

2032



Sensible

2023

History of the PDF

2008

1999

1995

1991

Read this article online →



01 A history of the PDF

1992

When was the last time you clicked a link, found you were opening a PDF, and didn't groan in pain?

The PDF is one of the most popular formats in the world – and has been for years – but it's also one of the most reviled. If you've seen the arrival and dominance of Microsoft Word, the rise of Google Docs across schools and offices, and the tides of companies and formats – Evernote and Notion; XML, Markdown, and HTML – the PDF likely stands out.



On the one hand, it's eerily well-supported.

You can open PDFs in your browser with or without an Adobe product and send them to others via Slack, iMessage, and more.

On the other hand, the PDF is decidedly anachronistic. PDFs are hard to read and edit on mobile devices. Worse, PDFs bear the increasingly surreal mark of clearly being a digital version of a physical object despite it being 2023.

The history of the PDF is three histories in one: a history of the PDF file format itself; a history of Adobe, the company that created the PDF and eventually released it as an

independent standard; and a history of the concept of the digital document – a concept that the PDF pioneered, exemplified, and eventually restricted.

The PDF is also a lens through which we can better understand how technologies evolve, die, and persist. This history is a jumping-off point for understanding how and why businesses always seem to lag behind technology advancements – from the paperless office that never really happened to the era of digital transformation that never seems to finish transforming.

By the end of this article, you'll have a greater appreciation of the cockroach-like format the PDF has proven itself to be. As ugly as it often is, it'll likely outlast all of us. And there's a lot of lessons to take from understanding why.

02 The PDF as an Idea

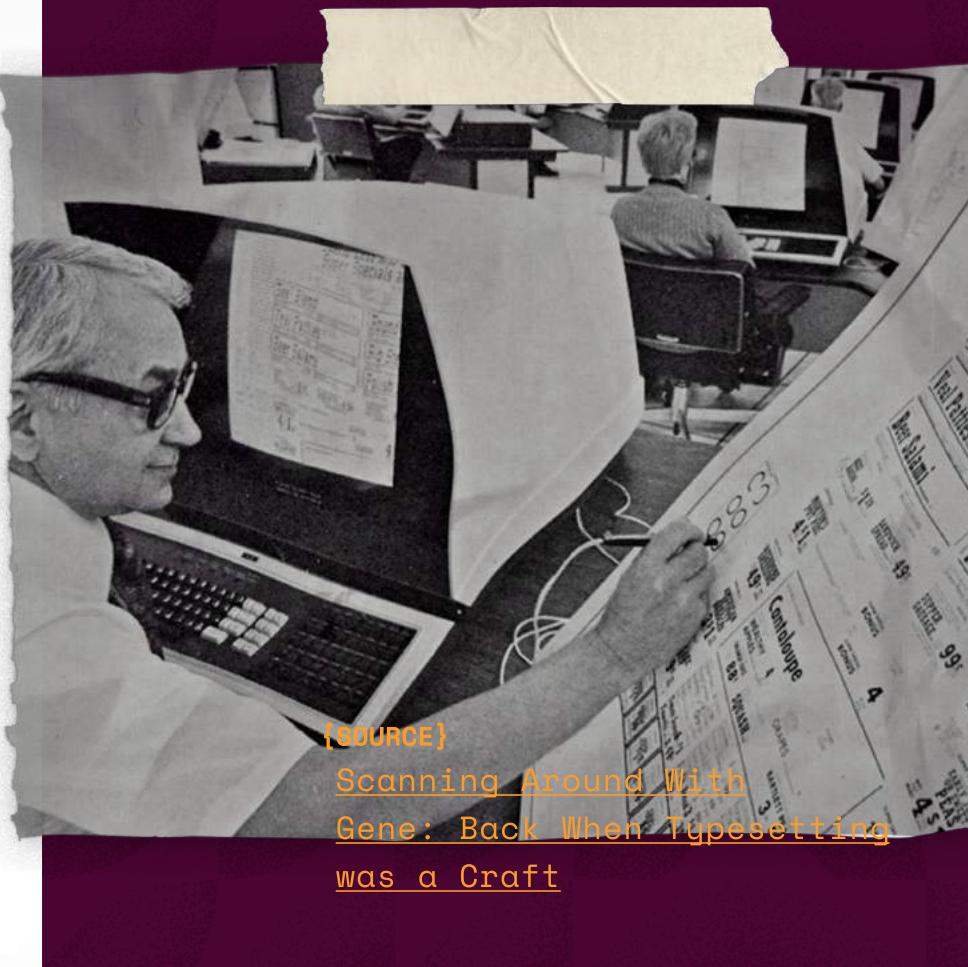
1992

We're starting at the beginning not because it's the beginning but because the first, grand promise of the PDF – made over three decades ago – has made the format last so long.

In the 1990s, business leaders were excited about a concept that was equal parts genuine innovation and buzzword: The paperless office. Enabling the paperless office, these businesses thought, would be the next major disruption, the next paradigm shift. And the company that heralded this shift would be rich.

The beachhead for this transformation was digital paper, and a crowd of companies were chasing it: DjVu, WordPerfect, Common Ground, and more. But Adobe, which announced the PDF format at a tech conference in 1992, eventually won (even though it turned out that this beachhead was more of an island and the office in 2023 still relies on paper).

The PDF won, among other reasons, because Adobe's founders started the company to create physical documents, not digital ones, and this background lent them the advantage they needed to make the ideal digital document.



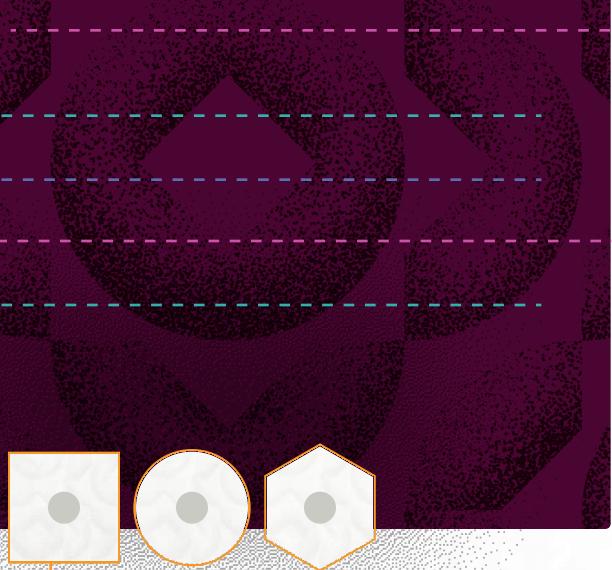
{SOURCE}

Scanning Around With Gene: Back When Typesetting was a Craft

We complain about ink cartridge prices and absurd printer DRM, but modern printers can at least reliably print a paper version of what we see on the screen.

But back then, people had to rely on a dot-matrix printer, including its screeching soundtrack and pixelated text, or enormously expensive typesetting machines.





Adobe's initial innovation was PostScript: a series of protocols that each desktop printer would carry and could, miraculously, render what the user wanted to print. PostScript debuted in 1985 on Apple's LaserWriter.

John Warnock, Adobe's co-founder, had a clear strategy – he wanted to make PostScript a universal standard. Resisting the norm of secrecy and private development at the time, Warnock pushed for openness. Years later, he said, "We had to publish it. We had to make it very, very open — because the trick was to get both [software] application developers and operating system developers to support it."

"The only way to make standards is to get them out and just compete," Warnock later said. Adobe got out there, competed, and a few years later, PostScript became the standard."

"At one point," Warnock said, "We had 22 PostScript competitors. There were 22 clones out there that were trying to undercut us in the market. And as far as I know, not one of them succeeded — including Microsoft's. I think they produced exactly one printer, of which they sold zero. It was a disaster."

By publishing a standard, Charles Geschke, Adobe's other cofounder, said, "You're taking the risk that someone will do a better job of implementing it. We had the self-confidence that we would always have the best implementation, and that has turned out to be true."

This strategy - outcompeting and then standardizing - became the blueprint for the PDF.

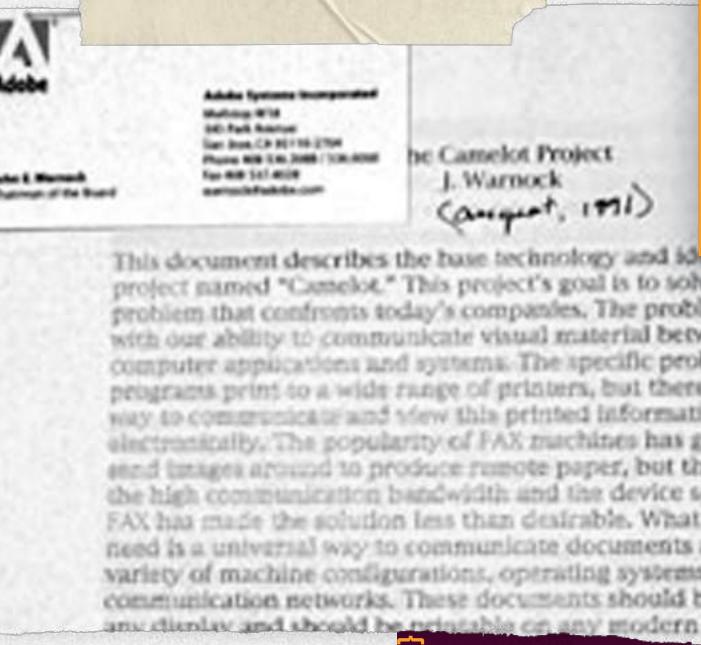


The mission that carried Adobe from PostScript through to PDF was to create, as David Parmenter, director of engineering for Adobe Document Cloud, put it, an “interchange format that preserved author intent.”

“Author intent” is the key idea here. Before the PDF, Mac, Windows, UNIX, and MS-DOS all interpreted files differently. If you were the rebellious sort and wanted to create a file in Windows but then move it to a Mac, your file “would likely have looked like Jackson Pollock got a hold of it.”

1991

The initial idea emerged in a paper written by Warnock called Project Camelot.



“This project’s goal,” he wrote, “is to solve a fundamental problem that confronts today’s companies.” The problem, he explained, was the lack of a universal way to “communicate and view printed information electronically.”

If documents could become viewable across all displays and printable across all printers, Warnock wrote, “the fundamental way people work will change.”

The vision exceeded the PDF, including “utilities, applications, and system software.” But the core idea that made the vision possible was that the PDF would be “completely self-contained.” It didn’t matter whether the receiving computer didn’t have the fonts the sending computer did. The PDF rendered the information as the author intended, regardless.

Warnock imagined a few possibilities as a result of the PDF, including the ability to send newspapers, magazine articles, and technical manuals over email and the ability to maintain databases of documents that people could access and print remotely. He imagined companies saving “millions of dollars in document inventory costs.”

{source}

Scanning Around With Gene:
Back When Typesetting
Was a Craft

Adobe started with two price points: a PDF-making program that cost about \$700 and a PDF-reading program (Acrobat Reader) that cost \$50.

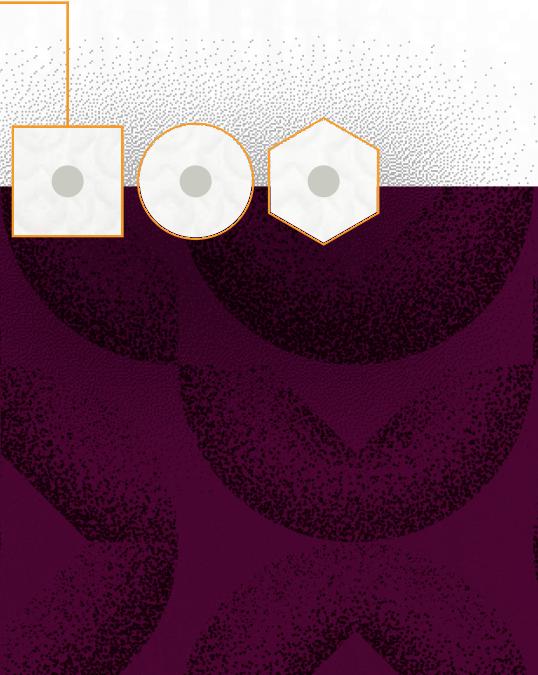
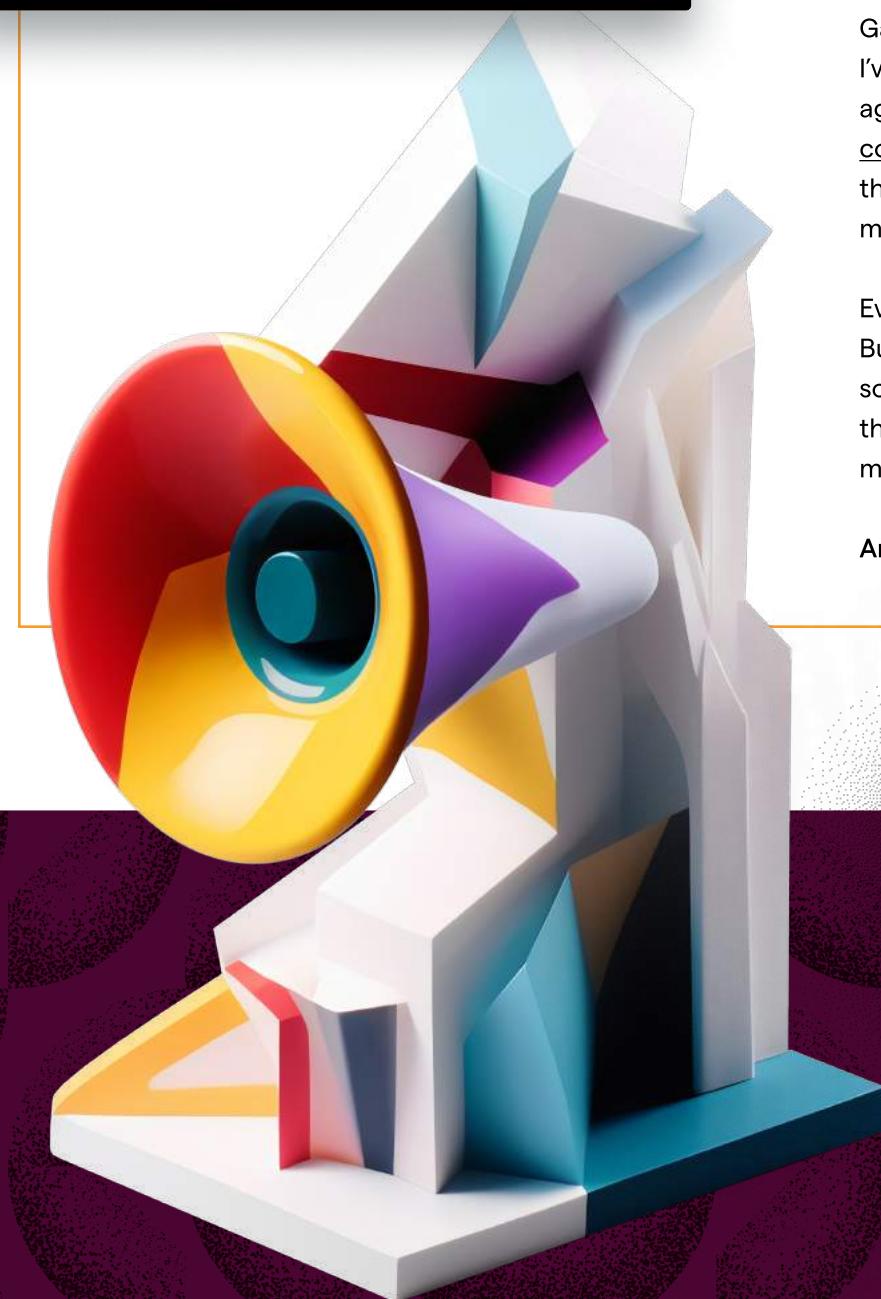
According to Warnock,
someone from Gartner told them,
“This is the dumbest idea I’ve ever
heard in my life.”

It was not a fast success. Reflecting, Warnock said, “When Acrobat was announced, the world didn’t get it. They didn’t understand how important sending documents around electronically was going to be.”

According to Warnock, someone from Gartner told them, “This is the dumbest idea I’ve ever heard in my life.” IBM executives agreed. James Fritz, who attended the conference where the PDF debuted, wrote that to many in 1992, the promise the PDF made was “heresy.”

Even Adobe’s board wanted to kill the PDF. But Warnock knew he had something good, something beyond good: “No one has to say this is a good idea or a bad idea. We can just make it a fait accompli.”

And they did.



On the day of the PDF's release, Adobe made the specs for the format freely available, and soon after, Adobe made its reader software free, too.

Rob Walker, senior writer for the business publication Marker, [explained the strategy well](#), writing that the company was “focusing entirely on the creation product as a revenue stream — but gambling that the more people who could read the format, the more attractive it would be to the creator side.” The PDF format would become a standard, and though this rising tide would lift many boats, none would rise higher than Adobe’s.

Of course, many other tides pushed the PDF higher, too. From the mid-1990s to the early 2000s, the Web became mainstream, and download speeds improved — making the already accessible, already compact PDF even more accessible and compact.

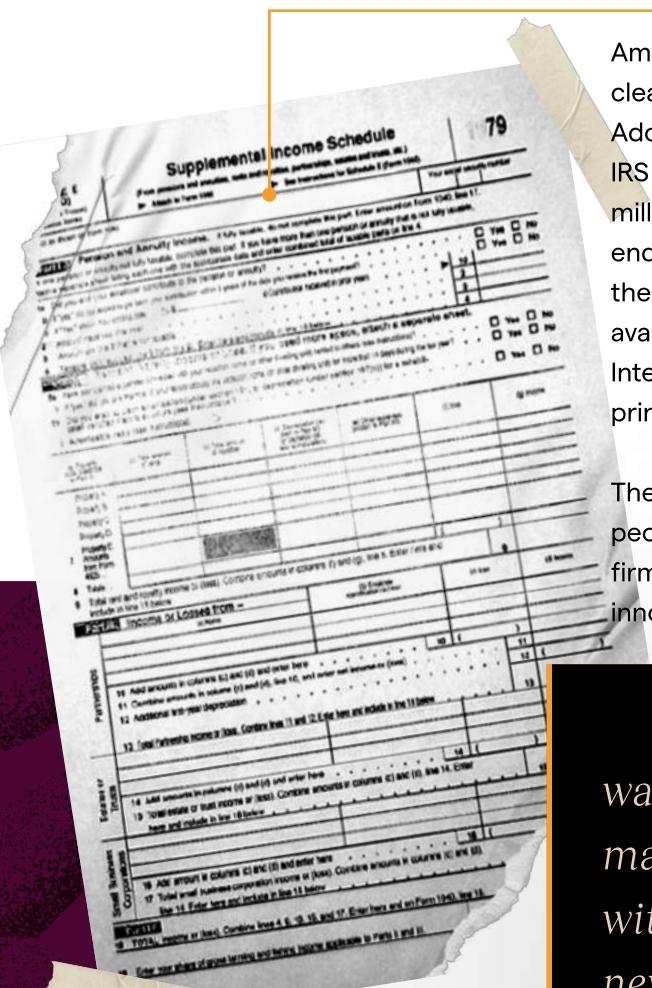
Amidst these trends, however, there’s still a clear pivot point: in 1996, the IRS became Adobe’s star customer. Before the PDF, the IRS was mailing tax forms to hundreds of millions of households, and the whole endeavor was complex and expensive. With the PDF, the IRS could make these forms available to the entire country via the Internet, and people could download and print them as they saw fit.

The IRS brought PDFs to everyone — average people, business leaders, academics, law firms, and more. It was a shift both innovative and familiar:

The magic of the PDF was that it truly was digital paper, and users could get many of the benefits of the Internet without having to parse a fundamentally new format.

{SOURCE}

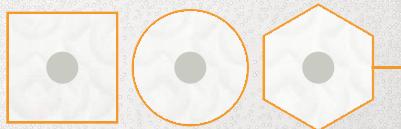
<https://creativepro.com/scanning-around-general-camelot-carousel-acrobat/>



Other tax software was available, including TurboTax and MacInTax, but asking your computer to do your taxes was a big leap for many people at the time. But downloading and printing forms? That they could and that they did.

In 1996, an Albuquerque journalist reported on the phenomenon: "If you need a form, forget about dragging yourself to the IRS office. Just point and click on the form on the Internet."

People could save a lot of time and energy – well, those with the Internet and desktop printers, at least – but this move was beneficial to the IRS, too. In a case study, IRS representatives write that "the agency saves millions of dollars annually by decreasing the money it spends on printing, storing, and mailing tax materials."



{SOURCE}

https://www.newspapers.com/article/17856587/first_article_about_irs_website/

"the agency saves millions of dollars annually by decreasing the money it spends on printing, storing, and mailing tax materials."

For example, say you are wondering if that old PC you donated to your church is tax deductible. You could read this online: Section 170(f)(8) provides that no charitable contribution deduction will be allowed under section 170(a) for a contribution of \$250 or more unless the taxpayer substantiates the contribution with a contemporaneous written acknowledgment from the donee organization.

Guess which tax code version this is?

Or you could drop by the sweet and simple tax code area online and read this: In 1993, Congress passed a law requiring people donating money or property to a charity to get a written receipt for each donation of \$250 or more. Donors must get the receipt before taking a deduction for the donation.

Let's all give the IRS a round of applause.

Flexibility is not one of the IRS's

Internet offers some easy and not-so-easy tax information

KIM KOMANDO



For the Journal

It's so very easy to lose track of time while you are working on a computer. At least that's what I tell the IRS when they ask me why my tax returns are late every year. Note: Don't use this excuse. It doesn't work.

There's not much you can get past the IRS and the federal agency we all love to hate is making it even tougher. Forget about using the lame-brain excuse you didn't have the right forms. Scratch you didn't know that although your kids attended a legitimate business dinner, their happy meals aren't tax deductible.

If you are depreciating a computer and also have an Internet account monthly charge as a business expense, you really better watch it. You see, the IRS put more than 600 forms and publications on the Internet for your browsing pleasure at <http://www.irs.ustreas.gov>.

As you might expect, despite its graphics and dry humor, the site is

you need as pdf files.

To use pdf files though, first get your free copy of the Adobe Acrobat viewing program. It allows you to view and to print the forms. Acrobat is available from the IRS site. You can also find Acrobat at Adobe's Internet home page at <http://www.adobe.com/Acrobat/free/read.html>.

Although you got the forms off the Internet, you can't submit your returns there, not yet anyway. The IRS is working on increased

allows it.

You can mail your old-fashioned way or electronically using software programs like TurboTax.

Make sure the program has the official IRS stamp so you can check this on the IRS website too.

Electronic filing isn't free, for example, charges

Turbo Tax users \$14.95. Still, if you are in for a huge refund, electronic filing may get your money back a little quicker.

You'll find tax help on the commercial on-line services too. America Online, CompuServe, Prodigy, and the Microsoft Network all have tax-specific areas.

While you're surfing on line for help with your taxes, here are some other Web pages for you to check out:

- The Income Tax Information: <http://www2.best.com/ftmexpat/html/taxsites.html>

- Text from J.K. Lasser's Your Income Tax 1995 book: <http://www.mcp.com/bookstore/jklasser/jklhome.html>

- Twenty-Five Most Common Tax Preparation Errors: <http://www.ey.com/us/tax>

- Tax Notes News-Wire: <http://205.177.50.2/news.htm>

- And lastly for some entertainment, How Would You

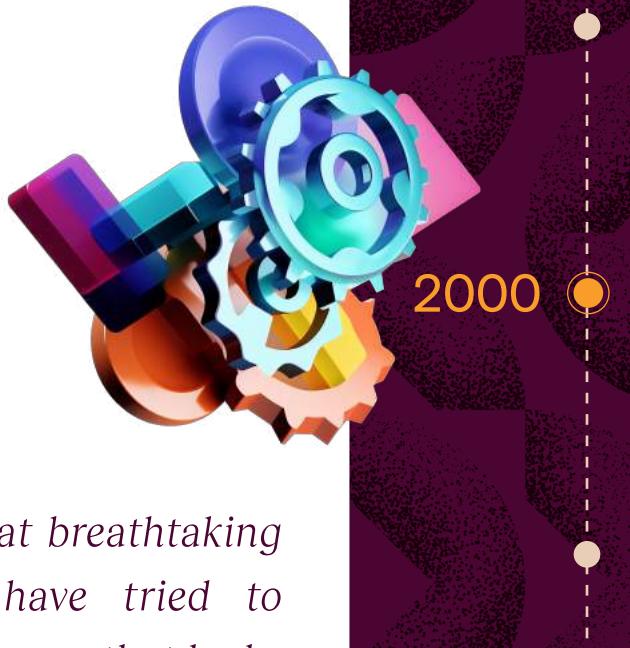
From here on, much of the progress was feature by feature. When Adobe released a plug-in that enabled Netscape users to view PDF files in the browser, adoption boomed. And when Adobe added the ability to link PDF files to and from HTML pages, the boom continued.

In 2000, Adobe released Acrobat 4.05, and by then, it was hard for anyone to dispute that the PDF had achieved the level of standardization Warnock and Geschke had pursued.

By then, people had downloaded over 100 million copies of Acrobat Reader, and even the industries that cared the most about preserving authorial intent – such as the graphics art and preprint industries – grew to accept the PDF. And their acceptance carried weight.

In 2001, The Wall Street Journal [reviewed Adobe Acrobat](#). Already, the PDF, which feels ancient today, left users with a sense of boredom that belied how impactful it had been and would be.

The reviewer wrote that the technology sounded about as exciting as a TV ad and that it “isn’t that breathtaking unless you have tried to design a Web page that looks the same regardless of the program it is viewed with, or you have sent a resume in Microsoft Word format to a potential employer only to discover that it doesn’t look quite as glitzy as when it left your desktop.”



[PDF] Isn't that breathtaking unless you have tried to design a Web page that looks the same regardless of the program it is viewed with, or you have sent a resume in Microsoft Word format to a potential employer only to discover that it doesn't look quite as glitzy as when it left your desktop.

But the reviewer extolled the benefits, writing that whether users created documents with “whiz-bang graphics” or simple text, Acrobat and the PDF could handle it. “This doesn’t sound like a huge leap for mankind,” he wrote, but it was a huge leap for businesses: “Indeed, most big companies already use Acrobat for exactly this purpose. But not enough do.”

This reviewer was right, but an even larger leap was coming soon.

Through the 1990s and the early 2000s, the primary strategy of Warnock, Geschke, and Adobe was to make the PDF a de facto standard. But in 2008, the company took a big step forward by making it an actual standard.

Adobe released the PDF format's specs to the independent nongovernmental organization International Organization for Standardization (ISO) and gave this body the royalty-free right to publish and control the patents and specs. Adobe maintained a seat on the ISO committee in charge, but otherwise, it stepped back from the PDF standard.

If the PDF was accepted before, it became undeniable afterward. "Once we made it available to everybody, there was a big halo effect," said Parmenter. Adobe built a natural association between itself and the PDF, but by making it a standard, Adobe could stand on the collective efforts of others, too. Microsoft Word added the ability to save Word documents as PDFs, and a flurry of other PDF-creation and reading tools emerged.

Adobe wasn't alone, but it didn't need to be by then—it was at the top.

Over the years, with the PDF remaining an independent standard that the ISO gradually iterated on, Adobe evolved and profited. According to Walker, "There's no question that the close association with the PDF has been vital to the long-running success of Acrobat, Adobe's document software."



The long run demonstrates success overall, but there were mistakes as well as victories.

As the Internet grew, Adobe both reaped some rewards and remained passive despite potential other rewards.

Better download speeds made the PDF more practical, but Adobe avoided working with HTML. According to Warnock, “The early versions of HTML — from a design point of view — were awful. There was nothing beautiful about it.” Geschke, the son and grandson of letterpress photo engravers, had similar sensibilities. And Adobe suffered for it.

But the victories were even greater.

There was “Liquid Mode” in 2020, an improvement that better adapted the format for readability on smartphones. Around the same time, Adobe made it easier for developers to embed PDFs into websites. These features pale in comparison, however, to how well Adobe survived the transition to the cloud and SaaS eras.

By 2020, Adobe’s Document Cloud offering – now central to Acrobat – had revenues of \$1.5 billion. And the COVID-19 pandemic, which ruined or damaged so many other businesses, boosted Adobe. According to a Forrester study, companies increased their spending on digital document processes and tools by more than 50%, leading to the share price of Adobe rising from \$333 to more than \$500.

Warnock—who was CEO until 2000 and chairman of the board (along with Geschke) until 2017—pushed a vital idea at Adobe, one that failed them when it came to HTML but helped when it came to smartphones and the cloud.

2020

“Companies build antibodies,” Warnock said.

“They build resistance to change. They get comfort zones where they want to work, and employees don’t want to try something new for fear that they are going to fail. So, they reject ideas. One of the hardest things about keeping a company innovative is killing off the antibodies and forcing change.”

But to the extent that Adobe successfully killed its internal antibodies, it profited mightily by – intentionally or not – introducing a format that came laced with its own antibodies, a format that has staved off change, challengers, and killers for decades.



05 The PDF as a zombie

In the intro, we shared a common reaction to opening a PDF in 2023: a groan, an eye-roll, a pained sigh. But this sentiment doesn't seem to affect the company or the format.

2023

Over the years, no one has taken the throne. Microsoft, a similar standard-bearer, has faced challenges from Google and Apple toward Word and PowerPoint, but a PDF challenger – much less killer – has yet to emerge.

Adobe reports that in 2020, about 303 billion PDFs were opened using its Document Cloud products. This popularity represented an annual increase of about 17%, and even then, this rate doesn't reflect the total amount of PDF usage due to its ISO-based standardization.

The persistent success and growth of the format comes from its original design: The PDF was designed to be compact and forward-compatible and to reflect the intent of the author across devices.

Does the PDF feel anachronistic? Yes, of course. But is that a bad thing? The printed book has lasted from the 15th century, and the Adobe founders, directly inspired by book printing, created a format meant to have a similar legacy.

But of course, it frequently is bad for users and businesses. Comments on places like HackerNews refer to it as “one of the worst file formats ever produced,” “soul-crushing,” and something that “should really be destroyed with fire.”



Hacker News

This sentiment, however, isn't a case of new users and developers not respecting their elders. In 1996, the research-based user experience group Nielsen Norman [criticized the PDF format](#). They were not wholly against PDFs, but they wanted the PDF to stay in its lane as digital paper and not encroach on the Web, where HTML remained the better format.

"PostScript and Acrobat files should never be read online," writes Jakob Nielsen in 1996. "PostScript viewers are fine for checking out the structure of a document in order to determine whether to print it, but users should not be tricked into the painful experience of actually spending an extended period of time with online PostScript."

Nielsen [restated the case](#) in 2001, writing, "PDF is great for distributing documents that need to be printed. But that is all it's good for. No matter how tempting it might be, you should never use PDF for content that you expect users to read online."

2020

"PDF is great for distributing documents that need to be printed. But that is all it's good for. No matter how tempting it might be, you should never use PDF for content that you expect users to read online."

Nonetheless, the PDF kept growing in popularity, and few limited how and where it was used. In 2020, Nielsen made [the case again](#), writing, "After 20 years of watching users perform similar tasks on a variety of sites that use either PDFs or regular web pages, one thing remains certain: PDFs degrade the user experience."

He couldn't be clearer – "PDF should never be used for on-screen reading. Don't force your users to suffer and slog through PDFs!" – but the lesson went unheeded.



In another [2020 article](#), Nielsen captured user responses that likely reflect some of the experiences you've felt yourself:

Information is outdated in those PDFs. So you're getting stuff that isn't current. They just haven't taken those links off.

I don't know if they [a PDF with email-signature templates] are updated. I can't confidently share it. Sometimes there are multiple versions of PDFs

All of the PDFs are horrible. There are so many old forms and version control is so difficult. We're starting to move them into a database but first have to audit them and track down people to ask them if they still need the form. We're taking the top-used forms and tackling those first.

We've come across problems with PDF forms. Others have to download the form in order to use it the way we want them to use it. You have to download it to get the features to work, so we always have to specify at the top of our documents that our partners are using these and they might not have the latest PDF readers.

Ultimately, the article's title makes the most forceful case: "PDF: Still Unfit for Human Consumption, 20 Years Later."

Of course, humans kept consuming PDFs, ranging from your average person trying to parse a PDF restaurant menu to the highest levels of Federal power.

In 2018, *Slate* reported that PDF usability was a significant reason why then Special Counsel Robert Mueller was able to indict Paul Manafort as part of the investigation into President Trump's ties to Russia.

Manafort had tried to defraud a potential lender by altering a profit-and-loss statement. Manafort emailed the PDF to an associate and asked him to convert it to a Word document so Manafort could make the fraudulent changes. Once he made the changes, Manafort's associate helped him convert the Word document back to a PDF.

income by more than \$4 million. The doctor...
MANAFORT by more than \$3.5 million. To create the false 2016 P&L, on or about October 2016, MANAFORT emailed GATES a .pdf version of the real 2016 DMI P&L, which showed a loss of more than \$600,000. GATES converted that .pdf into a "Word" document so that it could be edited, which GATES sent back to MANAFORT. MANAFORT altered that "Word" document by adding more than \$3.5 million in income. He then sent this falsified P&L to GATES and asked that the "Word" document be converted back to a .pdf, which GATES did and returned to MANAFORT. MANAFORT then sent the falsified 2016 DMI P&L .pdf to Lender D.

But as the PDF Association – yes, that's a thing – points out, *Slate* missed a detail: "Converting from PDF to Word for the purposes of surreptitiously altering text in the PDF document is a foolish way to commit fraud and break federal law at several levels" because the Word file won't perfectly resemble the original file and because PDF files are already editable.

"Manafort could have readily altered the PDF himself," the Association writes. "Had he done so, he would have avoided a key part of the paper trail that may land him in federal prison. He probably even had a PDF editor already on his computer."

{SOURCE}

<https://pdfa.org/federal-crimes-edition/>

06 The PDF as digital document

Over the decades, much of the frustration with the PDF has emerged because the format has subtly and slowly shifted from functioning as digital paper to functioning as digital documentation.

The PDF is a perfect format for paper made digital. Though we haven't reached the paperless office future, the need for paper has diminished, and the need for documentation has increased.

We have much more to document – think of all the SaaS contracts a business maintains, all the regulatory compliance work that needs to be written down, and all the processes for hiring, working, and communicating across offices, co-working spaces, and home offices – but we need documents to do so much more.

Documents were once outputs. Originally, PDFs outputted authorial intent for the sake of reader consumption via printer and digital paper. But over time, PDFs took on the role of inputs, too. As perfect as PDFs were for display, they became bad ways to store information and terrible ways to facilitate the interface between different functions and parties.

The effort to programmatically extract information from PDFs demonstrates this format is poorly suited for its modern needs. FilingDB, a company later acquired by Insig AI, has written in-depth about the struggle of extracting information from a format that was never really meant to serve as a medium for storage or interface.

A few examples include:

- Read protection (PDFs often have several access permissions flags that limit how content can be copied).

{SOURCE}
<https://web.archive.org/web/20230107085714/https://db.com/b/text-action>

SEB's core values

Customers first

We put our customers' needs first, always seeking to understand how to deliver real value.

Commitment

We are personally dedicated to the success of our customers and are accountable for our actions.

SEB's core values serve as the foundation for the bank's ways of working and culture, and in combination with the bank's vision – to deliver world-class service to our customers – they serve to motivate and inspire employees, managers and the organisation as a whole. These values are described in SEB's Code of Conduct, which provides guidance on ethical matters for all employees.

- Hidden text (PDFs frequently contain text outside the page's bounding box that's invisible to most PDF viewers, but that will show up during extraction).
- Too many and not enough spaces (PDFs often have extra spaces between letters in a word or too few spaces – usually for the sake of kerning).

- Embedded fonts (PDFs, meant initially to ignore font restrictions, sometimes have custom encoding and fonts that look fine to human eyes but confuse machines).
- Layout confusion (PDFs, always designed for humans first, often have layouts that a human might find readable but that can leave a machine bewildered, such as footnotes, asides, and varying column layouts).

“The main problem,” FilingDB writes, “is that PDF was never really designed as a data input format, but rather, it was designed as an output format giving fine-grained control over the resulting document.”

At the deepest level, they write, “The PDF format consists of a stream of instructions describing how to draw on a page [...] As a result, most of the content semantics are lost when a text or word document is converted to PDF - all the implied text structure is converted into an almost amorphous soup of characters floating on pages.”

Hacker News commenters [reacting to the article](#) wrote about the PDF in a much blunter fashion. But as the discussion continued, commenters also circled the primary problem, with one commenter writing that “Parsing pdf to extract data is like using a rock as a hammer and a screw as a nail” and another writing that “Actually, parsing text data from a pdf is more like using the rock to unscrew a screw, in that it was not meant to be done that way at all” and another still writing that “It’s closer to using a screwdriver to screw in a rock. The task isn’t supposed to be done in the first place, but the tool is the least wrong one.”

The PDF has outlived itself in many ways, but the revolution it created on the digital paper and digital document levels has had staying power that outstrips Warnock, the ISO standard, and Adobe itself. The PDF was built as a way of preserving an author’s aesthetic intentions, but software has eaten the world, APIs have eaten software, and information demands to be programmable, not beautiful.

The great irony is that the software and API revolutions hardly touched digital documents, which remain among the most important ways to communicate, store, and act on information in businesses worldwide.

What is a legally binding document, if not an API that connects an entity to a deliverable? And yet, even though a user can sign up for a service online, the enterprise version of that service will likely be codified in a PDF.



07 A future of the PDF

In 1991, a New York Times review of Adobe Acrobat – then called Carousel – touched on a future yet to come.

"If it succeeds," the reviewer wrote, "Carousel will alter the way computers are used in offices. Today, these machines are used primarily to create documents in word processors and spreadsheets. In the future, computers will increasingly be used to search for and view information."

"In the future, all documents might become information databases," the reviewer continued, and Adobe "could create a new market for corporate information systems."



This review was right when the PDF became a de facto standard; it was right when it became a real standard; it was right when PDFs became a mechanism for information storage; and it was right when businesses found themselves struggling to extract the information contained inside PDFs and turn digital documents into the interfaces and programs they needed them to be.

The future of the PDF remains unclear, but if the past decades have taught us anything, it's to bet on its survival and not its defeat. But that doesn't mean the PDF – the file format itself or the broader swath of digital documentation it represents – won't face disruption.

Adobe estimates that there are more than 2.5 trillion PDFs in the world today. As hard as it is to imagine a new technology finally dislodging the PDF, it seems just as hard to imagine no one ever finding a way to capitalize on and transform this market.

History of the PDF