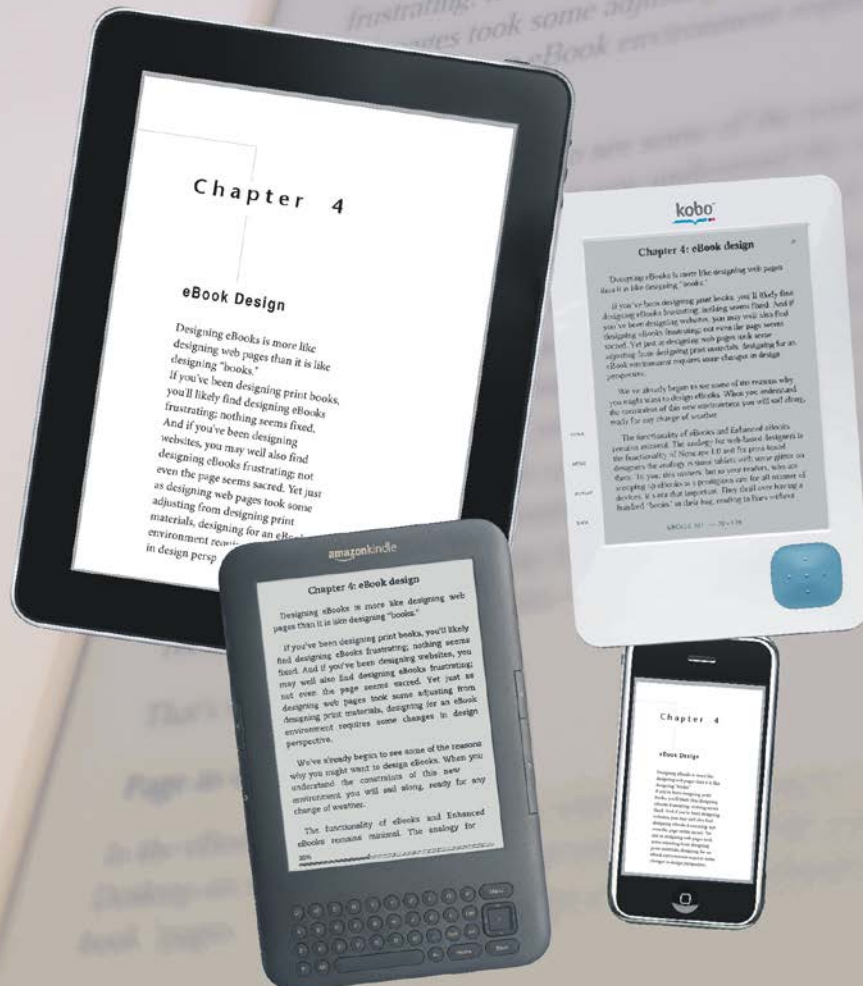


eBooks 101

The Digital Content Strategy for Reaching Customers Anywhere, Anytime, on Any Device



Ann Rockley and Charles Cooper

Foreword by Dev Ganesan, Aptara

Foreword

eBooks aren't just for books anymore.

Our authors, Ann Rockley and Charles Cooper, correctly present them as the ultra-convenient, flexible, mobile delivery formats that they are. eBooks provide rich and interactive user experiences, not just to customers of Barnes & Noble and of Amazon, but to *your* customers, prospects, employees, partners, and shareholders.

It doesn't matter if you sell books, tractors, cosmetics, medical devices, smart phones, or pharmaceuticals, or if you work in a university, consulting firm, or healthcare company. If you produce or distribute large volumes of content, you're a publisher. And as a publisher, your challenge is to provide your customers with the content they need when and where they need it, in the language they prefer, on the device of their choosing.

The problem when you're publishing an eBook is that there is no prescribed formula for success. Most traditional publishers are not good role models because they still employ production processes rooted in outdated (and often manual) print publishing practices that were never intended to tackle today's electronic challenges.

In order to profit - literally - from new digital markets, you must rethink the way you create, manage, publish, and deliver content. You must re-engineer your processes to create a more flexible and sustainable future. You must re-imagine a production process that frees your content to be easily transformed into whatever new formats and devices that your customers desire. And it all starts with recognizing the inefficiencies of tackling each new publishing challenge from scratch. For instance, instead of considering eBook creation as an add-on at the end of the print publishing lifecycle, you must adopt a content-centered publishing approach.

If you're considering a foray into eBook or app production, "eBooks 101: The Digital Content Strategy for Reaching Customers Anywhere, Anytime, on Any Device" is the best place to start your journey. Think of it as your go-to reference guide. It contains the secret sauce of eBook production, advice on avoiding pitfalls, device-specific and standard-specific information, design guidance, XML publishing tips, and a best-of-breed approach for creating, managing, and delivering eBooks and apps - in the most efficient way possible.

Rockley and Cooper call their approach the "Unified Content Strategy" - a repeatable, systematic process that supports the holy grail of publishing - separating content from its formatting, and making it possible to deliver content easily to any device, on any platform, in any format (print, web, mobile) - whenever and wherever your readers desire.

The authors have been helping organizations around the globe do just that, by developing repeatable, systematic, and strategic approaches to creating, managing, and delivering digital information. Their experiences over the past twenty years have made them some of the most highly sought after content management and digital enterprise publishing consultants in the world. They've distilled years of wisdom into this easy-to-read quick reference guide, the first eBook on eBook production

strategies, and a necessary one.

At Aptara, we help organizations leverage next-generation publishing approaches to ensure that they can meet the rapidly changing needs of digital consumers. Our customers are creating large volumes of content faster, more efficiently, and less expensively than ever before. And it's being published - often in parallel - to print, online, and mobile devices of all types, in formats that provide robust reader experiences.

Freeing your content from print-based workflows and developing device-independent content creation processes is not only possible, it's a highly achievable and profit-maximizing necessity. The bottom line: whatever type of content you publish, if you want to successfully compete in the digital age, you must design a process that enables profiting from its digital distribution.

"eBooks 101: The Digital Content Strategy for Reaching Customers Anywhere, Anytime, on Any Device" is not just a book about how to create eBooks. It's a quick and ready reference that will help you get your publishing processes in order so you can create engaging, interactive eBooks and apps that will dazzle your customers. And it's a handy instruction manual on how to do it right - the first time.

Ready? Set? It's time to ease into digital... the right way.

Dev Ganesan

President and Chief Executive Officer

Aptara, Inc.

Foreword	2
Introduction	6
Chapter 1: What is an eBook?	9
Chapter 2: Understanding eBook formats	15
Chapter 3: Understanding eReaders	28
Chapter 4: eBook design	45
Chapter 5: Understanding reuse	60
Chapter 6: Understanding structure	65
Chapter 7: Content analysis	69
Chapter 8: Using XML to future-proof your eBook	75
Chapter 9: Developing a Unified Content Strategy™	87
Chapter 10: eBook production	98
Chapter 11: Quality assurance	103
Glossary of terms	106
Acknowledgements and copyright	111
About the authors	112

Chapter 8: Using XML to future-proof your eBook

Earlier we discussed the importance of discovering and implementing “real” efficiencies in your eBook design process. The earlier in the workflow you incorporate these efficiencies, the more effective the process will be. XML is one of those efficiencies. XML provides a format that allows you to automatically publish to existing eBook formats as well as composition tools, and it “future-proofs” your eBooks by making them flexible enough to handle future formats - with no reworking your content to meet new requirements. That’s powerful stuff, and thankfully, it’s not magic.

What is XML?

In the beginning, back in the 1970s, there were disparate software applications that really did not play well together at all. Charles Goldfarb, a researcher at IBM, observed that many systems at IBM could not share information with each other, that they each used their own “language” (incompatible file formats) to format the text. SGML (Standard Generalized Markup Language) was born out of a project to build a system for creating, managing, and publishing legal documents.

A markup language is a set of annotations (tags) on text that describe how the text is to be structured, laid out, or formatted. A markup language is a modern implementation of instructions from editors to typesetters indicating how content should appear. HTML is the most widely known markup language. Any application that formats text has an underlying markup language. For example, RTF (Rich Text Format) was the underlying markup language for Microsoft Word until recently.

SGML

SGML was based on the following three principles:

- Computers need to be able to share files in a common format (they need to “speak” the same language).
- The markup of a document has to be extensible, for example, it must be able to be added to and handled by future software.
- There needs to be a way to identify the structure of content so that different pieces of content of the same type (for example, introduction) will share the same structure or rules.

HTML

The best known application of SGML is the language of the web, which is HTML. It was created in the early 1990s by Tim Berners-Lee at CERN (the European Laboratory for Particle Physics) as a way for scientists to share information. With the popularity of the web, HTML has become ubiquitous. So why not stick with HTML? Because HTML is limited in a number of ways:

- **HTML is a fixed tag set.** You cannot add your own tags to HTML. It's a fixed language.
- **HTML is designed for display.** HTML is perfectly suited for displaying documents in your browser, but that's all. It's not effective for print.
- **HTML is static rather than dynamic.** Static content makes it difficult to provide information in different ways based on user requests.
- **HTML is not structural.** With the exception of lists and tables, it doesn't contain any structural markup. Its focus on presentation makes it difficult, if not impossible, to process/manipulate chunks of HTML.

XML

As the internet came into being and developed, the community recognized that HTML had its limitations (insufficient tag set, could only describe format and not the content itself) and that SGML was essentially overkill. So a working group under the World Wide Web Consortium (W3C) began work on XML.

The goals of the W3C in developing XML included the following:

- Web-based delivery
- Open standard
- Based on SGML
- Formal and concise
- Easy to author and create
- Easy to develop applications for
- Extensible

The first point is extremely important. To be suitable for the web, the working group needed to create a streamlined version of SGML (SGML Lite) that would provide a lightweight markup standard, with all of the needed features and without the bulkiness that would overwhelm the web. XML was that streamlined version of SGML.

XML has been a great success. Use has spread beyond content markup to all sorts of other business and software applications.

Advantages of XML

XML provides a number of advantages:

- **Promotes consistency through structured documents.** Documents follow the same structures, with similar documents having the same content pieces.
- **Separates structure from format.** Structure in combination with stylesheets allows content to be published on multiple platforms without rework.
- **Enables reuse.** Structured information is easy to break into individual components for reuse or repurposing.
- **Enables dynamic documents.** Documents can be built from components, enabling you to select components dynamically.
- **Increases output flexibility.** Structured information is easy to manipulate to reconfigure or republish.

XML in publishing

From a content and publishing perspective, XML has become an extremely important technology for both big and small publishers. For complex content management with information reuse, XML is the technology of choice.

XML provides the ability to do a whole lot more than what can be done with traditional tools. The characteristics of XML that best support publishing are:

- Structured content
- Separation of content and format
- XSL (eXtensible Stylesheet Language) stylesheets

Publishing tools like Adobe InDesign and QuarkXPress both import and export to XML. And most importantly, XML provides the bridge between your content and print and multiple eBook formats.

You create one set of content with one set of XML tags. A different stylesheet is used to transform the content from one platform to another. For example:

- If you need to publish to InDesign, then you have a stylesheet that maps the XML tags to the InDesign tags.
- If you need to publish to the Kindle, then you have a stylesheet that maps the XML tags to the mobipocket format.
- If you need to publish to EPUBs, then you have a stylesheet that maps your XML tags to EPUB.
- If you need to publish to some new format and eReader, then you create a new stylesheet for that. You don't change your content; you only change your stylesheet.

It's almost magic! Of course there are design considerations for how you structure your content and build your content, but without XML you will continue to be subject to the tyranny of new technology.

Content separate from format

In XML, content is separate from format. That means instead of coding your content for Heading 1 and Heading 2, you simply use <title>. XML knows what level of heading it is, based on the heading's hierarchical position in the content. This means you can reuse the exact same component of content at a Heading 2 level in a chapter, but a Heading 1 level as part of a compendium. You don't have to change the tagging.

Here's another example: when creating content like a sidebar, instead of calling it a sidebar, you need to look at it from the perspective of what type of content it is. This is called semantic tagging. A sidebar could be a case study, it could be an example, more detail, or anything else, but whatever it is, your XML stylesheets know that when it goes into InDesign it's mapped to a particular spot on the page. When the content goes into EPUB, it's logically positioned at a natural break in the content (for example, before the next title).

One set of content, several sets of rules, and many, many outputs.

Intelligent Content: The real power of XML

XML enables us to design and prepare content in a way that is completely portable and open which in turn enables a wide range of applications to automate common content tasks such as formatting. If we make the content intelligent by tagging and structuring it, and by designing and preparing it for discovery and reuse, we can be freed from managing it within the "black boxes" of completed books.

We can move forward to actually managing the content itself once we take the step of making it intelligent. Intelligent content is the sum of all the recommendations we provide.

Intelligent content is content that is *structurally rich and semantically categorized, and is therefore automatically discoverable, reusable, reconfigurable, and adaptable.*

Let's look at this definition of intelligent content in a little more detail:

- **Structurally rich.** XML is structured content. Each and every element in your content is tagged with structural tags. Structure tells us how to write the content, enables us to automatically format the content for a particular output, and dynamically assemble content. We can perform searches or narrow our search to the particular type of information we are interested in (for example, look for all occurrences of a word in the context of a specific element such as case study).
- **Semantically categorized.** The word semantic means "meaning." Semantically categorized content is content which has been tagged with metadata to identify the kind of content within it. For example, you might tag your content with industry, role or audience, and subject area which allows you to automatically build customized information sets based on audience or subject. Without semantic metadata, it's very difficult to automatically, let alone manually, find the content we need.

- **Easily discoverable.** If the content has semantic tags and is structurally rich, it's a whole lot easier to find exactly what we are looking for. And when it is structurally rich, we can query the structure of the content to find specific information. Then when we add semantic tagging to the content we have a great deal of information that will allow us to zero in on exactly the content we are looking for (content mining).
- **Efficiently reusable.** Reusable content reduces the time to create, manage, and publish, also reducing translation costs. We can create modular structured content that can either be easily retrieved for manual reuse or automatically retrieved for automated reuse.
- **Dynamically reconfigurable.** Structured content is content where the words and the look and feel of the content are not embedded in the content. That makes it very powerful. When we know the structure of the content, we can output that content to multiple channels and reconfigure it to best meet the needs of the channel, or we can automatically mix and match content to provide what the customer wants when they need it. We can even transform (reconfigure) content from one structure to another, but only if we know what the structure is in the first place.
- **Completely adaptable.** We frequently create our content for a particular need or audience, but content can be adapted (used in a different way), often without our knowledge, to meet a new need. Content created for one title and one division could be reused for a completely new product, something we never originally envisioned.

More

This is only a portion of Chapter 8. The rest of the chapter includes:

- XML formats (TEI, NLM, DocBook, DITA)
- XML in the workflow
 - Traditional workflow
 - XML post-production
 - XML copyediting
 - XML authoring

For more on XML see the complete *eBooks 101: The Digital Content Strategy for Reaching Customers Anywhere, Anytime, on Any Device*.

Acknowledgements and copyright

A lot of people helped to make this eBook a success.

As always we would like to thank Scott Abel, The Content Wrangler, for his insightful edits, comments and suggestions.

We would also like to thank J. Michael Vore who provided editorial assistance and Laurel Simmons who provided a thorough copyedit.

And we would particularly like to thank our friends at Aptara, Inc., a digital publishing and content technology solutions provider, who provided technical insight along the way and kindly offered to transform this book into an eBook, the first of its kind. Not only did they provide conversion services for "eBooks 101" and our case study "Understanding the Glycemic Index," their design and development teams turned our storyboard into an app.

Disclaimer. Every attempt has been made to make eBooks 101 as accurate as possible; however, the industry is changing daily so some facts may have changed.

eBooks 101: The Digital Content Strategy for Reaching Customers Anywhere, Anytime, on Any Device Copyright © The Rockley Group Inc., 2011

The Unified Content Strategy is a trademark of The Rockley Group Inc.

Green Eggs and Ham, Dr. Seuss Properties™ & © 2010 Dr. Seuss Enterprises, L.P.
Concept, design, gameplay and technology © 2010 Oceanhouse Media, Inc.

Nixonland Copyright © 2008 by Rick Perlstein.

The Elements Copyright © 2010 by Theodore Gray.

Audio clip used with permission from Content Talks Episode #1: Intelligent Content
<http://5by5.tv/contenttalks/1>

All other Trademarks and Copyrights are the properties of their respective owners.

All rights reserved. No part of this publication may be stored, transmitted, used or reproduced in any manner whatsoever without prior permission from The Rockley Group Inc. except in the case of credited brief quotations embedded in articles, reviews and presentations.

ISBN 978-0-9865233-1-1

For information, contact:

The Rockley Group Inc.,
Email: moreinfo@rockley.com
Website: www.rockley.com

About the authors

The Rockley Group Inc.

The Rockley Group Inc. (TRG) has been helping organizations and publishers of all sizes to move to XML publishing strategies for more than 15 years. We help content managers and authors meet the increasing demands of creating, distributing and managing the content they create. Our team of experienced analysts brings a wide variety of expertise to the table and can help you avoid expensive pitfalls.

We've developed content reuse solutions that reduce the cost and effort to produce multi-platform information products including marketing collateral, documentation, online help, customer support materials, regulated content, eLearning, trade press, textbooks, eBooks, enhanced eBooks, and eBook apps.

The Rockley Group was established in 1995 to serve the information community. Our team has extensive experience at analyzing your content, your current workflow, identifying your pain points, and working with your goals and objectives to develop an XML digital publishing strategy that includes content strategy, technology recommendations and change management.

Our team of experienced analysts, information architects, project managers, information technologists, and technology partners provide our clients with the skills necessary to deliver multi-platform digital publishing solutions that work.

Ann Rockley

Ann Rockley is President of The Rockley Group, Inc. She has an international reputation for developing XML-based content strategies. She has been instrumental in establishing the field in eContent, content reuse, intelligent content strategies for multi-platform delivery, eBooks, and content management best practices. Rockley is a frequent contributor to trade and industry publications and a keynote speaker at numerous conferences in North America, Europe, and Asia-Pacific.

Ann's love of eContent began way back in 1983 when she worked with a start-up company that had developed an educational computer for grade schools. The computer had a Help key that did nothing. Ann set out to remedy this, developing GUI-based help materials nine years before WinHelp was first shipped. She then went on to work with Apple in 1987, where she helped develop HyperCard design guidelines (a Macintosh application for building hyperlinked content). She was hooked; eContent was her passion. Ann formed her own business in 1989 to focus on digital publishing.

Ann led Content Management Professionals, an international organization that fosters the sharing of content management information, practices, and strategies to a prestigious eContent 100 award in 2005. Ann co-chairs the OASIS DITA for Enterprise Business Documents subcommittee. Sometimes known as the "mother" of content strategy, she introduced the concept of content strategy in 2003 with her best-selling book, "Managing Enterprise Content: A Unified Content Strategy," New Riders Publishing, ISBN 0-7357-1306-5. Ann was ranked among the top five most influential content strategists in 2010.

Ann is a Fellow of the Society for Technical Communication and has a Master of Information Science from the University of Toronto. She is also the primary author of "DITA 101: Fundamentals of DITA for Authors and Managers," ISBN 978-0-557-69600-0.

Charles Cooper

Charles Cooper is Vice President of The Rockley Group, Inc.

Cooper has over 20 years of experience in quality assurance and over 15 years of experience in eContent, user experience, taxonomy, workflow design, composition, and digital publishing. He teaches, facilitates modeling sessions and develops taxonomy and workflow strategies.

Charles has assisted companies by analyzing their content, current workflow and taxonomy systems, helped to create new ones and worked to ensure that they are maintained on a consistent basis. He not only understands process, he understands the production tasks and can design a process that works for everyone in an organization. Charles always keeps the voice of the customer in mind when developing solutions.

He has a strong background in process and business planning and believes that taxonomies, structure, organization, workflow, and quality assurance processes must be designed to support the company as they work to provide products and services that their customers need.