

*Redaction of PDF Files Using Adobe Acrobat
Professional X*

**Enterprise Applications Division
of the
Systems and Network Analysis Center (SNAC)**

Information Assurance Directorate



**National Security Agency
9800 Savage Rd.
Ft. Meade, MD 20755**

Executive Summary

Redaction is the process of selectively removing visible and non-visible classified or sensitive information from a document for release to a recipient (s) not cleared to view that information. The goal of the process is to prevent *inadvertent* release of the information. The Portable Document Format (PDF) is a common format for publishing documents on the web and for exchanging files between government entities and government contractors. Redaction of PDF files is an ongoing challenge for these entities. This document describes how to use Adobe Acrobat Professional X for redaction of PDF files.

What This Document Addresses

This document describes a procedure using Adobe Acrobat Professional X to redact information from PDF documents. The original source of the document can be any application, but the process described applies to documents that are already in PDF.

What This Document Does Not Address

This document does not address the purposeful or covert release of classified information, nor does it address countermeasures for executable content vulnerabilities or any features of Adobe Acrobat Professional X other than those related to redaction. The process described does not address starting with documents in other formats or redaction using other tools.

Sources for Further Research

The NSA website, www.nsa.gov, contains related papers:

Redacting with Confidence: How to Safely Publish Sanitized Reports Converted From Word 2007 to PDF, I333-TR-015R-2005.pdf

Hidden Data and Metadata in Adobe PDF Files, I733-028R-2008.pdf

The Adobe website, www.adobe.com, contains many resources on the PDF format and Adobe products, including forums for discussion of product questions.

Table of Contents

1. Introduction	1
2. New Features in Adobe Acrobat Professional X	1
3. Recommended Procedure for Redaction Using Adobe Acrobat X	1
4. Detailed Steps	2
4.1. Turning Off JavaScript	2
4.2. Redacting Information and Apply Redactions	3
4.3. Sanitize Document	5
4.4. Changing Redaction Properties.....	5
4.5. Choosing Remove Hidden Information instead of Sanitize Document	7
4.6. (Optional) <i>File->SaveAs->Optimized PDF</i>	7
5. Conclusion.....	8

Table of Figures

Figure 1: Preferences Dialog	2
Figure 2: Protection Menu	3
Figure 3: Redaction Dialog	4
Figure 4: Redaction Outline Marks	4
Figure 5: Redaction Black Box Marks	4
Figure 6: Sanitize Document Dialog	5
Figure 7a: Default Redaction Properties Dialog.....	6
Figure 7b: Modified Redaction Properties Dialog.....	6
Figure 8: Redaction Mark with Custom Text	6
Figure 9: Remove Hidden Information Check List	7
Figure 10: PDF Optimizer Dialog	8

1. Introduction

Redaction of information from documents is an ongoing challenge. There are multiple tools to assist the user depending on the source file and the desired destination file type. NSA has released several papers on the topic of redaction and the removal of hidden data in Microsoft Word and PDF files. These papers are on the NSA.gov website. Adobe Acrobat Professional X (AAPX) includes improvements on the redaction process and a new feature, *Sanitize Document*, to greatly enhance redaction capabilities. This paper describes the use of AAPX in redaction.

2. New Features in Adobe Acrobat Professional X

Prior versions of Acrobat Pro had redaction functions spread across two menus, and the user had to perform multiple steps to remove hidden data. Even after completing the steps, some complex structures may have retained hidden data, such as layers in diagrams. In version X, Adobe added a *Sanitize Document* function to accomplish a one-button sanitization that removes metadata, embedded content and attached files, scripts, hidden layers, embedded search indexes, bookmarks, stored form data, review and comment data, hidden data from previous document saves, obscured text and images, comments hidden within the body of the PDF file, unreferenced data, links, actions and JavaScripts, and overlapping objects.

The new *Sanitize Document* feature is different from the *Remove Hidden Information* feature which allows the user to choose what content to remove and what to leave in. When the user selects *Remove Hidden Information*, AAPX creates a list of items to remove and the user can un-select things to keep in the document. The user must then click the *Remove* button to remove all of the selected items. On the other hand, *Sanitize Document* removes everything in one step creating a “what you see is what you get” document. In environments where the user must prepare many documents for declassification or for release where no advanced functionality is necessary in the final copy, *Sanitize Document* will save time and reduce the likelihood of process errors.

Complex images with hidden layers had to be handled individually in prior versions of Acrobat Pro. Documents with diagrams, embedded spreadsheets, etc. had to be saved as JPEGs to insure removal of hidden layers – the whole document had to be converted to JPEG images to ensure removal of all hidden layers. AAPX now handles hidden layers correctly with both *Sanitize Document* and *Remove Hidden Information*. The user is given the option to leave those items as is with the *Remove Hidden Information* option.

3. Recommended Procedure for Redaction Using Adobe Acrobat X

Assuming the user is starting the redaction process with a PDF document (rather than a Microsoft Word or other format), these are the high-level steps to perform redaction in AAPX (detailed steps are listed in the next section):

1. Start AAPX.
2. Turn off JavaScript. If you are trying to remove hidden malicious code from a document, it is best to not be hit with it yourself.
3. Open a copy of the original document in AAPX (always work with a copy).
4. Redact Information.
5. Apply Redactions.
6. Sanitize Document.
7. (Optional) Save Document with *File->SaveAs->Optimized PDF*.

4. Detailed Steps

4.1. Turning Off JavaScript

For documents created by the user or from a known source, turning off JavaScript is optional. However, this step is essential for sanitization of documents of unknown origin that might contain malware or other executable content.

JavaScript is one of the things that should be removed from PDF documents during the redaction process when the desire is to produce a “what you see is what you get” document. Since some PDF documents might contain malicious JavaScript, it is best to not be hit with such attacks while trying to prevent them. After opening AAPX and before opening a document, the user should turn off JavaScript:

1. On the top menu, select *Edit->Preferences* to bring up the Preferences dialog
2. Under *Categories*, select *JavaScript*. This brings up the JavaScript palate shown in Figure 1.

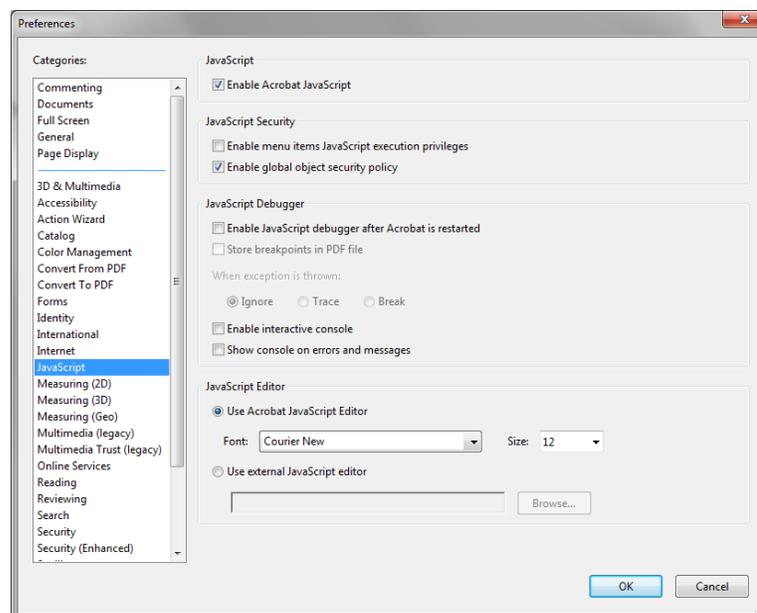


Figure 1: Preferences Dialog

- JavaScript is enabled by default as shown in Figure 1. Uncheck *Enable Acrobat JavaScript* and leave everything else as shown above, then click OK.

Now it is safe to open the document for redaction.

4.2. Redacting Information and Apply Redactions

After opening a document in AAPX:

- Open the Tools menu by clicking on *Tools* on the top right menu bar. This will open a menu that should look like Figure 2 without the Protection menu expanded. Another way to get to this menu is on the top menu bar, select *View->Tools->Protection*.
- Expand the *Protection* menu option which will look like the Figure 2.

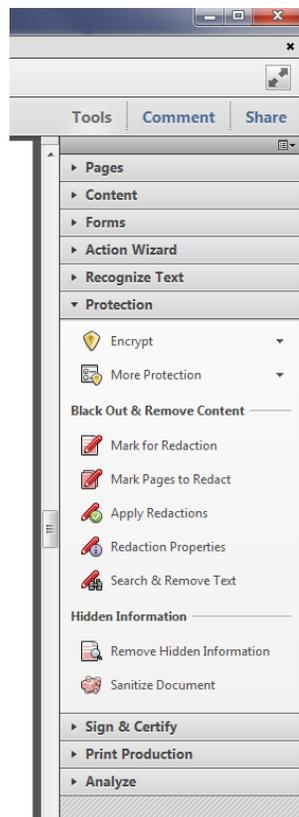


Figure 2: Protection Menu

- Select *Mark For Redaction*. This will display the dialog shown in Figure 3 stating that redaction requires two steps: *Mark for Redaction* and *Apply Redactions*. Click OK.

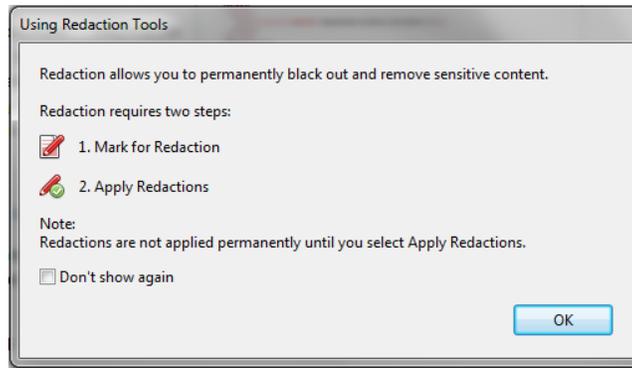


Figure 3: Redaction Dialog

- The cursor is now a redaction cursor. Place the cursor at the beginning of the area you want to redact. Click the left mouse button, hold it down, and drag the cursor over the information to redact. This will create a red outline around the information you are redacting as shown in Figure 4. When you mouse over an outlined area, it will turn black to indicate how the area will look in the final copy.

The newer XML metadata stream is shown in Figure 3. The XMP schemas define how the metadata will be stored, including what fields are available.

Following the XMP entry, a document information dictionary entry may also exist with much of the same data that was in the XML. For



Figure 4: Redaction Outline Marks

- After you have marked all of the areas for redaction, select *Apply Redactions* on the right panel menu. This will display a dialog stating “You are about to permanently remove all content that has been marked for redaction. Once the document is saved, this operation cannot be undone. Are you sure you want to continue?” Select OK.

After *Apply Redactions*, the red outlines are filled with black as shown in Figure 5.

The newer XML metadata stream is shown in Figure 3. The XMP schemas define how the metadata will be stored, including what fields are available.

Following the XMP entry, a document information dictionary entry may also exist with



Figure 5: Redaction Black Box Marks

6. After applying redactions, you will see a dialog box that says “Redactions have been successfully applied. Would you like to also find and remove hidden information in your document?” Select no. You can choose this option from the right panel later, or later chose *Sanitize Document* instead.
7. You can continue to Mark and Apply in the document. When you are finished marking all of the redactions, you are ready to sanitize the document.

4.3. Sanitize Document

After marking and applying all of the redactions for the document:

1. Select *Sanitize Document* from the right panel menu. This brings up a dialog shown in Figure 6.

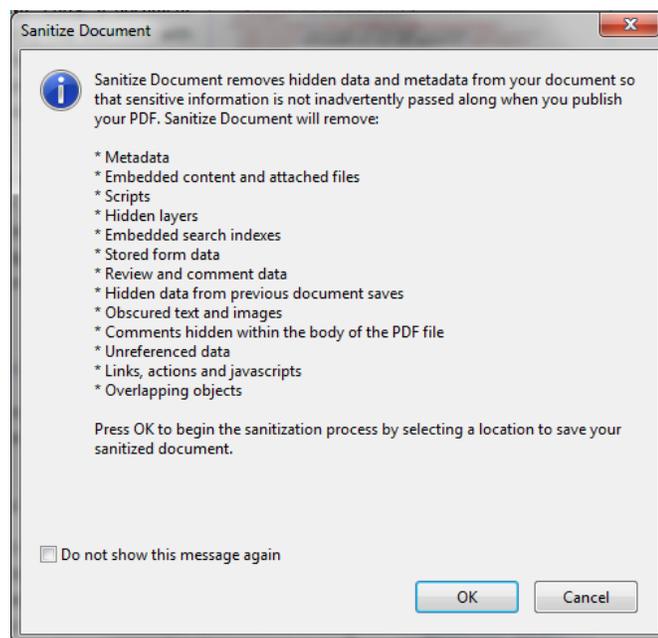


Figure 6: Sanitize Document Dialog

2. Select OK, you will be prompted to save a copy of the document. AAPX tries to minimize the risk of overwriting the original after you have removed all of the functionality of forms and scripts. Save the document. The changes are not final until you save the document. If you started with a copy, just save to whatever file you are working with. If you started with the original, be sure to save to a different filename.

4.4. Changing Redaction Properties

The user can customize the redaction properties rather than using the default black box mark. The right panel menu includes a *Redaction Properties* choice which displays the dialog in Figure 7a.

4.5. Choosing Remove Hidden Information instead of Sanitize Document

Sanitize Document removes all form and JavaScript functionality from the document, all tags, all comments, everything! In those cases where the user needs to leave some functionality in the document, the user can use the *Remove Hidden Information* option and select individual items to leave in the document; however, this reduces the assurance that all hidden data is removed and means the user must save the document as optimized PDF (see next section). *Sanitize Document* includes optimizing the PDF format. The *Remove Hidden Information* option should only be used for special circumstances. *Sanitize Document* is the recommended option otherwise.

When the user selects *Remove Hidden Information*, AAPX searches the document for hidden information and displays a checklist on the left menu pane as shown in Figure 9:

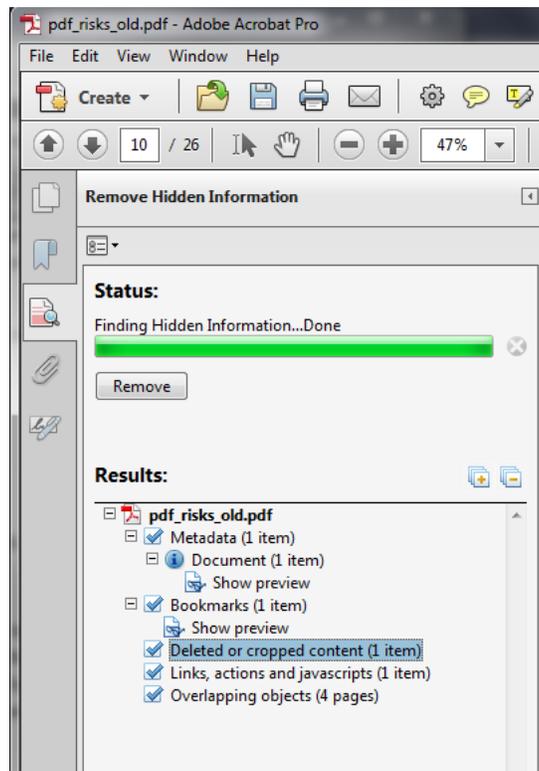


Figure 9: Remove Hidden Information Check List

The user can preview each piece of hidden information identified by AAPX and de-select items to leave in the document. When finished reviewing the checklist, the user has to hit the *Remove* button which brings up a warning dialog requiring the user to hit OK.

4.6. (Optional) File->SaveAs->Optimized PDF

Multiple vendors have applications that create PDF files. These files are not always well-formed or standardized PDF and could contain sensitive or classified remnant data within the file format. The *Sanitize Document* option includes optimizing the PDF format to correct these problems. However, if the

user chose *Remove Hidden Data* instead of *Sanitize Document* (see previous section), the user should then save the file as Optimized PDF.

1. Select *File->SaveAs->Optimized PDF* from the top menu. This brings up the dialog in Figure 10.

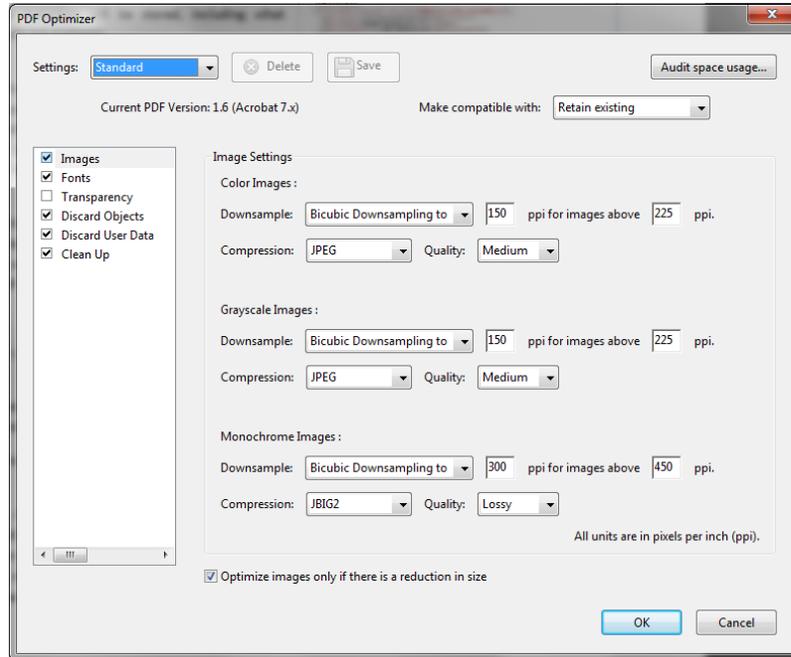


Figure 10: PDF Optimizer Dialog

2. Uncheck *Fonts* to retain embedded fonts. If the embedded fonts are removed and not available on the recipient's system, the system will substitute fonts which could affect the appearance of the document and the alignment of the redaction marks. Some situations may require removal of embedded fonts, but in general redaction tasks this is optional.
3. If you have already run *Remove Hidden Data* or plan to run it, you do not need to change any other options. Hit OK. You will be prompted for a filename, enter a filename and hit OK again.

5. Conclusion

When using AAPX for redaction of sensitive or classified information from documents meant for public release, the steps outlined in this paper with the *Sanitize Document* option are the recommended procedure. This produces a "what you see is what you get" final document.