Preservation with PDF/A (2nd Edition)

Betsy A Fanning
AIIM

DPC Technology Watch Report 17-01 July 2017

Series editors on behalf of the DPC
Charles Beagrie Ltd.

Principal Investigator for the Series
Neil Beagrie

DigitalPreservationCoalition

ISSN: 2048-7916

DOI: http://dx.doi.org/10.7207/twr17-01

First published in Great Britain in 2008 by the Digital Preservation Coalition. Second Edition 2017.

Foreword

The Digital Preservation Coalition (DPC) is an advocate and catalyst for digital preservation, ensuring our members can deliver resilient long-term access to digital content and services. It is a not-for-profit membership organization whose primary objective is to raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. It supports its members through knowledge exchange, capacity building, assurance, advocacy and partnership. The DPC's vision is to make our digital memory accessible tomorrow.

The *DPC Technology Watch Reports* identify, delineate, monitor and address topics that have a major bearing on ensuring our collected digital memory will be available tomorrow. They provide an advanced introduction in order to support those charged with ensuring a robust digital memory, and they are of general interest to a wide and international audience with interests in computing, information management, collections management and technology. The reports are commissioned after consultation among DPC members about shared priorities and challenges; they are commissioned from experts; and they are thoroughly scrutinized by peers before being released. The authors are asked to provide reports that are informed, current, concise and balanced; that lower the barriers to participation in digital preservation; and that are of wide utility. The reports are a distinctive and lasting contribution to the dissemination of good practice in digital preservation.

This report was written by Betsy A Fanning. The report is published by the DPC in association with Charles Beagrie Ltd. Neil Beagrie, Director of Consultancy at Charles Beagrie Ltd, was commissioned to act as principal investigator for, and managing editor of, this Series in 2011. He has been further supported by an Editorial Board drawn from DPC members and peer reviewers who comment on text prior to release: William Kilbride (Chair), Janet Delve (University of Portsmouth), Marc Fresko (Inforesight), Sarah Higgins (University of Aberystwyth), Tim Keefe (Trinity College Dublin), and Dave Thompson (Wellcome Library).

Acknowledgements

Many subject experts and their organizations have contributed countless hours of work and time to develop the standards that this report describes. Standards work requires a unique type of person, one who is not only an expert in their field but also a person of patience, as standards development takes time.

My appreciation goes to the many experts who have joined in this standards development effort. These include Stephen Levenson, who met me at an industry meeting and shared his 'crazy' idea, which eventually became Portable Document Format/Archive (PDF/A); many from Adobe Systems who willingly shared their knowledge and helped the committee form the requirements that make PDF/A an archival file format; and the archival community, who also shared their knowledge and helped to shape this standard.

Through the development of the PDF/A standard, two project leaders, Stephen Abrams formerly of Harvard University and Leonard Rosenthol of Adobe Systems, kept the project moving at a consistent pace, taking into consideration many differing points of view to develop this standard. The Association for Information and Image Management (AIIM, http://www.aiim.org) and the Association for Suppliers of Printing, Publishing and Converting Technologies (NPES, http://www.npes.org) jointly developed this standard. It was great to see these two organizations work together sharing knowledge and expertise.

I am sure there are many experts whom I should publicly acknowledge, but there is not time or space to do so adequately. I am particularly grateful to Tim Evans for his contribution to the case study, and to Sarah Higgins for the figures. I would be negligent if I did not acknowledge my employer, AIIM, for supporting this standards project. I also want to give special thanks to Neil Beagrie of Charles Beagrie Ltd, and the staff of the DPC, for their input and wise counsel in the development of this report.

Betsy A Fanning

July 2017

Contents

1. Introduction 3
1.1. Technology Watch Report Editions 3
1.2. Overview 3
1.3. Typical Uses for PDF/A 4
2. History and Features of PDF and PDF/A 5
2.1. History of PDF 5
2.2. What is PDF/A? 6
2.3. PDF/A-1 (ISO 19005-1:2005) 7
2.4. PDF/A-2 (ISO 19005-2: 2011) 8
2.5. PDF/A-3 (ISO 19005-3: 2012) 8
2.6. PDF/A-4 (ISO/CD 19005-4) 8
2.7. Why was the Standard Drafted in Multiple Parts? 9
2.8. How are Engineering and Dynamic Documents Archived? 9
3. Conformance and Conformance Levels 10
3.1. Overview 10
3.2. Level A Conformance 10
3.3. Level B Conformance 10
3.4. Level U Conformance 10
4. Metadata and Its Importance for Preservation 11
4.1. Overview 11
4.2. XMP 11
5. Challenges and Lessons Learned 12
5.1. Appropriateness for Use and Reuse: The end user perspective 12
5.2. Importance of Good Information and Preservation Management Practices 12
5.3. User Creation of PDF/A 12
5.4. Migration to PDF/A 13
5.5. Assessing Quality in PDF/A Construction 13
5.6. Electronic Signatures 14
5.7. Preservation Implications of Embedded File Streams in PDF/A-3 15
5.8. Fonts and Intellectual Property 15
5.9. File Size and Image Compression 15
5.10. Adoption 16
6. Future Development of the PDF Standard 17
7. Archaeology Data Service Case Study 18
8. Conclusions and Recommendations 20
8.1. Conclusions 20
8.2. Recommendations 20
9. Glossary 22
10. Further Reading 23

11. References 24
12. Appendix: Standards and Technical Guides 27

Abstract

This report discusses the digital document file format for long-term preservation that is known as PDF/Archive or PDF/A. As organizations have transitioned from a paper-centric to a digital document-centric way of operating, it has become necessary to develop a file format that will maintain the integrity of the information contained in the document and withstand the test of time. These documents need to be available for generations to come, due to their value for multiple purposes. As work began on the standard, it was found that this self-contained preservation file format independent of technology would provide benefits to many organizations and in many circumstances. The need for this preservation file format continues to exist, and new uses of the standard continue to be identified. This report discusses the file format and why it was developed, along with some of the issues and concerns organizations should consider when choosing to use PDF/A as one of their long-term preservation file formats. It is an updated edition of the original *Technology Watch Report 08-02, Preserving the Data Explosion: Using PDF* published in 2008.

Executive Summary

The focus of this report is the PDF/Archive (PDF/A) file format and standard. PDF/A is one of several file formats promoted in digital preservation. This report assesses the claims made for PDF/A and provides guidance on how the format's potential for digital preservation might be achieved.

The report begins with the history of the Portable Document Format (PDF) to better understand how this special file format, PDF/A, came to be. It examines PDF from when it was first created through to when Adobe Systems provided the PDF specification to the International Organization for Standardization (ISO) to formalize it as an ISO Standard.

PDF/A versions of PDF have been developed as a family of ISO Standards with the specific aim of addressing preservation. PDF/A is a restricted form of PDF intended to be suitable for long-term preservation by removing some features that pose preservation risks.

After the initial PDF/A-1 standard was developed, three other standards – PDF/A-2, PDF/A-3, and PDF/A-4 – were also developed to add functionality to the file format. This functionality extended the PDF/A file by enabling native files and XML to be contained in the PDF/A file while maintaining the archival nature of the file.

It is relatively simple to create a PDF, or more accurately, to create a file that for all intents and purposes appears to be a PDF. The same is true for PDF/A, but for preservation purposes it is important to know how closely a file conforms to the detailed requirements defined in the standard. This report will discuss the conformance levels that were developed by the working group. It will also discuss the validation methods that were developed to ensure those conformance levels.

Conformance to the standard is not a simple 'yes/no' binary state, in part because there are now four variants of PDF/A. One question that is often asked is: 'When should I use PDF/A, and which version should I use?' This report attempts to answer that question and to provide some guidance about the strengths, weaknesses, opportunities and threats associated with each. There are several conditions that make it beneficial to use PDF/A-3 rather than PDF/A-1, and vice versa. The report discusses these conditions and reviews practical considerations to make the most effective use of the file format.

Though important, the standards and validation methods described in this report comprise only part of a digital preservation strategy. The selection of a file format – even one carefully developed to support preservation – is not a complete digital preservation solution. The choice of file format is a component of a wider technical and organizational infrastructure which comprises a comprehensive digital preservation solution.

This report provides sufficient information regarding the standard and its use to help readers use the file format better to ensure the integrity of digital information. Through a member case study, it helps readers understand the practical issues involved and lessons learned, and to determine how best to implement PDF/A in their organization.

1. Introduction

1.1. Technology Watch Report Editions

This report is an updated edition of the original *Technology Watch Report 08-02, Preserving the Data Explosion: Using PDF* (Fanning, 2008)**.** Since its original publication in 2008, when only PDF/A-1 (ISO 19005-1) was available, the International Organization for Standardization (ISO) has published two more parts of the PDF/A standard which have added features to the file format: PDF/A-2 (ISO 19005-2) and PDF/A-3 (ISO 19005-3). The development of a fourth part is underway. This new edition of the *Technology Watch Report* will examine all four parts of the PDF/A standard and provide guidance on the appropriate part to use. The report will also take a brief look at the future for the PDF/A standard.

1.2. Overview

PDF became a ubiquitous file format for exchanging electronic copies of page-based documents because of the many benefits it confers on users. Some of its benefits include:

- compatibility across all platforms;
- ability to create compact and small files for easy exchange;
- the ability to create PDF files from source documents;
- easy-to-create PDF files;
- ability to be viewed within most web browsers;
- rich metadata-containing files;
- ability to have other files embedded within it.

However, because the PDF format is feature rich, it can cause difficulties for specific uses such as long-term preservation. And with the advantages of the PDF file format come some risks:

- any file type can be embedded;
- the primary document can be conformant as a static document, but the embedded files may not be static;
- embedded files may be infected by computer viruses;
- embedded files may have extended metadata requirements, may introduce unexpected dependencies or be subject to format obsolescence;
- embedded files may complicate matters relating to information security, data protection or the management of intellectual property rights.

PDF/A (A for Archive) versions of PDF have been developed as a family of ISO Standards with the specific aim of addressing preservation. PDF/A is a restricted form of PDF intended to be suitable for long-term preservation by removing features that pose preservation risks.

PDF/A seeks to maximize:

- device independence;
- self-containment;
- self-documentation.

PDF/A places some restrictions to reduce preservation risks:

- all fonts must be embedded and the fonts must be legally embeddable for unlimited, universal rendering;
- audio and video content are forbidden;
- JavaScript and executable files are prohibited;
- colour spaces must be specified in a device-independent manner;
- encryption is not allowed;

- use of standards-based metadata is mandated.

These restrictions make the PDF/A format a good option for long-term archiving of electronic documents, providing any restricted content is not present or is not required and can be removed.

However, users need to be aware of some other preservation challenges that remain and/or are in the process of being addressed. In particular, a variety of issues have contributed to uncertainty over PDF rendering and the impact this may have on long-term preservation (British Library, 2015). The variable quality and support provided by some PDF-creating software and third-party viewers means institutions have faced challenges in converting files to PDF/A, validating the conformance of files to PDF/A, and fixing faults with the format, particularly when files have been received from a wide body of external organizations and individuals. A robust vendor-independent mechanism for assessing full compliance of PDF/A files with the standards and the conformance levels they claim in their internal metadata will go a long way to addressing this challenge. The recent development and release of the veraPDF tool, a purpose-built, open source, PDF/A file-format validator, is a major step forward (see Section 5.5).

1.3. Typical Uses for PDF/A

There are many reasons why an organization might choose to use PDF/A to preserve their digital documents, including:

- its standardized format for storing digital documents for long periods of time;
- it allows for digitally signed documents using the very latest digital signature software;
- it reliably displays special characters for mathematics and languages since all are embedded within the file;
- it displays correctly on any device as the author intended, including the reading order;
- platform independence;
- provision of fully searchable documents through Optical Character Recognition.

PDF/A can be used in many situations where we want to preserve information, such as:

- scanning documents for archives;
- migrating existing document files into archives;
- digital mailroom processing and retention of incoming and outgoing mail;
- compliance with regulations and addressing regulatory concerns, e.g. in the financial, healthcare, or pharmaceutical sectors;
- eBilling and eProcurement processes where documents need to be entered into a workflow and archived;
- preservation of office documents or official documents;
- preservation of academic reports and publications;
- open government and long-term access to information by citizens.

2. History and Features of PDF and PDF/A

2.1. History of PDF

PDF is a file format originated by Adobe Systems in the early 1990s for the primary purpose of exchanging documents. It was intended to make digital documents essentially similar to their paper equivalents by being authentic, reliable and easy to use. The *PDF Reference* (http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf) is an open specification that defines the features and functions for the PDF file format, and encompasses a number of specifications relating to different uses. Some of these have evolved over time, leading to different versions of the specification.

Adobe made all the *PDF Reference* specifications freely available on their website (http://www.adobe.com/devnet/pdf/pdf_reference_archive.html) and allowed any software developer to use the specification in designing their own products. PDF quickly became a *de facto* standard.

As users became more proficient with digital documents, they began to request functionality that added new features to successive versions of the PDF specification, as shown in Table 1.

Table 1 – Features introduced in PDF

	PDF 1.1	PDF 1.2	PDF 1.3	PDF 1.4	PDF 1.5	PDF 1.6	PDF 1.7
External links	✓	✓	✓	✓	✓	✓	✓
Article threads	✓	✓	✓	✓	✓	✓	✓
Security features	✓	✓	✓	✓	✓	✓	✓
Device-independent colour	✓	✓	✓	✓	✓	✓	✓
Notes	✓	✓	✓	✓	✓	✓	✓
Support for OPI (Open Process Interface) 1.3		✓	✓	✓	✓	✓	✓
Support for CMYK (colour model for cyan, magenta, yellow, and key black)		✓	✓	✓	✓	✓	✓
Maintenance of spot colours in PDF		✓	✓	✓	✓	✓	✓
Halftone functions could be included as well as overprint instructions		✓	✓	✓	✓	✓	✓
2-byte CID fonts			✓	✓	✓	✓	✓
OPI 2.0 specifications			✓	✓	✓	✓	✓
DeviceN, a new colour space to improve support for spot colours			✓	✓	✓	✓	✓
Smooth shading, a technology that allows for efficient and very smooth blends (transitions from one colour or tint to another)			✓	✓	✓	✓	✓
Annotations			✓	✓	✓	✓	✓
Transparency support that allows text or images to be seen through				✓	✓	✓	✓
Improved security				✓	✓	✓	✓
Improved support for JavaScript				✓	✓	✓	✓
Improved compression techniques including object streams and JPEG2000 compression					✓	✓	✓
Support for layers					✓	✓	✓
Improved support for tagged PDF					✓	✓	✓
Improved encryption algorithms						✓	✓
OpenType fonts embedded						✓	✓
Ability to embed files to be a container file format						✓	✓
Ability to embed 3D data						✓	✓
Improved support for commenting and security							✓
3D support improvements							✓

In 2000, Adobe Systems initiated the first of what would become several efforts to standardize subsets of the *PDF Reference* for specific purposes. The first subset to be introduced was for document exchange and became known as PDF/X. After this came numerous other ISO PDF standards (see Figure 1).

The *Appendix* contains a listing of these and other standards pertinent to digital preservation, along with a brief statement of what they cover.

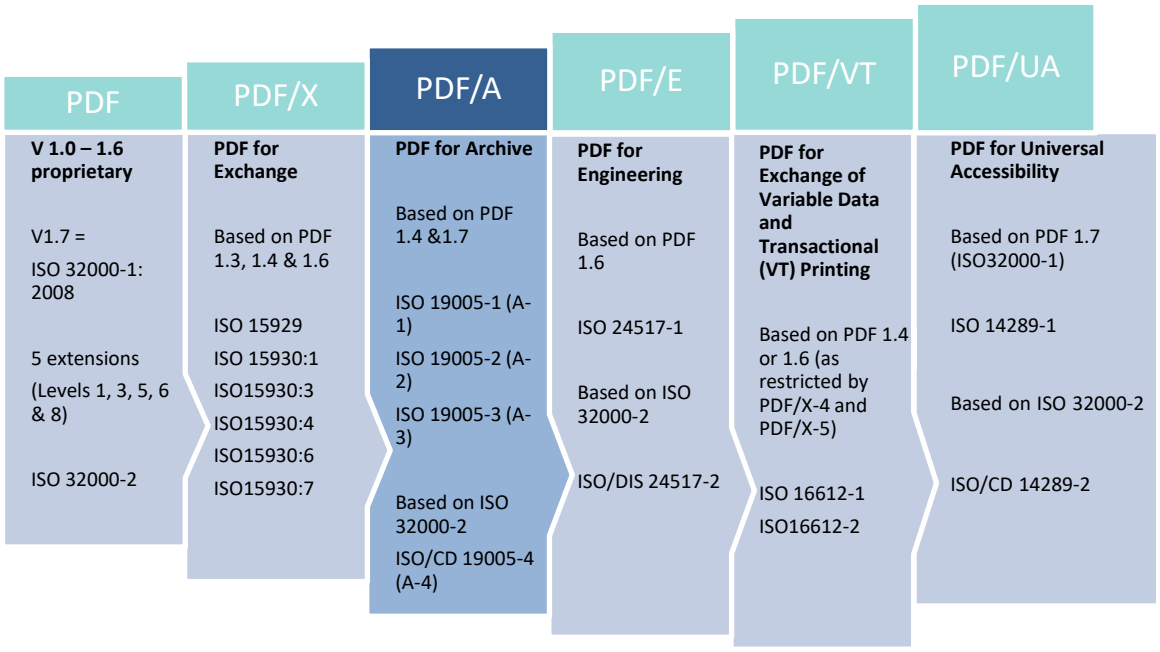

Figure 1 – Flavours of PDF. Illustration ©2015 Sarah Higgins, Aberystwyth University

In 2007, Adobe Systems went a step further to standardize their PDF Reference specification by providing it to ISO to become an open International Standard, ISO 32000-1.

At the time of writing, ISO 32000-2, referred to as PDF 2.0, was in the process of being published by ISO. This new version of PDF introduces many new features that continue to make the PDF file feature rich. These new features may have an impact on the various sub-set standards, including PDF/A.

2.2. What is PDF/A?

PDF/A is an open standard that defines a file format for the preservation of digital documents. The PDF/A file format restricts functionality with a potential for preservation risk. For example, functionality such as multimedia or transparency will be discarded when creating PDF/A files. Therefore, it supports a defined and large, but by no means comprehensive, set of preservation use cases. For example, it is well suited to relatively simple 'static page' documents, such as those produced by word processors or desktop publishing applications, where pagination, layout, graphics and text are important. It is less well suited to documents that include interactive, dynamic or multimedia aspects, such as may be found in spreadsheets or websites.

PDF/A is a self-contained file that includes text, graphics, fonts and metadata to ensure the accurate rendering of the document, regardless of the technology used. The file format makes it easier to undertake full-text search of an archive, as the inclusion of fonts, metadata and so on ensures that the digital document is easily searched.

The family of PDF/A standards was, and is, being developed by an ISO Joint Working Group, and is subject to rigorous technical review. This Joint Working Group brings together experts from the digital preservation and software development communities. It is important to note that the PDF/A standard specifies not only file format requirements, but also requirements for the viewing software.

File format assessment work conducted after the creation of the standard, such as that conducted by the British Library (May *et al.,* 2015), enforces the effectiveness of PDF/A for long-term preservation. This

work suggests that the functionality restrictions defined by PDF/A align closely with well-known preservation risks associated with the PDF format.

The family of PDF/A standards along with their functionality is summarized in Figure 2.

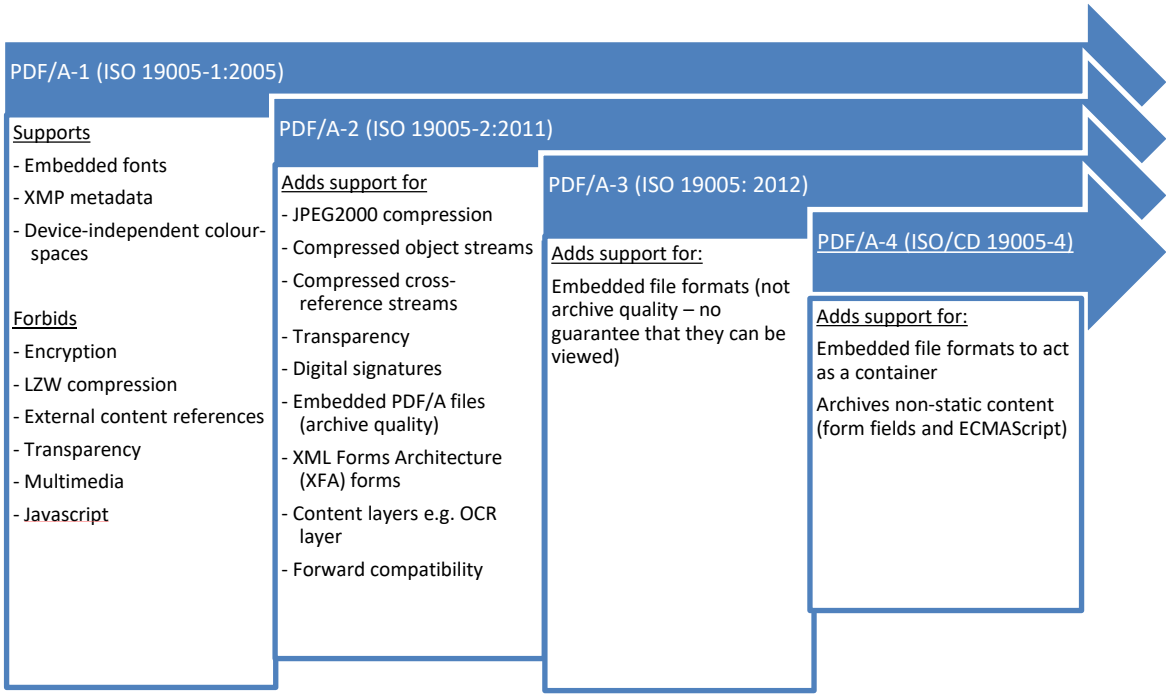

Figure 2 – PDF/A parts and functionality. Illustration©2015 Sarah Higgins, Aberystwyth University

2.3. PDF/A-1 (ISO 19005-1:2005)

The PDF/A-1 file format is based on, and includes, the foundational functionality included in the PDF version 1.4 (*PDF Reference 1.4)*. The committee (for the PDF/E Joint Working Group) was careful to state that PDF readers will ignore the features that are not documented in *PDF Reference 1.4*. However, in order to ensure long-term preservation, it was necessary to limit the functionality of PDF by establishing specific requirements. Therefore, the PDF/A standard specified the features that are allowed and those that are not allowed (AIIM, 2005, PDF/A FAQ) (Figure 2).

PDF/A-1 files allow:

- embedded fonts for unlimited, universal rendering. This includes only those fonts which may be legally embedded without a royalty fee;
- embedded colour spaces;
- device-independent colour;
- XMP (Extensible Metadata Platform) metadata;
- Unicode character map for Level A conformance;
- behaviour for NextPage, PrevPage, FirstPage and LastPage actions;
- tagged PDF.

PDF/A-1 files do <u>not</u> allow:

- encryption, as the method of encryption may not be supported when the files are opened at a later date;
- LZW (Lempel-Ziv-Welch) Compression due to intellectual property constraints;

- embedded files;
- references to external content because these may change or links may be broken;
- PDF transparency – a file can use techniques other than the use of transparency keys to provide the visual effect of partially transparent graphics;
- multimedia;
- JavaScript.

2.4. PDF/A-2 (ISO 19005-2: 2011)

PDF/A-2 extends the capabilities of PDF/A-1 and remains focused on 'static paper' by not including multimedia, 3D, etc. It is based on *PDF Reference 1.7* (as defined in ISO 32000-1) rather than *PDF Reference 1.4* (the basis for PDF/A-1). It takes into account the feedback the committee received from the industry, and new features added include:

- improvements to tagged PDF for enhanced accessibility;
- compressed object and XRef streams for smaller file sizes;
- PDF/A-compliant file attachments, portable collections and PDF packages;
- transparency;
- JPEG2000 compression;
- optional content (layers) for storage of multiple views of the document;
- digital signature enhancements.

This part of the standard is an extension to PDF/A-1 and does not require users of PDF/A-1 to migrate to it.

2.5. PDF/A-3 (ISO 19005-3: 2012)

PDF/A-3 further extends the capabilities of PDF/A-2 while maintaining the document as a 'static' page. It is also based on *PDF Reference 1.7* (as defined in ISO 32000-1). This version of PDF/A allows other files and attachments, for instance XML and/or native source files (word-processing files or spreadsheets), to be embedded in a PDF/A file. The standard provides sufficient information to interpret any conforming PDF/A-3 file. PDF/A viewers are not required to address the embedded files except to extract them if needed. There is no guarantee in the standard that the other files or attachments may be viewed or may be used in the future; PDF/A-3 is simply a container for the other files and attachments.

The inclusion of XML and native files in a PDF/A file was the result of the industry providing the direction for this version of the standard. It became apparent that it was important not only to preserve the collections of documents but also to preserve the native files. The committee heard from users of the file format who needed the native XML to be preserved for historical purposes. Unlike PDF/A-2, where files included in the collection had to be in PDF/A format, a file in a PDF/A-3 file did not need to be in PDF/A format to be included and to keep the file compliant.

As with PDF/A-2, this standard is an extension of the PDF/A-1 standard and does not require the user to replace or convert files. Each part of the PDF/A standard can be used for archival purposes without the use of the other parts.

2.6. PDF/A-4 (ISO/CD 19005-4)

PDF/A-4 extends the capabilities of the previous versions of PDF/A while continuing to maintain the document as a 'static' page. This version is based on ISO 32000-2. As with PDF/A-3, this part of the International Standard enables PDF documents to serve as containers for other file formats by allowing files to be embedded. This means that a single file can contain not only the visual representation, but also other representations, including the original authored version and rich semantic formats. The long-term suitability of the formats that may be embedded is not addressed in this part of the Standard.

This part of the standard introduces some new directions for archiving non-static content that may be represented in PDF documents, such as form fields and ECMAScript. Through inclusion of this, more information in the file can be preserved. A greater burden is placed on conforming viewers to ensure that this information does not alter the visual appearance of the file during consumption.

2.7. Why was the Standard Drafted in Multiple Parts?

The intent of the committee, as it developed the various parts to the PDF/A (ISO 19005) Standard, was to make it easier to implement the Standard. The availability of the parts allows end users to pick and choose the Standard with the features that best meets their needs. Each part is intended to stand alone (so that the files created using it do not become obsolete), while building functionality onto the previous version. This ensures that all PDF/A file formats, whichever version used, remain valid.

Unfortunately, the committee's philosophy of multiple parts resulted in confusion in the market place, making it more difficult for users to select the optimum file format. Many end users assumed they needed to implement the latest version when, in fact, PDF/A-1 was adequate for their needs. This confusion earned the Standard some negative press as being expensive and time-consuming, because users wrongly inferred that to implement it all files would need to be converted to meet the new Standard. While many end-user organizations may want to implement the latest version, the original (PDF/A-1) may meet their requirements. A needs assessment for file format requirements will provide guidance when deciding which PDF/A Standard to implement.

2.8. How are Engineering and Dynamic Documents Archived?

Sharing, publishing and exchanging engineering and mapping data has been a challenge for a sector with many incompatible toolsets, multiple data formats, and a variety of other barriers. This led to the formation of the PDF/E committee, and a focus on addressing these interoperability concerns with the development of PDF/E-1 (PDF/E Joint Working Group, 2008).

Subsequent developments have focused on supporting 3D content and on requirements for long-term preservation. PDF/A is designed to capture and preserve static documents. As such, it does not suit the complex, visual, and typically interactive, nature of engineering documents. It was therefore considered logical to address the unique challenges of preserving content of this kind within the dedicated PDF/E Standard. PDF/E-2 is expected to be published in 2018.

As stated in the Introduction of ISO/DIS 24517-2, *Document management – Engineering document format using PDF – Part 2: Use of ISO 32000-2 including support for long-term preservation, PDF/E-2*, PDF/E-2 is based on ISO 32000-2 rather than on PDF 1.6 as PDF/E-1 is. In addition to extending the capabilities of Part 1, this part of the Standard includes:

- support for numerous enhancements to 3D, including support for PRC
- support for Geospatial information (GIS) in both 2D and 3D
- compressed object and XRef streams (for smaller file sizes)
- transparency
- JPEG2000 compression.

3. Conformance and Conformance Levels

3.1. Overview

The PDF/A Standard establishes various levels of conformance for the file. This was intended to prevent the onerous requirements for full conformance presenting a barrier to software developers. At a bare minimum, the conformance levels require the file to comply with the version of PDF used as the basis for the PDF/A Standard.

The burden of compliance rests with the software developer, not the end user. Periodically, end users may examine representative files using a test suite to ensure that the software performs as expected, and that the organization is following the guidelines. Ultimately, users need to select software applications that generate compliant files. When organizations consistently select compliant software, an incentive is provided to vendors and developers to guarantee their products meet standard requirements, and to strive to maintain compliance through improving output quality. However, it is not a simple task for a user to select software that produces conformant file outputs, or to validate vendor claims of compliance. There are three PDF/A conformance levels: A, B, and U. Each of the first three parts of the PDF/A (ISO 19005) family of Standards uses the conformance levels A and B. The use of conformance level U began with PDF/A-2 (ISO 19005-2) and is carried forward for PDF/A-3 (ISO 19005-3). With the introduction of PDF/A-4 (ISO 19005-4), a PDF/A-4 conformance file does not need to specifically identify a conformance level, as this part of the Standard mandates the various requirements without a need for explicit identification. The multiplicity of the Standard's parts and detailed documentation for requirements make it a complex task for archival organizations to evaluate different software products.

3.2. Level A Conformance

Level A conformance is denoted as 'PDF/A-1a', 'PDF/A-2a', or 'PDF/A-3a'. Level A-conforming files or readers must meet all the requirements of the *PDF Reference 1.4* as modified by the PDF/A-1 (ISO 19005-1) Standard. Conforming Level A files may include any feature in *PDF Reference 1.4* or an earlier version of *PDF Reference* that is not forbidden in PDF/A-1 (ISO 19005-1).

3.3. Level B Conformance

Level B conformance is denoted as 'PDF/A-1b', 'PDF/A-2b', or 'PDF/A-3b'. Level B-conforming files must meet all the requirements in PDF/A-1 (ISO 19005-1) except those identified in clause 6.3.8 of the Standard pertaining to Unicode character maps, and clause 6.8 pertaining to the logical structure. The Level B requirements are intended to be the minimum necessary to ensure that the rendered visual appearance of a conforming file is preserved over the long term. Level B-conforming files may not be sufficiently rich to ensure the preservation of the document's logical structure and content text stream in a logical reading order. This conformance level is often used with documents that are scanned from paper. In the instance of digitized documents, files must be stored as 'b' compliant documents that are minimally conforming. If it is necessary to have the full conformance file created, then it will be necessary to run the file run through a separate process to capture the text using OCR technology, especially if the file is to be searchable.

3.4. Level U Conformance

Level U conformance is denoted as 'PDF/A-2u', or 'PDF/A-3u'. Level U conformance requires conformance to all the requirements of PDF/A-2 (ISO 19005-2) except those in clause 6.7, Logical structure. The Level U conformance requirements are intended to be those necessary to ensure that not only is the rendered visual appearance of the conforming file preserved, but that any text contained in the document can be extracted as a series of Unicode code points.

4. Metadata and its Importance for Preservation

4.1. Overview

Metadata is important for the effective management of a file throughout its life cycle as well as for document discovery in searches. Establishing a long-term digital document preservation system requires careful consideration of the metadata that will be needed to locate and render documents years from now. Such metadata should:

- be appropriate to the materials;
- support interoperability;
- use standardized controlled vocabulary;
- include clear statements on the conditions and terms of use;
- be authoritative and verifiable;
- support the long-term management of the document.

The collection of metadata for PDF/A documents is optional. A PDF/A file may be created without metadata describing the content of the document. However, if no metadata is associated with the file, problems may be encountered later when trying to locate and/or use the file.

One element of metadata is required – the PDF/A identifier. This is automatically generated when the PDF/A file is created. Other metadata requirements should be formalized. This means identifying the metadata you want to collect and store, and it may be helpful to explore predefined metadata schema such as PREMIS (Preservation Metadata: Implementation Strategies) and others to pick the useful metadata elements for your application. For further information and guidance see *Preservation Metadata* (Gartner and Lavoie, 2013).

The PDF/A Standard requires the metadata to be encoded in an XMP-compliant format. It would be wise for organizations to make the most use of what XMP has to offer when archiving documents, so that the retrieval of documents will be easier and more effective.

4.2. XMP

PDF/A requires the use of Adobe's Extensible Metadata Platform (XMP), an ISO Standard for the creation, processing and interchange of metadata (Wikipedia, 2016). XMP is used to encode the document's metadata within the PDF/A file, and is based on the Resource Description Framework (RDF). This was developed by the World Wide Web Consortium (W3C) and is the cornerstone to the semantic web. Using XMP, applications can access and understand the metadata of the documents that they manipulate.

XMP standardizes the definition, creation, and processing of metadata through data models, storage models and schemas. The XMP data model describes the metadata in the document. The storage model describes how XMP is embedded into the file. The schemas are predefined sets of metadata property definitions that are relevant for applications.

XMP metadata consists of a set of properties associated with a resource, which might be a file or a portion of a file. The schemas are defined by documentation and identified by an XML namespace URI (Universal Resource Identifier) which names or identifies a specific resource. The use of namespaces avoids conflicts between properties present in different schemas used to describe the data being included. XMP has been formalized as ISO 16684, *Graphic technology – Extensible metadata platform (XMP) specification*. This Standard was developed to have two parts. Part 1 covers data models, serialization and core properties while Part 2 discusses schemas using RELAX NG. XMP has been integrated into PDF files since PDF Specification 1.4. It is required in PDF/A for identification and to associate the metadata with the file. The XMP Standard provides a crosswalk for mapping the metadata information.

5. Challenges and Lessons Learned

This section provides a high-level overview of a number of the key issues that may be encountered when creating, handling, or preserving PDF/A files.

5.1. Appropriateness for Use and Reuse: The end user perspective

As the early sections of this report indicate, PDF was originally designed as a page description format. Considerable development of the format and related standards has expanded the target use cases considerably, and PDF is widely used for different purposes in different sectors. However, PDF does not necessarily suit all possible use cases, and therefore PDF/A will not be ideal for preserving all digital material.

PDF is ideal for capturing and consistently presenting a page-based document for reading by a user. It is less perfect for presenting data for reuse, extraction and further processing and analysis. For example, converting a spreadsheet from its native format to PDF is likely to remove most of the underlying meaning that connects the various spreadsheet cells. A PDF representation of a spreadsheet table presents the final figures, not the working. Furthermore, extracting the data in the table to a usable form for further analysis or automated processing may also be a challenge. The rapid growth in document scraping and analysis tools provides some hope, but is also indicative of the barriers that PDF can place on data reuse. To a human reader, these issues may not be a problem. To a data scientist seeking to computationally analyse the data further, or to an Internet service seeking to automatically process a bus timetable, publishing data in the PDF format can become a significant obstacle.

Section 5.7 discusses how the development of PDF/A-3 has sought to address these issues.

5.2. Importance of Good Information and Preservation Management Practices

PDF/A is sometimes presented as a 'magic bullet' for digital preservation, as if using PDF/A will deliver long-term access. This false impression ought to be vigorously refuted. Judicious selection of file formats alone – even one specifically and meticulously adapted for preservation purposes like PDF/A – will not deliver long-term access to digital materials. A robust information and preservation management programme with policies and procedures is also needed. ISO 19005, the PDF/A family of Standards, has been very strict about not dealing with, or making recommendations or requirements for, archival storage of document files. This silence is because such matters are outside the scope of the Standard, even though they are required (for guidance on the wider requirements of digital preservation see Digital Preservation Coalition, 2015).

5.3. User Creation of PDF/A

A number of solutions are available to convert digital documents to PDF/A format. PDF technologies have fundamentally two functions: to create documents in PDF/A format and to render PDF/A documents for reading or printing. The means to convert native documents from word processing, spreadsheets, and presentation files and even email to PDF/A format is integrated with many software packages, or can be accessed through dedicated PDF/A creation tools.

PDF and PDF/A creation functions are increasingly embedded in office equipment such as printers, scanners and photocopiers. Because such software is integrated into the device control system, there is limited scope for ensuring the extent to which files comply with the Standard. Consequently, it can be important to consider the impact of integrated software when selecting office equipment.

When preserving, organizations have a degree of control over the environment, technology and practices of those creating PDF and PDF/A files, small changes in practice have the potential to reap considerable dividends for preservation later in the life cycle. For example, standard desktop operating system images that are managed by an organization's IT department and installed on staff computers across the organization could quite easily be configured so that the office applications export to PDF/A, rather than PDF, by default.

The PDF Association maintains a list of software applications that can be used to capture or convert native documents into PDF/A format. You can find this listing on the PDF Association website, http://www.pdfa.org.

5.4. Migration to PDF/A

Wholesale migration of PDF files to PDF/A will not necessarily ensure their longevity, and done poorly could even be damaging to their content. When migrating files of one format to another, consideration should be given to an array of issues before proceeding. What is the goal of the migration? Can the content of the source format be represented with sufficient accuracy in the destination format? What will the impact of the migration be on users of the content? What will it cost? Are there any legal implications? How can the quality and completeness of the operation be verified? And so on. These considerations are typically considered as part of a formal Preservation Planning process. This process guides the user through a decision-making workflow, ensuring alternatives are considered, the best outcome selected, and the process and end result properly documented (see Becker *et al*., 2009).

Where PDF/A is considered as an appropriate format for preserving or providing access to a collection of files of another format (see Section 5.1), the details of the migration process equally require careful consideration.

The various PDF/A Standards require that certain features be present, whilst forbidding others from being used. The aim of this approach is to increase the likelihood of successful rendering, use and interpretation of information encoded as PDF/A in the long term. By ensuring PDF/A files are well structured and easily understood, ambiguities in their interpretation are minimized, external dependencies are removed, and functionality that may be difficult to preserve in future is excluded. Whilst these constraints may simplify future preservation action, they may significantly impact or even restrict what content can be preserved.

Particular issues relating to migration to PDF/A include:

- various PDF/A Standards forbid different features and functionality, so the target Standard should be chosen carefully. PDF/A-1 for example, forbids transparency, but this feature is permitted in PDF/A-2 and PDF/A-3.
- PDF/A forbids interactive content such as embedded multimedia and JavaScript. Restricting the use of these features entirely might make PDF/A documents easier to preserve in the long term, but if multimedia content is an essential part of the document to be preserved, PDF/A will not make an ideal destination format.
- PDF/A requires that fonts used within a document are embedded in the file. If the font is available as part of the migration source, it should be successfully embedded making it available for subsequent rendering. But if the font is not available (for example, when a PDF without embedded fonts is migrated to a PDF/A in a different environment to the one in which the PDF was created), a substitution will have to be made by the migration software. This rather negates the objective of embedding in the first place, particularly if a poor substitution is made. Either way, the result is likely to be a valid PDF/A (as would be assessed by a validation tool) but not necessarily a useful preservation result.
- Migration from PDF to PDF/A will break embedded electronic signatures, and so consideration must be given to when files are signed in an organization's workflow (see Section 5.6 for more information).

5.5. Assessing Quality in PDF/A Construction

Ambiguities in the PDF specification and resulting tolerances in viewing software such as Adobe Acrobat have led to a growing issue of the management and preservation of poorly constructed PDF files (Morrissey, 2012). Incompatibilities resulting from the large number of different applications that are utilized to create and to render PDF files could hinder preservation efforts. Inconsistencies in the creation of PDF/A files have also been encountered; just because a file purports (via a file extension or the metadata in its header) to be a PDF/A does not necessarily mean that it is.

Format validation is a process where a file is checked against the rules defined by a particular file format specification. Validating a file can increase confidence in the likelihood that a viewer closely aligned with the format specification will render the file correctly. In turn, this increases confidence that future users will be able to render the file without a loss of information. It also provides the possibility of identifying

problematic files. Resulting mitigation actions might include selecting alternative PDF creation software (and re-generating the file), or even altering the file where issues have been detected (see Section 5.4).

PDF/A validation can provide greater confidence that a particular PDF/A file does indeed meet the requirements defined within the Standard. A number of PDF/A validators are available, including:

- veraPDF (http://verapdf.org)
- Preflight (https://www.callassoftware.com/en/preflight-pdf-files)
- 3-Heights (http://www.pdf-tools.com/pdf/pdf-validator-pdfa-validate-iso.aspx)
- PDF/A Validator (http://www.validatepdfa.com)

Research has indicated that PDF validators do not always agree on the validity or otherwise of files (PDFlib, 2009; Koo *et a*l., 2012). However, test suites provide a means of assessing the compliance of software with relevant Standards. In particular, this enables validation tools themselves to be assessed. Test suites typically consist of a number of files that exhibit particular features that are defined by, or in cases such as PDF/A, restricted by, the format specification. A number of test suites are available including:

- Isartor (http://www.pdfa.org/2011/08/download-isartor-test-suite/)
- veraPDF Corpus (https://github.com/veraPDF/veraPDF-corpus)

It is perhaps unsurprising that PDF/A validators are likely to disagree to some extent on the validation results they generate, given the ambiguities present in the *PDF Reference* noted above. The veraPDF Project is, at the time of writing, aiming to address this challenge. It has developed a new PDF/A validation tool that provides comprehensive validity checks closely paired with precise format specification rules and a new test suite. It is hoped that this development will help to remove ambiguities and close loopholes in the format specification, and as a result improve the quality and consistency of PDF/A creation tools and the PDF/A files that they create.

See Section 7 for examples of organizations applying validation in practice.

5.6. Electronic Signatures

Electronic signatures have been a part of the PDF Standard since *PDF Reference 1.3*. PDF/A-2 permits the use of digital signatures compliant with the PDF Advanced Electronic Signatures (PAdES) Standard. Since PDF electronic signatures include a visual representation, the appearance must conform to all the requirements of PDF/A, including font embedding and the use of device-independent colour. Not all commercial electronic signature tools follow these requirements. Electronic signatures are handled as annotations in a PDF document, and must have an appearance stream that defines the visual appearance of the electronic signature, if it is to be visible. (There are electronic signatures that are invisible that are acceptable in a PDF/A file as well.)

Additionally, for PDF/A compliance, the font for the signature and signature information must be embedded. The colour of the signature graphic must also be compliant with the requirements in PDF/A, which means the ICC colour profile must be embedded in the file.

It is important to note that if a document was signed and the document is to be converted to PDF/A, the conversion will break the signature. A good practice is to not convert previously signed documents to PDF/A format, or to have a policy in place where the document is re-signed when it is converted to PDF/A. If the signature is important, you should maintain the original copy. The original signed document can be archived with the PDF/A version of the document, including a correction report with the PDF/A document. Over time the electronic signature will expire due to the associated certificates lapsing and changes in cryptographic algorithms. To resolve this issue, one alternative may be to have the documents re-signed on a periodic basis. This can be an arduous task, given the volume of documents in the repository and the risk of damage to files that may result from changing them. An alternative would be to use an archiving system that implements re-signing based on hash-tree algorithms. This alternative will not impact the PDF/A file once it is re-signed.

5.7. Preservation Implications of Embedded File Streams in PDF/A-3

PDF/A-3 (ISO 19005-3) permits the embedding of file streams of any file format, a significant change from PDF/A-2, which only permits other PDF/A files to be embedded. From the perspective of a PDF viewer, these embedded files are considered as supplemental and do not need to be rendered (but could be opened by a user in an alternative viewer application). There are, however, substantial implications for the creator, preserver and user.

The creator might embed a file of another format for a variety of reasons. Embedding the source file from which the PDF/A was generated provides a way of capturing information that might otherwise be lost in the conversion to PDF/A (see Section 5.1). Alternatively, an embedded file might simply represent some additional information that could, for example, provide context to the information in the PDF/A.

The introduction of this feature to the PDF Standards has proven controversial, with some touting the potential for enhanced preservation and others noting potential concerns as to how it will be utilized and the impact on preserving organizations.

A number of scenarios within which the application of PDF/A-3 might be beneficial are identified by Fresko, including that of capturing and preserving source content that would otherwise be lost on generation of a PDF/A-1 or PDF/A-2 (Fresko, 2013). For example, it may be desirable to archive a Word file containing an embedded Excel table. Converting the Word file to PDF will flatten the table to the numbers and letters present in the cells, and in doing so lose any hidden calculations or formulae. Applying PDF/A-3 to this scenario would enable a flattened and easily preserved PDF to be generated, whilst also including the source Word file (and embedded source Excel file), should it be needed. Maintenance of a source file prior to format migration is highly recommended within a preservation context (DPC, 2017 'Preservation Action', *Organisational Activities*). PDF/A-3 therefore has the potential to meet this requirement, even where these migrations are performed well before the influence of methodical, archival preservation planning is felt in the digital life cycle.

As Jenny Mitcham points out in her blog post 'Some Thoughts on PDF/A 3', embedding files '…provides a headache for digital archivists as any file that was deposited in an archive in PDF/A-3 format would then have to be assessed for the presence of embedded files and a separate check on both their value and longevity would need to be made' (Mitcham, 2013). Similar concerns were voiced by a US National Digital Stewardship Alliance (NDSA) Standards and Practices Working Group that stated: 'The PDF/A-3 specification does provide a required mechanism, the AFRelationship key, for expressing a relationship between the embedded file and the primary document, but the specification suggests no methods of verifying the stated relationship, nor is there any means in principle of doing so' (NDSA, 2014). It remains unclear how this metadata field will be utilized by PDF creation tools, and what input might be required from users to populate it.

Fresko has argued that much of the discussion of PDF/A-3 seems to be predicated on the idea that people might use it wrongly, which is not a sufficient argument against its use (Fresko, 2013). But it seems likely that preserving organizations will remain alert to the potential risks, whilst the necessary tool support and subsequent implementations begin to appear over the coming years.

5.8. Fonts and Intellectual Property

The PDF/A Standards require that fonts utilized within a particular PDF/A file must be embedded in that file so that a viewer application can reproduce the exact appearance of the text when it is rendered at a later date or in an entirely different environment to the one in which it was created. However, some fonts are subject to copyright, and there may be restrictions on how they 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.

5.9. File Size, and Image Compression

When PDF/A-1 (ISO 19005-1) was introduced, some concern was expressed with regard to the potential file size of typical PDF/A files, noting the need to embed content while also meeting restrictions on permitted compression algorithms. Subsequently, PDF/A-2 introduced compressed objects and XRef streams to minimize file sizes. JPEG2000 image compression was also permitted in PDF/A-2. This has the potential to dramatically reduce file sizes where bitmap images are present, but also introduces

additional complexity and potential long-term preservation concerns (Buckley, 2008). It remains to be seen whether the potential for embedding source files in PDF/A-3 will result in a significant increase in file sizes (see Section 5.7).

5.10. Adoption

A key criterion in determining the success of any Standards development effort is its adoption.

According to AIIM research, the adoption of PDF as a preferred and more flexible and searchable image format has progressed, but the take up of PDF/A has been surprisingly slow. As stated in AIIM's research (AIIM, 2014), only 17% of surveyed organizations created half or more of their scanned documents in the PDF/A format. On average, AIIM finds 22.9% of scanned documents are saved as PDF/A compared to 59.8% saved as PDF.

Figures from the UK Web Archive indicate that PDF/A files remain a niche format on public facing websites. Only 0.5% of the 23million PDFs archived in UK domain crawls from 2013 to 2015 were identified by DROID as PDF/A. Of these PDF/A files, only a handful were PDF/A-2 or PDF/A-3.

Anecdotal evidence from the veraPDF Project indicated slow adoption of PDF/A by memory institutions, and a lack of any significant numbers of PDF/A files acquired by deposit. As a result, the Project struggled to solicit test data from DPC and OPF members. Some of the larger libraries have employed PDF/A as an access format for digitized books, but typically retain the original component parts as preservation copies. Many memory organizations encourage deposit of PDF/A rather than PDF, but this has not resulted in significant PDF/A deposits. Few organizations have undertaken large-scale migration of PDF content to PDF/A, but an example is provided in Section 7.

6. Future Development of the PDF Standard

Development work on the PDF Standards is a continuing effort. At the time of publication, the following is the status of some of the PDF Standards work:

- PDF/A – during the International PDF/A Working Group meetings, the Working Group agreed to the development of a new part (PDF/A-4). Part 4 will be based on ISO 32000-2 and will include additional support for using the file format as a container for other file formats. This new part will provide recommendations on how to properly archive content that uses some of the newer features in ISO 32000-2, including page level output intents, associated files and tagged PDF improvements.

- PDF/E – The International PDF/E working group is finalizing part 2 for ISO 24517, which adds features for 3D models and features for the long-term preservation of engineering documents.

- PDF/UA – the PDF/UA International Working Group published the first part of their Standard in 2014, the first Standard published in an accessible format using its own Standard for guidance. The Working Group is working on the next part of the Standard that will add features and functions to the accessible format.

- PDF 2.0 – the ISO PDF 32000 Working Group finalized a draft that adds a number of features to the PDF file format. ISO 32000-2 (PDF 2.0) is to be published in 2017.

7. Archaeology Data Service Case Study

The Archaeology Data Service (ADS), established in 1996 and based at the Department of Archaeology at the University of York, has developed expertise in applying PDF/A technology to the preservation of archaeological records. ADS is a digital-only archive for UK-based fieldwork and research in archaeology. As academic research archives provide the most important surviving results of fieldwork that is often destructive in nature, the preserved digital record is a critical resource for future interpretation and study of the archaeological record.

At the time of writing, ADS's growing collection includes 1,100,000 metadata records, 30,000 reports and 700 rich archives. The documents preserved in these archives were normally deposited as file formats such as DOC, DOCX, RTF, ODT. The documents often comprise text and raster/vector elements created using a range of software including Adobe Illustrator.

ADS's file acceptance guidelines allow for text to be submitted in native Microsoft Word format, raster images in TIFF format and vector data as DXF. This guidance was the result of concern over the longevity of the PDF file format and the lack of suitable preservation format for PDF.

In line with the initial development and intended use of the PDF format, ADS used it as a mechanism for dissemination. However, on some occasions PDF was the only format that could be deposited for preservation, and this caused a problem for them. The document had to be copied and pasted into Microsoft Word or OpenOffice, saved in alternate formats such as XML or, as a last resort, each page saved in .TIF format. Problems posed by these methods included the potential loss of image quality, as well as lack of OCR in a raster format. The extra files created in the process provided headaches for managing future file migrations.

Luckily, ADS was dealing predominantly with academic archives, which were mostly controlled and had a small volume of PDF documents. In the early 2000s, they partnered with national and local governments and launched the Online Access to the Index of Archaeological Interventions (OASIS) system. OASIS is an online system for handling, sharing and archiving the records of archaeological investigations. It is aimed at primarily at development-led fieldwork undertaken as mitigation in the planning process, on average around 4,000 events a year in England alone. Many of these are small in scale but, taken collectively, are a vitally important account of the archaeological record as it is revealed each year. Each archaeological fieldwork 'event' produces a grey-literature report which is submitted to the local authority as part of 'preservation by record' initiative. At the time of writing, the resulting corpus contains around 39,000 reports from around the UK.

The reports contained text data but also images, tables, graphs, and vector data. As the use of OASIS has increased, it has been noted that the most common file type uploaded has been PDF, a file that to the user offers simple dissemination without the need for specialist software, and a consistency in appearance whether the report is read online or printed. Considering a lack of reliability in using other methods (above) for creating a file format for long-term preservation, and following the recent release of PDF/A-1 Standard, ADS decided to use PDF/A as a preservation format for these files.

Initially, all conversions to PDF/A of documents used Adobe Acrobat 8, then Pro 9 and 10. The files were manually validated using the included Preflight tool. It was almost impossible to migrate an archaeological report to the PDF/A-1a conformance level, so the Archaeology Data Service used PDF/A-1b as the default option. This proved to be problematic as there were issues with XMP, glyphs and fonts not being embedded, although most of these issues could be fixed with manual intervention. The uncertainty connected with 'Print to PDF/A-1b' led to a lot of time being spent in manual checks of the PDF files to ensure the significant preservation properties were retained.

Since mid-2011, ADS has used PDF/A manager from PDFTron for batch-level processing and automation of previously manual fixes to the files. Using this tool, on average 80% of the files processed were successfully converted to PDF as determined by the validation process. ADS continues to use a combination of PDFTron and Preflight to validate their documents. The work performed by ADS revealed a significant number of the files were not validating as PDF/A files, which was attributed to fine tuning of PDF/A-1 Standard throughout the various Adobe Acrobat versions 7-9 as well as with third-party tools.

Moving forward, ADS has found that using PDF/A-2 (ISO 19005-2) suits their needs as PDF/A-1 is too restrictive and doesn't deal well with aspects of PDF which have become available since PDF 1.4. ADS is

making use of both PDF/A-1 and PDF/A-2 with the understanding that some files will only convert to PDF/A-2 (and not to PDF/A-1).

In essence, ADS has adopted a 'best fit' policy using the two Standards. Some initial concerns over the formats and preservation of embedded content within the PDF/A-3 Standard meant that ADS will address if and how they use PDF/A-3 once the wider archival community has had a chance to consider and reflect on the format. For now, they are using the Callas PDF Toolbox, which provides them with the same validation tool as Adobe Acrobat X Pro and consistent validation, along with reporting and batch conversion. Where a 'straight' conversion to PDF/A-1 was not possible they allowed Preflight to turn pages into raster images and put the text into an invisible layer so that the documents were still searchable. This is not ideal, as it loses some aspects of the vector illustrations that are seen as significant to archaeologists (Evans, 2015).

.

8. Conclusions and Recommendations

8.1. Conclusions

PDF/A is one of a small number of file formats designed to enable digital preservation. This report has sought to provide guidance on how the format's potential for digital preservation might be achieved. In short, PDF/A does indeed facilitate digital preservation, but only if it is thoughtfully deployed in a wider digital preservation architecture.

Fundamentally, PDF/A is a subset of the Portable Document Format (PDF) so it exhibits many of the Format's properties and is most usefully deployed in use cases associated with PDF. Although proprietary in origin, it is now part of a family of ISO Standards. By excluding specific features that pose particular preservation risks, removing dependencies, and ensuring consistent interpretation, it is more robust for the long term than the generic PDF format. However, the PDF format is a pervasive form, so any technology or Standard which can reduce the preservation risks associated with it will have a profound effect on ensuring long-term access to digital documents.

There are currently four parts of the PDF/A Standard, with the initial PDF/A-1 format as a foundation and PDF/A-2, PDF/A-3 and PDF/A-4 adding functionality such as enabling native files to be embedded into the file. All four variants remain current simultaneously but their different functionalities mean subtly different preservation challenges arise.

PDF files are ubiquitous in part because it is relatively simple to create them. More accurately, it is easy to create a file that behaves as though it were a PDF, but which on closer inspection does not fully conform to a rigorous interpretation of the Standard. For preservation purposes, it is important to know how closely a file does confirm to the detailed requirements defined in the Standard, and processes have been developed to validate files accordingly.

Conformance to the Standard is not a simple 'yes/no' binary state, in part because there are now multiple variants of PDF/A. This report has provided guidance about the strengths, weaknesses, opportunities and threats associated with each of them. The final answer on which PDF/A version to use depends heavily on context, as does whether PDF/A is a viable preservation option. There are several conditions that make it beneficial to use PDF/A-3 rather than PDF/A-1, and vice versa.

Whatever the strengths and weakness of PDF/A, selection of a file format – even one carefully developed to support preservation – is not a complete digital preservation solution in itself. It is a component of a wider technical and organizational infrastructure, which comprises a comprehensive and continuous effort to ensure robust access to digital materials. As with all components of a digital preservation strategy, the selection of PDF/A is likely to be a contingent solution to an enduring but emerging challenge. This is especially evident in view of the ongoing development of functionality evident in PDF/A-2 and PDF/A-3. Consequently, the selection of PDF/A should be subject to periodic review. Moreover, digital preservation subject specialists are well advised to stay abreast of changes to the Standard and, where appropriate, to be involved in its development.

8.2. Recommendations

For those evaluating PDF/A as a digital preservation solution:

- The use of PDF/A depends on identifying your organizational requirements and it is not appropriate to every context. So before adopting PDF/A as a preservation solution it is essential to understand the organizational requirements and how PDF/A will support these.
- PDF/A is not a preservation solution on its own: rather it needs to integrate with other components of a wider preservation strategy. Consequently, it is necessary to ensure that it is consistent with other components of the preservation infrastructure, and indeed that there is a clear strategy to maintain and develop those other elements, such as backups, integrity checks and documentation.
- Different versions of PDF/A are fit for different purposes and have subtly different capabilities and preservation risks associated with them. These should be understood and a choice made as to which version is most appropriate. It may also be necessary to assert an appropriate conformance level. Where appropriate, these decisions should be documented and explained.
- Different vendors offer different tools to produce and manage PDF/A within an organization.

These should be compared against your requirements, including those that helped selection of the PDF/A version and the conformance levels.

For organizations collecting and preserving digital data:

- In some contexts, such as a collecting archive or library, it is not possible to control or restrict how documents are produced. However, it may be useful to give specific guidance or instructions prior to submission, such as on version and conformance level.

- Embed PDF/A validation tools into preservation workflows and record the results to help manage the digital preservation risks associated with PDF/A files received.

- Share PDF/A preservation experiences with relevant bodies and organizations to help improve our communal ability to understand, manage and work with PDF/A. This includes sharing problematic files, reporting bugs to tool developers and sharing preservation learning via communities such as the Digital Preservation Coalition.

For records managers and record creators:

- In other contexts, such as in government or business, the archive is directly managed by the agency that generates it. In these contexts, it may also be possible to configure or optimize the tools that generate or access PDF tools to ensure the consistency of PDF/A through the whole document life cycle. In this way, documents are 'preservation-ready' from the point of creation. This is likely to require an intervention at the point where enterprise systems, such as EDRMS, are procured and deployed, to ensure that format selection and conformance levels are built into system requirements.

- Produce a set of guidelines to help users and staff properly create their PDF/A files. Examples of such guidelines include the Koninklijke Bibliotheek in the Netherlands (Koninklijke Bibliotheek, 2007) and the UK Data Archive (UKDA, 2014).

- Digital signatures create a distinct challenge to documents migrated from PDF to PDF/A. A digital signature is likely to be an important property of a document, so it will be necessary to ensure that the workflow generating PDF/As does not inadvertently strip it out. Alternatively, it may be necessary to create a transparent mechanism to attach a new signature during the migration process. One way or another, a clear policy position should be developed to ensure the integrity of documents is not inadvertently compromised by migration.

For the digital preservation community:

- Track the development of the PDF/A Standard, intervening where necessary to ensure that it remains fit for purpose, advocate PDF/A for the preservation of office documents, and encourage the development of preservation-friendly tools to generate and process conformant PDF/As.

For the PDF developer community:

- Track standards and best practice within the digital preservation community to ensure that developments of PDF/A and related products remain closely fitted to emerging solutions and requirements in digital preservation.

9. Glossary

Conformance level – identified set of restrictions and requirements to which PDF/A files and readers must comply (ISO 19005-1, 2005).

Digital document (electronic document) – electronic representation of a page-oriented aggregation of text and graphic data, and metadata useful to identify, understand and render that data, that can be reproduced on paper or optical microform without significant loss of its information content (ISO 19005-1, 2005).

Digital preservation – the series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Digital preservation is defined very broadly for the purposes of this study and refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological and organizational change. Those materials may be records created during the day-to-day business of an organization; 'born-digital' materials created for a specific purpose (e.g. teaching resources); or the products of digitization projects. It should not be confused with digitization, which is related but distinct (Digital Preservation Handbook, DPC 2016).

Long term - period of time long enough for there to be concern about the impact of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository, which may extend into the indefinite future (ISO 19005-1, 2005).

Migration – a means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology (Digital Preservation Handbook, DPC, 2016).

Records – information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business (ISO 15489-1:2001).

Records management – field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including processes for capturing and maintaining evidence of, and information about, business activities and transactions in the form of records (ISO 15489-1:2001).

Standard – a document, established by a consensus and approved by a recognized body, that provides for common and repeated use, rules, guidelines, or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context (ISO/IEC, Guide 21,2005).

For other terms, their relationships and meaning see DPC, 2016, Glossary:
http://dpconline.org/handbook/glossary

10.Further Reading

Associations and Standards Developers:

AIIM (http://www.aiim.org) is the global community of information professionals. The mission of AIIM is to help information workers and their organizations survive and thrive in the era of information chaos.

In 2016, the Digital Preservation Coalition released the second edition of *The Digital Preservation Handbook* to provide comprehensive and up-to-date guidance on matters pertaining to digital preservation: http://dpconline.org/handbook

ISO (International Organization for Standardization) (http://www.iso.org) is an independent, non-governmental membership organization and the world's largest developer of voluntary International Standards.

NPES (http://www.npes.org) provides leadership in developing national and international standards for the printing industry. Their Committee for Graphic Arts Technologies Standards (CGATS) collaborated with AIIM in the development of the PDF/A Standard. For more information on NPES's standards activities, see http://www.npes.org/programs/standardsworkroom.aspx.

Open Preservation Forum (OPF) (http://www.openpreservation.org) sustains technology and knowledge for the long-term management of digital cultural heritage by providing reliable solutions to the challenges of digital preservation.

PDF Association (http://www.pdf.com) was founded as the PDF/A Competence Center, and exists to promote the adoption and implementation of International Standards for PDF technology.

White Papers and Websites Focused on PDF/A:

ADLIB Software Company, *PDF/Archive – Portable Document Format/Archive White Paper,* http://www.adlibsoftware.com/~/media/Files/Whitepapers/WhitePaper-PDFA-for-Archiving.pdf (last access 27/07/17)

British Library, 2015, *PDF Format Preservation Assessment*, version 1.2, http://wiki.dpconline.org/images/5/51/PDF_Assessment_v1.2_external.pdf (last access 27/07/17)

Johnson, Duff, *Talking PDF – a blog by Duff Johnson*, last access 2/10/ 2015, https://web.archive.org/web/20161202204353/https://talkingpdf.org/

NDSA Standards and Practices Working Group, *The Benefits and Risks of the PDF/A-3 File Format for Archival Institutions*, 2014, last access 27/07/17, http://www.digitalpreservation.gov/documents/NDSA_PDF_A3_report_final022014.pdf

PDFlib GmbH, *The PDF/A Family of Archiving Standards*, last access 27/07/17, http://www.pdflib.com/knowledge-base/pdfa/

United States Library of Congress, Digital Preservation (http://www.digitalpreservation.gov) is a United States website containing a lot of digital preservation materials and education. Last access 27/07/17

Vasilescu, Ramona, 'Tibiscus' University, Timişoara, *PDF/A standard for long term archiving*, *Annals. Computer Science Series 7th Tome 1st Fasc.,* 2009, last access 27/07/17 (http://arxiv.org/pdf/0906.0867.pdf)

11.References

Adobe Systems Inc., *How to Remove PDF/A Information from a file*, 2011, last access 27/07/17, http://blogs.adobe.com/acrolaw/2011/05/how-to-remove-pdfa-information-from-a-file/

Adobe Systems Inc., *Adobe Developer Connection/Adobe XMP Developer Center*, viewed 2 October 2015, http://www.adobe.com/devnet/xmp.html

Adobe Systems Inc., *Extensible Metadata Platform (XMP)*, last access 27/07/17, http://www.adobe.com/products/xmp.html

AIIM, 2014, *Information Chaos V Information Opportunity*, last access 27/07/17, http://www.aiim.org/Research-and-Publications/Research/AIIM-White-Papers/Information-Chaos-V-Information-Opportunity

AIIM, *ISO 19005-1 (PDF/A-1) Application Note*s, 2006, viewed 2 October 2015, https://web.archive.org/web/20160804082733/http://www.aiim.org/documents/standards/PDF-A/ISO19005AppNotes.pdf

AIIM, 2014, *Paper Wars 2014 – Update from the Battlefield*, last access 27/07/17, http://www.aiim.org/Research-and-Publications/Research/Industry-Watch/Paper-Wars-2014

AIIM, *PDF/A: Frequently Asked Questions (FAQs)*, viewed 2 October 2015, https://web.archive.org/web/20120926021935/http://www.aiim.org/documents/standards/19005-1_FAQ.pdf

Association for Digital Document Standards (ADDS), 2009, *PDF Up to Date: Long Term Archiving with PDF,* Germany: Association for Digital Document Standards (ADDS).

Ball, Alex 2013*, Preserving Computer-Aided Design (CAD), DPC Technology Watch Report 13–02*, York: Digital Preservation Coalition, last access 27/07/17, http://dx.doi.org/10.7207/twr13-02

Bavaria report from PDFlib, viewed 9 May 2016, https://www.pdflib.com/fileadmin/pdflib/pdf/pdfa/2009-05-04-Bavaria-report-on-PDFA-validation-accuracy.pdf

Becker, C, Kulovits H, Guttenbrunner M, Strodl S, Rauber A and Hofman H 2009, 'Systematic planning for digital preservation: evaluating potential strategies and building preservation plans, *International Journal of Digital Libraries,* 10: 133, DOI:10.1007/s00799-009-0057-1

British Library, 2015, *PDF Format Preservation Assessment*, version 1.2 25/02/2015, last access 27/07/17, http://wiki.dpconline.org/images/5/51/PDF_Assessment_v1.2_external.pdf

Buckley, R 2008, *JPEG2000: A Practical Digital Preservation Standard?, DPC Technology Watch Report 08-01*, York: Digital Preservation Coalition, last access 27/07/17, http://www.dpconline.org/docman/technology-watch-reports/87-jpeg-2000-a-practical-digital-preservation-standard/file

Cornell University, 2003, *Moving Theory into Practice: Digital Imaging Tutorial*, last access 27/07/17, https://www.library.cornell.edu/preservation/tutorial/preservation/preservation-01.html

DPC (Digital Preservation Coalition), 2017. *Digital Preservation Handbook*, 2nd edition, last access 27/07/17, http://handbook.dpconline.org/

Drümmer, Olaf; Oettler, Alexandra; and vol Seggern, Dietrich, 2007, *PDF/A in a Nutshell: Long-term Archiving with PDF*, Germany: Association for Digital Document Standards (ADDS).

Evans, Tim, 2015, 'Preserving PDF at the coalface: PDF/A at the Archaeology Data Service', Presentation to the DPC Forum Preserving Documents Forever: When is a PDF not a PDF? 15/07/2015, last access 27/07/17, http://www.dpconline.org/component/docman/doc_download/1420-pdf-oxford-15071

Evans, Tim and Moore, Ray 2014, 'The Use of PDF/A in Digital Archives: A Case Study from Archaeology', *The International Journal of Digital Curation,* 9:2, 123–38.

Fanning, Betsy, 2008, *Preserving the Data Explosion: Using PDF*, DPC Technology Watch Report 08-02, York: Digital Preservation Coalition, last access 27/07/17,

http://www.dpconline.org/component/docman/doc_download/86-preserving-the-data-explosion-using-pdf

Fresko, Marc, 2013, 'PDF/A-3: Do its Benefits for "Compound" Records Outweigh its Drawbacks?', a presentation to the DPC meeting Digital Preservation with Portable Documents: a workshop to introduce and discuss the PDF/A version, 13 March 2013, last access 27/07/17, http://www.dpconline.org/component/docman/doc_download/829-pdfa3dpcfresko

Gartner, Richard and Lavoie, Brian, 2013, *Preservation Metadata (2nd edition), DPC Technology Watch Report 13-3*, York: Digital Preservation Coalition, last access 27/07/17, http://dx.doi.org/10.7207/twr13-03

Higgins, Sarah, 2015, 'An Introduction to PDF', a presentation to the DPC briefing day Preserving Documents Forever: When is a PDF not a PDF? 15 July 2015, last access 27/07/17, http://www.dpconline.org/docs/miscellaneous/events/1418-pdf-oxford-150715-an-introduction-to-pdf/file

ISO, 2004, ISO Guide 2: 2004, Standardization and related activities – General vocabulary, https://www.iso.org/obp/ui/#iso:std:39976:en

ISO, 2001, ISO 15489-1:2001, *Information and documentation – Records management – Part 1: General*, viewed 2 October 2014, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=31908

ISO, 2005, ISO 19005-1:2005, *Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)*, viewed 2 October 2015, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=38920

ISO, 2007, ISO 19005-1:2005/Cor 1:2007, viewed 2 October 2015, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=45613

ISO, 2011, ISO 19005-1:2005/Cor 2:2011, viewed 2 October 2015, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=60603

ISO, 2011, ISO 19005-2:2011, *Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2*), viewed 2 October 2015, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=50655

ISO, 2012, ISO 19005-3:2012, *Document management – Electronic document file format for long-term preservation – Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3*), viewed 2 October 2015, http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57229

ISO/IEC, 2005, Guide 21-1:2005, Regional or national adoption of International Standards and other International Deliverables – Part 1: Adoption of International Standards, http://www.iso.org/iso/catalogue_detail.htm?csnumber=39799

Isartor from the PDF Association, last access 27/07/17, http://www.pdfa.org/2011/08/download-isartor-test-suite/

Koninklijke Bibliotheek, 2007, *Recommendations for the creation of PDF files for long-term preservation and access*, last access 27/07/17, https://www.kb.nl/sites/default/files/docs/pdf_guidelines.pdf

Koo, Jamin and Chou, Carol C, 2012, PDF to PDF/A: Evaluation of Converter Software for Implementation in Digital Repository Workflow, iPRES Proceedings of the 9th International Conference on Preservation of Digital Objects, Oct 1–5, 2012 pp 302–303, last access 27/07/17, https://ipres-conference.org/ipres12/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf

May, Peter, Pennock, Maureen, Wheatley, Paul and Whibley, Simon, 2015, *PDF Format Preservation Assessment*, last access 27/07/17, http://wiki.dpconline.org/images/e/e8/PDF_Assessment_v1.3.pdf

Mitcham, Jenny, 2013, *Some Thoughts on PDF/A 3*, last access 27/07/17, http://digital-archiving.blogspot.co.uk/2013/03/some-thoughts-on-pdf-version-3.html

Morrissey, Sheila, 2012, 'The Network is the Format: PDF and the Long-term Use of Digital Content', *Archiving 2012*, pp. 200-203 http://www.portico.org/digital-preservation/wp-content/uploads/2012/12/Archiving2012TheNetworkIsTheFormat.pdf

The National Archives, *Digital Preservation Guidance Note 1: Selecting File Formats for Long-Term Preservation,* last access 27/07/17, http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf

NDSA Standards and Practices Working Group,2014, *The Benefits and Risks of the PDF/A-3 File Format for Archival Institutions*, last access 27/07/17, http://www.digitalpreservation.gov/documents/NDSA_PDF_A3_report_final022014.pdf

Ockerbloom, John Mark, 2001, *Archiving and Preserving PDF Files*, University of Pennsylvania, last access 27/07/17, http://repository.upenn.edu/cgi/viewcontent.cgi?article=1054&context=library_papers

Oettler, Alexandra, 2013, *PDF/A in a Nutshell 2.0: PDF for Long-term Archiving*. Germany: Association for Digital Document Standards (ADDS).

Open Preservation Foundation (OPF), veraPDF (PDF/A validation), last access 27/07/17, http://www.openpreservation.org/about/projects/verapdf/

PDF/E Joint Working Group, 2008, *Frequently Asked Questions (FAQs)* ISO 24517-1:2008 PDF/E-1, last access 27/07/17, https://web.archive.org/web/20131007215735/http:/www.aiim.org:80/documents/standards/PDF-E/PDF_E_FAQ-Edits_Jan.pdf

PDFlib, Bavaria Report on PDF Validation Accuracy, viewed 14 September 2016, https://www.pdflib.com/fileadmin/pdflib/pdf/pdfa/2009-05-04-Bavaria-report-on-PDFA-validation-accuracy.pdf

Toptechtune, 2012, Pros and Cons of Keeping Files in PDF Format, last access 27/07/17, http://www.toptechtune.com/2012/04/pros-and-cons-of-keeping-files-in-PDF-format.html

UK Web Archive (UKWA), launched in 2004, https://www.webarchive.org.uk/ukwa/

United States Library of Congress, *Sustainability of Digital Formats – Planning for Library of Congress Collections*, last access 27/07/17, http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml

van der Knijff, Johan, 2015, *Understanding PDF Risks in Preservation – Why PDF/A validation matters (even if you don't have PDF/A*), a presentation to the DPC Forum Preserving Documents Forever: When is a PDF not a PDF? 15/07/2015, last access 27/07/17, http://www.dpconline.org/component/docman/doc_download/1421-pdf-oxford-15071

veraPDF*, Definitive PDF/A Validation*, viewed 2 October 2015, http://www.verapdf.org

Wikipedia, 2016, 'Extensible Metadata Platform', last access 27/07/17, https://en.wikipedia.org/wiki/Extensible_Metadata_Platform

12.Appendix: Standards and Technical Guides

The following Standards are of importance when using PDF/A for preservation.

ISO 14289-1: 2014, Document management applications – Electronic document file format enhancement for accessibility – Part 1: Use of ISO 32000-1 (PDF/UA-1)

This International Standard specifies how to use ISO 32000-1 to produce accessible electronic documents. It does not specify processes for converting paper or electronic documents to the PDF/UA format; technical design, user interface, implementation, or operational details of rendering; physical methods of storing these documents, such as media and storage conditions; or the computer hardware and/or operating systems that are required.

This Standard was developed primarily for use by software developers to develop the software that end users can use to create compliant files. With more assistive technologies in use, this Standard, when used to create a compliant file, will enable the assistive technologies to be able to read the document to the person. It emphasizes the use of tagging to ensure the technology will correctly read the document.

ISO 15489, Information and documentation – Records management

This establishes the concepts and principles from which approaches to the creation, capture and management of records should be developed. These concepts and principles address:

a) the nature of records, metadata for records and records systems;
b) the roles of policy, assigned responsibilities, and monitoring and training in supporting effective management of records;
c) the importance of ongoing analysis of business context and identification of records requirements;
d) the nature and uses of records controls;
e) the purposes of records processes.

The Standard applies to the creation, capture and management of records regardless of structure or form, in all types of business and technological environments, over time.

There are two parts to it: Part 1 deals with the concepts and principles while Part 2 is a technical report providing the methodology and examples of how the Standard is implemented.

This Standard is important to the implementation of PDF/A as it explains how to establish a good records management process. As established earlier in this report, it is not enough to implement PDF/A to ensure the long-term preservation of your digital documents, you must establish a records management program.

ISO 15801, Document management – Information stored electronically – Recommendations for trustworthiness and reliability

This Standard describes the implementation and operation of document management systems that can be considered to store electronic information in a trustworthy and reliable manner. It is for use by organizations that use a document management system to store authentic, reliable and usable/readable electronic information over time.

ISO 16684, Graphic technology – Extensible metadata platform (XMP) specification

This Standard consists of two parts: Part 1 covers the data model, serialization and core properties, while Part 2 provides a description of XMP schemas using RELAX NG.

The Standard defines the essential components of XMP metadata.

ISO/TR 17797:2014, Electronic archiving – Selection of digital storage media for long-term preservation

This Standard gives guidelines on the selection of the most appropriate storage media for use in long-term electronic storage solutions. It includes a discussion on magnetic, optical, and electronic storage.

ISO/TR 18492, Long-term preservation of electronic document-based information

This technical report provides practical guidance for the long-term preservation and retrieval of authentic electronic document-based information, when the retention period exceeds the expected life of the technology (hardware and software) used to create and maintain that information.

ISO 18829:2017, Document management – Assessing trusted systems for compliance with industry standards and best practices

This Standard identifies activities and operations an organization should perform to evaluate whether the electronically stored information is maintained in reliable and trustworthy environments. It should be used by organizations evaluating the trustworthiness of existing content/record/document management systems.

ISO 19005-1: 2005, Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)

This International Standard specifies how to use the Portable Document Format (PDF) 1.4 for long term preservation of electronic documents. It is applicable to documents containing combinations of character, raster, and vector data.

It does not specify processes for converting paper or electronic documents to the PDF/A format; specific technical design, user interface, implementation, or operational details of rendering; specific physical methods of storing the preserved document, such as media to use or storage conditions; or required computer hardware and/or operating systems. The Standard is for developers to develop software for end users to use to preserve their files.

In 2007, a technical corrigendum, a corrected version of the Standard, was produced to share corrections concerning the mapping of PDF's subject and keywords to XMP.

ISO 19005-2:2011, Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2)

The second part of the International Standard, ISO 19005, specifies the use of Portable Document Format (PDF) 1.7, as formalized in ISO 32000-1, for preserving the static visual representation of page-based electronic documents over time.

This is the first of the PDF/A Standards to be based on an ISO Standard and not a *de facto* specification like the first part.

ISO 19005-3: 2012, Document management – Electronic document file format for long-term preservation – Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)

This part of the PDF/A (ISO 19005) family of Standards specifies the use of the Portable Document Format (PDF) 1.7, as formalized in ISO 32000-1, for preserving the static visual representation of page-based electronic documents over time, in addition to allowing any type of other content to be included as an embedded file or attachment.

The industry recognized a need to be able to ensure authenticity, and to preserve not only the document but the underlying source for the document, and specifically XML. The Working Group took that requirement into consideration and added the capability to add the native file (XML) to an archived document without losing the ability to archive the original file.

ISO 24517-1:2008, Document management – Engineering document format using PDF – Part 1: Use of PDF 1.6 (PDF/E-1)

This Standard specifies the use of the Portable Document Format (PDF) Version 1.6 for the creation of documents used in workflows.

It defines the document file format for documents that are used in engineering workflows for purposes such as construction or manufacturing workflows. These documents will typically contain 2D drawings, dynamic information and numerous approval signatures.

ISO/DIS 24517-2, Document management – Engineering document format using PDF – Part 2: Use of ISO 32000-2 including support for long-term preservation (PDF/E-2)

This Standard specifies the use of the Portable Document Format (PDF) 2.0, as formalized in ISO 32000-2, for the creation and preservation of documents used in engineering workflows. Part 2 adds support for 3D, in addition to the 2D models defined in Part 1.

Engineering documents can be complex and have unique characteristics, given their dynamic nature. With these anomalies noted, the Working Group determined they needed to set the specific requirements for preservation of engineering documents. (See also the Digital Preservation Coalition-published *Technology Watch Report 13-02, Preserving Computer-Aided Design (CAD)*. Reports may be downloaded at http://www.dpconline.org/publications/technology-watch-reports.

ISO 32000-1:2008, Document management – Portable Document Format – Part 1: PDF 1.7

This specifies a digital form for representing electronic documents to enable users to exchange and view electronic documents independent of the environment in which they were created or the environment in which they are viewed or printed. It is intended for the developers of software that creates PDF files (conforming writers), software that reads existing PDF files and interprets their contents for display and interaction (conforming readers) and PDF products that read and/or write PDF files for a variety of other purposes (conforming products).

The Standard does not specify processes for converting paper or electronic documents to the PDF format; technical design, user interface or implementation or operational details of rendering; physical methods of storing PDF documents, such as media and storage conditions; methods for validating the conformance of PDF files or readers; or required computer hardware and/or operating systems.

ISO 32000-2:2017, Document management – Portable Document Format – Part 2: PDF 2.0

This is a revision of ISO 32000-1 that adds new features to the PDF file.

PDF/UA-1 Technical Implementation Guide: Understanding ISO 14289-1 (PDF/UA-1)

This guide is intended to provide implementers with information and examples to further explain ISO 14289-1 beyond the text of the Standard.

PDF/UA-1 Technical Implementation Guide: Understanding ISO 32000-1 (IPDF 1.7)

Tagging the content and content or document structure is important in order to provide an accessible file, and is only one of many features in ISO 32000-1. This guide is intended to provide information and examples to illuminate and clarify aspects of Section 14 of ISO 32000-1 for implementers of ISO 14289-1 (PDF/UA). This document is informative and does not state any requirements.

Achieving WCAG 2.0 with PDF/UA

This guide describes how WCAG 2.0 can be applied to other digital content technologies, including technologies that are not for web content. It provides an interpretation of how to determine the application of WCAG 2.0 Principles, Guidelines and Success Criteria.

The guide describes the alignment of WCAG 2.0 and ISO 14289-1:2012 (PDF/UA), the International Standard for accessible PDF technology. PDF software developers can achieve conformance with applicable WCAG 2.0 Success Criteria via implementations that follow this mapping to PDF/UA. The mapping provided in this document shows how to validate, in PDF file format terms, a PDF/UA document against WCAG 2.0.