

数字信息资源长期保存研究

SHUZI XINXI ZIYUAN CHANGQI BAOCUN YANJIU

主编 • 何晓庆

副主编 • 王圣洁 张雅晴 孙诗 蒋旭

非
外
借



四川大学出版社

数字信息资源长期保存研究

SHUZI XINXI ZIYUAN CHANGQI BAOCUN YANJIU

主编 • 何晓庆

副主编 • 王圣洁 张雅晴 孙诗 蒋旭



四川大学出版社

责任编辑:梁 平
责任校对:杜 彬
封面设计:璞信文化
责任印制:王 炜

图书在版编目(CIP)数据

数字信息资源长期保存研究 / 何晓庆主编. —成都:
四川大学出版社, 2018. 7

ISBN 978-7-5690-2093-9

I. ①数… II. ①何… III. ①数字技术—信息管理—
研究 IV. ①G203

中国版本图书馆 CIP 数据核字 (2018) 第 161984 号

书名 数字信息资源长期保存研究

主 编	何晓庆
出 版	四川大学出版社
地 址	成都市一环路南一段 24 号 (610065)
发 行	四川大学出版社
书 号	ISBN 978-7-5690-2093-9
印 刷	成都金龙印务有限责任公司
成品尺寸	148 mm×210 mm
印 张	10.625
字 数	285 千字
版 次	2018 年 8 月第 1 版
印 次	2018 年 8 月第 1 次印刷
定 价	48.00 元



版权所有◆侵权必究

◆读者邮购本书,请与本社发行科联系。

电话:(028)85408408/(028)85401670/
(028)85408023 邮政编码:610065

◆本社图书如有印装质量问题,请
寄回出版社调换。

◆网址:<http://press.scu.edu.cn>

序

随着 Internet 和计算机技术的迅猛发展,网络作为提供各种各样信息资源的全球化平台,正以一种史无前例的方式改变着人们生活的方方面面。网络已经成为人们搜索、利用信息资源的一个重要途径,呈几何级增长的信息资源以数字化的形式存在于网络之上,因此这些存放于网络之上的海量数字信息资源也成为非常宝贵的信息财富。但是,由于计算机技术的不断发展、网页内容的不断更新、软硬件的不断升级,再加上诸如机器意外损坏、病毒木马及意外灾害对数据的破坏等,使得这些网络信息资源消失的速度也相当之快,有价值的海量数字信息资源面临着随时都有可能突然永久消失的危险。如何通过有效的技术策略来保存各种各样的海量数字信息资源,使其不再面临消失的危险,能够得到长期有效的保存,供人类现在以及将来使用,已经成为一个日益受到学界关注,并且很有必要进行深入探讨研究的重要课题。

本书通过对数字信息长期保存领域现状的探讨,将重点更多地放在了法律、经济、技术策略以及风险管理、综合评价方法上,希望能对该领域的研究做出微薄的贡献。

本书第 1 章数字信息资源长期保存概论由何晓庆负责撰写,第 2 章数字资源长期保存影响因素由王圣洁负责撰写,第 3 章数字资源长期保存风险管理由孙诗负责撰写,第 4 章数字资源长期保存法律策略由张雅晴负责撰写,第 5 章数字资源长期保存技术

策略由何晓庆负责撰写，第6章数字资源长期保存的经济属性由王圣洁负责撰写，第7章云存储在数字信息资源中的应用由何晓庆负责撰写，第8章数字资源长期保存综合评价方法研究由蒋旭负责撰写。

如何在现有的信息技术条件下很好地实现数字信息资源的长期保存是一个引起社会高度重视的问题，我们应当让更多的力量参与到这个领域中，以期让这些宝贵的数字信息资源能够长期保存下去，为后代所用，为文明留痕。

目 录

1	数字信息资源长期保存概论	(1)
1.1	数字信息资源的概念	(1)
1.2	数字信息资源的特点	(3)
1.3	数字信息资源长期保存的主要内容	(4)
1.4	数字信息资源长期保存目标	(5)
2	数字信息资源长期保存影响因素	(8)
2.1	数字信息资源长期保存技术政策因素分析	(8)
2.2	我国数字信息资源长期保存政策制定现状	(17)
2.3	制定网络信息长期保存政策的原则	(24)
3	数字信息资源长期保存风险管理	(29)
3.1	数字信息资源长期保存风险	(30)
3.2	数字信息资源长期保存风险的识别与分析	(32)
3.3	数字信息资源长期保存风险的衡量与评估	(38)
3.4	数字信息资源长期保存风险对策选择	(53)
3.5	数字信息资源长期保存风险管理策略的选择过程	(58)
3.6	SPOT 模型在数字信息源长期保存风险管理中的 应用	(59)
3.7	案例分析	(75)

4	数字信息资源长期保存法律策略	(79)
4.1	数字信息资源长期保存面临的法律困境	(79)
4.2	数字信息资源长期保存中面临的知识产权问题 ...	(80)
4.3	国内外数字信息保存机制的版权问题解决方案分析	(95)
4.4	解决我国数字信息资源保存中的版权问题可采取的 对策	(120)
4.5	结语	(136)
5	数字信息资源长期保存技术策略	(140)
5.1	数字信息资源长期保存技术影响因素分析	(140)
5.2	数字信息资源长期保存技术策略分析	(152)
5.3	数字信息资源长期保存标准体系	(169)
5.4	一种基于 OAIS 的网络信息长期保存技术模型 ...	(172)
5.5	基于 OAIS 的数字信息资源长期保存技术策略模型 ——TMWRLP	(185)
5.6	基于 TMWRLP 模型的“5·12”汶川地震数字信息 资源长期保存实验	(202)
6	数字信息资源保存的经济属性	(218)
6.1	数字信息资源保存的基本经济属性	(218)
6.2	数字信息资源保存的经济方面风险与对策	(225)
6.3	数字信息资源保存供应的风险与对策	(227)
6.4	实现数字信息资源保存经济可持续性的原则	(233)
6.5	实现数字信息资源保存经济可持续性的对策	(241)
7	云存储在数字信息资源中的应用	(245)
7.1	云存储的性能	(245)
7.2	云存储在数字信息资源长期保存方面的发展前景 分析	(261)

7.3 云存储在数字信息资源长期保存方面的策略分析	(263)
7.4 国际应用案例研究	(272)
8 数字信息资源长期保存综合评价方法研究	(280)
8.1 数字信息资源长期保存评价研究现状	(280)
8.2 常用评价方法	(294)
8.3 数字信息资源长期保存综合评价方法	(308)

1 数字信息资源长期保存概论

1.1 数字信息资源的概念

数字信息资源是指以数字化形式记录的，以光介质、磁介质等多种形式表达出来的，并通过网络通信、计算机或终端再现出来的资源，其表现形式可以是文字、图像、声音、动画等多种形式信息资源的集合。简言之，数字信息资源就是数字形式的多种类型信息资源。相较于传统的文献资源，数字信息资源在技术上具有较大的进步并具有显著的特点。

1.1.1 数字信息资源长期保存的含义

数字信息资源长期保存是指为保证数字信息资源可长期存储和其内容可长期获取而进行的必要的管理活动，其具有两层含义：一是可供长期存储（storage），一是可供长期可获取（access）。而“长期”并非日常所指的长时间，属于一个概数；在数字信息资源中，“长期”是指数字信息资源被社会所需要的时间，是一个相对比较具体的时间。数字信息资源的长期保存与数字资源的存储属于两个不同的概念，存储通常以短时间使用为

目的,并不对其进行保护和长时间之后的获取有硬性的要求。而数字信息资源长期保存则是为了长期维护数字资源的真实性、可靠性、可用性、完整性以及可读性,避免因意外或者载体原因或软硬件过时的损坏而导致的数字信息资源的消失,确保人们能在所需的时间内进行获取。

1.1.2 数字信息资源的类型

数字信息资源多种多样,根据不同的标准可以将数字信息资源划分为不同的类型,如表 1-1 所示。

表 1-1 数字信息资源的类型

划分标准	类型	定义
媒体形式	文本信息	关于初始文献的数字信息资源
	图片信息	关于图片的数字信息资源
	视频信息	以电信号的方式加以捕捉、记录、处理、存储、传送与重现的信息
保存和获取的方式	邮件型	用电子邮件的方式发布的网络信息资源
	揭示板型	各种网络新闻和未命名的文件传输协议
	书目型	用于检索数字信息资源的各种检索工具
交流方式	非正式出版物	指流动性和随意性都较大且信息的数量也较大,而信息的质量却难以保证和控制的动态性数字信息资源
	半正式出版物	指受到一定知识产权保护,但是并没有被纳入正式出版信息系统的数字信息
	正式出版物	指受到一定的知识产权保护,信息质量相对可靠,利用率也较高的数字信息

续表

划分标准	类型	定义
资源内容	事实型	包括诸如网站播报的天气预报和节目预告等数字信息资源
	数值型	包括各种商品的价格和产品的规格等数字信息资源
	全文型	包括各种电子期刊和电子杂志、电子报纸、会议论文等数字信息资源
	实时型	包括网上聊天、网络会议等在内的数字信息资源

1.2 数字信息资源的特点

那些发布、传递和存储都使用计算机网络的种种信息资源的综合被称为数字信息资源。典型的数字信息资源包括：

- (1) 网上的电子图书、综合数据库等；
- (2) 政府机构发布的会议信息、消息法规、通知通告等；
- (3) 科研机构发布的论文集、研究成果、学术交流信息等；
- (4) 其他存放于网络上的动态数字信息如 E-mail、用户组等。

数字信息与传统印刷信息的特点对比如表 1-2 所示。

表 1-2 数字信息与传统印刷信息特点对比

	数字信息	传统印刷信息
设备依赖性	依赖性强。信息的创建、存储、传输等过程都需要通过计算机等外部设备来实现	依赖性较弱。只在信息创建的时候会用到相关印刷设备

续表

	数字信息	传统印刷信息
安全性	安全性较弱。除自然灾害之外，还可能遭受计算机病毒的威胁等	安全性较强。绝大多数情况下，面对的威胁都是自然灾害
真实性	真实性难确保。由于易遭受人为篡改，对于数字信息的持续完整和出处有较大难度	真实性易于确保。由于客观原因，印刷体信息的真实性容易得到保障
存储介质性能	绝大部分数字信息的存储介质天生就具有不稳定性，如果没有相适应的存储环境和管理会导致迅速失效	如果得当保存，纸质印刷品可以保存很长时间

由于数字信息资源的所有内容都是存放在各种网页、网站之上，所以数字信息资源还具有一些自己的独特特点，比如说网页寿命短暂、数字信息资源更新速度快，信息量巨大且多样，并且数字信息资源大都处于相对不集中管理状态，等等。

从表 1-2 及数字信息的相关特点我们看到，虽然数字信息已经非常普及，但是在信息存储和长期保存方面都存在很大的问题，很有必要对数字信息资源的长期保存进行深入研究。

1.3 数字信息资源长期保存的主要内容

数字信息资源长期保存的内容主要包括以下几个方面：

(1) 保存数字流，即通过对数字存储媒介的保护或转移，确保持存储在其中的物理数字信息能被准确完好地读出。

(2) 保存数字格式与处理信息，即通过保存有关数字资源编码、格式、标记、结构、压缩、加密等方面的技术方法的文件，确保能够识别和解析数字资源的内容。

(3) 保存数字资源的处理环境,如相关软件和硬件系统,确保能使用相应的技术工具来识别、利用数字资源。

(4) 保存数字资源的内容校验、身份认证、版本演变、知识产权管理机制等,确保能可靠、合法地鉴别和使用被保存的数字资源,保存数字资源的知识组织系统,保存数字资源的组织利用环境。

1.4 数字信息资源长期保存目标

对于数字信息资源,它的长期保存目标可以从以下几个方面来进行考虑。

1.4.1 数字信息资源的可读性

目前,由于数字信息资源的存储介质存在易碎、易损、易变质等问题,其使用寿命通常为几十年,并且随着时间的推移存储介质的可靠性会越来越低。哪怕数字信息资源的管理者采用了多种方式来对这些存储介质实施保护,但终究逃脱不了被更换的命运。存储介质的损坏将对数字信息资源的可读性产生很大的冲击。数字信息的可读性主要体现在能够对保存的数字信息资源根据所需信息进行查找、检索,满足人们对信息的有效使用。

1.4.2 数字信息资源的可用性

数字信息资源的可用性是指数字信息资源可以被人或计算机阅读和处理。借助计算机数字信息资源实现了从形成、传输到存储的整个过程,因此数字信息资源天然与计算机系统中的各种设备有着千丝万缕的紧密关系。计算机的各类软件与硬件设备是数

字信息资源产生的前提和基础,因此也决定了数字信息资源对计算机的软、硬件系统具有较强的依赖性。解决了原来的数字信息资源如何更好地应对新的电子环境的问题,也就意味着推动了数字信息资源的持久可用性。

1.4.3 数字信息资源的可理解性

数字信息资源的可理解性是指数字信息资源的内容可以被它的使用者所理解。无论是传统文件还是电子文件,都主要由格式与内容组成。传统文件的格式与内容固定在某一种载体例如纸张上,并不需要借助额外的辅助工具就可以使人理解;而数字信息资源的内容必须借助特定的软件才能供人阅读并理解。但是由于技术更新换代的速度日益加快,这些软件非常容易被新的版本所取代或者消失,这无疑造成了大量的数字信息资源由于内容或格式的丢失而产生了不可理解性。

1.4.4 数字信息资源的完整性

数字信息资源的完整性是指数字信息资源是齐备的,并且未加额外改动的。随着计算机技术的迅猛发展,对于数字对象的产生往往采用多种格式与多种技术相结合的方式以提高信息的方便性、直观性与可视性。与传统文献相比,数字信息资源能将文字、图形、声音、影像等多种类型的媒体信息组合在一起形成数字信息资源的内容,这必定导致数字信息资源的信息形态呈现出多样化和复杂化的特点。有别于传统文献,一个数字信息资源往往不再仅由一个文件构成,通常是由一组多种类型的文件组成。如果对这些文件处理不当则会直接产生对数字信息资源的完整性的负面影响。另外,不同的数字信息资源的呈现与查询接口也不同,这意味着需要有不同的对数据结构化的处理方式,如果不能

重现原有的格式，那信息也将因此变得不完整。

1.4.5 数字信息资源的真实性

数字信息资源的真实性是指在数字信息资源的保存寿命内，可准确完整地媒体内将信息读取出来，并且可被理解，同时对其内容要保证信息可信性和完整性。数字信息资源的真实性包含以下三重内涵：①文件与其用意相一致；②文件的形成及发送与其既定的形成者和发送者相符合；③文件的形成或发送与其既定时间相吻合。

参考文献：

- [1] 沈涌. 数字信息资源整合策略与服务共享模式研究 [D]. 长春：吉林大学，2009.
- [2] 丁丽鸽，程大帅. 我国数字信息资源保存研究综述 [J]. 四川图书馆学报，2007（5）：51—54.

2 数字信息资源长期保存影响因素

2.1 数字信息资源长期保存技术政策因素分析

2.1.1 联合国教科文组织的《数字遗产保存指南》

联合国教科文组织（UNESCO）发布文件指出：文化遗产的消失将造成全世界文化遗产的贫穷^①。而伴随着数字技术的发展，现在以数字形式进行保存的信息资源不断增加，使得全世界文化遗产的保存得到了新的发展。2003 年联合国教科文组织发布了《数字遗产保存指南》，其主要内容如图 2-1 所示。

^① 聂云霞. 数字遗产长期保存的 SWOT 分析 [J]. 湖北档案, 2013 (4): 14-17.

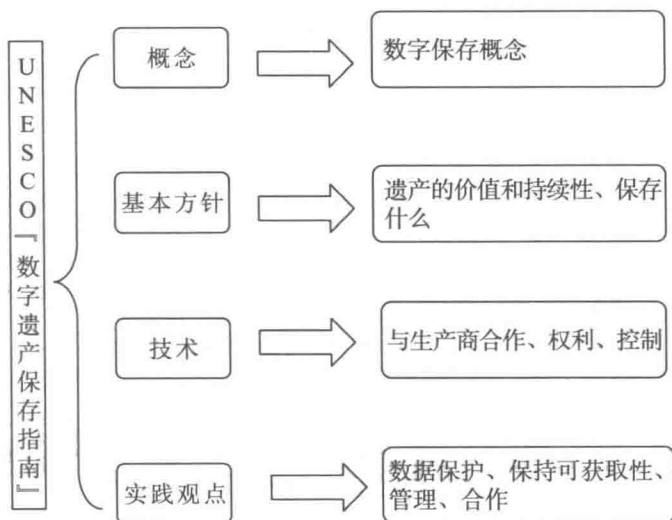


图 2-1 UNESCO《数字遗产保存指南》

联合国教科文组织出版的这份指南，不仅对世界文化遗产的数字化长期保存和利用提出了方案和方法，也为长期保存的实践者提供了行动的原则和指南。

2.1.2 OCLC 数字信息资源保存政策和支持文件

美国联机计算机图书馆中心（Online Computer Library Center, OCLC）从 2001 年就开始持续关注解决数字信息资源长期保存问题^①，于 2002 年开展了“数字存档计划”，根据两个重要指导性文件“可信赖数字仓库的属性和责任”及“OAIS 参考模型”于 2005 年推出了“OCLC 数字存档长期保存政策”，其主要内容如图 2-2 所示。

^① 牛爱菊，杨雪萍，卿蔚. 馆际互借系统平台管理与服务功能对比研究——以 OCLC 和 CASHL 为例 [J]. 大学图书馆学报, 2016, 34 (4): 63-68+12.