Hands on with EPUB (and DITA)

(workshop)

tcworld conference 2018 - Stuttgart, Germany

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.

- Consulting for custom Help systems, creative/functional web applications, and EPUB solutions.

Workshop tools

- Oxygen XML Editor – installer and trial license provided
- ADE ebook reader – installer provided
- Chrome+Readium plugin – installer not provided
- DITA-OT 2.5.4 – archive provided
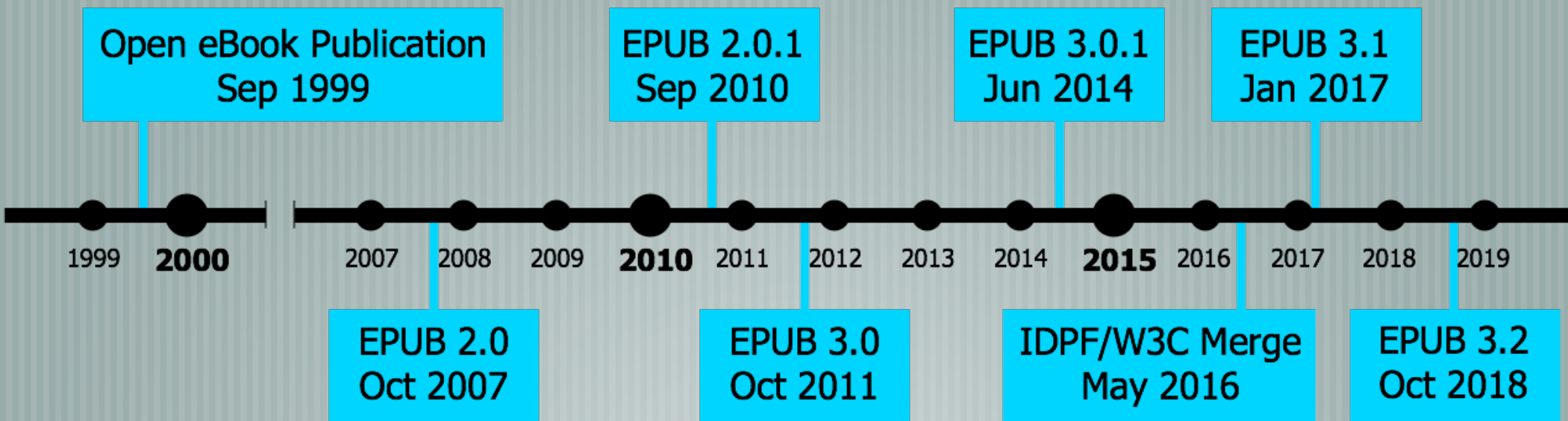- DITA for Publishers 1.0.0RC27 – archive provided
- Sample EPUBs – files provided

What is EPUB?

- EPUB: ebook file format for every device and platform

- Requires a "reader" application or dedicated device

- A collection of XHTML, XML, CSS, and media files wrapped in a "zip" archive

- Defines the content, navigation, and formatting of ebook
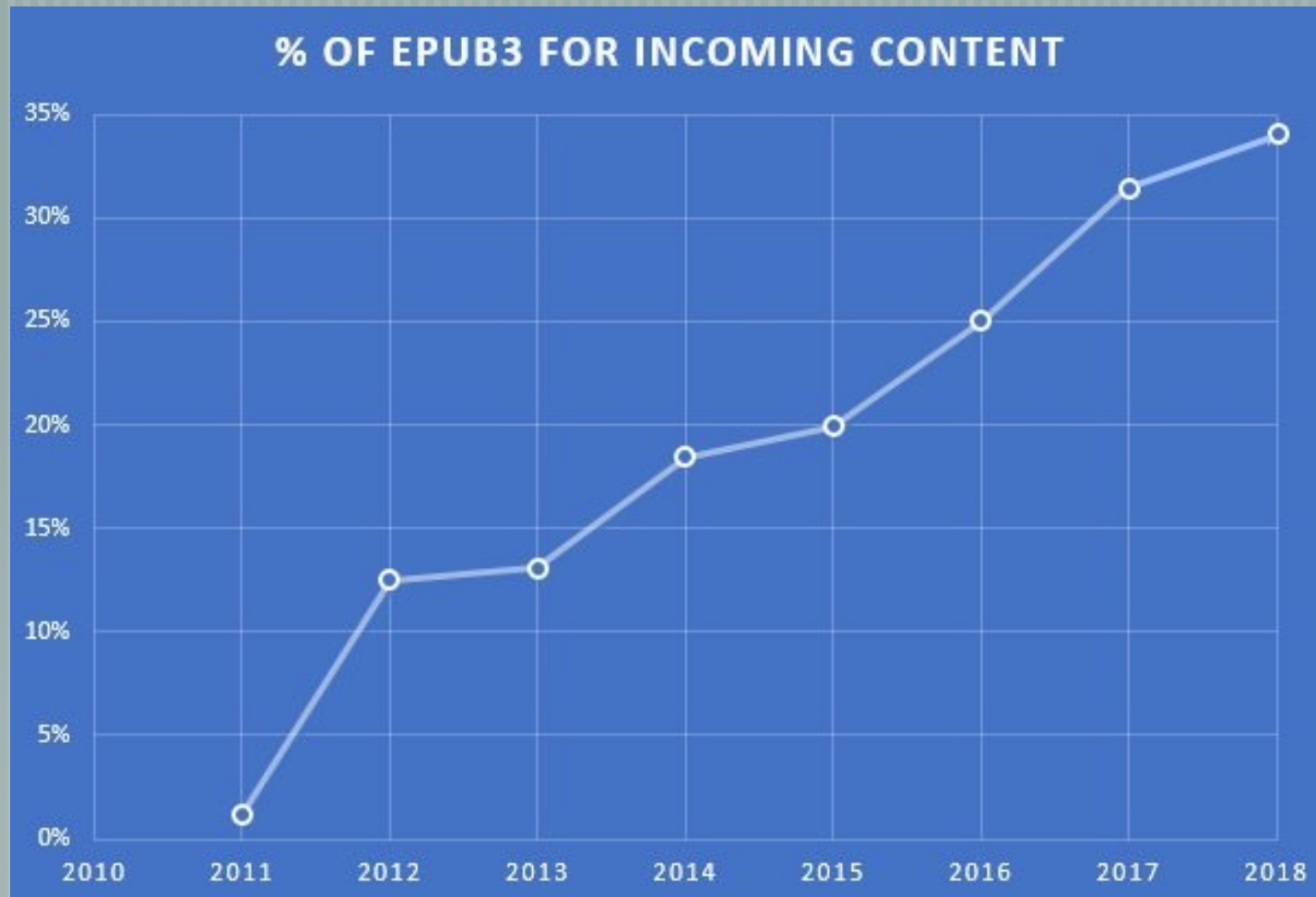
- "EPUB" is a registered trademark of the W3C (was IDPF)

[almost] 20 years of EPUB!

Open eBook Publication
Sep 1999

EPUB 2.0.1
Sep 2010

EPUB 3.0.1
Jun 2014

EPUB 3.1
Jan 2017

1999 **2000**   2007 2008 2009 **2010** 2011 2012 2013 2014 **2015** 2016 2017 2018 2019

EPUB 2.0
Oct 2007

EPUB 3.0
Oct 2011

IDPF/W3C Merge
May 2016

EPUB 3.2
Oct 2018

EPUB 2 or EPUB 3?

- EPUB 3 is current spec, but full support still weak

- 65% of new EPUBs are still EPUB 2 (but dropping!)

- EPUB 3 adds interactivity, multimedia, and much more

- Hand code to take full advantage of EPUB 3, no tools provide interactive features (yet)

- Some EPUB 3 features may limit the readers

- This workshop covers some of both

EPUB 3 at Kobo

% OF EPUB3 FOR INCOMING CONTENT

Posted on Twitter by Ben Dugas @bendu on Aug 16, 2017
data updated Nov 2, 2018

What about Kindle?

- Amazon uses proprietary ebook formats for Kindle readers

- MOBI (like EPUB 2) and KF8 (like EPUB 3)

- Download utility from amazon.com to convert EPUB to Kindle formats

- Mostly "works" but some situations may require extra effort

- This workshop will not cover Kindle conversion issues

EPUB 3 changes everything!

- Underlying technologies are familiar to techcomm: (X)HTML5 and CSS3

- EPUB 3 supports features you'll see on the web

- Supports most languages (text orientation and Unicode)

- MathML, SVG, audio, video

- Responsive design via media queries

- JavaScript support including libraries like jQuery

Page layout options

- Paged - typical ebook, default output for tools
    - Best for "linear" books
- Fixed layout (FXL) - like a PDF but in HTML
    - Graphic intensive books
- Vertical scrolling - per "topic" or entire book
    - References or online Help (EPUB Help!)

1 Yosemite National Park

Yosemite National Park is a United States National Park spanning eastern portions of Tuolumne, Mariposa and Madera counties in the central eastern portion of California, United States.

Figure 1-1: Yosemite Valley from Tunnel View

Location	Tuolumne, Mariposa, & Madera counties, California, USA
Nearest city	Mariposa, California
Coordinates	37°44′43″N 119°35′54″W
Area	761,268 acres (308,074 ha)
Established	October 1, 1890
Visitors	3,951,393 (in 2011)
Governing body	National Park Service
	UNESCO World Heritage Site

The park covers an area of 761,268 acres (3,080.74 km2) and reaches across the western slopes of the Sierra Nevada mountain chain. Over 3.7 million people visit Yosemite each year: most spend their time in the seven square miles (18 km2) of Yosemite Valley. Designated a World Heritage Site in 1984, Yosemite is internationally recognized for its

PDF

EPUB Help

iBooks - iOS

Readium+Chrome - Mac

EPUB updates since 3.0.1

- Resources may be loaded from remote locations

- HTML and SVG versions are now the "current" version

- CSS support removes EPUB-specific properties

- CSS position restriction removed

- EPUBCFI linking no longer "required" for readers

- NCX file (EPUB 2 ToC) to be removed in next major rev

- Guide element deprecated in favor of landmarks nav

Useful EPUB 3 readers

Desktop

- Chrome+Readium (macOS, Linux, Windows) .. WAIT!

- ADE (macOS, Windows, iOS, Android)

- AZARDI (macOS, Linux, Windows) .. maybe not?

- Readium Desktop macOS, Linux, Windows) .. SOON!

Mobile

- Apple iBooks (iOS)

- Gitden Reader (Android, iOS)

Authoring/publishing tools

- Adobe FrameMaker

- Adobe RoboHelp

- MadCap Flare

- MadCap DocToHelp

- oXygen XML Editor

- DITA-OT + DITA for Publishers

- Adobe InDesign

- Calibre

- Chapters

- and .. epubcheck!

EPUBs created from tools

- Tools only make "paged" EPUBs

- No interactivity

- Bloated HTML

- Bad CSS tagging

- Typically "invalid"

- Cleanup is almost always required!

Working with Oxygen

- Other tools can be used, but Oxygen makes it so easy!

- Open EPUB to view/edit content

- Changes are saved back to EPUB

- Validation errors are easy to locate and fix

- Easy to add/edit/delete files

- Perform global (regex-based) search and replace!

Review EPUB structure (demo)

- "ZIP" package

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

- OPF file contains metadata and manifest of all files

- Navigation files (EPUB 3: XHTML; EPUB 2: NCX)

- XHTML content files (with .xhtml extensions)

- Additional styling, scripting, resource files

Possible enhancements

- General cleanup of markup and CSS

- Update images and tables for small screen

- Integrate responsive CSS (media queries) techniques

- Change scrolling/pagination depending on content type

- Add JS library for formatting or scripting support

- Add video and audio (read-aloud?)

EPUB cleanup

Regardless of the tool, some cleanup will be needed ..

- Use Oxygen XML Editor

- Validate and correct errors

- Strip all @style attributes

- Add missing @class attributes so CSS works properly

- Delete unused JavaScript and other resources

Search/Replace

To add script (or CSS) to all files ..

Search: `</head>`

Replace: `<script src="js/cool.js"></script>\n</head>`

To delete all style attributes ..

Search: `style="[^"]+"` (regex)

Replace: (nothing)

Regex is your friend!

Demo — Basic cleanup

- Review EPUBs created from tools

- Fix validation errors

- Fix minor usability issues (typically tables and images)

- Add minor enhancements

- Review some table options

JavaScript in EPUB

- Apply formatting like code syntax highlighting

- Add/fix markup that doesn't exist in the HTML

- Modify markup or structure based on screen size, orientation, or user input

- Generate markup or graphics (SVG) on the fly from remote data

Useful JS libraries

- highlight.js – code styling, 185 languages + 89 styles

- jQuery – powerful scripting support

- fuse.js – lightweight fuzzy-search library

- moment.js – date/time support library

- Raphaël – vector graphic (SVG) support library

Integrating JS into EPUB

Add new files to EPUB; reference JS from HTML files

Add \<item> references to all new files in OPF manifest (be sure to use correct mime type) ..

```
<item id="highlight-js" href="js/highlightjs.js"
    media-type="text/javascript"/>
```

Add scripted property to each scripted HTML file's \<item> ..

```
<item id="topic1" href="topics/topic1.xhtml"
    properties="scripted"
    media-type="application/xhtml+xml"/>
```

Scripting issues

- Just because it works in a browser, doesn't mean it'll work in an EPUB (reality check)

- Don't go overboard

- Cookies definitely don't work

- Setting and retrieving data to/from external resources should work (someday)

Responsive EPUB

Media query loads alternate CSS based on media type and related expressions:

media-type and | not | only (features)

media-type: all, print, screen, speech, ...

feature: width, height, device-width, device-height, orientation, aspect-ratio, device-aspect-ratio, color, color-index, monochrome, resolution, scan, grid, ...

Responsive EPUB

Declare in HTML:

```
<link rel="stylesheet"
    media="screen and (max-width: 400px)"

    href="css/epub-help-small.css" />
```

Declare in CSS:

```
@media screen and (max-width: 400px)

{ ... }
```

EPUB Help

- Looking for an alternative to HTML Help?

- "EPUB Help" provides all of the features of Help

- Single-file Help deliverable on all platforms/devices

- Vertical scrolling "topics" (instead of paged)

- Context sensitivity (??)

- Hoping Readium Desktop will be "the" reader!

Demo — Advanced options

- Add highlight.js support to existing EPUB

- Test options with adding responsive techniques (media queries)

- Discussion/demo of EPUB Help prototype

- Explore options for adding charts and vector graphics with Raphaël library (time permitting)

EPUBs from DITA

- DITA-OT + DITA for Publishers is a great way to build EPUBs .. if you've got DITA content.

- Sensible file/folder structure in EPUB

- Easy to customize

- Builds easily automated

- "Free" (it's open source)

Installing OT and D4P

- DITA-OT – github.com/dita-ot/dita-ot/releases/tag/2.5.4

- D4P – github.com/dita4publishers/dita4publishers/releases/tag/1.0.0RC27

- Add path to dita-ot-2.5.4/bin to PATH system variable

```
$ cd OT-2.5.4
$ ./bin/ant -f integrator.xml
```

Generate output

- In a command shell, cd to DITA map folder, then ..

  $ dita –input=FILENAME.ditamap –output=OUTPUT-FOLDER –format=epub

- Additional options are available:

  - add a custom CSS to each file,

  - add custom JavaScript to each file,

  - generate a back-of-the-book index, and more!

Demo — EPUBs from DITA!

Generate a "paged" EPUB

Review command line options

Convert a paged EPUB into "EPUB Help"

Add full text search to your EPUB

Wrap-up

- EPUB really is like a "website in a box"!

- Decide what your customers need in their documentation and make it happen

- Don't be limited by the current tools

- Talk to tool vendors about what you want

JS Resources

- fuse.js: lightweight fuzzy-search library — fusejs.io

- highlight.js: syntax highlighting for the Web — highlightjs.org

- moment.js: parse, validate, manipulate, and display dates and times in JavaScript — momentjs.com

- Raphaël: simplify your work with vector graphics (SVG) — dmitrybaranovskiy.github.io/raphael

- jQuery: makes things like HTML document traversal and manipulation, event handling, animation, and Ajax much simpler — jquery.com

Resources

- EPUB 3.2 specification — w3c.github.io/publ-epub-revision/epub32/spec/epub-spec.html

- EPUB 3.1 specification — www.w3.org/Submission/epub31/

- HTML 5 specification — www.w3.org/TR/html5/

- CSS 3 specification — www.w3.org/TR/CSS/

- CSS 3 module status — www.css3.info/modules/

- Readuim Desktop — www.edrlab.org/readium/readium-2/

More resources

- On the Slow Adoption of EPUB 3 – epubsecrets.com/on-the-slow-adoption-of-epub-3.php

- Responsive Ebook Design – medium.com/@sandersk/responsive-ebook-design-a-primer-8bba01328219

- Follow #eprdctn on Twitter

- Scott Prentice <scott AT leximation.com> – www.leximation.com

Feedback

Your opinion is important!

Please tell us what you thought of the lecture. We look forward to your feedback via smartphone or tablet.

**Scan the QR code
or visit the URL:**

http://ta20.honestly.de

The feedback tool will be available even after the conference!