



“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

数字出版元数据基础

孙广芝 邢立强 张保玉 编著



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



“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

数字出版元数据基础

孙广芝 邢立强 张保玉 编著

电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

数字出版产业的发展,需要规模化的数字出版内容资源。作为内容组织和管理的有效工具,元数据对于出版内容资源整合及出版产业的可持续发展具有重要意义,按照元数据标准的要求对内容资源进行规范化描述,可以有效地促进内容资源的处理、交换、共享和开发利用。

本书首先介绍了数字出版元数据的基本概念、作用及其在国内外的的发展状况,然后针对数字出版全流程所涉及的元数据标准的内容、功能和特点给出了详细的介绍和分析,包括用于出版物及其内容资源标识的 ISBN、ISSN 和 DOI 等标准,以及用于出版物及其内容资源描述的 DC 元数据、机读目录格式和图书 ONIX 等标准,并给出一些元数据标准的应用示例。本书还对不同元数据的转换技术进行了介绍。

本书对数字出版元数据进行了系统、全面的论述,可以帮助读者快速了解数字出版元数据的全貌,适合对元数据、数字出版等感兴趣的技术人员阅读。

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

数字出版元数据基础 / 孙弘, 邢强, 张强编著. —北京: 电子工业出版社, 2013.9
(数字出版理论、技术和实践)
ISBN 978-7-121-21472-1

I. ①数… II. ①孙… ②邢… ③张… III. ①电子出版物—出版工作—研究 IV. ①G237.6

中国版本图书馆 CIP 数据核字(2013)第 215141 号

策划编辑: 李 弘

责任编辑: 张 峻

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: 21.25 字数: 410 千字

印 次: 2013 年 9 月第 1 次印刷

印 数: 2 000 册 定价: 72.00 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 zltz@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 敖 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

数字出版作为出版业的一种新兴业态,自问世以来,发展非常迅猛。随着全球信息化进程的推进以及数字技术向各个领域的不断延伸,我国数字出版产业规模迅速扩大,收入逐年大幅度递增,技术不断创新,产业形态日益完备,产业链趋于整合,数字出版已经成为我国出版业新的增长点。

在转型过程中,出版业积累和产生了大量高价值的出版内容资源,并以各种不同的形式存在。作为内容组织和管理的有效工具,元数据对出版内容资源整合及出版产业的可持续发展具有重要意义,按照元数据标准的要求对内容资源进行唯一标识和规范化描述,能够有效促进内容资源的处理、交换、共享及开发利用。为了满足数字出版不同业务环节针对不同类型资源的标识与描述需求,数字出版领域已出现并应用了多种元数据标准。随着数字出版产业规模的日益扩大,数字出版的技术人员更需要对相关的元数据标准具备完整的了解,本书的写作正是基于这一需求。

本书第1章首先介绍了数字出版元数据的基本概念、作用及其在国内外的的发展状况,然后在后续章节针对数字出版全流程所涉及的元数据标准的内容、功能和特点给出了详细的介绍和分析。第2章对用于出版物及内容资源标识的ISBN、ISSN、DOI等标准进行了介绍与对比分析。第3~8章分别介绍了用于出版物或内容资源属性描述的图书在版编目数据、中文图书标识规则、图书流通信息交换规则、中文机读目录格式、DC元数据和图书ONIX等标准,并给出这些元数据标准的应用示例及分析。第9章重点针对元数据互操作问题给出了元数据转换的技术途径。第10章对数字出版元数据的趋势和未来进行展望。

本书由孙广芝统稿,邢立强、杨秀丹审校。本书的第1~3章由孙广芝编写,第4章由杨玲、陈立娜编写,第5章由张保玉编写,第6章由杨秀丹、戚菲、汤艳莉编写,第7章由邢立强、杨建军编写,第8章由孙广芝、孙文峰编写,第9章由孙广芝、刘彦东编写,第10章由邢立强编写。

本书对数字出版元数据进行了系统、全面的论述，通过对数字出版涉及的各项元数据标准及相关技术的分析对比和应用介绍，可以帮助读者快速了解数字出版元数据的全貌，并根据需求选择对某一项技术和标准继续深入了解，适合对元数据、数字出版相关标准感兴趣的技术人员阅读。本书编写中引用了大量的专著文献，读者可以此作为研究数字出版元数据的参考资料，更加深入地了解各项技术和标准。希望本书能对那些想了解、学习和研究数字出版元数据标准和技术的人员有所启发，并吸引更多的人关注数字出版及元数据的研究与应用。

目 录

Contents

第 1 章 | Chapter 1

数字出版元数据概述	1
-----------------	---

1.1 数字出版概况和发展	2
---------------------	---

1.1.1 什么是数字出版	2
---------------------	---

1.1.2 数字出版在我国的发展	3
------------------------	---

1.2 元数据及元数据标准概况	6
-----------------------	---

1.2.1 什么是元数据	6
--------------------	---

1.2.2 什么是元数据标准	6
----------------------	---

1.2.3 元数据标准在数字出版中的作用	7
----------------------------	---

1.3 数字出版元数据标准概要	9
-----------------------	---

1.3.1 标识类标准	10
-------------------	----

1.3.2 描述型元数据标准	11
----------------------	----

1.3.3 元数据的互操作	13
---------------------	----

第 2 章 | Chapter 2

标识类标准	15
-------------	----

2.1 ISBN 及其在中国	16
----------------------	----

2.1.1 国际标准书号的起源及发展	16
--------------------------	----

2.1.2 ISBN 在中国的引入及发展	17
----------------------------	----

2.1.3 现行中国标准书号的构成	19
-------------------------	----

2.1.4 中国标准书号的实际应用	22
-------------------------	----

2.2 ISSN 及其在中国	26
----------------------	----

2.2.1 国际标准连续出版物号的起源及发展	26
------------------------------	----

2.2.2 ISSN 在我国的情况	28
-------------------------	----

2.2.3 数字出版环境下的 ISSN-L 的出现和发展	30
------------------------------------	----

2.3	ISRC 及其在中国	30
2.3.1	国际标准音像制品编码的起源和发展	31
2.3.2	中国标准录音制品编码	34
2.3.3	ISRC 的管理与应用	36
2.4	ISAN、ISTC、ISWC 及其在中国	37
2.4.1	ISAN 及其在中国	37
2.4.2	ISTC 及其在中国	39
2.4.3	ISWC 及其在中国	42
2.5	DOI 及其在中国	44
2.5.1	DOI 的起源及发展	44
2.5.2	DOI 在我国的情况	46
第 3 章	Chapter 3	
	图书在版编目数据	51
3.1	CIP 的起源与发展	52
3.1.1	CIP 含义的理解	52
3.1.2	图书在版编目的起源	52
3.1.3	图书在版编目在国外的发展	54
3.1.4	图书在版编目国际标准	54
3.1.5	中国图书在版编目的发展历程	55
3.2	标准主要技术内容	57
3.2.1	标准中的术语及定义	57
3.2.2	著录数据	58
3.2.3	检索数据	59
3.2.4	项目标识符	60
3.2.5	内容标识符	61
3.2.6	图书在版编目数据选取规则	61
3.2.7	图书在版编目数据的印刷格式	61
3.3	图书在版编目在我国的应用及其意义	62
3.4	CIP 的发展趋势	65
3.4.1	ECIP	65
3.4.2	CIP for E-Books	68

第 4 章	Chapter 4	
	中文图书标识规则	69
4.1	编制背景与过程	70
4.2	相关标准	71
4.2.1	GB/T 12450—2001 图书书名页	71
4.2.2	GB/T 12451—2001 图书在版编目数据	72
4.2.3	GB/T 5795—2006 中国标准书号	73
4.2.4	GB/T 3792.2—2006 普通图书著录规则	74
4.2.5	CY/T 39—2006 图书流通信息交换规则	75
4.3	本标准主要技术内容	76
4.3.1	范围	76
4.3.2	标识数据的构成	77
4.3.3	标识数据的描述结构	77
4.3.4	数据元素的描述	77
4.3.5	标识数据的构成	78
4.4	本标准的应用	95
4.4.1	中文图书标识数据应用实例	95
4.4.2	标识性编目	97
4.4.3	对数字出版的意义	99
第 5 章	Chapter 5	
	图书流通信息交换规则	101
5.1	CY/T 39—2006 图书流通信息交换规则概述	102
5.1.1	编制背景	102
5.1.2	编制过程	102
5.1.3	标准的定位	104
5.1.4	标准的特色	104
5.2	标准主要技术内容	108
5.2.1	概述	108
5.2.2	图书流通信息交换的内容	108
5.2.3	图书流通信息数据元素	111
5.2.4	文本文件交换规则	123
5.2.5	XML 文件交换规则	135
5.2.6	XML 文件输出实例分析	135

5.3	本标准的应用	139
5.3.1	打通行业信息链,提升供应链价值	139
5.3.2	掌握出版动态,满足市场需求	141
5.3.3	实现资源共享,加速商品流通	141
5.3.4	合理利用资源,降低交易成本	142
5.3.5	引用信息技术,规范业务行为	142
第 6 章	Chapter 6	
	中文机读目录格式	143
6.1	机读目录格式概述	144
6.1.1	机读目录的起源	144
6.1.2	MARC 的特点	145
6.1.3	CNMARC 的产生	146
6.2	CNMARC 主要技术内容	146
6.2.1	范围	146
6.2.2	CNMARC 格式结构	146
6.2.3	记录头标详细内容	148
6.2.4	数据字段详细内容	151
6.3	CNMARC 应用及记录样例	180
6.4	数字出版环境下 MARC 的发展	182
第 7 章	Chapter 7	
	DC 元数据	191
7.1	都柏林核心元数据概述	192
7.1.1	数字资源、网络资源对元数据标准的需求	192
7.1.2	都柏林核心元数据的起源及发展历程	193
7.1.3	历次 DC 会议的基本情况 & 主要成果	195
7.1.4	DC 标准化的情况	204
7.1.5	DC 在我国的引入 & 发展	206
7.2	标准主要技术内容	207
7.2.1	DC 规范文档概要	207
7.2.2	DC 的 15 个核心元素	210
7.2.3	DC 的修饰词	212
7.2.4	DC 元数据抽象模型	214

7.2.5 元数据应用纲要	218
7.3 DC 元数据的应用实例	219
7.3.1 DC 元数据在数字图书馆的应用	220
7.3.2 DC 元数据在跨国公司研究专题数据库的应用实例	222
7.3.3 DC 元数据在旅游资源数据库中的应用实例	223
7.4 DC 元数据的未来发展重点	224

第 8 章 | Chapter 8

图书 ONIX 标准	227
8.1 图书 ONIX 标准概述	228
8.1.1 ONIX 系列标准的起源	228
8.1.2 图书 ONIX 的版本更新及应用	230
8.1.3 图书 ONIX 在我国的标准化的	231
8.2 图书 ONIX 主要技术内容	232
8.2.1 图书 ONIX 产品信息的总体构成	232
8.2.2 图书 ONIX 数据元素总览	233
8.2.3 图书 ONIX 代码表	262
8.3 图书 ONIX 示例分析	266
8.4 图书 ONIX 的应用前景	281

第 9 章 | Chapter 9

元数据转换	285
9.1 元数据互操作问题的提出	286
9.2 解决元数据互操作问题的技术途径	286
9.2.1 语义互操作	287
9.2.2 语法互操作	288
9.2.3 结构互操作	289
9.2.4 基于协议的互操作	290
9.3 数字出版元数据的转换	292
9.3.1 MARC 的 XML 交换格式及其应用	293
9.3.2 DC 与 CNMARC 的转换	299
9.3.3 ONIX 与 MARC 的转换	309
9.3.4 元数据与本体之间的映射	310

第 10 章	Chapter 10	
	数字出版元数据的趋势与未来	313
10.1	元数据标准的协调与互操作	314
10.2	元数据服务的强化	315
10.3	元数据工作的专业化、自动化和社会化	317
10.4	结束语	319
参考文献		321

第 1 章

Chapter 1

▶ 数字出版元数据概述

数字出版作为出版业的一种新兴业态，自问世以来，发展非常迅猛。随着全球信息化进程的推进及数字技术向各个领域的不断延伸，我国数字出版产业规模迅速扩大，收入逐年大幅度递增，技术不断创新，产业形态日益完备，产业链趋于整合，数字出版已经成为我国出版业新的增长点。

在此背景下，数字出版领域的元数据标准越来越显示出其重要性和实用性。作为内容组织和管理的有效手段，数字出版元数据的应用既能实现出版物的唯一标识，也有助于对各类出版物从多方面进行准确规范统一的描述，能有效促进数字内容资源的处理、交换、共享及开发利用。

1.1 数字出版概况和发展

1.1.1 什么是数字出版

数字出版的含义是伴随着技术的不断进步而逐渐深化的。

《2005—2006 中国数字出版产业年度报告》指出：“数字出版就是用数字化（二进制）的技术手段从事的出版活动。”^[1]该定义认为，传统出版与数字出版的区别并不在于出版介质，而在于出版技术，只要采用了数字化的出版技术，无论最终的出版物是传统出版物还是电子出版物，该出版活动都属于数字出版。

学者张立在《数字技术与数字出版》一文中指出：“从广义上说，只要是用二进制这种技术手段对出版的任何环节进行的操作，都是数字出版的一部分。它包括：原创作品的数字化、编辑加工的数字化、印刷复制的数字化、发行销售的数字化和阅读消费的数字化。”^[2]该定义不但强调数字技术是数字出版的本质特征，还进一步明确了数字出版的业务范围。

《2007—2008 中国数字出版产业年度报告》在进一步阐释了数字出版概念的同时，还指出，“它包括传统出版业数字化的全部过程和结果，同时也包括新兴的数字媒体^[3]”，该报告不但划定了数字出版的边界，还指出了传统出版业的数字化和新兴的数字媒体产业已开始出现相互渗透、相互融合的趋势。

新闻出版总署《关于加快我国数字出版产业发展的若干意见（新出政发[2010]7号）》指出：“数字出版是指利用数字技术进行内容编辑加工，并通过网络传播数字内容产品的一种新型出版方式，其主要特征为内容生产数字化、管理过程数字化、产品形态数字化和传播渠道网络化。目前数字出版产品形态主要包括电子书、数字报纸、数字期刊、网络原创文学、网络教育出版物、网络地图、数字音乐、网络动漫、网络游戏、数据库出版物、手机出版物（彩信、彩铃、手机报纸、手机期刊、手机小说、手机游戏）等。数字出版产品的传播途径主要包括有线互联网、无线通信网和卫星网络等。”^[4]这一结论针对数字出版发展的实际状况，对数字出版的内涵及外延都做了相关界定，使得范围更加明晰，概念更加清楚。

[1] 郝振省. 2005—2006 中国数字出版产业年度报告 [M]. 北京：中国书籍出版社，2007.

[2] 张立，陈含章. 数字技术与数字出版 [J]. 编辑学刊，2006（3）：4-9.

[3] 郝振省. 2007—2008 中国数字出版产业年度报告 [M]. 北京：中国书籍出版社，2008.

[4] 新出政发[2010]7号：关于加快我国数字出版产业发展的若干意见 [EB/OL].

<http://www.gapp.gov.cn/cms/cms/website/zhrmghgxwcbzsww/layout3/index.jsp?channelId=2975&siteId=21&infoId=706056>.