# What to Expect When You're Converting

Source Content Preparation Best Practices

by Jean Kaplansky
Solutions Architect, Aptara

August 2012
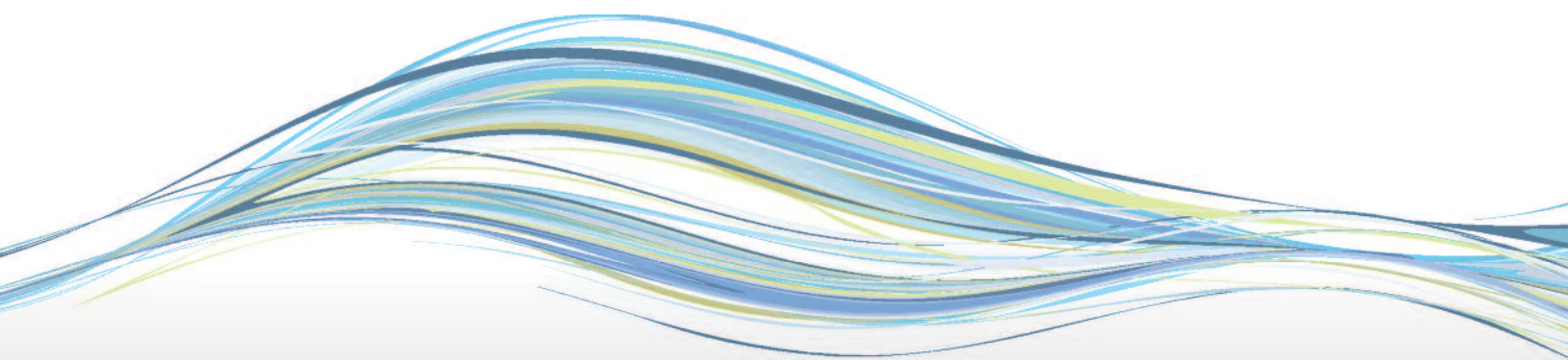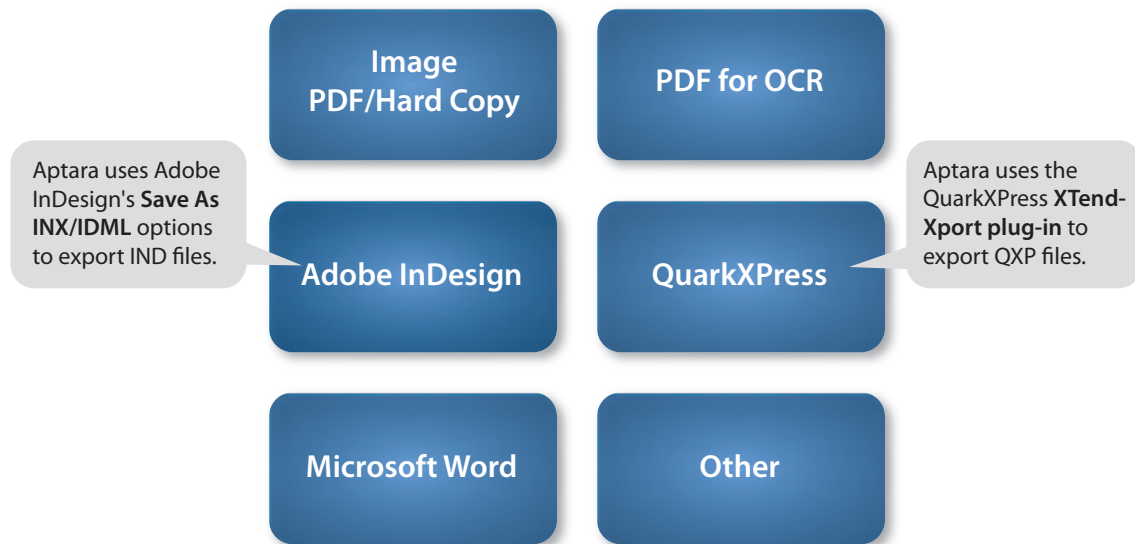
Table of Contents
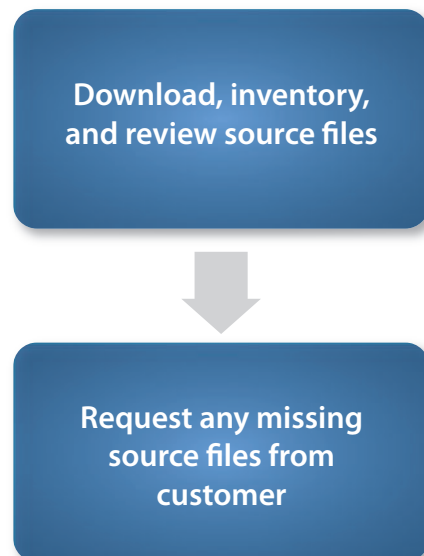
Starting from Scratch: A Brief Overview of the Content Conversion Process
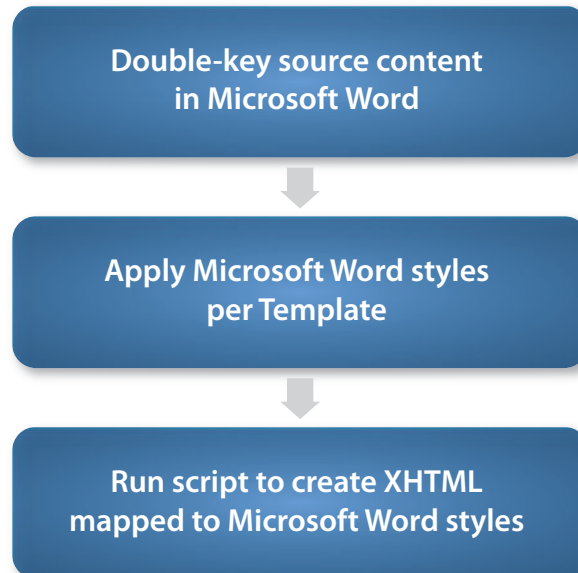
Step 1: Choose your source files for conversion.

Image PDF/Hard Copy	PDF for OCR

Aptara uses Adobe InDesign's **Save As INX/IDML** options to export IND files.

Adobe InDesign	QuarkXPress

Aptara uses the QuarkXPress **XTend-Xport plug-in** to export QXP files.

Microsoft Word	Other
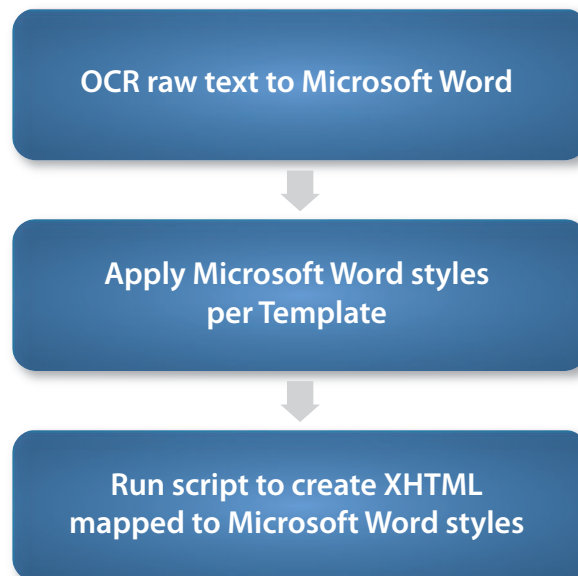
Step 2: Deliver your source files to your friendly Conversion Partner, who will then:

Download, inventory, and review source files

↓

Request any missing source files from customer
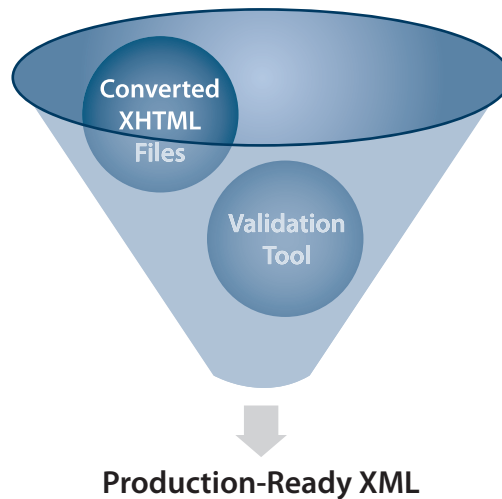
Step 3 – *Option 1***:** If your source files are image-based PDFs or hard copy, your Conversion Partner will:

**Double-key source content
in Microsoft Word**

**Apply Microsoft Word styles
per Template**

**Run script to create XHTML
mapped to Microsoft Word styles**

Step 3 – *Option 2*: If your source files are text-based PDFs, your Conversion Partner will:

OCR raw text to Microsoft Word

**Apply Microsoft Word styles
per Template**

**Run script to create XHTML
mapped to Microsoft Word styles**
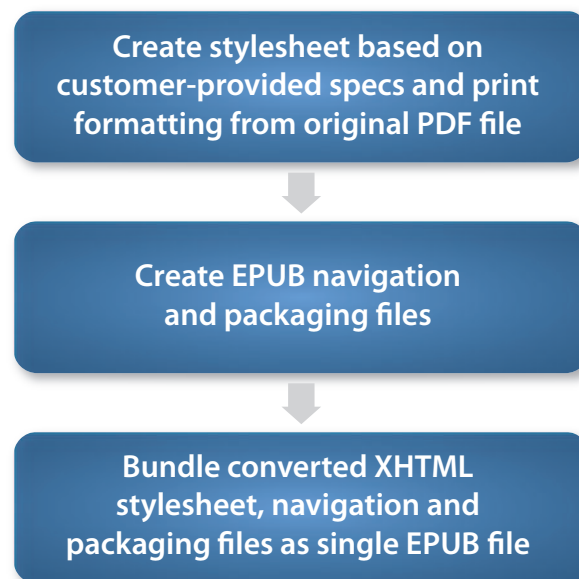
Step 4: Your Conversion Partner validates the converted XHTML files:

Converted
XHTML
Files

Validation
Tool

Production-Ready XML
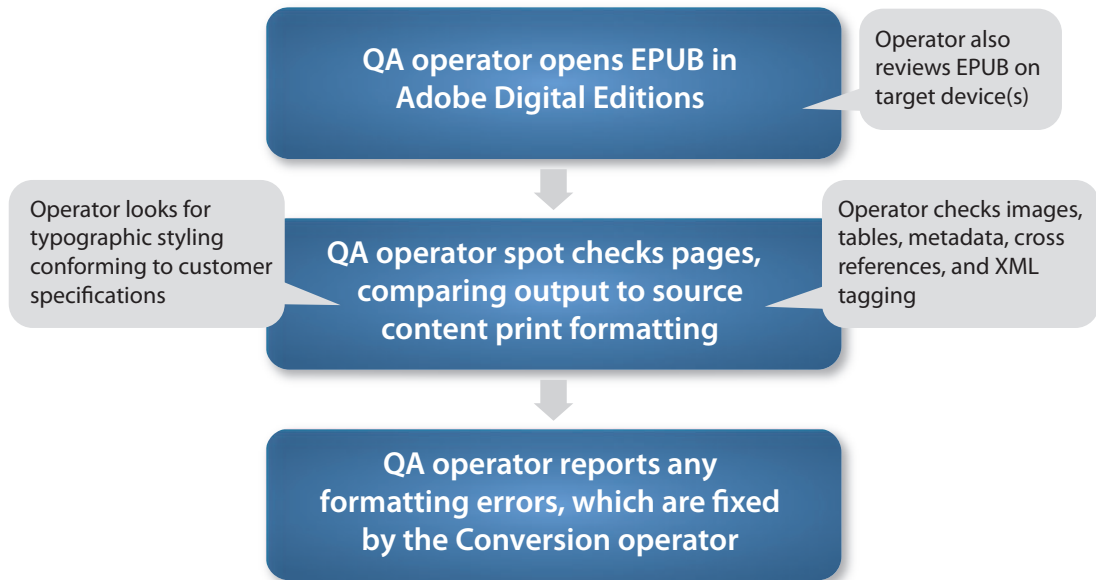
## Creating eBook EPUB Format Files from Production Ready XHTML
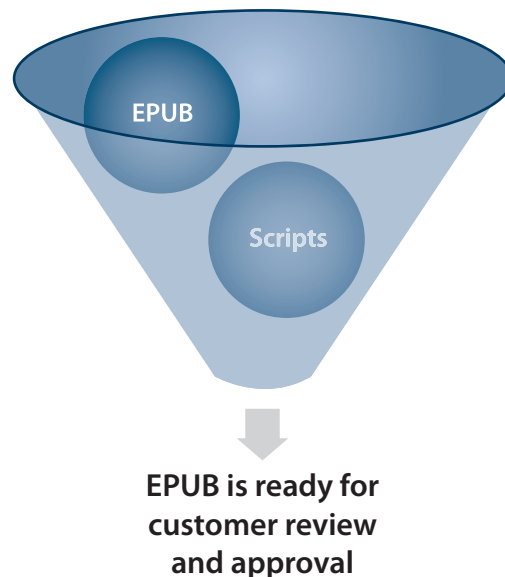
Step 1: Your Conversion Partner processes the valid XHTML files and creates additional files required to assemble the final EPUB.

**Create stylesheet based on customer-provided specs and print formatting from original PDF file**

Create EPUB navigation and packaging files

Bundle converted XHTML stylesheet, navigation and packaging files as single EPUB file
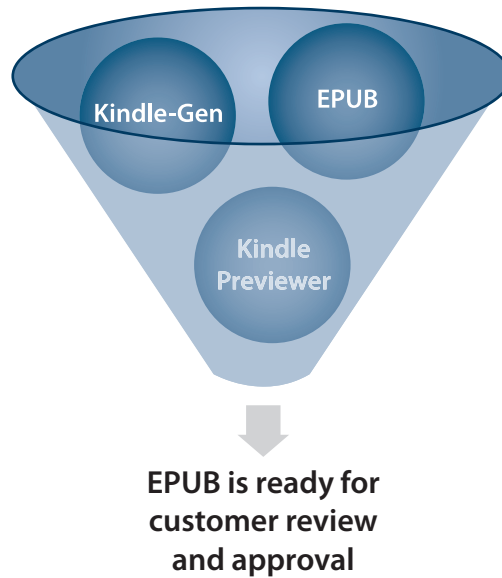
Step 2: Conversion Partner Quality Assurance (QA) staff reviews newly created EPUB files.

QA operator opens EPUB in Adobe Digital Editions

Operator also reviews EPUB on target device(s)

Operator looks for typographic styling conforming to customer specifications

QA operator spot checks pages, comparing output to source content print formatting

Operator checks images, tables, metadata, cross references, and XML tagging

QA operator reports any formatting errors, which are fixed by the Conversion operator

Step 3: QA moves corrected EPUB files through automated conversion checks.

EPUB

Scripts

EPUB is ready for customer review and approval

(Optional) **Step 3b:** Conversion Partner will use Amazon Kindle tools to create .mobi files corresponding to the internally approved EPUB.

EPUB is ready for customer review and approval
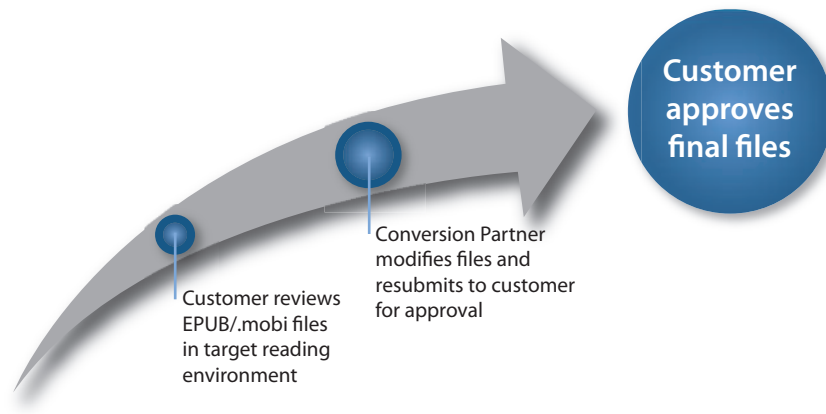
**Step 4**: QA staff spot checks individual EPUB pages on target reading device.

Step 5: Customer reports any issues to Conversion Partner.
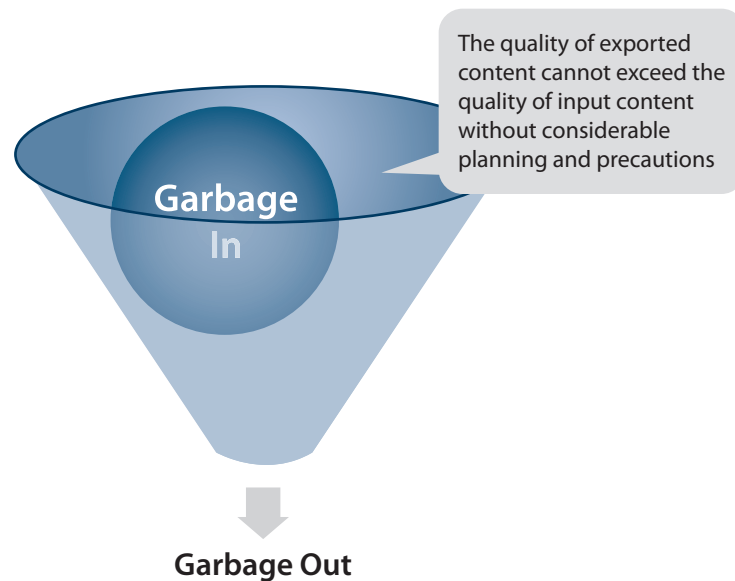
Step 6: Conversion Partner modifies files according to customer feedback.

Step 7: Customer approves final EPUB and/or .mobi files.

Customer approves final files

Conversion Partner modifies files and resubmits to customer for approval

Customer reviews EPUB/.mobi files in target reading environment

What Publishers Can Do to Facilitate Content Conversion

First Rule of Working with Digital Content

Garbage In

The quality of exported content cannot exceed the quality of input content without considerable planning and precautions

Garbage Out

2nd, 3rd, 4th and 5th Rules of Working with Digital Content

Establish Business Rules	**Know Your Content**
Have a Plan	**Be Consistent**

Notes about the 2nd, 3rd, 4th and 5th rules of working with digital content:

- Create a content catalog or inventory in order to learn your content inside out. Know what your source materials are, the file formats, the digital assets, etc.

- Learn about and apply required permissions metadata (this may be relevant to completing your project).

- Further sort your content by complexity (Easy, Medium, Hard).

- What is the quality of the source content? (This is especially pertinent for hard copy or image-based PDFs.)

- Is the copyediting correct?

Tips on 'knowing your content':

- What is the smallest reusable chunk/unit?

- Look at a widespread variety of content samples.

- Always involve editorial staff—they are your content "tour" guide.

- The sooner you start, the longer it takes…

Best Practices for Preparing Source Content for Conversion

If your source content is application-based (e.g., Microsoft Word, Adobe InDesign, QuarkXPress):

- Build and use templates. It's all about planning, business rules, structured templates, and making sure everything is done per the plan.

- Use application styles features.

- Use a consistent approach for applying title styles.

- Decide how to handle complex content.
  Complex content can include all of the following: math, chemistry, poetry, block citations, sidebars, annotations, arrows and callouts (especially in graphics), mixed content or accounting tables, code examples, multipart figures, position-based references (e.g., "see above", "see below").

- Pay attention to metadata.

- Build structure into the source content using named headings, parts, chapters, etc.

- Use application footnote and reference tools correctly.

Always verify:

- Reading order is properly configured.

- Each footnote or reference callout has a corresponding footnote or reference.

- Running header and footer text sources.

- Page numbers are properly configured.

- All Index entries have corresponding targets and working hyperlinks.

Note on Graphics: It is not a good idea to include text in graphics. Text in graphics is inaccessible. If you must include text in your graphics, make sure that all figure files are included for conversion (annotations, text, the image, or table itself). If you must implement a layered figure approach, always provide alternative text to display in accessible applications such as content readers (e.g., JAWS).

If your source content is already XML (or HTML):

- Choose a DTD/Schema and plan your markup hierarchy in advance of creating your content.

- Make sure your markup plan is based on business and content strategy rules.

- Validate, validate, and validate again.

Hooray if your content is already in XML! There are still some things to do to ensure that your XML is in tip top shape before sending it to your conversion partner. For example, you may want to consider another validation step using Schematron to ensure conformance to content business rules not directly enforceable by the DTD.

You may want to develop business rules based on title, or series, or discipline. You will also want to be familiar with how your particular valid XML should be mapped to XHTML.

If you are already creating source content in XML with a DTD or Schema, you should have based your application of XML upon a content strategy that defines business rules and points of integration required across production systems—authoring, storage, assets, rights, final production.

Best Practices for Working with a Conversion Partner

- **Be specific.** Don't leave guesswork for your Conversion Partner. Guesswork will slow the process. Most Conversion Partners will not make assumptions or make what you think may be logical conclusions. Operators are trained to follow specifications and directions implicitly and literally and to assume that the customer is always right—even if something is very obviously not right about the content. Do not rely on ambiguous specifications.

- **Beware of cultural communication differences if you are working with an offshore Conversion Partner.** Instead of asking for general opinions or advice, ask people for specific thoughts.

- **Build trust.** Get to know the people you are working with.

About Aptara

Aptara helps content publishers harness the rapid emergence of new media for a competitive advantage. For more than 20 years, our content development capabilities and technology innovations have helped publishers move to world-class digital content production for highly efficient, multi-channel publishing. Taking source content from any format and transforming it for distribution through any medium—from eReaders and smart phones, to tablets, PC's and print—Aptara enables leading global enterprises to unlock new top-line revenue growth in an evolving digital- and mobile-centric content marketplace.

aptara®

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

aptaracorp.com