

EXECUTIVE SUMMARY

There are numerous considerations to take into account when considering a scanning project but it all starts with two questions:

How will you be using these images?

This is probably the most critical consideration. What is the main business driver behind this scanning initiative? Will these images be used to produce true to life reproductions or is the purpose to create readable, reproductions of the original record?

What is the content that you will be scanning/ converting?

The quality and content of the information that you want to scan can affect the technology used and the quality that can be achieved when the scanning project is completed. The average number of pages, the total volume, the number of staples, sticky notes and hand-written notes in the margin are all considerations when determining which approach to converting the information is the most practical and will yield an acceptable result.

The following document will walk you through several considerations and the pros and cons of each. It is designed to help guide you to a better understanding of functions such as color modes, resolution, file formats and more so that once you answer the "How" question you can plan the rest accordingly.

GLOSSARY OF TERMS

Bare with us, we are about to get kind of "techy", but the below glossary will help shed some light on the more technical aspects of this whitepaper.

Compression

The act of reducing the file size of a scanned image typically by leveraging an automated software algorithm. There is typically a trade off between how much a file is compressed and the quality.

Document Generations

Refers to how many levels or steps a document is away from the original file. For instance, if you were to print an email, then later scan the paper, the scanned image would be the 3rd generation. Each generation typically loses some of its resolution.

Dots Per Inch (DPI)

Scanned images are made up of tiny squares (see Pixels below). The number of dots/squares/pixels within one inch of the scanned image is referred to as Dots Per Inch (DPI) or Pixels Per Inch (PPI). In theory, the higher the DPI the more information is captured during the scanning process and the larger the file will be.

Single Page File

One digital file that only contains a single image within it. For example, your digital camera produces one image file for every photograph you've taken.

Multi-Page File

One digital file that contains more than one image within it. For example, one PDF file that has 7 pages. Pages in this file refer to the number of front and back sides of documents that were scanned.

Document

A document is a collection of pages that are related to each other. A document can consist of one page or many pages. It may have been created by one person or many and could come from one source "system" or several different systems.

File

A file can refer to a File Folder or a digital document. It is important to be clear which type of "File" you are referring to when working with conversion of documents. Many will use the terms file and document interchangeably which can cause confusion. In addition, a file can contain one document or many documents.

Page

A page usually refers to a piece of paper. If the paper is one-sided, the number of pages and images are equal. If the piece of paper has information on two sides, there will be 2 images created during the conversion process. This is another area where using common terminology is important.

Image

The digital image that is the result of scanning or photographing one side of a piece of paper.

Simplex

When a document contains information on only one side of a piece of paper.

Duplex

When a document contains information on both sides of a piece of paper.

Pixel

If you zoom in close enough on a scanned document you will notice individual squares that make up the text and images. These are called pixels.

Raster

Raster images are ones that were created by capturing an image through a scanner or camera. It is a grid of tiny coordinates covering the entire image and each one of those coordinates (or pixels) are assigned a color value.

Index

Index refers to a word or phrase that helps to categorize or classify a document. Common Indexes are date created, author, document type (i.e.; contract, invoice, purchase order) and department (i.e.; Accounts Payable, Sales, Manufacturing). Indexes are typically created to help you filter down the results of your search. These words and phrases may or may not appear in the text on the document.

Text Search

A text search returns documents that contain specific words or phrases, much like a Google search. A text search can be very powerful when combined with document indices or keywords. Text Search only is not a recommended approach for storing searching through large volumes of business documents as it will take too long to look at each of the returned documents. Index values greatly reduce the number of documents returned from a search.

Document Preparation

"Doc Prep" refers to the labor and processing time required to prepare your documents for scanning. For example, staples and clips need to be removed, bent corners unfolded and sticky notes must be moved so they do not obscure important information and taped down so that they do not come off in the scanning process. Often each document is handled, sometimes multiple times during this process.

OVERVIEW OF COMMON DIGITAL FILE FORMATS

1. Portable Document Format (PDF)

PDF files can support a variety of compression options, as illustrated in the "File Formats and Commonly Supported Compression" section of this document. PDF supports multiple pages per file, allowing you to contain all the pages related to a single record within a single digital file.



2. Tagged Image File Format (TIFF)

TIFF or TIF are common file formats for storing raster based images (i.e. scanned documents, photographs, etc.). While TIFF can support multiple color modes, there have been compatibility issues with color and grayscale TIFF files in various viewing applications. Most commonly, TIFF files are used to store bitonal (black & white) scanned documents.

TIFF files can support a variety of compression options as illustrated in the "File Formats and Commonly Supported Compression" section of this document. TIFF also supports multiple pages per file, enabling you to contain all the pages related to a single record within a single digital file.

3. JPEG

The JPEG format is typically used with photographs and paintings of realistic scenes with smooth variations of tone and color. JPEG is not well suited for line drawings and other textual or iconic graphics, where the sharp contrasts between adjacent pixels is required.

COLOR MODES

Bitonal

Used on documents and drawings with only minor deviations in contrast throughout the collection. Bitonal is well suited for documents where there is no importance to the colors on the documents (i.e. logos). Business documents are typically scanned as bitonal files when they contain text, but no graphics. Bitonal files result in much smaller digital images which make them ideal for long-term archiving and active sharing/retrieval across intranets or the internet.

Bitonal is not suited well for poor contrast documents (i.e. faded thermal print pages) or documents where the identification of colors is important (i.e. photographs, drawing mark-ups)

Grayscale

Grayscale is recommended for files which would suffer information loss if scanned in bitonal mode. For example, poor contrast documents, photographs (where the color is not critical), and documents with a wide-range of variation in contrast, microfilm and microfiche. Grayscale document files are larger than bitonal image files, but smaller than color image files.

Why Should I Care?

Cost, cost and cost!

If there's no value in keeping the color, it will just end up costing you more to scan, store and manage.

Color

Color mode will typically not return more detail than grayscale on poor contrast files and is only recommended when color has a significant relevance to the information on the document. Examples of this may include: drawings with colored mark ups, photographs, charts/graphs which rely on color, etc.

Files scanned in color result in a much larger digital image, which can make it difficult to share images due to slow transfer speeds and can create network storage problems. Another option to consider to reduce file size is the use of automatic color detection. This is a process by which documents that contain color are saved as color files and documents that do not are saved as bitonal files. Note: Any color on a document other than black, such as a logo or a warning, will likely trigger the document to be scanned in color. This is often a consideration when a company's logo or other branding formatting is printed on the form.

RESOLUTION

Resolution, which is typically measured in Dots Per Inch (DPI), refers to the number of pixels captured during scanning. For instance, a document scanned at 300DPI would have 300 pixels per inch. In theory, the greater the number of pixels, the more information is captured. We say "in theory" because scanning at a higher resolution cannot capture data that was never there to begin with. As an example, most laser printers print documents at 300DPI, in that case scanning at 400DPI would not result in anymore content than scanning at 300DPI. This is also true for legacy media like microfilm and microfiche which have a realistic resolution of approximately 300DPI.

Another consideration to take into account with resolution is file size. As you increase the resolution, you increase the size of the digital image, which can lead to issues sharing and storing information.

When scanning large collections, 300DPI is more than sufficient for most applications. Adjusting the Color Mode (outlined above) is preferable to raising the resolution. A document scanned at 300DPI in grayscale mode will capture more data than the same document scanning at 400DPI in bitonal mode.

Why Should I Care?

Resolution is the backbone of quality. Go too low and you'll lose information. Go to high and it'll cost you a fortune

OPTICAL CHARACTER RECOGNITION (OCR)

OCR is the function that turns a scanned image into a fully searchable record. OCR does not replace a proper indexing/filing system; it only serves to enhance it. For example, you may have a legal file that contains over 200 pages. Flipping through the scanned images one page at a time trying to find the name "Thompson" would be extremely time consuming, but typing the name "Thompson" into the search box and being instantly shown all the instances of the name within the document is a huge benefit. Using only text search can be problematic. For example, if your company does a lot of business with ABC Co., you will likely have invoices, purchase orders, correspondence and checks from ABC Co. Searching for ABC Co. could return tens of thousands of documents. Searching for ABC Co. and filtering by a date range and/or by document type will greatly reduce the number of returned documents and speed up your access to the required information.

Finally, the results of OCR'ing are only as good as the original material. It's not a perfect technology, and as such, should not be relied upon as your only method of finding scanned documents. By nature, some documents are NOT well suited for OCR. These can include:

- *Photographs*
- *Drawings and maps*
- *Handwritten files*
- *Poor/low contrast documents*
Some microfilm and microfiche
Documents scanned at a low image resolution (typically less than 300 dpi)

Why Should I Care?

Findability!

OCR can make finding the right page on the right document a breeze -giving you more time to focus on the task at hand.

SUMMARY

As you can see, there is no hard and fast rule for "the best scanner settings". There are numerous variables to consider. However, in our experience, the clear majority of documents, drawings and photographs being scanned for their informational value are suitable to be captured at:

- 300
- Saved as PDF documents
- With the color mode being adjust based on the project needs.

Note: If your document management system is optimized around a different file format, such as TIFF, consult your service provider for a recommendation

BONUS CONTENT - SCANNING LABOR LEVERS

While the quality and accessibility of the final content is the most important consideration, few companies have an unlimited budget for their scanning projects. There are many factors that impact the cost to convert your documents. These can include:

- The number of documents and their condition
- The Document Preparation required - pulling staples and moving sticky notes is time consuming and labor intensive.
- Bitonal is the least expensive and color is the most expensive. The file size is the lever here as it takes longer to process the larger image formats generated by grayscale and color scans
- The higher the DPI, the larger the file size. Again, the file size is the cost lever.

Indexing required- more index fields might present the opportunity to receive a smaller returned search result, but it also requires more labor and more cost. Don't get carried away with the number of index fields. Use only what will be required to narrow down the search to a reasonable number of documents. Consider how everyone searches for the documents. For example, some may search by customer ID# while others use the customer name or the shipping location

- Another consideration is how easy it will be for a service bureau team member to locate the index information on the documents. Is it always on the first page of multi-page documents? Is it always in the same place?
- Index Field lengths - the more characters required, the longer it takes to key them
- Must individual document types be identified? For example, will you want the Application Form, Periodic Review Forms and other documents specifically indexed in a retired personnel folders each identified? If so, this requires the indexer to slow their keying rate and make decisions. Any time the indexing rate slows, the costs go up.

Turn-around time. If you need it fast, it will cost more as your vendor will have to move around other projects to meet your needs. If you can wait, they can fit it in during slow periods and they are more willing to give you a price break.

WE MAKE YOUR LIFE EASIER!

Most people take on their scanning and document management tasks infrequently and struggle to understand the best long term approach to dealing with these issues. Let us leverage our experience and expertise to help.

QUESTIONS? GET ADVICE FROM AN EXPERT

**BOOK YOUR FREE CONSULTATION TODAY BY
CALLING 1.800.443.4583**

