Beyond the Page

My story with PDF 2.0

Duff Johnson

Beyond the Page | My story with PDF 2.0

- Interoperability as a necessity
- Documents
- The ah-HA! moment
- End-user perspective
- Service-bureau perspective
- Software vendor perspective
- Creating a marketplace
- Standards development
- ISO 32000
- PDF 2.0
- What's different about PDF 2.0, and why the ownership thing matters
- What it's like at the standards table
- The next generation of PDF

Background

Why is a political consultant working with electronic document standards?

Interoperability first!

Born in the US

Grew up in the UK

Interoperability from first principles

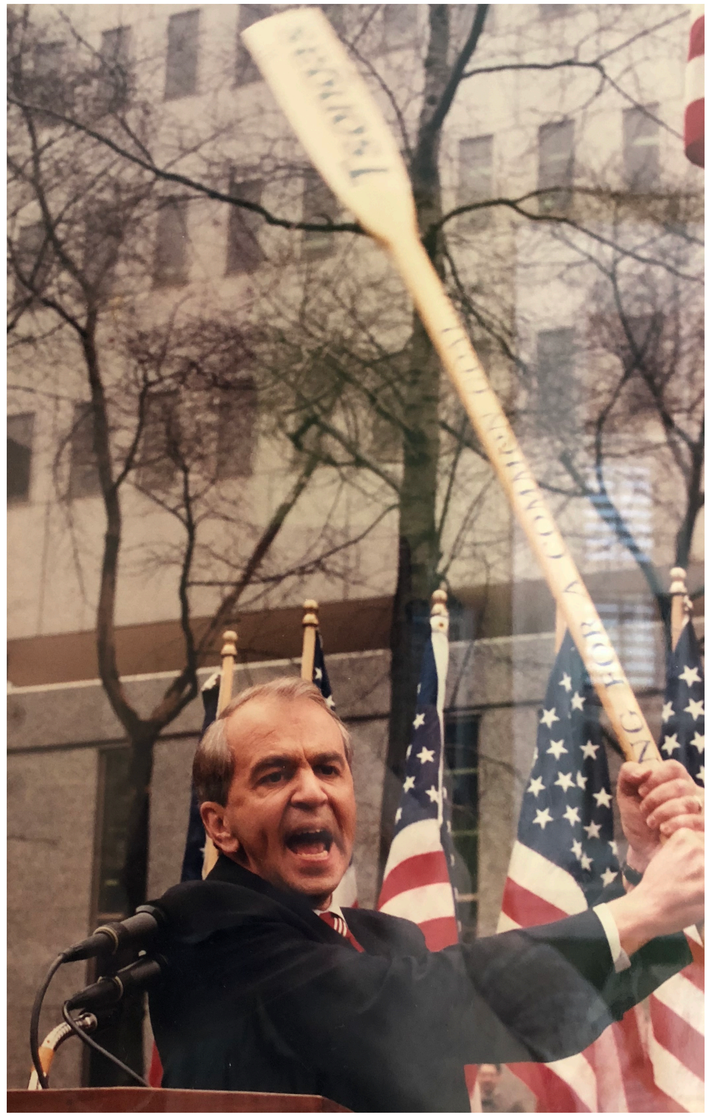
Interoperability from first principles

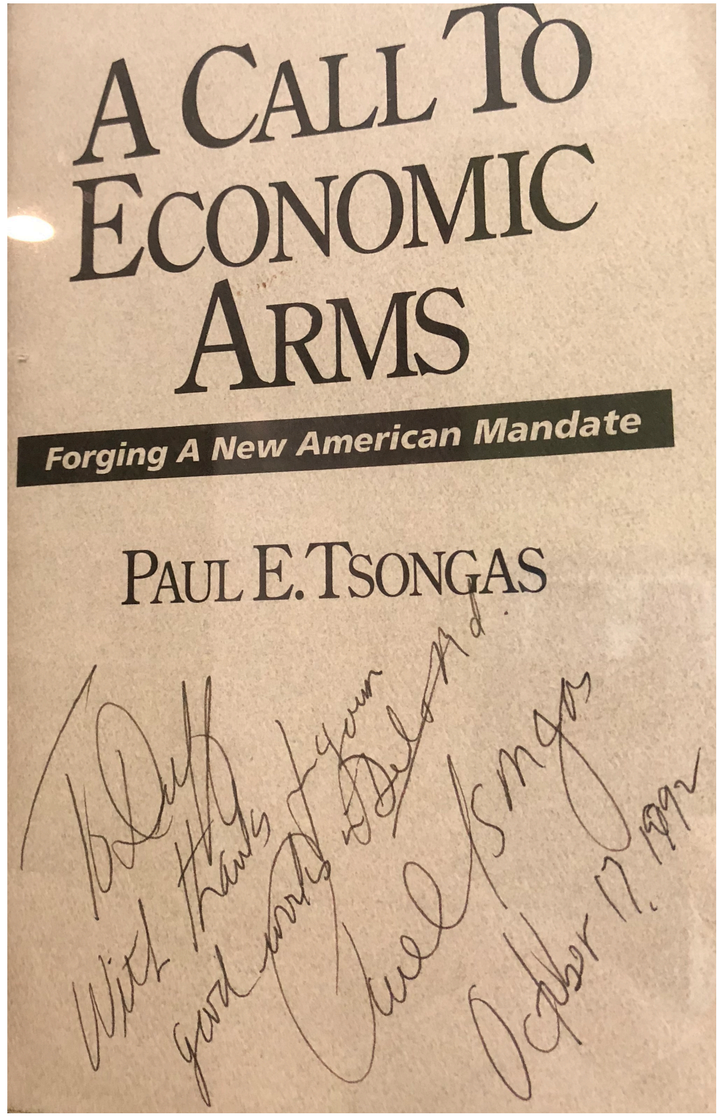
Trouble creator

Interoperability from first principles

A political consultant? Really?

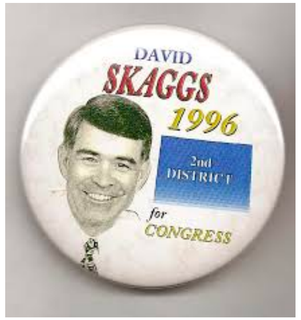
- Commercial fishing didn't work out for me
- In 1992 I started working for **Paul Tsongas for President**





A political consultant? Really?

- Tsongas lost, so….
- …naturally, I loved it, and worked on more campaigns!
- Colorado, Wisconsin, Virginia…
- Aruba, Antigua, Trinidad & Tobago…
- I got into **opposition research**
- I started to notice **documents**



DISTRICT 13 FOR MARSHALL AND EBERT, RACE IS ROUGH, COSTLY

Documents

- Drafting / collaborating
- Recording
- Sharing (communicating)
- Authenticating
- Referencing
- Deciding

Classic marketing imagery for “knowledge workers” - it’s all about documents

Ah-HA!

From politics to electronic document solutions

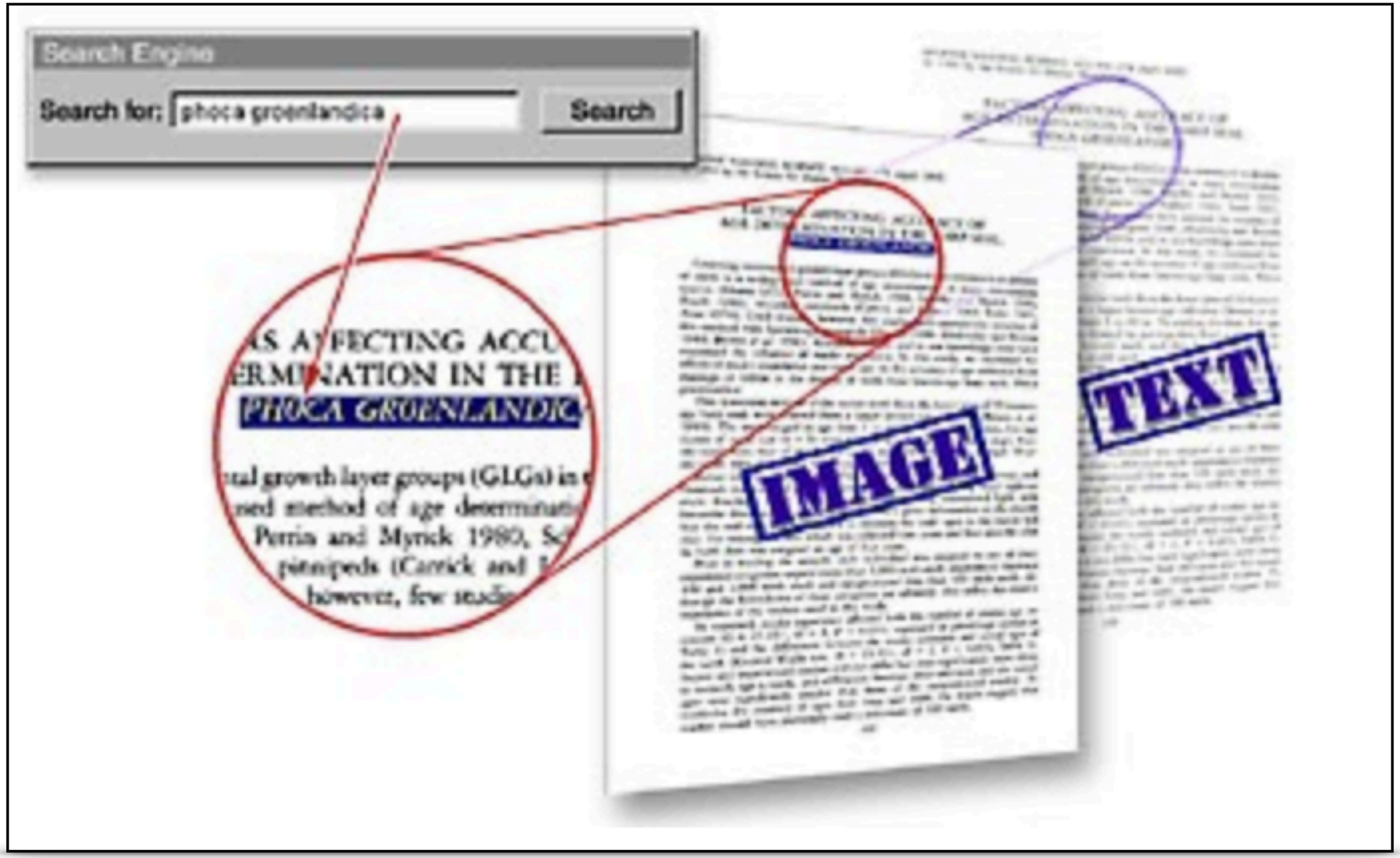
- Opposition research examines the record of candidates and opponents…
- Political researchers need effective means of analyzing collected documents…
- Campaigns need a means of packaging documents for efficient distribution…

It's obvious... right!

From politics to electronic document solutions

- Opposition research examines the record of candidates and opponents…
- Political researchers need effective means of analyzing collected documents…
- Campaigns need a means of packaging documents for efficient distribution…

It's obvious... right!


Search Engine
Search for: phoca groenlandica
Search
PHOCA GROENLANDIC
IMAGE
TEXT

INTERNET ARCHIVE
WayBackMachine

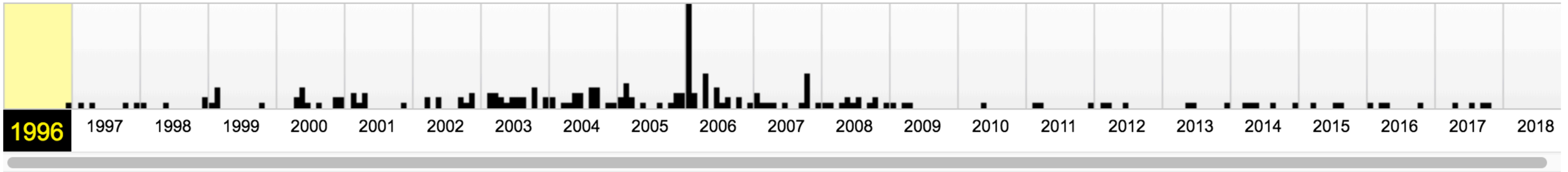
Explore more than 327 billion web pages saved over time

http://document-solutions.com:80/

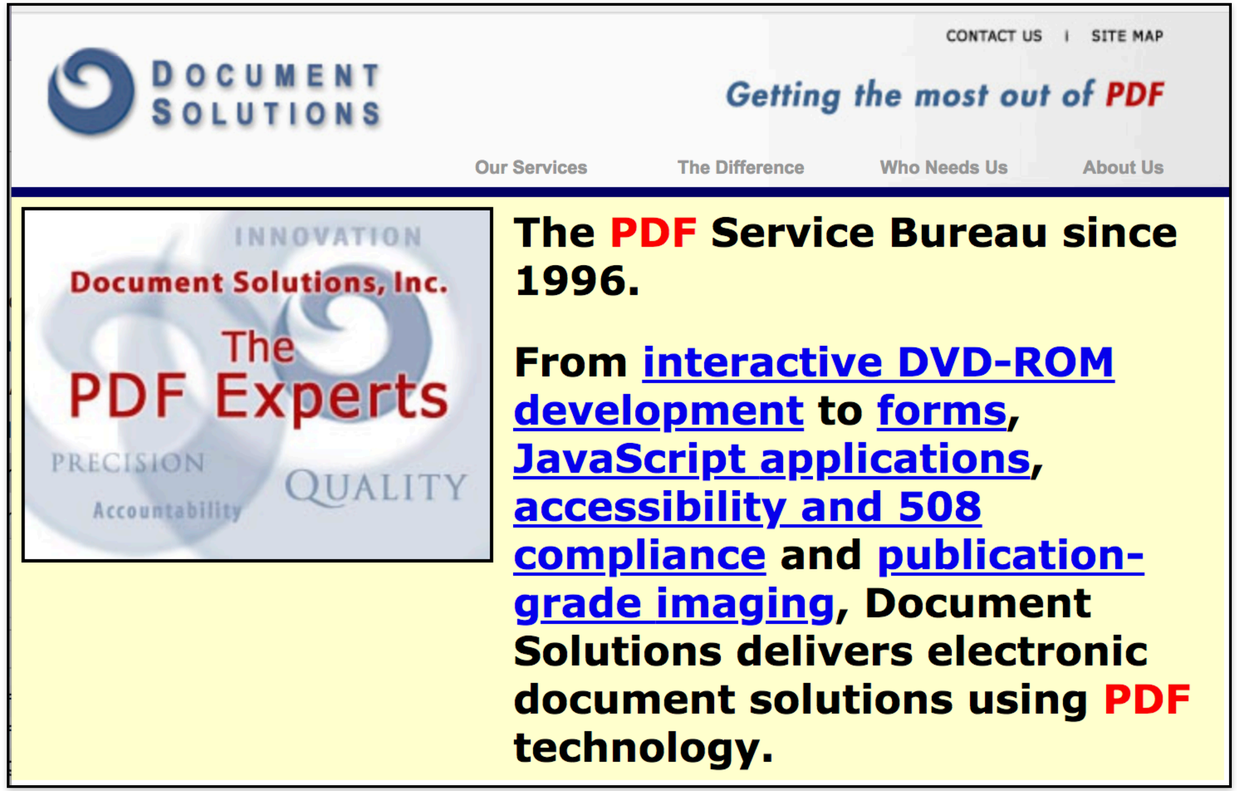
DONATE

Saved **202 times** between December 27, 1996 and October 31, 2017.

Summary of document-solutions.com · **Site Map of document-solutions.com**



Document Solutions, Inc. (1996 - 2008)



San Francisco Chronicle
NORTHERN CALIFORNIA'S LARGEST NEWSPAPER

MONDAY, SEPTEMBER 9, 1996

UNCONVENTIONAL TIDBITS: Here's some interesting trivia from the recent political conventions that you can use at the next cocktail party, courtesy of the political research folks over at Document Solutions in Oakland.

First the Republicans:

■ Only one Republican uttered the words "**Newt Gingrich**" in a speech down m San Diego — California Treasurer **Matt Fong.**

■ "Abortion" got a total of two mentions.

■ "Tolerance" was used once — in the context of "zero tolerance for drugs."

■ "Contract with America" was not mentioned at all.

■ "**Ronald Reagan**" was invoked no fewer than 44 times.

As for the Democrats:

■ "Universal health care" was cited only once.

■ "Assault weapons ban" figured 24 times.

■ "Deficit," 67 times.

■ Balanced budget was cited only 9 times.

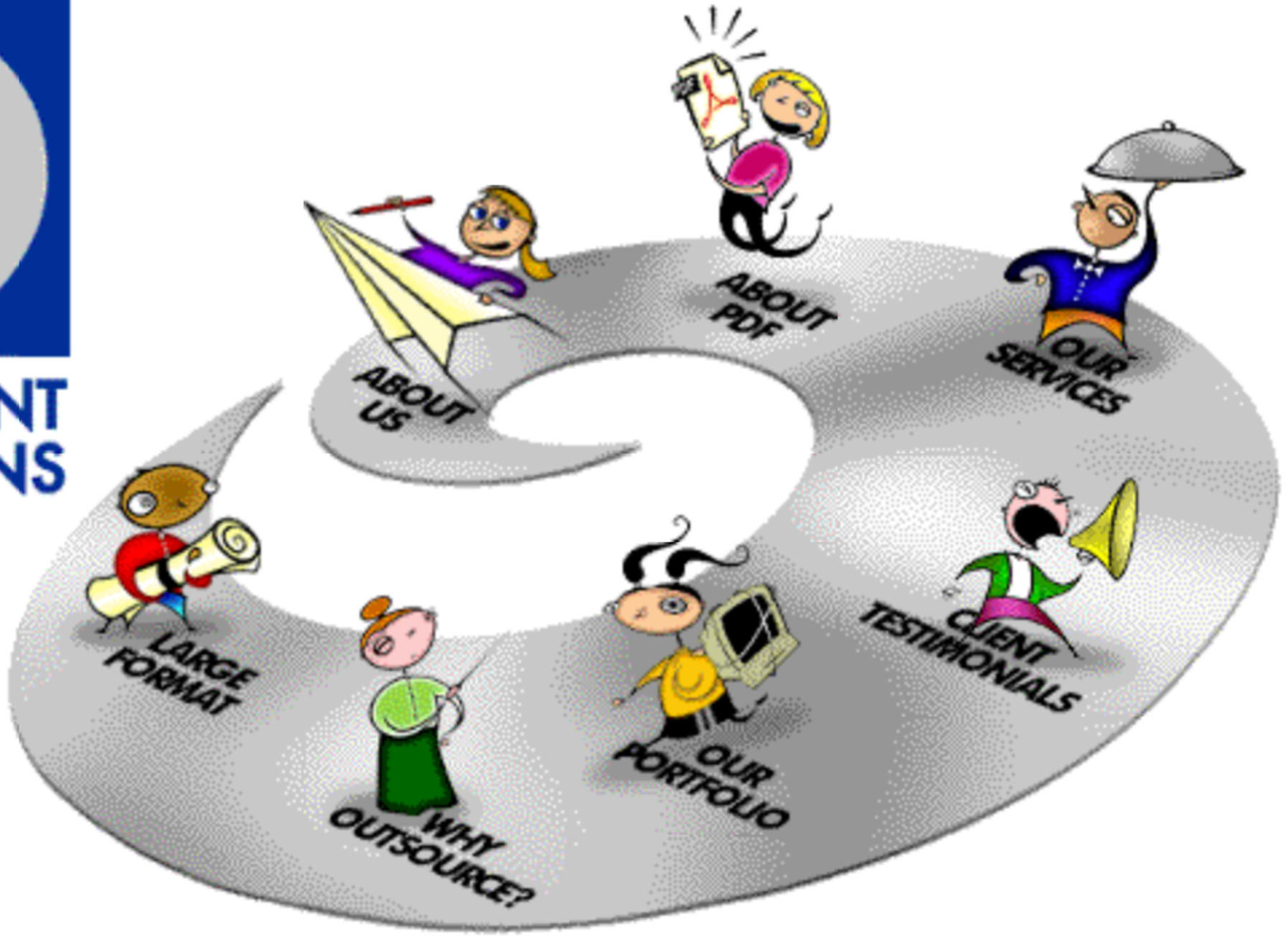
■ "Newt Gingrich," on the other hand, was invoked 105 times.

MATIER & ROSS

Chronicle columnists Phillip Matier and Andrew Ross appear Monday, Wednesday and Friday. They can be heard on Monday and Wednesday at 6.50 a.m. with Frank Dill on KNBR-AM (680) radio. Their e-mail address is MatierandRoss@sfgate.com.

PDF association Duff Johnson, Executive Director

DOCUMENT SOLUTIONS



From 1996 - “Scanning to PDF”

CONTACT US | SITE MAP

DOCUMENT SOLUTIONS

Getting the most out of PDF

Our Services | The Difference | Who Needs Us | About Us

PDF Scanning Services: "Publications Our Specialty"

Document Solutions provides state of the art publication-grade scanning services to magazine, newsletter, book and periodical publishers, printers, graphic designers, libraries, web site content managers, marketing departments, technical writers, government agencies and many others.

The common thread with all our scanning clients are ultra-strict quality requirements, advanced PDF conversion services and other special needs.

Through our new partnership with Strategic Office Solutions, DSI is able to offer our imaging services even more cost-effectively than ever before.

“Searchable Image PDF”

CONTACT US | SITE MAP

DOCUMENT SOLUTIONS

Getting the most out of PDF

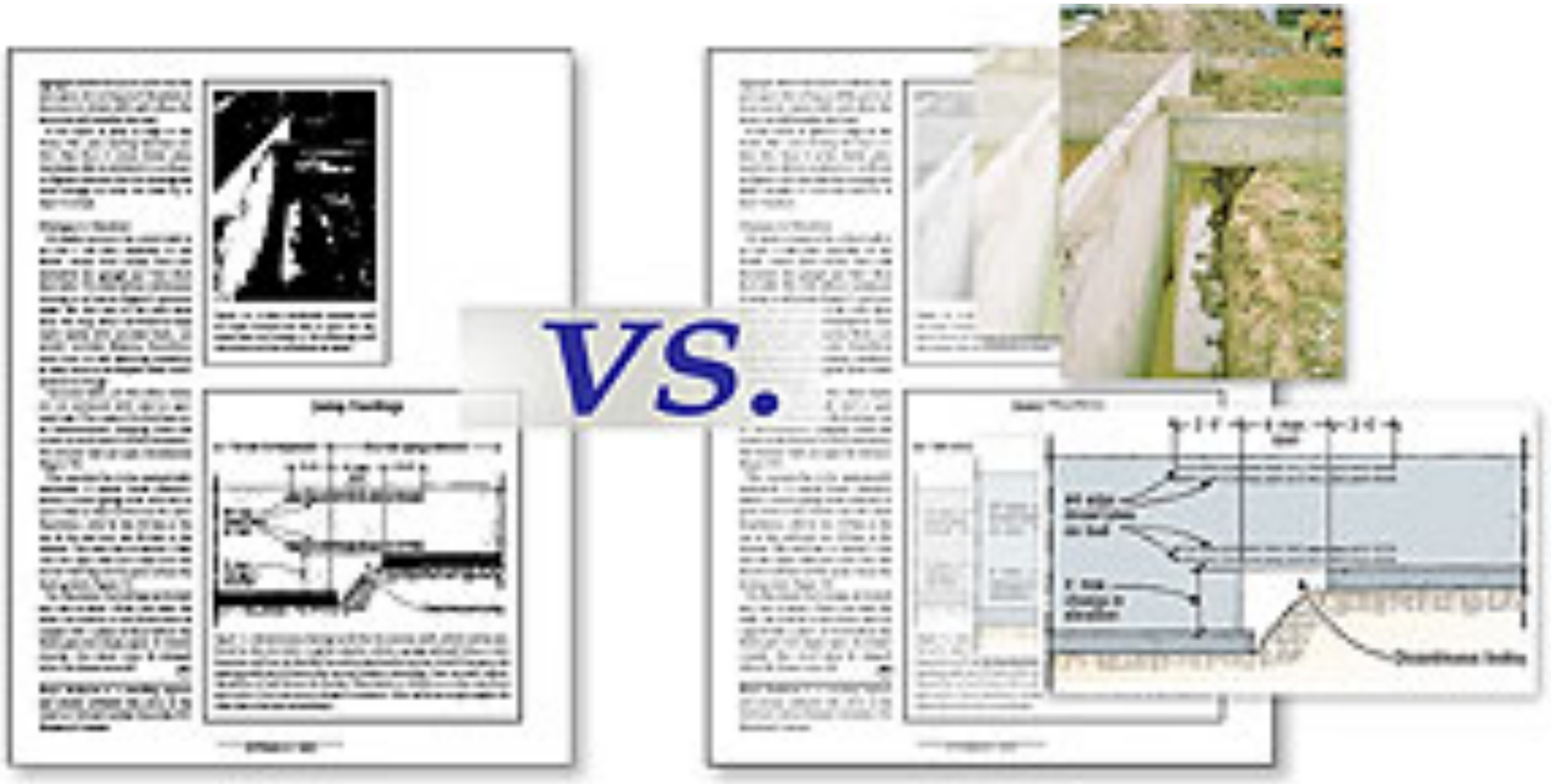
Our Services | The Difference | Who Needs Us | About Us

Searchable Image

Text resulting from an OCR (Optical Character Recognition) process may be "bonded"to the originating image to create a PDF/Searchable Image file. When you search for words or phrases, they will be highlighted in the image.

"Ex Machina"* OCR processing depends on the quality of your originals and other factors, but this "raw" conversion is accurate enough for many applications.

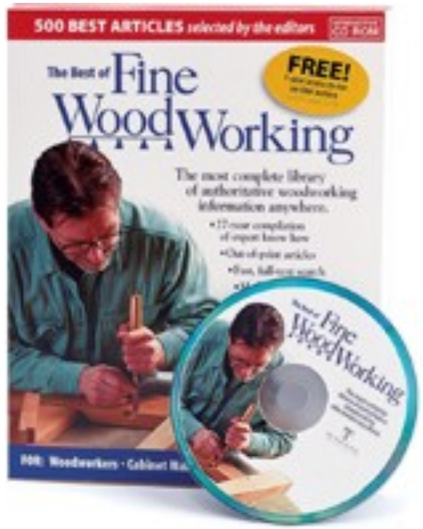
From 1997 - “MultiResolution PDF”

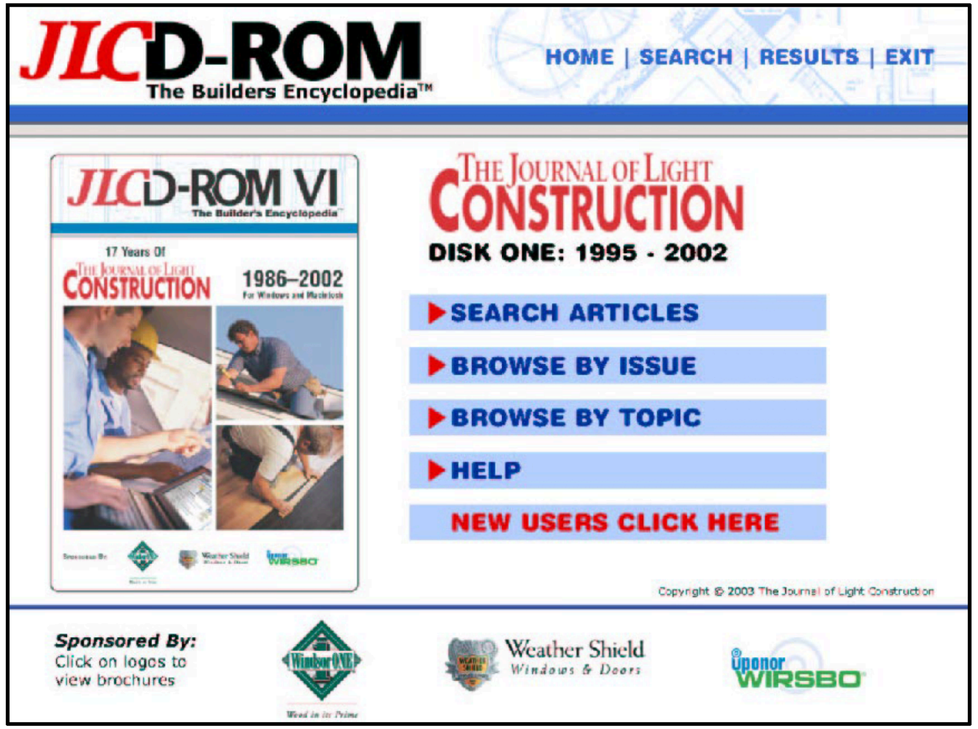


Significant imaging projects

- Scientific journal back-catalogs
- Trade magazine back-catalogs
- Product manuals and documentation for manufacturers
- Historical collections
- United Nations Yearbooks, United Nations ‘Law of the Sea’, Scientific American, and many others

From 1998 - disc-based collections





Writing about PDF

DOCUMENT SOLUTIONS

CONTACT US | SITE MAP

Getting the most out of PDF

Our Services | The Difference | Who Needs Us | About Us

PDF Online 101 - A Crash Course

The following guidelines for PDF authoring may not help you entirely avoid any problems, but the vast majority of your users and readers will be thankful nonetheless. Take good care of your PDFs when you make them, and everyone who ever uses the file will thank you for your consideration.

What is Tagged PDF?

PDF files can contain many things. At a minimum, a PDF contains the text, graphics, bookmarks, links and other elements of content that go to make up an electronic document.

DOCUMENT SOLUTIONS

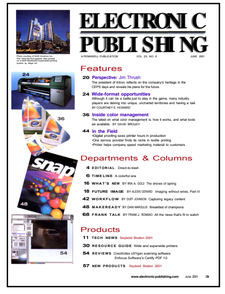
CONTACT US | SITE MAP

Getting the most out of PDF

Our Services | The Difference | Who Needs Us | About Us

Section 508 Policies for PDF content

It's not obvious how Section 508 applies to PDF content. What is for sure is that perceived accessibility and usability depend on making the document work well rather than work poorly, no matter how formally "compliant" it is. Here, we offer a few issues worthy of consideration when developing your Section 508 implementation policy.



Planet PDF classic

the home of the PDF community

Archive News Weblog Forum Archive

Find PDF Software | PDF Events | Newsletters | Free | About Planet PDF

REVIEW -- Adobe Acrobat 5.0: "Focus on the Document"

'The new 1.4 specification begins the long, slow transition to the complete XML enabling of PDF'

2 May 2001

By Duff Johnson, Planet PDF Contributing Editor and Susan Frank

Planet PDF

Planet PDF is brought to you by Debenu

Enterprise & Government | Creative & Print | Develop

Enterprise & Government News In-Depth Learning Center

Find PDF Software | PDF Events | Free | Newsletters | About Planet PDF

PDF IN-DEPTH Latest

Duff Johnson: First Impressions of Acrobat 7.0

By Dan Shea

Planet PDF

Planet PDF is brought to you by Debenu

Enterprise & Government

Enterprise & Government News In-D

Find PDF Software | PDF Events | Free | Newsletters | About Plan

PDF IN-DEPTH

PDF is green tech for your office

By Duff Johnson

Planet

Enterprise

Find PDF Software

PDF IN-DEPT

PDF on the iPhone: The first swipe

By Duff Johnson

G+

Duff Johnson is the CEO of Appligent Document Solutions

As the iPhone continues to grow in popularity, business applications for the iPhone OS are proliferating. Inevitably, a number of these applications focus on electronic documents, which is another way of saying (whether their developers recognize it or not) that they focus on PDF files.

ABOUT THE AUTHOR

Duff Johnson

Since 2011, Duff has worked in the ECM industry specializing in PDF technology since 1995. An industry leader. Duff co-chairs the international that manages ISO 320(specification for the PDI Responsible for marketi product... More

BEST OF Planet PDF

Talking PDF

AcrobatUsers.com

ADOBE ACROBAT USER COMMUNITY

Each PDF Page is a Painting

PDF association

Duff Johnson, Executive Director

Document Solutions, Inc. - service bureau perspective

- 1996 ~ The political research services idea was terrible…
- 1997 ~ …but services for electronic documents made sense!
- 1998 ~ Added PDF forms services
- 2000 ~ Added Tagged PDF services
- 2002 ~ Completed transition from volume imaging to PDF technology services
- 2007 ~ One more massive imaging project…

Service bureau perspective - lessons learned

- People really trust PDF
- Interoperability matters because stuff happens
- PDF matters because rendering matters
- Markup-language content is existentially distinct from documents
- PDF is almost perfect

Service bureau perspective - lessons learned

- OK, PDF isn’t perfect, but it’s not so badly broken that it can’t live forever

Service bureau perspective - lessons learned

- OK, PDF isn’t perfect, but it’s not so badly broken that it can’t live forever
- If we don’t screw it up

Software vendor

Now sign this NDA…

- 2009 - 2011 CEO of Appligent Document Solutions
- 2011 - 2013 President of NetCentric Technologies US

software vendor, lessons learned

- Software is hard
- PDF is really hard
- Software evolves, but documents are forever
- If you want to ship new features, you have to control support costs
- In documents, from paper size to PDF specification, standards enable business

Standards

- **2000** started experimenting with tagged PDF (beta program for Acrobat 5)
- **2001**, started offering “Section 508 compliance” services for PDF documents
- **2004**, became chair of AIIM’s committee to develop a best-practice guide for PDF accessibility, the project which became PDF/UA
- **2008**, present at the birth of ISO 32000-1
- **2011**, elected ISO Project Leader for ISO 32000
- **2012**, published ISO 14289-1 (PDF/UA-1)
- **2017**, published ISO 32000-2 (PDF 2.0)

Standards

- It only works if there’s community participation

Facts about PDF 2.0

- A community product
- Most significant feature: enhanced documentation
- Improves interoperability
- Reduces development and support costs
- Set out specific expectations for processing software
- Enhances PDF as a platform for business applications development

Five key features of next-generation PDF

Feature	Value
Ever-better interoperability	Reduced development and support costs Enhanced platform for new features
Transparency	A basis for future investment
Associated files	Deep and rich content capture
Document semantics	Rich options for content reuse
Authentication and encryption	Trust and privacy

Five key features of next-generation PDF

Feature	Value
Ever-better interoperability	Reduced development and support costs Enhanced platform for new features
Transparency	A basis for future investment
Associated files	Deep and rich content capture
Document semantics	Rich options for content reuse
Authentication and encryption	Trust and privacy

Technology-neutral

Counter-currents

PWP?

Web Publications for the Open Web Platform: Vision And Technical Challenges

W3C Interest Group Note 02 May 2017

This version:

https://www.w3.org/TR/2017/NOTE-pwp-20170502/

Latest published version:

https://www.w3.org/TR/pwp/

Latest editor's draft:

https://w3c.github.io/dpub-pwp/

Previous version:

https://www.w3.org/TR/2017/WD-pwp-20170309/

Web Publications for the Open Web Platform: Vision And Technical Challenges

“A book is a discrete collection of text (and other media), that is designed by an author(s) as an internally complete representation of an idea, or set of ideas; emotion or set of emotions; and transmitted to readers in various formats.”

“Books can learn from the Web how to be bounded, but open.
The Web can learn from books how to be open, but bounded.”

Abstract
Hugh McGuire, Medium, April 2016
https://www.w3.org/TR/pwp/#package

Web Publications for the Open Web Platform: Vision And Technical Challenges

W3C Interest Group Note 02 May 2017

This version:
https://www.w3.org/TR/2017/NOTE-pwp-20170502/

Latest published version:
https://www.w3.org/TR/pwp/

Latest editor's draft:
https://w3c.github.io/dpub-pwp/

Previous version:
https://www.w3.org/TR/2017/WD-pwp-20170309/

Is it easier?
Is is as capable?
Is it as flexible?
Is it as interoperable?

Web Publications for the Open Web Platform: Vision And Technical Challenges

W3C Interest Group Note 02 May 2017

This version:
https://www.w3.org/TR/2017/NOTE-pwp-20170502/

Latest published version:
https://www.w3.org/TR/pwp/

Latest editor's draft:
https://w3c.github.io/dpub-pwp/

Previous version:
https://www.w3.org/TR/2017/WD-pwp-20170309/

Is it easier?
Is is as capable?
Is it as flexible?
Is it as interoperable?
…as PDF?

Counter-counter-currents

- HTML is an experience, not a document
- The biggest platforms aren’t even using HTML anymore: Amazon, Facebook, Google and many other “web sites” are simply scripts with backend databases
- The idea of and need for documents persists; it’s the new stuff that ephemeral
- How do people ‘document’ the web? Screen-captures.

“Bounded but open” - and already accepted

PDFpackage – a lightweight content management platform

In a recent article I discussed **using PDF as a container** to organize, transport and archive collections of content. Since then I’ve had numerous discussions about this idea with members of the PDF technology and related communities. This article is an attempt to consolidate the substantial (and notably favorable) feedback. Written to stand alone, this article necessarily repeats several aspects of the **previous article**.

The basic facts

No-one owns PDF, but it’s accepted worldwide as the dominant end-state electronic document format.

The format includes rich capabilities in an ISO-standardized, vendor-independent specification.

PDFpackage - background

The “PDF Package” from a 2008 **post** in Adobe’s **Acrobat for Legal Professionals** blog.

- XMP metadata; onboard XML that may be associated with any PDF document, page, object or semantic structure element.
- ‘Embedded file streams’ and ‘associated files’ allow containment and characterization of arbitrary content (i.e., as is commonly found in email attachments)
- Rich interactive features to facilitate navigation, markup, annotations, JavaScript; and embedded content such as movies, 3D objects and MathML.
- Files generated from structured content are readily created with built-in navigation and accessibility
- Page-content may be reliably redacted to remove sensitive and / or GDPR-restricted content
- PDF files from “analog” workflows (e.g. scanned or faxed pages) are readily intermixed with PDF pages from other sources
- Encryption capabilities enabling new workflows
- Digital signature capabilities for tamper-proofing, authentication and more
- Facilitates rich experiences such as reflow and access to live data more commonly associated with HTML-based technologies

PDFpackage: PDF beyond electronic paper

- A standardized, vendor-neutral approach to retaining and future-proofing end-state user content such as email and case files
- A lightweight records and content-management technology platform enabling vertical solutions across every business sector
- A way to share rich, dynamic content experiences without giving up (or while leveraging) PDF's fundamental features
- A roadmap for affirmatively managing electronic content for authentication, long-term retention and security (inc. GDPR) purposes

PDFpackage: getting there

- Support PDF 2.0, especially PDFpackage-related features
- Develop industry-supported PDFpackage profiles describing appropriate usage in generalized applications (email archiving, case-folder management, etc.)
  - ZUGFeRD is an excellent example of a prototypical PDFpackage profile
- Stand back, and let the customer apply the content provisions or business rules they want

Communications technology over time

- Language | 100,000 – 200,000 BCE
- Painting | ~40,000 BCE
- Writing | ~3000 BCE
- Paper | ~200 BCE
- Printing press | 1440 CE
- Portable Document Format | 1993 CE

Communications technology over time

- Language | 100,000 – 200,000 BCE
- Painting | ~40,000 BCE
- Writing | ~3000 BCE
- Paper | ~200 BCE
- Printing press | 1440 CE
- Portable Document Format | 1993 CE
- **Up to you**

It’s a privilege to serve this industry

Thank you