Optimizing PDF files

PDF's have become quite popular on the Web but most are designed for high quality print output and are not optimized for the Web. Even PDFs designed for Web use can have problems with excess fonts and unoptimized images and forms. By optimizing your PDF files for the Web you can significantly shrink their size and boost the display speed, saving bandwidth and user frustration.

What Is a PDF?

Portable Document Format (PDF) files allow one to display the content of a document on or off the Web independently of any specific software except the Adobe Acrobat Reader. The reader is included in current Web browsers.

To create PDF files and optimize the files for Web delivery, one must purchase the Adobe Acrobat Professional software program. The software includes settings to compress and optimize the PDF file.

It is in how well you use these compression techniques, how efficiently the data is described (including image resolution) and the complexity of the document (read number of fonts, forms, images, and multimedia) that ultimately determines how large your resulting PDF file will be.

Creating Small PDFs

The main factors in creating small PDFs are image resolution, image type (bitmap or vector), the number of fonts used and how they are embedded, PDF version, and the level of compression. In general the higher the PDF version number, the smaller the file. Acrobat will usually display (with a warning) a more recent PDF version, but new compression schemes will spawn an error when opened in older versions of Acrobat.

To create the smallest possible PDFs for the Web minimize the number of fonts, bitmapped images, and substitute vector based-graphics instead. Minimize the number and complexity of forms in your PDF document and flatten form fields, and avoid the use of multimedia.

Graphics

Use the best quality images that you can at the output resolution of the PDF. Inserting compressed JPEGs into PDFs and Distilling them may recompress JPEGs, which can create noticeable artifacts. Use black and white images and text instead of color images for the best quality in compression. Be sure to turn off thumbnails when saving PDFs for the Web.

Minimize Fonts

How you use fonts, especially in smaller PDFs, can have a significant impact on file size. Minimize the number of fonts you use in your documents to minimize their impact on file size. Each additional fully embedded font can easily take 40K in file size.

Flatten Fat Forms

Acrobat forms can take up a lot of space in your PDFs. New in Acrobat 8 Pro you can flatten form fields in the **Advanced -> PDF Optimizer -> Discard Objects** dialog. Flattening forms makes form fields unusable and form data is merged with the page.

Use the RGB versus CMYK Color Space

For web-only PDFs if you have a choice, use the RGB color space for your PDFs versus the CMYK color space. RBG has one less data channel than CMYK, so files are that much smaller in size. Also, Microsoft applications all think in RGB, even when importing CMYK images.

Convert to Grayscale

If color is not required, you can convert your PDF to grayscale. In Acrobat 8 select **Advanced -> Print Production -> Convert Colors** menu. Under **Document Colors** select "Device Gray," and under **Destination Space** choose "Gray Gamma 1.8" or 2.2. A test on a color print ad saved 54% when converting to grayscale (save as).

Optimizing Existing PDFs

In many cases you won't have access to the original document, just the resulting PDF file. Many PDFs we've seen are not fully optimized for the Web, using conservative settings more appropriate to high-resolution printers. For computer monitors viewing web-based PDFs, you don't need high resolution images and exact reproduction of font faces, you just want to convey your information in an efficient way. Using the techniques outlined below, you can shrink your PDFs, while still maintaining the textual data for search engines, and reasonable quality for print output. Some webmasters offer two versions of their PDFs, once for fast web display, and one for printing.

Adobe's PDF Optimizer
http://www.websiteoptimization.com/speed/tweak/pdf/optimizer.html

Acrobat 8's PDF Optimizer (**Advanced -> PDF Optimizer**) is an improved version of Acrobat 7's PDF Optimizer - both of which are interfaces into Distiller's settings (see Figure 1). You can strip, flatten, downsample resolutions, and remove features from your PDF to minimize its file size.
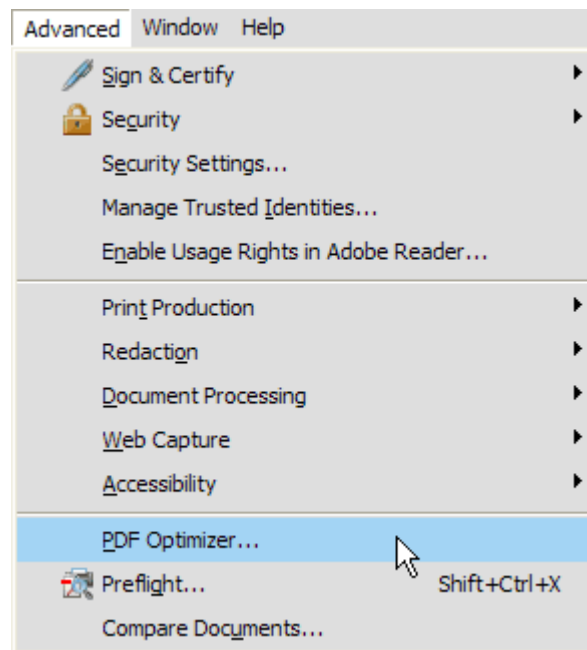

Figure 1: Selecting PDF Optimizer in Advanced Menu of Acrobat 8 Pro

Audit Space Usage

The next step when optimizing PDFs is to audit the size of the different components that make up your PDF. In Acrobat 8 Professional choose **Advanced -> PDF Optimizer** and click on the Audit Space Usage button (see Figure 2).
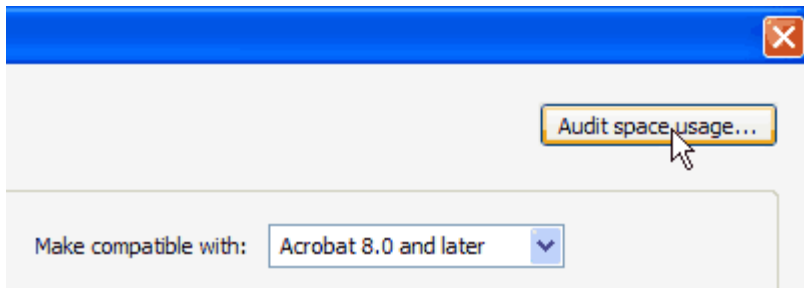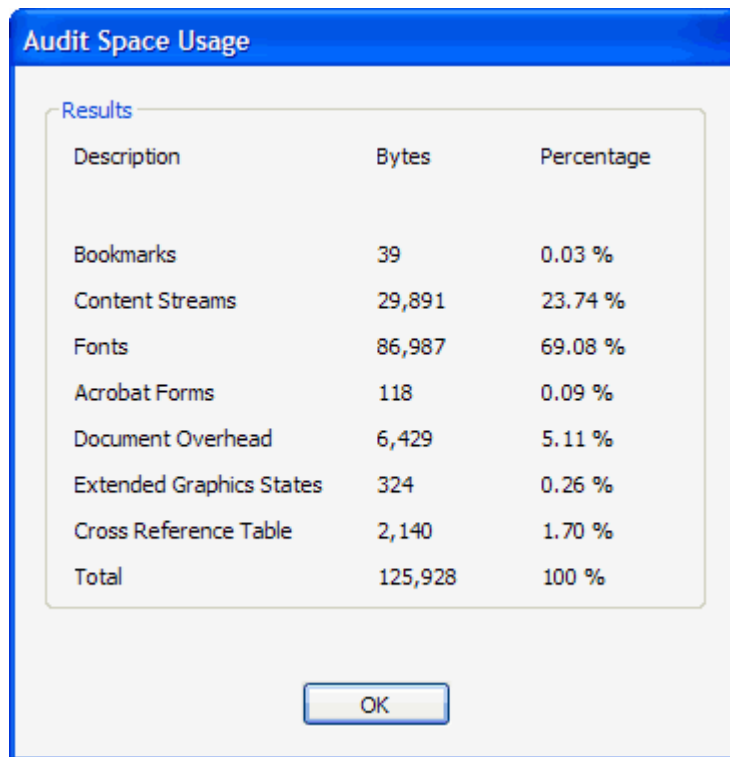
Figure 2:

This dialog shows which components of your PDF are taking up the most space, in bytes for each element and the percentage of total file size (see Figure 3). Concentrate your efforts on the largest areas. In our case fonts take up the most space, with content streams (text), and document overhead next in percentage of total file size.

Audit Space Usage

Results

Description	Bytes	Percentage
Bookmarks	39	0.03 %
Content Streams	29,891	23.74 %
Fonts	86,987	69.08 %
Acrobat Forms	118	0.09 %
Document Overhead	6,429	5.11 %
Extended Graphics States	324	0.26 %
Cross Reference Table	2,140	1.70 %
Total	125,928	100 %

OK

Figure 3: Audit Space Usage

Optimize Images

The first option in the PDF Optimizer settings panel is **Images**. The Image Settings panel lets you downsample images to lower image resolution. For web use you only need 72 dpi for color images, and 150 dpi is a good compromise for monochrome images. Choose JPEG compression for smooth-toned color images and ZIP for flat-color images that would normally be compressed as a GIF or PNG. Choose JBIG2 compression for monochrome images (supported in Acrobat 5 or higher) and lossy compression (see Figure 4). If you use

Acrobat 6 compatibility mode or higher you can use JPEG2000 instead of JPEG compression. JPEG2000 is slightly more efficient than JPEG compression.
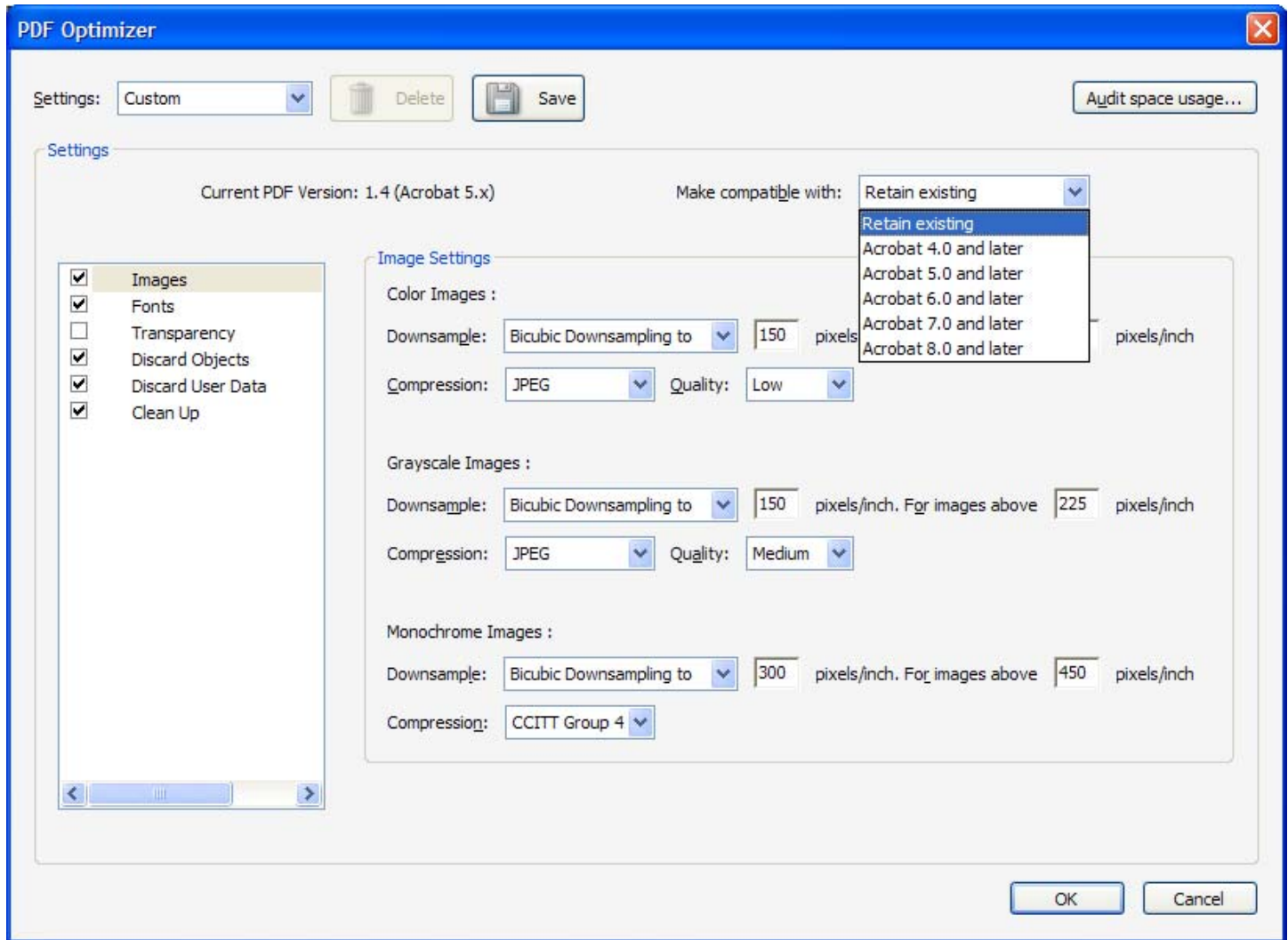

Figure 4: PDF Optimizing Dialog

Downsampling to lower resolutions

PDFs for web viewing need not be ready for offset printing. Many PDFs we tested on the Web were too large in file size for their intended purpose, viewing in a browser. Testing a random sample of large PDFs (1Mb to 15Mb in size) many of them had images and resolutions set too high. You can downsample your images, without changing the compression to realize significant savings in file size. After choosing **Advanced -> PDF Optimizer -> Images** and downsampling to a conservative 150 dpi for color images and preserving current compression (no visible change in appearance) the PDFs tested shrank by anywhere from 1/2 to 1/10 their original size, and still looked nearly identical on screen (see Figure 5).
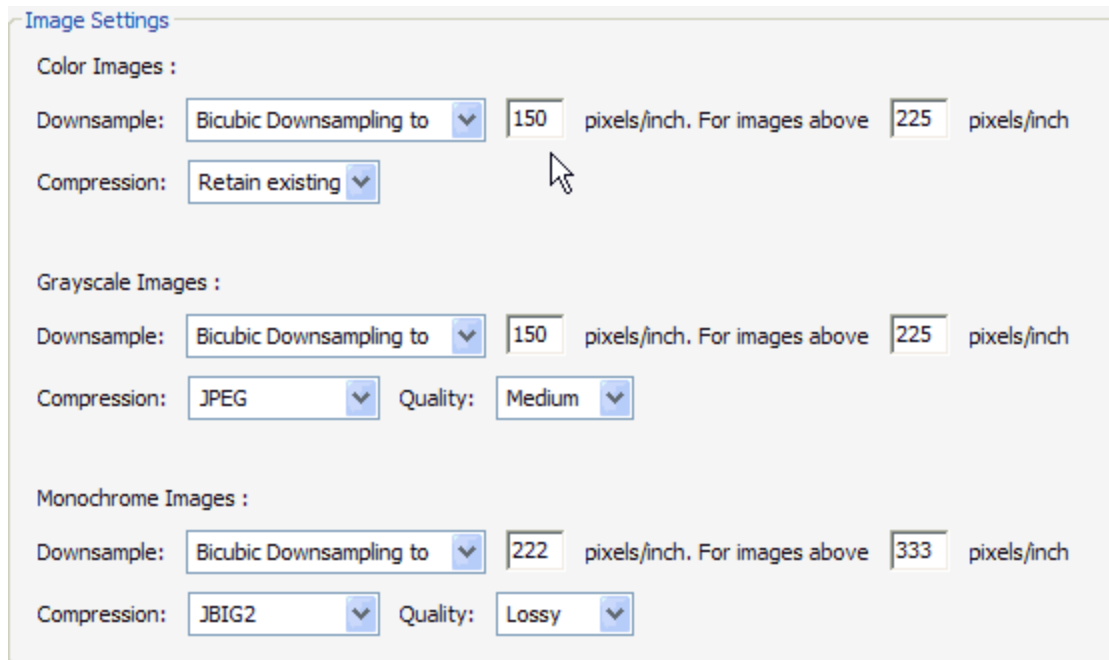
Figure 5: Downsampling in Image Settings Panel in Acrobat 8

Avoid Flatten Transparency and Outline Fonts

You can no longer flatten transparent artwork by choosing **Flatten transparency** in the **Transparency** panel in Acrobat 8. Note that flattening transparency will almost always create a larger PDF, but in some cases will create a smaller one (especially with outline fonts on). The flattening process makes the file less viable for both screen and print. Always keep transparency "live" for Web or screen documents, or avoid using it in the first place. Note in Figure 6, there is no check in the box in front of Transparency.
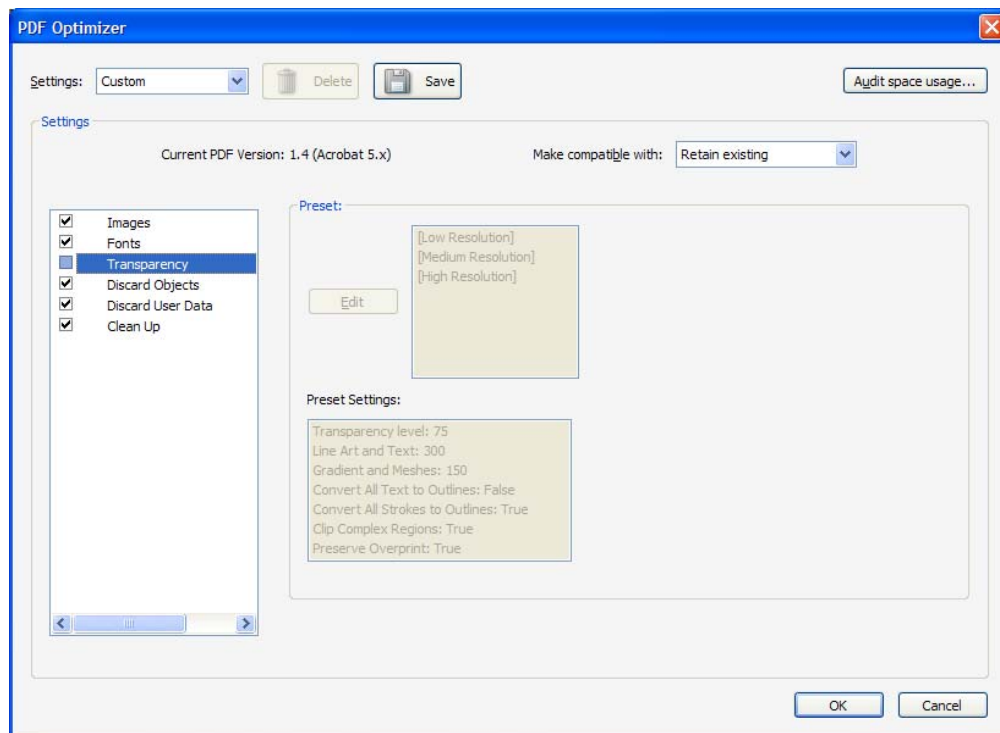

Figure 6: Transparency Settings Dialog in Acrobat 8 Professional

In the Flattener Preview dialog you can still check "Convert All Text to Outlines" but remember the caveats above. Outlined text is not recommended for web use in Acrobat 8, as it makes the document less accessible and removes textual information useful for search engine visibility.

Unembed or Subset Fonts

Depending on the font, each additional font face can easily add 40K or more of data to PDF files if fully embedded. For smaller PDFs, that can make a big difference in file size. Embedding fonts ensures that your document will be rendered exactly as you intended, but bulks up your PDF.

Adobe requires that all PDF applications include the so-called BASE 14 set of standard fonts that can be used unembedded. The fonts are the four faces of these Latin typefaces (Courier, Helvetica, and Times) plus Symbol and ITC Zapf Dingbats.

If Acrobat cannot find the font, it will substitute similar fonts for you. Without the original fonts, Acrobat first matches installed fonts (BASE 14), then "similar" fonts, then multiple master fonts, and finally a "last resort" font. Multiple master fonts (AdobeSansMM and AdobeSerifMM) are an extension of the Type 1 font format that allow the creation of many type faces from one font. Note that if there are any symbols, international glyphs, or non-roman languages you must embed your fonts for these characters to appear.

Delete Unused Objects

In the **Discard Objects** panel you can remove features that you don't use from your PDF (see Figure 7), and optimize curved lines in CAD drawings. Acrobat 8 Pro adds the "flatten form fields" option that makes form fields unusable and does not change their appearance. Your form data is merged with the page to become page content. **Discard Document Tags** removes tags from the document, which also removes the accessibility and reflow capabilities for the text. Leave the document structure alone, since it makes the document more accessible.
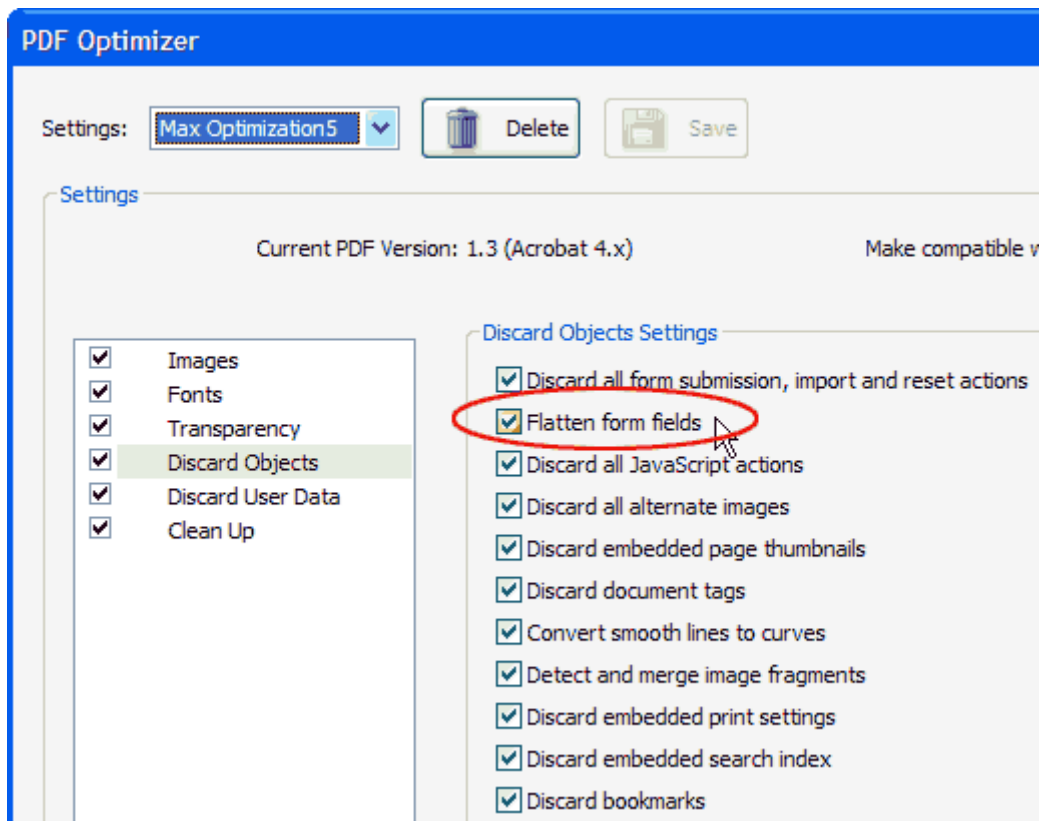

Figure 7: Discard Objects Settings Dialog

Discard User Data

Acrobat 8 has broken out user-specific cleanups to the new **Discard User Data** panel of the PDF Optimizer (see Figure 8). You can use this panel to delete any personal information you don't want to share with others. I usually leave "Discard Document Information and Metadata" unchecked, as this information is good for SEO and the semantic Web. This checkbox removes information in the document information dictionary and all metadata streams. You can use the **Save As** command to restore metadata streams to a copy of the PDF.
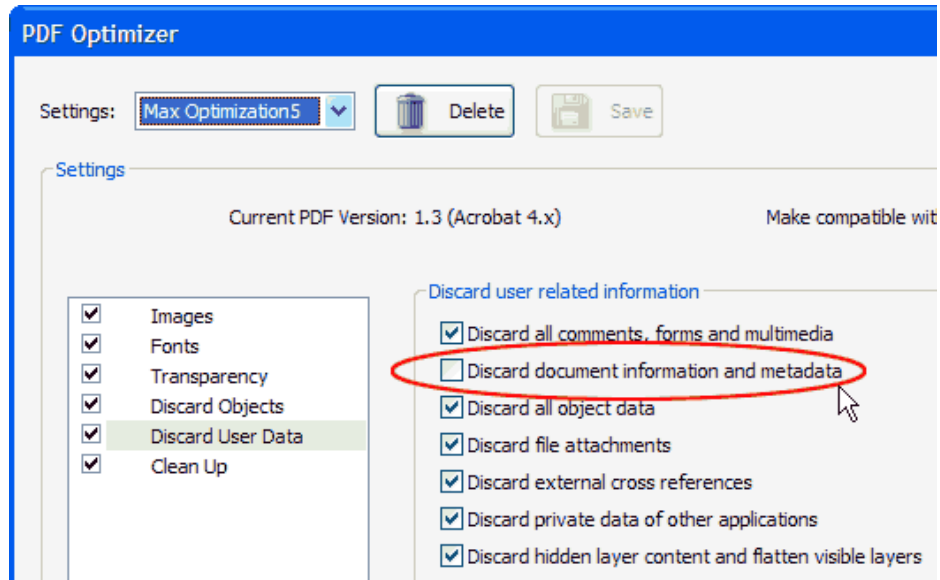

Figure 8: Discard User Data Settings Dialog

Clean Up Your Document

The **Clean Up** panel does some additional housekeeping to shrink your PDF and display the first page faster (Optimize the PDF for fast web view). The first two options use Flate to encode unencoded data streams (like outline fonts) and substitute Flate for LZW encoding (more efficient). Fast web view "linearizes" to make the first page selected display quickly, by essentially duplicating some data in the PDF. The PDF will be larger, but it will feel much faster (see Figure 9).
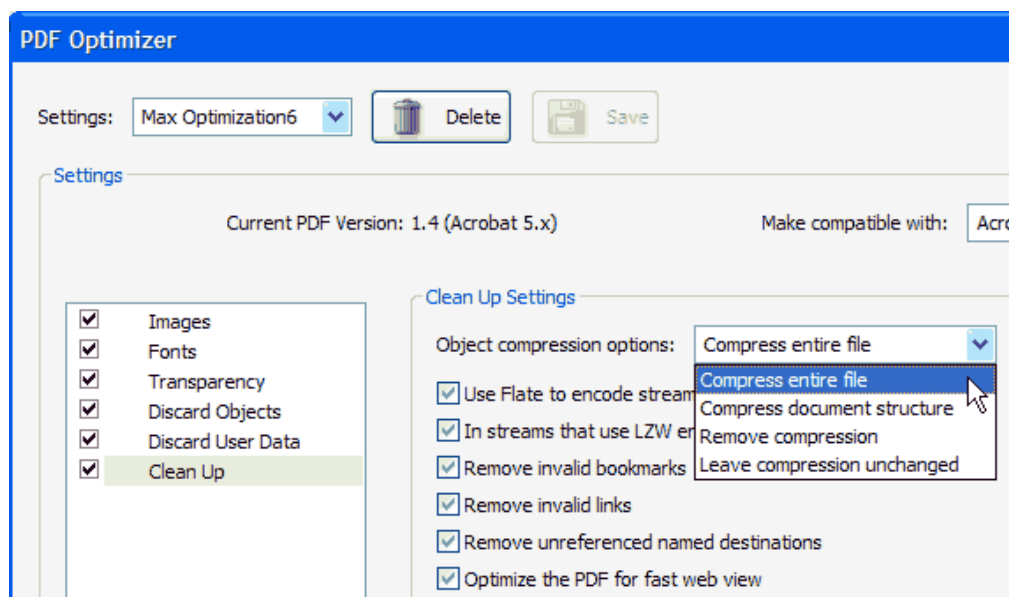

Figure 9: Clean Up Settings Dialog

Note that **compress entire file** is available only in Acrobat 6 or higher (PDF version 1.4). Using this option will spawn an error message in Acrobat 5 or lower (see Figure 10).
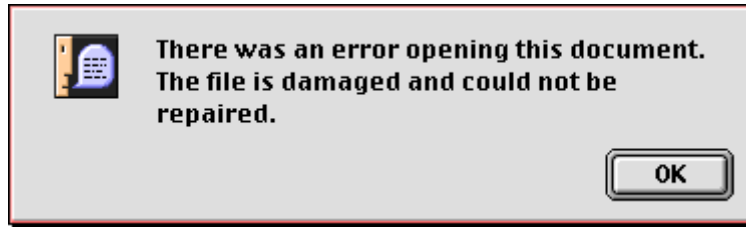

Figure 10: Error message in Acrobat 5 opening compressed Acrobat 6 document

Save the Optimized PDF

Once you've finished tweaking your PDF optimizer settings you can save them for reuse for subsequent PDFs. Save your optimized PDF as a new file to compare it to the original.

For more information see the WebSiteOptimization.com Web site at
http://www.websiteoptimization.com/speed/tweak/pdf/