

数字信息的长期保护问题

张晓林

文摘：本文简要分析数字信息长期保护的挑战和任务，介绍数字信息长期保护责任体系、保护系统结构、保护技术体系和长期保护元数据。

关键词：数字信息，长期保护，OAIS，数据更新，技术仿真，数据迁移，元数据

Preservation of Digital Information

Abstract: The paper briefly analyses the challenges and tasks of digital information preservation, describes the preservation responsibility mechanisms, preservation system structures, preservation techniques, and metadata.

Keywords: Digital information, Preservation, OAIS, Refreshing, Emulation, Migration, Metadata

随着数字信息系统逐步成为科研、教育、政府管理以及企业经营的主流运作环境，解决数字信息的长期保护问题已迫在眉睫，本文主要结合数字化学术信息的长期保护，对相关问题进行讨论。

1. 数字信息长期保护的挑战

根据美国加州大学对1999年全球数字信息发展的调查测算^[1]，全世界每年生产 1×10^{18} 到 2×10^{18} 字节(Exabytes)信息，其中所有印刷载体的信息量只占0.003%，磁介质(包括数字光学介质)所承载的信息量(即数字信息)达93%。而且由于它们增长速度几乎为100%，正迅速成为主流存储介质。而且网络中每年传递的Email达到11285TB(Terabytes, 1TB= 10^{12} 字节)，Usenet信息达到73TB，不重复公共网页21亿(每个网页平均10K数据，每天将会增加7百万新网页)，2001年网页数量将会达到40亿。并且众多网络数据库还以非网页形式存储了7500TB数据，其中包含5500亿条不同记录或文件。

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，它的长期保护面临许多新的问题^[2-3]，例如：

- a. 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微缩胶片介质。
- b. 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。
- c. 数字信息往往是动态变化的，例如一个动态出版并融入读者批注功能的电子期刊或一个不断更新的网站，因此对数字信息单元固定化(Fixity)、真实性考证、版本确认等带来很大困难。
- d. 数字信息往往是相互链接的，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，因此对数字信息单元的内容范围确认和版本内容组织、存储等造成很大困难。
- e. 由于数字信息在其产生和利用过程中的灵活性和动态性，致使数字信息单元的起源及其归属变化信息(Provenance)难以追踪和清晰理解。
- f. 数字信息交流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，因此迫切需要重新定义和建立数字信息长期保护的责任体系。
- g. 数字信息的知识产权法律和管理机制尚不完善，因此对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、备用服务以及相应的经济与管理问题还有待解决。
- h. 数字信息的庞大生产量将使上述问题更加突出，也使解决这些问题变得更加迫切。

2. 数字信息长期保护框架

为确定数字信息长期保护的政策与技术框架，需要全面分析数字信息长期保护环境，

(1)数字信息来源。

数字信息环境的一大特点是信息生产的方便性和由此带来的多元性^[4]。从原始数据形态来看,数字信息来源可以是由其它介质形态转移为数字形态(例如印刷图书的数字扫描或OCR版本),或直接由数字信息处理系统(例如文字处理软件、CAD、数字摄像设备、地理扫描设备)产生(称为 Born-digital)。从数据生产者角度,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,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。

(2)数字信息类型。

在数字信息长期保存中,人们面临多种信息类型和数据类型^[5](表1),它们各自有不同的编码机制以及相应的数据格式、处理软件、甚至硬件系统。

数字信息类型	数据类型	应用领域
数据集	文字数字	科学调查实验、事务或管理数据、著录数据
格式化文本	格式化文字数字	各种出版物、文本文件、网页等
多媒体文件	格式化文字数字、点阵或矢量图像	各种出版物、多媒体文件、网页、课件等
图表	矢量或点阵图像、文字数字	办公文件、商业文件、各种出版物等
设计数据	矢量或点阵图像、文字数字	计算机辅助设计
图像	点阵图像、文字数字	艺术、医学、建筑、历史与档案管理 等
音乐与语音	数字音频数据	语音处理、音乐、数字广播、音乐图书馆
影像	数字视频数据	数字录像或电视、影像图书馆、教育培训
地理空间数据	矢量和点阵数据、文字数字数据	地图、地理信息系统、空间测绘等
交互式多媒体出版物	交互插入的图表、图像、影像、文字与数值数据	交互式电子出版物、教育培训系统、商业营销系统、电子游戏等

表 1

(3)数字信息长期保护的的任务

由于数字信息依赖众多的不同的技术、方法、工具和管理机制才能被人们所利用,因此数字信息长期保护实际上需要保护数字信息本身及相关技术、方法和工具,这包括解决以下问题:

- a. 保存数字比特流,即通过对数字存储媒介的保护或转移,确保存储其中的物理数字数据能被准确完好地读出。
- b. 保存数字格式与处理信息,即通过保存有关数字信息编码、格式、标记、结构、压缩、加密等方面的技术方法信息,确保能够识别和解析数字信息内容。
- c. 保存数字信息处理环境,例如相关的软件甚至硬件系统,确保能拥有相应的技术工具来识别、利用数字信息。
- d. 保存数字信息的内容校验、身份认证、版本演变、知识产权管理机制,确保能可信赖、可靠和合法地鉴别使用被保存的数字信息。
- e. 保存数字信息的知识组织体系信息,保存数字信息的组织利用环境

(4)数字信息长期保护的环境框架

上述任务的解决需要一系列支撑机制,我们用图1给出这些机制的一种可能组合。

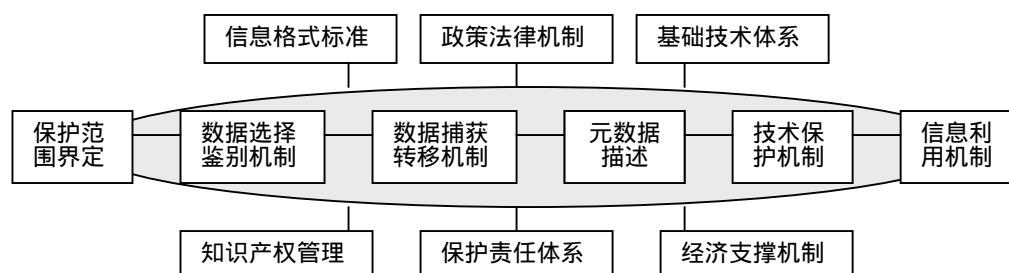


图 1

(5) 主要数字信息保护研究项目

许多国家和机构已经在学术性数字信息保护方面进行了深入的研究，其中影响最大的包括：

CEDARS (CURL Exemplars in Digital ARchiveS)^[6]，英国 JISC 资助，牛津、剑桥、里兹大学参加，致力于确认和描述有关数字信息收藏体系长期保护的策略框架和具体方法。

NEDLIB (Networked European Deposit Library)^[7]，由欧洲七个国家图书馆(荷兰、法国、挪威、德国、葡萄牙、瑞士、意大利)和三家主要出版社(Kluwer、Elsevier、Springer-Verlag)参加，主要研究数字版本存缴机制和长期保护系统机制。

PANDORA(Preserving and Accessing Networked Documentary Resources of Australia)^[8]，澳大利亚国家图书馆进行的有关选择、收集、保护和长期提供澳大利亚电子出版物的政策和程序体系的研究。

ERA (Electronic Records Archives Program)^[9]，美国国家档案管理局进行的关于建立数字记录和档案长期保护的完整体系结构。与此密切相关的是由美国匹兹堡大学进行的电子记录功能需求研究^[10]和由加拿大布列颠哥伦比亚大学进行的电子记录完整性研究^[11]。

PRISM(Preservation、Reliability、Interoperability、Security、Metadata)^[12]，由美国康乃尔大学承担，研究在分布式数字图书馆体系中保护数字信息完整性所需要的政策和机制。

PAERM (Persistent Archives and Electronic Records Management)^[13]，由美国国家档案与记录管理局和圣地亚哥超级计算中心联合进行，主要研究海量数字信息的技术平台迁移机制。

CAMiLEON(Creative Archiving at Michigan and Leeds: Emulating the Old on the New)^[14]，由美国密西根大学和英国里兹大学联合承担，主要研究应用仿真技术来保存数字信息处理环境的经济模型、技术工具、政策指南等。

Data Provenance Project^[15]，由美国宾西法尼亚大学承担，研究跟踪记录数字信息起源及归属变化所涉及问题、方法和工具。

LOCKESS (Lots of Copies Keeps Stuff Safe)^[16]，由美国斯坦福大学承担，并有哥伦比亚大学、哈佛大学、加州大学伯克利分校、田纳西大学、拉斯阿拉莫斯国家实验室等参加，研究通过在不同计算机上存储多份复本来保护数字信息的可行性。

3. 数字信息保护责任体系及保护内容选择

目前，许多发达国家正在积极研究建立数字信息保护责任体系及相应的选择标准，以改变数字信息保护的自流状态，争取实现系统、有序和可靠的保护。

美国研究图书馆协会提出建立分布式数字信息保护责任体系^[2]，其中提出：

- a. 数字信息生产者或提供者首先承担各自数字信息产品的存储和保护责任，作为数字信息保护的第一道防线，从而确保这些数字信息有可能被长期保存；
- b. 建立足够数量的可信赖的有足够技术能力的分布式的数字档案库来覆盖各类数字信息的长期保存，这些数字档案库需要经过严格的审查注册程序以保证其可信赖性和基本技术与管理能力；
- c. 数字信息生产者或提供者可以委托经过注册的数字档案库(Digital Archives)承担自己的部份甚至全部保护责任；图书馆或档案机构可以通过合同形式成为生产者或提供者的代理来管理它们的数字档案库；
- d. 在数字信息生产者或提供者不愿或不能承担、也不转移自己数字信息保护责任时，经过鉴定和注册的数字档案库有责任并被赋予权力作为后备保证机制(Fail-safe mechanism)采取主动抢救性措施来收集和保存相应数字信息；
- e. 数字档案库维护数字信息所有者提供相关信息服务并获取相应利益的权利，但在信息所有者因种种原因放弃提供服务时有责任并被赋予权力作为后备保证机制继续向公众提供有关服务。

美国全国科学院(National Academy of Sciences)在关于美国国会图书馆在数字化时代的任务的分析报告^[17]中提出，国会图书馆应积极承担数字信息的长期保护任务，并具体要

求：

- a. 国会图书馆应明确定义它将长期保存和提供的数字信息类型和范围；
- b. 国会图书馆应与其它机构合作来定义不在它自己长期保存范围的那些数字信息的保护和提供服务责任体系；
- c. 国会图书馆应在经过选择的领域建立门户机制(包括通过商业服务系统的链接机制),以便能提供对更大范围数字信息的链接和服务,并建立必要的法律、技术和管理保障以便在外部系统不履行义务时承担起相应的收集和服务责任；
- d. 国会图书馆应利用自己独特地位索取必要的法律支持来建立数字版本存缴制度,并争取信息生产者同意自己作为后备保证机制,以便在信息生产者无法履行义务时提供服务；
- e. 国会图书馆应积极争取澄清在版本存缴制度下进入和收集美国网站信息的法律权力,积极建立相应搜寻与存储实验系统,并尽快将实验系统经验转化为合适的美国网站信息收集政策机制。

澳大利亚国家图书馆已经建立具体的数字信息长期保护指南^[18],提出：

- a. 国家图书馆和各公共图书馆应积极推动版本存缴制度的法律改革,争取通过现有版本存缴机制收集和存储各类正式的数字出版物和政府机构数字信息产品；
- b. 国家图书馆希望与各省公共图书馆建立协作关系,由国家图书馆负责收集具有全国意义的数字信息,由各公共图书馆负责收集对本地区有意义的数字信息；
- c. 国家图书馆将主要收集那些关于澳大利亚或由澳大利亚作者生产的、权威的、具有研究价值的以数字形态为主的数字信息,目前主要收集 WWW、gopher 和 ftp 站点资源,主要采取对经过选择的站点定期进行拷贝的形式。

美国国家医学图书馆则制定了关于它自己生产的各类数字信息(例如数据库、网页等)的长期保护级别鉴定政策^[19],分析建立各类信息的长期保护级别、记载交换保护级别的方法、以及保障相应保护措施得以实施的程序。为了保障用户能连接和获取信息,NLM 保护级别体系考虑了三类对象：

国家医学图书馆数字信息保护级别	
信息内容类型	默认保护级别
通告通知类	不保证长期保存
申请表格式	不保证长期保存
书目	[无默认级别]
临床警戒通知	长期保存静态内容
数据库	[无默认级别]
数据库记录	长期保存稳定记录
数字图书馆图像	长期保存静态内容
数字图书馆收藏	[无默认级别]
展览数据	长期保存静态内容

性；

- a. 数字信息标识符,可能是过渡性或永久性；
- b. 数字信息可获得程度,可能不予保障或必须保障长期可获得；
- c. 数字信息内容的固定程度,包括持续变化的或者已经停止变化的动态内容、可能变化的或者已经停止变化的稳定内容、不会变化的静态内容。

对于不同类型的数字信息,NLM 建立了基本的默认保护级别。对于无默认级别的信息类别根据实际信息内容确定保护级别。创建相应数字信息内容的 NLM 工作人员将根据这些政策具体确定实际的保护级别。

4. 数字信息保存系统

美国空间数据系统咨询委员会于 1999 年提出的开放档案信息系统参考模型(Reference Model for an Open Archive Information System)^[20]已日益被接受为数字信息长期保存系统基本构架,并已作为 ISO 标准草案,图 2 给出 OAIS 系统基本结构。

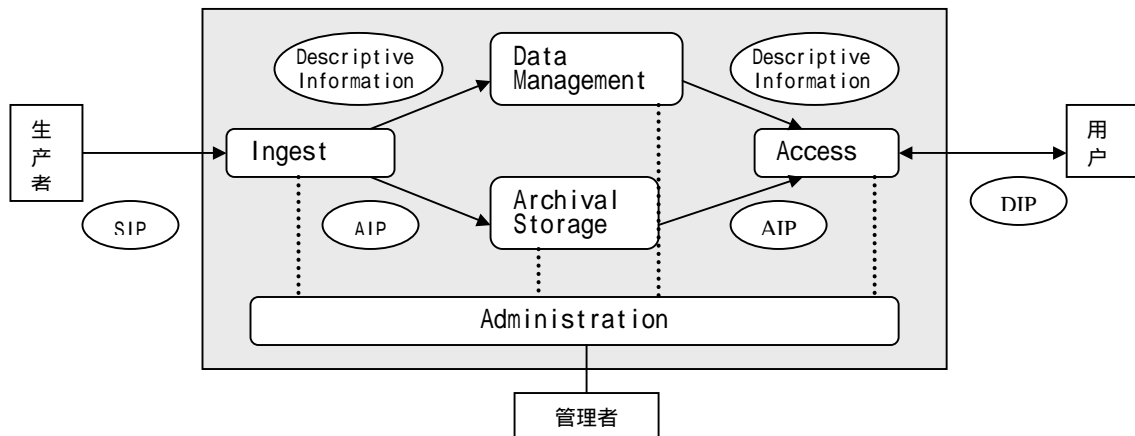


图 2

OAIS 系统由五个主要功能模块组成，其中：

a. 摄取模块(Ingest)从出版商或其它信息提供者处收集或接收按照一定格式组织的存交信息单元(Submission Information Package, SIP)，这些信息单元经过检验后，建立相应元数据，元数据交给数据管理模块，信息单元被转换为按照长期保护规定格式组织、包含专门的长期保护处理数据的存储信息单元(Archival Information Package, AIP)，然后交长期存储模块。

b. 长期存储模块(Archival Storage) 实际存储按 AIP 要求组织的数字信息，负责数据更新、技术仿真或数据迁移，负责建立具体存储与存取系统(例如梯次存储系统)，并在检索传递模块要求时将 AIP 提供给该模块。在技术仿真或数据迁移过程中，可能形成新的数字内容单元，可能需要与摄取模块协作重新建立有关元数据并送交数据管理模块。

c. 数据管理模块(Data Management)存储关于数字信息单元的元数据和关于长期保护处理政策、程序、技术和系统的元数据，并提供对这些元数据的基础检索与管理。

d. 检索传递模块(Access)提供用户检索元数据和索取数字信息单元的界面，提供检索机制，并将 AIP 转换为适合用户利用的传递信息单元(Dissemination Information Package, DIP)，还可能承担身份认证和授权管理等。

e. 系统管理(Administration)模块通过有关政策、规范、程序、工作流等来监测和控制整个长期保存系统的运行和各个模块的运行。

欧洲 NEDLIB 项目还研究了这类长期保存系统与其上层数字图书馆体系的关系，提出在欧洲数字版本存缴系统(DES)中建立相应的系统界面^[7]，其中一是位于生产者、元数据系统与长期保存系统间的信息捕获和传递界面，负责按照商定的格式和程序从生产者系统接收数字信息或从元数据系统检索接收元数据，传递给摄取模块或数据管理模块。二是位于长期保存系统与用户、元数据系统、其它数字图书馆系统等之间的封装和传递界面，负责按照使用对象需要析取、转换、组织信息，并传递给使用对象系统。

5. 数字信息保存技术

根据数字信息保护的要求，数字信息保存技术主要分为^[5, 21-22]：

(1) 数据更新(Refreshing)

数据更新指通过拷贝方法将数据流从旧存储介质转移到新存储介质上，保护数据本身不受存储介质质量恶化的影响。拷贝过程往往需要周密计划，以便有序地将不断积累的数字信息在它们各自存储介质可能超过安全存储期限前转移到新的存储介质上；而且，在拷贝过程中往往需要通过校验和(Checksum)或数字摘要(Digital digest)等技术来检验拷贝后数据流与原始数据流一致。但是，有些数字信息可能要求专用软件才能读取；有些数字信息虽然可直接转移存储但可能丢失相关的结构、链接或环境关系信息，或者没有同时和相关联地转移有关编码、压缩、加密信息。而且，数字信息的继续使用依赖相应软件、而这些软件又依赖于专门操作系统甚至硬件系统，如果这些软件及硬件系统没有同时保存，人们仍然不能利用保存下的数据内容。另外，目前数字技术的变化速度远远超过物理介质质量恶化速度，即使原始数据介质还完好无损，有关软件硬件可能早已过时甚至消失。因此，主要依赖数据更新来保存数字信息存在很大的危险性。

(2) 技术仿真(Emulation)

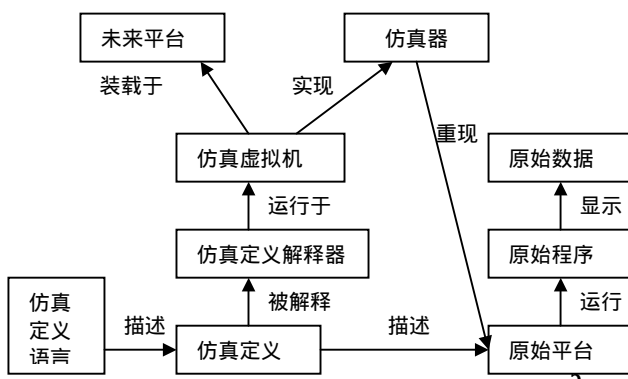


图 3

技术仿真是在实施数据更新保存数字信息内容的同时，通过保存数字信息利用环境(软件与硬件)来保障数字信息内容的可利用性。当然，我们可以物理地保存有关的软件版本和硬件设备，但由于计算机软硬件的迅速升级换代，对于绝大多数机构来说这种方法都不现实。实际上，人们更多

地是详细描述数字信息利用的各种技术参数和环境条件,以便未来计算机系统可以模仿原始系统环境来读取和使用数字信息内容。图3给出其中一种技术线路,仿真定义数据用标准的仿真定义语言详细说明原始平台和程序的条件参数,仿真定义解释器具体解析用文件或程序形式表示的仿真条件,仿真虚拟机作为未来平台的嵌入程序运行相应的解释器、从而实现仿真操作。由于我们总可通过新的仿真器来仿真上一代仿真虚拟机,因此不必持续更新针对特定数字信息的定义语言、仿真说明、解释器等。但是,技术仿真要求较高的和持续的技术水平和运行支持能力,这往往是一般机构难以独立做到的。

(3) 数据迁移(Migration)

数据迁移是持续地将数字信息从旧的计算机软硬件环境(例如旧版本旧格式)转换到新计算机环境(新版本新格式)的过程,从而保证当前存储的数字信息总能被当前系统读取。其中,也可以将原始数据格式转换为独立于具体原始软硬件平台的标准数据格式,例如XML格式,由于这些格式的开放性和标准性,往往可以保证总有相应软硬件系统来处理这些数据。数据迁移中并不要求绝对保存原始数据的全部“面貌”(Look and Feel),而主要是保存数据内容和内容关系。但也正因为每一次转换都可能丢失一定的结构、版式、链接、交互关系等信息,持续的转换可能将这些信息损失累积起来,从而使当前数字信息内容(包括结构、关系等)出现严重差异。另外,数据迁移的工作量大(每一新版本出现都带来新的迁移),尤其对于大规模数据系统(例如连续空间测绘系统和气象数据系统),由于不断累积的巨大数据量,数据迁移本身的操作负担可能远超过存储系统负担。

除了上述方法之外,还可通过将数字信息转换为模拟信息(例如打印)、使用后向兼容软件、提高各类系统互操作性等来保护数字信息,但目前趋势是将技术仿真和数据迁移作为数字信息保护的主要技术基础,当然具体选择什么技术取决于待保护的数据类型、数据格式、数据内容结构和关系复杂性、原始软硬件平台以及数字信息资源管理机制。

6. 数字信息长期保护的元数据

元数据作为描述数字信息本身及其保护环境的数据,在数字信息长期保护中起着重要作用。根据OAIS系统的信息模型,长期保护中涉及的元数据内容包括^[20,23]:

a. 内容信息(Content Information),包括数据内容本身和数据内容表现元数据,涉及编码、格式、显示样式、压缩、色彩等方面的技术和系统要求数据。实际上,保护描述、封装和内容描述数据也有相应表现信息。

b. 保护描述信息(Preservation Descriptive Information),包含确认信息(Reference),例如数字对象标识符;起源信息(Provenance),描述数据内容产生保存链;环境信息(Context),描述数据对象与其它数据对象或信息系统的关系;固化信息(Fixity),描述用以确认信息内容完整性的信息,例如校验和、数字摘要等。

c. 封装信息(Packaging and Representation Information),涉及将各种信息内容组合封装为一个完整独立信息单元的有关数据。

d. 内容描述信息(Descriptive Information),包括各种基本著录数据,支持信息内容发现。

目前在数字信息长期保护方面还没有统一或通用元数据格式,但有关研究项目已提出若干元数据格式,例如CEDARS^[24]、NLA^[25]、NEDLIB^[26]、LC Digital Repository^[27]、RLG^[28]等,表3给出CEDARS元数据的元素结构。

Preservation Description Information	Actors
Reference Information	Actions Permitted by statute
Resource description	Legislation text pointer
Existing metadata	Permitted by license
Existing records	License text pointer
Context Information	Fixity Information
Related information objects	Authentication indicator
Provenance Information	Content Information
History of origin	Representation Information
Reason for creation	Structure Information
Custody history	Underlying abstract form description
Change history before archiving	Transformer objects
Original technical environments	Platform
Prerequisites	Parameters
Procedures	Render/analyze engines
Documentation	Output format
Reason for preservation	Input format
Management history	Render/analyze/convert objects
Ingest process history	Platform
Administration history	Parameters
Action history	Render/analyze engines
Policy history	Output format
Rights management	Input format
Negotiation history	Semantic Information
Rights information	Render/analyze objects
Copyright statement	Platform
Name of publisher	Parameters
Date of publication	Render/analyze engines
Place of publication	Output format
Rights warning	Input format
Contacts or rights holders	Data Object

7. 结束语

由于数字信息长期保护涉及许多方面的复杂问题，因此在建立、接收、连接、利用和保存数字信息时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护的影响，建立系统化的策略体系和操作规范^[4]。例如，采用标准、开放的数据处理技术和格式，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。实际上，图书馆情报系统馆藏建设的许多原则和经验都可应用于此。

参考文献：

- [1] P. Lyman and H. Varian. How Much Information? 2000
<http://www.sims.berkeley.edu/research/projects/how-much-info/>
- [2] J. Rothenberg. Ensuring the Longevity Of Digital Documents. Scientific American, January 1995, p42-47.
- [3] RLG/CPA Task Force on Archiving of Digital Information. Preserving Digital Information: Final Report and Recommendations (May 1996) <http://www.rlg.org/ArchTF/index.html>
- [4] N. Beagrie and D. Greenstein. A Strategic Policy Framework for Creating and Preserving Digital Collections. Version 4.0, July 1998. <http://www.ahds.ac.uk/manage/framework.htm>
- [5] Comparison of Methods & Costs of Digital Preservation
<http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>
- [6] CEDARS (CURL Exemplars in Digital ARchives) <http://www.leeds.ac.uk/cedars/>
- [7] NEDLIB (Networked European Deposit Library) <http://www.konbib.nl/coop/nedlib/>
- [8] PANDORA (Preserving and Accessing Networked Documentary Resources of Australia)
<http://pandora.nla.gov.au/pandora/>
- [9] ERA (Electronic Records Archives Program) <http://www.nara.gov/era/index.html>
- [10] Functional Requirements for Evidence in Recordkeeping <http://www.lis.pitt.edu/~nhprc/>
- [11] The Preservation of the Integrity of Electronic Records <http://www.interpares.org/UBCProject/>
- [12] PRISM <http://www.prism.cornell.edu/main.htm>
- [13] PAERM <http://www.sdsc.edu/NARA/>
- [14] CAMiLEON <http://www.si.umich.edu/CAMiLEON/>
- [15] Data Provenance Project <http://db.cis.upenn.edu/Research/provenance.html>
- [16] LOCKESS (Lots of Copies Keeps Stuff Safe) <http://lockss.stanford.edu/>
- [17] LC21: A Digital Strategy for the Library of Congress. <http://www.nap.edu/html/lc21/>

- [18] Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library of Australia. <http://pandora.nla.gov.au/selectionguidelines.html>
- [19] M. Byrnes. Assigning Permanence Ratings to NLM's Electronic Publications. <http://www.rlg.org/events/pres-2000/byrnes.html>
- [20] Reference Model for an Open Archival Information System (OAIS). May 1999. <http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf>
- [21] R. Moore et al. Collection-Based Persistent Digital Archives, D-Lib Magazine, March 2000, 6(3)[Part 1] <http://www.dlib.org/dlib/march00/moore/03moore-pt1.html>
6(4)[Part 2] <http://www.dlib.org/dlib/april00/moore/04moore-pt2.html>
- [22] J. Rothenberg. An Experiment in Using Emulation to preserve Digital Publications, 2000 <http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf>
- [23] Preservation Metadata for Digital Objects: A Review of the State of Arts http://www.oclc.org/digitalpreservation/presmeta_wp.pdf
- [24] Metadata For Digital Preservation: The Cedars Outline Specification <http://www.leeds.ac.uk/cedars/MD-STR-5.pdf>
- [25] National Library of Australia: Preservation Metadata for Digital Collections. 15 October 1999. <http://www.nla.gov.au/preserve/pmeta.html>
- [26] Networked European Deposit Library (2000). Metadata for Long Term Preservation. <http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>
- [27] Library of Congress Digital Repository Development Core Metadata Elements <http://lcweb.loc.gov/standards/metadata.html>
- [28] RLG Working Group on Preservation Issues of Metadata. Final Report. May, 1998 <http://www.rlg.org/preserv/presmeta.html>

本文最初发表在《图书馆》2001年第5期第7-12页。