

Mobilizing your content

How the mobile hype affects the work of technical communicators

Finding the right TMS - how hard can it be?

Which Translation Management System is most suitable for your business?

Choosing the right words

An introduction to Search Engine Optimization and why it matters in technical communication

Is there an ePub in your future?

In April of 2011, the Association of American Publishers stated that eBooks were the Number 1 format among all trade categories for the month. If you have been watching the news, you are likely to see frequent reports of eBooks outselling printed books. Among all dedicated readers and applications, ePub is the most widely accepted eBook format. Let's take a closer look at what an ePub is and how it might help you produce more customer-oriented, cost-efficient documentation.

By Scott Prentice

If you want to make your documentation available on virtually all devices and platforms, an ePub may be the easiest way to provide a usable solution. By default you'll get a table of contents and full text search. Additionally, with most reader applications you'll also have the ability to place bookmarks and customize the user interface. This is almost equal to the level of functionality that you'll see in more popular Help systems. Granted, this may not be a replacement for online Help systems like HTML Help or WebHelp, but depending on your requirements it may be all that you need.



Photo: George Tsartsianidis

Some other benefits offered by eBooks are:

- Instant gratification - your customers can download and read your documentation right away without any complicated installation process.
- Lower cost - as compared to traditional physical books, eBooks are considerably cheaper to produce.
- Makes your documentation, "ultra portable" - your documentation will be available for use on any device at any time that it's needed.
- Nice for books with a "limited lifespan" - certain types of documentation can become out of date rather quickly. Rather than wasting the resources to produce books that will just be disposed of when the new version comes out, eBooks can provide a more viable alternative to print.

What is an ePub?

An "ePub" is a file format that defines an electronic publication or an eBook. It's a format which is a collection of files that specify the content, organization, and formatting of an eBook, conceptually similar to HTML Help (a CHM file). To view an eBook requires a reader application on your computer or a dedicated reader device. The ePub format is just one of many eBook formats, although it is by far the most widely used and accepted format. Technically this format name is cased "EPUB," but "ePub" seems easier on the eyes, so that is how you will see it in this article.

The ePub specification is maintained by IDPF (International Digital Publishing Forum). ePub 2.0 became an official standard in September of 2007, which superseded the older Open eBook standard (OEB) from 1999. ePub 2.0.1 was approved in May 2010 and, as of this writing, is still the current stable release. The ePub 3.0 first public draft was released in February 2011, and published as a IDPF Proposed Specification in May 2011. ePub 3.0 is scheduled for completion later in 2011. For the latest information, visit their web site at www.idpf.org.

The ePub specification is a combination of the following specifications:

- Open Publication Structure (OPS) - defines the standard for representing the content of electronic publications.
- Open Packaging Format (OPF) - defines the structure and semantics as well as the mecha-

nism by which the various components of an OPS publication are related.

- Open Container Format (OCF) - defines the mechanism by which all components of an electronic publication are packaged into a single deliverable.

Underlying technologies

The underlying technologies of the ePub format are XML, XHTML, and CSS. All of these are formats you are likely to be familiar with. The ePub format supports images, standard formatting, and linking. However, do keep in mind that there is often a disconnect between the features defined by a specification and how those features are supported by "real world" applications. This is very much the case with the ePub format and eBook reader applications.

Content in an eBook is intended to be re-flowable to fit the constraints of the rendering device or application. This is similar to the way simple content in a web browser will adjust to fit the width of the application window. Most eBook readers work on a screen-based "paged" format, meaning that the number of "pages" in the book will depend on the size of the device or application window and the size of the font. Typically the font size as well as color (both character and background) is user-definable. Because of this, you can never expect that a passage will be on the same page, in the same eBook, in different readers.

The content in an ePub may be "open" or locked through the use of DRM (Digital Rights Management). An open ePub can be copied to various devices and shared with others. It can also be converted into other formats if necessary. However, just because an ePub is unlocked, does not mean it is okay to share it with others; be sure to honor the copyright requests of the publisher. An ePub that is locked through the use of DRM cannot be copied or converted and will often be tied to a specific reader device.

What does an ePub look like?

As with most things, the answer to that question is "it depends." Fundamentally, an eBook reader application is just a web browser, so in general, ePub content will look a lot like a simple web

page. However, each reader and application may render the same ePub quite differently. If you're familiar with the problems that arise when trying to get the same web page to look good on all of the popular web browsers, this is a similar problem, but worse, because there are many more eBook reader applications to deal with.

To illustrate some of the issues that you may run into, let's review the rendering of the same "page" on different devices and applications.

Tablet and dedicated readers

Tablet and dedicated reader devices are a popular way to read eBooks. Let's compare the following three different devices:

- Apple iPad 1 (running the iBooks application)
- Amazon Kindle 3
- Sony PRS-600

Sample topic - cross-references

This topic shows a variety of cross-referencing methods, and the rendered results of those cross-references.

This is a paragraph in a topic, with an id attribute of `sample_xref_para`.

iPad/iBooks - inline formatting sample

Sample topic - cross-references

This topic shows a variety of cross-referencing methods, and the rendered results of those cross-references.

This is a paragraph in a topic, with an id attribute of `sample_xref_para`.

Kindle - inline formatting sample

Sample topic - cross-references

This topic shows a variety of cross-referencing methods, and the rendered results of those cross-references.

This is a paragraph in a topic, with an id attribute of `sample_xref_para`.

Sony Reader - inline formatting sample

The page we'll be discussing is titled "Sample topic cross-references" from the "DITA Style Guide" by Tony Self, published by Scriptorium Press. In general, the rendering of this page looks reasonably consistent on all three devices, but if you look carefully, you'll see a number of variations.

First, because the iPad has a color screen and the others are black and white (or grey tones), you'll see the headings and other text objects in color on the iPad. This is to be expected. If you look a little closer, you'll see that in the second paragraph, there is some text with inline formatting. On the iPad, only the "sample_xref_para" is formatted in a monospaced font, but on the other two devices the word "id" is formatted in addition to "sample_xref_para". It would appear that the underlying styling has been applied differently and on some devices that styling is honored, but not on others.

Elsewhere on this page is an image and image caption, which is rendered in a consistent manner on all devices. But when we look at the table that follows the image, we see that the Kindle has put the table title in an odd single cell where the iPad has rendered it in a more traditional manner.

Sample Table of photographic credits

Photo Description	Photographer
Charles Darwin	Julia Margaret Cameron

iPad/iBooks - table sample

Sample Table of photographic credits	
Photo Description	Photographer
Charles Darwin	Julia Margaret Cameron

Kindle - table sample

If we take a look at another page, one that contains code samples, we see more disparity. On the iPad, the indenting of the code sample is honored, and it wraps lines that are too long. On the Kindle, the indenting is not honored (all lines are left justified), and long lines wrap. While on the Sony, the indenting is honored, but long lines do not wrap, causing that content to be lost.

Mobile phones

While a mobile phone may not be the ideal device for reading an eBook, it can be convenient at times. The situation on mobile phones is similar to what we have seen on tablets. We will compare the following phones and applications:

- iPhone (1G) (running the Stanza application)
- Nexus / Android (running the Aldiko application)

As you'd expect, the headings and fonts will vary based on the user's settings, but interestingly, on Aldiko, we see that the first line in each paragraph is indented (other devices don't do this). As seen before, the text "id" is formatted on Stanza, but not on Aldiko while "sample_xref_para" is formatted on both.

Code samples on a phone can be a bit of a challenge due to the very small screen. In the samples reviewed, we see that both readers honor the indenting, and Aldiko wraps the long lines while Stanza does not (again causing that content to be lost).

Desktop reader applications

There are a number of eBook reader applications available for the desktop. We compare four of them here.

- Calibre - Mac
- Adobe Digital Editions - Mac
- Firefox, EPUBReader plugin
- FBReader - Windows

Three of these applications (Calibre, Digital Editions, and EPUBReader) render our page in a fairly consistent manner. The headings are similar, the inline formatting is handled the same way, and the image and figure title look the same. Digital Editions does seem to format the table title and headings a little differently, but with no serious problems. FBReader renders this page quite differently than the others. The "id" text isn't formatted, the "sample_xref_para" is italicized (all others render this in a monospaced font), and the paragraphs are indented (like Aldiko). FBReader also centers the image and doesn't even make an attempt at rendering the table (each cell displays as separate paragraphs).

Caveats?

Yes, there are some potential drawbacks with eBooks. One issue is the fact that the "paged" format may not be well suited to some types of documentation. However, assuming that your

Sample topic - cross-references

This topic shows a variety of cross-referencing methods, and the rendered results of those cross-references.

This is a paragraph in a topic, with an id attribute of sample_xref_para.

iPhone/Stanza - inline formatting sample

Sample topic - cross-references

This topic shows a variety of cross-referencing methods, and the rendered results of those cross-references.

This is a paragraph in a topic, with an id attribute of sample_xref_para.

Android/Aldiko - inline formatting sample

content works well in a printed book, it should work equally well in an eBook format.

At this time, eBooks have no capability for context-sensitivity with an associated software program you may be documenting. This is often a feature of more traditional online Help formats, but those formats are typically able to receive direct communication from the related program. This communication pathway does not typically exist for eBooks, especially since the documentation (ePub) is likely to be running on a different device from the software program. This capability may become a reality in the future.

As noted, the formatting will vary on each reader and device. For this reason, it is important to keep your formatting as simple as possible. In fact, some of the best looking eBooks are those that apply the least amount of formatting. If your content makes extensive use of tables, especially wide tables with more than two columns, you may not want to consider eBooks. Tables on a small screen are typically rendered quite poorly, and in many cases the data in multiple columns is lost to the reader. Hopefully this limitation will be addressed in future reader applications by allowing horizontal scrolling.

Additionally, some reader applications don't support linking. Although most do, you cannot be guaranteed that your links are available to all readers, so you may want to consider keeping links to a minimum.

The ePub specification does not natively support an Index. This would be the perfect companion to a table of contents (which is supported), and will hopefully be supported in future versions. For now if you want an index you will need to craft one by creating multiple pages of links to the indexed content. If done properly, this can work quite nicely (as seen in the "DITA Style Guide") but it does require a bit of extra work.

All reader applications don't support the same level of the ePub specification, and few are totally compliant. This is the drawback of providing content for a large number of different devices and applications. Even though your ePub content may adhere to the specification, you have no way of knowing what type of reader application is being used. Again, the best option is to keep your layout and formatting as simple as possible.

How do you make an ePub?

Because an ePub is just a collection of XHTML, CSS, and XML files, you might consider creating it "by hand," but using a tool is far more efficient and error-free. Your current authoring tool may export to ePub, if not, there are many conversion utilities available. As with the reader applications, tools for creating an ePub will vary in their support of ePub specification; try as many tools as possible before choosing one. It is a good idea to spend as much time as possible testing the conversion process. Be sure to have a document that includes a representative sampling of all of the "features" in your documentation. This doesn't have to be a real document, just one that will expose any potential weakness of the process or resulting file.

The easiest method for conversion is to save to HTML, then use one of the conversion tools to convert from HTML to ePub (many are free). Do keep in mind that you want your HTML to be as free from needless styling and formatting as possible. This will result in the most usable and best looking eBook.

There are many authoring tools that export to ePub as well as a large number of conversion tools, both desktop applications and online converters. The tools are constantly changing, so be sure to spend a little time searching on the Web for what is currently available.

Desktop authoring tools

The following authoring tools export to ePub:

- Adobe RoboHelp - (USD 1000), Windows
- Adobe InDesign - (USD 700), Windows/Mac
- Apple iWork Pages - (USD 70), Mac
- eCub - (free), Windows, Mac, Linux, FreeBSD, Solaris
- Jutoh - (USD 40), Windows, Mac, Linux, FreeBSD, Solaris
- Atlantis Word Processor - (USD 35), Windows
- Sigil - (free), Windows, Mac, Linux

Desktop conversion tools

The following tools convert from one or more formats into ePub:

- Calibre - (free), Windows, Mac, Linux
- DITA Open Toolkit + DITA for Publishers plugin - (free)
- eEscape ODT2ePub converter - (free), Windows, Linux
- Pincette ODT to ePub - (USD 55), Windows, Mac, Linux (Java)
- DNAML PDFtoePub - (USD 40), Windows.
- epub-tools (code.google.com) - (free), Windows, Mac, Linux
- epubcheck (code.google.com) - (free), Windows, Mac, Linux

Online conversion tools

There are many, many online conversion tools available, the following list provides a few options:

- Feedbooks (www.feedbooks.com)
 - author and develop eBook content directly in website
- EasyEPUB (www.easypress.com)
 - create ePubs from InDesign, Quark or MS Word files
- 2EPUB (www.2epub.com)
 - convert PDF, DOC, ODT, HTML, and eBook formats to EPUB, MOBI, and others
- Epub2Go (www.epub2go.com)
 - free PDF to ePub converter

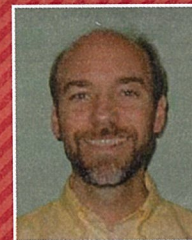
I've used the 2EPUB tool to convert HTML-based specifications (those found on web sites like the W3C and other organizations) into ePub

files for my phone and tablet. Perhaps there's a specification or document on the Web that you refer to frequently? This is a great way to produce a very usable eBook from that content.

So, is there an ePub in your future?

eBooks are definitely here to stay. If you produce documentation of any kind, it's worth considering how your customers would benefit from having your documents available as eBooks. If you already produce HTML-based documentation, taking the step to offering that as an ePub is not a huge effort. Just remember to keep it simple. The best looking and most usable eBooks are those that apply the least amount of formatting.

contact



Scott Prentice is the president of Leximation, Inc., based in San Rafael, California, USA. He provides consulting services for the development of custom online Help systems

and FrameMaker applications. He has been involved with DITA for many years and created the DITA-FMx plugin for FrameMaker.

www.leximation.com
scott@leximation.com

Stuttgart, 04.10.2011



Scott Prentice
Leximation, Inc.
122 H Street
San Rafael, CA 94901
USA

tcworld GmbH,
Rotebühlstraße 64
70178, Stuttgart
Germany

Copies of the tcworld magazine conference edition 2011

Dear Mr. Prentice,

As a gesture of gratitude, please find enclosed 3 copies of the tcworld magazine conference edition 2011.

Thanking you very much for your contribution,

Best regards,

tcworld GmbH
Rotebühlstraße 67
70178 Stuttgart