新技术的挑战: 电子信息存取与保护*

刘 家 真

作 者 刘家真, 武汉大学图书情报学院副教授; 武汉, 430072
关键词 电子信息 存取 保护 网络 英特网 新媒体
提 要 新技术在信息管理部门的利用, 特别是互联网的开通, 对信息的管理提出了许多新问题。本文仅就新媒体带来的信息保存寿命、信息技术过时与电子信息诚实性等问题展开了讨论, 提出为保证信息的长期存取, 必须对数字媒体进行迁移处理; 利用数字时间标记法保证被存取信息的安全。

在步入数字时代的今天, 计算机使用已十分普遍。在许多国家, 计算机数量已超过公办人员数额。而在我国, 它也已进入不少家庭。不仅如此, 目前, Internet网络还在向世界各国迅速蔓延, 计算机使用由单机逐步走向了网络化。仅仅几年内, World wide web已在发达国家家喻户晓, 电子 Landscape变化也相当迅速。World wide web的一个监视器显示, 在 Internet上有一个很快的增长部分, 表明每年有 500- 700个新的 home page被安装到网络上[①], 而每个 home page可以代表一个大的信息部门, 包括文本、图象、文件、声音与移动画面。在我国, 4条大型网络: 金桥网、科研网、教育网、电讯网已经在国内大城市铺开并汇集起来, 通向了 Internet。这就是说, 我国的信息网络已实现了与国际信息网络的互联。

在互联网上, 可以找到各种各样信息, 每个图书馆的馆藏可以为世界所共享; 只要有密码口令, 就可以在联网的任意终端上, 查阅所需要的档案信息等。当然, 利用各种现代化手段, 档案馆既可以使档案信息公开于网上, 又可以控制自如地保证其安全性与可公开程度。另外, 互联网不仅可以输出信息, 也可以由用户在终端向互联网输入信息, 使有关部门及时收到各个方面的信息反馈。因此可以说, 电子时代的新技术, 使信息资源最大程度、最大范围地实现了共享。信息技术已成为国家发展的必要工具。

信息技术在推动社会进步的同时, 也对信息管理不断地提出挑战。在此, 笔者拟就新媒体带来的新问题进行讨论。

* 国家教委人文科学研究“九五”规划基金研究项目 (96JAQ870003) 论文

一、新媒体* 带来新问题

从远古到现在，书写记录技术经历了持续不断地变化。今天，我们已经有能力去存贮详细的位图图象，也能将成千上万的信息内容存贮在海量光盘上，这是技术的一大进步。然而，令人啼笑皆非的是，进步伴随着后退。当媒体记录信息的能力随时间进展成指数增长时，存贮信息的媒体寿命却以相同比例下降。这是因为新技术并未将新载体的持久性作为首要考虑的条件去设计与生产。

事实表明，本世纪间，除缩微胶片外，新的记录媒体的耐久性均在下降。以下统计数字[②]表明了部分电子媒体的寿命：

媒体	9- 磁迹磁带	8mm 磁带	4mm 磁带	3480盒带	DLT 盒带	磁光	WORM
寿命	1~ 2年	5~ 10年	10年	15年	20年	30年	100年

纸张损坏后，容易被发现。穿孔卡片也可通过其外观来判断是否受损。但是，磁带及其他数字媒体是否损坏就难用肉眼判断，而只有放入机器检读。除媒体自身损坏外，影响信息存取的还有读出媒体的设备与软件的技术过时问题。必须注意的是，在数字世界里，机器与媒体总是交织在一起的。为了达到今天的高存贮密度，我们必须依赖于机器——由它产生信息，并由它读出信息。但这些机器与软件又随技术的不断进步而更新，结果，一旦硬件或软件遭到淘汰，媒体早先记录的信息便无法读出。从上面的统计数据可以看到，最新记录媒体CD的寿命，比其他数字记录媒体寿命都长。然而，为它首次记录上信息的计算机系统在这漫长的时期内，早已经历了数次更新换代，从而使得CD上的信息可能无法读出。这种情况是对信息技术进步的一大嘲弄，但这是信息管理人员必须面对的事实。

由新媒体带来的另一个更为严峻的问题是，数字信息可以完美地被复制，因而也可能不留痕迹地被修改。用户存取数字信息时，不得不对这样的问题担心：我正在阅读的信息是我需要的原始文献吗？同样一份文件，我所看到的信息内容与你所看过的信息内容完全一致吗？经过加工或多次阅读的文件，其信息内容还同原始文件一致吗？

从信息服务器供给终端用户的电子信息，特别是从档案服务器供给必须使用精确信息的用户的电子信息，其内容是否准确可靠，这是电子媒体对电子信息管理提出的另一挑战。

由此可见，信息技术的进步对信息的存取与保护带来三大问题：信息媒体的不耐久；读写信息媒体的硬件、软件的技术过时及电子信息的诚实性。

在高技术林立的今天，只要有雄厚的资金、科学的管理与维护技能，信息媒体基本上是可以达到预期寿命的。问题的关键是，大多数信息媒体的预期寿命均超过了读写它的硬件或软件的技术期限，这使得媒体寿命相对于技术过时而言，显得并不十分重要。世界上几乎没有一个厂家可以保证，它生产的硬件和软件不会过时。技术过时，使得各种信息的已知寿命变得几乎都不可靠。另一方面，如果不能保证输出的电子信息的诚实性，存取与保护信息又有什么意义呢？因而，新媒体带来的三大问题中，技术过时与信息的诚实性最为关键。

---

* 媒体：在这里借用媒体这一电子世界通用术语来复盖载体、档案制成材料等术语。

二、技术过时与对策

在信息的保存中，技术过时是个明显的挑战。数字技术的硬件和软件是在一定的技术平台上产生的，一般说来它也仅仅可以在该技术平台上使用。因而，一旦技术过时，厂家不再提供该产品的技术平台，则信息无法读出。一方面，技术使用期限是受到商业因素限制的。技术革新，使旧的存贮技术自然消失。如半英寸 9磁道的磁带密度在过去已由 200bpi发展到 556 800 1600与 6250bpi，今天再要找到 800bpi的磁带驱动器是很困难的。技术更新使半英寸磁带设备受到更密集的数字视频磁带及 8mm视频磁带设备的威胁。另一方面，当涉及到设备生产中的商业性问题时，如某种 WORM或磁光盘仅由某一个厂家生产或销售，这个厂家或遭到破产或出现意外或改变产品生产线时，再要购买这种产品作为替代就成为难题。

技术过时并不是一个新问题，它不只是发生在数字技术上，在模拟技术上也同样发生。

模拟技术过时的处理比数字技术过时处理复杂，它需要将模拟记录转化为当前数字记录，然后再按数字媒体处理。因为模拟记录若再以模拟方式转换格式，每转换一次，信息将质量下降一次。为保证转换后的数字格式的准确性，应在模拟记录尚未损坏前就将其转换成数字格式。

由于数学媒体的格式、软件、硬件往往混在一起，数字技术的过时相对模拟技术过时更令人头痛。因为即使数字媒体完好，如果打印、检索或编辑它的软件中任一项技术过时，都将影响数字信息的读出。

一般说来，数字存贮媒体的有效寿命都超过了它假设的技术寿命。因而，就这个意义而言，技术过时造成的危害比媒体损坏更严重。

针对技术过时，有以下 4种可供选择的解决方案:

1. 保护落后的旧技术 早期国外某些机构建议建立一个过时技术博物馆，以收集过时技术的软件、软件涉及的硬件及各种操作系统，供过时技术的媒体读出信息。显然，保护落后的旧技术是不明智的，且十分昂贵。因为，它不仅要保留落后的技术的全套硬件与软件，还必须保留操作这些落后技术的某些技能。另外，设备的老化与损失、文件编制消失、原生产厂家逐渐消失、贮存媒体在不断损坏等，均使这种措施无法做到长期保存信息。

2. 复制到更稳定的、人眼可读的硬拷贝上 将技术过时的媒体上的信息转移到纸或缩微胶片上，不再使用设备便可以读出。这种方法虽可以为长期保存信息提供方便并避免了技术过时带来的麻烦，但却存在两个现实问题。一是某些信息无法转移到硬拷贝上，如声音及某些交互信息、超文本信息、多媒体信息等。二是即便以上转换可行，也并不吸引人。将高技术的电子格式信息转移到稳定的硬拷贝上，对于某些档案材料，它可能是持久媒体的选择，它提供了长寿命、肉眼可读、不受技术淘汰的限制，但却失去了电子格式信息传递与使用的灵活性。因而，这种方式也并不总是可取的。

3. 仿效 (emulation) 仿效是制造一个运行过时硬件和其软件的软件。它是在软件中，对一个硬件与软件的模仿过程，以便其他处理认为原始设备与功能在原来形式下仍是可用的。仿效是延迟技术淘汰的一种必要方法。仿效器是一个软件，它可以在非原生平台上运行应用程序，有时是操作系统。说通俗点，仿效器应是升级了的软件。某些软件制造商曾在他们产品

中建立了对过时技术的兼容性，如 Microsoft Word 6可以读写 Microsoft Word 5文件。利用这种方法来挽救过时技术媒体的记录，为保证数字信息长期存取，必须把仿效器与原数据(metadata)套装在一起。虽然从技术上看仿效是可能的，但在实际工作中，它的兼容性是不可靠的，且表现也不好。实际上，仿效器自身也构成了档案的一个部分。因此，在使用期间，它的耐用性也必须得到维护。因而，如果用这种方式来保护信息的长期存取，尽管档案馆、图书馆并不必去维护实际的设备，但为了过时信息的读出，它必须要对仿效软件及数字信息同时进行维护。因此，虽然这种方法有一定的可行性，但对于新型硬件与软件不断涌现的今天来讲，制造一个执行过时硬件和其软件的软件，是不可能有效的。

4. 迁移（migration） 将数字信息从一种技术转移到另一种技术上的复制，包括硬件与软件，称为迁移。它是一种随技术变化定期改变数字信息格式的处理过程。这个过程使得信息将来也可以被存取。迁移，意谓着基于字符的数据，可以从一个存取媒体迁移到另一个存贮媒体上，以进行电子媒体信息的保护。要完成迁移，要求计算机既可以读出旧格式，也可以将它写在新格式上。迁移还包括更新，即拷贝数字信息而不改变它。但当利用更新来克服媒体不稳定时，它通常并不足以与技术淘汰同步。尽管如此，在操作系统改变后，它变得很必要。迁移并非新概念，在计算机工业上已有了很长时期的实践。1994年底，美国保护与存取委员会（CPA）与美国研究图书馆组组建了一个数字信息档案化特别工作组，负责调研与推荐能确保“数字式电子格式文件未来连续存取”的方法。在研究工作中，该特别工作组认为迁移是对付技术过时最好的良策，它应是数字材料完成定期转移的一系列有组织的任务，该任务涉及到维护数字对象的诚实与用户再检索、显示及其他利用能力。该工作组在总结报告中对某些迁移策略规划如下：(1) 将数字信息从稳定性低的媒体上迁移到稳定性更高的媒体上，从软件对此依赖程度高的格式迁移到软件对此依赖程度较低的格式上；(2) 将数字信息从任何时期已有的繁多的格式中迁移到为数较少的普通格式上；(3) 和工业界一起开发向后兼容性途径（backward compatibility paths），以此作为所有软件的标准；(4) 发展与利用迁移标准。[③] 该工作组认为，任何一个迁移策略的重要组成部分应是保存可能记录在总数据库上的迁移轨迹，它将记录连续迁移及它们的效果，把数字对象的原始格式与他们正在使用的格式的任何差别告诉用户。迁移工作应注意两个问题：一是迁移到不同的操作系统，常常意味着拷贝不是原件的“精确”复制件。因而，利用迁移技术时，即使在它不可能保持格式外观时，也应优先维护信息内容与其功能。二是，对待模拟技术的迁移问题，应首先将模拟信息转变为数字格式，目前声频与视频信息均可采用一定手段转换为数字格式。

三、信息的诚实性对策

对数字信息的大量存取是档案馆、图书馆的重要职能之一。用户要求被存取的信息是安全的、真实的并不被干扰，因而防止媒体损坏与技术过时，仅是保证信息存取的一部分工作，另一部分工作应是保证被存取信息的可靠性。这是新技术对信息管理的一个更大的挑战。

电子信息容易被改变或被巧妙地处置，如更换一个调制解调器就可修改拷贝。而被修改的拷贝容易在网络上广泛传开，结果造成多个难辩是非的版本。印刷的文献文本是固定的，稍有变动，我们可以觉察、判断出来，而电子文本的改变不受这些因素的暗示，这使我们难以

通过表面现象观察到电子文本信息的变化。因而,信息的诚实性是电脑代文本后出现的新的、紧要的问题。

以下几种情况可以导致电子信息内容的改变: 事故、有目的信息更新与蓄意破坏。

事故是电子信息在传送与操作过程中,由于疏忽而无意使数据丢失。如在网络传递中,数据有可能已经腐坏;或在计算机存贮器、磁盘间,数据已受损。这些均可造成数据丢失。尽管这类事故并不经常发生,但只要发生就会使信息的诚实性受到威胁。更常见的是,网络在更新过程中偶发性故障,使文献整个版本或整个段落丢失。许多网络工作者都有过这种经验,但读者往往未留意到这种丢失。

有目的的信息更新是网络的正常工作。动态数据库这一特性,使网络频繁更新。网络的更新将引起文献结构的改变和内在内容的更新,这将引起信息内容的改变,使我们现在看到的信息与上次看到的信息发生改变。这是一种正常的改变,它可能导致特定的新版式。预料中的结构更新,可能是交互文献的正常结果。

蓄意破坏是故意去改变网络上信息内容,如由于各种原因去覆盖别人或自己的文章或改变各种证据,以获得各种利益。

为确保用户需要的文献就是原始文献,即保证输出信息的诚实性,可以用某种方法固定文本或文献,这种固定文本或文献的手段称为鉴别。鉴别电子信息有三种方法: 密码法、散列法 (hashing) 及时间标记法 (time- stamping)。适用图书馆、档案馆使用的鉴别法,除应具有鉴别能力外,还应容易使用,且不妨碍网上其他文献的产生与存取;具有公开性,并可以在超越人类生命的漫长时期使用;同时要求其费用低。现有的三种鉴别法各有其特点。

1. 加密法 加密依靠文献的数学转换,建立一个特别数字,这个数字也称为密匙,是被要求加密的文本的译码,这些译码是一串很长的数字。现代加密依靠非常复杂的技术处理。在美国有两种最为熟悉的加密格式: DES与 RSA。DES是 1975年作为数字加密标准第一个建立起来的,现在为许多商业与政府机构采纳。RSA是三位数学家同时研制出来的加密方法,它优于 DES,现为私人销售。但任何一种加密方法未必能为档案馆、图书馆所欢迎。因为,加密要求读者知道密码或密匙,而用户在网上需要的信息量很大,用密码鉴别是不可能的。此外,图书馆、档案馆的大多数文献是长期保存的,在这漫长岁月里,密匙有可能丢失。这样加密后的文献就无法再利用了。另一方面,加密仅仅是提供了一个鉴别文本的可能,只是在文本没有发生改变时。若文献被修改后,再用同样密码进行加密,用户就无法鉴别了。由上可见,用加密法来鉴别档案馆、图书馆的电子文件显然是不现实的。

2. 散列法 (hashing) 另一种鉴别技术称为散列法。散列法取决于文献每一部分任意价值的确定,从该处产生具体价值而不是没有内容的值 (称为杂凑总数) 的电脑计算。由于杂凑总数的具体计算是没有价值的与没有内容的,因而从混乱信息返回到原始文献的计算是不可能的。这些混乱数据是一个或一百个或更多数据,它必须小于从计算机来的原始文献,它保存着原始文献信息的秘密。每产生一个文献信息,就会有混乱信号产生,保存它并分别进行补偿。如果该文献是电子出版物,它应随它的混乱信号一同出版;如果文献被引用,混乱信号也应是引文的一部分。如果读者使用文献并希望知道它是否真实,使用标准的算法计算的混乱信息出现在用户计算机上,对比用户正在使用文献的混乱信息,如果相同,用户可以确信使用的信息是真实的。档案馆、图书馆要使用这种手段来鉴别电子信息的诚实性,必须

以可行的方式广泛传播算法，使它成为微机上的一个菜单。因而这种方法虽然可行，但并不十分简便。

3. 数字时间标记法　数字时间标记（DTS　digital time- stamping）不仅可以鉴别电子信息的诚实性，而且还可以告诉用户，该信息产生的时间与地点。它类似于一个橡皮图章，用它对入网文献加盖印章。DTS是保护电子信息诚实性的一种最新的解决方法，它已广泛用于美国银行、法律部门、药剂公司及对政府有价值的电子信息上。

该技术是一个建立在软件基础上的算法，而不是建立在更新频繁的硬件上。它使用了单道杂乱信号的密码技术及使用“广泛被证明事情”的概念。该技术模拟用橡皮图章将它们接收到的数据与时间引入到文章中。在电子结构中，它通过文献创作者或对该信息负有责任的间接责任者，通过对该信息的最近一次使用，建立一个为后者鉴别该信息真伪的必要条件。

该方法仍使用了混乱信号。混乱信号的作用是将电子信息转变成一个单独的、确定数的规则系统。这个数并不是指数的自身，而是唯一代表这个数的一个集合。电子信息单道密文的杂乱信号可以用数字的复合来产生，并非由人或计算机技术产生，这就保证了该混乱信号的唯一性与非再生性，即要改变含有该混乱信号的电子信息是不可能的。可以允许电子信息与其单道杂乱信号同时在计算机上公开，产生杂乱信号的软件也可以被公开，但它的机制要保密。

在实际工作中，数字时间标记法要求存在时间标记服务软件，网络计算机上的用户软件同时要求：产生电子信息的杂乱信号，并能与该服务器连结。用户工作站在按键过程中就很快产生了一个电子信息的杂乱信号，并输向了印章服务器，印章服务器可以将刚收到的这一杂乱信号与先前收到的同一电子信息的杂乱信号相混合，以产生当前杂乱信号执照，该执照又被送回用户工作站。这一执照成为原始电子信息当前使用或将来（甚至半个世纪后）使用的鉴别信息诚实性的凭证。

时间标记法可以对档案馆、图书馆的电子信息提供长期的鉴别作用，保证了电子信息的诚实性。

注　释：

① 国家数字图书馆：http：//www. lyra. rlg. org/Arch TF/

② 保护数字实体：需要与挑战：http　//www. commnity. bellcore. com/lesk/auspres/aus. html

③ 迁移法：http：//WWW. nla. gov au/dnc/tf 2001//padi/migration. html

(责任编辑　张　琳)