

The accessibility of e-books

Luc Audrain,
Hachette Livre

Abstract:

Electronic books in EPUB format use standards of Web accessibility to assist persons with “print disabilities”. Thanks to encoding techniques for accessibility that have been standardized by organizations, such as DAISY Consortium (Digital Accessible Information System), traditional publishers of books, such as novels, have started systematically producing e-books accessible in the EPUB3 format.

An e-book is a computer file that complies with precise, standardized technical specifications so that it can be read on a wide variety of devices.¹ The already classical consumption pattern is immersion in the text displayed on the screen of a computer (less and less often), smartphone (more and more often) or e-reader (the favorite among big readers). Access via mobile devices implies an availability of the book even when the device is disconnected after downloading. An offer via streaming exists but seems marginal. For an e-book to be peripherally accessible, the distribution chain (the bookseller’s website, the payment system) and the devices themselves have to be accessible. Applications on smartphones or e-readers might have drawbacks when used by the disabled.

E-book formats

As for the access to e-books themselves, the technology used to encode the contents is the key for opening the door toward a full, immersive reading experience for everyone, even for persons with visual deficiencies. At the time of writing, three formats are found on the market:

- PDF, the historical format still widely used in academic circles, is losing its share of the market because it adapts poorly to different screen sizes and, for a decade now, has not benefitted from further development.
- Kindle (or .mobi), Amazon’s proprietary format, can be used only in this company’s closed environment (sales sites, proprietary devices and applications). Since its specifications have not been made public, its principles of accessibility are unknown.
- Specifications for the EPUB format have been worked out by international nonprofit organizations: at first in the IDPF (2006-2016) and, since 2017, in the Publishing@W3C department of W3C.² With hundreds of members from several fields in the digital realm, these organizations have publicly released specifications that are free and for-free. This format, a set of Web technologies, now benefits from the full force of innovation on the Internet. For this reason, let us focus on the standards of accessibility related to this format.

¹ This article has been translated from French by Noal Mellott (Omaha Beach, France). All websites have been consulted in April 2019.

² <http://idpf.org/>, <https://www.w3.org/> & <https://www.w3.org/publishing/>

The EPUB format is grounded on Web technology:

- Text is encoded with reference to the Unicode standard for publishing in all known writing systems on the planet with modern fonts.³
- The book's structure is encoded using the HTML standard for Web pages.
- The presentation of the text uses cascading style sheets (CSS).
- The images are included in the types of files recognized by browsers (jpeg, png).
- Multimedia features are called by a markup language specific to audio and video in HTML; they are encoded in the usual Web formats.

EPUB, a packaged website

Publication of an e-book calls for a set of resources that have to be downloaded so that the book can be read when the device is not connected to the Internet. Given that an e-book has a volume much larger than a Web page, it is divided into as many Web pages as there are chapters. A packaging procedure (at present, ZIP data compression) sees to the cohesion of all these resources. An EPUB file can thus be seen as the delivery of a coherent website (the e-book) in a compressed file that mostly contains:

- the HTML files with the text by chapter,
- the graphics (images, figures),
- the style sheets (CSS),
- the metadata that describe the publication (title, author, editor, date of publication, image on the cover),
- the metadata on accessibility, and eventually: multimedia files (audio, video) and/or interactions encoded in JavaScript.

Problems related to the accessibility of e-books in EPUB thus lead us back to the standards underlying the Internet.

For several years now, parties active on the Web have been striving within the W3C to specify the methods, codes and values so that the experience of browsing on a website will be fully intelligible and fluid for persons with impaired eyesight. The Web Accessibility Initiative (WAI) seeks to improve accessibility.

The accessibility of EPUB publications

Besides the accessibility of the HTML contents (ensured by WCAG), the accessibility of the publication involves: browsing: tables of contents and of pages; and the marking of a book's structural divisions. For the latter, a semantics is available, namely EPUB 3 Structural Semantics Vocabulary. A table of concordance has been established toward the values of ARIA "roles".⁴ For publications, new values have been added to these roles, as explained in Digital Publishing WAI-ARIA Module 1.0.

Example of a book's preface:

```
<section role="doc-preface" aria-label="Preface">
<p>When the author sent me the manuscript...</p>
...
</section>
```

³ <http://unicode.org/>

⁴ <https://www.w3.org/TR/dpub-aria-1.0/>

Accessibility was defined in 2017 in EPUB Accessibility 1.0 along with a concrete document on its application (EPUB Accessibility Techniques 1.0).⁵ A work group has made a French version and released a set of resources, including a translation of accessibility techniques.⁶

The DAISY Consortium, a major advocate of accessibility to electronic resources, has an Accessible Publishing Knowledge Base on good practices.⁷ This international organization backs EPUB version 3 as the format for e-books (in replacement of its own format).⁸ In late 2017, it developed a tool for evaluating the accessibility of EPUB publications: the Accessibility Checker (ACE) for EPUB.⁹ This for-free software enables the persons who produce EPUB3 files to verify the quality of the technical elements introduced when encoding contents.

Conclusion

Publishing houses are starting to produce “natively accessible” e-books by using these standards. For the production of e-novels, Hachette Livre has, since the start of 2018, used the EPUB3 “natively accessible” format. The ACE software systematically controls all the files from subcontractors and in-house production services. Any technical anomaly triggers rejection of the file and a demand for corrections. As a consequence, all persons in production are learning about the problems of e-book accessibility.

The French National Library (BNF), is producing accessible versions for the books available in its digital library, Gallica. Thousands of titles are thus made available every year in a natively accessible electronic format so that persons with reading impairments can imagine the end of the shortage of books for them.

⁵ <https://www.w3.org/Submission/EPUB-a11y/> & <http://www.idpf.org/EPUB/>

⁶ Groupe Normes et Standards in the Syndicat National de l'Édition: <https://www.sne.fr/numerique-2/normes-et-standards/> & http://www.edrlab.org/public/sne/TAE_HTML_V3/Techniques_d_Accessibilite_EPUB%201.0.htm

⁷ <http://kb.daisy.org/publishing/>

⁸ <http://www.daisy.org/baseline>

⁹ <https://inclusivEPUBLISHING.org/toolbox/accessibility-checker/getting-started/>