

●刘家真

缴存本制度的扩展与电子出版物的采集

摘 要 论述了馆藏数字资源建设的重要性,探讨了电子出版物的采集原则、采集方法、保证电子出版物可采集的举措,以及将已有的缴存本制度扩展到电子出版物中的构想。参考文献 11。

关键词 电子出版物 缴存本制度 文献采集

分类号 G253.1

ABSTRACT In this paper, the author discusses the importance of the development of digital collections, principles and methods for the acquisition of electronic publications and a new idea to extend the existing legal deposit system to include electronic publications. 11 refs.

KEY WORDS Electronic publications. Legal deposit system. Document acquisitions.

CLASS NUMBER G253.1

现代信息产业中任何一种技术都围绕一个核心,就是将现有的各种信息数字化。承担人类知识和信息存储与传播职能的图书馆,在人类交流方式发生巨大变革时期,应当充分利用现代信息技术,在网络时空中重新定位,加大所存文献的数字化转换与电子出版物的采集与传送,逐渐演变成数字化图书馆,为公民提供通向国际与全国信息资源共享的有效通路。为此,图书馆必须开展大规模的数字化馆藏资源建设,使自己拥有独立的和规模宏大的馆藏数字化文献资源,才可能拓宽服务内容,使资源共享成为可能。数字化馆藏资源建设可以通过以下两方面进行:一是积极开展数字化项目,将那些具有长期研究价值与文化价值的传统文献转化为数字化文献,使现场用户与非现场用户均可存取到回溯性资料。二是广泛采集电子出版物,无论是脱机文献还是网上文献,以扩充数字馆藏。要使采集电子出版物的工作顺利进行,必须扩展已有的缴存本制度,使其覆盖到电子出版物。

本文将要讨论馆藏数字化资源建设中缴存本制度的扩展与电子出版物的采集。

1 电子出版物的缴存本制度设想

当今,数字化技术已经能十分方便地将模拟信息转换成数字信息,从而使越来越多的文献以数字方式产生与存在。以数字方式出版的电子出版物有:杂志、报纸;静态文献,如电子文本库中的文本;具有链

接的动态文献,如 HTML 网页;Listserv 上的用户组与讨论组等。同传统出版物一样,众多的电子出版物中不少具有学术价值、研究价值、史料价值与管理价值,因而应同属人类文化遗产。与传统出版物不同的是,许多电子出版物至今尚未有缴存本制度的保护,一些光盘出版物一旦失去了自身的商业价值或出版社倒闭,它们就难以幸存。网上出版物则处于更大的风险之中。网络的频繁更新,使网上出版物寿命十分短暂,如果不及时采集,就会流失。

电子出版物可划分为两大范畴:一是以光盘、磁盘、磁带等数字媒体脱机出版的脱机数字资源;二是在网络上出版的网上资源。网上资源又分为两大类:一类是可自行存取的网上出版物,一类是限定使用的网上出版物。可自行存取的网上出版物必须满足以下条件中的任一个:它们是可以通过标准的 HTTP 协议进行存取的文献,或是这些文献是被其他文献所引用、在其他文献中被参见的;在网上存取这类文献不需任何口令,也不需付任何费用。

并非所有的网上文献都可自由存取。受出版商控制的网上资源,就是不可以自由存取的,属于限定使用的出版物。例如:出版商在 CD-ROM 与软盘中预先设入安全设置,或使用特殊的驱动器或硬编码进行拷贝保护,CD-ROM 与 DVD 还受地区码限制,防止任意存取。这些措施在保护知识产权的同时,也限制了图书馆的合理利用。

为了完整地保存我国文化遗产,图书馆应借助缴存本制度确保电子出版物为了当前与今后的使用、研究而存在。然而至今,国家的出版法中也尚未对电子出版物的保存责任进行明确界定。这种情况下,国家图书馆与版本图书馆有责任建议国家将现有的缴存本制度扩展到电子出版物。

有些西方国家已制定相应法规或扩展了传统的缴存本制度,使电子出版物能缴送到指定的图书馆。为了在更大范围内实行电子出版物资源共享,1998 年 11 月欧洲的国家图书馆就开展了一个合作项目 NEDLIB (Networked European Deposit Library)^[1],以构建欧洲网上版本图书馆的基本框架,确保电子出版物现在与将来的保存与合理利用。为开展这一研究,图书馆与 IT 开发者、出版者就电子出版物管理的方法、标准与技术措施展开了广泛的合作研究。

为使我国电子出版物受到缴存本制度的制约,笔者提出如下一些建议。

1.1 国家图书馆应起领导作用

在数字环境中,国家图书馆不仅应领导数字技术在图书馆内应用,而且应当在保存与存取电子出版物的管理中起领导作用,与其他图书馆、有关公共机构、信息产业部门等建立广泛协作与沟通,应向国家建议建立电子出版物的缴存本制度,还应组建相应工作组,共同构思电子出版物缴存本制度的具体条款。

1.2 缴送对象

电子出版物的缴送对象应覆盖脱机出版物与网上出版物,凡个人或部门每出版一种电子出版物都必须抽出一定数量样本,缴送给指定的图书馆。

1.2.1 脱机电子出版物

像印刷出版物一样,国家图书馆应负责对已送缴的脱机电子出版物统一编目,并输入国家电子书目数据库内。出版者通过对该数据库的访问,可了解某一类电子出版物的出版情况与被读者利用次数,以判断同类选题的市场效应。这项工作无疑会受到出版者欢迎与支持。

1.2.2 网上出版物

对网上自由存取的出版物,有些国家规定,出版者应将已经产生、正在产生或准备产生的网上出版物按一定方式缴送给指定图书馆^[2];有些国家则认为这类出版物对出版者、图书馆与任何用户一视同仁,可以自由存取,没有必要受缴存本制度的制约,以减轻图书馆工作人员的负担。例如,芬兰于 2000 年 1 月开始执行的新的缴存本法明确规定,网上可自由存取

的出版物将不再列入缴存本法,而授权给国家图书馆借助机器人全面采集这类文献。可自由存取的网上出版物经采集后,应编入芬兰国家全文数据库索引内,以便公众使用^[3]。

在我国当前网络管理环境中,应将可自由存取的网上出版物列入缴存本制度,要求出版者必须将其缴送到指定图书馆,减少图书馆采集的盲目性。对限定利用的网上出版物,出版者也应承担缴送责任,应在网上出版物出版后立即向指定的图书馆缴送。

1.3 缴送本的格式

电子出版物的格式直接影响着出版物的现在读出与今后长期保存。如果缴送本因其格式不能被图书馆使用与安装,则为无用的资源或相当于没有缴送。因而,不少国家对缴送本的格式提出了要求。例如,芬兰国家图书馆规定:国家图书馆主张以 SGML、XML、JPG 等格式缴送文献,脱机文献可以用 DAT 磁带、CD-ROM 盘、DVD 盘等数字存储装置缴送^[4]。对以其他格式形成的文献,芬兰国家图书馆要求出版者将其转换成国家图书馆可以接收的格式,或将它们拷贝到适当的可被国家图书馆接收到的存储装置中。对于一些无法转换成缴存本制度所要求格式的出版物,芬兰国家图书馆认为没有缴送的必要。但对于那些确有特殊价值的电子出版物,应资助出版者开发出转换成标准格式的软件以供缴送。

对于缴存本的采集与保存,我们最关心的应是保留电子文献最初的编排与特征。使用某一指定格式对多种电子文献进行格式转换有可能导致信息内容的丢失或文献外观、特征等发生明显改变。如果按标准格式要求,势必有一部分文献无法接受;而不按统一格式对缴存本进行限定又会加大图书馆处理众多格式文献的难度。为此,笔者提出 3 点建议。

1.3.1 鼓励出版者以标准格式出版

国家图书馆、版本图书馆应与信息产业部门、出版者合作,共同制定当前技术环境中的电子出版物格式标准,并鼓励出版者以标准格式出版电子资料,以便出版物被引用传送并长期保存。

1.3.2 适度放宽缴存本格式

在尚未完全使用电子出版物格式标准化之前,应适度放宽电子出版物缴存本的格式要求与存储装置要求。为了能处理较多格式的缴存本,国家图书馆与版本图书馆应寻求有关的仿真软件与仿真硬件。如果处理电子文献的程序有利于图书馆对其进行标引,或有利于人们查询该数据库,图书馆必须对这些文献

的处理程序也加以存取。

1.3.3 分别存储

对于一些复杂文献,为避免以某一格式存储带来信息丢失,可采用不同格式进行分别存储。例如芬兰的学位论文、新闻报导等电子出版物,其文本部分以 SGML 格式存放在国家图书馆的全文数据库,而图像单独存储在原文描述(textual descriptions)数据库^[5]。

1.4 缴送方式

与传统文献不同的是,电子出版物可以以多种方式缴送,这与电子出版物的出版形式有关。

1.4.1 以电子出版物的实在资料寄给或传送给指定图书馆

对于脱机电子出版物,十分适合采用这种方式缴送。

某些网上出版物也可采用这种方式将出版物的实体资料传送给指定图书馆的服务器,以减少图书馆的工作负担。但这种缴送方式存在两大问题:一是在传送过程中可能被人修改,或发生欺骗行为。对这一问题可采用数字签名方式进行控制。数字签名既可以证实缴送者身份,又可以证实传送内容的确是责任者所指定的。二是将实体资料传送到指定图书馆的服务器上,有可能使国家图书馆、版本图书馆收集到一些部门或个人在网上出版的、与图书馆收藏范围毫不相干的内容。

1.4.2 将电子出版物的报导材料缴送给指定图书馆

这样做,使图书馆可在出版者的服务器中检索到这些资料,采集需要保存的出版物。丹麦 1998 年始生效的版权法修订本中,就要求以这种方式缴送版权电子出版物^[6]。

这种缴送方法的特点:一是便于图书馆对已出版的电子出版物进行检索;二是便于图书馆对缴送本的保存进行选择;三是出版者可以在缴送报导材料时,将该文献的著录信息附在其上,一并缴送给指定图书馆。这种缴送方式有利于出版部门与图书馆间的连接网络的安装,也便于图书馆更新服务器内的资料。

自由存取的网上出版物适宜使用这种方式进行缴送,但限定的出版物不宜采用这种方式。

在缴存本制度中可不必硬性限定其中一种;应针对不同的电子出版物,指定不同的缴送方式,使之既有利于缴存本的获取,又有利于图书馆工作的顺利开展。

1.5 缴存本的数量

传统的印刷出版物、视听资料由于无法通过网络进行存取,相应地需要缴送多个拷贝给指定图书馆。

对于脱机出版物可以仿照传统出版物方式,至少缴送两个拷贝,分别给国家图书馆与版本图书馆。网上出版物存储在档案服务器内,只要使用适当工具就可以从国家图书馆与版本图书馆获取,因而不必缴送多份拷贝。

2 电子出版物的采集与归档保存

采集是传输接收的数据到一个文件中以进行备份或者以后分析的过程。采集电子出版物是创建数字馆藏的一种方式,其目的是为了长期利用这些数字资源。缴存本制度是图书馆可以对电子出版物进行全面采集的保证。

脱机电子出版物的采集,早已成为图书馆的日常工作之一。网络出版物的采集随着互联网的广泛利用而逐渐为人们重视。网上资源数量极大,对这类资源进行采集与标引需要采用现代化手段,仅靠人工采集十分困难。

2.1 网上出版物的采集原则

脱机出版物的采集原则可参照传统出版物。

对于网上出版物国外有两种不同的采集原则。

2.1.1 分工合作,有选择地采集

澳大利亚、加拿大等国家图书馆主张有选择地采集电子出版物,优先采集有价值的网上出版物^[7]。澳大利亚国家图书馆与地方图书馆就各州电子出版物的保存达成协议,如果网上出版物不具有广泛的价值,将不由国家图书馆保存,而由州和地方图书馆行使保存职责。我国许多有关论文中,也认为应有选择地采集网上出版物,其选择原则应是:针对性、系统性、重要性与适用性等^[9]。

2.1.2 全面地不加选择地采集网上出版物

某些欧洲国家,如芬兰、瑞士等认为国家图书馆应尽量广泛地采集本国的网上出版物,对于有永久保存价值的海外网上出版物也应采集。原因是:第一,有价值与无价值的判断标准是建立在现在的认识基础上,将来的价值标准究竟如何,谁也无法预料。因而,为了避免损失,应全面采集本国的所有网上出版物。第二,计算机存储器容量越来越大,体积越来越小,价格一再下跌,使人们有可能大量采集。第三,面对成千上万的网上出版物,对其价值进行逐一鉴别与挑选,要耗费大量的人力资源,图书馆无法满足^[10]。

笔者认为,全面、不加选择地采集网上出版物有一定道理,但目前我国的自动化水平有限,暂时还无法做到,因而我国的网上出版物还是以有选择地采集

为好。

2.2 网上出版物的采集方法

2.2.1 有选择地链接或粘贴到国家图书馆电子出版物档案库上

加拿大国家图书馆采用的就是此种方式。

2.2.2 逐一下载

我国对已选定的网上出版物一般是以逐一下载的方式进行采集。

2.2.3 使用机器人进行全面地采集

北欧许多国家利用“收割机器人”连续地寻找与采集已编入索引的网页^[11]。收割机器人由采集、分析、标引与存档 4 部分组成,它被正式地长期固定在网页上。网页一旦被采集,机器人内的软件就会自动分析网上出版物的内容与交互引用(如超级链接情况),随后输出 URL 表。该表是依采集策略被存储的,以保证网址经常被核对。然后对该出版物进行自动标引,并将它打包压缩,存储到当地存储库的磁盘上。

使用机器人进行采集时,只要把本国国家代码的域名编入索引,凡具有本国国家代码域名的网上出版物基本上可以收集齐全。

利用 URL 采集文献的缺点是缺少文献版本的概念,在某一时期在特定位置上的文献,有可能与下一时期相同位置文献毫不相干。为解决这一问题,欧洲开始使用 URL(统一资源命名法)生成元(URN generator)来确定文献的版本,以便机器人采集。限定使用的网上出版物难以被机器人采集,这时就要在缴存本制度的支持下由人工采集。

2.3 归档保存

归档是将数据文件存储到与系统兼容的媒体上,进行长时间存储,以备以后万一出现数据库差错而必须加以恢复时使用。

为了数据安全,以磁盘、磁带为载体的脱机电子出版物与网上出版物均应拷贝到档案服务器内归档保存。档案服务器的制作与维护相当昂贵,不可能每个图书馆都安装它,但国家图书馆与版本图书馆必须安装它,以维护数据安全。DVD、CD-ROM 这类脱机出版物,可以复制多个同一媒体的备份,安装进网络内,以提供利用。

存储在档案服务器内的资料,安全十分重要。为了使存档出版物可被授权使用,必须将服务器连到网上。图书馆应采取技术措施与管理策略,防止对电子出版物的非法存取,或受到第三者操纵与篡改等。

参考文献

- 1 Nedlib Project: <http://www.konlib.nl/nedlib/>
 - 2 <http://www.nlc.ca/initiatives/index.html>
 - 3, 11 EVA: <http://hul.helsinki.fi/tietoinja/0199/evaart.html>
 - 4, 5, 6 Electronic publications as Legal Deposit Copies: <http://hul.helsinki.fi/tietolinja/0199/leqaldep.html>
 - 7 Positioning the National Library of Canada in the Digital Environment: Strategic Directions.
 - 8 刘家真. 澳大利亚国家图书馆保存网上出版物的选择方针. 图书馆杂志, 1998(3)
 - 9 叶继元. 电子期刊收集策略探微. 中国图书馆学报, 1998(5)
 - 10 Kulturarw Heritage Project. <http://kulturarw3.kb.se/se/html/projectdescription.html>
- 注 本文为武汉大学人文社会科学重大项目。

刘家真 教授, 武汉大学传播与信息学院图书馆学档案学系副主任。通讯地址: 湖北武汉。邮编 430072。

(来稿时间: 2000-04-13)

(上接第 31 页) 企业经营术. 北京: 改革出版社, 1993

9 杜占远, 刘东. 政府在中小企业技术创新中的角色和作用. 中国软科学, 1998(9)

10 Berenika M. Webster, Catherine E. Hare, Julie McLeod Records management practices in small and medium-sized enterprises: a survey in North-East England. Journal Informaion Science, 1999v25 No4(283~294)

14 傅建华, 韩文亮, 黄文灼. 中小企业发展战略研究. 上海: 上海财经大学出版社, 1998

16 李冬梅. 孵化器—孕育中关村希望. <http://www.beijingdaily.com.cn>

17 <http://www.ccpit.org/index.html>

注 本文是国家社科基金九五规划黑龙江省自然科学基金项目“支撑中小企业的信息管理与网络建设研究”(99BTQ001)的成果之一。

冷伏海 中国科学院文献情报中心在读博士生, 黑龙江大学信息管理系教授、副主任。发表论文 40 余篇, 出版专著和教材 8 部, 获教育部霍英东优秀青年教师奖。通讯地址: 北京中关村科学院南路 8 号。邮编 100080。

(来稿时间: 2000-06-20)