

保护数字信息的长期存取策略^{*}

刘 家 真

(武汉大学大众传播与知识信息管理学院,湖北 武汉 430072)

作者简介 刘家真(1947-),女,湖北武汉人,武汉大学大众传播与知识信息管理学院图书档案系副教授,从事信息管理自动化与文献保护研究。

内容提要 本文论述了保护数字文献的必要性以及保护数字文献的关键在于维护数字信息的长期可存取性。为维护数字信息的长期可存取,作者提出了 3M 策略:数字媒体的选择与维护、科学管理以及技术迁移。在 3M 策略中,作者进一步分析了已取得的成绩与有待解决的问题。

关键词 数字信息 媒体 迁移 存取 管理

中图分类号 G 352.5 **文献标识码** A **文章编号** 1000-5374(1999)04-0146-04

数字技术对人类的生活与工作方式带来很大影响,由电脑与 Internet 产生的全新通讯方式,使今天的社会文化与记录正在被数字技术所浸透,越来越多的信息资源以数字方式来表示。

为提高服务质量,改进服务方法,提高信息资源共享的程度,图书馆使用数字技术揭示信息、传递信息、保护信息与进行信息存取控制。在大量收集数字文献并不断将传统文献转化为数字文献的过程中,图书馆自身也逐渐由传统图书馆时代平稳地转入数字图书馆时代。计算机的应用与发展,也使档案管理自动化得以实现。几年来,全国已有机电、建筑、轻工、纺织 4 个行业和 23 个省、自治区、直辖市全面开展 CAD(计算机辅助设计技术)的技术推广应用工作。此外,我国测绘资料档案馆与地理信息中心,已建立了地理资料数据库^②。地理信息系统(GIS)技术的引进,使我国的测绘资料实现了图-库数据互访。

数字文献与传统文献不同。在数字文献出现之前,文本、声音、图像及视频都依附在印刷物或其他模拟媒体上进行存贮与传送。数字信息的存贮、存取与传送是相互衔接的,它们受制于硬件、软件与服务分系统构成的技术系统与数字环境。无论是通过模

一数转换还是直接在计算机内形成的文献,一旦它们被编码数字化后,这些数字文献就共享技术层,并由该技术层承担数字信息的存贮与传递的任务。在数字世界里,机器、数字媒体、软件与数字信息四者总是交织在一起。

同传统文献一样,数字媒体的质量与寿命对数字信息的保存是至关重要的,然而,目前高质量的数字媒体(如某些光学媒体)的预计寿命已远远超过了产生该信息的系统的技术寿命。系统是用来查阅与解释存贮在媒体上的信息的,硬件、软件遭到了技术淘汰,媒体保存完好就会失去意义。因为,过时的媒体难以在新的系统上读出。

信息媒体的维护是一个老问题。发展至今,已经有许多方法可以测定出媒体的预计寿命,也有许多手段来创造理想的环境,保证媒体预计寿命的实现。然而,技术竞争十分激烈的当代社会,信息技术却是十分脆弱与不稳定的。目前,信息设备与软件的寿命估计不超过 10 年。技术环境在短期内快速的变化,使用户难以了解与掌握这种变化,更无法驾驭这种变化。因而,数字信息的长期保存与存取充满了风险。由于人们不可能准确地判断当前系统何时过时,

^{*} 国家教育部基金资助项目(96JAJ870003)

因此必须随时准备将有价值的数字文献连同标引、软件迁移到下一代新的系统上。

可见,对已记录到媒体上的数字信息,其长期保存的关键在于保证数字信息的长期可存取性。迄今为止,保证数字信息长期可存取性的关键在于数字媒体(Materials)选择、科学管理(Management)与成功的技术迁移(Migration)。由于这三个关键部分的英文单词都以M开头,故称为3M策略。

一 数字媒体

媒体是存储信息的实体或记录信息的材料。与传统文献一样,数字信息的长期保存与存取是与其依附的媒体材料有关的。造成信息丢失的原因之一,就是媒体腐坏。因此对需要长期保存的数字信息,必须对信息寄存的媒体进行选择,对保存媒体的环境进行控制。近几十年来,人类在这一问题上进行了不懈地努力,产生了许多媒体质量的行业规范与国际标准,并对媒体的保管、管理与灾祸控制进行了卓有成效的研究。

自1980年初以来,美国国家档案文件管理局(NARA)就已经着手了对光学媒体技术发展的监督。经研究,他们认为有三种光学媒体可用来作为联邦文件的数字记录存储媒体,即CD-ROM、WORM及可擦除式光盘。考虑到这类媒体的长期可存取性,NARA进行了两方面研究:光学媒体预计寿命的研究、计算机系统对数据老化的测定与补偿能力的研究。为统一规范光学媒体预计寿命的测定,1987年NARA与美国国家标准局、全美技术与标准协会(NIST)共同拟定了光学媒体预计寿命估计的试验技术标准,并于1992年正式发布生效。该标准的公布,使数字媒体的寿命有了可比较性。

光学媒体也会出现数据错误。数字数据的错误可由多个方面产生:由通讯系统传送数据引入;媒体变形、媒体上的斑点、媒体上微小的水平划痕以及与媒体有关的其他原因。在光学媒体上,那怕仅1.5微米(约1/50000英寸)的跟踪错误,就会引起已存储数据在读出时的较大错误。NIST对以上问题进行了研究,发现为减少光学媒体的读出差错,计算机系统必须具备提供存取错误检测与校正数据的驱动电子设备,用于监测与校正数字媒体的逐步老化。为了鼓励合作研究与开发光学驱动器的错误报告系统,NIST于1992年11月公布了有关研究报告,以便厂家在生产光学驱动器时,能校正媒体的数据老化,减少读出错误。

数字媒体随着信息技术的发展不断推陈出新。因而,紧跟技术发展的步伐,对新的数字媒体进行不断地监督与研究,以供文献收藏部门选用,显得十分必要与迫切。在这方面,美国国家媒体实验室(NML)走在了前列。

NML是由美国政府资助建立的用于高级存储媒体与系统的研究开发与协调的国家技术实验室。它可为政府、信息产业部门、商业部门、文献收藏馆所提供与媒体系统有关的当前最新的技术咨询服务。在它已有的研究成果中,涉及新型媒体稳定性研究、媒体保存环境研究、媒体与系统发展动态研究及媒体与系统故障与处理研究等。

此外,国际标准化组织也为数字媒体的标准化作出了自己的努力。

尽管我国在以上诸方面尚未达到国际先进水平,但我国在数字媒体的选用与媒体的保管方法上也已制定了一些行业标准与部门规范。例如,国家档案局即将出台的行业标准《办公自动化电子文件归档与电子档案管理方法》中要求:电子文件应“拷贝至耐久的载体”“本标准推荐采用的载体,按优先顺序分别是:只读式光盘、一次写入光盘、可擦式光盘、磁带,禁止使用磁盘作为归档电子文件的保管用载体”。此外,在我国正式标准尚未公布之前,有些地方与部门已对本部门、系统的数字媒体选择进行了规定。例如:黑龙江省档案局对本省党政机关电子文件归档的有关问题作出了具体规定:“可以将带有归档标识的电子文件集中,制成归档数据集,拷贝至可读光盘、一次写入光盘、可擦式光盘等耐久载体上,严禁使用软盘作为归档电子文件长期保管的载体。”

二 科学管理

传统文献的保存,无论是直接对文献形体的修复还是保存与利用过程中的维护,大多仅限于该文献的收藏部门内的行为。然而,维护数字文献的长期存取却是一个涉及到多因素的、有风险的管理过程,它与机构间合作、政策与法规、经济与技术以及商业竞争等因素有关。概括起来,维护数字信息长期存取的科学管理包括以下几个部分:

1. 全面开展合作与协调

传统文献的长期保存工作,其涉及面与影响范围是比较小的。例如某馆对破损文献重新装订、修裱、脱酸等处理,对其他馆的工作影响并不十分大。但是进入网络时代后,某一部分与环节出现问题,都会对整个共享的网络资源带来影响。又如缩摄,许多

馆可以找到多种理由拒绝购买他馆制作的拷贝。数字文献的拷贝与原件一样的特点,使文献收藏馆所希望通过合作与协调多渠道地实现最大资源共享,以减少成本,提高工作效率与服务质量。

维护数字信息长期存取的合作,不仅需要在本国进行,更需要在世界范围内跨行业进行。在这方面已经有了许多成功的例子,例如TF2001PADI调研组(澳)、数字归档特别工作组(美)与ECPA(欧盟)就是以保护数字信息长期存取为目的的合作组织。

TF2001PADI调研组是1992年12月由澳大利亚档案馆、澳大利亚图书信息服务委员会、澳大利亚国家保护办公室、澳大利亚博物馆、联邦科学与工业研究组织与澳大利亚电影录音档案馆共同组建的合作研究组织,合作的重点放在研究数字信息的长期保存与存取,为澳大利亚数字信息的长期安全存取,提供管理方法。1996年3月20日,该调研组发布了《保护数字信息:总结报告与建议》,为数字信息的长期存取提供了一套管理方案。

数字归档特别工作组是1994年美国保护与存取委员会(CPA)与美国研究图书馆小组(RLG)共建的合作研究组织。主要任务是对以数字格式存储的信息长期存取进行调研,并对其保护措施提出建设性意见。1996年5月1日,该工作组发布了《数字信息长期存取与保护的总结报告》,讨论了确保数字信息长期存取的关键问题,如机构、技术、法律、经济与迁移等,并逐一提出了相应的对策与措施。

ECPA是保护与存取欧洲委员会的英文缩写。一群学者、图书馆员与档案馆员出于对文献保护的关心,组建了这一区域性跨国协会。它们目前研究的热点是电子文件长期存取。任何人可以用任何文字与ECPA秘书部的电子邮箱(ecpa@bureau.know.nl)联系,寻求有关电子文件存取的咨询与合作。

数字信息长期存取的维护有赖多方合作与支持,涉及信息形成者、信息系统设计者、出版发行者、信息用户与信息提供者。只有通过合作才能更好地完成数字对象的定位、选择、鉴别、编目与保留,这些活动都与数字信息的长期存取息息相关。馆际之间的合作,有利于数字格式标准的统一、数字迁移路径的公布以及其他可以付诸实施的保护数字信息长期存取的措施。文献收藏馆所与技术部门、产业部门、标准部门的合作,共同研究数字文献长期存取的技术要求与保证措施,有利于开发活动的进行与技术市场的控制与协调。尽管图书馆、档案馆在复杂的数字技术市场中只是一个相关部分,但以上工作对于

技术市场的协调以及使技术市场朝着有利于数字文献长期存取的方向发展都是有一定作用的。例如,NARA与ISO及空间数据系统专业组咨询委员会就合作开发长期保存的数据与原数据类型,鼓励生产部门对数字信息的归档提供方便与资助,使数字信息的长期存取落到实处。

2. 管理策略的制定

管理策略的制定涉及到许多法令、制度与策略的出台,没有它们的支持,数字信息的长期存取的维护将难以实现。

数字文献来源不同于传统文献,它可以直接以电子方式产生,也可以是传统文献数字化后转换而成。有些数字文献不需要任何出版部门的认可,可直接在互联网上形成与传送。来源广泛的数字文献,已经不再是传统呈缴本法所能囊括的。维护数字信息长期存取的主要责任者,还有待新的法令来指定。

选择数字文献产生的技术平台,使其在产生之初就能满足当前标准与最新技术要求,是保证数字信息长期存取的第一道防线。这需要管理者把握技术发展脉搏,了解真实的商情与市场走向,正确地挑选产生信息的数字系统硬件、软件与媒体,避免产生的数字信息过早地被淘汰。

技术更新是不断进行的,迁移是绝对需要的,但其费用是昂贵的,涉及到人员培训、形体部分的维护、输入输出设备的选定与检查以及复杂文件的特殊技术等。怎样准确地把握技术淘汰与新旧技术交替时间,选定迁移的最佳时机,制定最佳迁移策略,使费用降至最低是管理中的一个风险最大又极为关键的问题。

技术又在不停地更新换代。21世纪是人工智能时代,怎样将人类超越时空地组织起来,制定各类相应的策略,确保今天产生的数字文献既能适合当前计算机的处理,又能适合21世纪的人工智能,这是一个涉及到许多已知与未知因素的艰巨工作,是对管理的一大挑战。

保护数字信息的长期存取,是对复制。为了维护数字信息的长期存取,人们在怎样才能将数字信息完整、

(无论是纸上或是缩微胶片上),它仅仅只能提供媒体的长寿命与避免

,但却失去了以电子方式存取数字信息的灵活性与特点。

,使前者的功能完全与后者相同。

,这个新的计算机系统可用仿真来执行原来计算机系统的程序,而不用重新编写程序。

,是延迟淘汰的一种方法,仿真的兼容性不可靠且表现也不好。

30年,仿真器对许多系统是有用的,但在新的硬件与软件不断涌现的今天,制造一个执行过时的硬件和其软件的软件,显然是不可能行之有效的。

,人们也曾提过“刷新”(refreshing)技术,即通过复制将数字信息拷贝到新的媒体上,将数字信息由一个媒体向另一个媒体拷贝,其对信息的保护程度是有限的。

,其维护数字信息存取的时间长度只能与当前用来操作该格式的软件一样长。只有在操作该格式的软件存在的情况下,其数字信息才能得到保护。这种复制的可能性还取决于软件对当前格式与旧格式的兼容性,旧一代硬件与当前硬件的兼容性,竞争硬件产品与软件产品的能力有赖于互操作。对于保护数字信息而言,更新并不能提供一种普遍的。

,这促使人们努力寻求其他解决方法。(migration)就在这一情况下产生,它是当前技术条件下,能维护数字信息长期存取的最好。

,或是从老的计算机技术向新一代的计算机技术转换。

“ ”,但在某种意义上是不同于更新的。并使用新一代技术维护该数字文献的兼容时,更新不可能保证产生一个和原文献绝对一致的复制件或精确的数字拷贝,而迁移可以保护原数字文献的完整性。

,即使不能保持原数字文献的外观,也可以优先维护原数字文献的内容及其功能。迁移在保护数字文献的范围方面比更新具有更深。

。迁移是当前技术条件下能确保今天的数字信息在将来可存取的最佳选择。

,如 US-MARC 目录数据库,信息向前迁移到新的标准与应用程序上,也比简单地进行更新更费时、

。虽然大多数数字文献的迁移是成功的,有些多媒体文献仍不能通过迁移成功地进行新旧系统的转移。迁移目前所遇到的最棘手的问题还不只是技术难题,更多的是非技术性问题,例如,由谁承担技术转换关键时刻的技术迁移责任与费用等。

,数字信息保存的关键在于维护其永久的可存取性。

,它与法律、,还需多学科,形成一个保护数字信息的社会氛围与技术环境。

(责任编辑 张琳)

Preservation Strategy for Digital Information

Liu Jiazhen

(School of Mass Communication & Information Management, Wuhan University, Wuhan 430072, China)

Author Liu Jiazhen(1947-), female, Associate professor, School of Mass Communication & Information Management, Wuhan University, majoring in automatization on information management and document preservation.

Abstract The author believes that the key of preserving digital information is access issues. She proposes “3M” strategy: materials, management, migration.

Keywords digital information; media; migration; access; management