

## Introduction

Almost immediately after the release of Barack Obama's "long-form" birth certificate PDF, pundits began publishing their analyses of the file. My curiosity was piqued, too. While the PDF seems to not be a pristine document, I took issue with the methods used by some researchers; I felt that they undermined their arguments by using Illustrator as the primary investigative tool. Their opponents maintained that the odd appearance of the file was the result of using OCR (Optical Character Recognition) software. I found myself disagreeing with both camps. So I started doing my own experiments.

My background is in printing; I have over 20 years of experience in Adobe Photoshop and Adobe Illustrator, and about 15 years of experience in Adobe Acrobat. I have extensive experience with Photoshop retouching. I have spent lots of time deconstructing problematic customer files.

## My Approach

I'll start with a couple of basic points, then describe what I've found by dissecting the "birth-certificate-long-form.pdf" file that I downloaded directly from this URL:

[http://www.whitehouse.gov/sites/default/files/rss\\_viewer/birth-certificate-long-form.pdf](http://www.whitehouse.gov/sites/default/files/rss_viewer/birth-certificate-long-form.pdf).

Then I'll describe my own attempts to replicate the type of content found in the official PDF. Some initial observations:

- It's obvious that it's been at least photocopied from a book, as evidenced by the shadowed page curl at the left. And yet it's on green patterned paper. It's possible that this is a standard procedure, to photocopy onto special paper rather than just white paper. Picture the book placed face down into a copier loaded with green patterned paper. Or perhaps the original bound document was photographed and the image stored on microfiche. That microfiche could be called up and printed, again, on patterned paper. Maybe some offices do that to make it an "official copy." I have no way of knowing. While this is a bit odd, it doesn't necessarily indicate fakery.
- A number of investigators have opened the PDF in Adobe Illustrator and poked around. I don't think this is a valid interrogation technique: We always recommend that people NOT use Illustrator to edit PDFs, because the original structure can be modified (fonts may go missing, and the document can come unraveled, depending on the nature of the original PDF and how it was generated). I think it's better to work in Acrobat. Opening in another application introduces uncertainties.

## Examining the Document In Acrobat

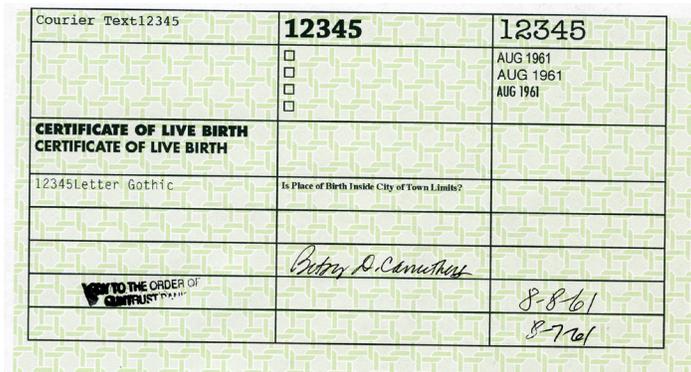
In Acrobat, if you examine the file properties (File > Properties), you can find out a lot:

- If the document had been run through OCR, as so many contend, you would see fonts listed under the Fonts pane: that pane is empty in the birth certificate. Additionally, there is no selectable text in the document. In my opinion, this has not been subjected to any OCR process.
- Under the Description pane, the birth certificate indicates that its creating application is the Macintosh application "Preview." Preview is a PDF and image viewer. It can scale images, and perform simple color adjustments (darkening, increasing contrast, etc.). It cannot edit images or combine them as Photoshop can. The fact that the creating application is Preview just says that someone opened a PDF (which had already been through the wringer) in PDF and chose File > Print, then chose the "Save as PDF" option in the Print dialog.
- In Acrobat, when you use the TouchUp Object tool to select parts of the page, it does start to come apart. One investigator suggested that the piecemeal nature of the file could be due to using PDF Optimizer in Acrobat, and, indeed, I could replicate the "layered" behavior with my own files. However, the layering was limited to a separation of the black lines, text, and handwriting from the patterned background. The black content did not break into segments as the birth certificate did.

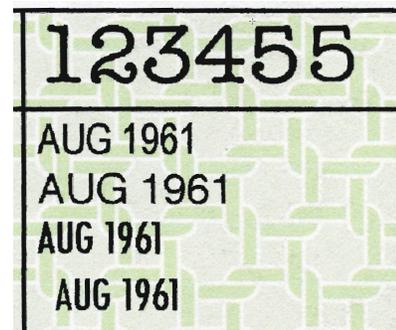
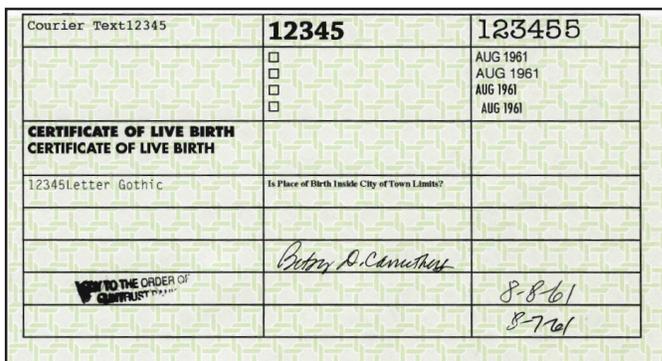
# An Attempt to Replicate the Features in the Long Form Birth Certificate PDF

Here's my own experiment. In Illustrator, I created a green pattern similar to the background of the long-form PDF, and printed it on my Epson inkjet printer. I created the black table and text in InDesign and printed it on the green patterned paper. I then added the rubber stamp and handwriting to the print.

1. I scanned the finished piece on an Epson 4870 Pro Photo scanner at 300ppi. I scanned it a bit crooked, to sort of mimic the curvature of the underlying scan in the birth certificate PDF.
2. I opened in the resulting TIFF in Photoshop CS5.1. To try to replicate what a retoucher might do, I selected the date "AUG 1961" (3rd instance, condensed text) with the Magic Wand and duplicated it into a new layer (Layer > New > Layer via Copy). I also duplicated the numeral 5 in the 12345 text at the right, using the same method.



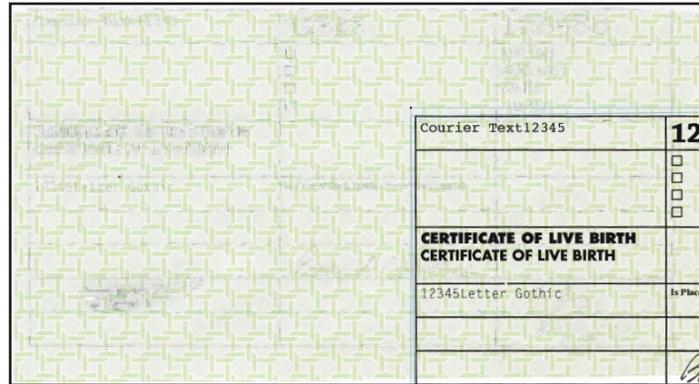
3. I saved the file to a Photoshop PDF in Photoshop CS 5.1.
4. I opened the image in Acrobat 9 Pro, then ran Document > Optimize Scanned PDF with default settings. This attempts to clean up the content, and deskews the document (that's why lines are now straight).



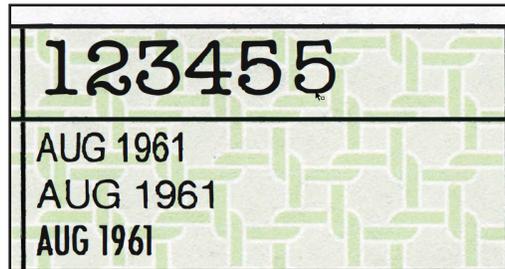
Left: PDF after optimizing in Acrobat. Note straightened lines.  
Right: enlarged section of optimized PDF.

I will note that performing document optimization on a PDF does produce multiple objects that are apparent if the PDF is then opened in Illustrator. So it's entirely possible that, at some point, Acrobat Optimization was performed. However, my results are not as piecemeal as the official document, so, in my opinion, additional processes have touched the file.

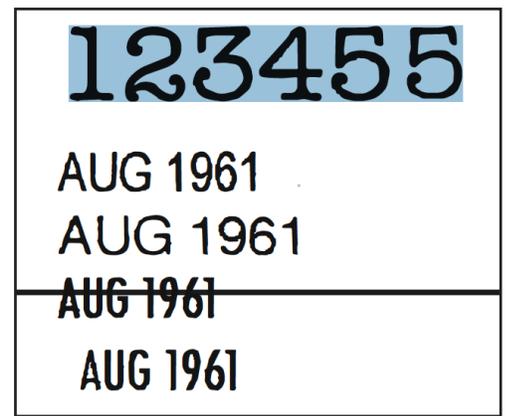
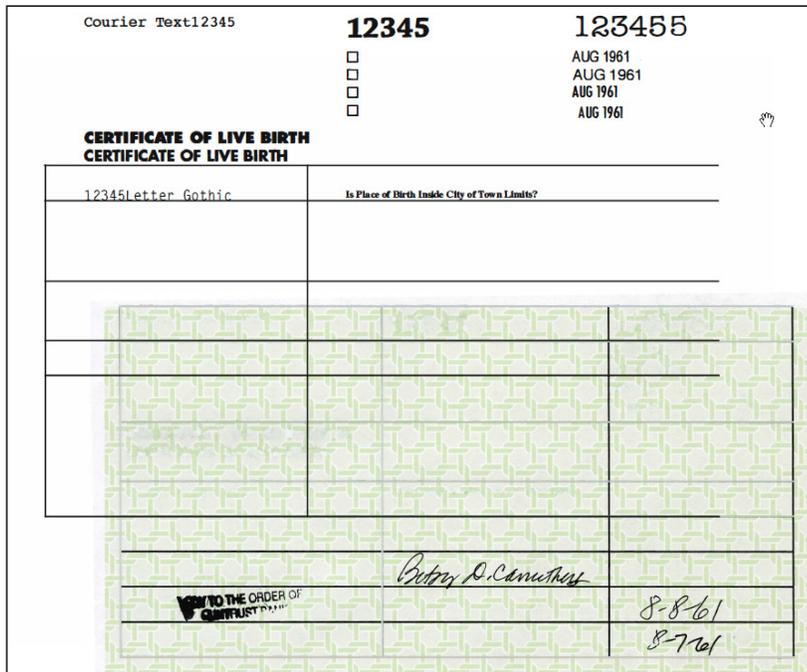
5. In Acrobat 9 Pro, I used the TouchUp Object tool to select and move the black content. *All* of the black content — lines, text, stamp, and handwriting — moved together, leaving smudgy remnants on the background. The black content did not break into pieces as the content of the long-form PDF does in Acrobat.



6. I ran Acrobat OCR (Document > OCR Text Recognition > Recognize Text Using OCR). See detail below; notice hard edge on text. The actual searchable text is invisible; the original scanned text has been replaced with vector artwork that is just a depiction of text. This is Acrobat's attempt to retain the appearance of the original art, while adding searchability.



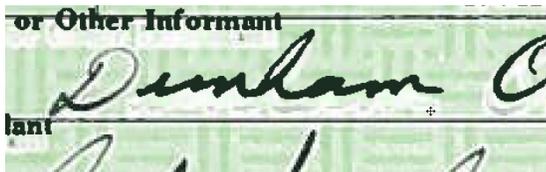
7. Once OCR has been run, text is selectable (below, right). Using the TouchUp Object tool, I can separate the artwork into several components. The black text is one component, most of the lines are each separate components, and the stamps and handwriting (and some of the lines) are part of the green background.



8. Note that the handwriting and stamp content is part of the background, and therefore it's the same resolution as the background and is composed of shades of color. The text and lines however, are vectorized content. When OCR is performed, the searchable text is not actually visible; it's held in an invisible layer above the artwork.



9. But notice the way the content is divided in the long-form birth certificate. Part of the signature is continuous-tone (the “D” is part of the background image) and part is bitmap (the remainder of the signature). In the text sample below, the “1” is part of the background image, and the remainder of the text is bitmap.



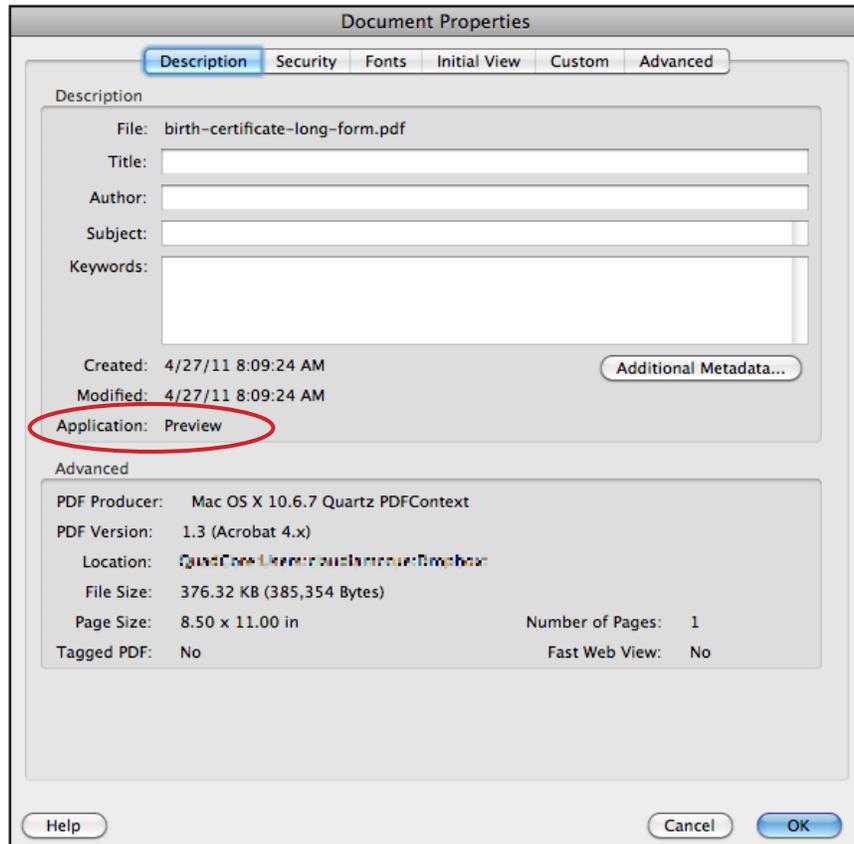
It might be helpful to compare *continuous tone*, *bitmap* and *vector* content.



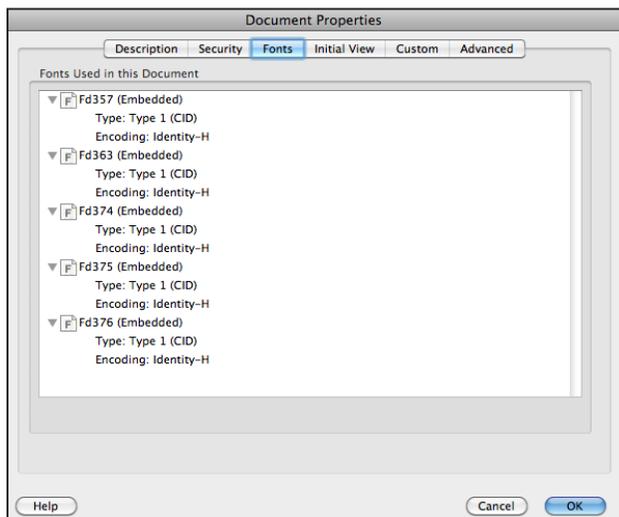
**Continuous-tone** content (*left*) consists of multiple levels of color (in this example, multiple levels of gray). **Bitmap** content (*center*) contains only black and white pixels, with no intermediate gray tones. **Vector** content is smooth, with no inherent resolution (*far right*).

## Examining Some Attributes of the Birth Certificate

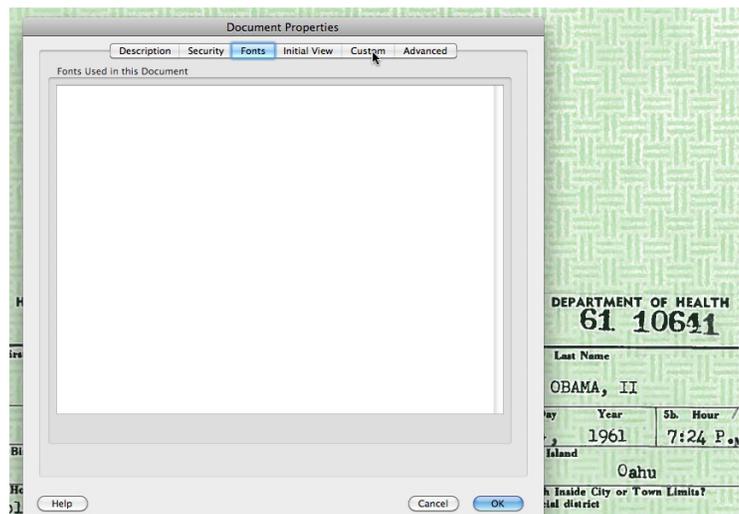
As I mentioned earlier, the Document Properties of the PDF indicate that it was generated from the Macintosh application Preview. The only way I've been able to create a PDF with that property is to open an image or a PDF in Preview, then choose File > Print, and select either "Save as PDF" or "Save as Adobe PDF." Either method produces a PDF with a Preview application stamp (circled below).



Finally, if the PDF were the product of an OCR process, at least some of the letterforms would have yielded searchable text. Even though the text is invisible, it still requires a font, and any fonts in the document would be listed in the Fonts pane of the Document Properties dialog.



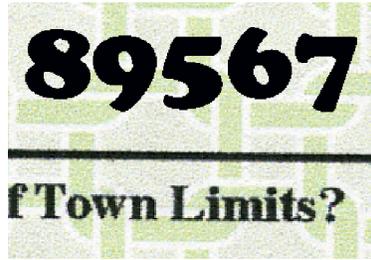
A PDF that has been processed through the OCR function lists embedded fonts in the Document Properties.



The long-form birth certificate lists no embedded fonts in the Document Properties.

## Using Adobe Illustrator to Assemble Image Content

The only way I have been able to replicate the piecemeal nature of the birth certificate (some continuous-tone content, some bitmap content) is by combining bitmap and continuous-tone images in Adobe Illustrator. In the image below, the small text, horizontal line, and green pattern are all one continuous-tone image. The 89567 numerals were saved from Photoshop as a bitmap TIFF image (also sometimes called bi-level). The images were combined in Illustrator.



I then saved the Illustrator file as an Adobe PDF, opened that PDF in Apple's Preview application, and printed to PDF. That approach comes closest to matching the attributes of the birth certificate PDF: bitmap plus continuous-tone content, some vector content, no embedded fonts, and the Preview application stamp. While Photoshop might be the first choice of an artist wishing to combine image content, it's possible to perform some compositing in Illustrator.

## Conclusion

To be fair, I don't have experience with every proprietary scanning and imaging solution. Some dedicated document-processing workflows may result in the piecemeal nature evident in the official birth certificate. But given what I know (as a Photoshop, Illustrator, and Acrobat ACE), I'd say that it was doctored.

Of course, the doctoring could have been performed with the most innocent of intentions, just to produce a clean-looking file. Given the scrutiny it was bound to receive, though, it would have been foolish and ill-advised to fiddle with it at all. It would have been far better to supply a TIFF or JPEG of the original scan. Or, to keep someone from doctoring it, the original scan could have been converted to a PDF with security settings, preventing any editing or dissection. But that PDF would likely show Acrobat or Photoshop (or some other imaging application) as its originating application — not Preview.

In my opinion, there would be no need to open an image or PDF in Preview and print it to PDF, unless doing so was an attempt to try to scrub off some ancestry.

I can only speculate on what's been done to produce the file as we see it. But, in my opinion, the supplied PDF is not a pristine scan simply enveloped in a PDF wrapper. It's been touched by other processes. We may never know the true genealogy (pun intended) of this document.

—CM  
July 18, 2011