Metadata in EPUB 3:
It's what makes it all work

Bill Kasdorf

Vice President, Apex Content Solutions
General Editor, *The Columbia Guide to Digital Publishing*
Metadata Subgroup Lead, *IDPF EPUB 3 Working Group*

EPUB is a package masquerading as a file format.

Metadata in EPUB 3

An EPUB 3 *can* contain (among other things):

- **XML content files**—XHTML, SVG
- **Images**—GIF, JPEG, PNG, SVG
- **Stylesheets**—CSS3 (more than one!)
- **Scripts**—Javascript
- **Audio and video**
- **Fonts**—OpenType or WOFF

- **. . . and METADATA!** (Lots of it)

Why metadata is so essential.

But your metadata isn't necessarily like my metadata.

EPUB 3 is not just for books.

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!

Should we be strict?

Or should we be permissive?

Answer: some of each.

For backwards compatibility, EPUB 3 __requires__ the same three metadata elements as EPUB 2:

dc:identifier, **dc:title**, and **dc:language**.

But it __permits__ lots more!

[Okay, there are a few additional technical requirements here and there . . . e.g., a "last modified timestamp" to help uniquely identify the package.]

Metadata in EPUB 3

Metadata in EPUB 3 *can* be associated with:

- **The package**—<metadata> in <package> (.opf)
- **The components**—<item>s in <manifest>
- **The spine**—whether spreads read l2r or r2l
- **The content elements**—epub:type attribute for semantics, epub:trigger element for multimedia
- **External resources**—either in the package (e.g., ONIX file, MARC record, PRISM metadata) or online (e.g., volatile metadata, streaming media)

Metadata in EPUB 3

EPUB 3 metadata is expressed in:

- **[DCMES]**—for identifier, title, and language
- **[DCTERMS]**—preferred vocabulary
- Uses small subset of **RDFa 1.1**
- **EPUB default profile**—small vocabulary of EPUB properties you can use without a prefix
- Add **other profiles for other vocabularies** (requires using a **prefix,** e.g. prism:, foaf:, xmp:)

Metadata in EPUB 3

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; + @property, @id, @about
- **<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 describes type of title, e.g., short title, acronym, subtitle, series title, volume title
- **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 @rel in <link>

Remember you can add others, with profiles/prefixes!

Metadata in EPUB 3

Three new features useful *inside* content docs:

- **epub:type** attribute can be used to associate semantics with any element (including <span>!); 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

Vice President, Apex Content Solutions

bkasdorf@apexcovantage.com

+1 734 904 6252