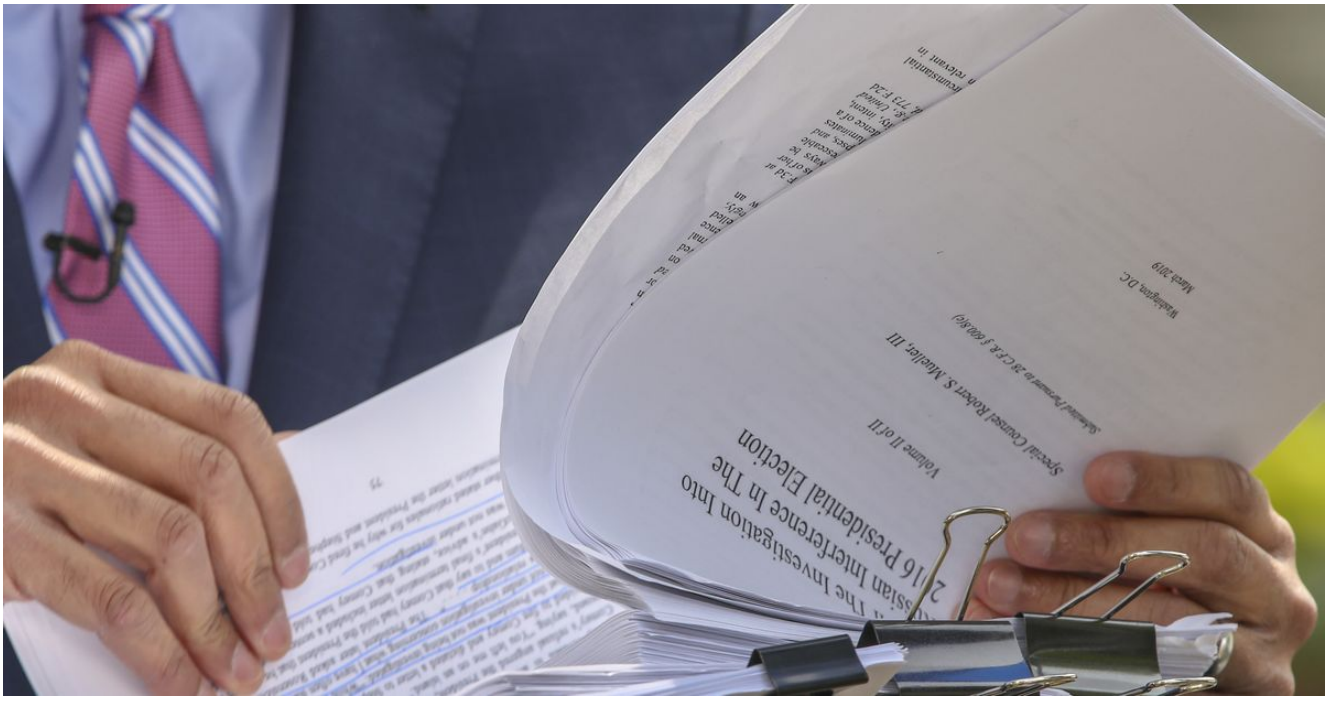
The PDF Association thinks the Mueller report sucks



When Special Counsel Robert Mueller released his report on Russian interference in the 2016 US presidential election yesterday and political wonks across the internet rushed to download it, many people noticed two things: you couldn't search for any text on the pages, and the whole file was *really, really large*. If you were annoyed by either of those things, you probably weren't nearly as ticked off as the PDF Association, which published a long explanation of just why the Mueller report PDF file was so bad.

"A Technical and Cultural Assessment of the Mueller Report PDF" is both an indictment of the Justice Department and a celebration of the venerable Portable Document Format. It starts with some basic facts: the 448-page document is "of acceptable quality," but it doesn't conform to archival standards. It was produced on April 17th on "probably a typical office network copier/printer," and it uses lossy compression "more appropriate to photographs than to text." The Justice Department might have gotten a high-quality PDF

from Mueller, printed it, and re-scanned it, or Mueller might have delivered a paper report that the department scanned and released.

PDF is the only format with "the cultural and technical requirements" for a document like this

As the post notes, re-scanning makes absolutely sure there's no inappropriate text data released, limiting people to the words they can see and the black redacted boxes. But it inflates the file size and makes the text unsearchable, unless people run it through their own optical character recognition software, a process that won't be as accurate as scanning the original source file.

Badly redacted searchable PDFs have occasionally revealed embarrassing secrets. A court filing in a Facebook lawsuit left text selectable under some hastily drawn black bars, and the Proud Boys extremist organization accidentally revealed its leadership with the same technique. Professional redaction software can prevent this, however, and the PDF Association post notes that an untagged and unsearchable PDF could violate the Justice Department's accessibility rules for people with disabilities.

Since this *is* the PDF Association, there's a detailed explanation about why the format is great and why "no one would have even suggested a Word file, or a set of TIFF images, or a website, or an XPS file, or EPUB, or plain text." The short answer is that PDFs preserve the original text and formatting of a document, they can include clear redaction, and they are supported by many platforms. "PDF is the only document format capable of carrying the cultural and technical requirements for important communications in the modern age," the post says.

The association just isn't too happy about the Justice Department's insult to this most honorable of file formats. As one expert puts it, the document is "really kind of sad." For

shame, Mueller report PDF. You have disappointed us all.

https://www.theverge.com/2019/4/19/18508148/pdf-association-mueller-report-scanning-redaction-sad-formatting