

21世纪 信息资源管理系列教材

数字信息保存

张美芳 等/著

中国人民大学出版社

21 世纪信息资源管理系列教材

数字信息保存

张美芳 等/著

中国人民大学出版社

• 北京 •

00000001 11001 110001 10101 11010 11011 10100 11000 0111 10001

图书在版编目 (CIP) 数据

数字信息保存/张美芳等著. —北京: 中国人民大学出版社, 2014. 12

21 世纪信息资源管理系列教材

ISBN 978-7-300-20520-5

I. ①数… II. ①张… III. ①数字信息-信息存贮-高等学校-教材 IV. ①G203②TP333

中国版本图书馆 CIP 数据核字 (2015) 第 000432 号

21 世纪信息资源管理系列教材

数字信息保存

张美芳 等/著

Shuzi Xinxì Baocun

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

邮政编码 100080

电 话 010-62511242 (总编室)

010-62511770 (质管部)

010-82501766 (邮购部)

010-62514148 (门市部)

010-62515195 (发行公司)

010-62515275 (盗版举报)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com>(人大教研网)

经 销 新华书店

印 刷 北京易丰印捷科技股份有限公司

规 格 170 mm×228 mm 16 开本

版 次 2016 年 1 月第 1 版

印 张 20

印 次 2016 年 1 月第 1 次印刷

字 数 341 000

定 价 36.00 元

版权所有 侵权必究

印装差错 负责调换

作者简介

张美芳，中国人民大学信息资源管理学院教授、博士生导师，曾赴美国匹兹堡大学做访问学者。国际标准化组织文献管理应用技术委员会工作组专家，中国文献影像技术协会第四届理事会常务理事，中国文献影像技术标准技术委员会第五届常务理事，中国感光学会影像保护专业委员会副主任，国家古籍保护中心评审委员会专家，中国档案学会档案保护技术委员会副主任。出版专著 5 部，教材 4 部，学术论文 70 余篇；主持或参与国家级科研项目十余项；参与制定、修订国际或国家标准 14 项。曾获省部级特等奖、一等奖、三等奖等，档案保护技术学课程入选 2005 年北京市精品课程。

内容简介

信息时代，数字信息保存是一项日常的、具有挑战性的任务。《数字信息保存》融合计算机技术、信息管理技术、文献保护技术、电子政务等，以数字信息为论述对象，以长期保存为目标，以案例形式介绍了国内外的数字信息保存项目和保存实践活动。

针对数字信息保存流程和面临的主要问题，本书较为全面地介绍分析了数字信息保存所使用的介质、数字信息长期保存工作流程、保存系统的构建方法和模式、濒危数字信息的抢救措施和数据恢复技术等。本书阐述了数字信息保存系统的管理与利用，从技术和管理方面介绍了开展数字信息保存工作可以采取的方法和措施，同时强调了加快数字信息保存相关标准建设、健全政策与法规对数字信息保存的重要性。

本书以全新的观点、系统的知识和合理的结构为特点，为构建数字信息保存学科奠定理论基础，积极迎接数字信息保存的未来挑战。

目 录

绪 论	1
第一章 数字信息保存的发展	3
1.1 数字信息保存的概念	5
1.2 数字信息保存目标及保存内容	8
1.3 国际开展主要数字信息保存研究及项目	19
1.4 我国数字信息保存基本状况	28
1.5 数字信息保存责任体系	35
第二章 数字信息保存的动机	42
2.1 数字信息保存的意义	43
2.2 数字信息保存的必要性	55
2.3 数字信息保存的主要问题	62
2.4 数字信息保存的风险	69
2.5 保存数字信息的可行性认证	72

第三章	数字信息保存系统的建设	78
3.1	数字信息保存工作流程的架构	79
3.2	数字保存系统分析	83
3.3	数字保存系统类型及功能	85
3.4	数字保存数据库开发	88
3.5	数字保存环境分析	91
3.6	数字保存系统使用权限管理	92
3.7	案例分析	96
第四章	数字信息存储方式与存储系统	99
4.1	概述	99
4.2	主流存储体系结构及对比分析	100
4.3	数字信息存储保护的软件系统	117
4.4	存储系统开发的关键技术与标准规范	129
4.5	技术方案选用原则与评价方法	134
第五章	数字信息长期保存的技术保障措施	139
5.1	概述	139
5.2	系统保存	141
5.3	仿真技术	142
5.4	介质更新	144
5.5	迁移技术	152
5.6	封套技术	156
5.7	保存元数据	157
第六章	数字信息保存相关标准	161
6.1	与数字信息保存的相关国际标准	161
6.2	数字信息保存标准的比较	195
6.3	我国数字信息保存标准研究进展	198
第七章	数字信息的容灾与抢救	206
7.1	概述	207
7.2	数字信息的容灾备份	210
7.3	数字信息的灾难恢复	215
7.4	数字信息容灾系统工程	219

第八章	数字信息长期保存的规划和管理	235
8.1	数字信息长期保存战略和政策的制定	235
8.2	数字信息长期保存模式的建立及运行	244
8.3	数字信息的质量控制	250
8.4	数字信息安全政策与风险评估	256
第九章	数字信息保存的未来挑战	265
9.1	数字信息存储技术	266
9.2	数字信息保存研究的拓宽和深入	275
9.3	数字信息保存与管理的关系	281
9.4	数字信息保存的协作、分工与共享	287
参考文献		299
附录	数字信息保存中常见术语中英文对照	307

绪 论

数字信息资源逐渐成为社会信息传播、交流和保存的主要方式，长期保存数字信息已成为数字图书馆、档案馆、电子政务和其他数字信息系统的战略性问题。因为它涉及在多变的数字环境中运行的信息的保存，以及怎样合法地控制这种变化并保持数字信息的长期存取，这是摆在我们面前亟待解决的现实问题。我国数字信息建设快速发展，但伴随发展而产生的数字信息保存问题还没有得到有效的解决。如果我们不在数字信息发展中认真考虑数字信息长期保存的问题，待数字信息成为整个社会发展支柱时，这一隐患将导致已建立的数字社会全面崩溃。因而应将数字信息保存视作数字信息建设的重要组成部分。数字信息的保存，涉及政策、技术、管理、法律、经济与组织机构等问题，本书就其中的主要问题、核心问题进行了较为全面的阐述和探讨。

科学和规范地存储与管理数量巨大的数字信息，建立社会化、程序化、科学化、专业化和集成化的管理方式，解决国家数字资源长期存储无保证、管理程序不规范、管理方法不科学、管理责任不明确、管理水平不一、管理效率低、技术与管理成本高、管理资源重复浪费等诸多问题，是数字信息存

储发展和研究的必由之路。构建数字信息保存技术管理规范、工作流程、存储体系，建立前瞻性的数字信息质量评价体系，融合国内外不同领域的管理思想和保护经验，建立数字信息保存技术，是摆在档案保护部门、信息管理部门、图书保护部门、文物保护部门、博物保护部门、IT 行业等科研人员面前的一个迫切需要研究的课题，也是本书的宗旨和核心所在。

《数字信息保存》是目前国内首部有关数字信息保存和长期维护的专著，它顺应时代发展，符合社会需要，满足各领域要求，填补档案保护、文献保护、计算机技术的理论空白，具有较高的学术价值和重要的理论意义。本书给自己设定了一系列的任务。这些任务既十分宏大，又具体深入。说它们十分宏大，是因为它们首次将研究的视野扩大，从国内到国际，从历史到现代，实现了跨地域、跨时空、跨领域、跨部门的研究；研究内容从技术拓宽到管理，研究方法由理论转向实践，实现理论与实践相结合。说它们具体深入，是因为它们对数字信息保存中的国内外动态、最佳实践的具体内容都作了较为详细的阐述，小到方案设计、数字信息保存流程、技术评估等，大到管理中的一些复杂问题如数字信息保存责任体系的建设、数字信息灾难恢复等。

本书共分为九章。其中一、二章由张美芳撰写，三、四章由王应解撰写，五章由夏天撰写，六章由张美芳、蒋琴、马丹宁、秦佳心撰写，七章由唐跃进撰写，八、九章由赵淑梅撰写，全书由张美芳做文字和内容的统一处理。书中引用了该领域许多专家学者的资料、数据和图片，在此谨致以衷心的感谢。

由于计算机技术的迅猛发展，数字信息保存技术的内容和知识需要随时更新、完善和补充，数字保存涉及的内容多、广而又复杂，加上作者的水平局限，书中难免有不足之处，敬请读者批评指正。

第一章

数字信息保存的发展

- 数字信息保存的主要内容
- 数字信息长期保存策略的目标
- 国外数字信息保存项目简介
- 我国数字信息保存进展
- 数字信息保存责任体系

随着计算机与网络技术的快速发展，数字信息正在大量地产生、存储与传播，人类社会也由此步入了信息社会和知识社会。数字资源的优点日益凸显，用数字化表示的各种信息具有体积小、容量大、可记载多媒体信息、容易复制和打印、更新速度快的特点。在网络环境下，数字资源的时间、空间得到进一步延伸。网络是继报刊、广播、电视之后的第四大媒体，它的传播范围广，可以提供不受空间限制的实时服务，可供多人同时阅读，可以即时更新信息。实际上，作为一种文献资源，数字信息有别于其他媒体信息的最重要特点是它的检索功能和方便利用，它可以在几秒时间内从几亿汉字中找到所需要的内容供阅读。但是，数字资源的广泛传播利用还需要逾越相当多

的技术障碍,如对阅读设备的依赖性。数字资源需要存储在计算机设备(硬件)中,在阅读、使用中需要借助操作系统、应用程序(软件),不同类型甚至不同来源的数字化信息可能有不同的应用软件。所用到的硬件、软件不是一成不变的,而是经常发生变化。硬件的升级有著名的“摩尔定律”,芯片容量每 18 至 24 个月翻一番;软件的升级换代更是频繁,从 DOS 到 Windows3.1、Windows95、Windows98、Windows2000、WindowsXP,一直到 Windows vista、Windows7,操作系统的变化随着微软的发展而不断推陈出新,应用软件开发商们则忙着升级、升级、再升级。使用者一方面要不断学习、学习、再学习,另一方面还要面对老数据、旧系统在新的软硬件环境中无法使用的尴尬局面。

计算机技术迅速变化的特性以及数字存储媒体寿命的短暂性,致使人类在长期保存已经形成或即将形成的各种数字信息时面临着巨大的挑战。

全世界每年生产的信息,其中印刷载体的信息量不到 1%,磁介质(包括数字光学介质)所承载的信息(即数字信息)量约 95%。而且由于后者的增长速度越来越快,正迅速成为主流存储介质;网络中每年传递的 E-mail、不重复公共网页、非网页形式网络数据库仍以几何级数在增长。由于数字信息在技术基础、利用方式和管理机制等方面的特殊性,它的长期保存面临许多新的问题。磁介质和数字光学介质的不稳定性,致使其有效保存期远低于纸介质和缩微胶片介质;数字信息技术迅速变化,致使人们借以利用数字信息的各种编码、压缩、格式化、加密解密、内容显示、计算运行等技术和软件甚至硬件迅速过时;数字信息往往是动态变化的,例如一个动态出版的电子期刊或一个不断更新的网站,因此,对数字信息的完整性确认、真实性考证、原始性认证等带来很大困难;数字信息与其他信息有这样或那样的连接,数字信息的产生需要一定的环境,例如嵌入许多多媒体数据文件的网页或链接许多参考文献的学术文章,因此,对数字信息存储内容的确认、对存储位置和存储格式的选择都具有一定的复杂性;数字信息在其产生和利用过程中的灵活性和动态性,致使数字信息的产生及所有权变化难以追踪和保证;数字信息发送传播机制已经打破传统的信息存储责任体系,例如出版商直接向用户提供数字期刊检索阅览、图书馆通过使用许可提供用户服务,但并不实际拥有数字信息,因此,亟须重新定义和建立数字信息长期保护的责任体系;数字信息的知识产权法律和管理机制尚不完善,数字信息长期保护中涉及的权利许可、权利转移、内容转移与存储、服务以及相应的经济与管理问题还有待解决。数字信息的巨大生产量和存储数字信息平台的不一致,导致信息

的交流分享不能成为现实，同时也将使上述问题更加突出，使解决这些问题变得更加迫切。

1.1 数字信息保存的概念

1.1.1 数字信息

根据联合国教科文组织的定义，数字遗产包括“从现在的任一种形式的知识转化成的数字产品”。这里所提到的数字产品既包括一开始就是以数字形式产生的文献资料，又包括由传统形式的文献转换而成的数字文献。对于后一类，我们称为数字再造资源，它指的是对以前存在的模拟载体上的信息进行数字化处理后形成的产品，也称为“digitalized”（数字化信息）。它们不一定与原物一模一样，只能算是原物的一种表现形式，但往往能利用数字化的特长将原物更好地展现和利用。譬如，印刷图书的数字扫描或 OCR 版本，很多名画的电子图片，书籍的电子扫描版本等。前一类则是直接由数字信息系统（例如文字扫描处理软件、CAD、数字摄像设备、地理扫描设备、数码拍摄等）产生的原生数字文献（称为 born-digital），包括档案材料，电子邮件，视听资料，数字艺术品和电子文学作品，电子游戏、虚拟现实和软件，电子手稿，图像，其他各种电子资料，联机信息和网络资源，数字资料的物理格式，科学、统计、研究数据（如太空、地理和类似数据集），文献通报资源等。据统计，越来越多的有价值的内容是“天生的数字型”，而且只能以数字形式来管理、保存和使用。从数据生产者角度看，数字信息来源可以是出版单位、各类机构、个人及其组织、社会文化组织、信息服务系统等。从长期保护角度看，人们主要关心的是以数字状态为唯一或主要形态的信息，其中又主要是由出版单位、各类机构、正式组织和信息服务系统生产的数字信息。美国国家信息基础结构和保存项目 NDIIPP（The National Digital Information Infrastructure and Preservation Program）将数字信息分为电子书、电子期刊、网站、数字电视等几大类。

为了使人类知识、文化得以长期保存，世界各地的图书馆、档案馆等保存机构经过不懈的努力，在纸张文献的保存方面积累了宝贵的经验，并制定了各种标准，取得了显著的成果，为保存一国文化做出了巨大的贡献。对于那些具有保存价值的数字信息，如何长期保存、谁来保存、怎样保存等是在信息时代出现的新问题，是人类要面对和解决的。数字信息的保存

方式、保存任务、保存结果、保存目标、存储载体等与传统模拟信息的保存有很大的区别,对其的保存不但是信息保管人员的任务,同时还涉及IT、法律、知识产权、政策等领域的相关问题。它不再仅限于关注承载信息载体的保护、环境的保护,而更多的是关注信息内容的保护和信息的有效读取问题。

由于数字信息长期保存涉及许多方面的复杂问题,因此,在数字信息的形成、接收、连接、利用和保存等过程中,应充分考虑所采用的有关政策、流程、技术、格式和工具对长期保存的影响,建立系统化的策略体系和操作规范。如采用标准、开放的数据处理技术和格式,优化存储系统、存储介质,建立知识产权确认和授权转移规范,建立数据检验鉴别标准和保护级别体系与处置程序,采用标准元数据格式并建立完善的元数据,建立长期保存技术选择模型和操作规程,建立严格的数据存储系统保障机制,等等。

在数字信息长期保存中,需要保护数字信息本身及相关技术、方法和工具,包括保存数字比特流、保存数字格式与处理信息、保存数字信息处理环境、建立数字信息的验证和管理机制、保存数字信息的组织利用环境。为了有效实施数字信息长期保存,需要建立数字信息保存责任体系及相应的选择标准,研究开发具体的数字信息更新、技术仿真和数据迁移技术与系统,研究数字信息长期保存系统,开发描述数字信息长期保存要求和政策的元数据等。

1.1.2 数字信息保存

保护、保存、典藏和存储等概念是近年来在数字信息管理中使用较多的概念,针对文献资料而言,这几个概念有相似之处,都靠“存”的行为,即采取的一些手段,使文献免遭损坏。“保存”是事物、性质、意义、作用等继续存在,不受损失或不发生变化;“保护”是尽力照顾,使不受伤害。在档案界、图书馆界多采用“保护”一词。对于传统的档案馆和图书馆而言,保护是指采取必要的预防措施以避免因使用不当或自然老化造成变质,对损坏文献采取抢救和修复,尽量使文献恢复原貌。档案保护和图书保护更多关注的是载体的变化和对其采取的相应抢救与修复措施。“典藏”是图书馆学的专用术语,从理论上讲图书典藏是指一种藏书的组织管理和安排。对收集的文献按一定的要求进行合理的布局、排列、保护、清点、存放和管理的过程,其目的是使藏书得到有序、完整、有效的利用。目前我国台湾多采用数位典藏

这一概念。“存储”侧重的是存放和安置的行为，数字信息的存储只是针对处理后的信息采取的一种存放的行为。存储并不是保存，存储较为容易，仅仅是在执行一种简单的命令，而保存相对较为复杂，意味着对存储的信息进行编目、建立系统环境、找到合适的位置、采用一定的技术和标准保证信息可存、可取、可利用。

由此看来，保护侧重的是载体，典藏侧重的是文献的有序排列和存放。针对数字信息来讲，面临的最大的困难不是对载体形态进行保护，而是解决由于技术的过时而造成的数据不能读取的问题，“保护”一词似乎面较窄，“典藏”一词又有些专门化。目前，“保存”一词对于数字信息的保护可能较为准确，包括行为、策略、方案和技术。

数字信息保存是指将有价值的实体或非实体数据，利用数字化科学技术保存在数字保存系统中，所有保存数据均以数字化形式存在。数字保存是为防止数字信息资源的损坏，或保证数字信息资源的可用性而对信息进行的采集、组织、分配、保护活动，也即是一种基于存档的活动，随着时间的流逝，在技术已经变化了的情况下，还能够对存档的数据进行存取。数字保存具有以下特点：就保存数据种类而言，数字保存的数据范围相当广泛，凡是代表人类文明，具有保存价值的数据，不论是数据产生的时间还是形式，如文化、科学、艺术、音乐、动植物等，都是数字保存的对象。就数字化数据的格式而言，只要是保存范围内的各种形式（如文件、器物、标本、录音、影片、建筑等）与数字化后的电子数据格式（如文字、语音、影像、2D/3D对象等），均属于数字保存的数据格式范畴。就数据的质量而言，数字信息保存的数据是后续研究与应用增值的基本素材，因此，所收纳的数字信息必须经过不断的累积与校订，以确保其保存数据的完整性与正确性，这种高质量的信息要求，与一般互联网上的网页数据有很大的差异。

除上述概念外，还有两个概念需要了解。

掌管（curation）：指从数据生产出来时就开始的管理和促进其被利用的行为，目标是使数据能够符合现实的需要，或能被用于发现和重用数据。

存档（archiving）：在掌管的基础上，确保数据是经过合理选择、得到良好保存的，并且这些数据可以被存取。随着时光的流逝，确保这些数据的逻辑和物理完整性得到维护，具有相应的安全和认证机构。

1.2 数字信息保存目标及保存内容

解决好数字化信息的长期保存问题是具有重要意义的,它反映了各个国家的信息技术发展水平,涉及不同行业和部门的切身利益,是信息技术发展中的核心问题。这些信息若保存得不好,就会永久丢失。目前数字信息长期保存已成为世界同行关注的焦点,并成立了众多的研究项目,形成了一系列研究成果。

1.2.1 数字信息的保存目标

可持续发展是数字信息长期保存的目标,它关系到过去、现在和将来历史文明能否永久保留下来的问题,是保存一国文化的迫切需要。建立可持续发展的数字保存生态环境,寻求科学的战略发展观是实现这一目标的基础和保证。

开放档案信息系统(Open Archival Information System, OAIS)认为,数字信息保存系统的目标是确保被保存的信息对于目标团体而言是独立可理解的,也就是说,应当确保目标团体在没有信息生产者(专家)帮助的情况下,仍能够独立地理解被保存的信息。为了实现这一目标,在保存系统中既要保存存档对象,也要保存针对对象的描述和说明信息。保存元数据实施战略(Preservation Metadata Implementation Strategies, PREMIS)认为,数字信息保存的目标是维持数字资源的长期可生存能力(要求被保存数字对象的比特流完整无缺并且能够从它所存储的载体之上读取)、可呈现能力(指的是能够将比特流转变为一种可以被人看,能够被计算机处理的信息的能力)和可理解能力(包括提供足够的信息使得这些被呈现的内容可以被目标用户理解和认识)。普里西拉·卡普兰(Priscilla Caplan)认为,从下到上、从较低要求到较高要求,数字信息保存的目标分别为:可获得性,即选择获取技术;可识别性,即描述;可理解性,即文档记录;完整性,即安全存储;持久生存能力,即媒体管理;可呈现能力,即格式技术策略;被保存对象的真实性,即真实还原技术。

艾比·史密斯(Abby Smith)在《未来意义上的保存》中指出:“一切保存方案的目的都是确保对一个机构信息资源的长期的、随时的存取。”保存数字信息的首要任务必须是形成该信息的各方面的信息能同时完整地、准确地再现。国内学者普遍认为数字信息的保存有四方面的含义:(1)保证数字信

息存储载体的稳定；(2) 保证信息内容逻辑上的准确；(3) 保证数字信息的功能性；(4) 保证数字信息的存取性。概括而言，保存的目的即是要保存数字信息的内容、物理形式和功能。

长期保存策略应确保数字信息在未来可读。为了达到此目标，构成数字信息的比特流应在利用计算机系统或设备而进行的以下过程中可用。数字信息在产生之初、存储之时、利用之时、将来被保存这四个过程中都有可能变得不可读。这主要由两种原因造成。

一是不适宜的保存环境。不适宜的环境包括载体保存物理环境、信息生成环境、存取环境、保存的技术环境等。目前用来存储电子文件信息的所有介质普遍容易受到不适宜环境的侵害，如温湿度的波动。这些不利条件会毁坏介质或加速其老化过程。不同类型的数字存储介质需要不同的保存环境以确保其最大寿命。一些存储介质在磁场干扰、灰尘和环境污染物的作用下容易产生信息损坏（磁性存储介质），而其他不易受到外界因素影响的介质（光学存储介质），只要稍加控制存储环境，就不易受损。不管使用何种存储介质，关键是要意识到不利的环境会引起各种形式的存储介质的损毁或老化。

二是介质过时也许导致信息无法读取。存储介质（如磁带或光盘）与使用的计算机硬件（如磁带或光盘驱动）若存在物理上不兼容，则导致信息无法读取。基于信息技术的发展趋势，介质将来过时是不可避免的，因为储存技术的进步会继续导致数字信息物理存储方式的改变（如唱片技术、光盘驱动硬件、软件接口的改变），存储介质形式和文件信息的比特流是用物理表示的（如纠错码），因此，随着时间的推移，存储介质将会与以后使用的介质不兼容。

因介质和保存格式、保存系统过时，长期保存策略宜特别重视将数字信息定期从老的介质转移到新的介质上的方法。

1.2.2 数字信息长期保存的真实而具体的含义

1.2.2.1 数字信息的可读性

作为比特的安全保护者，当今的数字存储介质至多只能维持几十年，而且存储介质存在易碎、易损、易变质等问题，会随着时间的推移而变得越来越不可靠。尽管数字信息管理者对这些存储介质采取了某些保护措施（例如建立良好的保护环境），但是随着时间的推移，存储介质最终还是需要更换，因为存储介质的损坏将直接影响数字信息的可读性。

1.2.2.2 数字信息的可用性

数字信息的可用性是指数字信息可以被人或电脑阅读和处理。数字信息

从形成、传输到存储都是通过计算机实现的,因此数字信息与计算机系统中的各种设备有着密不可分的关系,也可以说计算机操作系统与硬件设备是生成数字信息的前提和基础,这决定了数字信息对相应的软硬件系统的依赖性,而对其他软硬件环境则存在一定的不兼容。但是计算机硬件设备的迅速发展(如磁盘或磁带的驱动器)和操作系统的更新换代在给信息处理、信息存储、信息传输和信息阅读带来极大便利的同时,也造成旧有软硬件易于过时作废。因此,如何解决旧有数字信息适应新的电子环境的问题成为数字信息管理者关注的难题。如果不加以解决,最终的结果就是旧有的数字信息将不能被人或电脑阅读和处理,也就意味着阻碍了数字信息的持久可用性。

1.2.2.3 数字信息的可理解性

数字信息的可理解性是指数字信息的格式化内容可以被它的使用者理解。无论是传统文件还是电子文件,通常由格式与内容构成,只是传统文件的格式与内容固定在某一载体(例如纸张)上,一经结合就不易分开,而且无须借助于任何工具就可以使人理解。而数字信息的格式化内容则无法独立存在,必须依赖特定的软件才能阅读。然而由于技术的快速发展,这些软件却容易过时,被新的版本所替代,甚至消亡。尽管有些软件在更新换代时仍然考虑与以前的格式兼容,但是还有一些软件并不兼容以前的格式,而且不兼容的现实在计算机技术发展的短短几十年内尤其明显,大量的数字信息由于内容或格式的丢失造成了信息的不可理解,这些信息随着技术的进一步发展,最终将失去长期保存的价值。

长期保存策略应提供易理解的数字信息。计算机能通过描述底层的比特流而说明信息,数字信息只能被计算机理解。因此,数字信息的可理解性关涉比特流实际代表的功能和应用处理软件的能力及为此而采取的适当措施。

1.2.2.4 数字信息的完整性

数字信息的完整性是指数字信息是齐全的,并且未加删除、修正或改动。随着计算机技术的发展,人们为了追求使用信息的方便性、直观性与可视效果,往往采用多种格式与多种技术来产生数字对象。目前,文字与图形、声音、影像等多种媒体信息能够完美组合在一起构成数字信息的内容。与传统文献相比,数字文献的信息形态更具多样性和复杂性。一个数字信息不再仅仅由一个独立的文件组成,而是由一组具有独立格式和内容的文件构成。由于信息与载体具有分离性,当载体处于联机状态时也许不会出现问题,而如果说这些载体都是脱机保存的,则有可能造成混乱,处理不当会直接影响到数字信息的完整性。网络信息资源的完整性是指针对网络信息资源的多样性进