What is an eBook?

An eBook, or electronic book, is one which can be viewed using a reading system such as:

* a specialized device (Kindle, Nook, Kobo, …)
* software installed on a PC (Azardi, CoolReader, …)
* an app on a smart phone
* a plug-in for a web browser

An eBook could be a PDF file but more often refers to a specialized file containing the contents of a book. Besides text, it can contain images, audio, video, interactive features and apps. One major difference is that eBooks are not paginated like PDFs are. The text of an eBook is flowed as the user reads the book and the user can vary the point size or font if they want. Some eBooks, such as graphic novels, may limit, or even disallow, variations to their fixed format nature.

What is EPUB?

This paper will describe one eBook format, EPUB v2.0, which is an open source standard maintained by the industry consortium organization: the International Digital Publishing Forum (IDPF). Other eBook formats may be proprietary and all formats cannot be read on all eReaders. The example given here is ASI’s publication “Marketing Your Indexing Services”.

An EPUB eBook is at its simplest a single zip file:

MarketingYourIndexingServices.epub

While the file extension (in Windows) defines this file as an EPUB, it is really a zip file containing

* mimetype a file
* META-INF a directory
* OEBPS a directory

The mimetype file (MIME stands for Multipurpose Internet Mail Extensions) explicitly defines this zip archive as an EPUB and it contains a single line

application/epub+zip

The META-INF directory must be present. It contains the file

container.xml

This container file is XML that specifies the root file that describes the contents of the eBook. The container format is a standard called OCF (Open Container Format). In this example, the root file is specified as

OEBPS/content.opf

This file is found, not surprisingly, in the OEBPS folder. (OEBPS stands for Open eBook Publishing System.) Files in the OEBPS directory contain either the contents of the book’s components or information to be used by the reading system to render the contents for the user and discover features of the book that the reading system may use in special ways (for instance, which file is the Table of Contents or what is the order of the files in the book). Here is the complete list of files in the example EPUB’s OEBPS directory:

10-chap5.xhtml  
11-chap6.xhtml  
12-chap7.xhtml  
13-chap8.xhtml  
14-chap9.xhtml  
15-chap10.xhtml  
16-chap11.xhtml  
17-chap12.xhtml  
18-contributors.xhtml  
19-index.xhtml  
1-cover.xhtml  
2-titlepage.xhtml  
3-copyright.xhtml  
4-TOC.xhtml  
5-introduction.xhtml  
6-chapter.xhtml  
7-chap2.xhtml  
8-chap3.xhtml  
9-chap4.xhtml  
content.opf  
cover.jpg  
figure7-1.jpg  
styles.css  
title\_page.jpg  
toc.ncx

The reading system will look at the content.opf file, which is also XML, for

* the manifest element which lists the files that should be present (this is used to validate that the EPUB is complete and not corrupted), and describes, using MIME again, the type of each file, for instance
  + application/xhtml+xml
  + x-dtbncx+xml
  + image/jpeg
  + text/css
* the spine element which defines which file is the table of contents and the order in which the book’s content files should be presented, i.e., front matter files, chapters, index

Chapters of the book are usually type application/xhtml+xml because the text is tagged in XHTML according to the EPUB standard. They also may contain XML for some special text rendering such as math equations or chemical formulas. The file styles.css is type text/css which is a Cascading Style Sheet (CSS). CSS files tell the reading system how to display text using style attributes (font face, size, color, line spacing and indentation, etc.) for different paragraphs, titles, section headings, tables, etc.

The type image/jpeg is used for pictures or illustrations as are other graphic file formats. EPUB allows for many other file types used to add audio, video, interactive applications, etc.

Another standard employed in EPUB is DTB (Digital Talking Book) referenced by the type x-dtbncx+xml. This is used with the toc.ncx file. This XML file contains a navMap element that can be used by a reading system to provide the user with shortcut displays to navigate around the eBook.

Summary

In essence, an EPUB is very much like a web site in that it is comprised of a number of related files that the user can navigate through. It is a standard that is built on a lot of other standards while adding its own features into the mix. What makes it different from a web site is that a web site only contains a book’s contents with no information about its structure or discoverable features. It is the latter that allows reading systems to provide all the book’s features to a user.