



Which Scanning Standards Should I Use?

Getting the best image.

There are three scanning factors to consider:

1. Resolution.

Scanning resolution is often discussed in terms of PPI (Pixels Per Inch) and DPI (Dots Per Inch). **PPI** is most useful in preparing files for viewing on a computer screen. An image with a higher PPI tends to be higher quality because it has a greater pixel density. The industry standard for office documents is 300 PPI, but color, faded, or rare documents may be scanned at 600 PPI or even as high as 2400 PPI in order to capture every line of the image. **DPI** refers to the resolution value of a physical printer. Printers reproduce an image by spitting out tiny dots, and the number of dots per inch affects the amount of detail and overall quality of the print. Black and white documents that will be OCR'd (see below) should generally be scanned at 300 DPI. The larger the DPI, the bigger the file size which can eat up storage space.

2. Compression.

Compressing a scanned image means reducing the size of the image from its original size. Usually this is done by saving the image in a different format (saving a high resolution TIFF as a lower resolution JPEG, for example). Compression saves storage space, but it also reduces image quality, and that quality cannot later be recovered. If the documents are standard office documents, short a compressed file is usually adequate; other types of files, though, such as photographs or building plans may require an uncompressed high resolution image.

3. File Format.

It is best to scan your documents to a widely-used well known file format, such as PDF, TIFF, JPEG, etc. If you use a format that is specific to a product or to one vendor, you may one day have difficulty moving the file to a different storage and retrieval system or finding a reader that can open it. This becomes more important as the retention of the record increases.

Special considerations:

Archival/Permanent documents - If you plan to scan records that are considered archival (check your retention schedule), the files must be in PDF/A file format. Contact the State Archives if you have questions regarding this policy.

Optical Character Recognition (OCR)— The software that scans your documents often produces only an image of the text so that the text itself is not searchable (the computer only sees a picture, not words). If you plan to OCR your records (i.e., run a software that recognizes and indexes the text so that you can search for words in the document), you should test some scans to make sure your DPI is set high enough for the OCR software to recognize. For office documents, 300 DPI is generally adequate, but you may need to increase it to as much as 600 DPI to ensure accurate OCR'ing of the image.