

PDF White Paper - Final Revised – July 20, 2006

Prepared by PDF Subcommittee of the EPA Web Workgroup

Karen Reshkin, Chair

Amanda Babcock, Angela Palmer, Carol Cotton, Christine Dibble, Christine Tran, Donna Rickett, Jeff Tumarkin, Jennifer France, John Leitzke, Judy Dew, Karen Reshkin, Kathy Einwich, Lorien Wendt, Pat Burke, Peggy Heimbrock, Krista Carlson, Robert Johnson, Sally Martiny, Lawrence Lee

Section 1 - Introduction.....	2
Section 2 - When to use PDFs	2
Section 3 - Accessibility.....	3
Why are PDFs an accessibility problem?.....	3
How to make a PDF accessible (summary).....	3
Why it's sometimes hard to make a PDF accessible	3
What is a "reasonable effort" at making my PDF file accessible?.....	4
If you decide to post a PDF that is not fully accessible, how do you comply with Section 508?	4
Section 4 - Linking to PDFs	5
Required information.....	5
Other best practices for linking to PDFs.....	8
Linking to multiple file formats.....	8
Section 5 - Making large documents more user-friendly	9
Why are large documents a problem?	9
Add bookmarks within the PDF	9
Provide an abstract or an executive summary.....	10
Break up the document into chunks.....	10
Section 6 - Metadata.....	11
Guidelines for adding metadata to your PDFs	11
What is the Document Description?	12
Why is it important?.....	12
How do search engines use PDF metadata?.....	12
Indexing PDF files	13
Section 7 – Techniques for reducing PDF file size	13
Know what you want	13
Use RGB color or go grayscale.....	13
Use the "Save As" command	13
Audit the use of space in the file	14
Optimize images.....	14
Remove unused elements from the document	17
Section 8 – Dealing with scanned PDFs	19
General considerations	19
What about using a scanned image of a person's handwritten signature?	20
Section 9 – Which version of Acrobat Reader should our PDFs be compatible with?.....	20
Section 10 – Does it matter which version of Acrobat I use to create PDFs?.....	20
Section 11 – Security features.....	21
Appendix A – PDF reading order problems – an illustration	22
Appendix B: Sample Cover Sheet.....	23
Appendix C —Checklist for Using PDFs on the Web.....	24
Appendix D - References	26
From EPA.....	26
From Adobe.....	26
Other Resources	26
Appendix E - Results of link testing with visually impaired users.....	28

Note: Any instructions in this paper refer to Adobe Acrobat 7 Professional unless otherwise stated.

Section 1 - Introduction

Opinions on Adobe Acrobat Portable Document Files (PDFs) run the gamut from incredibly useful tool to diabolical monstrosity. People who object to PDF usually say that it's confusing for people to use, especially on the Web, since it requires reader software, either alone or as a browser plug-in (usability problems). They may also mention that PDFs are not inherently 508-compatible (accessibility problems). On the other hand, PDF is a widely used format for presenting documents electronically that look exactly like the printed originals. Although Adobe Acrobat is proprietary software, the reader software is free and available for every commonly used computer platform.

The accessibility aspect of this controversy has fueled many heated discussions over the years. As the federal government, we have an obligation to make our information accessible to everyone, regardless of their abilities or disabilities. More than many other debates in the government Web community, this one stirs up our ethical and moral senses. It is important to find solutions that are compliant, considerate, effective and achievable.

This paper is the result of over three years of research and testing by the PDF Subcommittee of the US EPA Web Workgroup. We want to share the best practices we've developed for addressing accessibility and usability concerns. We hope that EPA staff and contractors who work on our Web site will find this information useful and share it with contractors who prepare PDFs for eventual posting on the Web. We also hope the EPA Web Council will consider our research when developing policies and standards for the EPA Web site.

Timeline (how we got here)

- **December 2002** - First time the PDF Workgroup appears on a Web Workgroup Meeting agenda
- **July 2003** - Introduced Guidelines and Best Practices at Web Workgroup Meeting in Seattle
- **December 2003** - Released Draft PDF Guidance and Short Checklist for Using PDFs on the Web
- **Early 2004** - Our linking best practices were added to the EPA Web guide, along with our other documents. We researched how to make PDFs accessible and concluded that sometimes it takes enormous effort and might not even be possible (using Adobe Acrobat 5 or 6, the versions available at the time).
- **Late 2004** - Jerry Carrillo, then EPA's Section 508 Coordinator, joined our workgroup, and helped us draft an undue burden justification.
- **December 2004** - We presented the draft undue burden justification to the Section 508 workgroup of the CIO Council. They convinced us that an undue burden justification was not appropriate, said that our proposed way of addressing PDF accessibility was 508 compliant, and encouraged us to develop a PDF policy for our agency.
- **Spring 2005** - EPA Web Workgroup members commented on the first draft of this white paper.
- **April 2005** - Adobe Acrobat accessibility expert Greg Piskey gave his first presentations at an EPA Web Workgroup conference. He has presented hands-on training at every conference since then.
- **September 2005** - We presented our second draft of this white paper at the Web Workgroup conference and took further comments.
- **January 2006** - We tested different methods of linking to PDFs and tested them for accessibility and usability with blind and vision-impaired people.
- **April 2006** - We did follow-up testing of linking techniques with blind and vision-impaired people.

Section 2 - When to use PDFs

Using PDF on the Web is a choice; you could present the content in HTML or another file format. Determine whether PDF is the appropriate format to use to post your information on the Web. Reasons to use PDF include:

- the document is longer than 5 pages, or
- you need to preserve the formatting or layout of the original document (e.g. for printing), or

- you need to publish a paper document for which no electronic version exists.

Section 3 - Accessibility

Why are PDFs an accessibility problem?

PDFs generate a picture of how a printed document looks; this makes PDF a useful format for circulating picture-perfect replicas of printed documents electronically. This is extremely useful when you need to preserve the look of a document, whether because of graphics and design or because of official pagination and formatting. Unfortunately, because a PDF is a *picture of a page of text* rather than actual text, PDFs are not inherently accessible to visually impaired people.

You can tag PDFs (using Adobe Acrobat) so that screen readers will be able to recognize the content and read it aloud. This is much like adding alt text to graphics in HTML.

Depending on your document, tagging a PDF may go very smoothly, or it may be very difficult. For some documents it is practically impossible (or exorbitantly time-consuming).

How to make a PDF accessible (summary)

1. **Tag Your Document:** (NOTE: You can skip this step if you converted your original document as a tagged PDF, e.g. in Word.)
 - **If you are creating the PDF from a Microsoft Office application,** choose Adobe PDF > Change Conversion Settings and select "Enable accessibility and reflow with Tagged PDF". Then select OK. Now create the PDF using Adobe PDF > Convert to PDF (or the Convert to PDF icon).
 - **If you are using an existing PDF,** tag the file for accessibility using Adobe Acrobat (Advanced > Accessibility > Add Tags to Document)
2. **Run a Full Check:** (Advanced > Accessibility > Full Check), then look at the report it generates and see what you need to fix.
3. **Check the Reading Order:** Check the reading order and correct what doesn't make sense (Advanced > Accessibility > Touch Up Reading Order)
4. **Include Alt Text for Images:** Add alternate text for images (View > Navigation Tabs > Tags, select the image that needs alt text, then from the Tag window select Options > Properties and fill in the Alternate Text)
5. **Run a Full Check Again:** Repeat step 2 and see what you still need to fix.

Why it's sometimes hard to make a PDF accessible

A) The document was scanned to PDF

You'll have to OCR it, because it literally is a picture of the pages. OCR stands for Optical Character Recognition, and Adobe Acrobat has a built-in OCR feature. When it's done, your document will have a text layer that screen readers can access.

Note: If your document has a graphic that contains text, the Acrobat tagging utility cannot recognize that text, so you'll need to find the graphic in the tagging structure of the document and give it alt text.

B) Reading order is messed up

You can repair it, or you can go back to the source document and take steps to prevent reading order problems in the first place.

Reading order problems may occur when the document contains complex formatting features such as multiple columns, text boxes, tables and text within graphics.

We often lay out documents with columns and text boxes and tables so that they'll be visually appealing when printed. When Acrobat tags a PDF with these features, the resulting code doesn't necessarily follow the same reading order that appears on the page. While you may look at a document and read it top to bottom, column by column, the code in a tagged PDF may skip all over the page. (See Appendix A for an illustration of jumbled reading order.)

C) Long complex documents

You can tag these successfully, too, but it's going to take time and patience.

What is a "reasonable effort" at making my PDF file accessible?

This is the single most time-consuming question that the PDF Subcommittee has considered over the past three years. In the end, **it comes down to a personal judgment call** on the part of the person paying for the work. (If you are a federal employee and you are doing the work yourself, that person is you or your supervisor who is concerned about how you spend your time.) This workgroup cannot formulate an answer that will work well for everyone, but we can suggest things to consider in making your judgment.

1. Is it very hard to make this PDF fully accessible? If not, then maybe you should go ahead and do that. A small effort on your part could make a big difference to someone else.
2. Is this document unusually important or popular? If so, it may be reasonable to spend more effort on making it fully accessible.
3. Is it critical to get this information posted immediately? If so, then you may be able to justify posting a non-accessible version, at least temporarily, to get it up as fast as possible. On the other hand, if it's so critical, maybe it's also very important to have an accessible version out there immediately, too.
4. How much staff time or contractor money you have to spend will certainly figure into your decision.

If you decide to post a PDF that is not fully accessible, how do you comply with Section 508?

Your choices are:

1. **Don't use PDF.** Use HTML. If you must preserve the official pagination and structure of the document, or the graphics, or the signature, this option likely will not work for you. But for some documents it may be something to consider. Each Office and Region will need to decide whether they can offer documents only in HTML format.
2. **In addition to the PDF, provide the content in a more accessible format**, such as a text file (ASCII text) or HTML. (Note: The PDF Subcommittee doesn't think it is practical to post an alternate version of every PDF document. We don't recommend making this a requirement.)
3. **Provide contact information.** No one in the federal Web community seems to have much experience with this, so it's hard to predict what people will request. You'll probably have to handle this on a case by case basis, and work with your audience to give them what they need.

What kind of contact information should we provide, and who should we list?¹ We think it should be left up to your office. There is no "one size fits all" solution for this. A phone number or an email address are good choices. Possible contacts include your library, public information center, public affairs office or the person who created the document. Whatever information you provide, the document should be traceable back to the program which produced it. See Section 4 - Linking, for several ways to provide this information.

In December 2004, the EPA PDF Subcommittee presented these options to the Section 508 workgroup of the CIO Council. They confirmed that these options do satisfy Section 508.

Section 4 - Linking to PDFs²

The basic principle is: give people enough information so that they know what they're getting into when they click that link to your PDF.

Required information

The information you must supply whenever you a link to a PDF:

- title of the document
- PDF
- number of pages
- file size
- a link to the about PDF page

Here's an example:

[How they get those ships inside the bottles \(PDF\)](#) (24pp, 240K [About PDF](#))

Other information that you will provide sometimes

- dimensions for large pages (e.g. for a poster)
- contact information
- EPA publication number (if applicable) and document date

Let's go through the different elements of linking to PDFs.

Link text

Include the title of the document³ and the acronym "PDF" in parentheses in the link text. Our testing with visually impaired users showed that they are able to determine that a file is in the PDF format when links are formatted this way. It also helps users who have no visual impairment.

¹ Doug Wakefield (retired in 2005 from the Access Board), suggested that the contact to provide help on a PDF doesn't need to be the author or someone very familiar with the document. He thinks most requestors would probably only need a printed copy that they can run through a reader and study at their leisure; even someone who isn't familiar with the document could provide that. (We don't know whether most vision impaired people would be able to print the document themselves.) If they want to discuss the content of a document that's inherently visual (a graph or an aerial photograph), maybe it would be best if they could speak or correspond by e-mail with the person who created the image. This is one reason why it's important to include source information in a PDF's Document Description.

² This section has been revised to reflect the PDF linking standard adopted by the EPA Web Council on July 12, 2006.

³ Here's an example of a situation where common sense suggests an exception: a set of documents that are all related to a rulemaking and are grouped together in a list. Usually these documents (the fact sheet, the rule itself, the Regulatory Impact Analysis, etc., all have the same or similar title as the rule, so it seems quite redundant to list the titles -- **instead we use a description of the document as the link** (e.g., Fact sheet, Regulatory Impact Analysis, etc.).

If you're in doubt, read the link text aloud and ask yourself, "Would I know enough about this file if that were all I could hear?"

Tell them how big the file is

After the link, show the number of pages and file size as unlinked text, separated by commas, in parentheses.

For the number of pages, abbreviate using "pp."

For file size, use "K" for files smaller than 1000K and "MB" for files larger than 1000K. For MB, use one decimal place. The goal is to give people a general idea how big the file is, so you don't need to use a lot of decimal places. The standard abbreviation for megabytes is MB. The standard abbreviation for kilobytes is K. It's easier to understand 1.5MB than 1503K.

Document Dimensions

Provide as much description for the reader as possible. If the PDF is a poster, for example, you might provide the original dimensions:

[How Far From the Nearest Road? \(PDF\)](#) (poster 36 x 96 inches, 1 pg., 3 MB)

Contact Information

If you are posting a PDF that you know is not 508 compliant, and you are not posting the same content in some other accessible format (e.g., HTML or text), provide contact information. Add this at the file location or it can be part of the disclaimer text (in the box or paragraph). (If you supply an email address, follow the EPA standard for linking to email addresses: epa.gov/webgovernance/content/projects/standards/emailaddresses.htm)

Examples:

If you have difficulty accessing these PDF documents, please contact [Chris Robinson](#) (robinson.chris@epa.gov) at 202-123-4576 for assistance.

If you need assistance in obtaining the content of this document, please contact EPA's Water Permits Division at 202-564-9545.

Link to EPA's About PDF page

On every page which includes a PDF, you must link to EPA's PDF page (<http://epa.gov/epahome/pdf.html>), which explains what PDFs are and contains a link to the Adobe Web site where people may download the free Acrobat Reader.

There are two ways to do this.

The long way - disclaimer box or paragraph

Provide the following disclaimer text (or similar language) in a box or paragraph.

You will need the free Adobe Reader to view some of the files on this page. See [EPA's PDF page](#) to learn more.

Here are examples from pages that use disclaimer boxes.

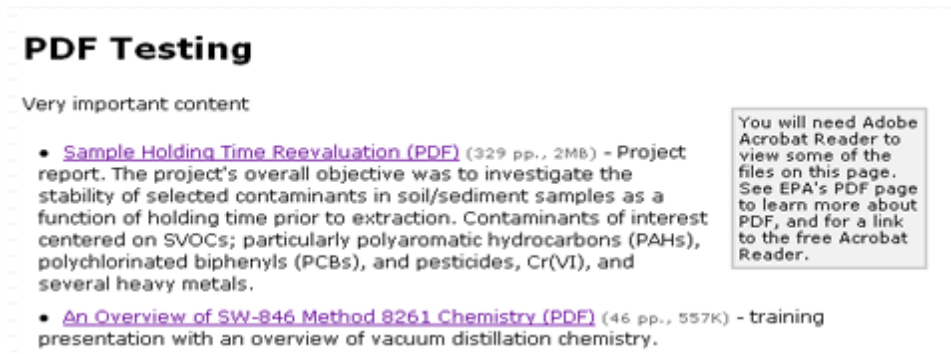


Figure 1: <p class="disclaimer">

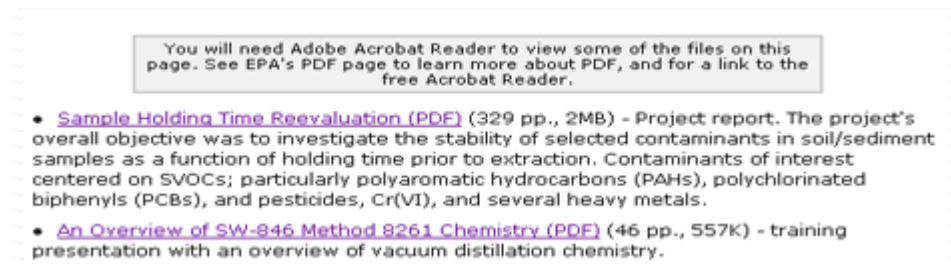


Figure 2: <p class="disclaimer-wide">

Whether box or paragraph, the long disclaimer may appear:

- at the top of a page that includes many PDFs,
- right before a paragraph or list of PDFs, or
- immediately before the first reference to a PDF

Don't put the disclaimer at the bottom of the page or anywhere below the PDF links because people won't see it there.

- The disclaimer text may be in a box but it's not a requirement.
- The style sheet for Template 3 includes styles for PDF disclaimer boxes. Beginning with Template 3, if you use a box for long disclaimers, use either the right-aligned box (Figure 1) or the centered box (Figure 2). <p id="disclaimer">You will need Adobe Acrobat(...)</p>
- You may include contact information in the same box.

When would you use this? When you have a lot of PDFs on one page and you don't want to clutter up the page with inline About PDF links. When the box would be problematic to the design of the page, use a paragraph or inline disclaimer.

The compact way - inline disclaimer (About PDF link)

You may link to the "About PDF" page by adding "About PDF" as the last item in parentheses after the link to the document itself. Here is an example from a page that uses an inline disclaimer.

[Vapor Extraction System Expanded Into Northern Area \(PDF\)](#) (2pp, 40K [About PDF](#))

[Diagram of the vapor extraction system \(PDF\)](#) (1pg, 156K, [About PDF](#)) If you need help accessing the content of this document, please contact the Superfund Division at 663-207-8942.

When would you use this? Use it when the long version is problematic, such as on a home page. Generally the inline disclaimer should be used for pages with five or fewer PDFs. If you have a long page, we recommend using the inline disclaimer text no matter how many PDFs there are, because it's harder for people to miss it that way.

Other best practices for linking to PDFs

- Make the file size information in parentheses appear in a smaller font; this helps the title stand out visually. The style sheet for Template 3 includes a style for PDF file information.
(114 pp, 611K)
- Include a title attribute At the time of writing, screen readers can be set to read the title attribute, but that is not their default setup. That's why we consider using the title attribute helpful but not required.
- It's considerate to include the date of the document. One good place for this is right after the size in parentheses.
- File naming conventions:
 - Use all lower case letters
 - Do not use spaces or special characters in file names. Use hyphens instead of spaces or underscores. (This is necessary for the EPA search engine; spaces and special characters can break your URLs.)
 - Try to make the file name meaningful. It is often returned as part of search results and a meaningful file name can be beneficial.
 - Having a unique file name is helpful when reading web statistics; using the EPA publication number as the file name makes it unique and easy to identify the document without having to open it (e.g., 420s06045.pdf).

Linking to multiple file formats

If the same content is available in both HTML and PDF versions (or PowerPoint, Word, Excel, or any non-HTML file format), one way to handle the situation is:

- link to the HTML page, and
- on the HTML page, offer a link to the PDF version

Here's an example:

Title of Document

EPA420-F-06-034, March 2006

Download a [PDF version of this fact sheet](#) formatted for print (4 pp, 125K, [About PDF](#)).

The advantage of this approach is that people are directed to the most accessible version by default. Once there, they may download a PDF version if they wish. There's no confusion about why there are two links to something with the same title. On the other hand, they have to visit another page before they can find out that there's a PDF if they want it.

Another approach is to link to both items from the same page.

[Haz-ED Glossary](#) - The Glossary defines many of the terms and concepts students will encounter in the exercises and activities. ([PDF version, 10 pages, 27K](#) [About PDF](#))

This lets your users see right up front that there's a PDF if they want it; they don't have to click to another page to find that out. The disadvantages are that multiple links may be confusing, or people may click on the About PDF link instead of the link to the PDF itself.

The usability testing we did with assistive technology users didn't show conclusively whether they had problems with this technique, but it did suggest that they understood we were offering multiple file formats.

If you've got more than one non-HTML version, try this:

[Impact of Best Management Practices on Water Quality of Two Small Watersheds in Indiana: Role of Spatial Scale \(PDF\)](#) (120 pp, 1.02 MB) | [Text Version \(textfile\)](#) (154 K)

This one uses an inline About PDF link:

[Impact of Best Management Practices on Water Quality of Two Small Watersheds in Indiana: Role of Spatial Scale \(PDF\)](#) 120 pp, 1.02 MB [About PDF](#)) | [Word Version \(MS-Word\)](#) (120pp, 540 K)

Section 5 - Making large documents more user-friendly

Following are best practices to help people use your PDFs.

Why are large documents a problem?

Large documents pose two main kinds of problems:

A person may not want to download a really large file. They may not have a fast Internet connection. They may not be willing to spend the time it takes to get the whole document. They may only want part of that enormous document.

It can be easy to get lost in a long document. For most people, flipping to the page you want is easier in a paper document than in a PDF.

You may be able to reduce the file size of large documents (see Section 7 of this paper for techniques).

This workgroup is not comfortable recommending a maximum file size for PDFs, but we do feel that some files are just too big to go on the Web. We recommend using restraint when posting large files.

Add bookmarks within the PDF

Bookmarks can really help people navigate through a long document. If the bookmarks pane displays by default when the document opens, it lets people see the menu of bookmarks.

- Include bookmarks in any PDF ten pages or longer. (You can use them in shorter documents if you think they'll be helpful.)
- Follow the document's Table of Contents, if one exists; otherwise use the major headings of the PDF document.⁴
- Set the initial view to "show bookmarks panel and page" (under Document Properties, Ctrl-D).

For detailed instructions on how to create and manage bookmarks, see the Using Bookmarks topic in Adobe Acrobat Help.

⁴ For an example of a PDF with bookmarks based on the table of contents, see: Town of Pines Administrative Order on Consent (PDF, 47pp, 3.3MB) April 2004 <http://www.epa.gov/region5/sites/pines/pdfs/pines-groundwater-AOC-200404.pdf>

Provide an abstract or an executive summary

On the Web page that links to the document, provide a description (abstract, executive summary, blurb) that will help people learn about the document before downloading it. See an example of an abstract:

Arsenic Removal from Drinking Water by Adsorptive Media (PDF, 1151 Kb, 55 pp)
<http://www.epa.gov/ORD/NRMRL/pubs/600r05159/600r05159.htm>

Break up the document into chunks

When creating a PDF from a very large source document, consider breaking your document up into chapters or sections and creating a separate PDF file for each. This allows your audience to download only the sections that they want.

We recommend that if you chunk a PDF, you should also offer it as one large document as well as one broken up into segments. Why? Because some people will want the whole thing, and it's tedious for them to grab it section by section.

If you decide to break a PDF into sections, it is important to **provide a table of contents or an executive summary on the Web page that links to the document**. Each link should be accompanied by some description of that section.

Here's an example of how to present a large, complex PDF broken up into segments:
<http://www.epa.gov/nerleerd/westmethman.htm>

When you chunk a bookmarked PDF, create the bookmarks first, then break up the document using "Delete Pages" (i.e., snip away the bits that aren't going to be in this chunk and Save As chunk number whatever). This way you won't lose bookmarks. (Using "Extract Pages" may cause you to lose bookmarks.) However, you will wind up with superfluous bookmarks in the smaller chunks. You'll have to go through and delete those by hand.

Identify the chunks

Suppose a chunk of one of your long documents turns up in Google search results, and someone opens it. How will you make sure they can find the other chunks of the document if they want them? Or at least know what original document they came from? How will they be able to tell it's an EPA document? ⁵

When you chunk a PDF, it's important to include document identification information on each file, including the full document file.

At a minimum, you should include the following information:

- EPA logo or "U.S. Environmental Protection Agency" spelled out
- Title of full document

⁵ It may be possible to hide individual chunks from search engines.

Split up the document into chunks and place them into a separate directory on the server that contains a .noindex tag. The .noindex tag is used to build a robots.txt file. That file is used by search engines such as Google, etc. The robots file tells them which directories they should not index. So any directory with a .noindex tag is not indexed. This method makes the PDF chunks invisible and the search only finds the whole PDF or the HTML page. The HTML page will contain links to all the chunks.

Unfortunately, not every search engine will respect the robots.txt file.

No matter how your customer obtains a PDF chunk, they should be able to tell where it came from.

- Title of chapter/section/appendix or description of what smaller file contains (e.g., "Final Rule: pages 750-828")
- EPA publication number of full document, if applicable
- Date of document (month and year)

What's the best way to present document source information?

There are three ways to present this information:

- Attach a cover sheet. You can create a template for your program. See Appendix B for a sample cover sheet.
- Add this information to the top of the first page.
- Add this information to the document as a header and/or footer.

You can add this information in the original program (e.g., Word) or you can add it in Adobe Acrobat to a document that was produced in any program.

Document > Add Headers & Footers

Lets you add formatted text in three columns. You can add a header or footer or both. You can make the information appear on all pages or on just one page.

Document > Add Watermark & Background

Lets you add formatted text or an image (jpg, bmp, or PDF) to a document. You can position this info in the margins and have it appear on all pages or on just one page.

In addition, be sure to fill in the Document Properties/Metadata fields for each file. For a chunk, you can use the same metadata as in the full document file, but add the specific information for that chunk (e.g., "Chapter 1, Calculation Methodology"). See Section 4 for more information on metadata.

Section 6 - Metadata

Guidelines for adding metadata to your PDFs

The most important part of using metadata is that you apply it uniformly to your document collection and that you use it accurately. When adding metadata, follow these guidelines:

- **Title** is important and should be understandable. Include a date if possible (Month, Year)
- **Author** is the organization that published the document. List the organization to a meaningful level. At a minimum include: EPA, AAship or Region, Office or Program, Division. Due to staff turnover, this can be more helpful than listing an employee's name.
- **Subject** should be a short, descriptive summary of the document. Can be the same as the Title field.
- **Keywords** should be specific and descriptive.
- Use lower case. You do not have to list terms in both upper and lower case (e.g., OSWER, oswer) because search engines will recognize the term no matter what case is used.
- Separate words with commas but no spaces (e.g: pesticides,chemicals,methods,study).
- Phrases (e.g., Memorandum of Understanding) can be used as one keyword
- Use at least two keywords and no more than ten.
- Hyphens are allowed, but be consistent

What is the Document Description? ⁶

While viewing or editing a PDF, press Ctrl-D to see the Document Properties window (also found under File/Document Properties). Choose the Description tab view or edit the fields. (Note: You'll also hear people call this Document Summary; that's what it was called in earlier versions of Acrobat.) The Document Description contains four metadata fields: **Title**, **Author**, **Subject**, and **Keywords**.

Why is it important?

It is very important to complete the Document Description information in PDF files for four reasons.

1. Search engine optimization: your document is more likely to be included in search engine results.
2. Clearer search engine description: the correct title will be displayed in search engine results.
3. Document source information: if the author field is filled in, users will be able to find out what office wrote the document if they are looking for contact information.
4. Improved accessibility: The World Wide Web Consortium (W3C) recommends the use of metadata as a Priority 2 Checkpoint for web accessibility.

How do search engines use PDF metadata?

EPA's new search engine, Northern Light, will use the title metadata for the title in its search results. If no title metadata is available, the search results will state "no title available." Likewise, major search engines such as Google and Yahoo use the title metadata for the title displayed in the search results. If there is no title metadata available, the search results will display something like:

C:\My Files\Policy\permits\faqs.wpd

Northern Light will use the subject (or description) metadata for the document description in the search results. ⁷ If no subject metadata is available, the search results will display a system-generated description from the document text. Typically, the document description in the search results is acquired from the body text in external search engines.

Northern Light will compare the keywords metadata to the document content to determine whether they are representative; if they are consistent, documents with keywords that contain search terms will get a higher score. Although not all search engines use the keywords, some do, so it is still worth completing the metadata for search engine optimization.

Completing the metadata for documents that are scanned directly into PDF is especially important if the text appears as an image instead of computer-readable text. Since the text of those documents is not searchable, the search engine will have to rely on the metadata.

⁶ Acrobat also has something it calls Document Metadata. You can view this feature Acrobat 7 by clicking on "Additional Metadata" from within the Document Description window. All the information you enter in the Document Description fields transfers to the Metadata fields.

Document Description (Ctrl-D) if you fill this field	Advanced > Document Metadata the contents will show up in this field
Title	Document Title
Author	Author
Subject	Description
Keywords	Keywords

⁷ At this point, this feature is not working properly, but they're working on it.

Although search engines don't use the Author field, this information can be helpful to users who want to find out what office or program published the document. This can be particularly important to users who don't know or remember what web site it came from. If the user needs to contact the office for an accessible version, this might be the only place to look for the contact information.

Indexing PDF files

EPA's previous search engine indexed new or changed PDF files on a daily basis, and all others weekly, although not perfectly. Northern Light will aim for any document to be indexed or re-indexed within 48 hours.

Check your PDF metadata reports

OEI has designed PDF metadata reports as a tool to help document owners improve the metadata in their PDF files. As stated on the EPA PDF metadata page: "Well-constructed, descriptive metadata improves search rankings, and the appearance and usefulness of search results. Good metadata also helps us to share data with other organizations." These reports are available on the EPA PDF metadata page at: epa.gov/webmast1/metareports/pdf/.

Section 7 – Techniques for reducing PDF file size ⁸

You can reduce file size by choosing File > Reduce File Size but you will not have control over what Acrobat does to your file. The quality of your graphics might be reduced too much and embedded fonts might be removed.

Know what you want

Before you take steps to reduce the size of a PDF file, consider how it will be used. If people will mostly view it online, then reducing the document's resolution is probably a good move. However, if you expect most people to print it and use the paper document (e.g. a map or pamphlet), then it may not print well if you reduce the document's resolution.

If your original document was created to be printed or sent to press, the graphics were probably set at high resolution. It is not ideal to use these files for a Web ready PDF. Once the original document is finalized, a copy of the document should be made to be used solely for creating a Web PDF. This way, the high resolution graphics can be replaced with lower resolution jpeg files or tiff files without affecting the quality of your original document.

Use RGB color or go grayscale

For Web-only PDFs, if you have a choice, use the RGB color space for your PDFs versus the CMYK color space. RGB 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. CMYK is usually needed only for documents that will be professionally printed.

And if you can get away with it, grayscale or black and white images will save even more space.

Use the "Save As" command

After you have made any final changes to the Adobe PDF document, choose File > Save As. Save the document with the same name to overwrite the original with your changes.

⁸ Note: Material in this section was adapted in part from Adobe's Acrobat Web tutorial.

When you choose Save As, Acrobat rewrites the entire PDF document as efficiently as possible. When you choose Save, changes are appended to the file, which may increase the file size. By default, Acrobat also optimizes a PDF document for Fast Web View when you save it using the Save As command. Documents optimized in this way can be downloaded one page at a time from a Web server or network, reducing the time it takes to access and view them. To check and make sure that this option is selected, do one of the following:

- *Acrobat 6 and 7:* Go to Edit > Preferences. From the categories on the left, select General. Make sure that 'Save As optimizes for Fast Web View' is selected.
- *Acrobat 5:* Go to Edit > Preferences > General. From the menus on the left, select Options. Make sure that 'Save As optimizes for Fast Web View' is selected.

Audit the use of space in the file

The audit results can help you identify the most effective ways to reduce the file's size by showing which elements are significantly large. The audit reports the total number of bytes used by fonts, images, bookmarks, forms, and comments. It also shows the total file size. To view the audit:

- *Acrobat 6 and 7:* Choose Advanced > PDF Optimizer. Select the Audit Space Usage button. The audit results list the bytes used by each element and its percentage of the document's total size.
- *Acrobat 5:* Choose Tools > PDF Consultant > Audit Space Usage.

Optimize images

There are two ways of optimizing your images. The first is done in Acrobat after the PDF has been created, and the other is done in the distiller settings before creating the PDF.

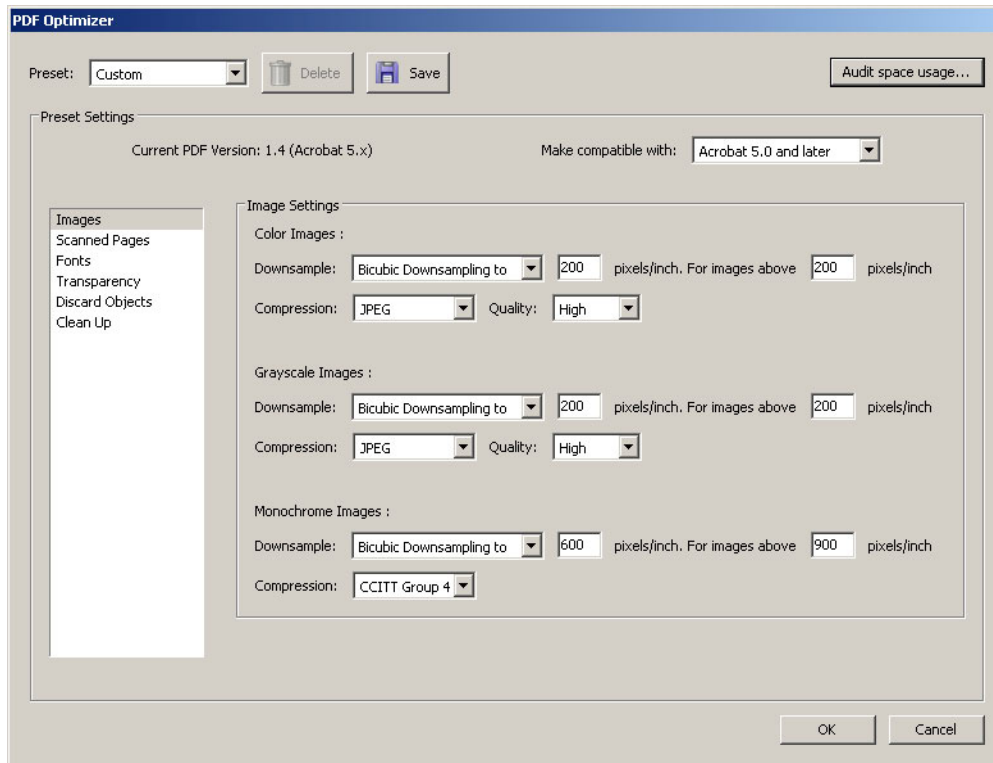
(see following page for illustration)

In Acrobat

When using PDF Optimizer to optimize the images, it is a good idea to save the optimized PDF file with a new name. This way you can test different settings and leave the original PDF document unchanged. If you reduce the pixels per inch too low, the graphics may be unreadable. It is a good general rule not to go below 100 pixels/inch. To use the PDF Optimizer, do one of the following:

Acrobat 6 and 7: Choose Advanced > PDF Optimizer.

On the Images tab, select compression options for color, grayscale, and monochrome images. Or select Enable Adaptive Compression and drag the slider to balance file size and quality.



Downsampling lowers the resolution of images. Compression eliminates unnecessary pixel data. In general, you should use a compression setting of JPEG or JPEG2000 for photographs and other images in which color changes gradually. ZIP is good for illustrations with large areas of solid, flat color. JBIG2 works well for monochrome images. If you choose JPEG or JPEG2000 compression, select a quality level to determine how much pixel data is removed. Lossless (available only for JPEG2000) retains all pixel data.

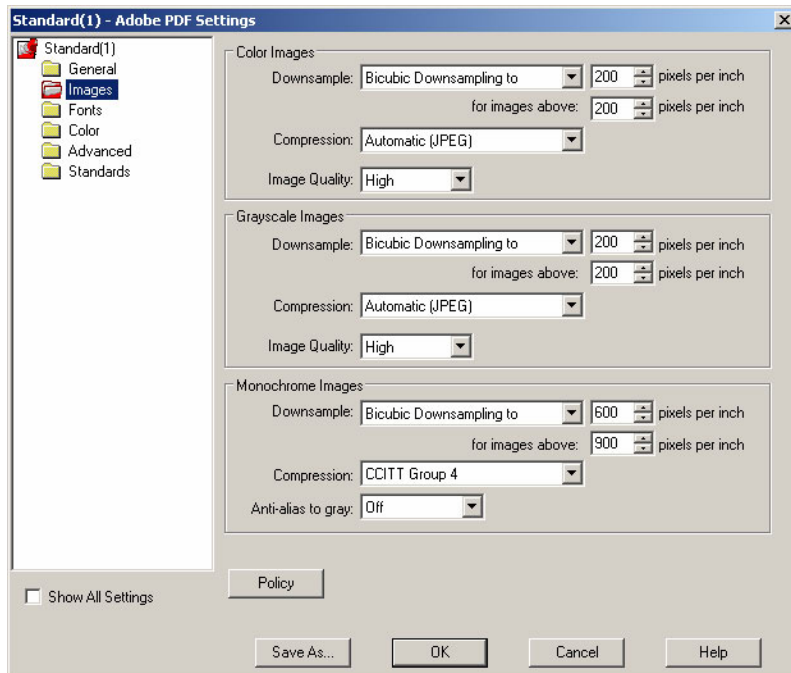
Adaptive Compression applies advanced image processing to images in the PDF file including segmentation and halftoning. Use the slider to set the balance between the smallest file size and the maximum image quality.

In Distiller

If you use Acrobat Distiller to create your PDFs, it has similar options for image compression.

Distiller 6 and 7: Choose Setting > Edit Adobe PDF settings, then select the Image tag.

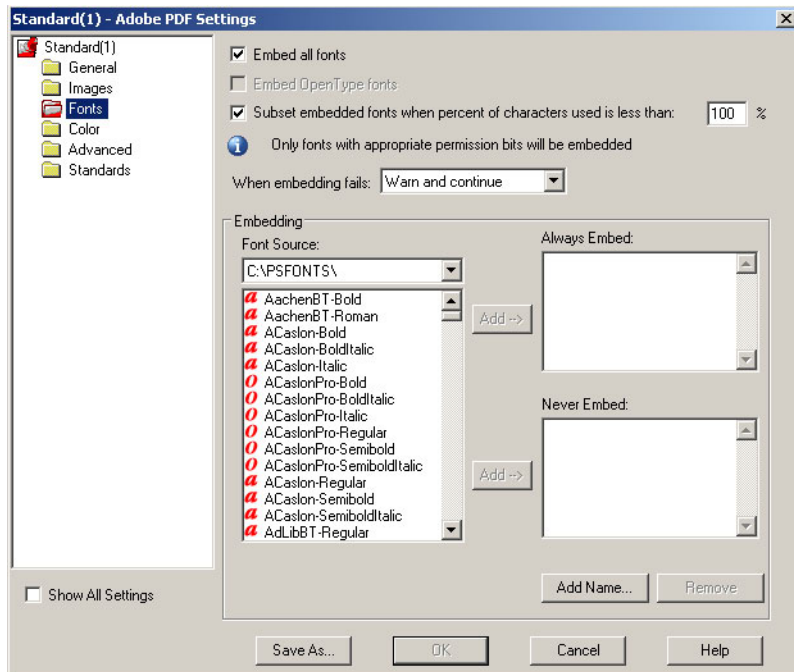
Distiller 5: Choose settings > Job Options, and then select Compression.



Embed only the fonts you need

Make sure you embed only fonts that are used in your document. Fonts often account for a large percentage of a PDF document's file size.

Distiller has an option that sets certain fonts to be always embedded. This can result in fonts being embedded that are not necessary. Acrobat 6 and 7 allow you to remove them easily by choosing Advanced > PDF Optimizer and selecting the Fonts tab. Also, make sure you select the check box labeled "Subset embedded fonts when percent of characters used is less than: 100%".



Distiller 6 and 7: Choose Settings > Edit Adobe PDF settings, then select the Fonts tag.

Sometimes you can safely remove fonts, but do this only if you know that they are already installed on the computers of the people who will read your PDF document. If you remove a font that isn't available, Acrobat will pick a substitute font when the document is opened on their computer. Unless this is a basic font, such as Arial, Helvetica, Times, or Times New Roman, or similar basic fonts, the result can be unpredictable.

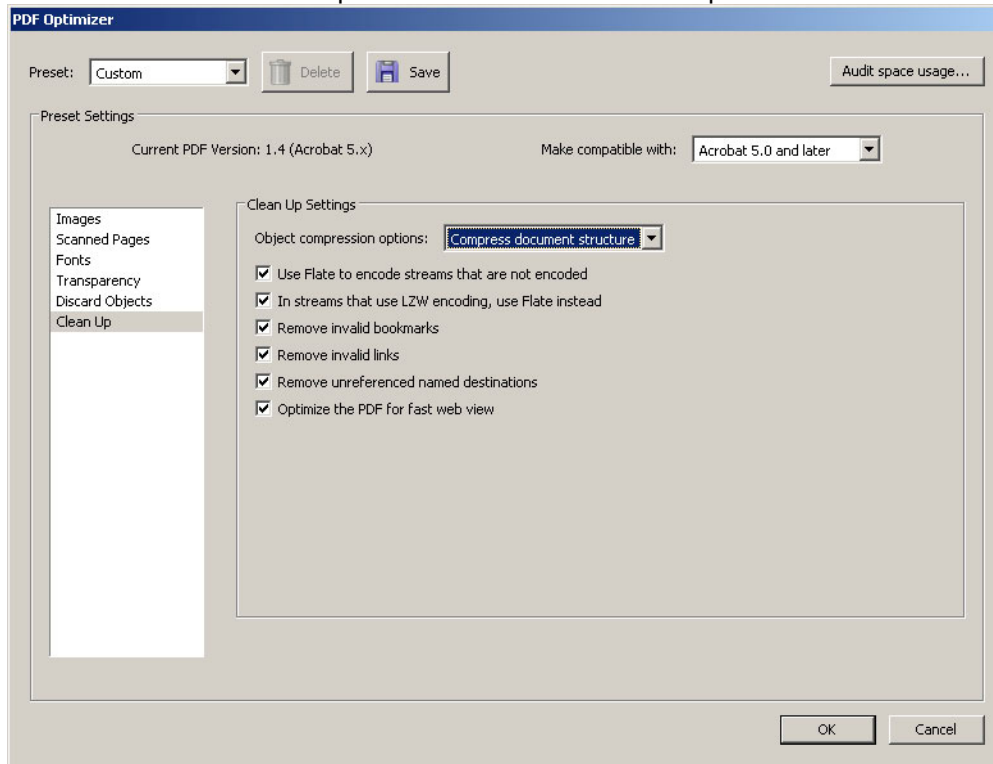
Try to minimize the number of fonts that you use in a document. Each font you use adds to the size of your document.

Remove unused elements from the document

Acrobat 6 and 7 and have a Clean Up tab and a Discard Objects tab, which lets you remove elements from the PDF document that you don't need. The options that are selected by default do not affect functionality, but other options may. Use caution when selecting unfamiliar options. If you choose to experiment, save the optimized PDF file with a new name so that you leave the original PDF document unchanged.

Clean Up Settings

Choose Advanced > PDF Optimizer and select the Clean Up tab.

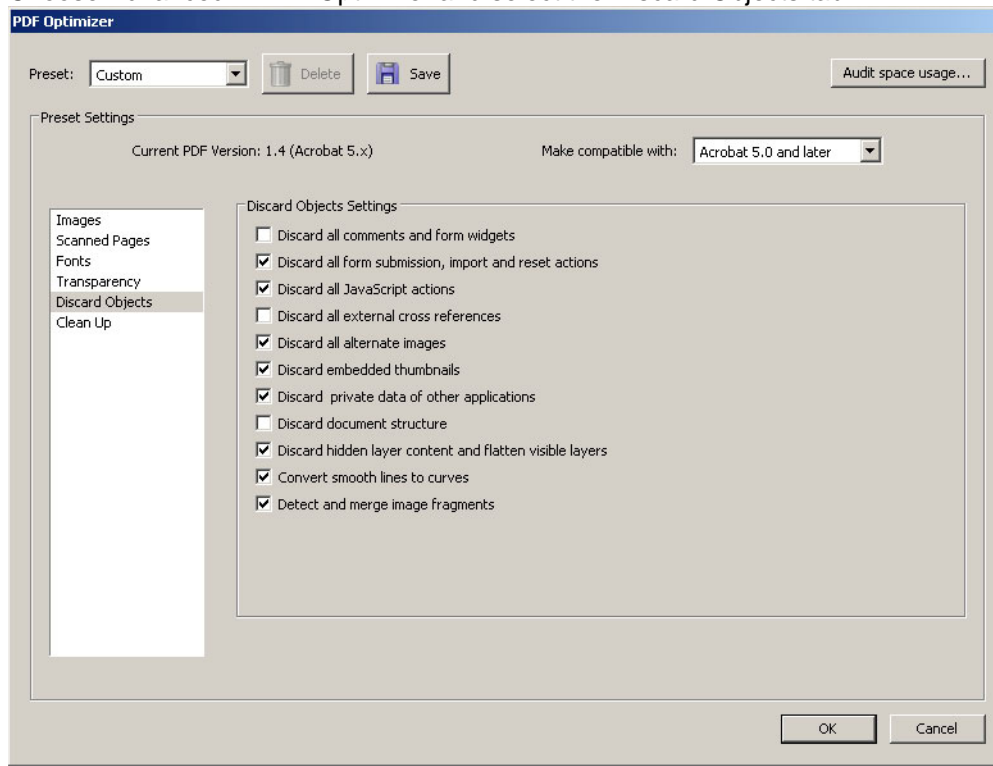


Acrobat 5 has a similar, but limited option. Acrobat 5 also has a feature that will remove unused bookmarks. Choose Tools > PDF Consultant. Select Optimize Space instead of Audit Space Usage.

After selecting your options, do a 'Save As.'

Discard Objects Setting

Choose Advanced > PDF Optimizer and select the Discard Objects tab.



Warning: If your PDF has hyperlinks or is tagged, **do not** select the following check boxes:

- Discard all comments and form widgets
- Discard all external cross references
- Discard document structure

Section 8 – Dealing with scanned PDFs

General considerations

We don't recommend scanning paper documents to post on the Web. You'll get better accessibility results with electronic PDF conversion (Convert to Adobe PDF or Print to Adobe PDF). Scanned documents are images instead of text, so anyone using assistive technologies will not be able to "see" the document unless you use Paper Capture in Adobe Acrobat or some other Optical Character Recognition (OCR) software.

- If a document is older than July 2001, you do not need to do anything to it. If it is a frequently used file, we recommend that you make the material accessible, but that's not a requirement under Section 508.
- If it is a recently scanned document, you will need to post an accessible version or provide contact information so that people can contact you to request an accessible version as needed. (See Section 1 on accessibility.)

What about using a scanned image of a person's handwritten signature?

This might sometimes be the only reason for a PDF, but not too often. We often make documents available electronically to satisfy the E-FOIA requirements⁹. Attorneys preparing for court are frequent requesters of EPA records and may need a copy of the signature page.

Our recommendations:

1. In cases where the record(s) has been requested 3 times or more or when the agency anticipates the documents may be requested frequently, we should **make all but the signature page accessible; in other words, only scan the signature page.**
2. If you're presenting the document as HTML, in place of the signature add "original signed by" before including the name of the signer. After the signer's name include a link to the PDF copy of the signature page.
3. It's especially important to add metadata to a scanned PDF, particularly if you haven't used OCR on it. Since the text of such documents is not searchable, search engines will have to rely on the metadata you provide.

Section 9 – Which version of Acrobat Reader should our PDFs be compatible with?

We recommend saving PDFs to be compatible with Version 5 Readers. As we write this document, multiple versions of both the Adobe PDF creation software and the free reader are available. (See <http://www.adobe.com/products/acrobat/acrrsystemreqs.html> for complete details on the reader.) The product changed its name from Acrobat Reader to Adobe Reader with version 6.0.1. For people who still run Windows 98 (not 98SE), the latest version they can use is 5.1.

Make compatible with Acrobat 5

How to: Choose Advanced > PDF Optimizer. Under "Make Compatible with" select "**Acrobat 5 and later**". Since PDF Optimizer keeps the last used settings it is a good idea to check Images, Scanned Objects, Fonts, Transparency, Discard Objects, and Clean up, to make sure the desired settings are selected.

You can also make your file compatible with Acrobat 5 by choosing File > Reduce File Size but you will not have control over what Acrobat does to your file. The quality of your graphics might be reduced too much and embedded fonts might be removed.

Section 10 – Does it matter which version of Acrobat I use to create PDFs?

We recommend getting the latest version of Adobe Acrobat Professional (currently Adobe Acrobat 7.0 Professional). Thankfully, Adobe has not changed the name of its PDF creation software; it remains Adobe Acrobat. However, there are now multiple editions of this product too: Professional, Standard, Elements, Capture... and versions 5 through 7 are in wide use at this time. The professional version has more accessibility features; you can compare them at <http://www.adobe.com/products/acrobat/matrix.html>.

⁹ Electronic FOIA Amendments of 1996: Three or more requests under FOIA for the same or substantially the same record qualify as a "frequent request".

Section 11 – Security features

Acrobat has security features you can use to prevent printing, copying, copying or extracting the content, and more. These can be defeated by people with readily available software and a little persistence. So while the security features may dissuade some people, all they really do is slow down a determined person who wants to defeat them.

Security restrictions can be a problem for legitimate users too, because if you need to go back and make changes to a document later on, you may not be able to if you don't have the password. This can happen when you inherit a document that someone else has created. Then you'll need to track down the password or the pre-PDF original, or hack the document yourself.

We don't recommend using security features to "lock down" some or all aspects of a document. If you have contractors preparing PDFs for you, specify that they should not use any document security restrictions.

Redacting PDFs is a different security issue. Some agencies have been embarrassed when they redacted documents in ways that could be defeated (e.g. masking words instead of deleting them).

You can read more about this in a paper published by the National Security Agency. There are also tools available for Microsoft Word that will remove change history from a document (so that people cannot see what edits were made). Please see Appendix D for these references.

Appendix A – PDF reading order problems – an illustration

1. This is what the page looks like.

Introducing...

The Nease Chemical site has a new remedial project manager. **Mary Logan** has spent almost 20 years with EPA in both Region 5 (Chicago) and Region 2 (New York). Mary has degrees in biology and environmental health sciences from the University of Chicago and University of Illinois respectively.

Contact EPA

If you have any questions, concerns, comments or want to be on the mailing list for news about the Nease Chemical cleanup, contact these EPA representatives:

Susan Pastor

EPA Community Involvement Coordinator
(312) 353-1325 or (800) 621-8431, weekdays 10 a.m. - 5:30 p.m.
pastor.susan@epa.gov

Mary Logan

EPA Remedial Project Manager
(312) 886-4699 or (800) 621-8431, weekdays 10 a.m. - 5:30 p.m.
logan.mary@epa.gov

As this project is being done in cooperation with Ohio EPA, questions can also be directed to:

Sheila Abraham

Site Coordinator
Division of Emergency and Remedial Response
Ohio EPA Northeast District Office
(330) 963-1290
sheila.abraham@epa.state.oh.us

2. When you try to select text, it doesn't highlight everything in the correct order.

Introducing...

The Nease Chemical site has a new remedial project manager. **Mary Logan** has spent almost 20 years with EPA in both Region 5 (Chicago) and Region 2 (New York). Mary has degrees in biology and environmental health sciences from the University of Chicago and University of Illinois respectively.

Contact EPA

If you have any questions, concerns, comments or want to be on the mailing list for news about the Nease Chemical cleanup, contact these EPA representatives:

Susan Pastor

EPA Community Involvement Coordinator
(312) 353-1325 or (800) 621-8431, weekdays 10 a.m. - 5:30 p.m.
pastor.susan@epa.gov

Mary Logan

EPA Remedial Project Manager
(312) 886-4699 or (800) 621-8431, weekdays 10 a.m. - 5:30 p.m.
logan.mary@epa.gov

As this project is being done in cooperation with Ohio EPA, questions can also be directed to:

Sheila Abraham

Site Coordinator
Division of Emergency and Remedial Response
Ohio EPA Northeast District Office
(330) 963-1290
sheila.abraham@epa.state.oh.us

3. This is what you hear:

"Introducing... The Nease Chemical site has a Mary Logan has spent almost 5 (Chicago) and Region 2 (New and environmental health sciences from the University of Chicago and If you have any questions, concerns, comments or want to be on the mailing list for news about the Nease Chemical cleanup, contact Susan Pastor Coordinator (312) 353-1325 or (800) 621-8431, weekdays 10 a.m. - 5:30 p.m. Mary Logan (312) 886-4699 or (800) 621-8431, weekdays 10 a.m. - 5:30 p.m. logan.mary@epa.gov As this project is being done questions can also be directed to: Sheila Abraham Site Coordinator Remedial Response (330) 963-1290 sheila.abraham@epa.state.oh.us new remedial project manager. 20 years with..."



EPAXXX-X-XX-XXX
Month Year

Appendix B: Sample Cover Sheet

**This is the Title
of the Full Document:
Including the Subtitle if there is one**

This is the Title of the Section
or Chapter or Appendix

Name of Division
Name of Office
U.S. Environmental Protection Agency

You can use this space to add a disclaimer or contract information
if the document is associated with or produced from a contract.

Appendix C —Checklist for Using PDFs on the Web

April 2006, EPA Web Workgroup PDF Subcommittee

Note: All directions assume you are using Adobe Acrobat 7 Professional.

1. **Determine whether PDF is the appropriate format to use** to post your information on the Web. Reasons to use PDF:
 - the document is longer than 5 pages, or
 - you need to preserve the formatting or layout of the original document (e.g. for printing), or
 - you need to publish a paper document for which no electronic version exists.
2. **If you are creating the PDF yourself, start with a "clean" original document.** Keep the format as simple as you can and use styling built into the word processor. This makes it easier to tag the document for accessibility.
 - Avoid placing graphics, text or tables side by side.
 - Use standard fonts.
 - Use styles, like Heading 1, Heading 2, etc.
 - Use the function for creating bulleted and numbered lists rather than formatting list items manually.
 - Use tabs and columns rather than spaces and tables for laying out your text.
 - Add alt text to graphics.
3. **Create a PDF using Convert to Adobe PDF.** Start with any authoring software (Word, Excel, etc.). If you are creating the PDF from an MS Office application, choose Adobe PDF > Change Conversion Setting and select Enable accessibility and reflow with Tagged PDF. Click OK. Now Select Adobe PDF > Convert to Adobe PDF. (Do not use File > Print > Adobe PDF that you used in previous versions of Acrobat unless the Adobe PDF menu is unavailable in your authoring software.)
4. **Create a Document Description.** Search engines use metadata from the Document Description in their search results.
 - Open the Document Description File > Document Properties (or Ctrl-D)
 - Select the "Description" tab.
 - Title** is important and should be understandable. Include a date if possible (Month, Year)
 - Author** is the organization that published the document. Include US EPA, AAship or Region, Office or Program, Division. Only rarely will you use an individual author's name.
 - Subject** should be a short, descriptive summary of the document. Can be the same as the Title field.
 - Keywords** should be specific and descriptive. Use lower case. Separate words with commas but no spaces (e.g: pesticides,chemicals,methods,study). Phrases (e.g., Memorandum of Understanding) can be used as one keyword. Use at least two keywords and no more than ten.
5. **Create Bookmarks.** Select text and use Ctrl-B to insert a bookmark.
6. **Set the Initial View to show bookmarks.** File > Document Properties. Select the "Initial View" tab. Under "Document Options, Show:" select Bookmarks Panel and Page.
7. **(Optional) Create Web links from URLs in text.** Advanced > Links > Create from URLs in document.
8. **If you break a large document into sections,**
 - A. **Identify each section on the first page. Include:**
 - EPA logo or "U.S. Environmental Protection Agency" spelled out
 - Title of full document
 - Title of chapter/section/appendix or description of what smaller file contains (e.g., "Final Rule: pages 750-828")

- EPA publication number of full document, if applicable
- Date of document (month and year)

B. Offer the whole PDF as well as the sections (and include source information on the full document file, too.)

Accessibility Steps

1. **Tag Your Document:** (NOTE: You can skip this step if you converted your original document as a tagged PDF, e.g. in Word.)
 - If you are creating the PDF from a Microsoft Office application, choose Adobe PDF > Change Conversion Settings and select "Enable accessibility and reflow with Tagged PDF". Then select OK. Now create the PDF using Adobe PDF > Convert to PDF (or the Convert to PDF icon).
 - If you are using an existing PDF, tag the file for accessibility using Adobe Acrobat (Advanced > Accessibility > Add Tags to Document)
2. **Run a Full Check:** (Advanced > Accessibility > Full Check), then look at the report it generates and see what you need to fix.
3. **Decide whether to finish making the PDF accessible** or make the content accessible another way.
4. **If you decide not to make the PDF fully accessible,**
 - Use HMTL instead of PDF or
 - In addition to the PDF, provide the content in a more accessible format, such as HTML or a text file (ASCII text), or
 - Provide contact information.
5. **If you *will* be making the PDF accessible, keep going:**
6. **Check the Reading Order:** Check the reading order and correct what doesn't make sense (Advanced > Accessibility > Touch Up Reading Order)
7. **Include Alt Text for Images:** Add alternate text for images (View > Navigation Tabs > Tags, select the image that needs alt text, then from the Tag window select Options > Properties and fill in the Alternate Text)
8. **Run a Full Check Again:** Repeat step 2 and see what you still need to fix.
9. **Fix any accessibility errors** identified. (see adobe.com/enterprise/accessibility/ for more info)

Follow EPA's PDF linking standards, posted at epa.gov/webguide/create/link_standard.html

[How they get those ships inside the bottles \(PDF\)](#) (24pp, 240K [About PDF](#)), November 1992

- title of the document
- PDF
- number of pages
- file size
- a link to the about PDF page (epa.gov/epahome/pdf.html)
- dimensions for large pages (e.g. for a poster)
- contact information if needed

Appendix D - References

From EPA

EPA's "About Portable Document Format Files" page:
<http://www.epa.gov/epahome/pdf.html>

EPA's PDF Subcommittee on the intranet has many resources:
<http://intranet.epa.gov/webgroup/committees/PDF/index.html>

How to link to a specific page within a PDF
<http://intranet.epa.gov/webgroup/resources/index.html#pdflink>

PDF file summary problems in your TSSMS account:
<http://www.epa.gov/webmast1/metareports/pdf/>

From Adobe

See what Adobe has to say for itself and its PDF products
<http://www.adobe.com/products/acrobat/main.html>

Adobe Acrobat accessibility resources
<http://www.adobe.com/enterprise/accessibility/main.html>

Creating Accessible PDF Documents with Adobe Acrobat 7.0 (PDF, 115pp, 10MB)
http://www.adobe.com/enterprise/accessibility/pdfs/acro7_pg_ue.pdf

Other Resources

Planet PDF – the home of the PDF community
<http://www.planetpdf.com/>

Section 508.gov
<http://www.section508.gov/>

Facts and Opinions About PDF Accessibility by Joe Clark, August 22, 2005
http://www.alistapart.com/articles/pdf_accessibility

Steps Toward Integrating Accessibility into Development of an Internet Option for the 2010 U.S. census
Elizabeth D. Murphy, October 2005
<http://www.census.gov/srd/papers/pdf/rsm2005-11.pdf> (PDF, 10pp, 161K)

Redacting with Confidence: How to Safely Publish Sanitized Reports Converted From Word to PDF
National Security Agency Report #I333-015R-2005, December 2005
<http://www.fas.org/sgp/othergov/dod/nsa-redact.pdf> (PDF, 14pp, 681K)

Federal Computer Week article about PDF/A
<http://www.fcw.com/article89385-06-27-05-Print>

Jakob Nielsen's Alertbox, August 29, 2005: Open New Windows for PDF and other Non-Web Documents
http://www.useit.com/alertbox/open_new_windows.html

PDF White Paper - Final - April 24, 2006

Jakob Nielsen's Alertbox, July 14, 2003: PDF: Unfit for Human Consumption
<http://www.useit.com/alertbox/20030714.html>

Jakob Nielsen's Alertbox, June 10, 2001: Avoid PDF for On-Screen Reading
<http://www.useit.com/alertbox/20010610.html>

Appendix E - Results of link testing with visually impaired users

(This is a separate PDF file.)