

# EPUB: Pulling it all Together



Conference for Software User Assistance 2012 - Memphis, TN

*14 March 2012 - Scott Prentice, Leximation, Inc.*

# Introduction

— [ Scott Prentice, President of Leximation, Inc.

— [ Specializing in FrameMaker plugin development as well as structured FrameMaker conversions, consulting, and development. FrameMaker user/developer since 1991.

— [ Developed DITA-FMx, a FrameMaker plugin for efficient DITA authoring and publishing.

— [ Developer of custom Help systems and creative/functional web applications.

# EPUB basics

- [ EPUB (not ePub or ePUB) is an ebook file format that can be read on almost every device and platform
- [ Requires a "reader" application or dedicated reader device
- [ A collection of XHTML, XML, CSS, and media files wrapped up in a "zip" archive
- [ Best way to learn about the EPUB format is to open your EPUB and review the contents

# Popular reader applications

— [ iBooks on iPad/iPhone/iPod

— [ Aldiko on Android devices

— [ EPUBReader plugin for Firefox

— [ Adobe Digital Editions on Windows or Mac OS

— [ Kindle and Nook desktop and mobile apps

# Dedicated reader devices

— [ Amazon Kindle

— [ Barnes & Noble Nook

— [ Kobo eReaders

— [ Sony eReaders

— [ Also .. Asus, Azbooka, Bookeen, Egriver, enTourage, Fujitsu, HanLin, Hanvon, iRiver, NUUTBook, Onyx, PAGEone, Pixelar, Pocketbook, Sagem, Spring Design, Viewsonic, WISEreader, ...

# Structure of an EPUB file

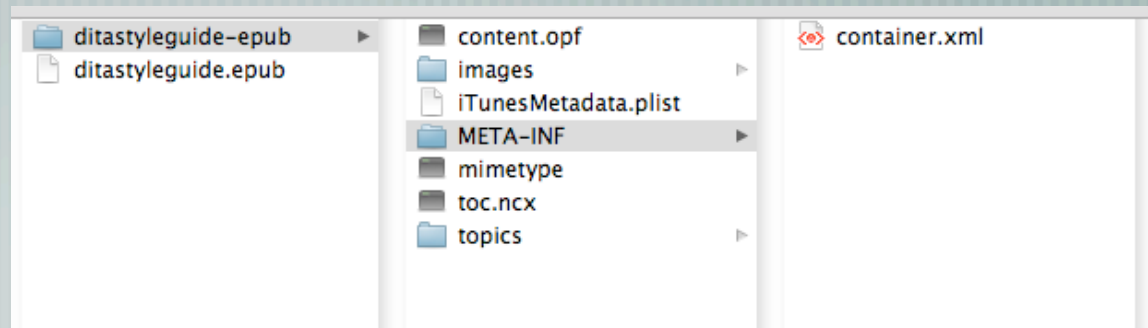
mimetype file (at root) contains "application/epub+zip"

META-INF/container.xml points to the OPF file

OPF file contains metadata, manifest, spine, and guide; points to an NCX file

NCX file is the TOC

XHTML, CSS, media, and image files are referenced by the OPF file



# Creating an EPUB

- [ Create content in an authoring tool then export to EPUB
- [ Convert to EPUB from existing content (PDF, HTML, DITA, etc.)
- [ People in techcomm will likely convert from existing content
- [ Each tool has its methods for authoring/conversion
- [ If possible, use named styles since these will typically map to CSS selectors in the XHTML

# Popular EPUB authoring tools

— [ Adobe InDesign, Apple Pages, Sigil

— [ Also .. Adobe RoboHelp, MadCap Flare, Microsoft Word, OpenOffice

— [ And others (see Resources page on [www.epubtest.com](http://www.epubtest.com))

— [ What about Apple iBooks Author? Nope, it doesn't create EPUBs



# Popular EPUB conversion tools

— [ Calibre

— [ DITA for Publishers (DITA-OT plugin)

— [ WebWorks Publisher

— [ PDF to EPUB converters

— [ Various utilities and plugins

(see Resources page on [www.epubtest.com](http://www.epubtest.com))

# EPUB2 or EPUB3?

— [ What's the difference?

— [ Tool/reader support for EPUB3 is lacking

— [ For now use EPUB2 unless you know that your target device/application supports EPUB3

# Got my EPUB, what now?

- [ Regardless of creation method, your EPUB will probably need tweaking
- [ Most tools create a single content file, best to split them up
- [ Update the content files to use efficient HTML/CSS tagging
- [ Add or clean up the cover image and title page
- [ Sanity check elements and metadata in OPF and NCX files
- [ Package / validate .. fix / package / validate ..

# oXygen XML to the rescue!

— [ Open your EPUB file in oXygen XML Editor to edit, revise, and save the file

— [ Uses the EpubCheck validation engine to ensure your changes are valid

— [ Diff two EPUB files to compare and merge their content

— [ Provides predefined templates to create a new EPUB file

— [ Puts it all back in to the EPUB file .. no muss no fuss

# An open source alternative?

— [ Sigil is an open source EPUB editor designed specifically for EPUB authoring

— [ Open compiled EPUB and edit/modify contents

— [ Nice TOC manager

— [ Reasonable WYSIWYG/code editor

— [ Some quirks, but a handy tool

— [ Validates and re-zips

# You can do it the hard way

— [ Change the ".epub" file extension to ".zip" and extract the contents with your favorite archive extraction tool

— [ Edit the files in a text editor

— [ Zip it back up with the proper tool

— [ Validate with the EpubCheck utility

# Split up content files

- [ If your conversion or generation process creates a single content file, it's best to split it up
- [ First heading in each file will always start a new page
- [ Any content or images that should start on a new page should be in its own file
- [ Too many files are usually better than too few

# Clean up HTML/CSS tagging

— [ Possibly the most time consuming part of an EPUB cleanup

— [ To be usable on the widest array of readers, aim for simplicity in your HTML coding

— [ Think of the early days of HTML coding, and you'll be all set  
<h1>, <h2>, <h3>, <p>, <b>, <i>, ...

— [ For styling, always use CSS classes

— [ Leverage “regular expressions” to make global changes



# HTML dos and don'ts

— [ Must be valid XHTML 1.1 - all tags must close

— [ Don't use @name anchors, use @id references

```
<p><a href="file.html#target">Link to Target</a></p>
```

```
...
```

```
<h2 id="target">Target Heading</h2>
```

— [ Don't use @font or other deprecated attributes

# Keep with next?

— [ Want to “encourage” tags to stick together?

— [ Headings should keep with the first paragraph and images should keep with their captions

**CSS:** `.nobreak { display:inline-block; margin:0; padding:0; }`

**HTML:** `<div class="nobreak">  
<h2>Some heading</h2>  
<p>The heading should always be above this paragraph and  
never alone at the bottom of a page.</p>  
</div>  
<p>Paragraph that follows the "nobreak" DIV.</p>`

# Code samples

— [ Use the `<pre>` tag carefully, it will often run off the page and the content will be lost

— [ Best to design code samples to wrap

— [ Apply a monospaced font

— [ Familiarize yourself with the CSS “white-space” options; ‘pre-wrap’ preserves whitespace but allows wrapping

# Tables

- [ Wide tables will often look really bad on small screens
- [ Some reader apps will lose extra columns
- [ More than 3 columns should be avoided
- [ If possible, try recasting as lists or multiple narrower tables
- [ When tables break over pages, they will often break in a bad location

# Cover image file

Put your cover image in its own file, perhaps "cover.html"

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
  "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Book Cover</title>
    <style type="text/css"> img { max-width:100% } </style>
  </head>
  <body>
    <div id="cover-image">
      
    </div>
  </body>
</html>
```

# Cover image size

— [ The size of the cover image may be dependent on how you plan to publish; some publishers have strict requirements

— [ Lulu.com requires 612 x 792 pixels

— [ Dimensions of around 600 x 800 should be fine

— [ JPG is recommended for photographs and PNG for other art

# Cover image OPF metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<package ... >
  <metadata>
    ...
    <meta name="cover" content="cover_img"/>
  </metadata>
  <manifest>
    ...
    <item id="cover_img" href="cover-image.jpg"
      media-type="image/jpeg"/>
    <item id="cover_html" href="cover.html"
      media-type="application/xhtml+xml"/>
  </manifest>
  <spine toc="ncx">
    <itemref idref="cover_html" linear="no"/>
    ...
  </spine>
</package>
```

# Sanity check OPF / NCX

- [ OPF/NCX files are the glue that holds your book together
- [ EPUB generation and conversion tools may not always do the right thing
- [ If you've made any modifications to the file structure, you'll need to reflect those changes in these files



# Update the OPF metadata

— [ metadata element contains publication-level metadata as defined by the Dublin Core Metadata Initiative and may contain meta elements for other publication-level metadata

```
<package ...>
  <metadata ...>
    <dc:title>The Great Story</dc:title>
    <dc:language>en</dc:language>
    <dc:identifier id="BookId" opf:scheme="ISBN">
      123456789X
    </dc:identifier>
    <dc:creator opf:role="aut">Joe Author</dc:creator>
    <dc:subject>History</dc:subject>
    <meta name="price" content="USD 19.99" />
    ...
  </metadata>
```

# Fix up the TOC (NCX)

- [ The NCX file defines the order and hierarchy of the TOC
- [ If you split up a large file or made any other structural changes, you'll need to update the navMap in the NCX
- [ The structure is fairly logical, nested navPoint elements define each TOC entry
- [ The navPoint/@playOrder attributes must be sequential and unique

# Partial NCX file

```
<ncx ...>
  <head>
    <meta name="dtb:uid" content="123456789X"/>
    <meta name="dtb:depth" content="2"/>
    <meta name="dtb:totalPageCount" content="0"/>
    <meta name="dtb:maxPageNumber" content="0"/>
  </head>
  <docTitle><text>The Great Story</text></docTitle>
  <navMap>
    ...
    <navPoint id="navPoint-12" playOrder="12">
      <navLabel><text>After the Fall</text></navLabel>
      <content src="afterfall.html" />
      <navPoint id="navPoint-13" playOrder="13">
        <navLabel><text>Evil lurks</text></navLabel>
        <content src="afterfall.html#evillurks" />
      </navPoint>
      <navPoint id="navPoint-14" playOrder="14">
        ...
```

# Packaging the EPUB

— [ Once all of the files “look” good, you’ll need to package (zip) and validate the EPUB

— [ If you’re using oXygen or Sigil, this will be done for you, otherwise you’ll need to do the packaging

— [ mimetype file must be first in the package and be uncompressed, remaining files follow

# Packaging the EPUB (2)

— [ On Mac or other UNIX systems, use the following commands  
(run from within EPUB project folder)

```
$ zip -0Xq thegreatstory.epub mimetype
```

```
$ zip -Xr9Dq thegreatstory.epub *
```

— [ On Windows systems, use the ePubPack utility  
[sourceforge.net/projects/epubpack/](http://sourceforge.net/projects/epubpack/)

# Validate your EPUB

— [ To ensure the EPUB meets the EPUB specification, use the EpubCheck utility - [code.google.com/p/epubcheck/](http://code.google.com/p/epubcheck/)

— [ Requires Java JRE 1.5 or later

```
$ java -jar /path/to/epubcheck.jar thegreatstory.epub
```

— [ Reports any violations

# Create Kindle MOBI file

— [ The Amazon Kindle doesn't read the standard EPUB file, there are a number of converters available, but it's best to use the official tool from Amazon

— [ Link on Resources page on [www.epubtest.com](http://www.epubtest.com)

```
$ /path/kindlegen thegreatstory.epub
```

— [ Creates a MOBI file from the EPUB

# Resources / questions

— [ IDPF EPUB 2 and 3 specifications - [www.idpf.org](http://www.idpf.org)

— [ Follow #eprdctn on Twitter

— [ Resources - [www.epubtest.com/resources.php](http://www.epubtest.com/resources.php)

— [ Questions?

— Scott Prentice <[scott AT leximation.com](mailto:scott@leximation.com)>

— Leximation, Inc. - [www.leximation.com](http://www.leximation.com)