



EPUBZone - The Community
and News Site for all things
EPUB


[Events](#)

[EPUB Event Reports](#)

[Forums](#)

[Resources](#)

Understanding EPUB 3

This article is a high-level introduction to EPUB 3, the latest version of the EPUB format, a widely used and easily manipulable format for representing digital publications. It describes what EPUB is, why it's needed, what kind of content EPUB is suited for and the many options for creating and distributing EPUB publications. If you're an author, editor, or other publishing industry professional, and want to understand the rapidly emerging open standard for next-generation portable documents – without having to understand the details of what's “under the covers” – you're in the right place.

NOTE - Companion articles discuss means of making and distributing EPUB 3 publications - for the impatient who don't care about the “what” and want to get straight on to the “how” you might want to start with one of these articles. A third companion article discusses the current adoption status of EPUB 3.

Even though EPUB publications are, like websites, amenable to hand-authoring and hand-tweaking (a core strength of the EPUB format), some people may appreciate an overview of EPUB from a functionality-oriented perspective (as one might cover PDF, JPEG, or ZIP functionality without delving into any of their internal file-format details). This article assumes that you have a very basic understanding of HTML and CSS, and that you know what JavaScript is.

TIP- If you prefer a more markup-centric introduction, the (free!) “[What is EPUB 3?](#)” from O'Reilly is highly recommended as it the overall O'Reilly “[EPUB 3 Best Practices](#)” book.

EPUB 3 In a Nutshell

- [EPUB 3.1](#)
- [EPUB 3 Overview](#)
- [EPUB 3 Packages](#)
- [EPUB 3 Content Documents](#)
- [EPUB 3 Open Container Format](#)
- [EPUB 3 Media Overlays](#)
- [EPUB Accessibility](#)
- [EPUB Canonical Fragment Identifiers](#)
- [EPUB 3.1 Changes from EPUB 3.0.1](#)
- [EPUB 3.0.1 Changes from EPUB 3.0](#)
- [EPUB 3.0 Changes from EPUB 2.0.1](#)
- [EPUB 3 Working Group Charter](#)
- [EPUB 2.0.1](#)
- [Activities and working group](#)


[Follow IDPF](#)

[Follow EPUB3Zone](#)
[IDPF Newsletter](#)

EPUB is a straightforward format. An EPUB file (extension: .epub) is simply a ZIP archive that contains what is in effect an entire website, including HTML files, all necessary images, CSS style sheets and other assets, plus additional structure and information (“metadata”) that makes the content able to be reliably consumed from beginning to end by any application or device that’s compatible with the EPUB specifications (“EPUB reading systems” or just “reading systems” for short).

EPUB 3 is the latest version of EPUB. EPUB 3 is based on the latest HTML5 standard, which means EPUB publications can now contain video, audio, and interactivity – just like websites in modern browsers.

An EPUB publication’s content is by default reflowable and most reading systems dynamically paginate that content. Adapting content display to the screens rather than forcing the reader to pan and zoom around pre-formatted content (and, more generally, making content accessible to different modes of consumption) is one of the key characteristics that distinguishes EPUB from PDF, a portable document format designed to represent print-replica content. But EPUB 3 can also, like PDF, represent fixed-layout, pre-paginated content. This can be useful for certain kinds of highly-designed content (illustrated children’s books, digital magazines, etc.) that are designed to be consumed only on a larger-screen device, such as a tablet.

NOTE - What makes EPUB different from a website in a ZIP file
To be more specific, and get only a bit more technical, an EPUB publication differs from an arbitrary website put into a ZIP archive in the following ways:

- Manifest - an EPUB publication contains a data structure detailing all of its constituent resources - in effect, a formal and complete “packing list”
- Defined reading order - an EPUB publication has a “spine” that determines a start-to-end reading order for its top-level content items (though nothing requires that the publication to be consumed in that order)
- Standardized table of contents and related navigational structures, including the ability to define links into content fragments down to the sentence or word level
- Metadata - publication and item level information about the publication and its components
- XHTML - the HTML content in an EPUB publication must use the XML-based serialization of HTML5, which makes it directly manipulable by XML tools
- Media Overlays - Well-defined publication-wide text+audio synchronization (an EPUB publication can be both an eBook and audio book)
- Core Media Types (such as image and audio formats) and minimum CSS features are well-defined, along with a requirement

for fallbacks when non-standard content types are used. This enables more reliable processing of EPUB publications, especially when distributed indirectly, than arbitrary websites which may use less standard content types without fallbacks and bleeding-edge CSS constructs that may not be universally supported.

Why Do We Need EPUB?

EPUB enables content to be created by an author or publisher once, via different tools and services, distributed through many channels, and viewed, online or offline, using many different devices and applications.

The EPUB specifications form a kind of “contract” between content creators and reading systems to enable this interoperability.

That all sounds great, but there are a couple of other well-known formats that deliver on similar interoperability promises; namely, PDF and HTML. We have PDF for portable documents and HTML for websites - so why do we also need EPUB? The short answer: PDF is limited to mimicking paper and thus too restrictive for digital readers in an increasingly multi-screen world. It is also an isolated technology silo that's not aligned with HTML5 and the modern Web Platform (for example using unique fillable form technology and scripting APIs). On the other hand, websites are too general, not building-in any way to reliably deliver multi-tier distributability, offline usability, or accessibility. EPUB delivers the portable document abstraction of PDF based on modern Web Standards but without PDF's limitation of representing only final-form pages. Portable documents aren't going away and EPUB makes them a first-class part of the Open Web Platform.

Another advantage of EPUB over PDF and arbitrary HTML is its easy manipulability. EPUB has its storage architecture (ZIP) completely separated from its content architecture (HTML, CSS, XML, etc.) whereas PDF entangles content and packaging together in a legacy binary format. Off-the-shelf ZIP, Web, and XML tools can be used to create and manipulate EPUB. Compared to HTML, EPUB provides significantly more structure and predictability. An arbitrary website has no established reading order (no beginning, middle, and end), no machine-readable table of contents, no well-defined metadata, and there is no guarantee that all of its constituent parts are present and accounted for. That's a strength of the Web as a hypertext architecture - the REST architecture of the Web is predicated on “just-in-time” dynamic content delivery and dynamic link discovery. But it means that navigating an arbitrary website is somewhat like navigating a complicated building without a map - a building where some of the “doors” (links) lead to other buildings (sites), and some of them may not even be materialized until you enter a given room (page). EPUB imposes a particular “map” on a publication that ensures that its content can be used reliably as a portable document. This is at the expense of

minimizing the dynamic nature of the content, an important tradeoff that has to be considered in determining whether to structure content as a portable document in the first place.

When you are going to be accepting content from multiple sources, and/or delivering it through multiple channels, the value of EPUB's document-level portability shines. Of course, if your content all lives in a Wiki or CMS system, it will likely have its own more regular structure imposed by that particular system. If that's OK for your needs, then this advantage of EPUB may seem unhelpful.

NOTE - a three-part article series focused specifically on the future of portable documents, EPUB, and the Web is available at

<http://toc.oreilly.com/2012/08/portable-documents-for-the-open-web-part-1.html>

What EPUB 3 Interoperability means for Content Creators and Consumers

The interoperability benefits delivered by EPUB are for both content creators and consumers.

Publishers can create a single EPUB file and deliver it to all distribution channels that accept the EPUB format. For eBooks, EPUB has become the most prevalent format used by publishers for content creation and distribution around the world and hundreds of different reading systems support this format. Interoperability, for content publishers, can be considered the most fundamental benefit of EPUB and it has, from its inception, facilitated this “supply chain” function.

Consumers can read EPUB files on their choice of device and software. Most EPUB-supporting reading systems allow consumers to import EPUB files into their libraries. This is sometimes called “sideloading”.

However, not all distribution channels that accept EPUB deliver actual EPUB files to consumers, and not all EPUB files can be read on different devices and software programs. Some content providers store EPUB files on their servers to enable online “cloud based” reading in browsers (Safari Books Online for example). Other (Amazon Kindle for example) convert EPUB files into their own proprietary formats and don't allow consumers to directly “sideload” EPUB files. Still others (such as Apple iBooks) apply a proprietary DRM - “Digital Rights Management” technology that encrypts the content of an EPUB file such that it can only be read on the vendor's own software and can't be “sideloaded” into other EPUB compatible reading systems - note that while Apple doesn't make their FairPlay DRM available to 3rd-party reading systems, even on Apple devices and computers, Apple iBooks does allow users to “sideload” non-DRM protected EPUB files that come from arbitrary sources.

But even though consumers may not always have EPUB files that are usable on multiple platforms, they are still receiving an indirect benefit of the EPUB interoperability at the publisher-distributor layer: content availability. With EPUB “greasing the wheels” of the distribution supply chain, simplifying life for publishers and reading system, there are now millions of eBooks available to all.



Copyright 2017 International Digital Publishing Forum.
All rights reserved.

[Terms of Use »](#)