PDF statistics –

the universe of electronic documents

Duff Johnson
Executive Director of the PDF Association

Duff Johnson,
Executive Director
PDF Association

Introduction

- PDF is:
  - Recognized worldwide
  - Broadly adopted
  - Accepted everywhere
  - Supported by a rich ecosystem
  - Etc…
- But what do we **know** about PDF adoption, utilization, etc.?

Duff Johnson,
Executive Director
PDF Association

Outline

- Context
  - Platform support
  - Web traffic
  - Competition
- Mindshare
  - “de facto”
  - Google Trends
- Hard data
  - On the web
  - Opened by Adobe software
- Guesstimates
  - On the web
  - Vs. other formats
  - Privately held
- Summary

Duff Johnson,
Executive Director
PDF Association

PDF platform support

Desktop

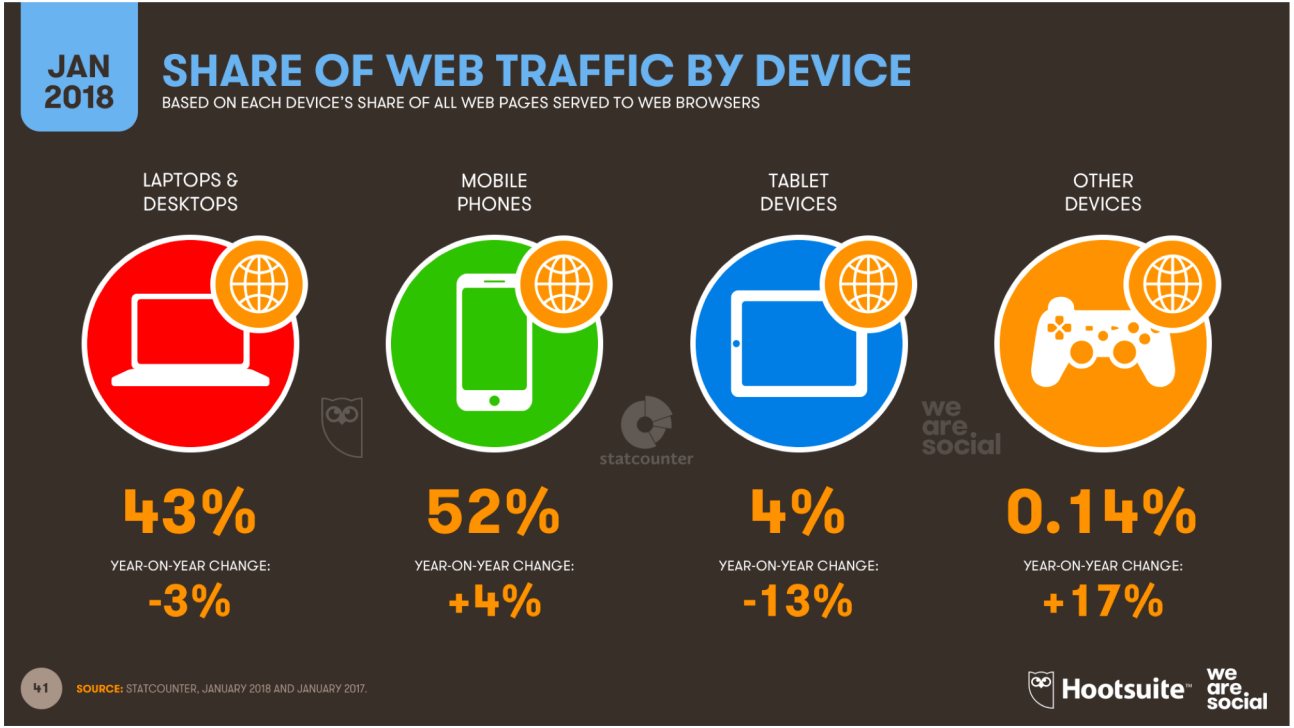
Browsers

Mobile *

Web

Duff Johnson,
Executive Director
PDF Association

Context: web traffic



2018

Source: wearesocial.com
https://wearesocial.com/blog/2018/01/global-digital-report-2018

Duff Johnson,
Executive Director
PDF Association

What would it take to compete with PDF?

- Self-contained (portable)
- Capable of being static (at least)
- Create from any source
- Generic (browser) support
- Easy-as-printing to create
- Published specification

Duff Johnson,
Executive Director
PDF Association

(Potentially) competitive technologies – 5 “families”

- Core W3C technologies (HTML / CSS / JavaScript)
- OOXML-based technologies (Microsoft universe)
- Other document formats (DjVu, EPUB, Pagis, etc.)
- Image formats (TIFF, PNG, JPEG)
- Non-electronic (Paper, fiche, film)

Duff Johnson,
Executive Director
PDF Association

Competitive technologies – 5 “families”

Type	Key benefits	Key failing
W3C	Dynamic, accepted	Ephemeral, limited workflows
OOXML	Editable, dynamic	Not fixed, M$ universe
Others	Varies	Poor acceptance
Images	Simple, easy	No text, poor feature-set
Paper / fiche	Eye-readable	Non-electronic

Duff Johnson,
Executive Director
PDF Association

Mindshare

Duff Johnson,
Executive Director
PDF Association

Examples of *de facto standards*

"…well-known and illustrative examples of *de facto* standards are… :

- HTML (computer file format) started as "*de facto*" (1993-1995) and became the "*de jure*" standard (1995–present day).
- PDF (computer file format) was first created in 1993 by Adobe…. PDF eventually became the *de facto* standard for printable documents. (…) In 2005, PDF/A became a *de jure* standard as ISO 19005-1:2005. In 2008 Adobe's PDF 1.7 became ISO 32000-1:2008."

https://en.wikipedia.org/wiki/De_facto_standard

Duff Johnson,
Executive Director
PDF Association

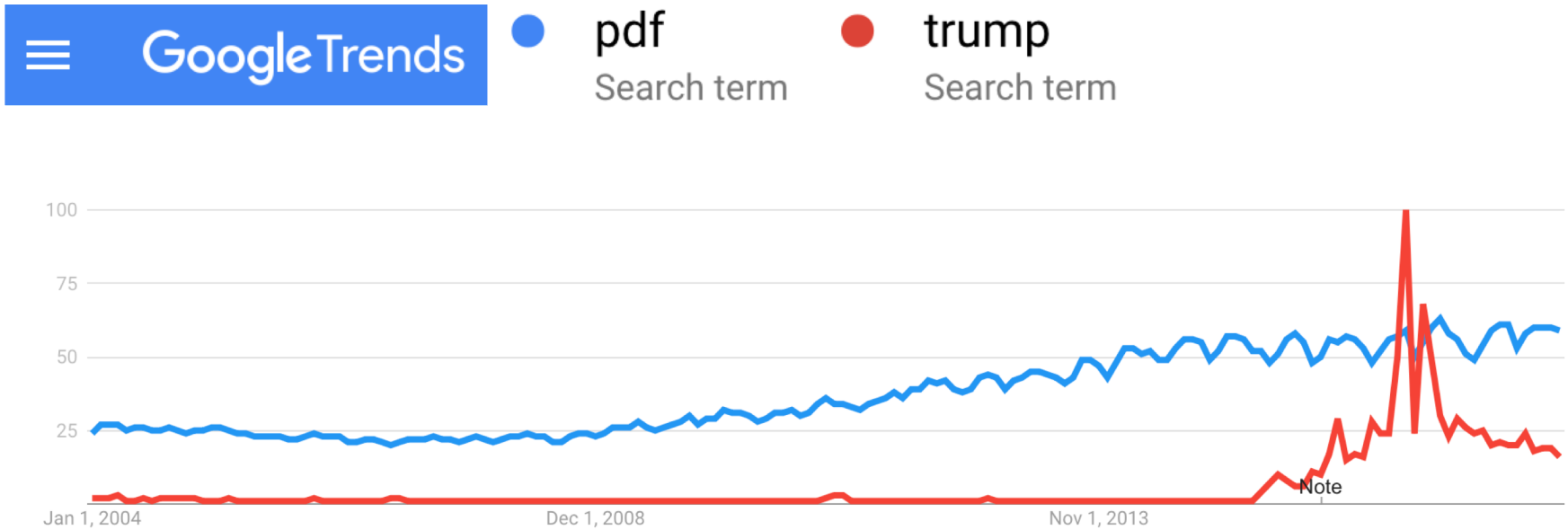
Over time, more people think about PDF more often

www.pdfa.org

Duff Johnson,
Executive Director
PDF Association

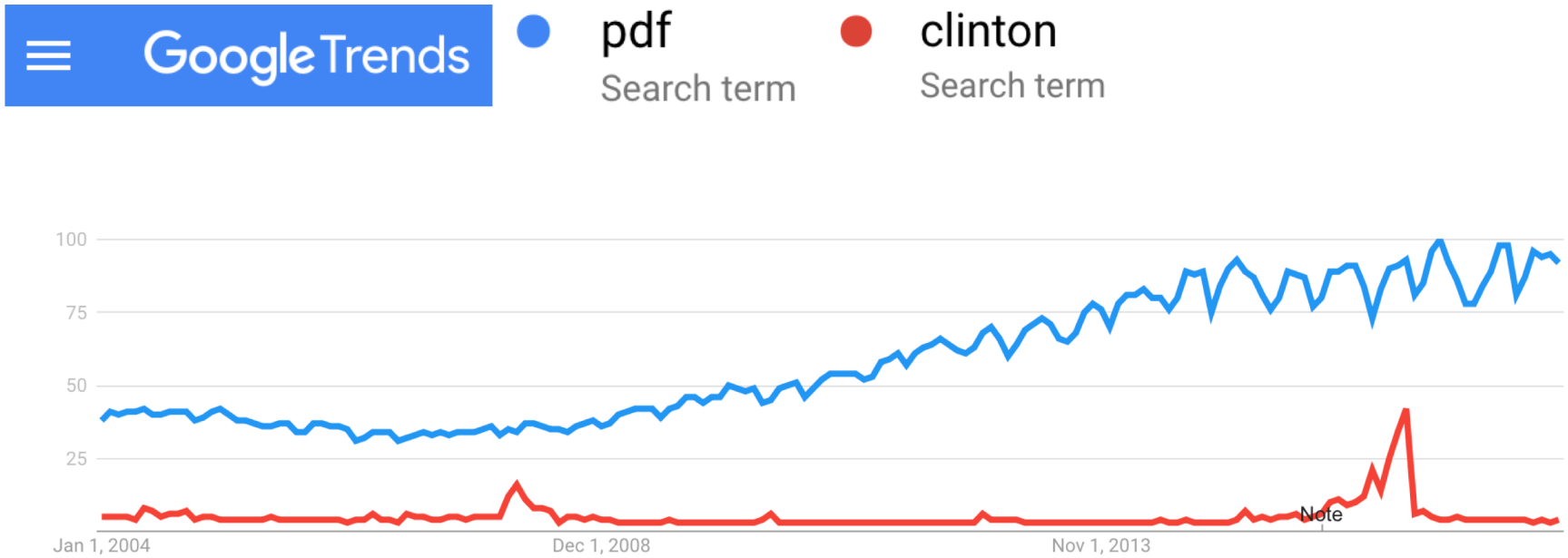
Even compared to some popular subjects…

www.pdfa.org



Duff Johnson,
Executive Director
PDF Association

Even compared to some popular subjects…



Duff Johnson,
Executive Director
PDF Association

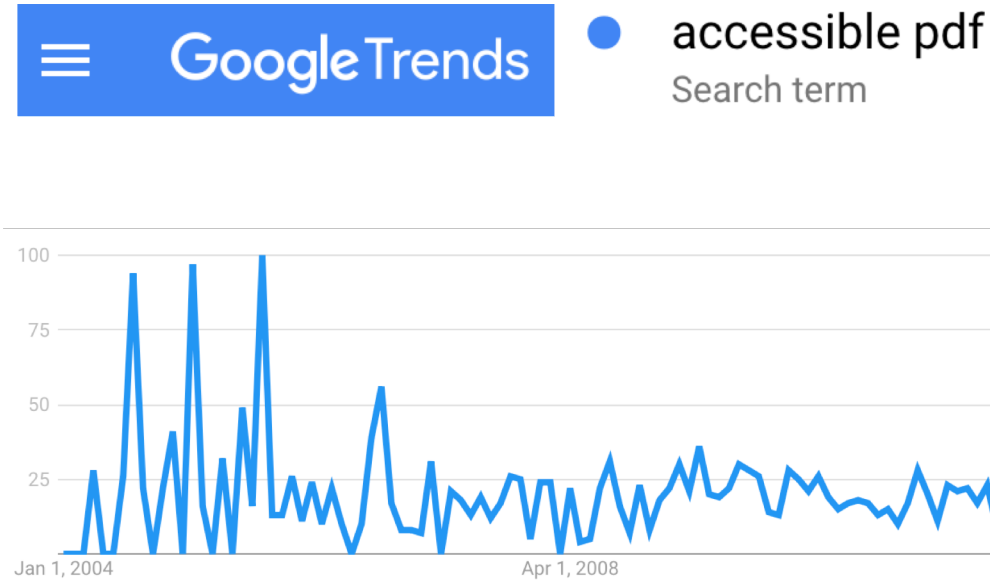
People want to do more with PDF…

www.pdfa.org

Duff Johnson,
Executive Director
PDF Association

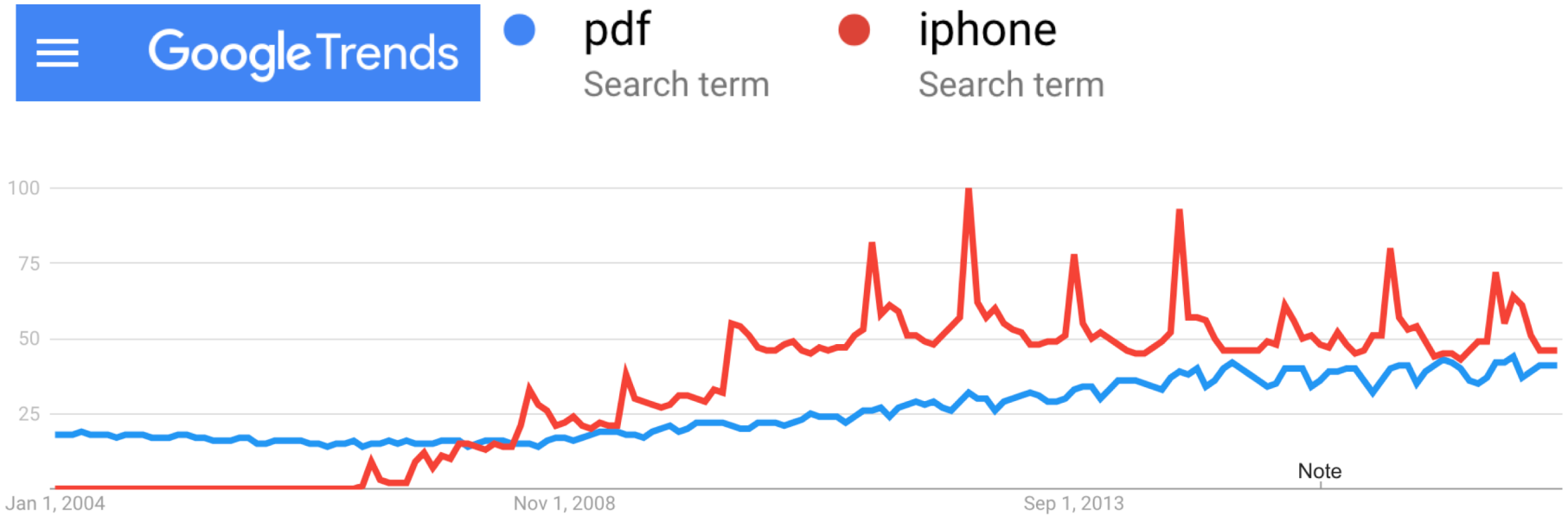
People want to do more with PDF…

www.pdfa.org



Duff Johnson,
Executive Director
PDF Association

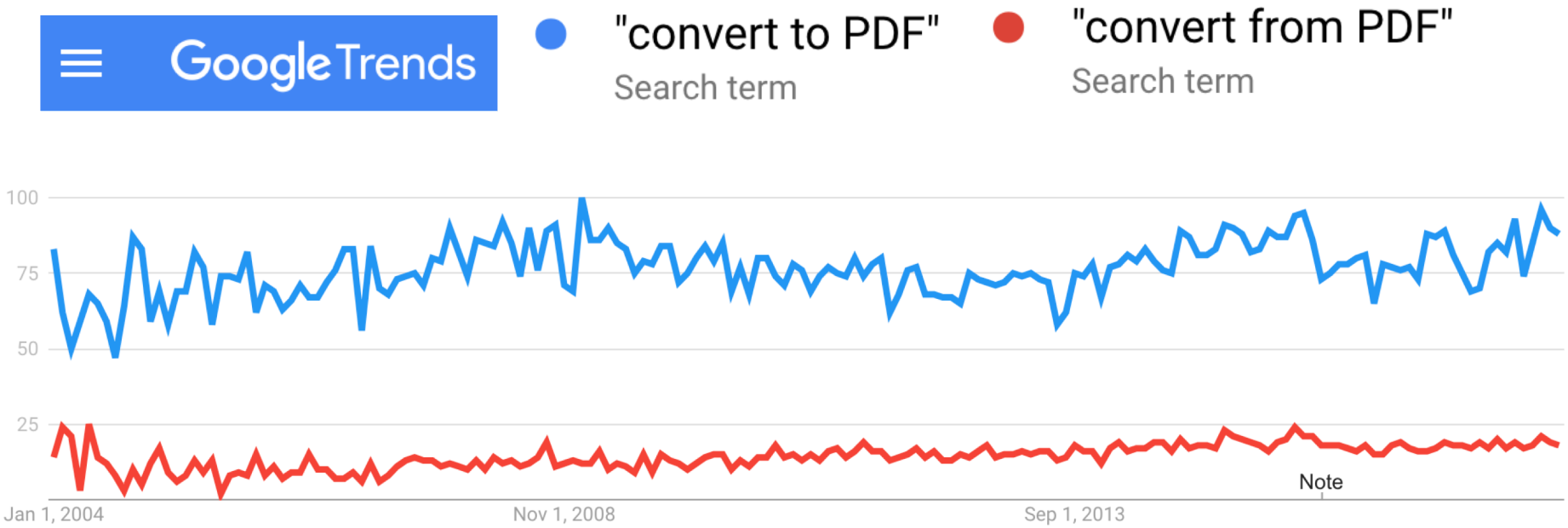
…compared to what?



Duff Johnson,
Executive Director
PDF Association

To and from PDF

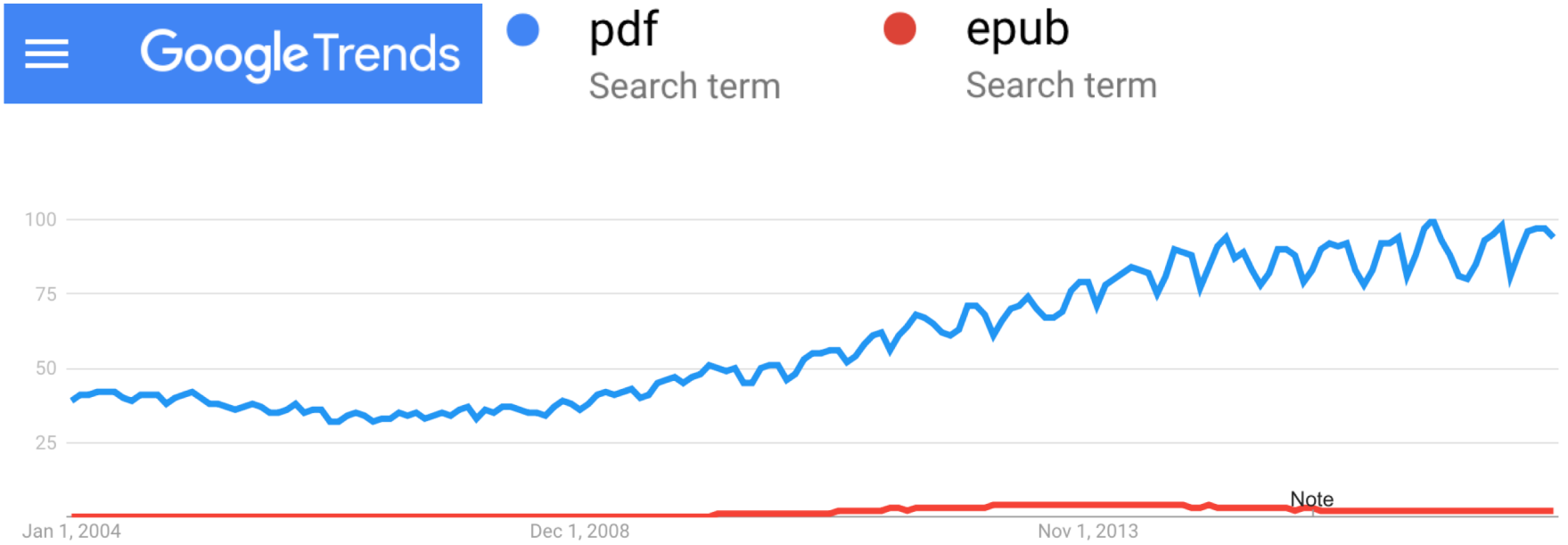
www.pdfa.org



Duff Johnson,
Executive Director
PDF Association

To and from PDF

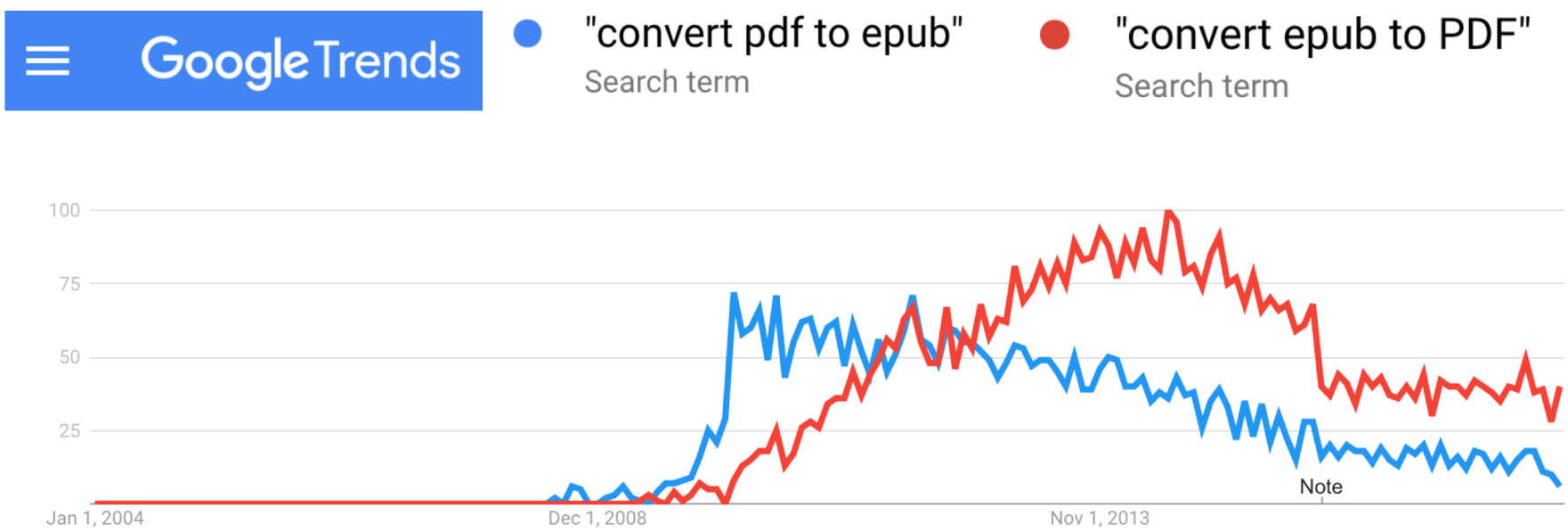
www.pdfa.org



Duff Johnson,
Executive Director
PDF Association

PDF to and from EPUB

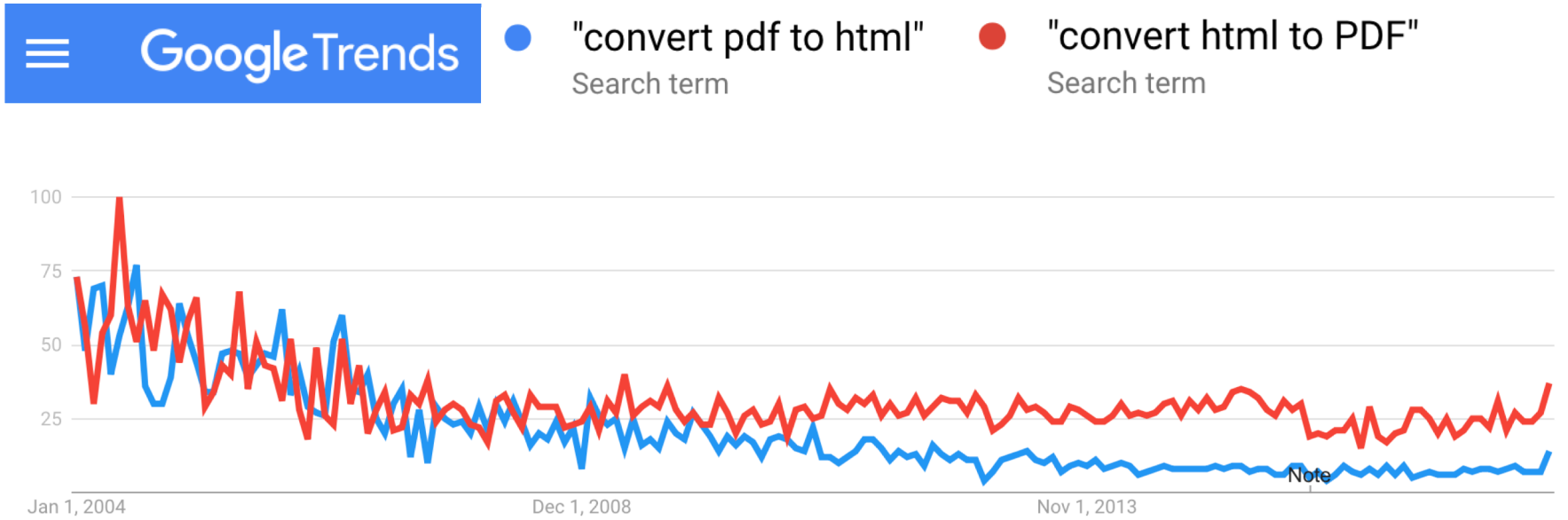
www.pdfa.org



Duff Johnson,
Executive Director
PDF Association

PDF to and from EPUB

www.pdfa.org



Duff Johnson,
Executive Director
PDF Association

Conversions are especially popular…

Duff Johnson,
Executive Director
PDF Association

Importance of e-book features to academic librarians

✓
n/a
✗
✓
n/a

US 2016

Source: Statistica
https://www.statista.com/statistics/715005/share-of-e-book-feature-importance-according-to-academic-librarians-us/

Duff Johnson,
Executive Director
PDF Association

Importance of e-book features to academic librarians

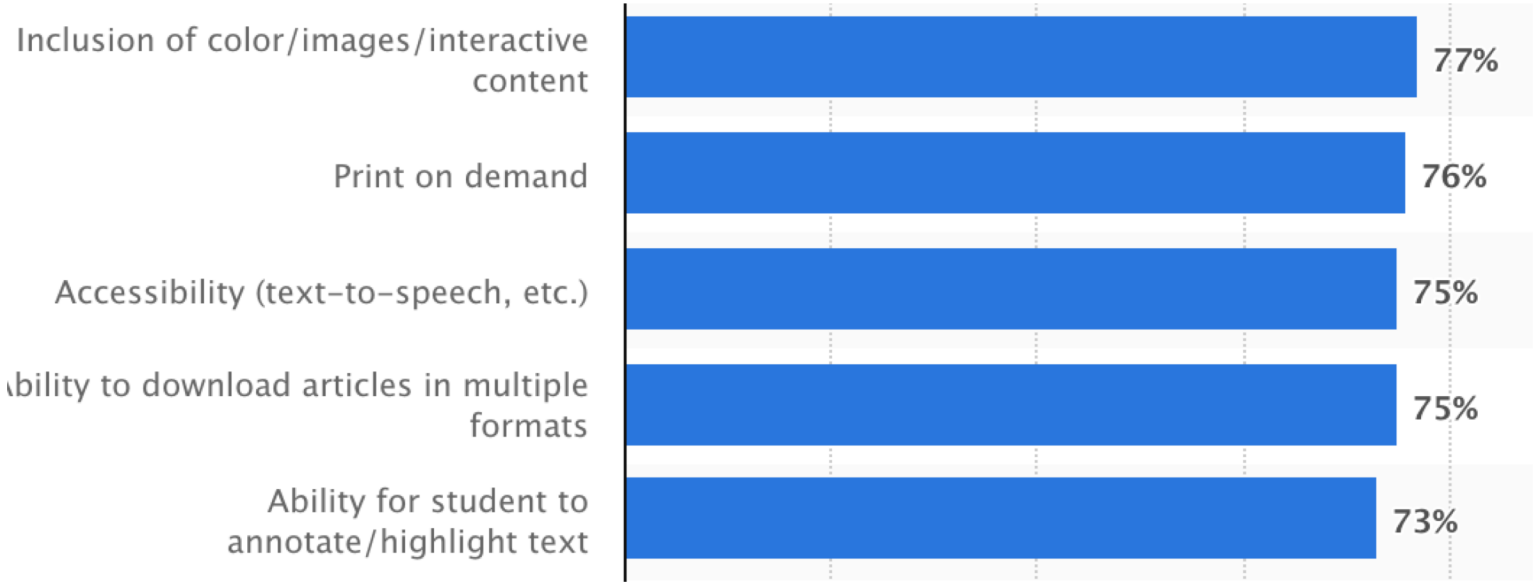
✓

✓

✓

n/a

✓

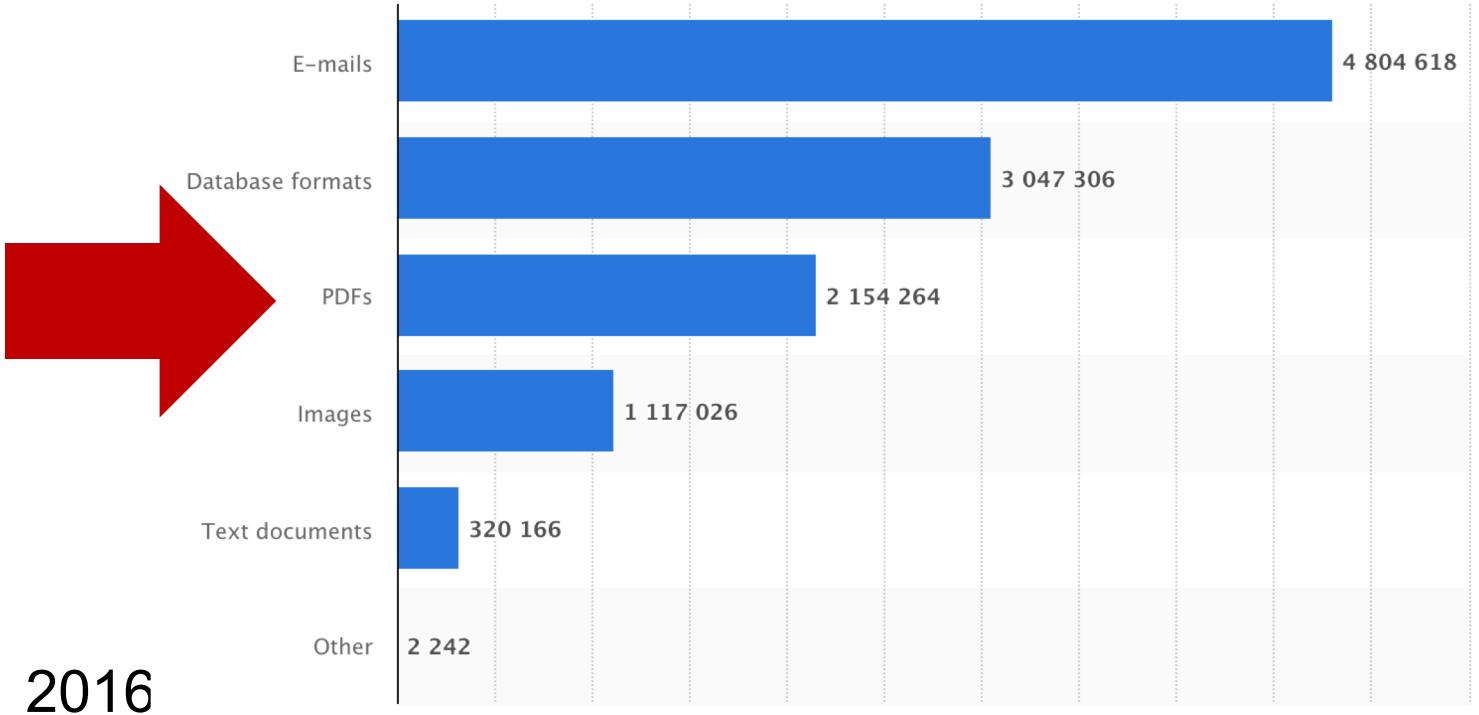


US 2016

Source: Statistica
https://www.statista.com/statistics/715005/share-of-e-book-feature-importance-according-to-academic-librarians-us/

Duff Johnson,
Executive Director
PDF Association

Type of files revealed in Panama Papers data leak



2016

Source: Statistica
https://www.statista.com/statistics/531286/panama-papers-data-type/

Duff Johnson,
Executive Director
PDF Association

How many PDF files?

Duff Johnson,
Executive Director
PDF Association

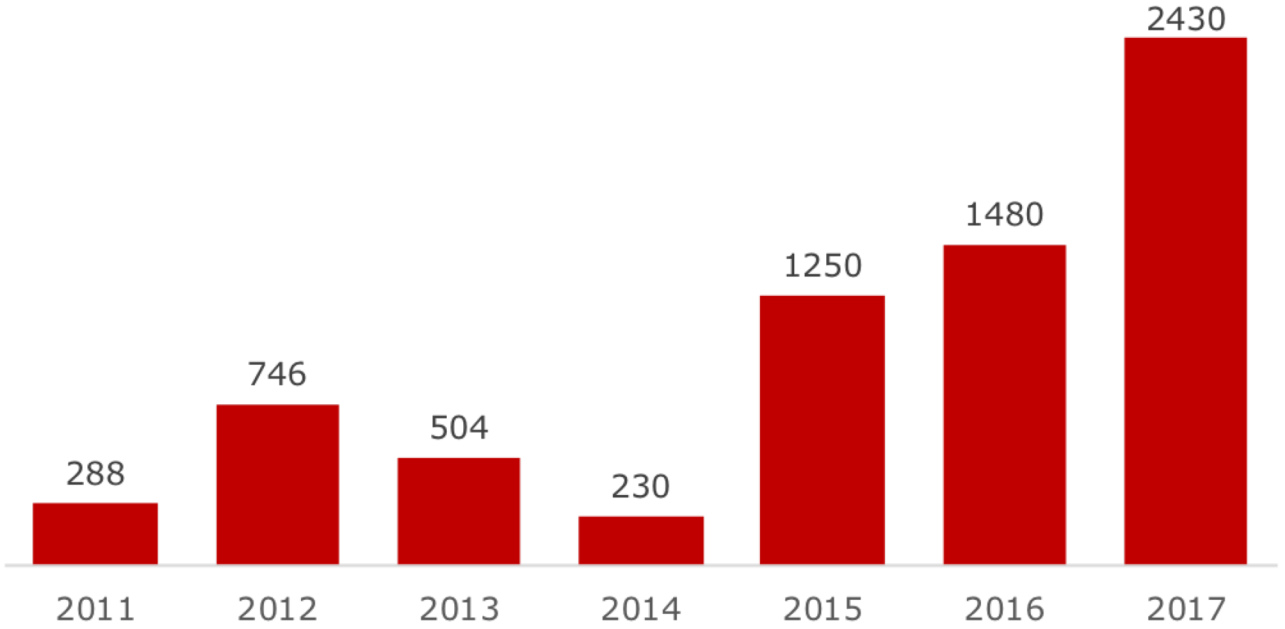
In April, 2016…

- 2.2 billion PDF files on the (public) web (Google)
- 20 billion PDF files in Dropbox
- Airbus, Boeing and the US Dept. of Justice each have over 1 billion PDFs
- 2 billion PDFs opened every year in Outlook.com
- 73 million new PDF files saved every day in Google Drive & Mail
- 60% of non-image attachments are PDFs in Outlook Exchange Enterprise

Duff Johnson,
Executive Director
PDF Association

In April, 2018

In 2018 Google no longer supports pure “filetype:” searches. 😢



Source: “filetype:” searches on Google. Figures are in millions.

Duff Johnson,
Executive Director
PDF Association

In April, 2018

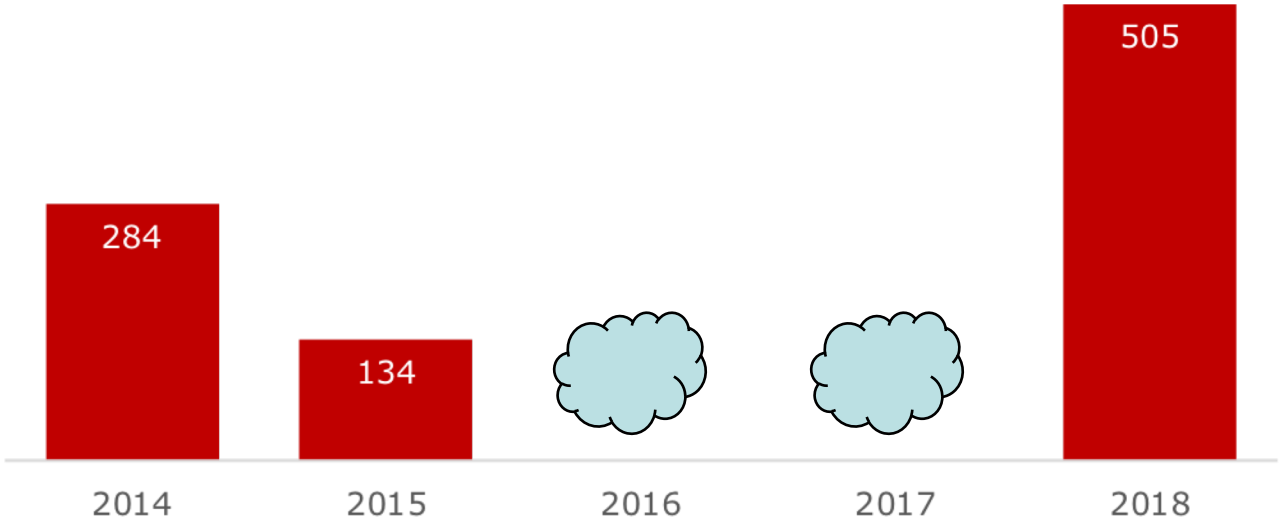
- Nothing has replaced PDF yet (or even threatened to do so)

Duff Johnson,
Executive Director
PDF Association

A PDF Association Presentation · © 2018 by PDF Association · www.pdfa.org

In April, 2018

Bing still supports pure “filetype:” searches…



Source: “filetype:” searches on Bing. Figures are in millions.

Duff Johnson,
Executive Director
PDF Association

How many PDFs??

- Trillions of PDF files
- Billions more PDF files every day

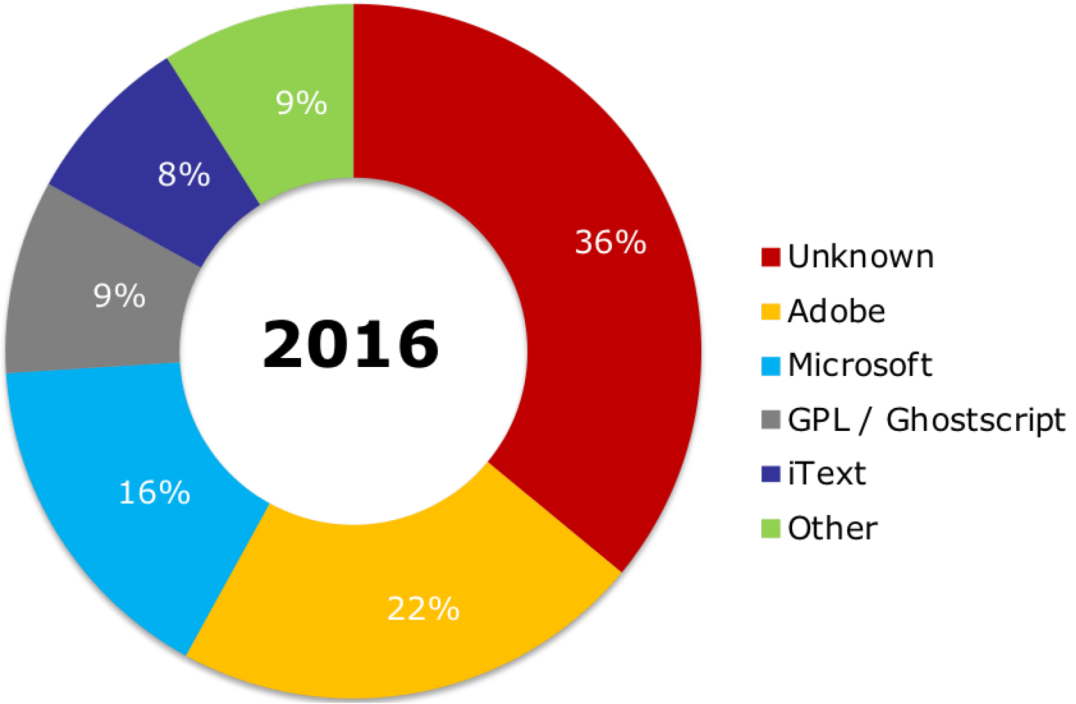
Duff Johnson,
Executive Director
PDF Association

Some hard data
(courtesy of Adobe)

Data was collected anonymously from **desktop** users who agreed to allow their usage to be tracked.

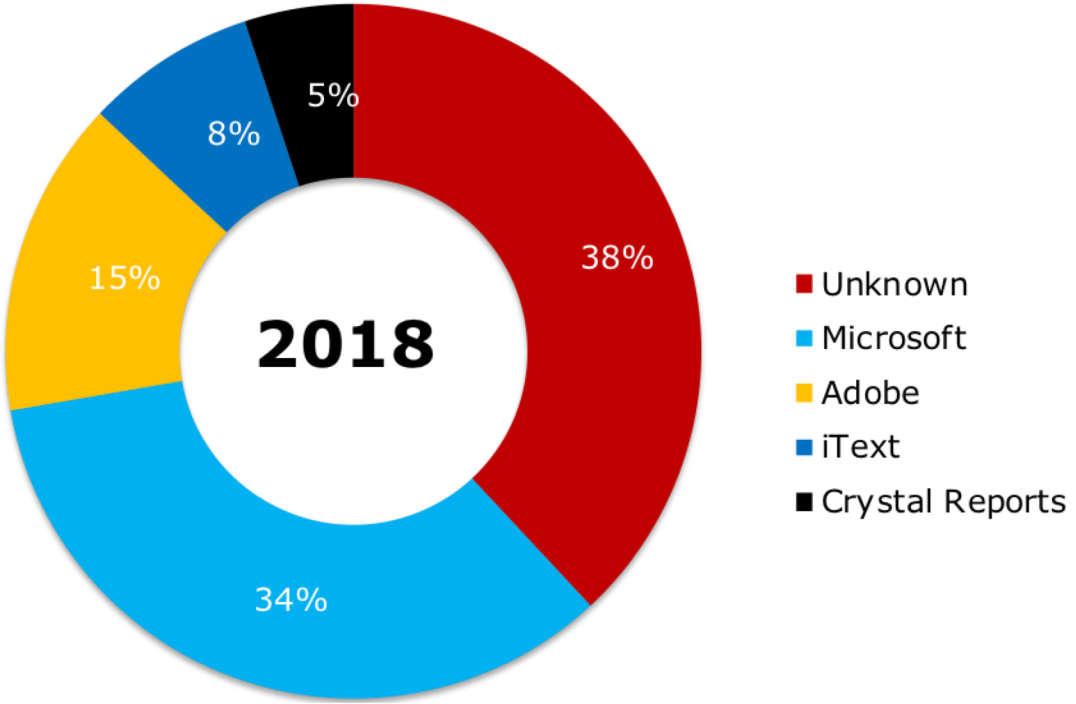
Duff Johnson,
Executive Director
PDF Association

Top producers of files opened in Acrobat/Reader



Duff Johnson,
Executive Director
PDF Association

Top 5 producers of files opened in Acrobat/Reader



Duff Johnson,
Executive Director
PDF Association

Users and documents – Adobe Reader

	2016	2018
Unique users	243,688,400	338,801,280 (+39%)
Files opened	12.3 billion	19.5 billion (+59%)
Biggest (pages)	226,882	816,960
1 page	43%	44%
90% of files	< 18 pp	< 20 pp
99% of files	< 272 pp	< 235 pp

Duff Johnson,
Executive Director
PDF Association

Users and documents – Adobe Acrobat

	2016	2018
Unique users	3,609,235	7,778,253 (+216%)
Files opened	298 million	867 million (+290%)
Image-only	23%	24%
Bookmarks	11%	13%
Secured	3%	3%
Acroform	3%	4%
Linearized	35%	37%

Duff Johnson,
Executive Director
PDF Association

PDF files randomly collected from the web
(courtesy of Dual Lab)

These are early results – more to come!

Duff Johnson,
Executive Director
PDF Association

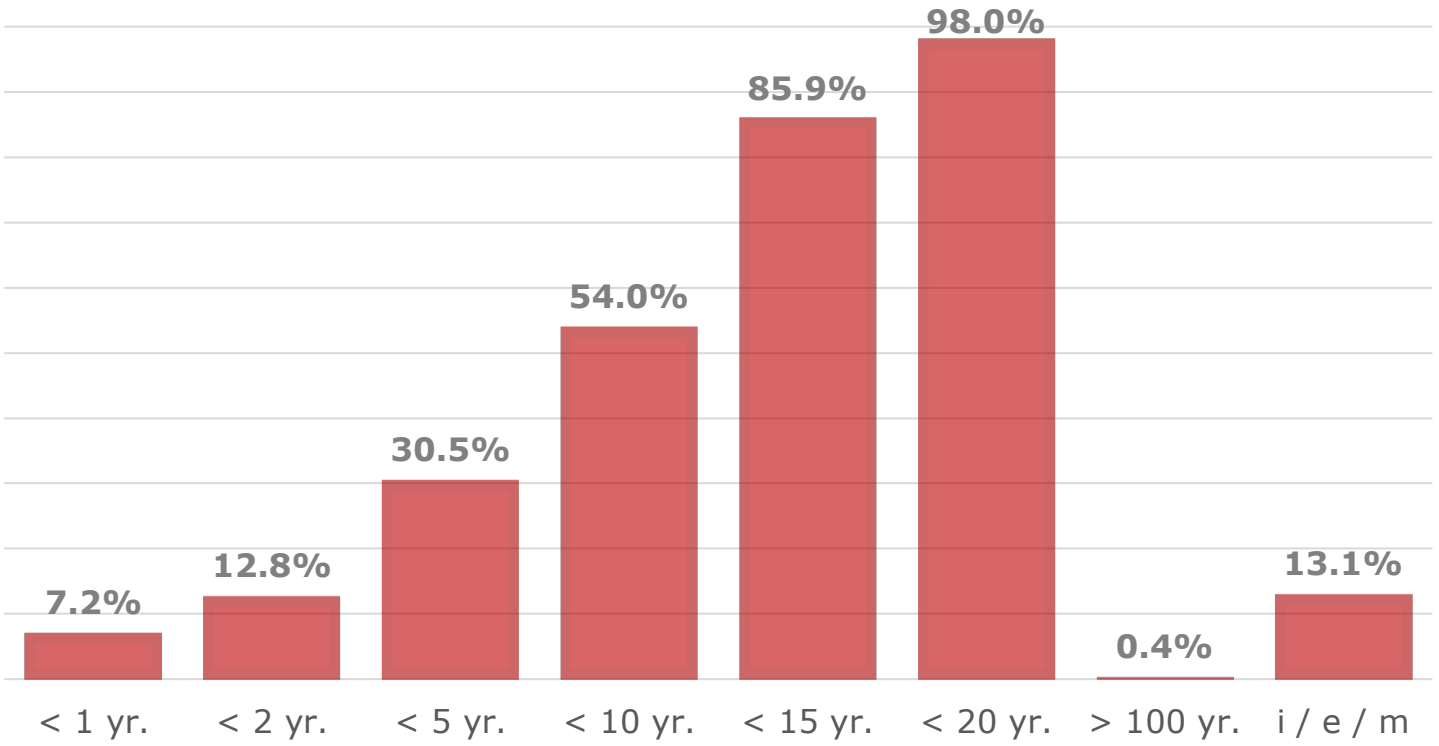
Dataset to-date

- 548,775 PDF files comprising 14,486,443 pages (26.4 pages/file)
- Method of collection:
  - Chose approx. 2,000-3,000 random words in English, German, French, Spanish, Russian (+ 100 words in Korean)
  - Perform a Bing search for each word + filetype:PDF Download the first 200 first search results
  - Remove duplicates based on hash

Duff Johnson,
Executive Director
PDF Association

Age of PDF files found online

www.pdfa.org



Duff Johnson,
Executive Director
PDF Association

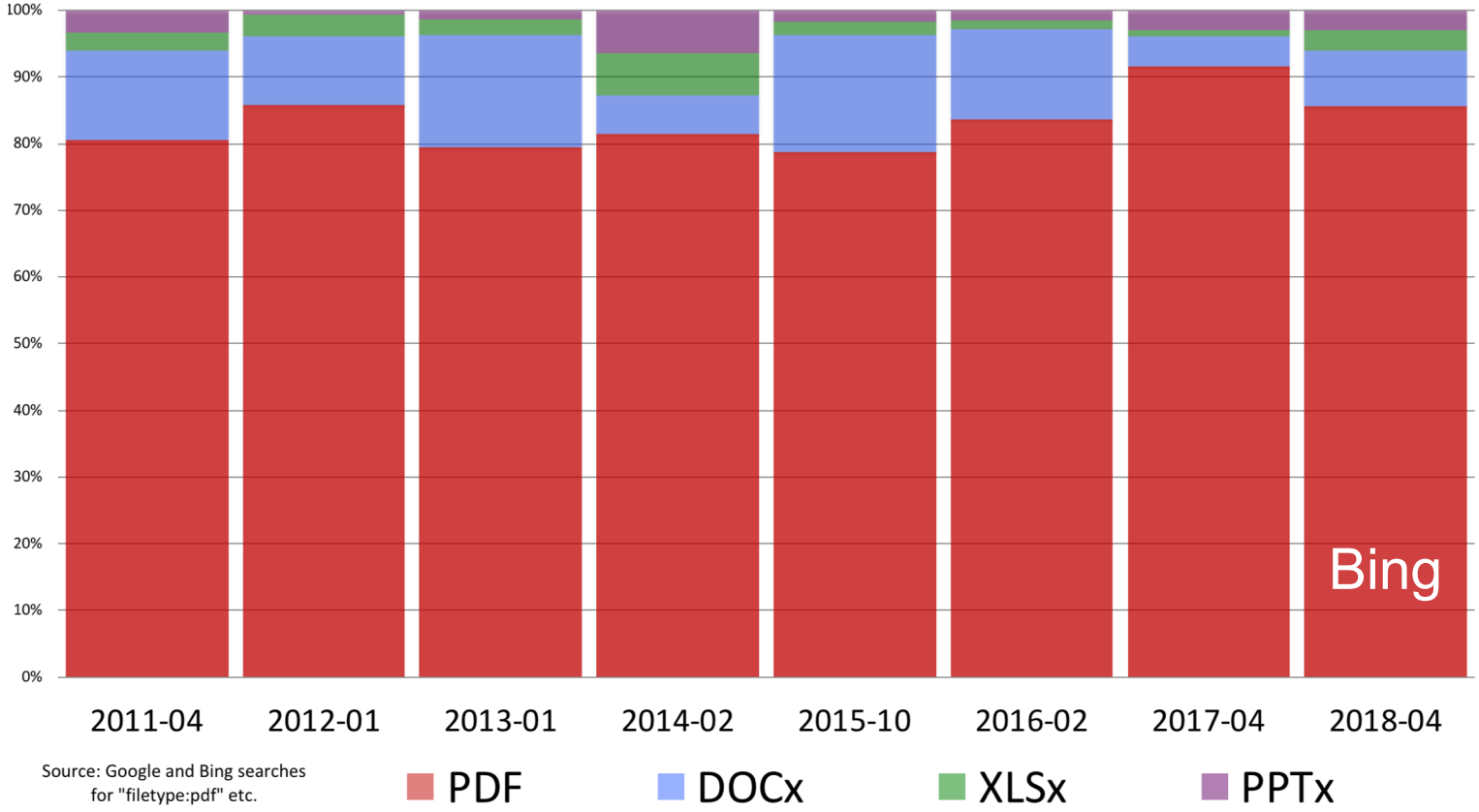
PDF subsets

- 0.73% identify as PDF/X
- 0.60% identify as PDF/A
- 0.01% identify as PDF/UA
- About 14% of files created after 2003 are tagged

Duff Johnson,
Executive Director
PDF Association

PDF as a percentage of document formats on the web

www.pdfa.org



Source: Google and Bing searches for "filetype:pdf" etc.

Duff Johnson,
Executive Director
PDF Association

2018-April - PDF vs. HTML on the web, according to Bing

- 9,360,000,000 “static” HTML files
  Does not include lots of web content!
  (Facebook, WordPress archives, etc.)
- 505,000,000 PDFs

1 in 20
static “pages”
are PDF files!

Duff Johnson,
Executive Director
PDF Association

Summary

- There’s a lot of PDF documents, and many more everyday
- PDF’s niche is pretty large
- PDF has no real competitor in its niche
- PDF continues to gain mindshare (as expressed in searches)
- The industry has a lot of room to grow

Duff Johnson,
Executive Director
PDF Association

Thank you! Any questions?

Get in touch: duff.johnson@pdfa.org
Web site: www.pdfa.org
Twitter: PDFassocation

Duff Johnson,
Executive Director
PDF Association