

# 디지털 리포지터리의 지속가능한 협력 모델로서 하티트러스트 사례 연구

## A Case Study on HathiTrust as a Sustainable Cooperative Model of Digital Repositories

이 유 경(You-Kyoung Lee)\* · 성 윤 아(Yunah Sung)\*\*  
정 영 미(Young-Mi Jung)\*\*\*

### 〈목 차〉

I. 서론	IV. 하티트러스트의 특징점과 가능성
II. 디지털 도서관 협력 모델 개요 및 현황	1. 저작권 관리
1. 디지털 리포지터리의 성장과 관련된 이슈	2. 효과적인 운영 비용
2. 디지털 리포지터리의 협력 모델	3. 디지털 콘텐츠의 공동관리 및 장기적인 보존
III. 하티트러스트 구축과 운영 사례	4. 공동 보존서고 운영
1. 설립 배경과 목적	5. 빅데이터 분석 토대 마련
2. 운영 및 주요 정책	V. 결론
3. 자료 축적 현황	
4. 이용자 서비스 및 기능	

### 초 록

세계의 많은 기관들은 학술 정보 유통을 위해 디지털 리포지터리들을 구축해왔다. 동시에 디지털 리포지터리들은 증가하는 디지털 자원을 어떻게 장기적으로 보존하고 지속가능한 정보 환경 구축할 것인가의 문제에 직면해 있다. 하티트러스트(HathiTrust) 파트너십은 미국 연구도서관들의 디지털 리포지터리의 지속가능한 협력 모델의 필요성에 의해 시작하여 전 세계의 리포지터리들과의 체결을 통해 점차 확장되고 있다. 이에 본 연구에서는 하티트러스트의 설립배경부터 구축, 운영 및 정책, 현황 및 이용 서비스에 이르는 전반적인 사례와 실질적인 운영 과정 등을 다루었고 또한 하티트러스트의 효과와 앞으로 기대되는 가능성 등을 파트너십 단위 기관의 시각에서 다루고자 하였다. 하티트러스트를 통해 파트너십 기관들은 효과적인 운영 비용, 디지털 콘텐츠의 공동 관리 및 장기 보존, 저작권 관리의 용이성, 접근성 확대 등의 특징점이 있으며, 향후에는 인쇄물의 공동보존서고 운영, 빅데이터 분석 토대 마련 등의 기회를 제공할 것으로 기대된다.

키워드: 리포지터리, 디지털도서관, 지속가능성, 하티트러스트, 장기적인 보존, 대학도서관

### ABSTRACT

A great number of institutions around the world have been building digital repositories to communicate scholarly information. Meanwhile, digital repositories have been struggling with how to preserve increased volume of digital contents for the long term and how to build a sustainable information environment. The HathiTrust partnership was established to meet the need of a sustainable collaborative model of digital repositories in research libraries, mainly in North America, and has been expanded globally by signing with other libraries around the world. This paper is dealt with the establishment, operation and policy, construction status, and user service of the HathiTrust. The results presented in this paper include the benefits and potential opportunities of the HathiTrust as a participating member. Partnership in HathiTrust would allow each member institution to provide more cost-effective operations, shared management and long-term preservation of digital content, ease of copyright management, and increased accessibility. In the future it is expected to provide a shared storage of printed materials and to facilitate a big data research center.

Keywords: Digital library, Sustainability, HathiTrust, Long-term preservation, University library

\* 이화여자대학교 정책과학대학원 기록관리교육원(youkyol@ewha.ac.kr)

\*\* Korean Studies Librarian, University of Michigan(yunahs@umich.edu)

\*\*\* 동의대학교 문헌정보학과 부교수(yomjung@deu.ac.kr)

•논문접수: 2016년 11월 20일 •최초심사: 2016년 11월 29일 •게재확정: 2016년 12월 23일

•한국도서관정보학회지 47(4), 443-464, 2016. [http://dx.doi.org/10.16981/kliss.47.201612.443]

## I. 서론

유통구조가 다양해지고 정보량이 증가하는 디지털 환경에서는, 정보를 공개적으로 이용하는 폭 넓은 공유 활동이 정보의 가치를 높이는 방법일 수 있다. 디지털 학술 정보 유통과 공유라는 측면에서 전 세계 많은 대학, 연구기관, 그리고 학회 등은 디지털 리포지터리를 구축 운영해왔다. 초기의 디지털 리포지터리들이 구축된 지 20여년의 시간이 흐른 지금, 이들을 지속적으로 유지, 성장시킴과 동시에 이용자에게는 보다 확대된 정보 공유 활동을 제공하기 위한 지대한 노력이 있어왔다. 특히 이를 위해서는 디지털 정보의 장기적인 보존과 보존된 정보로의 접근성 확보라는 두 가지 요소가 필수적이다(Piper 2013).

디지털 리포지터리는 이러한 요구의 측면에서 2004년 시작된 구글 도서관 프로젝트를 통해 많은 영감을 얻었다. 각각의 디지털 리포지터리에 산재되어 축적되어있는 디지털화된 정보를 공동으로 보존하고 저작권 문제에 함께 대응해서 정보 공유를 활성화 시키는 인프라 구축의 요구가 가시화되었다. 이에 2008년 하티트러스트(HathiTrust)라는 이름의 디지털 리포지터리를 운영하는 연구도서관의 파트너십이 시작되었고 이를 기반으로 하티트러스트 디지털 도서관이 설립되었다. 이 파트너십의 근본적인 목표는 장기적이고 지속가능한(sustainable) 정보 환경을 구축하고, 정보 접근의 영역을 전세계적으로 확장하는 것이다.

하티트러스트는 미국 중부 지역 13개 연구 기관 연합인 Committee on Institutional Cooperation(이하 CIC)와 캘리포니아 디지털 도서관(California Digital Library)에서 시작했지만, 2016년 현재 4개 컨소시엄과 150여 개 개별기관이 참여하는 대규모 디지털 도서관 연합(York 2015)이 되었으며, 미국뿐 아니라 전 세계의 연구 도서관과 관련 기관이 참여하고 있다. 하티트러스트 디지털 도서관은 참여 기관끼리 보유한 디지털 리포지터리의 축적된 자원을 공유하고, 거대한 규모의 디지털 자원을 형성하여 가시성과 접근성을 향상시킴과 동시에 협력적인 장기 보존체계를 구축하여 지속가능한 디지털 리포지터리 모델로 대두되고 있다.

이에 본 연구에서는 디지털 리포지터리를 운영중인 국내 도서관이 직면하고 있는 또는 앞으로 직면할 문제에 대비하기 위해, 접근성과 가시성이 높은 리포지터리의 확대된 정보서비스와 지속가능한 모델로서 하티트러스트의 운영사례를 살펴보았다. 여기에는 하티트러스트의 파트너십과 운영 전반을 포함하였으며, 또한 현재의 파트너십의 효과와 앞으로 기대되는 확장 및 활용 가능성을 포함하였다. 이것은 국내의 리포지터리를 둘러싼 정보 공유 환경의 문제점을 해결하는데 보다 실질적인 시사점을 제공할 수 있을 것이며, 국내를 넘어 국제적으로 지속가능한 하티트러스트와 같은 정보 유통 인프라 참여에 대한 전망 또한 가능해보는 것이다.

## II. 디지털 도서관 협력 모델 개요 및 현황

### 1. 디지털 리포지터리의 성장과 관련된 이슈

디지털 리포지터리는 일반적으로 디지털 콘텐츠의 저장소(container)로 정의되고 있으며 디지털 콘텐츠는 텍스트, 비디오, 이미지, 학습 객체와 데이터 세트 등을 포함한다(Gibbons 2004, 6). 오늘날 리포지터리는 학술정보 유통 환경을 재편하고 있다고 평가될 만큼 디지털 콘텐츠를 온라인상에서 광범위하게 확산시킬 수 있는 효율적인 도구이다.

리포지터리는 구축 주체와 대상에 따라 기관 리포지터리(institutional repository), 주제 리포지터리(subject repository), 그리고 포맷 리포지터리(format repository)로 구분할 수 있다(정영미, 배정희 2015, 531). OpenDOAR(<http://opendoar.org>)의 통계에 따르면 전세계의 리포지터리는 2005년 128개이었던 것이 2016년 현재는 3,200여개가 구축 운영되고 있을 만큼 상당한 규모의 성장을 해왔다. 이 중 본 논문에서 취급하는 기관 리포지터리는 전체 리포지터리의 85.1%로 다수를 차지한다. 기관 리포지터리(이하 리포지터리는 기관 리포지터리를 의미함)는 대학이나 연구소 등을 중심으로 그들 기관이나 소속 연구자가 생산한 디지털 콘텐츠, 또는 기관에서 보유하고 있는 자료들을 디지털화하여 축적하고 이를 공공에 제공하는 것이다. 리포지터리는 연구논문뿐만 아니라 회색문헌, 기관의 기록물, 디지털도서관 컬렉션 등과 같이 다양한 디지털 콘텐츠를 축적·제공할 수 있다.

초기의 리포지터리 운동은 온라인 상의 '접근'에 집중하였고, 이를 위해 대부분의 리포지터리들은 많은 양의 디지털 콘텐츠를 축적하여 어떻게 온라인 상에서 제공할 것인가에 고군분투했다(Association of Research Libraries 2009, 12). 그래서 이 분야 초기의 연구들은 리포지터리의 구축과 운영 관리에 관련된 기술적인 이슈에 관한 것이 대부분이었다. 새로운 리포지터리들은 계속해서 구축되고 있고, 기존의 리포지터리들은 새로운 디지털 콘텐츠를 추가·확장하고 있다. 또한 일각에서는 이들의 더 나은 서비스를 위해 다양한 문제들을 제기하고 있다.

미국연구도서관협회(Association of Research Libraries)의 디지털 리포지터리 이슈에 관한 전문가 그룹(Digital Repository Issues Task Force)은 리포지터리 개발자들의 경험을 토대로 앞으로의 디지털 리포지터리 발전을 위한 핵심적인 이슈들을 보고한 바 있다(Association of Research Libraries 2009, 15-22). 첫째, 오래된 콘텐츠와 새로운 콘텐츠를 포괄하는 서비스를 새로운 형식으로 구축하는 것이다. 초기의 리포지터리들은 소속 연구자들의 출판 전 또는 출판 후 논문과 석박사학위논문의 축적에 집중했었다. 이제는 리포지터리들이 기관이 생산하는 많은 양의 데이터, 이미지, 멀티미디어 저작물, 교수-학습 자료, 각종 디지털화 과정에서 발생한 디지털 레코드 등 매우 다양한 유형의 디지털 콘텐츠를 포

#### 4 한국도서관정보학회지(제47권 제4호)

함하는 것이 요구된다. 이것은 단순히 다양한 디지털 콘텐츠의 축적과 관련된 기술 이 외에 각 객체 별 제작, 수집, 변환 등의 기술적인 문제 또한 복잡해지는 것을 의미한다. 둘째, 리포지터리들을 더 큰 네트워크 환경에 결합하는 것이다. 리포지터리가 높은 수준에서 네트워크 환경에 결합되어 더 광범위하게 검색되고 제공되기 위해서는 공유된 서비스를 위한 다양한 지원과 오픈 API가 매우 중요하다. 셋째, ‘수요자 측면’의 서비스 개발에 관한 것이다. 이제 리포지터리 개발자들은 리포지터리 구축보다는 서비스 제공에 더 많은 관심을 가질 필요가 있다. 서비스 개발은 다양한 이용자 그룹에 의해 제기된 요구 사항과 제약 조건을 면밀히 검토하는 과정을 통해 가능할 것이다. 마지막 핵심 이슈는 리포지터리의 지속가능성에 관한 것이다. 지속가능성은 단순히 장기적인 자금 지원에 관한 것만을 의미하는 것이 아니라 리포지터리 서비스 및 기반 기술에 대한 계속되는 요구를 충족시킬 수 있는 지속적인 협력을 구축할 수 있는 역량과 조직적인 지원에 관한 것이다. 본 논문은 마지막 핵심 이슈인 지속가능성에 초점을 맞춘 차기 협력 모델에 초점을 맞추었다.

이보다 앞선 연구도서관그룹(Research Libraries Group)과 OCLC의 2002년 연구에서는 대규모의 또는 상이한 컬렉션을 결합한 지속가능한 디지털 리포지터리, 즉 신뢰할 수 있는 디지털 리포지터리(trusted digital repository)의 속성을 정의한 바 있다. 리포지터리의 다양한 조건이나 환경과 상관없이 모든 신뢰할 수 있는 리포지터리에 요구되는 기본적인 속성으로 다음과 같은 것을 들 수 있다(Research Libraries Group 2002, 13-15).

- OAI(Open Archival Information System) 참조모형 준수
- 관리상의 책임: 특히 가시성과 지속가능성에 영향을 미치는 물리적인 환경, 백업과 복구 절차, 보안 시스템을 위해 국가, 국제적인 표준을 따름
- 조직의 실행가능성
- 재정의 지속가능성
- 기술과 절차의 적합성
- 시스템 보안
- 절차에 대한 책임: 모니터링, 피드백

이들 속성에는 지속가능한 디지털 리포지터리를 위한 구조적이고 기술적인 요소뿐만 아니라 정책적이고 과정적인 요소 또한 중요하게 취급하고 있다. Li와 Banach(2011)는 미국 내 72개 연구도서관들의 리포지터리를 대상으로 한 디지털 보존에 관한 운영 및 정책 조사에서, 재정과 숙련된 직원에 초점을 맞춘 지속가능성을 검토하였다. 이들 중 63.3%의 리포지터리만이 지속가능한 재정을 확보하고 있었고, 48.3%의 리포지터리만이 훈련된 전담 직원을 고용하고 있는 것으로 조사되었다. Erway(2012)는 리포지터리 비용의 효율성과 이용의 극대화 등 지속가능한 리포지터리를 위한 방법으로 규모화(scaling)를 제안하고 있다. 규모화에는 유사

한 리포지터리와의 파트너십 체결, 웹 상에서의 가시성 확보 등이 구체적으로 제안되었다.

이와 같이 디지털 리포지터리가 다루는 다양한 디지털 자원과 네트워크를 통한 서비스의 지속가능성을 보장하기 위해서는 저작권, 비용, 질적 통제, 보안, 그리고 장기적인 디지털 보존 등 재정적, 기술적, 인적 요소까지 단일 기관이 해결하기 어려운 복잡한 문제들에 직면하게 된다. 문제 해결을 위한 하나의 방안으로 다음 절에서는 디지털 리포지터리 운영 도서관 및 기관의 협력 모델을 검토해 보았다.

## 2. 디지털 리포지터리의 협력 모델

이제까지 다양한 방식의 디지털 리포지터리 또는 디지털 도서관 협력 모델이 존재해왔다. 협력의 목적뿐만 아니라 협력체를 구성하고 있는 기관의 역할과 책임, 권한의 분배 또한 다양하다. 이에 디지털 리포지터리의 재정적 기술적인 지속가능성의 측면에서 몇 가지 협력 모델들을 살펴보았다.

대표적인 디지털 도서관 협력 모델로 NDLTD(Networked Digital Library of Theses and Dissertations)가 자주 언급된다. 이것은 ETDs(Electronic Theses and Dissertations)를 통해 대학의 전자 출판과 개방적인 공유를 지원하는 국제적인 비영리기구로 2008년 이후에는 MetaArchive Cooperation과 함께 보존에 대한 기술적인 지원 또한 수행하고 있으며, 재정은 개인이나 기관의 기부로 감당하고 있다.

보다 최근의 DPLA(Digital Public Library of America, 이하 DPLA)는 2010년 하버드 대학의 인터넷과 사회를 위한 버크만 센터(Berkman Center for Internet & Society)에서 알프레드 P. 슬론 재단(Alfred P. Sloan Foundation)의 지원으로 시작된 대규모 디지털 리포지터리 협력 모델이다. 이후 다양한 재원을 확보하면서 참여하는 그룹들이 확대되었다. 구글북스와 하티트러스트가 대부분 대학도서관과 파트너십을 가지는 반면, DPLA는 미의회도서관(Library of Congress), 미국국가기록원(National Archives), 스미소니언 기구(Smithsonian Institution), 미국저작권청(U.S. Copyright Office), 인터넷아카이브(Internet Archive), 샌프란시스코 공공도서관(San Francisco Public Library), 애플(Apple Inc.) 등 다양한 성격의 기관이나 조직과 파트너십을 가진다. DPLA는 이러한 다양한 유형의 디지털 컬렉션을 단일 플랫폼을 통해 제공함으로써 디지털화된 문화 유산에 보다 더 개방적이고 일관된 접근을 제공할 목적으로 만들어졌다. 그래서 이용자들은 DPLA를 통해 전자저널과 전자책으로 제한된 구글북스와 달리 수백만 권의 책, 팜플렛, 뉴스기사, 원고, 이미지, 녹음자료, 비디오 등 다양한 유형의 콘텐츠들에 접근할 수 있다. DPLA는 오픈 소스 코드, 링크드 데이터, 오픈 메타데이터, 콘텐츠의 디지털화, 새로운 콘텐츠의 형식, 참가자와 이용자를 위한 서비스와 도구들을 개발하기

위해 6개의 실무자 팀으로 구성되어있다.

이 외에도 국가나 비영리 기구의 주도로 이음새 없는 이용자 통합 서비스를 제공하고 기술적인 지원을 위한 분산 통합형 디지털 리포지터리 또는 도서관 협력 모델들은 다수 존재한다. 우리나라의 dCollection이나 국가전자도서관 등도 이 유형의 협력 모델에 해당한다. 본 논문에서는 이와 유사한 기능을 제공함과 동시에 재정이나 운영의 측면에서 단위 기관이 중심이 되는, 다른 유형의 협력 모델을 중심으로 살펴보았다.

구글북스(Google Books)는 2002년 ‘books’라는 시범적인 프로젝트를 통해 공식적으로 시작하였는데 구글이 스캔하거나 OCR(Optical Character Recognition)을 이용하여 텍스트로 변환한 수많은 책과 잡지들의 원문에 대한 검색과 접근을 제공한다. 구글북스는 이용자가 새로운 책을 발견하고 출판사가 새로운 독자를 발견하는 것을 돕는, 모든 언어로 된 모든 책의 편리하고 검색 가능한 가상의 목록을 만드는 것을 목적으로 시작된 프로젝트이다(Google Book Home Page 2016). 저작권이 소멸되었거나 퍼블릭 도메인인 경우 이용자들은 원문 전체를 보거나 다운로드할 수 있다. 여기에 포함된 디지털화된 책들은 구글북스 파트너 프로그램을 통해 출판사나 저자들로부터 제공된 것과 구글북스 도서관 프로젝트의 협약 도서관들에 의해 제공된 것이다. 구글북스 도서관 프로젝트는 2004년 12월에 구글과 5개 대규모 도서관(하버드 대학도서관, 미시간 대학도서관, 뉴욕 공공도서관, 옥스퍼드 대학도서관, 그리고 스탠포드 대학도서관)이 파트너십을 협약하고 이곳의 장서를 디지털화하면서 시작되었다. 이후 이 프로젝트의 파트너십 협약 도서관은 유럽과 일본의 도서관을 포함한 17개 도서관으로 점진적으로 증가하였으며, 2015년 말 이용 가능한 책은 2천5백만권을 넘어섰고 이 프로젝트는 지금도 진행 중이다(Google Book Home Page 2016). 구글북스의 경우 디지털화 자료 선정 및 제공은 협약 도서관들의 역할이고, 모든 예산과 디지털화 과정 및 서비스 제공, 사후 관리는 구글이 주도하고 있다. 일부 디지털화 콘텐츠의 질적 문제가 끊임없이 제기되고 있지만, 협약 도서관이 소장하고 있던 공개도서에 대한 접근성 강화와 대규모화를 통한 도서관 목록 가시성 확대에 기여한 바가 크다고 할 수 있다. 또한 디지털화된 도서의 사본을 구글이 해당 도서관에 제공하기에 도서관 입장에서는 디지털화에 소요되는 많은 비용과 시간을 절감하게 된다. 다만 디지털화된 장서를 유지하고 장기적으로 보존하여 서비스를 지속하는 것은 다시 도서관의 몫이라 볼 수 있다.

스콜라포탈(Scholars Portal)은 캐나다의 OCUL(Ontario Council of University Libraries)의 서비스로 토론토 대학 도서관(University of Toronto Libraries)이 1997년에 구현한 전자저널 서비스가 그 전신이다. 2002년에 시작된 스콜라포탈은 온타리오주에 있는 21개의 모든 대학도서관이 소장하고 있는 디지털 콘텐츠 및 관련 기술을 공유하는 것이다. 스콜라포탈은 전자저널, 색인 및 초록 데이터베이스, 연구데이터, 전자책 등이 디지털 리포지터리에 축

적되어 있는 다양한 유형의 디지털 콘텐츠에 대한 큐레이션, 보존, 그리고 접근 관리를 지원하고 있으며 인적자원 및 인쇄된 자원도 공유, 분담한다. 이를 위해 스콜라포탈은 모든 파트너 기관으로부터 콘텐츠를 제공받으며 서비스 제공 및 개발은 토론토 대학에서 맡고 있다. 스콜라포탈의 운영은 거버넌스 모델에 기반한 협력 모델이며, 서비스 비용을 분담하기 위해 파트너 기관들은 OCUL 비용 분담 모델을 채택하고 있다. OCUL 위원회에 기반한 거버넌스는 초기에는 원활하게 운영되었지만 서비스의 범위와 초점이 변경되고 핵심 서비스 및 선택 서비스에 대한 기대가 높아짐에 따라 오히려 비용 분담 모델이 파트너십 기관간의 충돌의 원인이 되기도 했다(Association of Research Libraries 2009, 26). 하지만 기술적으로 스콜라포탈은 OAI (Open Archival Information System) 참조 모델 및 TRAC (Trustworthy Repositories Audit & Certification) 체크리스트의 요구 사항을 바탕으로 장기간의 개체 무결성 및 유용성을 보장 할 수 있는 디지털 보존 시스템을 만들었고 이에 따라 2013년 2월 CRL (Center for Research Libraries)에서 신뢰할 수 있는 디지털 저장소로 인증 받았다 (Scholars Portal Home Page 2016).

하티트러스트는 2008년 10월 미국 내 연구기관 및 대학도서관이 각 기관의 리포지터리 내 축적된 디지털 콘텐츠를 장기간 보존하고 접근성을 높이기 위해 설립한 파트너십이다. 미국 중부 지역 대학 연합인 CIC와 캘리포니아 대학(University of California) 등 13개의 대학에서 시작된 하티트러스트는 현재 150여개의 파트너 기관으로 확대되었고 1,470만건 이상의 디지털 콘텐츠로의 접근을 제공하는 거대한 플랫폼이자 디지털 보존 리포지터리로 발전하였다. 모든 파트너 기관들은 하티트러스트에 디지털 콘텐츠를 제공하고 미시간 대학(University of Michigan)과 인디애나 대학(Indiana University)은 리포지터리 운영 및 서비스 개발을 수행하고 있다. 하티트러스트의 파트너십은 스콜라포탈과 같이 거버넌스 기반 협력 모델이다. 하티트러스트 초기에 중요한 이슈 중 하나가 파트너 기관으로부터 충분한 피드백을 얻는 방법 자체를 연구할 정도로 거버넌스 모델은 하티트러스트 운영에 중요한 요소이다(Association of Research Libraries 2009, 26). 하티트러스트는 이사회의 통제를 받으며, 이사회는 초기 설립 기관이 임명한 6명, 회원들이 선출한 6명, 의결권이 없는 최고경영자로 구성되어있다. 예산은 초기에는 미시간 대학 예산 시스템에서 지원하고 집행위원회에서 관리되었으나 현재는 분리하여 별도로 관리 되고 있고 일부 재정은 파트너십 기관에 의해 충당되고 있다 (HathiTrust Digital Library Home page 2016). 하티트러스트는 공공의 목적으로 운영 가능하면서 재정 및 기술적인 측면에서 지속가능한 디지털 리포지터리 협력 모델로 평가 받고 있다. 이에 본 논문에서는 하티트러스트의 기본 정책 및 운영에 대한 보다 자세한 내용을 다음의 3장에서 다루었다.

### Ⅲ. 하티트러스트 구축과 운영 사례

#### 1. 설립 배경과 목적

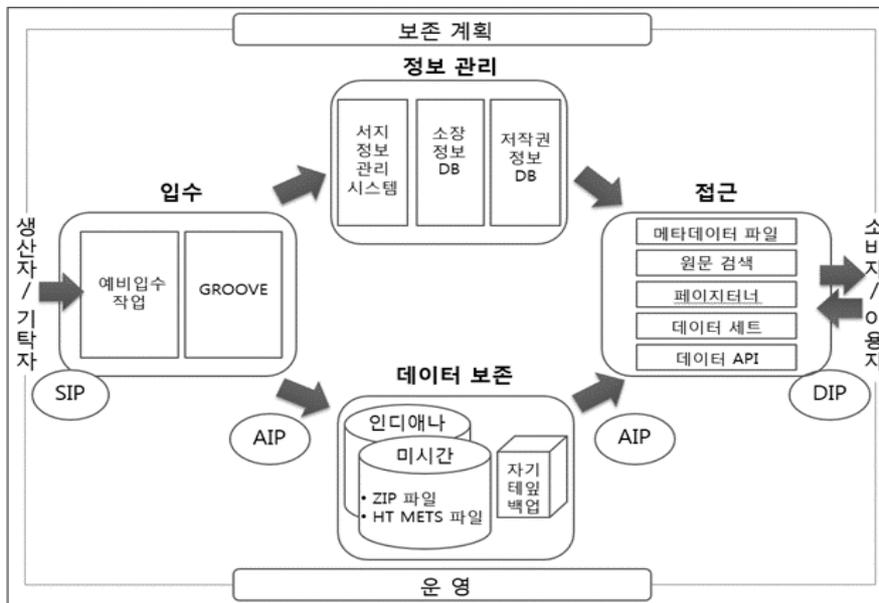
하티트러스트는 2008년 주요 연구기관과 도서관들이 파트너십을 기반으로 설립한 디지털 리포지터리로, 구글북스 디지털화에 참여한 도서관의 디지털 콘텐츠 장기 보존과 접근성 향상을 위해 시작되었다. 2004년부터 시행된 구글북스는 참여 도서관에서 제공한 도서를 스캔한 후에 디지털 사본을 각 도서관에 제공하였는데 스캔 자료의 품질이 낮고, 메타데이터 오류가 잦은 점이 문제점으로 지적되었고 무엇보다 상업 기관에서 반영구적으로 자원에 접근성을 제공할 것인가의 의문(Christenson 2010)이 도서관계를 중심으로 제기되었다. 또한 디지털 콘텐츠 양의 급증에 따른 디지털 보존에 투입되어야 하는 재정 지원문제도 도서관에 강한 압박으로 작용하였다. 당시 미시간 대학 도서관장이었던 Courant(2008)은 한 번의 접속으로 여러 도서관 자료의 검색과 이용을 가능하게 하는 대규모 디지털 도서관프로젝트의 필요성을 제기하면서, 연구도서관들의 협력을 기반으로 한 디지털 리포지터리를 구상하였다. “캠퍼스의 경계를 넘는(above the campus)” 디지털 리포지터리를 구축하려는 취지에 동참해 미국 중부 지역의 연구 기관 연합인 CIC 와 캘리포니아 대학이 중심이 되어 하티트러스트를 시작하였다(Courant and Wilkin 2010; Christenson 2010).

‘하티(hathi)’는 힌디어로 지혜로운 동물인 코끼리를 뜻하고, ‘트러스트(trust)’는 신뢰를 바탕으로 구축되는 협력 활동을 의미한다. 하티트러스트의 최종 목표는 지식과 지혜를 모으고 보존하기 위한 협력 기반 디지털 리포지터리 구축에 있다. 파트너 기관은 운영, 재정, 기술적인 부분 전체에 협력하고 있으며, 이 외에도 학술 연구 커뮤니티를 지원하고, 더 나아가 공익에 이바지 할 수 있도록 자원을 수집, 조직, 보존하며, 인류의 지식재산을 공유하기 위한 협력 활동(Furlough & Glenn 2016)을 수행하고 있다.

하티트러스트는 디지털 자원 보존을 목적으로 운영되는 여타의 기관들과 두 가지 요소에서 차별점을 갖는다. 첫째, 구조적인 면에서 하티트러스트는 멤버십 비용으로 운영되므로 상업기관인 구글, 비영리 기관인 인터넷 아카이브(Internet Archive)와 차별된다. 또한 정부 지원으로 운영되는 DPLA나 유럽 연합의 유로피아나(Europeana) 같은 국가나 지역단위 프로젝트와도 차별성을 갖는다. 둘째, 디지털 자원 보존 기능면에서도 다른 기관과 차별된다. 하티트러스트는 종종 Portico와 CLOCKSS와 비교되는데, 두 기관은 출판사로부터 제공되는 전자 저널과 전자책을 보존 대상으로 하며, 출판사가 더 이상 자료를 제공하지 못할 때가 되어야 접근을 제공하는 다크 아카이브(Dark Archive)의 역할을 한다(Christenson 2010). 반면 하티트러스트는 파트너 기관에서 보유한 모든 디지털 자원 중 퍼블릭 도메인, 즉 저작권을 적용 받지 않는 자료를 다른 파트너 기관들에게 최대한의 접근성을 제공하는 것을 목적으로 운영되고 있다.

## 2. 운영과 주요 정책

장기적인 보존을 보장하기 위하여 하티트러스트 디지털 리포지터리의 기본적인 프레임워크는 OAIS 참조모델을 따른다. 제출정보 패키지(Submission Information Package, 이하 SIP), 보존정보 패키지(Archival Information Package, 이하 AIP), 배포정보 패키지(Distribution Information Package, 이하 DIP)가 입수, 관리, 보존, 접근의 개념에 따라 순서대로 처리된다. 하티트러스트 디지털 리포지터리의 프레임워크는 <그림 1>과 같다. 첫째, 입수는 정보 생산자와 기탁자인 구글과 인터넷 아카이브, 개별 파트너 기관 등이 제출정보 패키지를 제공하면서 이루어진다. 제출정보 패키지 입수 시 예비 입수 과정을 추가로 거친다. 정보 제공처가 다양하고 그에 따라 정보 패키지의 포맷과 메타데이터 등이 다르기 때문에, 예비 입수를 거치며 하티트러스트 리포지터리 기준을 준수하게 된다. 제출정보 패키지는 MARC로 작성된 서지 정보와 콘텐츠로 구성된다. 둘째, 입수된 서지 정보는 데이터 관리의 대상이 된다. 데이터 관리는 서지 정보, 저작권 정보, 소장 정보로 구성된다. 서지 정보는 리포지터리 콘텐츠의 카탈로그가 되며 이를 통해 정보의 중복을 방지할 수 있다. 저작권 정보는 하티트러스트가 정리한 26개의 저작권 속성 중 하나로 분류되며 데이터베이스에 따로 관리된다. 소장 정보는 디지털 자료와 동일한 저작의 물리적 자료가 각 기관에 소장되어있는지 여부를 판단하는 기능을 한다. 셋째, 콘텐츠는 데이터 보존의 대상이 된다. 하티트러스트에 입수되는 기준 콘텐츠 포맷은 이미지는 ITU G4 (bitonal)



<그림 1> 하티트러스트 디지털 리포지터리 프레임워크  
(출처: York 2010, York 2015를 사용하여 재구성함)

TIFF, JP2, 텍스트는 Unicode Text다. GROOVE라는 자동 입수 절차를 거치며, 이 과정에서 개체에 부여된 바코드와 고정성 검사, 하티트러스트 METS(Metadata Encoding and Transmission Standards, 이하 METS) 파일 생성, 영구 URL 부여 작업이 진행된다. 저장소는 미시간대학과 인디애나대학 두 곳에 분산해서 구축했다. 한 저장소가 서비스하지 못할 경우 다른 하나의 저장소가 계속해서 서비스 할 수 있다. 두 곳의 저장소 외에도 자기 테이프 백업이 미시간 대학에서 별도로 진행되고 있다. 넷째, 이용자의 접근은 메타데이터를 기반으로 한 카탈로그, 원문 검색 페이지, 페이지터너 응용프로그램, 데이터 세트 및 데이터 API를 통해 이루어진다.

하티트러스트가 거버넌스 기반의 협력 모델인 만큼 파트너 기관간의 운영을 위한 공동의 약속과 역할은 매우 중요하다. 이를 위해 명문화된 정책으로는 접근과 이용, 서지 메타 데이터 수정, 저작권, 하티트러스트에서 삭제, 디지털 보존, 디지털 콘텐츠 보존에 대한 지침, 개인 정보 보호, 그리고 품질 요소 등이 있다. 이 중 지속가능한 협력 모델로서 주요한 정책을 정리해보면 다음과 같다.

#### 가. 저작권 관리 및 접근 정책

저작권은 접근성에 직접적인 영향을 미치는 중요한 요소이지만 개별 기관 단위에서 복잡한 저작권 상태를 추적하고 관리하기는 쉽지 않다. 이를 위해 하티트러스트는 저작권 데이터베이스를 구축하고 리포지터리에 저장된 개별 자료에 대해 퍼블릭 도메인, 저작권이 유효한 자료, 저작권 정보가 불분명한 고아 저작물 등 저작권 정보를 수집하고 갱신하여 제공하고 있다.

이용자가 자료에 접근하고자 할 경우, 접근은 IP 주소 추적, 이용자 인증, 접속 지역 추적과 저작권 데이터베이스를 이용해서 이루어지며 그 결과는 페이지터너 응용프로그램에서 구현된다. 저작권 상태는 자료가 입수되는 시점에 레코드에서 제공되는 정보에 의해 정해진다. 아래 <표 1>에 나와 있는 바와 같이 자료가 미국 내에서 1923년 이전에 발간되었거나 미국 외 지역에서 1876년 이전에 발간된 경우, 그리고 미국 연방 정부 문서인 경우 퍼블릭 도메인으로 분류된다. 미국 내 이용자에게는 미국 외 지역에서 1876년과 1922년 사이에 발간된 자료 역시 퍼블릭 도메인으로 분류되지만, 미국 외 국가에서 접속한 이용자들에게는 저작권이 살아있는 자료로 분류된다. 원문 다운로드를 살펴보면 아래 표1에 설명된 대로 구글에서 스캔한 자료일 경우 원문다운로드를 하는데 예외 조항이 있다. 저작권이 유효한 자료는 원문 다운로드를 일체 제공하지 않으며, 나머지 세 가지 저작권 유형에 속한 구글이 스캔한 자료는 파트너 기관, 또는 미국 내 파트너 기관에게만 제공된다. 저작권자가 하티트러스트에 접근 권한을 제공함과 동시에 구글 스캔 자료인 경우에는 크리에이티브 커먼즈 라이선스(Creative Commons License, 이하 CCL)를 통해서만 원문 다운로드가 가능하다(<표 1> 참조).

〈표 1〉 하티트러스트 저작권 및 접근 수준 구분. (출처: HathiTrust 2016)

저작권 유형	자료 검색	자료 보기	원문 다운로드
전 세계 저작권법에서 인정된 퍼블릭 도메인 저작 (1873년 이전 발간)	제한 없음	제한 없음	제한 없음 (구글에서 디지털화된 자료 제외- 파트너 기관만 가능)
미국 내에서만 퍼블릭 도메인인 저작 / 미국 외 지역에서 1873년과 1923년 사이에 발간된 저작	제한 없음	미국 내에서 접속할 경우만 가능	미국 내에서 접속할 경우만 가능 (구글에서 디지털화된 자료 제외 - 미국 내 파트너기관만 가능)
저작권자가 하티트러스트에 오픈 액세스를 허용한 저작	제한 없음	제한 없음	제한 없음(구글에서 디지털화된 자료는 CCL 하에 허용)
저작권이 유효한 저작 (저작권 정의가 완료되지 않은 자료 포함)	제한 없음	제공하지 않음	제공하지 않음

추가적으로 도서관과 기록관의 재사용을 정의하고 있는 미국 법률 108 조에 의거해서 훼손되었거나 활용할 수 없는 자료들은 도서관 내부의 온라인 접근만 허용된다. 빅데이터와 같이 컴퓨터를 통한 연구와 같은 모든 다른 유형의 접근은 미국 저작권법과 미시간대학교 법률 자문위원실의 가이드를 받아 이루어진다.

#### 나. 비용 정책

하티트러스트의 재정적 지속가능성은 파트너 기관에서 연간으로 지불하는 참여 비용을 통해 보장된다. 현재 적용되는 비용 모델은 2013년에 개정된 것으로 비용 항목은 다음과 같이 두 가지로 구성된다. 첫째, 하티트러스트가 보유한 퍼블릭 도메인 자료를 유지하기 위해 파트너 기관에게 동등하게 부여되는 지원 비용, 둘째, 개별 기관에서 보유하고 있는 실물 장서와 중복되는 저작권이 살아있는 자료의 공유에 대한 비용이다.

비용 산출은 다섯 가지 고정 요소와 한 가지 조정 가능한 요소가 영향을 미친다. 고정 요소는 저장 서버 사용료의 개념이고, 조정 가능한 요소는 다양한 활동에 기반한 비용 산출을 의미한다. 고정 요소 다섯 가지는 아래와 같다(Wilkin 2013, 2).

- 하티트러스트의 퍼블릭 도메인 자료의 수 (PD)
- 하티트러스트의 디지털 소장 자료와 중복되는, 파트너 기관에서 소장 중인 인쇄 자료 중 저작권이 유효한 자료의 수 (IC)
- 저작권이 유효한 특정 자료를 소장하고 있는 파트너 기관의 수 (H)
- 총 파트너 기관 수 (N)
- 하티트러스트의 기본 보존 인프라스트럭처 비용 (C)

조정 가능한 요소는 탄력적으로 적용되는 배수(X)로서 이사회에 의해 정해진다. 현재 기본 수익 배수는 1.5배이며 수익은 새로운 서비스와 기능을 개발하는 데 투자된다. 고정 요소와 조정 가능한 요소를 적용한 비용 산출 공식은 다음과 같다.

- 퍼블릭 도메인 자료를 지원하기 위해 파트너 기관에게 동등하게 분배되는 금액:

$$(PD*X*C)/N$$

- 하티트러스트의 디지털 소장 자료와 중복되는, 파트너 기관에서 소장 중인 실물 장서 중 저작권이 유효한 자료를 지원하는 금액:  $IC = (C*X)/H$

#### 다. 디지털 자원 공동 관리 및 보존 정책

하티트러스트는 장기적인 보존을 위해 기본적으로 디지털 보존 표준과 공개된 콘텐츠 포맷을 채택하고 있다. 세부적으로 그 전략을 살펴보면 다음과 같다. 첫째, 디지털 보존 표준에 맞는 표준화된 공개 콘텐츠 포맷을 이용한다. 해당 포맷들은 다양한 플랫폼에서 채택되었기 때문에 활용성이 높으며, 차후 새로운 보존 포맷으로의 마이그레이션에도 대응이 용이하다. 디지털 보존의 국제 표준인 OAIS 참조 모델과, METS와 PREMIS(Preservation Metadata Implementation Strategies, 이하 PREMIS)를 적용했다. 감사 증적은 TRAC 준수를 통해 리포지터리의 안전성, 마이그레이션, 접근성 역량을 평가한다.

둘째, 도서관 영역에서 기술, 표준, 모범 실무를 통해 정의된 비트 수준(bit-level)의 보존과 포맷 마이그레이션을 진행한다. 하티트러스트의 보존 포맷은 600dpi로 저장된 TIFF ITU G4 파일과 200에서 400dpi사이 해상도로 저장된 JPEG 또는 JPEG2000 파일, 유니코드 문자, 주로 METS로 정의된 XML 파일이다. 콘텐츠 입수 시에 OAIS 표준을 적용한 리포지터리 디자인에 적절한지, METS와 PREMIS 메타데이터 표준이 제대로 적용되었는지 등의 유효성 검사를 실시하고 TRAC을 통해 감사 증적을 실시한다.

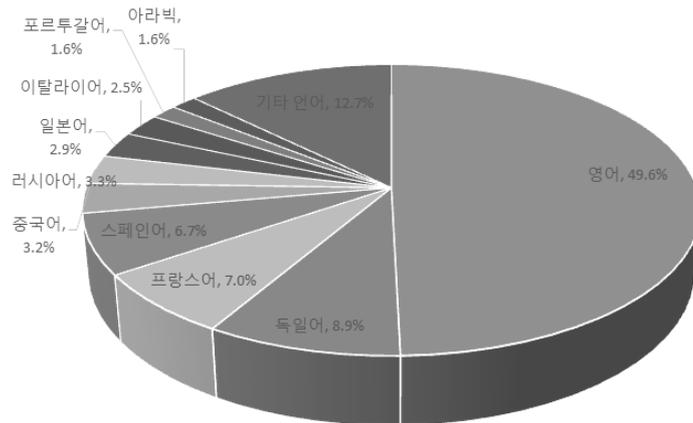
셋째, 분기마다 자동으로 입수 버전과 비교하는 콘텐츠 무결성 확인 작업을 정기적으로 실행한다.

넷째, 안전성 확보를 위해 미리 사이트를 구축한다. 주요 리포지터리는 미시간대학에서 보유하고 있으며, 동일한 데이터를 동시에 보관하는 미리 사이트를 인디애나대학에 따로 마련해서 보존의 안정성을 높이고 있다.

### 3. 자료 축적 현황

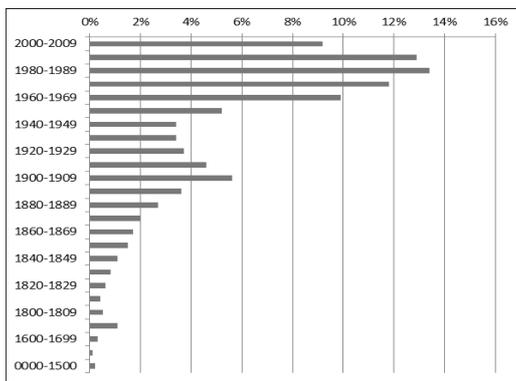
2016년 11월을 기준으로 하티트러스트 리포지터리에는 1,470만 건 이상의 디지털화된 도서와 정기 간행물, 문서 등이 축적되어 있다(HathiTrust Digital Library Home Page 2016). 이중 약 736만 건이 전자책, 약 40만 건이 전자저널이다. 이중 퍼블릭 도메인으로 분류된 자료는 전체 자료의 39%로 약 570만 건이다. 리포지터리 내 자료는 퍼블릭 도메인과 저작권이 유효한 자료, 정부 문서 등으로 구성되며, 파트너 기관에서 제공하는 디지털화된 자료의 입수작업은 매일 진행된다. 장서 구성은 언어별, 연대별로 살펴볼 수 있다. 언어별 구성을 살펴보면 총 460개의 언어로 작성된 자료들이 하티트러스트 리포지터리에 보존되어 있으며, <그림 2>와 같이 상위 10개 언어가 전체 자료의 87.3%를 차지하고 있다. 이중 영어 자료가 49.6%로 가장 많고, 그 뒤로

독일어 8.9%, 프랑스어 7%, 스페인어 6.7%, 러시아어 3.3%, 중국어 3.2%, 일본어 2.9%, 이탈리아 2.5%의 순서이며, 포르투갈어와 아라비아어는 각각 1.6%이다. 상위 10개 언어를 제외한 기타 언어 비율은 12.7%이며, 한국어 자료는 0.21%에 해당한다. 한국어 자료의 비율이 낮은 이유는 한국의 대학 도서관 중에서 하티트러스트 파트너 기관으로 참여하고 있는 곳이 없기 때문이다.

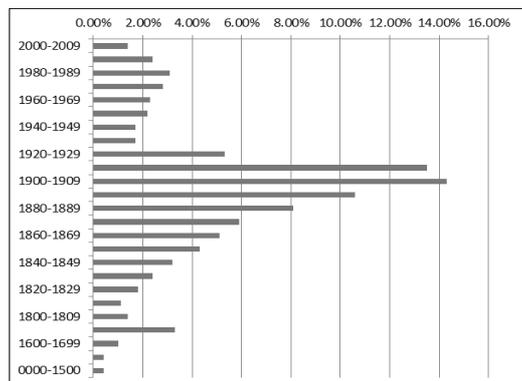


〈그림 2〉 언어별 자료 구성(출처: HathiTrust Digital Library Home Page의 통계정보 재구성)

연도별 구성을 살펴보면 하티트러스트에서 이용이 가능한 퍼블릭 도메인 중 1920년 이전에 발간된 자료가 전체 퍼블릭 도메인의 82.1%를 차지한다. 이는 1923년 이전에 발간된 자료와 미국 외 지역에서 1876년 이전에 발간된 자료를 퍼블릭 도메인으로 정의하는 미국 저작권법을 기준으로 하기 때문이다. 저작권과 상관없이 리포지터리에 입수된 전체 자료의 연도별 구성을 살펴보면 1960년대 이후부터 현재까지 발간된 자료가 전체 자료의 57.2%로 절반 이상에 해당된다(〈그림 3〉, 〈그림 4〉 참조).

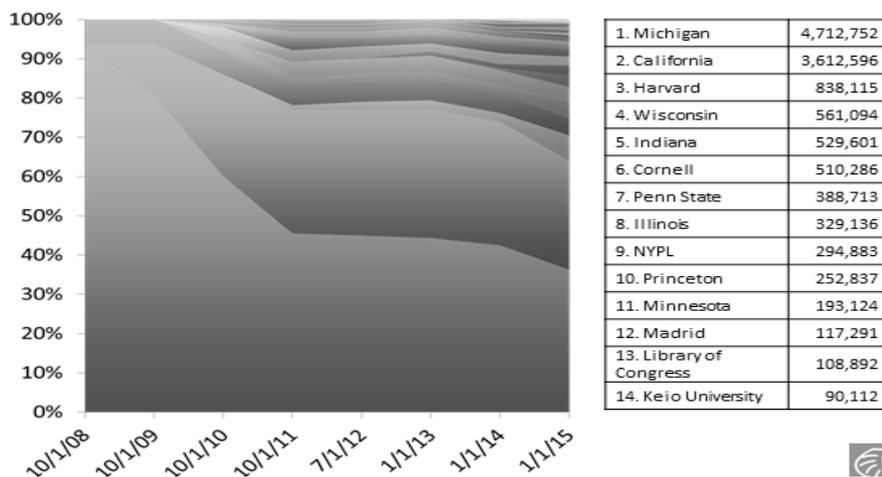


〈그림 3〉 연도별 자료 분포  
(출처: HathiTrust Digital Library Home Page)



〈그림 4〉 연도별 퍼블릭 도메인 분포  
(출처: HathiTrust Digital Library Home Page)

리포지터리내 입수된 자료의 상당부분이 1960년대 이후 발간된 자료이지만, 접근과 활용이 가능한 자료는 대부분 1920년대 이전에 발간된 자료이다. 그 이유는 초기 구글북스 도서관 프로젝트에 참여했던 5개 도서관 중 미시간대학 도서관을 제외한 나머지 도서관들은 퍼블릭 도메인 자료 위주로 선별해서 구글에 제공, 디지털화했기 때문이다. 반면 미시간대학 도서관은 전체 장서를 디지털화해서 보존하되 저작권이 유효한 자료는 이용자의 접근과 이용에 제한을 두었다. 해당 내용은 <그림 5>을 참조해서 해석할 수 있다. 하티트러스트가 시작되었던 2008년에는 미시간대학교 도서관이 자료 기탁의 90%이상을 차지하고 있다. 캘리포니아 디지털 도서관과 하버드대학교 도서관, 위스콘신대학 도서관 등의 자료 기탁이 증가하기 시작한 2010년까지도 미시간대학이 기탁한 자료가 하티트러스트의 상당부분을 이루고 있는 것을 알 수 있다.



<그림 5> 연도별 기탁기관 및 자료 제공량 (출처: York, 2015 HathiTrust overview)

#### 4. 이용자 서비스 및 기능

하티트러스트는 앞에서 살펴본 바와 같이 유동성 요소의 비용 산출을 통해 디지털 리포지터리의 이용자 서비스 확대 및 개발을 위해 지속적인 노력을 해왔다. 파트너 기관들은 자신의 디지털 콘텐츠들에 대한 보다 확대된 접근성과 가시성을 보장 받을 수 있으므로 이용의 활성화를 유도할 수 있고 이를 통해 부가적으로는 디지털 리포지터리 구축 및 운영의 정당성을 확보할 수 있다. 하티트러스트 디지털 리포지터리에서 제공하는 특징적인 이용자 서비스 및 기능을 해당 홈페이지(HathiTrust Digital Library Home Page 2016)를 조사해서 정리해보면 다음과 같다.

첫째, 이용자 인터페이스로 페이지터너(Page Turner) 응용 프로그램을 제공한다. 퍼블릭 도메인 자료들은 페이지터너 메커니즘을 통해 검색되고 이용될 수 있다. 메커니즘은 몇 가지 인증 로직의 조합에 의해 자동으로 매핑되어 이용자가 따로 로그인을 하지 않아도 자료에 접

근이 가능한지 여부를 자동으로 계산한다. 인증 로직은 COSIGN, 십볼렛(Shibboleth)인증 시스템, 접속자 IP주소, 하티트러스트 저작권 데이터베이스, 리포지터리 소장정보 데이터베이스, 접속국가 정보를 얻을 수 있는 GeoIP 데이터베이스를 활용한다(HathiTrust 2016). 검색된 화면에는 자료의 표제와 서지정보, 저작권 상태 정보를 제공하며, 또한 해당 자료를 입수 할 수 있는 다양한 옵션을 제공한다. 월드캣 연동을 통해 접속자와 가장 가까이에 있는 도서관 정보를 제공하는 ‘Find in a Library’, 구입을 원하는 경우 상업 웹사이트로 연결해주는 ‘Buy a Copy’ 기능을 제공한다. 다운로드 시에는 PDF 파일, OCR 텍스트 형태로 제공된다. 원문 보기는 상하 스크롤 방식, 넘겨보기 방식, 썸네일, 한 페이지씩 보이는 페이지뷰 방식, 이미지를 삭제한 단순 텍스트 보기의 다섯 가지 방식을 제공한다.

둘째, 하티트러스트는 정보 접근에 대한 사명을 가장 핵심으로 다루고 있다. 그 일환으로 카탈로그 검색과 원문 검색을 제공한다. 카탈로그 검색은 서명, 저자, 출판 년도 등의 서지정보를 대상으로 검색해서 소장 정보를 제공한다. 원문 검색 서비스는 OCR 도구를 통해 추출된 단순 텍스트만을 대상으로 키워드 검색이 가능하도록 설계되었으며, 리포지터리 내 전체 자료를 대상으로 이루어진다.

셋째, 가상 장서 구축(collection builder) 기능은 이용자 혹은 하티트러스트 관리자가 가상 장서를 구성하고 출판할 수 있도록 설계된 기능이다. 이 기능은 미시간대학교에서 제공하는 ‘Friends’계정을 만들면 파트너 기관 소속 이용자가 아니더라도 이용 가능하다.

넷째, 파트너 기관이 안전하게 접근하고 검색할 수 있는 도구 개발을 목표로 개방형 서비스를 제공하고 하티트러스트 내 콘텐츠에 대한 다양한 접근을 제공하기 위해 다종의 API를 제공하고 있다. 대표적으로 서지정보 API와 데이터 API가 있다. 서지정보 API는 자료 검색과 카탈로그 통합을 지원하며, 데이터 API는 디지털 객체에 내포된 데이터에 외부의 기계가 접근하는 것을 지원한다.

다섯째, 하티트러스트 연구 센터를 설립하고 리포지터리 코퍼스를 대상으로 광범위하고 다양한 분석이 가능한 데이터 마이닝 도구와 서비스를 제공하고 있다. 내부 자료뿐만 아니라 외부 자료까지 분석 가능한 데이터마이닝 도구를 개발했다. 이와 함께 컴퓨터를 이용한 작업과 분석을 하는 연구자들을 위해 샘플 데이터세트를 배포하고 있다. 샘플 데이터세트를 배포하는 목적은 연구자들에게 리포지터리의 구조에 대한 이해를 제공하기 위함이다.

여섯째, 시각 장애가 있거나 다른 이유로 인쇄 매체를 사용할 수 없는 이용자를 위한 서비스를 개발하여 보편적인 정보접근을 지원한다. 파트너 기관에 소속된 시각 장애 이용자들은 지정된 프록시를 통해 특별히 하티트러스트 내에 저작권이 유효한 자료도 합법적으로 사용할 수 있다. 이용자 인터페이스인 페이지터너는 장애인들의 정보 접근을 보장하는 미국 재활법 508조가 제시하는 기준을 통과했다.

일곱째, 일반 이용자뿐만 아니라 파트너 기관에서 콘텐츠를 제공하는 담당자도 이용자로 정의하고, 업무의 편의를 위해 디지털화된 도서와 정기 간행물 콘텐츠를 자동으로 입수하는 메커니즘 ‘GROOVE’를 개발했다. GROOVE는 “Google Return Object-oriented Validation Environment”의 약자로, 처음에는 구글의 디지털화 자료가 자동으로 입수되는 메커니즘이었으나 점차 파트너 기관의 디지털 콘텐츠도 자동으로 입수 절차를 규칙적으로 수행할 수 있도록 확장되었다.

#### IV. 하티트러스트의 특징점과 가능성

장기적이고 지속가능한 정보 환경을 구축하고, 정보 접근의 영역을 전세계적으로 확장하는 것을 목표로 2008년에 시작된 하티트러스트는 그 동안 14개의 파트너 기관에서 전세계의 150여개로 증가할 만큼 규모도 커지고 서비스도 고도화되고 있다. 이 장에서는 디지털 리포지터리들간의 대규모 협력에 의한 특징점과 파트너 기관으로 참여함으로써 얻을 수 있는 혜택, 현재 추진중인 새로운 프로그램을 통해 기대되는 가능성을 정리해보고자 한다. 특히 파트너 기관의 측면에서 얻을 수 있는 저작권 관리, 효과적인 운영비용, 디지털 콘텐츠의 공동 관리와 보존, 인쇄자원에 대한 공동 보존서고 운영 및 빅데이터 분석의 토대 마련 등을 중심으로 제시하였다.

##### 1. 저작권 관리

하티트러스트의 저작권 데이터베이스는 하티트러스트 콘텐츠의 합법적이고 적합한 접근권을 보장하기 위해 고도화된 저작권 관리 시스템을 통해 유지되고 있다. 콘텐츠의 저작권 정보는 각각의 디지털 콘텐츠와 함께 저장되어 있기에 쉽게 추적할 수 있으며 수시로 갱신된다. 저작권은 디지털 콘텐츠의 접근에 직접적인 영향을 미치는 중요한 요소이기도 하고, 실제 디지털 리포지터리를 운영하는 기관의 측면에서 문제가 발생했을 시 가장 대응하기 어려운 문제이기도 하다. 또한 개별 기관 단위에서 복잡한 저작권 상태를 추적하고 관리하기는 쉽지 않기 때문에 법적인 문제도 빈번하게 발생할 수 있다. 파트너십을 통한 저작권에 대한 공동 대응뿐만 아니라 다양한 기술과 방법을 이용한 저작권 데이터베이스의 관리는 단위 기관의 디지털 콘텐츠의 생산과 제공 활동을 고무시킨다.

그리고 파트너 기관은 하티트러스트 내에 축적된 퍼블릭 도메인, 고아저작물과 같이 저작권이 없는 디지털 콘텐츠에 대한 개방적인 접근과 더불어 저작권이 있는 일부 콘텐츠에도 확장하여 접근 가능하다. 예를 들면 소장하고 있는 인쇄자료가 너무 낡거나 파손 위험이 있어 이용이 제한되는 경우, 또는 예전에 소장했었으나 현재는 분실 또는 폐기되어 이용이 불가능한 경우는 다른 파트너 기관에서 디지털화된 콘텐츠를 중복해서 가지고 있다면 하티트러스트 내의 저작권 데이터베이스의 인증 과정을 통해 제공할 수 있다.

## 2. 효과적인 운영 비용

하티트러스트는 각 파트너 기관들이 부담하는 멤버십 비용으로 운영된다. 이는 영리기관인 구글에서 운영하는 구글북스나, 국가나 기구의 지원이나 개인의 기부로 유지되는NDLTD, 유로피아나, DPLA와는 차별되는 디지털 리포지터리 협력 모델이다. 특정한 재정에 의존하지 않고, 정책이나 주변 환경의 변화에 따른 재정 지원의 불확실성에 영향을 받지 않는 하티트러스트는 파트너 기관의 비용 제출과 이에 따른 적정한 수익으로 디지털 리포지터리를 유지 관리하고 있으며 새로운 서비스 개발을 위한 연구에 안정적으로 투자하고 있다. 이러한 재정적 자립은 하티트러스트의 지속가능성을 보장하는 주요 요인이 된다.

또한 하티트러스트 파트너십 비용을 지불하고 있는 파트너 기관은 실제로 직간접적인 예산 절감 효과를 보고 있다. 직접적인 예산절감의 사례로 예일대학 도서관(Yale University Library 2011, 9-10)을 들 수 있는데, 이 도서관은 하티트러스트를 통해 분실되거나 손상되어 대출이 불가능한 인쇄자료의 대체재와 소장되어 있는 인쇄자료의 대체재, 상호대차에 투입되는 비용의 절감 만으로도 연간 총 4,273,050달러의 혜택이 있었다고 분석했다. 간접적인 예산절감의 효과로는 퍼블릭 도메인의 이용에 대한 혜택, 공동 저장서버와 관리, 기술과 서비스 개발, 구축된 자원의 접근성과 가시성 확대 등으로 경제적인 가치로 환산하기 어렵지만 직접적인 예산절감 효과보다 클 것으로 예상된다.

하티트러스트는 기관의 참여의 확대와 자료의 추가를 통해 파트너 기관의 비용은 점차 낮추고, 반대로 혜택은 높아질 수 있도록 하기 위해 2013년에 비용모델을 수정해서 적용하고 있다.

2013년에 개정된 비용모델에 의하면 총 자료가 155만권이고 퍼블릭도메인 자료가 42만 권이라고 가정할 때, A도서관의 30% 가 저작권이 유효한 자료이고 파트너 기관의 수가 3인 경우 연간 비용은 약 200,000달러이다. 하지만 A도서관의 15%가 저작권이 유효한 자료이고 파트너 기관의 수가5인 경우는 연간 비용이 66,000달러로 감소한다. 물론 하티트러스트 파트너십 기관의 연간 비용을 산정에는 다양한 변수가 작용하지만 기본적인 가장 큰 변수는 참여 기관의 수와 자료의 추가에 있다. 앞으로 더 많은 기관이 파트너십으로 참여할 경우 각 기관의 부담 비용은 줄어들 것이 예상되기 때문에 개별 기관의 디지털 리포지터리 운영 비용 과 비교해서 하티트러스트 참여의 비용 효과는 극대화 될 것이 기대된다.

## 3. 디지털 콘텐츠의 공동 관리 및 장기적인 보존

하티트러스트 파트너십의 가장 근본적인 취지이자 특장점은 디지털 콘텐츠의 효과적이고 효율적인 관리, 보존, 그리고 이에 따른 디지털 콘텐츠의 신뢰성 높은 제공에 있다. 지금까지 전세계적으로 실험적인 디지털화 및 디지털 도서관 프로젝트들이 다양하게 진행되어 왔으나, 대부분의 프

로젝트는 단기적이었거나 실행 도중에 중단되었다. 그 이유 중의 하나는 여러 기관이 공동으로 참여한 프로젝트는 서로 상이한 표준의 디지털 콘텐츠를 통합하고 관리하는데 어려움이 많았기 때문이다. 또한 각 기관에서 보유하고 있는 디지털 콘텐츠는 적절한 시기에 적절한 절차를 걸쳐 지속적으로 보존해야 하는데 이런 보존 작업은 단위 기관에게는 상당히 복잡하고 어려운 일이었다.

그러나 하티트러스트는 파트너 기관들에게 디지털화 도구와 포맷 결정, 리포지터리 구축과 관리, 매체 노화에 대비한 마이그레이션과 같은 복잡하고 어려운 디지털 콘텐츠의 장기적인 보존과 관리를 상호협력을 통해 공동으로 대응함으로써 각 기관이 개별적으로 투입해야 하는 노력을 효율적으로 감소시켜줄 뿐만 아니라 이에 따르는 위험부담도 줄여준다.

하티트러스트는 이미 명문화된 보존 정책뿐만 아니라 OAIIS 참조모형 및 다양한 장기적인 보존 표준을 준수하고 이를 위한 안정된 시스템을 갖추고 있다. 앞에서 살펴본 바와 같이 지속가능한 디지털 리포지터리의 요건들을 충분히 충족하고 있으며 TRAC으로부터 인정받은 신뢰성 높은 디지털 리포지터리이다. 게다가 하티트러스트는 장기 보존을 위한 별도의 컨소시엄을 운영하고 있어 장기적으로 지속가능한 모델이 되기 위해 계속 노력 중이다.

#### 4. 공동 보존서고 운영

미국 도서관의 장서보유현황을 살펴보면 여러 기관에서 동일한 서적을 소장하고 있기도 하고, 한 기관에서 유일본을 소장하고 있기도 한다. 늘어나는 장서량에 따른 도서관 서고 공간 확보와 디지털 리포지터리의 역할이 중시되는 현실에 당면한 도서관계에서는 공동 보존서고에 대한 요구가 더욱 절실해지고 있다. 하티트러스트 목표 중 하나가 파트너 기관간의 공동 저장소로서 통합된 디지털 리포지터리 운영을 통해 실물장서의 공간과 관리에 드는 장기적인 자본과 운영비를 절감하는 것이었다. 하티트러스트에 인쇄자료와 동일한 디지털 콘텐츠가 존재한다면 도서관에서 해당 인쇄자료를 폐기하더라도 언제라도 디지털 콘텐츠에 접근 가능하다. 그러므로 어느 도서관이 해당 인쇄자료를 폐기할 것인가, 또 어느 도서관이 인쇄자료를 영구 보존할 것인가 등 인쇄자료에 대한 공동 보존서고 문제는 하티트러스트 내에서 여러 차례 논의되어 오던 중, 2011년 전체 회의에서 이에 대한 프로그램 개발지원계획이 채택되었다.

이 프로그램의 목표는 하티트러스트의 디지털 리포지터리에 있는 전체 단행본에 해당하는 인쇄자료를 공동 보존서고에 보존하고, 인쇄자료와 해당 디지털 콘텐츠를 연계시킴으로써 이 두 장서간의 보존을 보장하며, 하티트러스트 파트너 기관의 장서관리 비용을 절감시키는 것이다. 2014년에는 이 프로그램을 위해 하티트러스트 내에 실무팀(HathiTrust Print Monographs Archive Planning Task Force)이 조직되었고, 2015년에는 분산된 공동 보존서고를 위한 필수 정책, 운영 계획, 비즈니스 모델 개발하여, 현재 100여개 기관이 참여한 1단계(2016- 2017) 작업이 진행 중이다.

기본적인 정책으로는, 파트너 기관은 지정된 인쇄자료들을 향후 20년간 (2035년까지) 보

존하며, 요청 시 하티트러스트 파트너 기관에 대출해주어야 하며, 방화 방재 시설과 온도 습도 조절이 되어 있는 서고에 안전하게 보관하는 것이다. 또 어떤 도서가 어느 기관에 보존되어 있는가는 OCLC 월드캣 레코드 상에 표시하기로 결정했는데, 각 기관은 OCLC의 인쇄자료 보존을 위한 메타데이터 가이드라인에 따라서 소장 레코드를 생성하고 MARC의 583필드에 보존과 관련된 노트를 구체적으로 표기하도록 했다.

이 프로그램이 정착되면 복본이나 유일본에 대한 장기적인 보관이 공동으로 이루어지게 됨에 따라 각 파트너 기관마다 보존에 따르는 비용 절감은 물론이고, 서고 공간도 줄일 수 있어서 도서관 이용자들의 요구에 맞는 효율적인 도서관 공간을 제공할 수 있게 된다.

## 5. 빅데이터 분석 토대 마련

하티트러스트는 도서관들에서 구축한 다양한 디지털 자원들의 통합과 이들에 대한 광범위한 접근을 제공함으로써 도서관의 자원에 대한 빅데이터 분석의 토대를 제공한다. 하티트러스트 연구 센터(HathiTrust Research Center, 이하 HTRC)는 하티트러스트 디지털 리포지터리에 구축된 데이터 양이 늘어남에 따라 연구자들의 데이터 활용을 지원하기 위해 2011년에 설립되었다. 하티트러스트, 인디애나 대학(Indiana University), 일리노이 대학 얼바나삼페인(University of Illinois at Urbana-Champaign)의 참여로 설립된 HTRC는 고도화된 계량 분석을 제공하기 위해서 가상의 인프라와 최신의 소프트웨어 툴을 개발하고 있으며, 다양한 교육 프로그램을 통해 연구자들에게 방대한 양의 디지털 텍스트를 효과적으로 분석하는 방법을 전수하고 있다. 이용자들은 HTRC를 통해서 디지털자료에 접근할 수 있으며, 토픽 모델링, 태그 클라우드, 개체 추출 등을 할 수 있고, 또한 MARC 다운로드와 단어빈도수와 같은 기능도 사용할 수 있다. 현재 하티트러스트 디지털 리포지터리에는 1400만 건이 넘는 텍스트 코퍼스가 축적되어 있다.

이용자들에게는 하티트러스트 포털(<http://htrc2.pti.indiana.edu>)을 통해서도 텍스트 분석을 제공한다. 즉 여기에서 제공하는 텍스트 분석 툴을 실행하여 하티트러스트 디지털 리포지터리에 축적되어 있는 디지털 콘텐츠로 작업 세트를 구성하면 분석 결과에 따라서 시각화도 구현할 수 있다. 대표적인 활용 사례로 HTRC는 계량 분석 학자들을 위해 2015년에는 3개, 2016년에는 4개의 관련된 프로젝트를 지원하며 새로운 서비스 개발을 추진하고 있다. 예를 들면 2016년에는 카리브해 지역의 황열병에 대한 역사 탐구를 위해, 황열병을 참조하는 텍스트의 정서 분석 포함해서 유럽의 시각과 카리브해 주민들에 의해 묘사된 재해를 비교하는 연구를 수행하고 있다. 방법은 오염(filth)과 황열병에 관한 하티트러스트 디지털 도서관의 텍스트 코퍼스를 구축하여 1650년부터 1902년까지 'filth'와 'filthiness' 용어의 사용을 추적하는 것이다. 이와 유사하게 20세기 'creativity'와 'creative' 용어의 의미 변화와 사용 증가를 발견하기 위한 프로젝트, 시카고 건축 대학과 지난 75년간 역사를 조사하고 그것의

국제적 확산과 영향력 패턴을 알아보기 위해 텍스트 마이닝 기술을 사용하는 프로젝트, 그리고 낭만주의로맨틱 시대의 영국 소설을 연구하기 위해 디지털 인문학 접근법과 관련해서 정보 이론과 정보 밀도에 관한 두 가지 가설을 테스트하는 프로젝트 등이 수행 중이다.

하티트러스트를 통해 접근 가능한 많은 양의 다양한 유형의 디지털 자원은 주로 학위논문이나 학술논문의 텍스트 분석으로 제한되어온 이 분야 연구 지원 서비스를 보다 다양한 형태로 확장 가능하게 할 것으로 기대된다.

## V. 결론

디지털 리포지터리들이 등장 한 이후 20여년의 시간이 지났다. 그 동안 세계의 많은 대학 및 연구 기관들은 디지털 리포지터리의 구축 및 확산을 위해 노력해 왔고, 그 결과 디지털 리포지터리는 학술 정보 유통의 중요한 부분을 차지하게 되었다. 동시에 디지털 리포지터리는 디지털 콘텐츠의 장기적인 보존, 접근성 향상과 이용자 서비스의 확대에 따른 활발한 이용, 지속적인 서비스를 위한 안정된 재정 확보 등의 문제에 직면해왔다.

문제 해결을 위한 다양한 노력이 진행되고 있는 이 시점에, 본 연구는 지속가능한 디지털 리포지터리 협력 모델인 하티트러스트의 이익과 가능성을 검토하기 위해서 그 특징점을 살펴 보고, 설립배경 및 목적, 운영 및 주요 정책, 자료 축적 현황, 그리고 이용자 서비스 및 기능을 포함한 전반적인 운영 사례와 현재 진행중인 추가 사업 등을 검토하였다.

하티트러스트는 주로 대학 연구도서관 디지털 리포지터리들의 파트너십으로 거버넌스 구조에 기반한 협력 모델이며 파트너 기관이 일정한 비용을 부담하는 비즈니스 모델을 채택하고 있다. 지속가능한 디지털 리포지터리를 위해서는 축적된 디지털 콘텐츠의 장기적인 보존을 지원하는 기술적인 구조도 중요하지만 디지털 리포지터리 운영 기관의 독립된 의사결정과 지속 유지 의지를 반영할 수 있는 거버넌스 구조, 안정적인 재정 확보, 그리고 이용자들의 활발한 이용이 뒷받침되는 선 순환적인 모델이 매우 중요하다. 그리고

하티트러스트는 대규모화되고 통합된 서비스를 통해 이용자들에게 접근성을 향상시키고 디지털 콘텐츠의 가시성을 확대시킨다. 파트너 기관으로 참여할 경우, 저작권 관리 용이, 운영 비용의 효과 증대, 디지털 자원의 공동 관리 및 보존 등의 복잡하고 어려운 문제들을 해결할 수 있는 것으로 보인다. 또한 이미 대규모화된 디지털 컬렉션을 바탕으로 한 하티트러스트 연구 센터의 빅데이터 분석 및 부가적인 서비스 제공과 실물장서의 공공 보존서고 구축에 대한 해법도 잠재적인 가능성으로 기대된다.

본 논문은 지속가능한 디지털 리포지터리 협력 모델로서 하티트러스트에 대한 국내 학술계의 소개 글로, 개괄적인 내용과 특징점, 그리고 기대되는 가능성을 제시하였다. 기관별 객

관화된 경제 모델을 사용한 하티트러스트의 비용 효과 산출이나 연구자의 다양한 빅데이터 활용 방법 개발 등에 관한 더 많은 연구들을 통해 하티트러스트와 같은 협력 모델의 특징점과 가능성은 더욱 명확하고 구체화될 수 있을 것이다. 본 논문은 국내의 지속가능한 디지털 리포지터리 운영을 위해 새로운 협력 모델을 계획하거나, 국내 디지털 리포지터리 운영 기관이 파트너 기관으로서 하티트러스트에 참가 등을 고려할 때 도움이 될 것이라 본다.

## 참고문헌

- 정영미, 배정희. 2015. 기술확산 통합모델을 통한 개방형 기관 리포지터리 수용의 영향요인 분석. 『한국도서관·정보학회지』, 46(4): 529-549.
- Association of Research Libraries. 2009. *The Research Library's Role in Digital Repository Services*. <<http://www.arl.org/storage/documents/publications/repository-services-report-jan09.pdf>> [cited 2016. 9. 30].
- Christenson, Heather. 2010. "HathiTrust A Research Library at Web Scale." *Library Resources and Technical Services*, 55(2): 93-102. <<https://www.hathitrust.org/documents/christenson-lrts-201104.pdf>> [cited 2016. 8. 20].
- Courant, Paul and John Wilkin. 2010. "Building 'Above-Campus' Library Services." *Educause Review*. <<http://er.educause.edu/articles/2010/8/building-abovecampus-library-services>> [cited 2016. 8. 10].
- Erway, Ricky. 2012. *Lasting Impact: Sustainability of Disciplinary Repositories*. Dublin, Ohio: OCLC Research. <<http://www.oclc.org/research/publications/library/2012/2012-13.pdf>> [cited 2016. 10. 11].
- Furlough, Mike and Valerie Glenn. 2016. "Detecting US Federal Documents to Expand Access." *IFLA WLIC 2016*. <<http://library.ifla.org/1380/1/099-furlough-en.pdf>> [cited 2016. 7. 12].
- Gibbons, S. 2004. "Defining an Institutional Repository." *Library Technology Reports*, 40(4): 6-10.
- Google Book Home Page. <<https://books.google.com/>> [cited 2016. 10. 5].
- HathiTrust. 2016. "Access Determination for HathiTrust Objects." [https://www.hathitrust.org/access\\_determination](https://www.hathitrust.org/access_determination) [cited 2016. 11. 25].
- HathiTrust Digital Library Home Page. <<https://www.hathitrust.org/>> [cited 2016. 10. 25].
- Li, Yuan and Meghan Banach. 2011. "Institutional Repositories and Digital Preservation:

- Assessing Current Practices at Research Libraries.” *D-Lib Magazine*, 17(5/6).  
<http://www.dlib.org/dlib/may11/yuanli/05yuanli.html> [cited 2016. 10. 9].
- OpenDOAR Home Page. <<http://opendoar.org>> [cited 2016. 9. 22].
- Piper, Paul S. 2013. “The Library's Future is Digital: HathiTrust and the Digital Public Library of America.” *Online Searcher*, 37(2): 22–26.
- Research Libraries Group. 2002. *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA: RLG.
- Ross, Seamus and Margaret Hedstrom. 2005. “Preservation Research and Sustainable Digital Libraries.” *International Journal of Digital Libraries*, 5: 317–324.
- Scholars Portal Home Page. <<http://www.scholarsportal.info/>> [cited 2016. 11. 2].
- Wilkin, John. 2013. “Detailed Description of the Model.” <https://www.hathitrust.org/documents/hathitrust-cost-rationale-2013.pdf> [cited 2016. 11. 20].
- Yale University Library. 2011. “HathiTrust Working Group Green Paper.” <[https://www.hathitrust.org/documents/Yale\\_HathiTrust\\_GreenPaper.pdf](https://www.hathitrust.org/documents/Yale_HathiTrust_GreenPaper.pdf)> [cited 2016. 9. 12].
- York, Jeremy. 2010. “Building A Future By Preserving Our Past: The Preservation Infrastructure of HathiTrust Digital Library.” <<https://www.hathitrust.org/documents/hathitrust-ifla-201008.pdf>> [cited 2016. 9. 1].
- York, Jeremy. 2015. “The HathiTrust Digital Repository: Under the hood.” [presentation file] <<https://www.hathitrust.org/sites/www.hathitrust.org/files/HathiTrust-SI625-20150420.ppt>> [cited 2016. 11. 2].

국한문 참고문헌의 영문 표기

(English translation / Romanization of reference originally written in Korean)

- Jung, Youngmi and Junghee Bae. 2015. “Influencing Factors for the Acceptance of Open Access Institutional Repository Using the Integrated Technology Diffusion Model.” *Journal of Korean Library and Information Science Society*, 46(4): 529–549.