

# Scanning: Best Practices

*White Paper*

December 2005

**Laserfiche®**

*The information contained in this document represents the current view of Compulink Management Center, Inc on the issues discussed as of the date of publication. Because Compulink must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Compulink, and Compulink cannot guarantee the accuracy of any information presented after the date of publication.*

*This chapter is for informational purposes only. COMPULINK MAKES NO WARRANTIES, EXPRESS OR IMPLIED, AS TO THE INFORMATION IN THIS DOCUMENT.*

## **Table of Contents**

Scanning: Best Practices .....	i
Table of Contents .....	2
Before You Begin.....	3
Choosing your Scan Engine .....	3
Scanning Documents: Basic and Standard Mode.....	4
Enhancement Best Practices.....	4

Laserfiche Scanning is a key part of the Laserfiche document management system, allowing users to put paper documents in the electronic system easily and quickly. The Laserfiche scanning interface provides a wide variety of options for processing documents as they are scanned, to make them easier to find, easier to read, and generally more informative. A good scanning procedure is key to creating useful, well-organized repositories, and to keeping the scanning process fast and accurate.

In creating a scan procedure, you should keep several elements in mind. You will need to determine how you will capture your images, and what scanning engine you will use. Your scanning procedure should include steps for each stage of the process. Those parts of the process specific to Laserfiche Scanning involve assigning metadata to documents, processing images to improve aesthetics, readability and the OCR process, and sending the documents to Laserfiche.

### **Before You Begin**

If you are scanning a large batch of documents, try to make sure that each document is well separated from the documents around it. Slipsheets help scanner operators visually distinguish between different documents. Make sure, also, that each document has all of the necessary information in an easy-to-read and easy-to-locate spot, such as a cover sheet.

When you set up templates for scanned documents, don't include too many required fields. Since a scanner operator cannot file a document until all of the required fields have been filled in, they can slow down the scanning process considerably, particularly if required information about a document is missing or difficult to find.

### **Choosing your Scan Engine**

There are several possible options for interfacing with your particular scanner. Your organization may choose to use one or several, depending on your scanning procedure, available equipment, and preferred image processing methods.

**Kofax.** Laserfiche Scanning can communicate with any scanners that use Kofax Adrenaline cards and software to capture images. The Kofax scan engine comes with Laserfiche Scanning by default, and like ScanConnect may be used with ISIS scanners. However, no drivers are included with this engine. Use Kofax if you have Kofax Adrenaline Capture Engine (ACE) or Virtual ReScan (VRS) already installed.

**ScanConnect.** ScanConnect includes a wide variety of ISIS scanner drivers. It allows supported scanners to communicate directly with Laserfiche software.

**Photo Capture.** The Photo Capture engine allows you to capture documents with a digital camera—useful, for instance, in situations where a portable imaging solution is necessary. Once in Laserfiche Scanning, digital photo images are identical to normal images.

**Universal Capture.** Universal Capture creates a useful interface for image-capturing methods that don't use standard scanners. Universal Capture finds files on your hard drive or removable media and imports them through the scanning interface to be processed like other scanned images.

**TWAIN.** Laserfiche TWAIN is a beta scan engine compatible with some scanners that use TWAIN drivers. This engine will provide the user with the option of using the scanner's default software or the user-defined settings. For any given scanner, TWAIN is usually slower than ISIS, so if speed is a concern and ISIS is available you may prefer to use it. The ActiveX plugin for Web Access uses the TWAIN scan engine.

### Scanning Documents: Basic and Standard Mode

The first choice you make as you begin to scan is the choice between basic and standard scanning modes. **Basic mode** requires minimal configuration, providing a few basic enhancements and the option to set default document properties or fill in properties for each document. It is ideal for beginning and intermediate users with relatively simple scanning needs, or for short scanning jobs by any user. **Standard mode** provides much more flexibility, with a broad variety of available enhancements and processes giving users full control over images. Standard mode is best for more advanced users.

### Enhancement Best Practices

Scanned documents may not always be ideally formatted for electronic reading. They can be smudged or discolored, have margins that are too large or too small, or contain blank pages and irrelevant information. In order to fix these potential defects, Laserfiche Scanning allows users to perform enhancements on scanned documents. The Laserfiche Scanning help file describes the function of each image enhancement and processing operation. However, the order in which these operations are applied can have a significant effect on the end results, as well as on the time it takes to process the scanned images.

The following list describes the order in which image enhancements and processes should be performed, **if** you choose to perform them. It is not necessary, and in most situations not desirable, to perform every enhancement listed here; this list merely gives the preferred order for any enhancements and processes you do perform. For instance, you might want to despeckle, crop, and deskew your image, you would use this reference to see that you should deskew first, crop next, and despeckle last.

This list is only a suggested order; sometimes a different order might work better for your documents. Experiment with different combinations to see what works best.

1. **Resize.** This enhancement will always reduce your image quality, so it should be used very sparingly. If you need to change the image size, it's better to do it by scanning at a lower or higher resolution.
2. **Remove pages.** If you plan to remove unnecessary or blank pages, always do so before performing any other enhancement or process. Removing these first ensures that the scanning program won't waste time processing pages that will eventually be thrown away. Remember that deleting pages is permanent, so don't remove cover sheets until you have copied down any information you need from them.
3. **Rotate and deskew.** Some image enhancements, such as line removal or cropping, will act differently on horizontal images than on vertical images. Rotate and deskew images before performing other enhancements to ensure that all your enhancements behave in a predictable way.

If you are using photos of document pages instead of scanned images, **Photo Correction** should be applied at this stage, since it deskews and crops images.

4. **Crop and remove borders.** Once your images are rotated and deskewed, you can crop out any unnecessary regions; this can save some processing time when you apply other enhancements. Note that if you plan to use the Color Smoothing enhancement, you may wish to ensure that you leave a sufficiently large area of blank background to give Color Smoothing a good "reference" region; if your cropping operation will not leave an area of blank background for Color Smoothing to use as a sample, then perform Color Smoothing before you crop.
5. **Color Smooth.** This enhancement can only be applied to color images. Color Smooth improves the results of Dynamic Thresholding, so it should be performed first.
6. **Dynamic Thresholding.** Dynamic Thresholding converts color or grayscale images to black and white. This can greatly improve the performance of OCR processing.
7. **Invert.** If your image uses white text on a black background, use the Invert function to change it to black text on a white background. This enhancement will work on color or grayscale images, but it will probably have the best results on black and white.

- 8. Line removal.** Line removal is generally only necessary when you have lines running through your text. This enhancement will work on color as well as black and white images, but it will be more efficient on black and white.
- 9. Despeckle.** Despeckle works only on black and white images, and it expects images that have black text on a white background.
- 10. Smooth.** Use this enhancement to smooth out your text. Like Despeckle, this enhancement expects black text on a white background and works on black and white images. Running Smooth after Despeckle will slightly improve processing speed.
- 11. OCR.** Always OCR last. Most image enhancements are specifically intended to make text more readable both for users and for the OCR engine, so you should always make OCR your last step.

Remember, this list is just a suggestion. You'll usually be using just a few of the enhancements listed, and in some cases you might find that putting them in a different order gives you better results.



Scanning Best Practices  
December 2005

Author: Regina Carns  
Contributing Author: Ed Heaney  
Editor: Constance Anderson

Compulink Management Center, Inc.  
Global Headquarters  
3545 Long Beach Blvd.  
Long Beach, CA 90807  
U.S.A

Phone: +1.562.988.1688  
[www.laserfiche.com](http://www.laserfiche.com)

Laserfiche is a trademark of Compulink Management Center, Inc.  
Various product and service names references herein may be  
trademarks of Compulink Management Center, Inc. All other products  
and service names mentioned may be trademarks of their respective owners.

Copyright © 2005 Compulink Management Center, Inc.  
All rights reserved