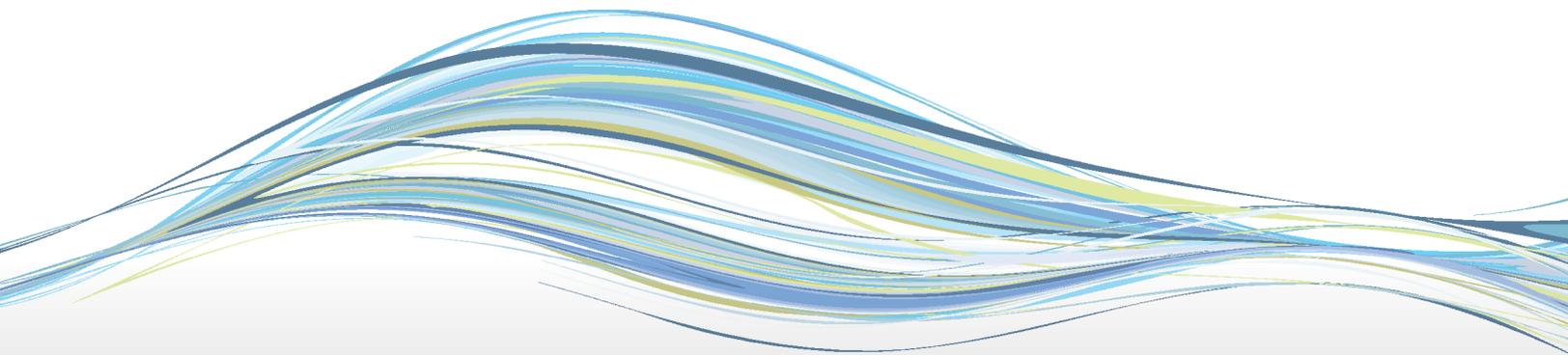




A Roadmap to Efficiently Producing Multi-Format/Multi-Screen eBooks

XML Tips

January 2013



This roadmap document is based on the November 8, 2012 webinar of the same name sponsored by Aptara. It offers a summary of the key takeaways and practical insights presented by Thad McIlroy and Stephen Driver on the topic of implementing XML, an unwieldy prospect for most publishers.

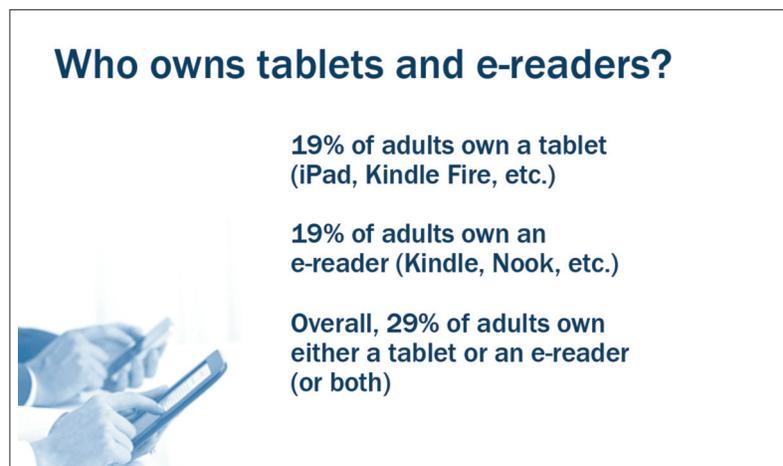
The webinar featured a case study on Rowman & Littlefield Publishing Group's successful adoption of XML for multi-format/multi-screen eBook production.

Thad and Stephen addressed three major questions facing publishers:

- Is there a way to simplify the complexities of formatting and optimizing content for the variety of tablets and eReaders?
- Does an XML-based workflow make sense for book publishers?
- If so, is there a realistic approach to implementing XML?

The Multi-Screen World

A 2012 Pew survey found that nearly one in 5 adults own either an eReader or a tablet, while many users own both an eReader and a tablet.

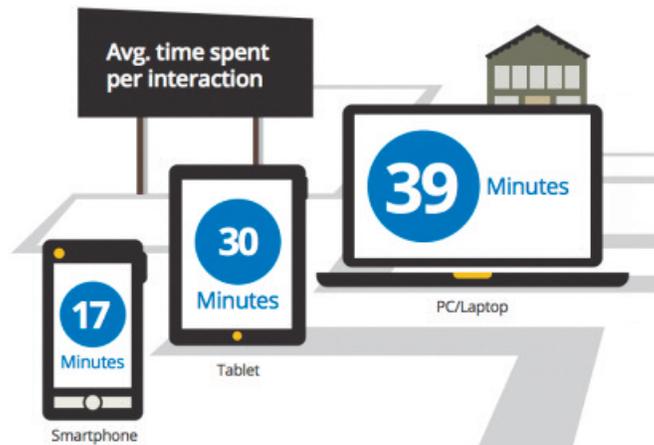
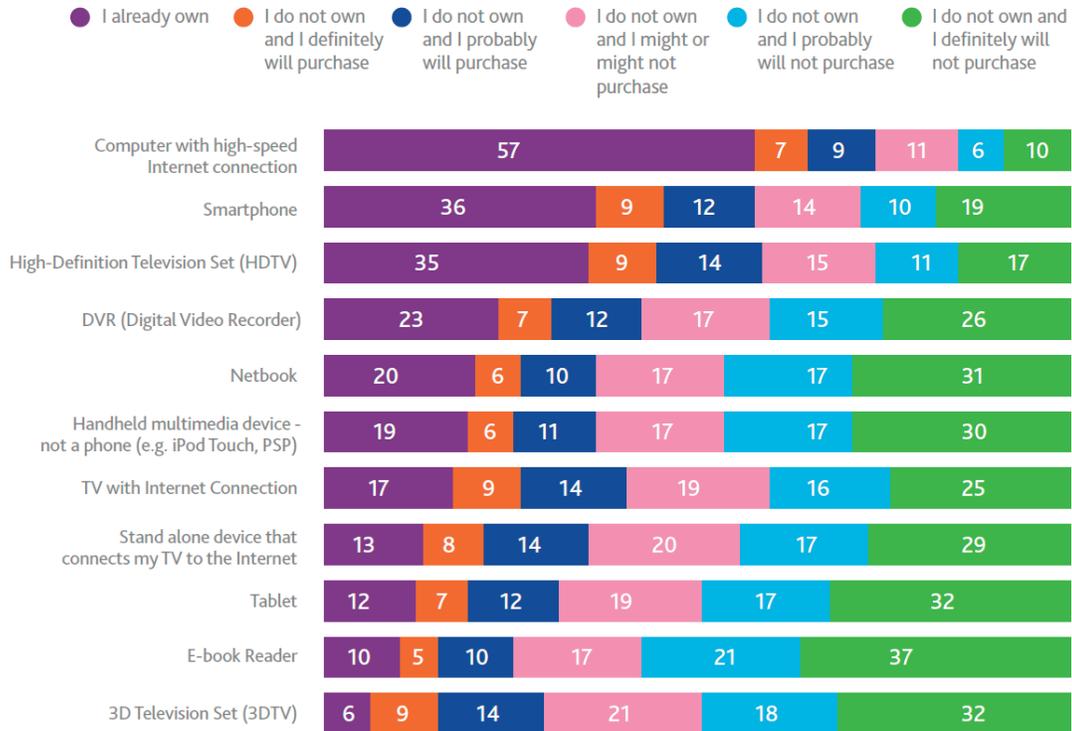


Source: Pew, "Tablet and E-book reader Ownership Nearly Double Over the Holiday Gift-Giving Period", January 23, 2012.

Consumers not only use multiple devices, but also expect to access the same content across all devices. For instance, readers expect to be able to move fluidly between a smartphone, tablet and PC, resuming wherever they left off reading an eBook.

Do you currently own or will you purchase any of these items in the next 12 months?*

Global Online Average



Source: Nielsen Global Survey of Multi Media Usage, Q3 2011

The Multi-Format World

As mobile devices have proliferated, so too have eBook formats.

The EPUB3 format, a free, open eBook standard from the International Digital Publishing Forum (IDPF), is the most widely supported by publishers and vendors. In response to EPUB3, Amazon introduced the KF8 format, designed for richer formatting of eBooks for tablet display. In the last year, Apple, Amazon and Kobo have each released file format variations for Fixed Layout eBooks.

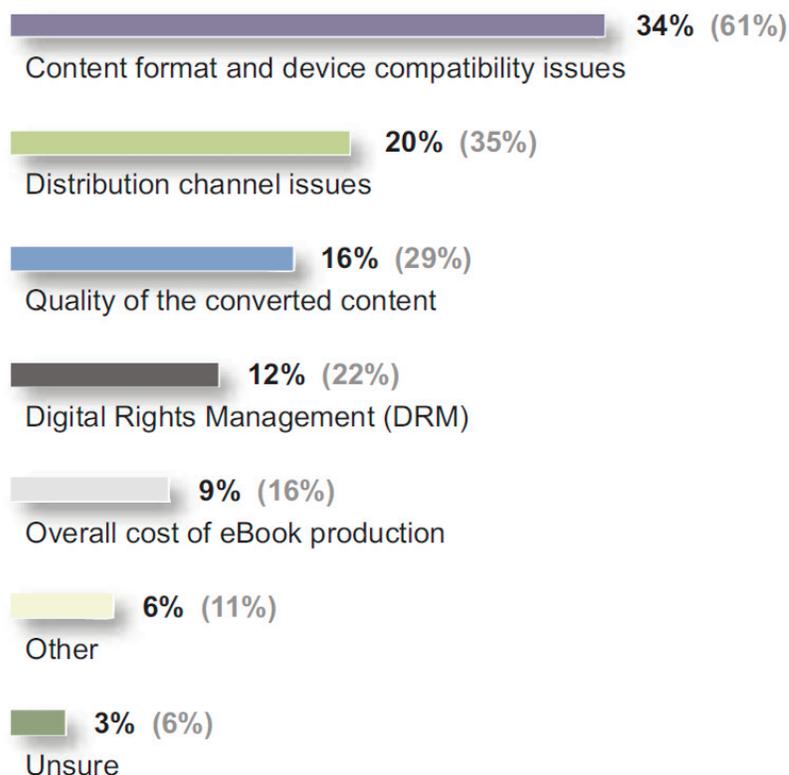
Even dominant formats are proving to be anything but standard. This chart illustrates their wide range of features:

	Kobo	iBook Author	iBooks	Nook	KF8
Region Magnification	●	●	●	●	●
Native Audio Support	●	●	●	◐	●
Native Video Support	●	●	●	◐	●
GIF/JPG/PNG/Support	●	●	●	●	●
Read Aloud	●	●	◐	●	●
Java Script Support		●	●		
Interactive Support		●	●		
3D Images		●	●		
Hot Spots		●	●		
Math ML		●	●		

Source: Aptara

eReaders' and tablets' characteristics and physical dimensions vary greatly, and therefore handle and display the same content very differently, truncating it if necessary. An eBook configured for the Amazon Kindle will look different on an iPad, Android tablet, or smartphone.

It's not surprising that the biggest eBook production issue stated by sixty-one percent of publishers is "content format and device compatibility."



Source: The 4th Annual eBook Survey of Publishers, Aptara & Publishers Weekly. September 2012

XML as a Solution

The diversity of eReaders and tablets and existence of multiple format standards seems to make producing multiple output versions of eBooks unavoidable. But coding individual book files for multiple formats and devices is impractical. It is slow, error-prone and too expensive for backlist or mid-list titles.

If handcrafting eBooks is unsustainable, then what's the solution?

Tags and style sheets enable efficient single-source, multi-channel publishing. And the most effective technology for employing tags and style sheets is XML.

There has never been a more compelling return on investment from implementing an XML workflow than exists in today's multi-format/multi-screen marketplace.

What is XML?

XML stands for Extensible Markup Language. It defines a set of rules for encoding documents in a **format** that is both **human-readable** and **machine-readable**.

- XML is a device-independent, system-independent method of storing and processing electronic text.
- XML guarantees that files encoded in XML will always be readable.
- XML is used to categorize and describe what data is, not how it should appear.
- XML works on any computing device using a wide variety of software.

What XML is Not

- XML is not a replacement for HTML, which is used to determine how data looks. (For more about the differences between the XML and HTML, visit Microsoft's short tutorial, [Going from HTML to XML](#).)
- XML is not software.
- XML is not a programming language.
- XML does not require a PC to interpret it.
- XML does not rely on a single company, system, or software. Relying on corporate systems places faith in a company to forever maintain their software program. (Remember Adobe Systems PageMaker or FreeHand?)

Because XML is device-independent, it works on PCs, eReaders, tablets, or any other hardware equipped with a microprocessor. Because it is system-independent it will run on Microsoft Windows, the Mac OS, Android, BlackBerry, and most other operating systems.

^ HTML File (DataIsland.htm)

```
HTML
<HTML>
<HEAD>
  <TITLE>Untitled</TITLE>
  <STYLE>
    .catalog_genre_head {background-color:darkGreen;
font-size:24pt;color:white;font-family:Impact;}
ont-family:Impact;}
    .catalog_head {background-color:green;font-size:
18pt;color:white;font-family
    .catalog_row0 {background-color:lightGreen;}
    .catalog_row1 {background-color:white;}
    .catalog_row_end {background-color:darkGreen;}
  </STYLE>
</HEAD>
<BODY>
```

^ XML File (book_catalog.xml)

```
XML
<?xml version="1.0"?>
<catalog>
  <book id="bk101">
    <title>XML Developer's Guide</title>
    <author>Gambardella, Matthew</author>
    <genre>Computer</genre>
    <price>44.95</price>
    <publish_date>2000-10-01</publish_date>
    <description>An in-depth look at creating
    applications with XML.</description>
  </book>
  <book id="bk102">
    <title>Midnight Rain</title>
    <author>Ralls, Kim</author>
    <genre>Fantasy</genre>
    <price>5.95</price>
```

A side-by-side comparison of HTML (left) and XML (right). Notice the similarities.

Why XML Works for Publishing

- **XML helps manage content**
Using a tagging system, XML separates content into chunks which content management systems can control and manage.
- **XML facilitates content discoverability**
Fully marked-up and tagged content is easier for search engines to parse and find.
- **XML separates form and content**
Separating form from content at the outset is a chief advantage to the multi-format/multi-screen output challenge. XML works as a foundational language. Formatting styles consistent with specific reader devices is applied when content is published.
- **XML is the language of XML metadata**
Metadata is now vital to online ecommerce and title discoverability. ONIX is the complex language of metadata and is expressed as 100% well-formed XML.
- **XML allows for semantic enrichment**
The XML workflow allows publishers to apply semantic tags to content to enhance discovery.
- **XML future-proofs content**
Because XML is a fully-documented plain Unicode text markup, its translation will be straightforward and precise no matter what language(s) succeeds it.
- **XML enables accessible texts**
XML supports Web accessibility efforts to help keep this legal obligation current.

"Ebook output became more functional at a fraction of the cost of multiple conversions. We began getting richer output at a lower cost."
~Stephen Driver, Rowman & Littlefield Publishing

XML Challenges

XML adoption is not without its challenges. They can include:

- **Unwieldy growth**
Anyone can define new tag sets within XML that can extend the markup to ever widening sets of data.
- **Jargon confusion**
Book production personnel will need to master a new vocabulary to master the technology.
- **Resistance to XML tools**
Authors, editors and designers may balk at the technical requirements and what they perceive as template limitations.

"If editors were introduced to this layer of complexity they'd start drinking at their desks!"

~Michael Tamblyn, Kobo

Tips for Overcoming XML Challenges

1. Incorporate Word into the XML workflow.

Allow authors to tag their manuscripts using macros within Microsoft Word. If non-technical authors still find this too difficult, apply tags at the publisher level before or after editing.

2. Recognize which books demand a designer's touch.

Not all books are suited to a templated design (aka XML). The production team needs to determine the difference.

3. A La Carte XML

Customize the implementation. Use only the parts of XML required to achieve your multi-format/multi-screen goals.

4. Standardize

The more standard and regular you can make your layouts the more likely you'll be successful with an XML-upfront process. Everything you put into the XML code to detail the product's look can be a compromise. The more processing instructions and attributes specific to a particular layout that you build into the XML, the less likely you'll be able to reuse that XML for other titles. It's easy to get to the point where you're handcrafting your XML to just one specific layout.

"We manipulated XML into what we wanted to do. We didn't feel like XML manipulated us."

~Stephen Driver, Rowman & Littlefield Publishing

Addressing the Human Factor

The human factor of an XML workflow is perhaps the most challenging piece of its implementation. Two common staffing issues can arise:

- **Training**
Without the necessary skills and training, an XML effort is doomed. Training is often overlooked because managers see it as expensive and a significant hit on short-term productivity. While this may be true, cost and productivity challenges will magnify without properly trained staff.
- **The cost to job security**
Used correctly, an XML workflow means that more books will get through production faster and cheaper. Unfortunately, staff reduction is sometimes an outcome of the introduction of new technology.

An XML Summary

1. The complexity of XML is offset by the tremendous value it offers.
2. Start small with XML. Implementation doesn't have to be "all the way."
3. Even if you start small, use an expert outside resource such as a consultant or technology vendor who has the benefit of experience.
4. Getting content onto multiple eReading devices is an absolute necessity for profiting in publishing today. At this time, XML offers the most workable and structured approach.

"With XML in place, output is up by 30-40% and staff morale is higher. People feel they have direct control over their work."
~Stephen Driver, Rowman & Littlefield Publishing

For Further Reference

Watch the 60-minute webinar with Stephen Driver and Thad McIlroy that this document summarizes:
[A Roadmap to Efficiently Producing Multi-Format/Multi-Screen eBooks](#)

Introductions to XML

1. [Introduction to XML](#): This IBM tutorial from 2002 is easy-going.
2. [A Gentle Introduction to XML](#): If you say so.
3. [An Introduction to XML Basics](#): Steven Holzner's 2003 guide on the Peachpit Press site.
4. [A Designer's Guide to Adobe InDesign and XML](#): This 2008 book by James Maivald and Cathy Palmer is extremely well written. It's specifically aimed at designers, which is to say that it's written in a language that a designer can (at least potentially) understand. So it's as good as it gets, but we think you'll still find it challenging. This book is the acid test of whether XML can be made safe for families with small children.
5. O'Reilly offers a superb set of XML learning and more advanced resources on its [O'Reilly XML site](#), along with a [\\$400 course](#): "XML Training for the Beginning or Intermediate Student."
6. There are a variety of XML videos online. [YouTube](#) has a large collection; [Vimeo](#) a more select set.

XML as a Failed State

1. [XML Can Go to H***](#): One Designer's Experience with the "Future of Publishing" (2004): According to Susan Glinert, who bears XML battle scars, the future is not bright.
2. [The Truth about XML](#) (2003): Systems powered by XML might someday prove to be the standard for information sharing between businesses, but not in the near future.
3. [What will it take to get \(end user\) XML editors that people will use?](#) (2011): Norman Walsh is a prominent proponent of XML, but here looks at some of the challenges surrounding the tools and user expectations.
4. [XML Fever](#) (2008): An in-depth piece, "this article is about the lessons gleaned from learning XML, from teaching XML, from dealing with overly optimistic assumptions about XML's powers, and from helping XML users in the real world recover from these misconceptions."

Beyond these you'll find a popular LinkedIn XML group, [XML-in-Practice](#). It's managed by IDEAlliance (<http://www.idealliance.org/about>) which plays an important role in standards-setting in the publishing industries. The W3C [XML section](#) has pointers to great resources, including XML conferences, books and training courses.

About the Webinar Presenters



Thad McIlroy is an electronic publishing analyst and author based in San Francisco and Vancouver, BC. His site, [The Future of Publishing](#), is the most in-depth on the industry. He has authored a dozen books and over 250 articles on publishing. McIlroy is on the editorial board of the journal [Learned Publishing](#) and the Canadian literary journal, [Geist](#). He is the co-author of the newly released [Metadata Handbook](#).



Stephen Driver is Vice President, Production Services for The Rowman & Littlefield Publishing Group, Inc., one of the largest and fastest-growing independent publishers and distributors in North America. The company publishes over 1,500 academic, reference, and general interest books each year under a dozen imprints in virtually all fields in the humanities and social sciences. Its backlist contains over 10,000 titles.

About Aptara

Aptara enriches content for capitalizing on all digital mediums. Providing full content lifecycle production – from content creation and design, to new media enhancements, content technology solutions, and production for all mobile devices and platforms – Aptara develops innovative digital products that deliver content how, when, and where recipients want it, while providing publishers renewed agility and revenue opportunities. Serving the 10 largest publishers in the world, Aptara is a US-based company founded in 1988 around its unique publishing technical expertise. Our over 5,500 professionals are deployed globally to help content providers create, enrich, publish, monetize, and manage their assets in the digital age.

For more information: <http://www.aptaracorp.com>



3110 Fairview Park Drive, Suite 900
Falls Church, VA 22042
+1-703-352-0001

aptaracorp.com

