The Portable Document Format: Archiving Our Traditions

Adam C. Powell

The digital age has generated the culturally uneasy notion of formless documents. Technologies such as RSS (Really Simple Syndication) make it possible for people to publish information to mass audiences without control over the form in which it will be received. Web services powered by XML allow information to be transformed in ways that its author had never conceived. Although there is a strong movement to separate form from content in the digital world, there is an equally strong opposing movement to marry the two. The evolution and rise of the paper-like Portable Document Format demonstrate the possibly illogical importance that society places on continuing to control the form and content of its documents.

In the years that elapsed between the invention of computerized text and the PDF, authors faced a set of concerns that they never before had encountered. Although electronic documents consisted of a set of instructions describing how to produce a copy of the original, they often had multiple prerequisites. While documents could recommend fonts, if the reader lacked the font, a substitution would be made. In this short-lived era, authors wishing to maintain the integrity of their documents had to both distribute their text and their typeface. The PDF has resolved this problem by encapsulating text, typeface, formatting, and graphics into one document. The only prerequisite to displaying a PDF in the manner in which the author had intended is a viewing application supporting an appropriate version of the standard. Once opened, the document appears onscreen formatted as if it were printed in a glossy circular. If the reader chooses to print the

document, an identical copy will be produced. In the words of the Census Bureau, a PDF "is a file that will look the same on the screen and in print, regardless of what kind of computer or printer someone is using and regardless of what software package was originally used to create it."[i] As was the case with traditional publishing, the publisher of the PDF has full control over where each and every line wrap will occur. What the reader sees is exactly what the publisher had intended.

At present, it is difficult to tell whether the Portable Document Format was created primarily for onscreen use or for the dissemination of documents to be printed. If the format were created merely as a means of distributing printed documents electronically, the motivation for its page-like format would be obvious. This, however, was not the case.

Insight into the purpose of the format can be gleaned through exploring its history. The precursor to PDF was introduced in 1991 under the name "Interchange PostScript." This name is telling, as it reveals the original intention of the format. The Portable Document Format was designed to assist document interchange between computer systems of different platforms. Regardless of whether a PDF is viewed on a Mac, a PC, or a UNIX machine, its author knows precisely how it will be displayed, as the format was designed to be completely platform independent.

The nascent name of the format also serves to remind us that the Portable Document Format was derived from PostScript, a technology that was created to specify the output of printers in a precise and system-independent manner. However, the first version of PDF was different from PostScript in an important way; PDF originally supported only the RGB color space, which is only suitable for onscreen display, while PostScript supported the CMYK color space, which is primarily used for printing.

Although the PDF was derived from a technology designed for printers, Adobe did not intend that the first implementation of it would be used for creating printable documents. This decision was further enunciated in Adobe's decision to formally name the format the "Portable Document Format," instead of "Interchange PostScript," deemphasizing the connection to the PostScript printing technology.

While the Portable Document Format was designed to function on any computer, it was clearly not created with printing in mind. If the PDF was not designed for printing, why do PDF documents often look like printed documents? There are PDF brochures, PDF forms, PDF articles, and PDF eBooks. Every popular use of the PDF has a paper analog. Since the first version of the PDF standard was not suitable for prepress work, the PDF must have been designed to create print-like documents for human convenience rather than for the benefit of printers. The most obvious explanation for this is that Adobe believed that people desired to create and interact with documents in fixed, print-like format, regardless of whether the documents were to be printed or exclusively viewed digitally.

The preference for print-like documents is not an intuitive preference, as when the formatting of a document is fixed, it often displays more poorly onscreen than if its formatting was easily adjustable. When the window of a RTF (Rich Text Format) document is resized, the text wrapping of the document changes to accommodate the window's size, as the format is rather flexible. Unlike the RTF, the text wrapping of the PDF does not change to accommodate window size. This often results in documents being displayed in a suboptimal manner. For instance, if a PDF representing a portrait-oriented, 8.5"x11" sheet of paper is displayed full-screen on a display with a 4:3 aspect ratio (such as at the resolution 1024x768), the page only fills 56% of the display. If the same amount

of information were conveyed within an RTF document, the text-wrapping would change so that the entire screen would contain content. When using the RTF, the user could either see more text on the screen than when using the PDF, or could see the same amount of text, but at a higher magnification. As there are clear disadvantages as well as advantages to specifying the formatting of a document, use of PDF is not without a cost.

If the Portable Document Format was created as a paper-like format to fulfill society's demand for traditional media onscreen, then its creation demonstrates that Adobe believed in the power of the printed form, even in circumstances where it is rather impractical. According to P. David Marshall, "Newer media often imply a supersession of the older form; but it should be made clear that new media have not historically obliterated their past."[ii] The PDF tames the traditional electronic document by freezing it into a format that does not vary from device to device. Corporations, students, and lawyers alike feel a discomfort with the unfettered digital document, and turn to the PDF in order to present their information in a comfortably fixed manner.

The PDF is a means of controlling the output of new digital technologies—anchoring them to the pre-digital past. Raymond Williams stated, "If the technology is a cause, we can at best modify or seek to control its effects."[iii] Through creating PDFs, people control the modification and use of their digital documents. While digital documents may in some respects be more useful if they can easily be modified and reformatted, people have a cultural preference for creating documents in an immutable format. Although RTF documents make better utilization of the surface area of the screen, PDF documents eliminate the publisher's fear of losing control. Unfettered digital texts are frightening to publishers, as they can be completely modified and recomposed by those

who receive them. The Touchup Text Tool provided within Acrobat only allows small changes to be made to PDFs, on one line of text at a time. As a result, it is very difficult to rewrite a paragraph of a PDF. While a devious individual could strategically inject the word "not" into a document, or change a "don't" to a "do," even these attempts at mischief can be prevented if the format's security features are used. The fixed nature of the PDF gives publishers the peace of mind that they need to disseminate portable documents in the digital domain.

In creating a relatively immutable format, Adobe realized that it was necessary, but not sufficient, for a document to be readable on a diverse set of platforms to be considered "portable." Although RTF files are platform-independent, they are not "portable," as their formatting varies slightly depending upon the platform on which they are displayed. The need for immutability (and thus a fixed form) arises from one of the definitions of portability. According to the Oxford English Dictionary, something that is "portable," within the domain of software, is "usable on different machines or on different systems; transferable from one machine or system to another."[iv] Meanwhile, an older definition of portability is "said of liquid substances congealed, and of gaseous substances liquefied, so as to be more conveniently carried or transported."[v] According to this classical definition, the fundamental state of a substance must be changed to achieve portability. In many ways, this still holds true in the case of PDF files. PDFs are "congealed" files, as they embed text, typefaces, and images into a single file.

Portability, by its very nature, requires the transformation of contents. Conversion to PDF is largely a one-way transformation. Although any type of document can be converted to a PDF, once made a PDF, it can only be viewed, printed, and to a slight

extent, edited. Just as it is easier to oxidize magnesium than it is to deoxidize magnesium oxide, it is easier to convert a document to a PDF than to return the contents of a PDF back to the original file format. Thus, PDF conversion is like an exothermic chemical reaction. The creation of a PDF transforms mutable, fluid documents into relatively immutable, solid ones. When the author has not suggested a form for the document to take, the PDF generator creates one, often by locking the text into a format in which it wraps at the edges of the page. In the absence of a format, PDF demands that a default be used, as a PDF cannot exist without a format.

The RTF document is not "portable" because it has not undergone a transformation during the creation process. Images are not recompressed when they are pasted into RTFs, and fonts are not embedded into the files. Acrobat, the program that creates PDF files, optimizes information as it generates a PDF by compressing images and subsetting fonts. Once an image has been compressed or a font has been subsetted, the full-quality original cannot be extracted from the resulting file. However, there is a benefit to the process, as the resulting file is lighter weight and easier to carry due to its smaller file size. It is portable.

There is nothing illogical about the Portable Document Format's insistence to marry form to content, although people may be illogical in their desire to use the Portable Document Format. If PDFs divorced content from form, they would no longer be able to tout themselves as portable, as they would cease to be "congealed." Although PDFs do not make the best use of screen real estate and are difficult to modify, there are compelling reasons for the use of the format. The Portable Document Format enables publishers to abate the modification of their documents and to optimize their content for rapid

download. While a fixed format may be an annoyance to onscreen readers, it is a boon to the creators of artistic and legal documents, who consider maintaining the layout of their documents essential. Perhaps, the only type of non-interactive content that does not benefit from being stored within a PDF is text whose security is irrelevant, layout is unimportant, and likelihood of (page-based) citation is low. As few people would like their documents to enter this category, only documents intended to undergo further revisions are commonly stored in other formats. Maybe, our cling to the page-based past is not so illogical after al

---

[i] "Portable Document Format (PDF)." U.S. Census Bureau. 31 Oct. 2005. <http://www.census.gov/main/www/pdf.html>

[ii] Marshall, P. David. "New Media Cultures." New York: Oxford University Press Inc. 2004. p. 4

[iii] Williams, Raymond. "The Technology and the Society." Electronic Media and Technoculture. John Thorton Caldwell. Piscataway: Rutgers University Press, 2004. p. 36

[iv] "*portable, a.: computing*" *Oxford English Dictionary Additions Series.* Vols. 1-2. Ed. John Simpson and Edmund Weiner. Oxford: Clarendon Press, 1993. *OED Online* Oxford University Press. 21 Nov. 2005. <http://dictionary.oed.com/cgi/entry/50184468>

[v] *"portable, a. 1b" Oxford English Dictionary.* Ed. J.A. Simpson and E.S.C. Weiner. 2nd ed. Oxford: Clarendon Press, 1989. *OED Online* Oxford University Press. 21 Nov. 2005. <http://dictionary.oed.com/cgi/entry/50184468>