 다운로드가 가능하다. 셋째, 신문 해제 서비스는 각 신문의 기사 건수, 연도별 데이터를 그래프로 시각화하고 감수자 정보와 함께 해제를 제공한다. 추가로 '100년 전 오늘', '주간 인기기사', '월간 인기기사' 등 대중적 관심을 유도하는 콘텐츠도 제공한다.

접근 방식은 완전 공개형으로, 별도의 회원 가입이나 로그인 없이 누구나 자유롭게 이용할

수 있다. 웹 브라우저 기반으로 설계되어 PC와 모바일 환경 모두에서 접근이 가능하며, 이는 공공 아카이브로서의 접근성 확보를 중시한 설계로 평가된다.

④ 메타데이터 구조

각 기사의 메타데이터는 신문명, 발행처, 발행일, 관련 주제어, 키워드, 기사위치, 기사유형, 기사주제, 주기사항 등으로 구성되어 있다. 주기사항에는 원소장처 정보와 함께 '한국언론진흥재단에서 원문 파일을 기증받아 이미지 파일로 변환하였음'과 같이 디지털화 구축 과정에 대한 설명이 포함되어 있어, 자료의 출처와 변환 이력을 제시하고 있다.

메타데이터 표준화 측면에서는 MODS(Metadata Object Description Schema) 형식을 제공하고 있으며, 국가서지 LOD(Linked Open Data)와 국가전자 데이터를 연계하여 신문 자료를 국립중앙도서관의 전체 지식정보 자원과 통합적으로 활용할 수 있는 구조를 구축하고 있다. 이는 근대신문 아카이브를 독립적인 데이터베이스가 아닌, 도서관 디지털 생태계의 일부로 설계하였음을 보여준다.

이용자 참여형 메타데이터 구조도 도입되어, 이용자가 직접 태그를 추가할 수 있도록 하고 있다. 다만 입력된 태그는 관리자의 검토를 거친 후 공개되는 정책을 적용하여, 메타데이터의 품질 관리와 이용자 참여를 동시에 고려하는 방식을 취하고 있다.

⑤ 저작권 및 이용 제한

대한민국 신문 아카이브의 저작권 및 이용 정책은 국립중앙도서관의 전체 디지털 자원 이

〈표 4〉 대한민국 신문 아카이브 주요 현황

구분	내용
운영 주체	국립중앙도서관
서비스 개시	2016년
수록 범위	1883-1960년대 신문 107종(864만 건 기사, 1,647만 건 색인)
협력 구조	한국언론진흥재단, 국사편찬위원회, 관훈클럽, 경남신문사 등
디지털화 수준	이미지 기반, 일부 OCR 텍스트 제공
접근 방식	무료 공개, 로그인 불필요
주요 기능	키워드 검색, 시대별/주제별 컬렉션, 통계 서비스, LOD 연계
메타데이터	MODS, 국가서지 LOD, 국가전거, 이용자 태그
저작권 정책	화면 캡처 제한, 메타데이터/통계 다운로드 허용

용 정책을 따르는 것으로 추정된다. 원문 이미지는 화면 캡처가 제한되어 있으며, 이는 저작권 보호와 무단 복제 방지를 위한 조치로 해석된다. 다만 메타데이터와 통계 정보는 엑셀 파일로 다운로드가 가능하여, 데이터 분석과 2차 활용을 일정 수준 허용하고 있다.

대한민국 신문 아카이브는 공공기관이 운영하는 아카이브로서 공개 접근을 원칙으로 하되, 원문 이미지의 직접적인 복제는 제한하는 절충적 정책을 적용하고 있는 것으로 평가된다. 이는 근대신문의 공공적 활용과 저작권 보호 사이의 균형을 추구하는 정책 방향을 보여준다.

3.1.2 고신문 디지털 아카이브²⁾

고려대학교 중앙도서관은 ‘아카이빙(Archiving)’ 메뉴를 통해 고신문 디지털 아카이브를 운영하고 있으며, 1898년부터 1910년까지 발행된 황성신문을 중심으로 서비스를 제공한다. 이 아카이브는 단일 기관 소장본에 한정되지 않고 고려대학교 도서관, 국립고궁박물관, 연세대학

교 학술정보원 등 복수 기관의 원본 자료를 통합하여 제공하는 협력형 구축 모델을 채택하고 있다는 점에서 특징적이다.

① 운영 주체 및 협력 구조

고려대학교 중앙도서관이 운영 주체로서 아카이브 시스템을 관리하며, 디지털화 자료의 원소장처는 고려대학교 도서관, 국립고궁박물관, 연세대학교 학술정보원 등 복수 기관으로 구성되어 있다. 서비스 화면의 개별 신문 호(號) 단위에는 원소장처가 명시되어 있으며, 예컨대 황성신문 여러 호의 원소장처가 연세대학교 학술정보원으로 표시되어 있어 외부 기관 원본을 바탕으로 한 연계 제공 구조임을 확인할 수 있다. 이러한 협력형 구축 방식은 단일 기관 중심의 아카이브와 차별화되는 ‘다기관 통합형 서비스 모델’로 평가된다.

② 콘텐츠 형태 및 디지털화 수준

고신문 디지털 아카이브는 1898년~1910년에 발행된 황성신문을 발간일자별로 제공하며,

2) <https://library.korea.ac.kr/datause/archiving/old-newspaper/old-newspaper/>

각 호는 면 단위로 세분화하여 원문 접근이 가능하도록 구축되어 있다. 목록 화면에는 원문 디지털 파일과 더불어 원소장처가 함께 표기되어 있어, 디지털화된 자료와 실물 소장처를 구분함으로써 보존 및 출처 확인의 신뢰성을 확보하고 있다.

면 단위 목록에서 '1면' 등 해당 버튼을 선택하면 원문이 PDF 형식의 고해상도 스캔 이미지로 열람되며, 서비스는 이미지 기반의 보존형 열람 구조를 전제로 한다. 다만 OCR(광학 문자인식) 기반의 텍스트 변환 및 기사 내 검색 기능은 제공되지 않아, 기사 수준의 키워드 검색보다는 발행일자·면 단위 중심의 탐색이 이루어진다. 이러한 제공 방식은 원형 보존의 측면에서 장점을 가지나, 학술 연구 및 데이터 분석 활용도를 높이기 위해서는 향후 OCR 적용을 통한 텍스트 검색 및 분석 지원 기능의 확장이 요구된다.

③ 서비스 및 접근 방식

고신문 아카이브는 신문명과 발행연도를 기준으로 한 연도별 리스트 검색 방식을 제공한다. 이용자는 특정 연도를 선택한 뒤 월·일 단위 발행 목록을 확인하고, 다시 면 단위로 접근할 수 있다. 각 발행일의 목록에는 신문명, 권호, 발간일자, 소장처, 원문 열람 버튼이 함께 제공되어, 디지털 서비스와 원본 소장처 정보를 병기하는 구조를 갖추고 있다.

'1면' 등 면 단위 버튼을 클릭하면 스캔된 원문이 PDF 형태로 열람되며, 뷰어 상에서 확대·축소, 페이지 이동 기능을 통해 지면 단위 열람이 가능하다. PDF 뷰어 상단의 다운로드 기능이 제공되어, 연구자는 원문 이미지를 개

별 파일로 내려받아 로컬 환경에서 보관·활용할 수 있다. 이는 단순 열람에 그치지 않고 데이터 셋 축적이나 연구 분석에 활용 가능성을 확장한다는 점에서 의미가 있다.

목록 화면에서 '엑셀 다운로드' 기능을 지원하여, 특정 연도나 전체 발행본의 서지 데이터를 일괄 내려받을 수 있다. 다운로드 된 파일에는 신문명, 발간일자, 권호, 소장처, 원문면수 등이 표 형식으로 포함되어 있어 연구자가 이를 데이터 셋으로 가공하거나 통계·시각화 연구에 활용할 수 있다. 다만 서비스 구조가 발행일·면 단위 나열 중심이므로 주제별 검색이나 기사 단위 검색은 제공되지 않으며, 이는 탐색 효율성과 연구 활용도를 제한하는 요소로 지적될 수 있다.

접근 방식은 도서관 메인 웹사이트 내 아카이브 메뉴를 통해 진입하도록 구성되어 있으며, 별도의 로그인 절차 없이 일반 이용자도 자유롭게 접근 가능한 공개형 서비스이다. 웹 브라우저 기반으로 최적화되어 있어 연구자·학생·일반 이용자 누구나 인터넷 접속만으로 이용 가능하며, 모바일 환경에서도 동일한 경로로 접근이 가능하나 열람 편의성은 PC 환경 중심으로 설계되어 있다.

④ 메타데이터 구조

고신문 아카이브는 발행 연도·월·일 단위의 서지정보를 기본 메타데이터로 제공한다. 목록 화면에는 신문명, 권호(양력·간지 표기 병행), 발간일자, 원본 소장처, 원문 열람 버튼이 체계적으로 병기되어 있어 이용자가 서지적 맥락 속에서 자료를 인지할 수 있다. 발행 연도 및 간지(예: 광무 2년)와 같은 역사적 연호 병

기는 한국 근대신문의 시기적 특수성을 반영하는 메타데이터로, 단순 연대기적 정렬뿐 아니라 당대 연구자에게 중요한 연호-서양력 대조 연구에도 활용 가능성을 제공한다.

메타데이터는 기관 소장처 정보를 포함하여 디지털화 자료와 원본 자료의 출처를 명확히 구분하는 구조를 갖추고 있다. 이는 연구자가 필요시 원본 열람이나 보존 자료 확인을 위해 소장기관과 직접 연계할 수 있도록 설계된 점에서 의의가 있다([그림 2] 참조). 그러나 메타데이터의 기술 수준은 기본 서지정보 중심으로 제한되어 있어, 기사 단위의 제목, 필자, 주제어 등 세부 기술 메타데이터는 부재하다. 이는 단행본·학술지 아카이브에서 활용되는 KORMARC, MODS, METS/ALTO 등과 같은 구조화된 기술 메타데이터와 비교할 때 활용 범위가 축소되는 한계로 지적할 수 있다.

⑤ 저작권 및 이용 제한

고신문 아카이브에서 제공되는 황성신문은 1898년~1910년 발행본으로, 저작권법상 보호기간(저자 사후 70년)을 이미 경과한 공공 영역(Public Domain)에 해당한다. 따라서 원문

열람과 다운로드는 저작권 제한 없이 가능하도록 설계되어 있다. 서비스 화면에서는 발행일자별 PDF 원문을 개별적으로 내려받을 수 있는 기능을 제공하고 있으나, 대량 다운로드나 API 제공 등 체계적 데이터 추출 기능은 지원되지 않는다. 이는 저작권 문제가 아닌 플랫폼 차원의 이용 통제로서, 서비스 안정성과 무단 상업적 활용 방지를 고려한 제한 조치로 해석된다.

별도의 이용약관이나 저작권 고지 문구가 상세하게 제시되지 않으며, '원본 소장처 명시'가 모든 자료에 병기되어 있다. 이는 디지털 이용자는 자유롭게 접근할 수 있으나, 학술 인용이나 2차 활용 시 반드시 출처 표기를 요구하는 간접적 저작권 관리 방식으로 볼 수 있다. 저작권 측면에서 고려대 아카이브는 공공저작물로서의 자유 이용을 보장하면서도, 데이터의 무단 복제·재배포에 대한 명확한 규정은 제공하지 않고 있다. 이는 국립중앙도서관의 공공누리 라이선스나 해외의 Creative Commons 유형 고지와 비교할 때, 저작권·이용 조건 고지의 표준화가 미흡하다는 한계가 존재한다.

〈표 5〉 고려대학교 중앙도서관 고신문 디지털 아카이브 주요 현황

구분	내용
운영 주체	고려대학교 중앙도서관
서비스 개시	구체적 개시 연도 미상
수록 범위	1898-1910년 『황성신문』
협력 구조	고려대 도서관, 국립고궁박물관, 연세대 학술정보원 소장본 통합
디지털화 수준	PDF 고해상도 스캔 이미지, OCR 미제공
접근 방식	무료 공개, 로그인 불필요, 웹 브라우저 기반
주요 기능	연도-월-일-면 단위 탐색, PDF 다운로드, 엑셀 메타데이터 다운로드
메타데이터	신문명, 권호(양력·간지 병기), 발간일자, 소장처, 기사 단위 부제
저작권 정책	공공영역(Public Domain), 개별 다운로드 허용, 대량 이용 제한

3.1.3 근대한국학 DB³⁾

연세대학교 근대한국학연구소는 인문한국플러스(HK+) 사업단과 공동으로 근대한국학 DB를 운영하고 있으며, 황성신문·대한매일신보·독립신문 등 근대신문과 잡지류를 통합 제공하는 연구 지향형 데이터베이스이다. 이 DB는 자체 소장 자료의 디지털화와 함께 외부 공공기관 DB와의 연계를 병행하는 협력형 수집 구조를 채택하고 있다는 점에서 앞서 살펴본 사례들과 차별화된 특징을 지닌다.

① 운영 주체 및 협력 구조

연세대학교 근대한국학연구소가 HK+ 사업단과 함께 DB를 운영·제공하며, DB 구축의 기본 전략으로 국사편찬위원회·국립중앙도서관 등 타 기관의 공개 DB 자료를 수집·활용하고, 신규 한국학 자료의 선별적 디지털화 및 미공개 자료 발굴·원문 디지털화를 병행하고 있다. 연구소는 연세대학교 중앙도서관 국학자료실과 연계하여 미공개 자료의 촬영·보정 등 원문 확보를 진행해왔으며, 이는 대학 내부(연구소-도서관) 간 역할 분담을 통한 소장·디지털화 협력 모델을 보여준다. 대기관 자료의 상호운용성과 공유를 위해 관계형 DB(RDBMS)와 더블린 코어(Dublin Core) 기반 저장 구조를 명시하고 있다는 점에서 표준 지향형 운영체제로 평가된다.

② 콘텐츠 형태 및 디지털화 수준

근대한국학 DB는 기존 공개자료뿐 아니라 미공개 자료까지 체계적으로 확보하여 다층적

디지털화 수준을 갖추고 있다. 1860년대부터 1890년대 말까지 발간된 단행본과 정기간행물의 원문·복사본을 확보하였으며, 국학자료실 소장 미공개 자료 176책을 직접 촬영·보정하여 디지털화 기반을 마련하였다. 특히 단순한 이미지 스캔에 그치지 않고 원문 텍스트 입력을 병행하여 기사 단위까지 검색 가능한 구조로 전환한 점이 주목된다. 시사총보, 한성신보, 친목회 회보 등의 자료는 1차 입력 완료 후 교정 단계를 거쳐 텍스트 데이터베이스로 구축되고 있다.

대규모 디지털화 프로젝트를 통해 한말·일제초기 발간된 일반간행물 31종(약 6,230면, 156만 자)과 잡지류 12종 107책(약 8,116면, 333만 자)을 원문 전체 디지털화 대상으로 삼아 전산화 작업을 진행하였다. 텍스트 데이터는 단순 OCR 처리에 그치지 않고 원문 입력과 교정을 통해 정확도를 높인 표준화된 형태로 구축되고 있으며, 향후 개체명 인식(NER)이나 텍스트 마이닝 기법을 적용할 수 있는 구조적 기반을 마련하고 있다.

③ 서비스 및 접근 방식

근대한국학 DB는 '전체검색'과 '상세검색'을 제공한다. 상세검색에는 신문/잡지명, 기사제목, 본문내용, 색인어 입력란이 제공되어 주제·기사 단위 탐색이 가능하도록 설계되어 있다([그림 3] 참조). 특정 간행물을 선택하면 목록에 번호, 신문/잡지명, 기사제목, 발행연월일, 원문소장기관이 제시되며, 기사 제목을 클릭하면 기사 형식, 호수, 색인어, 원문 소장기관 및

3) <https://cmks.yonsei.ac.kr/cmks/database.htm?ch=1>

식별자가 포함된 상세 화면이 열린다.

신문과 잡지는 제공 방식이 상이하다. 신문류는 결과 리스트와 상세 메타데이터 중심으로 제공되며, 원문 열람은 연계된 원문 소장기관(예: 국립중앙도서관 LOD)으로 안내하는 구조이다. 잡지류는 원문 전체 디지털화를 구축 방식으로 명시하여 원문 제공을 전제로 한 디지털화 정책을 채택하고 있다. 접근 방식은 연구소 공식 웹사이트 내 데이터베이스 메뉴를 통해 접근할 수 있으며, 별도의 회원가입이나 로그인 절차 없이 일반 이용자에게 공개되어 있다. 다만 대량 검색·데이터 다운로드보다는 개별 자료 검색과 열람 중심의 인터페이스로 설계되어 있어 연구 목적의 활용은 PC 환경에 최적화되어 있다.

④ 메타데이터 구조

근대한국학 DB는 더블린 코어(Dublin Core)를 기반으로 한 메타데이터 체계를 채택하여 모든 자료를 제목, 간행물명, 발행연월일, 원문 소장기관, 기사 형식 등의 공통 필드로 표준화하고 있다. 기사 단위 메타데이터에는 기사제목, 간행물명, 발행호수, 발행일자, 기사 형식

(논설, 보도, 광고 등), 색인어가 포함되며, 기사 상세 화면에서 색인어와 식별자(Identifier)가 명시되어 외부 소장처 연계 시에도 자료의 추적성을 보장한다.

관계형 데이터베이스(RDBMS)를 활용하여 동일 간행물 내 기사 간 연결, 간행물과 발행연도 간 관계를 구조화함으로써 검색 결과가 시계열적·주제적 맥락을 가진 데이터셋으로 제공할 수 있다. 현재 색인어는 주제 중심의 제한된 체계로 운영되고 있으나, 연구소는 자연어처리(NLP) 및 개체명 인식(NER) 기술을 적용할 수 있는 구조적 기반을 명시하고 있어 장기적으로 인명·지명·기관명 등 의미 기반 메타데이터 확장 가능성을 열어두고 있다.

⑤ 저작권 및 이용 제한

신문류(황성신문·대한매일신보·독립신문)는 DB 내에서 메타데이터 중심으로 제공되며, 원문은 소장기관 측 서비스로 연계되는 구조이다. 데이터베이스 검색 페이지는 별도 로그인 없이 접근·검색이 가능하며, 잡지류는 원문 전체 디지털화를 기조로 하여 원문 제공이 이루어진다. 다만 검색·결과 화면에 저작권·이용

<표 6> 연세대학교 근대한국학연구소 근대한국학 DB 주요 현황

구분	내용
운영 주체	연세대학교 근대한국학연구소(HK+ 사업단 공동 운영)
서비스 개시	HK+ 사업 기간 내 구축
수록 범위	1860년대~1910년대 신문·잡지류(황성신문·대한매일신보·독립신문 외 잡지 12종 등)
협력 구조	국사편찬위원회·국립중앙도서관 DB 연계, 연세대 중앙도서관 국학자료실 협력
디지털화 수준	신문: 기사 단위 메타데이터 + 소장기관 연계 / 잡지: 원문 전체 디지털화, 교정 텍스트 구축
접근 방식	무료 공개, 로그인 불필요, 웹 브라우저 기반
주요 기능	전체검색·상세검색, 기사 단위 탐색, 원문 소장기관 연계 안내
메타데이터	Dublin Core 기반, RDBMS 구조, 기사 단위(제목·형식·색인어·소장처 식별자)
저작권 정책	신문 원문 직접 제공 없음, 소장기관 연계, 별도 저작권 고지 미표기

조건에 관한 별도 고지 문구는 명시되어 있지 않으며, 이는 국립중앙도서관의 공공누리 라이선스 고지나 해외의 Creative Commons 유형 표기와 비교할 때 이용 조건 안내의 표준화가 미흡한 점은 한계라 할 수 있다.

조적 특성을 내포한다. 서비스 확장은 언론사와의 개별 협상에 의존하므로, 특정 언론사(예: 중앙일보)의 지면 아카이브는 지면보기가 유료로 제한되는 등 언론사별 계약 조건에 따라 제공 범위가 상이하다.

3.1.4 네이버 뉴스 라이브러리⁴⁾

네이버 뉴스 라이브러리(Naver News Library)는 국내 최대 포털 사이트 네이버(NAVER)가 2009년 4월 베타 서비스를 시작하여 2011년 4월 정식 서비스를 개시한 민간 신문 아카이브이다. 1920년부터 1999년까지의 주요 일간지 기사를 무료로 제공하며, 공공기관이 아닌 민간 기업이 운영하는 대규모 신문 아카이브라는 점에서 앞서 살펴본 사례들과 본질적으로 다른 성격을 지닌다.

① 운영 주체 및 협력 구조

네이버 뉴스 라이브러리는 주식회사 네이버가 단독 운영하며, 언론사와의 개별 계약을 통해 아카이브 서비스 권한을 확보하는 구조를 취하고 있다. 현재 서비스 중인 주요 언론사는 경향신문(1946년 창간), 동아일보(1920년 창간), 매일경제(1966년 창간), 조선일보(1920년 창간), 한겨레(1988년 창간)이며, 각 언론사의 창간호부터 1999년 12월 31일까지의 지면을 제공한다. 운영 재원은 네이버의 자체 사업 수익에서 조달되며, 별도의 정부 예산이나 공공지원 없이 민간 자본으로 구축·유지되고 있다. 이는 공공 아카이브와 달리 서비스 지속성이 운영 기업의 사업 판단에 종속될 수 있다는 구

② 콘텐츠 형태 및 디지털화 수준

네이버 뉴스 라이브러리는 1920년 3월 5일(동아일보·조선일보 창간일)부터 1999년 12월 31일까지의 기사를 제공하며, 수록 기간 80년에 걸쳐 국내 주요 5개 일간지의 전체 지면을 망라한다. 콘텐츠 형태는 신문 지면 스캔 이미지와 OCR 텍스트를 병행 제공하는 방식으로 구성되어 있다. 언론사별 아카이브 이미지 데이터를 네이버 자체 OCR 인식 솔루션(CLOVA OCR의 전신)으로 처리하여, 지면 섹션을 클릭하면 해당 기사의 OCR 텍스트가 즉시 추출되어 표시된다.

디지털화 수준은 이미지+OCR 텍스트 병행이라는 점에서 이미지 단독 제공인 고려대 고신문 아카이브보다 높으며, OCR 기반 키워드 검색이 가능하다는 점에서 연구 활용도가 높다. 다만 1920년대~1950년대 초기 자료의 경우 인쇄 품질 저하와 한자·구한글 혼용으로 인한 OCR 오인식 문제가 일부 확인되며, 1933년 한글 맞춤법 통일안 이전 자료에는 옛한글 표기가 포함되어 검색 정확도에 제약이 따른다. 또한 언론사 요청에 의해 일부 기사 내용이 삭제(***로 표기)된 경우가 있으며, 이는 서비스의 역사적 완전성에 한계를 초래한다.

4) <https://newslibrary.naver.com/search/searchByDate.naver>

③ 서비스 및 접근 방식

서비스는 낱자 검색과 키워드 검색의 두 가지 방식으로 제공된다. 낱자 검색은 특정 낱자를 선택하면 해당일 신문 지면 전체를 열람할 수 있으며, 키워드 검색은 특정 단어를 입력하면 해당 단어가 수록된 기사 목록이 제공되고 연도별 출현 빈도를 그래프로 시각화하여 추이를 확인할 수 있다. 이는 앞선 사례들과 달리 시계열 빈도 분석 기능을 기본으로 제공한다는 점에서 연구자에게 유용한 분석 도구로 기능한다.

접근 방식은 비로그인 상태에서도 낱자 검색과 기본 키워드 검색이 가능한 개방형 서비스를 기본으로 하며, 로그인 시 언론사·지면·낱자 범위 등의 상세 조건을 설정하는 고급 검색이 가능하다. 서비스는 PC 웹 브라우저 중심으로 설계되어 있으며, 초기 플래시(Flash) 기반에서 HTML5 기반으로 전환하였으나 모바일 전용 UI는 제공되지 않아 스마트폰 환경에서의 이용 편의성은 낮다. 저작권 보호를 위해 마우스 우클릭이 금지되어 있으며, API나 대량 데이터 다운로드 기능은 제공하지 않는다.

④ 메타데이터 구조

네이버 뉴스 라이브러리의 메타데이터는 신문명, 발행일자, 페이지(면) 번호, 기사 식별자(articleId), 판 번호(editNo), 발행 유형(publishType) 등 서비스 내부의 URL 구조를 통해 확인되는 기술적 식별자 중심으로 구성되어 있다. 기사 단위의 메타데이터는 OCR 처리 결과로 추출된 텍스트와 지면 이미지가 연동되는 구조이나, 기자명·주제 분류·색인어 등의 서지적 메타데이터는 별도로 제공되지 않는다. 이는 국립중앙도서관 신문 아카이브의 MODS 기반 메타데이터나 연세대 근대한국학 DB의 더블린 코어 기반 구조와 비교할 때 서지적 완성도 면에서 상대적으로 제한적이다.

키워드 검색 시 연도별 빈도 그래프가 제공되는 것은 암묵적인 시계열 색인 구조가 내부적으로 존재함을 시사하나, 이용자에게 노출되는 메타데이터 필드는 신문명과 발행일자에 집중되어 있다. 국제 표준 메타데이터 스키마(MODS, METS, 더블린 코어 등)의 명시적 적용이나 LOD 연계는 확인되지 않으며, 상호운용성보다는 자체 서비스 최적화에 집중된 구조로 평가된다.

<표 7> 네이버 뉴스 라이브러리 주요 현황

구분	내용
운영 주체	주식회사 네이버(민간 기업)
서비스 개시	2009년 4월 베타 / 2011년 4월 정식 서비스
수록 범위	1920년~1999년, 주요 5개 일간지(경향·동아·매일경제·조선·한겨레)
협력 구조	언론사별 개별 계약 기반(서비스 범위 언론사마다 상이)
디지털화 수준	지면 스캔 이미지 + 네이버 OCR 텍스트 병행, 기사 단위 텍스트 추출
접근 방식	비로그인 기본 열람 가능, 로그인 시 상세 조건 검색, PC 중심
주요 기능	낱자 검색, 키워드 검색, 연도별 빈도 그래프 시각화
메타데이터	신문명·발행일자·기사 식별자 중심, 서지 메타데이터·표준 스키마 미적용
저작권 정책	언론사 저작권 귀속, 우클릭 금지·다운로드 제한, 언론사 요청 시 부분 삭제 가능

⑤ 저작권 및 이용 제한

네이버 뉴스 라이브러리에서 제공되는 기사의 저작권은 각 언론사에 귀속되며, 네이버는 언론사와의 계약을 통해 디지털 아카이브 서비스 권한을 확보한 형태이다. 이용자는 기사를 개인적·비상업적 목적으로 열람할 수 있으나, 무단 복제·배포·상업적 이용은 금지된다. 저작권 보호 조치로 마우스 우클릭 금지, 이미지 저장 제한 등의 기술적 통제가 적용되어 있다. API나 대량 데이터 다운로드 기능이 제공되지 않는 것 역시 저작권 관리 차원의 이용 통제로 해석된다.

언론사 요청에 의해 특정 기사의 내용이 부분 삭제(***로 마스킹)되는 경우가 있어, 서비스 내 기사의 완전성이 언론사의 사후 요청에 의해 변경될 수 있다는 점은 아카이브로서의 신뢰성과 역사 기록의 무결성 측면에서 주목할 만한 한계로 지적된다. 또한 민간 기업이 운영하는 서비스이므로, 서비스 종료나 수록 범위 변경 등이 이용자 예고 없이 이루어질 가능성도 공공 아카이브와 구별되는 취약점이다.

3.2 국외 사례 분석

3.2.1 Chronicling America⁵⁾

Chronicling America는 미국 의회도서관(Library of Congress, LC)이 운영하고 국립인문기금(National Endowment for the Humanities, NEH)이 후원하는 미국 역사 신문 디지털 아카이브로, 2005년 National Digital Newspaper Program(NDNP)을 통해 구축을 시작하여 2007

년 일반에 공개되었다. 공공기관 주도의 완전 개방형 아카이브로서, 표준화된 메타데이터 체계와 연구자 친화적 API를 갖춘 국제적 모범 사례로 평가된다.

① 운영 주체 및 협력 구조

LC가 플랫폼 운영과 인프라 관리를 담당하고, NEH 기금을 바탕으로 각 주(state)의 주립 역사학회·주립 도서관·공립 대학 도서관 등 지역 문화유산 기관이 현지 신문을 선별·디지털화하여 LC에 제공하는 분산형 파트너십 구조를 채택하고 있다. 이 모델은 연방 차원의 자금 지원과 지역 기관의 직접 참여를 결합하여 광범위한 지역 커버리지를 확보하며, 단일 기관으로는 어려운 전국적 신문 아카이브 구축을 실현하였다.

② 콘텐츠 형태 및 디지털화 수준

수록 범위는 1770년대부터 1963년까지의 미국 역사 신문이며, 총 수백만 페이지에 달한다. 페이지 이미지는 JPEG, PDF 형식으로 제공되며, 고해상도 TIFF 파일은 별도 요청을 통해 열람 가능하다. OCR 텍스트는 단어 위치 좌표(bounding-box)를 포함하는 ALTO XML 형식으로 제공되어 전문 검색(full-text search)과 검색어 하이라이팅이 가능하다. 특히 LC는 Newspaper Navigator 프로젝트를 통해 1,600만 페이지 이상의 신문에서 사진·광고·만화·지도 등 시각 자료를 머신러닝 기반으로 자동 추출·분류하여 텍스트 중심 아카이브를 시각 정보 탐색까지 확장하였다.

5) <https://www.loc.gov/collections/chronicling-america/about-this-collection/>

③ 서비스 및 접근 방식

간편 검색과 고급 검색을 병행 제공하며, 고급 검색에서는 Titles · Issues · Pages(Full Text) 단위 선택, 주 · 카운티 · 도시 · 민족성(Ethnicity) · 언어(Language) · 날짜 범위 등 다양한 패시(facet) 필터를 지원한다([그림 4] 참조). 또한 “These words within 5 words of each other”와 같은 근접 검색기능은 역사적 표현의 다양성을 고려한 정밀 탐색을 가능하게 한다. 로그인 없이 누구나 PDF · JPEG 다운로드가 가능하며, 연구자 · 개발자는 API와 bulk 데이터 다운로드를 통해 대규모 데이터 분석, 텍스트 마이닝, 토픽 모델링 등 디지털 인문학 연구에 활용할 수 있다. 웹 접근성 지침(Section 508)을 준수하여 시각장애인의 스크린 리더 접근을 보장하며, 이민자 커뮤니티가 발행한 독일어 · 프랑스어 · 이탈리아어 등 다언어 신문을 별도 컬렉션으로 제공한다.

④ 메타데이터 구조

METS/ALTO 국제 표준을 기반으로 배치(batch)-릴(reel)-이슈(issue)-페이지(page)의 계층적 구조를 기술한다. METS XML은

신문 제목(LCCN), 발행일, 발행 지역, 언어, 판형 등 서지 데이터를 계층적으로 연결하며, ALTO XML은 OCR 텍스트를 문단 · 줄 · 단어 단위로 세분화하고 좌표값을 포함하여 원문 이미지와의 정밀 대조를 가능하게 한다.

LCCN(Library of Congress Control Number) · OCLC(Online Computer Library Center) 번호 등 영구 식별자(Persistent Identifier)를 활용하여 개별 발행호와 페이지를 고유하게 식별하며, 이 메타데이터는 API의 JSON 출력에도 반영되어 외부 데이터 셋과의 상호운용성을 확보하고 있다.

⑤ 저작권 및 이용 제한

수록 자료는 주로 1924년 이전 발행 신문으로, 저작권이 만료된 퍼블릭 도메인(Public Domain) 자료이다. 이용자는 별도 허가 없이 원문 이미지와 OCR 텍스트를 자유롭게 열람 · 다운로드 · 재활용할 수 있으며, 학술 인용 · 데이터 분석 · 교육 자료 제작 등에 제한 없이 활용 가능하다. 1925년 이후 자료는 저작권 보호를 이유로 컬렉션에서 제외되어 있으며, 사이트 내 Rights & Access 페이지에서 개별 신문 타이

<표 8> Chronicling America 주요 현황

구분	내용
운영 주체	미국 의회도서관(LC), NEH 기금 지원
서비스 개시	2007년(NDNP 2005년 시작)
수록 범위	1770년대~1963년, 미국 역사 신문 수백만 페이지
협력 구조	NEH 기금 기반 주 · 지역 문화유산 기관 분산형 파트너십
디지털화 수준	JPEG · PDF · TIFF 이미지 + ALTO XML OCR, ML 기반 시각 자료 추출(Newspaper Navigator)
접근 방식	완전 무료 공개(Open Access), 로그인 불필요, Section 508 접근성 준수
주요 기능	간편 · 고급 검색, 패시 필터, 근접 검색, API · bulk 다운로드, 다언어 컬렉션
메타데이터	METS/ALTO 표준, LCCN · OCLC 식별자, JSON 출력, 계층적 구조
저작권 정책	1924년 이전 퍼블릭 도메인, 자유 이용 허용, Rights & Access 페이지 개별 고지

틀마다 저작권 상태와 재사용 조건을 명시하여 이용자의 법적 불확실성을 최소화하고 있다.

3.2.2 Newspapers.com⁶⁾

Newspapers.com은 세계 최대 계보(genealogy) 서비스 기업인 Ancestry.com LLC(이하 Ancestry Inc.)가 운영하는 상업적 디지털 신문 아카이브이다. 2012년 Ancestry 플랫폼에서 분사하여 독립 서비스로 운영되고 있으며, 1700년대부터 2000년대까지 10억 페이지 이상의 신문을 유료 구독 기반으로 제공한다. 앞서 살펴본 Chronicling America가 공공기관 주도의 완전 개방형 모델이라면, Newspapers.com은 동일한 신문 아카이브 영역에서 상업적 구독 모델이 어떻게 구현되는지를 보여주는 대비적 사례이다.

① 운영 주체 및 협력 구조

Ancestry Inc.가 단독으로 운영하며, 2020년 사모펀드 블랙스톤 그룹(Blackstone Group)에 인수되어 현재에 이르고 있다. 콘텐츠 수급은 신문사·출판사·도서관 등과 제휴를 맺고 마이크로필름 또는 원본을 무상으로 제공받아 디지털화하는 방식으로 이루어지며, 제휴 기관은 디지털화 비용 없이 아카이브를 구축하고 Ancestry는 그 콘텐츠에 대한 서비스 접근권을 확보하는 상호이익 구조를 취한다. 기관 이용자를 위해서는 ProQuest와의 협력을 통해 Library Edition을 제공하여, 대학도서관·공공도서관이 IP 인증 방식으로 접근할 수 있는 별도 채널을 운영하고 있다.

② 콘텐츠 형태 및 디지털화 수준

1700년대부터 2000년대까지 미국 내 지역지·전국지와 해외 일부 신문을 포함하여 10억 페이지 이상을 제공한다. 모든 콘텐츠는 고해상도 이미지(PDF·JPEG)로 디지털화되며, OCR을 통해 기사(article) 단위로 자동 분리된 텍스트와 이미지를 병기 제공한다. 특히 인물 이름·날짜·키워드 중심으로 색인이 구축되어 가계·계보 연구에 최적화된 인텍싱이 강점이다. 다만 디지털화 전략은 공개·학술적 오픈 액세스보다 구독 서비스 모델에 종속되어 있어, 무료 이용자는 페이지 미리보기 수준만 가능하고 고화질 확대·전문 검색·PDF 다운로드·클리핑 기능은 유료 구독자에게만 제공된다.

③ 서비스 및 접근 방식

키워드·인물명 검색과 발행일자·주(State)·도시·신문명 필터를 결합한 기본 검색 외에, 주(State)→도시→신문명→연도→월→날짜→페이지 순으로 드릴다운하는 Browse 기능과 발행지역별 신문 분포를 지도 위에서 탐색하는 Map 기능, 연대표(Timeline) 슬라이더를 통한 시기 필터링을 제공한다. 기사 뷰어(Article Viewer)에서는 검색어 노란색 하이라이트, 이미지 확대·축소·색 반전·회전, 하단 필름스트립 네비게이션이 지원되며, 상단의 CLIP·PRINT/SAVE·SHARE 버튼을 통해 기사 단위 클리핑·저장·소셜 공유 및 Ancestry 계정 연계가 가능하다. 단, 클리핑·저장 기능은 구독 회원에게만 활성화된다([그림 5] 참조).

6) <https://www.newspapers.com/>

접근 방식은 개인 유료 구독(Standard/Publisher Extra) 중심이나, ProQuest Library Edition 을 통한 기관 구독 채널을 병행하여 개인·기관 이용자를 분리 서비스하는 다층적 구조를 갖추고 있다.

④ 메타데이터 구조

Newspapers.com은 METS/ALTO, MARC 등 국제 표준 메타데이터를 적용하지 않으며, Ancestry Inc. 내부에서 구축한 상업적 색인 체계를 사용한다. 인명(Name Index)·지명·발행일자·신문명을 중심으로 색인이 구성되어 있어 가계 연구 목적의 개인 단위 추적에 특화되어 있다. API나 대규모 데이터 다운로드 는 제공되지 않으며, 데이터는 구독 기반 웹 인터페이스를 통해서만 접근 가능하다. 구독자가 기사를 클릭할 경우 제목·날짜·신문명·페이지 등 최소한의 메타데이터가 자동 부착되지만, 이는 학술적 상호운용성보다 개인 사용자 중심의 맞춤형 기록 관리에 초점을 맞춘 구조이다.

⑤ 저작권 및 이용 제한

저작권이 만료된 고신문도 서비스 접근 자체는 유료 구독으로 제한되며, Publisher Extra 요금제에 포함된 20세기 이후 신문은 저작권 보호 상태에서 다운로드·인쇄·재사용 범위가 추가로 제한된다. 기관 구독(ProQuest Library Edition) 환경에서도 도서관 계약 범위 내 열람만 허용되며 대규모 다운로드와 외부 재배포는 금지된다. 이처럼 Newspapers.com은 퍼블릭 도메인 자료조차 상업적 서비스 모델 안에서 접근을 통제하는 구조로, Chronicling America 의 완전 개방형 모델과 대비되는 저작권 운용 방식을 보여준다.

3.2.3 British Newspaper Archive⁷⁾

British Newspaper Archive(이하 BNA)는 DC Thomson 계열사인 Findmypast Newspaper Archive Limited가 단독 소유·운영하는 상업적 디지털 신문 아카이브이다. 2010년부터 British Library와 파트너십을 맺고 영국 국립 도서관이 소장한 방대한 신문 컬렉션을 디지털

<표 9> Newspapers.com 주요 현황

구분	내용
운영 주체	Ancestry Inc.(2020년 블랙스톤 그룹 인수, 민간 상업 기업)
서비스 개시	2012년(Ancestry 플랫폼에서 분사)
수록 범위	1700년대~2000년대, 10억 페이지 이상(미국 지역지·전국지, 해외 일부 포함)
협력 구조	신문사·도서관 제휴 콘텐츠 수급, ProQuest Library Edition 기관 배포
디지털화 수준	고해상도 이미지(PDF·JPEG) + OCR 기사 단위 자동 분리, 인명·날짜 중심 인덱싱
접근 방식	유료 구독 중심(Standard/Publisher Extra), ProQuest 기관 구독 병행
주요 기능	키워드·인물명 검색, Browse·Map·Timeline 탐색, 기사 클리핑(Clip)·공유·저장
메타데이터	상업적 내부 색인 체계, 국제 표준 미적용, API·대규모 다운로드 미지원
저작권 정책	퍼블릭 도메인 포함 전 자료 구독 접근 제한, Publisher Extra는 저작권 보호 자료 추가 포함

7) <https://www.britishnewspaperarchive.co.uk/>

화하여 웹 플랫폼으로 제공하고 있다. 앞서 살펴본 Newspapers.com이 순수 민간 기업 단독 운영 모델이라면, BNA는 공공기관(국립도서관)이 원본 자료를 제공하고 민간기업이 디지털화와 서비스를 담당하는 공공-민간 파트너십 모델의 대표 사례이다.

① 운영 주체 및 협력 구조

British Library는 열람실에 스캐닝 팀을 상주시켜 보관 중인 원본 신문을 제공하고, 디지털 처리 및 품질 관리는 Findmypast의 Dundee 오퍼레이션 센터가 담당하는 분업 체계를 갖추고 있다. 초기에는 약 10년 단위의 디지털화 계획으로 시작하였으나, 2025년 11월 기준 9천만 페이지 이상이 디지털화되었으며 매주 수십만 건씩 추가 업데이트가 이루어지고 있다. 이 구조는 공공기관의 소장 자료와 민간 기업의 기술·플랫폼 역량을 결합하여 어느 한쪽만으로는 실현하기 어려운 대규모 디지털화를 가능하게 한 협력 모델이다.

② 콘텐츠 형태 및 디지털화 수준

1700년대부터 2000년대 초까지 영국 및 아일랜드 신문 약 9천만 페이지를 고해상도 이미지(PDF·JPEG)와 OCR 텍스트 병기 형태로 제공한다. 검색어는 페이지 내에서 하이라이트로 표시되어 원문 이미지와의 교차 검증이 가능하다. BNA의 주목할 특징은 신문 전체를 일괄 스캔하는 방식이 아니라, 발행 지역·연도·신문명·주제별 컬렉션 단위로 디지털화 범위를 전략적으로 선별한다는 점이다. 특정 신문사는 전 기간을 커버하는 반면 일부 신문은 특정 시기·지역에 한정하는 등 선택과 집중을 통해

디지털화 자원을 효율적으로 배분하고 있다.

③ 서비스 및 접근 방식

이용자는 신문명→발행 연도→월·일→페이지 순의 단계적 Browse 탐색, 영국·아일랜드 전역의 신문 발행 분포를 시각화한 지도(Map) 탐색, 상단 타임라인 슬라이더를 통한 시기 필터링을 복합적으로 활용할 수 있다. 고급 검색(Advanced search)에서는 AND·OR·NOT·정확한 구문 검색 조합과 함께 발행 지역(Place)·신문명(Title)·발행일·데이터 추가일·접근 유형(Free to View/Subscriber Access)·기사 유형(광고·기사·가족 공지·삽화·뉴스·표 등)·Public Tags를 복합 지정할 수 있어 다층적 탐색이 가능하다. 기사 뷰어에서는 검색어 하이라이트, 이미지 확대·축소·회전·전체화면, 저장·인쇄·북마크 기능을 지원한다.

접근 방식은 개인 유료 구독(월별·연간) 기반이나, 'Free to View'로 지정된 일부 콘텐츠는 비회원에게도 공개된다. British Library 열람실에서는 모든 BNA 콘텐츠를 무료로 열람할 수 있으며, 영국 내 일부 공공도서관·대학도서관도 Findmypast와의 라이선스 계약을 통해 방문 이용자에게 무료 접근 권한을 제공한다.

④ 메타데이터 구조

BNA의 메타데이터는 British Library가 보유한 전통 서지정보(발행일·발행지·제목 등)를 기반으로 하되, Findmypast가 상업적 플랫폼에 맞게 가공한 혼합 구조이다. 특히 Advertisement·Article·Family Notices·

Illustrated · News · Table · Frontpage only 등 기사 유형(Article Type)별 메타데이터 필드를 제공하여 내용 · 형식 기반 탐색을 가능하게 한다는 점이 다른 상업 서비스와의 차별점이다. 고급 검색의 모든 필터 항목이 메타데이터와 직접 연동되어 이용자가 시공간적 범위와 기사 성격을 복합적으로 지정할 수 있도록 설계되어 있다. 다만 Chronicling America와 달리 API나 대규모 데이터 다운로드는 제공되지 않아 상호운용성보다는 상업적 웹 인터페이스 중심의 구조를 유지하고 있다.

⑤ 저작권 및 이용 제한

저작권이 만료된 자료는 뷰어 상단에 'Public domain' 아이콘을 명시적으로 표시하고, 저작권이 유효한 자료는 'Subscriber Access'로 구분 표시하여 이용자가 각 자료의 법적 상태를 직관적으로 인식할 수 있도록 한다. 저작권이 유효한 자료는 출판사 · 권리자와의 계약 범위 내에서만 서비스되며, 다운로드 · 재배포 등 2차 활용은 허용되지 않는다. 이처럼 BNA는 Newspapers.com과 동일한 유료 구독 기반이

면서도 퍼블릭 도메인 라벨링을 통해 이용자의 권리 인식을 명확히 지원한다는 점에서 저작권 고지 체계 측면의 모범 사례로 평가된다.

4. 분석 종합 및 특성 도출

4.1 운영 주체별 서비스 구조 비교

7개 사례의 운영 주체는 크게 네 가지 유형으로 분류된다. 첫째, 공공기관 단독 운영형(대한민국 신문 아카이브, Chronicling America)은 국가 예산에 기반하여 안정적 서비스 지속성을 확보하나, 확장 속도가 예산 주기에 종속되는 한계를 지닌다. 둘째, 공공-민간 파트너십형(BNA)은 공공기관의 소장 자료와 민간의 기술 · 자본을 결합하여 대규모 디지털화를 신속하게 실현하였으나, 민간 기업의 사업 판단에 따라 서비스 구조가 변동될 수 있다. 셋째, 민간 기업 단독형(네이버 뉴스 라이브러리, Newspapers.com)은 수익 모델에 종속된 서비스 지속성과 콘텐츠 완전성의 취약성을 내포한

〈표 10〉 British Newspaper Archive 주요 현황

구분	내용
운영 주체	Findmypast Newspaper Archive Limited(DC Thomson 계열, 민간 상업 기업)
서비스 개시	British Library(원본 제공), Brightsolid(스캐닝 참여)
수록 범위	1700년대~2000년대 초, 약 9천만 페이지 이상(영국 · 아일랜드)
협력 구조	2010년(British Library-Findmypast 파트너십)
디지털화 수준	고해상도 이미지(PDF · JPEG) + OCR 텍스트, 기사 유형별 전략적 선별 디지털화
접근 방식	유료 구독(월별 · 연간) + Free to View 혼합, 국립 · 공공 · 대학도서관 무료 열람
주요 기능	Browse · Map · 타임라인 탐색, 고급 검색(AND/OR/NOT · 기사 유형 · 접근 유형 필터), 저장 · 북마크
메타데이터	BL 서지정보 기반 + 상업적 인텍싱, 기사 유형별 필드, API 미지원
저작권 정책	"Public domain" 아이콘 명시, "Subscriber Access / Free to View" 구분 표시

다. 넷째, 대학 연구소 기반 사례(고려대, 연세대)는 학술 연구 목적에 특화된 반면, 사업 종료 시 서비스 유지에 위험이 따른다는 구조적 공통점을 지닌다.

4.2 콘텐츠 형태 및 디지털화 수준 비교

디지털화 수준은 이미지 단독 제공(고려대), 이미지+OCR 병행(대부분), 교정 텍스트 구축(연세대), 구조화 XML 텍스트(Chronicling America)의 단계적 스펙트럼으로 분포한다. 국내 사례는 근대 자료의 인쇄 품질 저하와 구한글·한자 혼용으로 인한 OCR 오인식 문제가 공통 과제로 나타나며, 기사 단위 분리 수준은 서비스 목적에 따라 상이하다. 국외 사례는 모두 이미지+OCR 텍스트 병기를 기본으로 제공하며, Chronicling America의 ALTO XML 기반 좌표 연동과 ML 기반 시각 자료 추출(Newspaper Navigator), BNA의 기사 유형별 분류는 단순 스캔을 넘어선 고도화된 디지털화의 방향을 제시한다.

4.3 서비스 방식과 접근 방식의 차이

서비스 방식의 가장 두드러진 차이는 접근 조건과 연구 지원 수준에서 나타난다. 국내 공공 아카이브(대한민국 신문 아카이브, 고려대, 연세대)는 무료·비로그인 공개 접근을 원칙으로 하나 API·데이터 재사용 지원이 미흡하며, 국외에서는 Chronicling America가 완전 무료 공개와 API·bulk 다운로드를 동시에 제공하는 유일한 사례로, 연구 인프라로서의 완성도가 가장 높다. 상업 서비스(Newspapers.com, BNA)는 이용자 경험(클리핑, 지도, 타임라인 등) 측면에서 국내 사례보다 앞서 있으나, API 개방은 없어 학술 데이터 분석에는 구조적 한계를 지닌다.

4.4 메타데이터 구조 및 저작권 정책 비교

메타데이터 구조에서 가장 뚜렷한 대비는 표준 기반 여부와 외부 개방성이다. Chronicling America는 METS/ALTO 국제 표준과 LCCN·OCLC 영구 식별자, JSON API 출력을 갖춘 유일한

〈표 11〉 국내외 근대신문 콘텐츠 및 서비스 종합 비교

사례	운영 유형	수록 시기	OCR	접근 조건	메타데이터 표준	API	저작권 고지
대한민국 신문 아카이브	공공 (국립중앙도서관)	1883-1960년대	△ (일부)	무료·비로그인	MODS, LOD	×	출처 표기
고신문 디지털 아카이브	대학도서관 (고려대학교)	1898-1910	×	무료·비로그인	비표준	×	소장처 명시
근대한국학 DB	대학연구소 (연세대+HK+)	1860-1910년대	○ (교정 텍스트)	무료·비로그인	Dublin Core	×	별도 고지 없음
네이버 뉴스라이브리	민간 (네이버)	1920-1999	△ (오인식 있음)	비로그인(기본)/ 로그인(고급)	비표준	×	기술적 통제만
Chronicling America	공공 (LC+NEH)	1770년대- 1963년대	○ (ALTO XML)	무료·비로그인	METS/ALTO, LCCN	○ API·bulk	자료별 개별 고지
Newspapers.com	민간 (Ancestry Inc.)	1700년대- 2000년대	○ (기사 단위)	유료 구독	비표준 (상업 색인)	×	미고지
British Newspaper Archive	공공-민간 (BL+Findmypast)	1700년대- 2000년대 초	○ (유형 분류)	유료+ Free to View	BL 서지+ 상업 색인	×	Public domain 아이콘

완전 표준형 사례이며, 대한민국 신문 아카이브와 연세대 DB는 MODS·Dublin Core 기반으로 일정 수준의 표준화를 달성하였다. 반면 상업 서비스(Newspapers.com, BNA)와 민간 기업 운영 사례(네이버)는 내부 색인 중심의 비표준 구조로 상호운용성이 낮다.

저작권 정책 측면에서는 고지 체계의 명확성에서 큰 차이가 나타난다. Chronicling America는 Rights & Access 페이지를 통해 자료별 저작권 상태를 개별 고지하며, BNA는 'Public domain' 아이콘과 'Subscriber Access' 구분 표시를 통해 직관적 라벨링을 제공한다. 반면 국내 사례 대부분은 이용 조건에 대한 표준화된 고지 체계가 미흡하며, 특히 네이버 뉴스 라이브러리는 언론사 요청에 의한 사후 기사 삭제가 가능하다는 점에서 아카이브 무결성에 구조적 취약성을 내포하고 있다.

5. 결론 및 제언

본 연구는 국내외 근대신문 디지털 서비스 7개 사례를 대상으로 운영 주체 및 협력 구조, 콘텐츠 형태 및 디지털화 수준, 서비스 및 접근 방식, 메타데이터 구조, 저작권 정책의 다섯 가지 항목을 분석 틀로 적용하여 구조적 특성을 비교·분석하였다. 이를 통해 근대신문 디지털 서비스의 유형별 특징을 규명하고, 공공성과 접근성을 균형 있게 확보하기 위한 논의의 기초를 마련하는 것을 목적으로 하였다.

분석 결과, 7개 사례는 공공기관 단독 운영형, 공공-민간 파트너십형, 민간 기업 단독형, 대학 연구소형의 네 가지 운영 유형으로 구분되며,

유형에 따라 서비스의 지속성, 콘텐츠 규모, 개방성 수준이 뚜렷하게 달라짐이 확인되었다. 디지털화 수준은 이미지 단독 제공에서 구조화 XML 텍스트에 이르는 단계적 스펙트럼을 형성하고 있으며, 국내 사례에서는 구한글·한자 혼용으로 인한 OCR 오인식이 공통 과제로 나타났다. 메타데이터 구조에서는 공공기관 운영 사례일수록 국제 표준 준수 경향이 강한 반면, 상업 서비스는 이용자 경험 중심의 내부 색인 체계에 집중되어 상호운용성이 낮은 특징을 보였다. 접근성 측면에서는 국내 공공 아카이브가 무료·비로그인 원칙을 충실히 구현하고 있으나, 연구자 대상 API 및 데이터 재활용 지원은 전반적으로 미흡한 것으로 나타났다.

이상의 분석 결과는 근대신문 디지털 서비스의 발전을 위해 다음과 같은 시사점을 제공한다. 첫째, 운영 유형에 따른 구조적 강점과 한계를 고려하여, 공공성과 지속 가능성을 동시에 확보할 수 있는 협력 모델의 설계가 필요하다. 둘째, 디지털화 품질 향상과 함께 기사 단위 메타데이터 표준화를 통해 학술 연구와 대중적 활용을 아우르는 데이터 기반을 강화해야 한다. 셋째, 공개 접근 원칙을 유지하면서도 API 등 개방형 데이터 채널을 마련하여 디지털 인문학 연구 환경을 지원할 필요가 있다. 넷째, Chronicling America와 British Newspaper Archive의 저작권 라벨링 사례를 참고하여, 이용자가 각 자료의 법적 상태와 활용 가능 범위를 명확히 인식할 수 있는 표준화된 저작권 고지 체계를 마련해야 한다. 본 연구는 근대신문 디지털 서비스의 구조적 특성을 유형별로 비교·분석한 기초 연구로서, 향후 근대신문 아카이브의 구축 및 운영 방향을 논의하는 데 참고 자료로 활용될 수 있을 것이다.

참 고 문 헌

- 고려대학교 중앙도서관 (발행년불명). 고신문 디지털 아카이브.
출처: <https://library.korea.ac.kr/datause/archiving/old-newspaper/old-newspaper/>
- 곽승진, 노영희, 정대근 (2018). 이용자 인식조사를 기반으로 한 근대문학자료 서비스 활성화 방안 연구, 49(1), 47-75. <https://doi.org/10.16981/kliss.49.1.201803.47>
- 국립중앙도서관 (2024). 한국 근대문헌 온라인 통합서비스 개발 및 운영 방안 연구(11-1371029-100004-01).
- 국립중앙도서관 (발행년불명). 대한민국 신문 아카이브.
출처: <https://www.nl.go.kr/newspaper/>
- 네이버 (발행년불명). 네이버 뉴스 라이브러리.
출처: <https://newslibrary.naver.com/search/searchByDate.naver>
- 송주형 (2016). 온라인 신문 아카이브 연구: 국내외 구축 사례를 중심으로. 기록학연구, 48, 93-139.
- 연세대학교 근대한국학연구소 (발행년불명). 근대한국학 DB.
출처: <https://cmks.yonsei.ac.kr/cmks/database.htm?ch=1>
- 유춘동, 오영식 (2021). 시대 변화에 따른 근대문헌 범주의 재설정 문제와 방안. 서지학연구, 87, 51-65. <https://doi.org/10.17258/jib.2021..87.51>
- 윤정원 (1997). 한국 근대 정기간행물에 관한 서지학적 연구: 1889-1945. 서지학연구, 14, 355-361.
- 이고은, 최예은, 이유진, 김경민, 이지연 (2025). 근대문헌 통합 온라인 서비스 제공을 위한 이용자 요구 및 인식에 관한 연구. 정보관리학회지, 42(2), 177-203.
<https://doi.org/10.3743/KOSIM.2025.42.2.177>
- 홍종욱, 장문석, 류준필 (2025). 한국 근대문헌 디지털 아카이브의 설계: 『진단학보』를 중심으로. 인문연구, 110, 79-122. <https://doi.org/10.21211/JHUM.110.3>
- Cook, T. (2013). Evidence, memory, identity, and community: Four shifting archival paradigms. *Archival Science*, 13(2-3), 95-120. <https://doi.org/10.1007/s10502-012-9180-7>
- Huvila, I. (2008). Participatory archive: Towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Archival Science*, 8(1), 15-36. <https://doi.org/10.1007/s10502-008-9071-0>
- Lee, B., Mears, J., Jakeway, E., Ferriter, M., Adams, C., Yarasavage, N., Thomas, D., Zwaard, K., & Weld, D. (2020). The newspaper navigator dataset: Extracting headlines and visual content from 16 million historic newspaper pages in *Chronicling America*. Proceedings of the 29th ACM International Conference on Information & Knowledge Management,

3055-3062. <https://doi.org/10.1145/3340531.3412767>

Library of Congress (n.d.). *Chronicling America: About this collection*. Available:

<https://www.loc.gov/collections/chronicling-america/about-this-collection>

Newspapers.com (n.d.). *Newspapers.com*. Available: <https://www.newspapers.com/>

Ringel, S. (2023). Digitizing the paper of record: Archiving digital newspapers at the New York Times. *Journalism*, 24(2), 245-261. <https://doi.org/10.1177/14648849211023849>

Schwartz, J. M. & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science*, 2(1-2), 1-19. <https://doi.org/10.1007/BF02435628>

The British Newspaper Archive (n.d.). *The British Newspaper Archive*. Available:

<https://www.britishnewspaperarchive.co.uk/>

• 국문 참고자료의 영어 표기

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

Hong, Jong Uk, Jang, Moon Seok, & Ryu, Jun Pil (2025). Design of digital archive of modern Korean bibliography: Focusing on The Chin-Tan Hakpo. *Journal of Humanities*, 110, 79-122. <https://doi.org/10.21211/JHUM.110.3>

Korea University Library (n.d.). *Old newspapers digital archive*. Available:

<https://library.korea.ac.kr/datause/archiving/old-newspaper/old-newspaper/>

Kwak, Seung Jin, Noh, Young hee, & Jeong, Dae Keun (2018). A study on the revitalization of modern literature service based on user awareness survey. *Journal of Korean Library and Information Science Society*, 49(1), 47-75.

<https://doi.org/10.16981/kliss.49.1.201803.47>

Lee, Go Eun, Choi, Yae Eun, Lee, Yu Jin, Kim, Gyeong Min, & Lee, Jee yeon (2025). A study on user needs and perceptions for an integrated online service for Korean modern literature. *Journal of the Korean Society for Information Management*, 42(2), 177-203.

<https://doi.org/10.3743/KOSIM.2025.42.2.177>

National Library of Korea (2024). A study on the development and operation strategies of an integrated online service for modern Korean documents (11-1371029-100004-01).

National Library of Korea (n.d.). *Korea newspaper archive*. Available:

<https://www.nl.go.kr/newspaper/>

Naver (n.d.). *Naver news library*. Available:

<https://newslibrary.naver.com/search/searchByDate.naver>

- Song, Zoo Hyung (2016). A study on the online newspaper archive: Focusing on domestic and international case studies. *The Korean Journal of Archival Studies*, 48, 93-139.
- Yonsei University Institute of Modern Korean Studies (n.d.). Database of modern Korean studies. Available: <https://cmks.yonsei.ac.kr/cmks/database.htm?ch=1>
- Yoo, Choon Dong & Oh, Young Sik (2021). Redefinition and conservation of the modern publication according to changing periods. *Journal of the Institute of Bibliography*, 87, 51-65. <https://doi.org/10.17258/jib.2021..87.51>
- Yoon, Jung Won (1997). A bibliographical study of modern Korean periodicals, 1889-1945. *Journal of Bibliography*, 14, 355-361.

