

# Preparing Files for pdfExpress

Version 1.0.5, 5/20/2000

## 1 Overview

This document defines general procedures that should be used when creating files for pdfExpress. It is assumed that the reader is familiar with traditional pre-press issues such as fonts, layout, etc.

This document is targeted at designers familiar with desktop publishing design and general pre-press.

Note that the strategy outlined in this document is only for documents that will act as replacement templates in pdfExpress. If a PDF file will be used in a pdfExpress merge, but will be substituted whole, i.e., without mark up, then the issues outlined in this document do not apply.

## 2 General Strategy

Creating a document for subsequent data substitution by pdfExpress requires planning and preparation beyond what might normally be done in traditional pre-press.

The following design elements should always be kept in mind when creating a document for pdfExpress:

First, you, as the designer, must anticipate the types of values that will be used during substitution. In traditional pre-press, once the design is complete, the textual content of the document will not change. This is not true in the pdfExpress model. A subsequent user will be able to apply text to any marked up elements you have defined. As such, you must consider *at the time you design the document* the attributes of the values that might be attached to the field, e.g., the maximum length of the text, any special characters, and so on.

Second, you must understand *that pdfExpress is not itself a page layout program*; text which is substituted at merge time can only follow the attributes of the original text. At merge time the user will not be able change the font size, leading, character spacing and so on. Therefore you, the designer, must *anticipate* how pdfExpress will perform substitutions so that you do not create a document which is impractical to use as a merge template.

Third, pdfExpress has *some practical limitations* in its operation that you must consider at design time and anticipate for merge time. Though not complex or limiting, you still must remain aware of how the program works and what strategy it employs to do its job so that the final, merged result will be acceptable.

Finally, you must understand the details of how to prepare files for pdfExpress. There are certain requirements that must be followed in order to prepare a poster for sale.

### **3 Working With pdfExpress**

This section will provide an overview of the pdfExpress process. Although the details of how the process should be followed are provided in later sections, you should read this section carefully in order to gain an overall understanding of the pdfExpress process.

#### **3.1 The Design Process**

pdfExpress works by modifying a PDF file you create through a traditional, pre-press design process. Typically you will use Quark, Illustrator, InDesign, PageMaker, PhotoShop, or some other high-end design application to create the document. This design will include static elements as well as replacement elements. Once your design is complete, you must print the file to PostScript.

After the PostScript file is created, it must be converted to PDF with Acrobat Distiller.

Once a file is converted to PDF, the mark up process may begin. Note that until this point, things have proceeded more or less exactly as they normally would for any pre-press project.

#### **3.2 Converting to and Working in PDF**

The distillation process reduces the PostScript information in the file to a series of PDF operators. Some of these PDF operators cause text and graphics to appear in the output, either in print or on the screen. Note that pdfExpress is concerned only with text defined as actual characters in a font and with bitmapped images. Characters created in programs like Adobe

Illustrator that are then converted to outlines will not be accessible to pdfExpress.

Each PDF text display operator is responsible for displaying some or all of the text in the document. In addition, separate operators are responsible for the position, size, style, slant, and other text attributes.

There are two types of bitmapped images in a PDF file that are accessible to pdfExpress: XObjects and in-line graphics. Acrobat Distiller converts most bitmapped images into XObjects unless the number of bits required to represent the image is very small; in that case the in-line graphics format is used. Both formats are basically identical, except that the in-line graphics format requires less space for small images.

During the distillation process, the Acrobat Distiller may make certain decisions about how to convert the rendering of the graphical and textual components of your design into PDF operators. Since pdfExpress works as a plug-in to Acrobat, pdfExpress cannot know or understand your original design intent if the distiller modifies it. Therefore, when you work in pdfExpress you are *working with the Distiller's perception of what the page layout program intended*.

### **3.3 Doing the Mark up**

The next step in the pdfExpress process is to mark up graphical and textual elements in the PDF file with pdfAuthor. Positioning the pdfAuthor markup tool over or near a graphic or a piece of text and double clicking selects an element. pdfAuthor draws a red box around the marked object. The first thing that you will notice is that the red box does not always choose a graphic or sequence of characters that you might intuitively think goes together.

In the case of graphics, the image might be *cropped*. Cropping means that some or all of the image is not visible. The red selection box encloses the entire image, not the cropped portion that you see.

Usually its easy to make sure you have selected the correct image. However, if the image takes up the entire document, for example as is the case with a poster, it may be difficult to select text or graphics because the

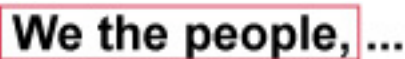
selection tool will tend to select the background image and not the text or graphic that you want.

In this case, hold down the Alt- (or Cmd-) key when selecting; then draw a box around the object you wish to select. Note that you must select the entire region that Acrobat believes is part of the text or graphic.

In the case of text, there are more selection issues to consider. For example, suppose you entered the text

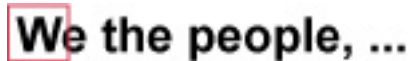
**We the people, ...**

as part of your document. When you go to mark this text up for replacement in pdfAuthor, you double click on the “W”. Instead of seeing



**We the people, ...**

as you might expect, you see



**We the people, ...**

which you don't expect. The reason you see the second selection version is that the Acrobat Distiller has decided that since your page layout application has kerned the “W” and the “e” so that the “e” sits slightly under the right-most stroke of the “W” it should break the text up into two parts, namely “W” and “e the people, ...”

The marked red box above tells pdfAuthor that the text associated with the box may become variable, i.e., when the document is merged the user may want to make a substitution there.

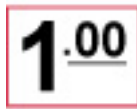
pdfAuthor puts some additional information into the file indicating that the text operator that draws the “W” will be marked for replacement.

In order to complete the mark up on this element (don't worry, before the end of the document we'll explain how to make this particular line work correctly), you need to single click the mouse inside the red box. This will bring up the pdfAuthor dialog box that lets you associate a replacement name with the item.

### 3.4 Markup Considerations

In general, you will have to work with whatever decisions Distiller has made about the text you have entered. These decisions can also affect your design, both in terms of what you see as well as what will happen when your document participates in a merge.

For example, suppose you want a poster showing the price of bananas. So, somewhere on the poster you put “1<sup>.00</sup>,” intending that the purchaser of the poster can set the price of bananas to “2<sup>.39</sup>”. What you will find is that instead of seeing markup that looks like



you will actually see three fields:



Probably not what you want to see, nor is it what you might want the person substituting values into the poster to see because now there will be two instead of one: one for dollars and one for cents.

Later on, we'll discuss ways to ensure that Distiller does not do things that make the mark up process more difficult that it has to be as well as design approaches to resolve issues such as the one above.

#### 3.4.1 Other Markup Concerns

Currently pdfAuthor handles two of the four PDF operators for imaging text. These are TJ and Tj.

When you mark up text in a PDF file, pdfAuthor internally locates the appropriate command that causes that text to be rendered, typically a TJ or Tj.

The ‘Tj’ command takes a single string as an argument, e.g., (We the people, ...) Tj and renders the words ‘We the people, ...’ directly to the output. In this case the corresponding markup box would select the entire string.

The ‘TJ’ command takes a more complex argument than a single string: a list of strings and movement distances. For example, “We the people, ...” *might* be converted by Distiller to “[ (W) 10 (e the people, ...) ] TJ”. In which case the pdfAuthor would select the entire string. However, there is no guarantee that Distiller would chose to use the ‘TJ’ command. It might capriciously chose two ‘Tj’ commands, e.g., (W) Tj ... (e the people, ...) Tj, which leads to the undesired markup described above.

‘TJ’ can also make even more trouble. For example, suppose you wanted to create a table where you want the user to supply a date, time and location for a particular event such as the one below:

<u>Date</u>	<u>Time</u>	<u>Location</u>
Date	Time	Location
Date	Time	Location

You might find that distiller will use ‘TJ’ to make each line of Date/Time/Location a single markup item, i.e., “[ (Date) -200 (Time) -200 (Location)] TJ”, so that regardless of whether you click on Date, Time or Location on a given line, pdfAuthor will highlight all three with a single **red** box.

Date	Time	Location
Date	Time	Location

Though you, the designer, intended to make each item separate

Date	Time	Location
Date	Time	Location

so that each element could be marked up individually, the Distiller has thwarted your efforts by using a single PDF command to render them.

This feature of Distiller is particularly annoying when the Distiller takes text which is actually completely unrelated and realizes that it can be linked with a single ‘TJ’ command, e.g., a date and page number in a heading:

1/1/2001

Page 5

is actually rendered by a single ‘TJ’ command (most likely because they have the same baseline and there is no horizontally intervening text).

We have had enough experience with PDF to determine that most of the time the behavior described above is actually what you want, i.e., ‘TJ’ is frequently used for kerning. Making each separate string in the ‘TJ’ command accessible to the markup process would actually be more annoying. Fortunately, there are some relatively easy tricks to circumvent the propensity of the Distiller to generate this command.

### **3.5 Publishing for pdfExpress**

Once you have marked up all the fields you want the merge process to modify, you will need to save the file. The process of saving modifies the PDF file further to include information that is collected during the markup process.

Note that you should save the file with a slightly different name than that of the original file. The reason for this is that after you have spent a lot of time marking up a file someone might find a typo in the original file. If you change the file and re-distill, the distiller will happily overwrite the original file without so much as a warning; in which case you will have to mark the file up completely a second time.

Its also entirely possible that, on rare occasions, pdfExpress will crash or lock up Acrobat. Again, if you save with a different name, your more likely to avoid trouble.

Generally, we recommend adding a “\_g” to the name when saving markup, as in “foo.pdf” becomes “foo\_g.pdf”. That way, if you need to redistill “foo.ps” to fix a typo, it won’t overwrite the markup in “foo\_g.pdf”. If you want to save time, you can open the new “foo.pdf” and copy and paste the markup from “foo\_g.pdf”, the close “foo\_g.pdf” and save “foo.pdf” as

“foo\_g.pdf” (the copy and paste dialog for pdfAuthor can be invoked by holding down “Shift” and “Ctrl” while clicking on a spot in the PDF without any text or graphics).

## **4 Guidelines for Design**

This section outlines many considerations for designing a pdfExpress document. Please read them carefully before creating your first document.

Some issues listed below will not be apparent until you attempt to mark up the file for the first time. It will be your responsibility as the designer to test your document before publishing it for general merge use.

### **4.1 Fonts**

PC’s and Mac’s encode fonts differently. This means that the encoding in the PDF is dictated by the platform that created it. When a document is set up for publishing through pdfAuthor, you must consider the possible characters a user would want to enter into a replacement field.

#### **4.1.1 Characters in a Font**

Most letters, numbers, and traditional ASCII values, e.g., ‘?’ are in the same place on both platforms, e.g., replacing a name with “Harry” will do the right thing regardless of the platform on which the document was created. However, many special characters including fancy quotes “””, characters with diacritical marks, and so forth may not be. These characters may be located anywhere in the range of ASCII values from 0 through 255 or beyond (in the case of UNICODE or Asian fonts).

You, as the designer, must consider this issue if you are designing on a platform that is different from the PC. In general, if a replacement string will contain only letters and numbers, e.g., a name field, you will probably not have to worry. If, for example, you create a document that uses special characters around a name, e.g., ○◆ JOE ◆○, you have to make sure that the characters you want to use will be available and will render correctly at merge time. The best way to do this is to test the merge.

This issue also applies to internationalization. Many international character sets have extended ASCII values to reference special characters. Many foreign countries have keyboard mappings (on both PC and Mac) which are different than those found in the US. This means that typing the ‘\$’ in England might not produce the correct currency symbol.

#### 4.1.2 Distillation and Fonts

Distillation can affect fonts by causing a font to be “subsetted”. This means that the Distiller changes the character set from normal ASCII to some other encoding. Thus if ‘a’ was at position 95 in a particular font, it now might be 1, and so on. You will not necessarily be able to determine if the Distiller has done this without testing the document after its creation.

At merge time, the individual using pdfExpress to create a merged document can select these non-standard characters by determining their actual ASCII value and entering a special value, e.g., “\371” to generate a ◆.

#### 4.1.3 Complete Fonts

When creating the original PDF you must ensure that the PDF contains the complete font. Using the required Distiller options is one way, but you can also force the proper inclusion of the font by using most of the characters in the font when defining a markup variable *at design time*. See ‘Graphic Arts Guidelines’ below for details.

### **4.2 Distiller Settings**

When distilling a file, you must use the Distiller settings below.

To check or adjust your Distiller settings, launch the distiller and open the “Settings” menu item and choose “Job Options”. The subsections below reflect the tabs in the dialog you will see. Make sure to check each tab.

#### 4.2.1 General

Compatibility: Acrobat 3.0

ASCII Format: Checked

Optimized PDF: NOT Checked

Generate Thumbnails: NOT Checked

All Other Settings: Don't Care

#### 4.2.2 Compression

Compress Text and Line Art: NOT Checked

All Other Settings: Don't Care (though you will for your customers).

#### 4.2.3 Fonts

Embed All Fonts: Checked

Subset All Embedded Fonts Below: NOT Checked

The base 14 fonts should not be embedded, any fonts the user wants to use in a Poster should be moved to the "Always Embed" box. "Never Embed" should be empty.

#### 4.2.4 Color

More than likely you will want to leave the color alone.

#### 4.2.5 Advanced

Again, we don't care about these settings, but you may.

### **4.3 Graphic Art Guidelines**

When creating document for pdfExpress users should follow these general guidelines.

#### 4.3.1 What Fonts to Use

Replacement text should be defined only in Type 1 fonts (either the base 14 or any valid Type 1 font from any source) or TrueType fonts.

The fonts should be appropriate for the platform you are on. *Don't try to convert a Mac font to a PC font with Fontographer or other font editing program*, the person performing a merge will have no end of problems.

In addition, certain programs, such as Illustrator, don't get along well with TrueType fonts. Illustrator 8 happily converts characters in TrueType fonts into paths; leaving the text impossible to mark up.

In general, we have found PageMaker to be the most reliable form of creating well-behaved PostScript with regard to fonts. In particular, documents created with PageMaker and TrueType fonts work well with pdfExpress.

#### 4.3.1.1 Checking Fonts

You can use Acrobat to check a PDF (this applies to users of pdfAuthor). Open the PDF in Acrobat. Select "File>Document Info>Fonts" from the menu. A dialog box should appear listing all fonts used in the document.

The dialog box should present a table that has the following general form, except that the fonts appropriate for your poster should be listed:

<u>Original</u>	<u>Type</u>	<u>Encoding</u>	<u>Used Font</u>	<u>Type</u>
AvantGarde-Book	Type 1	Windows	Embedded	Type 1

If you see "Custom" listed under the **Encoding** heading (instead of "Windows" or "Macintosh"), then you will in all likelihood have a problem with the character encodings. All fonts must also be "Embedded".

When using bold or italic text, make sure that the bold or italic version of the font is installed. Some fonts (particularly TrueType) will apply adjustments to characters that are incompatible with pdfAuthor.

Tools such as *EnFocus PitStop* will not reliably fix font problems.

#### 4.3.1.2 Creating Markup Text

In general, text that will be marked up must exist in the document as characters in a font. Text embedded as part of an image, e.g., a JPEG, cannot

be selected or marked up. In addition, do not use the “Create Outlines” feature of programs such as Adobe Illustrator on text that is to be marked up.

Text included as part of an EPS file (as long as the text is not embedded in an image in the EPS file) may be marked up so long as the complete font is contained in the EPS.

When creating text in a page design program to be marked up and replaced by pdfExpress, use short, simple text strings, e.g., “123” or “abc”, and try to avoid text with characters that are typically kerned, e.g., “W”.

When creating replacements with more than one line of text that will be centered or right justified, *use the same text for each line* and center or right justify the text in the design program.

For example, to create text with a replacement on the left, center and right, you could enter the following:

000

000

000

Note that the “000” text is centered by the page layout program and the “000” is right justified in the page layout program. When these items are marked up in pdfAuthor, it is necessary to check the appropriate left, right or centered radio button (note that this is because by the time Acrobat sees the text, it cannot tell what its design-time justification was).

The color of the text will be retained during the substitution process. Text attributes such as style (bold, outline, etc.), character width adjustments, skews and slants will be retained through substitution time.

Sometimes you will notice that a markup string appears to be converted to numerous individual characters. This is controlled by Acrobat Distiller. Unwanted characters, i.e., those after the first (which would carry the mark up name), can be made “invisible” as necessary to resolve this problem.

#### **4.3.1.3 Embedding Fonts**

You can “force” font embedding by creating a markup item that uses virtually all the characters in a font, e.g., “abcd...ABC...1..0..!@#..//”. Don’t forget to include the space character. There are three ways to do this. First, you can create such text as part of a left-justified markup item so that it will not appear in the final output. Second, you can place it outside the crop marks of the page but still touching the output area. Third, you can mark the text invisible.

Check the “File>Document Info>Fonts” dialog as described above to make sure the encoding is “Windows” or “Macintosh”.

#### **4.3.1.4 Page Design Applications**

Our customers have used virtually all sorts from Quark to InDesign to Word to PageMaker to Illustrator. The key thing to remember is that if there is a “Save to PDF” the PDF must not be “Optimized.”

*Also, it is very important not to use the “PDF Writer” from Adobe. This program causes all sort of problems because it the PDF it generates is not always suitable for pdfExpress. It is always desirable to print to PostScript and then distill.*

#### **4.3.1.5 Page Size**

It is critically important when using a design program to create an appropriately sized page for Distillation. This can be difficult if you are unfamiliar with the process. We have adopted the following guidelines for successfully creating pages of the proper size.

First, make sure you have the proper PPD and/or printer selected. Most design programs limit the size of the output pages to those that are supported by the printer. It may also be necessary to create an appropriate printer in Windows to support this.

Second, when printing, make sure to check the appropriate paper size in the print dialog. Often design programs will default to an inappropriate size. This should be checked every time you print.

Third, when the PDF is first distilled, make sure to check the output size when you first open it for mark up. If it is not correct, you will have to check you settings in the page design program and re-distill the file.

#### **4.4 *Miscellaneous Techniques***

There are many strategies and techniques to control uncooperative fonts, layout programs and pdfAuthor when working on a project. This section is focused on many common how-to's.

##### **4.4.1 Selections**

Often it can be difficult to select the exact text or graphic you want. The easiest way to do this is to hold down the “Alt-” key and drag a box around the text you wish to select. Note that the box does not always exactly fit the text. This is due to mismatches in Acrobat between the size the text actually is and what is being displayed.

If you are having trouble selecting text, try double clicking first to get an idea of the bounding box first, then dragging with “Alt-” to select the text you actually want.

Note unselectable text, e.g., text in an image or outline text, can never be selected by pdfExpress.

Similarly, you can use “Alt-” to select a particular image. Note that if one image exactly overlays another, you may not be able to select the top image.

##### **4.4.2 Fragmented and Joined Text**

As described above, what you enter at design time for markup may be fragmented into pieces or combined with other, unwanted text.

In the case of fragmentation, e.g., the “W” in “We the people, ...” is separated from the “e the people, ...”, simply make the “W” the replacement variable and hide the “e the people, ...”.

When two pieces of separate text with the same baseline is joined capriciously by Distiller, the easiest way to make the individual items

separately selectable is to introduce (at the design level) a design element which breaks up the joining.

There are several ways to do this.

If the alignment is not critical, moving one element up or down a few 1,000<sup>th</sup>'s of an inch will usually do the job.

Another approach is to add a super or subscripted character in front of or after one of the two elements, e.g., this <sup>breaks</sup> things up.

#### 4.4.3 Combining Elements

If you need to create multi-sized fields that appear to be a single variable, try the following approach. Suppose you want to create a price as in the previous example, e.g., \$1<sup>.00</sup>. How would you do it?

First, consider the price itself. Will it always be of the form X.XX? Or will prices beyond \$9.99 be possible, e.g., \$14.95? This will determine if the '\$' sign is fixed or is part of the variable.

Next, think of the price as two separate fields: a right justified whole dollar field and a left justified cents field.



The left-most example would be marked up for whole dollars, the right-most field for cents. The justifications will ensure that the price looks correct when substituted. You could make the '.' a variable (center example) if the poster would be used in Europe where the US notions of "." and "," are exchanged.

Note, however, that the user of the template will have to enter two (or possibly three) values: one for dollars and one for cents.

#### 4.4.4 Problem PDF's

If your having trouble marking up a PDF that someone else has supplied, particularly a PDF not created according to the published Distiller parameters, then try the following:

- 1) Create a printer for the intended output device associated with FILE:. Print the troublesome PDF from Acrobat to PostScript or EPS.
  - A) Redistill the PDF. See if the resulting PDF will work for markup.
  - B) Try to open the PostScript or EPS file in another application such as PhotoShop or Illustrator. If you can, then you can change it there.
  - C) Convert the PostScript or PhotoShop to a CMYK TIFF. Then import it into Quark or Illustrator as a background. Place new text for markup in a layer in front of the image, print to PostScript and distill again.
- 2) Edit the file with Enfocus PitStop (download a trial version for free from [www.enfocus.com](http://www.enfocus.com)).
- 3) Try an alternate Distillation such as GhostScript (free for download) or Transverter Pro.

#### **4.5 pdfAuthor Don'ts**

When designing variable documents, it is possible to create situations where what you want is not what you will get. This section tries to cover some common mistakes pdfAuthor users may encounter.

##### **4.5.1 Don't Use Special Styles**

Do not use underlines (pdfExpress can't generate an underline because it doesn't know how long the replacement text will be). If you need to draw a line, use a line or pen tool. If you line needs to be the exact length of the text, try using another font or font style instead. If that doesn't work, change your design approach.

Don't use super/sub scripts or all caps modes in fonts.

#### 4.5.2 pdfExpress is not a Layout Program

The easiest way to understand what pdfExpress will not do during a replacement is to think of the following model. When the original design is being created place an example text element for later markup. You can think of the replacement process (at design time) as the equivalent to placing the text cursor into the replacement example text, highlighting the entire field, and typing a new value.

If you don't see what you expect, then neither will someone using your document during a merge.