

## Il formato EPUB 2

### Editoria digitale

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

## Il formato EPUB

- Il formato EPUB, 2007, è definito sugli standards Web
- Garanzia di sopravvivenza nel tempo del formato
  - Il Web, fin dalla sua prima versione nel 1999, è stato visto come l'ambito naturale in cui implementare la lettura digitale
  - HTML aveva già assunto un ruolo primario nel panorama dei formati re-flowable che adattano il contenuto ai diversi strumenti di output, contrapposto ai formati a layout fissi, come il PDF
- Essere legati ai Web e ai browser offre indiscutibili vantaggi per lo sviluppo dei sistemi di lettura
  - Con l'uso dei CSS si sono garantiti una grande ricchezza di visualizzazione

## Supporto a EPUB

- Non tutte le caratteristiche possibili in ePub vengono supportate in uguale misura da piattaforme sw ed eReader dedicati
  - Amazon non supporta ePub
  - Apple iBooks al momento supporta sia il formato liquido sia fisso al meglio
  - Kobo buon supporto
  - Barnes & Noble non supporta ufficialmente ma i file epub vengono visualizzati
  - Google Play Libri solo parzialmente
  - Sony solo minimalmente
- Tutte le informazioni aggiornate vengono mantenute da Book Industry Study Group BISG alla pagina [epubtest.org/](http://epubtest.org/)

## Radium foundation

- La fondazione [Radium.org](http://Radium.org) ha l'obiettivo di favorire lo sviluppo e la diffusione del formato ePub
- Il plugin Radium per la visualizzazione di file epub per il browser Chrome
- L'ambiente SDK, Software Development Kit

## Struttura file EPUB 2 - 1

EPUB 2 usa sostanzialmente solo 3 tecnologie:

- XHTML per definire i file di contenuto
- CSS per definire le specifiche di visualizzazione
- ZIP per compattare in un unico file

La struttura della cartella .epub compressa è così composta:

- Il file **mimetype**
- La cartella **META-INF**
- La cartella **OEBPS**

## Struttura file EPUB 2 - 2

Più in dettaglio la cartella .epub è così strutturata:

- Il file **mimetype** con la dichiarazione del formato
  - application/epub+zip
- La cartella **META-INF** obbligatoria contenente
  - Il file **container.xml** obbligatorio, scritto in XML
  - Altri file non obbligatori, come la firma digitale o il manager dei diritti, DRM
- La cartella **OEBPS** obbligatoria contenente
  - Il file **content.opf**
  - Il file **toc.ncx**
  - Le cartelle **Images Styles Text** con i contenuti

## La cartella META-INF: il file container.xml

- **container.xml** è il primo file che un applicativo di lettura deve interpretare
- Descrive come le varie parti di contenuto devono essere archiviate per realizzare un unico file
- Segue la specifica **OCF**, OEBPS Container Format
- Indica i file root del documento e se l'ePub file contiene più di un documento allora questo file deve indicare più di un file root
  - L'istruzione importante è la dichiarazione **rootfile**

## Il file container.xml

nell'esempio il file root è content.opf

```
<?xml version="1.0"?>
<container version="1.0"
xmlns="urn:oasis:names:tc:opendocument:xmlns:container">
  <rootfiles>
    <rootfile full-path="OEBPS/content.opf" media
              type="application/oebps-package+xml" />
  </rootfiles>
</container>
```

- Importante è la dichiarazione del **rootfile** content.opf
- content.opf il file di metadati che descrive il contenuto del documento/libro

## La cartella OEBPS

OEBPS ??

- Sta per **Open eBook Publication Structure**
  - Sviluppato inizialmente da Open eBook Forum, chiamato poi **International Digital Publishing Forum** (IDPF), ora aderente al W3C
- Costituisce la specifica, in XML, del contenuto, della sua struttura e della presentazione della pubblicazione elettronica

## La cartella OEBPS

Più in dettaglio la cartella **OEBPS** è così strutturata:

- Il file **content.opf**
  - Il documento che descrive la struttura dell'eBook, i file OEBPS, l'ordine con cui devono essere visualizzati, i metadati
- Il file **toc.ncx**
  - La tavola dei contenuti utilizzata dagli applicativi per navigare nel contenuto
- Le cartelle **Images Styles Text**
  - Contengono i contenuti dell'opera divise ovviamente per immagini, testo e i file di stile per la visualizzazione

## La cartella OEBPS: il file content.opf

Il documento principale descrive la pubblicazione e contiene i metadati . Se il file ePub è strutturato in più di un libro/documento ciascuno deve avere il suo file content.opf

La radice del file è l'elemento `<package>`. Ad esempio:

```
<package xmlns=http://www.idpf.org/2007/opf
  unique-identifier="bookid"
  version="2.0">
  ... Le sezioni metadata manifest spine guide ...
</package>
```

Inoltre il file content.opf contiene le sezioni **metadata manifest spine guide**

## Il file content.opf: la sezione metadata

La sezione **metadata** contiene tutti i metadati descrittivi del documento come: titolo, autore, editore, creatore, soggetto, descrizione... E' racchiusa dal tag `<metadata>`

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

  <dc:identifier>almeno 1 unico, ISBN, DOI</dc:identifier>
  <dc:title>Il titolo che vuoi</dc:title>
  <dc:language>it</dc:language>
  <dc:date opf:event="data stampa">2012-09-10</dc:date>
  <dc:date opf:event="data epub">2012-09-10</dc:date>
</metadata>
```

## Il file content.opf: la sezione manifest

La sezione **manifest** lista precisamente tutti e solo i file che compongono l'eBook. Ogni elemento di questa lista è un file. E' racchiusa nei tag **<manifest>** e ogni elemento è istanziato con il tag **<item>**

Ad esempio un'istanza di file della sezione:

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

Ogni item ha 3 attributi: **id**, **href**, **media-type**. Rispettivamente un identificatore **unico**, l'URI in cui si trova il file relativo alla risorsa e il tipo, i cui valori possibili sono:

```
application/x-dtbncx+xml (riservato al file .ncx)
application/xhtml+xml, text/css, image/png, image/gif,
image/jpeg, image/svg+xml, font/svg, font/opentype
```

M.A. Alberti. Editoria digitale, AA 17/18

13

Il formato EPUB

## Il file content.opf: la sezione spine

La sezione **spine** definisce la sequenza di lettura di tutti i file che possono essere visualizzati indipendentemente. E' racchiusa dai tag **<spine>** Ogni elemento è introdotto dal tag **<itemref>** e attributo **idref**

```
<spine toc="ncx"> ...
  <itemref idref="title">
  <itemref idref="copertina">
  ...
</spine>
```

L'attributo **toc** ha come valore l'id dell'item **toc.ncx** della sezione **manifest**

M.A. Alberti. Editoria digitale, AA 17/18

14

Il formato EPUB

## Il file `content.opf`: la sezione `guide`

La sezione **guide** ha la funzione di segnalare all'applicativo di lettura il ruolo dei diversi file. È racchiusa dai tag `<guide>`. Ogni elemento è introdotto dal tag `<reference>` con attributo **type**

```
<reference type="cover" href="cover.xhtml" />
```

```
<reference type="text" href="chapter-01.xhtml" />
```

L'attributo `type` può avere i valori:

`cover`, `colophon`, `title-page`, `copyright page`, `toc`,  
`loi`, `lot`, `foreword`, `preface`, `text`,  
`acknowledgements`, `dedication`, `epigraph`, `notes`,  
`index`, `glossary`, `bibliography`

## Il file `toc.ncx` in OEBPS

Contiene la tavola dei contenuti del libro. Il nome `toc` non è obbligatorio ma è consuetudine, mentre lo è l'estensione `.ncx`.

È costituito da 3 sezioni: **head**, **document title** e **navigation map**.

La sezione **head** contiene i metadati sul libro, la profondità dei titoli di sezioni e sottosezioni, il numero delle pagine (se rilevante per la stampa)

La sezione **document title** contiene il titolo del libro che deve essere lo stesso in `content.opf`



## Il file toc.ncx: la sezione navigation map

La sezione **<navMap>** deve essere coerente con l'ordine e l'id degli elementi in lista nella sezione **<spine>** del file **content.opf**

```
<navMap>
  <navPoint class="chapter" id="title" playOrder="1">
    <navLabel>
      <text>Title Page</text>
    </navLabel>
    <content src="Text/title.xhtml" />
  </navPoint>
  ...
</navMap>
```

I numeri d'ordine devono essere successivi e non devono presentare salti

## Per la pubblicazione

- Includere sempre la copertina: un'immagine di dimensioni di 548x800 per una resa ottimale nelle diverse piattaforme
- Includere sempre: la tavola dei contenuti (...ncx), la pagina del titolo, la pagina del copyright
- Includere i metadati nella sezione **<metadata>** del file **content.opf**

## Per la pubblicazione

- Un file .epub scritto correttamente in XHTML e secondo le specifiche per l'organizzazione del contenitore.
- Può essere successivamente validato
  - EpubCheck 4.0.2 ultima versione del nov 2016
  - <https://github.com/IDPF/epubcheck/releases>
  - Per essere accettato dalle diverse piattaforme di distribuzione o diversi eReader non basta avere un file validato

## La pagina copyright

Come prima riga deve avere l'indicazione di chi detiene i diritti. Le frasi successive sono specifiche e possono variare da caso a caso

Copyright © 2011 O'Reilly Media, Inc.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://my.safaribooksonline.com>). For more information, contact our corporate/institutional sales department: (800) 998-9938 or [corporate@oreilly.com](mailto:corporate@oreilly.com).

Nutshell Handbook, the Nutshell Handbook logo, and the O'Reilly logo are registered trademarks of O'Reilly Media, Inc. *What Is EPUB 3?*, the image of a Thick-Knee, and related trade dress are trademarks of O'Reilly Media, Inc.

This book was published in collaboration with the **International Digital Publishing Forum (IDPF)**, the global trade and standards organization for electronic publishing that is responsible for development of the EPUB format. EPUB is a registered trademark of the IDPF.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc., was aware of a trademark claim, the designations have been printed in caps or initial caps.

## I metadati

Quelli minimi richiesti

```
<dc:title>Vostro titolo</dc:title>
<dc:language>it</dc:language>
<dc:identifier id="uuid_id" opf:scheme="uuid">
  id univoco es ISBN
</dc:identifier>
<dc:creator>Nome autore</dc:creator>
<dc:publisher>Nome editore</dc:publisher>

<dc:date>2015-11-10</dc:date>
```

La data non è richiesta da tutte le piattaforme  
**dc** sta per **Dublin Core**: un formato per la metadattazione