

# *Publication Date:* *September 11, 1998*

FOUR REASONS PDFS ARE EVERYWHERE

BY LISA GITELMAN

---

**In all of our pocket histories and self-congratulatory palaver about the antiquity of “print” and the triumph of digital forms, we’ve neglected the history of page images.**

---

It was September 11th. Office workers and wonks remember just where they were when it happened. It was 1998. That’s right, three years to the day before the attacks of 9/11, before the so-called war on terror could be declared, and long before Harvey Weinstein even started to get what he had coming, the Government Printing Office (GPO) published the Starr Report. Officially called *Communication from Kenneth W. Starr, Independent Counsel, Transmitting a Referral to the United States House of Representatives*, it detailed the case against then President Bill Clinton. He was supposed to have been a corrupt Arkansas politico (Whitewater), a sexual harasser (Paula Jones), or the victim of a vast right-wing conspiracy, depending on whom you believed, but now it turned out that maybe he was just a run-of-the-mill philanderer, albeit one who may or may not have obstructed justice, perjured himself, and suborned perjury. The lengthy investigation into the failed real estate deal known as Whitewater had gone nowhere; the case brought by Jones had been delayed, dismissed, appealed, and then settled (\$850,000) with no acknowledgement of wrongdoing; but here was something else. Shaken loose along the way was the Monica Lewinsky affair. Clinton had been dicking around in the Oval Office. Had he also committed impeachable high crimes and misdemeanors?

After eight long months of work on the Lewinsky matter, Starr’s report arrived at the House of Representatives on September 10th, catching members of Congress off guard. Two vans pulled up outside the Ford

House Office Building where thirty-six boxes of material were unloaded by Capitol Police and put under lock and key. According to the *Washington Post*, House officials also reached out to Starr's office to request the report on computer disks. The next day the House would debate and then pass House Resolution 525, agreeing—even without reading it—to make Starr's 453-page narrative public immediately, while holding back all of the supporting materials for review. Already the morning *Post*, citing "sources," had published leaked details from the narrative ("cigar as a prop in a sex act," etc.). Already the GPO was poised to print the report and post it to the web. Already the Library of Congress, the House, and the House Judiciary Committee were also prepared to post it to the web. Already government websites were clogged with traffic. Already journalists had staked out and reserved machines at a nearby Kinkos shop, where a precious early hard copy of the report was promised. The whole nation seemed poised to read.<sup>1</sup>

So this is what publication looked like in the fall of 1998: After H.Res. 525 passed, the GPO received one version of Starr's report in its original WordPerfect 6.0 format and another version imperfectly converted to HTML by the House clerk's office. The HTML file was online within thirty minutes, according to a later press release by the GPO. By one estimate web traffic that day resulted in an 89% hang rate among users trying to access the report at .gov sites. House members also gave some diskettes of the HTML to reporters, and news outlets began to republish and report on the material. Meanwhile the WordPerfect version was computer typeset and then printed by the GPO as a 221-page government document. Five hundred loose-leaf copies were delivered to the House by 6 pm that day, and an edition of thirteen thousand bound copies was ready on Saturday, September 12th for distribution to Congress, distribution to federal depositories, as well as for sale to the public. The retail price was \$14, and there was a line stretching down the block when the GPO bookstore opened. A corrected HTML version went up on September 12th too, fixing the footnotes and other incidental errors caused in the conversion from WordPerfect.<sup>2</sup>

---

1. Susan Schmidt and Peter Baker, "Impeachment Report Contends Clinton Lied, Obstructed Justice; Alleged Deceit is Outlined" *Washington Post* (11 September 1998) A1; Linton Weeks, "Starr Report's Net Challenge; House Scrambles for Online Release" *Washington Post* (11 September 1998) A38.

2. GPO Press Release (18 September 1998) gpo.gov; estimate of hang time is from Kurt Foss, "PDF Version of Starr Report Nets Place in Global Communications History," *Planet PDF* (16 January 2003) planetpdf.com. See also David Kravets, "Sept. 11, 1998: Starr Report Showcases Net's Speed" *Wired* (11 September 2009) wired.com; and Fedwa Malti-Douglas, *The Starr Report Disrobed* (New York: Columbia University

The publication of the Starr Report was a drama—comedy or tragedy, take your pick—staged in multiple acts by a huge company of players with recourse to many props. Did it start with the initial posting of the imperfect HTML version? Or would it be more accurate to say that it started with the two vans, the boxes and then the computer disks that issued forth from the independent counsel’s office and headed toward Congress? And when did it end? With thirteen thousand printed copies in inventory, or after that, when the GPO bookstore actually opened to the public? Perhaps this event of publication extended even further still, since a PDF version of the Starr Report would be made a bit later, apparently by scanning a printed copy. Today you can read the corrected HTML version at [Washingtonpost.com](http://Washingtonpost.com) and elsewhere; you can buy a mass-market paperback version on Amazon for cheap; and you could also go find the report (print or microfiche) at a federal depository library. The GPO now makes a .txt version and the PDF—dated 28 September 1998—freely available on its website, as it does other government documents.

The very idea of publication hypothesizes instantaneous availability. A book, a movie, a song, a newscast is public upon release, as if everyone everywhere were suddenly and simultaneously in the know. Hypothesis or conjecture enters in partly because of the logistics of production and distribution and partly because the public that is being addressed is rarely paying much attention. Books languish and get remaindered, movies flop, someone flips the channel. But there are moments of collective cognition, when the operative fantasy of instantaneous availability seems almost to come true, and for all of its messiness and attenuation the publication of the Starr Report was one of those moments. In this it was reminiscent of old-fashioned daily newspapers in their heyday, those grey ladies that Benedict Anderson has described as “one-day best sellers,” only the Starr Report sold well for weeks in the GPO edition and editions rushed into print by other publishers. It was certainly also reminiscent of galvanizing broadcasts, one of Franklin Roosevelt’s fireside chats, say, election returns, the Super Bowl, or the much-anticipated final episode of *Seinfeld* that had aired that May, only with an additional dose of prurience (the semen-stained blue dress, etc.) and perhaps half a dose of Schadenfreude. “Within an hour of the report’s release, thanks to the Internet,” the *Post* reported, “the nation began history’s first simultaneous reading of smut.” It was like something

out of a Magic Realist novel, in which everyone everywhere is reading the very same book on the very same day; only this was also somehow about the internet.<sup>3</sup>

The *Post* at the time called the internet “the mammoth network of computers [that] is playing a greater and greater role in American politics.” Congressman Solomon (R-NY), the author of H.Res. 525, was less clear on this point, since according to the *Congressional Record* he said that the Starr Report would be “available to the Internet and other Websites” as soon as possible after his resolution passed. Certainly the internet itself was and remains part fantasy, a great gathering of users engaged in the imagination of their own interconnectivity, true not least for those many Americans who waited helplessly that day as their browsers tried and failed to load the Starr Report.

In retrospect the whole episode can be framed as a proof-of-concept moment: yet another lesson in the arrival of the internet, the importance of bandwidth, and the emergent utility of online publication. But it was also a lesson in formats—print formats and electronic ones—calling attention to an array of technical variables and structural conditions that help to determine how publication works. As such this was also a reminder that fantasies related to publication are also fantasies related to politics. If American democracy is built partly upon the ideals of an informed public and an unfettered press—that is, on real news, not fake—it must also depend partly upon transparent governance and a self-documenting state apparatus. Kudos to C-SPAN, then, for its soporific live coverage and its video archive, but the Starr Report is a closer cousin to the *Congressional Record* than to any broadcast. Even if its initial publication was broadly cast, that is, it was and remains a document, part of the enormous and ongoing documentary record of the United States government.

As a lesson in formats, the publication of the Starr Report offers a particularly good opportunity to think about what documents are and what they have become in the era of electronic publication. Paper is old hat, yet the document survives and even flourishes. Whether documents today are paper or not, the idea of paper persists as part of the way that documents function. This is clearest in relation to the *page image* as an interface con-

---

3. Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso, 1983) 35; Marc Fisher and David Montgomery, “Public Reacts to Details with Anger, Amusement,” in *The Starr Report: The Findings of Independent Counsel Kenneth W. Starr on President Clinton and the Lewinsky Affair with an Analysis by the Staff of the Washington Post* (NY: Public Affairs, 1998).

vention: those onscreen pictures of pages that make the PDF file such an evident descendant of microfiche, microfilm, and even of photocopies and facsimiles. WordPerfect and HTML files also open as onscreen pictures of pages, to be sure, but these do not count as page images because they are not stable across platforms or devices; like .txt files, they are great for transmitting alphanumeric content but not for transmitting form. Page images are ideal for documents because preserving the form of a document—its layout and typographic design—is both an important signal of authenticity/authority and also a crucial enabler of shared citation, allowing readers to arrive intuitively “on the same page” as one another since their copies are presumably all alike. Page images—PDFs, microforms, photocopies, and facsimiles—are intuitively self-identical with the pages they represent and reproduce, stable carriers of content *and* form. This is the first reason why PDFs are now so useful and ubiquitous. It may have taken the GPO a week or so to publish a PDF of the Starr Report, but already in 1998 Adobe’s PDF format had “strong government use,” according to one expert, since the format was so good for viewing, handling, and printing multipage documents.<sup>4</sup>

In all of our pocket histories and self-congratulatory palaver about the antiquity of “print” and the triumph of digital forms, we’ve neglected the history of page images. It’s a history that can deepen and enrich our understanding of what we call print, especially since so many modern techniques of printing involve page images. When the GPO used its computer composition system to typeset the independent counsel’s WordPerfect files, for instance, it didn’t set old-fashioned letterpress type for relief printing, instead it created galleys, page images, formatted for proofing against the original. The corrected galleys were then used to create etched metal plates for offset printing, either by the old labor-intensive photographic process (shoot film and “strip” negatives, “imposing” them onto carrier sheets for a plate-making machine) or by means of the new computer-to-plate system that the GPO had announced in April that it was acquiring. The plates were then used to print five hundred copies (for the House) plus thirteen thousand copies (for everyone else) using the GPO’s battery of giant web offset presses.

---

4. Anne R. Kenney, “Digital Benchmarking for Conversion and Access” 24–60 in Anne R. Kenney, Oya Y. Rieger, ed. *Digital Imaging for Libraries and Archives* (Mountain View, CA: Research Libraries Group, 2000) 51.



Looked at one way, page images were a necessary intermediary step in the printing process, since the computer-generated galleys pictured the form that the ultimate printed document would take on paper. Looked at another way, page images were the end result of printing processes like this, since the offset presses effectively reproduced pictures of pages. The pictures had been etched into metal plates and were then offset onto giant rolls of paper to be cut into pages and assembled into booklets. Complicated, yes, but this is the second reason why PDFs are so useful and ubiquitous: Page images are handy in workflow. Today *any and every printed book or ebook is almost certain to have been an intermediary PDF somewhere in its production process*. And while PDF page images have become a routine element of publishers' workflow, we know that they are also very useful as part of our own personal workflow, as printable derivatives, like the page images we download from digital archives (databases) of high-resolution TIFFs, for instance, or like the parts of an Excel spreadsheet we capture for future reference. TIFFs are much higher resolution, spreadsheets can keep accounts, but PDFs are great for doing some of what we need to do with both of these formats in the everyday course of white-collar labor.

The history of page images can complicate our understanding of publication processes, then, so it can enrich our sense of what we call print. But the history of page images is also full of fascinating false starts and dead ends. Notable among these are broadcast newspapers (radio facsimiles) of the 1920s–1940s: technically possible but commercially unsuccessful. Imagine broadcasting through a scanner instead of a microphone, using the airwaves to communicate with home receivers that print facsimiles. No one remembers the Sacramento *Radio Bee*, an experimental newspaper broadcast to 580 homes on a voluntary, trail basis.<sup>5</sup> But the *Radio Bee* does hint at the third reason why PDFs are so useful and ubiquitous today: They are designed to make page images easily transmissible across networks, even in low bandwidth environments. It's not just that the Starr Report would have been downloaded with more success on September 11, 1998, if it had been a PDF—it certainly might have—it's that the networked circulation of PDF page images has been part of a massive change in the ways that documents are handled and thus how they operate within and beyond organizations. The simple email attachment is a radical innovation, while

---

5. Jonathan Coopersmith, *Faxed: The Rise and Fall of the Fax Machine* (Baltimore: Johns Hopkins University Press, 2015) 58–61, 81–85.

online document-retrieval is even more of one. Not only have faxes and filing cabinets largely disappeared from American offices, so have interoffice envelopes, snail mail, and diskettes. The history of PDF has been partly the history of paper and paperwork, of course, of envelopes, file folders, and boxes. But it has also been partly the history of transportation (for example, vans) and labor (for example, van drivers). Sacramento may not have been ready to restructure itself around the broadcast *Bee*, but much indeed has been restructured around PDFs and the networks that transmit them.

The fourth and final reason why PDFs are so useful and ubiquitous today has less to do with how they contrast with other digital file formats or how they continue the history of page images; it has more to do with how they contrast with paper documents. PDFs are (mostly) searchable. Today that means they are findable by Google, but in September 1998 it meant that—like HTML or WordPerfect files—you could search them for “sex” without having to skim every page. Or with the *Post* as your guide, you could search for “cigar,” parachuting in to several mentions of cigars and cigarettes. It was “reading” by a whole new route, more prurient for its parachuting quality somehow, disabling of reflection and enabling of snark. This is an important or perhaps simply an ironic detail. After all, legal proceedings—including impeachments—do not turn on the occurrences of certain words or character strings; they turn rather on what words mean and what was intended by their utterance. Case in point, the Clinton impeachment famously involved, as he put it, “what the meaning of the word ‘is’ is.”

Rereading the Starr Report today, like researching its publication history, means travelling back in time to a simpler era. Pundits decried a decline in public morals, of course, and the president was an embarrassment, but we know now where things were headed on both counts. Observers were wary of “spin” in 1998, but no one had yet broached “alternative facts.” *Seinfeld* (with its self-indulgent trivialities) had ended, while *The Apprentice* (with its pathological narcissism) had not yet begun. Clinton was ultimately impeached by the House and then acquitted by the Senate. History must judge him harshly, especially now that we have passed the Weinstein marker. Documents like the Starr Report remain crucial, and they still pile up at dizzying rates at the GPO, sensible—that is, historical—according to norms of publication for which the PDF format has proved such a successful support. Documents like Barak Obama’s birth certificate, by con-

trast, arise and persist in public consciousness according new norms, those hatched in the id-infested cauldron of social media, where alternative facts (that is, lies) boil and bubble.

---

LISA GITELMAN is a media historian whose research concerns American book history, techniques of inscription, and the new media of yesterday and today. She is particularly concerned with tracing the patterns according to which new media become meaningful within and against the contexts of older media. Her second book, *Always Already New: Media, History, and the Data of Culture*, was published by the MIT Press in 2006. She has a recent edited collection, *"Raw Data" Is an Oxymoron* (MIT 2013), and a recent monograph: *Paper Knowledge: Toward a Media History of Documents* (2014).