



国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

“十二五”国家重点图书出版规划项目
数字出版理论、技术和实践

数字出版与数字图书馆

黄肖俊 吕肖庆 等编著



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

“十二五”国家重点图书出版规划项目
数字出版理论、技术和实践

数字出版与数字图书馆

黄肖俊 吕肖庆 汤 帆 方中华 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书系统论述了数字图书馆的发展历程、性质和特点，重点描述了数字资源建设与管理，以及新兴数字化服务的典型代表和发展趋势；深入分析了数字出版、数字图书馆，以及正在崛起的新兴信息服务商三方关系，探讨了能够融合各方利益的解决方案；总结了出版商提供数字资源和服务的各种商务模式，以及在实际工作中的操作建议。本书有选择性地介绍了数字出版和数字图书馆可以相互借鉴的先进技术，侧重讲述了数字图书馆的资源服务系统的设计，以及一些极具参考价值的技术标准。

本书的编写目的是总结和 research 数字出版和数字图书馆在新的数字化环境中如何相互支撑、共同发展。本书适合编辑出版学和传播学类学生、出版工作者，以及信息服务工作者等人员阅读。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目(CIP)数据

数字出版与数字图书馆 / 黄俊等编著. —北京：电子工业出版社，2013.9
(数字出版理论、技术和实践)

ISBN 978-7-121-20522-4

I. ①数… II. ①黄… III. ①电子出版物—出版工作—研究②数字图书馆—研究
IV. ①G237.6②G250.76

中国版本图书馆 CIP 数据核字 (2013) 第 111188 号

策划编辑：李 弘

责任编辑：秦绪军

印 刷：北京天来印务有限公司

装 订：北京天来印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：720×1 000 1/16 印张：24.5 字数：480 千字

印 次：2013 年 9 月第 1 次印刷

印 数：2 000 册 定价：78.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888。

质量投诉请发邮件至 zltts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：(010) 88258888。

指导委员会

主 任 孙寿山

委 员（按姓氏笔画排序）

王关义 王志成 方中华 田胜利 朱伟峰 李宏葵

余昌祥 张志强 张增顺 张毅君 郝振省 教 然

聂震宁 谢俊旗 谢新洲 薛松岩

编辑委员会

主 任 魏玉山

副主任 刘九如

委 员（按姓氏笔画排序）

王 强 王晓光 王德胜 方 卿 邢立强 吕肖庆
刘成勇 刘拥军 刘锦宏 孙 坦 孙广芝 李 弘
沈艳波 张 立 张 峻 张宝元 陈 丹 陈源蒸
郝黎明 秦绪军 徐丽芳 高 昂 黄肖俊 程三国

序

Introduction

数字出版方兴未艾。作为新闻出版业的重要发展方向和战略性新兴产业，数字出版近年来发展迅速，已经成为当前我国新闻出版业转型发展的助推器和新的经济增长点。基于互联网、移动通信网、有线电视网、卫星直投等传播渠道，并以 PC 机、平板电脑、智能手机、电视、iPad 等阅读终端为接收载体的全新数字出版读物，已成为人民群众精神文化生活不可或缺的组成部分。

从毕升的活字印刷到王选的激光照排系统问世，技术元素始终是出版业发展壮大的重要源动力。进入 21 世纪，信息通信技术（ICT）的飞速发展成为新经济发展的主要引擎，使得以思想传播、知识普及、文化传承、科学交流和信息发布为主要功能的出版业可以持续、广泛地提升其影响力，同时大大地缩短了信息交流的时滞，拓展了人类交流的空间。计算机芯片技术、XML 及相关标记语言技术、元数据技术、语义技术、语音识别和合成技术、移动互联技术、网络通信技术、云计算技术、数字排版及印刷技术、多媒体技术、数字权利管理技术等一大批数字技术的广泛应用，不但提升了传统出版产业的技术应用水平，同时极大地扩展了新闻出版的产业边界。

如同传统出版业促进了信息、文化交流和科技发展一样，数字出版的多业态发展也为 20 世纪末期开始的信息爆炸转变为满足个性化需求的知识文化服务提供了技术上的可能。1971 年，联合国教科文组织（UNESCO）和国际科学联盟理事会（ICSU）便提出了 UNISIST 科学交流模型，将出版业所代表的正式交流渠道置于现代科学交流体系的中心位置。进入 21 世纪，理论界又预见到，网络出版等数字出版新业态的出现正在模糊正式交流和非正式交流的界限，更可能导致非正式交流渠道地位的提升。随着以读者（网络用户）为中心的信息交流模式，比如博客、微博、微信和即时通信工具等新型数字出版形态的不断涌现，理论构想正在逐渐变为现实。

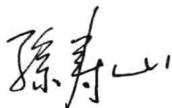
通过不断应用新技术，数字出版具备了与传统出版不同的产品形式和组织特征。由于数字出版载体的不断丰富、信息的组织形式多样化以

及由于网络带来的不受时空限制的传播空间的迅速扩展,使得数字出版正在成为出版业的方向和未来。包括手机彩铃、手机游戏、网络游戏、网络期刊、电子书、数字报纸、在线音乐、网络动漫、互联网广告等在内的数字出版新业态不断涌现,产业规模不断扩大。据统计,在2006年,我国广义的数字出版产业整体收入仅为260亿元,而到了2012年我国数字出版产业总收入已高达1935.49亿元,其中,位居前三位的互联网广告、网络游戏、手机出版,总产出达1800亿元。而与传统出版紧密相关的其他数字出版业务收入也达到130亿元,增长速度惊人,发展势头强劲。

党的十七届六中全会为建设新时期的社会主义先进文化做出战略部署,明确要求发展健康向上的网络文化、构建现代传播体系并积极推进文化科技创新,将推动数字出版确定为国家战略,为数字出版产业的大发展开创了广阔的前景。作为我国图书出版产业的领军者之一,电子工业出版社依托近年来实施的一批数字出版项目及多年从事ICT领域出版所积累的专家和学术资源,策划出版了这套“数字出版理论、技术和实践”系列图书。该系列图书集中关注和研究了数字出版的基础理论、技术条件、实践应用和政策环境,认真总结了我国近年发展数字出版产业的成功经验,对数字出版产业的未来发展进行了前瞻性研究,为我国加快数字出版产业发展提供了理论支持和技术支撑。该系列图书的编辑出版适逢其时,顺应了产业的发展,满足了行业的需求。

毋庸讳言,“数字出版理论、技术和实践”系列图书的编写,在材料选取,国内外研究成果综合分析等方面肯定会存在不足,出版者在图书出版过程中的组织工作亦可更加完美。但瑕不掩瑜,“数字出版理论、技术和实践”系列图书的出版为进一步推动我国数字出版理论研究,为各界进一步关注和探索数字出版产业的发展,提供了经验借鉴。

期望新闻出版全行业以“数字出版理论、技术和实践”系列图书的出版为契机,更多地关注数字出版理论研究,加强数字出版技术推广,投身数字出版应用实践。通过全社会的努力,共同推动我国数字出版产业迈上新台阶。



2013年8月

前言

Preface

出版的繁荣使得源源不断的文稿得以流传，图书馆的兴起让普通大众享受到了公益性的信息服务，两者都具有悠久的历史，时至今日，迅猛发展的信息技术和通信技术使两者都发生了深刻的变革，数字出版、数字图书馆在加速发展的同时，也开始发生一些碰撞和冲突。

本书的前两章简要回顾了数字图书馆的发展历程，并对其资源管理和新兴的数字化服务进行了系统的分析。

从第 3 章开始，转入本书关注的重点问题：数字出版和数字图书馆的关系。我们试图通过对出版业、图书馆在理念、业务、技术和版权等多方面的细致梳理来总结出导致两者关系变化的原因，同时也将目光投向了与两者有密切关系的新型信息服务商。经过本章的分析我们不难发现，现存问题的解决之道离不开对多方利益的权衡。

今天，数字出版和数字图书馆在各自领域都有了长足的发展，积累了大量的先进技术，其中有些从一开始就是共用技术，有些是在一个领域中率先发展，第 5 章我们选取了一些属于后者的技术热点，希望通过我们的介绍让不同领域的从业人员了解对方的技术发展状况，相互借鉴。并在分析数字图书馆的服务模式和系统设计原则的基础上，阐述了其资源服务系统的设计方法，并给出了多个重点子系统的设计指导意见。我们的介绍比较偏重应用，原因在于本书的定位主要是为出版从业人员了解图书馆界如何使用数字资源，而不是用来指导实现其复杂的内部结构。

标准是技术的结晶，我们将与数字出版相关的数字图书馆领域的重要标准汇编在第 6 章。由于图书馆对于信息服务开展得最早，在技术方面也应用得最多，因此大量与信息服务有关的标准都会率先在该领域出现。换言之，数字图书馆领域的很多标准，如资源描述标准、互操作标准、资源检索标准等，都是某一方面技术先进性的集中体现。在知识经济快速发展的今天，这些标准无疑为出版等其他行业提供了宝贵的经验和参考。

出版单位在为图书馆提供数字资源或技术支持时难免会遇到各种各样的问题，本书的第 7 章依据笔者长期的研发经验，为出版机构能够避免或克服实际业务中的具体问题提供了大量中肯的建议。

最后，本书的第 8 章对数字出版与数字图书馆进行了反思与展望，预计数字出版将加大对数字图书馆的支持力度，分析了推动数字图书馆发展的一些深层原因，并提出了数字图书馆的发展趋势。

本书是集体智慧的结晶，成书之际感谢多方帮助。感谢北京方正阿帕比技术有限公司赫思佳总经理、中科院国家科学图书馆孙坦副馆长、中国高等教育文献保障系统（CALIS）王文清总工程师、万方数据股份有限公司技术研究院郭晓峰老师、北京大学的高良才副教授等在百忙中抽出时间审读书稿并提出宝贵意见。同时，感谢本套丛书的其他分册的编写人员、方正国际 BPO 事业部周长岭总工程师、方正阿帕比魏丕、北京大学贾爱霞老师、杨森博士以及焦喜音、李佳林、王弢、叶文等同学的大力支持。本书引用了大量国内外学者的研究成果，没有他们的细致分析，就不可能有本书的问世，在此，对他们的辛勤劳动和贡献表示崇高的敬意和感谢！

本书的出版得到了电子工业出版社的大力支持，数字出版中心主任李弘、副主任王德胜、副主任张峻，以及编辑徐蔷薇老师、张莹老师为本书内容编排等许多方面提出了宝贵的意见，在此表示衷心的感谢！

由于编写者水平有限，特别是涉及的领域较多，书中难免存在缺点和错误，敬请读者以及关注数字图书馆和数字出版的人士批评斧正。

编 者

2012 年 10 月

目 录

Contents

第 1 章 | Chapter 1

数字图书馆概论	1
1.1 图书馆的发展简史	2
1.2 数字图书馆的概念	2
1.3 数字图书馆的发展	4
1.3.1 早期的数字图书馆	4
1.3.2 20 世纪前后的数字图书馆	6
1.4 数字图书馆管理系统	12
1.4.1 仓储管理系统	12
1.4.2 基础设施、虚拟研究环境和生态系统	15
1.4.3 当代图书馆集成管理系统	16
1.5 数字图书馆属性演变过程	24
1.5.1 传统业务自动化升级改造阶段	24
1.5.2 数字资源的大规模建设阶段	25
1.5.3 以用户为中心的知识服务阶段	25
1.6 数字图书馆的特点	26
1.6.1 从封闭空间到开放的云服务	26
1.6.2 从纸本书到多媒体的有序数字化存储	26
1.6.3 从借阅柜台到远程互动服务	27
1.6.4 从书架到语义网	27
1.7 数字图书馆的作用	27
1.8 数字图书馆面临的环境变化	28
1.8.1 政策环境	28
1.8.2 周边产业环境	30
1.8.3 技术环境	34

第 2 章 Chapter 2	
数字图书馆的资源建设与新型服务	37
2.1 数字资源的建设与管理	38
2.1.1 找准定位, 把握平衡	38
2.1.2 持续更新, 保证质量	40
2.1.3 互联互通, 内外有别	40
2.2 数字化服务的新趋势	43
2.2.1 借助互动技术, 不断提升服务的主动性	43
2.2.2 多管齐下, 全方位推动个性化服务	44
2.2.3 融合社会化网络, 聚集更多服务外援	45
2.2.4 紧跟移动互联, 将服务做到无微不至	46
2.3 具有代表性的数字化服务	46
2.3.1 虚拟咨询服务	47
2.3.2 定题服务	51
2.3.3 数字图书馆对 E-Learning 的支撑和学习共享空间的营造	55
2.3.4 虚拟研究环境中学科服务的升级与创新	59
第 3 章 Chapter 3	
图书馆、出版社与新兴信息服务商的关系	65
3.1 传统图书馆与出版社的关系	66
3.1.1 图书馆与出版社	66
3.1.2 传统图书馆与出版社的平衡关系	66
3.2 数字化变革引发的关系变化	67
3.2.1 数字化对图书馆和出版业的影响	67
3.2.2 调整中的数字出版与数字图书馆的关系	67
3.3 新兴信息服务商对数字图书馆和数字出版商的影响	69
3.3.1 新兴信息服务商	70
3.3.2 新兴信息服务商对数字图书馆的冲击	70
3.3.3 信息服务商对数字出版的冲击	71
3.4 解决之道	72
3.4.1 协调三方关系	72
3.4.2 技术保障	74
3.4.3 法律保障	75

3.5	数字出版支撑图书馆的转型与和谐关系建立的原则	77
3.5.1	数字图书馆与数字出版的辩证关系	77
3.5.2	数字出版企业为数字图书馆服务的基本原则	79
第4章	Chapter 4	
	面向数字图书馆的数字出版商务模式	89
4.1	单本销售模式	90
4.1.1	电子图书	90
4.1.2	其他形式	92
4.1.3	国内外成功案例	92
4.2	连续出版物销售模式	96
4.2.1	数字期刊	97
4.2.2	数字报纸	98
4.2.3	国内外成功案例	100
4.3	数据库销售模式	114
4.3.1	条目式数据库	114
4.3.2	学位论文数据库与会议论文数据库	115
4.3.3	电子书库	115
4.3.4	专题数据库	116
4.3.5	国内外成功案例	116
4.4	开放存取	123
4.4.1	开放存取的基本概念	124
4.4.2	开放存取的发展历程	124
4.4.3	开放存取的商务模式	129
4.4.4	开放存取对图书馆的影响	129
4.4.5	开放存取对数字出版模式的颠覆与重建	130
4.4.6	开放存取面临的问题	131
第5章	Chapter 5	
	数字出版与数字图书馆的技术互鉴与资源服务系统设计	133
5.1	数字出版领域中可供借鉴的重要技术	134
5.1.1	文档技术	134
5.1.2	数字版权保护技术	141
5.1.3	跨平台阅读技术	144
5.1.4	版面理解和智能标引技术	150

5.2 数字图书馆领域中可供借鉴的重要技术	151
5.2.1 语义网.....	151
5.2.2 本体.....	154
5.2.3 资源描述框架 (RDF)	158
5.2.4 网络本体语言 (OWL)	162
5.2.5 信息检索.....	166
5.2.6 自然语言处理.....	180
5.2.7 自动文摘.....	189
5.2.8 云计算.....	200
5.3 面向图书馆的数字资源服务系统的设计	211
5.3.1 数字图书馆服务模式.....	211
5.3.2 系统组成与总体结构设计的原则.....	216
5.3.3 重点子系统设计.....	218
5.3.4 接口设计.....	226
5.3.5 数字图书馆的发现服务.....	228

第 6 章 | Chapter 6

与数字出版相关的数字图书馆标准.....	231
6.1 数字图书馆标准建设的概况	232
6.2 国际标准的研究情况	233
6.3 我国标准的研究情况	237
6.3.1 “我国数字图书馆标准与规范建设”项目 (CDLS)	237
6.3.2 CADLIS 技术标准与规范建设简介.....	240
6.4 资源描述格式标准	242
6.4.1 元数据与元数据标准.....	242
6.4.2 METS 元数据编码术传输标准.....	245
6.5 互操作标准	255
6.5.1 OAI 标准之一: OAI-PMH.....	256
6.5.2 OAI 标准之二: OAI-ORE	263
6.5.3 OpenURL	270
6.5.4 开放式出版发布系统 OPDS.....	276
6.6 数字对象唯一标识符 (DOI) 协议	281
6.6.1 DOI 的历史.....	281
6.6.2 DOI 的内容.....	283

6.6.3	DOI 对数字图书馆和出版的影响	286
6.6.4	DOI 的局限性与发展前景	287
6.7	信息资源检索协议	288
6.7.1	Z39.50 协议	289
6.7.2	过渡性标准——ZING	291
6.7.3	查询与检索的 URL/Web 服务 (SRU/W)	292
6.7.4	情境查询语言 CQL	294
6.7.5	ZeeRex	296
第 7 章	Chapter 7	
	出版机构为数字图书馆服务的实操建议	299
7.1	出版机构数字出版机制的建立	300
7.1.1	出版社开展数字出版的瓶颈	300
7.1.2	出版社开展数字出版业务最紧迫的事项	301
7.1.3	出版社该如何建立和健全数字出版机制	302
7.1.4	出版社电子文档收集与管理样例	303
7.2	数字出版的准备工作	309
7.2.1	内容选择与组织	309
7.2.2	电子图书出版制作加工流程	309
7.2.3	排版制作与加工工具	311
7.2.4	元数据加工与 MARC 编目	312
7.2.5	资源加工原材料合格标准与成品合格标准	320
7.2.6	数字版权保护及出版发行平台的选择	327
7.3	其他资源为数字图书馆服务的准备工作	329
7.3.1	期刊数字出版为数字图书馆服务的准备工作	329
7.3.2	报纸数字出版为数字图书馆服务的准备工作	331
7.3.3	条目式资源库数字出版为数字图书馆服务的准备工作	332
7.3.4	图片数字出版为数字图书馆服务的准备工作	333
7.3.5	数字出版与图书馆采访系统的对接	335
第 8 章	Chapter 8	
	反思与展望	337
8.1	数字出版将加大对数字图书馆的支持力度	338
8.2	推动数字图书馆发展的一些深层原因	338

8.3 数字图书馆的发展趋势	340
8.3.1 资源构建社会化和服务交互化.....	341
8.3.2 知识组织的语义化和关联化.....	341
8.3.3 知识环境的复杂化和泛在化.....	342
8.3.4 服务虚拟化和行业融合化.....	343
8.3.5 公益性与商业化.....	343
8.4 展望	344
参考文献	345

第 1 章

Chapter 1

▶ 数字图书馆概论

自从有文字记载以来，人类社会的发展悠悠数千载，会聚了灿若星河的文化典藏，图书馆一直承担着保存和传承人类文明的重大职责。近代意义的图书馆在百年前告别了古代藏书楼，走向了公共服务。20 世纪 40 年代以来，计算机技术、网络技术和信息处理技术迅猛发展，深刻地改变了人们的学习、工作、生活，甚至思维方式。

近年，网络作为一种新的信息交流和通信工具日益普及，成为了人们获取信息的重要来源。越来越多的文字、图片、声音、影像资料以数字形式大量涌现，成为影响信息化社会发展的重要力量。越来越多的国家认识到信息对于提高国际竞争力、增强综合国力的重要性，相继提出了“信息高速公路”计划，建立信息网络，支持国家创新与经济社会发展，人类社会快速进入了一个前所未有的信息化社会。在此背景下，数字图书馆作为网络环境下一种新的信息资源组织与服务形式应运而生。

数字图书馆是网络环境和数字环境下图书馆新的发展形态，它利用现代信息技术，对海量、分布、异构的数字资源进行整合，形成有序的整体，通过各种媒体提供友好、高效的服务，使人们随时随地获取信息和知识。

在当今的知识社会中，数字图书馆发挥着举足轻重的作用，它不仅能将全球知识仓储中的海量信息提供给任何地点的用户，服务于他们的学习、工作和生活，同时，也起着重要的文化传播作用。数字图书馆的出现使知识管理的各个环节发生了翻天覆地的变化。

1.1 图书馆的发展简史

人类在社会交往中创造了文字，用以表达思想、记录信息，伴随着文献的积累，产生了收集、整理、贮存和使用文献的专门机构，这就是最初形态的图书馆。

公元前 3000 多年，美索不达米亚地区使用方头铁笔将楔形文字内容记录在湿泥版上，晒干或焙烧后制成泥版书，亚述巴尼拔在尼尼微创建的图书馆中收藏了数万块泥版书。公元前 5 世纪，希腊的作家和学者逐渐开始拥有私人图书馆。公元前 3 世纪，托勒密一世建立了古代最大的图书馆——亚历山大图书馆，馆藏多达 70 万卷。中国自汉朝开始，历代皇家藏书都在不断地更替、继承和发展，经过了唐、宋、明、清几个重要的历史时期，形成了官府藏书、私人藏书、寺院藏书和书院藏书四大体系。自 12 世纪起，欧洲各大城市陆续出现了大学，并建立了自己的图书馆。18 世纪末，免费、开放的公共图书馆逐步建立并发展起来。20 世纪初，国家图书馆有了长足的发展，成立于 1800 年的美国国家图书馆现已发展成了全球馆藏量最大的图书馆。

图书馆的发展不仅是规模的日益扩大，其功能也在不断完善和变化。早期图书馆一直以“藏”为主；从 17 世纪初开始，“藏用结合”的图书馆逐渐增多，特别是到了 20 世纪中后期，“用”的功能越来越受到重视，这也体现了图书馆从以“物”为中心到以“人”为中心的转变。

现代信息技术的发展使图书馆进入一个崭新的发展时期。图书馆的大量文献资源正在从纸质资源向数字资源转变，数字化的服务方兴未艾，图书馆比以往任何历史时期都更加注重知识服务和个性化服务，可以说，今天的图书馆正在从纸质图书馆时代迈入了一个崭新的数字图书馆时代。

1.2 数字图书馆的概念

从字面上看，数字图书馆可以简单地理解为传统图书馆的计算机化，但实际上，数字图书馆涉及的领域众多，包括图书馆学、情报学、信息学、计算机科学、通信科学等，是一个跨学科、多层次而且还在快速发展变化的新领域。因此，在不同的专家看来，其内涵和特点有着不同时期、不同层次和不同角度的多样性理解。

美国密歇根大学在 1990 年率先提出了“数字图书馆”这一概念，当时给出的定义是：数字图书馆是若干联合机构（Federated Structure）的总称，它使人们能够