Adventures with ePub3: When Rendering goes Wrong

Maureen Pennock
The British Library
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
+44 (0)1937 546302
Maureen.Pennock@bl.uk

Michael Day
The British Library
96 Euston Road
London NW1 2DB
+44 (0)330 333 1144 ext. 3364
Michael.Day@bl.uk

ABSTRACT

The role of standards in digital preservation is widely acknowledged. The current version of the ePub standard, used for publishing and disseminating eBooks, is ePub3, specifically 3.1 (January 2017). A marked difference from ePub2 is support for fixed layout files and, whilst several different ePub readers are available, not all have upgraded to provide full support for ePub3. In late 2017 and early 2018 the British Library's digital preservation team undertook research into the impact of using an ePub viewer without explicit support for ePub3 on a mixed sample of ePub3 files. The sample comprised 54 files: 20 of these were of fixed layout, the remainder utilized reflowable layouts. For the analysis, content was accessed using two different open source ePub readers: Calibre, which has a wide user base but does not currently explicitly support ePub3, and Readium, which does explicitly support ePub3.

ePub3 files with a reflowable layout broadly rendered to an acceptable standard using both readers. There was one notable exception for both readers, and investigations indicated this was likely due to a problem with the file itself rather than the rendering software. Problems manifested in a more serious way when using Calibre to access just under half of the fixed layout ePub3 files. The rendering of these items inhibited access to intellectual content for example by overlaying it on other content, misrepresenting it, or not including it at all. Other issues that were initially apparent with the other half of the fixed layout sample were mostly resolved by the simple quick fix of switching from 'page' view to 'flow' view. By contrast, Readium was able to display all of fixed layout files correctly.

The research serves as a reminder that whilst standards remain an essential tool in the digital preservation toolbox, updates to a standard may necessitate changes to the rendering software in use. It further underlines the importance of accurate characterization so that repositories can identify formats at a version level or be able to identify items with explicit rendering needs beyond those served by their default rendering viewers.

CCS Concepts

• **Information Storage and Retrieval→Digital Libraries -- Collection.**

Keywords

Digital preservation; epub; ebooks; tools

1. INTRODUCTION

ePub is an open, XML-based publishing standard format maintained by the International Digital Publishing Forum (IDPF).

It was developed by a consortium of publishers and technology companies as an open and cross platform format for representing electronic books, or eBooks. It is designed to offer an optimum viewing experience on a wide variety of devices and screen sizes by allowing reflowable text while also providing support for embedded metadata, raster and vector images, audio and video. [1]

ePub3 is currently the closest thing available to an open standard for eBooks. In 2013, Bläsi and Rothlauf concluded that ePub3 had the "highest expressive power" of all formats in the eBook ecosystem, and that it included the superset of all features used in proprietary eBook formats like KF8, Apple's Fixed Layout ePub, and iBooks. [2] The ePub3 format is widely used by publishers and therefore increasingly received by memory institutions.

2. CONTEXT

2.1 Fixed layout rendering in eBooks

Historically, most eBook formats have been designed to support reflowable content, in that content automatically adapts to the size of screen being used. One of the main principles of the ePub format is that "content presentation adapts to the user, rather than the user having to adapt to a particular presentation of content." [3]

However, being able to define a fixed layout has been considered important for certain classes of eBook. Commonly given examples are comic books, children's books, and publications that include mathematical equations.

ePub3, therefore, added the possibility of supporting fixed layout content. As the standard itself notes:

Sometimes content and design are so intertwined they cannot be separated. Any change in appearance risks changing the meaning, or losing all meaning. Fixed-Layout Documents give Authors greater control over presentation when a reflowable EPUB is not suitable for the content. [3]

While ePub2 did not officially support fixed layout content, proprietary formats based on the format did sometimes include a fixed layout option. For example, Apple developed functionality for iBooks that allowed ePub2 files to be displayed as fixed layout. [4] This, in turn, was later adopted by Kobo. It also became the basis of the fixed layout support incorporated in the ePub3 standard.

Support for creation of fixed layout eBook content was embedded in ePub3. It was contained within an extension specification for ePub 3.0 and 3.0.1, and as part of the main specification in ePub 3.1. The inclusion of fixed text layout only in version 3.1 of the main ePub specification suggests that some ePub 3.0 and 3.0.1 readers would not need to support fixed layout unless they wanted to implement the extension specification. However, readers that

explicitly support ePub 3.1 should support delivery of fixed layout content.

Fixed layout support within ePub3 files was broadly welcomed by publishers, who have seen them as useful for certain kinds of children's books and for complex non-fiction, e.g. textbooks and cookbooks. The functionality to create fixed layout eBook content has since become incorporated into some of the business tools used by publishers to create eBooks. For example, Adobe's InDesign desktop publishing software application added the facility to export fixed layout content in 2014. This was seen to be of benefit to eBook designers, who previously mostly used specialist companies to create fixed layout ePubs for publishers, with a corresponding lack of flexibility. [5]

2.2 Applications that support the rendering of fixed layout ePub3 content

The ePub3Test website lists several applications that support fixed layout ePub3 content.[1] This indicates that fixed layout content is supported by a broad range of reader applications available for different platforms and devices, e.g. Windows, iOS, Android, Kindle and Kindle Fire. The majority of these relate to apps available on eBook readers or tablets, but some do have desktop equivalents. For example, the Windows based apps listed include versions of: Adobe Digital Editions, Kindle for PC, the Kobo Desktop and Windows Apps, and Readium for Chrome. Readium is particularly important as it was used by the IDPF to implement the ePub3.0 specification.

3. METHODOLOGY

A statistically relevant sample set of ePub2 and ePub3 files was generated for the research from the Library's existing collection of eBooks. ePubCheck v4.0.1 was used on the sample to confirm format versions and further to identify whether the file contained a fixed layout declaration.[2] A smaller sample of 351 *ePub3 only* files was then generated. ePubCheck indicated that only approximately 5% of these utilised a fixed layout. Given the relative disparity between the number of fixed and reflowable layout files in this sample, another sampling exercise of the ePub3 reflowable layout files took place in order to produce a final test set of 54 files as follows:

- Fixed-Layout ePub3: 20 files all identified by EPUBCHECK as ePub3 files with a fixed-layout declaration
- Reflowable-Layout ePub3: 34 ePub3 files for which EPUBCHECK did not report a fixed layout declaration.

The content types represented in the sample included illustrated children's books (including comics), manga, illustrated non-fiction (including musical scores and a songbook), academic/professional publications, and trade publications such as novels, poetry and travel guides.

Visual checks took place using Calibre 3/14 [64 bit], running as a standalone program, and Readium 2.29.0, installed as an extension to Chrome; the browser used for this was Google Chrome Version 63.0.3239.132 (Official Build) (64-bit). Readium was selected as an alternative to Calibre, due in part to its origins from the IDPF and its known support for ePub3 fixed layout files.

Each ePub file was opened and (at least) the first 30 pages were visually inspected for potential rendering issues. For items under 30 pages in length the whole of the object was assessed. This visual inspection was one of the main reasons for keeping the sample set relatively small.

4. RESULTS
4.1 Fixed layout ePub3 files
4.1.1 Sample content
The fixed layout sample set consisted of twenty items that represented a variety of content types:

- illustrated children's books (9 items, including comics)
- manga (4 items, all from the same publisher)
- illustrated non-fiction (7 items)

All items were primarily image based, at least for viewing, although most seemed to include also some searchable text and a table of contents.

4.1.2 Analysis
Eight of the fixed layout items in the sample could be rendered more-or-less acceptably in Calibre using 'flow' viewing mode. This does not mean that all pages rendered perfectly, but that all of the page images would load and could be viewed - although this might be over more than one logical page in the viewer. Several of these items exhibited other minor rendering issues, e.g. poor text formatting on selected pages or the display of extraneous underlying text outside the page margins, but in these cases, this was considered not to significantly impair the understanding of the underlying item (in these cases, it usually only occurred on the colophon or copyright pages). In some of the items, the page image appeared to be very slightly cropped in the left or right-hand margins, although this did not happen on every page (and all of the text parts typically remained perfectly viewable).

Three of these items used a wider-format than was optimal for rendering in Calibre. A work-around was to view the items in Calibre's 'flow' mode, and then to use the horizontal scroll bar to navigate to the parts of the page not easily viewable otherwise. Needless to say, this was not an ideal solution and would result in a sub-optimal user experience.

The remaining twelve items manifested text rendering issues. Sometimes this only affected particular pages (e.g. the colophon), while in others the effects were more thoroughgoing.

Nine of the twelve items did not render to an acceptable standard. Five of these were illustrated books (three illustrated children's books, a large format illustrated children's book, and an illustrated non-fiction book), where the following rendering problems were variously evident:

- Text incorrectly superimposed over images, either superimposing when not intended, or the loss of an image behind a block of text
- Text extending beyond page margins and occasionally overlapping
- Strange font rendering
- Text appearing 'scrambled' (e.g. like a word cloud)
- Images extending beyond page margins.

---

[1] ePub3Test website reading system evaluations: http://epubtest.org/testsuite/epub3/

[2] EPUBCHECK is a tool to validate ePub files, coordinated by the W3C ePub3 Community Group: https://github.com/IDPF/epubcheck

The other four items that rendered poorly included songbooks and non-fiction works with heavy usage of images. Problems included music staves being stretched horizontally, song titles superimposed upon the music staves, rights information not being rendered, and text overlapping with diagrams.

Of the twenty items in this set, only one incorporated any "enhanced" functionality. The item metadata indicated that the item included "OverDrive (READ) format features optional read-along narration." This functionality was not available within Calibre and the sound component of the content was therefore wholly inaccessible.

When accessed via Readium, all twenty items were viewed acceptably using default Readium settings. The reader automatically fitted the page size to the viewer, although for fixed-width content this could also be customised by the user. The four manga items were correctly delivered with pages from right to left. It was also possible to activate the OverDrive audio read-along track on the item with "enhanced" functionality. Readium provided an acceptable viewing experience for all fixed layout items even without customisation. There was a single query regarding whether text had been cropped slightly on a left hand margin on a single page, but it was difficult to be certain about that.

4.2 Reflowable layout ePub3 files

4.2.1 Sample content
This test set comprised 34 files and was noticeably different in composition from the fixed layout ePub3 files in the previous set. The content types represented were:

- Academic or professional publications (25 items)
- Trade publications such as novels, poetry or travel guides (9 items)

The majority of these items contained both internal links (e.g. tables of contents, notes and references) and external links or references to companion web pages, alongside images (e.g., figures, photographs) of varying quality and size.

4.2.2 Analysis
All but one of the 34 items rendered acceptably in Calibre. A single travel guide item did not render well, specifically the rendering of the images and maps, although the text elements were mostly legible. It is perhaps worth noting that this item did not render very well in either of the tested readers and is discussed in more detail in section 4.2.3.

Some other minor rendering issues were noted within the sample. The table of content links in one item could not be activated. Other minor rendering issues, such as text looking compressed and slight margin cropping, could be fixed by changing the size of the reader window.

As with Calibre, Readium was able to render all but one of the 34 items acceptably; the exception again being the travel guide item referred to above. Two other items displayed some minor rendering issues, including the link activation problem that had been noted in the Calibre tests, and one instance of slight cropping of text within margins.

4.2.3 Travel guide rendering issues
A single travel guide item in the reflowable layout sample did not render well in either Calibre or Readium. This item combined images and text with maps that were intended to utilise viewer zoom technology.

In the Calibre viewer, several serious rendering and usability issues were evident, the main ones being the inability to change the size of the maps and the distortion of images. Of these, the map enlargement problem was probably the most significant, potentially resulting in usability issues for a guide book. Some of the map images were small and there were no discernible means available to enlarge or reduce them. In addition, the "zoom left" and "zoom right" functionality indicated within the item could not be activated. Another apparent problem was that of image rendering. Many of the images were distorted in regard to their aspect ratio, which produced some odd results. The image ratio seemed to be fixed, as resizing the page or fonts made no difference to the way that the image rendered (or to the size of the image).

There were some additional problems, especially with chapter title pages when using full-screen mode. Additional text would appear superimposed upon the page and the background images became distorted.

With the exception of the image distortion, many of the same problems manifested when the content was accessed in Readium. The maps could not be usefully enlarged and there were some obvious rendering problems with pages, including the overlapping of text and image elements.

The item was a valid ePub3 file and a correct implementation of the specification. It is possible therefore that these problems are simply down to file construction methods, such as incorrect aspect ratio specifications. This would require further analysis to confirm.

5. CONCLUSIONS
There are thousands of software applications on the market that could potentially be used by memory institutions to provide access to content stored in different formats and versions of those formats. These applications provide varying degrees of rendering support for formats and format characteristics. This research represents only a tiny corner of the format and application ecosystem and explores how just two different and widely used applications render a given format, in this case ePub. It found distinct differences in how the two applications performed when rendering some files in a newer version of that format, though notably not all of the files in that version.

Both applications are considered to support ePub, though Calibre only explicitly supports ePub versions up to ePub2. Our research showed that, despite this, Calibre was broadly capable of rendering ePub3 files with a reflowable layout. However, it has its limitations as a viewer for ePub3 fixed layout content, especially that which is mainly image-based. Changing the viewing mode resolved the presentation problems in some fixed layout ePub3 files, although page navigation could be a bit tedious (i.e., there were lots of blank "pages"). However, where the format of the underlying content was non-standard - e.g. square sized, inordinately large – significant parts of the page were not able to be viewed appropriately within Calibre, either in paged mode (the default) or flow mode.

These issues can be summarised as:

a) Rendering difficulty that does not impact on access to intellectual content but that does impact on the reader's viewing experience.
b) Rendering difficulty that impacts or inhibits access to intellectual image-based content
c) Rendering difficulty that impacts or inhibits access to intellectual textual content.
d) Rendering difficulty that inhibits access to intellectual audio content

The research did not consider whether these problems also manifested in other ePub viewers that did not explicitly support ePub3. It is quite possible therefore that these problems may manifest with other viewers aside from Calibre, though it was out of scope of the research to investigate this. We take pains to stress that at the time of writing, Calibre does not claim to support ePub3 and that the observations in this paper should be taken as indicative of rendering problems in general rather than being a particular criticism of Calibre, which as open source software could be updated to more explicitly support ePub3 should anyone wish to do so.

More important is what we can conclude from this research from a preservation perspective. It serves as a stark reminder that the suitability of a rendering application can change when a format is updated, *particularly if the application is not updated to explicitly support the new version of the format*. Whilst this may not necessarily manifest for all files in a collection, it may manifest for those files that utilize newly updated or different features of the format specification, and in ways that not only impact on the reader experience but also impede access to intellectual content. It might be argued that this is an access challenge rather than preservation per se, but if the objective of preservation is to provide reliable access over time then perhaps it cannot be one without also being the other.

We further note that reliance upon updating software to the most recent version of the application does not necessarily mean it will support new or revised versions of formats. The version of Calibre tested in this research was the most up to date available at the time, having been released in December 2017. ePub 3.1 was published in January 2017, some eleven months earlier.

It is essential for repositories to ensure they provide software that will accurately render files in their collections. This research indicates that retaining format metadata alone may not be sufficient for preservation and access even in the short term. Characterisation may be needed at a more granular level if it is to serve its purpose, with format identification at a version or even feature level. More work is needed to understand the wider implications of this with formats other than ePub3.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Whibley, S., et al. 2015. *EPUB format preservation assessment*, v 1.2. British Library, London. URL=http://wiki.dpconline.org/images/a/a9/EPUB_Assessment_v1.2.pdf

[2] Bläsi, C., and Rothlauf, F. 2013. *On the interoperability of eBook formats*. European and International Booksellers Federation, Brussels. URL= http://wi.bwl.uni-mainz.de/publikationen/InteroperabilityReportGutenbergfinal07052013.pdf

[3] International Digital Publishing Forum. 2017. EPUB Packages 3.1: Recommended Specification, 5 January 2017. URL=http://www.idpf.org/epub/31/spec/epub-packages.html#sec-package-metadata-fxl

[4] eBook Architects. 'Learn about eBooks: Non-fiction fixed layout eBooks.' URL=http://ebookarchitects.com/learn-about-ebooks/non-fiction-fixed-layout/

[5] Schmidt, G. 2014. 'Why fixed-layout EPUB from InDesign is a big deal for designers,' InDesign Secrets, 25 August 2014. URL=https://indesignsecrets.com/fixed-layout-epub-indesign-big-deal-designers.php