

Surveying ISO Standards for PDF: archive, accessibility, engineering, metadata, 3D data and PDF itself. History, pain points, solutions and the dream.

Duff Johnson
PDF Association
Neue Kantstrasse 14
14057 Berlin, Germany
+1 617 283 4226
duff.johnson@pdfa.org

ABSTRACT

This workshop provides high-level information on the features, development and current status of ISO standards for PDF relevant to archiving. It then provides an opportunity for participants to clearly identify and prioritize related requirements and solutions.

Presented by the ISO Project Leader of ISO 32000, an independent consultant, the workshop will begin by reviewing the features and history of the technology and its associated standards. Thereafter, a set of breakout sessions will challenge participants to clearly specify their ideas and concerns about PDF and suggest a variety of solutions leveraging commercial software interests.

Finally, workshop participants will come together to discuss the breakout session results, identify common themes and build consensus on specific, actionable requests for PDF producer and processor developers in the short and medium term.

General Terms

strategic environment, preservation strategies and workflows, specialist content types, digital preservation marketplace

Keywords

PDF, PDF/A, PDF/E, PDF/UA, XMP, PRC, software, standards

1. OUTLINE

Following a series of short presentations introducing ISO-standardized PDF technology, the balance of time is occupied by breakout sessions followed by a group discussion.

1.1 Presentations (80 minutes)

This segment begins with a review of the procedures to be followed in the workshop together with a brief review of the proposed breakout session questions.

Next the workshop will hear a set of short presentations surveying PDF features, history and standards development, with ample time for questions. The segment closes with a group discussion in which three topics are selected for the breakout sessions.

The presentation segment covers:

- 1) workshop organization, release form;
- 2) more than rendering: PDF's value proposition & feature-set

iPres 2014 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

- 3) ISO Standards for PDF: File format standards of interest to digital preservation professionals, including:
 - a. ISO 32000 parts 1 and 2 (PDF 1.7 and PDF 2.0);
 - b. PDF/A parts 1, 2 and 3;
 - c. PDF/E parts 1 and 2;
 - d. PDF/UA;
 - e. XMP (PDF's XML-based metadata standard);
 - f. PRC (PDF's ISO-standardized 3D file-format).
- 4) PDF history: from Adobe to ISO to PDF 2.0;
- 5) the PDF future: standards development and the marketplace;
- 6) review some options, then create the breakout session topics.

1.2 Break (10 minutes)

A 10 minute break.

1.3 Breakout sessions (90 minutes)

This segment provides a means for workshop participants to collaborate on sketching a common understanding of key pain points related to some controversial, technical, conceptual or other participant-chosen topic specific to PDF technology.

Just before the break the workshop participants selected three breakout session topics. Following the break participants choose and join one of the three breakout sessions. After 30 minutes the groups come back together and present each breakout session's results to the rest of the workshop participants for review.

In addition to reviewing and characterizing these concerns, participants will be asked to propose specific actions they'd like to see industry take to address them.

The objective of each breakout session may be stated thusly:

- 1) 5 minutes for online research or just focused thinking;
- 2) 1 minute to select a minute-taker and facilitator
- 3) 24 minutes to discuss and achieve consensus on:
 - a. **Up to 3 key pain points or big opportunities** identified, characterized, ranked and rated for significance to digital preservationists;
 - b. **3 conceivable market-driven solutions** (conservative, plausible, shoot-the-moon) that would address one (or more) pain points;
 - c. **3 key reasons to be pessimistic or optimistic** about each proposed solution. If there's time, address both perspectives.

Following discussion, and after the workshop ends, the breakout session results will be anonymized, and distributed to participants and disseminated within the industry.

1.3.1 Breakout session topics

With a target of 3-4 breakout sessions, each session's topic is decided by a vote of all workshop participants. During the break, participants choose which session to attend.

Centered on PDF technology, sessions may be entirely technical, practical, theoretical, marketplace-specific, or otherwise. The presenter will visit each breakout session to provide feedback, quick answers (if available) or other assistance.

1.3.2 Breakout topic examples

- PDF/A-3: The context, challenge and opportunity of embedded non-archival formats
- The roles and responsibilities of PDF/A processors
- Should there be a PDF/A-4? What should it look like?
- Are we missing any conformance levels from PDF/A?
- Archiving dynamic data: is PDF/E the right way?
- Assume PDF 2.0 is great, with no ambiguities... what should PDF 3.0 look like? Or should we not bother?
- Why TIFF when there's PDF?
- Why PDF when there's EPUB? Or vice-versa
- What's PDF worth when Acrobat includes text-editing
- The PDF icon Adobe products install is actually the Adobe Acrobat icon. No wonder users still call PDF files "Adobe"! Does it matter? Discuss
- How could conformance level "a" become useful?
- "Enough with this theory and standards: 80% of my problems with PDF are due to X. Let's talk about X!"
- What should a PDF/A validator do, exactly? Just syntax, or rendering too? Certification? Does PDF validation (as opposed to PDF/A) matter to you, and why?
- Is PDF a nearly-perfect set of features for digital preservation of textual content? What's it missing?

2. INTENDED AUDIENCE

Archives policy-makers and practitioners, records-managers, ECM implementers.

3. EXPECTED LEARNING OUTCOMES

Tutorial attendees will gain authoritative high-level knowledge of the archival-relevant PDF standards and respective conformance levels. They will be invited to experience and participate in a product-management exercise conducted from the industry perspective. Participants will gain deeper understanding of the commercial view of PDF technology from exposure to industry and standards-development orientations. They will learn how to access industry-sponsored activities and resources and participate in PDF standards development and best-practices efforts.

4. PRESENTER'S BIOGRAPHY

Duff Johnson began his career with PDF in 1996 when he founded an electronic document service bureau dedicated to PDF technology-based solutions. Having managed three companies in the PDF technology space, he is now an independent consultant.

Johnson began working in PDF standards development in 2005 as chairman of the AIIM committee developing PDF/UA, now ISO 14289. Elected Project Leader for ISO 32000 in 2011, Johnson chairs the committee developing PDF 2.0, the next generation of the world's chosen fixed-layout electronic document format.

Johnson serves as Vice Chairman of the PDF Association (www.pdfa.org), the vendor organization for the worldwide PDF software industry. He is currently serving as Standards Committee Chairman and as a member of AIIM's Board of Directors.