



2019年新信息环境下 CASHL 资源与服务拓展设计研究

基于 Google Book Search 与 HathiTrust 运行模式的 CASHL 转型发展策略研究

课题单位：兰州大学图书馆

课题负责人：宋戈

课题组成员：张继忠（兰州大学图书馆）、孙林（兰州大学图书馆）、魏清华（兰州大学图书馆）、王海花（兰州大学图书馆）

结项时间：二零二零年七月

摘要：Google Book Search 及 HathiTrust 项目宗旨均为与全球重要的科研机构及图书馆取得合作，共同致力于为人类文化资源的永久性保存及获取提供保障，是本世纪在世界范围内发展迅速、效益显著的两大电子文献仓储。CASHL 是国内第一个为人文社会科学教学和科研提供文献服务和最终保障的国家人文社会科学信息资源平台，是高校三大信息资源共享体系(CCC)之一。CASHL 与 Google Book Search、HathiTrust 在建设目标、资源发展、服务宗旨等诸多方面具有共通性、参照性。面对国家“一带一路”、“丝绸之路经济带”的重大战略、“双一流”建设规划要求，以及网络信息技术发展、用户需求的变化，对比分析 Google Book Search、HathiTrust 的运行模式，针对 CASHL 转型发展面临的扩大共建共享规模，加强全领域、特色性学科文献布局，多途径广泛开展数字化资源建设，开展个性化知识共享服务，加强技术与服务的深度融合，战略规划和管理体系调整，解决图书馆联盟间同质化竞争问题，版权管理，建设国家社科文献安全保障体系等问题，提出了：（1）建立兼具前瞻性、科学性与现实性的 CASHL 战略规划；（2）联盟化协同发展；（3）合作共享；（4）以全球特色人文社会科学文献资源建设为重心；（5）重视数字资源的长期保存和安全性指标；（6）强化资源、服务与技术的深度融合；（7）数据质量和数量是 CASHL 持续发展的关键；（8）科学高效的管理体系是 CASHL 发展的保障；（9）提升数字资源的服务功能；（10）助推版权策略等转型发展策略。

关键词：Google Book Search HathiTrust CASHL 运行模式 转型发展 策略

目录

0 引言	5
1 Google Book Search 及 HathiTrust 项目概况	5
1.1 Google Book Search 项目起源	5
1.2 HathiTrust 项目起源	6
1.3 Google Book Search 及 HathiTrust 目标使命及相互联系	6
2 CASHL 及其转型发展需求	7
2.1 CASHL 项目概况 ^[9]	7
2.2 CASHL 管理、资源建设与服务	7
2.2.1 CASHL 建设目标和基本原则	7
2.2.2 CASHL 管理体系	8
2.2.3 CASHL 资源体系	8
2.2.4 CASHL 服务体系	9
2.3 CASHL 建设成效与转型发展	9
2.3.1 建设成效	9
2.3.2 CASHL 转型发展需求	10
3 Google Book Search、HathiTrust 运行模式	10
3.1 运行管理机制	10
3.2 资源组织	11
3.2.1 资源来源	11
3.2.2 资源质量	11
3.3 版权管理	12
3.4 平台功能	13
3.4.1 运行系统	13
3.4.2 数据格式	13
3.4.3 数字化扫描	13
3.4.4 数据存储与安全性	14
3.5 服务内容	14

3.5.1 海量的数字资源.....	14
3.5.2 强大的检索功能.....	15
3.5.3 输出引文功能.....	16
3.5.4 个性化服务.....	16
3.5.5 其他个性化服务功能.....	17
3.5.6 隐私问题.....	17
3.6 成效影响.....	17
4 基于 Google Book Search 与 HathiTrust 运行模式的 CASHL 转型发展策略.....	18
4.1 建立兼具前瞻性、科学性与现实性的 CASHL 战略规划.....	18
4.2 联盟化协同发展——CASHL 发展重要机制.....	19
4.3 合作共享——CASHL 资源建设的必由之路.....	19
4.4 以全球特色人文社会科学文献资源建为重心.....	20
4.5 重视数字资源的长期保存和安全性指标.....	20
4.6 强化资源、服务与技术的深度融合.....	20
4.7 数据质量和数量是 CASHL 持续发展的关键.....	21
4.8 科学高效的管理体系是 CASHL 发展的保障.....	21
4.9 提升数字资源的服务功能.....	22
4.9.1 强化检索功能.....	22
4.9.2 深化个性化服务.....	22
4.10 助推版权策略.....	22
5 结语.....	23
参考文献:	23

基于 Google Book Search 与 HathiTrust 运行模式的 CASHL 转型发展策略研究

0 引言

Google Book Search 及 HathiTrust 项目宗旨均为与全球重要的科研机构及图书馆取得合作,共同致力于对人类文化资源的永久性保存及获取提供保障,是本世纪在世界范围内备受关注、发展迅速、效益显著的两大电子文献仓储,它们给各国图书馆事业的发展带来巨大的影响和深刻的启示。^[1]

CASHL 是国内第一个为人文社会科学教学和科研提供文献服务和最终保障的国家人文社会科学信息资源平台,是高校三大信息资源共享体系(CCC)之一。CASHL 与 Google Book Search、HathiTrust 在建设目标、资源发展、服务宗旨等诸多方面具有共通性、参照性。面对国家“一带一路”、“丝绸之路经济带”的重大战略、“双一流”建设规划要求,以及网络信息技术发展、用户需求的变化,当前 CASHL 面临转型发展的新时期,他山之石,可以攻玉,Google Book Search、HathiTrust 的成功运行及在世界范围内的影响力和地位对 CASHL 未来发展具有良好的借鉴意义。本报告主要依据文献调查法和网络调查法进行资料搜集,通过对 Google Book Search 与 HathiTrust 全方位的调查分析和研究,为 CASHL 的转型发展策略提供理论参考。

1 Google Book Search 及 HathiTrust 项目概况

1.1 Google Book Search 项目起源

Google Book Search,即 Google 图书搜索计划,其宏伟目标为扫描所有图书、组织全球信息,并使之在全球范围内可获取和利用。正式起源于 2004 年,包括两个子项目:出版商计划(Partner Program),和图书馆计划(Library Project)^[2]。

谷歌做为一个发展历史较长商业性的机构,2003 年 12 月谷歌推出对书摘搜索服务的 Google Print(谷歌打印)试用版。2004 年 10 月谷歌与美国的兰登书屋合作,正式启动搜索图书全文服务的谷歌图书搜索计划。它扫描了书的封面页、目录页、版权页及部分内容页建立自有的仓储,并通过网络为全世界公众提供在任何地方都能搜索的便捷的在线全文检索服务。谷歌还先后推出了与出版商合作的“Google Print Publisher(谷歌打印出版商)”子项目和与图书馆合作的“Google Print Library(谷歌打印图书馆)”子项目^[3],为丰富资源的来源找到了解决途径。目前谷歌已与至少 28 家重要的图书馆、100 多个国家或地区的 10000 多个出版商和作者建立了合作关系。谷歌对其合作伙伴所拥有的数以百万计的文献资源进行扫描及数字化并加入到谷歌的数据库中,而合作伙伴也可以获得该资源的一个数字拷贝。2010 年,谷歌还宣布启动了数字书店,称作 Google Editions(谷歌版本),所有的内容均被放在云端。谷歌扫描了超过 120 万本的书。

1.2 HathiTrust 项目起源

2008年,Hathitrust由美国机构合作委员会(Committee on Institutional Cooperation, CIC)的12所大学联盟及加利福尼亚大学(University of California, UC)系统所属的11所大学图书馆发起,使命为通过合作收集、组织、保存、沟通和分享人类知识记录,为研究、学术和共同利益作出贡献^[4]。同时也是为应对Google Book Search所引起的“一场垄断的噩梦”,以及学术研究的未来不能掌握在商业机构手中^[5]。这些研究图书馆合作包括数百万图书在内的大量数字馆藏创建机构库,将承担大学图书馆数百年的信息保存和获取。Hathitrust是一个非盈利机构,致力于支持数字化各图书馆的资源,并创建一个庞大的资源库。后来HathiTrust成员逐渐拓展到包括美国国会图书馆、哈佛大学、杜克大学、十大学术联盟等在内的近130个。

1.3 Google Book Search 及 HathiTrust 目标使命及相互联系

Google Book Search项目的执行者Google公司是一家商业性机构,是以赢利为目的的。Google Book Search项目所拥有的千万册数字化的图书能为Google公司吸引来更多的用户,用户数量的增加还能带来谷歌广告的升值,从而促使各方利益的最大化。Google公司将此项目的使命确定为“组织全球信息,并使之在全球范围内可获取和可利用”,“将继续发展其技术并扩大与全世界出版商和图书馆的合作关系.....并从数字图书馆向整个图书馆业、互联网信息服务业乃至世界文化的传播方面扩展。”^[6]谷歌开始从一个单纯检索服务提供商向知识提供商转型,逐步向数字化知识资源进军。

Hathitrust项目执行者CIC和UC是非营利性的组织机构,其创建数字资源仓储的目标不是为了寻求盈利,其项目宗旨为:(1)建立可靠的、日益全面的成员机构纸质资源的数字化档案,供成员机构共享;(2)首要改进数字化资源的获取途径,以满足成员机构尤其是阅读障碍者的需求;(3)致力于开发低成本、高质量的基础设施,保障学术研究人员数字资源需要,包括各种格式的资源和本地数字资源;(4)发展合作伙伴关系和服务,以确保HathiTrust纸质资源和数字学术资源的长期保存;(5)通过加倍努力协调图书馆间的共享存储策略,以减少长期保存和纸质资源存储的成本;(6)建设基础设施,提高成员机构成本效益和合作效率,降低获取校园知识资产的成本;(7)提供一系列服务支持使用HathiTrust料库进行研究;(8)创建一个技术框架,支持各类工具和服务的集中或分布式创建;(9)维持HathiTrust的“公益性”,同时提供一系列有益于成员机构的服务^[7]。项目经费来源于参与Hathitrust的各个资金雄厚的高校。项目初建之时,会要求各成员机构根据各馆要被数字化的文献资源的数量交纳一笔初始经费,以后每年再缴纳保存数字资源所需的年费。

Hathitrust项目与Google Book Search有很深的联系,CIC和UC组织内的图书馆都曾经通过加入Google Book Search项目数字化了它们的部分馆藏资源。Hathitrust项目成立后,这些图书馆也在继续和谷歌公司合作,共同数字化它们的馆藏资源,目前Hathitrust数据库中有超过200万册的图书是由Google数字化的。这些图书资源和Hathitrust扫描

的其他资源一起都可以在 Hathitrust 里被检索获取到。两者在信息技术、数字图书馆、项目管理等领域都拥有明显优势及成熟经验^[8]。

2 CASHL 及其转型发展需求

2.1 CASHL 项目概况^[9]

2002 年,为繁荣发展哲学社会科学,教育部设立了“中国高校人文社会科学文献中心(China Academic Social Sciences and Humanities Library,简称 CASHL)”CASHL 是在教育部领导下,为我国哲学社会科学教学科研提供外文文献及相关信息服务的最终保障平台,其建设目标是“国家人文社会科学文献信息资源平台”。2004 年 3 月 15 日,CASHL 项目作为教育部“繁荣发展哲学社会科学计划”的一部分正式开始服务。负责项目组织协调的 CASHL 管理中心设在北京大学,复旦大学图书馆参与管理。

2006 年,在教育部的指导下,“文专项目”图书采购相关工作统一由 CASHL 管理中心组织协调。自此,全国高校图书馆在资源建设上形成了一个庞大的外文人文社科文献联合保障体系,由北京大学、复旦大学、武汉大学、吉林大学、中山大学、南京大学、四川大学、北京师范大学、东北师范大学、华东师范大学、兰州大学、南开大学、山东大学、清华大学、厦门大学、浙江大学、中国人民大学等 17 所综合性高校图书馆,70 余所教育部文专院校及 851 个成员馆组成。

2.2 CASHL 管理、资源建设与服务

2.2.1 CASHL 建设目标和基本原则

CASHL 以“国家人文社会科学信息资源平台”为建设目标,通过组织国内具有学科、资源和服务优势的高等学校图书馆,有计划、有系统地整体引进国外人文社会科学图书、期刊和电子资源,借助现代化的服务手段,整合国内高校人文社科领域研究成果,服务国家发展战略,为人文社会科学教学科研、人才培养提供全面和最终的文献信息资源保障。同时,CASHL 努力打造中国人文社会科学走向世界的成果平台,加强中文特色文献资源建设,积极参与中国特色哲学社会科学话语体系的建设。

CASHL 在资源建设上本着整体建设、统筹安排、相对集中、讲求效益的原则,通过建设全国中心、区域中心和学科中心三级体系,实现高效图书馆馆际文献收藏与服务的分工合作,建设我国人文社会科学领域最高水平、最全面和最可持续的人文文献资源中心。

CASHL 在服务建设上本着“共建、共知、共享”的原则,由 CASHL 中心馆和高校文科图书引进专款项目院校以集中式平台和分布式服务相结合的方式共同合作,建设集中式的 CASHL 网络服务体系,揭示报道 CASHL 收藏的印本期刊、电子资源、文专图书、大型特藏及其他非 CASHL 馆藏的人文社科学术资源,面向全国高校、哲学社会科学研究机构和工作提供综合性文献信息服务。

2.2.2 CASHL 管理体系

CASHL 设立了由“CASHL 项目指导委员会”、“CASHL 管理中心”、“CASHL 中心馆馆长联席会”、“CASHL 专家咨询组”、“CASHL 全国中心”、“CASHL 区域中心”、“CASHL 学科中心”、CASHL 各业务工作组组成的管理职能清晰、层级分明的管理体系。

其中，“CASHL 项目指导委员会”负责指导 CASHL 开展各项，审核 CASHL 资源和服务发展规划和实施方案、制定 CASHL 项目管理办法及经费使用办法等。成员包括北京大学和复旦大学主管校长、CASHL 管理中心主任和副主任等。“CASHL 管理中心”负责落实成立专家咨询组，召开中心馆馆长联席会议及举办工作会议，管理与协调 CASHL 各中心，推动项目的整体发展。“CASHL 中心馆馆长联席会”由 CASHL 各中心馆馆长或主管馆长组成，参与 CASHL 的管理和协调工作。“CASHL 专家咨询组”主要职责加强对 CASHL 整体规划的协调和指导，使 CASHL 的发展规划、工作评估、资源发展、服务发展等更加科学合理。

在管理主导层级之下，“CASHL 全国中心”和“CASHL 区域中心”、“CASHL 学科中心”则分别负责全国、区域、学科的资源整体规划、建设和服务，收藏印本期刊、电子资源，提供本校外文人文社科期刊馆藏数据，配合 CASHL 管理中心开展数据加工工作，维护本地文献传递服务系统，面向全国、区域开展培训、宣传、文献检索和原文提供等服务并进行相关评估。

“CASHL 全国中心”设在北京大学、复旦大学，“CASHL 区域中心”设在武汉大学、吉林大学、中山大学、南京大学、四川大学、北京师范大学、兰州大学等 7 所大学，“CASHL 学科中心”设在东北师范大学、华东师范大学、南开大学、山东大学、清华大学、厦门大学、浙江大学、中国人民大学等 8 所大学。此外，还有 6 个 CASHL 各业务工作组，包括文专图书馆藏发展小组、电子资源工作组、期刊工作组、CASHL/CALIS 联合成立的高校馆际互借协调组、服务工作组、新媒体工作组等，其职责是深入开展相关业务工作、提供专业咨询服务、进行创新性业务研究等。

CASHL 还与 CALIS 管理中心、中国社科院图书馆、上海图书馆、北京高校网络图书馆、澳门科技大学等相关行业机构，在技术支持、馆际互借与文献传递、高校人文社科外文期刊目次数据库和高校人文社科外文图书联合目录建设等方面开展业务合作。与中国教育图书进出口有限公司在 CASHL 资源采购和资源发现和服务方面开展合作。

2.2.3 CASHL 资源体系

近二十年来，CASHL 资源建设经历了品种从单一到丰富、学科从少数到全面、载体从纸本到电子、采购从分散到集中、资源类型从单一到丰富（重要期刊、图书、开放获取资源、数据库等）的过程。CASHL 资源主要包括：（1）完整的人文社科类外文资源体系^[10]。现有：11100 多种国外人文社会科学领域的核心期刊和重要期刊（包括 JSTOR、PAO 等国外著名人文社科数据库在内的电子期刊逾 1370 种，过刊的回溯年代早至 1665 年）；国外人文社科领域的重要印本图书 44 万种，以文科专款图书为主（包括 ECCO、EEBO 等国外著名人文社科在线全文数据库在内的电子图书 27 万种，以 15—18 世纪图书为主）；（2）整合全国 17 家重点高校及中国社会科学院的馆藏人文社科资源（包括北京大学、复旦大学、武汉大学、吉林大

学、中山大学、南京大学、四川大学、北京师范大学、东北师范大学、华东师范大学、兰州大学、南开大学、山东大学、清华大学、厦门大学、浙江大学、中国人民大学、中国社会科学院); (3) 覆盖所有人文社会科学一级学科(法律、教育、文学、经济/商业/管理、语言/文字、哲学/宗教、政治、军事、历史、人物/传记、社会科学、社会学、体育、统计学、地理、图书馆学/信息科学、区域学、文化、心理学、艺术)。

2.2.4 CASHL 服务体系

CASHL 构建了由 2 个全国中心、7 个区域中心、8 个学科中心、34 个服务馆、851 个成员馆组成了分工明确、协同高效的遍布全国的网格化人文社科文献服务体系,为系统收藏资源、有效开展服务提供保障。同时,科研院所系统的中国社会科学院图书馆、公共图书馆系统的上海图书馆也参与其中,统一以“中国高校人文社会科学文献中心”(CASHL)的名义对外开展服务。

在 CASHL 各相关高校共同努力及与高等教育文献保障系统(CALIS)、中国社科院图书馆、上海图书馆成为战略合作下,CASHL 外文可服务文献量得到进一步扩大。迄今为止,可供服务的人文社科核心期刊和重要期刊达到 2.46 万余种、印本图书达 200 余万种、电子资源数据库达 12 种,累计提供文献服务近 2200 万件,其中手工文献服务已突破 130 万,文献平均满足率达 96.29%左右,服务时间缩短为 1.87 天,大大提高了外文图书的利用率,充分发挥其效益。除此之外,CASHL 还提供“高校人文社科外文期刊目次库”和“高校人文社科外文图书联合目录”等数据库,提供数据库检索和浏览、书刊馆际互借与原文传递、相关咨询服务等。CASHL 服务辐射面也进一步拓展,CASHL 目前已拥有 878 家成员单位,个人注册用户愈 13.6 万多个,CASHL 服务惠及上千万用户。

2.3 CASHL 建设成效与转型发展

2.3.1 建设成效

CASHL 项目在不同历史时期,承担了不同历史任务,从早期的硕博基地建设,到繁荣发展哲学社会科学使命,对国家的经济、社会、文化、人才培养等的发展都起到了文献信息的关键保障作用,对国家哲学社会科学的繁荣发展和中国学术的国际话语权建设具有重要现实意义。CASHL 项目为中国高等教育的人才培养、教育教学、学术研究等方面作出了重大贡献。

(1) 建立了一套覆盖全国的、完整的“共建共享”机制。(2) 站在国家百年大计的高度,建立了我国最大、最全面的人文社科文献保障体系。(3) 为国家多个归国人才引进计划提供了强有力的基础支撑。CASHL 项目所引进外文文献起到了基础智力支撑的作用。(4) 保障了全国人文社科教学科研人员获取资源的公平性和公正性。CASHL 已经成为我国教学科研人员高度依赖、不可或缺的资源获取渠道,已经成为全国高校人文社科领域具有影响力的品牌服务。(5) 中国哲学社会科学国际话语权进一步扩大。(6) CASHL 项目的存在,对培养具有国际视野人才起了巨大推动作用。

2.3.2 CASHL 转型发展需求

随着通讯网络技术和国际国内图书情报事业的迅猛发展,CASHL 在取得可喜成绩的同时,也不可避免地遇到一系列与可持续发展密切相关的问题:(1)如何在同行业、跨行业间扩大共建共享规模;(2)加强全领域、特色性学科文献布局;(3)多途径广泛开展数字化资源建设;(4)开展个性化知识共享服务;(5)加强技术与服务的深度融合;(6)按事业的发展需要调整战略规划和管理体系;(7)解决图书馆联盟间同质化竞争问题;(8)协调东西部地区图书馆联盟发展不均衡;(9)版权管理;(10)建设国家社科文献安全保障体系,等等。在国家“十四五”规划到来之际,CASHL 转型发展成为必然,在近 20 年间与 CASHL 并行成长发展的 Google Book Search 及 HathiTrust 项目在建立大型数字仓储并在世界范围内提供便捷服务所取得的卓越成绩和深刻影响力不失为我们学习借鉴的成功案例。

3 Google Book Search、HathiTrust 运行模式

3.1 运行管理机制

(1) 资金来源。Google Book Search 资金由 Google 公司全权负责;HathiTrust 各方面建设资金由各成员馆分摊,一些具体项目也得到基金会支持。

(2) 目标设置。Google Book Search 公布了总目标并相继推出两个子项目,但缺乏明晰的阶段性目标;HathiTrust 官网上发布了非常清晰的终极目标、长期目标、短期目标及每月数据更新计划。

(3) 性质规模。作为商业机构的 Google 公司利用雄厚资金和先进技术将 Google Book Search 项目建设得规模庞大;HathiTrust 作为高校图书馆数字联盟更注重和具有服务科研和教学的特色。

(4) 组织管理。Google Book Search 为常规的企业管理模式,服从 Google 公司整体战略安排^[11]。HathiTrust 管理采用多个委员会和工作组,设置了科学规范的整体管理与职能框架——常设职能机构+项目工作组,其中主要是执行董事委员会和项目指导委员会。执行董事委员会主要由各成员馆馆长及首席信息官担任委员;项目指导委员会下设收藏委员会,联邦文件咨询委员会,共享打印咨询委员会,元数据的政策、策略、使用和共享咨询组;工作组包括:质量保证和标准工作组,权利和访问工作组;根据项目管理形式,下设多个工作小组,项目任务完成后小组随即解散;建立了柔性的人力资源管理,以及弹性工作模式。HathiTrust 还设立了一个由各高校教职工、学生以及其他用户等组成的代表团,向委员会提交建议。

3.2 资源组织

3.2.1 资源来源

Google Book Search 的图书来源有二：一是出版商计划，通过与全球出版商及作者等版权所有者的洽谈合作获得已出版或计划出版图书授权，同时允许他们自行通过 Google 开发的平台登录，上传其版权所有图书扫描图像或电子文档；二是图书馆计划，通过与全球著名大学图书馆及一些公共图书馆、科研图书馆签署协议获准对其馆藏图书数字化。Google 利用其专利性光学识别扫描技术及强大物流，把图书、期刊运输到图书扫描中心，每台仪器每小时数字化 1000 页左右的图书，截至 2015 年 10 月，已扫描图书 3000 万册^[12]，来自 100 多个国家、地区的 10000 多个出版商和作者及 40 多所图书馆参与其中。

HathiTrust 数字资源来源多样，除了与 Google 和微软的合作项目享有同等数字化资源备份，还与数字图书馆互联网档案馆（Internet Archive）以及联盟内部合作。数字化资源类型包括传统馆藏数字化、自建特色数字资源、商业数据库。它与全球重要的科研机构及图书馆取得合作，共同致力于为人类文化资源的永久性保存及获取提供保障，资源增长稳定。

根据 HathiTrust 网站介绍(<https://www.hathitrust.org/home>)可知，HathiTrust 所收藏的很多作品在 Google Book Search 里没有呈现，这是因为二者资源不同的获取渠道和不同的版权决定过程导致的。HathiTrust 独有最大的文献类型包括美国联邦政府文件和 1923 年后在美国发表的公共领域的作品，还有如各个成员图书馆的珍贵馆藏的数字化资源、大学的学术论文及研究成果数据库以及自建的特色资源数据库等。HathiTrust 大约拥有每一个北美研究型图书馆收藏印刷品总量 50% 的电子版本。

3.2.2 资源质量

根据大型数字图书馆和数字化创新的质量要求，Google Book Search 曾因数字化的质量和元数据的质量问题而遭受用户批评。关注点主要在一些页面扫描的质量较差、错误很多、难以识别的 OCR (Optical Character Recognition, 光学字符识别) 上。因为 Google Book Search 是依靠 OCR 技术和电脑处理来提取记录的元数据，是先数字化再考虑质量的，虽然谷歌也在提高数据质量上花费了大量的精力。

HathiTrust 2011 年 3 月得到 TRAC (Trustworthy Repositories Audit and Certification, 值得信任的仓储审计和鉴定) 的认证，这一认证使得数字仓储鉴定得以保证。HathiTrust 在网站上发布了关于质量保证的声明，承诺为确保内容质量的优化，在它的仓储中对所有文献内容实施全面的质量审核。HathiTrust 为了应对大规模数字创新所带来的元数据挑战，在给它的收藏提供元数据时承受着巨大的压力，它依托图书馆合作，不断提高现存的印刷目录质量并且为数字世界优化书目元数据。比如 HathiTrust 网站上可以获得的索书号、日期和语言的一个数据可视化标识^[13]。

3.3 版权管理

Google Book Search 采用其独创的“选择-退出 (Opt-out)”模式和“默示许可”模式,即“先使用,后授权”。由版权所有者列出拒绝其作品进入 Google Book Search 的作品清单,Google 依照要求进行规避^[14]。“默示许可”模式主要是指版权人虽未进行明确授权,但他人可从版权人的行为或沉默中推定其获得了授权许可^[15],版权所有者可在使用者利用其作品过程中,主动向使用者提出付酬请求,使用者必须按照适当标准支付,在使用者恶意不支付的情况下,版权所有者可禁止使用者使用作品并要求经济赔偿^[16]。Google Book Search 是在未经版权人同意情况下将上千万册图书扫描入库,已扫描的图书中 80% 是处于版权保护期内的,对于版权限制图书,Google 图书利用片段预览 (20% 内容) 形式进行限制获取,对于那些已过版权保护期的图书,用户能够浏览全文^[17]。谷歌收益的产生通过以下几种方式:“和扫描书籍的预览页面同时出现的商业广告的发行,图书馆和其他机构对于谷歌扫描的所有书籍制作的数据库的订阅,对于受版权保护的书籍的阅读权的销售。谷歌将得到这些收益的 37%,剩下的 63% 给作家和出版商。”^[18]

HathiTrust 采用主动且严格的版权管理模式。为了保存和跟踪每个数字化文档的版权信息,Hathitrust 与版权所有者签订版权许可协议,建立了单独的版权数据库,该数据库与书目数据库相连,随着书目数据的变化而自动更新相关内容。目前可以免费获取全文的资源包括已进入公共领域的出版物(主要是 1923 年前在美国境内和 1870 年前在美国境外出版的作品)、不受版权保护的出版物(如政府出版物)以及已获得版权许可协议的出版物。^[19]基于高校和联盟属性,HathiTrust 利用校园网及 VPN 限制访问和获取。只有在版权所有者许可的情况下,该图书资源才会对外开放。对于仍然受版权保护而没有取得许可开放的文献,只提供有限制的获取。从而避免了版权纠纷,维护了版权所有者的利益。Hathitrust 的绝大部分文献,只在校园网内可以提供全文,在公众网上只能检索到这些文献的一些章节片段,或者仅仅只有书目信息。而各成员高校的学者,可以采取用户验证的方式,通过专门的渠道,在校外使用这些数字资源。根据 Hathitrust 的版权状况,它的内容大约 68% 在版权保护内,32% 在公共领域。在 32% 的公共领域内,21% 是世界范围内的公共领域。其中 4% 是联邦政府文献,11% 是在美国公共领域。大约 12000 卷或 0.1% 的内容被许可为开放存取,包括有创新的被公众许可的内容。

Google Book Search 和 HathiTrust 受到极大关注和争议的著作权侵权案分别于 2016 年 4 月和 2014 年 6 月被驳回上诉,终判定为“合理使用”^[20]。

Hathitrust 也面临法律上的挑战,在 2011 年 9 月,作者协会对 Hathitrust 和密歇根大学等各大学提出了一个联邦版权侵权控诉,控诉它们存储了数以百万计的书的数字副本。2012 年 10 月,判决宣判了支持图书馆。Hathitrust 因此在它的网站上关于宣判发表了声明。

3.4 平台功能

3.4.1 运行系统

Hathitrust 和 Google Book Search 的运行系统都采用基于 OAIS (Open Archival Information System, 开放档案信息系统) 的框架结构, 包括信息摄取、存储、数据管理、获取等功能模块, 使用元数据编目及传输规范 (Metadata Encoding and Transmission Standard, METS) 和保存元数据实施战略规范 (Preservation Metadata Implementation Strategies, PREMIS)。”^[21] 很多其他数字图书馆项目也使用这样的数据规范, 因此有利于与其他数字化项目进行共享数据。

Hathitrust 基本上采用了 SaaS 云服务模式, 以数字资源知识库为中心, 围绕数据包构建了一个云框架, 方便成员机构通过互联网托管、部署或接入, 同时开放性地支持成员机构贡献应用程序、分析工具等, 在保证成员机构海量数字资源安全存储的同时, 可以按需为成员机构和用户提供个性化的服务。如密歇根大学开发了一个新的全文检索模型; 加州大学负责全文检索的拼写检查, 克服希伯来文无法识别的问题等^[22]。

Hathitrust 以广泛收集、整合、长期保存和获取数字学术资源为主要目标, 因此, 在搭建云平台时, 综合考虑了数字资源规模、长期保存和获取、开放性三个因素, 并依据 OAIS 框架, 进行了知识库建设。其将知识库分为四个工作模块^[23]: (1) 数字资源摄取: 检验存入知识库的数字资源对象的固定性, 如果需要将它们转换为 HathiTrust 规定的长期保存格式, 执行严格的验证、打包, 存入知识库。(2) 数字资源存储: 存储在地理上分开的三个数据中心。(3) 数据管理: 存储关于数字资源对象的元数据, 以及与 HathiTrust 保存的数字资源相对应的成员机构的纸质馆藏信息。(4) 访问获取。包括数字资源的题名和全文检索; 公共领域和开放存取数字资源的在线阅读和下载; 书目数据和存储内容的各种 API 等。同时, Hathitrust 注意到: 不同的数字资源对象 (数字文本、图像、视频、音频、多媒体、模拟等) 在保存过程中是有明显差异的, 为了保证这些数字资源对象在长期保存过程中的一致性和完整性, Hathitrust 还定义和保存了区分不同数字资源对象的特征信息, 具体包括数字资源对象的内容、固定性、标识参考、信息来源和前后关联^[24]。

3.4.2 数据格式

HathiTrust 的书目数据采用自行开发的元数据格式 (HathiTrust Metadata), 而没有使用 GBS 等通常的 MARC 格式, 以更好地在各成员馆及 OCLC 之间传输转换数据。

3.4.3 数字化扫描

印刷型文献的数字化扫描是电子书仓储最主要也是最艰巨的工作任务。为了保护珍本、善本乃至孤本图书, “Google Book Search 和 Hathitrust 扫描计划采取了区别于传统扫描方法的三维红外立体照相技术, 可以探测放在扫描仪里的书籍页面的三维形状及角度, 然后

将探测信息转给光学字符识别软件，进行变形失真调整，以便光学字符识别软件能更精确地识别文字。”这样的扫描方式有很多的优点，既不会损坏书籍的装订，效率也非常的高。如 Google，“他们的扫描速度大约为每分钟 2.25 本书，仅在斯坦福大学图书馆中，每天就扫描高达 5 万页，而且将来会进一步提高速度。”^[25] HathiTrust 的数字资源每天都在高速更新，在这个项目成立后的连续两个月内，每月都有超过 30 万册文献被添加到数字仓库中^[26]。

3.4.4 数据存储与安全性

HathiTrust 数据的存储包括书目和全文数据两部分，合作伙伴通过依靠联盟标准和实践经验制定管理内容、服务政策与程序，以及维护模块化开放的基础架构，确保数字图书馆的可靠性和效率。

HathiTrust 旨在为数字资源提供持久性和高可用性存储。为了达成这一目标，专为容错和长期资源保留设计了存储架构。存储架构的核心是同步在密歇根州安娜堡和印第安纳波利期两个数据中心进行数据存储，并在第三个数据中心以加密磁带的方式备份了数字资源 6 个月前的数据版本。所有数据中心均符合国际权威机房认证机构 Uptime Institute 建立的机房建置等级 Tier II 要求，只有授权的 IT 人员才能进入。每 3 个月工作人员会对所有存档数字资源进行完整性检查，在存储内部执行数据完整性检查以及所有静态数据的定期完整性检查，并使用奇偶校验和冗余检验来修复遇到的任何错误，以确保数字资源与存档时相同，从而确保资源存储的可靠性。在存储系统外部，HathiTrust 完全遵循 OAIS 框架要求和可信度标准摄取数据，并定期验证数据，以确保数据被正确摄取并保持不变。同时，HathiTrust 还将存储设备的使用寿命设定为 3~4 年，每年对已达到寿命的设备进行更换^[27]。HathiTrust 保存了完整齐备的日志文件，并以唯一的 ID 进行标识，以便于后期进行事件追溯或其他用途。

3.5 服务内容

3.5.1 海量的数字资源

(1) 数量。截至 2015 年 10 月，Google Book Search 已扫描图书 3000 万册^[28]，来自 100 多个国家、地区的 10000 多个出版商和作者及 40 多所图书馆参与其中。

截至 2020 年 5 月，HathiTrust 数字化资源共 15125873 卷，图书 7482362 册，其中可直接被用户使用的资源为 38%，62%处于保存收藏状态^[29]。美国联邦政府文件收藏 1358802 卷，其中可完整阅读的有 1218263 卷，不完整公开的有 140539 卷^[30]。

(2) 质量。HathiTrust 更受好评。Google 是先数字化再考虑质量，被批评数字化质量和元数据质量有问题，关注点主要在一些页面扫描质量较差、错误很多、不值得信赖的光学字符识别技术。而 HathiTrust 在 2013 年 1 月获得 TRAC(Trustworthy Repositories Audit and Certification, 值得信任的仓储审计和鉴定) 认证，其官网上关于资源质量的声明称，对所有内容进行全面质量审核，确保优质的资源内容。(3) 文献形式的多样性。HathiTrust

优于 Google Book Search。前者拥有常见的数字化资源类型，如图书、杂志、音频、会议文件等，后者主要定位为图书和杂志。(4) 内容更新。由于扫描技术先进，两者更新速度迅速且稳定。

3.5.2 强大的检索功能

两者都基于高品质资源开展优质化服务。Google Book Search 依托 Google 搜索引擎强大高效的检索功能，将图书索引与网页搜索索引进行整合，实现一站式高效检索。在检索策略上，Google 图书检索主页保持其搜索引擎简洁搜索框一贯风格，对用户检索技能没有过高要求，支持标题、作者、ISBN、关键字或任意字段检索，2007 年已提供 35 种以上语言支持。其高级检索除常见的语言、名称、作者、出版商、出版日期、ISBN、ISSN 限制，搜索结果还包含全部字词、包含完整、包含至少一个字词、不包含字词字句的限定，以及部分预览、全文预览，电子书可按图书或期刊区分类型。

HathiTrust 最初需通过各成员馆检索平台逐一检索。2011 年与 OCLC 合作开发 WorldCat Local 用户界面，支持多语言、多途径跨库检索，实现了统一检索平台上的跨库检索，用户通过统一检索平台上即可获得联盟馆馆藏服务^[31]；该平台可以一次返回多个数据库的检索结果，并显示在哪一个馆中存储，避免用户逐个登录。读者不仅可以访问本馆的资源，还能获得合作馆的馆藏服务。但是只有 HathiTrust 成员馆的读者可以获得全文下载，其他注册用户只能浏览目录。这个检索平台实现了资源一站式检索、导航和全文获取^[32]。还与 Serial Solutions 合作，通过 Summon 全网发现服务系统实现 HathiTrust 全文检索，并可通过 HathiTrust 直接访问公共领域资源。HathiTrust 的形式限制还包括音像声像和地图资料，还有额外的检索领域“丛书标题”^[33]。检索主页相对复杂，检索框右侧上方有全文和目录选项。目录选项的下拉菜单中包括所有领域、题名、作者、主题、ISBN/ISSN、出版商、系列标题选项，其高级目录检索要求用户运用布尔逻辑进行以上选项限定，同时语言和输出内容类型以列表列出供选择。相较于 Google 而言，HathiTrust 在输出内容类型上极其丰富，包括图书、杂志、音频、传记、会议、地图、数据文件等，这是由两者数字资源建设内容决定的。

在图书检索输出结果上两者相似，均为三种：书名没有找到、图书可以全文获得、图书部分预览或获取。部分预览或获取说明该图书已数字化，但因版权限制不能全文获取。而 Google Book Search 部分预览中有的图书是显示片段文字、书封等，有的只显示基本的有限书目信息。对于一些图书获取不到原文的，Google Book Search 的片段显示是有用的，它在部分被数字化的页面上显示了检索词，以及这个检索词在这本图书中首个片段图像之上出现了多少次。Google Book Search 比 HathiTrust 的有限获取更有用，因为有限显示仅仅显示给使用者检索词出现在特定页面的次数，而 Google Book Search 显示这个检索词定位在这本图书的准确位置^[34]。不能在线获取全文的图书资源，两者均可链接到 OCLC 浏览目录。其中，Google 还拥有自身的电子书商店，单击“购买此书”或“获取此书的印刷版”链接，可显示购买或借阅的相关地址。

3.5.3 输出引文功能

引文输出对于所有数据库来说是一个重要的特征，谷歌图书和 Hathitrust 都提供了这个服务^[35]，不过都很有限，用户会被提醒信息输出是不完整的。谷歌图书可以输出到 Bibtex、Endnote 和 Refman 这些书目引文管理软件。而 Hathitrust 有一个“cite this”的功能，可以返回针对文章是 MLA 或者 APA 格式的格式化的引文，并且仅仅输出到 Endnote(提醒连续出版物的引文可能是不完整的)。如果 Hathitrust 可以给经常引用芝加哥形式的政府出版物的历史学家提供芝加哥形式引文的选择，功能就更全面了。谷歌图书和 Hathitrust 的使用者被限制每次输出一个记录。因为输出引文选项仅仅出现在单个标题或目录记录里。当然两者都正确地 and 一致性地提供了标题、出版者、URL 数据在输出引文的随机性的样本中^[36]。

3.5.4 个性化服务

Google Book Search 个性化服务包括：(1) 用户可利用 Google 账号建立“我的书架”，把搜索结果添加到我的收藏、正在阅读。(2) 纸本获取。如果想获取感兴趣的书的纸本，可点击“获取此书的印刷版”，除点击在线书商链接购买，还可“在图书馆中查找此图书”，访问 OCLC WorldCat 输入自己的位置，找到收藏该图书的最近的图书馆。(3) 提供全部用户书评及评分。评分采用五星制，后台统计每星级人数，并以条形图显示结果，把同星级的评论整合为集合。集合按星级排列并显示评论数量，直观反映出用户评价。(4) 提供相关搜索。“在此书中搜索”功能可输入任何片段，如内容有匹配即返回有背景颜色标识的所有片段内容、页码。对于只显示片段预览的图书，有该书中常见术语和短语形成的词云，该书热门引用章节、引用次数，检索内容在预览片段中提到的位置，使用户虽不能获取全文，但可进一步发现该图书是否符合需求。(5) 每种书可生成专门链接网址供分享。(6) 强化细节服务。图书、计划阅读的图书、已读图书。记录搜索历史和书评，并相应推荐可能喜欢的图书，所有信息可设置公开或私密。

HathiTrust 的个性化服务包括：(1) 发挥“集合”功能。为用户提供个人收藏库，允许用户把记录添加到一个新的或现有的永久性收藏，如果用户没有登录会自动创建临时集合。集合内还能对其他数据库资料进行搜索，方便资料备查，并可下载元数据集合。(2) 建立知识社区。永久性收藏可以是私人用户或公开与他人共享，用户通过添加标签、完善资源主题信息，与他人分享所收藏内容、个人观点、专业知识等，从而建立相似用户圈，聚合成知识社区。(3) 创建定制数据库。如果用户对非 Google 或 Google 数字化资源不满意，列出所需卷册标识，HathiTrust 能帮助用户创建定制数据库。(4) 开展深层次学术服务。HathiTrust 研究中心 (HTRC) 还开发了系列主题馆藏和应用工具。主题馆藏以国会图书馆分类为基础，基于公共领域材料收藏提炼主题。工具方面，SEASR 可以为前沿的数字人文活动提供动力，并通过授权学者在虚拟工作环境中共享数据和研究来促进合作；Author Search 在检索书目信息时提供了多种文本分析和可视化的工具，通过调用 SEASR 服务进行数据分析^[37]。其研究中心还与先进的协同支持系统 (ACS) 项目合作，助力学术服务，为外部学者和中心工作人员之间解决具有挑战性的计算分析问题提供合作，方便大家获得 HathiTrust 基于个人学

术的服务。

3.5.5 其他个性化服务功能

Google Book Search 对于仅仅显示 snippet 的图书提供了一些额外的功能。第一个是“普通词组”特征，可以在图书的文本中创建一个词云，这提供了一个有用的发现工具。其次是谷歌图书还在 snippet 显示中提供地图特征，被称作“在书中提到的位置”。这个提到的位置上还显示了谷歌地图，和一个能启动 Google earth(谷歌地球)的链接。Google Book Search 有一个 QR 号在 snippet view 里，能提供简单的 URL 到图书记录里。Google Book Search 还有个功能叫“添加到我的图书馆”，在那引文可以被添加进去产生列表。添加到我的图书馆中的标题列表，默认是公开可看的，但是可以被设置为隐私。

Hathitrust 有一个共享特征，允许用户共享图书永久的链接，或者出版物中特定的页面。它也有“添加到收藏”的功能，为用户提供个人收藏夹，满足用户专题需要，允许终端用户把记录添加到原来创建的收藏里或创建一个新的收藏。个人收藏夹可以被公开，用户通过添加标签、完善资源的主题信息，与他人分享所收藏的内容和个人观点、专业知识等，从而实现了用户聚合。此外，Hathitrust 提供了一些 Google Book Search 所不具备的服务，如用户自定义搜索、学术研究工具、便捷的资源获取途径等。Hathitrust 还采用一些新科技为特殊群体服务，如为盲人读者提供的有声阅读和盲人阅读等。

3.5.6 隐私问题

Google Book Search 与 HathiTrust 的服务中都涉及到用户隐私问题，与典型的图书馆出版商的数据库不同，这两个图书仓储对于任何在线用户都是可以自由获取的。在每一个仓储里面都有一些特征对任何人均可见，但使用这些资源的用户期望他们的隐私能够被保护。

在 2009 年 Google Book Search 争端解决的讨论中，Google 收到了很多要求提供隐私保护的请求。因为当用户通过 Google 商店购买图书时，一些用户信息会被发送给第三方共享。出版商接受的销售信息数据被连接到了用户的 Google 账户。当用户登录到他们的 Google 账户中使用谷歌图书，然后他们使用“我的图书馆”或者购买图书时，他们的活动都将会被记录。当用户把图书添加到“我的图书馆”时，他们必须把列表设为公开，以可以和其他人分享这些链接。因此这些个人信息都被公开了。

而 Hathitrust 的隐私政策宣称：它仅仅登录到与用户有关系的交易中，且持续很有限的一段时间。这些登录被用来解决困难和问题。并宣称没有个人信息会被共享给第三方。此外它宣称，当一个问题被解决后，登录信息就会被销毁。Hathitrust 使用 Google 的一个分析工具，使用了一个 cookie，并且把 IP 地址段传递给 Google。Hathitrust 的隐私政策解释了用户如何通过关掉 cookies 或者使用 Google 分析退出浏览器来退出 Hathitrust 的^[38]。

3.6 成效影响

3.6.1 在世界范围内促进了知识资源的长期保存和获取。两者都立足于对人类文明成果和

知识资源的长期保存和便捷获取,对于数字图书馆的探索都取得了让全世界瞩目的成绩,促进了知识组织、存储、获取、服务的长足进步。

3.6.2 为网络时代数字资源开放使用提供了可资借鉴的法律依据。二者均是大规模著作权侵权案被告,且终在法律层面上被认可为“合理使用”。

3.6.3 以非盈利为目的的基于合作与联盟体系的数字资源长期保存与共享前景广阔。(1)截至2017年4月,HathiTrust合作伙伴共有5个联盟体系和127个包括大学图书馆、公共图书馆、研究机构图书馆在内的个体机构,涵盖世界众多知名高校及研究机构,且发展势头良好。而Google Book Search作为以盈利为最终目的的商业项目,其发展受到一定制约,但对全球用户同等开放,其扩大业务领域和业务深度的进取心使得Google Book Search被称为“本世纪野心勃勃的计划”^[39]。(2)法国国家图书馆馆长Jean-Noël Jeanneney声称“Google电子书仓储中,英语图书占据绝对优势,反对英语文化的垄断”^[40],但不可否认这项计划的超前意识及巨大现实利好性,也让它被称为“值得握手的敌人”^[41]。(3)符合用户需求。2011年,美国著名非营利性组织伊萨基战略与研究部(Ithaka S+R)就通过网络调研了52所HathiTrust成员馆及部分非成员馆,反馈率达88%,详尽收集了关于HathiTrust用户满意度、功能规划、申请加入原因等方面数据^[42]。

一是用户对HathiTrust“我的图书馆”功能认可度较高;二是HathiTrust在做好成员馆用户服务同时兼顾全球其他公共用户也得到了一定程度的认可;三是对与HathiTrust共享资源原因的调查,两个突出的重要原因是使数字资源得到长期保存和希望自己的数字资源成为能被检索发现的数字资源集合的一部分,即不断提升的数字资源长期保存及被更广泛利用的意识,促使用户寻求更稳定和强大的数字平台,HathiTrust在这些方面具有相当的前瞻性和安全性;四是对HathiTrust功能运作偏好是在本地机构还是集中处理的调查显示,数字图书馆基础设施管理、数字化、推广和营销、研究支持、权力决策、工具开发、用户支持七项功能中除资源数字化偏向个体机构自主进行,其他六项功能都倾向于集中处理,抱团发展,特别是研究支持、工具开发对技术要求较高的两项功能各馆更希望借助联盟的整体实力。

4 基于 Google Book Search 与 HathiTrust 运行模式的 CASHL 转型发展策略

4.1 建立兼具前瞻性、科学性与现实性的 CASHL 战略规划

CASHL自成立以来即向着“以深入促进哲学社会科学繁荣为目标,以达到世界一流的文献保障水平为核心,努力建设国家人文社会科学信息资源平台”的总体目标而发展。不同时期不同战略规划决定着不同使命、目标、行动及后期发展,HathiTrust官网公布的系列中长期目标规划有着良好的启示。在“十四五”即将来临之即,树立“大资源与大服务”观,遵循共建“一个中国数字人文社会科学图书馆”的理念,以国际化的视野从国家角度、区域角度、联盟角度等对中国人文社会科学信息资源建设与服务进行顶层设计,建设覆盖全球重要学术资源的文献资源战略体系,打造基于现代化通讯网络技术的公共文献信息服务共享平台,

统一规划与评估,实现国家人文社科资源建设的一体化和资源间的互通互联,提高资源保障率,加速精准化服务。

4.2 联盟化协同发展——CASHL 发展重要机制

Google Book Search 与 HathiTrust 的高效发展历程中,充分体现着普遍性的联盟意识、联盟文化。目前,国内以 CALIS、CASHL、CADAL、NSTL 为代表,形成了影响力最大、运行最为成功的典型的全国性文献资源建设与服务体系。在这些全国性服务体系的带动下,基于区域性的文献资源共享联盟也随之发展壮大,如 BALIS、JALIS、TALIS、FULINK 等。这些全国性的文献服务体系、地域性的文献资源共享联盟加上以中国国家图书馆和上海图书馆为代表的文献服务公共机构,形成当前国内较为完备的文献保障体系。此外,超星百链云图书馆为代表的商业信息服务机构、百度学术文献互助平台为代表的网络文献互助机构也成为国内用户获取文献的有效补充。以上这些不同目标任务、不同规模层次、不同内容范围、不同运营模式的联盟,不断按照自行的轨迹成熟、规范、有序发展。然而从国家层面上,在资源建设、资源共享、统一服务平台、统一标准、保障评估等尚未凝聚成强大的合力乃至形成大的国际影响力,同时也缺少与国际各类联盟或机构的合作。在此情况下,加强联盟化协同发展^[43]势在必行,整合各级各类联盟和机构,在战略规划(国家层面)、协同机制(运营体制机制)、资源共建(CHAIR,自建数据库,特色资源等)、长期保存(国外引进数字资源、开放获取、数据等)、学科服务(竞争力评价、前沿态势分析)、深化共享(文献、数据、人力)、统一标准(资源、技术、评估)、联合仓储等诸多方面进行深入的联盟化协同发展。

4.3 合作共享——CASHL 资源建设的必由之路

优势互补、共建共享、联合发展是 HathiTrust 的成功经验,其成熟的管理体系和运行机制,效益显著的联合采购、馆际互借、密集仓储等联盟合作项目值得我们深入学习和借鉴。首先,高校图书馆领域内合作共享,经过多年发展,馆际互借与文献传递、统一集成管理系统、资源联合采购、资源联合仓储、数字化特色资源、图书馆员继续教育等在项目中不断应用发展。在传统合作项目不断走向成熟的同时,新的合作内容也在不断拓展,出现了联盟数据统计服务、联盟需求驱动采购、区域机构知识库建设等新兴内容^[44]。其次,突破地域、系统乃至图书馆的限制,建立广泛的跨界合作关系^[45]。应当立足于图书馆职业,将内部单一、浅层合作拓展到对外多元、前沿合作的层面。如 OCLC,尽管有着来自不同国家、地区图书馆的支持,但面对近年来文化遗产领域保护工作的兴起,它仍寻求与联合国教科文组织、蓝盾(Blue Shield)等组织的通力合作,以实现文化遗产的风险登记和保护。通过跨界合作将专业的形象和服务“带出去”,又将跨领域的知识、标准和资源“带进来”。此外,更广泛意义上的合作还包括:支持创新的科研合作,加强应用的技术领域合作,实现共赢的出版合作,提升服务的用户合作等。

4.4 以全球特色人文社会科学文献资源建为重心

Google Book Search 与 HathiTrust 的资源建设中, 突出显示了做为重点的国际一流特色资源的建设, 如 HathiTrust 收录的独具特色的美国政府文件、研究数据等。2015 年 OCLC 重新定义了图书馆资源建设的边界, 指出资源建设应由六部分构成: ①自有资源, 包括购买和保存的资源; ②可借用的资源; ③许可使用的资源; ④可共享的印本文献资源; ⑤需求驱动的资源; ⑥免费的资源, 包括谷歌学术上的资源、免费提供的电子书、Web 资源导航等^[46]。当前 CASHL 信息资源建设是自有资源、网络资源、共享资源、开放资源等充分融合的多元化的资源体系, 全球化特色资源建设要以重点学科领域、重点专题领域为依托, 实现特色资源的深度整合, 并从学科信息门户入手, 实现将学科特色资源建设与学科知识服务的无缝结合。

4.5 重视数字资源的长期保存和安全性指标

一直以来, HathiTrust 都将安全和可持续性做为联盟评价的重要指标^[47], 从理念、政策、管理到技术、工具、平台, 处处都体现了安全。可持续性信息安全是信息的延伸, 也是联盟运行需重点考虑的问题。在保存政策上、物理位置上、服务架构上、技术应用上都以可持续性为前提, 尤其是分布式多点数字资源备份的应用, 提升了数字资源保护的强度和联盟的可持续性发展。

随着数字资源的大量衍生和用户对数字资源的日益依赖, 数字资源长期保存问题日益突出。如何抵抗风险, 保证数字资源能够长期、合法和经济地被广泛获取已成为国家发展战略。虽然国内的数字资源长期保存研究和实践步入了快车道, 但国家层面的有关政策仍未出台, 国家层面的战略规划缺乏; 多个机构进行了实践探索, 但多处于起步阶段, 资源量小且共享有限, 数字资源的规模性、系统性长期保存未形成, 国家整体的数字资源容灾能力偏低^[48]。同时, 即使是已经进行长期保存的数字资源也无法保证即时可获取。未来, 还需要国家顶层设计, 制定有关的国家政策和发展规划, 建立起国家主导的数字资源合作长期保存体系, 逐步推进数字资源长期保存的规范化、系统化、常态化建设。

4.6 强化资源、服务与技术的深度融合

HathiTrust 十分重视技术在资源与服务中的应用。将自己的服务职能分为短期和长期两种。其中, 短期服务功能包括: (1) 翻页机制; (2) 品牌推广; (3) 格式验证、迁移和错误检查; (4) API 的开发应用; (5) 阅读障碍用户的访问机制制定; (6) 一站式发现服务; (7) 发布虚拟馆藏; (8) 非 Google 数字内容的直接采集机制。长期服务功能包括: (1) 遵守可靠存储库审计和认证 (TRAC) 标准及清单中的要求; (2) 强大的全文本跨库检索发现功能; (3) 创建数量众多的 API; (4) 支持图书和期刊以外所有数字资源格式的存储和获取服务; (5) 关注了对阅读障碍用户的服务; (6) 建立 HathiTrust 的研究中心 (HTRC), 进行先进软件工具和技术开发, 并广泛采用其他来源的分析工具, 以为学者提供一个安全的计算和数据环境^[49]。

作为数字资源的保存联盟, HathiTrust 具备数字资源存储和访问的许多基本服务功能,

包括支持图书和期刊以外所有数字资源格式的存储和获取服务；支持存储和访问数字资源的多种格式和多个版本；支持对数字资源的一站式发现；支持用户和平台认证、授权服务；支持对操作或事件的跟踪；数据注释服务；记录验证服务等。并且将构建开放协作的环境作为主要的服务特征之一，通过开放 API 的形式，为用户提供更丰富、更有创意的平台应用^[50]。

HathiTrust 云存储联盟以云技术为依托搭建了数字资源仓储系统，基础设施由密歇根大学承担，联盟成员馆不仅能够安全、快速、经济地完成数字资源的长期存储（资源所有权归各成员馆），而且会员用户能够不受限制地获取联盟资源，实现了资源存储与开放服务的统一。它为图书馆实现服务模式的转化、降低运营成本、快速灵活地响应社会 and 市场需求提供了新的思路和解决方案^[51]。

此外，HathiTrust 还开发了支持手机、ipad、平板电脑等移动终端设备连接、下载、保存资源的功能，并且提供可视化工具，便于用户对资源进行可视化分析，以开放服务的视角保证服务与技术的同步发展、深度融合。

4.7 数据质量和数量是 CASHL 持续发展的关键

资源数据质量是 Google Book Search 和 HathiTrust 建设的优先事项，关系到用户需求的满足度和满意度，是项目或联盟有效运转和可持续发展的重要基础。二者都竭尽全力致力于提高资源数据质量，它们先进的数字化扫描技术，大型的数据存储设备以及运作模式，都保障了仓储内电子书的质量。数据质量是一个复杂的概念，既包括对数据规范性、一致性、准确性、完整性等的要求，也包括对数字资源描述和揭示的规范性、可获得性、关联性等的限定。HathiTrust 重视解决多层次、宏观上的数字对象及其元数据的质量问题，并直接与合作机构、规划指导委员会和相关团队合作，推出了一系列评估方法，持续地推进数据质量^[52]。

海量的资源才能带来规模效应。Google Book Search 和 HathiTrust 都具有海量优质的电子书，并且对内容进行了深度加工整合，具有高度整合的优质内容资源是它们发展的战略核心。它们把上千万册的处于离散状态的图书全文资源高度集成，给全球亿万用户带来了便捷的资源获取。巨大的用户资源又能吸引上游的内容提供商和下游的终端设备商加盟合作，从而可以实现对全行业的主导与整合。

4.8 科学高效的管理体系是 CASHL 发展的保障

Google Book Search 和 HathiTrust 都制定有较完善的联盟（机构）章程和科学的管理体系。联盟章程是联盟发展的纲领性指导文件，其明确规定了各个委员会、各个成员在联盟中的权责以及如何进行联盟的管理，是联盟日常运行及其他一切文件制定的依据和参照。同时，随着联盟的发展，还会适时对章程进行修订、补充和调整，以保证章程的与时俱进的适用性。董事会管理制是国外数字资源保存联盟普遍采用的一种管理制度，其特点是高效实用，可以灵活多变地进行组织架构、治理方式、部门分工等的调整，能够综合考虑政府、社会、成员机构、其他合作伙伴等多方的利益。董事会下还设立了其他委员会，以负责联盟具

体业务的运行和发展，此外，根据业务发展需要，常设或临时设立了一些工作组，用于研究和解决一些专门问题，处理联盟发展中遇到的各种问题。总之，这种统一规划、科学分工、有序管理、灵活协调的联盟管理体系既保障了各方的利益，充分调动了各方的积极性，又推动了联盟的可持续发展。

4.9 提升数字资源的服务功能

4.9.1 强化检索功能

Google Book Search 和 HathiTrust 都有强大的数据库功能，两者的检索策略和界面显示都非常先进。Google Book Search 沿用搜索引擎式便捷检索策略，允许布尔检索而不需要用户知道布尔检索式。HathiTrust 的统一检索平台，一次登录，支持多语言、多途径、跨库检索，返回多个数据库检索结果并显示存储具体位置，直接享有成员馆馆藏服务。在检索结果显示上注重完整性，包括纸质资源信息、数字文献及相应配套数字资源在线获取情况，还未完全数字化的文献可先提供书封，试读片段；对资源的下载、在线阅读、借阅热度等数据提供可视化图表分析，进一步揭示资源利用动态；未检索到文献推荐其他获取途径，如国内外其他现有文献保障系统内资源、网络开放资源、网上书店购买等有效链接；注重相关文献推荐，形成知识专题，利于用户拓展知识网络并进行知识发现。

4.9.2 深化个性化服务

HathiTrust 同时兼具强大的个性化功能，“集合”功能形成知识社区，并不断开发新技术工具为用户服务。Google Book Search 提供了很好的知识发现工具，如显示该书中常见术语和短语形成的词云，该书热门引用章节及引用次数，以及检索内容在书中提到位置。并在“我的图书馆”中深入个性化知识服务，实现知识互动，深度发掘文献资源。营造知识交流新阵地，建立读者观点表达和交流栏目，如文献评价和评分，发挥读者参与馆藏建设、评价资源与服务的作用，促使知识流动与创新，让读者成为知识受益者、生产者；开发及完善移动端服务，如手机 OPAC，APP 等；还可利用学术分析工具和知识管理工具，加强知识服务，并开设相关工具应用的教学讲座，或专门录制使用教学视频；RSS 定制或推送功能定期提供所需的新书入库功能，同时提供个人定制和个人虚拟书架功能；提供多种引文输出的方式，以实现个性化的使用；在提供深化知识服务及可视化工具的运用中，同时应注意对用户隐私的保护。

4.10 助推版权策略

Google Book Search 和 HathiTrust 都在经历了版权纠纷后，十分注意实行严格的版权制度。借鉴二者成功经验开展版权管理^[53]：①借鉴“选择-退出 (Opt-out)”模式，使其成为践行合理使用制度的重要手段；②引入版权信息库，实时更新版权授权信息及版权期限

满情况，加强版权动态管理；③完善限制获取，已过版权期限文献，高校用户通过校园网或VPN即可全文获取，还在版权期限文献，允许在线全文阅读，获得相关允许协议后可复制、下载；④从合理使用过渡到法定许可，还在版权期限的文献也可全文获取，已收藏文献按一定方案一次性给予版权所有者补贴，未来购入文献在批量购买时一次性付清后期补贴。⑤尝试“默示许可”模式，声明所有资源不得用于盈利活动，默认许可非盈利性质的网络传播。综上，要实行电子图书的仓储战略，既要注意尊重图书的版权，保护版权所有者的利益，又要灵活的使用版权策略，为不同的版权许可转让签订不同的版权协议，为电子图书的仓储战略铺平道路。

5 结语

近20年来，高校联盟式HathiTrust、校企或企业间合作式Google Book Search的成功案例，验证了多元化联盟的强大优势。同时，做为国家主导式的CASHL项目，围绕国家重大战略和高校“双一流”建设需求，以服务哲学社会科学教学科研为目标，在开展大规模有特色成体系的文献建设，扩大共享规模，优化基础服务，创新知识服务，构建文献安全保障机制，建成资源丰富、服务一流、便捷高效的国家人文社会科学信息资源平台方面做了大量的工作，众多高校成员馆在资金上有国家财政扶持，技术力量和人才资源有一定保证，文献资源历史积累丰厚，服务网络 and 用户遍及全国，取得了良好的效益和一定的国际影响力。然而，在目标规划设定、资金雄厚程度、资源数量质量、技术创新能力、深入合作程度、个性化服务、版权管理策略等方面较HathiTrust和Google Book Search还有一定的差距。因此，Google Book Search和Hathitrust对于CASHL的发展大型电子书仓储业务具有很好的借鉴，把握好海量的内容资源、强大的数据库功能、版权策略的助推、合作发展的优势和资源质量的保障等几大成功要素，紧跟时代发展趋势，牢牢抓住机遇，汲取Google Book Search和Hathitrust的先进理念和成功模式，更好地推动CASHL的可持续发展，让每个参与项目的高校在信息技术、数字图书馆、项目管理等领域都发挥其就有的力量，聚合资源，消除障碍，提供有价值的研究工具，为中国的人文社会科学研究提供坚实有力的文献资源保障。

参考文献：

- 1 陈邦，徐军华，张帆. Google Book Search 与 HathiTrust 案例特点及启示, 大学图书馆学报, 2018 (10): 79-86
- 2 Google Books [EB/OL]. [2020-03-06]. <https://books.google.com/>.
- 3 徐跃权, 董贺, 孔悦凡. 管理学视角下的 Google 图书搜索项目解析[J]. 图书馆学研究(理论版), 2010(1): 37-40.
- 4 HathiTrust [EB/OL]. [2020-04-20]. <https://www.hathitrust.org/>.
- 5 Wilkin J. BackTalk: HathiTrust and the Google Deal [J]. Library Journal, 2009, 15.
- 6 周军兰. Google 数字图书馆计划及其影响分析[J]. 情报资料工作, 2006(2): 69-71.

- 7 Mission and Goals [EB/OL]. [2020-04-05]. https://www.hathitrust.org/mission_goals.
- 8 陈铭. 谷歌图书与Hathitrust的比较与启示. 图书馆杂志, 2016(12):80-88
- 9 关于CASHL. [EB/OL]. [2020-04-05]. <http://www.cashl.edu.cn/index.php/node/45>
- 10 CASHL 资源体系. [EB/OL]. [2020-04-05]. http://lib.lzu.edu.cn/cashl_zyyfw/info-6738.shtml
- 11 Functional Framework[EB/OL]. [2020-04-06]. https://www.hathitrust.org/functional_framework.
- 12 Google's audacious project that scanned millions of books has been declared legal [EB/OL]. [2020-04-2]. https://thenextweb.com/google/2015/10/16/googles-audacious-project-that-scannedmillions-of-books-has-been-declared-legal/#.tnw_Cdk5VDpk.
- 13 Naomi Eichenlaub. Checking in with Google books, Hathitrust and the DPLA[J]. Computers in Libraries, 2013, 33(9): 6-9.
- 14 童万菊. Google 数字图书馆“选择退出”(opt-out)机制的分析及思考 [J]. 图书馆杂志, 2014(6): 92-96.
- 15 李捷. 论网络环境下的著作权默示许可制度[J]. 知识产权, 2015(5): 67-71, 108.
- 16 周丽霞, 张偶. Google 数字图书馆版权获取模式及对我国的启示[J]. 情报科学, 2012(9): 1325-1352.
- 17 陈铭. Google 图书与HathiTrust 的比较与启示 [J]. 图书馆杂志, 2016(12): 80-88.
- 18 夏立新, 金晶. 从Google 网络图书馆计划的成功启动看图书馆数字化发展[J]. 情报科学, 2009, 27(4): 485-488.
- 19 李咏梅, 袁学良, 唐李杏. 美国HathiTrust 项目及其对我国高校图书馆数字化的启示[J]. 四川图书馆学报, 2011(3): 35-37.
- 20 Authors Guild Drops HathiTrust Case [EB/OL]. [2017-05-12]. <http://www.publishersweekly.com/pw/by-topic/digital/copyright/article/65217-authors-guild-drops-hathitrust-case.html>. 6. Challenge to Google Books Is Declined by Supreme Court [EB/OL]. [2017-05-12]. <http://www.scoop.int/googletfreely>.
- 21 同 19
- 22 单蓉蓉, 陆铭, 魏可. 云环境下HathiTrust 的用户服务研究 [J]. 图书馆论坛, 2012, (6): 69-73.
- 23 刘万国, 周秀霞, 姜雷. 数字资源长期保存联盟研究. [J]. 情报资料工作, 2016, (3): 75-79
- 24 A Preservation Infrastructure Built to Last: Preservation, Community, and HathiTrust [EB/OL]. [2018-03-25]. <https://www.hathitrust.org/documents/york-MemoftheWorld-201209.pdf>.

- 25 周小文. 谷歌数字图书馆的运作模式对我国数字图书馆发展的启示[J]. 新世纪图书馆, 2011(1): 68-69.
- 26 苏海明. HathiTrust 数字仓库项目概述 [J]. 数字图书馆论坛, 2009 (7): 60-65.
- 27 Technological Profile [EB/OL]. [2018-03-25]. <https://www.hathitrust.org/technology>.
- 28 Google' s audacious project that scanned millions of books has been declared legal [EB/OL]. [2020-04-2]. https://thenextweb.com/google/2015/10/16/googles-audacious-project-that-scannedmillions-of-books-has-been-declared-legal/#.tnw_Cdk5VDpk.
- 29 Deposited Volumes by Original Source of Content [EB/OL]. [2020-04-2] [https://www.hathitrust.org/statistics_deposited_volumes_monthly]
- 30 Statistical Overview of HathiTrust's Federal Documents Collection [EB/OL]. [2020-04-2] [https://www.hathitrust.org/usdocs_registry/stat_overview]
- 31 HathiTrust Digital Library 和 OCLC 发布 World Cat Local 原型 [EB/OL]. [2017-05-12]. <https://www.oclc.org/zh-Hans/news/releases/2011/20114.html>.
- 32 Laura Sare. A Comparison of HathiTrust and Google Books Using Federal Publications[J]. Practical Academic Librarianship: The International Journal of the SLA Academic Division, 2012, 2(1): 1-25.
- 33 Serials Solutions to Enable Full-Text Search of the HathiTrust Collection from the Summon™ Service [EB/OL]. [2017-05-12]. <http://www.proquest.com/about/news/2011/serialssolutions-summon-full-text-hathitrust-collection.html>.
- 34 同 17
- 35 同 17
- 36 Laura Sare. A Comparison of HathiTrust and Google Books Using Federal Publications[J]. Practical Academic Librarianship: The International Journal of the SLA Academic Division, 2012, 2(1): 1-25.
- 37 同 22
- 38 同 22
- 39 WHAT EVER HAPPENED TO GOOGLE BOOKS? [EB/OL]. [2017-05-12]. <http://www.newyorker.com/business/currency/what-ever-happened-to-google-books?mbid=rs>.
- 40 Jeanneney J N. Google and the myth of universal knowledge: A view from Europe [M]. University of Chicago Press, 2008.
- 41 王陆军. 睁眼看世界: 我们向国外图书馆学习什么 [M]. 北京: 海洋出版社, 2010: 127.

- 42 Ithaka S+R .Briefing Paper on Progress and Opportunities for HathiTrust[EB/OL]. [2017-06-02]<http://www.hathitrust.org/documents/hathitrust-3year-review-2011.pdf>.
- 43 朱强. 联盟化协同——图书馆发展的重要机制[PPT]. [2018-04-26]. 西安
- 44 孙波, 刘万国, 倪煜佳, 姜凯予. 美国典型区域图书馆联盟研究[J]. 现代情报, 2018, 38(3):116-123
- 45 肖鹏, 陈润好. 从资源共同体到命运共同体: 中国图书馆联盟的四种建设路径及其未来趋势[J]. 图书馆建设, 2018, (94)12:31-36
- 46 The library in the life of the user: some contextual remarks lircan Dempsey[EB/OL]. [2020-04-20]. <https://slideplayer.com /slide/8719768/>
- 47 刘万国, 周秀霞, 姜雷. 数字资源长期保存联盟研究[J]. 情报资料工作, 2016, (3):75-79
- 48 刘万国, 周秀霞. 数字学术信息资源保障的国家协同布局模式研究 [J]. 情报理论与实践, 2016, 39 (8): 40-43
- 49 Functional Objectives [EB/OL]. [2020-04-22]. <https://www.hathitrust.org/objectives>.
- 50 同 47
- 51 刘华, 许新巧. 从共享到共有: 学术图书馆云存储联盟研究[J]. 图书馆, 2014, (4):127-129.
- 52 Quality [EB/OL]. [2020-04-22]. <https://www.hathitrust.org/quality>.
- 53 蒋逸颖, 周淑云. Google 数字图书馆版权获取模式对我国的启示[J]. 图书馆学研究, 2016 (13): 33-36.