

PDFs and Accessibility

Last Modified on Friday, 23-Dec-2022 09:17:59 GMT

This is a guide to PDF, how it affects people trying to use assistive technology (A.T.) with it, and what can be done. It assumes familiarity with A.T., whether a screenreader or magnifier or reading toolbar like ClaroRead, and with Adobe Reader. It is designed for organisations producing PDF files and for A.T. practitioners to better-understand the issues. It may be of use to Adobe Reader users who want to understand how best to approach PDFs.

Chapter 1 describes PDF, how it fits into business processes, and why it is popular. It tells you how you can approach the production of accessible PDF files.

Chapter 2 describes the specific technical problems with PDF for the “traditional” screenreader user, why it can be hard to use, and the solution (tags) that Adobe has invented to help.

Chapter 3 describes how Adobe Reader handles a PDF and the implications for how a PDF should be structured.

Chapter 4 catalogues the visual display options for Adobe Reader and how they interact with the accessibility features.

Chapter 5 provides a set of “best” settings for Adobe Reader or for accessible PDF files.

Chapter 1: The Adobe Portable Document Format (PDF)

PDF does a great job for what it is designed to do: it describes what should be printed on paper. It was invented so that people creating documents on the new desktop publishing technology in the 1980s and 1990s could send them to printers and have them appear as they intended – font, colours, layout.

So PDF works in terms of pages, and what images and text goes where on each page. It has no idea of things like headings, or columns, or chapters, or words or sentences.

PDFs fit into traditional business print processes very well, and still does:

- Writers write text in a word processor (Microsoft Word, Pages, OpenOffice). Graphic artists create images (Adobe Photoshop, Adobe Illustrator, Inkscape, GNU Image Processor).
- Graphic designers and editors lay out the text and images in a desktop publishing program to create the brochure or report or book (Adobe InDesign, Microsoft Publisher, Quark Express, Scribus).
- The ready-to-print document is exported from the desktop publishing program as a PDF file and sent to managers and stakeholders to review. Minor corrections are made and the final PDF sent a commercial printer. Because PDF cares about font, and colour, and layout, the editor can be confident that the printed document will look just as they intended.

The problems arose when people started trying to use PDFs not for print but for online distribution and eBooks and other non-print purposes. There were and are good reasons for this. The existing business processes had already produced millions of PDF files and continued to produce new PDF files as the final, edited format, so it was no extra work. PDF is an open format, so every platform has a PDF reader and they all display documents the same – you can’t send a Microsoft Word document to someone on a different platform and count on it looking the same. Adobe has always provided a free reader, Adobe Reader, for most users. PDFs let you keep your publications in your corporate font and colours and style, so your manager is happy. There is an incorrect belief that PDF files cannot be edited or changed, so people are

confident that their documents will always present their intended message. And finally, PDF gives you a way to collect images and text together in a single file, which gave it the edge over that other ubiquitous format, HTML, when you want to email your report.

It is likely, then, that any organisation produces PDFs as their main form of content production, second only perhaps to their website. Often the substantive content on the website itself is in PDF – reports, brochures, newsletters.

But while PDF is convenient for creators and distributors of content, it is not for some consumers of content, notably people who are not in a position to use Adobe Reader to display PDF on a standard desktop system (or print the document!) and read the content visually off the page. People who use assistive technologies like screenreaders and magnifiers have considerable problems with PDF, described in more detail in Chapter 2.

What can be done?

There are two options: stop using PDF and make the PDFs you produce accessible.

For the reasons given above you are probably not in a position to stop producing PDFs. You will still be delivering files to printers. You will still be doing your final editing and approval in PDF. The final PDF file is your canonical document for distribution. You could change your business process: for example, when the text for your documents are created in a word processor, before it is laid out, it is in a format that makes sense for assistive technology – big lumps of text. But that means duplicating effort in your business process. The flow of content from writer and artist to final document has to be performed twice, once with PDF as the intended output and once with an alternative format in mind. This can be expensive. It could backfire if you choose the “wrong” alternative format. And the extra delay to produce the alternative format is likely to be a problem.

So you are probably going to want to keep creating PDF files, and make them accessible after the fact. There are several ways of doing this.

- You can take just one step back, to the desktop publishing program that produced the PDF, and use its ability to support accessible PDF output. For example, Adobe InDesign lets you set reading order and created tagged documents when you output to PDF. However, as noted above, this may not be just one step: it may be many steps. This probably requires the most in-house investment in time and skills.
- You can use another application to take the PDF and make it accessible. They vary in terms of cost, ease-of-use, and the ability to automate the process. This can be easy or hard but is quite cost-effective depending on the tool.
- You can send your PDF files to an external agency, which will use one of the afore-mentioned tools to make the PDF more accessible. This is the easiest but most expensive option. Your printer may offer this service, although PDF accessibility is a specialised field.

Whichever approach you take, if you are interested in the accessibility of your PDFs, the rest of this document will help you understand the issues.

Chapter 2: The basics of PDF Accessibility – Tags

PDF is a format designed for printing and displaying on the screen, so it often does not work well with speech and other assistive technologies.

These are the problems with PDF in the context of trying to use them with A.T.:

1. **Images.** It can have bitmap images in them – pictures of text, not text itself. This means that the content cannot be read at all. This must be solved by OCR – optical character recognition – turning the image into text.
2. **Reading order.** It does not know in which order things should be read – so it does not understand that you read down columns, then to the top of the next column, for example. This means that when you try to read a PDF it reads across columns, or jumps about the page, or reads out of order.
3. **Structure.** It does not know anything about content, like “headings” or “lists” or even “sentences” and “paragraphs” – so you can’t skip to headings and when you try to read by sentence the highlighting goes wrong. Not being able to skip around is okay for sighted mouse users because they can see the headings and chapters and suchlike, scroll around and click where they want to play. It’s difficult for blind screenreader users, who have no way to skip around the document and make sense of all the text. The lack of “sentence” and “paragraph” structure can lead to odd reading for everybody, like sentences being split in the wrong place or highlighting not matching the sentence being read. (In formal terms this is referred to as “semantics” – PDF files do not have any semantic information.)
4. **Zoom.** It does not reflow well. That is, if you are reading text on a mobile ‘phone, or zooming in because your vision is poor, it is hard to read a page if you have to scroll left and right to see the whole of every line. It is much easier if the page reflows – rearranges to cope with the larger text size – and you still only have to scroll up and down.

To solve Problem 1 (images) you have to do OCR – there is no way round that. But even if all the text in your PDF file is text, not images of text, you still have Problems 2, 3 and 4 (reading order, structure, zoom)

To solve these problems Adobe added features called **tags**. These tell Adobe Reader what order should be used for the text and provides chapters, headings and other styles. So a **tagged document** is much more accessible for reading. To be exact, a document *must* have tags to be read. These tags provide the information to allow reading order, structure and zooming to work.

However, most documents are *not* tagged. Adobe Reader will therefore attempt to work out tags when it opens these documents. This has two problems: first, it can take a long time. That dialogue that comes up when you open a PDF file, “Preparing document for reading” – that is Reader working out the tags. A hundred-page PDF file, which is really quite common, can take a few minutes on a fast machine, which is very off-putting if you just want to quickly check out a document. Second, it may not be very accurate. Columns, textboxes, and captions can all confuse Reader when it tries to create a reading order. No attempt at structure is made. Reflowing the PDF file may not work well. Finally, the tags worked out by Reader are not stored anywhere for future use, so if you close and open the PDF again then the same process must be followed.

It is clearly much better to define the tags beforehand, rather than leaving it to Adobe Reader. In fact, this is essentially what is meant by “creating an accessible PDF” – it means “adding correct document tags so reading order, structure and zooming are all correct” (although OCR’ing of images is required in some cases too, of course.)

(Note that Google takes account of reading order when it indexes PDF files it finds online, which means that your PDF files will work better and produce better search results if you take the time to set tags to identify the reading order!)

Other PDF features that affect Accessibility

Bookmarks

These can form a “contents” section, and are turned on in View, Show/Hide, Navigation Panes, Bookmarks. (The option is not there if there are no bookmarks in the document.) These are examples of structural (semantic) tags in the document. They are mainly of use to screenreader users, who can use them to quickly skip around the document using shortcut keys, and mouse users, who can easily explore the document tree and work out where they want to read.

Fast Web View

Fast Web View enables PDF files displayed in a web browser to start displaying before the whole file is loaded. The file is structured so that the first few pages can be displayed as soon as they are received. This requires some re-ordering of the PDF file, but does not necessarily have any impact on accessibility. However, some tools may, at the same time as processing the PDF file so that it can support Fast Web View, discard all the tags to “optimise” the PDF file, making it smaller and faster to download to view – but losing all the accessibility information.

Punctuation

The punctuation of a PDF file affects how usable it is. A fully-tagged document correctly identifies paragraphs and headings and other text structures, so assistive technology can punctuation speech and output correctly – highlighting sentence by sentence for example. In documents where tags are missing, or the document has been transformed or lost its tags, Adobe Reader and other tools are better able to identify sentences if they end with full stops. This mostly applies to headings, which generally do not have full stops at the end of them. This can mean that Adobe Reader gets confused and runs the heading into the next paragraph, which can be confusing to the user. Full stops help your document read correctly no matter what happens to it.

Forms in Adobe Reader

Forms allow users to fill in textboxes and select buttons and checkboxes in the PDF document, and even submit it to an online service. For accessibility the form elements should be tagged to indicate their name and function, and their order should be specified for clarity and to enable tabbing around the form in a sensible order.

If you do not provide tags for forms then Reader will try to infer the tab order from the document structure. If “Use document structure for tab order when no explicit tab order is provided” is unchecked then Reader will go across the page and down, or down the page and across, or some other order specified in the PDF file.

Adobe PDF files in web browsers

By default Adobe Reader makes PDF files appear within web browsers. This is generally less accessible, since you don’t have the menu, and the space given over to the display of the PDF must be smaller, and many access programs will not work with the PDF when it is within the browser. This is especially the case for browsers other than Internet Explorer.

You can make Adobe Reader launch PDF files encountered by your web browser by unchecking Preferences, Internet, Web Browser Options, Display PDF in browser. This means Adobe Reader will pop up when you open a PDF, which may be more confusing, but is simpler.

Adobe Reader Multimedia options

The Multimedia tab of Preferences has an Accessibility Options section. These apply to embedded video content and provide options for subtitles and other features.

Chapter 3: How Adobe Reader handles Accessibility

This chapter explains how Adobe Reader handles PDF accessibility. This lets you understand how accessible a given PDF is, in the context of someone requiring text access through a screenreader or other A.T..

(It is assumed you are using Adobe Reader on Windows. Mac users can employ VoiceOver and the native PDF display tools on their Apple machine.)

Note that Adobe Reader does nothing for embedded images – it has no built-in OCR feature. So these will always remain inaccessible (for anyone requiring text access, that is).

First, Reader sees if there is a “screenreader” running. This means a program that has told Windows that it needs special access to content so it can read it out, so programs like ClaroRead count as well as JAWS and WindowEyes. If there is no screenreader running, it just displays the PDF and nothing more is done.

Second, it checks the PDF file for any text that is not tagged. This doesn't include text that is stored as pictures of text. And it doesn't matter if any of the text is tagged: so long as there is *any* text that is not tagged then the document must be processed to create tags. So a completely-tagged document can become “untagged” if one person edits the PDF file in Acrobat and adds one caption or textbox. If a PDF file is completely tagged then nothing more is done.

Third, it processes the PDF file and adds tags for PDF files where this is needed – documents containing any untagged text. This is when you see one of these dialogues:

Again, you will **not** see either of these dialogues if all the text in the document is already tagged **or** you do not have a “screenreader” running.

You can check whether a document is tagged by opening it in Adobe Reader and going to File, Properties. The Document Properties window shows if the PDF is tagged under “Tagged PDF”:

Fourth, it takes the tags, generated or originally in the document, and works out the reading order (technically it build the document object model.) Even for documents that started off fully tagged this can take a long time. Also, you don't necessarily see any progress dialogue if the document was tagged: Adobe Reader may stop responding while it is working, but there is no other sign. You can't get it to read, of course.

Note that while by default Adobe Reader uses a reading order if one was already defined, the user can change this. If “Reading Order” is set to “Tagged reading order” then the pre-defined reading order is used. But if “Reading Order” is set to “Infer reading order from document” then the reading order worked out by Adobe Reader will be used. However, these two options are only visible where you have opened a tagged PDF file. When Adobe Reader opens an untagged PDF file (where it has to work out the tags itself) it hides the “Tagged reading order” option from the four “Reading Order” options and defaults instead to “Infer reading order from document”. You have to open a tagged document to check that “Tagged reading order” is indeed the default for tagged documents.

When the document has been fully-tagged then it is ready to be read aloud. Reading order is fully-

determined. Structure (semantics), like sentences and paragraphs and headings, has been identified, and bookmarks identified. If the document is to be zoomed it can be reflowed in a sensible pattern (see Chapter 4).

However, there are a number of complications.

- Because processing the document is so slow, Adobe Reader by default only completely processes documents if they are shorter than 50 pages. If they are longer than this then only the current page is processed. Only when you scroll or read past the processed area is processing triggered again. You cannot read, again, until this is completed. If your reading program cannot get to the next page itself when it finishes reading the current page you have to intervene manually, switch to the next page and start reading again.
- There are actually three options for reading order, found in Adobe Reader Preferences, Reading, Reading Order Options, Reading Order. We have been describing the default, “Infer reading order from document (recommended)”. This means the tags, whether generated by Adobe Reader or already in the document are used to work out the order. A second option, “Left-to-right, top-to-bottom reading order” means that Adobe Reader works out the reading order according to the layout of text on the page, as described. This may work better in some pages, but may go completely wrong, for example with columns. Finally, the third option, “Use reading order in raw print stream”, does not look at the layout of the text on the page at all, but the order of text inside the PDF file. This can actually be better for some documents than Adobe Reader’s attempt to work out the reading order, but can be completely wrong for other documents – whole paragraphs can be out of order, for example.
- Even if a document is fully tagged then it is possible to ignore this provided order – Reading Order Options, Override the reading order in tagged documents. We can assume that someone has set this for a reason: it should be unset to provide a good experience with a correct accessible PDF.

At a technical level, the external interfaces used by A.T. programs, like the Microsoft Active Accessibility (MSAA) interface, and the Adobe Reader Document Object Model (DOM) are now generated and available for use. Again, these all depend on the tags of the document.

For non-mouse users without sight, this is generally the point where their A.T. takes over. For example, JAWS screenreader users do not care about the zoom level of the PDF they may not be able to see at all. Other users may employ the built-in Adobe Reader Read Out Loud function, which reads the text using synthesised speech but otherwise does not change the display of the PDF. For users with some sight, many other Adobe Reader and PDF features come into play, described in the next chapter.

Chapter 4: How Adobe Reader handles PDFs for a sighted user

This chapter describes key Adobe Reader features for people with some sight who need zooming, or high-contrast, or low-contrast, or other visual features to make a PDF easier to use.

Traditional accessibility concerns are often about blind screenreader users. But some people simply want the text scaled up a bit (zoomed in) or the colours set to something that is comfortable to read, such as a high-contrast setting for people with poor vision or a low-contrast setting for people with some print impairments like dyslexia.

Setting the colours to something that is suitable is quite simple. Preferences, Accessibility, check Replace Document Colours. This overrides the colours in the original PDF files with a choice of the Windows Colour Scheme (for example, if you have selected the High Contrast yellow-on-black Windows scheme), one of four High-Contrast combinations provided by Reader, or a completely Custom Colour combination for text

and background. The custom option may be useful if you want a lower-contrast colour combination, for example if you have dyslexia.

The problem with zooming in a PDF is that it is designed around fixed pages, so when you zoom in you have to start scrolling left and right to read whole lines across the PDF, just like using a magnifying glass on a real piece of paper:

This does not only affect people with visual impairments. It also affects people trying to read PDF files on small devices, like mobile 'phones. The solution is to make PDF files more like web pages, so when you zoom in the text rewraps or **reflows**, making new shorter lines, so it fits across the width of the available display area and you do not have to scroll left and right:

Images are centred and scale to fit too. One drawback is that you cannot do a text search in Reflow mode. You can switch to this view with Ctrl+4 at any time, or select Reflow from the Zoom submenu of the View menu.

This ability to reflow depends on the content of the PDF being text, not images before OCR, and the tags in the document. Also, some types of PDF file will not reflow: PDF files with images and text behind them cannot reflow, because the text is locked to the position of the image. This is important because these PDF files, often called "searchable PDF" files, are very popular since they allow a scanned page from a book to be perfectly captured in a PDF with text behind them so you can read it aloud. This is great for, for example, dyslexic students who can refer to the original textbook and use the mouse to decide what to read where on a desktop computer. It is not the right format for people with a visual impairment who want to scale the text or people reading PDF files on a small mobile device.

Finally, when a document is Reflowed, Adobe Reader does not update the accessibility information available to external applications, so you will have to close and open the PDF file to work with it in this new format.

Zoom, Page Display and Reflow settings in Adobe Reader

By default PDF files tell Adobe Reader how they should be zoomed and displayed when they are opened. You can change this when they are opened with shortcut keys, menu items, or the toolbar. But you can also change Adobe Reader to display all PDF files in your preferred format permanently.

Note that if the setting Preferences, Documents, Open Settings, "Restore last view settings when reopening documents" is checked then you will not see your changes to the defaults have an effect for PDF files you have already opened. Uncheck this before you try experimenting with the default display options.

Page Display

There are three default settings for Page Display:

- That set in Preferences, Page Display, Default Layout and Zoom, Page Layout.
- That specified by the PDF file itself. This overrides the setting in Preferences, Page Display.
- That set in Preferences, Accessibility, Override Page Display, Always use Page Layout Style. This overrides the other two settings if you check Always use Page Layout Style.

So you need to set your preferred Page Display in Preferences, Accessibility if you want it respected. "Single Page Continuous" is probably best for seeing most of the document at a decent size. The complete set of options in Preferences, Accessibility are:

Single Page

Only one page is

shown. If it is zoomed in, then you may have to scroll around, but when you scroll to the edge of a page it “flips” to the next page.

Same as menu item View > Page Display > Single Page View

Single Page Continuous

One page width, you can scroll up and down and see each page on top of each other.

Same as menu item View > Page Display > Enable Scrolling

Two-up

Two pages shown at once. If they are zoomed in, then you may have to scroll around, but when you scroll to the edge of a couple of pages it ‘flips’ to the next two pages.

Same as menu item View > Page Display > Two Page View

Two-up continuous

Two pages wide, you can scroll up and down and see each two pages on top of each other.

Same as menu item View > Page Display > Two Page Scrolling

Zoom

You can set the default **Zoom** by going to Preferences, Page Display, Default Layout and Zoom, and setting Zoom to your preferred setting. You can also set the default Zoom by going to Preferences, Accessibility, Override Page Display, Always use Zoom Setting. These have slightly different options. The Accessibility Zoom setting will override the Page Display setting. Both will override the PDF file setting.

“Fit visible” (Page Display) or “Fit width” are good if you have a small display and limited sight, and “Fit page” is good if you have a big display (so you don’t have any zooming or scrolling) or you have no sight (so you don’t care what size it is but it simpler to understand what page you are on.)

“Reflow” (Accessibility) is good if want to zoom in a document but don’t want to have to handle horizontal scrolling. Some pages reflow better than others: tagged accessible PDF files reflow best.

The complete set of options are:

Automatic (Page Display only)

The PDF file is allowed to specify how it should be zoomed. This is the default. It usually means ‘display the whole page, zoomed out to fit’.

Fit Page

The document is zoomed so that exactly one page fits horizontally or vertically in Adobe Reader, whichever is smaller. If “Single Page” is selected as the Page Display then only one page will be visible, but all of it will be visible. If “Single Page Continuous” is selected as the Page Display then you can see more than one page, but the pages will still be zoomed as though only one is being fitted to.

Fit Width

The document is zoomed so that the width of the whole page fits across the whole of Adobe Reader. This means you don’t have to scroll left or right, which can cause confusion, but can mean that you have to scroll up and down to read a whole page.

Fit Height

The document is zoomed so that the height of the whole page fits vertically the whole of Adobe Reader. This means you don’t have to scroll up or down, but since most documents are portrait orientation, the

PDF will appear very small.

Fit Visible (Page Display only)

PDF documents normally have blank borders, suitable for printing. "Fit Visible" means the borders are not displayed, so the page content (text, images) can take up more of the Adobe Reader display and therefore be a little larger.

Reflow (Accessibility only)

Every document is reflowed.

10% – 6400%

Zoom by a fixed ratio.

Colour in Adobe Reader

The Accessibility section of Preferences has a Document Colors Options section that allows you to set the desired colour of text and background colour. You can select from the Windows selection, a pre-defined high-contrast palette and a custom palette.

One of the problems, though, is that embedded images and other complex drawing elements like textboxes may not change colour "cleanly" – they may develop odd shapes, or disappear entirely. So by default "Only change the color of black text or line art" is checked, meaning that sections of many documents will not recolor but everything should still be visible. A further checkbox, "Change the color of line art as well as text" lets you decide if graphics should be recolored – again, for anything but plain text, this may create visual confusion.

Chapter 5: Suggested "best" settings for Adobe Reader

Clearly, it is necessary to get the Adobe Reader settings right to maximise the chance of any PDF reading correctly. Here is the best set for best performance (at the expense of slower loading time for large documents). It doesn't take into account the many settings for multimedia and colour. It's primarily to ensure that external A.T. applications have the best access to the PDF file and have the best chance of working well.

Preferences, Reading, Reading Order

Set to "Infer reading order from document (recommended)".

Preferences, Reading, Override the reading order in tagged documents.

Uncheck this checkbox.

Preferences, Reading, Screen Reader Options, Page vs Document

Set to "Read the entire document".

Preferences, Reading, Screen Reader Options, Confirm before tagging documents

Uncheck this checkbox.

Preferences, Accessibility, Override page display, Always use Page Layout Style.

Check this checkbox. Set to "Single Page Continuous".

Preferences, Accessibility, Override page display, Always use Zoom Setting.

Check this checkbox. Set to "Single Page Continuous".

Preferences, Accessibility, Tab Order, Use document structure for tab order when no explicit tab order is specified.

Check this checkbox.

Preferences, Internet, Web Browser Options, Display PDF in browser

Uncheck this if you want PDF files to open in Adobe Reader directly. This lets you see more document content and works better with assistive technology.
