

Content Management Bible, 2nd Edition

Bob Boiko



WILEY

Wiley Publishing, Inc.

Content Management Bible, 2nd Edition

Content Management Bible, 2nd Edition

Bob Boiko



WILEY

Wiley Publishing, Inc.

Content Management Bible, 2nd Edition

Published by
Wiley Publishing, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Copyright © 2005 by Wiley Publishing, Inc., Indianapolis, Indiana

Published simultaneously in Canada

Library of Congress Control Number: 2004114477

ISBN: 0-7645-7371-3

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

2B/RQ/RR/QU/IN

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Legal Department, Wiley Publishing, Inc., 10475 Crosspoint Blvd., Indianapolis, IN 46256, (317) 572-3447, fax (317) 572-4355, e-mail: brandreview@wiley.com.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services or to obtain technical support, please contact our Customer Care Department within the U.S. at (800) 762-2974, outside the U.S. at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Trademarks: Wiley, the Wiley Publishing logo and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, Inc., is not associated with any product or vendor mentioned in this book.



WILEY is a trademark of Wiley Publishing, Inc.

About the Author

Bob Boiko is a teacher, consultant, writer, programmer, and itinerant businessman. Bob is currently President of Metatorial Services, Inc. (www.metatorial.com) and Associate Chair of the Masters of Science in Information Management (MSIM) program in the iSchool at the University of Washington (www.ischool.washington.edu). Bob teaches information systems design, organizational management, and content management. He also conducts seminars and lectures around the world as part of his business. He has consulted on content management to a number of the world's top technology and publishing firms, including Microsoft, Boeing, Motorola, Honeywell, and Reed Elsevier. In addition to this book, Bob has written more white papers, articles, and reports than he cares to remember. Bob is helping to found and is serving as the first president of CM Professionals (www.cmprofessionals.org), a content management community of practice.

Bob began programming in 1977 and has practiced it since (it was always a great way to make money when he was broke). He entered the modern computer age, however, not as a programmer but as a writer. After earning undergraduate degrees in physics and oceanography and a Master's degree in human communication, Bob got his start in electronic information as a technical writer on contract at Microsoft. Among other projects, he wrote more than half of the MS DOS 5.0 User's Guide and one of Microsoft's first all-electronic User's Guides. From there, he began to develop electronic information systems on local networks, floppy disks, CD-ROMS, and when it was invented, the Web. In pursuit of electronic information and then of content management, he has created scores of applications and three businesses.

Bob lives in Seattle, Washington, U.S.A., with his wife Laura and sons Scotty and Corey.

Credits

Executive Editor

Chris Webb

Development Editor

Sara Shlaer

Production Editor

Angela Smith

Copy Editor

Mary Lagu

Editorial Manager

Mary Beth Wakefield

**Vice President & Executive
Group Publisher**

Richard Swadley

Vice President and Publisher

Joseph B. Wikert

Project Coordinator

Erin Smith

Graphics and Production Specialists

Beth Brooks

Andrea Dahl

Kelly Emkow

Carrie Foster

Lauren Goddard

Denny Hager

Joyce Haughey

Jennifer Heleine

Heather Pope

Heather Ryan

Mary Gillot Virgin

Quality Control Technicians

John Greenough

Joe Niesen

Carl W. Pierce

Proofreading and Indexing

Christine Sabooni, TECHBOOKS

Production Services

To Laura, Corey, and Scotty, for even more space

Foreword

Over the past three years I've encountered dozens of people around the world who purchased the first edition of the CMS Bible. Invariably, their copies are dog-eared and well-thumbed, because like any good reference, this book delves deeply into a complex subject that trails a surprisingly long history behind it.

Although many people associate “CMS” with the Web, the practice of content management goes back at least a couple of decades. For as long as we have had electronic documents, people have struggled with how best to manage the information those documents contained.

Early adherents of content management could be found especially among the technical documentation and scientific publishing communities. They had important reasons for wanting to reuse snippets of information at a very elemental level and, along the way, they redefined the entire notion of a document. But at the same time, these pioneers labored in relative obscurity, providing finished information products to enterprises that often had scant understanding of what went into assembling those publications.

Then the World Wide Web hit, and content management went mainstream — but this time for the specific purpose of automating the process of corporate Web site publishing. Content management became a front-office concern, drawing in marketing, design, and technology specialists, as well as line-of-business leaders.

Now the two communities — traditional information managers and Web content managers — are beginning to converge: core enterprise content increasingly gets exposed via the Web, whereas Web site managers seek to manage content further upstream in the organization. But convergence has not been easy and the process has brooked more than a little confusion among nearly all concerned. So the world needs a canonical written resource to explain, advise, and even exhort.

In short, we need a Content Management Bible.

I can't think of anyone better suited to the task of writing that bible than Bob Boiko. As a former technical documentation specialist, content management systems implementation leader, and now CMS strategy consultant, Bob brings a unique mix of practical experience to his explanations and advice.

The CMS Bible starts with an important discussion about content — what is it, how it differs from information and data, and why you should care about it. The book then delves into process. It then addresses technology. That is exactly the order in which you should undertake any CMS project.

Even if you already use a content management system (or more than one), some emerging trends in electronic information profusion should give you pause. According to recent studies, the amount of information generated worldwide doubles every three years and is likely continue to do so, with attendant compounding effects. Stop and think about that expansion

for a moment in connection with your own enterprise. Surely, success in all your other initiatives — from knowledge management to enterprise search to Internet publishing to customer care to e-commerce — will be increasingly predicated on your ability to manage the growing flow electronic content into and out of your organization. Whether you are about to implement a content management system or not, the CMS Bible can serve as your guide to connecting all that content to real business objectives.

So take the time read the entire book through once, but then keep it close at hand. Like its namesake, the CMS Bible is really a collection of smaller books. Each passage will come in handy at some point as you work to manage content more effectively.

Tony Byrne
Founder, CMSWatch.com

Preface

I originally created this book because I had to. For more than 10 years, I'd been stuffing my head so full of the design, programming, management, and content of information systems that I had to let some out before I could learn any more. Seriously, from the first time I matched a printed user's guide against the capabilities of Windows 3.0 Help, to the last time I sat with a dot-com client and discussed the impact of massive content management on the architecture of an e-commerce site, I have been living the transition from print to the computer screen. I've seen a ton of technologies and a slew of systems. I've learned enough to know that there is a lot to discuss and figure out. When my thinking on what I had experienced reached some sort of embryonic maturity, it hatched as the first edition of this book. Now, three years later, the second edition seems to me like a toddler; not entirely steady on her feet but up and walking and talking up a storm! The discipline, as well, is just beginning to recognize itself as a distinct entity with common practices and a community of practitioners.

My big points are:

- ◆ As an organization, you have groups of people who consume the information and services you provide. To be an effective organization, you need to treat these groups as distinct audiences and make use of all the available channels of communication with them. To communicate effectively with a variety of audiences through a variety of channels, you need the organization and focus that a content management system provides.
- ◆ Content is the information and interactivity that organizations must harness in order to deliver value to their audiences.
- ◆ Content management systems (CMS) collect, manage, and publish this information and interactivity.
- ◆ A CMS is not a CD-ROM that you install, start, and forget about. Rather, it is an ongoing process of knowing your information and your audiences and how to match the two in a set of publications.

This book attempts to lay a comprehensive foundation under these concepts and create a solid methodology for the practice of content management. Content management is not an end in itself but rather a means to becoming the most effective organization you can be.

About the Second Edition

Since the first edition of this book was published I have:

- ◆ Given dozens of talks at conferences based on its content
- ◆ Given at least a dozen half- and full-day workshops based on its content
- ◆ Taught eight 10-week classes based on its content
- ◆ Spoken formally and informally with scores of people about the book.

These interactions have led me to the conclusion that the information in the book is still pertinent for the average consumer. I still spend almost all of my time explaining the simplest parts of the book. The information is still new to most people, and none have said that they “know all that already” and would like newer information. Thus, I am not motivated to update the book (generally) because the information is out of date. Clearly, as a rewrite happened, I found much to update and topics to articulate more fully. But the main thrust of the book still stands.

In general, the comments I have received about the first edition have been overwhelmingly positive. Here is a summary of what seemed to work:

- ◆ **Book size:** While people love to kid me about the size of the book, it is clear that they regard this as a mark of comprehensiveness and definitiveness.
- ◆ **The mix of upfront theory and later practical information:** People have been happy to have the background that I provide in the first part of the book. It still seems to provide the only conceptual background that is available on the subject.
- ◆ **The style and readability:** Despite its size, I often hear that people really appreciate the open style and especially the sidebars and personal anecdotes in the book.
- ◆ **The variety of audiences that are addressed:** The book seems to work well for managers, information architects, and technologists.
- ◆ **The technology-neutral approach:** People like the fact that the information in the book stands regardless of the product or platform they use.

What I have heard on the negative side is that:

- ◆ **The book is too exhaustive** (or exhausting): It is difficult to get through it all.
- ◆ **The book details the largest of projects:** If you have a smaller project, where do you start and how do you use the book?
- ◆ **Not enough specifics in the technology:** Although people do not seem to want product-specific information, they do want (in Part V) a more comprehensive approach to the technology that CM systems use.
- ◆ **Vocabulary.** Some of the vocabulary I use (particularly the word *component*) is hard to understand and doesn’t match well with what is becoming the accepted standard.

Given the positive and negative reactions above, I have crafted the new edition in the following way:

- ◆ **Bringing it up to date:** I have reviewed the complete work to find any part that was out of date or where I have new material to add. As part of this, a **terminology review** has assured that the words I am using are the best for the discipline and the book.
- ◆ **Quickstart chapters and sections:** In Parts III, IV, and V, I provide new introductory chapters that give you a summary and top-line methodology for these parts. This approach helps users who do not have the time or the project size to justify an in-depth reading of these sections. They also serve to focus users who are interested in a smaller, minimal CMS project on the barest essentials of what they must do (and can safely defer) in their project. In many of the chapters in Parts III and IV, I provide a “Jumpstart” section that overviews and highlights the methodology presented in the chapter.

- ◆ **Invited sidebars:** From industry experts and CM practitioners around the world, I have solicited real-world examples and short case studies that appear in the book. These commentaries serve to enrich and solidify the concepts of the book and provide an array of opinion on the topics of the book. Because each sidebar represents the views of its author, the opinions and perspectives of these invited commentators may differ from mine. (Sidebars that are unsigned are my own.)
- ◆ **Technology taxonomy:** I have completely rewritten the final part (Part V) of the book to be a more comprehensive view of the technology behind content management. This part is not a guide to CM products, but rather a hierarchy of all the systems and sub-systems that could be included in a CMS. The systems include overviews as well as lists of features that the systems could have.

I believe that these new additions (in the neighborhood of 250 new pages) bring this work up to date and keep it at the front of this emerging discipline.

Who Should Read This Book

This book presents a complete model of content management as well as a practical guide to doing content management. It brings together current trends in content management, with my own experience and thinking, into a unified framework. It will be of foremost use to people tasked with designing or implementing content systems (programmers, designers, writers, and managers). It will also be of use to decision makers within an organization who need to develop an electronic communication or content management strategy.

The people who have the most to gain from this book fall broadly into these categories:

- ◆ **Managers** in charge of a content project. These managers need to know how to get a project initiated and completed. The information in Part III, “Doing Content Management Projects,” is most directly targeted to managers. In addition, senior level managers in charge of e-business, Web or communications strategy will benefit from this book. These people need to figure out how to do business in a wired world and, then, how to create an organization that can build and run an e-business system.
- ◆ **Practitioners** in areas such as content creation (writers and editors), content administration (database and information managers), and content publishers (site masters, designers, publishers, and Web designers). For these people there are relevant examples, methods, and overviews.
- ◆ **Programmers**, IT staff, and other technical professionals who need to understand how their skills fit into the broader framework of content management. The material in Part V, “Building a CMS,” will be of particular interest to technical folks.
- ◆ **Information architects** who want to better understand the nature of a large, organization-wide content management initiative. For these people, Part IV, “Designing a CMS,” may be particularly interesting.
- ◆ **Project staff**, who consist of the in-the-trenches creative, administrative, or publishing staff who want insight into their own jobs and to understand how their contribution fits into a larger framework.

- ♦ **Students** of business, technology, or information who want to keep up with the latest trends in information management. For them, content management represents a viable future occupation.
- ♦ **Business analysts and consultants**, who are called upon to help others figure out a content management strategy. With few standard sources to draw on in a rapidly changing field, those who help others will find help of their own in this book.

I also believe that there is value in this book for anyone else who wants to know what the new age of information management might look like. Even if you are not personally faced with creating a CMS, you may want to know what one is and how one is put together.

Why You Need This Book

To obtain maximum benefit from this book, you need to have felt the pain of too much content and not enough system to handle it. For those with a small site or only a little content to distribute, this work is overkill. For those who have tried and failed to, as we say, “tame the information beast,” this book can help. It provides tools that you can use to win the battle next time. For those who are faced with helping to create a large Web site or other publication type for the first time, this book can help you approach the problem in a coherent and reasoned way.

Content management is important. It can underlie today’s most significant digital technologies (including, among others, e-commerce, customer relationship management, personalization, advanced Web sites, and electronic communities). By understanding and properly implementing a CMS, organizations will have laid the groundwork on which the rest of these systems can stand. In so doing, they can save a tremendous amount of time and money and can unite these disparate systems with a single, enduring infrastructure.

This book matters for two additional reasons.

- ♦ The field of content management is in its infancy. I believe that this book helps define it. In my work I daily experience the confusion and frustration of people who need to define or implement a content management system for their organization and do not know how to approach it. These people are being bombarded by product-centered white papers and superficial ad-speak that present an all-too-simple picture. In contrast, this book provides a thoroughgoing and impartial framework upon which to base an understanding of the problems and the solutions of content management. It is part of a small but significant library of works on content management that are available today.
- ♦ There is a definite need for the kind of practical knowledge this book provides. The processes and practices that I, along with my colleagues, have developed can be of great use to people who need to implement and staff a content management system. From job descriptions to conversion code samples, you will find a good supply of methodologies, pointers, and insights. This practical knowledge, woven into an overall framework for implementing a content management system, should provide a powerful resource to anyone needing to understand or do content management.

When I began speaking on the subject of content management, my audiences consisted mostly of writers, marketing people, managers, editors, and librarians who were tasked with putting together a large Web site. Most had in-depth knowledge in their respective disciplines and some experience creating Web sites, but few had the resources for tackling the job they had taken on.

Today, my audiences consist of much the same people, but now they have job titles like Content Manager, Director of Web Strategy, and Chief Information Officer. In addition to creating an Internet presence, many are being tasked with developing an entire enterprise system for controlling the creation and dissemination of information. They have bigger titles and greater responsibility, but few extra tools to help them meet these new responsibilities. I hope that people in this position will find a useful methodology (or at least a kindred spirit) in this work.

How This Book Is Organized

This book consists of five parts.

Part I: What Is Content?

Content is information and functionality that has been harvested and organized toward some particular use. In this section, I dive into this definition and describe what is meant by organization and use. I use this definition to show why content is just information, but also more than information.

Part II: What Is Content Management?

Today, many people see content management as a way to create Web sites. As I present it, content management is a much broader process of collecting, managing, and publishing information to whatever medium you need. In this section, I present a full definition of content management and a complete model that you can use to understand and scope your own content management problems and solutions.

Part III: Doing Content Management Projects

This section describes content management from the manager's perspective. Given my model of content management, I turn to the practical problem of how you go about creating a CMS. Unfortunately, it is not simply a matter of buying a suitable product, installing it, and turning it on. No CMS you create or buy will tell you what content your organization should manage or why. Rather, doing content management is a process of getting your organization behind your project, designing all the details behind the system, selecting an appropriate system, and finally, implementing it.

Part IV: Designing a CMS

This section describes content management from the designer's perspective. Usually called information architect, the designer's job is to study and develop an interconnected system of collection, management, and publishing. The heart of this analysis is a metadata framework that ties all content together. In this section, I talk about this process in depth. I detail the exact sorts of information you have to amass to fully understand the system you want to create. I break the information to be collected down into a set of interrelated entities that rest on what I call the wheel of content management.

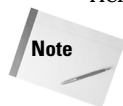
Part V: Building a CMS

This section describes content management from the perspective of the builder. Generally, CMS builders are content handlers, programmers, and information technology professionals. This section has information of interest to all three types of builders. For content handlers (who convert and tag content), it provides an under-the-hood picture of how content is constructed and how it can be processed. For programmers, it provides details on constructing publishing templates and programming content conversion systems. For IT professionals, it describes the software and hardware behind a CMS and how you might decide what is right for your organization.

Conventions Used in This Book

The book uses a variety of conventions to help you scan through it and quickly find information of interest. Most of the code samples in the book are either HTML or XML. In both of these sorts of code, I use uppercase letters for tag names and upper- and lowercase letters for tag attributes. In XML the capitalization of tag names really matters (it is a case-sensitive language). So, in a few cases, I break my all-uppercase convention where the sample code needs to be consistent with a system that uses a different convention.

Here are some of the other conventions you will see in the book:



Note

This icon presents a quick aside to the general topic. You will find information of particular note or important caveats to the current discussion.



Tip

This icon presents a short idea that you may want to implement as part of your own content management system design or implementation. There is advice, of course, throughout the book, but these paragraphs highlight “advice bytes” that you can immediately use.



Cross-Reference

This icon introduces a link between the current discussion and another that is related to it. Cross-references also appear within sentences to more closely link them to the ideas to which they apply.

This Is a Sidebar

Sidebar presents an extended aside. They contain stories and other devices that give you another perspective or angle on the discussion at hand. My sidebars are written by me or by invited experts, often in a lighter tone, and emphasize some of the quirkier aspects of content management.

Acknowledgments

Jim Larkin, president of Resources Online (www.ronline.com) is as responsible for this edition as I am. Despite a more than full time job and a new baby, Jim worked tirelessly to edit and oversee this edition. His fortitude and attitude made this edition happen. Thanks Jim. Along with Jim, Tom Richards helped shepherd the sidebars from vision to reality.

To everyone who wrote a sidebar for this edition, I would like to say thanks for continuing our great conversations in print and contributing to the larger conversation that is beginning to congeal into a community. And speaking of the community, I'd like to thank the people of CM Professionals (the original 23 and the thousands to come) for their contributions to and critiques of my thinking.

I'd also like to thank my many clients who over the years have seen the value in this approach and have supported it through its many twists and turns. I'd like to extend thanks, too, to Sara Shlaer and the production staff at Wiley for making my prose consistent, bringing to life my graphics, and putting up with my last-minute changes.

Finally, I'd like to thank Mike Eisenberg and the faculty and staff of the University of Washington iSchool for their confidence and advice, and their support of my work.

Contents at a Glance

Foreword	ix
Preface	xi
Acknowledgments	xvii
Introduction	xxxix

Part I: What Is Content? 1

Chapter 1: Defining Data, Information, and Content	3
Chapter 2: Content Has Format	13
Chapter 3: Content Has Structure	21
Chapter 4: Functionality Is Content, Too!	31
Chapter 5: But What Is Content Really?	41

Part II: What Is Content Management? 63

Chapter 6: Understanding Content Management	65
Chapter 7: Introducing the Major Parts of a CMS	85
Chapter 8: Knowing When You Need a CMS	113
Chapter 9: Component Management versus Composition Management	131
Chapter 10: The Roots of Content Management	147
Chapter 11: The Branches of Content Management	173

Part III: Doing Content Management Projects 199

Chapter 12: Doing CM Projects Simply	201
Chapter 13: Staffing a CMS	219
Chapter 14: Working within the Organization	243
Chapter 15: Getting Ready for a CMS	271
Chapter 16: Securing a Project Mandate	289
Chapter 17: Doing Requirements Gathering	303
Chapter 18: Doing Logical Design	315
Chapter 19: Selecting Hardware and Software	341
Chapter 20: Implementing the System	399
Chapter 21: Rolling Out the System	429

Part IV: Designing a CMS	451
Chapter 22: Designing a CMS Simply	453
Chapter 23: The Wheel of Content Management	459
Chapter 24: Working with Metadata	491
Chapter 25: Cataloging Audiences	517
Chapter 26: Designing Publications	537
Chapter 27: Designing Content Types	559
Chapter 28: Accounting for Authors	609
Chapter 29: Accounting for Acquisition Sources	629
Chapter 30: Designing Content Access Structures	647
Chapter 31: Designing Templates	683
Chapter 32: Designing Personalization	733
Chapter 33: Designing Workflow and Staffing Models	755
 Part V: Building a CMS	 791
Chapter 34: Building a CMS Simply	793
Chapter 35: What Are Content Markup Languages?	805
Chapter 36: XML and Content Management	821
Chapter 37: Processing Content	845
Chapter 38: Building Collection Systems	867
Chapter 39: Building Management Systems	917
Chapter 40: Building Publishing Systems	993
 Appendix: Epilogue.	 1061
 Index	 1065

Contents

Foreword	ix
Preface	xi
Acknowledgments	xvii
Introduction	xxxix

Part I: What Is Content?

1

Chapter 1: Defining Data, Information, and Content 3

What Is Data?	4
Content Is Not Data	5
Content Is Information Put to Use	7
Content Is Information Plus Data	11
From Data to Content and Back	11
Summary	12

Chapter 2: Content Has Format 13

Storage Formats: Storing Information	13
Rendering Format: Presenting Information	14
Dealing with Formatting	15
Categorizing Formatting	17
Formatting for effect	17
Formatting by method	18
Formatting by scope	18
Summary	19

Chapter 3: Content Has Structure 21

Structure Is Important	21
Structure Can Be Difficult to Create	24
How to Categorize Structure	26
Structure by purpose	27
Structure by type	27
Structure by scope	28
Summary	29

Chapter 4: Functionality Is Content, Too! 31

What Is Functionality?	31
Monolithic versus Mix-and-Match Functionality	32
Functionality Interspersed with Information	33
Managing Functionality Like Information	35
Publishing functionality on the Web	36
Where information and functionality diverge	38
Summary	40

Chapter 5: But What Is Content Really? 41

Content, Context, and Meaning	41
Creating Context Rules	43
Content Organization Starts with Purpose	45
Content Is Named Information	47
From Data to Wisdom	49
Data is raw and discrete	49
Data is nondiscursive and out of context	51
Information is processed and continuous	52
Information is discursive and full of context	54
Knowledge and wisdom can be information	57
Why Does Text Get All of the Attention?	59
Text is codified language	60
Text has the lead	61
Summary	62

Part II: What Is Content Management?**63****Chapter 6: Understanding Content Management 65**

Defining Content Management	65
CM Is Distributing Business Value	67
CM Is a Balance of Organizational Forces	69
CM Is the Combination of Content-Related Disciplines	70
CM Is Collection, Management, and Publishing	72
CM Is a Computer Infrastructure	74
The static Web site	75
The dynamic Web site	75
The Web CMS	77
The full CMS	79
The enterprise CMS	81
The Content Management Industry	82
Summary	83

Chapter 7: Introducing the Major Parts of a CMS 85

A CMS Overview	86
The Collection System	87
Authoring	88
Acquiring	89

Converting	92
Aggregating	93
Collection services	99
The Management System	100
The repository	100
The administration system	103
The workflow system	105
Connections	105
The Publishing System	106
Publishing templates	106
Publishing services	108
Connections	108
Web publications	109
Other publications	111
Summary	111
Chapter 8: Knowing When You Need a CMS	113
Gauging the Amount of Content	114
Managing the Size of the Contribution Base	115
Anticipating the Amount of Change	117
Knowing the Number of Publications	119
Estimating Complexity	121
Vacation company “A”	123
Vacation company “B”	124
Vacation company “C”	125
Vacation company “D”	126
Evaluating your own need	128
Summary	129
Chapter 9: Component Management versus Composition Management	131
CM Systems Can Be Modular or Linear	131
Component CM Systems	133
Composition Systems	135
Collection in a composition system	136
Management in a composition system	137
Publication in a composition system	138
Schema-Driven Systems	141
Collection in a schema-driven system	143
Management in a schema-driven system	144
Publishing in a schema-driven system	144
Which System Is Right for You?	145
Summary	146
Chapter 10: The Roots of Content Management	147
The Knowledge Base of the Publishing Industry	148
The publication	148
Content collection	149
Abstracting the process	149

The Principles of Document Management	151
Files contain content	152
Files store BLOBs	153
Publications are files	153
Document management systems versus content management systems	154
The Limits of Information Technology Groups	155
IT departments have avoided the Web	156
Content begins where IT groups leave off	156
The Legacy of the Multimedia Industry	157
Electronic publications	159
End-to-end content management	159
The Challenges of Technical Communication	160
Huge information bases	160
Simultaneous publications	162
Communication Theory	163
The Traditions of Library and Information Science	164
Information behavior	164
User services	165
Knowledge representation	165
Information retrieval	166
The Technology of Software Development	168
Collection technologies	168
Management technologies	168
Publishing technologies	168
The functionality in electronic publications	169
The Premises of Marketing	171
Summary	172

Chapter 11: The Branches of Content Management 173

Personalization	174
What is personalization?	174
Content management underlies personalization	176
Advanced Web Sites	177
Multiple Publications	179
E-commerce	182
Catalogs and content management	183
E-commerce functionality and content management	184
Knowledge Management	185
What is knowledge management?	185
Knowledge is content to manage	186
Online Communities	187
What is a community?	188
How are online communities constructed?	188
Other Kinds of Management	193
Digital asset management (DAM)	195
Learning object management (LOM)	195
Source management	195
Digital records management	196
Digital rights management (DRM)	197
Summary	198

Part III: Doing Content Management Projects**199****Chapter 12: Doing CM Projects Simply 201**

Why Create a Minimal CMS?	201
Staffing on a Shoestring	203
The business person	203
The content person	204
The publications person	204
The technology person	205
Getting Ready for the Project	205
Exploring the organization	206
Finding the right project	207
Finding the right sponsors	207
Getting a mandate	208
Key planning deliverables	209
Doing Design	209
Getting minimum requirements	209
Logical design essentials	210
Key design deliverables	210
Implementation	211
How low-tech can you go?	211
Slimming down product selection	214
Saying no	214
Key implementation processes	215
Key implementation deliverables	216
Deployment	217
Key deployment processes	217
Key deployment deliverables	218
Summary	218

Chapter 13: Staffing a CMS 219

About CMS Jobs	219
Managers	221
Content manager	221
Project manager	223
Production manager	223
Business Analysts	224
Information Architects	225
Content analyst	226
Metator	227
Infrastructure Staff	228
CMS administrator	229
Deployment analyst	230
Trainer and documentation specialist	231
Software Developers	231
Software analyst	231
Template and CMS developer	232
Custom application developer	232
Software integrator	233
Test analyst and test engineer	233

Publications Staff	234
Publication analyst	234
Publication designer	235
Page developer	236
User interface specialist	237
Content Processing Staff	237
Conversion analyst	237
Tool developer	238
Content processor	239
Content QA specialist	239
Content Creation Staff	239
Acquisitions manager	240
Traffic cop	240
Writers and other content creators	240
Editor	241
Summary	241

Chapter 14: Working within the Organization 243

Content Management and the Organization	243
Content Management versus Other Systems	247
Tracking Information Flow in the Organization	247
Understanding your information	249
Understanding your functionality	249
Understanding Organizational Roles	251
Business units generate value	251
Editorial teams unify content	251
Marketing teams direct and unify publications	252
IT groups build and maintain infrastructure	253
What Can Be Shared?	253
Product	253
Code	254
Content	254
Publications	255
How do you decide?	255
Exploring Organizational Models	256
Collection variations	257
Publishing variations	258
Management variations	261
Using Functional Collection and Publishing	263
Organizing collection systems and publications by type	263
Creating functional teams	264
Identifying Your CMS Hurdles	267
Summary	270

Chapter 15: Getting Ready for a CMS 271

Readiness Jumpstart	271
Understanding the CMS Project Process	272
Techniques for Getting the Job Done	274
Start with the project team	275
Look for pain in the organization	277
Assess the current mandate	278
Assess the organization's assumptions	278

Taking Stock of the Deliverables	281
The readiness assessment	281
Document inventory and analysis	283
The state-of-the-content system report	284
The education plan	285
A preliminary project plan	286
A risk assessment	287
Taking Stock of Your Staffing Needs	288
Summary	288
Chapter 16: Securing a Project Mandate	289
Mandate Jumpstart	289
What to Agree on	290
Techniques for Getting the Job Done	291
Recognize sponsors	291
Learn about your sponsors	293
Taking Stock of the Deliverables	294
Sponsor profiles	294
A hierarchy of issues	295
Notes and minutes	295
The mandate process	296
The project mandate	299
Taking Stock of Your Staffing Needs	302
Summary	302
Chapter 17: Doing Requirements Gathering	303
Requirements Jumpstart	303
What Are Requirements?	304
Techniques for Getting the Job Done	304
The requirements process	305
The requirements	306
Approaching the organization	310
Taking Stock of the Deliverables	311
The requirements plan of attack	312
The requirements document	312
Taking Stock of Your Staffing Needs	313
Summary	313
Chapter 18: Doing Logical Design	315
Logical Design Jumpstart	315
What Is Logical Design?	316
Techniques for Getting the Job Done	318
Why do logical design?	319
From business to system	320
Iterating through your design	322
Triangulating on constraints	323
Managing the details	325
Taking Stock of the Deliverables	325
The design plan of attack	326
The collection design document	327
The management design document	328
The publication design documents	330

An audience analysis	332
A localization plan	333
A risk assessment plan	335
A revised project plan	336
An executive summary	337
Taking Stock of Your Staffing Needs	338
Summary	339

Chapter 19: Selecting Hardware and Software 341

System Selection Jumpstart	341
The Product Paradox	343
Build, Buy, or Rent?	348
Building a CMS	348
Buying a CMS	349
Renting a CMS	349
Techniques to Get the Job Done	351
How to select decision makers	351
How to select a product	353
Taking Stock of the Deliverables	364
Product files	364
The selection criteria and RFP	364
The score card	365
The design diagrams	365
The decision report	366
Risk assessment update	366
Project plan update	367
Taking Stock of Your Staffing Needs	367
Sorting Through the CMS Selection Criteria	369
Business criteria	370
Overall criteria	371
Collection criteria	377
Management criteria	382
Publishing criteria	390
Summary	397

Chapter 20: Implementing the System 399

Implementing the System Jumpstart	399
Looking at the Process So Far	401
Looking at the Project So Far	401
Techniques for Getting the Job Done	402
Cutting back	402
Do you have one project or many projects?	403
Detailing the implementation process	405
Taking Stock of the Deliverables	418
Collection specifications	420
Management specifications	421
Publication specifications	422
The project plan	424
The risk assessment plan	426
Taking Stock of Your Staffing Needs	427
Summary	427

Chapter 21: Rolling Out the System 429

Rolling Out the System Jumpstart	429
What Is Deployment?	430
Techniques to Get the Job Done	431
Creating documentation	431
Doing training	437
Powering up the system	440
Preparing content	444
Revising the system	444
Taking Stock of the Deliverables	446
The deployment specification	446
The staffing plan	447
The training plan	447
The documentation plan	448
The maintenance plan	449
Summary	450

Part IV: Designing a CMS**451****Chapter 22: Designing a CMS Simply 453**

Logical Design Essentials	453
The Entities at a Glance	454
Logical Design: An Example	455
PLAN International — an example organization	455
A very simple logical design	455
Getting Beyond the Simple Logical Design	458
Summary	458

Chapter 23: The Wheel of Content Management 459

The Content Management Entities	459
Introducing the CMS wheel	461
Entities add a layer of abstraction	463
Goals and requirements	464
Audiences	468
Publications	471
Content types	474
Authors	478
Acquisition sources	480
Access structures	483
Workflow and staffing	484
Think, Plan, Integrate	488
Summary	489

Chapter 24: Working with Metadata 491

What Is Metadata?	491
What does meta mean?	491
What does metadata mean?	492
What does metatorial mean?	494

- The narrow view of metadata 496
- The wide view of metadata 496
- Metadata and content management 497
- Understanding the Types of Metadata 498
 - Structure metadata 498
 - Format metadata 500
 - Access metadata 502
 - Management metadata 503
 - Inclusion metadata 504
- Categorizing Metadata Fields 506
- Metatorial Processing 508
 - The metator 510
 - The metatorial guide 511
- Localizing Metadata 514
- Summary 515

Chapter 25: Cataloging Audiences 517

- Cataloging Audiences Jumpstart 517
- Serving versus Exploiting an Audience 518
- What Is an Audience? 520
 - Audiences and communicators 520
 - Audiences and marketing 521
 - Audiences and users 522
 - How many audiences do you have? 522
- Audiences and Localization 523
 - What is localization? 524
 - What are your localities? 525
 - Audiences and locality 526
 - What gets localized? 526
 - Localization and content management 527
- An Example Audience Set 528
- Analyzing Audiences 529
 - Think 529
 - Plan 530
 - Integrate 535
- Summary 536

Chapter 26: Designing Publications 537

- Analyzing Publications Jumpstart 537
- What Is a Publication? 538
 - Publication purpose 539
 - Publishers 541
 - Audiences 542
 - Messages 543
 - Authorship 545
 - Publication format 546
 - Publication structure 546
 - Publications can be good or bad 547
 - Publication gotchas 549

Analyzing Publications	552
Think	552
Plan	553
Integrate	557
Summary	558

Chapter 27: Designing Content Types 559

Content Type Jumpstart	559
The Idea of a Content Model	560
What Are Components?	563
Components are like objects	564
The basic unit of content management	565
How do you divide content?	566
Components versus pages	568
Content has types and components	569
Content Types Have Elements	570
Elements have types and values	571
Unique identification	572
Locality in components	572
What Do Components Look Like?	573
Components in flat files	573
Components in structured files	574
Components in relational databases	575
Components in object databases	575
The Relationship between Functionality and Components	576
A functionality component	577
Functionality content types and components	580
Analyzing Content Types	581
Think	582
Plan	582
Integrate	589
A Sample Set of Content Types	590
Affiliations	592
Locations	593
Solutions	593
Offering	594
Annual Reports	595
SEC Filings	596
Events	596
Press Releases	597
Newsletters	598
Outside News	599
Outside Links	600
Articles	601
FAQs	601
Customers	602
Jobs	603
Downloads	604
Bio	604
Standard management elements	605
Summary	608

Chapter 28: Accounting for Authors 609

Authoring Jumpstart	609
What Is Authoring?	610
Who can become an author?	611
Harvesting creative product	611
Authoring tools	612
Authoring prior to the CMS	612
The Author's Attitude	613
Changing the author	613
Changing the content	615
Your Attitude Toward Authors	615
Technical savvy	616
Influence	618
Incentives	618
A savvy/influence matrix	619
Analyzing Authors	621
Think	621
Plan	621
Integrate	627
Summary	627

Chapter 29: Accounting for Acquisition Sources 629

Acquisition Jumpstart	629
What Is Acquisition?	631
Found sources	631
Syndicated content	632
To syndicate or connect?	635
Acquiring functionality	635
Is It an Author or an Acquisition Source?	637
Analyzing Sources	638
Think	638
Plan	639
Integrate	645
Summary	645

Chapter 30: Designing Content Access Structures 647

Access Structure Jumpstart	647
Understanding Access Structures	650
Publication navigation versus CMS access structure	652
The content domain	653
Hierarchies	656
Indexes	659
Cross-references	662
Sequences	666
Full-text search	668
Analyzing Access Structures	669
Think	669
Plan	670
Integrate	680
Summary	681

Chapter 31: Designing Templates 683

Templating Jumpstart	683
Publications and Templates	684
Bridging worlds	685
Mixing the static and the dynamic	687
Building pages	689
Creating a system of publications	689
Using a template processor	690
Understanding template logic	691
Using templates within templates	692
Web templating	695
Print templating	703
Fax templating	706
E-mail templating	707
Analyzing Templates	707
Think	707
Plan	708
Integrate	715
A Sample Set of Templates	715
Component templates	717
Navigation templates	723
Summary	731

Chapter 32: Designing Personalization 733

Personalization Jumpstart	733
Publications and Personalization	734
Personalization and the audience	734
Personalization and components	736
Personalization and rules	737
Personalizations in templates	738
Customization versus personalization	740
Dynamic and static personalization	742
Analyzing Personalization	742
Think	743
Plan	743
Integrate	752
Summary	753

Chapter 33: Designing Workflow and Staffing Models 755

Workflow Jumpstart	755
Understanding Workflow	757
Workflow triggers	761
Workflow objects	762
A perspective on workflow	764
Push-and-pull workflow	765
Introducing Tasks, Jobs, and Steps	766
Staff modeling	767
Staff, jobs, and tasks	768
Task-time calculations	770
Staffing calculations	771
Fudge factors	772

Fitting Localization into Your Workflow	773
Analyzing Workflow	777
Think	777
Plan	778
Integrate	784
Analyzing Staffing	784
Think	785
Plan	785
Integrate	788
Summary	788

Part V: Building a CMS

791

Chapter 34: Building a CMS Simply 793

What Is Physical Design?	793
Physical Design Essentials	794
A Simple Physical Design Process	796
Introducing the Technology Taxonomy	798
An index to the technology taxonomy	799
Summary	803

Chapter 35: What Are Content Markup Languages? 805

A Brief and Selective History of Markup Languages	805
What Is a Markup Language?	808
A Taxonomy of Markup Languages	809
ASCII versus binary	809
Format versus structure	810
Extendable versus nonextendable	811
Range of coverage	812
Working with Markup	812
Don't be baffled by syntax	812
The language versus the interpreter	814
Representing representation	815
The concept of nesting	815
The benefits of white space	816
People play in the margins	818
Summary	818

Chapter 36: XML and Content Management 821

What Is XML?	821
XML and data interchange	822
XML tagging	822
Management by Schema or DTD	827
Adding formatting	828
Using XML in Content Management	830
XML in collection	830
XML in management	830
XML in publishing	831
XML in integration	832
Help from the rest of the XML gang	832

Programming in XML	834
Who needs to know XML?	835
Introducing DOM XML programming	836
Summary	843

Chapter 37: Processing Content 845

What Is Content Processing?	845
Stripping	846
Mapping	847
Content Processing and the CMS	848
Focusing on the long-term benefits	848
Focusing on the short-term benefits	851
Distilling the essence of a process	852
Bringing people into processing	852
Tracking the master copy	853
Managing the Processing Project	855
Taking stock of the content inventory	856
Drafting the processing specification	856
Defining testing methods	857
Getting to the Core of Content Mechanics	858
Understanding the principles of mapping content	860
Summary	865

Chapter 38: Building Collection Systems 867

Getting Started	868
Minimize Disruption, Maximize Specific Value	868
Forms, Files, and Batch Processes	869
Driven by a Content Model	870
Authoring System	871
Integrated search	872
Integrated workflow	873
Spawning applications	873
Upload support	873
Advanced media support	874
Spell checking	874
Integrated editorial and metatorial guides	875
Off-line content creation	876
Preview	876
Feedback	877
Metadata support	878
Web forms	881
Other authoring applications	884
Conversion System	887
Basic import	889
Mapping files to content types	889
Batch processing	890
Integration with the aggregation system	890
Process management	890
Quality control (QC) and monitoring	892
Process and review tools	894

Acquisition System	896
Process and partner management	897
Rights and usage	897
Attribution management	898
Acquiring database records	898
Acquiring Web site content	900
Aggregation System	902
Automatically applied metadata	903
Metadata standards support	904
Workflow triggers	904
Segmentation	905
Editorial processing	906
Metatorial processing	909
Repository Interface	914
Updates and deletes	914
Submitting and storing files	914
Support for WebDAV	915
Summary	915

Chapter 39: Building Management Systems 917

What's in a Management System?	918
The Repository	918
Fitting into the CMS and the organization	919
Getting content in and out	920
A repository-wide schema	920
Global search and replace	921
Locating content	922
Bulk processes	923
Field type support	923
General storage requirements	924
Relational database systems	926
Object databases	942
File systems	951
Versioning System	953
Automatic versus manual versioning	954
Granularity of versions	954
Restoring	955
Differencing versions	955
Source branching	956
Source Control System (Content and File Sharing)	956
Sharing granularity	957
Locking	957
Check in and out	958
Notification	958
Audit trails	959
Localization System	959
Collection localization	960
Management localization	962
Publication localization	966

Workflow System	968
End user interface	969
Workflow administration	972
Workflow mechanics	974
CMS Administration System	976
Administrative dashboard	977
User administration	978
Logging	978
Reporting	979
Security	980
Link support	982
Media support	982
Intellectual property tracking and costing	983
External connections	984
Robustness	987
Summary	990

Chapter 40: Building Publishing Systems 993

Templating System	994
Producing target formats	995
Producing target units structure	996
Layout and surrounds	996
Template programming	997
Producing static and dynamic publications	1000
On-the-fly conversion of text and media	1001
Integration with publication authoring tools	1002
Mix-and-match templates	1002
Navigation building	1003
Personalization System	1017
A personalization dashboard	1017
Collecting data	1018
Building rules	1021
Delivering content	1024
Deployment System	1027
Staging	1027
Content-based deployment	1028
Distributing files	1028
Scheduled publication	1029
Web System	1030
Integration with an existing Web infrastructure	1030
Searching and indexing	1032
Browser-independence	1034
Distributing files across servers	1034
Web platform support	1035
Print System	1037
Technical publications	1037
Section and subsection support	1039
Narrative support	1040
Navigation support	1044
Dynamic print publications	1045
Producing Word files	1045

E-mail System	1048
E-mail types	1049
Destination pages	1050
Integration to an e-mail server	1050
Template chooser	1051
Personalization support	1051
Syndication System	1052
Syndication subscriber management	1053
Selecting content	1054
Building and distributing feeds	1055
Other Publication Systems	1056
Multiple preview	1058
Repository Interface	1058
File and directory creation	1058
Runtime dependency resolution	1059
Database and metadata output	1059
Summary	1059

Appendix: Epilogue.	1061
------------------------------------	-------------

Index.	1065
----------------	------