FileCenter Automate User Guide

Updated: 2020-05-28

Table of Contents

Section 1: Get Up and Running

Chapter 1: Introduction

FileCenter Automate is, simply put, a converter. It takes files in one end, switches them to a different format, then pushes them out the other end – a simple concept with a lot of power behind it. What kind of power? Three things in particular: First, you have a lot of control over the process. Second, FileCenter Automate works on massive quantities of files – hundreds or even thousands a day with no problem. And third, FileCenter Automate does it all autonomously. It requires no user interaction.

That last point can make FileCenter Automate difficult to grasp at first. We're used to programs like Microsoft Word where we open files and make our own changes to them. FileCenter Automate doesn't work like this. You can't "open files" in FileCenter Automate. Instead, you tell FileCenter Automate what you want it to do to *every* file, point it at a clump of files, and let it go. FileCenter Automate will go through all of the files on its own and do the conversions for you.

This sounds complicated to set up, but you will be surprised how easy it is. This guide will walk you through the process step by step. You can easily master the fundamentals in an hour. But first, let's go over some basic concepts. These concepts underlie everything you'll do in FileCenter Automate.

1.1 Key Concepts

You may now have a question stirring in the back of your mind: What kinds of "conversions" does FileCenter Automate do? Ninety percent of users employ FileCenter Automate the same way – to convert files into *searchable PDF* files. While FileCenter Automate is capable of plenty of other conversions, making searchable PDFs was its original design and continues to be its primary purpose.

Why PDF?

You may wonder why there's so much buzz about PDF files. It's because PDF has become the *de facto* standard for archiving documents. A PDF file has a few attributes that make it desirable for archival.

First, it's a true representation of the original document.

Second, it doesn't matter what program created the original document. For example, suppose you create a report with an obscure accounting program. Once upon a time, you could only share that report with other people if they had the same program. The PDF format solves that problem. Any user can open any PDF file, regardless of where the original document came from.

Third, PDF files remain valid over time. File formats are constantly evolving. For most files, there's no guarantee that ten years down the road any program will still be able to read them. PDF, on the other hand, was designed with long-term archival in mind. You will still be able to open your PDF files decades from now.

Fourth, besides supporting things like annotations, PDF files can be made *searchable* unlike regular scans.

What is a Searchable File? OCR?

For a file to be *searchable,* it must have *real text*. You know a document has real text if you can either edit it or copy/paste it into a word processor.

Let's take a scan as an example. To your computer, a scan is nothing more than a digital picture. While you can discern letters and words in the scan, to your computer it's no different than a picture of, say, a mountain. So let's suppose you want to do a keyword search for text in the scan. You can't. Why? Because, as far as your computer's concerned, there is no text in the scan.

FileCenter Automate plays a key role in making your files searchable: It performs OCR. "OCR" stands for "Optical Character Recognition". OCR is a very sophisticated process where the software closely analyzes a digital picture and tries to identify the shapes of letters and words. Wherever it can spot images of text it exposes *actual* text.

In the case of FileCenter Automate, as it converts scanned images to PDF, it can optionally perform OCR at the same time. If it finds any text, it will invisibly layer it underneath the scanned image. This makes the scan searchable while at the same time ensuring that it continues to look exactly like the original.

In the future, you will be able to search the body of this scan for keywords. Thus, the scan has become a *searchable PDF*.

What is Routing?

In addition to converting files and making them searchable, FileCenter Automate can perform another critical function: *routing* converted files.

Routing is just a fancy way of saying that FileCenter Automate can intelligently choose where to save files and how to name them. FileCenter Automate has a number of tools to help you set up dynamic file naming rules. It also has a *Separator* feature that adds intelligence to the way FileCenter Automate splits up stacks of documents and saves out the individual files.

Chapter 2: Installation and Setup

This chapter will take you step-by-step through installation. *If you plan to run FileCenter Automate as a service, please read through that section carefully*.

2.1 Install FileCenter Automate

Before you install FileCenter Automate, you need to make a couple of determinations:

Which Machine Should I Put FileCenter Automate On?

As a rule of thumb, install FileCenter Automate on the same machine that holds the files you want to convert. While FileCenter Automate can convert files on other machines, you avoid unnecessary complications by putting it on the machine where the files are. You will also improve conversion performance and minimize network traffic.

Perform the Installation

To begin your installation, locate the setup file and double-click it.

If you're not sure where to find the file and you downloaded it, look on your Desktop or in your Downloads folder (usually at C:\Users\[username]\Downloads). Look for a file called *FileCenterSetup_.exe*, where the underscore "_" will be a version number.

Once the installation starts, click through the screens, paying special attention to these:

All Users vs. Current User vs. Unique Profiles

You may or may not see an option asking whether you'd like to install for *All Users*, the *Current User Only*, or *All Users with Unique Profiles*.

Unless you have a compelling reason to do otherwise (and you know the ramifications), you should select *All Users*. If you plan to run FileCenter Automate as a service, you *must* select *All Users*.

Terminal Server and Citrix Considerations

If you plan to install FileCenter Automate on a Terminal Server (or in a Citrix environment), you should *not* give each user their own instance/profile of FileCenter Automate.

FileCenter Automate is not end-user software. It is a background process that should run independent of user profiles. If multiple users have their own FileCenter Automate profiles, you will easily get into a situation where FileCenter Automate starts consuming excessive system resources and where conversion collisions start occurring.

Product Key

If you already purchased a license of FileCenter Automate, your place of purchase will have given you a *Product Key*. Enter it now.

If you're doing a free trial of FileCenter Automate, click *Install as a 15-Day Trial*.

2.2 Launch FileCenter Automate

After the installation finishes, open FileCenter Automate. You can launch from the *FileCenter* icon on your Desktop or from the *Start* menu > *All Apps/Programs > FileCenter > FileCenter*. After FileCenter opens, select the *Automate* tab.

2.3 Run FileCenter Automate as a Service

If you install FileCenter Automate on a server, you'll most likely want to run it as a *service*. A "service" is a background program that runs whenever the computer is on. No-one needs to be logged in for the service to run. As a result, it's ideal for FileCenter Automate. You want your file conversions to run quietly in the background any time the server is running.

This kind of convenience does come at a price. Services carry restrictions that can keep FileCenter Automate from being able to access your files. In this section, we'll see how to set up the service and how to avoid its common pitfalls and limitations.

Important: Even if you run FileCenter Automate as a service, you can still use the main interface to control your jobs and the queue. Only the conversion engine itself runs as a service.

Pro Only: This feature is only available in FileCenter Automate Pro and Pro Plus.

Step 1: Activate the Service

To activate the service, open FileCenter Automate and go to *Settings*. Select *Automate* on the left. Select *Run as a Service* then click *OK*.

You'll need to restart the machine for the change to take effect.

Step 2: Give the Service Elevated Privileges

Next you need to allow the service to run with elevated privileges. *This step is critical*.

Here's the problem: services don't have permission to access protected folders. As a result, the service won't be able to reach the files it needs to convert unless you elevate its privileges. Do the following:

1. Go to the Windows *Control Panel > Administrative Tools* (if you don't see *Administrative Tools*, select *View By > Small Icons*)
2. Double-click on *Services*
3. Locate *FileCenterAutomateSvc* in the list
4. Right-click on it and select *Properties*
5. Go to the *Log On* tab
6. Under the *Log On As* option, select *This Account*
7. Enter the username and password for an administrator or someone with enough rights to access all of the files
8. Click *Apply* and *OK*

Now, back on the list of services, select the *FileCenterAutomateSvc* service then click *Stop* and, once it has stopped, *Start*. If the service fails to restart, it means the username and/or password you entered is incorrect.

> *Tip:* Failing to give the service elevated privileges is the most common support issue we receive. Users think that FileCenter Automate has stopped working when, in fact, the service can't reach their files because it lacks the security clearance it needs.

Step 3: Use UNC Paths

This step ONLY applies if FileCenter Automate needs to access a network drive. For example, if the files you want to convert are on a different machine than the one where you installed FileCenter Automate, or if FileCenter Automate will be saving the converted files to a different machine, you must follow this step.

The second limitation of a service is that it can't use mapped network drives. What is a *mapped drive?* If you have a network drive that uses a drive letter (like "S:" for example), you have a mapped drive.

> *Important:* Follow this rule. *Local drives* reside on your machine. *Mapped drives* reside on <u>other machines</u> and use a <u>drive letter</u>.

If FileCenter Automate needs to access a drive on another machine, you have to use the network address instead of a mapped drive letter. This is also known as a *UNC path*. A UNC path looks like this:

\\server_name\share_name`\folder\file.pdf`

The part in bold takes the place of the mapped drive letter. For example, suppose you have a mapped drive called "S:" where you access the shared network files. Here is the path in two formats – as a mapped drive and as a UNC path:

`S:\Johnson\Agreement.pdf`

`\\OfficeServer\SharedFiles\Johnson\Agreement.pdf`

Both formats point to exactly the same file. The first one has simply "mapped" the server and share address to a drive letter ("S:"), hence the name *mapped drive*.

How do I find the UNC path? The easiest place to find the UNC path is in Windows Explorer. Open Windows Explorer and look at the mapped drive in the tree on the left (if you don't see the tree, click the *Navigation Pane* button).

Locate the network drive in the tree. Next to the drive letter, you'll see a name and a path in parentheses, like this:

```
SharedFiles (\\OfficeServer) (S:)
```

The first item, *SharedFiles*, is the share name. The second item, *OfficeServer*, is the server name. So instead of using the drive letter "S:\" you'll use "\\OfficeServer\SharedFiles\" followed by the folder(s) and filename. Here are a few more examples:

```
S:\Client Files\Thacker, Robert J.\Old Will.pdf
\\OfficeServer\SharedFiles\Thacker, Robert J.\Old Will.pdf

S:\Staff\Jill\Personal\Address List.xlsx
\\OfficeServer\SharedFiles\Staff\Jill\Personal\Address List.xlsx

S:\New File.docx
\\OfficeServer\SharedFiles\New File.docx
```

Hopefully you get the idea. Again, both formats point to the same file.

Where do I need to use UNC paths? You must use UNC paths instead of mapped drive letters any time the location is on a different machine than the machine where FileCenter Automate is installed. These features in FileCenter Automate all involve paths in one way or another:

- Job source
- Job destination
- Job backup directory
- Separator destination
- Shared separator database location

> *Important:* Again, let us emphasize that you *do not* need to use UNC paths for drives that reside on the same machine where FileCenter Automate is running. Use regular drive letters for local drives.

Step 4: Take Note of Conversion Limitations

If you're running FileCenter Automate as a service, you are limited to converting into/from *image formats* like TIFF, JPEG, PDF, etc.

The problem is that many types of files require third-party software that can't run under a service. The classic example: Microsoft Office files (Word documents, Excel Spreadsheets, etc.). These require a working copy of Office on the machine. Unfortunately, Office will not run under a service. So FileCenter Automate can't convert Office files to PDF if it runs as a service.

Restrict your conversions to image formats.

Chapter 3: The FileCenter Automate Interface

At this point, we need to make a distinction between the FileCenter Automate *interface* and the FileCenter Automate *engine*. You should now understand that the actual FileCenter Automate engine runs invisibly. It works autonomously behind the scenes locating files and getting them converted for you.

You, however, need a way to tell the engine what to do and monitor its progress. This is where the FileCenter Automate *interface* comes into play. **It is critical** for you to understand that the engine will go on working, whether or not you have the interface open. The interface is for your convenience only.

In this chapter, you will become familiar with the layout of the interface and the options available to you. In later chapters, we will cover the specifics.

3.1 Jobs

Before we look at the interface, it's time to introduce a fundamental element of FileCenter Automate: *Jobs*. You can think of jobs as conversion instructions. For example, you might set up a job that tells FileCenter Automate to convert all of the TIFF files in a folder to PDF. FileCenter Automate will work away at the job until it finishes. Or you might set up job that tells FileCenter Automate to monitor the network scanner's folder for new scans. This is an ongoing job.

Most of the interface elements you'll learn about relate, in one way or another, to jobs. We will cover jobs in Chapter 5: Job Basics, but for now simply be aware of them.

3.2 The Main Toolbar

You control your jobs and the FileCenter Automate engine from the main toolbar.

New, Edit, Delete

Use the *New/Edit/Delete* buttons to create and manage your jobs. We'll cover job creation in detail in Chapter 5: Job Basics.

Get Files, Pause, Refresh

Use the *Get Files, Pause,* and *Refresh* buttons to control the *queue* – the active conversion list (see p. 17, The Queue).

Get Files will execute the current job and fetch the next batch of files to convert. Note that this isn't necessary if you assign the job to run on a regular schedule.

The *Pause* button will temporarily suspend all conversions. Conversions will remain paused until 1) you un-pause them by clicking *Pause* again, or 2) you close the FileCenter Automate interface.

Refresh simply refreshes the display. This is primarily for updating the log information.

Other Buttons

The *Tools* and *Separators* buttons control various functions and will be covered at a later point. Use the *Settings* button to change FileCenter Automate preferences. The *Guide* and *Help* buttons give you access to the product documentation.

3.3 The Job Tabs

Below the toolbar you'll see a row of tabs. These tabs are for your jobs. At the moment you probably only see one tab: *Default Job*.

What do the tabs do? They control what information you see on the screen. All of the information below the tabs is specific to the selected job. Once you set up more jobs, you'll find that clicking through the tabs changes what you see in the interface – queue, logs, etc.

The *Jobs* tabs also affect the *Edit, Delete,* and *Get Files* buttons on the main toolbar. In other words, if you want to edit a job, first select its tab, then click *Edit*. The same thing applies to the *Delete* and *Get Files* buttons.

3.4 The Queue

Below the *Jobs* tabs you'll find the *Queue*. Think of the queue as the list of files waiting to be converted.

Here's how FileCenter Automate uses the queue:

Each time a job runs, FileCenter Automate looks for file candidates that match the criteria in the job. Every file it locates gets placed in the queue.

In the meantime, the conversion engine works away at the queue, taking the files sequentially, one at a time (unless you use FileCenter Automate Pro Plus, which can do multiple, simultaneous conversions), and converts them as dictated by the job.

If you look at the queue while conversions are running, you'll see files work their way through the list.

Important: The queue list is specific to the selected job. In other words, it will only show the files for the tab you have open. To see another job's queue, go to its tab.

Tip: You can bump a file from the queue. Just select it, then click the *Remove* button above the queue. To clear the whole queue, click *Clear*.

3.5 The Status and History Views

The right side of FileCenter Automate is your information center. Here you can get a quick glimpse at the job's settings, what files are in the job folders, when the job has run, and the history logs for the job.

History/Failures

The *History* tab lists the files that FileCenter Automate tried to convert along with some detailed information, including the number of pages in the file, how many files were produced, and how many *separators* were found in the file (see Section 2: The Automation Guide). Most importantly, it gives you the status of the file: whether the conversion succeeded or not.

You can filter the history to show you only the files that match a specific status. Select the status codes you want to see below the list of files.

To get a quick look at just the failed conversions, look at the *Failures* tab instead. This tab only shows the history items that have a status of *Failed*.

Tip: To see more info about a specific file, select it in the list. Detailed information will appear below the list of files.

Skip List

The *Skip List* plays a critical role in your conversions. As files get converted, FileCenter Automate adds them to the skip list. This ensures that FileCenter Automate won't try to convert them again the next time that the job runs.

Note: FileCenter Automate won't use the skip list if the original file gets moved or removed as part of the job. In this case, the skip list isn't necessary since the original file no longer exists.

Tip: To re-do a file, select the file and click *Remove* above the list. To redo the entire job, click the *Clear* button.

Inputs/Outputs/Backups

The *Inputs* tab shows the source folder(s) for the job. This is a handy way to see what files are available for converting.

The *Outputs* tab similarly lets you browse the converted results. You can open files from the *Outputs* tab by double-clicking them.

Important: If you delete a file from any of these tabs, it will delete the <u>actual file</u>.

Tip: To quickly convert a specific file, drag it from the *Inputs* tab to the *Queue* list. This adds it to the queue. Likewise, to re-convert a file, drag it from the *Backups* tab to the *Queue* list.

Scheduler

The *Scheduler* tab shows you every instance when the job was fired, how long it ran, and how many files were converted in that batch. It's a diagnostic aid to help you monitor what's happening behind the scenes.

3.6 Active Conversions

In addition to the information panels and lists we just discussed, you can see which files are actively in the process of getting converted. These show up at the bottom of FileCenter Automate. You will see a progress bar for the current file along with certain details like the job name, page count, and current phase.

If you use FileCenter Automate Pro Plus, you will see a separate progress bar for each file that's being converted. To designate the number of concurrent conversions, go to *Settings > Automate* and set the *Number of Concurrent Documents*.

Pro Plus Only: Multiple concurrent conversions are only available in FileCenter Automate Pro Plus.

Daily Page Count

FileCenter Automate Standard comes with a daily page limit. To help you keep track of how far you've gone into your daily limit, FileCenter Automate will display the current page count in the lower left-hand corner.

Once the daily page limit is reached, FileCenter Automate will finish whatever document it's on then go to sleep until midnight. It will then pick up where it left off.

Chapter 4: Prepare Your System

Before you get to work setting up your jobs, it might be helpful to make a quick assessment of your system: where the files are that you want to convert, what your goal is, and how the conversion process might impact your system resources.

4.1 Where Are Your Files?

On page 10, Which Machine Should I Put FileCenter Automate On?, we suggested that the best place to install FileCenter Automate is on the machine where the files are. Now it's time to determine the exact location of those files. Locate the drives and folders that hold the files and note the following:

- Do you have a single, common folder that contains all of the data, or are there a few scattered folders? Where are they?
- Do your data folders include subfolders?
- Do you have a network scanner or fax service that outputs to specific folders? Where are they?
- Are any of the folders on a mapped drive? Do they require access rights?

If your folders are on a mapped drive and/or require access rights and if you plan to run FileCenter Automate as a service, see p. 11, Run FileCenter Automate as a Service for important guidelines.

Considerations with Multiple Input Sources

If you have multiple locations to convert, you should mentally rank them from top priority to the lowest priority. This will help you when it comes time to set up your job scheduling later.

For example, suppose that you have a massive repository of files to convert. You also have a network scanner that needs top priority. You will set up your ongoing conversion job with a low priority and your network scanner job with a high priority so that, when new network scans show up, they get handled immediately. We will discuss this in more detail on p. 36, Schedule. For now, it's enough to make a mental note of how your conversion goals rank against each other.

4.2 Where Do You Want the Converted Files to Go?

Now it's time to make an important decision. Where do you want FileCenter Automate to put the converted files? Consider each of these options and their ramifications:

Convert Your Files In-Place

With an in-place conversion, FileCenter Automate will replace the original file with the converted copy, keeping it in exactly the same place. This is a good option if you have a collection of files that you work out of daily and you want FileCenter Automate to keep them updated and searchable.

For example, suppose that your office does a heavy volume of scanning. You prefer to have FileCenter Automate handle all of the OCR to make the scanning as quick as possible. This situation would be a perfect candidate for in-place conversion. FileCenter Automate will comb through your working files looking for new scans and make them searchable.

The downside of in-place conversions is that it can be difficult to see which files have been converted and which ones haven't. Additionally, if you need to re-do a conversion, you have to pull the original out of backup.

Mirror to a New Folder

The safest conversion option, and the one we normally recommend, is that you mirror the files to a new folder. With this option, FileCenter Automate puts a converted copy of the file into a new folder leaving the original file untouched, exactly where it is. If the original location includes subfolders, FileCenter Automate can optionally mirror the original folder structure.

This option works well because it is easy to see what has been converted. Additionally, it gives you the safety net of preserving the originals in-place. This option also makes it easy to reconvert files if necessary, since the originals are still available.

Route to a New Folder

Routing to a new folder is similar to mirroring, except that the original file gets moved to the new folder as it's converted. If the original location includes subfolders, FileCenter Automate can preserve that structure in the new location.

A classic use for routing is a network scanner. Network scanners typically deposit new scans in a specific folder (or folders). You can set up FileCenter Automate to watch for each user's new scans, OCR them on the spot, and move them to an inbox where the user can then access them.

This is the best option for monitoring the conversion process, since it accurately shows not only what has been converted, but also what is left to be converted.

On the downside, if you need to reconvert a file you have to pull it out of backup.

Intelligent Routing

Finally, FileCenter Automate can intelligently route converted files on a document-by-document basis. This includes not just choosing the output location intelligently, but also choosing how the file is named. We will discuss this at length in Section 2: The Automation Guide.

Note: Intelligent routing can happen in conjunction with the other options. If a document gets routed, it can override the normal output settings.

Chapter 5: Job Basics

Now it's time to start setting up your jobs. As we mentioned earlier, you can think of a job as a set of conversion instructions. Normally you will have one job for each folder (or folder tree) that you want to convert. In fact, *two jobs should never draw from the same set of source files*.

If you took our advice in Chapter 4: Prepare Your System and made a careful assessment of the folders you need to convert, you should have a fairly clear view of what jobs you need to create. For the rest of this chapter, we'll go step-by-step through setting up your first job.

5.1 Create a New Job

To add a job, click *New* on the main toolbar, then provide the following information:

Job Name. The job name is for your reference only. The name will appear on the job's tab.

Type. There are two types of job: *Manual* and *Scheduled*. A manual job will only run when you explicitly launch it (see p. 39, Running a Job). A scheduled job runs on a timer, firing whenever the schedule dictates. For example, you can schedule a job to run every night when the office is closed, or to run only on weekends. Most jobs are scheduled.

Priority. When multiple jobs can potentially run at the same time, FileCenter Automate needs to know which one has precedence. Use the *Priority* option to assign a priority to this job. High priority jobs will fire before low priority jobs.

> *Note:* The priority setting has nothing to do with FileCenter Automate's priority over other software. It only deals with how one conversion job ranks in priority over another one.

Edit/Delete a Job

We'll pause here to describe how to edit or delete an existing job. It's simple. Select the job's tab, then click *Edit* or *Delete* on the main toolbar.

5.2 Inputs Tab

Next go to the job's *Inputs* tab if you're not there already. This tab is where you tell FileCenter Automate what to look for.

The Key Settings

Location of Input Files. First and foremost, set the *Location of Input Files*. This is the folder that contains the files you want to convert.

> *Important:* If you're running FileCenter Automate as a service, there are important limitations around the input folder. See p. 11, Run FileCenter Automate as a Service.

Input Types. Odds are you don't want to convert/OCR *every* file in the input folder. For example, most Office files (Word, Excel, etc.), emails, and web pages are already searchable. Use the *File Types* option to designate which file types FileCenter Automate should look for. If your only concern is making scans searchable, select: *PDF, TIF, TIFF*.

File Dates. Sometimes you may want to limit your job to files created/modified on a certain date or within a certain date range. For example, you may want to convert only the files which showed up in the system yesterday. You have a number of flexible date options allowing you to easily focus on any window of dates you'd like.

> *Note:* If you don't see this option, go to *Settings > Automate > Use Advanced Input File Filters*.

Also Include Subfolders. If your source folder has subfolders and you want to include those in the job, select this option. If this option isn't selected, FileCenter Automate will only look at the main folder.

Delete Input Files after Conversion. This is an important setting to get right. There are two situations where you will want to delete the original file. First, if you're *moving* the converted files into a different folder (see p. 22, Route to a New Folder), and second, if you're converting *in-place* (see p. 22, Convert Your Files In-Place).

In the case of in-place conversions, if you don't delete the original file you can end up with a situation where both the original file and the converted file are sitting right next to each other, which can confuse users. For example, suppose all of your source files are TIFF files that you're converting to PDF. If you don't delete the TIFF files after they're converted, every file in the input folder will be doubled: one copy that's a TIFF and one that's a PDF.

> *Note:* With in-place conversions, FileCenter Automate won't delete the original file if the converted file is replacing it. For example, if you convert a PDF to a searchable PDF, the name won't change. As a result, FileCenter Automate won't delete the original (because the original has simply been replaced).

> *Important:* FileCenter Automate will only delete a source file after the backup file is successfully created AND the converted file is successfully saved.

Less Common Settings

You won't normally need to change the following settings except in very specific situations:

Skip Input File if Output Destination Already Exists. This setting assumes that you're *mirroring* the source folder (see p. 22, Mirror to a New Folder). In the event you have to re-start the conversion job from the beginning, this setting can help ensure that you don't re-convert files that have already been converted. In a nutshell, before FileCenter Automate attempts to convert a file, it looks to see if a converted copy already exists in the output folder (based on the filename and path it would have used for the file). If a copy already exists, FileCenter Automate will skip the file.

Copy Log of Failed Files to Input Folder. To make it easy to check the failed files log (see p. 18, History/Failures), FileCenter Automate can keep a copy of the log in the input folder.

Backup Folder. By default, FileCenter Automate puts backup files at "C:\FileConvert Backups". Use the *Backup Folder* option to change the location.

PDFs with Existing Text

By default, FileCenter Automate will skip any PDF that already has text in it. You can override this behavior:

Skip and Do Not Route. This is the default behavior. FileCenter Automate will ignore the file completely.

Skip but Route to Destinations. When you're trying to *move* the converted files into a different folder, this policy can leave some files behind (see p. 22, Route to a New Folder). In this case, you'll want to select *Skip but Route to Destinations*. This option will move PDFs that would otherwise be skipped into the same folder where the rest of the converted files are going, as if they had been converted.

Process and Add OCR Text to Existing PDF. FileCenter Automate will OCR the file and add the new OCR text to whatever is already there. Use this option for PDF files that contain a mixture of text pages and scanned pages. The text pages will be left as they are and the scanned pages will be OCRed.

Process and Add OCR Text to New PDF. Use this setting with caution. This option creates a *new* PDF file out of the old file by converting all of the pages into images. It then runs OCR on this new file. This has some serious repercussions:

Text-based documents will be converted into image-based documents. Bookmarks will be lost. Customizations will be lost. Stamps and annotations will become permanent parts of the page.

You should only use this option if you want to <u>re-do</u> the OCR text on <u>scanned documents</u>.

Process but Skip Pages with Existing Text. FileCenter Automate will go through the document page-by-page and check whether each one has text. If it encounters scanned pages with no text, it will OCR just those pages.

Because this method needs to analyze each page individually, it is the slowest method. It is, however, the best way to deal with the situation where your users append new scanned pages to existing documents.

5.3 Outputs Tab

On the *Outputs* tab you control where the converted files will go.

Replace Input File

As we discussed earlier on p. 22, Convert Your Files In-Place, you can optionally convert your files in-place. In other words, the converted files will replace the original files. To do this, select *Replace Input File*.

> *Note:* Files will always be converted to PDF. If you want to convert to a different format, use the *Route to Fixed Destination* option instead along with *Use Same Subfolder Structure*. This will put the converted files in a new folder tree, but it will give you the option of choosing a different output format.

Route to Fixed Destinations

If you're not converting your files in-place, you need to tell FileCenter Automate where to save them. This is called a *Destination* – an output location along with all of the other information FileCenter Automate needs to save the file, like filename, file format, etc.

FileCenter Automate can save a converted file to more than one destination. In fact, there's no theoretical limit to how many places you can have FileCenter Automate save a file (though there's certainly a practical limit).

To add an output location, or *destination*, click *Add* under the list of destinations then provide the following information:

Location. Browse out to the folder where you want the converted files saved.

File Type. Select the format to save the converted files in. Available formats:

PDF. In most cases, you'll want to convert into PDF format. This is the *de facto* standard for archiving files. What if your original files are already PDFs? That's not a problem. Some PDF files are not searchable. FileCenter Automate will turn those PDFs into searchable PDFs.

TIFF. The TIFF file format was once the standard for all document scanning. Some users still prefer the simple TIFF format over PDFs for archiving files. The downside: TIFF files cannot be made searchable. A TIFF file is simply a digital picture of the document.

TEXT. If all you care about is the OCR text of your documents, you can choose to convert them into simple text format. This text file will keep nothing but the raw, unformatted text. All formatting and images will be lost.

Use Name Builders. FileCenter Automate allows you to have a great deal of control over how the converted files are named. We'll discuss this in depth in Chapter 7: Automate File Naming.

Route according to Document Rules

Instead of fixed destinations, you can use *intelligent* destinations. What do we mean by that? You can have FileCenter Automate analyze the content of each document according to *rules*. Does it look like a contract? Does it have an invoice number? Does it contain a specific doctor's name? The answers to these questions will then dictate where the document is saved

We call these *Document Rules*. We'll discuss document rules in depth in Section 2: The Automation Guide.

Preserve the Folder Structure

If your input folder includes subfolders, you can mirror that folder structure in the output location. In other words, if a file came from a subfolder, FileCenter Automate will put it in an equivalent subfolder in the output location. Once the conversion job is done, the output folder will have exactly the same folder structure as the input folder.

How Does FileCenter Automate Name Files?

By default, FileCenter Automate will keep the original filename, regardless of whether you're converting in-place or routing the converted file to a different folder.

Whenever there's a filename collision (there's already a file with the same name in the same location), FileCenter Automate will put an increment on the new file. For example, if FileCenter Automate tries to save "Insurance Policy.pdf" but finds there's already a file with that name, it will save the file as "Insurance Policy 01.pdf".

Tip: You can change the way FileCenter Automate names the files, even including dynamic fields in the filename. See Chapter 7: Automate File Naming.

5.4 OCR Tab

As we discussed on p. 7, Key Concepts, one of FileCenter Automate's main roles is to make documents searchable through OCR. The following OCR options are available.

The Key Settings

Make Searchable PDF. This is the most important of the OCR settings. *Make Searchable PDF* does two things: First, it forces OCR to run. Second, it takes the resulting OCR text and embeds it in the PDF to make it searchable.

Note that the OCR text does not change the appearance of the document. Rather, it gets inserted invisibly behind the image.

Auto-Rotate Pages. Sometimes scans have both portrait and landscape pages. FileCenter Automate can automatically detect what direction a page's text is running then rotate the page upright. If you want FileCenter Automate to correct page rotation, select *Auto-Rotate Pages*.

OCR Engine. You might have more than once OCR engine available to you, depending on your system. Each engine has its own strengths and weaknesses:

Standard. The biggest advantage of this engine is speed – the engine is very fast. Its accuracy is typically 80-90% on clean documents – high enough to make your documents searchable. It handles poor images gracefully, but its accuracy degrades quickly as the image quality goes down. This engine does not support columns. It only recognizes English characters.

Tesseract. The open-source Tesseract engine was originally developed by Hewlett Packard in the 1980s. It is now maintained by Google. Tesseract is capable of very good accuracy – well over 90% -- and also supports columns. Its weakness is speed. It's the slowest of the engines in FileCenter Automate.

Advanced. The *Advanced* engine is faster than the *Tesseract* engine and has better accuracy. It handles difficult documents very, very well. It does <u>not</u> recognize columns. This engine supports and automatically detects Danish, Dutch, English, French, German, Italian, Norwegian, Portuguese, Spanish, and Swedish.

ReadIris. ReadIris is arguably the best OCR engine available today. Its accuracy, even on problematic documents, is unrivaled by any other engine in FileCenter Automate. It is at least as fast as the *Advanced* engine. It handles columns and Western European languages with no difficulty.

Important: You will only see these engines as options if they are available on your system.

Less Common Settings

Line Breaks. When you're converting into a plain text file, FileCenter Automate needs to know where to put line breaks. You have three options:

By Paragraph. FileCenter Automate will try to figure out where paragraphs end based on punctuation. For example, if a period falls at the end of a line, it's probably the last sentence in the paragraph. FileCenter Automate will insert two returns wherever it thinks a paragraph ends. If you choose this option, you should proofread the text to make sure all of the line breaks were handled correctly.

By Line. FileCenter Automate will preserve the original lines from the document. This means that one horizontal line of text in the document will be one line of text in the resulting file. In other words, wherever a line wraps in the document, FileCenter Automate will insert a line break. FileCenter Automate will not try to figure out paragraph endings.

None. FileCenter Automate will not insert any line breaks. The text will come out as one continuous line of text.

Page Text. When FileCenter Automate creates a searchable PDF, it puts the text character-aligned behind the words in the scan. Some users prefer to have the text lumped in a single hidden paragraph at the top of the page, which reduces the file size slightly and reduces the

number of text objects in the document. To do this, change the *Embedded Text* option to *Top of Page*.

Most users should leave this set to *Word-Aligned*.

Page Timeout. Sometimes an OCR engine can get stuck on a really tough page – poor resolution, poor contrast, a lot of images, etc. The *OCR Timeout* option sets the "give up" point. If the OCR engine hasn't been successful after __ seconds, FileCenter Automate will give up and move on to the next page. For reference, most pages, even tough ones, can be recognized in under 30 seconds.

Page Threads. To speed up OCR, you can have FileCenter Automate OCR more than one page at a time. Most modern hardware can handle this without bogging down. Use this setting to specify the number of pages you want to OCR simultaneously (current limit: 2).

Limit OCR To __ Pages. If your main reason for running OCR is to make the PDF searchable, you might find that most of the important information is in the first few pages of the document. If that's true, you can save a lot of time by limiting OCR to the first few pages, especially if you scan documents that are dozens or hundreds of pages long.

Use the *Limit OCR* option to choose how many pages of the document you want to OCR. OCR will quit once those pages are done.

Turn Off Column Recognition. Some OCR engines attempt to recognize columns. In other words, text will wrap at the column end instead of the end of the line. Use this option to disable column recognition. Note that you will only see this option if the OCR engine supports column recognition.

5.5 Separate Tab

Some of the files you convert may in fact be collections of documents – a stack of documents scanned into a single file. FileCenter Automate has a few different mechanisms that can split these documents out into their own, individual files. This is known as *Document Separation*. The document separation options are available on the *Separate* tab.

Separate Document Every __ Pages. If you want to automatically break the document up every few pages, select *Separate Document Every __ Pages* and provide a page interval. For example, suppose you've scanned a stack of three-page forms. Enter "3" as the page interval. You'll get a series of numbered files, each three pages long. If your *Filename* is "Scanned Form.pdf" you'll end up with files named *Scanned Form [1].pdf*, *Scanned Form [2].pdf*, etc.

For more advanced separation, you can use *Separators*. Think of separators as rules for splitting up stacks of documents. You can use special cover pages to show where a new document begins, or you can base separation on the content of the document itself paired with logic. We'll cover separators in detail in Section 2: The Automation Guide.

Separate on Barcodes. You can have FileCenter Automate treat your own barcodes as separators. It can separate a document every time it encounters a barcode, route the separated file to a path contained in the barcode, or even treat the barcode like a full-fledged separator. We discuss using barcodes as separators in Section 2: The Automation Guide.

Tip: If your barcodes aren't being recognized, go to the *Advanced* tab and select *Force Barcode Processing*.

Separate on Unique Barcodes Only. If a specific barcode occurs more than once in a document, separation will only happen on the first instance.

Route to Path Contained in Barcode. If the barcode contains a valid path (drive + folder + filename + file extension) you can have FileCenter Automate route the document to that path.

Route via Separators Database. You can base a FileCenter Automate separator on a barcode. Select this option to have FileCenter Automate try to look up barcodes in the separator's database.

Separate on Page Separators. Think of *Page Separators* as special cover pages that show where a new document begins. If you're scanning a stack of documents and you put page separators between them, FileCenter Automate can automatically split up the stack into individual documents. We'll go over this in Section 2: The Automation Guide.

If you want FileCenter Automate to look for page separators in the document, select *Separate on Page Separators*, otherwise it will ignore them.

Keep Page Separators in Documents. Normally page separators are removed from the resulting documents after they're split up and saved. This option will retain the separators.

Separate on Document Rules. You can separate documents *intelligently* based on whether the document contents meet certain criteria. These are the same document rules that we mentioned on p. 28, Route according to Document Rules. In that context, the job routes a *whole file* based on rules. In this context, FileCenter Automate *splits apart* documents into their own files based on similar rules. We'll cover document rules in Section 2: The Automation Guide.

Copy Routed Documents to Job Destinations. Separators can play two roles: separating documents and, as we alluded on p. 23, Intelligent Routing, routing documents to specific locations. When a separator routes a document, FileCenter Automate will *not* save it to the job's output folder. If you'd like it saved to *both* places, select *Copy Routed Documents to Job Destinations*.

5.6 Advanced Tab

The options on the *Advanced* tab can normally be left at their default settings unless you have a specific need. The following options are available.

PDF/A-1b. Select this option to create a PDF/A-compatible file. PDF/A is an archival-quality PDF intended for long-term storage. Some government agencies require the PDF/A format.

> *Note:* PDF/A does not support the full range of PDF features. You should not modify a PDF/A file or you might introduce elements that aren't compatible with PDF/A. PDF/A will also produce larger file sizes.

Remove Blank Pages. FileCenter Automate can try to identify blank pages in scans and drop them. To identify a blank page, FileCenter Automate looks at how much of the page is white. If it's almost completely white, FileCenter Automate will consider the page blank and drop it from the scan. To adjust the sensitivity, use the *Blank Sensitivity* setting, where a low number tolerates no stray marks on the page and a high number will tolerate some speckles and black marks.

Use Image Size as PDF Page Size. When you're converting from an image file to PDF, FileCenter Automate needs to know what size to make the PDF pages. You have two options: select a specific page size, or base the page size on the size of the original image.

If you select *Use Image Size as PDF Page Size*, FileCenter Automate will make the PDF the same size as the original image based on its dimensions and resolution.

The *Force Image DPI* option then lets you optionally specify the image's resolution, which will in turn affect its scale. Resolution is given in *dots per inch*. For example, assume an image is 400 pixels wide by 300 pixels tall. If you put the DPI as *100*, your resulting PDF will be 4" x 3".

> *Important:* Most images already contain a setting with their ideal DPI. To use the image's default DPI, set *Force Image DPI* to *0*. Only use the *Force Image DPI* setting if you want to make the PDF page size larger or smaller.

Specify a Default Page Size. Instead of basing the PDF size on the size of the original image, you can force the PDF to specific page size. Note that FileCenter Automate will *scale* the image to fit within the page size.

Force PDFs as Black and White. During the conversion, you can force the PDF into black and white. Some offices prefer black and white PDFs because they're smaller than full-color PDFs.

Force Images as Black and White. Alternatively, if you're converting to TIFF format rather than PDF, you can force the TIFF image to black and white.

Force Barcode Processing. If your barcodes aren't being found, you can force FileCenter Automate to look for them.

Remove Barcode Pages when Separating on Barcode. Normally pages containing barcodes will be left in the resulting document. This option causes them to be removed. The classic case is when you use a barcoded cover sheet for the document.

Date/Time Stamp. When a file gets converted, Windows will normally want to give the converted file a new timestamp – the current date/time. Often, however, it's desirable to preserve the *original* file's timestamp for historical purposes. You can do this by changing the *Date/Time Stamp* option to *Original Date/Time*.

Flatten Comments. If your source documents are PDFs, they may contain comments or annotations. You can flatten all annotations in the converted PDF.

> *Important:* If you don't know what annotation flattening is and don't understand the ramifications, you should NOT select this option.

PDF Compression. The compression settings control how your scanned images will be compressed if you scan to PDF. Most users will get the best compression if they use *JBIG* when scanning in black and white, and *JPEG* or *JPEG + ZIP* when scanning in grayscale or color. JPEG compression also lets you set your JPEG quality, which has a dramatic effect on file size. The default quality of *75* strikes a good balance between file size and image quality. For a sharper image, at the cost of file size, increase the quality. For a smaller file size, at the cost of graininess in the image, decrease the quality.

5.7 Metadata Tab

Think of *metadata* as tags and keywords that can improve the searchability of a document. FileCenter Automate can add metadata to your converted documents.

This *Metadata* tab allows you to use four different metadata fields: *Title, Author, Subject,* and *Keywords*. There is no hard and fast rule about which field(s) to use. You can use any one of them in isolation or all of them together, however it fits your workflow best.

You can hard-code metadata text in these fields. FileCenter Automate also supports some dynamic tags to provide common, repetitive values:

%FileName%. The tag "%FileName%" (including the percent symbols) will be converted to the name of the file once the document gets saved (excluding the drive and folders).

%FilePath%. The tag "%FilePath%" resolves to the complete file path, including the filename.

%DateCreated%. The "%DateCreated%" tag gives the date the document was created (formatted using the default date format on your system). This will normally be the current date unless you're appending the scan to an existing document, in which case it will use the existing document's creation date.

%DateToday%. The "%DateToday%" tag gives you the current date formatted using the default date format on your system.

%UserName%. The "%UserName%" tag will insert the Windows login name for the current user. Note that this isn't the display name, but the login name they type when logging onto the computer.

%ComputerName%. The "%ComputerName%" tag will insert the network name for this computer.

Important: Any metadata you set in the job will be added to *every* file that the job converts.

5.8 Map Tab

FileCenter Automate can optionally export specific information from the document's content to an external data file. This is accomplished through *Content Maps*. See Section 2: The Automation Guide for more information.

5.9 Variables Tab

The *Variable* naming option field allows you to provide the field value at runtime. See p. 47, Dynamic Fields.

The place to provide field values is on the *Variables* tab. Values you provide here will feed into the naming options when the file gets saved. Enter the values in this format:

```
variable_field_name=value
```

For example, suppose you set up a variable field called *DocumentType*. To set it to "Letter" you would enter the following:

```
DocumentType=Letter
```

You can set more than one variable. Simply enter one per line:

```
Department=Estate Planning
DocumentType=Letter
User=Bob Johnson
```

Note: Variable names cannot contain spaces (but the values can). Variable names are *not* case sensitive.

5.10 Schedule

You have a lot of flexibility in scheduling conversion jobs. You can specify how often FileCenter Automate searches for new files and you can limit conversions to specific blocks of time. To see the scheduling options, set the job *Type* to *Scheduled*.

Important Guidelines – Scan Frequency vs. Queue Size

In order to get the best possible performance out of FileCenter Automate, you need to balance two critical settings: the job's *Scan Interval* and the *Queue Limit*.

Queue Limit. You set the queue limit in FileCenter Automate's *Settings > Automate* tab > *Queue Limit for Scheduled Jobs*. The queue limit is the maximum number of files that can sit in the queue.

Scan Interval. The job's scan interval is simply a setting to tell FileCenter Automate how often to scan for files and replenish the queue.

It's important to understand the interplay between the *Queue Limit* and the *Scan Interval*. Whenever the job runs, this sequence happens:

1. FileCenter Automate begins scanning the job's input folder looking for files to convert
2. Candidates are added to the queue

3. The queue continues filling until the queue limit is reached
4. Scanning stops
5. When the scan interval lapses, the process begins again

Scanning for files can drastically affect your system's performance, so it's very important to minimize this impact. Follow these guidelines:

- Determine how often the job needs to scan for new files
- Set the queue limit to the number of files FileCenter Automate can convert in this time interval

Let's go through two examples that will help you better understand how to set your scan interval and queue limit.

Small Source Folder

If the input folder never has more than a few dozen files in it and the folder gets emptied every time FileCenter Automate runs, set the scan interval *low* – somewhere in the range of every minute to five minutes.

The classic example here is a network scanner. Network scanners save new scans to a folder. You would want FileCenter Automate to check this folder frequently, and you'd usually have FileCenter Automate move the converted scans into their final location, which means the scanner's folder is always getting emptied.

Scanning a single folder for a small number of files has very little impact on system performance, so you can afford to run this high-priority job frequently. The queue limit doesn't have much relevance in this case, since you'll rarely have very many files in the queue.

Large Source Folder

If your source folder has thousands and thousands of files in it, you're better off having FileCenter Automate only refill the queue once or twice a day. Crawling through huge numbers of files has a significant impact on system performance, so you don't want that happening every couple of minutes.

The classic example here is when you're converting a huge repository of files. There is no reason to check this very often for new files. Instead, have it check for files just once or twice a

day, and set the queue limit at a reasonably high level – about as many files as can be converted in that interval.

When you're converting a huge volume of files, you want to keep FileCenter Automate busy *converting* files, not looking for them.

Scheduling Options

There are three schedule types:

Always. Think of this as a timer. Every time the scan interval goes off, FileCenter Automate checks for new files and loads them into the queue. The conversions themselves run constantly, as long as there are files in the queue.

Daily. Think of this as a window. You set a start time and an end time. FileCenter Automate will only run the job during this block of time. In other words, FileCenter Automate will completely pause the job outside of the schedule – no scanning and no conversions.

Weekly. This is also a window, just like a *Daily* schedule, except that it can span multiple days.

Daily and Weekly schedules have a *Start Time* and an *End Time*. The job will begin operating at the start time and continue operating until the end time – scanning for new files and processing the queue. For weekly jobs, the end time can be on a different day. This lets a job run only on the weekend, for example.

If your start time is later than your end time, the job will begin at the start time, continue running past midnight, then pause at the end time. For example, if your start time is 6:00 PM and your end time is 6:00 AM, the job will wake up in the evening, run through the night, then pause again at 6:00 AM. This is useful when you want to have a job run after-hours, but not run during the day.

When the end time arrives, FileCenter Automate will pause any further conversions for that job until the job becomes active again. If it is in the middle of a file, it will finish that file before it pauses.

Note: Each job has its own schedule. This means that one job could be paused while others continue running.

Schedules A & B. Note that if you have a *Daily* or *Weekly* schedule, you can set up an optional second schedule. For example, you could use this option to convert files overnight on weekdays

and all day on weekends: set *Schedule A* to *Daily* from 6:00 PM - 6:00 AM and set *Schedule B* to *Weekly* from 12:00 AM Saturday - 11:59 PM Sunday.

FileCenter Automate will ignore any overlap between the schedules. To omit the second schedule, set it to Never.

5.11 Running a Job

If you've set up a schedule for your job, FileCenter Automate will run the job automatically according to the schedule, scanning for files and converting them without any interaction on your part. FileCenter Automate doesn't have to be open for this to work. It will all happen behind the scenes.

If you set your job to only run manually, or if you want to force a scheduled job to scan for more files, follow these steps to launch it:

1. Select the tab for the job you want to run
2. Click *Get Files* on the main toolbar
3. If you want to limit how many files you add to the queue, select *Limit Search To* and provide the limit
4. Click *Begin Search for Files*
5. FileCenter Automate will begin scanning for files. When it's done …
6. Click *Add to Queue* to begin conversion

It's safe to close the FileCenter Automate interface while files are converting. FileCenter Automate will continue working in the background.

Chapter 6: Advanced File Conversions

Most of what you need to know about file conversions was covered in Chapter 5: Job Basics. This chapter covers the outlying topics.

6.1 Convert Individual Files

While FileCenter Automate was designed to convert files in bulk, it can also handle single file conversions.

This is done through drag-and-drop. The *Queue* window lets you drop in files from outside. The dropped files will immediately go into the queue.

You still need to set up a job to set the conversion parameters and tell FileCenter Automate where to put the converted file. Here's the process for drag-and-drop conversions:

1. Set up a job with the conversion options you want to use
2. Go to that job's tab
3. Drop outside files onto the *Queue* window
4. Conversion will begin immediately

6.2 Redoing Conversions

If you need to redo a conversion, you have a few options.

Re-Convert Individual Files

If the original files are still available, select the *Inputs* tab, locate the files, then drag/drop them onto the *Queue* list. This will force a re-conversion.

Alternatively, locate the files in the *Backups* and drag/drop them onto the *Queue* list.

Redo Entire Jobs

To redo an entire job, first get the original files back in place if they're not already there. You'll find the originals under the *Backups* tab. This backup preserves the original folder structure. If

you need to locate the backup folder in Windows, go to the *Details* tab and look at the *Backup Folder Location*. Copy all of the files from the backup folder and paste them back in their original location.

Next you need to purge the *Skip List*. Go to the *Skip List* tab and click *Clear*. This will purge the list and allow your conversion job to run from a fresh start.

6.3 Convert from Word, Excel, etc.

It is possible to convert Office files, including Word, Excel, PowerPoint, and Outlook messages *which have been saved as individual MSG files*.

This requires you to have a working copy of Office on the machine where FileCenter Automate is installed.

Additionally, you *cannot* run FileCenter Automate as a service.

6.4 Convert from Other Text Formats

Similarly, you can convert other text formats like TXT, HTML, EML, etc., as long as these prerequisites are met:

- There must be a program on the machine that can open that type of file
- The program must be set as the Windows default for that type of file
- The program must expose a *Print* option to the shell; in other words, if you right-click on that type of file, you will see a *Print* option on the menu
- You must not run FileCenter Automate as a service

Chapter 7: Automate File Naming

The key to locating files quickly is in the way you name and organize your files. A predictable filing structure and intuitive filenames will allow users to drill down to the file they want without a second thought. This becomes even more important if you're sharing files on the network. Every network user should be able to locate the files they need with minimal hunting. This requires *consistency* and *predictability*.

FileCenter Automate has the ability to alter filenames as it converts according to custom naming rules. These rules can draw from information already known about the file, like its name, location, date created, etc., as well as information within the file itself, like OCR or barcode text. But before we show you how to start creating rules, let's take a few minutes and help you establish some guidelines for naming files.

7.1 File Naming Conventions

If you want your files to be easy to find, you need to be consistent in how you name them, and you need sufficient identifying information in your filenames. If you consistently follow the same pattern every time you name your files, you'll easily predict what a file's name will be and you'll find it without hunting.

Wise users set up a *file naming convention* and stick to it religiously. In other words, they decide how they're going to name their files – what elements a filename will contain and how to format them – and they follow those self-imposed guidelines religiously.

Elements of a Naming Convention

As we've worked with various companies over the years, we've noticed the ones that use clear, predictable filenames. While the filename styles vary widely, we see certain elements show up again and again:

Begin with a date in the format *YYYY-MM-DD*. If being able to sort your documents chronologically matters to your business, consider this a requirement. The *YYYY-MM-DD* format makes it so files will always sort by date. Other date formats that we've seen:

```
YYYYMMDD
YYYY.MM.DD
```

```
YYYY.MMDD
YYYY-MMDD
```

We recommend always using a four-digit year; it makes the date format unambiguous and faster to read, and leaves no doubt about the century. If you don't need the day, or even the month, you can obviously leave those off.

What date should you use when naming a file? For correspondence, the correspondence date. For official documents, like filings and court pleadings, use the filing or service date. For lab reports, the report date. For transactions, the transaction date. For any dated document, the date in the document. Otherwise, use the date you create the document.

Name of the case, account, client, patient, etc. Consider including the person or matter that this document pertains to. This will often correspond to a folder name if you use one folder per case, client, patient, or account. A few examples:

```
1998-12-15, Estate of Esther Jones ...
2012.01.28 John Browning ...
19640129 – Frame, Roddy ...
2001-05-14, Joseph v. Samuels ...
2005.11.06 ~ Johnson Electric ...
```

Transaction number, account number. Alternatively, if you run a transactional business rather than a business based on clients, customers, or patients, you might include a transaction identifier or account number as one of the first elements in the filename.

Document type. Identify what kind of document this is: patient intake form, lab report, purchase order, court brief, opinion letter, official response, invoice, etc. While we don't encourage abbreviations, this is one place where clear, standardized abbreviations can help. See our discussion on abbreviations below.

Correspondence. For correspondence, consider including the party's name along with a "T" or "F" to indicate "to" or "from". Some use the abbreviations *LT/LF* for "letter to/from", *FT/FF* for "fax to/from", and *ET/EF* for "email to/from". Some formats we've seen:

```
... Robert Johnson (LT) ...
... LT Johnson, Sara ...
... ET Jose da Silva – email subject line ...
... FF Sara Bingham, fax description ...
```

Subject matter. If you feel you need to describe the document further, add a short description or a couple of keywords. In the case of email, you might include the original email subject line.

Increment. If you could potentially end up with multiple files sharing identical names, like a lengthy email exchange, put an increment or "counter" at the end. Decide how many

characters you'll need and use a zero-padded number. For example, if you could end up with ten or more versions, number them like this: 01, 02, etc. If a hundred or more, use three digits: 001, 002, etc.

Field separation. Finally, decide how you're going to separate the different elements of your filename. Some users who prefer conciseness will just string everything together with spaces, while others will use special characters to separate the chunks of information. It doesn't really matter what you choose as long as each piece of information "pops" out easily and as long as you're consistent.

You can't choose any of these: **< > : " \ / | ? ***. Common separators are a dash (with or without spaces), commas, tildes, parentheses, and periods. Examples:

```
1998-12-15, Estate of Esther Jones, LT Sara Johnson.pdf
2012.01.28 John Browning PO 01.pdf
19640129 – Frame, Roddy – Lab Results.pdf
2001-05-14, Joseph v. Samuels, Motion to Dismiss.docx
2005.11.06 ~ Johnson Electric ~ LF Human Resources.pdf
```

As a rule of thumb, don't choose a character that's used in any of the fields. For example, if you separate your dates with a dash, don't use dashes to separate the fields:

```
1996-06-15 – Alice Frame – Living Will.pdf
```

… isn't as clear as:

```
1996.06.15 – Alice Frame – Living Will.pdf
```

Or if you display your party names as "Last, First" you won't want to use commas to separate fields:

```
1996-06-15, Frame, Alice, Living Will.pdf
```

… isn't as clear as:

```
1996.0615 ~ Frame, Alice (Living Will)
```

Important: Some long-time PC users still avoid using spaces and truncate their filenames, like this: *SomeFileName.doc*. In the days of DOS, when we were limited to 8 character filenames, that was almost required. But those days have long passed. Use spaces! If you don't, your filenames will be less readable, and worse, indexers will treat the whole filename as a single word, making filenames harder to search.

Other Considerations

Should you include a date? Some files don't need a date context. In those cases, you can obviously omit the date. Research libraries, contact information, and form templates are obvious examples.

When a date does matter, some users prefer to just rely on the file's *Created* date in Windows. You'd be well advised to include that date in the filename as well. Why? Because you can't assume that the Windows date won't change. As a general rule, programs should never change a file's *Created* date, but be aware that some programs will anyway. File synchronization software in particular has been known to fiddle with dates.

Another pitfall to be aware of: You can't assume that the Windows *Created* date is the right date to use for the file. For example, what if an old file was scanned? The Windows *Created* date won't match the internal date in the document. In these cases, you can set up rules that will look inside of the document for a date and include that date in the filename.

Should you use abbreviations? Abbreviations can help shorten the name, and sometimes even make it more readable ... to a point. But if you decide to use abbreviations, limit them to your most common terms and establish a standardized list. Don't let your staff use their own idiosyncratic abbreviations; limit them to the standardized ones, otherwise your filenames will get cryptic and hard to search. Also, don't go overboard with abbreviations. Readability and clarity matter more than brevity.

Good abbreviations will expand easily in your mind. If you're having to pause and decipher every abbreviation, it's going to slow you down. So while "PO" is an obvious abbreviation for "Purchase Order", "MT" could mean either "Motion", "Meeting", or "Minutes". "MOT", "MTG", and "MIN" would be better. Sometimes just taking out the vowels produces a good abbreviation, or taking the first three, four, or five characters of the word, or the first letter of each word.

For a good rule of thumb when setting up your abbreviations, think about a new staff member coming into the office. How quickly and easily will they understand your filing system?

Do I need to include the client, patient, case, etc. if it's in the folder name? Some would argue that if you're putting the file in well-named folders, you don't need to duplicate that information in the filename. For example, if you're filing something in the "ABC Industries" > "Widget Contract Renegotiation" folder, do you need to include those elements in the filename too?

Arguments for: If you remove a file from its current location, it loses some of its identifying information unless that information is also in the filename. For example, if you put a file on a thumb drive, you won't know what client/matter it belongs to unless that information is in the filename. The same argument applies to misfiled items. If the filename holds the file's complete information, it's easy to find missing files and return them to the right location.

Arguments against: In Windows, the safe working filename length is 260 characters. If it gets longer than that, some programs won't be able to work with the file. Note that when we say *filename*, we're actually referring to the *full* filename, folders and all. So if you tend to use really long, descriptive folder names and really long filenames, you could easily surpass 260 characters.

There's also a practical limit on how many characters Windows will display. For example, Windows will rarely show more than about 35 characters of the filename before truncating it.

You can mitigate these limitations with the careful use of abbreviations.

7.2 Intelligent Filenames

FileCenter Automate supports the ability to rename files according to intelligent rules called *Naming Options*. You can set up one default rule for your job, then apply other rules on a case-by-case basis with *Separators* (see Section 2: The Automation Guide).

These rules can contain dynamic elements in them, like dates, client/customer names, counters, etc.

We call these pre-defined filenames *Naming Options*.

Create Naming Options

You'll set up your naming options when you add destinations to your jobs or separators:

1. Add a destination
2. Select *Use Filename Builder*
3. Next to the *Filename* field, select *Add/Edit*

You'll now find yourself in an expression builder where you can construct the name from a combination of static text and dynamic fields. You'll build your name in the *Expression* field. You can type in text by hand:

```
Engagement Letter
IRS Form SS-4
Monthly Newsletter
```

You can also add dynamic fields to the expression. Here are some examples of how fields can be translated into filenames:

```
«FOLDER Parent 1» Engagement Letter
Robert Johnson Engagement Letter

«DATE mmmm» Monthly Newsletter
August Monthly Newsletter
```

To insert a field, do the following:

1. In the expression, position the cursor where you want to insert the field
2. In the lists above, select the *Field Type* and, if applicable, the *Field Option*
3. Click *Insert Field into Expression*

This will insert your field into the name expression. You can add text before or after the field, or cut and paste the field into a different part of the expression. With dates, you can even modify the date format in the field.

To remove a field from the expression, simply delete it using your keyboard. You can also clear the entire expression using the *Clear Entire Expression* link or just the increment portion of the expression using the *Clear Increment link*.

Dynamic Fields

There are a number of dynamic fields available to you. We describe each one in detail at the end of Section 2: The Automation Guide, in the *Reference* section.

Character Formatting within Fields

For advanced users, it's possible to apply some character formatting to an individual field, like altering the capitalization or limiting the field to a certain number of characters. We describe the formatting options, with examples, at the end of Section 2: The Automation Guide, in the *Reference* section.

Section 2: The Automation Guide

Chapter 8: Introduction to Automation

Stop and consider how many times a day you click your mouse. For the average office user, that's 5,000 to 7,000 times a day. Some of that mouse work is purposeful, but how much of it amounts to little more than mindless repetition? And how much of that mindless repetition is really *necessary?* Putting it another way, if we now have computers that can drive cars, can't we get computers to take on some of these monotonous, repetitious, *predictable* tasks for us?

Let's take scanning, for example. Most scanning methods require two things: first, that you scan one document at a time, and second, that you manually name and save each scan. What if you could eliminate both of these steps?

Most offices that deal with a high volume of scanning find themselves scanning the *same documents* over and over and over. Granted, a few key details change from document to document, but the forms themselves are standardized. Why can't a computer be taught to know what kind of document you're scanning, where to find relevant details on the page, and what to do with the document?

Enter FileCenter's automation tools. Both FileCenter DMS and FileCenter Automate share a common set of features that can help you take redundant mouse clicks out of scanning and filing documents – from automatic document separation to naming and saving documents based on text captured within the documents themselves.

8.1 Core Concepts

Separating and Routing

Two primary concepts underlie almost all of the automation features: *separating* and *routing*. While those terms may seem intimidating at the outset, they're not hard to learn.

When we mention *separating* we're talking about automatically splitting up a scan into multiple documents. This implies that instead of scanning documents one at a time, we've scanned a whole stack of documents at once. We're letting the software take care of separatng those documents out into their own files.

When we mention *routing* we're referring to automatically saving a document where it needs to go. This includes formulating a proper filename and choosing the location where the document needs to be saved.

8.2 Which Editions Support Automation?

Unless otherwise specified, the features that follow are supported in:

- **FileCenter Automate:** all editions
- **FileCenter DMS:** the *Pro Plus* edition <u>only</u>

One noteworthy exception is the *Separators* feature, described in the next chapter, which is available in FileCenter DMS *Professional* as well as *Pro Plus*.

Chapter 9: Page Separators

The easiest automation feature to get started with is *Page Separators*. It requires no technical know-how, to the point that nearly any user can generate a separator and put it to use. Yet despite how easy it is to set up and use, the separator feature can give your office a significant productivity boost.

The concept behind page separators is simple. You place a special, printed cover sheet on top of each document you scan. You can then scan the documents individually or as a stack. FileCenter will split the documents apart at the cover sheets. If the cover sheets also contain *routing* instructions, FileCenter will even name and save the documents as instructed.

Using page separators, it's possible to scan, separate, and save a whole stack of documents without touching the keyboard. Just get the scan started and FileCenter will take it from there.

> *FileCenter DMS Pro:* Page separators are available in FileCenter DMS Professional in addition to FileCenter DMS Pro Plus and FileCenter Automate (all editions).

9.1 FileCenter DMS vs. FileCenter Automate

When it comes to using page separators, it's important to be aware of how each program handles them.

Set Up Separators

Setting up new separators is identical in both programs and both programs can share the same separators. So if, for example, your users create separators in FileCenter DMS, FileCenter Automate will recognize those separators if it is set up properly. For more information, see section 19.5, Sharing Separators.

Use Separators

In FileCenter DMS, you can have the software look for separators every time you scan. For existing documents, FileCenter DMS can look for separators when you OCR the document.

In FileCenter Automate, you can instruct your conversion jobs to look for separators in each document that gets processed.

We'll go over the specifics later in this chapter.

Differences in How Documents Are Saved

In FileCenter DMS, you'll probably use separators at the time you scan. In most cases, the scans will be saved into whatever drawer you currently have selected.

In FileCenter Automate, however, things are handled by the conversion job. To understand where documents will go after they get separated, see section Chapter 10:: About Destinations.

9.2 The Separators Dialog

To access the page separator feature, click the *Separators* button on the main toolbar (in either FileCenter DMS or FileCenter Automate).

Here you can view, manage, and print your separators. You'll find that there are four kinds of separators:

Page Separators (Default). Page separators act like cover sheets. You place one of these special cover sheets on top of each document. FileCenter will split up the documents at the separator sheets and follow any routing instructions on the sheet.

User Select. A variant of the page separator, a *User Select* separator lists different routing options like a multiple-choice question. The user can mark the routing choice they want to use.

Bookmark. Another variant of the page separator, a *Bookmark* separator inserts a PDF bookmark in the resulting document instead of separating the document.

Barcodes. If you use barcodes, FileCenter can split up documents based on the occurrence of specific barcodes instead of the presence of a cover sheet.

We'll discuss each of these below.

9.3 Get Started with Page Separators

As we mentioned, it's easiest to think of *page separators* as special cover sheets you put on top of documents. The separators show FileCenter where each document starts so that FileCenter can break the documents off into their own files.

But page separators don't stop there. They can also contain *routing* instructions. In other words, they can instruct FileCenter *where* to save the file, *how* to name it, and *what* options to use (like file type, whether to run OCR, etc.).

Page Separator Basics

Here's an example of how you'll use your page separators in FileCenter DMS:

1. Print off as many copies of the separators as you'll need (they are reusable)
2. Place the appropriate separator on top of each document
3. Put your documents in a stack
4. Click *Scan* on the main toolbar of FileCenter DMS
5. Select the *Separate* tab and make sure *Separate on Page Separators* is selected
6. Proceed with the scan

FileCenter DMS will scan the whole stack of documents. Then, during OCR, it will look for separator pages. Each time it runs across a separator, it will pull that document off of the stack into its own file and, if the separator has other instructions, save the file wherever and however the separator instructs.

The Generic Separator

FileCenter already has one page separator set up for you: the *Generic Separator*. This page separator only has one purpose – to split apart documents. It doesn't have any other options.

Let's get our feet wet on the generic separator:

1. Open FileCenter DMS
2. Click the *Separators* button on the main toolbar
3. The only separator on the list will be *Generic*
4. Select it and click *Print*
5. Close the dialog

Now look at the printed separator. It has a series of codes along the top. These codes are what identifies the separator, and every separator will have a different set of codes. Now let's take the separator for a spin:

1. Sandwich the separator between two sheets of paper (sheet + separator + sheet)
2. Put the stack in your scanner
3. In FileCenter DMS, go to the *Edit* view

4. Click *Scan*
5. On the scan dialog, go to the *Separate* tab and select *Separate on Page Separators*
6. Proceed with the scan

After the scan comes in, you'll see OCR run, then the magic happens – FileCenter DMS will pull the pages apart, each in its own document tab. The separator itself will disappear. A single scan has now given you two documents.

Let's repeat the same experiment in the *Manage* tab:

1. Put the pages back in the scanner
2. Go to the *Manage* view in FileCenter DMS
3. Click *Scan*, and once again, make sure *Separate on Page Separators* is selected
4. Give the scan a name
5. Proceed with the scan

This time you'll end up with two individual files. They both use the same filename, but with counters after them.

> *Tip:* By default you'll also get a file with the original, unsplit scan in it. If you don't want that file, select *Delete Original File after Successful Separation* when you scan. You'll find it on the *Separate* tab with the other separation options.

9.4 Add Routing Instructions

Separators spring to life when you add routing instructions. We refer to a routing instruction as a *destination*. A destination provides three pieces of information: *where* to save the file, *how* to name the file, and *what* file format to use (PDF, TIFF, etc.).

Route into Subfolders

Let's imagine a doctor's office. The office uses FileCenter DMS to store patient records, one drawer per patient. Inside each drawer they have a standard folder layout: one folder for patient information, one folder for billing and insurance records, and one folder for doctor's notes.

When new patients come into the office, they fill out a patient history form and a separate form for insurance/billing info. The receptionist then scans these forms into the system.

This is a perfect case for separators because it's a repetitive filing task.

The office sets up two separators: one for the patient intake form, and one for the insurance form. Each separator's *destination* specifies 1) a filename for the document, 2) a subfolder where it should go, and 3) that it should be saved as a PDF. The receptionist puts the appropriate separator on top of each form, selects the patient's drawer, then scans the stack. The forms automatically get split apart, sent to the correct folders, and named.

How to Set it Up. Let's set up a page separator that will save a file to a subfolder. Click the *Separators* button on the main toolbar in FileCenter DMS, then do the following:

1. Click *Add*
2. Give the separator a unique name; this is just to help you identify it
3. Set the *Separator Type* to *Default*
4. You'll notice a tab for *Destinations*; a destination is a routing instruction
5. Click *Add* in the grid of destinations
6. Select the option called *Relative to Selected Location or Destination*
7. For the *Optional Sub Folder*, type in "Test Scans"
8. Click *OK*, then *OK* again
9. You'll see a prompt to print the separator; click *Yes*
10. Print two copies
11. Close the *Separators* dialog

How to Use It. Now make sure you're in *Manage* view. Select a drawer, then do the following:

1. Get two printed sheets of paper and stack them with the separators in this order: separator + sheet + separator + sheet (in other words, a separator *on top of each sheet*)
2. Put the pages in the scanner
3. Select a drawer
4. Click *Scan* and, once again, make sure *Separate on Page Separators* is selected
5. Proceed with the scan

This time, FileCenter DMS will separate the two sheets and automatically file them into a folder called "Test Scans", as specified on the separator. Now let's add just a little bit more power by including a *Naming Option*.

Include Naming Options

We covered naming options earlier. They're a powerful way to automatically name files. Let's edit your separator to use a naming option.

1. Click the *Separators* button on the main toolbar
2. Select the separator you just made and click *Edit*
3. Select the destination in the *Destinations* tab and click *Edit*
4. Select the option called *Use Filename Builder*
5. Click *Add/Edit* next to the filename
6. We're now in the *Expression Builder*; select *Folder Name > Folder Parent 2*, and *Insert Field Into Expression*
7. Add the text "New Scan" to the end of the expression
8. Select *Auto Increment > 01, 02* then *Insert Field Into Expression*
9. Click *OK* and *OK* again

You don't need to reprint the separators. Now go back to the drawer and re-scan the stack, making sure to use the *Separate on Page Separators* option.

This time, the two files should pick up the name of the drawer, followed by "New Scan" and an increment, like this:

```
Some_Drawer New Scan 01.pdf
Some_Drawer New Scan 02.pdf
```

Separate into Pre-Defined Locations

It's also possible to have a separator send a file to a fixed, pre-determined location. While this is less flexible, it has its uses. Take office mail as an example. The receptionist can set up separators for each person in the office, routing their mail to their own, personal *Inbox* folder. When the mail comes in, the receptionist opens it and puts the appropriate separator on each piece of mail, then scans the whole stack at once. FileCenter DMS routes each piece of mail to the correct inbox automatically.

How to Set it Up. The setup is identical to Route into Subfolders, above, with one difference: when you set up the *Destination*, choose *Fixed Location* and browse to the folder where you want the scans to go.

Let's do a quick experiment:

1. Click the *Separators* button on the main toolbar
2. Select the separator we've been working with and click *Edit*
3. Select the destination in the *Destinations* tab and click *Edit*
4. Select the option called *Fixed Location*
5. Browse to the location where you'd like to save the file

6. Click *OK* and *OK* again

You don't need to re-print the separator.

How to Use It. Now in either *Manage* or *Edit* view, scan the same stack of pages and separators you used for the last tests, again making sure you've selected *Separate on Page Separators*. This time, notice that when everything's done processing, a window opens up at the bottom of FileCenter DMS.

This shows where the documents were sent. Use this to verify that everything went to the right place. Select an item on the list and click *Locate* on the toolbar right above that window. FileCenter DMS will jump to the file. Notice how it went to the exact folder you specified.

> *Tip:* To toggle the *Processed Files* window on and off, go to the *Tools* button on the main toolbar > *Show Processed Files*.

9.5 User-Select Separators

With the page separators you've discovered so far, each separator is tied to a single *destination*, or routing instruction. As a result, you'll end up with quite a few different printed separator sheets. As the number of separators continues to grow, keeping your separator sheets organized can get unwieldy.

FileCenter gives you a way to simplify that: *User-Select Separators*. The idea is this: instead of maintaining dozens of separators with distinct destinations, put all of the destinations into a *single* separator and choose the destination you want at scanning time.

User-Select Separator Basics

If you look at a printed user-select separator, you'll see that it presents a list of destination option almost like a check-list. You put a mark next to the destination(s) that you want to use; the rest will be ignored.

The workflow goes something like this:

1. Print off many copies of your user-select separator
2. For each document in your scan, mark the destination you want to use
3. Place the separator on that document
4. Do the same for other documents in the scan

5. Scan the stack just like you would with regular separators

After you scan, FileCenter will separate each file at the separator and save it to whichever destination(s) you picked.

As you can see, this method allows you to keep a small number of separators that each have a high amount of flexibility.

Create and Use a User-Select Separator

To create a user-select separator:

1. Add a new separator
2. Select *User-Select* as the separator *Type*
3. Add as many *Destinations* as you'd like; each destination will be a separate option
4. Create and print a few copies of the separator

Look at the printed separator. You'll notice that each destination begins with: (-)

To select a destination, put a *short vertical line* through the hyphen: (ǂ)

Your goal is simply to <u>mark out the hyphen without affecting the parentheses</u>. Also, try not to stray above or below the parentheses. A short vertical line is all you need. Once you've marked the separators and placed them on the documents, you'll scan the stack just like you would with page separators.

Important: To mark a destination, your goal is just to destroy the hyphen ("-"). A short vertical line through the hyphen is all it takes. A small "x" through the hyphen will also work. But it's *critical* that your mark not touch the parentheses and not extend above or below the parentheses.

Tip: If you're having trouble getting the destinations recognized, try selecting the option called *Use Alternate Detection Method for User-Select Separators* on the *Separate* tab when you scan or run OCR.

Practical Example

You work in a three-attorney office. Each attorney has a separate area of practice: estate planning, LLC formation, and civil litigation. In addition, there are client intake forms which all clients fill out.

You will be able to automate the majority of the firm's repetitive scan jobs with just four separators: an *Intake Forms* separator, an *Estate Planning* separator, an *LLCs* separator, and a *Civil Litigation* separator.

Start by creating the *Intake Forms* separator. Add a separate destination for each type of document you scan at client intake, setting the folder path and filename as described in section 9.4, Add Routing Instructions.

Do the same for the other three separators, adding a destination for each kind of document that you deal with regularly in that area of practice.

After you've created your separators, print or photocopy a large quantity and leave them next to the scanner. Whenever you need to scan a document, grab the separator that deals with that area of law, mark the appropriate destination, and scan the separator with the document.

9.6 Bookmark Separators

A *Bookmark Separator* isn't really a separator at all. Its job isn't to *split* documents. Rather, a bookmark separator *inserts a bookmark* into the resulting PDF. This feature is included with page separators because it uses the same mechanism: a cover sheet.

Bookmark separators are designed for users who prefer to collect multiple documents into a single PDF. The bookmarks make it easy to jump to the page where a given document begins.

Bookmark Separator Basics

A bookmark separator works similarly to a page separator:

1. Print your bookmark separator
2. Place it on top of the page that should receive the bookmark
3. Scan your document

When FileCenter processes the document, it will place a PDF bookmark on the page following the separator then remove the bookmark separator from the document.

Create and Use a Bookmark Separator

To create a bookmark separator, do the following:

1. Add a new separator
2. Select *Bookmark* for the separator *Type*
3. Enter the bookmark text as the separator *Name*
4. Click *OK* to create the separator
5. Print the separator

Use this separator just like a page separator. Since you'll probably be scanning a stack of documents, your workflow will look something like this:

1. Place the appropriate bookmark separator on top of each document
2. Start your scan in FileCenter DMS
3. On the *Separate* tab, make sure that you select *Separate on Page Separators*
4. Proceed with the scan

FileCenter DMS will scan the document then look for separators. Instead of getting multiple files like you would with page separators, your scan will come in as a single file. When you open it in a PDF editor, you'll find that bookmarks have been inserted on the pages following each bookmark separator. The separators themselves have been removed.

> *Tip:* You can mix page separators and bookmark separators in the same scan. Everywhere FileCenter finds a page separator, it will split the document. Everywhere it finds a bookmark separator, it will create a bookmark.

9.7 Barcode Separators

You have the option to use your own barcodes as separators. This can take three forms: you can do simple separation, where FileCenter starts a new file every time it sees a barcode, you can add routing by saving the file to the path contained in the barcode, or you can associate a barcode with a separator to take advantage of the full power of separator destinations.

Simple Separation

Suppose that every document you deal with has a barcode on the first page. The presence of a barcode clearly signals the beginning of a new document, so you'd like FileCenter to automatically split the scan every time it encounters a barcode. This is easy to do:

1. Click *Scan*
2. Select the *Separate* tab
3. Select *Separate on Barcodes*

Now proceed with your scan. After the scan finishes, FileCenter will quickly look through the pages and split off a new file every time it encounters a barcode. The barcoded pages will be treated as the *first page* of the new file.

The separated files will all show up together in the same folder (unless you scan in the FileCenter DMS *Edit* tab, in which case, they'll each open in new tabs). The files will all have the same name you used for your scan, with an increment: *Intake Form 01.pdf, Intake Form 02.pdf,* etc.

Important: In this scenario, FileCenter does not look at the value of the barcode. It's merely looking for the *presence* of a barcode.

What if every page of the document has a barcode? Some documents include the barcode on every page. Fortunately, that won't stop you from using this feature, as long as every document has a *unique barcode*.

For example, suppose that, in your system, every document gets stamped with a unique document id in the barcode. That barcode, however, shows up on every page of the document. FileCenter can limit separation to unique barcodes. When you scan, select the option called *Separate on Unique Barcodes Only* on the *Separate* tab.

When FileCenter is looking for barcodes, it will only separate a document if it hasn't seen its barcode before *in this scan*. Once it separates on a barcode, it will ignore that barcode for the rest of the scan.

Note: FileCenter only remembers barcodes during the current scan. Once you start a new scan, FileCenter takes a fresh look at the barcodes.

Route to a Barcode Path

In addition to separating documents based on the presence of a barcode, FileCenter can also route documents if their barcode contains a valid file path. To do this:

1. Ensure that the barcode contains a *complete* path
2. Click *Scan*
3. Select the *Separate* tab
4. Select *Separate on Barcodes*
5. Select *Route to Path Contained in Barcode*

Note that a "complete path" includes a *drive*, *folder(s)*, and a *filename* (including an extension). If the path isn't valid or accessible, FileCenter DMS will still separate the document, but will save it to the same folder and filename as the original scan (with an increment).

If the path in the barcode happens to point to an existing file, FileCenter will save the new file with an increment.

> *Tip:* One advantage to this method is that you can share these barcodes among multiple users without having to share a separator database, since this kind of barcode is completely self-contained.

Create a Barcode Separator

The most flexible and powerful way to separate with barcodes is to create a barcode separator. All of the naming and routing power that comes with a FileCenter separator becomes available whenever you use the barcode. Yet you don't need to mess with printed separator sheets. It's the best of both worlds. To use this option:

1. Either edit an existing separator or create a new one
2. Set the separator *Type* to *Barcode*
3. Enter the barcode's *value* in the *Barcode Text* field
4. Add a destination
5. Set any other values you want for the separator, then click OK
6. Start a new scan
7. Select the *Separate* tab
8. Select *Separate on Barcodes*
9. Select *Route Via Separators Database*

FileCenter DMS will now look for barcodes in the scan and, if it finds one, attempt to look it up in the separator database. If the barcode matches a separator, FileCenter will act on the separator's settings as if it had found the separator in the document. If it doesn't find a matching separator, FileCenter will skip the barcode.

Make Reusable Barcode Separators

It's possible to create reusable barcode separator pages that you use like cover pages. These sheets can contain any other text or information you'd like; it won't interfere with their function. To make this work:

1. Either edit an existing separator or create a new one
2. Set the separator *Type* to *Barcode*
3. Enter the barcode's *value* in the field provided
4. Add a destination
5. Set any other values you want for the separator, then click OK
6. Start a new scan
7. Select the *Separate* tab
8. Select *Separate on Barcodes*
9. Select *Route Via Separators Database*
10. Select the *Advanced* tab
11. Select *Remove Barcode Pages in Destination Documents*

9.8 Other Separator Topics

Using Separators with Existing Files

In FileCenter DMS, it's possible to insert page separators into an existing file. When would you want to do this? Imagine a medical practice. A patient gets referred by another doctor, and all of the patient's records get sent over in a single PDF file. The front office staff can quickly browse through the file and insert page separators wherever a new document starts. Then they process the separators in the file. FileCenter DMS automatically splits up the file into individual documents and stores each one where it belongs.

Here's how to do it:

1. Open the PDF in the *Edit* tab
2. Find a page where a new document starts
3. Click the drop-arrow under the *Separators* button on the main toolbar
4. Select the separator you want and click *Insert Before* or *Insert After*
5. Do the same for other documents in the file
6. When you're done, click the *OCR* button on the main toolbar
7. Select *Process Separators Only*
8. Click *Start OCR*

FileCenter DMS will go through the file, locate the separators, then break up and route the individual documents accordingly.

> *Note:* If you use separators with *relative* destinations, the original file's directory will be treated as the active directory.

What about Naming Conflicts?

FileCenter will never overwrite an existing file, even if there is a name conflict. When a name conflict happens, FileCenter will use counters to keep names unique. For example, suppose that the file *scan.pdf* already exists. If FileCenter needs to place a new file, also called *scan.pdf*, in the same folder, it will save the new file as *scan_2.pdf* to prevent a conflict. If another *scan.pdf* comes along later, FileCenter DMS will save it as *scan_3.pdf*", and so on. You can customize the format of the increment value by using the *Filename Builder* option on the *Destination* dialog (*Separators > Add/Edit >* add or edit a *Destination > Use Filename Builder*).

Common Mistakes with Separators

If separators don't seem to be working for you, see if one of these scenarios applies.

Using Separators Created by Another User. You can only use a separator on the machine where it was created (unless you've set up shared separators, like we describe in section 19.5, Sharing Separators). If you try to use a separator on another machine, FileCenter will see that it's a separator but won't know what to do with it.

To explain, when you use a separator, FileCenter extracts the separator codes from the top of the page and looks the separator up in its database. If the separator isn't there, FileCenter will treat the separator like a *Generic Separator* and just separate the documents.

Continuing to Use a Deleted Separator. Once you've deleted a separator from FileCenter, you can't continue using the printed copy. If you do, FileCenter will see that the page is a separator, but when it looks up the separator codes, it won't find anything and, consequently, won't know what to do with the separator. In this case, it will treat it like a *Generic Separator*.

Re-Creating a Separator. If you delete a separator in FileCenter and then re-create it using the same name, you must re-print the separator and destroy the old copies. If you try to use an old copy, FileCenter won't recognize it, even though the new separator uses the same name.

Multiple Destinations. When users want a separator that routes to multiple destinations, sometimes they make the mistake of selecting the option called *Choose among Destinations by Marking on Printed Separator*. This creates a multiple-choice separator, as explained in section

9.5, User-Select Separators. If you enable that setting, FileCenter won't route to *any* destinations by default.

If you simply want to send a document to multiple destinations, all you have to do is add the destinations to the separator.

All Separator Options Explained

Besides the features we've already touched on, separators offer a number of useful options. Here is the full line-up of separator capabilities. You'll see these when you create a separator or add a destination to a separator.

Separator Name. This is for your reference. It will get printed on the paper separator. Try to use a name that describes what the separator does.

Separator Type. Page separators which we've described above:

Default. A page separator that separates and routes. You'll select this most of the time.

Bookmark. If you select this option, the separator will insert a PDF bookmark instead of splitting up the document. The bookmark will use the name of the separator. No other options are available for bookmark separators.

User Select. This creates a multiple-choice separator. FileCenter will print the separator as a check-list of destinations. You can choose the destination you want by marking it on the separator. After you scan with this kind of separator, FileCenter DMS will only save to the destination(s) that you picked.

Begin a New Document Every __ Pages. A single separator can optionally split a document at fixed page intervals. For example, suppose that you scan a whole stack of 3-page documents. A single separator will split it into individual files every three pages.

To enable this, go to the *Destinations* tab and select *Automatically Begin New Document Every __ Pages* and enter the number of pages you want in each document. If using this with a printed separator, put it at the *beginning of the stack*. Only insert it once. It will remain in effect until the end of the file, or until FileCenter runs into another separator.

Destinations – Fixed Location. A separator can always route the file to exactly the same place (fixed location), or it can go to your current location in FileCenter (relative location). The *Fixed Location* option for a destination does just what it says: sends the scans to that folder every time.

Destinations – Relative Location. Relative locations are designed for scanning in the *Manage* view. The separator will either save the scan to the selected drawer/folder or, if your separator specifies a sub folder, it will create the sub folder (if necessary) and put the scan there.

Destinations – File Type. Select the format that you want the scan saved in. It can be any of:

PDF – This option produces a PDF file. If you use the *Make Searchable PDF* option when you scan or run OCR, this will be a searchable PDF.

TIFF – This format is simply an image of the original file. There is no actual text. The file will look like the original, but you won't be able to search its contents or copy any text from it.

Text – A text file is simply the OCR text without the scanned image itself. Text files are searchable, and you can edit them in any word processor. You might select this option if you only care about the text of the document and the formatting does not matter.

Destinations – Use Subfolder Name Builder. Use this option to dynamically create subfolders off of the main folder using naming option fields. The classic use is creating dated subfolders, which will put the files in dated subfolders within the destination folder.

Destinations – Use Filename Builder. Select this option if you want to use dynamic filenames. See p. 47, Dynamic Fields for help building a filename.

Note: If you don't use the filename builder, FileCenter DMS will use the original document name, if available, or a generic document name.

Metadata. Your separator can automatically populate some of the PDF metadata fields. For more information, see the FileCenter DMS chapter on scanning options or the FileCenter Automate chapter on job options.

Map. It's possible to mine data from the document's contents and export it to an external data file. This is accomplished through *Content Maps*. See Chapter 17: Data Mining with Content Maps for more information.

Chapter 10: About Destinations

Before moving on, let's spend an extra minute talking about *Destinations*.

Recall that two key concepts underlie everything in FileCenter's automation features: *separation* and *routing*. Any given document may or may not need separation, but *every* document needs routing. In other words, every document must get saved somewhere.

Let's break that down a little bit further. In order to save a document, FileCenter needs to know a few things:

Where Will It Go? FileCenter needs a path for the document. The current folder? A subfolder of the current folder? A predetermined folder? A new folder created from dynamic information?

How Will It Be Named? Documents also need a filename. Do you want to keep the original document's name? Do you want to construct a name from pre-determined rules? Do you want to build a name out of text in the document itself?

What Format Will Be Used? Finally, FileCenter needs to know what kind of document you want to create. Do you want a PDF document? A TIFF document? A plain text file?

10.1 Destinations Are the Answer

Every question above is handled by a single feature: *Destinations*. Destinations are the heart of routing. You already got a taste of destinations back in the chapter on separators. Destinations will continue to play a role going forward as we cover the other automation features. Any time FileCenter routes a file using automation, it needs a destination.

At its simplest level, a destination is simply an instruction that tells FileCenter *where* to save the document, *how* to name it, and *what* format to use.

The Name Builder

When you instruct FileCenter *where* to save the document and *how* to name it, you can provide a static folder or file name. For example, if you have a separator for saving an intake form, you may want to always put it in a folder called *Client Info* and call the document *Intake Form.pdf*. Simple and straightforward.

You also, however, have the ability to generate folder names and file names *dynamically*. This includes everything from dated subfolders to filenames that draw from the body of the documents.

When you add a destination, you'll notice two options: *Use Subfolder Name Builder* and *Use File Name Builder*. This is the same name builder that's used to create *Naming Options* in FileCenter DMS. If you've already used naming options, you'll be familiar with them.

The name builder gives you a wide range of naming power. For example, to create a dated subfolder:

1. Select *Use Subfolder Name Builder*
2. A *Subfolder* field will appear
3. Next to the field, click *Add/Edit*
4. The *Name Builder* dialog will appear
5. Select the *Date Today* option
6. Select the format you want to use
7. Click *Insert Field Into Expression*
8. Click *OK*

The naming expression is inserted into the *Subfolder* field. Now whenever FileCenter uses this destination, it will automatically create the dated subfolder if it doesn't already exist.

A few things to note:

Static Text. Your naming expression can be a mixture of dynamic fields and static text. You can enter the text either in the *Name Builder* as you're building the expression or directly in the *Subfolder* or *Filename* fields.

Nested Subfolders. You can go as deep as you'd like with subfolders; just put a backslash "\" between each one. DON'T include an initial backslash.

Naming Conflicts. If your destination creates a filename that conflicts with an existing file, it will add an increment. Note that one of the dynamic fields available to you is a formatted increment. Use this to control the format of the increment.

Available Fields. There are a number of dynamic fields available to you. For a full reference, including examples, see 0 Dynamic Fields in the Reference section at the end of this guide.

Character Formatting. In addition to the dynamic fields, there are some character formatting flags that you can use, for example, to force text to uppercase or lowercase, capture only certain characters from a field, etc. For a full reference, including examples, see 0 There are a

number of dynamic fields available to you. We describe each one in detail at the end of Section 2: The Automation Guide, in the *Reference* section.

Character Formatting within Fields in the Reference section at the end of this guide.

10.2 Relative Destinations

When you add a destination, you can either use a *fixed* location or a *relative* location. The difference is simple: a *fixed* location includes an unchanging drive and path (even if it adds subfolders on the fly). A *relative* location, on the other hand, changes according to your context. This behaves differently between FileCenter DMS and FileCenter Automate.

Relative Destinations in FileCenter DMS

In FileCenter DMS, relative locations are determined according to the active drawer/folder.

For example, suppose you're in the drawer for "Abbott, Scott" and you scan a document that contains an intake form and some estate planning documents. There are separators for each document. The separators split up the documents and save them into <u>relative</u> subfolders called *Input Forms* and *Estate Planning*.

Because "Abbott, Scott" is the active drawer, the documents will get separated and routed as follows:

```
R:\Client Files\Abbott, Scott\Input Forms\Client Intake.pdf
R:\Client Files\Abbott, Scott\Estate Planning\Trust.pdf
R:\Client Files\Abbott, Scott\Estate Planning\Will.pdf
etc.
```

If you were to scan the *same documents* in a different drawer – like "Davis, Sarah" – the documents would still get their own subfolders, but they'd be in Sarah Davis' drawer instead:

```
R:\Client Files\Davis, Sarah\Input Forms\Client Intake.pdf
R:\Client Files\Davis, Sarah\Estate Planning\Trust.pdf
R:\Client Files\Davis, Sarah\Estate Planning\Will.pdf
etc.
```

In other words, a *relative* destination will change according to where you're at in FileCenter DMS.

Relative Destinations in FileCenter Automate

In FileCenter DMS, relative locations will be based off of whatever drawer or folder you currently have selected. In FileCenter Automate, however, relative locations become a lot less clear. You don't have a "current location" in FileCenter Automate … or do you?

In fact, you do. In FileCenter Automate, the current folder is the folder where the converted file *would be saved by default*. The easiest way to understand this is with an example.

Suppose you're converting all of your old client files, which are separated into folders by client name. The converted copies are going onto a different drive, but they're retaining the same folder structure.

Example input folders:

```
E:\Client Files\Abbott, Scott
E:\Client Files\Brown, Rufus
E:\Client Files\Davis, Sarah
etc.
```

The converted files will show up in:

```
R:\Client Files\Abbott, Scott
R:\Client Files\Brown, Rufus
R:\Client Files\Davis, Sarah
etc.
```

In the "Abbott, Scott" folder, there's a file called "Scanned_Documents.pdf":

```
E:\Client Files\Abbott, Scott\Scanned_Documents.pdf
```

As in the last example, this file contains an intake form and some estate planning documents. There are separators for each document. The separators split up the documents and save them into relative subfolders: *Input Forms* and *Estate Planning*.

When FileCenter Automate processes this file, by default it should stay in the same location, but on the new drive:

```
R:\Client Files\Abbott, Scott\Scanned_Documents.pdf
```

So the *current folder* is "R:\Client Files\Abbott, Scott".

In our case, the file has separators in it – separators that want to put the documents into *relative* subfolders. So because "Abbott, Scott" is the current folder, the documents will get routed as follows:

```
R:\Client Files\Abbott, Scott\Input Forms\Client Intake.pdf
R:\Client Files\Abbott, Scott\Estate Planning\Trust.pdf
R:\Client Files\Abbott, Scott\Estate Planning\Will.pdf
etc.
```

10.3 Additional Options

Anywhere you can use a destination, you'll also see these options:

Metadata. FileCenter provides a number of metadata fields you can use to automatically add searchable keywords to the document when you route it (if it's a PDF). If the document already has its own metadata, FileCenter will *append* the new values to each field unless you select the *Replace Existing* option.

Map. You can optionally mine text from the document and save it to a CSV file. We discuss this in Chapter 17: Data Mining with Content Maps.

Chapter 11: Content-Based Automation

One of the most powerful features available to you is the ability to tap into a PDF's content and use its text to automate portions of your workflow. For example, you can pull names, dates, invoice numbers, and other information from a PDF and use them to create dynamic filenames. You also have the ability to route new PDF documents based on what kind of information they contain.

The mechanics of these operations are explained in the next few chapters, so we won't get into them here. Instead, this chapter devotes itself to the overarching concepts behind FileCenter's automation features.

11.1 Fields and Rules

At the heart of FileCenter automation sits the *Content Field*. This is exactly what it sounds like: a mechanism for extracting a very specific piece of document content, which you can then use in a number of different ways. Every aspect of automation begins with a content field.

You can use content fields on their own. For example, you'll notice that the FileCenter DMS *Naming Options* can use content fields. This means that you can pull document text into your filenames.

It doesn't stop there, though. Beyond simply pulling text into filenames, you can make *decisions* based on what you find in the document. This is where *Content Rules* come into play. Content rules *answer questions* about the document. Does this document look like a Will? Does it have an invoice number? Does it have the header of a pleading? Does this patient belong to Doctor Jones? Content rules always give a yes/no answer to a question.

Finally, *Document Rules* take those answers and put them to action: If this is a will, extract the client name from the document and save it to the client's drawer in FileCenter DMS. If this is a patient intake form and the patient belongs to Doctor Jones, create a new patient drawer in Dr. Jones' cabinet in FileCenter DMS and save the form there.

In summary, *Content Fields* extract information from a document, *Content Rules* answer questions from those fields, and *Document Rules* perform actions based on the answers.

Content Maps

There is an additional automation feature that doesn't fulfill any function in FileCenter. Rather, it exists solely to benefit your other programs.

Content Maps take the information that was gathered in the content fields and exports it in a format that outside integrations can read. Why would you want this? Take a medical practice, for instance. The client intake form has most of the information that your practice management database needs. FileCenter can capture that information in the form (as long as it is typewritten) and export it to file. Your practice management software can then pull that information into the database.

Thoughtful use of this feature can eliminate a lot of the data entry that weighs down your staff.

11.2 How Does Automation Happen?

We'll go into the specifics later, but in a nutshell:

You can use *Content Fields* in your FileCenter DMS *Naming Options* and in your separator *Destinations*. These fields will work any time that you scan or OCR a document.

In FileCenter DMS you can also perform *Jobs* which apply the full power of *Document Rules* when you perform a new scan or when you OCR/convert an existing document.

In FileCenter Automate, your conversion jobs can call upon *Document Rules* to intelligently decide what to do with each document you process.

Chapter 12: The Automation Setup Tool

We have provided a tool to help you create your *Content Fields, Content Rules,* and *Document Rules*. This tool lets you test and debug interactively against sample documents. You can see the result of any content field, content rule, or document rule in real-time without affecting your live documents.

To access the tool, go to the *Admin* tab in FileCenter, then click the *Automation Setup* button.

12.1 Packages

You'll see tabs for *Packages*. There will always be at least one package: the default *FileCenter* package.

What is a package? A package is simply a bundle of FileCenter settings and options. For example, the *FileCenter* package has all of your local settings in it, including any FileCenter DMS cabinets or naming options, any FileCenter Automate conversion jobs, etc.

We'll devote a whole chapter to packages later on, but for now you may want to create a special package you can use to experiment with the automation features:

1. Click *Packages* on the main toolbar
2. Click *New*
3. Call the package *Sandbox* (or anything else you'd like)
4. For the *Folder Location*, browse to any folder where you'd like to store these settings
5. Click *OK* to create the package then *OK* to close the *Packages* dialog

You'll now see a tab called *Sandbox*. Make sure that you're always on this tab as you experiment with the automation features. That will keep your experiments from affecting your live FileCenter settings.

12.2 The Sample Document

One of the most powerful features of the automation setup tool is the interactive Sample Document window.

If you haven't already activated the *Automation Setup* button on the main toolbar, click it now. A *Sample Document* window will open on the right.

You can open any PDF document in this window and not only use it as a reference while you create your fields and rules, but actually draw zones directly on the document. You can also test your fields and rules on the document.

Go ahead and open a sample document that you'd like to experiment with. The automation tool won't make any changes to it.

Tip: If the document is a scan, make sure that it has already been OCRed, otherwise it won't have any text in it to work with.

12.3 The Text Window

Because your content fields rely so heavily on the document text, we've provided a number of ways to view the text.

Taking a step back, it's important to understand how PDF text works, especially with scans.

PDF scan has two aspect to it: the visual *image* of the document and the actual *text*. The image is just that – a digital photo. It's nothing more than a picture of the original document. The image has no text in it. While you see words, you're only looking at a picture of words. To your computer, it's just a picture, no different than a picture of mountains or flowers.

If, however, you've OCRed the scan, your computer compared the shapes in the image against its library of letter shapes and took an educated guess as to what text might be in the image. It then took this text and placed it invisibly behind the scanned image.

So now your PDF is really two documents: a scanned image on top, and a text document underneath.

The problem is, you can't see the text portion – because it's invisible.

The Automation Setup tool solves this problem. It allows you to view the underlying text. With a *Sample Document* open, click the *Text Window* button. A new panel will open that shows the underlying document text for the current page.

Note: To see the text for a different page, select it on the *Page* dropdown above the document preview.

FileCenter can show the PDF text formatted in four different ways. Note the *Text Option* dropdown in the text window. Change it to *Continuous*. All of the text on the page will be reformatted into one continuous blob of text. Now change it to *Lines*. The text is now presented in lines that match the horizontal lines of text in the original document.

Here are the four formatting options:

Continuous. All of the text is presented in a single, continuous paragraph, each word separated by a single space.

Paragraphs. FileCenter will attempt to reformat the text into paragraphs. FileCenter regards a period (".") followed by a return as the end of a paragraph.

Lines. The text is presented in lines as it appears in the original document. Note that some OCR engines recognize columns. In this case, a line will be limited to the column it's in.

Words. The text is presented one word per line.

Why does this matter? When you start writing rules to pull text out of your document, you'll find that sometimes the text is easier to target if it's broken up into exactly the same format found in the document (like address blocks, for example), and other times it's easier to target in paragraph format, like when you need to locate information based on the words around it.

When you create your content fields, you'll be able to select what kind of text formatting to use for each one.

As you create these fields, you'll find it invaluable to be able to see exactly what the document text looks like in each of the different formats.

12.4 Zones

One type of *Content Field* you'll rely on heavily is the *Zone* field. As its name implies, this field grabs text from a specific "zone" or rectangular region on the page.

The Automation Setup tool makes it easy to define those regions.

Add a New Zone

With a sample document open, do the following:

1. Click *New Zone* above the document preview
2. The mouse cursor will change to crosshairs
3. Draw a rectangle around the region you'd like to capture in the document
4. The *Content Field* dialog will open
5. Give the field a name
6. Click *OK*

Now select *Content Fields* on the left-hand side of FileCenter. You'll see the field you just created. Select it. The zone will appear on the sample document.

Update a Zone

To modify a zone, do the following:

1. Select *Content Fields* on the left, if you're not there already
2. Select the content field you want to modify
3. Its zone will appear on the sample document
4. Click on the zone in the sample document
5. It will change to resizable borders
6. Drag the borders to resize them, OR …
7. Click the circle in the middle of the zone to move it

FileCenter will update the zone in real-time.

12.5 The Results Window

At the bottom of your screen you'll also see a *Results* window. You'll use this window to test your *Content Fields, Content Rules,* and *Document Rules*.

For example, do the following:

1. Select *Content Rules* on the left if you're not there already
2. Select the zone field you created in the prior step
3. The text that falls within the zone will appear in the *Results* window

We will describe the *Results* window in more detail over the next few chapters as we explore each of the automation features. For now, it's enough to say that the *Results* window will show you, in real time, the results of any field or rule. If it's not already obvious, the *Results* window, the *Text* window, and the *Sample Document* are going to become your best friends.

Chapter 13: Content Fields

Content fields extract text from a document. You can then use that text to name documents, you can use it to choose or create folders, or you can use it to make decisions with rules. Content fields come in a few different flavors, each of which uses a different technique to extract text:

Barcode Fields. Barcode fields pull the information out of a barcode. The only thing you need to specify is which barcode to use. FileCenter will extract the barcode data and return it in the field.

Zone Fields. Zone fields pull text from a document based on specific coordinates. In other words, you define a rectangle on the page. Any text that falls within that rectangle will be returned. Zone fields are the perfect choice if the text you're targeting never moves on the page. For example, you could use a zone field to pull an invoice number from a standardized invoice.

Regular Expression Fields. A *regular expression* pulls text from the document based on a pattern. For instance, suppose you want to capture a social security. You'll set up a pattern that looks for three digits, a dash, two digits, a dash, then four digits. You can also extract text based on neighboring text. For example, a will often begins with the text "Last Will and Testament of _____." You can capture the text that follows this phrase.

13.1 Add a Content Field

We recommend that you do all of your content field development in the *Admin* tab, which we described in Chapter 12: The Automation Setup Tool.

We also recommended that you set up a special *package* called "Sandbox" for experimenting with the automation features. For instructions, see page 74, Packages.

Go to the *Admin* tab now and select your *Sandbox* package. We also recommend that you open a sample document to experiment with, as described on page 74, The Sample Document.

On the left side of the screen you'll see all of the different kinds of items you can add to your *Sandbox* package. Select *Content Fields* then click *Add* above the list of content fields.

Common Field Options

All content fields include the following options:

Name. Each content field requires a unique name. Use a name that helps you identify the field easily. This name will be displayed everywhere that you use the content field.

Categories. If you're dealing with a large number of content fields, it helps to break them up into categories. Enter any categories you'd like to assign to your field, comma-separated. You can then filter your fields by category in the content fields list. The categories are for display purposes only.

Description. The description is for your purposes. Use it to supplement the name, giving further information about what this field does.

Formatting. You can apply some character formatting to the text that gets pulled from the document. For example, you can convert it to all uppercase letters. For a complete list of formatting codes, see 0 There are a number of dynamic fields available to you. We describe each one in detail at the end of Section 2: The Automation Guide, in the *Reference* section.

Character Formatting within Fields.

Page. Since a content field pulls information out of the document, it needs to know which page to look on. You will specify this with the *Page* option.

Regular Expression & Return Expression. Every kind of content field supports pattern-matching. For example, you may want to target a specific zone of text, then use a pattern to further refine the exact text you want to pull out of the zone. Pattern-matching is accomplished through *Regular Expressions* and *Return Expressions*. We'll go into more detail below.

Text Option. With *Zone* and *Barcode* fields, you need to specify how the text should be formatted. We describe the text formatting options on page 76, Text Formatting. Text formatting comes in handy when, for example, you need to target information based on lines. If you capture an address block, you need to know where the returns are. Conversely, if you need to target information based on the surrounding text, you'll want to choose *Paragraph* or *Continuous* text formatting.

13.2 Regular Expression Fields

Regular Expression or *RegEx* fields are probably the most difficult to understand but also the most powerful. In short, regular expressions allow you to extract text from the document based on a pattern. For example, you can write a pattern that looks for things like:

- Dates
- Addresses
- Invoice numbers
- Social security numbers
- Names, based on surrounding text

Regular expressions aren't unique to FileCenter. They are widely used in the computer world as the standard way to capture text based on patterns. Because of this, you can find a lot of help on the Internet.

Important: Many regular expression examples that you'll find on the Internet put quotes and/or forward-slashes around the expression like this: "/expression/". **Do not** put the expression in quotes or slashes. The only thing you need to enter is the expression itself.

Your First Regular Expression

We devote a whole chapter to writing regular expressions in Chapter 22: Regular Expressions. For now, we'll cover just enough to give you a taste for regular expressions and how to use them.

If you don't have a sample document open in the FileCenter *Admin* tab, open one now (see p. 74, The Sample Document). With a sample document open, do the following:

1. Select *Content Fields* on the left-hand side of the *Admin* tab
2. Click *Add* to add a new content field
3. Give the content field a name
4. Set the *Type* to *RegEx*

In the *Regular Expression* field, enter this value:

```
[A-Z][a-z]+
```

Explanation: This pattern looks for a single capital letter followed by any number of lower-case letters. In other words, it will give you the first capitalized word in your document.

Click *OK*. Now look at the *Results* window at the bottom of FileCenter. It should display the first word in the document. That's the value that your regular expression field is capturing.

The Text Option

In the sample document you have open, click the *Text Window* button. This will display the document text, as described on page 75, The Text Window. Change the *Text Option* in the text window to see how it changes the way the text flows.

With that in mind, now do the following:

1. Select the content field you just created and click *Edit*
2. Change the *Regular Expression* to: **^.*$**
3. Change the *Text Option* to *Lines*
4. Click *OK*

Explanation: The pattern you entered gives you all of the text on the first line of your document.

In the *Results* window you should see the first line of text from your document. To compare, change the *Text Option* under the sample document to *Lines*. The first line in the *Results* window will match the first line in the *Text* window.

Edit your content field again and change the *Text Option* to *Continuous*. Click *OK*. The *Result* window will now show you all of the text on the page. Why? Because with *Continuous* text, all of the text comes back in one continuous line.

As you create your regular expressions, you'll want to keep in mind how the text flow affects the way you target your text. Choose the flow that makes it easiest to isolate the text you're looking for.

In most cases, the text flow is irrelevant and you can leave it set at *Continuous Text* or *Paragraphs*. If, however, you need to target text based on its **position in a line**, you'll want to set the *Text Option* to *Lines*.

For example, suppose that the customer's name is always the last text on the line and always follows "Bill To". You can set your regular expression to grab all of the text between "Bill To" and the end of the line:

```
Bill To (.*)$
```

Format Your Output

Edit your content field again. This time, set the *Formatting* option to: **-UC**

Click *OK*. The *Results* window will now show all of the text in UPPERCASE letters. The *Formatting* option gives you quite a bit of power to reformat and manipulate the text you get back from the field. We cover all of the formatting options in 0 There are a number of dynamic fields available to you. We describe each one in detail at the end of Section 2: The Automation Guide, in the *Reference* section.

Character Formatting within Fields.

The Return Expression

By default, a regular expression field will give you back whatever text the pattern matched (if there are multiple matches, you'll get the first one back). You can, however, use parentheses in your regular expression to isolate specific pieces of text. Further, you can rearrange those pieces of text and even add additional text and punctuation.

Edit your regular expression field once again, and change the *Regular Expression* to the following:

```
([A-Z][a-z]+) ([A-Za-z]+)
```

Interpretation: This pattern looks for a word that begins with a capital letter followed by a space and another word (which can contain either capital or lower-case letters).

Notice that the two words are placed in parentheses. The parentheses give you specific access to those bits of text. To access the text in parentheses, put a backslash "\" in the *Return Expression* followed by a number indicating which set of parentheses to access. Going back to our example, put this in the return expression:

```
\2 \1
```

Click *OK*. The *Result* window is going to show the first two words in your document *reversed*. Why? Because you asked FileCenter to give you the second word first.

You can also put additional text and punctuation in the return expression. For example, edit your content field and change the *Return Expression* to this:

```
The first word is \1 and the second word is \2.
```

Click *OK* and examine the results.

Getting Back the Whole Match

Normally if you want the whole match, you should leave the *Return Expression* blank. But if for some reason you need to include the whole match in the *Return Expression*, use "\0" (a backslash and the digit 0).

Suppose you have the following regular expression to extract a date from the document:

```
([0-9]+)/([0-9]+)/([0-9]+)
```

Interpretation: This pattern looks for three sequences of digits separated by slashes, so it will find a date in this format: 7/20/1969. Note that there are parentheses around each date element to isolate the values.

Here are all of the placeholders you could use in the return expression along with the value each would give you for the date 7/20/1969:

- **\0** = 7/20/1969
- **\1** = 7
- **\2** = 20
- **\3** = 1969

Working with Dates

One of the most powerful uses for regular expression fields is locating the document date. For example, many office like to use a date in the filename. Using a regular expression field, you can pull the document's actual date and use it in the filename.

But there's a problem with this approach. The date in the document is rarely formatted the same way you'd want to use it in a filename. For example, suppose that the date in a document shows up in the typical format: *July 20, 1969*. That's not terribly useful for a filename. We always recommend that filenames use the format *1969-07-20*. Which raises the question, how do you reformat document dates for use in filenames?

You can do this through the field's *Formatting* option. Just enter *-DT* as the formatting flag, followed by the date format you want to use:

```
-DT yyyy-mm-dd
```

The "-DT" signals that you want to use a different date format, and the "yyyy-mm-dd" indicates what format you want to use.

The date formatting codes are explained on p. 112, Dates.

13.3 Zone Fields

Zone fields are exactly what they sound like. You define a rectangle on the page. Any text that falls within that rectangle will come back in your field. Zone fields are the perfect choice if the text you're targeting never moves on the page. For example, suppose you want to grab the date from the top of a letter. The date never moves – it's always centered at the top of the page. You put a rectangle around the date. As you scan similar letters going forward, FileCenter will always give you the date that falls within that rectangle.

Before you create a zone field, you will need a sample document that you can open in the preview window. You will draw your zone on the sample.

Your First Zone Field

If you don't have a sample document open in the FileCenter *Admin* tab, open one now (see p. 74, The Sample Document). With a sample document open, do the following:

With a sample document open, do the following:

1. Click *New Zone* above the document preview
2. The mouse cursor will change to crosshairs
3. Draw a rectangle around the region you'd like to capture in the document
4. The *Content Field* dialog will open
5. Give the field a name
6. Click *OK*

Now select *Content Fields* on the left-hand side of FileCenter. You'll see the field you just created. Select it. The zone will appear on the sample document.

Also look at the *Results* window. It will display the text from your zone.

Now do the following:

1. Edit your content field
2. Change the *Text Option* to *Continuous*
3. Click *OK*
4. Note how the text changes in the *Results* window
5. Edit the field again

6. Change the *Text Option* to *Lines*
7. Click *OK*
8. Note, once again, how the text flow changes in the *Results* window

Tip: Grab as much space around the text as you can in order to give yourself some buffer since page alignment can vary from scan to scan. If the zone picks up fragments of outside text, they'll be ignored if at all possible.

Update a Zone

To modify a zone, do the following:

1. Select the content field you want to modify
2. Its zone will appear on the sample document
3. Click on the zone in the sample document
4. It will change to resizable borders
5. Drag the borders to resize them, OR …
6. Click the circle in the middle of the zone to move it

FileCenter will update the zone.

Tip: If you add a zone field but don't give it a zone, you can add one later. Select the field on the *Content Fields* list then draw the zone on your sample document. FileCenter will add the zone to your field.

Add a Regular Expression to a Zone

One thing you'll notice is that you can add a regular expression to a *Zone* field. Pairing a zone with a regular expression creates powerful possibilities.

When you create a zone field, you are, in essence, narrowing in on an isolated section of the page text. A regular expression can then target specific text within that zone. The field will only give you the text that matches the regular expression.

In other words, the zone limits the text to a specific region of the page, then the regular expression further limits the text to whatever matches the pattern. Judicious use of zones paired with regular expressions can vastly simplify the task of isolating the text you want on the page.

For example, suppose that a document contains multiple dates but you only want to capture the date that falls in the lower 10% of the page. Draw a zone around the bottom of the page then add a regular expression to extract the date from that zone.

Alternatively, suppose that you want to pull the zip code from an address block. Draw your zone around the address block then add a regular expression to target the zip code. Only the zip code will come back in the field.

13.4 Barcode Fields

A *Barcode* content field does one simple thing: it extracts the information out of a barcode and gives it back to you. Barcode fields are arguably the easiest fields to use. The only thing you need to specify is which barcode to use. For example, suppose a page has three barcodes. You simply need to specify which of the three barcodes you want. FileCenter will locate the barcode and pass you back its value.

Note: If you have more than one barcode on the page, don't assume what order they fall in. They aren't always identified in the order you'd expect. Test out your barcodes carefully to be sure you've targeted the right one.

Tip: If your barcode isn't being recognized, you probably need to adjust your barcode settings. Not all barcode types are enabled by default. Go to the FileCenter *Settings* button > *Barcodes*.

Add a Regular Expression to a Barcode

As with *Zone* fields, you can add a regular expression to a barcode field. This is useful if you only need a portion of the barcode's value.

For example, suppose that the barcode contains a full file path, like this:

```
D:\Network Scans\230469.pdf
```

If you only care about the filename ("230469.pdf") you could write a regular expression to extract only the filename. Here are a few ways to do it:

```
\\(.*)$
[^\\]+$
\w+\.pdf
```

These will make more sense once you've studied Chapter 22: Regular Expressions.

Apply Character Formatting

Note that you can also use the field's *Formatting* option to control what you get back from the field. This is sometimes easier than using a regular expression. For example, suppose that a barcode contains a lengthy invoice number and you only need the last six characters. You can use a formatting option to have the field give you just the rightmost six characters. To do that, enter this in the *Formatting* option: **-R6**

The formatting options are fully explained in 0 There are a number of dynamic fields available to you. We describe each one in detail at the end of Section 2: The Automation Guide, in the *Reference* section.

Character Formatting within Fields.

Chapter 14: Content Rules

While *Content Fields* are useful for pulling text out of documents, they become powerful when you wrap them in logic. In other words, when you can use content fields to make decisions, your ability to automate reaches a new level.

This is where *Content Rules* come into play. Content rules are, simply put, a way to ask yes-no questions about the information coming out of your content fields. Does the title contain the word "contract"? Is there a social security number? Is the invoice marked "paid"?

Where content fields *gather* information, content rules *analyze* that information.

14.1 Where Do I Use Content Rules?

Content rules are the backbone of *Document Rules*, which we'll discuss in the next chapter. You'll recall that automation is built on three tools:

- *Content Fields* extract information from your documents
- *Content Rules* answer questions about that information
- *Document Rules* take those answers and put them into action

Content rules can also be paired with *Destinations* – dynamic paths that control where a document is saved and how it's named. Any time you use a destination in FileCenter, you can make the destination *conditional* on a content rule. We'll come back to this in Chapter 15: Conditional Destinations.

14.2 Add a Content Rule

First, we'll echo what we said for content fields: We recommend that you do all of your content rule development in the *Admin* tab, which we described in Chapter 12: The Automation Setup Tool.

We also recommended that you set up a special *package* called "Sandbox" for experimenting with the automation features. For instructions, see page 74, Packages.

Go to the *Admin* tab now and select your *Sandbox* package. Open a sample document to experiment with, as described on page 74, The Sample Document.

On the left side of the screen you'll see all of the different kinds of items you can add to your *Sandbox* package. Select *Content Rules* then click *Add* above the list of content rules.

Content Rule Options

Content rules have the following options:

Name. Each content rule requires a unique name. Use a name that helps you identify the rule easily. This name will be displayed everywhere that you use the content rule.

Categories. If you're dealing with a large number of content rules, it helps to break them up into categories. Enter any categories you'd like to assign to your rule, comma-separated. You can then filter your rules by category in the content rules list. The categories are for display purposes only.

Description. The description is for your purposes. Use it to supplement the name, giving further information about what this rule does.

Conditions. Now you need to add *conditions* to your rule. Each condition answers one "yes-no" question about either a content field or the overall text in the document.

Tip: Include plenty of conditions! The biggest problem with content rules is false positives: rules that are so generic that they end up matching almost every document you scan. The best way to battle false positives is to include enough conditions to narrow in on the unique criteria of each document.

14.3 Conditions Explained

A content rule is nothing more than a series of *Conditions* or "yes-no" questions about the document text: *Does the document have a date? Is there an invoice number? Is the doctor's name "Jones"?*

To add a condition, click *Add* above the list of conditions. You have the following options:

Content Field. Look at the value of a specific content field. For example, you might set up a content field that looks for an invoice number in a precise region. You can then check to see if an invoice number was found there.

Page Content. In lieu of a content field, you can examine the entire page body to see if it has (or doesn't have) a certain value.

Important: Page content tests are case-*insensitive* and also match *partial* words. For example, a test for the word "Will" also matches "<u>will</u>ing" and "un<u>will</u>ing".

And/Or. If you set up more than one condition, you need to specify the logical relationship between the conditions. An *AND* condition means that it *AND* the condition before it must *both* be *TRUE*. An *OR* condition creates an either/or scenario. We'll discuss this more shortly.

Operation & Value. Here's where you test whether the text you're analyzing *contains* or *equals* a specific value (or does *not* contain/equal). For example, you'd probably want to test if a page's entire content *contains* a value rather than *equals* a value. On the other hand, you may want to see if a zone *equals* a certain value.

Tip: Use content fields to get more precise matches. When you create a condition, you can base it on *Page Content* or a *Content Field*. Remember that looking for a value in *Page Content* is very broad: the search is *not* case sensitive and partial word matches are allowed. A *Content Field*, on the other hand, is much more targeted. It can examine text from a single, specific region (*Zone*), or it can do a very specific pattern match (*Regular Expression*).

Check if a Content Field Was Found

Sometimes all you need to know is whether a content field resulted in a match. For example, suppose you want to know if the document contains one of the firm attorneys' names. You have this regular expression:

```
(Susan Rogers|Bill Goodwin|Kris Blake|Sandra Yo|Theron Wills)
```

You don't care *which* attorney is mentioned in the document, only that *one of the attorneys* is. You can handle this two different ways in your content rule:

Option #1: Use the Return Expression. In the underlying content field, set the *Return Expression* to the text "TRUE" (without quotes); then in the content rule, check whether the field *Equals* "TRUE" (without quotes)

Option #2: Is Not Empty. Alternatively, set the content rule to check whether the field *Does Not Equal* then leave the value empty; if any value comes back, it will be *true*. In other words, you're checking if the field isn't blank.

Get the Logic Right

If you have multiple conditions in your rule, the logical relationship between them becomes very important, particularly if you have a mix of *AND*/*OR* conditions. Keep these two rules in mind:

- Conditions are analyzed top-down
- *OR* conditions will be grouped together if possible and then compared to *AND* conditions

Things are fairly straightforward if all of your conditions are *AND* conditions: the rule will come out as *TRUE* if **A and B and C** are all *TRUE*. But what happens when you throw a series of *OR* conditions into the mix? This is where the order becomes critical.

Let's illustrate this with a very simple example. We have a document with this text:

```
My favorite desserts are apple pie, ice cream, chocolate cake,
and pumpkin cookies.
```

We have a content rule that tries to identify documents that discuss desserts. It has the following conditions:

```
    [PageContent] contains "cake"
OR  [PageContent] contains "eclair"
OR  [PageContent] contains "ice cream"
OR  [PageContent] contains "cookie"
OR  [PageContent] contains "pie"
AND [PageContent] contains "dessert"
```

FileCenter will *group* the *OR* conditions together if at all possible, then compare them against the *AND* conditions. So the logic behind the rule will be analyzed like this:

```
(cake OR eclair OR ice cream OR cookie OR pie) AND dessert
```

Which, of course, resolves to *TRUE*.

Use the Results Window

While you're creating your content rules in the *Admin* tab, always have the *Automation Setup Tool* activated with a sample document loaded. When you do this, the *Results* window will show you the result of whatever content rule you have selected.

The first line in the *Results* window will show you whether the overall rule came out as *TRUE* or *FALSE*.

Below that, you'll see each individual condition in the rule, the value being analyzed, and whether that condition came out as *TRUE* or *FALSE*.

When your content rules aren't producing the outcome you'd expect, the *Results* window can make it easy to track down exactly where things are going wrong.

Chapter 15: Conditional Destinations

We talked about *Destinations* in Chapter 10: About Destinations. To summarize, a destination is an instruction that tells FileCenter *where* to save a document, *how* to name it, and *what* format to use.

We've already seen one place where destinations come into play: *Separators*. We also saw that a separator can have multiple destinations. Normally if you assign more than one destination to a separator, the separator will save the document to <u>all</u> of the destinations. What if you could let FileCenter decide which destination to use based on a content rule?

15.1 Example: Office Mail

Suppose a law practice scans all incoming mail and sorts each letter into the appropriate attorney's FileCenter DMS cabinet. They used to have a distinct separator for each attorney to route the mail into the right cabinet. The problem was that they had a pile of different separators to juggle, not to mention that they had to be careful to put the right separator on every piece of mail.

Thanks to content rules, they now accomplish the same job with just *one* separator.

First, the office set up a series of content rules aimed at identifying the various attorneys. The first rule is called "For Bill". It uses a combination of zones and regular expressions to determine if Bill Goodwin's name appears anywhere in the top 25% of the document. The office set up similar rules for Susan Rogers, Kris Blake, Sandra Yo, and Theron Wills.

Next, the office created a single separator called "Office Mail" and gave it multiple destinations – one for each attorney in the office.

Finally, the important step: for each destination, they clicked the *Add Rule* option and assigned that attorney's content rule to the destination – the "For Bill" rule to Bill's destination, the "For Susan" rule to Susan's destination, etc.

When they scan letters now, they use the same separator for all of them. FileCenter recognizes the separator then starts looking at its destinations one at a time, checking the rule assigned to each destination. If the rule comes back *TRUE*, FileCenter routes the document to that destination; *FALSE* destinations get ignored.

15.2 Assign a Rule to a Destination

The following instructions apply *anywhere* that you use a destination. For this example, we'll use a separator:

1. Add or edit a separator
2. Under *Destinations*, click *Add*
3. Give your destination any options you'd like
4. Click *OK* to add it
5. Back on the *Destinations* list, select the destination
6. Click *Add Rule* above the list of destinations
7. Choose the content rule you want to assign to the destination
8. Repeat for any other destinations

Now whenever you use this separator, FileCenter will look at the rules assigned to each destination. FileCenter will only save the document to any given destination if its rule comes back as *TRUE*.

Note that you don't have to assign a rule to every destination. If a destination doesn't have a rule, FileCenter will *always* route to that destination.

Chapter 16: Document Rules

While *Content Rules* make decisions based on document content, *Document Rules* take that logic and put it to action. A document rule pairs the logic of content rules with the dynamic paths of destinations. When applied to documents, document rules are able to examine a document, make decisions about what kind of document it is, and save/route the document accordingly. Document rules bring a powerful level of automation to your document flow.

16.1 Add a Document Rule

In the FileCenter *Admin* tab, select *Document Rules* on the left, then *Add* above the list of document rules.

Document rules have the following options:

Name. Each document rule requires a unique name. Use a name that helps you identify the rule easily. This name will be displayed everywhere that you use the document rule.

Description. The description is for your purposes. Use it to supplement the name, giving further information about what this rule does.

Content Rule. Assign a content rule to trigger the document rule. If the content rule is TRUE, the document rule will be executed. To use an existing rule, click the Library link.

Destinations. Add a destination to control where the document is saved and how it's named. These act just like separator destinations. For more information, see Chapter 10: About Destinations.

Conditionally Route among Multiple Destinations

Like we saw in the last chapter, you can have as many destinations as you'd like. Additionally, you have the option to make each destination conditional. In other words, you can use the logic of content rules to determine whether or not a given destination should be used.

This gives document rules two layers of logic:

Main Content Rule. The main content rule assigned to the document rule triggers whether or not the document rule gets executed. If it comes back as *TRUE*, FileCenter will proceed to examine the destinations. If it comes back as *FALSE*, FileCenter will skip the document rule and move on.

Destination Rules. Each destination can also have its own rule. FileCenter will only route the document to any given destination if its rule comes back *TRUE*.

16.2 Conversion Jobs

Document rules are available in both FileCenter DMS and FileCenter Automate. They come into play when you run a *Conversion Job*.

Jobs in FileCenter Automate

Conversion jobs in FileCenter Automate don't require any explanation. They are the heart of program. The job tells Automate *where* to look for files to convert, *what* files to look for, *how* to convert them, and *where* to save them.

The last point is where document rules come into play. When you tell FileCenter Automate *where* to save the converted documents, you can use document rules in place of an output folder.

Here's how that works in practice:

1. For *each document* that gets converted …
2. FileCenter Automate compares the document against each document rule
3. The first rule that comes back as *TRUE* wins
4. FileCenter Automate will then route the document to the destination(s) in that rule (which may be conditional themselves)

Document rules, then, provide a high level of *intelligent, dynamic* routing to your documents.

Jobs in FileCenter DMS

FileCenter DMS jobs have exactly the same options as jobs in FileCenter Automate, except for the *Input* options. In FileCenter DMS, you always run a job against *the file you have selected* or a *new scan*.

To run a job in FileCenter DMS:

1. If you want to convert/route an existing file select it
2. Click *Jobs* on the main toolbar
3. Select whether you want to process the *Selected File* or perform a *New Scan*

4. Select the conversion job *Profile* you want to use
5. Click *Start Job*

FileCenter DMS will then either perform the scan or take the existing document and treat it as if it were getting converted and routed by FileCenter Automate, including optionally routing it through document rules.

Set Up a FileCenter DMS Job

You can set up your conversion jobs in the FileCenter *Admin* tab. Select *Job Profiles* on the left then *Add* above the list of job profiles. For a description of the job options, see the FileCenter Automate User Guide.

16.3 Example: Office Mail

Let's go back to the office mail example from the last chapter. Recall that the office used *Content Rules* to simplify things to the point that they only needed a single separator for all of their mail routing. With *Document Rules* they can eliminate the separator altogether.

The office sets up one additional content rule called "Is Mail". This content rule looks for features typically found on a piece of mail: *Does it have a date at the top? Does it have a return address, does it have a salutation? Etc.*

Next they set up a document rule called "Mail Routing". They assign the new content rule, "Is Mail", to trigger this document rule. Finally, they add in all of the destinations just like they had them in the separator – one destination for each attorney along with a content rule to trigger that destination.

The document rule works like this:

1. It examines the document to see if it looks like a piece of mail
2. If it does, it then looks through the rule for each destination to see where the document should be saved

To put their new document rule into use, they set up a conversion job that uses document rules for its output.

They can now either:

- Scan into a folder that the FileCenter Automate job watches, OR …

- Start a new scan from the FileCenter DMS *Jobs* button and select the job

FileCenter will now recognize the scan as a piece of mail, figure out which attorney it's for, and route it accordingly.

16.4 Test and Debug Your Document Rules

As with *Content Fields* and *Content Rules*, the *Automation Setup Tool* on the FileCenter *Admin* tab is your friend. It will help you test and debug your *Document* rules.

Open a sample document, then select the document rule that you want to test.

The *Results* window will tell you:

- Whether the document rule as a whole would be triggered (*TRUE*)
- Whether the associated content rule came back as *TRUE*
- Whether each destination would be triggered
- Where the document would be saved

Note: FileCenter doesn't actually save and route the document. It only tells you where it *would* be saved and how it *would* be named.

Chapter 17: Data Mining with Content Maps

By now you've seen the power of *Content Fields* for extracting text from the document body. Despite this power, we've only used content fields in two ways: to name/route documents, and to make decisions about the page.

What if your other programs could benefit from content fields too?

For example, most businesses use some type of database, usually to manage customer, client, patient, or case information. Databases require data, and data, by and large, is entered by hand. What if content fields to provide that data instead?

FileCenter's final automation feature, *Content Maps*, can do just that: mine data from your documents and feed it to an outside program, like a database.

17.1 Content Maps Explained

Most databases and outside programs can read a *CSV* or "Comma-Separated Values" file. It's a standard format for sharing data between programs.

A content map is simply a pre-defined list of content fields that you would like to export to a CSV file.

As you scan or convert documents in FileCenter, FileCenter will simultaneously query each of the content fields in the map and export the results to a CSV file. Your other programs can then grab that information and import it.

How Is the CSV File Formatted?

The first line of the CSV file will contain the field names, which are the names of the content fields. The next line will contain the values, if found, in standard CSV format.

17.2 Add a Content Map

The best place to create content maps is on the FileCenter *Admin* tab. Select *Content Maps* on the left then click *Add* above the list of maps. As with the other automation features, you can test content maps in real-time against actual documents.

Content maps have these options:

Name. Each content map requires a unique name. Use a name that helps you identify the map easily. This name will be displayed everywhere that you use the content map.

Description. The optional description is for your purposes. Use it to supplement the name, giving further information about what this map does.

Categories. If you're dealing with a large number of content maps, it helps to break them up into categories. Enter any categories you'd like to assign to your map, comma-separated. You can then filter your maps by category in the content maps list. The categories are for display purposes only.

Content Fields. Next add content fields to your map. You can re-arrange the order of the fields using the *Up/Dn* links. The order of the fields in the list is the order that will be used for the final CSV file.

Folder. This is the folder where the CSV file will be saved.

17.3 Where Can I Use Content Maps?

Content maps are available in FileCenter DMS when you scan, run OCR on an existing document, use separators, or run a document job. In short, anywhere that a document can be OCRed and processed. You'll see a *Map* tab in each of these contexts.

In FileCenter Automate, maps are available as a job option, in document rules, and in separators.

Each individual file that gets produced will produce its own CSV file. For example, if you scan a stack that gets separated into 12 individual files, it will produce 12 individual CSV files – one per saved document.

Chapter 18: Barcodes Revisited

Because many offices are already using barcodes, it's worth a minute to briefly recap all of the ways that you can use barcodes in your automation strategy.

18.1 Use Barcode Text in Filenames and Folder Names

You can use the barcode values in your filenames and folder names. To do this, set up a *Content Field* to grab the barcode value. You can then use that field any time that you generate a filename or a folder name, like in FileCenter DMS *Naming Options*, separator *Destinations*, and destinations in other contexts like *Conversion Jobs* and *Document Rules*.

18.2 Use Barcodes to Separate & Route Documents

Suppose that every document you scan has a barcode on the first page. FileCenter can watch the incoming scan for barcodes and automatically start a new file whenever it finds a barcode. This is available when scanning or running OCR in FileCenter or as a job option in FileCenter Automate.

Associate a Barcode with a Separator

A more powerful option is associating a barcode with a separator. This gives you the convenience of using a barcode on your documents, but all of the naming and routing power you can get from a separator. In a nutshell, if FileCenter encounters a barcode, it will look it up in the separators database then act on it as if it were a full-fledged FileCenter separator.

Use Barcodes to Route Documents

Besides using barcodes to separate documents, if a barcode contains a valid path and filename, FileCenter can save the document to that path. In this sense, barcodes can completely replace separators.

18.3 Trigger Rules with Barcodes

The most powerful option is using *Content Rules* and *Document Rules* with barcodes to automatically detect what type of document it is and how it should be named and routed. Document rules work with Conversion Jobs in both FileCenter DMS and FileCenter Automate.

Chapter 19: Packages

Back in Chapter 12: The Automation Setup Tool, we sent you to the FileCenter *Admin* tab to do all of your automation development. If you recall, we advised you to create a new *Package* there called *Sandbox*. We also briefly explained a package as a bundle of FileCenter settings and options. In this chapter, we'll take a closer look at packages and how you can use them to distribute not just your automation development but *any kind* of FileCenter setup.

19.1 Package Basics

As we said, think of a *Package* as a "bundle" of setting and features. This bundle can contain any sort of configuration items: cabinet layouts, naming options, folder templates, separators, scanning profiles, favorites, content fields, content rules, etc.

The magic of packages comes from two aspects of them. First, you can have any number packages in active use at the same time, and second, packages are completely portable. This means that you can, for example, easily share a package on the network or distribute a package to third-party users. And when other users activate the package, the package items seamlessly flow into their running configuration.

19.2 What Does a Package Do?

When a package is installed and activated, the items in the package become available in FileCenter. For example, suppose that a package contains some pre-defined folder templates developed for tax professionals. When a user installs the package, all of those folder templates will now show up with the user's other folder templates.

It's important to note that a package does not alter a user's active configuration. Rather, the package's features get displayed *side-by-side* with the user's other features and settings.

To get a sense for how this works, go to the *Admin* tab in FileCenter. You'll see a row of package tabs there. The first tab is called, simply, *FileCenter*. This package represents your local settings. If you click through the items on the left, you'll see everything from your local configuration: cabinets, folder templates, naming options, conversion jobs, etc.

When FileCenter launches, it first loads everything from the *FileCenter* package. Next it loads any other packages that you have activated. The *combined contents* of these packages are what

you see in FileCenter. They display together seamlessly without actually affecting each other behind the scenes.

There is no limit to the number of packages a user can have installed.

What about Conflicts?

It's possible that a package may have an item that has the same name as an item in your local configuration or in another package. Suppose, for example, that you have a local cabinet called *Clients*. You then connect to a shared package that also defines a *Clients* cabinet.

When conflicts like this happen, the local item gets *hidden*. If you want to see the hidden cabinet, go to *Cabinets* in FileCenter DMS and select the *Hidden* option. This will force hidden items to show. You will now see both cabinets listed. You can unhide any item by selecting it then clicking *Unhide*. This also applies to other items like naming options and folder templates.

19.3 Install a Package

End users will install packages from the FileCenter *Settings* button > *Packages*.

Packages are self-contained files with an *FCDB* extension, like this: *filecenter.fcdb*

You have two options for installing a package: *opening* or *importing*

Open a Package

When you *Open* a package, the package stays where it's at. FileCenter will always access it where it currently resides. Use this option, for example, with a shared package on the network.

First you should place the package in a folder where it won't be disturbed. On the network, you may want to create a network folder dedicated to the purpose, like *FileCenter Packages*. On your local machine, you may want to make a similar folder on your C: drive.

Next follow these steps:

1. Go to the FileCenter *Settings* button > *Packages*
2. Above the *Packages* list, click *Open*
3. Click *Browse for File*
4. Browse out to the package and click *Open* then *OK*

5. The package will be added to the list
6. Make sure it has a checkmark next to it so that it will be activated

Import a Package

When you *Import* a package, the package gets *copied* to your settings folder alongside the rest of your local FileCenter settings. This give you your own, permanent copy of the package. You may want to use this option, for example, when installing a third-party package.

To import a package, follow these steps:

1. Go to the FileCenter *Settings* button > *Packages*
2. Above the *Packages* list, click *Import*
3. Click *Browse for File*
4. Browse out to the package and click *Open* then *OK*
5. The package will be added to the list
6. Notice that the package's *Path* is in your local FileCenter settings
7. Make sure the package has a checkmark next to it so that it will be activated

Enable/Disable a Package

You can temporarily enable or disable a package. In FileCenter's *Settings > Packages*, simply check or uncheck the box next to the package. Only packages with checkmarks will be loaded.

19.4 Add/Edit a Package

You can create your own packages and edit any package that isn't marked *Read Only*.

When you go to the *Admin* tab in FileCenter, you'll see tabs for all of the packages that you currently have installed. Any changes you make to any of these packages will be applied immediately.

> *Important:* Be careful when editing shared packages. The changes will immediately affect every user accessing that package.

Create a New Package

To create a package, go to the FileCenter *Admin* tab. Click the *New* button on the main toolbar then supply these options:

Name. Give the package a unique name. This is both the filename and the display name for the package. Use only filename-safe characters.

Folder Location. Select the folder where you'd like the package to be stored.

Read Only & Administrators. If you only want certain users to be able to edit the package, select *Read Only* and provide the users' Windows username in the Administrators field, comma-separated.

> *Important:* If you mark a package Read Only, you won't be able to edit the package unless you set yourself as an Administrator.

Click OK to create the package.

You should now see a tab for the package. With this package selected, you can now build the package by adding things like cabinets, folder templates, naming options, content fields and rules, etc. Select categories on the left then add items. Items are saved to the package as they're created.

Open a Package for Editing

You can edit any package that you have the rights to. To open a package for editing, click the Open button on the main toolbar and browse out to the package. The package will open in its own tab.

Note that this will also activate the package in FileCenter. This allows you to see the effects of your changes immediately.

A Note on Naming

As you add items to a package, be careful how you name them to avoid naming conflicts with other packages or the user's local configuration. For example, if a user has a local cabinet called *Invoices* and a package also supplies a cabinet named *Invoices*, this will hide the user's cabinet.

Copy Items Between Packages

You can copy items between packages or even clone entire packages. Do the following:

1. Go to the *Admin* tab
2. Make sure that both packages are currently open in their own tabs
3. Go to the tab for the package that contains the items you want to copy
4. Select the item you want to copy
5. Click the *Copy To* button on the main toolbar
6. On the drop-list, select the package that should receive the copy
7. Note that any dependent items will also be added automatically
8. Click *OK* to create the copy

To copy *everything* from the source package, click *Copy* on the main toolbar, then select *Copy Entire Package* and proceed as above.

> *Tip:* You can also use this technique to copy items from your local configuration to a shared package.

Use Custom Location for Separator Database

Normally, separators will only be recognized on the machine where they were created because the separator database is kept with the rest of FileCenter's local settings. However, in an office environment you may want to be able to share separators between users. For this to work, the users need to share a common separator data file so that separators created by one user are available to the other users.

To share separators, select the Use Custom Location for Separator Database option and select a folder on the network where the shared separator database should be stored. This location must be visible to every user that will share separators.

NOTE: You'll need to set this option for every user that will share separators.

19.5 Sharing Separators

Normally, separators are only valid when you scan on the machine where they were created because the separator database is part of the FileCenter DMS local settings. However, in an office environment you may want to be able to share separators between users. For this to work, the users need to share a common separator data file so that separators created by one user are available to the other users.

To share separators, go to the machine where the separators are set up and do the following:

1. Click the *Settings* button on the main toolbar
2. Select the *Separators* section on the left
3. Select *Use Custom Location for Separator Database*
4. Select a network folder that all users will be able to access
5. Click *OK*

Now repeat these steps for the other users.

Section 3: Reference

Chapter 20: Naming Fields

Anywhere you use a *Naming Option* or a *Destination* you can employ dynamic fields. This chapter outlines all of the available fields. Note, however, that not every field is available in every context. For example, the *Original Filename* field is only available if there *is* an original filename. If, for example, you're saving a new file in Microsoft Word, you can't use the *Original Filename* field.

Important: When you're using naming options in FileCenter DMS, FileCenter will only show you the naming options that are *valid* in your current context. All others will be hidden.

Note: Some fields are only available in specific editions of FileCenter DMS, as noted below. All fields are available in FileCenter Automate.

20.1 Available Fields

Here are the dynamic fields available to you:

General

ORIGINAL Name. When you are renaming a file or saving a file under a new name, you can use the original filename as part of a naming option. Select the *General* option > *ORIGINAL Name*.

For example, if a file is named *c:\files\resume.doc*, the original filename is *resume*.

If you pair the original filename with an *Increment*, you can make it easy to save new versions of a file:

```
resume 01
resume 02
resume 03
... etc.
```

USER Name or COMPUTER Name. Sometimes a good way to identify a file is by the user who created it or the machine where it was created. Under the *General* section, you'll see two options: *USER Name* and *COMPUTER Name*. These fields will insert either your Windows username or the name of your computer on the network.

Folder Name

Folder Name. When you're working in FileCenter DMS, you'll usually be saving files into cabinets, which means the file will go into a drawer and possibly folders. You can use the drawer and folder names as part of a naming option. For example:

```
e:\cabinet\Johnson\Trusts
```

This path has a drawer name (*Johnson*) and a folder name (*Trusts*). You can use those names as part of the filename. Why would you want to? Many businesses use client or patient names for drawer names, or maybe a case name or case number. If that's true for you, you'll be able to pull the client name, case number, etc. into your filenames.

For example, if you use one drawer for each client, the drawer name will have the client's name. You could then use a naming option such as this one to include that name in the filename:

```
«FOLDER Drawer» - Rollover Trust.pdf
```

If the drawer were named "Jones, Linda" the resulting filename would be:

```
Jones, Linda – Rollover Trust.pdf
```

If, on the other hand, the first folder in a drawer contains the client name, you may be more interested in using that folder as a field.

The *Drawer Name* field option will always refer to the drawer, if there is one. Folder names are supported up to nine levels deep. You can reference them either down from the drawer, or up from the file:

```
Drawer\Level 1\Level 2\Level 3\File.doc
Drawer\Parent 3\Parent 2\Parent 1\File.doc
```

If you save to a location outside of a cabinet, or you don't use drawers, you'll only be able to use the *Parent* locations:

```
e:\...\Parent 3\Parent 2\Parent 1\File.doc
```

Important: FileCenter Automate can't use the *Drawer* or *Folder* fields because there is no cabinet context in FileCenter Automate. Instead, you'll need to use the *Parent* fields.

Note: The *Level* and *Parent* fields can go up to nine levels deep. To use the deeper levels, insert any *Level* or *Parent* field and manually change the level number.

Dates

You have a few date fields available to you:

Date Today. The current date/time.

Date Created. If you're processing an existing file, the *Date Created* field will give you the Windows timestamp for when the file was created.

Date Modified. Like the *Date Created* field, except it gives you the date/time the file was last modified.

For all three fields, you have a list of pre-defined date formats to pick from. Note, however, that you're not limited to those choices. If you want to come up with your own format, insert any date format from the list then modify it however you'd like with these formatting identifiers:

- **d** – Day of month without a leading 0
- **dd** – Day of month with a leading 0
- **ddd** – Abbreviated day of the week (*Sun, Mon, etc.*)
- **dddd** – Full day of the week (*Sunday, Monday, etc.*)
- **m** – Month as a number without a leading 0
- **mm** – Month as a number with a leading 0
- **mmm** – Abbreviated month name
- **mmmm** – Full month name
- **yy** – Two-digit year
- **yyyy** – Four-digit year
- **h** – Hour without a leading 0
- **hh** – Hour with a leading 0
- **n** – Minutes without a leading 0
- **nn** – Minutes with a leading 0
- **s** – Seconds without a leading 0
- **ss** – Seconds with a leading 0

You can use spaces, commas, and dashes between the date elements and arrange the date elements however you want. You can also omit any of the date elements. For instance, you may want to use only the current month or the current year.

Important: You cannot use a forward slash "/" or a colon ":" in your date format, since Windows won't allow those in filenames.

Tip: You can add your favorite date formats to the list. Go to *Settings > Date Formats*.

Auto Increment

Think of an *increment* as a counter. You'll want to use increments when you'll end up with a lot of files with the same name in the same folder.

The auto increment will automatically add a number to the end of the filename. For example, if a folder will contain many files called *Report.pdf*, the auto-increment will name them *Report 01.pdf*, *Report 02.pdf*, *Report 03.pdf*, etc.

You have a few increment formats to choose from. Each one pads the number with more 0s.

> *Note:* When you insert the increment field, it must always go at the <u>end</u> of the expression. An increment cannot be customized or edited.

Content

The *Content* field pulls text out of the document to use in your filename. You can pull values from barcodes, text from specific regions on a page, or text that matches certain patterns or criteria. For example, it can grab names, dates, invoice numbers, etc. to create truly dynamic filenames. The content field is sophisticated enough to deserve its own chapter. To learn more, see Chapter 13: Content Fields.

> *Pro Plus Only:* This feature is only available in FileCenter DMS Pro Plus. It is available in all editions of FileCenter Automate.

Variables

A *Variable* field gets its value from you at the time you scan or run OCR in FileCenter DMS, or from the conversion job in FileCenter Automate. Think of a variable as a named placeholder that gets populated later.

You insert the placeholders in your naming options. Later when, for example, you scan a document, you'll provide the value you want FileCenter to use instead of the placeholder.

Set Up Variables. To create a variable, first put it in your naming expression through the *Variable* field. You'll be prompted to provide a name for the variable. The name is *not* case-sensitive.

Use as many variable fields as you'd like in your naming options. You can re-use the same variable name in different naming options.

Use Variables. On the *Scan* and *OCR* dialogs in FileCenter DMS and the job settings in FileCenter Automate, you'll see a *Variables* tab. Here you enter the names of the variables you used in your naming options and give them the values that you want to them have. Each assignment needs a *name* and a *value* in this format:

```
variable_name=variable_value
```

For example, suppose you have a naming option that uses a variable called *DocumentType*. To set it to "Letter" you would enter the following in the *Variable Assignments* box:

```
DocumentType=Letter
```

Now proceed with the scan, OCR, or conversion job. Wherever the *DocumentType* variable appears in your naming options, it will now show "Letter" instead of the placeholder for the current file.

Note that you can set more than one variable. Simply enter one per line:

```
Department=Estate Planning
DocumentType=Letter
User=Bob Johnson
Attorney=Alison Black
```

The name is *not* case sensitive. It can contain any characters you want except a space or an "=" sign.

Important: Variable assignments are temporary. They only last for the current document.

Tip: You can include variable assignments in your scanning and OCR *Profiles*. This can make variable assignments almost automatic.

Pro Plus Only: This feature is only available in FileCenter DMS Pro Plus. It is available in all editions of FileCenter Automate.

Chapter 21: Character Formatting within Fields

For advanced users, it's possible to apply some character formatting to an individual naming field, like altering the capitalization or limiting the field to a certain number of characters. This applies equally to content fields.

The following formatting flags are available:

- **-DT** *format* – Converts the value into the date format specified
- **-L***num* – Returns a specific number of characters from the *left* side of the text
- **-LC** – Puts all of the text in lower case
- **-LN** – Converts certain letters to numbers, like "S" to "5" and "O" to "0"
- **-NS** – Removes all spaces from the text ("no spaces")
- **-R***num* – Returns a specific number of characters from the *right* side of the text
- **-RC{***characters***}** – Removes individual characters from the text (case sensitive)
- **-RT{***text***}** – Removes a string of text from the text (case sensitive)
- **-SR{***find***,***replace***}** – Perform a search and replace within the text (case insensitive)
- **-TC** – Capitalizes the first letter of each word ("Title Case")
- **-UC** – Puts all of the text in UPPER CASE
- **-X** *expression* – Pulls out a portion of the field's text based on a regular expression

Note: For naming options, these formatting codes only apply to the *specific field* that they're placed in. They don't affect the rest of the naming option.

21.1 Additional Tips

A few of the fields deserve more discussion.

Date/Time

The *-DT* flag converts a date into a different format. Note that with *date* fields, this isn't necessary because they have their own formatting option. Other field – like *Content* fields, for example – may give you dates that need to be reformatted. The formatting codes are explained on p. 112, Dates.

Left/Right Characters

The *-Ln* and *-Rn* flags are used to trim down the text to a fixed number of characters. When would you need this? Suppose that your drawer names also contain a client ID number at the end, like this: *Johnson, Robert – 65298*. If you want to use this ID number in your filename, you'd include the flag "-R5" which would give you the last five characters of the drawer name, or *65298*:

```
«FOLDER Drawer [-R5]»
```

Remove Characters/Text

You can strip out individual characters or strings of text from the field. To remove individual characters, use the *-RC{ }* flag and list out all of the characters you want removed. For example, to strip out all punctuation:

```
«FOLDER Drawer [-RC{,.:;?!}]»
```

Text removal with the *-RT{ }* flag works the same way. The only difference is that you're providing an exact string of text to remove. For example, to remove the text "Invoice":

```
«CONTENT Zone 1 [-RT{Invoice}]»
```

Search and Replace

The *-SR{ }* flag allows you to do a search and replace within the text. For example, to replace all underscores "_" with spaces:

```
«COMPUTER Name [-SR{_, }]»
```

To do multiple search-replace operations, chain them together like this:

```
-SR{search_a,replace_a,search_b,replace_b}
```

For example:

```
«CONTENT Zone 1 [-SR{v.,versus,et al.,}]»
```

This replaces "v." with "versus" and removes the phrase "et al." (replaces it with nothing). There's no limit to the number of search-replace operations you can perform. For example, to convert numbers to letters (to compensate for OCR mistakes), you might do this:

```
«CONTENT Zone 1 [-SR{0,O,1,I,2,Z,5,S,6,G,7,T,8,B}]»
```

Regular Expressions

Special rules apply to the regular expression option:

If you use more than one formatting flag, the regular expression must be the *last* one. For example:

```
«COMPUTER Name [-TC -X ^\w+]»
```

If you want to capture text from a set of parentheses in the expression, indicate the set number after the *–X*. For example, to capture the text between two commas in the drawer name, use the following:

```
«FOLDER Drawer [-X1 , ([a-zA-Z\.\- ]+),]»
```

The *-X1* means we capture only the text between the first set of parentheses.

We cover regular expressions in depth in Chapter 22: Regular Expressions.

> *Important:* Some programs require regular expressions to begin and end with forward slashes */like_this/*. Don't do this in FileCenter.

21.2 How to Use Formatting Flags

Naming Options

To use any of these formatting options, put the flag in square brackets "[ ]" at the end of the field, before the closing chevron. A few examples:

```
«FOLDER Drawer [-LC]»

«COMPUTER Name [-TC -L5]»
```

If the original filename is a date stamp, you can re-format it:

```
«ORIGINAL Name [-DT yyyy-mm-dd]»
```

Note that you can include your own comments in the square brackets. This is a nice way to identify obscure fields:

```
«FOLDER Drawer [-LC The client name in lower case]»

«CONTENT Zone 1 [INTAKE FORM DATE]»
```

```
«CONTENT Zone 1 [-L25 ~ INTAKE FORM DATE]»

«COMPUTER Name [-TC -L5 (first five chars)]»
```

> **Tip:** You can set off the comment from the formatting codes any way you want. Just make sure you don't use any characters that are invalid in a filename and that you don't use any characters with special significance in naming options (like apostrophes, single quotes, double quotes, chevrons, commas, dashes, and square brackets).

Content Fields

Every content field has a *Formatting* option that accepts formatting flags. Place as many formatting flags as you'd like in this field. Unlike naming options, you don't need to use square brackets:

```
-LC

-TC -L5

-RT{incoming} -RT{outgoing} -SR{correspondence,letter} -TC
```

If the content field captures a date, you can reformat it:

```
-DT yyyy-mm-dd
```

Note that the formatting will be applied to the *entire result* of the content field.

Chapter 22: Regular Expressions

22.1 Regular Expression Syntax

A regular expression is a special pattern used to *match* then *extract* text. Maybe the best way to understand is with an example. Let's suppose you want to capture a date in the document's body. You know the date is formatted like this: *YYYY/MM/DD*.

To capture the date, you create a regular expression *pattern:* four digits+ a slash + two digits + a slash + two digits. FileCenter looks at the page text to see if anything matches this pattern. If it finds a match, it will capture that text and give it to you.

With regular expressions, your goal is to pattern the text that you want to extract. You create this pattern through a very specific syntax:

any text	That exact text must be present
[a-zA-Z]	Matches any letter of the alphabet
[0-9]	Matches any number
[^*characters***]**	Must not match any of the characters listed in the brackets
\d	Any digit; identical to [0-9]
\D	Any non-digit; identical to [^0-9]
\w	Any "word" character: a-z, 0-9, or underscore "_"; identical to [a-zA-Z0-9_]
\W	Any non-word character; the inverse of \w
\s	Any "whitespace" character: space, tab, return, etc.
\S	Any non-whitespace character; the inverse of \s
.	A period matches any single character (a wildcard)
*****	Matches zero or more of the preceding character
+	Matches one or more of the preceding character
?	The preceding character is optional
!	Must not include the preceding character in that position
{*n***}**	Matches a specific number of the preceding character
\b	Requires a match to be at a "word boundary"
^*expression*	Requires the match to occur at the beginning of the line of text
*expression***$**	Requires the match to occur at the end of the line of text
\A*expression*	Requires the match to occur at the very beginning of the page/section text
*expression***\Z**	Requires the match to occur at the very end of the page/section text

Either-Or

You can list alternatives by separating them with a pipe "|". For example, this will look for any common household pet:

```
cat|dog|goldfish|parakeet
```

To isolate the options from the rest of the pattern, surround them in parentheses:

```
I want a (cat|dog|goldfish|parakeet) for a pet
```

You can also use the pipe with patterns. For example, to look for either a complete word or a complete sequence of digits:

```
[a-zA-Z]+|[0-9]+
```

How to Use the Special Characters Literally

If you want to use one of the special characters literally – as an actual character in your pattern – put a backslash "\" in front of it. For example, if you want to use a period, plus symbol, or exclamation mark as part of the pattern, do this:

```
\.
```

```
\+
```

```
\!
```

The special characters that would require a backslash are: **\ ^ \$. | ? * + ( ) [ ] { }**

If you're using a content field, you'll see an option called Return Expression. Use this to indicate which substring(s) you want and, optionally, reformat them.

22.2 Regular Expression Examples

The best way to learn regular expressions is through examples. So here are some simple and real-world examples.

bat   Matches the *bat* in bat, battle, combat, unbathed, etc.

b.t   Matches *bat, bit, bot, b9t, b&t, b t* but not *boat* (will also match any of those in the middle of a word)

b.*t	Matches *bt, bat, boat, b39-26xt,* etc. (will also match any of those in the middle of a word)
\bword\b	Ensures that word is a complete word; it won't match sword, swordplay, or wordy, only *word*
b.+t	Matches bat, boat, b39-26xt, etc. but not bt
[a-zA-Z]+	Matches any complete word: axe, house, hippopotamus but not double-take
\w+	Matches any complete sequence of letters or numbers: bat, B2384, 229A, etc.
^[a-zA-Z]+	Matches any complete word at the beginning of a line of text
[a-zA-Z]+$	Matches any complete word at the end of a line of text
[0-9]+	Matches a sequence of numbers: 12345 but not 12,345
[0-9,]+	Matches a sequence of numbers, including commas: 12,345
[0-9,\.]+	Matches a sequence of numbers, including commas and decimals: 12,345.67
\d+	Identical to [0-9]+
[aeiou]	Will match any single vowel
[aeiou]+	Will match any sequence of vowels
[^aeiou]+	Will match any sequence of consonants

22.3 Capture Only a Portion of the Pattern

Normally FileCenter will give you back *all* of the text that matches the pattern. If, however, you only need a portion of that text, you can target it using parentheses. For example, this expression will match last name, first name:

```
[A-Z][a-z]+, [A-Z][a-z]+
```

To explain, this is looking for a capital letter followed by any number of lower-case letters, a comma and space, and another capital letter followed by lower-case letters.

Let's suppose that you only want to use the *first* name. To capture only that part of the pattern, put it in parentheses:

```
[A-Z][a-z]+, ([A-Z][a-z]+)
```

You can even isolate multiple sections of the text. For example, a date might look like this:

```
(\d{4})/(\d+)/(\d+)
```

This gives you individual access to the year, month, and day values.

The question now is, how do you target the value from a specific set of parentheses? The answer depends on whether you're using the regular expression in a *Content* field or in a naming option field. We describe both techniques below.

22.4 Pull Body Text Using *Content* Fields

Each of the examples below requires you to create a *Content* field, as described in Chapter 13: Content Fields. Use each example as the *Regular Expression* value.

\b\d\d\d-\d\d-\d\d\d\d\b

Extracts a social security number. It looks for exactly three digits + a dash + exactly two digits + a dash + exactly four digits. The *\b* are word boundary markers. In other words, there can't be any other alphanumeric characters before or after the social security number. This prevents false positives, like 00000-00-00000.

\b\d{3}-\d{2}-\d{4}\b

Identical to the last example.

\b[0-9]{3}-[0-9]{2}-[0-9]{4}\b

Identical to the last example.

Capture Text from Parentheses

As we mentioned above, it's possible to capture only a portion of the matched text using parentheses. To do this, you'll use the option called *Return Expression* on the content field. Here you can indicate which sets of parentheses you want and, optionally, reformat them.

To indicate a substring, use a backslash "\" plus the string number: *\1, \2, \3, etc*.

Let's go back to this example, where we attempt to capture a name that's in the format *Last, First:*

```
[A-Z][a-z]+, ([A-Z][a-z]+)
```

To get only the first name, put "\1" as the *Return Expression*. Going a step further, you can put parentheses around both the first and last name:

```
([A-Z][a-z]+), ([A-Z][a-z]+)
```

You can now set up a *Return Expression* that reformats "Johnson, Robert" into "Robert Johnson" like this:

```
\2 \1
```

In other words, you've just asked FileCenter to give you the text from the *second* set of parentheses followed by a space and the text from the *first* set of parentheses.

Here are a few more examples:

Dear ([a-zA-Z\-\.]+),

Pull the recipient's name from a letter that's in this format:

```
Dear John Hanks,
```

It looks for the literal text "Dear " followed by a sequence of letters and spaces leading up to a comma. It tolerates hyphens and dots in the name, so "Ann-Louise P. Barnacle" would be captured.

Important: To use this example, you need to set the content field's *Return Expression* to: \1

(January|February|March|April|May|June|July|August|September|October|November|December) (\d+), (\d+)

Matches a date in the format "January 1, 1990". You can either accept the whole date back, or take out individual portions by putting any combination of the following in the content field's *Return Expression*:

- **\1** = the month
- **\2** = the date
- **\3** = the year

So to get only the month and year, use this *Return Expression:*

```
\1, \3
```

assign ([a-zA-Z\-\.]+) as my agent

Gets the text between "assign _____ as my agent". For example, suppose your document consistently states the agent name with this language: "I hereby assign Bob Johnson as my agent." This pattern will pull out the name. Again, to use this example, you need to set the content field's *Return Expression* to "\1".

Deal with Multi-Line Blocks

A common task is to separate a multi-line block into separate lines, like an address block. This expression will separate three lines:

```
\A(.*)$\s+^(.*)$\s+^(.*)\Z
```

Explanation:

\A	From the very beginning of the block of text …
(.)$*	Grab everything up to the end of the first line
\s+	Account for the return(s)
^(.)$*	Grab everything from the beginning to the end of the next line
\s+	Account for the return(s)
^(.)\Z*	Grab everything from the beginning of the line to the end of the block

The lines will be in \1 … \3. Note that the *\s+* is necessary; it accounts for the line break characters (which aren't considered part of any line).

To capture only the first line in a block, use this:

```
\A.*$
```

To capture only the last line in a block, use this:

```
^.*\Z
```

> *Important:* To use these examples, you need to use a *Zone* content field to isolate the block, then set the content field's *Text Style* to *Lines*.

22.5 Pull Text from a Naming Option Field

In 0 There are a number of dynamic fields available to you. We describe each one in detail at the end of Section 2: The Automation Guide, in the *Reference* section.

Character Formatting within Fields, we discussed the *–X* formatting option which can employ a regular expression to pull out only a portion of the field text. Place the regular expression immediately after the *-X* flag like this:

```
«ORIGINAL Name [-X expression]»
```

The *-X* flag will give you back all of the text that matches the pattern. If, however, you want to isolate text from a set of parentheses, indicate which set of parentheses after the *-X*. For example, to capture the text from the second set of parentheses, use *-X2:*

```
«ORIGINAL Name [-X2 expression]»
```

In the following example, the first set of parentheses looks for either "will" or "trust" or "deed". The second set of parentheses grabs the document number. Because we've used *-X2*, the only text we'll get back is the document number from the second set of parentheses:

```
«ORIGINAL Name [-X2 (Will|Deed|Trust) #(\w+)]»
```

If the requested substring doesn't exist, you'll get the entire match.

Practical Example – Parse a Drawer Name

Let's suppose that our drawer names have the client's *last name, first name - id number* like this:

```
Johnson, Robert - 55236

Kirby, LeAnn - 238967

Larson, Steve - 1296
```

You could use the following expressions to pull out each piece of information:

«FOLDER Drawer [-X1 ^(.*),]»

This takes everything from the beginning of the drawer name up to the comma. In other words, it captures the last name. The important character is the "^" at the beginning, which requires the match to occur at the beginning of the drawer name. We've also used parentheses to capture only a portion of the pattern, so we also used -X1 to give us back the value of the first set of parentheses.

«FOLDER Drawer [-X \d+$]»

Captures the client ID by taking the sequence of numbers from the end of the drawer name. The important character is the "$" at the end, which requires the match to happen at the end of the drawer name.

«FOLDER Drawer [-X \w+$]»

Also extracts the client id from the end of the drawer name. This version tolerates letters and numbers in the id.

«FOLDER Drawer [-X [0-9\-]+$]»

Also extracts the client id from the end of the drawer name. This version tolerates hyphenated numbers like "555-21397".

«FOLDER Drawer [-X [^]+$]»

Also extracts the client id from the end of the drawer name. This version tolerates any kind of character except a space.

«FOLDER Drawer [-X1 , (.*) -]»

Captures everything between the comma and hyphen. In other words, it captures the first name. We only want the portion in parentheses (the first substring), so we use *-X1* where the "1" indicates we want the text from the first set of parentheses.

22.6 Model Expressions

The model expressions in this section can be used as-is, often without modification. With that said, the specifics of your documents and/or naming option fields may require that you modify these models slightly. At the very least, these models should help you get a better sense for how regular expressions work and what they're capable of doing.

If you're using regular expressions to capture text in a document, we **strongly recommend** that you create a fairly generous *Zone* field around the text you want to capture, then use a regular expression to isolate the text within that zone.

You should only use a *Regular Expression* field if the text tends to move around on the page.

SSN/EIN

Social Security Number:

```
\b\d{3}-\d{2}-\d{4}\b
```

Employer Identification Number:

```
\b\d{2}-\d{7}\b
```


Dates

January 31, 2000:

```
(Jan.*|Feb.*|Mar.*|Apr.*|May|Jun.*|Jul.*|Aug.*|Sep.*|Oct.*|Nov.*|Dec.*)\s+(\d+),\s
+(\d+)
```

This pattern tolerates either full month names ("January") or abbreviated month names ("Jan"). It doesn't matter if the numbers are padded ("01") or if the year has two or four digits. This pattern will give you the whole date back. You can also get just part of the date. In a content field, use any of these in the Return Expression:

   \1   The month
   \2   The day of the month
   \3   The year

In a naming option field, you can capture any **single** element of the date:

   -X1   The month
   -X2   The day of the month
   -X3   The year

1/31/2000, 31/1/2000, 01/31/2111, etc.

```
\b(\d+)/(\d+)/(\d+)\b
```

This is a very flexible date pattern, where the day, month, and year can be in any position. It doesn't matter if the numbers are padded ("01") or if the year has two or four digits. Like the previous example, you can capture the individual date elements with *\1…\3* or *-X1, -X2,* or *-X3*.

2000-01-31, 1-31-2000, 31-1-2000, etc.

```
\b(\d+)-(\d+)-(\d+)\b
```

Identical to the last pattern, but with dashes ("-") instead of forward slashes ("/").

31 JAN 2000

```
(\d+)\s+(JAN.*|FEB.*|MAR.*|APR.*|MAY|JUN.*|JUL.*|AUG.*|SEP.*|OCT.*|NOV.*|DEC.*)\s+
(\d+)
```

Very similar to the first example. Note that **regular expressions are case-sensitive**. If your months aren't in all-caps, you'll need to change the expression.

Plaintiff/Petitioner and Defendant/Respondent

Assuming that you're using these patterns with a *Content* field, do the following:

1. Open a sample document in the *Automation Setup* tool on the *Admin* tab
2. Click *New Zone* above the document preview
3. Draw a generous zone around the area in the pleading header where the parties are listed – wide enough and long enough to accommodate any case; it doesn't matter if some of the pleading body bleeds into the zone, but DON'T capture line numbers
4. Set your content field *Type* to *Zone*
5. Set the *Text Option* to *Lines*

Format #1

```
                    John Doe v. Jane Doe
```

Plaintiff/Petitioner

```
        ^([\w\s,-\.]+)\s+v\.
```

To get the name, set the *Return Expression* to: /1

Defendant/Respondent

```
        \bv\.\s+([\w\s,-\.]+)$
```

To get the name, set the *Return Expression* to: /1

Format #2

```
          John Doe, Plaintiff v. Jane Doe, Defendant
```

Plaintiff/Petitioner

```
        ^([\w\s,-\.]+),\s+(Plaintiff|Petitioner)
```

To get the name, set the *Return Expression* to: /1

Defendant/Respondent

```
\bv\.\s+([\w\s,-\.]+),\s+(Defendant|Respondent)$
```

To get the name, set the *Return Expression* to: /1

Format #3

```
                          John Doe
                             v.
                          Jane Doe
```

Plaintiff/Petitioner

```
        ^(.*)$\s+^v\.$
```

This pattern gives you the complete line of text before the "v." in the document. To get the name, set the *Return Expression* to: /1

Defendant/Respondent

```
        ^v\.$\s+^(.*)$
```

This pattern gives you the complete line of text after the "v." in the document. To get the name, set the *Return Expression* to: /1

Format #4

```
                          John Doe
                          Plaintiff
                             v.
                          Jane Doe
                          Defendant
```

Plaintiff/Petitioner

```
        ^(.*)$\s+^(Plaintiff|Petitioner)$
```

This pattern gives you the complete line of text before "Plaintiff" or "Petitioner" in the document. To get the name, set the *Return Expression* to: /1

Defendant/Respondent

```
^(.*)$\s+^(Defendant|Respondent)$
```

This pattern gives you the complete line of text before "Defendant" or "Respondent" in the document. To get the name, set the *Return Expression* to: /1

Address Blocks

It's possible to extract the name, street address, city, state, and ZIP from an address block. For purposes of these patterns, we're assuming a standard US address in one of the formats that follow.

Assuming that you're using these patterns with a *Content* field, do the following:

1. Open a sample document in the *Automation Setup* tool on the *Admin* tab
2. Click *New Zone* above the document preview
3. Draw a generous zone around the area in the pleading header where the parties are listed – wide enough and long enough to accommodate any case; it doesn't matter if some of the pleading body bleeds into the zone, but DON'T capture line numbers
4. Set your content field *Type* to *Zone*
5. Set the *Text Option* to *Lines*

Format #1

```
John Doe
1234 Some Street
My City, State 12345
```

Addressee

```
^.*$
```

This pattern grabs all of the text from the first line in the block.

Street Address

```
\A.*$\s+^(.*)$
```

This pattern grabs all of the text from the second line in the block. To get the address, set the *Return Expression* to: \1

City, State, ZIP

```
^([a-zA-Z \-\.]+),\s+([a-zA-Z ]+)\s+([\d\-]+)\Z
```

This pattern captures all three elements. The state can be an abbreviation or a full state name. The ZIP code can have five ("12345") or nine ("12345-6789") digits. Retrieve a specific piece of information in the *Return Expression* as follows:

\1   City
\2   State
\3   ZIP code

Format #2

```
John Doe
1234 Some Street
My City NY 12345
```

Addressee, Street Address

(same as above)

City, State, ZIP

```
^([a-zA-Z \-\.]+),?\s+([A-Z]{2})\s+([\d\-]+)\Z
```

With this pattern, the state MUST be a two-letter abbreviation in UPPERCASE. The comma after the city is optional. Retrieve specific pieces of information in the *Return Expression* as follows:

\1   City
\2   State
\3   ZIP code

Format #3

```
Bla bla bla John Doe, who resides at 1234 Some Street, My City, NY 12345.
```

In this example, the address appears in the middle of a sentence. That means we have to handle things a little bit differently:

1. Create either a *RegEx* field if this is the only address on the page, OR …
2. A *Zone* field if you need to target a specific part of the page
3. Set the *Text Option* to *Paragraphs* or *Continuous*

Party Name

```
(\s[A-Z][\w\.\-]+)+, who resides at
```

This pattern captures any number of capitalized words preceding the text "who resides at.…" To get the name, set the *Return Expression* to: \1

Street Address

```
who resides at ([^,]+),
```

This pattern captures everything between the text "who resides at …" and the comma. To get the text, set the *Return Expression* to: \1

City, State, Zip

```
\b([a-zA-Z \-\.]+),?\s+([A-Z]{2})\s+([\d\-]+)\b
```

This pattern REQUIRES that the state be a two-character, capitalized abbreviation, just like in the example. The comma after the city is optional. Retrieve specific pieces of information in the *Return Expression* as follows:

\1 City
\2 State
\3 ZIP code

Trim a 9-Digit ZIP Code to 5 Digits

Any of the patterns above can be modified to capture only the first five digits of a nine-digit ZIP code. Note that this works with both five-digit and nine-digit ZIPs.

Replace the final:

```
([\d\-]+)
```

With:

```
(\d+)(-\d+)?
```

Invoice Number

Invoice numbers can take on a wide range of different formats. We'll illustrate a couple of examples here which you should be able to adapt to your specific situation.

> *Tip:* If the invoice number never moves on the page, use a *Zone* instead. Regular expressions are the last resort when a zone can't get you what you need.

Format #1

```
                         Invoice #1234567890
```

Expression

```
        Invoice #(\d+)
```

To get the text, set the *Return Expression* to: **\1**

If using the -R character formatting flag in a field, use: **-R1**

Format #2

```
                       Invoice: AZX-1234567890
```

Expression

```
         Invoice:\s+([A-Z0-9\-]+)
```

To get the text, set the *Return Expression* to: **\1**

If using the -R character formatting flag in a field, use: **-R1**

Salutation

```
Dear Dr. Nelson,
```

Set the field's *Text Option* to *Lines*. Optionally draw a zone around the salutation, unless the text tends to move up and down on the page.

Expression

```
^Dear\s+[a-zA-Z\.]+\s([\w'\- ]+),
```

Explanation: This pattern looks for the text "Dear" followed by letters and a period – the title, which will be ignored. It then takes the rest of the text up to the comma. This tolerates multi-part last names like "Van Der Berg", names with apostrophes like "O'Toole", and hyphenated names like "Smith-Warner".

22.7 Other Resources

While regular expressions can seem intimidating, they are worth learning if you need that extra bit of power.

Regular expressions are widely used in the computer world. As a result, you have a lot of resources available to you on the web. Here are just a few:

- rexegg.com
- regular-expressions.info
- regex101.com

Some of these have excellent tutorials. Some have cheat sheets. Some have "sand boxes" where you can test regular expressions in real-time. For example, regex101.com lets you experiment with different regular expressions to see what they'll produce. It gives you a real-time explanation about what each part of your regular expression is doing.

If you need a regular expression to capture common information, like social security numbers, dates, phone numbers, etc., you can search the web and find good examples.

TIP: Many samples you'll find on the web put the regular expression in forward slashes like this: */expression/*. **Don't include those slashes** in FileCenter.