

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook Summary Assessment

Document History

Date	Version	Author	Change Details
03/02/17	1.0	Simon Whibley, Peter May, Paul Wheatley	Initial release.

British Library Digital Preservation Team
digitalpreservation@bl.uk



This work is licensed under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

1. Introduction

This document provides a high level assessment of eBook file formats with respect to long term digital preservation. It is intended to provide a broad but shallow overview of formats within the eBook sector in order to give context to more detailed single format assessments.

1.1 Scope

This document will primarily focus on the main eBook formats, taking into account eBook format frequency in BL collections. It focuses primarily on preservation considerations with respect to file formats and does not cover the broader preservation issues relating to managing eBooks (such as licencing/ownership of commercially published eBooks).

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 What is an eBook?

Wikipedia describes an eBook as “a book-length publication in digital form, consisting of text, images, or both, readable on computers or other electronic devices.” Whereas the PDF format captures the content and layout of a printed page in digital form, an eBook format is designed to capture content in a manner that will be suitable for viewing on an electronic device, whether that might be a desktop computer, a mobile phone or a dedicated eBook reader. Predominantly text based material tends to be represented in a reflowable text [1] form, allowing the rendering of an eBook to be tailored to viewing devices with different screen sizes or resolutions. However, material that is more image-based or structured – such as newspapers or graphical novels – is sometimes captured in a non-reflowable form (sometimes known as fixed-layout or FXL [2]).

2. Background and market context

A broad impression of the evolution of these file formats and the market drivers influencing their future

A variety of eBook file formats have emerged with the aim of representing a written work in a form suitable for display on an electronic device, rather than the printed page. Beginning primarily as a form dedicated to textual content, such as with the landmark Project Gutenberg [3] in 1971, the market for eBook content has continued to grow and formats have been developed to support still image, sound, video and other multimedia content.

The emergence in the late 2000s of dedicated eBook readers [4] with e-paper [5] displays such as Amazon's Kindle [6] kick-started a commercial market for published eBooks. The academic world has embraced this market, as have memory organisations who have published content digitised from their archives. There is also a sizeable self-publishing movement [7] taking advantage of the easy ability to create and publish eBooks. This has led to the generation of supporting information, tools and services for creating and working with eBooks [8].

In general, the market for eBooks has grown quickly with publishers announcing ever greater proportions of sales via e rather than paper books. The latest Association of American Publishers Bookstats survey stated that “eBooks grew 45% since 2011 and now constitute 20% of the Trade market, playing an integral role in 2012 Trade revenue” [9]. Predictions from Price Waterhouse Coopers suggest eBook sales will overtake printed book sales by 2018 [10]. Despite these predictions, though, eBook sales fell for the first time in 2015 for the UK's “big five” publishers according to the Bookseller [11] [12]. Such reports should be interpreted in context however - these sales are only part of the larger eBook publishing market, with independent self-publishers not included in these figures; in other words they don't necessarily represent the *overall* state of eBook sales.

3. eBook format evolution

A list of the main formats in this sector, descriptions of their key characteristics and notes on relevant preservation issues

There are a considerable number of eBook file formats that exist within an evolving marketplace. Although most formats utilise open standards to some degree, many also deploy proprietary modifications and/or DRM schemes [13] to ring-fence their ecosystems which can cause problems when an eBook publisher ceases trading [14]. Wikipedia provides a high level overview of 30 different eBook formats [15] and more detailed descriptions of many of the more popular formats can be found on the MobileRead wiki [16].

Despite Steve Jobs announcement in 2010 that "We use the epub format: It is the most popular open book format in the world." [17], the current market situation has been widely acknowledged as a 'format war', with a variety of suppliers, publishers and hardware manufacturers pushing competing formats [18]. Kirchoff and Morrissey note the preservation risks associated with this situation: "There are sometimes quite serious preservation risks associated with the formats in which eBooks are created. This is particularly the case for proprietary formats, particularly those closely tied to a particular commercial vendor's hardware platform and distribution system. The variants, deliberate and inadvertent, in even standard eBook formats such as PDF and EPUB bring to mind the 'browser wars' which took place in the early days of the Internet or the even earlier videotape format wars of Betamax versus VHS" [13].

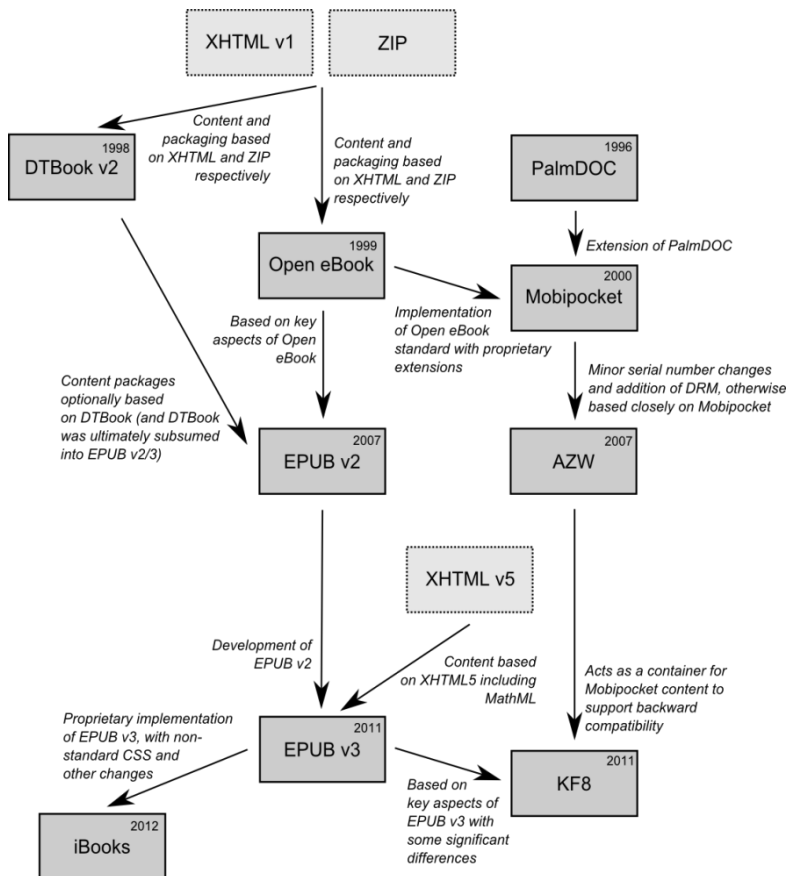


Figure 1: eBook file format family tree, centred on EPUB

A number of the major eBook formats share their roots in the same open standards, such as XHTML, but divergence and incompatibility has subsequently been introduced in proprietary versions. The origins of the EPUB format and its relationships with a number of its significant contemporaries highlights the complexity of the eBook format ecosystem. This is outlined in Figure 1, although it should be noted that for practical reasons a number of related formats have not been included!

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

In 2011, the impact of the eBook format wars was viewed as a confusing situation for consumers [19], with significant restrictions on usage of eBooks across different devices [20], and an uncertain future for reader devices, publishers and the actual eBook file formats themselves. Since then it could be argued that EPUB, PDF and Mobi are all strong formats in their own right and by virtue of the popularity of Amazon's Kindle device, the AZW/KF8/KFX format.

4. eBook formats

A summary of the key formats with notes on their status and an impression of their preservation worthiness

4.1 Main formats

eBook format	Notes on format origin and design	Status	Notes on preservation
EPUB (.epub)	Based on Open eBook. ZIP container with a mix of XHTML, CSS, XML and other embedded content (e.g. SVG images).	The closest there is to an open and well supported eBook standard. Forms the basis for several other proprietary eBook formats. EPUB 3 allows the creation of multimedia items as it is based on HTML5 and CSS3 [20]	Encouraging outlook for preservation with some caveats such as potential for DRM. Possibility of dropping support for EPUB2 in EPUB 3.1+ with the removal of NCX [21]. Good tool support including characterisation/validation. See (BL) EPUB File Format Assessment [22]. DRM issues depending on the book store (see Nook [14]). Fixed-layout epub not reflowable [23] and some fixed layout files cause access problems [24]
Mobipocket (.mobi, .prc)	Has an origin in both the PalmDOC format (the .prc version) and Open ePub [25].	Official support for Mobipocket hardware and software ended when Mobipocket was acquired by Amazon [26]. It is currently used by Amazon with a slightly different DRM scheme and called AZW. Ereader / PalmDOC was originally an extension of this format [27].	Potential for DRM to be a concern [25]. Lack of standardisation and gradual development of this proprietary format is a concern. Viewing and conversion support provided by open source Calibre tool, although tool support in general not as widespread as many other eBook formats.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook format	Notes on format origin and design	Status	Notes on preservation
AZW (.azw)	Based closely on EPUB version 2 with the addition of DRM [28].	The Amazon Kindle's AZW format is basically just the Mobipocket format with a slightly different serial number scheme.	Potential for DRM a concern. Appears to be in the process of being replaced by KF8. Amazon uses AZW as a generic term for Kindle eBooks so they may identify only AZW as the format supported but expect that the device will also support Topaz files (also called AZW1) and perhaps others (AZW4). However KF8 (also called AZW3) will be specifically identified if it is supported. [28]
KF8 (.aw3, .kf8)	Based to some degree on EPUB version 3 but divergent from that standard [29], with proprietary extensions, some non-standard HTML and support for backward compatibility with AZW [30]. Also known as AZW3 [31]	Introduced with Amazon's Kindle Fire in 2011, and acts as a replacement for Amazon's AZW/Mobipocket format [32]. In August 2015, all the Kindle e-readers released within the last two years were updated with a new typesetting and layout engine that adds hyphens, kerning (improved word spacing), and ligatures to the text; e-books that support this engine require the use of the "Kindle Format 10" (KFX) file format. E-books that support the enhanced typesetting format are indicated on the book's description page. [6] KFX's new features are only visible on devices that support the new format [33].	Potential for DRM and proprietary nature of the format as well as the use of non-standard HTML are concerns. Can be converted to KFX using Calibre [34].
iBooks (.ibooks)	Based on the EPUB version 3 with non-standard CSS and other proprietary extensions [35].	Latest format launched from Apple that diverges from the EPUB standard. Described by some as a deliberate tactic to sabotage EPUB [36]. The iBooks format is created using the authoring application iBooks Author [20].	Potential for DRM and proprietary nature of format a concern. Supplied via iTunes and can only be viewed in iBooks application.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook format	Notes on format origin and design	Status	Notes on preservation
Portable Document Format (.pdf)	Created by Adobe and subsequently ISO standardised.	<p>Designed as a way of representing a printed page in a digital file.</p> <p>Often provided by publishers as an alternative to their preferred format.</p> <p>Support for reflowable text is uncommon on mobile devices [37].</p>	<p>Some concerns (DRM). See (BL) PDF File Format Assessment [38].</p> <p>PDF files are supported by almost all modern eBook readers, tablets and smartphones.</p> <p>However, PDF reflow based on Tagged PDF, as opposed to reflow based on the actual sequence of objects in the content-stream is not yet commonly supported on mobile devices. Such reflow options as may exist are usually found under "view" options, and may be called "word-wrap" [37]</p>

4.2 Other formats

Other less common, obsolete, unsupported or more document-based eBook formats include:

eBook format	Notes on format origin and design	Status	Notes on preservation
Apabi (.xeb, .ceb)	APABI is a format created by Founder Electronics which is popular in the Chinese eBook market [15].	<p>The format can be read using the Apabi Reader software, and created with the Apabi Publisher.</p> <p>Both .xeb and .ceb files are encoded binary files [15].</p>	
Apple Fixed Layout EPUB (.epub)	Apple proprietary epub variant released in 2010 as part of iBooks 1.2 [20].	Obsolete as its features were replaced by EPUB 3 [20].	DRM issues.
BBeb (.lrf, .lrx)	Broadband eBooks (BBeb) is a Sony proprietary format.	Replaced by a conversion to EPUB in 2010 [39].	DRM issues.
Comic book archive (.cbr, .cbz, .cbr, .cba, .cb7)	Type of archive file designed to aid the sequential viewing of comic books	Not a distinct file format, it is normally made up of JPEG or PNG [40]	

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook format	Notes on format origin and design	Status	Notes on preservation
DAISY DTBook / Digital Talking Book	Designed to support accessibility. Later versions based on open standards [41].	Via DAISY's participation in the IDPF [42], DTBook and its functionality has been subsumed into EPUB.	Still supported by creation, migration and viewing applications [43]. Software support could decline if EPUB replaces DTBook, but the existence of open source tools is likely to mitigate major preservation concerns. DAISY is already aligned with the EPUB open standard, and is expected to fully converge with its forthcoming EPUB3 revision [42]
DJVu (.djvu, .djv)	Format specialized for storing scanned documents	A compressed image format [44], it includes advanced compressors optimized for low-colour images, such as text documents. Individual files may contain one or more pages [45].	Cannot be reflowed [45].
DOC (.doc) / DOCX (.docx)	Word processor format relating to Microsoft's Word application.	Not designed as an eBook format but supports reflowable text and is well supported by a few eBook readers. DOCX is an XML-based version [15].	Some concerns including possible security problems.
Ereader / PalmDOC (.pdb)	A special eBook version of the format used in many of Palm's application.	In 2009, Barnes & Noble implied that eReader would be their preferred eBook format but subsequently sold mostly in EPUB format [15].	Concerns with DRM.
FictionBook (.fb2)	An open XML-based eBook format which originated and gained popularity in Russia	Last release was in 2008 [46].	Single XML file. No DRM. Can be converted by Calibre [47].

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook format	Notes on format origin and design	Status	Notes on preservation
HTML	<p>HTML is the mark-up language used for most web pages.</p> <p>eBooks using HTML can be read using a Web browser.</p> <p>The specifications for the format are available without charge from the W3C [15].</p>	<p>HTML is not a particularly efficient format to store information in, requiring more storage space for a given work than many other formats.</p> <p>“However, several eBook formats including the Amazon Kindle, Open eBook, Compiled HTML, Mobipocket and EPUB store each book chapter in HTML format, then use ZIP compression to compress the HTML data, images, metadata and style sheets into a single, significantly smaller, file” [15]</p>	<p>Displaying HTML files can be complicated due to the wide range of standards it encompasses.</p> <p>Many of the features supported, such as forms, are not relevant to eBooks [15].</p> <p>Potential for DRM issues.</p>
INF (.inf)	<p>IBM created this eBook format and it was used for its operating systems [15].</p>	<p>The INF files were often digital versions of printed books that came with some bundles of OS/2 and other products.</p>	<p>Compact and fast, it supports images, reflowable text, table and various list formats.</p> <p>Initially limited to a dedicated viewer and compiler; open source viewers did appear later [15].</p>
Microsoft Compiled HTML Help (.chm)	<p>A Microsoft proprietary online help format, consisting of a collection of HTML pages, an index and other navigation tools [48].</p>	<p>The files are compressed and deployed in a binary format with the extension .CHM, for Compiled HTML.</p> <p>The format is chiefly used for software documentation [48].</p> <p>Replaced by Microsoft Help 2.</p>	
Microsoft LIT (.lit)	<p>Is only useable with the Microsoft Reader application in Windows [15].</p>	<p>Was discontinued at the end of August 2012 [15].</p>	<p>Concerns with DRM.</p> <p>Reliance on the MS Reader application where DRM is present. Other third party readers can read unprotected LIT files [15].</p>
Newton eBook (.pkg)	<p>More popularly referred to as the Apple Newton book.</p> <p>All systems which run the Newton Operating System can access these (e.g. the Newton Messagepad) [15].</p>	<p>The Apple Newton was cancelled in 1998.</p>	<p>Open format which was released to the public prior to their takeover by Apple [15].</p>

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook format	Notes on format origin and design	Status	Notes on preservation
Open eBook/Open Electronic Package (.opf)	Legacy XML-based format. Forerunner of EPUB based on open standards [49].	Now somewhat obsolete in terms of current usage, but support remains in key software applications [49].	Still supported by migration and viewing applications. Can have DRM.
Open XML Paper Specification (.oxps, .xps)	OpenXPS is an open specification for a page description language and a fixed-document format developed by Microsoft in 2006 [50].	In June 2009, Ecma International adopted it as international standard ECA-388 [51].	It is deliberately limited to sequences of: Glyphs (a fixed run of text), Paths (a geometry that can be filled, or stroked, by a brush), and Brushes (a description of a shaped brush used to in rendering paths). The chance of the accidental introduction of malicious content is therefore lessened [15].
Plain text (.txt)	No formatting options (contains only ASCII or Unicode Text) [15].	The first eBooks in history from Project Gutenberg were in this format.	Does not support DRM. Easy to convert to other formats [15].
Plucker (.pdb).	Offline Web and free eBook reader for Palm OS based handheld devices, Windows Mobile (Pocket PC) devices, and other PDAs [52].	Website doesn't seem to have been updated since 2007 [53].	Concerns with DRM.
Postscript (.ps)	A page description language used in the electronic and desktop publishing areas for defining the contents and layout of a printed page [54]	Developed by Adobe but has been around since 1982. Many office printers directly support interpreting PostScript and printing the result [54].	Common in Unix environments.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook format	Notes on format origin and design	Status	Notes on preservation
Rich Text Format (.rtf)	A document file format from Microsoft that is supported by many eBook readers	Format still maintained but not updated since 2008 and has discontinued enhancements [55].	Widely supported, reflowable and easily edited. It can be easily converted to other eBook formats, increasing its support. Some concerns with malware issues [55].
SSReader (.pdg)	The digital book format used by the Beijing Superstar Electric Company digital library company [15].	A proprietary raster image compression and binding format, with OCR plug-in modules. The company scanned a huge number of Chinese books in the China National Library [15].	A detail of the format has not been published [15].
TEI Lite (.xml)	A XML-based text electronic text format which is a customisation of the TEI (Text Encoding Initiative) encoding scheme [15].	TEI is a consortium which collectively develops and maintains a standard for the representation of texts in digital form and has been running in some form since 1987. Has adoption amongst a large number of projects and universities [56]	
TomeRaider (.tr2, .tr3)	EBook reader and cross-platform reference viewer for handheld devices. Versions available for Android, Windows Mobile (aka Pocket PC), Palm, Psion, Symbian, iPhone as well as Windows PCs [57] [15].	TomeRaider was created by UK company, Yadabyte. First version of TomeRaider was developed in 1999, developed up to TomeRaider version 3 [57]. No longer supported or developed [58]	Proprietary format with DRM.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

eBook format	Notes on format origin and design	Status	Notes on preservation
Topaz (.tpz)	Rare Amazon format for Kindles (also known as AZW1). It is the output of an automatic conversion of scanned text images with each page being a separate document [59].	The format is unrelated to AZW or MOBI despite both using the PalmOS container format and contain Amazon DRM [59].	Proprietary format with Amazon's DRM. Reflowable. [59]

4.3 Multimedia and Enhanced eBooks

A multimedia (also known as complex or enhanced) eBook is a combination of media and book content formats. The term is used in contrast to media which “only utilise traditional forms of printed or text books” [15]. Multimedia eBooks could include a combination of text, audio, images, video or interactive content formats.

Some of the formats being used to produce enhanced eBooks in this changing landscape include:

- ◆ Next generation EPUB (EPUB3) and Kindle (including KF8) formats
- ◆ Interactive PDF format
- ◆ Apps for mobile devices (e.g. Apple OS/Android OS/Windows OS)
- ◆ Fixed layout eBooks
- ◆ Web apps [60]

All the above five formats/technologies can integrate multimedia content (audio and video) with restrictions placed on them by which audio or video formats the reader application associated with it can play.

The KF8 format is a special case because despite audio and video integration being possible, it cannot be played back on Kindle devices - only via the iOS app on Apple devices (strangely not on the Android version). It does have a proprietary function allowing image magnification as pop-ups though (particularly useful for digital graphic novels) [20].

EPUB3 offers options for dynamic content (dialogues, animation) but only as an option due its probable lack of support across all reader devices.

Apple's iBooks offers generally the same options as EPUB3 [20].

5. eBook Readers

A summary of main portable hardware devices for viewing eBook formats

The main devices currently available are:

- ◆ Amazon Kindle: Kindle (AZW, TPZ), TXT, MOBI, PRC and PDF natively; HTML and DOC through conversion [17]
- ◆ Apple iPad: EPUB, PDF, HTML, DOC (plus iPad Apps, which could include Kindle and Barnes & Noble readers) [17]
- ◆ Barnes & Noble Nook: EPUB, PDB, PDF (no longer on sale in the UK as of March 2016 [61])
- ◆ Kobo: EPUB, PDF, HTML, DOCX, TXT [4]
- ◆ Sony Reader: EPUB, PDF, TXT, RTF; DOC through conversion [17]

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

Wikipedia provides a more detailed and intensive comparison of eBook readers including file format support information for readers current and obsolete [4].

6. Adoption and Usage

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

To date, few national libraries have experience with ingesting and processing eBooks, often due to a lack of inclusion within legal deposit provision. Here are some examples of current practice:

On behalf of the UK Legal Deposit Libraries, the British Library has been collecting eBooks under legal deposit since 2013 [62]; with PDF or EPUB the preferred formats. The Library's Publisher Submissions Portal guidance states the Portal currently supports the following file formats: PDF, Microsoft Word and EPUB 2 [63].

The Bibliothèque nationale de France (BnF – National Library of France) does have a complete legal deposit chain for eBooks. Their e-distribution agent acts as a facilitator for the library, ensuring the files are maintained in an open format (PDF and/or EPUB) so that conversion to proprietary formats never occurs and DRM is not applied [65].

The National Library of the Netherlands (KB) has no legal deposit legislation to support the receipt of eBooks, but their policy of voluntary collection asks that in exchange for making the eBooks restricted to on-site users DRM is removed from eBooks deposited from contributing publishers. Current formats received include PDF and Microsoft Word and an expectation to receive content EPUB2 and EPUB3 [13].

The Library of Congress only has a small percentage of their eBook collection available in digital form [64]. It is still at the start of introducing eBooks to its collections, mainly because they are not included in their interim regulations for electronic legal deposit. It does also receive some PDFs, and their collections include content in HTML/XHTML, XML/TEI and EPUB2 formats [13].

7. Preservation Risk Summary

A summary of preservation risks

As previously stated, this is a high level assessment providing an overview of eBook formats and not intended to provide detailed risk accounts for each; consult the individual format assessments for risk summaries specific to each format.

A few main formats, both open source and proprietary in nature, have risen to the forefront of the eBook formats landscape. These formats continue to develop, however, as they do there is potential for divergence from existing standards or updates that affect backwards compatibility.

Use of DRM and the proprietary nature of some formats both pose a long-term preservation concern, hindering access. Rendering support may be tied to specific eBook reader devices/applications; conversion from these formats to an open eBook flavour (e.g. EPUB) may not be trivial.

8. References

1. Reflowable document. *Wikipedia*. [Online] [Cited: 26 February 2015.] http://en.wikipedia.org/wiki/Reflowable_document.
2. Hoffelder, Nate. The Problem with Fixed Layout eBooks. *The Digital Reader*. [Online] 22 February 2015. [Cited: 4 April 2016.] <http://the-digital-reader.com/2015/02/22/the-problem-with-fixed-layout-ebooks/>.
3. *Project Gutenberg*. [Online] [Cited: 27 February 2015.] <http://www.gutenberg.org/>.
4. Comparison of e-book readers. *Wikipedia*. [Online] [Cited: 27 February 2015.] http://en.wikipedia.org/wiki/Comparison_of_e-book_readers.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

5. Electronic paper. *Wikipedia*. [Online] [Cited: 27 February 2015.] http://en.wikipedia.org/wiki/Electronic_paper.
6. Amazon Kindle. *Wikipedia*. [Online] [Cited: 27 February 2015.] http://en.wikipedia.org/wiki/Amazon_Kindle.
7. **Campbell, Lisa**. Self-published titles '22% of UK e-book market'. *The Bookseller*. [Online] 23 March 2016. [Cited: 1 February 2017.] <http://www.thebookseller.com/news/self-published-titles-22-e-book-market-325152>.
8. Self-publishing. *The Guardian*. [Online] [Cited: 27 February 2015.] <http://www.theguardian.com/books/self-publishing>.
9. Trade Publishing Increases Nearly 7%, Ebooks Grow Nearly 45%, According to "BookStats" Industry Survey for 2012. *BookBusiness*. [Online] 15 May 2013. [Cited: 9 March 2016.] <http://www.bookbusinessmag.com/article/trade-publishing-increases-nearly-7-ebooks-grow-nearly-45-according-bookstats-industry-survey-2012/>.
10. Ebooks on course to outsell printed editions in UK by 2018. *The Guardian*. [Online] 4 June 2014. <http://www.theguardian.com/books/2014/jun/04/ebooks-outsell-printed-editions-books-2018>.
11. **Tivnan, Tom**. E-book sales abate for Big Five. *The Bookseller*. [Online] 29 January 2016. [Cited: 25 January 2017.] <http://www.thebookseller.com/blogs/e-book-sales-abate-big-five-321245>.
12. Ebook sales falling for the first time, finds new report. *Guardian*. [Online] 3 February 2016. [Cited: 9 March 2016.] <http://www.theguardian.com/books/2016/feb/03/ebook-sales-falling-for-the-first-time-finds-new-report>.
13. **Kirchoff, Amy & Morrissey, Sheila**. Preserving eBooks. *Digital Preservation Coalition*. [Online] June 2014. [Cited: 1 April 2016.] <http://dx.doi.org/10.7207/twr14-01>.
14. **Novak, Matt**. You Don't Own Your Ebooks. *Gizmodo*. [Online] 7 March 2016. [Cited: 8 March 2016.] <http://paleofuture.gizmodo.com/you-dont-own-your-ebooks-1763333576>.
15. Comparison of e-book formats. *Wikipedia*. [Online] [Cited: 27 February 2015.] http://en.wikipedia.org/wiki/Comparison_of_e-book_formats.
16. E-book formats. *MobileRead Wiki*. [Online] [Cited: 27 February 2015.] http://wiki.mobileread.com/wiki/E-book_formats.
17. **Buchanan, Matt**. Giz Explains: How You're Gonna Get Screwed By Ebook Formats. *Gizmodo*. [Online] 10 March 2010. [Cited: 23 February 2016.] <http://gizmodo.com/5478842/giz-explains-how-youre-gonna-get-screwed-by-ebook-formats>.
18. Analysis: Ebook Format Wars, Round 2. *Acitivity Press*. [Online] 26 March 2012. <http://activitypress.com/2012/03/26/analysis-ebook-format-wars-round-2/>.
19. Which is the best format for ebooks? *The Guardian*. [Online] 15 September 2011. [Cited: 4 April 2016.] <http://www.theguardian.com/technology/askjack/2011/sep/15/ebook-format-drm-kindle>.
20. **Bläsi, Christoph & Rothlauf, Franz**. On the Interoperability of eBook Formats. *wi.bwl.uni-mainz.de*. [Online] April 2013. [Cited: 1 April 2016.] <http://wi.bwl.uni-mainz.de/publikationen/InteroperabilityReportGutenbergfinal07052013.pdf>.
21. **Van der Knijff, Johan**. The future of EPUB? A first look at the EPUB 3.1 Editor's draft. *KB Research*. [Online] 10 March 2016. [Cited: 4 April 2016.] <http://blog.kbresearch.nl/2016/03/10/the-future-of-epub-a-first-look-at-the-epub-3-1-editors-draft/#more-1666>.
22. **Day, Michael, May, Peter & Wheatley, Paul**. EPUB Format Preservation Assessment. <http://wiki.dpconline.org/>. [Online] 4 September 2015. [Cited: 2 March 2016.] http://wiki.dpconline.org/images/a/a9/EPUB_Assessment_v1.2.pdf.
23. Fixed layout ePub. *MobileRead Wiki*. [Online] 4 March 2016. [Cited: 4 April 2016.] http://wiki.mobileread.com/wiki/Fixed_layout_ePub.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

24. **Van der Knjiff, Johan.** Valid, but not accessible EPUB: crazy fixed layouts. *KB Research*. [Online] 4 April 2016. [Cited: 4 April 2016.] <http://blog.kbresearch.nl/2016/04/04/valid-but-not-accessible-epub-crazy-fixed-layouts/>.
25. MOBI. *MobileRead Wiki*. [Online] 14 May 2015. [Cited: 2 March 2016.] <http://wiki.mobileread.com/wiki/MOBI>.
26. Mobipocket. *Wikipedia*. [Online] [Cited: 27 February 2015.] <http://en.wikipedia.org/wiki/Mobipocket>.
27. PalmDOC. *MobileRead Wiki*. [Online] 5 November 2014. [Cited: 4 April 2016.] <http://wiki.mobileread.com/wiki/PalmDOC>.
28. AZW. *MobileRead Wiki*. [Online] 12 August 2015. [Cited: 18 February 2016.] <http://wiki.mobileread.com/wiki/AZW>.
29. The New Kindle Format 8 (KF8). *Digital Pubbing*. [Online] 6 March 2012. <http://www.digitalpubbing.com/the-new-kindle-format-8-kf8/>.
30. KF8. *MobileRead Wiki*. [Online] [Cited: 27 February 2015.] <http://wiki.mobileread.com/wiki/KF8>.
31. Kindle Format 8. *Amazon*. [Online] 2016. [Cited: 18 February 2016.] <http://www.amazon.com/gp/feature.html?docId=1000729511>.
32. **Jerney, John.** Exploring Ebook Formats—Amazon Kindle (AZW/MOBI). *ePublish Yourself!* [Online] 25 February 2012. <http://www.epublishyourself.com/exploring-ebook-formats-amazon-kindle-azwmobi/>.
33. **Henkel, Guido.** The horrors of Kindle Format X. *Guido Henkel*. [Online] 27 October 2015. [Cited: 26 January 2017.] <http://guidohenkel.com/2015/10/the-horrors-of-kindle-format-x/>.
34. How to Use Calibre to Convert eBooks to KFX Format (for the Enhanced Kindle Typography). *The Digital Reader*. [Online] 28 March 2016. [Cited: 26 January 2017.] <http://the-digital-reader.com/2016/03/28/how-to-use-calibre-to-convert-ebooks-to-kfx-format-for-the-enhanced-kindle-typesetting/>.
35. **Bjarnason, Baldur.** The iBooks 2.0 textbook format. [Online] 19 January 2012. <http://www.baldurbjarnason.com/notes/the-ibooks-textbook-format/>.
36. **Bott, Ed.** How Apple is sabotaging an open standard for digital books. *ZD Net*. [Online] 22 January 2012. [Cited: 4 April 2016.] <http://www.zdnet.com/article/how-apple-is-sabotaging-an-open-standard-for-digital-books/>.
37. Comparison of e-book formats: Portable Document Format. *Wikipedia*. [Online] http://en.wikipedia.org/wiki/Comparison_of_e-book_formats#Portable_Document_Format.
38. **Wheatley, Paul, May, Peter & Pennock, Maureen.** PDF Format Preservation Assessment. <http://wiki.dpconline.org/>. [Online] 4 September 2015. [Cited: 2 March 2016.] http://wiki.dpconline.org/images/e/e8/PDF_Assessment_v1.3.pdf.
39. BBeB. *Wikipedia*. [Online] 15 December 2013. [Cited: 1 April 2016.] <https://en.wikipedia.org/wiki/BBeB>.
40. Comic book archive. *Wikipedia*. [Online] 3 February 2016. [Cited: 29 February 2016.] https://en.wikipedia.org/wiki/Comic_book_archive.
41. About Us. *DAISY Consortium*. [Online] [Cited: 27 February 2015.] http://www.daisy.org/about_us.
42. DAISY Standard. *DAISY Consortium*. [Online] [Cited: 27 February 2015.] <http://www.daisy.org/>.
43. Conversion Tools and Services. *DAISY Consortium*. [Online] [Cited: 27 February 2015.] <http://www.daisy.org/tools/conversion>.
44. DJVU to PDF. *Online Converter*. [Online] 2016. [Cited: 26 January 2017.] <https://www.onlineconverter.com/djvu-to-pdf>.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

45. DjVu. *Wikipedia*. [Online] 27 February 2016. [Cited: 29 February 2016.] <https://en.wikipedia.org/wiki/DjVu>.
46. Fiction Book. *Wikipedia*. [Online] 19 March 2016. [Cited: 5 April 2016.] <https://en.wikipedia.org/wiki/FictionBook>.
47. FB2. *MobileRead Wiki*. [Online] 18 March 2016. [Cited: 4 April 2016.] <http://wiki.mobileread.com/wiki/FB2>.
48. Microsoft Compiled HTML Help. *Wikipedia*. [Online] 8 February 2016. [Cited: 5 April 2016.] https://en.wikipedia.org/wiki/Microsoft_Compiled_HTML_Help.
49. Open eBook. *Wikipedia*. [Online] 21 January 2015. [Cited: 19 February 2016.] https://en.wikipedia.org/wiki/Open_eBook.
50. Open XML Paper Specification. *Wikipedia*. [Online] 24 November 2016. [Cited: 26 January 2017.] https://en.wikipedia.org/wiki/Open_XML_Paper_Specification.
51. Standard ECMA-388 Open XML Paper Specification. *Ecma International*. [Online] June 2009. [Cited: 26 January 2017.] <http://www.ecma-international.org/publications/standards/Ecma-388.htm>.
52. Plucker. *Wikipedia*. [Online] 28 February 2016. [Cited: 5 April 2016.] <https://en.wikipedia.org/wiki/Plucker>.
53. Plucker. *plkr.org*. [Online] 2 November 2007. [Cited: 5 April 2016.] <http://www.plkr.org/news>.
54. PostScript. *Wikipedia*. [Online] 5 April 2016. [Cited: 5 April 2016.] <https://en.wikipedia.org/wiki/PostScript>.
55. Rich Text Format. *Wikipedia*. [Online] 1 April 2016. [Cited: 5 April 2016.] https://en.wikipedia.org/wiki/Rich_Text_Format.
56. TEI Lite. *Text Encoding Initiative*. [Online] 25 October 2012. [Cited: 11 April 2016.] <http://www.tei-c.org/Guidelines/Customization/Lite/>.
57. TomeRaider. *Wikipedia*. [Online] 9 March 2016. [Cited: 19 April 2016.] <https://en.wikipedia.org/wiki/TomeRaider>.
58. TomeRaider. *Infodisiac*. [Online] [Cited: 5 April 2016.] <http://infodisiac.com/Wikipedia/TomeRaider.html>.
59. Topaz. *MobileRead Wiki*. [Online] 3 February 2015. [Cited: 19 April 2016.] <http://wiki.mobileread.com/wiki/Topaz>.
60. **Taylor, Martyn**. How to produce illustrated, multimedia ebooks. *DigitalPublishing101*. [Online] 8 July 2015. [Cited: 19 April 2016.] <http://digitalpublishing101.com/how-to-produce-illustrated-multimedia-ebooks/>.
61. **Williams, Zoe**. Goodnight and good Nook: farewell to a beloved e-reader. *Guardian*. [Online] 7 March 2016. [Cited: 8 March 2016.] <http://www.theguardian.com/technology/shortcuts/2016/mar/07/goodnight-good-nook-farewell-e-reader-ebook-kindle>.
62. Legal deposit for websites and electronic publications. *The British Library*. [Online] [Cited: 20 April 2016.] <http://www.bl.uk/aboutus/legaldeposit/websites/>.
63. Depositing electronic publications. *The British Library*. [Online] [Cited: 20 April 2016.] <http://www.bl.uk/aboutus/legaldeposit/websites/elecpubs/>.
64. **Hohman, J. Cheyenne**. Finding E-books: A Guide. *Library of Congress Web Guides*. [Online] 18 March 2011. [Cited: 2 February 2017.] <https://www.loc.gov/rr/program/bib/ebooks/>.
65. BNF Digital Roadmap March 2016. *Bibliothèque nationale de France*. [Online] March 2016. [Cited: 2 February 2017.] http://www.bnf.fr/documents/digital_roadmap.pdf.
66. ISO/DIS 32000-2 PDF 2.0. *ISO*. [Online] [Cited: 24 February 2015.] http://www.iso.org/iso/catalogue_detail.htm?csnumber=63534.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

67. **Albertini, Ange.** A PDF 101 document walk through. [Online] [Cited: 24 February 2015.] <http://imgur.com/a/PbN8H#7>.
68. Portable Document Format. *Wikipedia*. [Online] [Cited: 24 February 2015.] http://en.wikipedia.org/wiki/Portable_Document_Format.
69. **Johnson, Duff.** The 8 most popular document formats on the Web. *Duff Johnson Strategy and Communications blog*. [Online] 17 February 2014. [Cited: 24 February 2015.] <http://duff-johnson.com/2014/02/17/the-8-most-popular-document-formats-on-the-web/>.
70. *Characterising and Preserving Digital Repositories: File Format Profiles*. **Hitchcock, Steve and Tarrant, David.** 66, January 2011, Ariadne. <http://eprints.soton.ac.uk/273241/>.
71. *Formats over Time: Exploring UK Web History*. **Jackson, Andrew N.** Toronto : s.n., 2012. iPRES 2012. <http://arxiv.org/abs/1210.1714>.
72. **Johnson, Duff.** PDF Readers - 5 readers compared. *Talking PDF*. [Online] 30 November 2010. [Cited: 24 February 2015.] <http://talkingpdf.org/pdf-readers-5-readers-compared/>.
73. List of PDF Software. *Wikipedia*. [Online] [Cited: 24 February 2015.] http://en.wikipedia.org/wiki/List_of_PDF_software.
74. Adobe Reader XI. *Adobe*. [Online] [Cited: 24 February 2015.] <http://www.adobe.com/uk/products/reader.html>.
75. **van der Knijff, Johan.** Adobe Portable Document Format: Inventory of long term preservation risks. *Open Preservation Foundation*. [Online] 20 October 2009. [Cited: 24 February 2015.] http://www.openplanetsfoundation.org/system/files/PDFInventoryPreservationRisks_0_2_0.pdf.
76. Migration: Context and Current Status. *Digital Preservation Testbed*. [Online] 5 December 2001. [Cited: 24 February 2015.] http://www.nationaalarchief.nl/sites/default/files/docs/kennisbank/migration_0.pdf.
77. *The Network is the Format: PDF and the Long-term Use of Digital Content*. **Morrissey, Sheila.** Copenhagen, DK : s.n., 2012. Archiving 2012. Vol. 8, pp. 200-203. <http://www.portico.org/digital-preservation/wp-content/uploads/2012/12/Archiving2012TheNetworkIsTheFormat.pdf>.
78. **Johnson, Duff.** Are Your Documents Readable? How Would You Know? *Duff Johnson Strategy and Communications Blog*. [Online] 24 January 2014. [Cited: 24 February 2015.] <http://duff-johnson.com/2014/01/24/are-your-documents-readable-how-would-you-know/>.
79. **van der Knijff, Johan.** Identification of PDF Preservation Risks: Analysis of GovDocs Selected Corpus. *OPF Blog*. [Online] 27 January 2014. [Cited: 24 February 2015.] <http://www.openplanetsfoundation.org/blogs/2014-01-27-identification-pdf-preservation-risks-analysis-govdocs-selected-corpus>.
80. JHove PDF-hul Module. *JHove Sourceforge*. [Online] [Cited: 24 February 2015.] <http://jhove.sourceforge.net/pdf-hul.html>.
81. **Owens, Evan.** Automated Workflow for the Ingest and Preservation of Electronic Journals. [Online] 2010?? <http://www.portico.org/digital-preservation/wp-content/uploads/2010/01/Archiving2006-Owens-pres.pdf>.
82. SPRUCE Characterisation Hackathon. *OPF Wiki*. [Online] 2013. [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/SPR/SPRUCE+Hackathon+Leeds%2C+Unified+Characterisation>.
83. **Cliff, Peter.** Visual Analysis of PreFlight Output. *OPF Wiki*. [Online] [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/SPR/Visual+Analysis+of+Preflight+Output>.
84. PDF-A Validation and Conversion in Florida Digital Archive. *Florida Virtual Campus*. [Online] 19 September 2013. [Cited: 24 February 2015.] https://share.fcla.edu/FDAPublic/Affiliates/FDA_PDF-A_validation_conversion.pdf.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

85. PDF/A Manager. *PDFTron*. [Online] [Cited: 24 February 2015.] <http://www.pdftron.com/pdfamanager/index.html>.
86. KOST-Val. *COPTR Digital Preservation Technical Registry*. [Online] [Cited: 24 February 2015.] <http://coptr.digipres.org/KOST-Val>.
87. KOST-CECO. [Online] [Cited: 24 February 2015.] <http://kost-ceco.ch/cms/>.
88. PDF/A Pilot. *Callas Software*. [Online] [Cited: 24 February 2015.] <http://www.callassoftware.com/callas/doku.php/en:products:pdfapilot>.
89. *PDF to PDF/A: Evaluation of Converter Software for Implementation in Digital Repository Workflow*. **Koo, Jamin and Chou, Carol C. H.** 2012. iPRES 2012. pp. 302-303. https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres_2012_Conference_Proceedings_Final.pdf.
90. 3-Heights(TM) PDF to PDF/A Converter. *PDF Tools*. [Online] [Cited: 24 February 2015.] <http://www.pdf-tools.com/pdf/pdf-to-pdf-a-converter-signature.aspx>.
91. **Gilham, Jo and Cliff, Peter.** PDF/A SPRUCE Scenario. *OPF Wiki*. [Online] 2013. [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/SPR/PDFA+Validation+tools+give+different+results>.
92. **Johnson, Duff.** PDF Validation Dream or Yawn? Waking up to the possibilities of an open-source PDF validator. *Duff Johnson Strategy and Communications Blog*. [Online] 2013. [Cited: 24 February 2015.] <http://duff-johnson.com/wp-content/uploads/2014/01/PDFValidationDreamOrYawn.pdf>.
93. Flint. *Github*. [Online] [Cited: 24 February 2015.] <http://openpreserve.github.io/flint/>.
94. Apache Tika. *Apache Software Foundation*. [Online] [Cited: 24 February 2015.] <http://tika.apache.org>.
95. Metadata Extraction Tool. *National Library of New Zealand*. [Online] [Cited: 24 February 2015.] <http://meta-extractor.sourceforge.net/>.
96. JHove. *JHove Sourceforge*. [Online] [Cited: 24 February 2015.] <http://sourceforge.net/projects/jhove/>.
97. 3 Heights(TM) PDF Extract. *PDF-Tools*. [Online] [Cited: 24 February 2015.] <http://www.pdf-tools.com/pdf/pdf-extract-content-metadata-text.aspx>.
98. **Shea, Dan.** Acrobat and PDF Developer Libraries. *Planet PDF*. [Online] 13 November 2013. [Cited: 24 February 2015.] http://www.planetpdf.com/developer/article.asp?ContentID=acrobat_pdf_developer_librar&gid=6218.
99. QPDF. *QPDF Sourceforge*. [Online] [Cited: 24 February 2015.] <http://qpdf.sourceforge.net/>.
100. *Archaeology Data Service*. [Online] [Cited: 24 February 2015.] <http://archaeologydataservice.ac.uk/>.
101. **Mitcham, Jenny and Wheatley, Paul.** PDF to PDF-A conversion. *OPF Wiki*. [Online] 2012. [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/REQ/PDF+to+PDF-A+conversion>.
102. PDF Reference and Adobe Extensions to the PDF Specification. *Adobe*. [Online] [Cited: 24 February 2015.] http://www.adobe.com/devnet/pdf/pdf_reference.html.
103. Adobe PDF Reference Archives. *Adobe*. [Online] [Cited: 24 February 2015.] http://www.adobe.com/devnet/pdf/pdf_reference_archive.html.
104. *PDF Association*. [Online] [Cited: 24 February 2015.] <http://www.pdfa.org/>.
105. **Johnson, Duff.** Is PDF an Open Standard? *Planet PDF*. [Online] 10 June 2010. [Cited: 24 February 2015.] http://www.planetpdf.com/enterprise/article.asp?ContentID=Is_PDF_an_open_standard&page=1.

Digital Preservation Team	Preservation Assessment: eBook Summary Assessment	Date: 03/02/2017
		Version: 1.0

106. **Wolf, Julia.** OMG-WTF-PDF Denouement. *FireEye*. [Online] 2 February 2011. [Cited: 24 February 2015.] <http://www.fireeye.com/blog/technical/cyber-exploits/2011/02/omg-wtf-pdf-denouement.html>.
107. **van der Knijff, Johan.** What do we mean by "embedded" files in PDF? *OPF Blog*. [Online] 9 January 2013. [Cited: 24 February 2015.] <http://www.openplanetsfoundation.org/blogs/2013-01-09-what-do-we-mean-embedded-files-pdf>.
108. **Wolf, Julia.** OMG-WTF-PDF [PDF Ambiguity and Obfuscation]. *Troopers*. [Online] 31 March 2011. [Cited: 24 February 2015.] https://www.troopers.de/wp-content/uploads/2011/04/TR11_Wolf_OMG_PDF.pdf.
109. **Albertini, Ange.** PDF Tricks. *Corkami*. [Online] 2014. [Cited: 24 February 2015.] <https://code.google.com/p/corkami/wiki/PDFTricks>.
110. **NDSA Standards and Practices Working Group.** The Benefits and Risks of the PDF/A-3 File Format for Archival Institutions. *Digital Preservation*. [Online] February 2014. [Cited: 24 February 2015.] http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_PDF_A3_report_final022014.pdf.
111. **Johnson, Duff.** Archivists: No flowers for PDF/A-3. *Duff Johnson's Strategy and Communications Blog*. [Online] 28 February 2014. [Cited: 24 February 2015.] <http://duff-johnson.com/2014/02/28/archivists-no-flowers-for-pdf-a-3/>.
112. Detect, extract and analyse embedded objects in PDFs. *OPF Wiki*. [Online] 2011. [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/AQuA/Detect%2C+extract+and+analyse+embedded+objects+in+PDFs>.
113. PDF Format Issues: JavaScript. *OPF Wiki*. [Online] [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/TR/JavaScript>.
114. PDF Format Issues: References to external files. *OPF Wiki*. [Online] [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/TR/References+to+external+files>.
115. PDF Format Issues: Fonts missing, damaged or incomplete. *OPF Wiki*. [Online] [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/TR/Fonts+missing%2C+damaged+or+incomplete>.
116. SPRUCE PDF Solution: PDF Characterisation Tool. *OPF Wiki*. [Online] 2011. [Cited: 24 February 2015.] <http://wiki.opf-labs.org/display/AQuA/PDF+Characterisation+Tool>.
117. *Born Broken: Fonts and Information Loss in Legacy Digital Documents*. **Brown, Geoffrey and Woods, Kam.** 1, s.l.: University of Edinburgh, 2011, International Journal of Digital Curation, Vol. 6, pp. 5-19. <http://dx.doi.org/10.2218/ijdc.v6i1.168>. ISSN 1746-8256.
118. Tools and Utilities. *PDFTron*. [Online] [Cited: 24 February 2015.] <http://www.pdftron.com/pdftools.html>.
119. **Library of Congress.** PDF (Portable Document Format) Family. *Sustainability of Digital Formats*. [Online] [Cited: 24 February 2015.] <http://www.digitalpreservation.gov/formats/fdd/fdd000030.shtml>.
120. PDF Current Threats. *Malware Tracker*. [Online] [Cited: 25 February 2015.] <http://www.malwaretracker.com/pdfthreat.php>.