

● 刘家真

我国数字资源保存状况调查*

摘 要 调查表明,我国文献保存部门的数字资源,大多数数据在无控状态下产生,忽视保存数据有效读出的维护,欠缺维护数字信息资源的相关知识。应加强宏观政策的导向和微观的维护指导。表 16。

关键词 数字资源 保存 图书馆 档案馆 信息中心

分类号 G253

ABSTRACT Based on a survey of preservation of digital resources in China, the author thinks that most data are created in uncontrolled conditions without the preservation of effective output mechanisms. There is a lack of related knowledge concerning the maintenance of digital information resources. The author recommends putting emphasis on the macroscopic policy guidance and microscopic maintenance guidance. 16 tabs.

KEY WORDS Digital resource. Preservation. Library. Archive. Information center,

CLASS NUMBER G253

随着我国电子文件、电子出版物与各类网络资源以及宽带通信的普及,越来越多的图书馆、档案馆与信息中心积累与保存数字资源,并通过网络为社会服务。IT 技术的迅猛发展,使数字信息的弱点暴露无遗。今天,数据的脆弱性与它的便捷性一样,已越来越多地被人们认识。在这种背景下,数字资源的管理者是如何对待蕴含人类文化遗产的数据的,传承数字信息资源的过程中存在哪些重大隐患,我们该及时地采取哪些应对措施避免重大损失?针对这些问题,我

们在图书馆、档案馆和信息中心展开了调查。

1 调查对象

调查分为 3 个部分:数字资源类型、数字资源保存状况和数字资源的灾备措施。调查表散发到我国图书馆、档案馆与信息中心,收到 57 个单位的回函。这 57 个单位散布在 14 个省市(见表 1),大多数为高校图书馆、城建档案馆和综合性档案馆(见表 2)。

表 1 数据来源的地区分布

(%)

省市	北京	上海	天津	黑龙江	江苏	浙江	福建	江西	山东	河南	湖北	广东	四川	贵州
比例	29.80	5.40	1.70	15.80	5.40	3.50	1.70	7.00	1.70	1.70	14.10	8.80	1.70	1.70

表 2 参加调查的机构类型 (个,%)

机构类型	数量	所占比例
大学图书馆	20	35.09
大学档案馆	5	8.77
企业档案馆	2	3.51
城建档案馆	11	19.30
综合档案馆	10	17.54
政府部门图书馆、档案馆	1	1.75
信息中心	8	14.04
总计	57	100.00

对各类档案馆的调查使用了同一调查表。在对图书馆与信息中心调查时,对调查表作了细微修改,以便可能提供对某些问题更为清楚的答复。

2 调查发现

与国内众多的文献收藏机构相比,所采集到的数据可能不够全面。但这些信息已经凸现了我国数字资源保存中存在的主要问题。

* 本文为国家自然科学基金课题“中国文化数字资源保存库的国家战略研究”(编号 70373048)研究成果之一。

2.1 我国数字馆藏资源复杂

我国数字资源馆藏的复杂性主要表现在:数字馆藏类型较多(见表3),馆藏数字资源格式复杂(见表5、表6、表7、表8),数据存储介质多样(见表9)。表3是按系统对我国图书馆、档案馆与信息中心的馆藏类型的宏观统计。由表3可见,数字图片、数据库和多媒体在我国以上3种机构中均有收藏;所有的图书馆和信息中心都收有数据库,绝大多数的机构都有数字图片,档案馆保存多媒体馆藏没有图书馆与信息中心普遍。图片数据收藏的普遍与近些年来我国馆藏文献的数字化有关。

表3 馆藏类型 (%)

类型	数字图片	数据库	多媒体
档案馆系统内部	75	60.71	46.43
图书馆系统内部	100	100	100
信息中心系统内部	82.5	100	62.5

表5 数据格式的选用 (%)

类型	数字图片				数据库		多媒体			
	jpg	tif	pdf	其他 ^①	Sql-server	其他 ^②	mp3	avi	wav	其他 ^③
图书馆	80.95	66.67	80.95	61.90	95.24	71.43	57.14	57.14	47.62	61.90
信息中心	62.5	62.5	75	12.5	62.5	62.5	75	12.5	62.5	62.5

注:①图片的其他格式包括 GIF、方正与超星等。②数据库的其他格式包括 sybase、mysql、oracle、access 与 包库等。③多媒体的其他格式还有 rm、rmvb 等。

虽然各机构储存格式选用复杂,但主流格式开始浮出,如更多的图片选用 jpg、tif 与 pdf 作为储存格式;数据库更多选用 Sql-server;多媒体选用 mp3、avi、wav 为多。

我们还对同一单位内同类资源格式的选用状况进行了分析(见表6、表7、表8)。同一单位内同类资源采用多种格式储存的情况较为普遍。这可能与馆藏资源的管理目的不同有关(如用于保存、传输或检索等),也可能与管理不规范有关。

表6 图片储存格式选用状况 (%)

格式种类	仅1种	采用2种	采用3种	采用3种以上
图书馆	9.25	23.81	28.57	38.1
信息中心	12.5	25	0	50

表7 数据库储存格式选用状况 (%)

格式种类	仅1种	采用多种
图书馆	28.57	71.43
信息中心	25	75

表4是一个单位内所拥有的数字馆藏种类。许多图书馆都同时具有以上3类馆藏。近一半的档案馆只有两类数字馆藏,而大多数信息中心三类馆藏均有。

表4 一单位内数字馆藏类型的比例 (%)

类型	仅有一类馆藏	仅有两类馆藏	三类馆藏都有
档案馆	14.28	42.86	21.43
图书馆	0	0	100
信息中心	12.5	0	62.5

2.2 存储主流格式初见端倪

在对图书馆与信息中心的调查中,我们增加了对数据储存格式的调查。有21家图书馆和8家信息中心提供了相关数据。储存格式选用的种类较多,主流格式初见端倪(见表5)。

表8 多媒体储存格式选用状况 (%)

格式种类	仅1种	采用2种	采用3种	采用3种以上
图书馆	28.57	28.57	28.57	14.29
信息中心	0	40	0	60

2.3 储存介质多样化

表9说明两个问题:一是图书馆、档案馆的数据储存在多种介质上,包括光盘、软盘、磁带、硬盘和其他;二是大多数图书馆与档案馆都保存了大量光盘。需要说明的是,图书馆收藏的大多数光盘是电子出版物,其储存介质是出版商提供的,大多为 CD-ROM;档案馆收藏的光盘则大多是 CD-R,一般是档案馆将馆藏档案数字化数据刻录在光盘上,储存介质的选择权大多数掌控在档案馆。

表9 数据储存介质选用状况 (%)

储存介质	光盘	软盘	磁带	硬盘	其他
图书馆	95.24	47.62	61.90	100	23.81
档案馆	89.29	10.71	53.57	50.00	3.57

2.4 保管状况并不乐观

表 10 为参加调研单位的数据储存介质保管条件宏观统计。我国大多数数字馆藏的保管条件是不错

的,绝大多数单位的数据存储介质处于单独保管、并配置有环境温湿度控制与防火、防水、防盗设备的环

表 10 图书馆、档案馆与信息中心保管数字馆藏环境 (%)

存放情况	与其他馆藏同处	单独管理	有空调	有防盗	有防火	有防水	有其他防护
所占比例	1.75	92.98	87.72	84.21	85.96	71.93	26.23

表 11 以机构为单位,对图书馆与档案馆的数据馆,至今仍然还有馆藏数据处于较差环境中。储存环境进行了比较分析。无论是图书馆还是档案

表 11 数字资源存储介质保管条件 (%)

环境	与其他馆藏同处	单独管理	存放处有空调	有防盗	有防火	有防水	有其他防护
图书馆	4.76	90.48	90.48	85.71	80.95	52.38	19.05
档案馆	0	92.86	82.14	78.51	85.71	82.14	21.43

以上统计是以机构为单位进行宏观比较,不能说明每个单位的保管条件,因为“单独管理、有空调、有防盗、有防火与有防水”等 5 项措施中,是否所有单位都同时具有,表 11 反映不出。为了更为清晰地观察到数字资源的保管环境,表 12 以馆为单位统计了图书馆、档案馆的数字资源保管状况。经过进一步的数据分析发现,大多数单位 5 项保护措施全部到位,但仍然有任何防护措施也没有的单位(占调研数量的 1.75%)。档案馆比图书馆更加关注数据存储介质的保管条件,绝大多数单位是 5 项防护措施俱全,但仍然有少数档案部门仅有 3 项或 2 项防护手段。

表 12 图书馆与档案馆数据存储介质保管条件比较 (%)

防护措施	共 5 项	共 4 项	共 3 项	共 2 项	共 1 项	0 项
档案馆	78.57	10.71	7.14	0.00	2.57	0.00
图书馆	61.90	23.81	2.76	4.76	2.76	4.76

2.5 数据存储介质物理损伤严重

数据存储介质物理损坏带来的数据丢失往往是无法恢复的。保管条件仅仅是维护其形体完好的一个方面,保管方式不当、维护措施不力等都可能带来数据存储介质的物理损伤。所有参与调研的单位都给出了相关数据。分析表明,图书馆、档案馆与信息中心都存在数据存储介质物理性损伤,而且比例不低,其中尤以图书馆数据存储介质损伤更为严重(见表 13)。数据存储介质的损坏在我国是相当普遍的问题,

图书馆数据存储介质损伤严重可能与其利用率高相关,但这完全可以通过管理减少或避免;档案馆的数据存储介质损伤达 32%,必须引起高度重视。

表 13 数据存储介质物理损伤统计分析 (%)

机构	档案馆	图书馆	信息中心
损伤比例	32.10	71.40	25

2.6 数据无法读出并不罕见

保存数字资源的目的是为了长期、有效地获取信息。调查发现,除信息中心外,图书馆和档案馆均有数据难以有效读出的问题(见表 14),其中图书馆问题较为严重。

表 14 馆藏数字资源物理状况分析 (%)

机构类别	介质物理损伤	介质无法读出	介质已经过时
档案馆	32.10	14.30	10.70
图书馆	71.40	47.60	33.30
信息中心	25	0	12.50

图书馆数据存储介质损伤严重,与其利用率高及维护不当有关;介质过时比例大,与图书馆收藏数字资源历史较长相关;介质有效读出率低,与以上两个因素都息息相关。针对数据存储介质物理损坏、技术过时、数据无法有效读出等不良状况,我们逐一类机构的每一单位的不良状况同时存在比率做了进一步分析。发现以上 3 类不良状况同时发生比例最高的在图书馆(见表 15)。

表15 馆藏资源不良状况存在分析 (%)

机构类别	一类	二类	三类
档案馆	35.70	10.70	0
图书馆	28.60	19.00	28.60
信息中心	37.50	0	0

2.7 数据备份管理环节薄弱

图书馆、档案馆与信息中心的绝大多数数字资源

是我国的文化瑰宝,这类资源大多具有长久的价值,是应为当代人和后代人而加以保护和传承的文化遗产,应当加强风险意识与灾备管理,尽最大可能将人为与自然的不测之灾带来的毁灭性损毁降到最低。

备份是为数据制作一份或多份拷贝,通过数据的拷贝减小数据损毁的几率,降低因某储存技术过时而带来数据不可读的风险。我们对3类机构的数字资源备份状况进行了调研(见表16)。

表16 备份措施统计 (个,%)

类别	总计数 (共57个)	占总体比例	图书馆 (共21个)	占图书馆 系统比例	档案馆 (共28个)	占档案系 统比例(%)
未采取任何备份措施的单位	4	7.02	1	4.76	3	10.71
同时采取离线备份与其他备份措施的单位	19	33.33	7	33.33	9	32.14
非离线数据备份单位	16	28.07	11	52.38	2	7.14

由表16可见,仍然有少数单位的数据并未备份,尚有少数单位采取在线备份手段。

上,无异于保留垃圾,毁灭信息。诸多现象表明一个问题:忽视数据的有效读出管理。

3 我们的思考

调查数据显示,我国数字资源在保存问题上存在三大隐患。

3.1 大多数数据在无控状态下产生

绝大多数单位,一般都是先有了数字对象,并依赖其内容提供服务,在服务过程中遇到问题才考虑到数据的安全与维护。很少有数字对象是在适当规范的控制下产生的,并在利用、保存等环节得到事先的关怀。因而造成数据储存格式复杂、储存介质稳定性差(含技术稳定性与物理稳定性)。这使得须以数字形式传承的信息必须进行再处理,增加了数字信息保存的难度与成本。

3.2 忽视保存数据有效读出的维护

由表10可知,我国数字馆藏的储存环境受到一定程度重视,大多数图书馆、档案馆都在保存数据的环境方面进行了硬件投资。但在维护数据读出方面却未受到重视,例如图书馆、档案馆尚未将软盘数据迁移到稳定的介质上,仍然有图书馆与档案馆保留着软盘(见表9)。又如,保留形体损伤的数据储存介质(见表14)而不将其信息及时更新或迁移到新的介质

3.3 欠缺维护数字信息资源的相关知识

许多调查数据说明,我国文献管理部门欠缺维护数字信息资源的相关知识。

尽管大多数图书馆、档案馆和信息中心都注意改善数据储存环境,但数据储存介质损坏仍十分普遍,这大多与操作、保管和日常维护有关。储存介质是脆弱的,但许多损伤可以避免或是可以通过科学管理减少。其原因,除管理不善,更多可能就是缺乏基本的维护知识。介质损伤后不及时更新以抢救信息,而是保留受损的介质,不是缺乏对可能永久丢失数据后果的认识,就是责任心缺乏与管理不力。

数据储存技术过时并不少见,过时技术的数据并没有及时迁移到新一代储存介质上。这不能简单地归结于缺乏相关技术或技能,关键是对过时技术造成的数据无法读出的后果缺乏认识或缺乏责任心。如果认识到可能永久失去这些信息的严重后果,就会寻求技术去解决问题,国内有相当多的数据恢复公司就可以承担这类业务。

主要职能不同的图书馆和档案馆大都选用光盘储存数据。人们究竟是选择了光盘的哪些特点而放心地储存数据呢?如果仅仅将光盘用于数字馆藏常

用数据的拷贝,还可以理解;如果将它作为可靠的储存器就大大错误了。究竟该如何选择数据储存介质,似乎还缺乏清醒认识。

有的图书馆或档案馆并没有进行数据备份,这说明人们对数据的脆弱性还缺乏认识,或欠缺对数据的安全管理。

以上问题的存在,表明我国在数字资源的维护上,最缺乏的是宏观政策的导向和微观的维护指导。

随着大量信息将主要以数字方式产生,传承人类文明的图书馆、档案馆必须加强数字资源维护,以应对数

字时代的挑战。然而,维护数字资源的传承并非是国家文献管理部门一家可以独撑的。依据我国国情,应由国家信息化办公室牵头,组成相关的政策制订小组,以数据生命周期为依据,制定一套管理方案,使数字资源从产生、利用、维护到保存,整个生命周期都有适当的管理政策指导与控制,以减少数字资源损毁的风险与传承数字信息资源的成本。

刘家真 武汉大学信息管理学院教授、博士生导师。通信地址:武汉。邮编 430072。(来稿时间:2006-01-18)

(上接第 70 页)

- 5 范并思,胡小菁. 图书馆 2.0:构建新的图书馆服务. 大学图书馆学报,2006(1)
- 6 Library 2.0. In: Wikipedia. http://en.wikipedia.org/wiki/Library_2.0 (2006-02-10 查询)
- 7 O'Brien, A. Online catalogs: enhancements and developments. In M. E. Williams (Ed.). Annual Review of Information Science and Technology. Medford, NJ: Learned Information, 1994, 29 (转引自 Borgman)
- 8 Tennant, Roy. What I wish I had known. Library Journal, 2005-11-15. <http://www.libraryjournal.com/article/CA6282632.html> (2006-02-10 查询)
- 9 IFLA Task Force on Guidelines for OPAC Displays. Guidelines for Online Public Access Catalogue (OPAC) displays: draft. 2003-09-30. <http://www.ifla.org/VII/s13/guide/opacguide03.pdf> (2006-02-10 查询)
- 10 IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Records; final report. München, K. G. Saur, 1998
- 11 FRBR: OCLC Research Activities and IFLA's Functional Requirements for Bibliographic Records. <http://www.oclc.org/research/projects/frbr/default.htm> (2006-02-10 查询)
- 12 Bibliographic Services Task Force. Rethinking how we provide bibliographic services for the University of California.

2005. 12. <http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf> (2006-02-10 查询)
- 13 Casey, Michael. OPAC Wishlist, more. 2005-10-11. http://www.librarycrunch.com/2005/10/opac_wishlist_continued_1.html (2006-02-10 查询)
- 14 West, Jessamyn. OPAC Manifesto. <http://www.librarian.net/opac> (2006-02-10 查询)
- 15 Blybert, John. ILS Customer Bill-of-Rights. 2005-11-20. <http://www.blyberg.net/2005/11/20/ils-customer-bill-of-rights/> (2006-02-10 查询)
- 16 Jackie's Weblog. 由 ILS Customer Bill-of-Rights 引发的一系列对话. 2006-01-24. <http://jackiege.wordpress.com/2006/01/24/> (2006-02-10 查询)
- 17 LOLA Suggest. <http://www.plymouth.edu/library/bibinfo/suggest.html> (2006-02-10 查询)
- 18, 19 Bission, Casey. Designing an OPAC for Web 2.0; ALA Midwinter IUG SIG Presentation. 2006-01-20. <http://homepage.mac.com/misterbisson/Presentations/ALAMidwinter-2006Jan20.mov> (2006-02-10 查询)
- 20 OCLC. Perceptions of Libraries and Information Resources: a report to the OCLC membership, 2005. <http://www.oclc.org/reports/2005perceptions.htm> (2006-02-10 查询)

胡小菁 华东师范大学图书馆副研究馆员。通信地址:上海。邮编 200062。(来稿时间:2006-03-03)