

Redaction of Confidential Information in Electronic Documents

How to safely remove sensitive information from Microsoft Word documents and PDF Documents Using Adobe Acrobat

CONTENTS

[Typical Causes of Redaction Problems](#) 1

[Application Tools for Removing Data](#) 2

[Redacting a Word Document](#) 3

[Setting PDF conversion parameters](#) 9

[Redacting a PDF Document](#) 11

[References](#) 13

Redaction, which means removing information from documents, is necessary when confidential information must be removed from a document before final publication. Problems can arise when editors use an improper method such as trying to obscure information rather than deleting it, or if they are unaware of sensitive metadata in a document. They can find out, too late, that the information can later be extracted from the document.

Documents are typically authored in an application such as Microsoft® Word® or PowerPoint®, and converted to PDF for final distribution. As with many publishing operations, redaction is best accomplished in the authoring application.

Using Microsoft Word as an example, this document explains how to set preferences for safe conversion to PDF. The general principles can be applied for use with other word processing or page layout applications.

When only a PDF version of a document is available, it is necessary to redact using Acrobat. The section [“Redacting a PDF Document” on page 11](#) describes a procedure for that purpose. Again, every effort should be made to redact in the authoring application before converting to PDF.

NOTE: This document addresses redaction for documents that will be distributed as PDF files. Publishing documents in, for example, Microsoft Word or PowerPoint format can involve issues that are beyond the scope of this document.

Typical Causes of Redaction Problems

There are two main causes of failing to remove confidential information from a document:

- *Attempting to hide confidential content by obscuring or covering the information:* Editors may try to cover sensitive information with a colored rectangle or by highlighting text in black. While these methods work for hard copy documents, they are not appropriate for electronic documents because there are ways to extract the information from the resulting PDF document.

It is also possible that sensitive information might be covered, either intentionally or not, by a non-sensitive image. Since it might be unintentional, the need for redaction might not be obvious to the editor.

- *Being unaware of document Metadata, or not knowing how to properly remove it:* Both Word and PDF documents can carry metadata

information about the document, such as author, subject, keywords, and title. The author may be unaware of metadata generated by the application, and it may not be apparent unless the user knows where to look for it.

Application Tools for Removing Data

Microsoft Word XP/2003

Microsoft has provided some tools for redaction for their Office 2003 suite of products; see “References” on page 13 for links to the Microsoft Web page. The descriptions for these tools does not claim to remove all metadata or sensitive information from the source document. You should make your own assessment regarding the effectiveness of these tools.

Adobe Acrobat

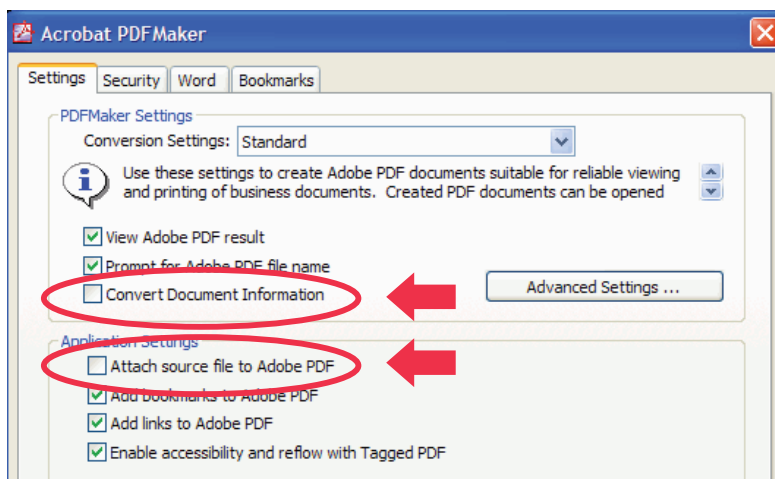
Adobe® Acrobat® does not have tools specifically for redaction, but it is important that you correctly set the conversion parameters for converting Word files.

If you have only a PDF version of a document that requires redaction, there are two choices. You can obtain an Acrobat third-party plug-in such as Redax from Appligent (<http://www.appligent.com>), or you can use the procedure explained in “Redacting a PDF Document” on page 11.

Conversion settings for Adobe® Acrobat® PDFMaker are accessible through the Microsoft Word user interface. PDFMaker works with Acrobat Distiller’s; its operation can be modified by settings selectable within Distiller or PDFMaker (Select **Adobe PDF** > **Change Conversion Settings** > **Advanced Settings**, see “Setting PDF conversion parameters” on page 9).

Most of the conversion settings adjust the size and resolution of the resulting PDF document. PDFMaker has a number of settings related to conversion from Word, as shown in Figure 1, two of which are for controlling confidential information.

FIGURE 1 PDFMaker Settings in Microsoft Word



You must verify the following settings:

- The checkbox **Convert Document Information** controls the conversion of Microsoft Word metadata to PDF and is *checked* by default. Unchecking **Convert**

Document Information removes one source of metadata transferring to the PDF document, but is not a complete solution.

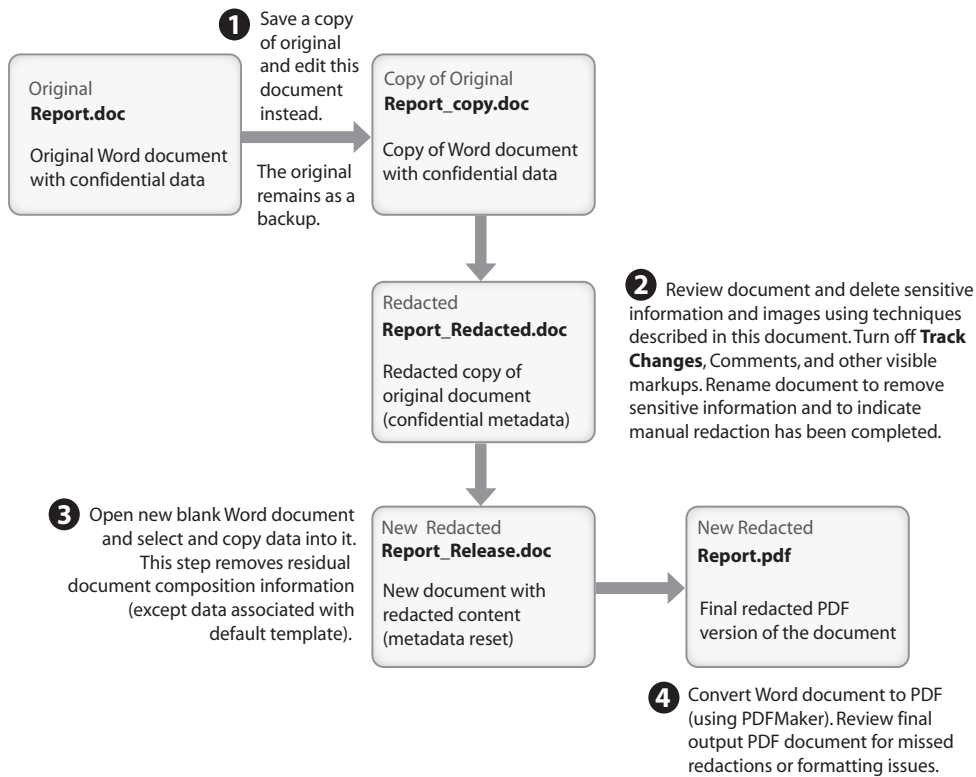
- **Attach source file to Adobe PDF** inserts a copy of the original Word document into the output file, which is rarely what is wanted when redacting a Word document. It is *unchecked* by default, and should remain unchecked for most purposes.

Redacting a Word Document

The key to understanding how sensitive data can be embedded in a PDF document is that information hidden or covered in an electronic document, can easily be recovered. The solution is to ensure that sensitive information is not just visually hidden or made illegible, but is actually deleted from the source file.

In some documents, deleting sections can cause an undesirable reflow of text and graphics. If document formatting is a critical issue, the procedures below discuss some methods for maintaining that formatting.

FIGURE 2 Redaction Process Workflow



Detailed Sanitizing Procedure

The following procedures are described for use with Microsoft Word, but they can be easily adapted for use with other word processor products.

NOTE: The step numbers in [Figure 2](#) above correspond to the step numbers below (white Arabic numerals on a black circle background).

❶ Create a New Copy of the Document

a. Create a new copy of the file

Open the document and select **File > SaveAs** from the top menu bar; give the file a new name. Make sure the new name is not sensitive. All redaction will be done with the new copy, preserving the original as a backup.

b. Turn Off “Track Changes”

The Track Changes feature is a toggle. Selecting **Tools > Track Changes** from the top menu bar toggles the feature on or off. The quickest way to determine if Track Changes is on or off is by looking at the bottom status bar. The letters TRK are dimmed if Track Changes is off, and bold if Track Changes is on.

❷ Review and Delete Sensitive Content

- a. Select each chart, diagram, image, or segment of text to be redacted and delete that item. Delete all comments. Resizing an image, covering a section with a black box, or changing the color of a font to make it invisible, will not work; the item **must** be deleted.

If deleting an item changes the format or structure of the document in an unacceptable way, replace the item with meaningless content of a size that retains the desired formatting.

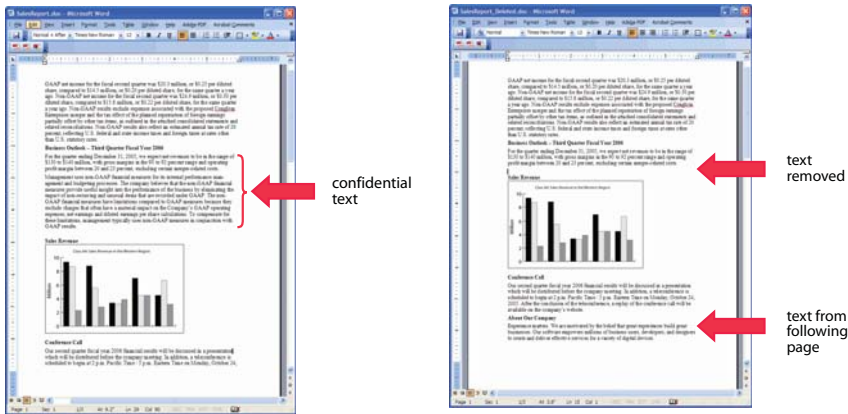
If the redacted item is text, you can replace the text with a single character, such as all Xs, repeated to fill the equivalent number of lines. If the redacted item is an image, you can replace the item with an appropriately colored rectangle (for example, white or gray) of the same size. For detailed procedures, see [Redacting Text](#), below; for redacting images, see [“Redacting an Image” on page 5](#).

Redacting Text

Figure 3 shows a page of a document before redaction (left), and after the sensitive paragraph has been deleted (right).

The sample page on the right side of [Figure 3](#) shows the same page after deleting the text (indicated by the top arrow). Notice that additional text from the next page moved up in the document because deleting the text caused text from the following page to move up onto this page (text indicated by the second black arrow on the right side).

FIGURE 3 Original and Redacted Document

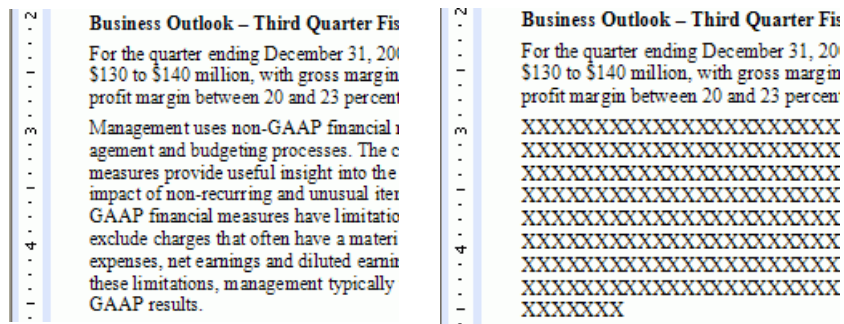


For some documents, the potential text reflow will not be a problem. In other documents, the reflow can cause changes to all following pages, so some reformatting may be necessary to ensure that illustrations stay with the appropriate text and that page breaks are in the correct place. This could be time consuming for a large document.

NOTE: For Microsoft Word, using the free *Microsoft Office 2003 Add-in: Word Redaction*, enables you to redact without changing the layout of the document. See “References” on page 13.

If formatting changes are a concern, you can replace the redacted text with meaningless text of the same size, rather than delete it. Figure 4 shows a before-and-after close-up of the replaced text. Notice that the paragraph following the replaced text did not shift position, thus preserving the formatting of the rest of the document.

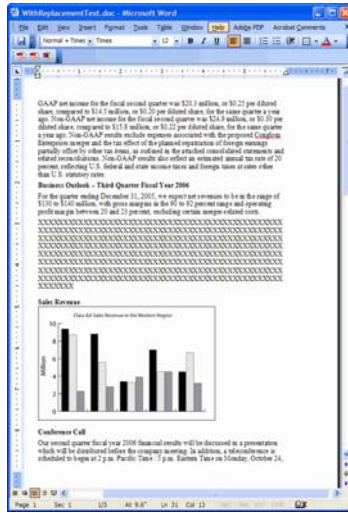
FIGURE 4 Replacing Text with an Equal Amount of Meaningless Text



Redacting an Image

Figure 5 shows the page from Figure 3 after redacting the text as described above. The following procedure describes how to delete an image (in this example it is a chart which was imported as an image), and to retain existing page layout and page breaks.

FIGURE 5 Redacting an Image

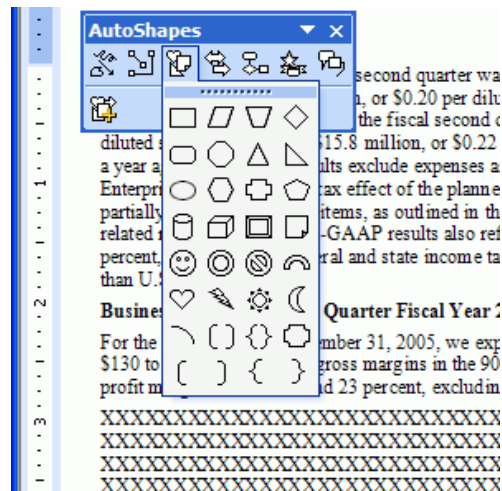


To preserve the formatting, insert a rectangle the same size as the chart (or image) into the vacant space as detailed in the following steps.

NOTE: A confidential graphic may contain an additional title, caption, or other text that must be removed but whose dimensions may not be included in the information on the **Size** tab. If you remove the text, you might want to adjust the dimensions of the replacement rectangle to compensate for the removal of that text, or replace the text with Xs as described above.

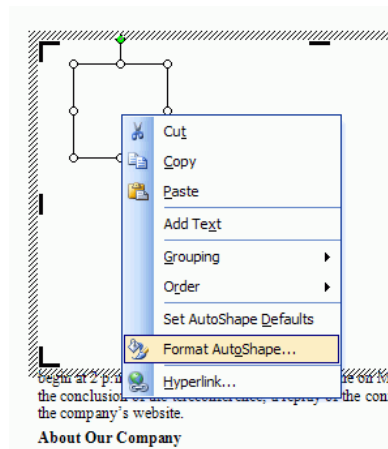
- i. Before deleting the image, determine its size: select the image, select **Format > Picture** from the top menu bar, and then select the **Size** tab in the Format Picture dialog box. Make a note of the height and width of the image.
- ii. Select the image and delete it.
- iii. Insert a rectangle in the same space: select **Insert > Picture > AutoShapes** from the top menu bar (this opens the AutoShapes toolbar). Select the rectangle from the AutoShapes toolbar as in [Figure 6](#). Left click the rectangle which creates a drawing canvas. Move the cursor somewhere on the drawing canvas, left click again to drop a rectangle on the canvas.

FIGURE 6 Choosing a Replacement Shape



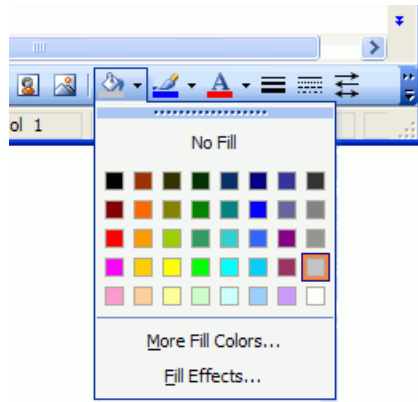
- iv. Resize the rectangle by selecting the rectangle, move it to the upper left corner of the drawing canvas. Right click and select **Format AutoShape** from the menu as shown in Figure 7 (this opens the Format AutoShape dialog box), select the **Size** tab, enter the desired height and width of the rectangle. If you deleted additional text such as a caption for the image, you can add to the height to compensate, or you can replace that text with Xs as described above. Click **OK**.

FIGURE 7 Setting the Rectangle Size



- v. Fill the rectangle with gray (or any other appropriate color): select the paint bucket on the bottom toolbar which brings up the **Fill** toolbar shown in Figure 8. Left click on one of the chosen colors.

FIGURE 8 Setting the Fill Color



- vi. Figure 9 shows the page after redacting both the text and the image; notice that the format and page layout of the subsequent text in the document is preserved.

FIGURE 9 Redacted Text and Image



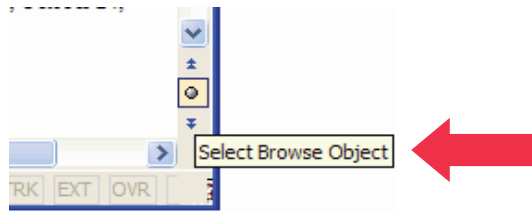
b. Double Check the Redacted Document

Carefully read over the document to ensure that all material to be redacted has been deleted and, if necessary, replaced with nonsense filler text. Check all headers, footers, captions, section titles, footnotes, endnotes, labels, etc. Verify that all comments have been deleted.

To easily flip through the document, use Word's "Browse By" feature for some of the different objects. By default, the arrows at the bottom of the right scroll bar cause you to browse through the document *by page* when clicked. To change this:

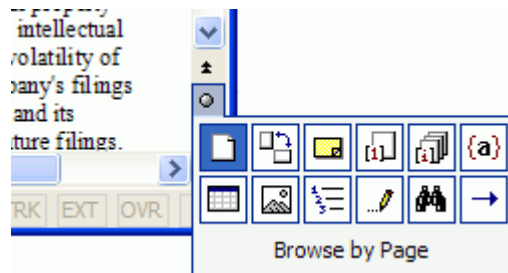
- i. Click the dot at the bottom of the scroll bar shown in Figure 10. This brings up a toolbar with the **Browse By** options, including **Browse By Comment**, **Browse By Footnote**, **Browse By Graphic**, etc., as shown in Figure 11.

FIGURE 10 The “Browse By” Button



- ii. Select one of the options.
- iii. Use the double arrows on top and bottom of the dot to search through the document, moving to each instance of the chosen type of object. Repeat for all types of objects. This is a quick way to make sure all comments are deleted, all images have been redacted correctly, all footnotes have been checked, etc.

FIGURE 11 Selecting “Browse by Page”



④ Create a New Blank Document

a. Select and copy document contents to a new file

Select the entire contents of the document using **Edit > Select All** from the top menu bar. Copy the selected contents using **Edit > Copy** from the top menu bar. This will copy the contents, the formatting, the headers and footers and all necessary information into Word's buffer.

Open a new Word document using **File > New** from the top menu bar.

Paste the redacted document contents into this new document using **Edit > Paste** from the top menu bar.

Save the document with a new name using **File > SaveAs** from the top menu bar.

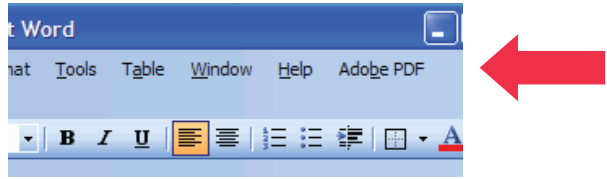
b. Setting PDF conversion parameters

One reason to convert a Word document to PDF is that the conversion redacts some information or hidden data from the document that is intrinsic to the Word format. However, some PDF software has the ability to automatically copy document metadata and properties from Word into the PDF file. This feature, among others, must be disabled when sanitizing documents.

The following procedure is the most common way to disable the data transfer of sensitive document content. However, other methods can be used if the PDF software is configured to prevent converting anything but the visible contents of the file.

NOTE: If Adobe PDF is not on the menu bar, try reinstalling Adobe Acrobat or consult your system administrator.

FIGURE 12 Selecting “Adobe PDF” from Word Menu

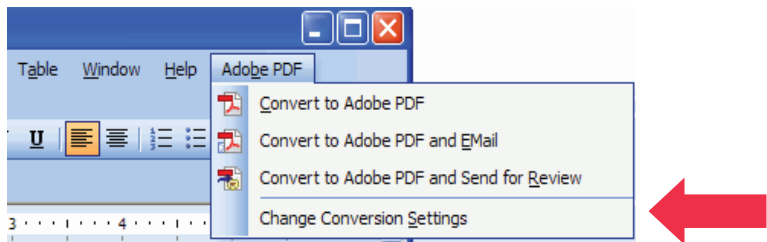


There is also a configuration tool accessible from the Word menu bar (using Acrobat 6 or higher) that allows the user to change numerous settings for PDF conversions. The following procedure describes the settings for converting redacted Word documents, and these options must be verified for each document conversion.

To select the settings for PDF from the Word menu bar, do the following:

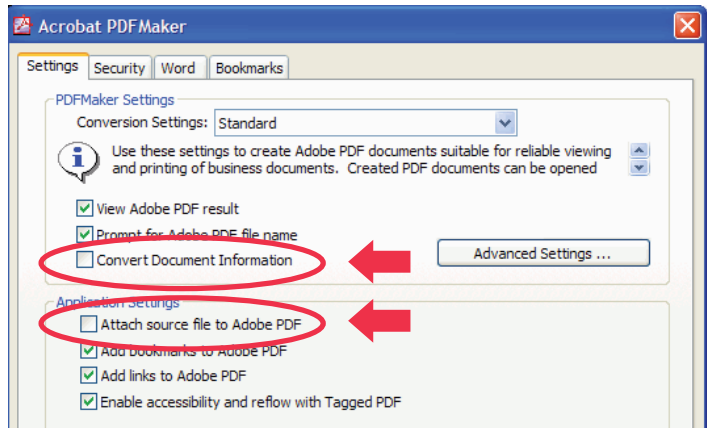
- i. From the Word menu bar, select **Adobe PDF > Change Conversion Settings**, as shown in [Figure 13](#).

FIGURE 13 Selecting “Change Conversion Settings”



The Acrobat PDFMaker dialog opens, as shown in [Figure 1](#).

FIGURE 14 Setting PDFMaker Preferences

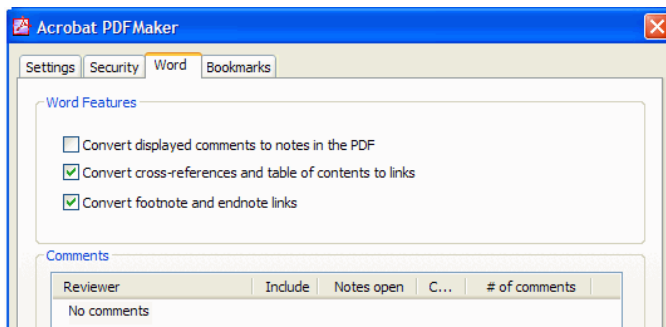


- ii. The PDFMaker dialog contains four tabs labeled **Settings**, **Security**, **Word**, and **Bookmarks**. On the **Settings** tab, ensure that **Convert**

Document Information and **Attach Source File to Adobe PDF** are both unchecked as shown in Figure 14.

- iii. On the **Word** Tab, shown in the close-up in Figure 15, uncheck **Convert displayed comments to notes in the PDF** and click **OK**.

FIGURE 15 PDFMaker Word Tab



4 Convert to PDF

- a. Select **Adobe PDF > Convert to Adobe PDF** from the menu bar.
- b. Lastly, review the output PDF for formatting issues and omissions. If any are found, return to step 2.

Redacting a PDF Document

Although redaction is ideally performed using the authoring application, it may sometimes be necessary to redact when only a PDF document is available. One solution would be to obtain an Acrobat redaction plug-in, or you can use the procedure below.

As mentioned above, it is not sufficient to use PDF annotations such as black rectangles to cover sensitive text and images. The following procedure is somewhat slow for documents with many pages, and you must be very careful in using it.

Briefly, the process of redacting in PDF involves:

- covering each item of confidential information with a black rectangle or by using black text highlighting
- converting the PDF document to multiple TIFF image files
- converting and reassembling the files into a single PDF document.

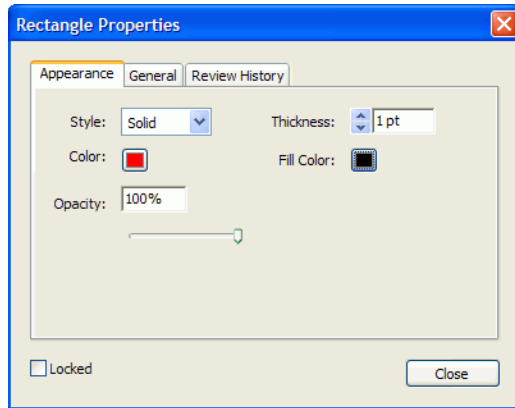
Using this procedure, each page is saved as a TIFF (bitmap) image, and all covered text and metadata will have been lost.

NOTE: Converting to TIFF and back to PDF has the unfortunate consequence that the file will no longer be searchable, and accessibility is lost because document structure and tags are lost in the process. Using a third-party redaction plug-in would avoid such problems.

To remove confidential information from a PDF document:

1. Go to **Tools > Drawing Markups**.
2. Draw a rectangle to cover the sensitive text or graphic.
3. Right click on the rectangle and select *Properties* from the pop-up menu. Set the **Fill Color** to *Black* (see Figure 16).

FIGURE 16 Setting Rectangle Fill to Black



4. Click OK.
5. Right click the rectangle and choose Make Current Properties Default, from the pop-up menu, so that each additional rectangle will be solid black.
6. Alternately, you can use the Highlight tool to highlight text (if you have a PDF Normal or Searchable PDF). Set the highlight color to *Black*. Highlight all sensitive text in the document.

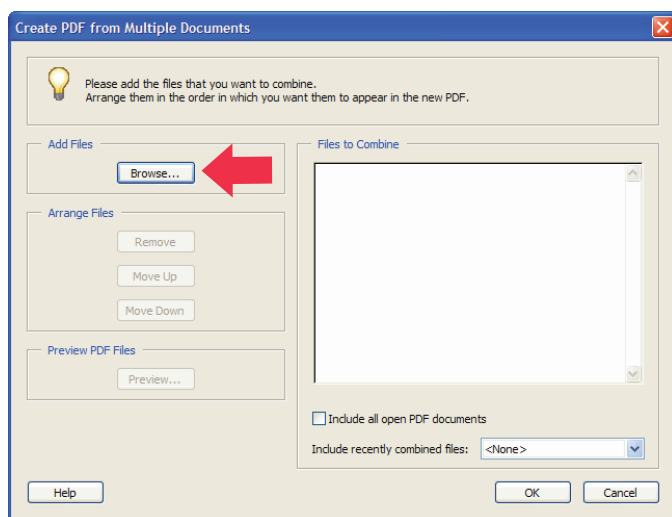
IMPORTANT: *At this point, all you have done is to cover up the confidential information. To remove the information, you need to “flatten” the file by converting to TIFF images using Steps 7 through 11 below:*

7. Select **File > Save As**.
8. Choose *TIFF* from the Type pop-up at the bottom of the dialog box.

Acrobat will save each page of the document as a separate, sequentially numbered TIFF file. You can now reassemble those files into a single PDF document (steps 9 through 11, below). Each page will be a bitmap image, so there will be no text information underlying the black rectangles and highlighting.

9. In Acrobat 7 Standard or Professional, click the **Create PDF** button on the toolbar.
10. Choose *From multiple files*.
11. Click the browse button (Figure 17) and navigate to the location where you saved the TIFF files. Select all TIFF files that you want combined and converted to PDF, and click **OK**.

FIGURE 17 Creating PDF from Multiple Documents



NOTE: If you have a lot of documents to redact, you can save to TIFF and create PDF from TIFF in batch mode using Acrobat Professional.

References

Microsoft Word 2003 Redaction Add-in Tool

<http://office.microsoft.com/en-us/assistance/HA012193991033.aspx>

Office 2003/XP Redaction Add-in (free download)

<http://www.microsoft.com/downloads/details.aspx?FamilyId=144E54ED-D43E-42CA-BC7B-5446D34E5360&displaylang=en>

“How to Minimize Metadata in Office Documents” Microsoft, January 28, 2005.

<http://support.microsoft.com/default.aspx?scid=kb:EN-US:Q223396>

“Redacting with Confidence: How to Safely Publish Sanitized Reports Converted From Word to PDF” National Security Agency. (The report on which this document is based.)

<http://www.fas.org/sqp/othergov/dod/nsa-redact.pdf>

U.S. District Court web page on Redaction:

<https://ecf.cand.uscourts.gov/cand/faq/tips/redacting.htm>

Copyright 2006 Adobe Systems Incorporated. All rights reserved.

Adobe Systems, Inc.

345 Park Avenue, San Jose, CA 95110-2704 USA

<http://www.adobe.com>

Adobe, the Adobe logo, Acrobat, Adobe LiveCycle, and Reader are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. Mac OS is a trademark of Apple Computer, Inc., registered in the United States and other countries. Linux is a registered trademark of Linus Torvalds. Microsoft, Windows, and Word are either registered trademarks or trademarks of Microsoft Corporation in the United States