

10 Electronic Books and Magazines

10.1 Terminology, History

10.2 Formats for Electronic Books

10.3 Open Standard for Electronic Books: ePUB

Literature (German):

V. Wang: E-Books mit ePUB, mitp Verlag 2011

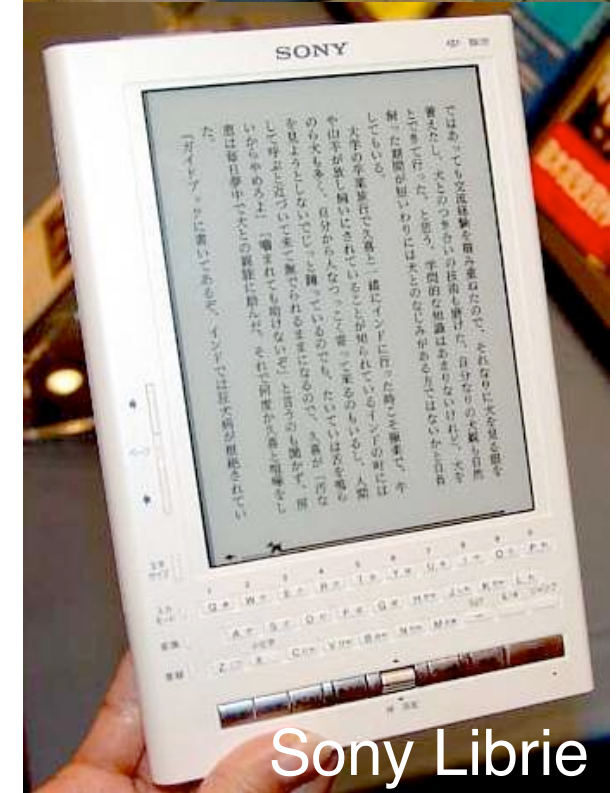
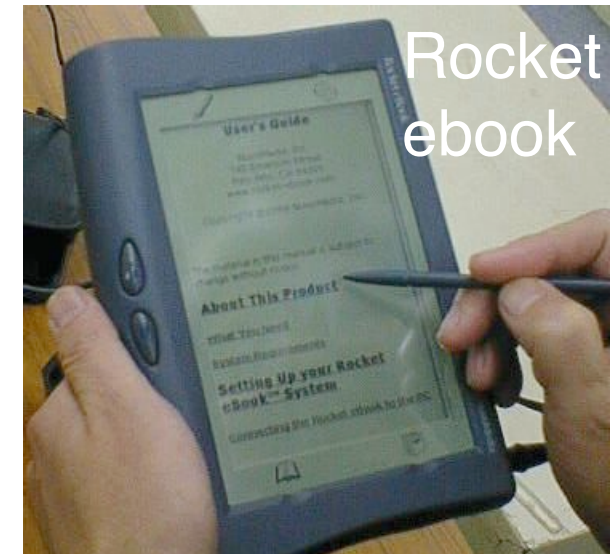
Electronic Book

- An **electronic book (e-book)** is a publication which
 - contains text and pictures as main content, possibly enhanced with further media (audio, animations, video)
 - is distributed in the form of a digital code
 - is optimized for the presentation with appropriate reading systems (hardware and software)
- Examples for proper e-books:
 - products made for viewing on specific e-book readers like Amazon Kindle
 - products of the Apple iBook store
- Examples for products similar to e-books, sometimes also called e-book:
 - Audio books
 - PDF versions of a printed book (one-to-one representation)
- An **electronic magazine** or **electronic newspaper** is a publication with the same characteristics as an e-book, but with a regular publication schedule of individual issues.

History of Electronic Books

- 1971: Michael S. Hart, *Project Gutenberg* (until today)
- 1985–1992: Robert Stein, *Voyager Company*
- 1993: Company *Digital Book*, Digital Book Format (DBF), books on floppy disk
- 1998: *Rocket ebook* and *Softbook*, first e-book readers
- 2004: *Sony Librie*, first e-book reader with e-ink
- 2007: Launch of *Amazon Kindle*
- 2010: Launch of *Google eBooks*
- 2010: Amazon.com reports that e-book sales have for the first time outnumbered hardcover book sales (2nd Quarter 2010)

Source: Wikipedia,
Images: harpowoman.wordpress.com, www.yesky.com



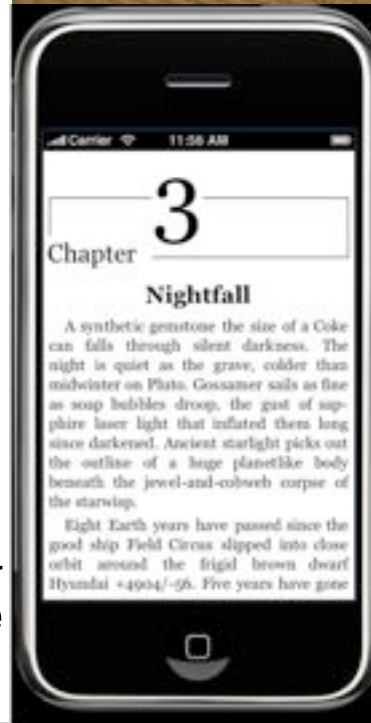
E-Book Readers

- An **e-book reader** is a (usually portable) device which is designed for being used to read electronic books.
- Classification:
 - Specialized e-book reading devices, mostly using **e-ink** display technology
 - PDAs or smartphones with reader software
 - Notebooks or netbooks with reader software
 - Tablet PCs with reader software

Stanza
on iPhone
(lexcycle.com)

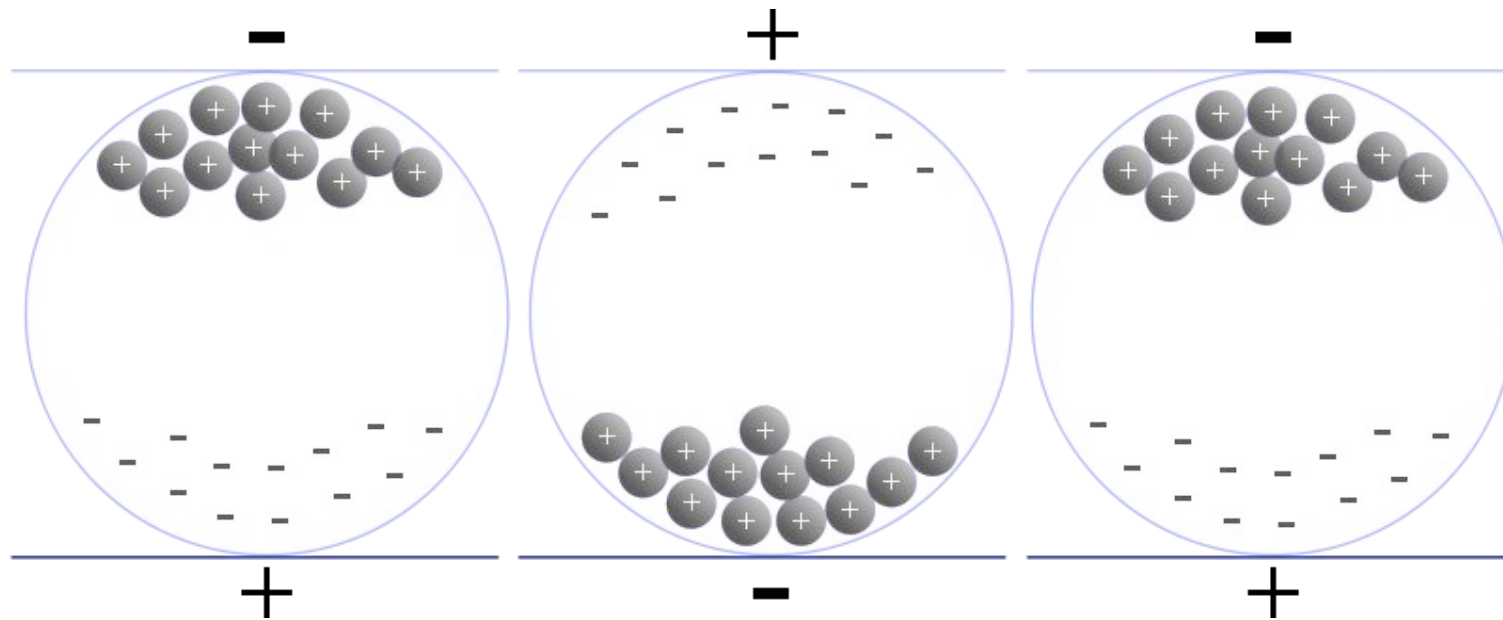


Various readers
(Wikimedia)



Electrophoretic Displays (E-Ink)

- Principle: Small coloured particles (one micrometer diameter) dispersed in dyed hydrocarbon oil, particles migrate to electrodes on opposite sides
- Relatively slow, only global updates (needs "flashing" to avoid "ghost images"), currently mostly black&white (color prototypes exist)
- Extremely low energy consumption, extremely high contrast, image stays



(Wikipedia)

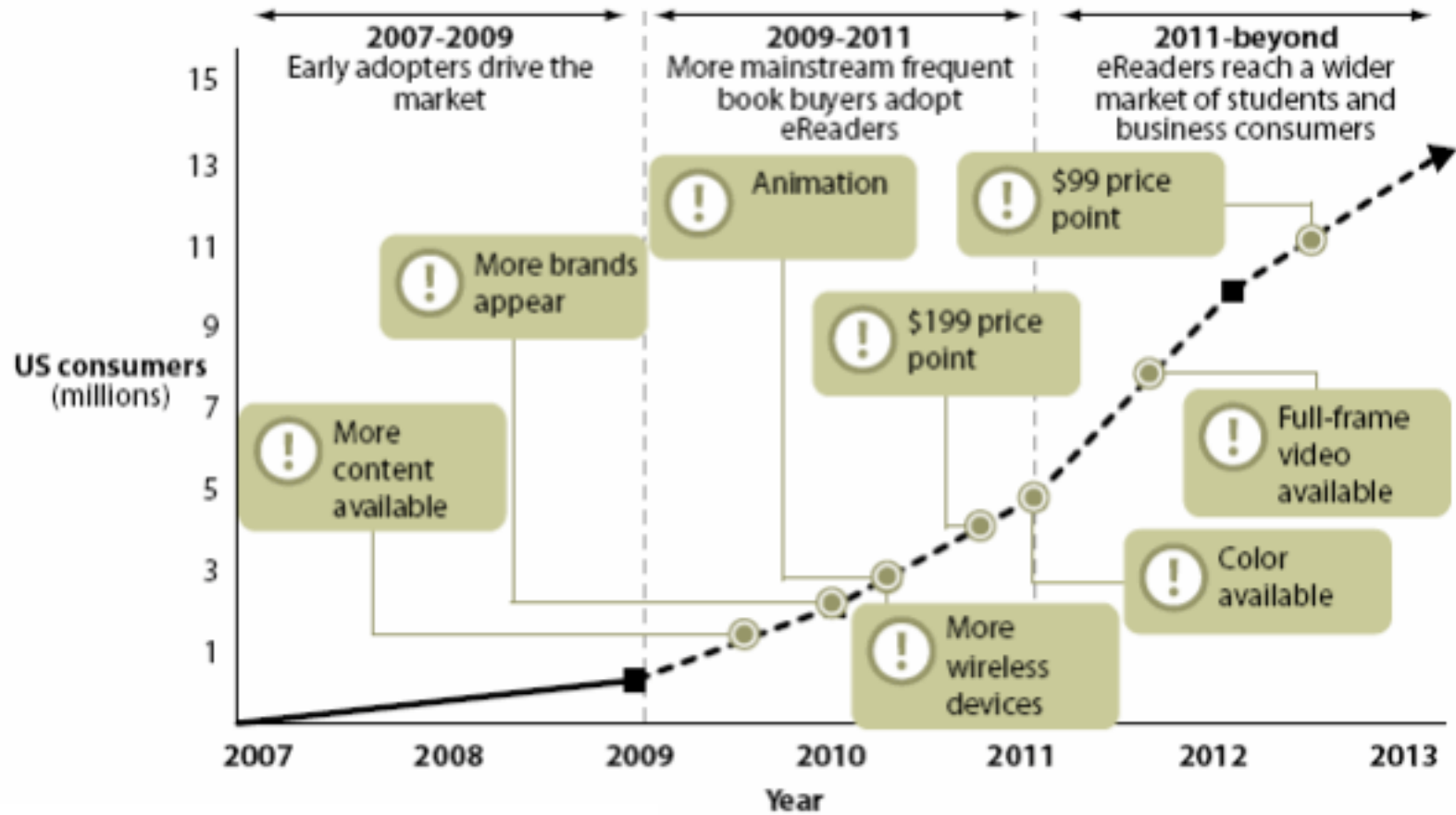
E-Reader Market USA 2011



Kindle:
Estimated 62% market share

Images: ereaderleader.com, market data: electronista.com

Hype Cycle for E-Book Readers



www.readwriteweb.com, June 2009

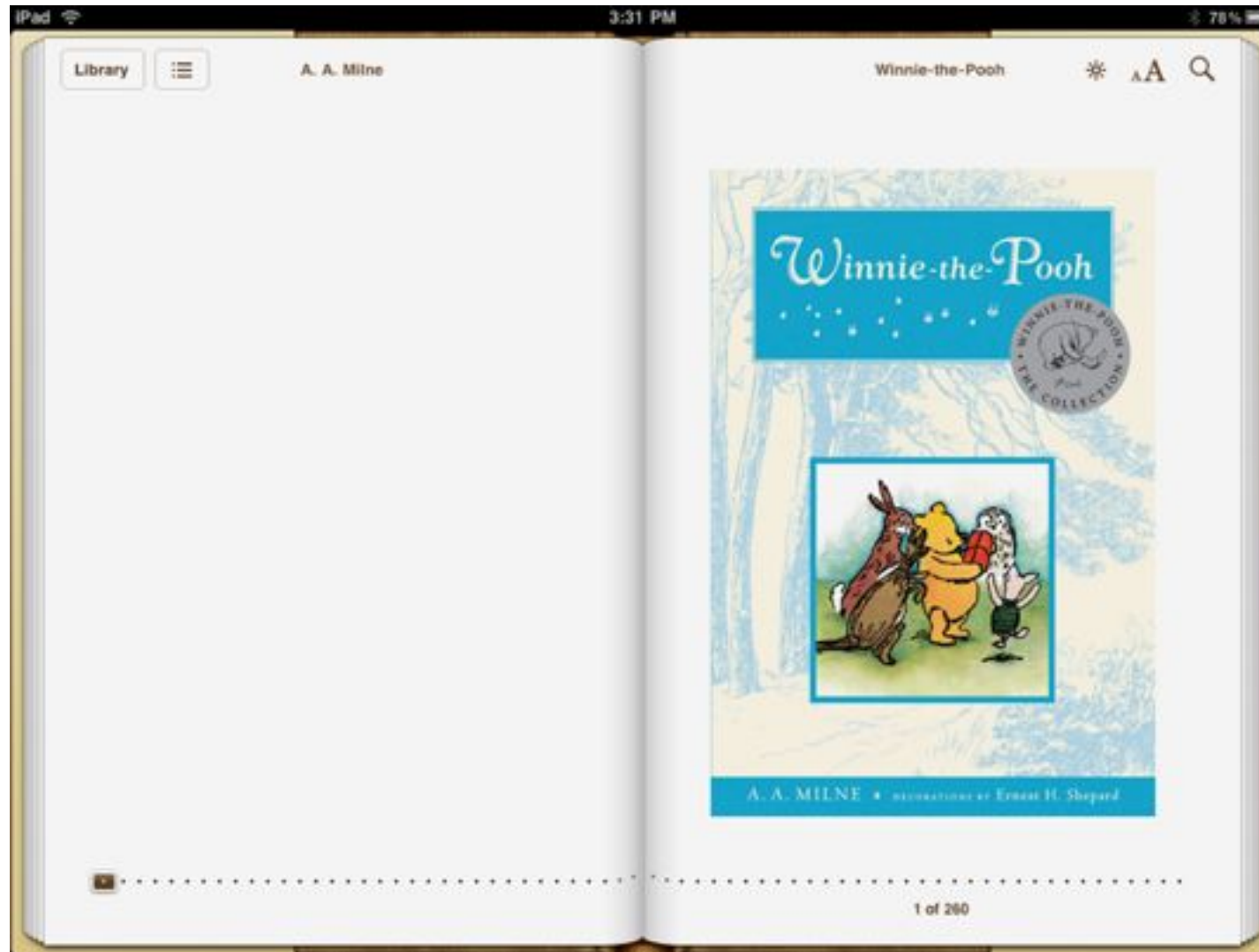
Source: Forrester Research, Inc.

In-Stat, Sep 2010: 12 million units sold in 2010, to triple by 2014

Conceptual Difference E-Book / Printed Book

- E-Book:
 - optimized for display on viewers
 - flexible e.g. for different font sizes
 - navigation by hyperlinks and bookmarks
 - search functions
- Many of the traditional elements of book pages are redundant or not usable:
 - chapter and section headers
 - (fixed) page numbers
 - margin columns
 - index
- E-Books are produced from traditional books by extracting the pure text and removing unneeded elements

Example: Apple iBook



www.freeipadapps.net

Copyright, Viewing Restrictions and DRM

- Freely available e-books are usually copyright-free material
- Most e-book formats support Digital Rights Management (DRM)
- Electronic previews of printed books (*no e-books, strictly spoken!*) are restricted in viewing volume (fair use doctrine)

Google Books:



Sie haben entweder eine Seite erreicht, die nicht angezeigt werden kann, oder die Anzeigebeschränkung für dieses Buch erreicht.

10 Electronic Books and Magazines

10.1 Terminology, History

10.2 Formats for Electronic Books

10.3 Open Standard for Electronic Books: ePUB

Literature (German):

V. Wang: E-Books mit ePUB, mitp Verlag 2011

H. Reibold: E-Books selbst gemacht, bomots Verlag 2010

HTML

- HTML is platform-independent and browsers exist for many platforms
- HTML can be used for simple e-books
- HTML is the basis for the chapter content in most current e-book formats
- HTML is not memory-efficient
 - HTML code needs to be compressed, eg. using ZIP

Open E-Book

- 1998: First e-book conference in the US (NIST)
- Working group "Open eBook Authoring Group" founded
 - 25 representatives of international companies and organisations
 - Developed Open eBook Publication Structure (OEBPS, OeB), 1999
- OEBPS:
 - Based on HTML 4 and XML
 - Uses subset of Cascading Style Sheets (CSS)
 - Package file with metadata (in Open Package Format OPF)
 - Chapters in HTML
- 1999: Open eBook Forum founded to maintain the standard
 - 2005 renamed to "International Digital Publishing Forum" (IDPF)
- OEBPS 1.01 (2001):
 - Fully compatible with XHTML 1.1
 - Works only with external style sheets
- Impact mainly indirectly through other formats

Microsoft LIT Format (.lit)

- DRM-protected proprietary format, since 2000
 - Originally developed for PocketPC
- Readable with
 - Microsoft Reader (protected files)
 - Lexstyle Stanza and other tools (unprotected files)
- Compressed form of Microsoft Hep Module Format (CHM)
 - Essentially based on OEBPS and HTML
- Uses patented *ClearType* technology for improving readability on LCD screens
- Supports user annotations
- "Verbosity functions" enable speech output

Mobipocket-Format (.prc)

- French company Mobipocket, 2000:
 - develops e-book reader software for PDAs, in particular for Palm devices
 - eBook Technology award 2001 (Frankfurter Buchmesse)
 - Has been ported to many operating systems
(including Windows, Windows Mobile, Blackberry, Symbian OS)
- File extension .prc just for compatibility with PalmOS
- Based on OEBPS and HTML
- Development tool *Mobipocket Creator*
 - Compiles into binary and compressed file format

Amazon Kindle Format (.azw)

- 2005: Amazon buys Mobipocket
 - Kindle devices on the market starting 2007
- "AZW" probably relates to "Amazon Whispernet"
 - Amazon network for wireless distribution of content
- Amazon-specific DRM, but otherwise compatible with Mobipocket format
- Conversion tool (command line) *Amazon KindleGen*
 - from HTML, XHTML, XML (OPF/IDPF), ePUB
- Topaz:
 - Proprietary format used by Amazon/Kindle, not many details known

More E-Book Formats

- Examples:
 - DAISY
 - » Digital Accessible Information System
 - » Mainly oriented towards accessibility for challenged people
 - » Used e.g. for audio books for the visibly impaired
 - » Being harmonized with ePUB
 - FictionBook (.fb2)
 - » Russian book format, used in particular for Russian literature
 - » XML-based
 - LRF
 - » Proprietary format by Sony
 - ...

10 Electronic Books and Magazines

10.1 Terminology, History

10.2 Formats for Electronic Books

10.3 Open Standard for Electronic Books: ePUB

Literature (German):

V. Wang: E-Books mit ePUB, mitp Verlag 2011

Why ePUB?

- ePUB is the most likely candidate for a long-lasting e-book standard
 - based on open standards
 - itself open standard
 - supported by major industry players
- Possible breakthrough event:
 - iBook store on iPad/iPod/iPhone by Apple is based on ePUB

OPF/OPS and ePUB

- 2005: IPDF tries to harmonize the various similar but divergent formats
 - Single final format for e-books
 - Creation of a single flexible container format (Open Container Format OCF)
 - Compatibility with latest versions of related standards (e.g. XML, CSS)
 - Integration of accessibility aspects based on DAISY
- 2007: Definition of "ePUB" (.epub)
 - Container in OCF
 - Contents in Open Packaging Format (OPF), following the Open Publication Structure (OPS)
 - (all standards in version 2.0)
- Plans for ePUB 3.0:
 - Rich Media, interactivity (based on HTML5?)
 - Newspapers and magazines
 - User annotations
 - ...

ePUB: Three Parts of the Standard

- OPF (Open Packaging Format)
 - Metadata
 - Handling of related files, serialization
 - Navigation, accessibility
- OPS (Open Publication Structure)
 - Structuring of contents
 - Representation of contents
- OCF (Open Container Format)
 - Single file for publication
 - Logical file system within the (compressed) single file
 - DRM, watermarking



Open Publication Format (OPF)



- Two XML files:
 - Open Publication Format file (.opf)
 - Navigation Control File for XML Applications (.ncx) (from DAISY)
- Contents of OPF file:
 - <package> element containing:
 - » <metadata> based on Dublin Core Standard
 - » <manifest>: list of files
 - » <spine>: order of documents (serialization)
 - » <tours> (optional, deprecated): Alternative sequences
 - » <guide> (optional): Entrance point to overview parts of publication

Open Publication Structure



- OPS: Definition of the actual language in which the content is written.
- Two alternative vocabularies:
 - XHTML, strongly based on vocabulary for Web sites (frequently used)
 - DTBook, based on the DAISY standard, stressing accessibility (rarely used)
- DTBook gives the more sophisticated structuring aids
- For a first start, XHTML knowledge is sufficient to write ePUB documents

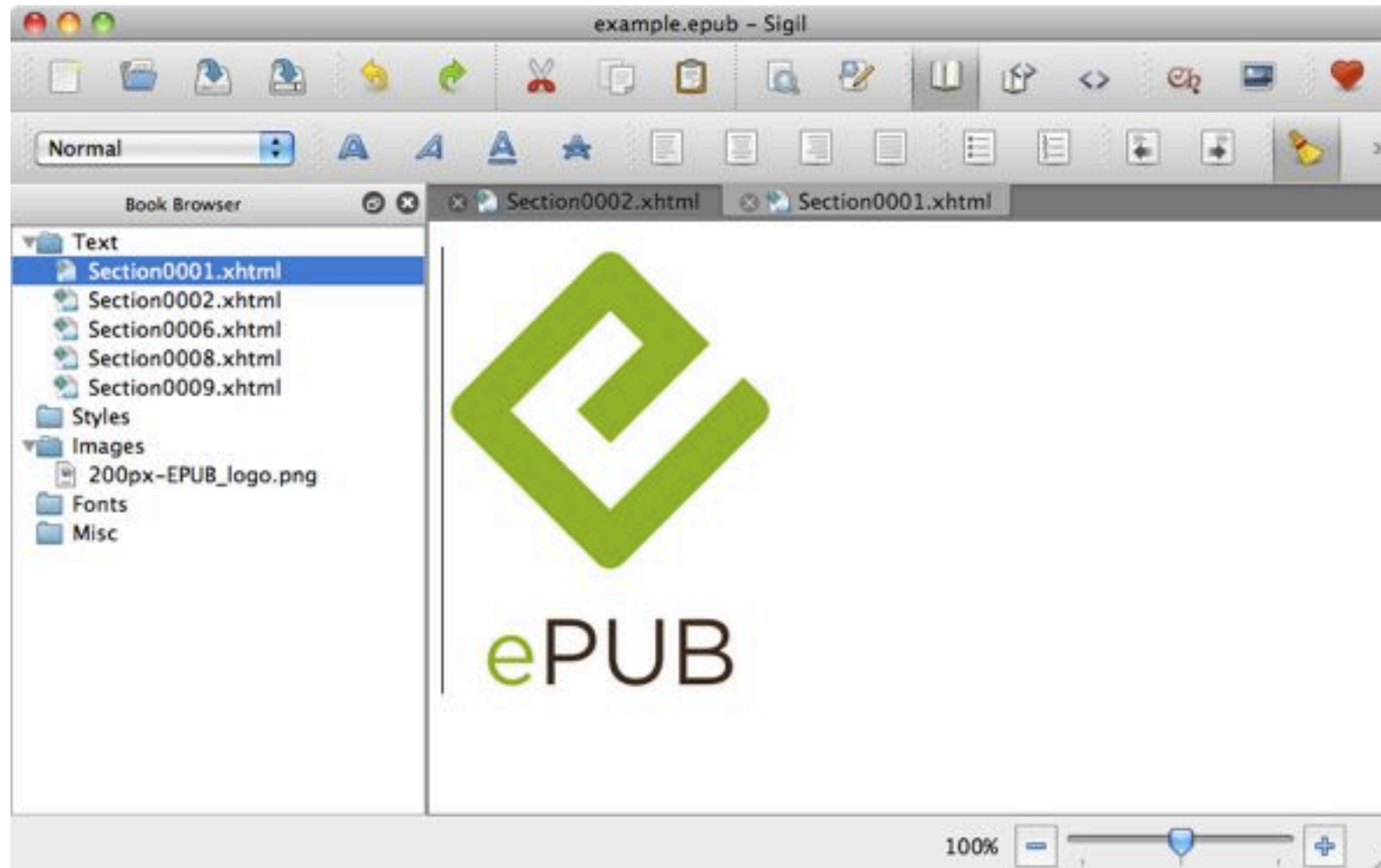
Open Container Format OCF



- OCF defines the structure of the ePUB file
- File is a ZIP file containing a file system (similar to e.g. jar files)
- META-INF folder:
 - container.xml: defines root directory and OPF file of publication
 - manifest.xml (optional), metadata.xml: required only for non-OPS contents
 - signatures.xml, encryption.xml, rights.xml (all optional): DRM info
- mimetype file:
 - Defines MIME type of file, usually "application/epub+zip"
- OEBPS folder:
 - Contains actual content files (XHTML)

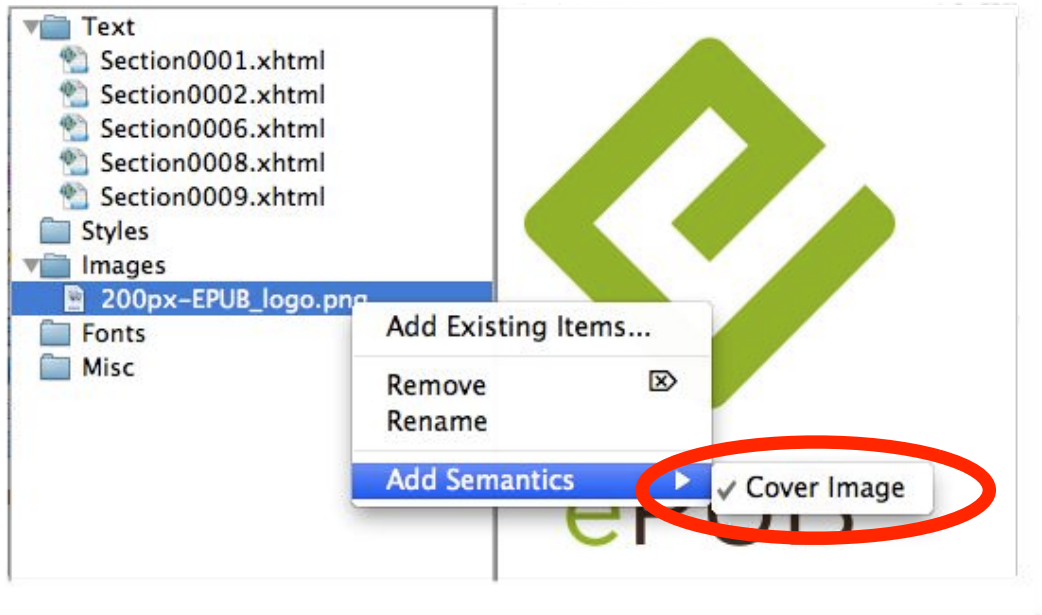
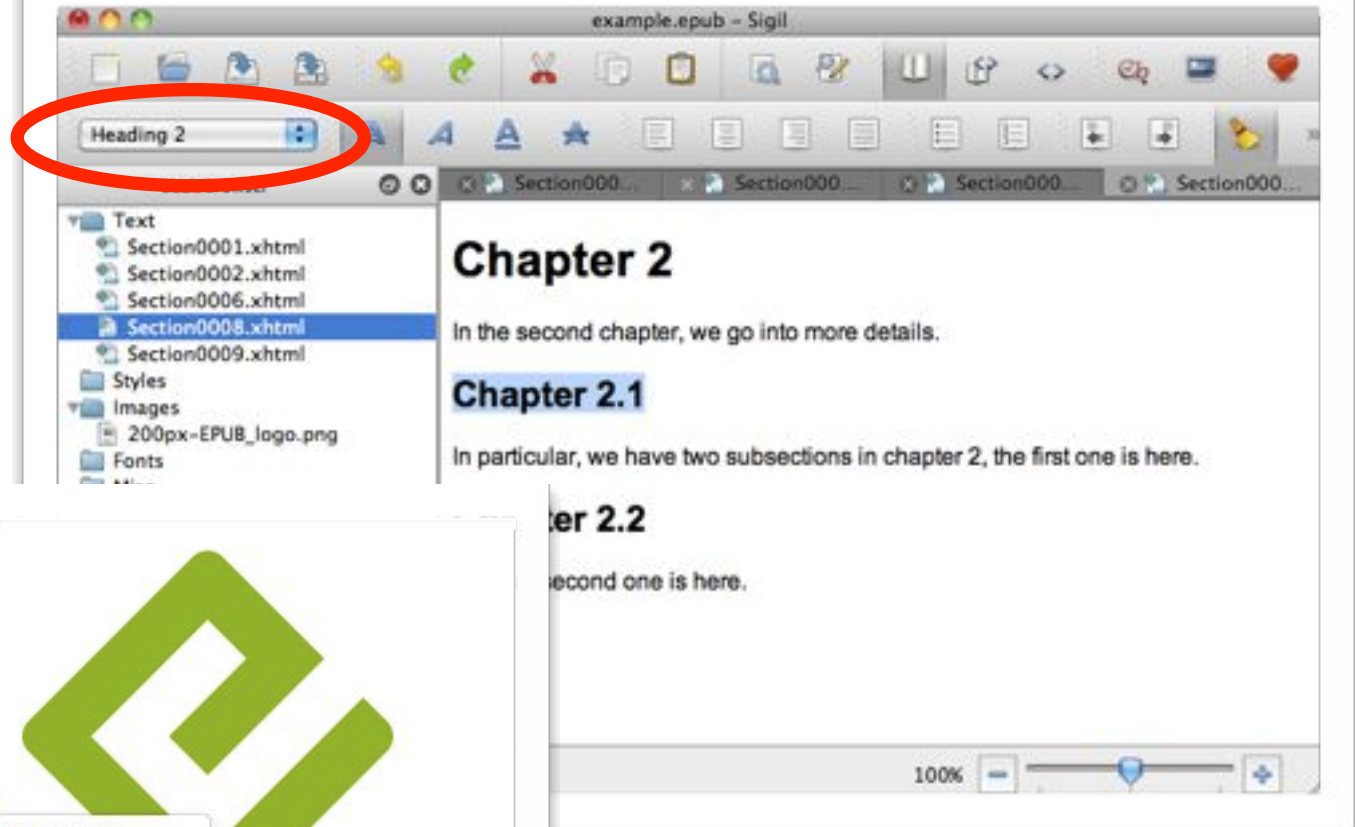
Example Using ePUB Editor (1)

- Tool used for the example: "Sigil" editor (<http://code.google.com/p/sigil/>)



Example Using ePUB Editor (2)

Declaring semantic attributes



Example Using ePUB Editor: Archive Structure

▼	Folder	META-INF	--	Folder
	File	container.xml	4 KB	Text document
	File	mimetype	4 KB	Document
▼	Folder	OEBS	--	Folder
	File	content.opf	4 KB	Document
▼	Folder	Images	--	Folder
	Image	200px-Epub_logo.png	8 KB	Portab...image
▼	Folder	Text	--	Folder
	Text	Section0001.xhtml	4 KB	TextW...ument
	Text	Section0002.xhtml	4 KB	TextW...ument
	Text	Section0006.xhtml	4 KB	TextW...ument
	Text	Section0008.xhtml	4 KB	TextW...ument
	Text	Section0009.xhtml	4 KB	TextW...ument
	File	toc.ncx	4 KB	Document

Example Using ePUB Editor: container.xml

```
<?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>
```

Example Using ePUB Editor: content.opf (1)

```
<?xml version="1.0" encoding="UTF-8"?>
<package xmlns="http://www.idpf.org/2007/opf" unique-identifier="BookID"
version="2.0">
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:opf="http://www.idpf.org/2007/opf">
    <dc:title>An example e-book</dc:title>
    <dc:creator opf:role="aut">Heinrich Hussmann</dc:creator>
    <dc:language>en</dc:language>
    <dc:identifier id="BookID" opf:scheme="UUID"> ... </dc:identifier>
    <meta name="cover" content="x200px-EPUB_logo.png"/>
    <meta name="Sigil version" content="0.3.4"/>
  </metadata>
  <manifest>
    <item id="ncx" href="toc.ncx" media-type="application/x-dtbnx+xml"/>
    <item id="x200px-EPUB_logo.png" href="Images/200px-EPUB_logo.png"
media-type="image/png"/>
    <item id="Section0001.xhtml" href="Text/Section0001.xhtml"
media-type="application/xhtml+xml"/>
    <item id="Section0002.xhtml" href="Text/Section0002.xhtml"
media-type="application/xhtml+xml"/>
    ...
  </manifest>
  ...
</package>
```

Example Using ePUB Editor: content.opf (2)

...

```
<spine toc="ncx">
  <itemref idref="Section0001.xhtml"/>
  <itemref idref="Section0002.xhtml"/>
  <itemref idref="Section0006.xhtml"/>
  <itemref idref="Section0008.xhtml"/>
  <itemref idref="Section0009.xhtml"/>
</spine>
<guide>
  <reference type="cover" title="Cover" href="Text/Section0001.xhtml"/>
</guide>
</package>
```

Example Using ePUB Editor: Content file

```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
  "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title></title>
</head>

<body>
  <h1 id="heading_id_2">Chapter 2</h1>

  <p>In the second chapter, we go into more details.</p>

  <h2 id="heading_id_3">Chapter 2.1</h2>

  <p>In particular, we have two subsections in chapter 2,
    the first one is here.</p>

  <h2 id="heading_id_4">Chapter 2.2</h2>

  <p>And the second one is here.</p>
</body>
</html>
```

Example Using ePUB Editor: toc.ncx

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ncx PUBLIC "-//NISO//DTD ncx 2005-1//EN"
    "http://www.daisy.org/z3986/2005/ncx-2005-1.dtd">

<ncx xmlns="http://www.daisy.org/z3986/2005/ncx/" version="2005-1">
  <head>
    <meta name="dtb:uid" content="47e865c1-81d2-42f6-ace6-de4953f3f222"/>
    <meta name="dtb:depth" content="2"/>
    <meta name="dtb:totalPageCount" content="0"/>
    <meta name="dtb:maxPageNumber" content="0"/>
  </head>
  <docTitle>
    <text>An example e-book</text>
  </docTitle>
  <navMap>
    <navPoint id="navPoint-1" playOrder="1">
      <navLabel>
        <text>Chapter 1</text>
      </navLabel>
      <content src="Text/Section0006.xhtml"/>
    </navPoint>
    <navPoint id="navPoint-2" playOrder="2">
      <navLabel>
        <text>Chapter 2</text>
      </navLabel>
      <content src="Text/Section0008.xhtml"/>
      <navPoint id="navPoint-3" playOrder="3">
        <navLabel>
          <text>Chapter 2.1</text>
        </navLabel>
        <content src="Text/Section0008.xhtml#heading_id_3"/>
      </navPoint>
    </navPoint>
  </navMap>
  ...
</ncx>
```


Example Viewed in E-Book Reader

Using *Adobe Digital Editions*:

