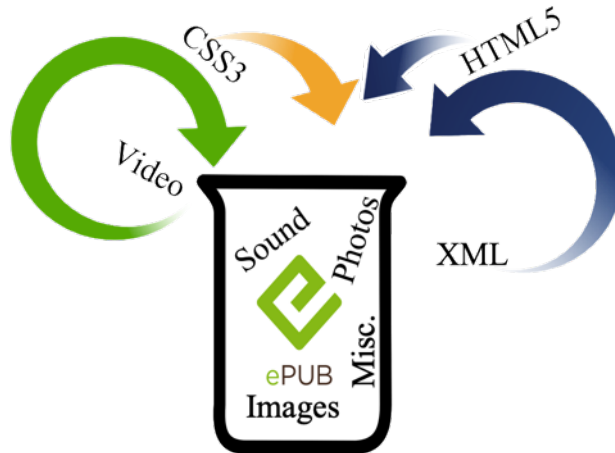


What is an EPUB file?

To simply put it, the .epub format is a container. Much like a .zip and a .tar.gz file. It's used to hold all of the assets that make up your EPUB3 manuscript. Every file format we use is a container; .docx, .doc, .txt, .pdf, .rtf, .mobi, .exe, .dmg, .rpm and so forth.



Everything that is used in your EPUB3 manuscript goes into this container. This keeps all of your assets in one neat place. By validating that the correct EPUB standard has been applied, is what's known as making the manuscript "Print Ready".

For example, before I send this manuscript to my printer, I must make it print ready according to my printer's requirements. This has to do largely with the type of printing press and software used by my printer. I have to send the photos, images, fonts, page layout designs and any other assets used to make my manuscript print the way I designed it.

If a photo or a special font is missing, I will waste valuable time and resources. The printer will have to stop working on my manuscript, until I send over the missing assets. This can delay a project, by days, weeks or even months. It depends on how busy the printer is. My project is sideline and requeued once the printer has all of my assets. This also cost me additional money. I have to pay for the time the printer wasted on my project.

This is why the validation process for EPUB3 is so vital. Your manuscript can't go to the distribution channels until your manuscript meets EPUB3 validation standards. If it doesn't meet standards, your project is side lined, until it passes validation. Some agitators will validate your manuscript for you, for a hefty fee.

An agitator is the go between for the publisher and the digital distribution chain. You can bypass the agitator, but you will have to deal individually with each distribution channel. We will go deeper into this when we get to distribution channels.

EPUB3 Standards

We touched a little bit on EPUB standards. Now we are going to take a closer look into them. Before we dive into it, let's take a look as to why EPUB Industry Standards were developed.

Before EPUB standards, it was the wild west for digital publishers and a nightmare for the consumer. Everyone had their own format.

1. **PDF** – Adobe® Portable Document Format (pdf), was once believed to be the crown prince apparent. It's still widely used and dominates when it comes to digital documents, which need to look and print exactly as the traditional print format. This format is still the preferred format for legal business documents. The problem with developing eBooks or eZines (Electronic Magazines), is scalability. PDF documents are not able to adjust to changing screen sizes very well. You can scale the height or width in some cases, but the overall scaling stayed relational. This means it retained its original size. An 8in X 11.5in document retained the 8:11.5 relationship, regardless of the device or the angle the device was being held at. This makes reading a pdf hard, without having to zoom in and out constantly.
2. **RFT** – Rich Text Format. Still compatible with every eBook reader on the market today. Unlike the simple text (txt) format it allows the use and retention of HTML tags. It's scalable. It can wrap its text to fit screen size. The problem is that the author had little to zero control over the unwanted editing of their published manuscript. Anyone with a basic word processor could rewrite the manuscript. You may be asking, what is the likelihood of anyone doing that? The answer is zero to none. But think about the work a writer puts into their manuscript. There is the research, validation of facts, editing, setting goals to write a certain number of words per day (1,000 to 3,000 word per day), formatting and making your manuscript print ready. This format never really caught on with novels, nonfiction and magazines.
3. **LIT** – Microsoft® Literature Format. This is a proprietary format created by Microsoft® for use with Microsoft® eBook readers and apps. It can be encrypted to prevent tampering and restrict sharing between devices. When we get to publishing, I will go deeper into encrypting your manuscript, the pros and cons.
4. **FB2** – Fiction Book version 2.x. Originally an eBook format developed in Russia. Now it's primarily used to store eBooks. It can easily be converted into various formats while retaining the metadata.
5. **DJVU**- A digital document format used today by the scientific community to digitalize large scientific documents. It's similar to pdf but has 10 times the compression ratio. You can find out more about this format by going to: <http://www.djvu.org/>. Never really caught on outside of the scientific community. But is still widely in use.
6. **LRS** - LRF and LRX were developed by Sony and Cannon for use on the Broad Band eBook readers and the Sony Librie. These formats are no longer in use. Sony now uses the EPUB format. LRS is available to the public, free. LRF and LRX remain proprietary formats. Why release one format and not all three? It's easier to track and update one format than trying to track and update three formats. Those are resources Sony can't afford to waste. By making it Open Source, the open source community can handle any updates.
7. **IBA** – A proprietary format for the Apple iBook Reader®. IBA actually stands for (Image Format Unknown). Apple® takes trade secrets extremely seriously. Sometimes a little

too serious. By looking at the extension, I would have guessed something like (iBook Approved). It's the proprietary format for the Apple iBook®.

8. **AZW3** – Amazon Kindle Fire® Readers format. AZW is the format used by the first-generation Amazon Kindle® Readers. It's Amazon's® proprietary format.
9. **MOBI** – Also used by Amazon Kindle® Readers. Originally developed by a French company, it was bought out by Amazon®. MOBI is sometimes referred to as KF7 and KF8. The mobi format was on track to becoming the Industry Standard, before EPUB emerged victorious. I used the mobi format almost exclusively in some of my earlier works. Imagine if a private company was the gate keeper to Industry Standards. Would there be a licensing fee, to develop digital documents outside of their platform? I'm not suggesting that Amazon® or anyone else for that matter, had any grand design on how the world purchase and consume product online (digital and tangible). I'm just speculating. For additional information about the mobi format go to:
<https://www.mobilereader.com/>.
10. **KOBO** – Not truly a format. Kobo is an eBook reader produced Kobo Inc., a Canadian based corporation. It can read almost every nonproprietary format. Barnes & Noble sell the Kobo readers in the United States.
11. **CBZ** – Comic Book formats. Other formats are; CBR, CB7, CBT and CBA. As you see there are no clear forerunners at this time. EPUB3 can easily handle comic books. However, the Big Boys on the block such as Marvel Comics® publish in CBZ or CBR. So, if you are a comic book distributor you will want one of these formats. These are primarily nothing more than your basic zip files.

If you are a comic book collector, you can use Calibre, a F/OSS. Calibre can read and organize multiple digital formats. It can pull missing metadata from the Internet for your books. It can also convert from one format to another. Don't use it to convert for commercial purposes. It will not pass validation in most cases. For additional information and to download Calibre go to:
<https://calibre-ebook.com/>.

As you can see, in the beginning there was chaos. Then the IDPF said "Let there be EPUB"! And it was so. However, Amazon® is the last holdout. The Amazon Kindle® Readers will not natively read the EPUB format, without tweaking from the end-user. Amazon will convert EPU3 formatted manuscripts into the mobi and azw3 formats.

Amazon has a good reason for not abandoning the mobi or azw3 formats. We will explore that later, when we compare EPUB3 to mobi.

Let's head over to the website for the International Digital Publishing Forum (idpf) at:
<http://idpf.org/epub/30>.