



APPLIED DISCOVERY WHITE PAPER

File Formats for Electronic Document Review

Why PDF Trumps TIFF

LexisNexis™
 Applied Discovery®

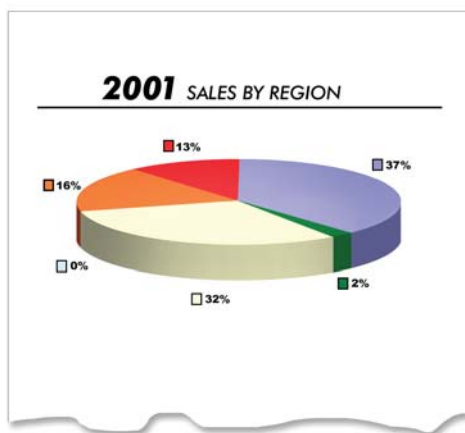
What is the difference between PDF® and TIFF, and why should lawyers care? The short answer is that lawyers need an efficient way to review and share different kinds of electronic documents, and PDF is the best way to do so.

A PDF (Portable Document Format) file is a self-contained electronic document that any computer user can view or print, regardless of the hardware, software, or operating system used to create the original document. Traditional software applications such as Microsoft® Word or Excel are used to create and edit documents. PDF is a universally accepted format for distributing, viewing, and printing such documents.

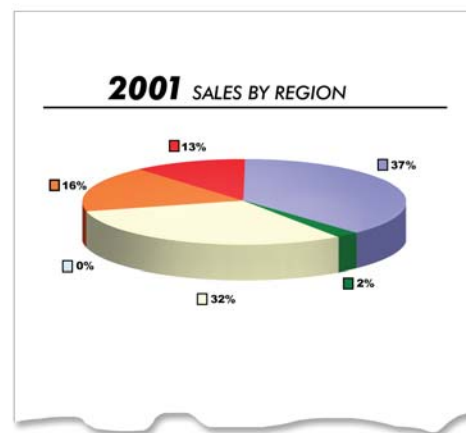


PDF is a universal format for transmitting and viewing almost any file type.

PDF files always display and print exactly as the original documents were created. The PDF file format preserves graphics, letterheads, fonts, layout, and color. Because the PDF file format utilizes Adobe's® Postscript printer control language, PDF documents provide unmatched visual clarity.



Original Document

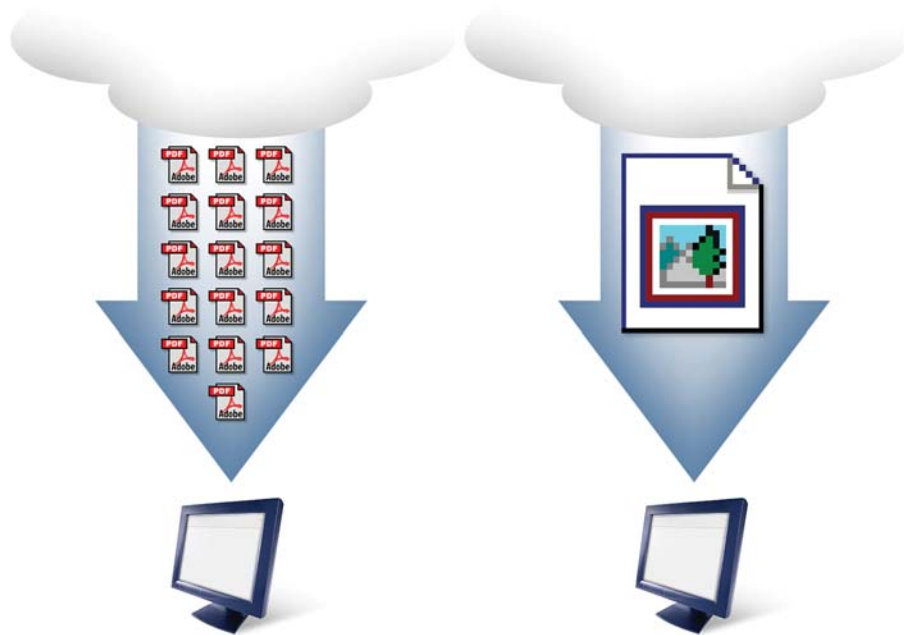


PDF Document

A PDF file maintains the visual integrity of the original document.

PDF is more than just a picture of a document. PDF files retain all the content and electronic history of the original document, including the formatting and "meta data" (more on this later). But, unlike the original application file, the PDF document can be shared, viewed, and printed by anyone using the Adobe Acrobat Reader®—which is available as a free download.

Despite the fact that PDF files contain considerably more information than TIFF images, PDF files are one-tenth the size of TIFF images. This is because the PDF format has built-in file compression, resulting in much smaller file sizes. These compact files result in fast transmissions and downloads, making PDF ideal for use on the Internet.



Small file sizes make PDF ideal for fast transmission and downloading.

What's a TIFF?

In simplest terms, a TIFF (Tagged Image File Format) image is a picture of a document. TIFF is an older file format, created in 1987, and was one of the first designed for scanning paper documents. The TIFF file format came into being just as lawyers began looking for a way to manage huge volumes of paper in large cases. With TIFF, these paper documents could be converted to electronic form by using a scanner and saving the file as a TIFF image.

In its day, scanning represented a significant gain in efficiency over manual paper review. Throughout the late '80s and in to the early '90s, scanned images of documents were the only alternative to manual review processes—but that is no longer the case. Although Adobe holds the copyrights to both TIFF and PDF, it has not released new specifications for TIFF since 1992, and is no longer investing in its development. Instead, Adobe developed PDF to supersede TIFF.

What's a PDF?

The PDF file format, first introduced in 1992, addressed a far more ambitious objective than the TIFF format: to enable any electronic document to be viewed and printed on any computer, regardless of the operating system or software on that computer. As the name suggests, the purpose of PDF is to make documents "portable"—easy to move about.

Widespread use of the Internet and resulting document sharing among computer users prompted the need for PDF. The variety of platforms and software programs used in today's business world can cause difficulties when users attempt to share documents. A recipient may not be able to review or edit a document without the application used to create it. Even then, incompatibilities between operating systems and different versions of software programs can cause critical content and formatting to be lost in transmission.

PDF has emerged as the answer: documents can be universally viewed, downloaded, and printed, regardless of how they were created and without loss of document attributes. The fundamental benefit of PDF is that anyone can view and print a document on any platform with the free Acrobat Reader software. No aspect of the original document is lost. And—importantly for transmission on the Internet—PDF files are compressed so they can be transferred and opened quickly.

PDF vs. TIFF

PDF's universal compatibility is not its only advantage over TIFF. Others include:

Searchability

Unlike a TIFF image, the content of a PDF file can be searched electronically. When a document—say a memo written in Microsoft Word—is converted to a PDF file, all of the original document's text and meta data are captured. Adobe's Acrobat software (the software used to create PDF files) not only creates a visual image of a document, but also embeds the text and meta data; the image, text and meta data are all integrated into one searchable file.

By contrast, a TIFF image is simply an opaque picture of a document, and therefore cannot be electronically searched. Because of this, these documents are usually scanned a second time using "OCR" (Optical Character Recognition) software to capture the document's text. (OCR technology is, on average, only about 85 percent accurate.) The scanned text is then saved into a second, unrelated text file, which is in turn loaded into a database to enable keyword searching.

Meta Data

In addition to its other limitations, the OCR process fails to capture the original document's meta data. In the simplest terms, meta data is hidden data about the document: date created, original author, date last modified, and so forth. Email files include more than 90 fields of meta data, including when and by whom the email was received, opened, and forwarded, and the email addresses of those on the "bcc" list. Unlike PDF, the TIFF file format cannot capture any meta data.

File Size

Portability means small file sizes. The built-in, state-of-the-art compression algorithms in PDF creation software produce compact, lightweight files. Smaller files can be transferred, opened, and downloaded more quickly, and require a fraction of the storage space.

By contrast, unwieldy TIFF images occupy about ten times the space, and consequently take significantly longer to transmit, open, and download.

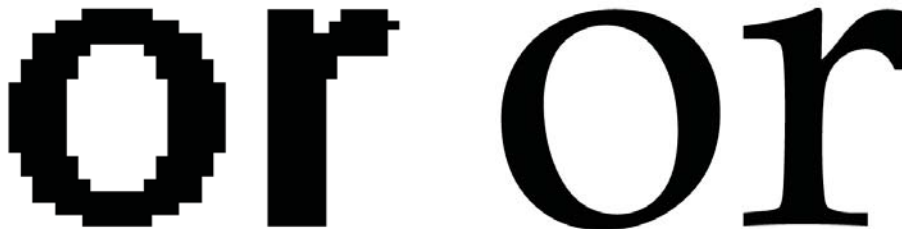
Clarity

Despite their smaller file size, PDF files look better on screen and print with dramatically higher clarity. How is this possible?

The TIFF file format is based on "bitmap" technology: the scanner divides the image into a grid of boxes and represents each box with either a zero or a one, depending on whether the box is filled in. The resulting matrix of bits—called a bitmap—can then be stored in a file and displayed on a screen. This process produces a TIFF image.

By contrast, PDF is based on Adobe's PostScript printer control language. Rather than relying on a grid of dots, PostScript creates a mathematical formula that tells the printer (or computer monitor) precisely how to draw the image. The resulting image perfectly represents the original and allows for enlargement without loss of clarity.

PDF files are much clearer. Zoomed-in images of the same characters show why the difference matters:



TIFF zoomed to 400%

PDF zoomed to 1600%

PDF provides unmatched visual clarity.

Page Orientation

Another benefit of PDF is that it maintains the page-to-page integrity of the original electronic file. Each original page is displayed as a separate page in the PDF file, but all of the pages are neatly organized together within the same file. This is not the case with TIFF images.

Because TIFF was originally designed as a single-page scanning technology, each page of a document is a separate TIFF image; a ten-page contract ends up as ten individual TIFF images—none of which can be electronically linked.

Security

Unlike TIFF, PDF security protects documents from unauthorized changes. After a PDF file is created, the author can "lock" the file, requiring a password. The PDF file format offers several security options, enabling the author to prevent editing, copying—even printing—of the file.

Future Prospects

The federal courts are in the process of replacing their outdated docketing and case management protocols with a new system based on PDF technology. The new system encourages attorneys to file court documents—all pleadings and their attachments—electronically via the Internet. Implementation is already well underway in district and bankruptcy courts, with appellate courts following their example. Millions of federal cases are already online.

This new system accepts electronic documents only in PDF. Parties who maintain their key documents in PDF form will have a distinct advantage over those who scan documents to TIFF. Without the added time and expense of converting documents from some other format to PDF for filing with the court, these parties will operate more efficiently, saving time and money in the process.

TIFF was invented in response to the proliferation of paper and the need to scan it. PDF was invented in anticipation of today's all-electronic business environment. The significant technological benefits of PDF over TIFF, coupled with the mandate for PDF filings in federal court, makes it clear that PDF is the document format of choice for legal professionals.

The information contained herein is not intended to provide legal or other professional advice. Applied Discovery encourages you to conduct thorough research on the subject of electronic discovery..

For updated summaries of electronic discovery case law,
visit Applied Discovery's online Law Library at
www.lexisnexis.com/applieddiscovery.

