

●宛玲

国外数字资源长期保存的最新发展及对我国的启示

摘 要 在国外,数字资源长期保存的研究和实践活动都有了新进展:制订了标准规范,国家图书馆、高校图书馆和出版机构成为主要实施者,合作项目是发展重点,有多种运作模式。参考文献 2。

关键词 数字资源 长期保存 保存机制 资源建设

分类号 G253

ABSTRACT In this paper, the author introduces the latest developments of long-term preservation of digital resources in foreign countries, such as standards drafting, cooperation programs and new operation patterns, which can be learned by Chinese librarians. 2 refs.

KEY WORDS Digital resources. Resource development. Long-term preservation. Preservation mechanism.

CLASS NUMBER G253

数字资源长期保存在国际上已不是新话题,但具体实施并没有大范围展开,主要集中在几个发达国家的一些大型图书馆。了解国外的最新发展,对于国家数字资源长期保存机制尚未建立、数字资源长期保存活动尚未广泛开展的我国来说,具有重要的启示和借鉴意义。

1 国外数字资源长期保存的最新发展

1.1 标准规范的制定与引用

近年来 ISO 组织、美国国家信息标准组织 NISO 以及其他一些国家政府和民间机构都不同程度地研制了有关标准和规范,并在实践领域得到应用。这些无疑对开展数字资源长期保存活动起到了很好的规范和引导作用。

在数字资源长期保存领域,最有影响的标准当属开放档案信息参考模型 OAIS。1999 年美国空间数字系统咨询委员会提出了 OAIS,并提交 ISO 成为 ISO 标准草案。经过 3 年讨论修改,2002 年 1 月出版了 ISO 蓝皮书 CCSDS 650.0-B-1 版本,现已正式成为 ISO 标准(ISO 14721:2003.)。实际上在 OAIS 还是标准草案时,几乎所有大型的长期保存项目如英国的 CEDARS 项目等都将它作为实施标准。许多机构结合自身特点和需求,建立了基于 OAIS 的本地长期保存基础结构,如美国哈佛大学 1998 年开始的 5 年规划项目 LDI(Library Digital Initiative)。虽然 OAIS 是由美国空间数据系统委员会提出,但图书馆

界为 OAIS 的修改做出了相当大的贡献,如 2000 年欧洲网络存储图书馆项目 NECLIB 的成员与 ISO 工作组一起参与了 OAIS 的修订。目前美国研究图书馆工作组 RLG 正在进行一个 3 年期的项目(2002~2004 年),推广 OAIS 的应用。

目前还没有一个统一通用的数字资源长期保存元数据规范格式,许多长期保存项目建立了项目内部通用的元数据格式,如英国 CEDARS 项目、澳大利亚 PANDORA 项目、欧洲 NECLIB 项目等。但也有的元数据格式已在国际上被广泛引用,如美国国会图书馆制定的专门用于多媒体保存描述的 METS 元数据格式。近期最有影响的是由 RLG 与美国联机计算机服务中心 OCLC 联合成立的 OCLC/RLG 工作组所建立的标准元数据格式。2002 年 OCLC/RLG 工作组在研究了上述几大项目的元数据格式的基础上,发布了基于 OAIS 的数字资源长期保存元数据格式。目前 OCLC/RLG 又建立了由图书馆、学校、博物馆、政府、商业等领域代表组成的元数据研制工作组 II,计划用 1 年的时间,即 2003 年 6 月至 2004 年 6 月,执行一个称之为 PREMIS(PREservation Metadata: Implementation Strategies)的项目,其目的是进一步发展前工作组所制定的元数据格式,并在实践领域中推广^[1]。

其他相关的标准还有美国国家信息标准组织 NISO 制定的 Z39.87:数据字典—描述数字图像的技术元数据。实际上数字资源长期保存并不是一种孤



立的活动,数字资源建设中的许多标准如内容描述标准 DC、唯一标识符标准、XML 等都是工作中需要引用的。除了标准,还出现了许多如下具有重要指导意义的文件。

(1)RLG 与 OCLC 合作,于 2002 年发表了《可信赖存储体的属性与责任》(*Trusted Digital Repositories Attributes and Responsibilities*)。该报告从可信赖者定义、应具备的基本属性、应承担的责任、选择与评价标准、选择评价和建立可信赖者的规范流程,以及为可信赖者数字资源长期保存工作有效长久进行应提供的相应支持等 6 个方面,界定了符合 OAIS 要求的数字资源长期保存系统应该具备的基本条件和责任体系^[2]。

(2)2000 年 12 月美国国会通过了一项立法,即建立国家数字信息基础结构和保存规划 NDIIPP (National Digital Information Infrastructure and Preservation program),并责成国会图书馆联合主要的国家机构,在数字内容存档的收集和管理方面有实践经验的团体,以及其他致力于数字格式工作的私有机构等,带动全国开展长期保存数字内容的活动。NDIIPP 定义了数字保存基础结构(digital preservation infrastructure)。此结构确定了数字资源长期保存的基本框架,规定其功能是用来支持长期的多种信息交流需求,能够应对快速的技术发展以及保持透明性和可信程度。基础结构包括数字保存网络和数字保存结构两大部分。2003 年初美国国会又通过了国会图书馆的 NDIIPP 实施计划。NDIIPP 的设计思想和方法对各国都有指导性作用。

(3)2003 年 4 月 OCLC 发布了关于数字资源长期保存经济激励因素分析报告(*The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*)。该报告将长期保存涉及的利益者分为权利拥有者、存储者和受益者,并据此将长期保存运作模型分为向心模型、离心模型、单边供给模型、单边需求模型和统一模型。

(4)2003 年澳大利亚国家图书馆受联合国教科文组织信息社会部的委托,起草了 UNESCO 数字遗产长期保存指南(*The UNESCO Guidelines for the Preservation of Digital Heritage*),该指南提出了 39 点长期保存的基本原则。

1.2 国家图书馆、高校图书馆和出版机构等成为保存运作的主要实施者

对于印本资源的长期保存,各国都有完善的缴

存本制度,有法律支持的规范的责任体系。但对于数字资源,尤其是网络资源的保存,目前各国还普遍缺乏法律上明文的规定。尽管如此,传统的保存机构——图书馆,在数字资源保存活动中,责无旁贷地成为了主要实施者,同时出版机构和一些非营利信息服务机构也发挥了相当的作用。

1.2.1 国家图书馆

澳大利亚国家图书馆,早在 1996 年就制定了《澳大利亚电子出版物的国家策略》,当时的重点是放在光盘文献上,随后启动了关键的数字保存项目 PANDORA (Preserving and Accessing Networked Documentary Resources of Australia),目前已经建设成用户可以直接访问的国家联机出版物知识库。澳大利亚图书馆认识到长期存储因特网上重要的联机出版物绝非单个图书馆可以完成的,因此联合了 7 家州级图书馆形成了分布式的保存网络——国家模型,PANDORA 站点成为其重要的组成部分。

美国国会图书馆已开始收集同时代的数字形式的文献。2003 年初国会图书馆通过了 NDIIPP 实施计划,后者获得 1 亿美元来发展、设计和实现国家的数字资源长期保存基础结构。

荷兰国家图书馆早在 1994 年就将保存电子出版物作为其重要职能之一。1998 年参加了欧洲国家图书馆合作项目 NEDLIB,积极研究相关机制。与 IBM 公司合作,从 2001 年起开始建立电子出版物存储系统 E-DEPOSIT,最新系统于 2003 年 1 月启用。1999 年与荷兰出版协会 NUV 达成了关于存储联机 and 脱机出版物的一个总协议,为国家图书馆开展存储电子出版物活动奠定了基础。2002 年 8 月和 2003 年 5 月,分别与 ELSEVIER SCIENCE 公司和 KLUWER 公司正式签署长期保存其联机期刊和电子图书的协议,这是一种典型的建立国家 Archive 的行为,最大程度地保障了国家的公共利益。

英国、法国、加拿大等国的国家图书馆,以及一些专业性公共图书馆,如美国医学图书馆等也都在其数字图书馆建设中不同程度地开展了数字资源长期保存活动。

1.2.2 高校图书馆

网络上的海量数字资源,最具保存价值的应当是学术性资源。而这些资源通常来源于和应用于高校,因此高校图书馆成为长期保存活动中不可忽视的主力。比较有典型代表意义的有以下图书馆。

加利福尼亚大学建设有世界著名的数字图书馆

CDL。CDL目前正在实施 E-SCHOLARSHIP 计划,使其成为全校数字资源保存中心。在该计划中,CDL在各学院图书馆、研究机构、实验室以及博物馆建有节点,为用户提供本地化专门的保存服务,同时依据 CDL 分布式系统结构支持公共保存需要。该系统保存的内容包括其员工所创造的数据集、预印本以及其学院图书馆不收藏的资料。

麻省理工学院图书馆与美国惠普公司合作,于2002年底建立并正式启动了新型多媒体数字图书馆系统 Dspace,一个获取、传播和保存该校智力产出的数字图书馆。该系统具备处理该校教师和研究人員每年完成的总计1万多份数字化科研成果的能力。本学校的员工按照 Dspace 提供的提交页面或直接与图书馆工作人员联系,将自己的成果数字化并长期存储在该系统中;最终用户可以通过它访问和查询图书馆收藏的论文;而研究单位可以对本单位的数字化科研产出进行统一的管理和使用。目前该图书馆正在运用 Dspace 概念发展研究图书馆 DSpace 联盟。

英国剑桥大学图书馆,与本学校的计算机服务中心合作,加入了 DSpace 联盟启动项目,正在建立 'DSpace@Cambridge' 存储库。该项目从2002年底开始,计划到2005年建成。此系统将存储剑桥大学图书馆自己数字化的资料和本校其他机构创造的数字资源。该馆还参加了已经结项的 CAMiLEON、CEDARS 等长期保存联合项目。

此外还有哈佛大学、斯坦福大学等大学的图书馆开展了数字资源长期保存的活动。哈佛大学有100来个图书馆,每个图书馆都有责任选择需要保存的资源。斯坦福大学图书馆建立了一个长期保存自产和外购数字资源的数字知识库,还开发了 LOCKSS 系统,致力于解决电子期刊的保存和利用的相关技术问题等。

1.2.3 出版商

在数字环境下,出版商也开始涉入保存领域,而且在保存方面起到关键的作用。对于数字资源,出版商通常仅提供给各订购者对数字资源的使用权。在这种情况下,出版商无疑就需要承担对数字资源的维护任务。但是否提供长期保存,并没有普遍的做法。目前有不少数据库商提供若干年的回溯数据,在合同期内他们必须要保证这些资源的可使用性,如 ELSEVIER、EI、IEEE 等公司都保证用户可以回溯检索其一定时期的出版物,这种保证是其商业运作的一部分,具有一定的保存意义,但还不能说是

一种长期保存行为。在订购方越来越重视数字资源长期存取的发展趋势面前,不少出版商开始寻求可靠的第三方承担长期保存的责任。也有的出版商应要求给予订购方对所购资源的永久使用权,实施了一种长期保存的行为。

1.2.4 非营利信息服务机构

并非所有大学甚至包括一些著名大学都提供长期保存服务,许多出版商既不承诺进行长期保存,也没有选择第三方实施或授权与订购方,而目前社会上又没有一种提供这种服务的机制,因而许多数字资源找不到长期保存的归宿。在这种背景下,一些非营利机构开始了有益的尝试。最有名的主要有美国 OCLC 和 RLG,他们正在为其成员单位开展长期保存成员单位数据的服务。

1.3 合作项目仍是发展的重点

在数字资源长期保存中,很难看到单一机构独自进行,多方合作是主要模式。这种合作不仅仅是指图书馆与出版商、技术公司之间的合作,更主要的是直接进行长期保存机构之间的合作。合作通常以项目形式出现。

(1) CEDARS (CURL Exemplars in Digital ARchiveS),1998年4月开始,2002年3月结项。由英国信息系统联合委员会 JISC 资助,牛津大学图书馆、剑桥大学图书馆、里兹大学图书馆参加。其成果是研究了数字资源长期保存的战略框架和具体方法,并建立了一个分布式的长期存储系统。

(2) NEBLIB (Networked European Deposit Library),1998年1月启动,2001年1月结项。荷兰国家图书馆牵头,欧洲7个国家图书馆(荷兰、法国、挪威、德国、葡萄牙、瑞士、意大利)以及3个主要出版社 KLUWER、ELSEVIER、SPRINGER-VERLAG 参与,以合作项目形式研究长期保存基础结构。其目的是研究欧洲存储图书馆机制,以保证电子出版物的长期获取。成果主要有:建立了一种电子出版物长期存储系统模型;研究了实际应用的最佳实践、技术标准以及运作方法;研制了建立存储系统的小型工具包,元数据搜索程序等。

(3) JISC 下的 DPRM (Digital Preservation and Records Management)项目。JISC 是英国一个支持教育科研的系统,实施了多个相关数字信息项目,其中就包括数字信息长期保存 DPRM 项目。DPRM 启动于2000年6月1日,计划2006年5月31日结项。

(4) ERPANET (The Electronic Resource Preserva-

tion and Access NETwork)项目,启动于2001年11月,计划2004年10月结项。该项目主要由欧洲委员会资助,由来自英国、荷兰、意大利和瑞士的学术性机构合作进行。ERPANET的目的是建立一个可扩展并能自我持续的欧洲启动项目,成为保存文化遗产和科学数字对象领域的虚拟数据交换中心和知识存储库。

还有前面提到过的美国NDIIPP、LOCKSS、PRESTO项目(一个致力于发展长期保存视听资料的欧洲合作项目)等正在进行中。

1.4 多种运作模式

(1)国家的资助。数字资源长期保存如果以商业模式运作,必须具有足够的商业激励因素存在才可以驱动团体和个人从事这项活动,形成相应的市场。但在目前,这种激励因素远不能超越数字资源长期保存巨额的费用以及可持续发展的不确定性带来的压力。从国家利益、长久利益着眼,国家的资助是必须和现实的。如美国启动的NDIIPP,获得了9980万美金的国家资助,其他如荷兰国家图书馆的许多项目、澳大利亚的PANDORA等都得益于国家资助。

(2)其他途径的资助。也有许多数字资源长期保存是由各种基金会、协会等资助的。如英国CEDARS项目是由JISC资助建成;NEBLIB、ERPANE和PRESTO项目是由欧洲委员会资助。最为突出的是MELLON基金会资助了许多大型图书馆在本地建立E-journal存储库。

(3)出版商转移存储权给特定订购方。如2000年ELSEVIER公司与YALE大学图书馆签订了后者长期保存前者期刊全文数据库数据的协议。这种模式还包括出版商许可订购方对所订购数字资源的永久使用权。如2002年9月JISC通过了数字期刊订购许可模式,该模式确定出版商应当通过许可协议,允许JISC建立本地服务的拷贝版,并在双方停止订购商业活动后,允许JISC对以前所购置的数字资源的永久使用。

(4)出版商转移存储权给第三方。如JSTOR, JSTOR是一个提供过期期刊检索服务的非营利组织。它首先从出版商那里获得长期保存以及提供检索服务的许可,然后用户付费从JSTOR订购ON-LINE检索服务。这些期刊内容的版权仍然属于出版商。出版商可以中止协议或不再续签,但以前已经签过的内容JSTOR仍然可以为用户提供服务。为了保障出版商和版权者的利益,JSTOR采取了一种

“活动墙”方式,即只有出版了2~5年以后的期刊才可进入JSTOR系统中。在这种运作模式中,长期保存费用通过提供检索服务实际上转嫁到用户身上,对出版商来讲,自己免去了保存的责任,还可保持现有正常的收入渠道。

(5)出版保存一体。典型的例子是American Geophysical Union。它长期保存其出版的所有内容,保存费用是由读者和作者承担:作者按页码缴纳保存费用,读者通过订购使用权缴纳使用费。

(6)经费使用重新分配。许多国家通过对图书馆经费的使用重新分配,已将数字资源长期保存费用从项目经费支持转到图书馆的正常经费使用,目的是使保存工作真正长期正常地维持。

2 对我国的启示

(1)积极推进建立国家数字资源长期保存战略机制和基础结构。建立国家数字资源长期保存战略机制,有利于在全国范围内有组织有计划地构建一个国家数字资源长期保存协作网,最全面、最经济、最合法、最有效地实施对国内外重要数字资源的长期保存和使用。建立国家数字资源长期保存战略机制,需要联合各部委、各省市的信息管理部门和新闻出版部门等,构建一个国家的数字资源长期保存体系结构,从法规、经济、方针政策和行政管理等方面来支持和约束。

(2)呼吁、研究与实践并行发展。数字资源长期保存制约因素主要涉及法律、经济和管理体制等方面,如缺乏类似印本缴存制的法律支撑机制,缺乏全国性的长期保存管理体制,缺乏规范持久的经济支持模式等。我国应当在这些方面充分进行研究,并尽早建立有关的法规和体制。这些问题不是短时期内可以全面完善解决的,应当学习国外经验,积极开展具体的实践活动,为形成全国性乃至国际间的协作网打下基础。

(3)建立不同层次的合作机制。一是国际合作。包括标准规范的制定、问题的研究、具体实践等,参与国际合作是我国吸收和紧跟世界最先进思想、理念、技术等的有效措施,也是向世界展现我国发展的窗口。二是国家级别机构之间的合作。如教育部、国家图书馆、中国科学院、科技部等,是我国学术性数字资源收藏最多的几个国家级别机构,相互之间应当也可以进行协调,技术上合作攻关、资源保存上合理分工、经费支持上实现最佳分配等。三是根据

●王芬林

关于全国文化信息资源共享工程数字 资源建设的思考和对策

摘 要 文化信息资源是一个范围很广泛的概念。就力所能及而言,全国文化信息资源共享工程的资源建设应以文化部可控制资源和实施方案中既定的资源建设目标为内容。工程应着手建设数字化图书馆,可为实现其数字图书馆服务的终极目标奠定基础。在资源共建模式和选题策划上,也都应有科学的规划。今后的全国文化信息资源共享工程建设应在总体思路、建设目标、技术支持、知识产权配合、组织实施、用户服务、工作流程等方面进行调整和改进。

关键词 文化信息资源 全国共享工程 资源范围 共建模式 发展对策

分类号 G253

ABSTRACT Cultural information resource is a broad concept. Practically speaking, the development of digital resources for the National Cultural Information Resources Sharing Project should have the contents of resources controllable by the Ministry of Culture and resources clearly specified in the plan. The project should be focused on the development of digital libraries and to realize its final objects of services via digital libraries. In this paper, the author proposes some recommendations concerning the overall plan, objectives, technical supports, intellectual property, implementation, customer services, work flows, etc.

KEY WORDS Cultural information resources. National Cultural Information Resources Sharing Project. Scope of resources. Pattern of cooperation development. Development strategies.

CLASS NUMBER G253

数字资源建设是全国文化信息资源共享工程(以下简称共享工程)的建设核心,是工程的命脉,这已是共识,无需再来论证。笔者从2001年10月开始共享工程试验系统研发至今一直负责资源建设工

我国机构特点,采取纵向——行业内、横向——地区内双管齐下的合作,建立分布式存储系统。四是形成上下合作链,即个人与机构进行合作、基层机构与省级机构合作、省级机构与国家机构合作、国家机构与国际机构合作等。当然这不是固定的模型,基层机构如一般大学图书馆在有条件时也可直接参与国际合作,但从总体上形成这样的合作链有利于层层带动,也有利于为将来的国家长期保存体系结构的构建打下基础。

(4)积极推进学术性资源永久使用权的建设。我国各类图书馆近几年购置了大量网络数据库。如CALIS目前已组团引进了近20个厂商的90多个数据库(含电子刊),参加的高校遍布全国各地;中国国家科学数字图书馆项目管理中心在中国科学院院内组织引进了10多个网络数据库。对于网络数据库,使用权与拥有权通常是分离的,一旦终止订购,或受社会不可抗力因素影响,订购方都有可能瞬间就失去了所有资源。在目前法律机制下,订购方一定

要积极向商家争取所购置数字资源的永久使用权。永久使用权的获取是国家长期保存战略机制建立的基础、实施的重要组成部分。

另外还包括建立引进、借鉴和研究有关标准规范的机制,建立责任分配与认证机制,加强相关法律、经济运作模式、管理模式等问题的研究,积极开展自建数字资源的长期保存活动等。

参考文献

- 1 A Metadata Framework to Support the Preservation of Digital Objects Final Report (June 2002). http://www.oclc.org/research/pnwg/pm_framework.pdf
- 2 Task Force on Digital Repository Certification. <http://www.rlg.org/longterm/certification.html>.

宛玲 河北大学管理学院副教授,中国科学院文献情报中心在读博士生。通讯地址:北京中关村北四环西路33号。邮编100080。(来稿时间:2003-08-22)