

机构知识库HathiTrust的版权管理研究及其启示

□ 王亚凤 / 江苏科技大学图书馆 镇江 212003

摘要: HathiTrust是由美国机构合作委员会 (CIC) 图书馆联盟下的13所大学合作发起的机构知识库项目。HathiTrust采用了严格、详尽的版权管理政策,建立了专门的版权数据库。版权数据库里存储并持续追踪HathiTrust每卷作品的版权信息,实现版权状态的自动更新。文章列举了HathiTrust版权数据库中定义的19种属性类别和17种原因代码,并结合HathiTrust中具体的版权处理实例进行了说明。最后阐述了HathiTrust版权数据管理对我国机构知识库建设的启示。

关键词: HathiTrust, 机构知识库 (IR), 版权数据库, CALIS

DOI: 10.3772/j.issn.1673—2286.2014.01.011

1 前言

HathiTrust Digital Library创始于2008年,最初是美国机构合作委员会 (Committee on Institutional Cooperation, CIC) 图书馆联盟下的13所大学合作发起的机构知识库项目。Hathi在印地语中是“大象”的意思。Hathi与Trust两词结合,传递的核心理念是帮助研究者能从海量的数字存储中获得并分享第一手宝贵资料^[1]。HathiTrust的成员馆不断增加,截止到2013年5月,已有美国麻省理工学院、杜克大学、普林斯顿大学、斯坦福大学等72所图书馆机构参与合作^[2]。资源总计有1074万卷册 (volumes), 包括563万册书 (book titles), 28万册丛书 (serial titles), 37亿多页 (pages), 容量482TB, 其中32%的资源属公共领域^[3], 规模非常宏大, 当属翘楚。

HathiTrust发展之迅速与它完善的版权管理政策密不可分,其版权数据库存储并追踪每卷作品的版权信息,以实现版权状态的自动更新。本文旨在详细分析HathiTrust先进的版权数据管理方法与技术,为我国机构知识库及数字图书馆建设提供一些参考。

2 HathiTrust的版权管理

2.1 HathiTrust版权数据库

HathiTrust采用了严格、详尽的版权管理办法,建立

了专门的版权数据库 (HathiTrust Rights Database)。版权数据库存储并持续追踪HathiTrust每卷的版权信息。其意义在于: 1) 为版权信息建模以减少维护工作; 2) 确保版权术语解释的准确性; 3) 追踪超期后版权信息的变化等^[4]。频繁地更新数百万条记录,以及频繁地改变数据结构对于大型数据库的长期运行来说几乎是行不通的。最安全的做法是尽可能地使版权数据库建立在简单、成熟的版权政策及版权术语的基础之上^[5]。所以HathiTrust将版权数据库里的属性类别事先定义,属性内容与版权政策保持一致,并在元数据编目时就标注其权利状态,例如定义出版物的版权状态的属性是公共领域 (pd) 还是在版权保护期 (ic) 内,或是被权威认定为是孤儿作品 (orph),或是知识共享署名许可 (cc) 等等。版权数据库中版权状态的属性类型,具体如表1^[6]。

HathiTrust在版权数据库中将原因代码 (reasons) 分成17种,用来描述版权状态的形成原因^[7],如书目数据衍生信息 (bib), 作品中没有版权声明页 (ncn), 作品的版权状态取决于版权持有人的合同协议文件 (con) 等等,详见表2。HathiTrust利用属性类别和原因代码二者结合的方式来表征版权状态,访问权限的优先级别依据版权状态而定。权限模型中的优先级分成四个级别: 最低为1级,最高为4级^[8]。例如第1级pd/bib、ic/bib、und/bib、pdus/bib。第2级ic/unp、pd/ncn、ic/ren等。第3级nobody/pvt、cc-by/con等。第4级pdus/man、ic-world/man等。级别越高,作品的版权状态越

表1 HathiTrust版权数据库属性列表

ID	名称	类型	描述	describe
1	pd	版权	公共领域	public domain
2	ic	版权	版权保护期	in-copyright
3	op	版权	绝版（受版权保护）	out-of-print (implies in-copyright)
4	orph	版权	孤儿版权（受版权保护）	copyright-orphaned (implies in-copyright)
5	und	版权	未确定版权状态	undetermined copyright status
6	umall	访问	UM分支机构和赞助人（所有校区）	available to UM affiliates and walk-in patrons (all campuses)
7	ic-world	访问	版权保护和被版权持有人许可的世界范围内可见	in-copyright and permitted as world viewable by the copyright holder
8	nobody	访问	无人有权限，对所有用户封锁	available to nobody; blocked for all users
9	pdus	版权	仅限在美国的公共领域	public domain only when viewed in the US
10	cc-by	版权	知识共享署名许可	Creative Commons Attribution license
11	cc-by-nd	版权	知识共享署名许可-禁止演绎	Creative Commons Attribution-NoDerivatives license
12	cc-by-nc-nd	版权	知识共享署名-非商业性使用-禁止演绎	Creative Commons Attribution-NonCommercial-NoDerivatives license
13	cc-by-nc	版权	知识共享署名-非商业性使用	Creative Commons Attribution-NonCommercial license
14	cc-by-nc-sa	版权	知识共享署名-非商业性使用-相同方式共享	Creative Commons Attribution-NonCommercial-ShareAlike license
15	cc-by-sa	版权	知识共享署名-相同方式共享	Creative Commons Attribution-ShareAlike license
16	orphcand	版权	孤儿作品的候选作品-在90天的持有期内（受版权保护）	orphan candidate - in 90-day holding period (implies in-copyright)
17	cc-zero	版权	知识共享无许可（PD）	Creative Commons Zero license (implies pd)
18	und-world	访问	未确定版权状态，存储人允许世界范围内可见	undetermined copyright status and permitted as world viewable by the depositor
19	Icus	版权	受版权保护，仅限在美国境内	in copyright in the US

表2 HathiTrust版权数据库原因代码列表

ID	名称	描述	describe
1	bib	自动化进程下书目数据的衍生信息	bibliographically-derived by automatic processes
2	ncn	没有版权声明	no printed copyright notice
3	con	版权持有人的合同协议文件	contractual agreement with copyright holder on file
4	ddd	勤勉档案文件	due diligence documentation on file
5	man	手动访问控制；详情见附注	manual access control override; see note for details
6	pvt	私人信息可见	private personal information visible
7	ren	版权续期研究	copyright renewal research was conducted
8	nfi	需要进一步调查（版权研究部分完成；遇到不明确、不清楚或其他耗时的情况）	needs further investigation (copyright research partially complete; an ambiguous, unclear, or other time-consuming situation was encountered)
9	cdpp	标题页或封底包含版权日期、和/或书目记录中无出版地信息	title page or verso contain copyright date and/or place of publication information not in bib record
10	ipma	已出版和市场调研已进行	in-print and market availability research was conducted

表2 HathiTrust版权数据库原因代码列表 (续)

ID	名称	描述	describe
11	unp	未发表的作品	unpublished work
12	gfv	谷歌可视性设置 (VIEW_FULL)	Google viewability set at VIEW_FULL
13	crms	通过内部版权解决政策, 版权审查管理系统 (CRMS) 中生成多种评论; CRMS记录详情咨询	derived from multiple reviews in the Copyright Review Management System (CRMS) via an internal resolution policy; consult CRMS records for details
14	add	作者死亡日期研究已实施或收到权威方的通知	author death date research was conducted or notification was received from authoritative source
15	exp	有合作者的非美国作品的版权期限届满	expiration of copyright term for non-US work with corporate author
16	del	从机构库中删除; 详情见附注	deleted from the repository; see note for details
17	gatt	非美国的公共领域作品恢复成在美国境内受版权保护 (GATT)	non-US public domain work restored to in-copyright in the US by GATT

严苛, 访问权限越高, 作品越难获取。

2.2 版权数据管理案例

针对上述列表中的内容, 笔者结合HathiTrust中具体的版权管理实例^[9], 就应用属性类别和原因代码的使用规则进行一些说明:

根据美国版权法, 符合以下条件的作品, 被视为公共领域 (属性=pd), 访问不设限制: ①美国联邦政府的文件; ②1923年之前在美国境内的出版的作品; ③1870年之前在美国境外出版的作品。而对于不符合上述标准的作品 (如1923年后美国境内出版的, 而不是政府文件的作品) 被视为受版权保护 (属性=ic及原因=bib)。HathiTrust还规定将附加属性用于以下的情況作品: 1870年至1923年之间在美国境外出版作品。其版权状态取决于用户的IP地址, 即IP地址源于美国的用户有权获取1870年至1923年之间在美国境外出版的作品, 而非美国IP地址的用户则无权限获取 (属性=pdus和原因=bib)。HathiTrust将是否属公共领域的规则制订得十分周密, 为后续版权状态判定提供了重要标准。

例如Bell所著的*Edward Carpenter, The British Tolstoi*于1932年在美国出版。根据判定规则, 该作品受版权保护, 访问应受限制。HathiTrust在调查中发现该作品中没有版权声明页。根据美国版权法, 该作品应属公共领域。因此, 其版权状态为公共领域 (属性=pd)

且缺乏版权声明 (原因=ncn) 状态, 访问不受限制。笔者进入HathiTrust的搜索界面, 检索到了该书*The British Tolstoi*, 版权状态显示的是“copyright: public domain”, 可下载整本书的PDF文档。

例如Stirling Bowen所著的*Wishbone*于1930年在美国出版。根据判定规则, 其应属版权保护期, 访问应受限制。经调查发现作品中版权人已更正, 现在的版权持有者是Penguin。HathiTrust和Penguin商议后, Penguin同意对HathiTrust的机构用户开放访问, 条件是HathiTrust不得提供重印本, 并要求协议必须在五年内续期。所以, HathiTrust对合作机构用户开放该作品的访问权限, 但不提供页面图像的打印服务。因此, 该作品的版权状态为受版权保护, 且仅对合作机构用户开放访问权限 (属性=umall; 原因=con)。

再如Alfredo Candia Guzman所著的*Bolivia: un experimento comunista en la America*, 20世纪50年代在玻利维亚出版。根据判定规则, 属版权保护期, 访问权限受限。然而HathiTrust调查发现作者已去世, 出版社也不复存在。因此, 该作品被认定是一个孤儿作品 (属性=orph, 原因=ddd)。

上述实例列举了HathiTrust在美国现有司法制度下, 针对不同版权状态作品的不同处理方法, 并对作品、版权持有人或被转让人实行了勤查制度。调查结果以属性类别与原因代码二者组配的方式来实现, 用于表征版权的不同状态。根据权利状态的不同, 设置访问权限, 实现版权数据的自动化管理。

3 启示

HathiTrust的合作机构用户绝大部分是美国的高校图书馆,其与我国CALIS的成员馆构成、项目宗旨等极为类似,都是高等教育数字图书馆联盟,目的都是为实现数字资源的共建与共享。CALIS三期规划的高校机构知识库建设及推广项目,于2011年6月开始筹备^[10],由北京大学、清华大学、厦门大学等五个示范馆联合建设。项目组与各高校协同工作,开发完成三套机构知识库系统。但笔者发现,不论是中心系统还是本地系统,以及CALIS机构知识库主页(<http://ir.calis.edu.cn>)下的各个栏目,均未提及版权管理的内容^[11]。2012年4月27日,CALIS三期机构知识库建设项目验收报告中的系列标准/规范/说明/手册^[10],也未述及版权管理方面的内容。在知识产权法律关系日益全球化的今天,CALIS三期作为国家级机构知识库的开发平台,居然没有任何涉及版权管理的内容或声明,笔者颇感失望。

CALIS三期机构知识库项目可以借鉴和参考HathiTrust数字化版权管理的经验,结合我国现有的版权法律法规及适合于图书馆享有著作权法规定的具体限制与例外,建立CALIS自己的版权数据库。例如在元数据中定义版权的基本属性类别,包括公共领域、版权保护期、知识共享署名许可,或是受版权保护仅限在中国境内等等,这些在技术上并不难实现,关键是规则的具体内容如何根据我国的情况制订。

例如根据《中华人民共和国著作权法》规定权利的保护期为“作者终生及其死亡后五十年,截止于作者死亡后第五十年的12月31日。对于作者身份不明的作品,其著作权法第十条第一款第(五)项至第(十七)项规定的权利的保护期截止于作品首次发表后第50年的12月31日。作者身份确定后,适用著作权法第二十一条的规定”。作品保护期届满,版权不再予以保护,作品进入公共领域。而美国版权保护期限是作者死亡后七十年,且我国并未像美国那样制订具体的进入公共领域的年限(如1923年之前在美国境内出版;1870年之前在美国境外出版)。因此,制订版权管理政策时要结合我国的著作权法及相关法律法规,具体情况具体分析。版权状态调查过程中可能需要大量的人力、物力对作品、版权

持有人进行清查制度,会遇到各种棘手的问题,但是不能因为过程繁复,而讳莫如深。尤其是CALIS三期作为国家级的机构知识库平台,其发展理念与建设构想应当具有一定的前瞻性和宏观性。

4 结语

HathiTrust详细、周密的版权管理方法与措施为我国机构知识库的建设提供了良好的借鉴与参考平台,但值得注意的是完全照搬照套国外的方法也是不可取的。例如,目前我国法律还没有“孤儿作品”概念,即这类作品归国家所有,而不是让其进入公共领域。但是,2012年7月6日,国家版权局在网站发布《中华人民共和国著作权法》(修改草案第二稿),其第26条系对第一稿第25条规定的孤儿作品法定许可条款进行了细化和拓展,增加了其他机构使用的相关内容。第42条增加了第13款作为合理使用的开放性规定,弥补了无法穷尽列举豁免情形的缺陷^[12],这些法律条款的变化为我国机构知识库建设及数字图书馆的版权政策提供了重要依据与参考,构建时需注意。

参考文献

- [1] What is HathiTrust? [EB/OL]. [2013-05-30]. http://www.hathitrust.org/help_general.
- [2] Individual Institutions [EB/OL]. [2013-05-30]. <http://www.hathitrust.org/partnership>.
- [3] Currently Digitized [EB/OL]. [2013-07-05]. <http://www.hathitrust.org/about>.
- [4] Introduction [EB/OL]. [2013-05-30]. http://www.hathitrust.org/rights_database.
- [5] Philosophy [EB/OL]. [2013-05-30]. http://www.hathitrust.org/rights_database.
- [6] Database Layout-ATTRIBUTES [EB/OL]. [2013-05-30]. http://www.hathitrust.org/rights_database.
- [7] Database Layout-REASONS. [EB/OL]. [2013-05-30]. http://www.hathitrust.org/rights_database.
- [8] Precedence of Rights Information-Using rights types and reason

- codes to infer precedence [EB/OL]. [2013-05-30]. http://www.hathitrust.org/rights_database.
- [9] Use Cases [EB/OL]. [2013-05-30]. http://www.hathitrust.org/rights_database.
- [10] 聂华,韦成府,崔海媛.CALIS机构知识库:建设与推广、反思与展望[J].中国图书馆学报,2013(3):46-52.
- [11] CALIS机构知识库[EB/OL]. [2013-05-30]. <http://ir.calis.edu.cn>.
- [12] 肖燕.追寻著作权保护与权利限制的平衡:图书馆界对《著作权法》修改草案的建议与期盼[J].中国图书馆学报,2013(5):18-23.

作者简介

王亚凤 (1979-), 女, 硕士, 馆员, 研究方向: 参考咨询与科技查新。E-mail: w7911w@126.com

Research about Copyright Management of HathiTrust Institutional Repository

Wang Yafeng / Library of Jiangsu University of Science and Technology, Zhenjiang, 212003

Abstract: HathiTrust was initially conceived as an IR collaboration of the thirteen universities of the Committee on Institutional Cooperation (CIC). HathiTrust used strict, detailed copyright policy, and specially established copyright database. Copyright database stores and keeps on tracking copyright of each volume which copyright status is updated automatically. We elaborate nineteen attributes and seventeen reasons of HathiTrust copyright database with the examples. Finally we discussed the HathiTrust Rights Management effect on institutional repositories of China.

Keywords: HathiTrust, Institutional repository (IR), Copyright database, CALIS

(收稿日期: 2013-11-18)