

Guideline

Technical considerations for scanning university documents



Records Management and Privacy Office
Kingston, ON K7L 3N6
Tel. (613) 533-6095
recordsmanagement@queensu.ca

This guideline is designed to assist units when undertaking a scanning project or establishing a scanning standard for a department, office, or Faculty.

Considerations

Prior to initiating a scanning project, units should refer to the fact sheet [From Paper to Digital: Scanning University Records](#) to explore the legal and operational implications of scanning hard copy documents.

Determine the optimal file format

In order for scanned documents to remain usable for their entire retention period (as determined by the university's [records retention schedules](#)), they must be scanned in such a manner that they can reliably replace the hard copy originals. If there are issues of readability, unsupported file formats, or missing information then the scan cannot be said to be a reliable replacement for the original.

To remain legible, a high enough resolution needs to be selected, but the resolution in dots per inch (dpi) must be balanced against the final file size in Mega Bytes (MB). Scanning at the highest available resolution will result in the greatest fidelity and readability, but the file size may be so large as to make the file cumbersome to transmit or to store.

Another consideration is colour. If the colour elements of a document are key to their readability or legality they must be preserved, but if that colour information is not essential, the file size can be greatly reduced by selecting greyscale during the scanning process. Choosing black and white (1-bit scanning) can cause too much loss of information and can make the document difficult to read and is not normally advised unless the original document contains nothing more than black text on a white page.

Most scanners will default to a few file formats. Choose the appropriate file format(s) for the unit's needs.

- **PDF** (Portable Document Format) is a standard file format for text-based records. PDFs can be created as Optical Character Recognition (OCR) scans which allow for text searching, or as non-OCR PDFs which lack text searchability. Any PDF file can be compressed when it is created or saved, and information can be lost during this compression; therefore, users should review their final scanned document to ensure it meets their requirements.
- **TIFF** (Tagged Image File Format) is a standard file format recommended for image-based records. TIFF is a widely supported uncompressed file format commonly used in desktop publishing, 3-D applications

Scanning Standards

and medical imaging applications. Due to their lack of compression, TIFF files are large and should be used only when complete fidelity of the image being scanned is essential.

- **JPEG** (Joint Photographic Experts Group) is a standard file format used for storing scanned image-based records. JPEG files typically employ an irreversible compression which causes a loss of quality in order to achieve a smaller file size. Differences in quality between the original and scanned image should be considered before choosing this file format.

Minimum technical standards for textual documents

Use the table below to determine the minimum standard for scanning textual documents. If maps, technical drawings, photographs or other non-textual documents are to be scanned, contact the Queen's University Archives or the Records Management and Privacy Office to discuss a best course of action for these documents.

Resolution Guideline Minimums

Scan resolution is 100% scan ratio (1:1)

Documents with text only, black and white	200 dpi, black and white (1 bit)
Documents with watermarks, grey shading, grey graphics, handwritten notes/markings, low contrast, half-tone illustrations, or images	300 dpi, Grey Scale (8 bit)
Documents with discrete colour used in text or diagrams where colour is important for accurate representation	300 dpi, Full colour (24 bit)

Conduct quality control

Most scanners will assign default names for scans; these automatically generated file names are of little use and a file naming standard should be agreed upon within each unit. See the guideline on [Creating and Maintaining File Naming Standards](#).

Once a file has been renamed it should be stored on a Queen's University server which is backed up on a regular basis. Documents should be filed in the appropriate folder or subfolder within the unit's recordkeeping system. See the guideline on [Creating and Maintaining a Folder Structure](#).

To ensure the reliability of the scans, a sampling of pages scanned should be reviewed to check for quality. A 30% sample is recommended, and high-risk documents may necessitate 100% review. Evaluate the digital output quality to verify that the digitized version accurately represents the content of the original document. All scans must be legible to the smallest font on the record, regardless of colour and markings. If the document is not legible at the minimum standard, adjust scanner settings and thresholds, or take other actions as required to ensure accuracy and legibility of the scanned document. Check the pages to examine:

Scanning Standards

- smallest detail legibly captured (e.g., smallest type size for text; clarity of punctuation marks, including decimal points);
- completeness of detail (e.g., acceptability of broken characters, missing segments of lines if present);
- dimensional accuracy compared with the original (i.e., is the document still readable?);
- scanner-generated speckle (i.e., speckle not present on the original);
- completeness of overall image area (i.e., missing information at the edges of the image area);
- density of solid black areas; and
- colour fidelity (if colour is required).

Document the process

Document your unit's scanning process in a [Recordkeeping Protocol](#) (template available from the Records Management and Privacy Office), including the scanning resolution and standards, the file type, naming convention(s), folder locations, and quality assurance minimums applied to the documents.