

White Paper – Global eBook Standards and Applicability for Indian Language Publishing

GreatMinds Media Solutions Private Limited

Introduction

eBook is a term used used to describe an electronic or digital version of printed matter. The printed matter can be anything, such as Newspaper, Magazine, Book, Greeting Card etc. In the rest of this document the term eBook covers all content and is not restricted to Books alone.

An eBook is defined as a collection of written text , images and optional multimedia that can be read on one or all of the viewing platforms such as print , mobile phones , tablet PC's , e-Ink Readers , Desktop PC's , laptops , and regular internet browsers. Most of the formal eBook standards have an auto – reflow of text and images that ensures that the pages are automatically formatted to fit the screen of the device that it is read on.

eBooks have been around a while since the start of the computing age and people have been using eBooks in formats such as .txt, .doc, .pdf etc for some time and reading them on their computers. However, the real widespread use of eBooks came in after the invention of the eBook Reader device, based on the e-Ink technology revolutionized the world of Books.

In the following sections I would try to describe the various types of eBook standards that are prevalent today and the eBook standards that are evolving in 2011, as well as future standards to look for.

We would also speak about some of the most popular eBook reader platforms and devices in the market, across the world and in India. We would also see the trends in Indian language eBooks and the challenges that face Indian language publishers as they migrate to new digital platforms.

EBook standards

Some of the most popular eBook standards are described below. Please note that some of the eBook standards have been discontinued since 2010 and they are not mentioned. I have provided a link to Wikipedia website that talks about legacy standards.

Amazon Kindle Old Format

Format: Kindle

Published as: .azw

With the launch of the Kindle eBook reader, Amazon.com created the proprietary format, AZW. It is based on the Mobipocket standard, with its own Digital Rights Management (DRM) formatting. Because the eBooks bought on the Kindle are delivered over its wireless system called Whispernet, the user does not see the AZW files during the download process. The Kindle format is now available on a variety of platforms. Only supports Latin Languages and does not support embedded fonts.

Amazon Kindle Format – 8

Format: Kindle

Published as: .kf8 (to be confirmed)

Amazon recently announced the support of a new format called KF8 based on HTML5 and CSS3 .KF8 is the next generation file format for Kindle books – replacing Mobi 7 and is launched with the new Kindle Fire eBook tablet to be launched in November 2011. KF8 enables publishers to create great-looking books in categories that require rich formatting and design such as children’s picture books, comics & graphic novels, technical & engineering books and cookbooks. Kindle Format 8 replaces the Mobi format and adds over 150 new formatting capabilities, including fixed layouts, nested tables, callouts, sidebars and Scalable Vector Graphics, opening up more opportunities to create Kindle books that readers will love. Supports embedded fonts and would in future support non Latin languages.

EPUB

Format: IDPF/EPUB

Published as: .epub

The .epub or OEBPS format is an open standard for e-books created by the International Digital Publishing Forum (IDPF). It combines three IDPF open standards:

- Open Publication Structure (OPS) 2.0, which describes the content markup (either XHTML or Daisy DTBook)
- Open Packaging Format (OPF) 2.0, which describes the structure of an .epub in XML

- OEBPS Container Format (OCF) 1.0, which bundles files together (as a renamed ZIP file)

The EPUB format, as of 2011 is the most widely supported vendor-independent XML-based e-book format. The format can be read at least by the Kobo eReader, Apple's iBooks app running on iOS devices such as the iPhone, iPod Touch and iPad, Barnes and Noble Nook, Sony Reader, BeBook, Adobe Digital Editions, Lexcycle Stanza, BookGlutton, AZARDI, FBReader, Aldiko, Moon+ Reader and WordPlayer on Android, Freda on Windows Mobile and Windows Phone 7, and the Mozilla Firefox add-on EPUBReader. Several other desktop reader software programs are currently implementing support for the format, such as dotReader, FBReader, Mobipocket, uBook and Okular.

The only notable device lacking integrated support for the EPUB format is the Amazon Kindle which does not support .epub.

Adobe Digital Editions uses .epub format for its e-books, with DRM protection provided through their proprietary ADEPT mechanism. The recently developed ADEPT framework and scripts have been reverse-engineered to circumvent this DRM system.

The .epub standard is the most popular eBook standard for non Latin languages including Indian languages as it supports embedded fonts like PDF files.

Mobipocket

Format: Mobipocket

Published as: .prc; .mobi

The Mobipocket e-book format is based on the Open eBook standard using XHTML and can include JavaScript and frames. It also supports native SQL queries to be used with embedded databases. There is a corresponding e-book reader. A free e-book of the German Wikipedia has been published in Mobipocket format.

The Amazon Kindle's AZW format is basically just the Mobipocket format with a slightly different serial number scheme (it uses an asterisk instead of a Dollar sign), and .prc

publications can be read directly on the Kindle. The Kindle AZW format also lacks some Mobipocket features such as JavaScript.

Mobipocket has developed an .epub to .mobi converter called KindleGen (supports IDPF 1.0 and IDPF 2.0 epub format, according to the company).

Portable Document Format

Format: Adobe Portable Document Format

Published as: .pdf

A file format created by Adobe Systems, initially to provide a standard form for storing printable documents containing a set of page images. The format derives from PostScript, but without language features like loops, and with added support for features like compression, passwords and DRM. Because PDF documents can easily be viewed and printed by users on a variety of computer platforms, they are very common on the World Wide Web. The specification of the format is available without charge from Adobe.

Multiple products support creating and tagging PDF files, such as Adobe Acrobat, PDFCreator, OpenOffice.org, iText, and FOP, and several programming libraries. Third party viewers such as xpdf are also available. Mac OS X has built-in PDF support, both for creation as part of the printing system and for display using the built-in Preview application.

PDF files are supported on many e-book readers including: Mobipocket, iRex iLiad, iRex DR1000, Sony Reader, Bookeen Cybook, Foxit eSlick, Amazon Kindle (2, 3, International & DX)], Barnes & Noble Nook[citation needed], the iPad, PocketBook Reader, Bebook Neo and the Kobo eReader. Also, pdf files can be read on the iPod Touch using the free Stanza app. On iPad, there are also many readers as Stanza, iBook, FlipReader, etc.

Scientific papers and books are often found in PDF format. In this case, the problem of showing tables and equations correctly has been dealt with, when the PDF was created. When trying to re-flow such text, ereaders typically obtain very poor results.

PDF supports embedded fonts and all languages. The main disadvantage of PDF format is the lack of proper text reflow as mentioned above, making it not very reader friendly on small screen mobile devices.

Plain Text files

Format: Plain Text

Published as: .txt

eBooks that are distributed in plain text exist and are used for free distribution. These files are in all languages and do not have text formatting, images, table of contents etc.

Flipping eBooks - Flash

Format: Flash eBooks

Published as: .swf, .exe

A Flash-eBook magazine, catalogue, or book is a book with which users can flip pages naturally just like they would do with a real book. By dragging any of the 4 page corners to the opposite location, appropriate pages are revealed. This is a new intelligent way of interacting with information which will definitely impress your visitors. It is often times burdensome to attach a magazine in the form of a PDF document or image files on a website since the user will have to download a large file size although the user may only be interested in reading specific pages. Through such hosted flash eBooks, smaller optimized versions of the pages are displayed, leaving the user to flip through desired pages and choosing to reveal the page in high resolution.

Within the pages of a Flash-eBook, you may have text, images, video, animation, or even flash-based interactive applications (e.g. children electronic coloring book). Certain media can be selected per page. A Flash-eBook can then be incorporated into your website, or embedded on a CD. A flash eBook is recommended for online reading or distribution via CD's of rich content.

Flash eBooks support all languages as fonts are embedded in the content.

One of the disadvantages of Flash eBooks is that they cannot be read on the Apple iPhone and iPad. However they can be read on Mac PC's.

EPUB3 or Enhanced EPUB

Format: EPUB3

Published as: .epub,.epub3

The new EPUB3 is a major revision to the current standard and with a move to HTML5 based content, now provides support for key emerging technologies including video, audio, scripting, interactivity, styling and layout enhancements, vertical writing, improved accessibility and MathML support.

The standard is being finalized in November 2011 and would be released for distribution by end of the year. The EPUB3 standard adds the following features to EPUB:

- HTML5 & CSS support
- Rich Media and Interactivity
- Layout Improvements
- Global Languages
- Accessibility Improvements

Fixed Layout ePub

Format: EPUB

Published as: .epub

A fixed layout book, sometimes called an “illustrated book”, has a pixel specific page size that you can use to precisely control the layout of the objects on the pages. You can embed fonts, choose particular text sizes, and precisely position images. You can even create a two-page spread in which a single image spans both pages. A fixed layout eBook is normally used for illustrated children's books , Travel Books with colourful pages , Cook Books etc.

These books can be zoomed but it is like zooming a picture. The text does not reflow and you will have to pan around to see the page. The books are primarily images with text overlays and in this fashion they resemble PDFs however they are peculiarly unique in that they begin as ePub style documents.

The fixed Layout ePub is supported only on iPad & iPhone and a few other devices.

HTML5 eBooks

Format: HTML5

Published as: .html

HTML5 eBooks are the new generation of eBooks, and is being adopted by worldwide standards such as EPUB3, Amazon etc. HTML5 brings the power of interactivity, animations, multimedia etc with Flash-like performance on a regular HTML browser and can also integrate security. HTML5 eBooks can be read on regular browsers. HTML5 eBooks cannot be read on regular eBook readers, however they can be read on PC's and smartphones/tablets running on Apple OS and Android platforms.

HTML5 is an open standard and is more used for rich, animated books that can be built into a flipping book and distributed via hosted websites or through CD's.

HTML5 is language agnostic and supports embedded fonts like other HTML files.

Legacy standards

Some of the text in this whitepaper is taken from Wikipedia's site comparing eBook standards. The various standards and the legacy standards can be viewed in the link given below:

http://en.wikipedia.org/wiki/Comparison_of_e-book_formats

eBook Creation & Delivery

eBook Creation Software

Due to the wide variety of standards for eBooks and ePublications, there are quite a large number of software suites and workflow solutions to create these Book formats. A summary of the main software solutions are given below:

1. Adobe Digital Publishing Suite - Licensed
2. Adobe Indesign (included in the Publishing Suite which is a workflow software for publishing companies that want to unify their print and digital publication creation process) - Licensed
3. Quark Xpress Digital Publishing Suite - Licensed
4. Calibre eBook creation suite – Open Source
5. Librios Digital Publishing Suite – Licensed
6. KindleGen Software – Creator for Amazon Books , Licensed
7. Baker framework – Open Source HTML5 eBook creation Software
8. Laker Framework – Open Source HTML5 eBook creation software
9. Circularflo Fixed Layout ePub Software – Licensed
10. Flipping Book creation software – From many sources
11. Adobe Acrobat Professional – Licensed
12. Proprietary eBook Creation Solutions

eBook Reader Software

There are a wide variety of eBook reading software, almost all of them being freely available for download on all platforms. There are also web based eBook readers that can be used from Mobile devices.

A summary of the eBook reader software available is given below:

1. Adobe Digital Editions – PDF , ePUB , ePub3 (Beta) : PC, Linux and Mac
2. Calibre Reader - All formats : PC, Linux and Mac
3. Apple iBooks Reader – ePub , fixed layout ePub Mac
4. Sony , Kindle , Nook Digital reader Software : PC , Mac and Linux
5. Ibis HTML5 eBook Reader
6. Google Web Reader
7. Amazon Cloud Reader
8. MagicScroll HTML5 Web eBook Reader
9. Adobe/Macromedia Flash Renderer Plug-in for Flash eBooks
10. Adobe Acrobat Reader

eBook Reader Devices

An e-book reader, also called an e-book device or e-reader, is a portable electronic device that is designed primarily for the purpose of reading digital books and periodicals.

An e-book reader is similar in form to a tablet computer. A tablet computer typically has a faster screen capable of higher refresh rates which makes them more suitable for interaction. The main advantages of e-book readers are better readability of their screens especially in bright sunlight and longer battery life. This is achieved by using electronic paper (e-Paper) or electronic Ink (e-Ink) technology to display content to readers.

Any device that can display text on a screen can act as an e-book reader, but without the advantages of the e-paper technology.

Some of the most popular eBook Reading platforms are given below:

1. Amazon Kindle – based on e-Ink technology
2. Amazon Kindle Fire tablet – based on Android OS
3. Barnes & Noble Nook Tablet & Nook Readers – Android OS and e-Ink based versions are available
4. Sony Reader – based on e-Ink technology
5. Wink Reader – India focussed device supporting Indian language eBooks – developed by EC Media , a division of DC Books
6. Infibeam Pi – India focussed device supporting Indian language eBooks
7. Akash Tablet – Developed in partnership with Government of India , Android based tablet
8. Apple iPad tablet
9. Samsung Papyrus e-Ink Reader
10. Nile Reader
11. Kobo e-Ink and tablet
12. OEM eBook Readers (BOOX, Hanlin are examples). The Indian eBook readers mentioned above are actually rebadged OEM eBook readers.

eBook Security

eBook security is a large topic of debate in the ePublishing industry. Most of the Digital Rights Management Software (DRM) used by the Industry leads have been compromised by freely available software, though DRM is still quite prevalent in the industry.

The most commonly used system of securing eBooks is by tying up an eBook that is purchased to a specific eBook reader device , a specific PC or a specific Reader application. This is done via DRM software such as Adobe DRM, Amazon Kindle DRM, Apple DRM , DNL DRM etc.

In this case the eBook is encrypted with the buyer credentials, credit card details, device details or login details and the Book will open only view a web login interface or by device authentication or Application authentication. A re-authentication can be done if a device fails, Hard drive crashes etc by re-entering purchase information and personal details.

There are also many eBook stores that sell DRM free eBooks (www.smashwords.com is an example). In addition, simple methods as password protection , watermarking an eBook with personal details of the owner etc are done in some cases.

Going forward with HTML5, open standards for eBooks coming up, GreatMindsTM feels that DRM through device authentication, App based authentication (eBook only opens in your reader's App screen) and Web based authentication of books would be the way to go. In addition web browser based eBook reading (Cloud eBooks) would be in the long run an approach that will continue to grow.

Indian eBook Scene

The Indian eBook publishing industry is relatively nascent and is currently dominated by a few players such as EC Media, ChilliBreeze, Cinnamon Teal, etc. In addition some of the major Book publishers in India such as India Book House, Penguin etc are working on eBook delivery platforms for their books. In addition it is rumoured that Flipkart.com would be soon launching an eBook store to complement their Book offerings.

Most of the eBooks in India are read on Mobile phones and laptops , with a small percentage being read on tablet PCs such as iPad , Galaxy Tab etc. There is also a section of travellers who read on the Amazon Kindle , though these are limited to English language eBooks and Indian language PDF file books.

The Indian ePublishing market is estimated to be 0.005% of the total publishing Industry in India. However, this is expected to grow to 30% of the market in the next 10 years according to various market reports. GreatMindsTM would soon be coming out with a market report on the Indian ePublishing Industry that would be available to a target audience for review.

In terms of regional language eBooks, the market is very small though is growing at a steady pace. The Books that are selling well today are adult fiction, Children's book and technical books along with eComics and travelogues.

Regional language eBooks are mostly read in the ePub format, though PDF versions and Secure Flash or HTML5 flipping versions read on a hosted website or distributed as an executable in memory card or CD/DVD is widespread as well.

References

<http://idpf.org/epub>

<http://www.amazon.com/gp/feature.html?docId=1000729511&tag=vglnk-c4-20>

<http://www.20thingsilearned.com/en-GB>

<http://www.slideshare.net/smlowe/ebook-standards>

<http://ebook-reader-review.toptenreviews.com/>

http://en.wikipedia.org/wiki/Comparison_of_e-book_readers

<http://www.thewinkstore.com/>

<http://www.infibeam.com/eBooks/>

<http://www.dlib.org/dlib/january01/mooney/mooney-notes.html#ref4>

<http://copyrightandtechnology.com/2009/02/04/76/>