

What Is a Book?

For the W3C Workshop on Publishing and the Open Web Platform
September 2013

Casey Dougherty, iBooks, Apple Inc.
Edward O'Connor, WebKit, Apple Inc.

The title of this paper may seem rather strange. Don't we all know what a book is? But publishing is going through a big revolution...

In the print landscape, the format of a publication is based on time and money constraints combined with the printing technology available within those constraints. For example, a newspaper needs to be printed expeditiously and inexpensively. As a result, printed newspapers are rich with text, but have few images of poor quality on thin paper. Newspapers are excellent for conveying a story, but have little extra media to expand on that story. Next, compare a newspaper to a magazine: A magazine is printed far less frequently, and readers will pay much more for it. Therefore, a magazine publisher can use superior printing with more advanced layouts and many full-color images. A final example is a television program. Traditionally, television specialized in video, but its ability to present text was poor, and images, limited.

The above described publication landscape has changed dramatically as digital technology evolves. One place this change is evident is on the websites of typical newspaper, magazine, and television companies which are ever-increasingly a mix of text, images, audio, and video. These companies have re-oriented to identify themselves around what they present, and less about what format they publish in; instead, they are news companies, fashion editors, food aficionados, etc.

This change of focus – from the vehicle to the content – is happening across the publishing industry and includes books. Traditionally, books have primarily been vehicles for text, and sometimes images (at a higher production cost). Now, digital books can easily include rich media like images, audio, and video. Digital books are vehicles that are increasingly defined by their content, not by the process used to create them. In enriching digital books, media are blending together, and thus blur the lines that divide books from other traditional publications like newspapers, magazines, television, and websites. Taking this a step further into the digital sphere, a digital book is very similar to a website not just in its content (text, images, media), but also in its construction (HTML, CSS, JS). With these lines of demarcation blurring, we ask again, what is a book?

One key differentiator that identifies a book is how it is used by the reader. A book differs from a website in two major ways: the reader expects the book to be self-contained, and usable offline.

What do we mean by “self-contained”? While it is common for a book to reference outside of itself, books have traditionally had a clear division between the content contained within the book (an internal resource), and content referenced that is not within the book (an external resource). The clarity between internal and external resources in a printed book is in stark contrast to a reader’s experience on the World Wide Web. As its name suggests, content in the World Wide Web is interconnected and the user can easily flow from resource to resource without any clear indication of content boundaries.

Currently, digital books exist somewhere between the expectations of a printed book and the World Wide Web and contain elements of both media. For example, similar to printed books, digital books have a spine that dictates a linear reading order. However, unlike printed books, digital books can contain links (Uniform Resource Locators or URLs) that allow the reader to jump to an unknown destination either internal or external to the book. URLs unfortunately, have no clear distinction between internal or external: the closest we have is that relative URLs are often used for internal content, and absolute URLs for external content. Therefore, it is currently the responsibility of the book’s creator and the reading system to manage the expectations of where these URLs lead the reader.

Another example of how a book is different from a website is a reader’s expectation to use a book offline in locations like the beach or on a plane. While offline access is a key feature for digital books, most digital book formats rely on web standards, and it’s not obvious that the web groups – principally W3C working groups – have offline use in mind. Instead, the presumption that web standards are used strictly on the web (and hence when online) can be quite strong. It is not always clear that the working groups consider how a given feature would work in an offline application.

For publishing formats based on the open web platform, we think it’s critical that the community both embrace and explore the digital publishing revolution by fostering the relationship between the book and core web specifications. The introduction of various features in HTML 5 (notably the media elements <audio> and <video>) has driven a profound re-thinking in the publishing industry and is a good example of how web technology can benefit many media types in the increasingly inter-connected world of digital publishing. In developing new web features digital publishing and reader’s methods of consuming that content must also be considered.