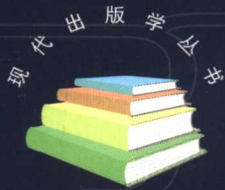


The Columbia Guide to Digital Publishing

哥伦比亚数字出版导论

〔美〕威廉·E.卡斯多夫 (William E. Kasdorf) 主编

徐丽芳 刘萍 译



苏州大学出版社



哥伦比亚 数字出版导论

The Columbia Guide to Digital Publishing

威廉·E.卡斯多夫 主编
William E. Kasdorf

徐丽芳 刘萍 译

哥伦比亚数字出版导论

The Columbia Guide to Digital Publishing

主编

威廉·E.卡斯多夫

译

徐丽芳、刘萍

责任编辑

李幸

出版发行

苏州大学出版社

(苏州市干将东路200号 邮编 215021)

印

苏州市民生印刷有限公司

(苏州市西门外寒山桥南首 邮编 215200)

开本 787mm×960mm 1/16

印张 11.38 字数 303千

版次 2007年8月第1版 2007年8月第1次印刷

◆ 苏州大学出版社

图书在版编目(CIP)数据

哥伦比亚数字出版导论/(美)卡斯多夫(Kasdorf, W. E.)主编;徐丽芳,刘萍译.—苏州:
苏州大学出版社,2007.8

(现代出版学丛书)

书名原文: The Columbia Guide to Digital Publishing

ISBN 978-7-81090-920-4

I. 哥… II. ①卡…②徐…③刘… III. 电子出版物—出

版工作—研究—哥伦比亚 IV. G237.6

中国版本图书馆 CIP 数据核字(2007)第 122235 号

著作合同登记号 图字: 10-2005-261 号

The Columbia Guide to Digital Publishing by William E. Kasdorf

Copyright © 2003 by

Columbia University Press Simplified Chinese translation copyright ©(year) by Soochow

University Press Published by Arrangement with Columbia University Press All Rights Reserved



现代出版学丛书

Contemporary Publication Studies Series

总策划

吴培华

策划

李寿春

哥伦比亚数字出版导论

The Columbia Guide to Digital Publishing

主编

威廉·E. 卡斯多夫

译

徐丽芳 刘萍

责任编辑

李寿春

出版发行

苏州大学出版社

(苏州市干将东路 200 号 邮编 215021)

印刷

丹阳市民生印务有限公司

(丹阳市西门外陵川绿岛南首 邮编 212300)

开本 787mm×960mm 1/16

印张 41.25 字数 693 千

版次 2007 年 8 月第 1 版 2007 年 8 月第 1 次印刷

书号 ISBN 978-7-81090-920-4

定价 55.00 元

前言



当今的数字图景令大多数出版人觉察到了危险。技术以惊人的速度演进,新发展层出不穷。这迫使我们重新思考出版业究竟要去往何方,而同时一些广为人知的障碍又羁绊了我们前进的步伐。信息风暴模糊了我们的视线,往往地图还没有画出来,道路已然改变。

我们需要指引,这正是编写《哥伦比亚数字出版导论》的初衷。本书力图做到有益、权威而且全面,作者们都是专家,拥有与其写作的主题相关的丰富学识。

一本好书可以帮助你达成目的。本书并不是一般的手册或指南,它并不是告诉你“如何做”,而是帮助你探索、发现最适合你的道路。它展示了众多相关的主题和看法,经过精心设计的结构使其超越了对具体问题的快速回答,而能够将答案置于更有意义的上下文中,并引导你思考许多也许你尚未意识到的问题。

数字革命最深远的影响是扩大了出版商这一角色的内涵与外延。在模拟时代,对“出版商”的定位是很清楚的——尽管普通人常常混淆了“出版商”和“印刷商”。尽管那时出版过程也涉及许多相关人员——作者、社外编辑和设计人员、生产人员、服务人员、印刷商、经销商、零售商——但是他们之间是彼此独立的,通常并不卷入也不需要了解其他人的工作。顾客也一样:我们将他想象成在火炉边的摇椅上悠闲地翻阅小说的人,而忘记了购买出版物的广大教师、学生、商人及其雇员、各种公私立机构(尤其是图书馆)。然而在数字时代,所有人——从创造出版物的作者、编辑和设计人员,到生产出版物的排版人员、印刷商和网站开发人员,再到帮助人们获取图书的书店、企业和图书馆——都联结在一起:他们所做的事情既会影响到别人,也受到别人的影响。本书的目的是要帮助所有这些人。

尽管我们鼓励作者们对各种手段和观念兼收并蓄,但同时也鼓励他们





采用合适的立场。好的向导不会说：“好吧，要上山有两条路，A 路和 B 路，你想要走哪条？”导游的责任是指出走 A 路要多花一倍时间，但是沿途有壮丽的瀑布和优美的风景，或者指出许多选择 B 路的人丢掉了性命。在出版领域，人不会因为走错路而丧命，但是公司会。

本书的作者们在写作过程中不约而同地达成了某种一致。有许多共同的主题，通常是作者在阐述某一个话题的时候自然地延伸过去的，这种一致性并非仅仅是要吸引读者的注意力。大多数作者认为未来会有多种形式的出版物，包括印刷的和电子的。大多数作者承认出版正变得越来越民主，越来越国际化，他们强调更新工作流程的重要性，建议采用创作、编辑、生产和发行环节高度整合的流程；他们还看到了可扩展标记语言 XML (Extensible Markup Language) 是达成上述目标的基础；他们认为技术的发展能够同时促进生产工艺和产品，而且随着时间的推移会日益提高两者的效率和有效性。这就是所谓的“殊途同归”。

一本好书应该使用读者习惯的语言。尽管讨论的主题是技术性的，但是本书的作者们仍用出版和相关领域里的大多数人都能理解的语言来写作。权威性是本书的追求之一，作者的目的是要清楚地解释数字时代与出版有关的诸多话题。一名读者可能是数字出版某个方面的专家，但是需要通过本书来了解其他方面的情况。一位经验丰富的图形设计师早就了解了本书所说的图形文件格式，但是也许打算更多地了解 XML 和元数据。一名编辑也许已经掌握了上述知识，但是对于 TIFF、JPEG 和 SVG 这三种图形文件格式的区别不甚明白。如果本书不仅能给那些人提供必要的信息，而且可以指明其他可用的印刷和网络资源，那么本书的编写目的也就实现了。

本书帮助出版工作者在遇到问题时快速地查找答案，但同样本书也是用来“读”的——如果你不仅需要信息，同时还需要某种洞察力的话。它不是无所不包的百科全书（尽管术语表部分提供了这样的帮助）；它的章节之间互相联系、彼此观照，共同构成了内容丰富的文本整体。下面的部分讨论了使用本书的一系列方法。本书同时出版印刷版和网络版并非偶然：这两种版本互相配合，既提供了动态的、链接丰富的电子资源，也方便读者将印刷品当做方便查询的手册和可详细阅读的书籍。

最后，本书将一路伴随着你。本书在出版过程中努力的目标是既充分地利用数字出版的技术与技巧，同时要使书本身成为这些技术与技巧的极佳例证。创作、编辑和生产本书印刷版与电子版的整个过程都是数字化的，





这不仅仅是“赶时髦”，而是想强调今天能够获得的技术的实际应用。在“如何使用本书”部分，对此将有更为详细的介绍。

本书的基本特点就是将持续地扩充和更新。尽管并不打算成为新闻和产品信息源，但是本书将提供数字时代影响出版的技术进步、行业发展和市场变化的最新信息。网络版的用户可以看到最近更新的版本；在有足够需求的时候，也将定期地发行本书的印刷版。这意味着本书的印刷版本“过时”了吗？应该说，编辑和制作本书的数字工艺是非常高效的，这就意味着一旦决定要印刷新版本，很快就可以把刚刚出版的内容“快照”下来以制作成书；而且本书的重点在于讨论和解释数字出版的根本问题和核心问题，而不是进行新闻报道，因此它会有长久的生命力。

如何使用本书

我们提供了很多使用本书印刷版和网络版的方法。首先，本书是以前后连贯的章节形式写作的，因此从头至尾阅读是理解每一个章节所讨论的主题的最佳方法；同时，每一章又被设计成足够小的“粒度”，它包含四级标题，以便读者能够方便地浏览并快速定位于某个专门的主题。

另一个很好的入口就是术语表。其中每个术语都有简短的定义。术语表的每一个词条都包含了本书中涉及到它的章节的序号，以方便读者回头参阅上下文。

还有，本书同时为印刷版和网络版的用户提供了更多的信息源。文中包含许多统一资源定位器（URL），以方便读者访问本书引用的网站（当然，在网络版本中可以直接点击进入）。

此外，每章末尾还提供了很多印刷资源和网络资源的书目信息。对于网络版本来说，随着时间的推移，不断增加的书目信息资源将会具有很高的价值。

本书的形成

如果说数字时代的出版有什么特点的话，那就是合作。本书的创作、编辑、生产和出版，就是在出版商哥伦比亚大学出版社，排版公司图书与期刊印制服务公司，内容管理系统开发商，开放图书系统（OBS）公司这些机构之间动态合作的例子，它生动地展示了当今数字技术的实际应用。我们不是因为一种技术本身的原因而去尝试它和使用它，而是要利用当今大多数出版商可以获得的方法和工具，来尽可能高效地、经济地和有效地生产本书的印刷版和网络版，同时要保证以一种适应未来技术发展的方式使用今天的





技术。

在这个项目背后有两个关键技术:OBS 开发的内容管理系统,以及印制服务公司开发的 XML 基础架构(利用相关的编辑和排版工具)。这两方面的结构和功能则是 OBS、印制公司和出版社之间从最初提出概念到最后实施期间持续不断合作的结果。在传统的工作流程中,这三方的工作相对独立,一般既不了解也不关心其他环节的方法和工艺。但在本书的形成过程中,各方工作人员不仅要了解,也要持续地参与他方的工作。

本书的原始内容是用 Word 软件写作的,使用的定制创作模板提供了一系列对应于本书各种元素的样式,以及方便实现这些样式的定制工具条;其次,编辑模板——与印制公司为哥伦比亚大学出版社开发的编辑系统配合使用——被用于清除作者文件中不符合规则的内容,它为编辑提供了更多功能,并且能够制作最终的 XML 结果文件。(与一般认为由编辑使用这些模板比作者使用更好的看法一致,全部编辑都大量地使用编辑模板,但是只有一半作者利用了创作模板的功能。)

本书使用了四种 XML 文档类型定义(Document Type Definition, DTD)。它们之间的区别是有启发性的。

首先,Word 模板的输出结果叫做结构良好的 XML(Well-Formed XML)。基础文档类型定义(Base DTD)是开始时的评估工具,它非常简单、随意,可以容纳很多章内容,而且可以容许结构上的不规则。在将内容输入内容管理系统(Content Management System, CMS)之前,使用对文本结构要求更高的输入文档类型定义(Import DTD)来分解文件,同时允许编辑在内容进入系统以前修订所有不符合规则的地方。在 CMS 管理内容的时候,不是为内容加标签,而是将内容组织进一个数据库。这为系统提供了强大的功能。CMS 在段落水平上跟踪内容,提供了修改、添加内容,以及管理一系列元数据和结构关系的机制,如本书副标题的号码、交叉引用链接和术语表链接等。它同样管理对内容的访问权限:章作者只能修改他/她自己的章节(通过万维网浏览器访问);一般编辑可以访问所有章节;而哥伦比亚大学出版社的责任编辑不仅可以访问所有内容,而且可以批准内容的最终出版。这些错综复杂的结构都由第三种文档类型定义即档案文档类型定义(Archival DTD)来描述。如果要把本书的内容从该 CMS 输入到其他系统,档案文档类型定义将会帮助定义所有的元素是什么,它们间的相互关系怎样。它也是生成其他形式出版物的基础,如未来可能的电子书版本等。





到了制作印刷版本的时候, CMS 生成由第四种文档——排版文档 (Composition DTD) 定义的 XML。设计这种 XML 的目的是要充分利用本书所使用的基于 XML 的高端排版系统的能力。该系统可以将标签丰富、高度结构化的 XML 文件作为输入代码, 而排版 DTD 要求本书内部复杂的关系表达为处理章节号码、交叉引用等元素的属性。XML 处理结构; 排版系统利用 XML 来显示设计者预先规定的内容和结构。利用 CMS 生成的排版 XML 文件可以高效地对章节进行排版, 因为它是百分之百的编码, 而且 DTD 保证了编码的正确性。

本书的打样过程也是完全数字化的。排版过程产生的 PDF 文件利用文件传输工具 (File Transfer Protocol, FTP) 传给哥伦比亚大学出版社, 并通过电子邮件发送给一般编辑和作者, 后者检查并订正以后发给出版社。出版社的责任编辑在 CMS 系统中完成所有的修改后, 输出正确的排版 XML 文件并用 FTP 传给印制公司。版面确定后, 将最终的 PDF 文件再用 FTP 传给图书制造商爱德华兄弟公司 (Edward Brothers)。

本项目另一个有趣的方面是索引。它使用印制公司开发的一种专门工艺, 使得专业索引人员能够用高端索引软件来创建索引 (对于本书, 是 CINDEX™), 而且可以在流程的任意阶段完成索引 (本书是在第一次排版以后的校样阶段完成的)。索引以 XML 或文本形式独立保存在 CMS 中, 不是在 XML 中嵌入索引条目, 而是在 CMS 保存的内容中包含特定的标记, 该标记描述了索引条目的出处——不是页码, 而是相关的正文部分——无论是一段还是几页。然后这种整合索引 (Intergrated Indexing™) 工艺在排版系统分页结束以后, 从 XML 文件中为特定版本的书籍生成索引。这不仅使得本书的索引与时俱进, 也使它能够适应不同的文件格式和分页系统。

数字技术不仅用于写作、编辑、生产和在线出版, 同时它还用于本书的印刷出版。数字时代为出版商提供了一系列选择。对于本书来说, 由于对印刷版的需求量较高, 因此采用胶印是合理的 (尽管充分利用了现代的数字印前技术)。由于爱德华兄弟公司提供了胶印和数字印刷相结合的方式, 因此有可能通过优化印刷技术来适应环境。本书的后续版本将会根据需求情况少量印刷——也是通过 PDF 文件——在爱德华兄弟公司的数字图书中心。

致谢

为了发扬本项目的合作传统, 我要感谢在此过程中作出贡献的许多人。



哥伦比亚大学出版社的斯蒂芬·斯特恩斯(Stephen Sterns)是本项目完美的编辑:懂技术(他对 DTD 的理解比我强),工作勤奋(术语表主要由他完成,在许多最后期限来临前加夜班),总是精神饱满,给人鼓舞,而且最重要的是,他掌握这本书的愿景。同样在哥伦比亚大学出版社,琳达·塞艮达瑞(Linda Secondari)从头至尾大力支持本项目,将团队团结在一起共同前进;她是一切活动运转的稳定的轴心。最后,要感谢詹姆士·雷米斯(James Raimis),他在退休前构想并启动了该项目。在开放图书系统公司,劳拉·菲尔莫(Laura Fillmore)的贡献远远超过了公司 CMS 的要求;她的洞察力和经验在本书出版过程中的各个方面都发挥了重要作用,本书的成功是与她的创造力和灵活性分不开的。她的工作团队中特别要提到的三个人是:大卫·卡斯蒂(David Cassidy),他以令人惊叹的专业精神和良好的工作状态来管理本项目,迪·兰德格兰恩(Dee Landergren)和萨尼·格里什(Sunny Gleason)在系统设计和编程方面作出了贡献。在印制公司,戴维·尼尔森(David Nelson)不知疲倦地为基础架构工作,随着项目的形成和发展,他反复写作 DTD,总是希望将工作做得更好。我所了解的数字时代的出版,尤其是 XML 方面的知识,是从戴维和他的同事那里学来的;其中四位贡献最大的人是戴维·尼亚加斯(David Niergarth,他的专家意见和富有思想性的见识是无价的)、丹尼尔·格罗斯伯格(Daniel Grossberg,他与两个戴维一起对书目作出了贡献)以及克里斯蒂娜·苏考斯基(Christina Sulkowski)和贾森·凯维拉(Jason Kivela)对网络版的案例作出了贡献,还要感谢克里斯·考特(Chris Cott)无可比拟的 XML 排版技能。我还要感谢克里斯蒂·品特(Christie Pinter)和詹尼弗·盖博(Jennifer Gibeaut),在本项目中充分显示了他们的才干。当然我最最要感谢的是创作了本书内容的作者们:他们是一群知识渊博的人,非常感谢他们能在百忙之中抽出时间来与我们分享他们的知识和远见。

比尔·卡斯多夫(Bill Kasdorf)

2002 年 11 月 10 日



目 录

前 言

1



1 引论：数字时代的出版

- | | |
|----------------------|----|
| 1.1 数字出版既是已知条件也是发展目标 | 1 |
| 1.2 不同类型的出版商,不同的解决方案 | 3 |
| 1.3 下一步是什么 | 22 |
| 1.4 开始吧 | 28 |



2 技术基础

- | | |
|-------------|----|
| 2.1 概述 | 29 |
| 2.2 计算机架构基础 | 30 |
| 2.3 处理环境 | 33 |
| 2.4 显示 | 40 |
| 2.5 数据存储 | 44 |
| 2.6 数据通信 | 47 |
| 2.7 其他重要资源 | 55 |



3 标记：XML 及相关技术

- | | |
|-------------------|----|
| 3.1 概述 | 57 |
| 3.2 HTML: 超文本标记语言 | 69 |



3.3 XML: 可扩展标记语言	76
3.4 交流、配合、协作	135



4 组织、编辑和链接

4.1 回顾: 从传统出版到电子出版的转型	138
4.2 组织内容	139
4.3 电子出版对传统出版发行模式的影响	143
4.4 有关内容的信息: 元数据	146
4.5 链接	150
4.6 总结	153



5 数据获取和转换

5.1 概述: 进入结构天地	160
5.2 理清结构与内容	164
5.3 你的数据来自何处	167
5.4 转换过程	173
5.5 分析问题	187
5.6 总结	193



6 排版、设计和图形

6.1 概述	195
6.2 文本、图形和页面版式: 页面三要素	196
6.3 设计和生产	197
6.4 三种重要技术	197
6.5 PostScript: 印刷出版的语言	198
6.6 PDF——Adobe 的可移植文档格式文件	201
6.7 印刷排版	203
6.8 图形类型和文件格式	219



6.9 色彩	234
6.10 页面制作	246
6.11 图像获取与处理	251
6.12 工作流程	257
6.13 印刷工艺	279
6.14 与印刷商合作	286
6.15 参考资料	287



7 可访问性

7.1 概述	290
7.2 近距离的观察	302
7.3 结语	322
7.4 资源和文献	323



8 数字印刷

8.1 概述	329
8.2 数字印刷技术	330
8.3 数字印刷的使用	336
8.4 现有的印刷系统	341
8.5 数字图书印刷的销售渠道	348



9 多媒体出版

9.1 概述	350
9.2 什么是多媒体	352
9.3 多媒体决策	352
9.4 体验多媒体	358
9.5 多媒体业务	359
9.6 多媒体技术	359





9.7 权利问题	370
9.8 结语	372



10 内容管理与网络出版

10.1 内容管理简介	374
10.2 内容管理系统的类型	382
10.3 内容管理系统的优点	387
10.4 内容管理中要考虑的问题	391
10.5 评估内容管理系统	399
10.6 实施后的问题	405



11 电子书和开放电子书出版结构

11.1 引言	407
11.2 OeBPS 简述	409
11.3 电子书概述	415
11.4 清晰地思考电子书	420
11.5 格式问题	425
11.6 OeBF 开放电子书出版结构	436
11.7 结论	457
11.8 对电子出版商的一些建议	458
11.9 更多信息	461



12 存档

12.1 存档的重要性	465
12.2 存档的其他问题	468
12.3 存档的位置	470
12.4 技术问题	473
12.5 有关发展和实施的问题	480



12.6	结论	485
12.7	资料	486



13 法律框架：版权和商标

13.1	版权	489
13.2	商标法	510
13.3	其他法律	516
13.4	诉讼：有没有全国通行的司法权	521
13.5	和顾客签订合同	522
13.6	结语	525



14 国际问题

14.1	纵览	526
14.2	全球互联网用户	527
14.3	STM 出版行业的先锋	532
14.4	建立网站	534
14.5	了解全球的市场环境	538
14.6	地理和国家优先	539
14.7	英语和其他语言	540
14.8	信息服务的新经济	542
14.9	全球网络广告	543
14.10	互联网上的市场	544
14.11	国际信息资源	547
14.12	国际出版法	549
14.13	结论	550



15 数字权利管理

15.1	概述	558
------	----	-----



15.2	基于权利的商业模式	560
15.3	DRM 技术	569
15.4	DRM 标准	579
15.5	法律方面的进展：重要的立法	587
15.6	供应商：为出版商提供 DRM 解决方案	598
15.7	市场状态	597
15.8	实施 DRM 过程中的问题	599
术语表		602
译后记		642



引论： 数字时代的出版

威廉·E. 卡斯多夫 (William E. Kasdorf)

当今,几乎所有的出版都在某种程度上数字化了,无论最后它是以电子的方式还是印刷的方式来传播内容。但是具体到每一类出版,则有它自己数字化的方式与道路。像学术期刊出版商和参考资料出版商,它们已经在电子出版的道路上走得很远,几乎要开始放弃印刷出版了。其他的,如大众杂志出版商和商品目录出版商,它们更加关注数字化生产技术;还有报纸出版商,它们则集中精力整合印刷出版和网络出版的工作流程。电子书尽管具有很强的优势,这一点对教科书出版商和学术出版商来说尤其如此,但它仍然处于实验阶段。在本书的引言部分将主要审视当今出版的方方面面,并就数字时代出版的许多问题提出看法,这些观点将在本书随后的章节中更深入地加以探讨。本章的最后部分将就当今出版商必须特别关注的重大技术问题提出建议。

§ 1.1

数字出版既是已知条件也是发展目标

对出版的数字革命人们一方面估计过高,一方面又估计不足。

如果与谷登堡时代的革命相比——一般也倾向于做这样的比较——人们的期望值很高。但是现实并不符合这样的期望。人们蔑视现在的电子书并且宣称他们不想通过屏幕来阅读,尽管我们中的许多人都花费了许多的时间在做这件事情。我们中的大多数人更喜欢阅读纸介质的报纸和杂志



(我们的孩子们情况可能不同);同时我们也理所当然地认为万维网是最现成的信息源,无论我们是要查询股票价格、统计数据,还是要寻找一条轰动的新闻或者一件衬衫。当我们购买以数字形式出版的内容时,无论它是故事、歌曲、电影还是一本书,我们都希望能够轻松享用它而不必为已经存在或者即将发明的数字设施支付任何额外的费用;尽管我们认为同样的音乐作品购买CD、磁带乃至黑胶唱片都是顺理成章的事情。对于如今的大多数人来说,“数字”就意味着“免费和容易”。然而一旦碰到障碍,无论是技术障碍、法律障碍还是商业障碍,我们都会将之作为证据,来证明数字革命并没有大家说的那么好。

我们并没有充分理解数字革命对印刷型出版物的生产方式造成的深远影响。我们惬意地带到床上或者带去海滩阅读的纸介质图书——那也许是我们通过万维网发现和购买的——很可能是作者在电脑上用文字处理软件创作出来的。图书的页面排版几乎肯定也是数字化的。具有讽刺意味的是,由于全数字化的工作流程并没有充分普及,很可能是先对作者纸介质的手稿进行编辑,然后在印度或者菲律宾扫描或重新录入稿件——作品以数字化的方式穿越了半个地球以后又回到了美国。尽管有些书连印刷也是数字化的,但带到床上或者海滩阅读的书更有可能是通过胶印机印刷的,至于将墨水由印版转印到纸上之前的整个印前过程则都已经数字化了。每天阅读的报纸和杂志也是如此:我们今天早上在安阿伯或者奥马哈阅读的《纽约时报》,它的所有文字和图片都是昨晚以数字文件的方式从卫星上下载的——当然,如果我们是和万维网一起成长的一代人,那么我们很可能直接在线阅读报纸。

1.1.1 真正的革命已经来临

真正的革命已经来临,最根本的变化就是内容与承载它的物理载体的分离。虽然,这两者本来就是相互独立的;但是,直到数字时代我们才充分地认识到,文章或者书并不一定要和一堆纸张联系在一起。

这是深刻的解放。

它导致两种后果,每一种都包含许多选择和许多机会,即:有许多新内容可以出版;有许多新方法来出版它们。

即使我们只打算出印刷版本,数字出版也提供了许多方式,而且这些方