PDF Format Preservation Assessment Part 2: PDF/A Profile

Document History

Date	Version	Author(s)	Circulation
03/06/2019	1.0	Akiko Kimura, Peter May	External (Restructured as a two-part document. See PDF Format Preservation Assessment, Part 1: PDF, Version 1.5)

British Library Digital Preservation Team
digitalpreservation@bl.uk

1. Introduction

This document provides a high level, non-collection specific assessment of the PDF/A 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 is Part 2 of a two-part document. It will primarily focus on the most relevant (for digital preservation) and commonly encountered sub-formats - or profiles - of the PDF family: PDF/A (parts 1 through 3). This document should be read in conjunction with the core PDF (versions 1 through 2) assessment captured in the separate Part 1 document.

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 PDF/A Summary

PDF/A is a profile of the PDF format. It is a multi-part ISO standard that restricts functionality considered problematic for long-term archiving. Three parts of the standard have been published so far: PDF/A-1 through 3.

Restrictions placed across the PDF/A family include:

- non-embedded fonts
- JavaScript
- audio and video content
- LZW compression
- non-embedded colour spaces, and
- encryption.

Each part of the PDF/A standard is an independent profile with various levels of conformance:

> **PDF/A-1** (ISO19005-1:2005) is based on PDF version 1.4 (not ISO standard). As well as the above mentioned restrictions, PDF/A-1b prohibits additional functionality including attachment of files in any formats, JPEG2000 compression and transparent elements. Two levels of conformance are provided: level B (basic) – satisfies minimum requirements necessary for visual appearance reproduction; level A (accessible) – adds additional requirements on document structure in order to support text extraction and accessibility (such as the use of Tagged PDF and Unicode character maps).
>
> **PDF/A-2** (ISO19005-2:2011) is based on PDF version 1.7 (ISO32000-1:2008). Unlike PDF/A-1, it allows: attachment of other PDF/A files, JPEG2000 compression and transparent elements. In addition to the conformance level B (basic) and level A (accessibility), PDF/A-2 provides level U – level B plus the requirement for all text to have Unicode equivalents.
>
> **PDF/A-3** (ISO19005-3:2012) is also based on PDF version 1.7 and provides 3 levels of conformance: A, B and U. The profile is basically identical to PDF/A-2 but with a significant difference reflecting end-user/industry requirements — in particular, PDF/A-3 allows attachment of files in *any format*.

In essence, PDF/A-1 is the most restrictive profile. The newer parts, PDF/A-2 and 3, have additional functions that came with the newer version of base PDF format as well as the reinstated functionally to attach other files (other PDF/As for A-2 and any format for A-3).

2. Assessment

2.1 Development Status

A summary of the development history of the format and an indication of its current status

PDF/A was created against the backdrop of growing popularity of the PDF format, which is highly complex and not suitable for long-term archiving of electronic documents. Wide-ranging organisations such as administrative bodies, archiving communities and industries worldwide pressed for the development of the standard, which resulted in the formation of an ISO Joint Working Group and the release of ISO19005-1 (PDF/A-1) in 2005. The latest of the PDF/A family, PDF/A-3, was released in 2012. PDF/A-NEXT (aka PDF/A-4) is currently under development[1].

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

From the very start, the PDF/A standard has been recommended by many public and industrial bodies as the archival format for electronic documents. Early adopters include the European Commission, several European governments, and U.S. Courts [1]. The Library of Congress also lists a growing number of organisations, many in the U.S., recommending or requiring PDF/A [2]. Despite the enthusiasm, the true picture of PDF/A adoption is difficult to quantify due partly to the lack of reliable PDF/A identification/validation tools (discussed below). In her 'DPC Technology Watch Report', Betsy Fanning suggests that, based on available figures, actual take up of PDF/A format may be slow [3][2]. This might be because end users are still preferring general PDF format over PDF/A for its convenience and by necessity (e.g. academic reports requiring inclusion of audio visual elements, attachment of data source files etc.). Also, even though PDF/A-2 and 3 allow attachment of files, many tools currently claiming to create PDF/A (especially the ones aiming at non-enterprise users) only support creation of PDF/A-1 [2]. However, we may see a rapid change in PDF/A's popularity depending on the improvement in software support and how the evolution of PDF/A-4 pans out.

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

Support for rendering is generally good as all PDF viewers are capable of rendering PDF/A files. Not all PDF viewers are, however, conforming to the PDF/A standard. The 'Whitepaper: A technical introduction to PDF/A' by PDFlib lists possible issues with non-conforming viewers, for example, they may ignore embedded elements such as fonts and ICC-based colour spaces, and use locally available options instead [4][3]. In general, PDF viewers consider any attached files as supplemental and do not attempt to render them. Conforming PDF/A viewers should, however, provide a mechanism for attached files to be extracted and saved (but not necessarily rendered) so that, if necessary, the files can be viewed on different software appropriate for the format. Beyond that, how exactly a PDF/A conforming viewer should behave is still not well defined. Such uncertainty can undermine the premise of PDF/A that it 'provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files[4]'.

[1] According to the placeholder on the ISO website, PDF/A-NEXT (ISO19005-4) is based on PDF version 2.0 (https://www.iso.org/standard/71832.html).
[2] Page 16.
[3] Page 4.
[4] Quote from the introduction of the PDF/A-1 standard (ISO19005-1:2005).

In addition, PDF/A viewers inherit any rendering issues native to PDF viewers owing in part to the sheer complexity and range of functionality added to the PDF specifications over the years. See the Part 1 assessment for further details on potential PDF rendering issues.

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

Being a profile of PDF, PDF/A is not readily distinguishable from the PDF specification it is based on. Identification of the specific PDF/A variant could be considered more a matter of characterisation and validation, and this is discussed below.

Validation, Conformance Checking and Detecting Preservation Risks

veraPDF [5], released in 2017, is the first complete open source PDF/A validator, covering all parts and conformance levels of the PDF/A standard. The development was initiated by the EU-funded PREFORMA project and led by the Open Preservation Foundation [6] and the PDF Association [7]. Fanning describes veraPDF as 'a new PDF/A validation tool that provides comprehensive validity checks closely paired with precise format specification rules and a new test suite' and expresses her anticipation that 'this development will help to remove ambiguities and close loopholes in the format specification [3][5]'. veraPDF also has a useful feature to validate the conformance of documents against institutional policies.

In her 2018 thesis, Anna Oates conducted a case study in which she migrated a set of files from various formats including PDF to PDF/A using various migration tools[6]. Of the 698 (supposedly) successfully migrated files, only 483 files passed validation with veraPDF [8][7]. It should be noted that the scope of veraPDF is restricted to the clauses within the PDF/A standard themselves due to the limited resources. It means that some elements only described in the PDF specifications are not validated to the full extent but only in the manner in which conformance to the PDF/A standard is satisfied. Examples of affected functionality include JPEG 2000, ICC profiles, and Tagged PDF. Users are encouraged to use veraPDF's extensible architecture to develop necessary plugins to fulfil their own needs for PDF/A validation [9][8]. It should be noted that some issues affecting rendering of the content, such as corrupt images, are out of the scope of validation tools including veraPDF.

Up until the release of veraPDF, the lack of effective PDF/A validation tools has been highlighted in several experiments: Florida Virtual Campus reported on the shortcomings of JHOVE PDF/A validation [10]. PDF/A Manager [11] (part of the PDF Tron suite [12][9]) was assessed (along with pdfaPilot [13] and 3-Heights [14]) by Carol Chou and Jamin Koo, with all products achieving PDF/A validation accuracy results of between 90 and 95% [15]. Also, Jo Gilham and Peter Cliff reported disparity between the validation results of different tools such as PDF/A Manager and Adobe Preflight [16] [17]. More independent case studies to better ascertain the comparative accuracy and effectiveness of veraPDF would be extremely beneficial.

Metadata Extraction

Any metadata extraction tools which handle PDFs, such as Apache Tika [18], the NLNZ Metadata Extraction Tool [19], JHOVE [20], or 3-Heights™ PDF Extract [21] should be applicable to PDF/A files.

[5] Page 14.
[6] The migration tools Oates used in her case study were: Adobe Acrobat DC, Callas pdfaPilot, Intarsys PDF/A Live!, LibreOffice, PDF Studio, PDFForge PDFCreator, and PDFTron PDF/A Manager CMD migration software.
[7] Page 32.
[8] Page 161.
[9] PDF/A Manager is a commercial offering providing PDF validation used by the Kost-val [70] toolset at Kost [53].

Content Extraction

Reusability of the content of a PDF/A document is a concern for memory organisations. As PDF was developed fundamentally as the format for visual representation of page-based documents, extraction of text and other PDF objects to reuse them on different format/media is a challenge. Various types of assistive technology (AT) introduced to assist visually impaired users, such as Tagged PDF, could also support content of the document to be extracted in a semantically logical order, and helps search engines to discover the content reliably. The PDF/A standard's conformance level A (accessible) clearly indicates that such AT measures were applied to the document, which could pave the way for effective content extraction tools for PDF/A content to be developed. However, Duff Johnson argues that the level A conformance of PDF/A is totally insufficient: 'Technically, it's possible to comply with PDF/A-1a with a single tag for each page, irrespective of the document's contents. That's the key reason why claims of conformance or validation of PDF/A-1a are, by themselves, essentially meaningless' [22][10].

Migration

Migration from PDF/A to other formats may be greatly assisted by the fact that the source files are not encrypted and some elements, such as fonts and colour spaces, are embedded in the files to prevent the loss of key components of the documents.

There are a number of dedicated PDF focused applications which support migration to and from PDF/A. The PDF Association maintains a list of its members' relevant products [23]. Not all the tools, however, support all the variants of PDF/A. Even some popular products of Adobe, the inventor of the PDF format, only support export to PDF/A-1b (e.g. InDesign version 14.0 [24]).

Support for quality checking via comparison between source and destination files is generally poor, and may not provide sufficient confidence that significant properties have survived the transformation. Fanning adds particular issues concerning migration to PDF/A which include: a loss of embedded digital signatures, attached files, transparency, and also a scenario that, when fonts in the source files are not available to embed, a migration tool may make a poor font substitution which may result in not only a loss of appearance of the document but also loss of meaning if specialised fonts such a mathematical font was substituted without sufficient warning [3][11]. Jenny Mitcham also notes various challenges with PDF to PDF/A migration [25].

As discussed above, the nature of the restrictions in PDF/A means that wholesale migration of a PDF collection to PDF/A is unwise. Before committing to any file migration, thought should be given as to the aspects of the original file that must be preserved, and whether or not PDF/A can support them.

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

All the parts of PDF/A (1 through to 3) are ISO standards and the documentation is available for download (with a fee for non-ISO members). PDF 1.7, which is the underlying format of PDF/A-2 and 3, is also an ISO standard, and the copies of the specifications are provided by Adobe [26] along with an archive of legacy documentation [27]. PDF/A-1 is based on PDF 1.4, which is not an ISO standard but its documentation 'PDF Reference, third edition, Adobe Portable Document Format, Version 1.4 [28]' is made freely available to the public by Adobe. In addition, support for PDF and PDF/A in general is available from a number of sources, such as the PDF Association.

As mentioned in Part 1, PDF viewers' tolerances to invalid PDF files is often considered as a consequence of ambiguities in PDF documentation over time. While this is a matter affecting all the current PDF/A variants, PDF/A-4 (currently under development) should benefit from the fact

[10] Johnson goes on to discuss PDF/UA (Universal Accessibility) which sets a clear standard to describe accessible PDF in technically complete terms: 'a really good PDF/A-1a file is one that also complies with PDF/UA'.

[11] Page 13-14.

that it is based on PDF 2.0 (ISO 32000-2) released in 2017. According to Johnson, PDF 2.0 'has evolved' in terms of clarity in its documentation [29].

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?

PDF/A is designed to overcome the long-term preservation risks identified in the PDF format. Nonetheless, PDF/A is fundamentally a PDF and therefore inherits the issues derived from PDF's complexity and uncertainty (see more in the Part 1 assessment). In particular, PDF/A-3 files require careful handling because of its file attachment functionality as discussed below.

2.6 Embedded or Attached Content

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

PDF/A-3 removes a single but contentious restriction to the PDF/A-2 standard by allowing *any* file format to be embedded as a file stream. This may prove beneficial in some use cases, and especially from a preservation point of view where, for example, a source document for the PDF/A itself can be attached. It has been observed, however, the attachment facility has the potential to be used for a variety of purposes, such as embedding additional information or even information of a more critical nature than the primary document. This leaves potentially difficult questions for an archive. Is an attached file of critical importance for preservation (despite potentially being of any format and hence potentially a significant preservation risk), merely a secondary object with optional or additional data, or, as noted above, the source data for the primary document?

Given such concerns, PDF/A-3 standard imposes additional requirements for attached files (termed 'associated files' in the standard) including: each embedded file stream has to have a header stating the MIME type (although 'application/octet-bitstream' can be used if a precise MIME type is not known); relationships between attached files and the primary PDF/A-3 file must be explicitly expressed within the file specification dictionary (predefined key values are: source, data, alternative, supplement or unspecified) [30]. Although such measures could help preservation tools and procedures to improve, would PDF/A creators take the time to provide that accurate metadata where 'unknown/unspecified' statements are allowed? How will the PDF/A creating software influence this process?

veraPDF can be configured to detect PDF/As with file attachments. Although it is a significant improvement from the PDF/A collection management point of view, safeguarding such files tacked inside the PDF/A container remains a challenge for archiving organisations. The NDSA PDF/A-3 Working Group published a detailed discussion focused on these concerns [31], which considers a number of use cases and makes some sensible recommendations (such as checking received PDF/A-3 files for embedded files, and treating PDF/A-3 files separately from the other PDF/A variants in terms of format preference lists and related format action plans). On the whole, however, this remains a significant area of concern.

Also it needs to be mentioned that, the size of a PDF/A-2 or 3 file can be significantly large due to file attachments. Even with no attachment, a PDF/A file may be larger than a PDF file of the same content due to embedded fonts etc. According to the PDF Association, 'Embedded fonts can slightly increase the size of a PDF/A file', and it can be in some cases problematic when archiving a very large number of documents [32][12]. On the other hand, Fanning notes that 'PDF/A-2 introduced compressed objects and XRef streams to minimize file sizes [3][13]'. The true impact of file size issues for archiving organisations remains to be seen.

[12] Page 18.
[13] Page 15.

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

Removing external dependencies is one of the key measures applied to the PDF/A standard in an effort to reduce preservation risks. Key components such as fonts, colour space, and images are required to be embedded in PDF/A. However, there is a common misconception that PDF/A does not allow any external references at all; Johan van der Knijff notes, 'contrary to some (incorrect) popular belief [sic], the PDF/A standards do not rule out references to external files completely. The following mechanisms for referring to external content are allowed: URI actions: these refer to Internet resources (i.e. a clickable hyperlink), GoToR actions: these refer to an external PDF file (i.e. a clickable link to a locally stored PDF) … From a preservation point of view, GoToR actions may nevertheless be a risk (e.g. in case of a collection of PDFs that refer to each other), even though the rendering of the files that contain the reference is not affected' [33].

2.8 Legal Issues

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

As far as the PDF/A format is concerned, there is no conspicuous legal issue relating to preservation risks. All variants of PDF/A profiles are ISO standards as well as PDF 1.7, which PDF/A-2 and 3 are based on. Even though patents relating to PDF 1.4, which PDF/A-1 is based on, is owned by Adobe, the company has issued royalty free rights on a significant number of patents in order to encourage uptake and third party development of PDF tools [34]. And a change of policy in at least the short term appears to be highly unlikely.

A potential legal issue with PDF/A creation/migration is copyright restriction with fonts. The PDF/A standard requires fonts to be embedded in the document to reduce preservation risks. However, the requirement may unwittingly open up other types of preservation risk. Firstly, copyrighted specialised fonts may prevent electronic documents to be saved as PDF/A at all. Chou and Koo note that: 'There are ways to circumvent possible copyrights infringement through font substitution but some specialized fonts may prove to be difficult not only to procure but also to use in PDF/A conversion, as their makers can prohibit embedding of fonts' [15]. Secondly, where a restricted font has been embedded without copyright clearance, this may risk copyright infraction for the preserving organisation. Fanning suggests that there may also be restrictions on how fonts are distributed: 'Organizations sometimes use their own bespoke families of fonts to help authenticate documents. These restrictions may impact on where PDF/A can be applied, how it can be preserved and where access can be provided' [3][14]. Furthermore, as mentioned earlier, a PDF/A migration tool may make a poor font substitution and embed copyright-protected fonts into the document without issuing a sufficient warning.

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

There is no encryption issue with PDF/A as the functionality is prohibited in the standard.

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

All the variants of PDF/A standard permits embedded digital signatures, while the use of the PDF Advanced Electronic Signatures (PAdES) standard is permitted in PDF/A-2 and 3. Digital signatures are a mixture of PDF objects and strings in a cryptographic message syntax [35]. According to Fanning, migration to PDF/A will 'break' the signature [3][15].

[14] Page 15.
[15] Page 14.

Unlike migration, receipt or deposit of PDF/A-1 and 2 files may not raise significant preservation concerns as the standard prohibits functionality associated with the preservation risks identified in this report. Assuming, of course, that the source is trusted and the files do indeed conform to the restrictions described in the standard.

2.11 Preservation Risk Summary

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

PDF/A is a widely recognised profile of PDF, the world's most ubiquitous electronic document format, designed for long-term archiving by restricting specific functionality identified as preservation risk.

Although some of the most prominent risks for preservation identified in the PDF assessment (Part 1 of this document) such as encryption and missing fonts are eliminated from PDF/A, it still inherits deep-seated complexity and uncertainty from the underlying PDF format. Many preservation issues may be unnoticed as a result of viewers tolerating invalid or badly formed files. Unlike earlier flavours of PDF/A, PDF/A-3's capability to have file attachments in any formats reintroduced the risk for potentially key elements of the document being unidentified and/or lost during the migration process. Furthermore, there is no guarantee that software to render such attachments in various formats is available now, or in the future.

The definition as to what a PDF/A conforming software should consist of is poorly defined meaning, for example, a mathematical font embedded in PDF/A may be substituted with a locally available font by the viewing tool, potentially impairing the visual appearance and understanding of the document. Similarly, when migrating to a PDF/A, the tool may not be clear when a copyright protected font has been embedded in the document, which risks copyright infraction for the preserving organisation.

Although the lack of reliable PDF/A validation tools have been a major issue widely recognised within digital archiving communities, the recent development of veraPDF has changed the landscape. It has made it possible for an archiving organisation to assess its PDF/A collections more reliably, and plan for effective long-term preservation strategies.

The following list summarises risks for PDF/A.

- **Invalid or badly formed PDF files**
  - May affect ability to render files now or in the future
- **Legal issues**
  - Embedded copyrighted fonts may pose copyright infraction threats for the preserving organisation
- **File attachments**
  - Attachments of any file format could pose a variety of preservation risks in themselves (PDF/A-3 only)
- **File size**
  - Embedded fonts and colour spaces, and attached files (PDF/A-2 and 3 only) may increase the file size
- **Loss of significant properties**
  - Embedded video, sound, JPEG2000 streams[16], digital signatures, and/or interactive elements such as JavaScript will be discarded on conversion to PDF/A
  - Attached files may be discarded on conversion to PDF/A-1 and 2
- **External references**
  - References to locally stored PDF files (via GoToR action) may affect rendering if they change or move the location.

[16] Note that JPEG2000 compression is forbidden in PDF/A-1 but not in PDF/A-2 and 3.

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 non-preservation issues such as resourcing.

As software support for PDF/A has improved, most notably with veraPDF, there are more opportunities to mitigate preservation risks identified in this assessment. This section should be read in conjunction with the Recommended Action section in the core PDF assessment (Part 1).

Handling Recommendations

- It is recommended that PDFs are created to one of the PDF/A standards (ideally PDF/A-1) and validated using a suitable PDF validator.
- Batch migration from PDF to PDF/A, without deep analysis of the collection, is not recommended. It could potentially do more harm than good, and risk the loss of significant properties.
- Receipt or deposit of PDF/A is recommended to prefer the PDF/A-1 profile rather than PDF/A-2 and 3 to reduce the risk concerning attached files.
- It is recommended that PDF/As are validated with a suitable PDF validator at the point of receipt or deposit to check their conformity to the PDF/A standards.
- Files confirmed to be PDF/A-3 with file attachments should be, where possible, managed in such a way that specific preservation actions can be applied. Files should be checked with a suitable PDF validator to confirm the presence of embedded files, but be aware the PDF user-created metadata (see below).
- User-created metadata concerning attached files (MIME type and relationships to the primary document) should not be trusted; deeper inspection should be performed.
- Digitally signed documents should not be migrated to PDF/A, or should be re-signed after migration [3][17].

Knowledge Recommendations

- Understand the frequency and impact of PDF/A risks in specific collections in order to identify genuine problems, or discount identified risks whose frequency or impact is not significant.
- Undertake independent case studies to better ascertain the comparative accuracy and effectiveness of validation tools, especially veraPDF.

Software Recommendations

- Support enhancements to validation tools to improve detection of attached files (e.g. MIME type).
- Support enhancements to migration tools to improve detection of significant properties lost or substituted during the process of transformation.

Monitoring Recommendations

The preservation risks faced by PDF/A are unlikely to change rapidly and so review of this document should not be considered a high priority. However awareness of new software developments, particularly to veraPDF, and to new PDF/A variants should be maintained:

- Monitor veraPDF for new features and capabilities.
- Monitor development of PDF/A-4.

[17] Page 13-14.

4. References

1. PDF/A archiving standard. *Adobe.* [Online] 23 12 2012. [Cited: 19 June 2019.] http://web.archive.org/web/20121223184242/http://www.adobe.com/enterprise/standards/pdfa/.

2. PDF/A Family, PDF for Long-term Preservation. *Library of Congress.* [Online] [Cited: 19 June 2019.] https://www.loc.gov/preservation/digital/formats/fdd/fdd000318.shtml.

3. Fanning, Betsy A and AIIM. *Preservation with PDF/A (2nd Edition).* s.l. : Digital Preservation Coalition, 2017. ISSN: 2048-7916.

4. PDFlib GmbH. *Whitepaper: A Technical Introduction to PDF/A.* s.l. : PDFlib, 2013. p. 4.

5. veraPDF. *veraPDF.* [Online] [Cited: 18 June 2019.] https://verapdf.org/home/.

6. Open Preservation Foundation. [Online] [Cited: 19 June 2019.] https://openpreservation.org/.

7. *PDF Association.* [Online] [Cited: 24 February 2015.] http://www.pdfa.org/.

8. Oates, Anna. *Navigating the PDF/A Standard: A Case Study of Theses in the University of Oxford's Institutional Repository.* s.l. : Graduate College of the University of Illinois at Urbana-Champaign, 2018.

9. *veraPDF: building an open source, industry supported PDF/A validator for cultural heritage institutions.* Wilson, Carl, McGuinness, Rebecca and Jung, Joachim. 2, s.l. : Emerald, 2017, Digital Library Perspectives, Vol. 33.

10. PDF-A Validation and Conversion in Florida Digital Archive. *Florida Virtual Campus.* [Online] 19 September 2013. [Cited: 20 June 2019.] https://libraries.flvc.org/documents/181844/502298/PDFA+Validation+and+Conversion+in+FDA/.

11. PDF/A Manager. *PDFTron.* [Online] [Cited: 20 June 2019.] https://www.pdftron.com/documentation/cli/guides/pdfa-manager/overview/.

12. PDFTron: Functionality. *PDFTron.* [Online] [Cited: 20 June 2019.] https://www.pdftron.com/pdf-sdk/features.

13. pdfaPilot. *Callas Software.* [Online] [Cited: 20 June 2019.] https://www.callassoftware.com/en/products/pdfapilot.

14. 3-Heights PDF Validator – PDF and PDF/A standard conformance validation. *PDF Tools.* [Online] [Cited: 18 June 2019.] http://www.pdf-tools.com/pdf20/en/products/pdf-converter-validation/pdf-validator/.

15. *PDF to PDF/A: Evaluation of Converter Software for Implementation in Digital Repository Workflow.* Koo, Jamin and Chou, Carol C. H. 2013. iPRES 2012. pp. 4-5. https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres 2012 Conference Proceedings Final.pdf.

16. Adobe. Correcting problem areas with the Preflight tool (Acrobat Pro). *Adobe.* [Online] [Cited: 18 June 2019.] https://helpx.adobe.com/acrobat/using/correcting-problem-areas-preflight-tool.html.

17. Gilham, Jo and Cliff, Peter. PDFA SPRUCE Scenario. *OPF Wiki.* [Online] 2013. [Cited: 24 February 2015.] http://wiki.opf-labs.org/display/SPR/PDFA+Validation+tools+give+different+results.

18. Apache Tika. *Apache Software Foundation.* [Online] [Cited: 20 June 2019.] http://tika.apache.org.

19. Metadata Extraction Tool. *National Library of New Zealand.* [Online] [Cited: 20 June 2019.] http://meta-extractor.sourceforge.net/.

20. JHove. *JHove Sourceforge.* [Online] [Cited: 20 June 2019.] http://sourceforge.net/projects/jhove/.

21. 3 Heights(TM) PDF Extract. *PDF-Tools.* [Online] [Cited: 19 June 2019.] http://www.pdf-tools.com/pdf/pdf-extract-content-metadata-text.aspx.

22. Johnson, Duff. Accessibility: What PDF/A-1a Really Means. *PDF Association.* [Online] 1 October 2010. [Cited: 26 June 2019.] https://www.pdfa.org/accessibility-what-pdf-a-1a-really-means/.

23. About our members products. *PDF Association.* [Online] [Cited: 20 June 2019.] https://www.pdfa.org/products/.

24. Adobe InDesign User Guide: Export to Adobe PDF. *Adpbe.* [Online] [Cited: 20 June 2019.] https://helpx.adobe.com/uk/indesign/using/exporting-publishing-pdf.html.

25. Mitcham, Jenny and Wheatley, Paul. PDF to PDF-A conversion. *OPF Wiki.* [Online] 2012. [Cited: 20 June 2019.] http://wiki.opf-labs.org/display/REQ/PDF+to+PDF-A+conversion.

26. PDF Reference and Adobe Extensions to the PDF Specification. *Adobe.* [Online] [Cited: 24 February 2015.] http://www.adobe.com/devnet/pdf/pdf_reference.html.

27. Adobe PDF Reference Archives. *Adobe.* [Online] [Cited: 24 February 2015.] http://www.adobe.com/devnet/pdf/pdf_reference_archive.html.

28. Adobe Systems Incorporated. *PDF Reference, third edition: Adobe Portable Document Format, Version 1.4.* s.l. : Adobe Systems Incorporated, 2001. ISBN: 0-201-75839-3.

29. Johnson, Duff. PDF 2.0: The worldwide standard for electronic documents has evolved. *PDF Association.* [Online] 30 August 2017. [Cited: 12 June 2019.] https://www.pdfa.org/pdf-2-0-the-worldwide-standard-for-electronic-documents-has-evolved/.

30. Library of Congress. *PDF/A-3, PDF for Long-term Preservation, Use of ISO 32000-1, With Embedded Files.* [Online] [Cited: 25 June 2019.] https://www.loc.gov/preservation/digital/formats/fdd/fdd000360.shtml#notes.

31. 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: 25 June 2019.] http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_PDF_A3_report_final022014.pdf.

32. Oettler, Alexandra. *PDF/A in a Nutshell 2.0: PDF for long-term archiving.* Berlin : Association for Digital Document Standards, 2013. pp. 13-15.

33. PDF Format Issues: References to external files. *OPF Wiki.* [Online] [Cited: 24 June25 2019.] http://wiki.opf-labs.org/display/TR/References+to+external+files.

34. Library of Congress. PDF (Portable Document Format) Family. *Sustainability of Digital Formats.* [Online] [Cited: 25 June 2019.] http://www.digitalpreservation.gov/formats/fdd/fdd000030.shtml.

35. Bärfuss, Hans. Digital signatures in PDF/A. *PDF-TOOLS.* [Online] 25 June 2019. http://blog.pdf-tools.com/2015/01/digital-signatures-in-pdfa.html.

36. ISO/DIS 32000-2 PDF 2.0. *ISO.* [Online] [Cited: 24 February 2015.] http://www.iso.org/iso/catalogue_detail.htm?csnumber=63534.

37. Albertini, Ange. A PDF 101 document walk through. [Online] [Cited: 24 February 2015.] http://imgur.com/a/PbN8H#7.

38. Portable Document Format. *Wikipedia.* [Online] [Cited: 24 February 2015.] http://en.wikipedia.org/wiki/Portable_Document_Format.

39. 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/.

40. *Characterising and Preserving Digital Repositories: File Format Profiles.* Hitchcock, Steve and Tarrant, David. 66, January 2011, Ariadne. http://eprints.soton.ac.uk/273241/.

41. *Formats over Time: Exploring UK Web History.* Jackson, Andrew N. Toronto : s.n., 2012. iPRES 2012. http://arxiv.org/abs/1210.1714.

42. 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/.

43. List of PDF Software. *Wikipedia.* [Online] [Cited: 24 February 2015.] http://en.wikipedia.org/wiki/List_of_PDF_software.

44. Adobe Reader XI. *Adobe.* [Online] [Cited: 24 February 2015.] http://www.adobe.com/uk/products/reader.html.

45. 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.

46. *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.

47. 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/.

48. 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.

49. JHove PDF-hul Module. *JHove Sourceforge.* [Online] [Cited: 24 February 2015.] http://jhove.sourceforge.net/pdf-hul.html.

50. 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.

51. SPRUCE Characterisation Hackathon. *OPF Wiki.* [Online] 2013. [Cited: 24 February 2015.] http://wiki.opf-labs.org/display/SPR/SPRUCE+Hackathon+Leeds%2C+Unified+Characterisation.

52. 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.

53. *KOST-CECO.* [Online] [Cited: 24 February 2015.] http://kost-ceco.ch/cms/.

54. 3-Heights(TM) PDF to PDF/A Converter. *PDF Tools.* [Online] [Cited: 24 February 2015.] http://www.pdf-tools.com/pdf/pdf-to-pdfa-converter-signature.aspx.

55. 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.

56. Flint. *Github.* [Online] [Cited: 24 February 2015.] http://openpreserve.github.io/flint/.

57. 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.

58. QPDF. *QPDF Sorceforge.* [Online] [Cited: 24 February 2015.] http://qpdf.sourceforge.net/.

59. 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.

60. 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.

61. 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.

62. 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.

63. Albertini, Ange. PDF Tricks. *Corkami.* [Online] 2014. [Cited: 24 February 2015.] https://code.google.com/p/corkami/wiki/PDFTricks.

64. 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 .

65. PDF Format Issues: JavaScript . *OPF Wiki.* [Online] [Cited: 24 February 2015.] http://wiki.opf-labs.org/display/TR/JavaScript.

66. 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.

67. SPRUCE PDF Solution: PDF Characterisation Tool. *OPF Wiki.* [Online] 2011. [Cited: 24 February 2015.] http://wiki.opf-labs.org/display/AQuA/PDF+Characterisation+Tool.

68. *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.

69. PDF Current Threats. *Malware Tracker.* [Online] [Cited: 25 February 2015.] http://www.malwaretracker.com/pdfthreat.php.

70. KOST-Val. *Community Owned digital Preservation Tool Registry (COPTR).* [Online] [Cited: 20 June 2019.] http://coptr.digipres.org/KOST-Val.

71. *Archaeology Data Service.* [Online] [Cited: 20 June 2019.] http://archaeologydataservice.ac.uk/.

72. *Archaeology Data Service.* [Online] [Cited: 24 February 2015.] http://archaeologydataservice.ac.uk/.

73. Adobe PDF Reference Archives. *Adobe.* [Online] [Cited: 24 June 2019.] http://www.adobe.com/devnet/pdf/pdf_reference_archive.html.

74. PDF Reference and Adobe Extensions to the PDF Specification. *Adobe.* [Online] [Cited: 24 June 2019.] http://www.adobe.com/devnet/pdf/pdf_reference.html.

75. Johnson, Duff. Archivists: No flowers for PDF/A-3. *Duff Johnson's Strategy and Communications Blog.* [Online] 28 February 2014. [Cited: 25 June 2019.] http://duff-johnson.com/2014/02/28/archivists-no-flowers-for-pdfa-3/.

76. van der Knijff, Johan. Adobe Portable Document Format: Inventory of long term preservation risks. *Open Preservation Foundation.* [Online] 20 October 2009. [Cited: 25 June 2019.] http://www.openplanetsfoundation.org/system/files/PDFInventoryPreservationRisks_0_2_0.pdf .

< - end - >