Digital Preservation Team	**Preservation Assessment:** EPUB Format	**Date:** 26/06/2019
		Version: 1.4

EPUB Format Preservation Assessment

Document History

Date	Version	Author(s)	Circulation
01/07/2015	1.0	Michael Day, Peter May, Paul Wheatley. Review: Johan Van der Knijff	External
02/07/2015	1.1	Peter May	External
04/09/2015	1.2	Simon Whibley	External
20/12/2016	1.3	Peter May	External
26/06/2019	1.4	Simon Whibley, Michael Day	External

Digital Preservation Team	**Preservation Assessment:** EPUB Format	**Date:** 26/06/2019
		Version: 1.4

1. Introduction

This document provides a high level, non-collection specific assessment of the EPUB file format with regard to preservation risks and the practicalities of preserving data in this format.

This format assessment is one of a series of assessments carried out by the British Library's Digital Preservation Team. An explanation of criteria used in this assessment is provided in italics below each heading.

1.1 Scope

This document will primarily focus on two main releases of the EPUB format:
- EPUB version 3 -- the current version of EPUB. The most recent specification version is EPUB 3.2 [1], which will be published by the W3 EPUB 3 Community Group and International Digital Publishing Forum (IDPF) in 2019 [2]. It supersedes the minor revision of 3.1 in January 2017 and the maintenance update version 3.0.1 published in June 2014.
- EPUB version 2 -- a release of EPUB **initially** approved by IDPF in October 2007 with a maintenance update (2.0.1) issued in 2010. While now superseded by EPUB 3, EPUB 2 is still used by a number of eBook publishers. EPUB 2 superseded, in turn, the proprietary Open eBook Publication Structure (OEBPS) format.

Note that this assessment considers format issues only, and does not explore other factors essential to a preservation planning exercise, such as collection specific characteristics, that should always be considered before implementing preservation actions.

1.2 EPUB Summary

EPUB has been developed by a consortium of publishers and technology companies to provide an open and cross platform file 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 [3] while also providing support for embedded metadata, raster and vector images, audio and video.

EPUB is based on a variety of open standards. An EPUB file is a ZIP container that includes a mix of XHTML files (eBook content), CSS files (formatting), XML files (metadata) and other embedded content such as SVG.

EPUB 2.0.1 is defined by three specifications as follows:
- Open Publication Structure (OPS) 2.0.1 defines the formatting of the contents
- Open Packaging Format (OPF) 2.0.1 is an XML description of the EPUB structure
- Open Container Format (OCF) 2.0.1 defines the collection of all EPUB files as a ZIP archive

EPUB 3.0.1 is defined by the following specifications:
- The EPUB 3 Overview, which provides context and a roadmap for the other documents
- EPUB Publications 3.0.1, which defines semantics and conformance requirements for EPUB publications
- EPUB Content Documents 3.0.1, which defines profiles of XHTML, CSS and SVG
- EPUB Open Container Format (OCF) 3.0.1, which defines how the EPUB components are packaged within a container file (ZIP)
- EPUB Media Overlays 3.0.1, which defines a model for the synchronisation for text and audio

EPUB 3.1 underwent a major reorganisation and was organised by the following specifications:
- EPUB 3.1 Specification [4] [5], which is an umbrella specification acting as the point of entry and incorporating EPUB publication and reading system requirements.
- EPUB Packages 3.1 [6], which is a renaming of the Publications 3.0.1 document "to better reflect that it defines Renditions of content" [7]. This specification now includes the EPUB Navigation Document (moved from the Content Documents 3.0.1) which replaces the EPUB 2 NCX document.

- EPUB Content Documents 3.1 [8] continues to define HTML, SVG and CSS profiles.
- EPUB Media Overlays 3.1 [9] continues to define a model for synchronisation of text and audio.
- EPUB Open Container Format 3.1 [10] continues to define the EPUB file format and encapsulation.
- EPUB Accessibility 1.0 [11] which defines requirements to evaluate the accessibility of EPUB content.
- Alternate Style Tags 1.1 [12], which defines an approach for tagging alternate style sheets, for example to enable horizontal-vertical layout switching.

EPUB 3.2 is still awaiting full publication later this year (2019) but similarly underwent a further reorganisation. The specification includes:

- EPUB 3.2 Specification [1] acts as the new umbrella specification representing the top specification in the family for both EPUB Publications and EPUB Reading Systems which were formerly defined in 3.01 and have been now moved to the top-level specification along with the section on Publication Resources
- EPUB Packages 3.2 [13] continues to define requirements for each rendition of the content.
- EPUB Content Documents 3.2 [14] continues to define XHTML, SVG and CSS profiles.
- EPUB Media Overlays 3.2 [15] continues to define the file format and a processing model for the synchronisation of text and audio.
- EPUB Open Container Format 3.2 [16] continues to define the file format and processing model for encapsulation of a set of related resources in a single (ZIP) EPUB container
- EPUB Accessibility 1.0 has not been updated.

EPUB 3 is currently the closest thing available to an open standard for eBooks. In 2013, Bläsi and Rothlauf concluded that EPUB 3 had the "highest expressive power" of all formats in the eBook ecosystem, and that it included the superset of all features used in proprietary formats like KF8, Fixed Layout EPUB, and iBooks [17].

2. Assessment

2.1 Development Status

A summary of the development history of the format and an indication of it's current status

The EPUB format was created as an open standard by the International Digital Publishing Forum (IDPF) [18] in September 2007, based on the previously proprietary Open eBook format developed by SoftBook Press [19]. It utilised the same underlying standards, namely XHTML version 1.1 packaged within a ZIP container. Minor revisions were made in version 2.0.1 (the final EPUB 2 release, approved by IDPF in May 2010) and more significant structural and functional changes were made in EPUB version 3 (released by IDPF in October 2011), including support for equations via MathML (Mathematical Markup Language) and greater control of layout and typography including fixed-layout documents. A detailed description of changes can be found on the IDPF website [20].

At the time it was first developed, EPUB version 3 was based on technologies that were themselves, at the time, still under development, such as XHTML5 [21] (now usually seen as part of HTML5, published in 2014 [22]) and CSS3 (this update was split into modules and has a variety of release dates at different statuses) [23]. A minor maintenance release, version 3.0.1, was issued by IDPF in June 2014. EPUB 3.1, a minor revision of EPUB 3, was then issued by W3C in January 2017 [5], but did not receive wide adoption [24]. The focus of the W3C's EPUB 3 Community Group then shifted to the development of EPUB 3.2, a second minor revision of EPUB 3. EPUB 3.2 was specifically designed to be backwards-compatible with EPUB 3.0.1, whilst retaining many of the changes made in EPUB 3.1 [25]. In terms of compatibility, this means that all valid EPUB 3.0.1 files will also be valid EPUB 3.2 files. EPUB 3.2 is currently still in the final stages of development, but is expected to be published later in 2019 [26]. The latest

version of EPUB 3.2 at the time of writing is a W3C Final Community Group Specification from May 2019 [27].

In late 2014, EPUB 3.0 was submitted to ISO/IEC JTC 1 SC34 (Joint Technical Committee 1, Subcommittee 34, Document description and processing languages) and published as ISO/IEC TS 30135 (parts 1-7) [28]. In December 2014, IDPF proposed submitting an EPUB 3.1 update for consideration as a full international standard [29], but it is not entirely clear whether this was followed through. In 2018, however, EPUB 3.0.1 was submitted to ISO under a fast-track procedure as a Draft International Standard, largely so that it could be referenced normatively by the emerging EPUB Accessibilty (EPUB A11Y) specification [30]. The most recent officially-published version of the ISO technical specification remains ISO/IEC TS 30135 from 2014 (it was most recently reviewed and confirmed in 2018 [31].

IDPF merged with the W3C (World Wide Web Consortium) in early 2017 [32], becoming part of Publishing@W3C [33]. The W3C's EPUB 3 Community Group is the current forum for ongoing technical development of EPUB 3 and for ancilliary tools like EPUBCheck [34].

Over the development cycle, EPUB has had significant changes in its relationship to the core web specifications of HTML, CSS and SVG, in that, as defined by the W3C, EPUB 3.2 now officially supports the current versions of the specifications, rather than pointing to dated versions, as it had in the past [24]. A major change introduced in EPUB 3 was the ability to include metadata allowing the creation of fixed layout documents, enabling the dimensions of the page to be controlled and content to be fixed [13]. The metadata does not just flag whether content is be fixed or reflowed, but also enables content creators to specify the orientation of the pages and how to position pages, providing control over the presentation of the publications. Fixed-layout has been considered important for certain classes of ebook such as comic books or children's books and, depending on implementation, could help facilitate the external referencing of EPUBs. Fixed-layout content does though cause rendering issues with some current viewers (see below).

Competing suppliers have produced eBook formats broadly based on EPUB, but which also incorporate DRM and other modifications that, in some cases, restrict use to particular platforms or users. For example, the Kindle Fire tablet utilises it own proprietary ebook formats, Kindle Format 8 (KF8) or the more recent KFX [35], based in part on EPUB version 3, with Amazon DRM. Apple has its own proprietary brand of EPUB, the iBook format. The use of subtle changes to EPUB 3 and non-standard CSS restrict its use to Apple software. Martin Taylor characterised the situation as it stood in 2012: "So, even before the first EPUB3 eBook goes on sale, we have three (incompatible) variations of the industry standard." [36]

2.2 Adoption and Usage

An impression of how widely used the file format is, with reference to use in other memory organisations and their practical experiences of working with the format

EPUB adoption

EPUB currently enjoys reasonable (if not universally widespread) support across e-reader devices, PC based software, eBook creation services, publishers and online suppliers [37] [38].

Many of the suppliers, publishers and application developers who supported EPUB version 2 have now also developed support for EPUB version 3 [39]. Support for some aspects of the more recent versions of the format has now been verified [40].

The Book Industry Study Group (BISG) recommended EPUB 3 as their format of choice for packaging of content in 2015 [38].

EPUB in Memory Institutions

On behalf of the UK Legal Deposit Libraries, the British Library has been collecting eBooks under legal deposit since 2013. This includes content in EPUB[1] which can be either received directly from publishers or distributors (by arrangement) or uploaded by publishers via a deposit portal.

Similarly, the German National Library (DNB) receives eBooks in EPUB format through legal deposit. While covered in principle by the German legal deposit regulation, the DNB does not collect proprietary eBook formats like those used by Kindle [41]. However, it does provide advice to authors that retain distribution rights on how to create and deposit self-published eBooks in unencrypted EPUB, e.g. by using conversion tools like Calibre [42].

Of the four memory organisations examined by Kirchhoff and Morrissey as case studies in a 2014 DPC technology watch report, only the Library of Congress at the time were already receiving files in EPUB format, having received content in EPUB 2, and were expecting to receive files in EPUB 3. HathiTrust, Portico and the National Library of the Netherlands had not received EPUB files, although the latter were expecting "...to be receiving content in EPUB 2 and EPUB 3 in the near future..." [43]. As of 2019, the Library of Congress reports it has an EPUB 3 collection but no current holdings of EPUB 2 [44]. Its Recommended Formats Statement indicate EPUB 3 as a preferred format for textual works in digital form [45]

The popularity of EPUB for the publication or supply of material by memory organisations is unclear, but the Kirchhoff and Morrissey report notes that HathiTrust supplies some content in generated EPUBs. Similarly, the Bibliothèque nationale de France makes some of its digitised content available in EPUB format via its Gallica portal [46]. Several thousand books are downloadable in the EPUB format from Gallica and, since 2014, it has used the EPUB 3 format [47].

2.3 Software Support

2.3.1 Rendering Software Support

An overall impression of software support for rendering the format with reference to: typical desktop software; and current support on British Library reading room PCs

A considerable number of hardware e-readers [48] and software applications provide support for viewing EPUB files. Most of the leading dedicated e-reader devices are able to render EPUB files, for example, e-readers produced by Kobo, Barnes and Noble (Nook) and Sony (Sony Reader), but with the significant exception of Amazon's Kindle.

On most platforms, EPUB files will primarily be rendered by software applications, which range from browser plugins [49], to open source eBook suites such as Calibre [50], free readers such as Azardi [51], and commercial applications like Adobe's Digital Editions [52]. Reader software for EPUB files is also widely available for tablet computers and mobile telephones, including the iBooks app (iPhone or iPad), and Aldiko Book Reader (Android).

IDPF had been supporting the development of a reference implementation for EPUB version 3 in the form of the Readium rendering engine and software development kit [53]. This had progressed sufficiently to enable the launch of an EPUB reader for IOS7 based on Readium [54]. Unfortunately, the final release of the Readium Chrome app was announced in October 2018 (noted on app store as version 2.31.1) [55]. "Version 1.0" (in actuality version 0.16) of the Readium SDK was released in January 2015, described as an "open source library designed to support rendering of EPUB 2 and EPUB 3 content in native applications for mobile and desktop operating systems" [56]. The JS and SDK versions (latest version 0.31) are now in maintenance mode with no plans for further releases other than fixing high priority bugs [57].

EPUB content collected by the British Library through legal deposit provisions are made available on reading room PCs through Ericom's AccessNow client, which provides remote access through a web browser (and includes an eBook viewer).

---

[1] In EPUB 2 and EPUB 3. As of June 2019, approximately 30% of the ebooks intake has been in EPUB 3

Issues

The BISG EPUB 3 Support Grid (part of the EPUBTest.org website) evaluates reading systems based on a test suite of EPUB 3 documents [58]. The results reveal the limited support of EPUB (in terms of functionality coverage) offered by many applications. The extent of any correlation between functionality in eBook files and areas of functionality not supported by major eBook applications is unknown, and the impact of these shortcomings on long-term preservation is therefore hard to estimate.

EPUB version 3 files may not render correctly on viewers that only support EPUB version 2 as EPUB version 3 is based on XHMTL5 rather than XHTML1.1. Publishers such as O'Reilly have produced EPUB 3 files that will render correctly on EPUB 2 viewers, but they commented that achieving complete backwards compatibility was not trivial [59].

AZARDI is specifically an EPUB 3 reader but hasn't been updated since 2016.

EPUB version 3 dropped support for the DTBook (DAISY Digital Talking Book) format and some EPUB viewers (e.g. Calibre, Readium) are unable to process EPUBs that contain DTBook content [60].

Fixed-layout EPUB 3 files have their own issues depending on the reader used, Calibre especially has issues rendering large fixed-layout within the screen. Readium, AZARDI (post-2011) and Adobe Digital Editions (version 4.52 onwards) both offer support for fixed-layout content [61].

The requirement for an NCX (the short name for a Navigation Control file for XML applications) was removed in the move from EPUB 2 to 3 and replaced by the Navigation Document though they can co-exist. EPUB 3 readers must ignore the NCX file but there appears to be no evidence that has created any rendering issues [62] [63] [64].

Audio/Video content can be included (or referenced in EPUB 3), however there is no requirement for EPUB readers to support any video codecs at all [65]. Consequently, there are no guarantees that a given EPUB reader will be able to display any, or all, video files within an EPUB.

2.3.2 Preservation Software Support

An impression of the availability and effectiveness of software for managing and preserving instances of the file format

Format identification

EPUB has a registered media type (formerly MIME-type) of "application/epub+zip" [66]. Identification of this is supported by DROID and Apache Tika but, as at the time of writing (in DROID v6.4 and Apache Tika 1.21), the EPUB version is not identified; given the potential difficulties for a single reader to render various versions of EPUB (see Rendering Software Support, above), understanding which versions are in a collection may be important.

Although not a generic file format identification tool, EPUBCheck [67] does provide EPUB version information in its output, so this information could be extracted along with validation results.

As an EPUB file is composed of a number of XHTML and other files within a ZIP container, it is possible that format identification could also – and may want to – be run at lower levels of aggregation.

Validation, Conformance Checking and Detecting Preservation Risks

Both EPUBCheck [67] and FlightCrew [68] provide open source validation facilities for EPUB. EPUBCheck outputs a detailed array of information (using the JHOVE schema) that including basic descriptive metadata, reports of validation errors, a report on the presence (or otherwise) of encryption, and a summary report on validity. As with other validation tools and formats,

interpreting the validation reports is not always straightforward and intuitive; a commercial tool called Flightdeck [69] provides extra interpretation on EPUBCheck output and may offer some value here [70].

The SCAPE Project [71] developed a modular tool, FLint (previously called DRMlint), which wraps EPUBCheck and Calibre (amongst others) and uses them to validate files against format specific profiles, such as whether there is DRM/Encryption in eBook formats [72].

EPUB3 support within these tools is mixed. FlightCrew (v0.9.2) does not currently support EPUB3 validation [73]; EPUBCheck (v4.0.2 onwards) and, by virtue, Flightdeck and FLint do.

In general, tools should be monitored and actively evaluated for EPUB coverage, as development efforts appear to have been sporadic over the last few years. FlightCrew, in particular, saw some development effort in August 2011, and then nothing until April 2015 when it migrated to GitHub hosting; since then there has only been two to three minor code commits but the most recent dates back to December 2016. EPUBCheck has been more active with moderate development effort levels since late 2012 with peaks in 2013-2015 through to increased contributions recently (June 2019) [74]. There was at least one bug in EPUBCheck which resulted in some EPUBs erroneously passing validation but this was resolved in 2017 [75]. The latest beta of EPUBCheck (version 4.2.0) will provide support for conformance checking of EPUB 3.2, updates to the vocabularies and support for new Core Media Types [76].

The IDPF provides a range of sample EPUB 3 files that could be used to facilitate tool evaluations [77].

The market situation for eBooks, while creating uncertainty, also provides some potential for optimism with regard to preservation. Application support for the quality assurance of files created by the self-publishing movement is driven by the submission requirements from publishers and the complex situation for the eBook market described above.

Metadata Extraction
The open and XML based nature of EPUB, makes metadata extraction straightforward [78]. A variety of tools, including EPUBCheck (see above), therefore support metadata extraction.

Migration
A number of applications support migration to and from EPUB formats (version 2 and 3), including the open source Calibre tool [79]. As with the market drivers for quality assurance tools, the need for authors to provide eBook versions in different formats to meet the requirements of publishers and suppliers has created a market for eBook migration tools and services. However the quality of migration is unknown and quality assurance may be challenging without extensive manual testing.

2.4 Documentation and Guidance
An indication of the availability of practical documentation or guidance with specific reference to the facilitation of any recommended actions

File format specifications for all major versions of EPUB are available either from the IDPF or W3C websites. Specifications for versions of EPUB published prior to 2017 (i.e., EPUB 2.0.1, EPUB 3.0, and EPUB 3.0.1) are available from the (now mostly archived) IDPF website [64], but with comprehensive links from an active "EPUB Specifications and Projects" page [80]. EPUB 3.1 was published as a W3C Member Submission in 2017 [5], while the specifications themselves are also available from the IDPF website [80]. The latest version of EPUB 3.2 (W3C Final Community Group Specification) has been published by W3C [26], but with other draft documents available from the W3C GitHub site [80]. As noted above, the EPUB 3.0 specifications were also republished in 2014 by the International Organisation for Standardisation as ISO/IEC TS 30135 parts 1-7 [28].

Guidance for supporting users of EPUB is not in short supply due to the market for self-publishing. The development of the Readium reference implementation for EPUB version 3 [53],

mentioned above, should also be noted despite there being little chance of future support. The Mobileread wiki has detailed format descriptions, recommendations for working with the format and notes on tool support [81].

2.5 Complexity

An impression of the complexity of the format with respect to the impact this is likely to have on the British Library managing or working with content in this format. What level of expertise in the format is required to have confidence in management and preservation?

The choice of human readable open web standards (such as XHTML and CSS) on which EPUB is based considerably simplifies the accessibility of the EPUB format to non-experts. For this reason, expertise in management and preservation of web content, as well as eBooks specifically, may be helpful with EPUBs.

2.6 Embedded or Attached Content

The potential for embedding or attaching files of similar or different formats, and the likely implications of this

EPUB files may include a variety of embedded media types (or *publication resources* as the EPUB specification refers to them as), such as raster graphics, audio or video. EPUB readers must, according to the specification (but see the section on Rendering Software Support, above), support all media types on the specification's *core media types* list.

The *core media types* for EPUB version 3 [65] were extended, with some deprecations and replacements over those allowed in EPUB version 2 [82]. In particular, EPUB 3 allows embedded JavaScript, however, there is no core media type for video formats (and therefore no requirement on EPUB readers to support any video codecs).

If the media format to be embedded is not on the list of *core media types*, an alternative that is on the list (a *fallback*) must be included as well. Whilst the specification's requirement to include a *fallback* version of embedded media (via the manifest file) should enable an EPUB reader to present an appropriate rendering to a user, there appears to be no guarantees that *fallback* versions will present the exact same rendition as intended by the original embedded file. Indeed it turns out that few EPUB readers support manifest fallbacks at all [83]. Both the lack of support for fallback media and the potentially different rendition from them presents potential preservation problems in terms of access and authentic rendition of the EPUB.

2.7 External Dependencies

An indication of the possibility of content external to an instance of the file format that is complimentary or even essential to the intellectual content of the instance

In EPUB version 2 all *publication resources* (see Embedded or Attached Content, above) must be embedded (i.e. included in the EPUB ZIP container) with the exception of fonts. The same applies to EPUB version 3 with an additional exception for audio and video resources which may be externally referenced with only embedded image or text based *fallbacks* required. EPUB 3.1 also allows for resources retrieved by scripts to be located outside the EPUB (with restrictions on the HTML and SVG script elements). In all cases, EPUB authors are encouraged to include resources within the EPUB Container to enable offline (no internet connectivity) access.

All publication resources, whether embedded or externally referenced must be declared in the package manifest.

2.8 Legal Issues

Legal impediments to the use, management or preservation of instances of the file format

A facility is provided in EPUB to provide basic protection of embedded commercial fonts through encryption of the first 1040 bytes of a font file, known as font obfuscation. Its aim is to provide basic protection against casual font piracy, as embedded fonts can otherwise easily be removed

from EPUB ZIP containers. The presence of font obfuscation could cause difficulties in migrating EPUB files from one format to another. While defeating the encryption would not be challenging, it may not be legal to do so.

The EPUB specification requires that font obfuscation is specified in the embedded encryption.xml file. Van der Knijff notes that EPUBCheck is able to indicate the presence of font obfuscation encryption, however it does not distinguish between this and encryption used more generally [60] [84].

Publication resources, particularly ones located externally to the EPUB, may have copyright restrictions of their own to consider.

2.9 Technical Protection Mechanisms

Encryption, Digital Rights Management and any other technical mechanisms that might restrict usage, management or preservation of instances of the file format

Beyond font obfuscation, files within an EPUB (versions 2 and 3) container can be encrypted (with some exceptions, such as the container.xml file pointing to the root files for the EPUB publication). Encryption is, obviously, a long-term access concern if the necessary decryption keys are lost or unavailable. An encryption.xml file must be included to provide the necessary details about which files have been encrypted.

The EPUB standard does not prescribe a specific DRM scheme, though EPUB files may use DRM that requires a registration key or password to access them. Various publishers and suppliers use their own DRM technology, with Amazon's DRM system, Apple's FairPlay DRM [85] and Adobe's ADEPT system [86] being the common varieties. A major consequence of such publisher/supplier specific DRM is that eBooks tied into one DRM system are restricted to that publisher/supplier's rendering devices and/or applications.

2.10 Other Preservation Risks

Other evidence based preservation risks, noting that many known preservation risks are format specific and do not easily fit under any of the sustainability factors above

None known.

2.11 Preservation Risk Summary

A summary of preservation risks and recommended actions (where possible).

EPUB is widely used and supported within the eBook sector and is encouragingly based upon an array of open standards. However, the eBook market is still developing and this means that EPUB remains just one choice within a wider range of available formats. EPUB exists within a complex market situation that many have described as a format war. Adaptations of EPUB have created competing formats and incompatible DRM schemes that have left the eBook consumer in an unsatisfactory situation. This suggests that eBook formats are not yet at the end of a period of rapid evolution. EPUB, and other eBook formats, will need to be carefully monitored over the next few years at the very least.

The existence and continued development of open source viewers increase confidence in EPUB preservation considerably, but the degree of completeness of support for functionality described in the standard is a concern. EPUBCheck provides vital validation and metadata extraction capabilities, however it is not complete in coverage and responsiveness to fixing errors seems low.

The widespread use of DRM, implemented in an array of schemes, is another significant area for concern for EPUB preservation.

Of the remaining risks, the potential for obfuscated or non-embedded fonts appears to be the most concerning and the impact and incidence of this risk will need to be monitored.

- **Lack of EPUB format stability**
  - Evolving standards, context and the format itself places uncertainty on the future preservation situation.
  - Proprietary changes and non-standard use of specifications may be used to restrict access to specific manufacturer hardware/software.

- **Incomplete support in EPUB viewers**
  - Support for all aspects of the standard appears to be mixed, although the impact of this is unclear.
  - Approach to Fixed-Layout content not universal
  - With no requirement for video codec support, there is potential for embedded/referenced video content to be unrenderable in a given EPUB reader.

- **Invalid or badly formed EPUB files**
  - May affect ability to render files now or in the future, but there is some cause for optimism that this will not be as significant a problem as with (for example) PDF.
  - Bugs in validation software (e.g. lack of tests for compressed mimetype.xml files) may mean that invalid EPUB files are not always detected.

- **DRM and Encryption**
  - Multiple widely used DRM schemes have the potential to prevent viewing and preservation.
  - Obfuscated fonts could hamper preservation actions.

- **Missing font information**
  - Where not easily substituted, non-embedded fonts could lead to loss of critical information. Equally the ePub specification allows for fonts to be encrypted which would potentially pose a long term preservation issue with future migrations. Encrypted fonts must be described in META-INF/encryption.xml, so checking for this is advisable.

- **Legal issues**
  - Embedded copyrighted fonts could be obfuscated, hampering preservation actions; the legality of defeating this encryption is not clear.

- **JavaScript**
  - EPUB 3 allows embedded JavaScript which enables complex interactive behaviour that could be challenging to preserve.

- **Embedded content**
  - Embedded content not on the core media type list and not support by an EPUB reader should – according to the specification – be rendered using a specified fallback that is a core media type. There are no guarantees this will present the exact same rendition (if any is presented at all).

- **External References**
  - Externally referenced audio or video content may be difficult to preserve (EPUB 3 only) if this is not captured and ingested alongside the EPUB. Detecting this is not straightforward however.
  - Embedded fallbacks should provide some degree of graceful failure, however these are only required to be image or text (and not all readers support fallbacks).

3. Recommendations for Action

Recommended actions in usage and handling of the format. Recommend actions in the support or development of software applications that provide, or have the potential to provide, significant risk mitigation for the format. Note that these recommendations do not take into account other requirements such as those driven by specific British Library collections, or preservation issues like resourcing.

The eBook market is an evolving space; changes to the standards are expected and this is reflected in the monitoring recommendations below.

Handling Recommendations

- Deposited EPUBs should be validated and, in particular, the absence of encryption and DRM verified (subject to Software Recommendations, below).
- External resources referenced in the EPUB should be identified and plans made for these to be preserved also.

Knowledge Recommendations

- Consider developing and maintaining expertise in management and preservation of web content (in particular enhancing knowledge of open web standards such as XHTML and CSS upon which EPUB is based).
- Develop and maintain knowledge about the capabilities of EPUB readers, particularly in relation to the supported features of the standard.

Software Recommendations

- Assess suitable validation tools, such as EPUBCheck, to gain a better impression of their effectiveness in identifying characteristics associated with the preservation risks described above; however this must be considered against the early stage in the lifetime of EPUB.
    - It is vital to confirm that confidence can be held in EPUBCheck's ability to detect issues with fonts and detect DRM. Whilst there evidence it can detect obfuscated fonts further investigation is required around non-embedded fonts
    - Enhancements and bug fixes may require development support from the community.
- Support enhancements to format identification tools to improve version detection.
- Further investigation of rendering software for fixed-layout content.

Monitoring Recommendations

The lack of stability of EPUB and the standards on which it is based, plus the rapidly evolving EPUB market suggests that this assessment should be reviewed on an annual basis.

4. References

1. EPUB 3.2 Final Community Group Report 21 May 2019. *W3C on GitHub.* [Online] 21 May 2019. [Cited: 11 June 2019.] https://w3c.github.io/publ-epub-revision/epub32/spec/epub-spec.html.
2. **International Digital Publishing Forum.** EPUB 3.0.1. *International Digital Publishing Forum.* [Online] 2017. [Cited: 7 June 2019.] http://idpf.org/epub/301.
3. Reflowable document. *Wikipedia.* [Online] [Cited: 26 February 2015.] http://en.wikipedia.org/wiki/Reflowable_document.
4. EPUB 3.1 Specification. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 14 February 2017.] www.idpf.org/epub/31/spec/epub-spec.html.
5. **W3C.** EPUB 3.1: W3C Member Submission. [Online] 25 January 2017. [Cited: 25 June 2019.] https://www.w3.org/Submission/2017/SUBM-epub31-20170125/.
6. EPUB Packages 3.1. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 5 February 2017.] www.idpf.org/epub/31/spec/epub-packages.html.
7. EPUB 3.1 Changes from EPUB 3.0.1. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 14 February 2017.] http://www.idpf.org/epub/31/spec/epub-changes.html.
8. EPUB Content Documents 3.1. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 14 February 2017.] www.idpf.org/epub/31/spec/epub-contentdocs.html.
9. EPUB Media Overlays 3.1. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 5 February 2017.] www.idpf.org/epub/31/spec/epub-mediaoverlays.html.
10. EPUB Open Container Format (OCF) 3.1. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 5 February 2017.] www.idpf.org/epub/31/spec/epub-ocf.html.

11. EPUB Accessibility 1.0. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 15 February 2017.] www.idpf.org/epub/a11y/accessibility.html.

12. Alternate Style Tags 1.1. *International Digital Publishing Forum.* [Online] 5 January 2017. [Cited: 15 February 2017.] www.idpf.org/epub/altss-tags/altss-tags.html.

13. EPUB Packages 3.2. *W3C Github.* [Online] 21 May 2019. [Cited: 12 June 2019.] https://w3c.github.io/publ-epub-revision/epub32/spec/epub-packages.html.

14. EPUB Content Documents 3.2. *W3C on GitHub.* [Online] 21 May 2019. [Cited: 12 June 2019.] https://w3c.github.io/publ-epub-revision/epub32/spec/epub-contentdocs.html.

15. EPUB Media Overlays 3.2. *W3C on GitHub.* [Online] 21 May 2019. [Cited: 12 June 2019.] https://w3c.github.io/publ-epub-revision/epub32/spec/epub-mediaoverlays.html.

16. EPUB Open Container Format (OCF) 3.2. *W3C on GitHub.* [Online] 21 May 2019. [Cited: 12 June 2019.] https://w3c.github.io/publ-epub-revision/epub32/spec/epub-ocf.html.

17. **Bläsi, Christoph and Rothlauf, Franz.** *On the Interoperability of eBook Formats.* Brussels : European and International Booksellers Federation, 2013.

18. About us. *International Digital Publishing Forum.* [Online] [Cited: 26 February 2015.] http://idpf.org/about-us.

19. Open eBook. *Wikipedia.* [Online] [Cited: 26 February 2015.] http://en.wikipedia.org/wiki/Open_eBook.

20. EPUB 3 Changes from EPUB 2.0.1: Recommended Specification. *International Digital Publishing Forum.* [Online] 11 October 2011. [Cited: 14 February 2017.] http://www.idpf.org/epub/30/spec/epub30-changes.html.

21. Plan 2014 (HTML5): Proposed plan. *W3C.* [Online] [Cited: 26 February 2015.] http://dev.w3.org/html5/decision-policy/html5-2014-plan.html#plan.

22. HTML 5.2: W3C Recommendation. *W3C.* [Online] 14 December 2017. [Cited: 10 June 2019.] https://www.w3.org/TR/html5/introduction.html#html-vs-xhtml.

23. Cascading Style Sheets. *Wikipedia.* [Online] 20 May 2019. [Cited: 10 June 2019.] https://en.wikipedia.org/wiki/Cascading_Style_Sheets#CSS_3.

24. EPUB 3.2 Changes. *W3C on GitHub.* [Online] 21 May 2019. [Cited: 12 June 2019.] https://w3c.github.io/publ-epub-revision/epub32/spec/epub-changes.html.

25. **W3C.** EPUB 3.2 Changes: Final Community Group Specification. [Online] 8 May 2019. [Cited: 25 June 2019.] https://www.w3.org/publishing/epub3/epub-changes.html.

26. —. EPUB 3.2: Final Community Group Specification. [Online] 8 May 2019. [Cited: 25 June 2019.] https://www.w3.org/publishing/epub32/epub-spec.html.

27. EPUB 3.2: Final Community Group Specification. *Github.* [Online] 8 May 2019. [Cited: 25 June 2019.] https://w3c.github.io/publ-epub-revision/epub32/spec/epub-spec.html.

28. ISO/IEC TS 30135. *International Organization for Standardization.* [Online] November 2014. [Cited: 19 June 2019.] https://www.iso.org/search.html?q=30135.

29. **International Digital Publishing Forum.** EPUB 3.0 Published as ISO Technical Specification. [Online] 21 12 2014. [Cited: 26 05 2015.] http://idpf.org/news/epub-30-published-as-iso-technical-specification.

30. **W3C EPUB 3 Community Group.** ISO Standardization. *GitHub.* [Online] 13 November 2018. [Cited: 25 June 2019.] https://github.com/w3c/publ-cg/wiki/ISO-Certification.

31. ISO/IEC TS 30135-1:2014. Information technology -- Digital publishing -- EPUB3 -- Part 1: EPUB3 Overview. *ISO.* [Online] [Cited: 25 June 2019.] http://www.iso.org/standard/53255.html.

32. New Roadmap for Future of Publishing is Underway as W3C and IDPF Officially Combine. *W3C.* [Online] 1 February 2017. [Cited: 19 June 2019.] https://www.w3.org/2017/01/pressrelease-idpf-w3c-combination.html.en.

33. **W3C.** *Publishing@W3C.* [Online] [Cited: 25 June 2019.] https://www.w3.org/publishing/.

34. —. EPUB 3 Community Group. [Online] [Cited: 25 June 2019.] https://www.w3.org/publishing/groups/epub3-cg/.

35. **Yan, Iris.** All Things about Kindle KFX Format. *Epubor.* [Online] 16 April 2019. [Cited: 19 June 2019.] https://www.epubor.com/all-things-about-kindle-kfx-you-may-want-to-know.html.

36. Analysis: Ebook Format Wars, Round 2. *Activity Press.* [Online] 26 March 2012. http://activitypress.com/2012/03/26/analysis-ebook-format-wars-round-2/.

37. **Fisher, Tim.** What is an EPUB file? *Lifewire.* [Online] 4 May 2019. [Cited: 17 June 2019.] https://www.lifewire.com/what-is-an-epub-file-2621084.

38. EPUB. *Wikipedia.* [Online] 7 June 2019. [Cited: 17 June 2019.] https://en.wikipedia.org/wiki/EPUB.

39. Elsevier To Produce All Ebooks In Epub3. *Booktrade.info.* [Online] 23 April 2014. http://www.booktrade.info/index.php/showarticle/53556/nl.

40. BISG EPUB 3 Support Grid: Current Results. *EPUBTest.org.* [Online] [Cited: 26 February 2015.] http://www.epubtest.org/results/.

41. **Deutsche Nationalbiblothek.** Online publications - FAQ. [Online] 31 03 2015. [Cited: 27 05 2015.] http://www.dnb.de/EN/Header/Hilfe/netzpublikationenFaq.html.

42. Müssen Kindle-E-Books an die Deutsche Nationalbibliothek abgeliefert werden? [Aktualisierung]. *literaturcafe.de.* [Online] 15 01 2013. [Cited: 27 05 2015.] https://www.literaturcafe.de/muessen-kindle-e-books-an-die-deutsche-nationalbibliothek-abgeliefert-werden/.

43. **Morrissey, Amy and Kirchhoff, Sheila.** Preserving eBooks. *DPC Technology Watch Report 14.* [Online] 14 June 2014. http://dx.doi.org/10.7207/twr14-01.

44. Sustainability of Digital Formats: EPUB, Electronic Publication, Version 2. *The Library of Congress.* [Online] 31 March 2011. http://www.digitalpreservation.gov/formats/fdd/fdd000278.shtml.

45. Sustainability of Digital Formats: EPUB, Electronic Publication, Version 3. *The Library of Congress.* [Online] 18 June 2018. [Cited: 16 June 2019.] https://www.loc.gov/preservation/digital/formats/fdd/fdd000308.shtml.

46. Focus sur...rles livres téléchargeables en ePub dans Gallica. *Bibliothèque nationale de France.* [Online] 10 2013. [Cited: 27 05 2015.] http://lettre-gallica.bnf.fr/les-livres-en-epub.

47. Télécharger les documents. *Gallica.* [Online] 2019. [Cited: 17 June 2019.] https://gallica.bnf.fr/html/und/telecharger-les-documents.

48. EPUB: Hardware reading systems. *Wikipedia.* [Online] [Cited: 26 February 2015.] http://en.wikipedia.org/wiki/EPUB#Hardware_reading_systems.

49. Firefox Add-ons: EPUBReader1.5.0.4. *Mozilla.* [Online] [Cited: 2015 February 2015.] https://addons.mozilla.org/en-US/firefox/addon/epubreader/.

50. About calibre. *Calibre.* [Online] [Cited: 26 February 2015.] http://calibre-ebook.com/about.

51. AZARDI ePub3 Reader. *Infogrid Pacific.* [Online] 28 July 2016. [Cited: 19 June 2019.] http://azardi.infogridpacific.com/.

52. Adobe Digital Editions Home. *Adobe.* [Online] [Cited: 26 February 2015.] http://www.adobe.com/uk/products/digital-editions.html.

53. Readium Project Goals. *Readium.org.* [Online] [Cited: 26 February 2015.] http://readium.org/readium-project-goals.

54. txtr delivers EPUB 3 reader for iOS 7 with Readium SDK. *EPUBZone.* [Online] 22 May 2014. http://epubzone.org/news/txtr-delivers-epub-3-reader-for-ios-7-with-readium-sdk.

55. Readium. *Chrome Web Store.* [Online] 18 October 2018. [Cited: 6 June 2019.] https://chrome.google.com/webstore/detail/readium/fepbnnnkkadjhjahcafoaglimekefifl.

56. **Readium.** Readium SDK reaches Version 1.0. [Online] 05 01 2015. [Cited: 05 05 2015.] http://readium.org/news/readium-sdk-reaches-version-10.

57. **Meur, Laurent Le.** Readium 0.31 has been released. *EDRLab.* [Online] 25 September 2018. [Cited: 6 June 2019.] https://www.edrlab.org/2018/09/25/readium-0-31-has-been-released/.

58. EPUBTest. [Online] [Cited: 18 June 2019.] http://epubtest.org/.

59. **Kleinfeld, Sanders.** O'Reilly's journey to EPUB 3. *O'Reilly.* [Online] 07 02 2013. [Cited: 27 05 2015.] http://toc.oreilly.com/2013/02/oreillys-journey-to-epub-3.html.

60. **Knijff, Johan van der.** Policy-based assessment of EPUB with Epubcheck. *KB Research.* [Online] 13 March 2015. [Cited: 18 June 2019.] https://researchkb.wordpress.com/2015/03/13/policy-based-assessment-of-epub-with-epubcheck/.

61. **Pennock, M., and Day, M.** Adventures with ePub3: When Rendering goes Wrong. *Open Science Framework.* [Online] 28 December 2018. [Cited: 18 June 2019.] https://osf.io/e3gcf/.

62. Remove the NCX. *Digital Aladore - Wordpress.* [Online] 24 October 2015. [Cited: 18 June 2019.] https://digitalaladore.wordpress.com/2015/10/24/remove-the-ncx/.

63. NCX. *MobileRead Wiki.* [Online] 26 October 2017. [Cited: 18 June 2019.] https://wiki.mobileread.com/wiki/NCX.

64. EPUB 3 Overview. *International Digital Publishing Forum.* [Online] 11 October 2011. http://www.idpf.org/epub/30/spec/epub30-overview.html.

65. EPUB Publications 3.2 Final Community Group Specification: Core Media Types. *W3C.* [Online] 8 May 2019. [Cited: 19 June 2019.] https://www.w3.org/publishing/epub3/epub-spec.html#sec-core-media-types.

66. IANA Media Types. *IANA.* [Online] [Cited: 19 12 2016.] http://www.iana.org/assignments/media-types/media-types.xhtml.

67. Validation tool for EPUB. *Github.* [Online] [Cited: 17 June 2019.] https://github.com/w3c/epubcheck.

68. *flightcrew.* [Online] [Cited: 9 February 2017.] https://github.com/Sigil-Ebook/flightcrew.

69. *FlightDeck.* [Online] [Cited: 26 February 2015.] https://ebookflightdeck.com/.

70. Checking out the FlightDeck checker for epub files. *Roger Packer.* [Online] [Cited: 26 February 2015.] http://rogerpacker.com/blog/checking-flightdeck-checker/.

71. *The SCAPE Project.* [Online] [Cited: 26 February 2015.] http://www.scape-project.eu/.

72. Flint. *Community Owned digital Preservation Tool Registry (COPTR).* [Online] [Cited: 26 February 2015.] http://coptr.digipres.org/DRMLint.

73. Flightcrew Issue #58079c2: Use bk.epub_version() to restrict to epub2. *Github.* [Online] [Cited: 10 February 2017.] https://github.com/Sigil-Ebook/flightcrew/commit/58079c2de42e2f143ef2ba850ebe99266428a4d4.

74. Contributors to epubcheck. *GitHub.* [Online] 17 June 2019. [Cited: 17 June 2019.] https://github.com/w3c/epubcheck/graphs/contributors.

75. Not checking mimetype is uncompressed in Zip #303. *GitHub.* [Online] 10 October 2018. [Cited: 17 June 2019.] https://github.com/w3c/epubcheck/issues/303.

76. **Deltour, Romain.** EPUBCheck v4.2.0-beta. *GitHub.* [Online] 25 February 2019. [Cited: 17 June 2019.] https://github.com/w3c/epubcheck/releases/tag/v4.2.0-beta.

77. EPUB 3 Samples Table. *GitHub.* [Online] 9 June 2017. [Cited: 17 June 2019.] http://idpf.github.io/epub3-samples/30/samples.html.

78. Python library to extract 'epub' information. *Stack Overflow.* [Online] [Cited: 26 February 2015.] http://stackoverflow.com/questions/3114786/python-library-to-extract-epub-information.

79. ePub. *MobileRead.* [Online] [Cited: 26 February 2015.] http://wiki.mobileread.com/wiki/EPub#Conversion_tools.

80. **W3C.** EPUB Specifications and Projects. [Online] 12 February 2019. [Cited: 25 June 2019.] http://www.idpf.org/epub/dir/.

81. ePub. *MobileRead.* [Online] 7 May 2019. [Cited: 17 June 2019.] http://wiki.mobileread.com/wiki/EPub.

82. Open Publication Structure (OPS) 2.0.1 v1.0.1 Recommended Specification: MIME Media Types. *International Digital Publishing Forum.* [Online] 4 September 2010. http://www.idpf.org/epub/20/spec/OPS_2.0.1_draft.htm#Section1.3.7.

83. Epubcheck Issues: Handling of foreign resources with fallbacks #511. *GitHub.* [Online] [Cited: 18 12 2016.] https://github.com/IDPF/epubcheck/issues/511.

84. **Knijff, Johan van der.** EPUB for archival preservation. *Open Preservation Foundation.* [Online] 20 July 2012. http://openpreservation.org/system/files/epubForArchivalPreservation20072012ExternalDistribution.pdf.

85. FairPlay. *Wikipedia.* [Online] [Cited: 26 February 2015.] http://en.wikipedia.org/wiki/FairPlay.

86. Adobe Digital Editions. *Wikipedia.* [Online] [Cited: 26 February 2015.] http://en.wikipedia.org/wiki/Adobe_Digital_Editions.