Validated PDF

It’s Time to Get Canonical

Prepared for
the PDF Association

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Background

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Statements of Fact

- Today, approximately 80% of non-HTML documents on the public web are in PDF; the world's chosen electronic document format for final-form content
- PDF remains the dominant exchangeable electronic medium for 2D design and print renderings
- There is no likely replacement for general-purpose final-form electronic documents on the visible horizon
- Reliability and consistency are the core of PDF's value proposition
- Today, few ECM / EDM / ERM systems leverage PDF; almost all still treat PDF as functionally equivalent to TIFF

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

However...

- PDF isn’t perceived to be as reliable or as universally supported as one might hope for a truly generic electronic document format
- Adobe is stuck fixing files made by sloppy vendors while 3rd party vendors are judged against Adobe software instead of the specification. This mess might be OK for other, less universal, less significant, less relied-on file-formats - but it’s not OK here
- The perception that Adobe continues to “own” PDF retards 3rd party investment (e.g. ECM vendors) in PDF technology, and thus, retards integration opportunities and cost-savings
- It ought to be push-button easy to move records from agency X to NARA; it isn’t, but we can get there from here

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

About PDF/A…

PDF/A-1 is based on PDF 1.4, a proprietary Adobe specification

PDF/A-2 and PDF/A-3 are based on ISO 32000, an open specification

If scanned pages are all you care about PDF/A-1 is reasonable. Modern “born digital” content often requires PDF/A-2 for transparency, JPEG 2000 and other features of ISO 32000-1

PDF/A sets rules for using PDF. While PDF/A itself may be validated there is no generally accepted means of validating PDF file syntax, objects or rendering besides Adobe Reader

So, how do we know we’ve got a good PDF? After all…

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

It’s a jungle out there

PDF Association End User Survey

34% say they personally encounter bad files or think they are common.

31% think that bad PDF files cause “significant” business problems.

38% say that differences in PDF software cause problems.

(This survey of electronic document professionals was conducted over March-April 2013 with 66 responses, more information is available on request)

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

The Problem, The Solution

- The PDF specification has changed over time; not all PDF files are created equal

- The PDF specification isn’t always straightforward; not all software is equally capable with respect to a given feature

- Invalid, incapable or marginal files and software cause errors and data-loss, triggering business-process failures every day

- **Validation** detects problems, enabling corrective action
  Validation fosters software that meets ISO specifications

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

It’s Time

- At 20 years old PDF is a mature technology; it’s the implementations that are spotty
- Variations in file quality and software capability continue to retard EDMS solutions, slowing ROI for current and planned implementations while limiting over-the-horizon opportunities
- PDF/A-1 and/or PDF/A-2 are the wrong way to resolve the fundamental reliability and functionality concerns; they aren’t substitutes for full-scale PDF validation
- All newer PDF subset standards are based on ISO 32000-1, not 1.4

- Canonical syntax and object validation are prerequisites of canonical rendering validation
- ISO 32000-2 won’t be a free download as is PDF 1.0 - ISO 32000-1! Adoption will improve dramatically with a free canonical validator
- To produce an ISO 32000-2 validator in a timely fashion it’s probably essential to produce an ISO 32000-1 validator first
- Governments may soon require procured software to conform with relevant ISO standards in a readily verifiable manner

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

It’s Time

- At 20 years old PDF is a mature technology; it’s the implementations that are spotty
- Variations in file quality and software capability continue to retard EDMS solutions, slowing ROI for current and planned implementations while limiting over-the-horizon opportunities
- PDF/A-1 and/or PDF/A-2 are the wrong way to resolve the fundamental reliability and functionality concerns; they aren’t substitutes for full-scale PDF validation
- All newer PDF subset standards are based on ISO 32000-1, not 1.4
- Canonical syntax and object validation are prerequisites of canonical rendering validation
- ISO 32000-2 won’t be a free download as is PDF 1.0 - ISO 32000-1! Adoption will improve dramatically with a free canonical validator
- To produce an ISO 32000-2 validator in a timely fashion it’s probably essential to produce an ISO 32000-1 validator first
- Governments may soon require procured software to conform with relevant ISO standards in a readily verifiable manner

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

The ISO 32000 Validator

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Executive Summary

- A software project driven by professional product and technical management to design, develop, produce, promote and support the canonical ISO 32000 validator, eventually under an appropriate open source license (e.g., Apache)
- Free web-based and downloadable PDF validation software with the capacity (via end-user opt-in) to deliver data about files processed to interested vendors
- A project ownership, management and technical model designed to facilitate rapid industry-wide acceptance

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Usage

Developers
Use the source code, APIs and opt-in data from end users via the Validator SaaS to support software development and reduce customer support costs

End users
Use free web-based or downloadable software to test files

Procurement authorities
Develop policies referencing conforming PDF files and software

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Who

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

The PDF Association (PDFa.org)

- A vendor-neutral entity that will not feel forced to protect its own products while serving the industry as a credible project driver
- Prior experience in developing PDF standards compliance models (Isartor Test Suite, Matterhorn Protocol)
- The existing working relationship and Category A liaison status between the PDF Association and TC 171 WG 8
- Adobe's newly-released 'formal representation' for its Dictionary Validation Agent (DVA) needs a third-party "home" the ISO 32000 Committee cannot provide but the PDF Association can

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Benefits

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

New Options to Reduce Support Costs

- Customer support organizations gain new options for automated and semi-automated means for addressing incoming requests
- 3rd party validation may be integrated into end user support workflows (“Step 1: Test your file at VeraPDF.org…”)
- End users are more easily re-directed to the real source of their problems; problems are more readily solved overall
- “Early warnings” from end-users performing opt-in tests on the public-access SaaS and via downloadable software
- Solutions for customer problems become easier and faster to produce; more data provides potential for insights to help drive software enhancements

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Benefits for Developers

- An immediate independent feedback mechanism to drive software QA
- Helps move developer time from addressing problems to product enhancement
- Provides credible independent confirmation that an unreadable file was invalid
- Provides new ways to trim support loads and encourage customer self-support
- Eventually, separates real-world PDF validation from Adobe Reader
- Creates new possibilities for products that create, check or transcode validated PDF
- Encourages customer adoption, ensuring the future of the technology
- Energizes open source development to resolve specific pain-points (such as fonts) as well as objects and resources defined elsewhere (such as images, XMP, video, 3D, etc)

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Benefits for End Users

- Reduced costs due to improved quality and reliability
- Improved interoperability and functionality in PDF software
- Deeper implementation of PDF features in ECM systems
- Better support from vendors, better options for self-support
- Eventual wholesale transcoding to “validated PDF”
- A means of establishing key quality criteria in procurement

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Benefits of Sponsorship

VeraPDF sponsors may access a variety of branding options limited to project supporters, including:

- Use of “Validated PDF”, “VeraPDF” and other licensed graphics or content in software, on websites and elsewhere
- Representation on PDF Association, VeraPDF and SaaS sites as a contributing sponsor of the PDF Validator
- Feature placement or listing in VeraPDF and PDF Association press-releases and other communications

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

“VeraPDF”

A Service of the PDF Association

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

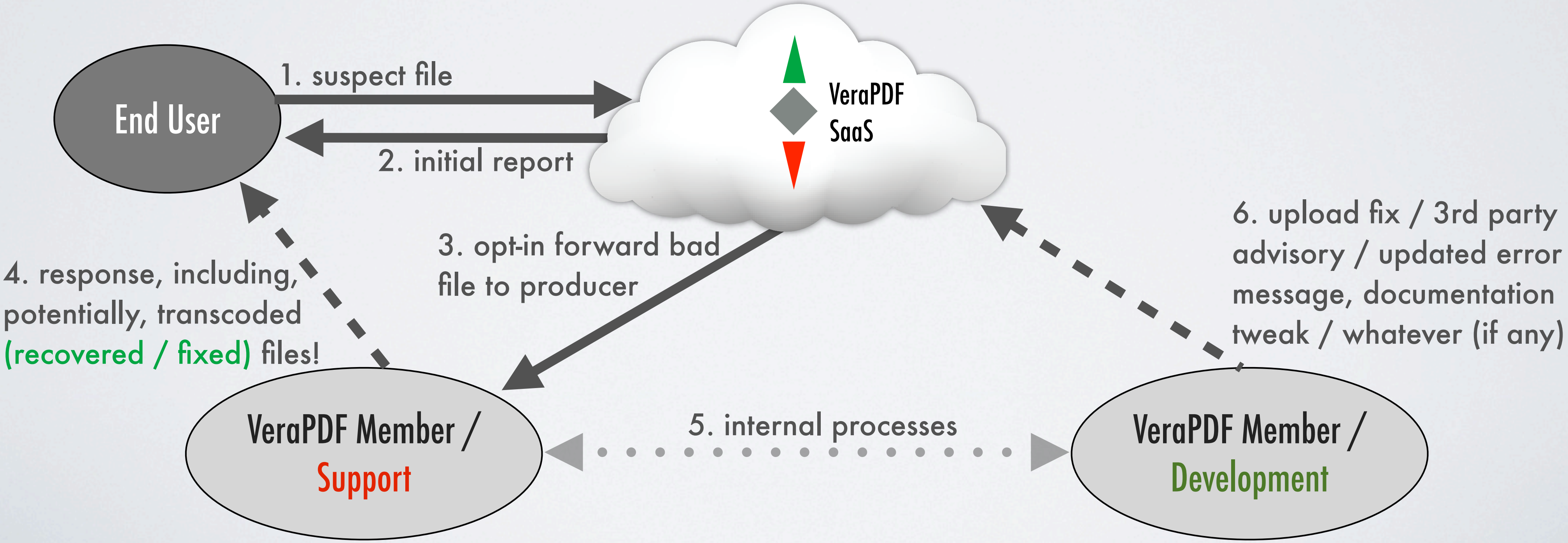
“VeraPDF”

- “VeraPDF” is the proposed name of a new non-profit membership organization dedicated to developing canonical open-source ISO 32000 validation
- VeraPDF membership is available to any interested party. PDF Association (which organizes and supervises VeraPDF) members receive a substantial discount on VeraPDF membership Association with VeraPDF
- Project mechanisms, policies, priorities and conflict resolution are managed by a Steering Committee elected by the VeraPDF membership
- A senior software engineer oversees development while a project manager provides member and sponsor support, coordination, communications, product management and marketing to drive adoption

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

VeraPDF SaaS:
Benefits for Developers & End Users

VeraPDF's SaaS can report error data, heuristics or actual files to members and serve as a programmatically-accessible clearing-house for information (e.g. about font encodings) that benefits all.



End User

1. suspect file

2. initial report

VeraPDF SaaS

3. opt-in forward bad file to producer

4. response, including, potentially, transcoded (recovered / fixed) files!

5. internal processes

6. upload fix / 3rd party advisory / updated error message, documentation tweak / whatever (if any)

VeraPDF Member / Support

VeraPDF Member / Development

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

Finances

Revenue

VeraPDF is funded by annual membership dues plus sponsorship fees. PDF Association members get a discount on VeraPDF membership equal to PDF Association annual dues for standard members

Expenses

Hosting, development and support infrastructure, SaaS implementation development, project management, marketing, promotional, legal (inc. code review), overhead, funding maintenance mode following active development

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

VeraPDF can only happen in stages

Each stage requires both member and sponsor (technical and promotional) participation to meet project objectives.

1. **Initialization** (10 members + $x0,000 in sponsorship)
2. **Infrastructure** ($y00,000 in sponsorship)
3. **Go from closed to open-source** (Steering Committee)
4. **Set maturity / utility targets** (Steering Committee)
5. **Launch SaaS** (30 members + $z00,000/year sponsorship)

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.

What you can do

- Let them know you care. Show up in Vienna (and generally, at TC 171 SC 2 WG 5, 6, 8 & 9 meetings to bang the table and tell the committees that NARA, LoC and USG in general want to know the quality of the pens and paper they write on, and that the customer plans to use the fact of ISO 32000 to make this request stick.

- Recognize that “full” PDF validation implies real validation of... font programs, JPEG, TIFF, etc, etc and eventually, rendering. ultimately, it’s a big job, but you know you want it... and PDF is the appropriate catalyst for industry. This is a highly leveraegable opportunity with great ROI; a classic example of ISO standards implementation

- Support the project - or support a better approach to the same objective - by creating appropriate rules and procurement recommendations for USG

- Project Sponsorship - a simple, effective and cost-effective way to say it all

Copyright © 2013 Duff Johnson Strategy & Communications, all rights reserved.