Creating ePubs

Copyright © 2014 by Steve Litt
No warranty, use at your own risk

What We'll Cover

- What's an ePub, and why bother
- How's it built
  - Required files
  - Directory structure, required and optional
  - Splitting into Xhtml files
  - Inter-file links
  - Zipping the ePub
- Conversions
- Preview on making them programmatically

What is ePub

- Format especially for reading devices
- Flowing text
  - Linebreaks at read time, not creation time
- Directory structure zipped up

```
ePub Root
  ├META-INF
  │  └container.xml
  │
  ├mimetype
  ├opf file
  ├ncx file
  ├css file(s)
  ├xhtml
  │  └xhtml files
  └images
     └image files
```

ePub Paging

- Pages are meaningless with flowing text
- Each Xhtml file starts at top of device

It Sounded Like a Good Idea At the Time

- Convert your print book source, no human intervention
- Complex/big diagrams
- Convert from PDF

Why Not Use...

- Sigil is slow for authoring
- Calibre is a conversion utility, not for authoring
- MSWord must be authored just right to make a passable ePub
- LyX Xhtml conversion stinks for ePub
- LibreOffice? You gotta be kidding me!
- Conversion services make you wait and pay for every tiny edit
- If you try to convert from PDF, you deserve every bad thing life throws at you, and you'll get it.

Bluefish Plus Converter Program

- Makes high quality ePubs
- Not that difficult to make Xhtml to LaTeX
- For a text/tag editor, Bluefish is pretty fast to use.
  - Zen coding
  - Completion

Why Make an ePub Manually?

- So you'll learn how ePub really works, so that:
  - You can use Sigil or Calibre right
  - You can write your own ePub exporter
  - You can write your own exporter from your favorite authoring software to input to an existing ePub maker.
- So you learn the difference between good and bad ePub

Required ePub Directory Structure

- Epub Root
  - mimetype
  - META-INF
    - container.xml
- Other files can go anywhere in ePub root tree
  - container.opf (OPF file)
  - toc.ncx (NCX file)
  - Xhtml files
  - CSS files
  - Graphics files

ePub Zipping

- An ePub file is just the directory tree zipped.
- The mimetype file must NOT be compressed.
- mimetype MUST be first file.
- The rest of the directory MAY be compressed.

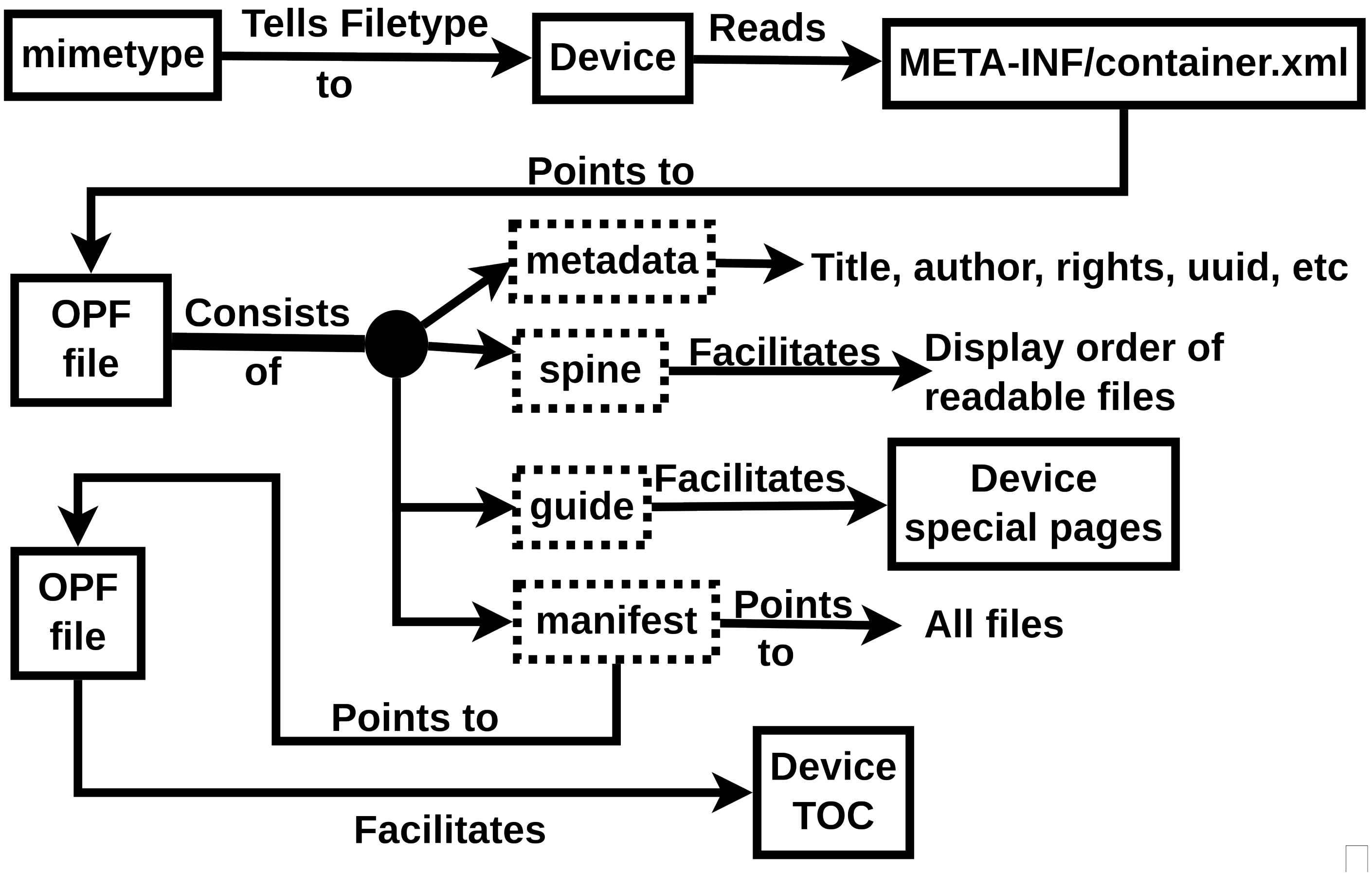
```
zip -0Xq /tmp/hello.epub mimetype
zip -Xr9Dq /tmp/hello.epub *

ebook-viewer /tmp/hello.epub
```

- -0=no compress
- -X=no extra attribs
- -q=quiet

- -r=recurse paths
- -9=max compression
- -D=don't create empty directories

★ -D is a little more complicated than stated.

ePub Startup



content.opf

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<package xmlns="http://www.idpf.org/2007/opf" unique-identifier="BookID" version="2.0">
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:opf="http://www.idpf.org/2007/opf">
<dc:title>My First Epub</dc:title>
<dc:identifier id="BookID" opf:scheme="CustomID">HelloWorld</dc:identifier>
</metadata>

<manifest>
<item href="xhtml/cover.xhtml" id="cover_page" media-type="application/xhtml+xml"/>
<item href="xhtml/titlepage.xhtml" id="title_page" media-type="application/xhtml+xml"/>
<item href="xhtml/toc.xhtml"    id="toc_page" media-type="application/xhtml+xml"/>
<item href="xhtml/chap01.xhtml" id="chap01_page" media-type="application/xhtml+xml"/>
<item href="xhtml/chap02.xhtml" id="chap02_page" media-type="application/xhtml+xml"/>
<item href="mystyles.css" id="css" media-type="text/css"/>
<item href="toc.ncx" id="ncx" media-type="application/x-dtbncx+xml"/>
<item href="images/epubrocks.svg" id="epubrocks.svg" media-type="image/svg"/>
<item href="images/cover.svg" id="cover.svg" media-type="image/svg"/>
</manifest>
```

content.opf (part 2)

```
<spine toc="ncx">
<itemref idref="cover_page" linear="yes"/>
<itemref idref="title_page" linear="yes"/>
<itemref idref="toc_page" linear="yes"/>
<itemref idref="chap01_page" linear="yes"/>
<itemref idref="chap02_page" linear="yes"/>
</spine>

<guide>
<reference href="xhtml/cover.xhtml" type="cover" title="Cover"/>
<reference href="xhtml/titlepage.xhtml" type="title-page" title="Title Page"/>
<reference href="xhtml/toc.xhtml" type="toc" title="Table of Contents"/>
</guide>

</package>
```

toc.ncx

```xml
<?xml version='1.0' encoding='utf-8'?>
<ncx xmlns="http://www.daisy.org/z3986/2005/ncx/" version="2005-1" xml:lang="eng">
  <head>
    <meta content="HelloWorld" name="dtb:uid"/>
    <meta content="1" name="dtb:depth"/>
    <meta content="Me, myself and I" name="dtb:generator"/>
    <meta content="0" name="dtb:totalPageCount"/>
    <meta content="0" name="dtb:maxPageNumber"/>
  </head>
  <docTitle>
    <text>My First Epub</text>
  </docTitle>
  <navMap>
    <navPoint id="chap01_page" playOrder="1">
      <navLabel>
        <text>Chapter 1: Overview of ePub</text>
      </navLabel>
      <content src="xhtml/chap01.xhtml#chaptertitle"/>
    </navPoint>
    <navPoint id="chap02_page" playOrder="2">
      <navLabel>
        <text>Chapter 2: Calibre</text>
      </navLabel>
      <content src="xhtml/chap02.xhtml#chaptertitle"/>
    </navPoint>
</ncx>
```

Sample Xhtml File

```
<?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" xml:lang="en">
<head>
<title>Lame, Moot Title, Chap2</title>
<meta http-equiv="content-type" content="application/xhtml+xml; charset=UTF-8"/>
<meta http-equiv="content-style-type" content="text/css"/>
<meta http-equiv="expires" content="0"/>
</head>
<body>
<h1 id="chaptertitle">Chapter Two: Calibre</h1>
<p>When it comes to converting between various eBook formats, and sometimes
even importing and exporting to some of these formats, Calibre is a powerful tool.
Calibre also comes with great utilities like ebook-viewer to view an ebook,
and epub-fix to fix and check an ebook, and find subtle errors.</p>

<a href=”calibre.explain.xhtml”><img src=”../images/calibre_screen.svg”></a>

<p>Be careful with epub-fix though: It changes the file rather than just
checking it. Use epub-fix on a copy, not on the original.</p>

</body></html>
```

Hello World Directory Structure

```
ePub Root
  ┃ META-INF
  ┃   └container.xml
  ┃
  ┣━mimetype
  ┣━opf file
  ┣━ncx file
  ┣━xhtml files
  ┣━image files
  ┗━css file(s)
```

Steve Litt's Recommendation

```
ePub Root
 ┣━META-INF
 ┃   ┗━container.xml
 ┃
 ┣━mimetype
 ┣━opf file
 ┣━ncx file
 ┣━css file(s)
 ┣━xhtml
 ┃   ┗━xhtml files
 ┗━images
     ┗━image files
```

Traditional Directory Structure

```
ePub Root
  ├─META-INF
  │   └─container.xml
  │
  ├─mimetype
  └─OEBPS
      ├─opf file
      ├─ncx file
      ├─xhtml
      │   └─xhtml files
      ├─images
      │   └─image files
      └─CSS
          └─css file(s)
```

Making ePubs Programmatically

- Insist on styles-based, hierarchical authoring
- Map <h1>-<h6> to Part, Chapter, Section, etc.
  - Map them in config file
- Split off chapters into individual Xhtml files
- Build OPF and NCX in memory as XML trees
- Use Python's lxml.etree
- I'll cover in detail in a later presentation

ePub Troubleshooting

- Tools:
  - ebook-viewer
  - epub-fix
  - file-roller
  - Litt's XML checker
- Techniques:
  - Really read the error message
  - Narrow to smallest working example
  - Use good troubleshooting procedure. See Troubleshooters.Com

Conversions

- ePub is viewable by the most devices
- Calibre
  - Converts to .mobi and other formats
- Kindlegen
  - Converts .mobi to Kindle file

For More Information

- http://www.troubleshooters.com/ebooktech/epub_demystify.htm
- http://www.troubleshooters.com (for general troubleshooting)

What We Covered

- What's an ePub, and why bother
- How's it built
  - Required files
  - Directory structure, required and optional
  - Splitting into Xhtml files
  - Inter-file links
  - Zipping the ePub
- Conversions
- Preview on making them programmatically