

图书馆数字资源 保护与服务研究

董晓莉 著



中国书籍出版社
China Book Press

图书馆数字资源 保护与服务研究

董晓莉 著



中国书籍出版社
China Book Press

图书在版编目 (CIP) 数据

图书馆数字资源保护与服务研究 / 董晓莉著. ——北京 : 中国书籍出版社, 2016.4

ISBN 978-7-5068-5414-6

I. ①图… II. ①董… III. ①数字图书馆—信息管理—研究②数字图书馆—图书馆服务—研究 IV. ①G250.76

中国版本图书馆CIP数据核字(2016)第034103号

图书馆数字资源保护与服务研究

董晓莉 著

-
- | | |
|------|---|
| 策 划 | 安玉霞 |
| 责任编辑 | 安玉霞 |
| 责任印制 | 孙马飞 马 芝 |
| 封面设计 | 展 华 |
| 出版发行 | 中国书籍出版社 |
| 地 址 | 北京市丰台区三路居路 97 号 (邮编: 100073) |
| 电 话 | (010) 52257143 (总编室) (010) 52257140 (发行部) |
| 电子邮箱 | chinabp@vip.sina.com |
| 经 销 | 全国新华书店 |
| 印 刷 | 北京京海印刷厂 |
| 开 本 | 789 毫米 × 1092 毫米 1/32 |
| 字 数 | 200 千字 |
| 印 张 | 8.125 |
| 版 次 | 2016 年 4 月第 1 版 2016 年 4 月第 1 次印刷 |
| 书 号 | ISBN 978-7-5068-5414-6 |
| 定 价 | 32.00 元 |
-

序 言

20 世纪 80 年代以来，由于计算机的广泛普及，信息技术快速而全面地渗透到社会生活的各个方面。数字信息被广泛应用，其数量呈现出几何级数的增长态势并以传统信息资源难以比拟的优势逐渐成为信息资源的主体。数字资源作为信息资源的重要组成部分，在有些国家已经被置于国家战略资源的高度，被誉为国家的“数字资产”（Digital Assets）。一个国家的科技创新能力以及与此相关的国际竞争力都依赖于其快速、有效地保存、开发与利用数字信息资源的能力。20 世纪 90 年代中期以来，欧美发达国家和地区对数字信息资源的保护给予了高度重视。美国把数字信息资源的生产、传播、获取、保存和利用作为国家信息化建设的关键和重点。

大量事实证明，实现数字资源的保护与服务有利于人类文明传承，有利于国家有效控制具有战略意义的信息资源，有利于充分发挥信息资源对经济发展、社会发展、文化发展的战略价值。然而，受多种复杂因素的影响，大量有价值的数字资源正在迅速消失或者失去使用价值。如何对数字数据进行管理和保存并使其实现服务增值是关涉国家经济、社会发展大局的战略性问题，对这一问题的有效应对刻不容缓。数字信息资源的保护与服务并不是一个简单的问题，它涉及诸多方面。本书重点介绍了数字资源保护与服务所涉及的社会背景、基础理论、研究进展和关键技术，

旨在为读者提供将数字资源保护应用于实际问题所必须的知识。

本书涵盖八个主要主题：社会信息变革；数字资源保护的概念；研究进展分析；图书馆数字资源保护策略；图书馆数字资源保护平台构建；数字资源保护的安全保障；数字资源保护的技术保障；服务拓展。目的是通过深入浅出的分析和论述，使读者在透彻地理解数字资源保护基础的同时，还能了解更多重要的高级主题。

本书首先从数字资源保护的时代背景入手，分析了社会信息环境的变革及其对图书馆的影响，即新的社会信息环境下图书馆所面临的挑战、图书馆文献资源体系的变化、用户需求的变化等关键问题，并探讨了新信息环境下图书馆的应对措施。进而分析了数字资源保护的内涵，并对保存与保护等重点概念进行了深入浅出的辨析，使读者更清晰地理解数字资源保护的概念，在此基础上对数字资源保护所涉及的重要模型进行了分析和探讨。结合国内外研究进展，从组织机构、实践项目、研讨会及国际刊物、职业教育等方面对数字资源保护的研究现状进行对比分析，使读者从各个方面更好地了解国际上数字资源保护的研究现状。目前，国内对 DC 的研究探索处于刚刚起步阶段，尚以理论研究为主，虽有部分科研机构针对其产生的科学数据开展了一些实践性的探索，但图书馆界在该领域的涉足仍然比较少。本书在介绍图书馆在数字资源保护中的重要作用及所承担的角色等内容的基础上，重点介绍了图书馆数字资源保护体系建设和平台建设所涉及的内容，从保护策略，管理体系、平台内容和构成、平台要素、平台体系架构、模型及其实现等方面，较为全面地分析了图书馆数字资源保护的实现。数字资源的安全是数字资源保护重点关注的核心问题之一，其影响因素颇多，本书在分析数字资源保护的影响

因素和风险的基础上,从数字资源的不变性控制、可信性控制、数字资源的备份和容灾备份等方面进行了多角度、多层次的深入分析,并结合每项技术探讨了实现方案和实施效果。数字资源长期保护过程涉及的相关技术分布在计算机、网络、通信、标准化等多个领域,本书结合技术前沿对数字资源保护所涉及的关键技术进行了介绍和分析,重点分析了存储技术、云技术、大数据和用户行为等技术。最后,以 MOOCs 和 ACGN 为例,对图书馆数字资源保护在服务方式和资源类型两方面的拓展进行了分析。

本书具有一定的创造性、实践性和广泛性,可以为广大从事数字资源保护的图书馆工作者提供理论和经验参考;同时可使社会公众全面了解数字资源保护领域的技术特点。

由于笔者水平有限,书中难免存在不足之处,衷心希望读者批评指正。

董晓莉

2016 年 1 月 20 日

目 录

第一章 时代背景	1
1.1 社会信息环境的变革	1
1.1.1 信息资源逐渐成为社会发展的基础性资源	2
1.1.2 信息资源爆炸性增长与快速消亡的矛盾	3
1.1.3 信息技术的飞速发展	6
1.2 社会信息环境变化对图书馆的影响	10
1.2.1 图书馆面临新的挑战	10
1.2.2 图书馆文献资源体系的变化	12
1.2.3 图书馆用户需求的变化	14
1.3 新信息环境下图书馆的积极应对	16
第二章 概述	19
2.1 内涵	19
2.2 概念辨析	24
2.2.1 数字资源保存与长期保存	24
2.2.2 Digital Curation 和 Digital Preservation	26
2.2.3 数字对象和表征信息	30
2.2.4 归档信息的表现信息与利用信息的表现信息	31
2.3 相关模型	31
2.3.1 DCC 数据保护生命周期模型	32
2.3.2 Athena 研究中心数据保护模型	35

2.3.3	开放档案信息系统 (OAIS) 参考模型	37
2.3.4	USAEP 项目的数字资源保护战略模型	40
2.3.5	数据保护项目 (Data Conservancy) 的 概念框架.....	41
2.3.6	美国加州大学嵌入式网络传感中心 (CENS) 的数据生命周期模型.....	41
第三章	研究进展分析.....	42
3.1	美国.....	42
3.1.1	组织机构.....	42
3.1.2	实践项目	44
3.2	英国.....	47
3.2.1	组织机构.....	47
3.2.2	实践项目	48
3.3	其他国家.....	51
3.3.1	组织机构.....	51
3.3.2	实践项目	52
3.4	研讨会与国际刊物	53
3.5	职业教育	55
3.6	国内图书馆界研究现状	60
3.6.1	理论研究.....	60
3.6.2	实践探索.....	61
第四章	图书馆数字资源保护	62
4.1	现状分析.....	62
4.1.1	国外数字资源保护策略已比较成熟	62
4.1.2	国外数字资源保护法规比较完善	63
4.1.3	图书馆在数字资源保护中发挥着越来越大的	

作用	64
4.2 图书馆的重要作用及其所承担的角色	66
4.2.1 图书馆的重要作用	66
4.2.2 图书馆所承担的角色	68
4.3 图书馆数字资源保护策略	73
4.3.1 内容选择策略	73
4.3.2 平台建设策略	75
4.3.3 组织结构保障	76
4.3.4 人员保障	77
4.3.5 运作模式	78
4.4 图书馆数字资源保护管理体系	79
4.4.1 内容构成	79
4.4.2 管理机制	80
4.4.3 协作机制	80
4.5 我国图书馆数字资源保护的挑战	81
第五章 图书馆数字资源保护平台	83
5.1 图书馆数字资源内容及构成	83
5.1.1 数字资源界定	83
5.1.2 数字资源内容分类	85
5.1.3 数字资源类型	87
5.1.4 数字资源构成	88
5.2 图书馆数字资源生命周期	90
5.2.1 图书馆数字资源生命周期阶段分析	90
5.2.2 图书馆数字资源保护活动与数字资源 生命周期	94
5.3 图书馆数字资源保护平台要素	96

5.3.1	平台资源	96
5.3.2	管理主体	98
5.3.3	标准体系	99
5.3.4	相关职责	99
5.3.5	计算机软件及硬件	100
5.4	图书馆数字资源保护平台体系	102
5.4.1	平台模型分析	102
5.4.2	功能模块分析	104
5.4.3	平台特征分析	109
5.5	图书馆数字资源保护平台建设	110
5.5.1	数字资源的处理流程管理	111
5.5.2	数字资源的质量保证	111
5.5.3	数字资源存储平台的可靠性	112
5.5.4	数字资源保护系统的实现	112
5.6	典型数字资源保护平台分析	114
5.6.1	约翰·霍普金斯大学 Data Conservancy	114
5.6.2	哈佛大学 Dataverse Network	115
第六章	安全保障	119
6.1	数字资源保护影响因素分析	119
6.1.1	保护价值分析	119
6.1.2	保护成本分析	120
6.2	数字资源保护的风险分析	123
6.3	数字资源的不变性控制	125
6.3.1	SIRF 存储模型分析	126
6.3.2	SIRF 应用	132
6.4	数字资源的可信性控制	141

6.4.1	可信性分析	141
6.4.2	可信性管理流程	142
6.4.3	可信性管理实现	144
6.5	数字资源的备份	148
6.5.1	备份系统建设需求	149
6.5.2	备份方案的选择	150
6.5.3	从系统设计角度考虑数据分层	154
6.6	容灾备份	155
6.6.1	容灾定义	156
6.6.2	容灾的技术分类	156
6.6.3	需要考虑的相关因素	157
6.6.4	需求分析	158
6.6.5	容灾规划	159
6.6.6	容灾实现	164
第七章	技术保障	171
7.1	存储技术	171
7.1.1	存储介质	171
7.1.2	存储系统	174
7.2	用户行为分析	175
7.2.1	用户与网络用户信息行为	176
7.2.2	网络用户信息行为的类型	177
7.2.3	图书馆的用户行为数据	180
7.2.4	用户信息行为分析方法	181
7.3	大数据	183
7.4	云存储	185
7.4.1	概念	186

7.4.2 基于云存储的数字资源保护平台特点	189
第八章 服务拓展	191
8.1 ACGN	191
8.1.1 图书馆 ACGN 服务的提出	192
8.1.2 我国图书馆 ACGN 资源服务现状	195
8.1.3 图书馆与 ACGN 资源建设	198
8.1.4 图书馆与 ACGN 推广	200
8.2 MOOCs	205
8.2.1 认识 MOOCs	206
8.2.2 MOOCs 环境下的图书馆合作与服务创新	210
8.2.3 MOOCs 与图书馆信息技术平台整合	213
8.2.4 图书馆与 MOOCs 平台的资源服务整合	218
参考文献	224

第一章 时代背景

任何时代，图书馆总是存在于一定的社会信息环境之下。社会信息环境的变化对图书馆来说，既是一种挑战，给图书馆带来不断创新和变革的动力；也是一种机遇，可以使图书馆在新的社会信息环境下主动变化以适应社会对图书馆的需求，促使图书馆从旧的模式中走出来，去迎接更大的发展空间。在新的社会信息环境下，信息被无限泛在化，用户信息行为在变化，社会信息需求在变化，图书馆的社会信息保护和服务的功能已经被弱化。图书馆处于一个飞速变化的社会信息环境中，必须重新定位其在数字资源保护与服务中的作用以及其与服务对象之间的关系，顺应互联网思维的发展趋势，优化图书馆数字资源保护和服务结构，合理定位，塑造新的信息资源保护与服务环境。

1.1 社会信息环境的变革

20 世纪后期，在信息科学技术的推动下，信息产业有了飞速的发展。信息技术的广泛应用，将人类由工业社会带入了信息社会。全球信息化正在引发当今世界的深刻变革，宽带无线、智能网络、超级计算、虚拟现实、网络制造与网络增值服务极大地丰富了信息的形式与内容，拓展了信息的传播范围，提升了信息的应用价值。信息化可以使人们突破语言文字的屏障，更便捷地

进行跨文化、跨地域的信息交流。

1.1.1 信息资源逐渐成为社会发展的基础性资源

国际上对信息资源的定义有广义和狭义两种解释。广义的信息资源包括信息和其他相关资源,如人员、设备、资金和技术等。狭义的信息资源是指信息内容的总体,是可供人类作为生存发展基础而利用的信息集合,包括来自各种载体、各种表现形式的信息内容。本书所指的信息资源是指代狭义的信息资源,即可供人类作为生存发展基础而利用的信息集合。

人类进入信息社会,信息资源逐渐成为重要的战略资源,是现代社会发展的基础性资源。中共中央办公厅、国务院办公厅颁布的《关于加强信息资源开发利用工作的若干意见(中办发〔2004〕34号)》中明确提出,信息资源作为生产要素、无形资产和社会财富,与能源、物质资源同等重要,在经济社会资源结构中具有不可替代的地位。从资源属性看,信息资源与能源、物质资源之间存在相互依存、相互转化的内在联系。信息资源的开发利用可以节约能源和物质资源,避免能源和物质资源的浪费,实现能源和物质资源的增值。从资源的作用看,信息资源已经在促进经济增长与社会发展、提高国家综合国力和竞争力、实现绿色增长及可持续发展等方面发挥了重要作用。

在所有信息资源中,数字资源由于其无所不在、分布广泛、随时随地可取等特点,在科学研究、终生教育、提升生活品质等方面都发挥着不可替代的作用。数字资源具有通用性、开放性和标准化的数据结构,不受复本量的限制,在网络环境下,终端读者可以共同使用这些数字资源,实现信息资源共享;它不受时间、空间的限制,大大缩短了信息传递时间,加快了信息交流和反馈

的速度;数字资源存储信息密度高、容量大,占用空间少,检索速度快,检索功能齐全,能提供各种检索途径,从而提高了文献的利用率;另外它还能存储文章、声音、图像等多媒体信息,并具有打印输出及磁盘输出等多种输出方式等特点和优势。现今,数字资源已经在数量上占据了人类信息资源总量的核心,并逐步成为人类社会信息资源的主体。^①今天的数字资源比以往任何时候都更加深入地与我们的生活交织在一起,人类期待着用数字资源解决各种问题,改善生活,推动经济繁荣。

1.1.2 信息资源爆炸性增长与快速消亡的矛盾

人类需要通过信息进行日常交流、知识传递、协作活动、娱乐消遣等。在过去,由于受社会生产力的限制,信息的产生、传播都十分有限。然而随着社会的发展和科学技术的进步,信息的产生、传播变得非常容易,信息的总量也不断扩大,并呈现出加速的趋势。

国家统计局《2012年中国统计年鉴》显示,2011年我国出版的图书种类达369523种,总印数为77.1亿册;期刊种类达9849种,总印数为32.9亿册;报纸种类达1928种,总印数为467.4亿份。全年制作公共电视节目达3274套,总播出时间为16753029小时;全年播出电视剧为247060部,共6636255集;全年动画电视播出数为280255小时;全年电影生产数量为689部;全年录音制品出版品种数量为9931种^②。这些数据显示,我国文化事业得到蓬勃发展,同时伴随而来的是大量的信息得以产生,

① 向锋.提高新升本科院校数字资源利用率之对策[J].科技情报开发与经济,2005(3)63-64.

② 中华人民共和国国家统计局.2012年中国统计年鉴2012. <http://www.stats.gov.cn/tjsj/ndsj/2012/indexch.htm>.

然而这些信息只是冰山一角。

随着信息技术的不断发展,特别是互联网的广泛应用,大量的信息得以涌现。互联网应用,如微博、博客、论坛、门户网站等,使得用户不仅可以浏览网页获取信息,还可以自由地发布信息,而不受时间和空间的限制。这些技术使得信息的产生和传播极其便利,促进了信息的爆炸。瑞典 Royal Pingdom 发布的 2013 年全球互联网络产业发展状况报告显示,2012 年,全世界有 6.34 亿个网站,24 亿网络用户,使用谷歌搜索引擎查找信息的次数为 1.2 万亿^①。中国互联网络信息中心 2013 年发布的“第 31 次中国互联网络发展状况统计报告”显示,截至 2012 年 12 月底,我国网站总数为 268 万个,网民规模达 5.64 亿,手机网民规模为 4.20 亿,搜索引擎网民规模为 4.511 亿。^② 这些数据显示了互联网网站及用户的数量极其庞大。此外,随着社会的进步,科学技术不断得到发展,科学工作者人数不断增加。据 2012 年中国科技论文最新统计结果显示,中国近 10 年来在国际上发表的论文就达到 102 万篇。^③ 据汤森路透集团 2009 年发布的《全球科研报告》系列估计,每年世界新增发表论文达百万量级。^④ 在我国,科技工作者已经接近 5800 万。^⑤ 面对学术信息的需求,大量科学文献信息已经广泛分布于互联网络中,这也促成了一些大的搜索引擎

① Royal Pingdom. Internet 2012 in Numbers. 2013. <http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/>.

② 第 31 次中国互联网络发展状况统计报告. CNNIC. 2013.

③ 中国科学技术信息研究所. 2012 中国科技论文统计结果. 2012. <http://www.istic.ac.cn/ScienceEvaluateArticalShow.aspx?ArticleID=94466>.

④ Thomson Reuters. Global Research Reports. 2009. <http://sciencewatch.com/global-research-reports>.

⑤ 光明网. 科技工作: 做社会的良心. 2011. Retrieved in 2013-03-29. http://theory.gmw.cn/2011-10/10/content_2758694.htm.

公司对学术信息提供特定的搜索，例如谷歌学术搜索、微软学术搜索。总之，互联网数字化信息资源已经非常丰富，用户几乎可以在其中找到自己感兴趣的所有内容。

此外，随着数字和网络技术的飞速发展和迅猛普及，公众对数字阅读从适应到依赖，推动了数字出版产业的快速发展。目前国际上较为成功的数字出版企业有哈泼·柯林斯出版集团、约翰·威立-布莱克维尔出版集团、培生教育出版集团、麦格劳·希尔教育出版集团和圣智教育出版集团、里德·艾尔思维尔出版集团、汤姆森集团等。国内的北大方正、万方数据、维普数据库、书生之家、超星数字图书馆、中文在线等内容提供商也迅速发展^①。同时，一些行业外巨头也试图进入这一领域，如以游戏起家的盛大、淘宝网，以卖家电为主的京东以及亚马逊中国、苹果中国等。截至2011年1月，中国移动的手机阅读基地的月收入已经超过8千万元^②。2008年法兰克福图书博览会报道，参展产品的1/3以上是数字化产品，期间400场活动1/2以上与图书数字化潮流有关。由此可见，越来越多的出版商和行业巨头，认识到数字出版的广阔前景，并借助网络、手机、数字阅读器、电子书、电纸书等手段有效推动数字出版的发展，使得原生全媒体资源呈现几何级爆炸性增长的态势。

信息总量的增加为满足用户信息需求提供了可能，然而当前绝大多数信息以数字形式存在，数字信息由于其脆弱、容易消失、容易变化等特点，特别是由于人们认知局限以及管理策略不当，

① 平原. 中文电子书厂商四大家族. http://blog.sina.com.cn/s/blog_48dcd6f9010005g1.html, 2006-8-8.

② 蒋庆. 传统出版进入“半衰期”. Finance.newssc.org/system/2011/01/13/013037642.shtml, 2011-01-13.