PDF’s mindshare

Environment, end user awareness, usage

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF is…

- Recognized worldwide
- Broadly adopted
- Accepted everywhere
- Supported by a rich ecosystem

But what do we know about PDF adoption, utilization, etc.?

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF association

www.pdfa.org

This presentation

- Context
- Mindshare
- Some data

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

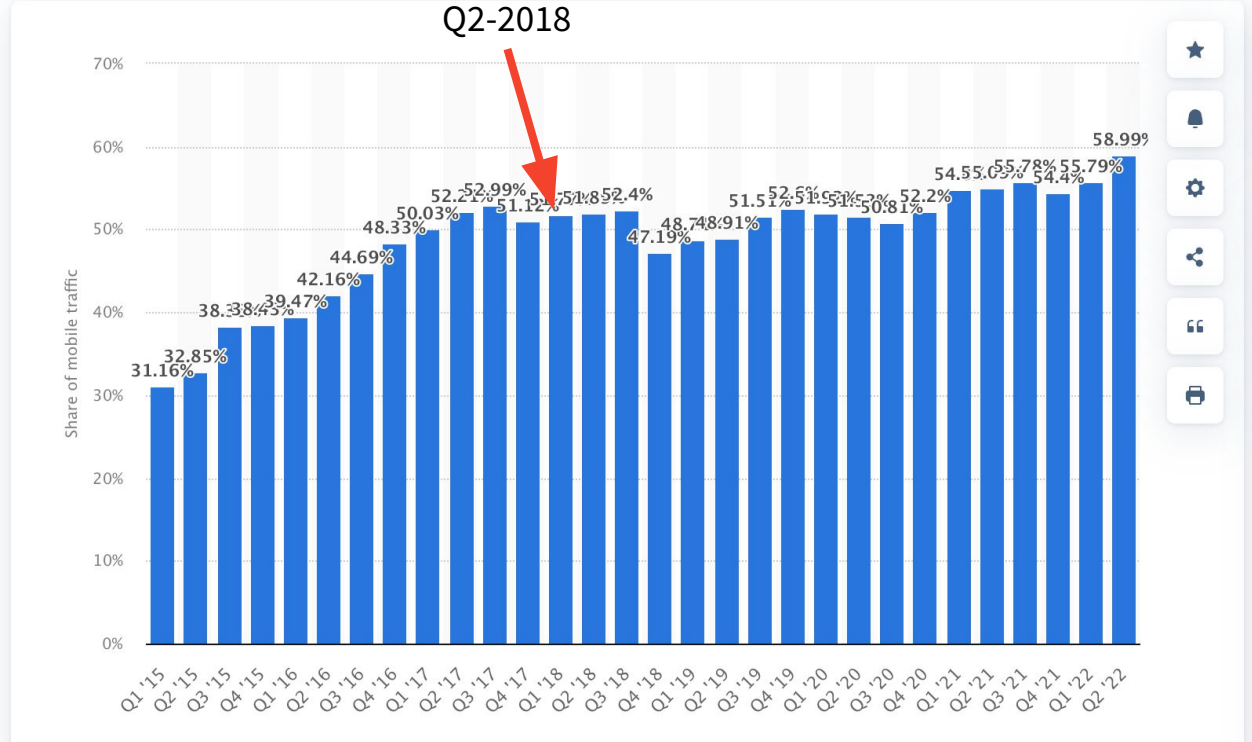
Platforms are essentially unchanged since 2018

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Mobile devices increased traffic share to ~60%

www.pdfa.org

Percentage of mobile device website traffic worldwide 2015 - 2022



Source: Statista / https://www.statista.com/statistics/277125/share-of-website-traffic-coming-from-mobile-devices/

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF association
www.pdfa.org

What would it take to compete with PDF?

The value proposition in 2022; same as in 2018:

- Self-contained (thus portable)
- Fixed-layout, but accommodates annotations, forms, reuse, etc.
- Create from any source
- Generic support, including basic support in all browsers
- Easy-as-printing to create
- Published specification

Competing formats: Cloud-based office software and storage (EPUB and PWP have shown little traction)

Competing ideas: Reinventing the wheel

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Potentially competitive technologies in 2022

4 classes of “competitive” technologies

Class	Key value	Key failing
Web-based (inc. EPUB, PWP)	Ubiquitous, adaptive	Ephemeral, limited workflows
Application (inc. OOXML, ODF)	Editable, dynamic	Not portable, application-centric
Images	Simple, easy	The lack of text
Paper / fiche	Eye-readable	Various limitations

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF’s mindshare

SWOT and PESTLE analysis

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF association

www.pdfa.org

Strengths, **W**eaknesses, **O**pportunities, **T**hreats

Strengths

- Reliability
- Ubiquity
- Broad range of capabilities to accommodate general document needs

Weaknesses

- Variations in support
- Perceptions of mainstream web developers
- Complexity (and a paywalled specification does not help!)

Opportunities

- Digital transformation
- Semantic structure
- Workflow integration
- Packaging

Threats

- Bad / low-quality / unreusable files
- Erosion of “simulated paper” workflows
- Vulnerabilities (redaction, cybersecurity)
- Commodification at a low common denominator that fails to address the n-device challenge

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Political, Economic, Social, Technical, Legal, Environmental

Political

PDF is used by government; government operations are increasingly regulated.

Due to challenges with accessibility PDF use is increasingly incorrectly perceived by governments (e.g., Australia, UK) as a “problem”.

Economic

PDF is ubiquitous throughout the economy as “digital paper”, with no identifiable competition in that role.

In addition to inspiring new web-based technologies to address modern business needs, the expansion in remote working has expanded use of PDF.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Political, **E**conomic, **S**ocial, **T**echnical, **L**egal, **E**nvironmental

Social

For end users, PDF files for documents as well as image and video data have supplanted paper and celluloid as typical means of preserving and sharing information.

“PDF it” is increasingly replacing “print it”.

That said, PDF also presents a challenge in terms of accessibility to users with disabilities, with most authors ignorant of what’s needed and most tools offering basic - or inadequate - support.

Technical

Ubiquitous as it is, many developers and IT professionals regard PDF as a legacy technology that doesn’t play well with the web and mobile devices and is slow to update to the latest crypto, HDR and other tech. PDF is not perceived as “agile”.... and that’s both good and bad.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Political, **E**conomic, **S**ocial, **T**echnical, **L**egal, **E**nvironmental

Legal

As ISO-standardized technology PDF is increasingly required in legislation or recommended by government entities worldwide. There remains, however, much room for improvement in terms of specific policies.

Accessibility mandates present ongoing challenges for digital content; PDF in particular.

Environmental

PDF is in some ways regarded as a “green” technology in that it can replace paper in many workflows. To this end, highly optimized, smaller files can also reduce storage requirements and thus, energy costs.

However, there’s also a perception that delivering PDF encourages end users to print. Industry can counter this with increased emphasis on robust reuse.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF’s mindshare

A defacto standard

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

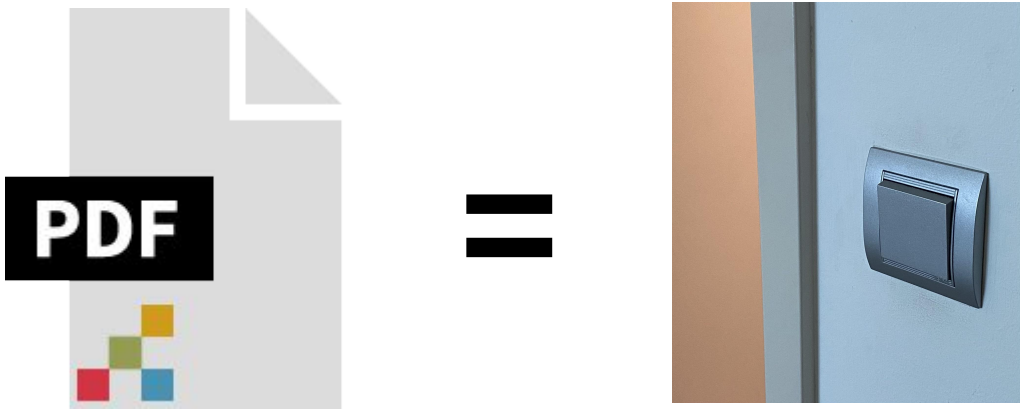
A “de facto” standard

“…well-known and illustrative examples of de facto standards:

- QWERTY (arrangement of keyboard letters)
- USB (serial interface for computers)
- MP3 (audio format)
- HTML (computer language)
- DOC (computer file format for creating content)
- PDF (computer file format for sharing content)

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF is like electricity…



You never want to think about how it works.

You can’t live without it.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

How do you recognize a “de facto” standard?

What does a “de facto” standard look like in real life?

We’ll use Google Trends - statistics from activity on Google’s search engine.

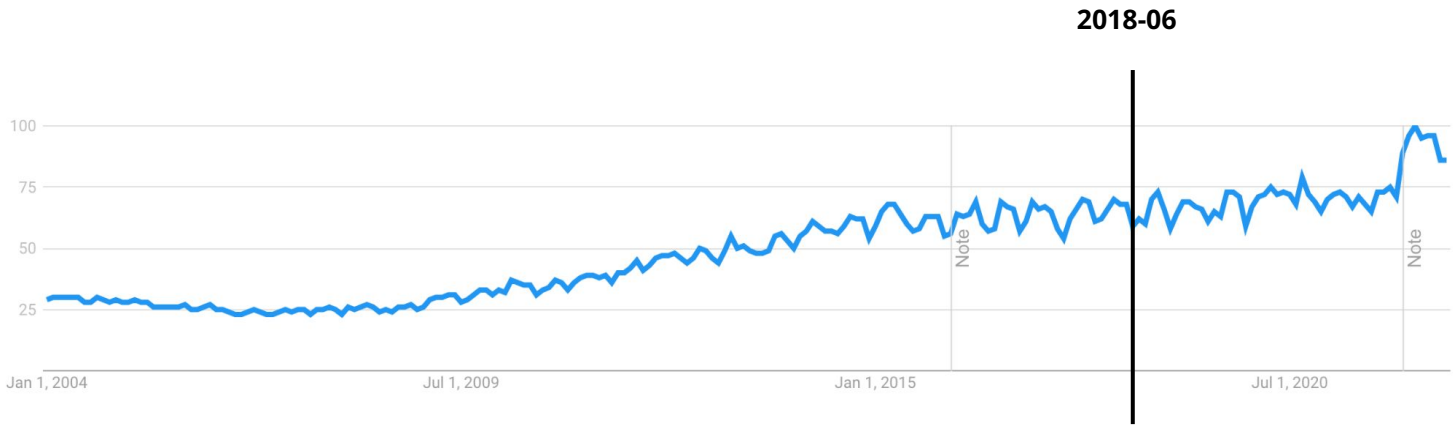
https://trends.google.com

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF keeps going up…

Google Trends **PDF**

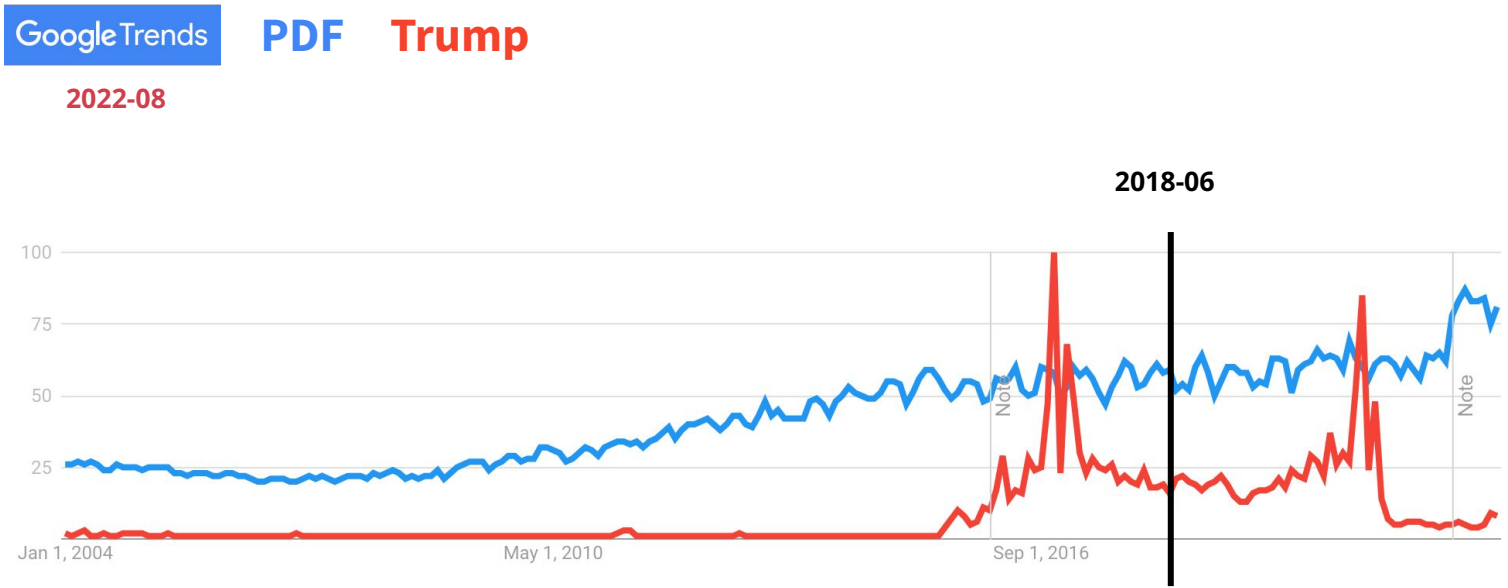
2022-08



Google applied an "improvement to our data collection system" in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

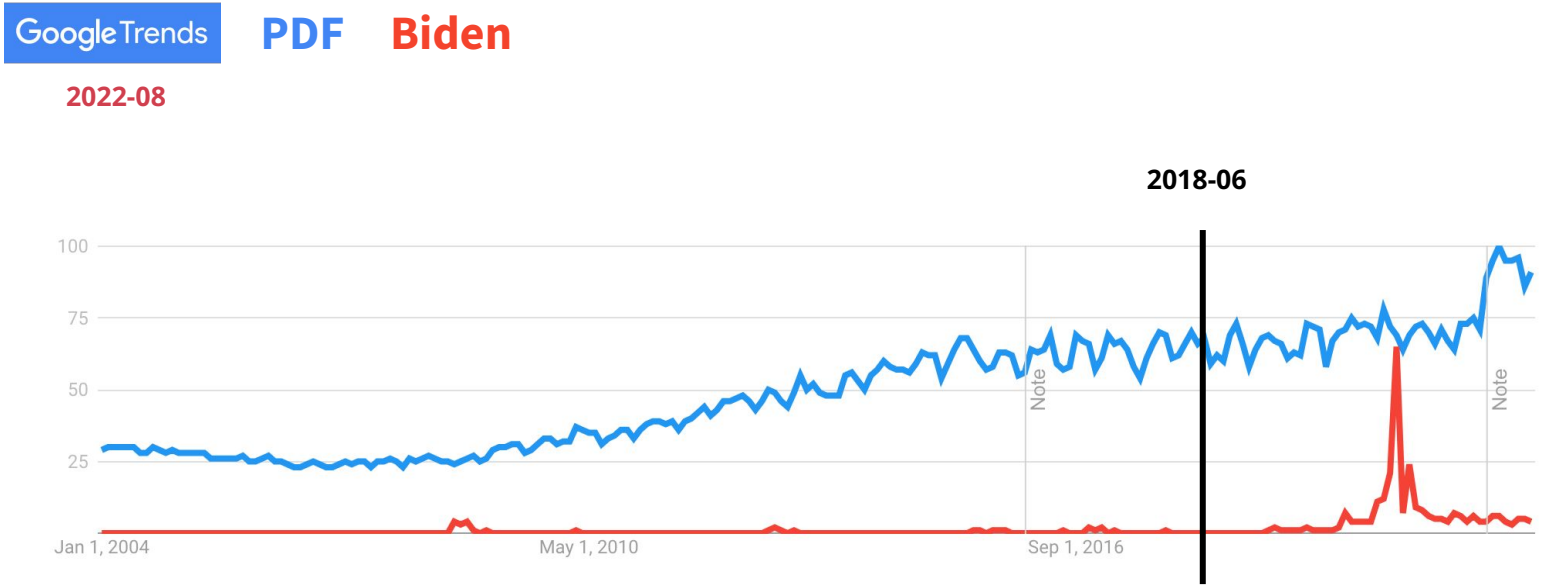
And compares well to other popular subjects…



Google applied an "improvement to our data collection system" in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

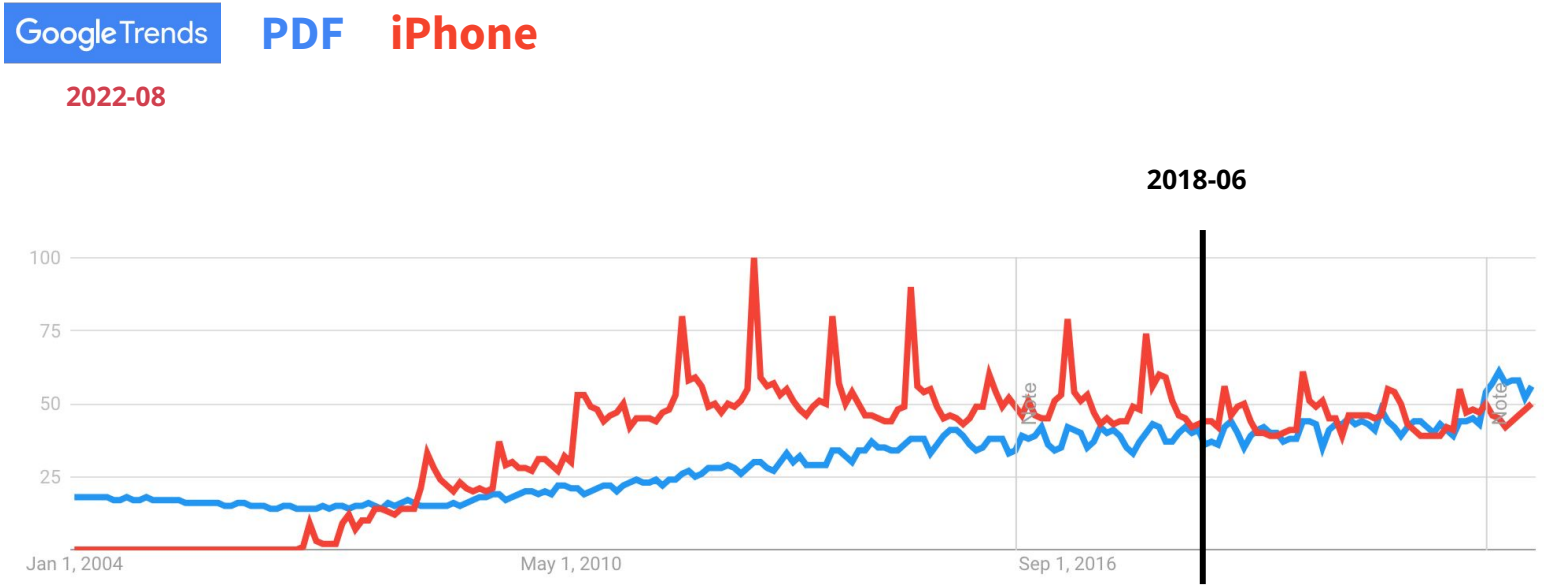
Biden won the US election but he didn't beat PDF



Google applied an "improvement to our data collection system" in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

What else gets PDF's relative search volumes?



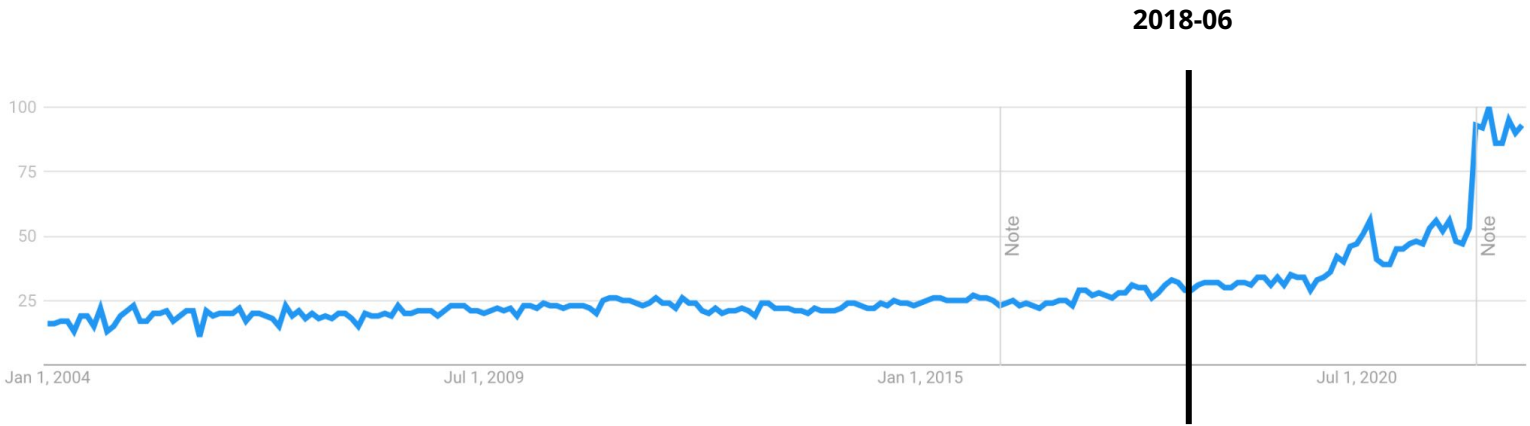
Google applied an "improvement to our data collection system" in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Users want to do more than just view PDFs…

Google Trends “edit PDF”

2022-08

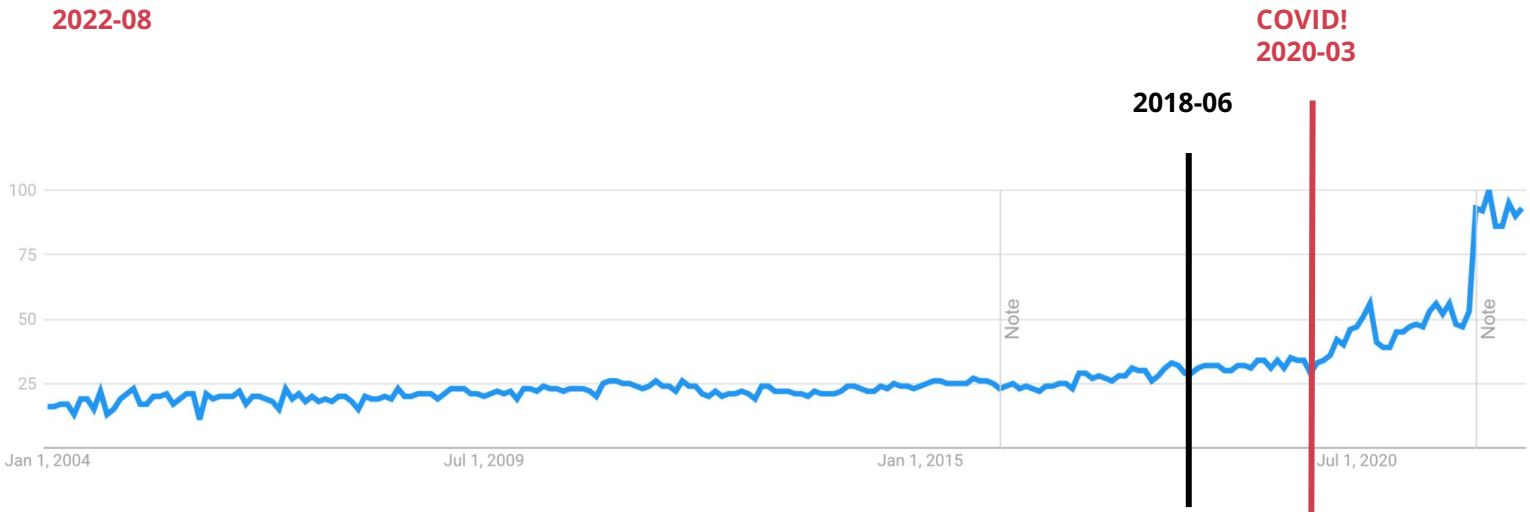


Google applied an “improvement to our data collection system” in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

And even more so during a pandemic!

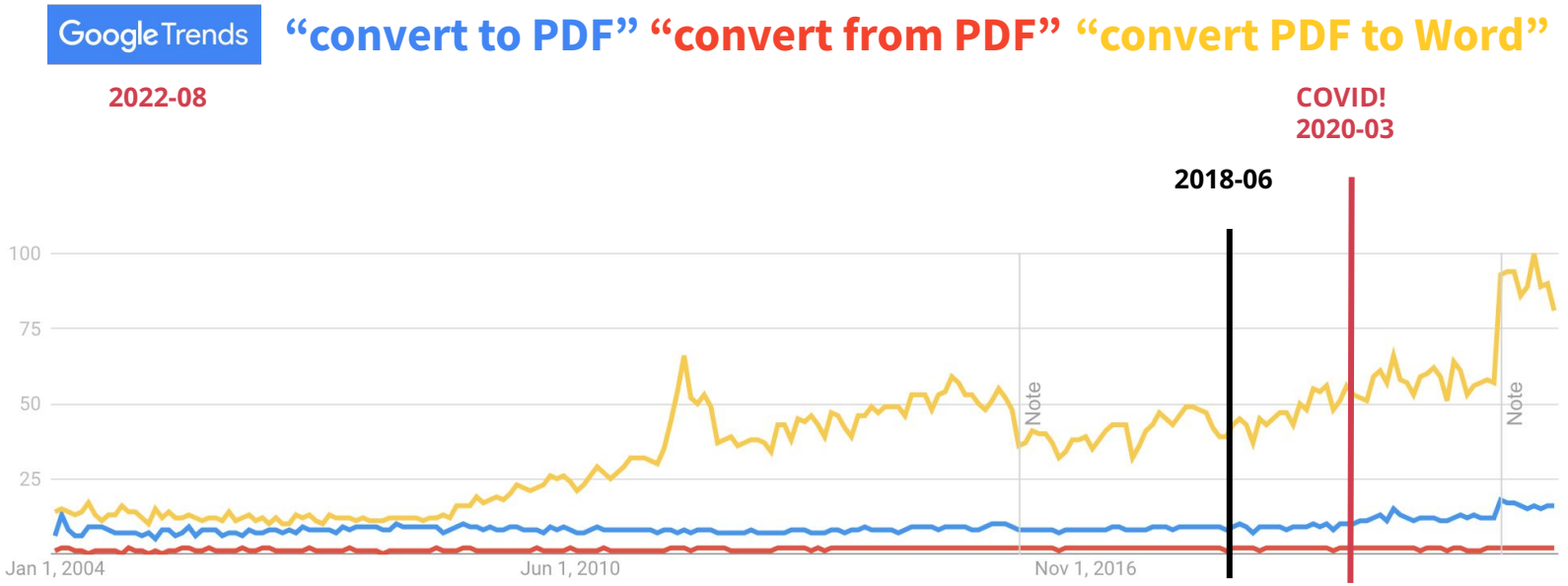
Google Trends “edit PDF”



Google applied an “improvement to our data collection system” in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

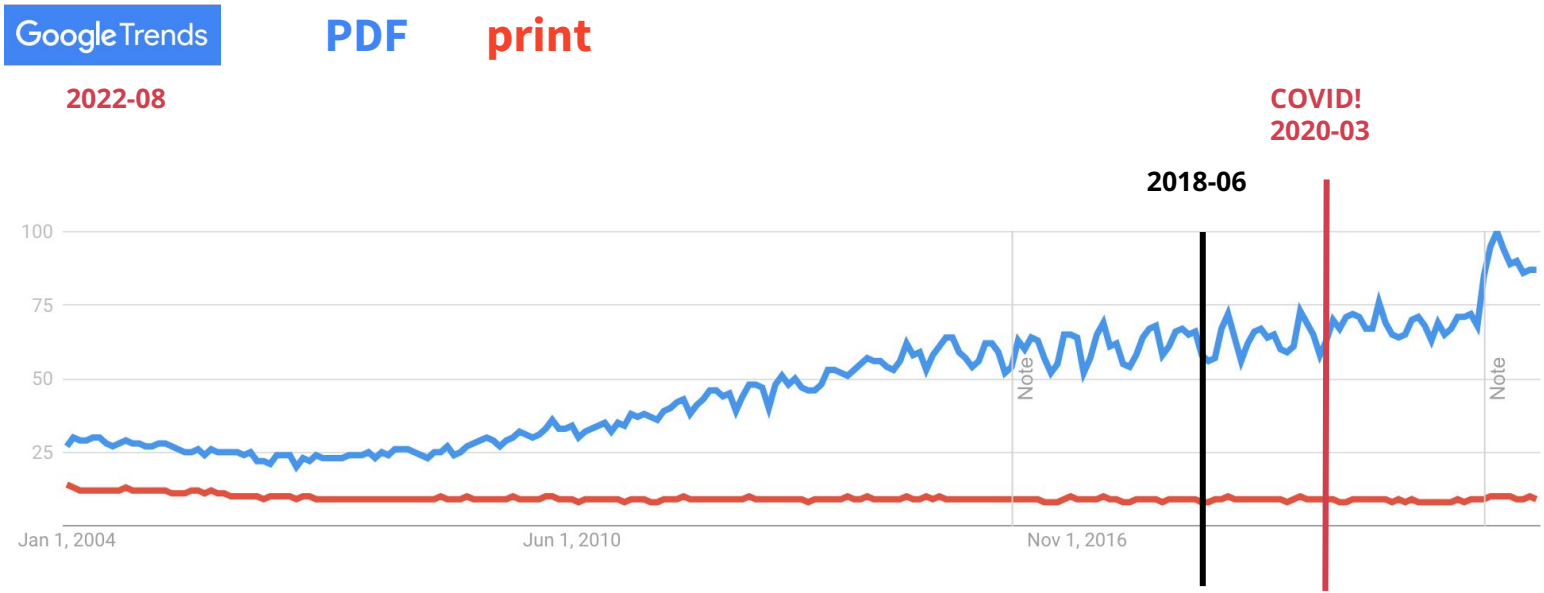
Work From Home: +1 to PDF reuse!



Google applied an “improvement to our data collection system” in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

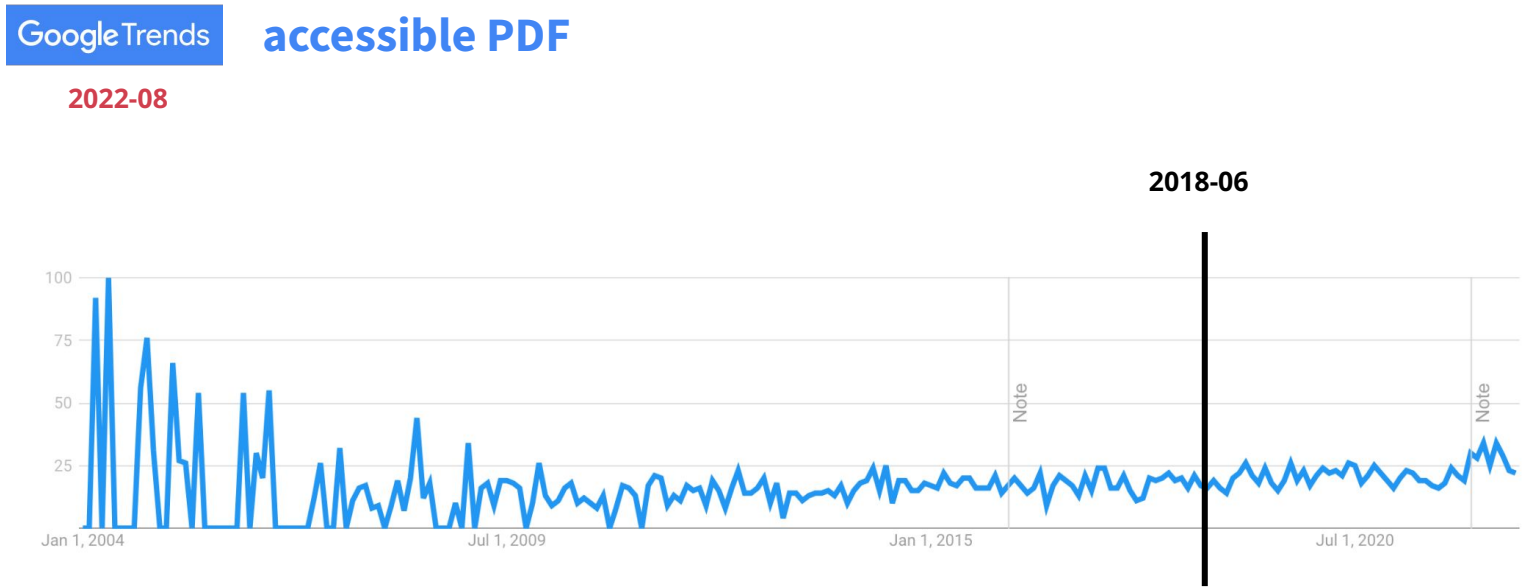
Work From Home: -1 for print



Google applied an "improvement to our data collection system" in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

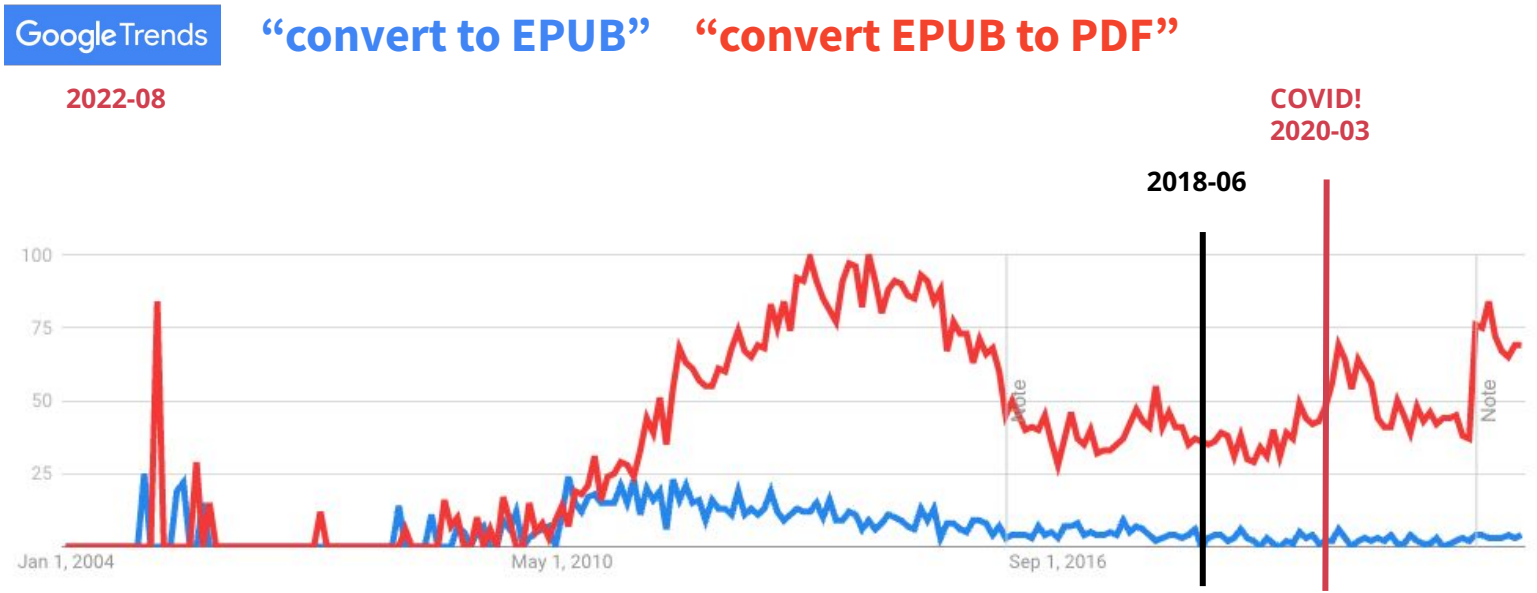
“Accessible PDF” still grows relative to other searches



Google applied an “improvement to our data collection system” in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Another COVID symptom..



Google applied an “improvement to our data collection system” in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF remains competitive among conversion objectives

www.pdfa.org

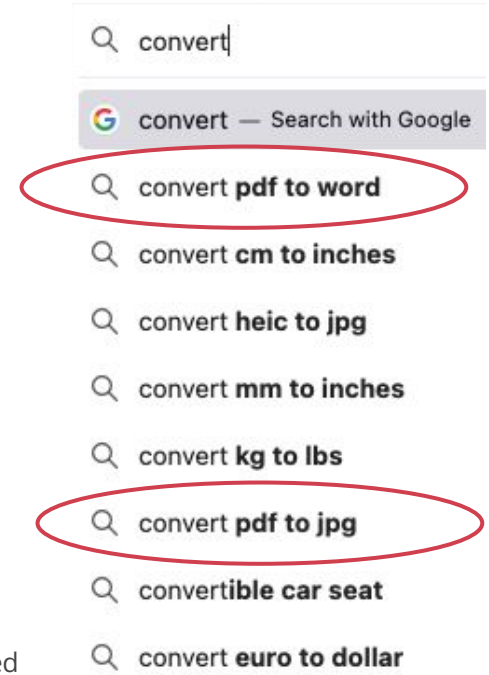
2018-06

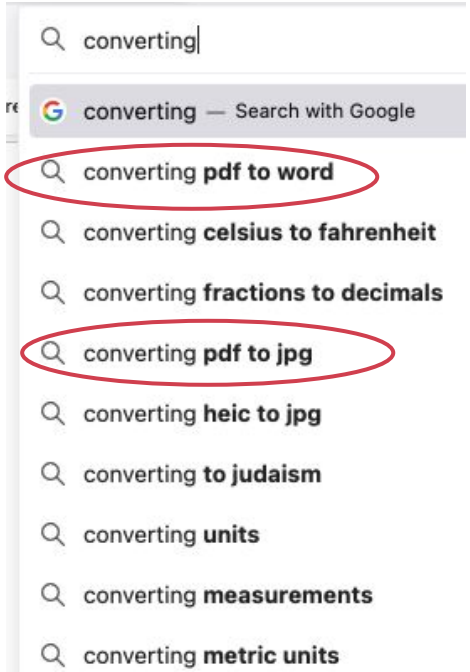
2022-08

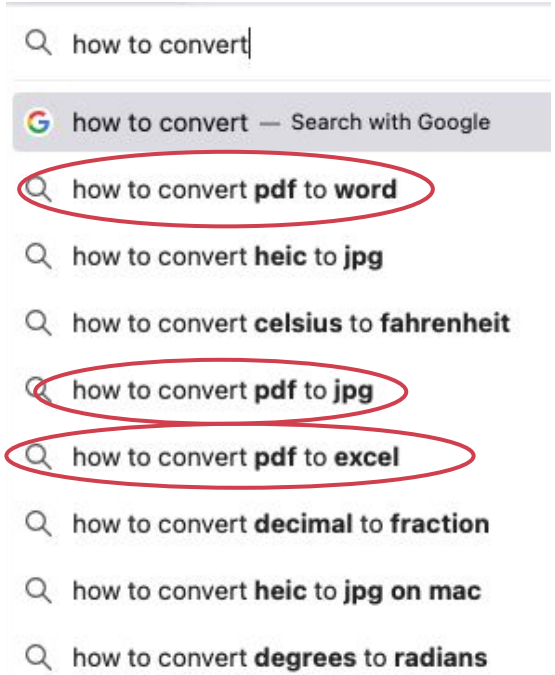
Google applied an "improvement to our data collection system" in 2022-01.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Searches for “conversion” are often about PDF







Searches conducted 2022-08 using Firefox with an empty cache.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF demographics

What’s out there?

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Four years later…

Four years ago I presented Dual Lab's data based on their own crawl of files captured randomly from the web in Q2, 2018.

548,775 PDF files

For 2022 Tim Allison has kindly shared data from his analysis of PDF files found in the August 2021 CommonCrawl.

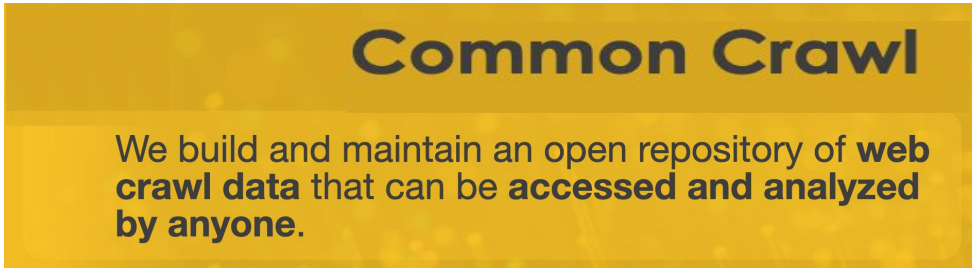
7,932,836 PDF files

Adobe's Leonard Rosenthol has also shared data gathered by Adobe's analytics and from their own "Deep Reservoir" corpus of files collected over time.

59,674,054 PDF files

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF Observatory



https://commoncrawl.org/2021/08/july-august-2021-crawl-archive-available/

Tim's 2022 PDF Days presentation is in the "Implementation" track!

Tim Allison's team at NASA built the **File Observatory** to support DARPA's SafeDocs program by enabling parser developers to understand features of PDF files in the wild at scale.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Cumulative percentage of PDF files by age

www.pdfa.org

2021-08 CC

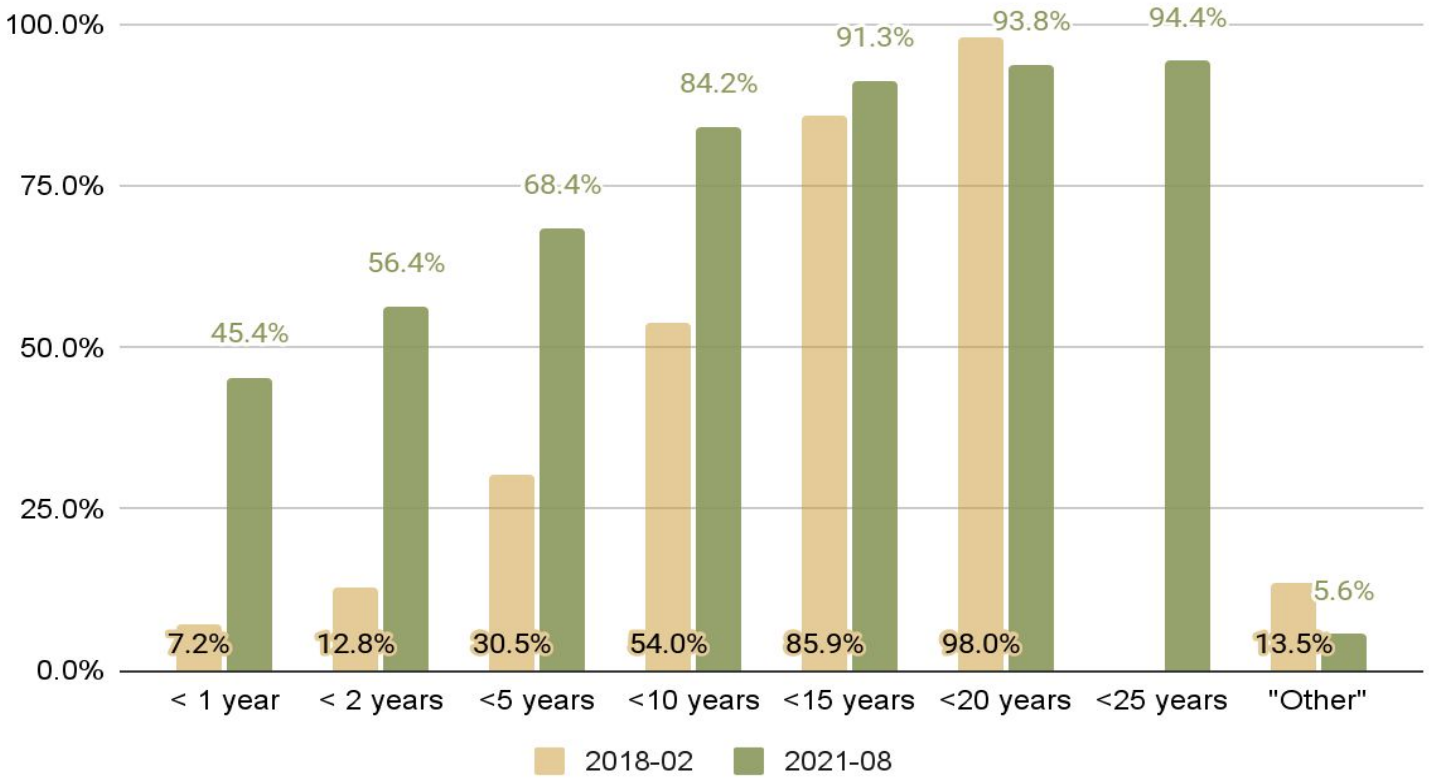
7,932,836 PDFs

Source: CommonCrawl 2021-August, Tim Allison (NASA)

2018-Q2 Dual Lab

548,775 PDFs

Source: Dual Lab, Q2-2018



A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Proportion of ISO subsets and “Tagged PDF”

www.pdfa.org

2021-08 CC

7,932,836 PDFs

Source: CommonCrawl 2021-August, Tim Allison (NASA)

2018-Q2 Dual Lab

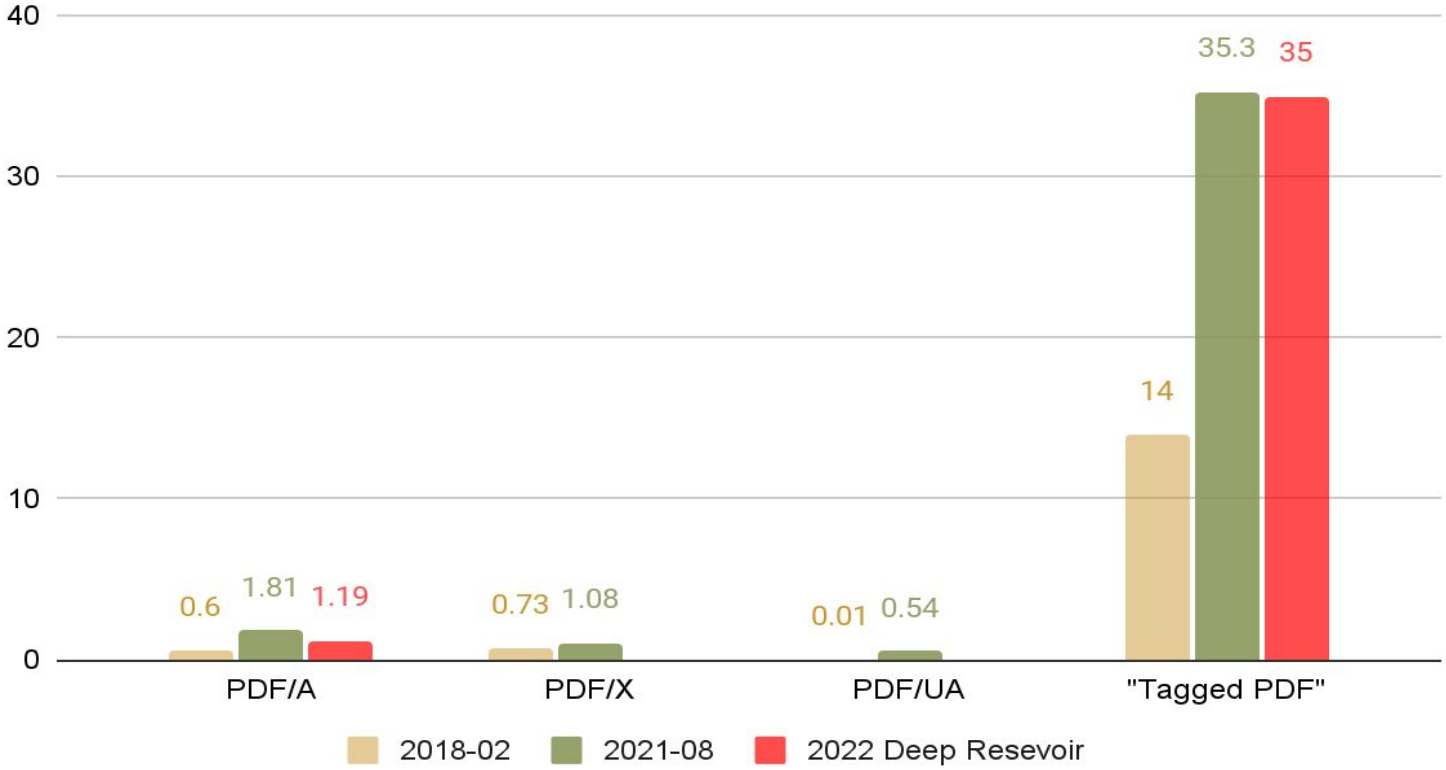
548,775 PDFs

Source: Dual Lab, Q2-2018

Deep Reservoir

59,674,054 PDFs

Source: Adobe (2022-09)



A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

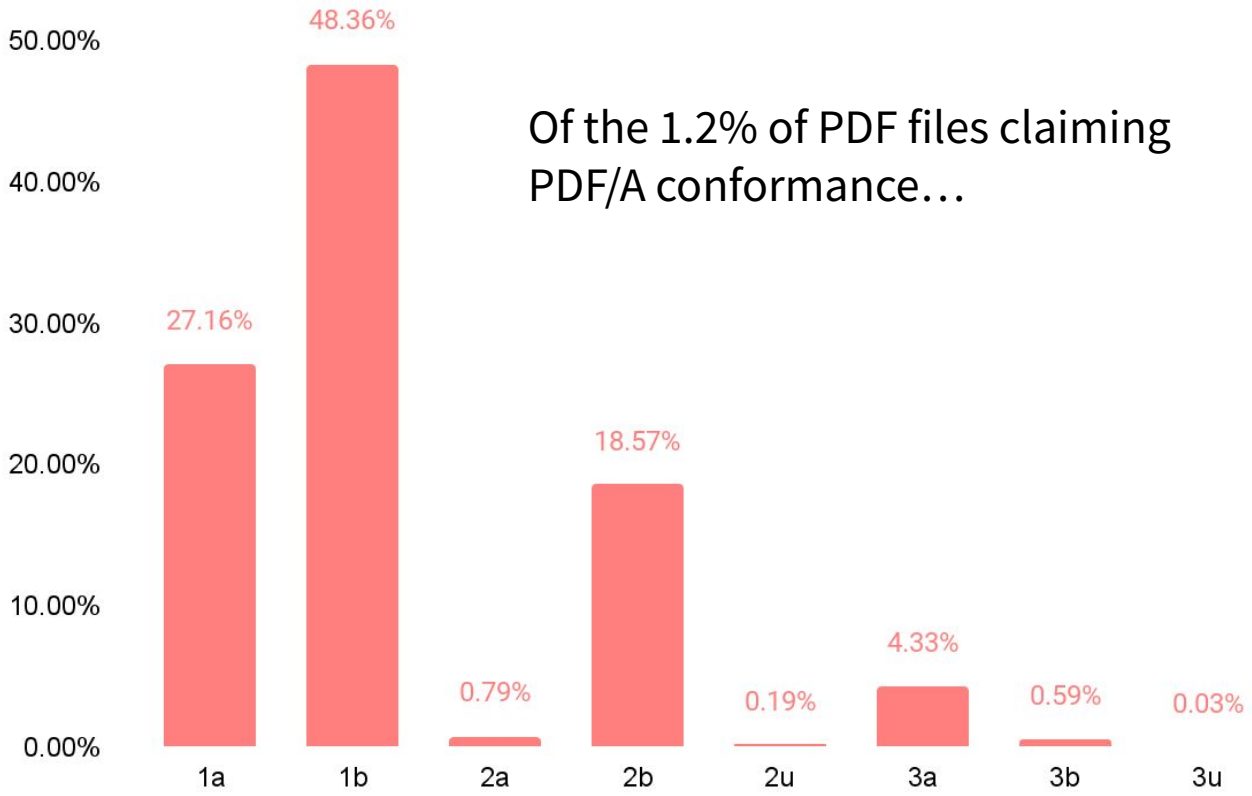
Flavors of PDF/A

www.pdfa.org

Deep Reservoir

59,674,054 PDFs

Source: Adobe (2022-09)



Of the 1.2% of PDF files claiming PDF/A conformance…

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Proportion of PDF files by file header version

www.pdfa.org

2021-08 CC

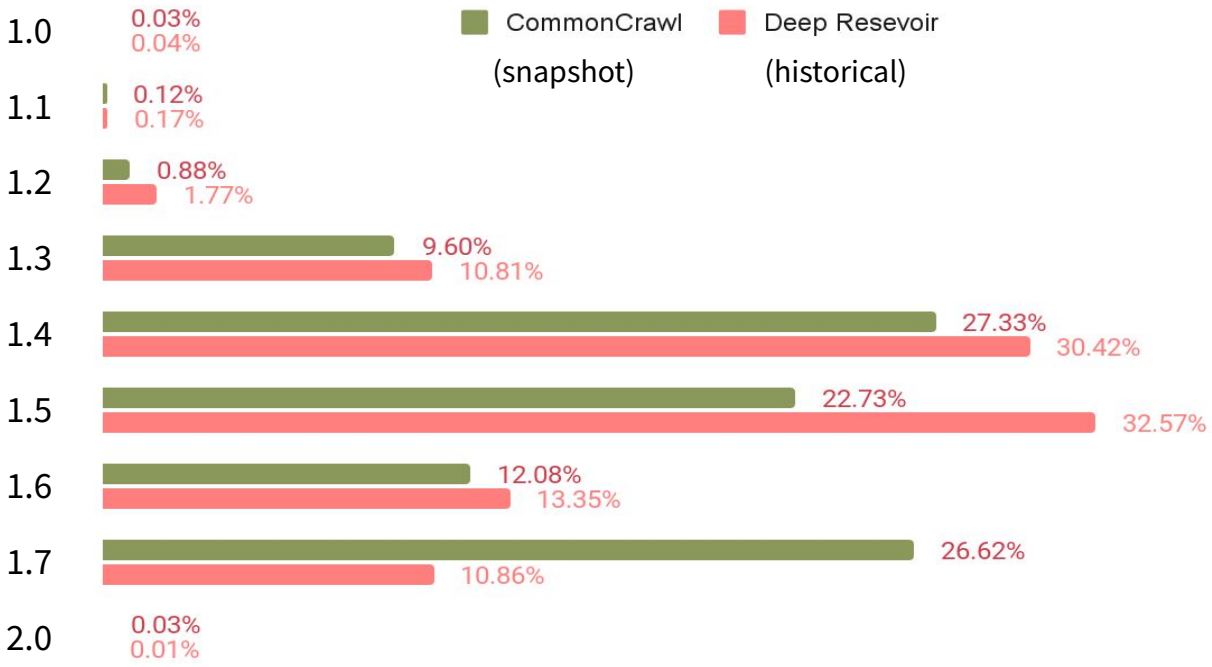
7,932,836 PDFs

Source: CommonCrawl 2021-August, Tim Allison (NASA)

Deep Reservoir

59,674,054 PDFs

Source: Adobe (2022-09)



A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Proportion of PDF files by file header version

www.pdfa.org

2021-08 CC

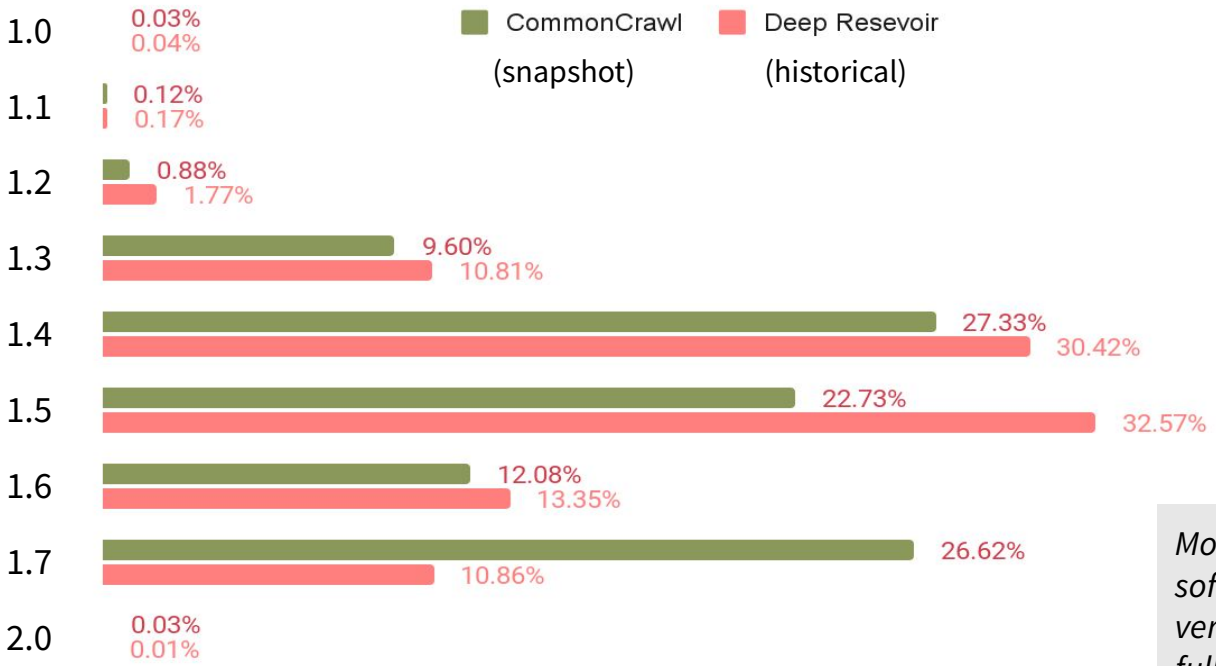
7,932,836 PDFs

Source: CommonCrawl 2021-August, Tim Allison (NASA)

Deep Reservoir

59,674,054 PDFs

Source: Adobe (2022-09)



Most PDF creation software uses the oldest version necessary to fully support the source document's features.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF demographics

As seen by a “well-known implementation”

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF association

www.pdfa.org

Unique users of Adobe's desktop software (1 month)

This data is presented thanks to Adobe.

Adobe collects data anonymously from desktop users who agree to allow their usage to be tracked.

	2016	2018	2022
Reader	243,688,400	338,801,280 (+39%)	177,500,000 (-48%)
Acrobat	3,609,235	7,758,253 (+116%)	4,847,200 (-38%)
Files opened	12.3 billion	19.5 billion (+59%)	10.0 billion (-48%)

Key dynamics since 2018

- Browsers have developed capable PDF viewers
- Users have moved to cloud-based workflows and file-storage
- A broad trend towards reading on mobile devices
- Other vendors' viewers have gained

BUT: Many large organizations block analytics.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

PDF association

www.pdfa.org

In the files viewed by Acrobat and Reader

This data is presented thanks to Adobe.

Adobe collects data anonymously from desktop users who agree to allow their usage to be tracked.

	2016	2018	2022
Image-only	23%	24%	22%
Bookmarks	11%	13%	15%
Secured	3%	3%	4%
Acroform	3%	4%	6%
Tagged	35%	37%	22%

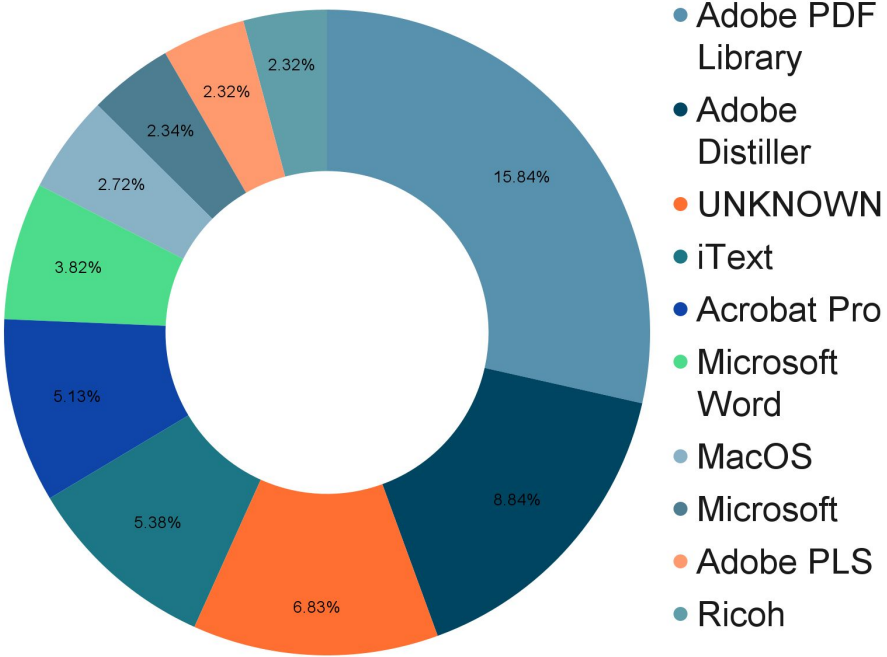
BUT: Many large organizations block analytics.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Top 10 PDF “Producer” opened by Adobe (2022)

This data is presented thanks to Adobe.

Adobe collects data anonymously from desktop users who agree to allow their usage to be tracked.



BUT: Many large organizations block analytics.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.

Thank you

The meeting-place of the PDF industry.

A PDF Association Presentation © Copyright 2022 by PDF Association, e.V.