

Introduction and Rationale

The very beginnings of RUCORE had its roots in the preservation of scanned digital content. The very first accepted file formats and architectures were built around the premise that the majority of content within RUCORE would be printed documents, manuscripts, paper photographs and books which would have to be digitized, or scanned from the analog original into a digital file format.

In order to standardize imaging specifications throughout RUCORE and other archival imaging projects, the following minimum standards and requirements have been established for preservation-grade digital images.

Color photos and documents:

- Should be scanned at minimum 600 dpi for archival purposes
- Color space should be sRGB, at least 24 bit truecolor. 8 bits per channel is preferred.
- Archival files should be in TIFF format, encoded with no compression or (where available) a lossless compression algorithm such as LZW.

Black and white documents with gray tones, or black and white photos:

- Should be scanned at a *minimum* of 600 dpi.
- Color space should be grayscale. RGB is also acceptable, but contributes nothing to pure grayscale images and only increases file size.
- Archival files should be in TIFF format. Archival files should be in TIFF format, encoded with no compression or (where available) a lossless compression algorithm such as LZW.

Black and white documents with pure text and line drawings; *no* photographs or items with greyscale tones of any kind:

- Should be scanned at a *minimum* of 400 dpi.
- Color space should be 2 bit color (black and white), with a 50% threshold. No dithering of any kind, as this will impede the performance of Optical Character Recognition (OCR) software to recognize text.
- TIFF format for archived images, with no or lossless (LZW) compression.

Additional Considerations - the 3,000-pixel rule

- Scanning at resolutions much higher than 600dpi may be required if the document is small in physical size.
- All digitized documents should be scanned such that at least one dimension of the visible area of the document (width or height) is comprised of a minimum of 3,000 pixels each. If a scan at 600 dpi does not meet this minimum, then the resolution should be increased and the document re-scanned accordingly. *It is not acceptable under any circumstances to artificially increase the dpi by resampling the image. The original scan **must** meet the 3,000 pixel minimum.*
- Extraneous image information (such as a calibration strip, record labels, or any part of the scanned image that does not directly pertain to the document itself) should not be included in this calculation. Measurements should be made using only the visible area of the document.
- Example: a 3"x4" color photograph scanned at 600 dpi will render an image that is 1800 x 2400 pixels. As the width does not meet the minimum, a 600dpi scan will be unacceptable. The scan should be increased to 800 dpi to render a 2400 x 3200 pixel image, and thus bring the image to an acceptable level of detail.

On the other hand, an 8.5"x11" sheet of paper with black and white text will scan acceptably at 400dpi, rendering a 3400 x 4400 pixel image.