



LOCKSS 还是 Portico

——谈图书馆电子资源的长期保存

LOCKSS or Portico: Long-Term Preservation of Electronic Resources in Libraries

孙 艳 马炳厚 王 栋 (第四军医大学图书馆 陕西 西安 710032)

[摘 要] 目前国际上有多种比较成熟的网络电子资源长期保存模式,其中比较有影响力的是 LOCKSS 和 Portico 模式。通过对 LOCKSS 和 Portico 的组成模式、运作方式、工作原理、可靠性和费用等方面的对比分析,可以明确二者的优势和适用范围,并从中吸取经验,为我国图书馆电子资源长期保存模式的研究和实践提供参考。

[关键词] LOCKSS Portico 电子资源 长期保存

[中图分类号] G253.73 [文献标识码] B

[Abstract] There are several mature long-term preservation models of network electronic resources in the world. LOCKSS and Portico are main patterns with large influence. Compared LOCKSS with Portico in organization model, operation mechanism, working principle, reliability and cost, we could clarify advantages and adaptable scope, absorb experience from them, and provide reference for the research and application of long-term preservation models of network electronic resources in China.

[Key words] LOCKSS; Portico; Electronic resource; Long-term preservation

1 引 言

随着信息技术的发展和互联网的普及,电子资源已成为科研与教育领域的主流资源。而且由于电子资源便利与实用的特点,越来越多的研究人员已经或正在改变着查找资源的习惯,即更加倾向于使用电子资源。大多数图书馆购买电子资源所占的经费比例也逐年增加。但由于电子资源不像纸质文献那样具有有形的实体可以保存,而且大多数资源是通过网络获取的,出版商对电子资源也没有长期保存的义务,因此,图书馆无法确保正在使用的资源能为未来的读者服务。

2008年11月在北京召开了主题为“国家数字文献资源长期保存服务体系建设的数字图书馆高层论坛首届年会”,会议呼吁国家教育科研机构 and 文献情报机构充分认识数字文献长期保存的战略重要性和现实紧迫性,行动起来,促进我国数字资源长期保存的顺利开展^[1]。其实从上世纪90年代开始,国外就已经开展了比较正式的对数字资源长期保存的研究活动,该领域经历了从基础理论研究到个体实验再到最佳实践的发展过程。目前相对成熟并已经在实施的模式包括:机构独立保存、合作保存和第三方委托保存^[2]。

机构独立保存是由保存机构在拥有对被保存资源的使用权的条件下,获得出版商许可后进行的保存。这种保存模式目前一般都有国家收藏的背景。如英国国家图书馆(British Library,简称BL)、荷兰国家图书馆(Koninklijke Bibliotheek,简称KB)等。本文将重点放在合作保存和第三方委托保存上。目前这两种模式比较成熟的代表是LOCKSS和Portico,并且两种模式也都运行了10年左右,都拥有一定的用户群体。本文下面将从两种模式的组成模式、工作原理和费用3个方面对LOCKSS和Portico分别进行比较。

2 LOCKSS 和 Portico 的比较

2.1 组成模式

2.1.1 LOCKSS (Lots of Copies Keep Stuff Safe)是由MELLON基金(Andrew W. Mellon Foundation)、美国国家自然科学基金(National Science Foundation)、美国国会图书馆资助,斯坦福大学图书馆实施的一个致力于解决电子资源长期保存的项目。该项目意为通过建立出版商与图书馆之间的协作关系,允许图书馆在本地收藏、管理图书馆订购的电子资源并提供用户服务的系统^[3]。LOCKSS从1999年开始运行至今已有11年的历史了,截至2008年秋季,

LOCKSS 成员包括全球超过 400 家主要的出版商和 200 家图书馆, 并且 LOCKSS 每天新增 700 万页信息^[4]。

2.1.2 Portico 始于 2002 年, 其宗旨是保存电子形式的学术资源, 以备将来的研究人员查找与使用。该项目得到 MELLON 基金、ITHAKA、美国国会图书馆和 JSTOR 的资助^[5]。Portico 作为第三方的资源保存机构为图书馆和出版商服务。截至 2009 年 8 月 7 日, Portico 已拥有授权电子期刊 10 324 种、电子书 6 488 种, 总计 13 566 761 篇文章。合作出版商 75 家, 其中不乏全球知名学术期刊出版商 Elsevier、Springer、John Wiley&Sons. Inc.、IEEE.Inc.、OUP 等, 囊括了商业性出版商、大学出版社及学协会出版社^[6]。其拥有合作图书馆 491 家, 其中 334 家是美国图书馆^[7]。

2.2 工作原理

2.2.1 LOCKSS 采取的模式是为图书馆提供开放性源码的分布式存储系统, 图书馆只要具有资源的认证, 或者说得到了出版商的许可或授权, 就可以在本地收藏、管理电子资源。LOCKSS 对运行环境要求很低, 一般的个人计算机就可以完成整个存储任务。图书馆员可对文献的保存、使用、访问等进行管理和监控。

LOCKSS 运作时, 首先由出版商给图书馆授权, 图书馆利用网络爬虫一类的工具将目标资源下载到本地可存档的 LOCKSS 存档箱 (LOCKSS Boxes)。同时 LOCKSS 存档箱内的内容会持续不断地与其他 LOCKSS 存档箱中相同的内容进行比对, 以保证内容的正确与完整。又由于 LOCKSS 采用的是 Peer-to-Peer 对等方式的分布式保存系统, 一旦某个存档箱中的内容有误或者丢失, 它可以通过 LOCKSS 内部其他存档箱中的内容进行修复^[8]。在 LOCKSS 中, 每个专题只要有不少于 6 份的备份, 就能保证资源的完整与安全, 确保整个系统的正常运行。

当出版商由于网络崩溃、取消订购或出版商倒闭等不能够再提供服务时, 授权用户可以使用 LOCKSS 存档箱的信息。

2.2.2 Portico 存档采用的模式是迁移 (Migration), 即当原有的文件格式不再被使用时, 存档内容可以转换成其他经常被使用的格式。同时, Portico 还采用基于字节保存源文件的模式。

Portico 对电子内容的保存方法有一系列指导原则, 包括: 重视期刊学术内容的保存, 忽略期刊的外观 (如: 扉页、版权页、目次页等), 对出版商的网站或传播平台不予保存。Portico 对内容质量的保证采取对保存内容实行自审的方式, 同时接受图书馆、出版商审校, 但一般情况下不允许图书馆及用户访问已存储的资料。只有当出版商停止运作、停止某期刊的出版、不再提供回溯期刊、平台遭受破坏且资源无法再从其他源头获得时, 图书馆具有出版商的授权, Portico 允许图书馆通过采用 IP 控制或密码为用户提供所保存内容。在是否允许访问这一点上, LOCKSS 和 Portico 具有相同的要求。由于 Portico 是第三方电子资源长

期保存模式, 为证明其可靠性, Portico 承诺如果自身停止运行, 该组织会将已保存的内容转交美国国会图书馆或其他非营利性机构加以保存并继续提供服务。同时由荷兰国家图书馆线下保存 Portico 的全部资源, 以增强资源对外部影响的抵抗力^[9]。

2.3 费用

2.3.1 LOCKSS 系统是一个分布式的电子资源存储系统, 而作为一个分布式的系统必须制定一些关于软件开发与技术支持以及收藏管理方面的协调框架。LOCKSS 作为一个非营利的为图书馆和出版商服务的组织, 其系统开发的成本来自多项基金的支持, 并通过合作机制均摊开发费用。所做的工作包括: 软件的开发、修复漏洞、系统更新、安装与使用的技术支持、协调成员的合作, 等等^[10]。由于系统应用的是成熟的对等网组网技术和开源系统^[11], 其运行和开发成本较低。LOCKSS 与组织中成员共同参与规范相关政策, 要求每个成员为系统提供必要的支持。鼓励成员参与到软件开发的过程中, 并在开发中给予技术支持。通过成员的广泛参与, 做到成本、风险、利益均摊。

对于图书馆来说, 除了参与到 LOCKSS 中并尽相应的义务外, 图书馆自己建立的存储系统对硬件的要求非常低, 只要是普通的个人计算机, 就可以转化为一个数字信息存档箱。目前, 我国清华大学、中国科学院、香港大学、香港中文大学、香港科技大学和香港城市大学已参加了 LOCKSS 组织。而出版商加入到 LOCKSS 中是免费的。

2.3.2 Portico 作为第三方的电子资源存储机构接受了多方基金的支持。为了避免对单一资金的依赖, Portico 对接受其服务的图书馆和出版商采取收取年费的方式维持机构的正常运行。图书馆缴纳的费用取决于图书馆的年度采购经费 (Library Materials Expenditure, 简称 LME)。由于每个图书馆的年度采购经费差别很大, 因此 Portico 公布了不同年度采购经费的不同的保存费用, 原则上, 年度采购经费越高的图书馆, 其保存费占年度采购经费的比例越低。如当图书馆 LME 低于 15 万美元时, 其保存费为 LME 的 1%。当 LME 为 500 万美元时, 其保存费为 LME 的 0.2%, 当 LME 为 3000 万美元时, 其保存费为 LME 的 0.08%^[12]。出版商缴纳的费用取决于其年度期刊的订购收入, 包括电子版、印刷版期刊的订购收入、授权收入和广告收入等。原则上收入越高, 其缴纳的费用占收入的比例越低, 大致每年缴纳的费用在 250-75 000 美元之间^[13]。

3 讨论

3.1 项目运行时间和规模

从项目开始运行时间来看, LOCKSS 始于 1999 年, Portico 始于 2002 年, 都有 10 年左右的发展历程。

从运行规模来看, LOCKSS 中参与的出版社有 400 多家, 远远超过 Portico 的 75 家。当然 Portico 的 75 家中也包括了世界上主要的出版商。出版商的数量差别也许与 LOCKSS 对出版商免费进入的政策有关。尤其对于中小型

出版商来说,这种只需要给图书馆提供相应的保存文献的权利,不用花费额外的资金就将长期保存电子资源的任务赋予给图书馆的模式,还是很吸引人的。全球参加 LOCKSS 的图书馆超过 200 家,其中近 100 家为美国以外国家的图书馆,包括我国内地和香港地区的 5 家图书馆。Portico 则有 491 家图书馆,美国以外国家的图书馆有近 160 家。两个项目都以美国图书馆为主。出版商和图书馆是电子资源长期保存的两个重要方面,缺一不可。

3.2 项目运作方式

两个项目的启动资金都来自几种基金的支持。LOCKSS 采取的是开放性源码的分布式存储系统。即出版社提供保存的认证和许可,参与的每一个图书馆都要进行具体的存储工作。而 Portico 则是作为独立于出版商和图书馆之外的第三方组织来保存电子资源的。两者都是在不可抗拒的因素(如出版商变化或倒闭、资源网站受到恶意攻击、自然灾害等)影响下,才允许组织内的图书馆可以继续使用和索取存储的资源。

3.3 保存模式及可靠性

LOCKSS 采用分布式存储方式,每个图书馆收集的内容是得到授权的资源,通过与其他存档箱的比对来保证资源存储内容的正确性。由于系统内的全部备份数据损坏的可能性都极低,因此,参加的图书馆越多,其可靠性越有保证。Portico 是一个集中的资源存储库,存储的内容包括加入系统的出版商出版的所有资源,并且采用自审的方式。同时,图书馆和出版商也具有对内容审校的权利。

3.4 费用

LOCKSS 和 Portico 都有多项基金的支持,但是对于后续的维护与升级,LOCKSS 和 Portico 的方式不一样。LOCKSS 采用联盟机制,参加的成员将得到免费的后台管理、操作平台和工具插件软件等,各成员购置各自的存储设备,对于项目的开发费用采用均摊的方式,同时 LOCKSS 鼓励成员参与到软件开发中。Portico 则按照年度收取服务费,即每个加入的出版商和图书馆都根据保存的内容每年缴纳一定的年费。

4 结论

对比两种不同的存储模式,笔者认为,首先要得到出版商的支持,只有最大限度地保护出版商的利益不受侵害(即以最小的代价获得最大的利益),才能调动出版商参与的积极性。因此,LOCKSS 吸引大出版商参与、同时也让小出版商感兴趣的做法无疑夺得头筹。其次,LOCKSS 保存活动中不断与系统内相同内容的备份进行比对并不断修复的模式与 Portico 以自审为主的模式相比,保存内容的正确性具有更可靠的保证。第三,LOCKSS 通过图书馆直接参与电子资源的保存,将资源以较低的成本保留在本地,使图书馆从真正意义上拥有电子资源。这也比较符合图书馆传统的“收藏”职能。第四,相对于 Portico 每年需要缴纳的年费,LOCKSS 由于主要是自身的设备投入,图书馆更加容易

接受。当然,Portico 也有其优势,如避免重复性资源存储,节省人力、物力,符合社会化分工趋势等。

综上所述,电子资源的长期保存不仅涉及到介质的自然损耗,还涉及到技术的更新与淘汰、数据的更新与迁移等,这些都是绝大多数个体图书馆难以承担的,而图书馆一般又有长期保存电子资源的愿望。因此,笔者认为我国的高校、科研院所的图书馆及文献信息中心应该成立一个全国性的联盟,以类似于 LOCKSS 的模式,由国家投入一笔启动经费,由清华大学、中国科学院国家科学图书馆、国家图书馆等牵头单位组织软件开发、升级、维护,从而组成类似于 CALIS (China Academic Library & Information System, 中国高等教育文献保障系统)的核心领导小组,并由其负责与出版商的谈判和图书馆的组织、协调与培训等。各个图书馆都应积极参与自身文献的长期保存建设,笔者相信通过图书馆界同仁的共同努力,一定会推进我国电子资源长期保存的进程。

参考文献:

- [1] 全国图书馆界呼吁建设国家数字文献资源长期保存服务体系[J]. 现代图书情报技术, 2008(11):71.
- [2] 张 玫, 李 麟, 张晓林, 等. 中国图书馆数字文献资源长期保存现状调查[J]. 图书情报知识, 2009(2):47-51, 59.
- [3] LOCKSS>About Us[EB/OL].[2009-08-06].http://www.lockss.org/lockss/About_Us.
- [4] LOCKSS.Public VS Private LOCKSS Networks[EB/OL].[2009-08-06].http://www.lockss.org/lockss/Private_LOCKSS_Networks.
- [5] Portico>About Portico[EB/OL].[2009-08-06].<http://www.portico.org/about/>.
- [6] Portico. Participating Publishers [EB/OL].[2009-08-06].http://www.portico.org/about/part_publishers.html.
- [7] Portico. Facts & Figures: Portico at a Glance [EB/OL].[2009-08-06].http://www.portico.org/about/participating_libraries.html.
- [8] LOCKSS.How It Works[EB/OL].[2009-08-06].http://www.lockss.org/lockss/How_It_Works.
- [9] 丁艳君, 郑建程. Portico: 第三方数字资源长期保存模式初探[J]. 图书情报工作, 2009(2).
- [10] 王若琳. LOCKSS——实现网络电子资源的持久保存[J]. 图书馆杂志, 2007(2):58-60.
- [11] 黄田青, 陈清文. 多备份资源保存: 科技电子期刊长期保存的新模式[J]. 出版发行研究, 2007(7):16-20.
- [12] Portico.Annual Archive Support[EB/OL].[2009-08-17]. http://www.portico.org/libraries/aas_payment.html.
- [13] Portico.Annual Archive Contribution[EB/OL].[2009-08-17]. http://www.portico.org/publishers/pub_contribution.html.

[作者简介]

孙 艳 女, 1967 年生, 工作于第四军医大学图书馆, 副研究馆员, 研究方向: 文献资源建设。

马炳厚 男, 1951 年生, 工作于第四军医大学图书馆, 研究馆员, 研究方向: 文献缩微存储。

王 栋 男, 1970 年生, 工作于第四军医大学图书馆, 副研究馆员, 研究方向: 网络化信息服务。

[收稿日期: 2009-08-20]