

Il formato EPUB 3

Editoria digitale

Maria A. Alberti
Informatica per la comunicazione digitale
AA 17/18

Il documento EPUB

- Il formato EPUB 3 può essere visto come il collante che lega tutte le tecnologie che usa
- Un documento EPUB costituito da numerosi file che contengono separatamente i dati. Ad esempio:
 - Il testo suddiviso in capitoli o sezioni
 - Importante usare i tag semantici (article, section, figure ...) e gli attributi epub-type per specificare note, rimandi...
 - File di stile come CSS
 - Le immagini e la copertina
 - Altri file di media: video o audio
 - File di scripting per gestire l'interazione
 - L'indice
 - I metadati

Da EPUB2 a EPUB3

EPUB2 è basato sulle precedenti tecnologie XHTML 1.1 e CSS 2

EPUB3 aggiunge la possibilità

- di sviluppare interazione con JavaScript
- di inserire risorse multimediali e overlay multimediali
- di affrontare il tema dell'accessibilità

EPUB 3, una collezione di formati

Alcune tecnologie usate da EPUB3 sono state sviluppate ad hoc dall'IDPF ma la maggior parte erano già state sviluppate e sono standard pubblici, nel dettaglio:

- **XHTML5** per rappresentare il testo, incluse le formule matematiche in **MathML**, markup per la pronuncia e immagini in grafica vettoriale **SVG**
- **SVG 1.1** per rappresentare grafica (ad esempio per fumetti o vignette)
- **CSS 2.1 e 3** per la resa visuale del contenuto
- **JavaScript** per creare interazione
- **TrueType** e **WOFF** per fornire supporto per i caratteri oltre a quelli di base
- **SSML/PLS/CSS 3** speech per la resa della funzione di lettura automatica text-to-speech
- **SMIL 3** per sincronizzare il testo con l'audio
- **RDF** per associare vocabolari e tesauri per le informazioni semantiche sulla pubblicazione
- **XML** per la definizione dei contenuti
- **ZIP** per la compressione dei contenuti in un unico file

La versione EPUB 3 rispetto alla precedente contiene diverse aggiunte essenzialmente legate alla possibile introduzione di contenuti multimediali

EPUB 3, documenti di specifica

Il formato EPUB 3 è definito da 4 documenti di specifica, ciascuno dei quali dettaglia come vengono usate le diverse tecnologie e come interagiscono tra loro.

- EPUB Publications 3.0
- EPUB Content Document 3.0
- EPUB Media Overlays 3.0
- EPUB Open Container Format 3.0

EPUB specifica quali risorse la pubblicazione include e come sono organizzate.

- Non solo il formato del contenuto ma anche come usare il contenuto e quindi non solo dati ma anche metadati o informazioni sui dati
- *A web site in a box*

Queste informazioni sono consistenti e coerenti per tutti gli strumenti di lettura

EPUB Publication 3.0

La specifica che definisce il formato XML usato nel **package document** per archiviare le informazioni sulla pubblicazione-

Il **package document**

- contiene i metadati della pubblicazione (come titolo, autore, lingua ...)
- lista tutte le risorse (testi, immagini, grafica, audio, video, file smil ...)
- definisce la sequenza di default della lettura e
- indica dove si trova il navigation document

EPUB Content Document 3.0

Specifica il profilo XHTML5, SVG 1.1, CSS2 e CSS3 usati per definire il contenuto della pubblicazione.

profilo significa definire il sottoinsieme delle funzionalità che si possono usare in ePub.

Definisce ad esempio le nuove funzionalità circa la multimedialità e l'interazione di ePub3 vs ePub2

Definisce anche il formato del navigation document, che può includere strumenti di navigazione con tavole, figure.

EPUB Media Overlay 3.0

Specifica come arricchire la pubblicazione con audio, non inserito nelle pagine, ma attraverso un file SMIL che costituisce uno strato aggiunto al documento.

Il file SMIL specifica come rendere la pubblicazione nei sistemi di lettura sincronizzandoli all'audio

EPUB Open Container Format – OCF 3.0

Specifica il contenitore indicando come assemblare tutte le risorse insieme in un singolo file.

Quindi verrà compattato con **zip**

Il documento EPUB...

- Il **package document** con le risorse associate e il **content document** sono quindi compressi in un unico file con estensione .epub contenente anche
 - Lo speciale file **mimetype** con l'indicazione che il file zip contiene un EPUB file
 - La cartella **META-INF** con il file **container.xml** usato dai lettori per trovare il package document e sapere come sono aggregate le risorse

Struttura file EPUB 3

La struttura della cartella è così composta:

- Il file **mimetype** con la dichiarazione del formato
 - application/epub+zip
- La cartella **META-INF** obbligatoria contenente
 - Il file **container.xml** obbligatorio, secondo la specifica **OCF**, OEB Container Format descrive come le varie parti di contenuto devono essere archiviate per realizzare un unico file
- La cartella **OEBPS** secondo le specifiche
 - **OPF**, OEB Package Format, che descrive la struttura del file .epub. La specifica è rispettata dai file:
 - **content.opf** elenca i contenuti e i metadati relativi al contenuto
 - **toc.xhtml** per la generazione della tavola dei contenuti
 - **OPS**, OEB Publication Structure che descrive la formattazione dei contenuti di testi in XHTML, immagini, font e stili in CSS

Il file content.opf

Descrive la pubblicazione

- Un file XML, **.opf** OEB Package Format
- Viene usato dai lettori per stabilire come renderizzare il contenuto

Contiene 3 sezioni obbligatorie:

- La sezione **metadata** con metadati specificati secondo Dublin Core
- La sezione **manifest** lista tutte le risorse per controllare se il documento epub è completo
- La sezione **spine** stabilisce l'ordine di default delle risorse

Elemento package . . .

Gli elementi obbligatori:

- **metadata** con le meta informazioni sul documento
- **manifest** con la lista delle risorse che costituiscono il doc, ciascuno indicato da un elemento **item**
- **spine** che definisce l'ordine di lettura di default degli elementi **item** della sezione **manifest**, ciascun indicato con l'elemento **itemref**

Gli elementi opzionali:

- **guide** deprecato ma utile per compatibilità con ePub2
- **bindings**

Elemento package: il codice

La radice del package con gli attributi richiesti:

- **xmlns** Lo spazio dei nomi.
- **version** delle specifiche EPUB
- **unique-identifier**

```
<package  
  xmlns="http://www.idpf.org/2007/opf"  
  unique-identifier="uuid_id"  
  version="3.0">
```

Struttura del file content.opf

```
<package ...  
  version="3.0" xmlns="http://www.idpf.org/2007/opf">  
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">  
    ...  
  </metadata>  
  <manifest>  
    ...  
  </manifest>  
  <spine>  
    ...  
  </spine>  
</package>
```

Sezione metadata

La dichiarazione dello spazio dei nomi del Dublin Core con i 3 attributi obbligatori.

```
<metadata  
  xmlns:dc="http://purl.org/dc/elements/1.1/">  
  <dc:identifier>...</dc:identifier> anche più di uno, ma  
  uno deve essere unico: ISBN, DOI o altro  
  <dc:title>titolo</dc:title>  
  <dc:lang>it</dc:lang> ripetere se più di una lingua  
  ...  
</metadada>
```


Altri elementi DCMES opzionali

`dc:creator` un secondo livello di collaborazione alla produzione, che può essere specificato ulteriormente tramite la proprietà `refines`

```
<metadata> ...  
<dc:creator id="creator">Gustave Doré  
</dc:creator>  
<meta refines="#creator" property="file-as">  
  Doré, Gustave  
</meta>
```

Sezione manifest

Elenca tutte le risorse che compongono l'epub, tranne `mimetype`, `container.xml` e `content.opf`

```
<manifest> ... Per ogni risorsa un ingresso  
<item  
  id           essenziale per riferimenti interni  
  href         il file effettivo  
  media-type  il MIME che specifica tipo e formato  
  ... questi 3 attributi obbligatori, altri opzionali come  
  media-overlay con valore il file .smil dell'item  
</item>
```

Sezione spine

Fornisce l'ordine di lettura ed è usato dallo strumento di lettura per fornire l'ordine di default di lettura

```
<spine> ... Per ogni elemento
  <itemref idref="capitolo 1" />
  <itemref idref="capitolo 2" />
  <itemref idref="capitolo 3" />
  ...
</spine>
```

Sezione spine

La sezione può contenere un elemento con attributo `toc` che fa riferimento al file `toc.ncx` deprecato in `epub3` per ragioni di compatibilità con `epub2`

```
<manifest> ...
  <item href=id="ncx"
    media-type="application/x-dtbnex+xml" />
</manifest>
<spine toc="ncx" >
  ...
</spine>
```

Differenza tra spine e il documento di navigazione toc.xhtml

- La sezione `spine` fornisce l'elenco dei capitoli o documenti principali nel loro ordine di lettura di default possibilmente diverso da quello scelto dal lettore umano
 - Fondamentale per lo strumento di lettura per sapere quali sono le risorse primarie nella loro successione logica
- Il documento di navigazione può fornire uno strumento di orientamento nel documento epub molto più ricco esplicitandone la struttura interna
 - Elenchi di tavole, immagini, punti salienti del documento