

EPUB 3: **Not your father's EPUB**

Bill Kasdorf

Vice President, Apex Content Solutions

General Editor, *The Columbia Guide to Digital Publishing*

Metadata Subgroup Lead, IDPF EPUB 3 Working Group

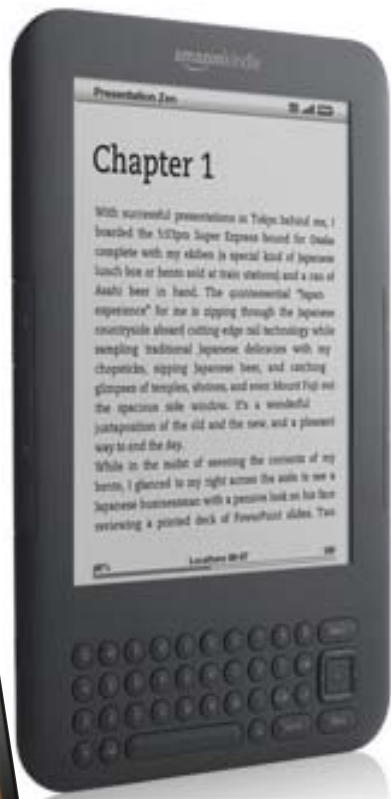


Trying to get ahead of the developers.



It's not just for books.

**Our how-you-want-it,
where-you-want-it,
when-you-want-it,
what's-the-next-new-thing
environment.**



**EPUB 3 has to work
in lots of different
environments—
particularly OFFLINE.**



Devices are
for purposes of
illustration and do
not imply EPUB 3
compatibility!



**EPUB is a package
masquerading as
a file format.**

What's In an EPUB 3?

An EPUB 3 can contain (among other things):

- **XML content files**
- **Images**
- **Stylesheets**
- **Scripts**
- **Audio and video**
- **Fonts**
- **Metadata**
- **Links to external content**



EPUB 3 organizes all the stuff!



Should we be strict?



Or should we be permissive?



Answer: some of each.





**To get people to conform to a standard,
it's good to make the standard conform
to what people are doing.**

**When multiple options exist,
agreeing on one or two of them
contributes to interoperability
and avoids obsolescence.**

**Where possible,
adopt widely used,
browser-based
standards.**

Technologies



- **HTML5**—A defined profile as XHTML 
- **SVG** —A defined profile
- **CSS3**—A defined profile as EPUB style sheets 
- **JavaScript** for scripting 
- **Dublin Core** for metadata





Core Media Types don't require a fallback

In this presentation, Core Media Types are marked with 

Fallbacks must be provided for Foreign (e.g. non-Core) resources

Technologies **HTML**



- **HTML5**—A defined profile as XHTML 
- **SVG** —A defined profile
- **CSS3**—A defined profile as EPUB style sheets 
- **JavaScript** for scripting 
- **Dublin Core** and a subset of **RDFa** for metadata


Core Media Types don't require a fallback

In this presentation, Core Media Types are marked with 

Fallbacks must be provided for Foreign (e.g. non-Core) resources

Content Markup




- **XHTML**  (The XML serialization of **HTML5**)
- **MathML** (mostly the presentational aspects)
- Eliminated **DTBook** model
- Uses aspects of **HTML5** and modules of **CSS3** that are sufficiently stable or considered essential
- **CFI (Canonical Fragment Identifier)**
enables linking *into* an EPUB publication

*EPUB 3 reading systems must process **EPUB 2***



Layout



- **CSS3**  multicolumn layout module
- **EPUB Style Sheets** are a profile of CSS based on **CSS2.1** + selected CSS3 modules + several EPUB-specific properties
- Support for **multiple style sheets** and CSS3's **media queries**
- More control of **hyphenation**, **linebreaks**, **alignment**, and **word-** and **letterspacing**

Fonts



- **Embedded fonts** enabled and recommended —esp. useful for math and non-latin alphabets
- **OpenType**  is a Core Media Type
- **WOFF fonts**  is a Core Media Type
- Fonts can be obfuscated for license issues

Images



- **GIF** 
- **JPEG** 
- **PNG** 
- **SVG** 

are the four Core Media Types for images

Global Language Support



- **Vertical writing**
- **Right-to-left pagination**
- **Ruby**

Rich Media

- Audio: **MP3** and **MP4 AAC LC**
- **HTML5 <video>** + “informative recommendation”; H.264 vs. VP8/WebM couldn’t be resolved
- Scripting: **JavaScript** per **HTML5** and **SVG**
- **epub:trigger** attribute associates controls with multimedia elements
- **Media Overlays** synchronize text and audio
- **Text-to-speech** (PLS, SSML, CSS3 Speech)

Metadata



Metadata in EPUB 3 *can* be associated with:

- **The publication**—<metadata> in <package>
- **The components**—<item>s in <manifest>
- **Spine items**—whether spreads read L2R or R2L
- **The content elements**—**epub:type** attribute for semantics, **epub:trigger** element for multimedia
- **External resources**—in the *container* (e.g., ONIX file, MARC record, PRISM or XMP metadata) or *online* (ONLY for streaming audio and video)

Metadata



EPUB 3 metadata uses **DCMES** and **<meta>** elements with *@property*, *@refines*, *@id*, *@scheme*

- **Required:** dc:identifier, dc:title, dc:language and `<meta property="dcterms:modified">` (new)
- Small **default vocabulary**—use without prefix
- **Reserved vocabularies**—dcterms, marc, onix, media, xsd—don't have to be declare these prefixes
- **Can specify other prefixes** by declaring them in the `<package>` (e.g. prism:, foaf:, xmp:)

Metadata



Package <metadata> element contains:

- **<dc:identifier>**—one or more
- **<dc:title>**—one or more
- **<dc:language>**—*exactly one* (ISO per RFC5646)
- **<meta>**—one or more
- **<link>**—zero or more (to internal or external resource)

Identifiers in EPUB 3

- **Multiple identifiers** can be included (e.g., ISBNs, DOIs, etc.; none are mandated)
- **@scheme** may be used to specify scheme
- One ID must be designated **@unique-identifier** for the **publication**; intended to be **persistent**
- **<meta property="dcterms:modified">** is a **timestamp** for this exact “bag of bits”
- **“Package Identifier”** is the combination of unique-identifier and modification date

Titles in EPUB 3



- **Multiple versions of title** can be included
- **title-type** property includes *main*, *subtitle*, *short*, *collection*, and *edition*; others with spec'd @scheme
- **display-seq** property specifies display sequence (e.g., series title first, then volume title)
- **file-as** property for proper alphabetization
- **group-position** property for sequence in a collection, e.g. the third Harry Potter book
- **alternate-script** property for non-Latin version

Metadata in EPUB 3



Properties for components (<item>s in <manifest>):

- **cover-image** specifies which image is the cover
- **math-ml** alerts that the item has math-ml content
- **nav** identifies the navigation document (XHTML)
- **remote-resources** alerts to external resources
- **scripted** alerts to scripted content in the item
- **svg** alerts that there is SVG content in the item

Metadata in EPUB 3



Other properties in the EPUB default profile:

- **publication-type** and **source-identifier**
- **meta-auth** to specify a metadata authority
- **role** to describe creators and contributors
- **scheme** with datatype (string or URI)
- **marc21xml-record**, **mods-record**, **onix-record**, **xml-signature**, **xmp-record** to specify properties for @rel in <link>

Metadata in EPUB 3



Three new features useful *inside* content docs:

- **epub:type** attribute can be used to associate semantics with any element (including `!`); functionally equivalent to standard W3C `@role`
- **epub:switch** enables embedding non-native XML fragments (e.g., ChemML) in EPUB content
- **epub:trigger** element enables images or elements to control multimedia content

EPUB 3 metadata—from the package to the phrase!

Thanks!



Bill Kasdorf

bkasdorf@apexcovantage.com

+1 734 904 6252

@BillKasdorf

Go to **<http://idpf.org/epub/30>** for
Overview, Publications 3.0, Content Documents 3.0,
OCF 3.0, Media Overlays 3.0, EPUB 3 Changes
from EPUB 2.0.1, & Structured Semantics Vocab.