

# PDF/A for RFCs

Leonard Rosenthol  
PDF Architect, Adobe Systems  
Project Leader, ISO 19005



# What Is Needed for Archiving?

- A document format that
  - Can convey critical information
  - Can be rendered accurately - "consistent and predictable"
  - Rich metadata support
    - Standard schemas such as DublinCore
    - Custom schemas
    - Document providence, version history, audit trails
  - Can incorporate associated "marginalia"
    - Notes scrawled in the margins on documents
    - Comments & markup
  - Digital Signatures for "tamper-proofing"
- A definition of retrieval devices (readers) behavior

**Abschrift**

Militärpolizeilicher Dienst      Wien, 1.10.1948  
10. Aufsertigung

**R u n d s c h r e i b e n    Nr. 31/48**

1. Die Alliierten Untersuchungskommissionen haben bisher festgestellt, dass in folgenden Konzentrationslagern keine Menschen mit Giftgas getötet wurden:  
Bergen-Belsen, Buchenwald, Dachau, Flossenbürg, Gross-Rosen, Mauthausen und Nebenlager, Natzweiler, Neuengamme, Niederhagen (Wevelsburg), Ravensbrück, Sachsenhausen, Stuthof, Theresienstadt.  
In diesen Fällen konnte nachgewiesen werden, dass Geständnisse durch Folterungen erpresst wurden und Zeugenaussagen falsch waren.  
Dies ist bei den KV-Erhebungen und Einvernahmen zu berücksichtigen.  
Ehemalige KZ-Häftlinge, welche bei Einvernahmen Angaben über die Ermordung von Menschen, insbesondere von Juden, mit Giftgas in diesen KZ machen, ist dieses Untersuchungsergebnis zur Kenntnis zu bringen. Sollten sie weiter auf ihre Aussagen bestehen, ist die Anzeige wegen falscher Zeugenaussage zu erstatten.

2. Im RS 35/48 kann P. 1 gestrichen werden.

RECEIVED  
MILITARY POLICE  
NOV 11 1948

Der Leiter des MPD.:  
Müller, Major

Für die Richtigkeit  
der Ausfertigung:  
Lachout, Leutnant

L.S.

P.d.R.d.A.:  
Republik Österreich  
Waffenamt Wien  
Kommando

Ich bestätige hiemit, dass ich am 1. Oktober 1948 als Angehöriger des militärpolizeilichen Dienstes beim Alliierten Militärkommando die Richtigkeit der Rundschreiben-Ausfertigung gemäss § 18 Abs. 4 AVC beglaubigt habe.

120

Wien, 27. Okt. 1948

bestätigt:

und aus 1/2 Dogen bestehen:  
1/2 gegenmilt. zur Zahl  
anzahl                  angezeig-  
mit                  gegeben

Bezirksgericht Favoriten  
1101 Wien, Angalgasse 35  
Wien, am 27. Okt. 1948

RECEIVED  
MILITARY POLICE  
OCT 27 1948

120 120

RECEIVED  
MILITARY POLICE  
OCT 27 1948

# The result -> PDF/A (ISO 19005)

- Long term preservation of black and white and color compound documents as electronic data
  - Combinations of character, raster, vector and other data
  - Provisions for capturing semantic information
  - Preservation and retrieval of appropriate metadata
- "static electronic paper"
  - Plus
    - Annotations & Marginalia
    - Attachments
    - Metadata
    - Digital Signatures

# Core PDF/A Features

- Restrictive “coding” of PDF details
  - Ensures less ambiguity when implementing
- Reliable and consistent color
- All data must be self-contained
  - No external resources
  - Fonts MUST be embedded and meet strict technical requirements
- Metadata based on Adobe XMP
- Digital Signatures compliant with EU legislation (PAdES – ETSI 102778)
- No Security or Encryption
- No interactive content (eg. video, audio or 3D)
- No JavaScript

# PDF/A and the Open Web

- PDF is an essential part of the Open Web Platform
  - It's an **open international standard** (ISO 19005)
  - There are **thousands of independent implementations**
    - **Open Source:** iText, Ghostscript, FOP, pdfTeX, OpenOffice/LibreOffice, Scribus, etc.
    - **Commercial:** Adobe, Microsoft, Corel, Nuance, FoxIt, NitroPDF, etc.
  - It's **used on the web today** (and has been for many years)
  - PDF/A is **mandated by over 50 countries** around the world

# PDF/A + XML = A Powerful Combination

- PDF/A provides native integration with XML in various ways
  - XML as attachment
    - Similar to an email message or ZIP file, files can be embedded into a PDF
  - XML and PDF Content living along side each other
    - Tagged or Structured Content
      - Similar to HTML where content is organized by a series of "tags" that identify the semantic nature of the content (eg. header, paragraph, table, etc.)
    - Associated Files
      - Enables association of XML directly with a piece of PDF content.
        - For example, the MathML representation of an equation or the SVG representation of a graphic.

- Xml2rfc currently supports the creation of PDF using two methods
  - Text->Postscript->PDF
  - XML->HTML->PDF
- The Postscript method
  - Relies on Ghostscript to convert the Postscript to PDF
  - Current versions of Ghostscript support generation of PDF/A
    - <http://svn.ghostscript.com/ghostscript/trunk/gs/doc/Ps2pdf.htm#PDF/A>
- The HTML method
  - HTML is produced via an XSLT on the XML
  - Relies on wkhtmltopdf to convert HTML to PDF
    - Wkhtmltopdf relies on the Qt libraries for PDF generation
    - No support for PDF/A at this time nor currently announced
  - OPTION 1: Could print to Postscript and then go the GS route
  - OPTION 2: Replace HTML->PDF component with PDF/A supporting alternative

# Wrap Up

- PDF/A provides an excellent foundation for the long-term preservation of “static” content consisting of text, vector and raster images.
- PDF/A is a recognized and/or mandated open standard by over 50 countries around the world including numerous national archives.
- PDF/A is not the only solution, but part of the solution.
  - PDF + XML provides the best of both worlds
  - Human + machine readability in a single package that can be digitally signed to guarantee authenticity.





고맙습니다 谢谢 תודה!  
*mahalo* *děkuji*

**Thank You**

شكرا köszönöm gracias  
Ευχαριστώ merci  
どうもありがとう danke

