



Guideline on Digitizing UW-Madison Paper Documents

Digitising is the conversion of paper-based documents to digital images and making them readily accessible, thereby enhancing the business process and workflow for university departments and units on campus

Your path towards successful Imaging Project:

Analyze your document workflow and determine if the digital copies will be used instead of or in addition to paper documents. Decide will the digital items be separated from one source item into multiple documents.

Organize records series that your paper documents may create. Check if they may already be scheduled under Records Disposition Authorization (archives.library.wisc.edu/records/rda.html) or UW-Madison General Records Schedule (archives.library.wisc.edu/records/rda.html#general-records). Save time by consulting with campus Records Officer - recmgmt@library.wisc.edu (*Peg Eusch*).

Consult with the CIO Office of Campus Information Security (cio.wisc.edu/security-risk.aspx) to ensure that project practices will comply with applicable state and federal regulations and standards regarding public records. Remember that risk of security breach is much higher because of ease of reproducibility. Consider incorporating visible or metadata-based watermarks to prevent casual copying.

Arrange order of your documents for scanning. Deliberate documents preparation ensures that all of them will be scanned and scanning will be quick. It is best to prepare all documents in batches before they are going to be scanned. It will take about 1 hour to prepare 1000 pages. This figure increases for non-standard file sizes.

Prepare documents carefully removing all clips, brads, staples, and other fasteners from folders and from multi-page documents. Unfold paper and check proper sequence of files and pages. Remove sticky notes or tape them if they contain important information. Identify documents to be enhanced by using special scanning techniques.

Identify a metadata scheme that will be used for describing and capturing information in the documents. This will allow the records/documents be searched and retrieved for later use. At minimum, the information system should capture: *Date Created, Format, Unique Identifier, Official Title* of the image, *Creator/Organization* that originated the document. Dublin Core metadata allows search. Administrative metadata identifies rights, statutes, etc. Structural metadata ties records together.

Choose appropriate resolution and decide on the files format for your images. Prefer 300 dpi - Grayscale - TIFF or PDF for preservation of text documents, and 600 dpi - Grayscale/RGB - TIFF or JPG for preservation of photographs, maps or other graphics.

*Visit the University Records Management Program
www.archives.library.wisc.edu/records/
for more details*

Index your new digital files with consistent naming system, metadata elements, controlled vocabulary. All this will greatly increase accessibility. File names should be easily browsable (*Example: Date_DepName_UniqueProjectName*). Every efficient retrieval mechanism should include mandatory metadata elements, such as *Title, Creator, Date, Identifier* and other, as well as may contain *Subject, Tags, Controlled Vocabulary* elements. More information on metadata Standards see at dublincore.org/documents/2008/01/14/dces/.

Convert SOURCE FILES that may be used only as the preservation official copies to subsequent CONVENIENCE COPIES with lesser resolution, more diverse formats and that are compressed for office use. Use 150 dpi/PDF for text documents, and 300 dpi/JPG for graphic files.

Inspect newly digitized documents on the quality of images and metadata integrity. Make sure that *the Optical Character Recognition program* reads the text accurately. Examine images on specific digital defects: color and brightness balance, orientation, resolution, lines and defects on scanner screen.

Train all the personnel who will be given access to the digital document system. Only personnel executing their regular daily job duties with digital documents may be granted access to imaging system.

Provide a secure access to digital files and a convenient public or admin retrieval point. Communicate with the UW Records Officer (*Peg Eusch, recmgmt@library.wisc.edu*) to ensure choosing an appropriate document management system. UW-Madison may have a campus-wide license for this type of content management system and highly recommends projects to be done through them.

Design backup procedures to create security copies of files. This copies should become available in cases of disaster, theft, malfunction. The copy files should be stored in an offsite, environmentally controlled location. Campus IT Department (doit.wisc.edu/about/departments/#security), Risk Management stuff (busssvc.wisc.edu/risk_mgt/risk_mgt.html) and the UW System Office of General Counsel (legal.wisc.edu/about/coordination.html) must be consulted for advice.

Prepare your digital files for a long-term storage or data migration from older to newer hardware and software platforms when a storage method becomes outdated in order to comply with Wisconsin Administrative Rule 12.

Destroy and delete image records that exceeded retention periods. The actual record, besides its indexes, must be deleted from media. Check *The University Archives and Records Management* webpage for retention information (archives.library.wisc.edu/records/rda.html#records-schedules)

Digitization is NOT a solution for reducing storage costs for paper documents. The cost of software, hardware, server space, training, and maintenance of systems exceeds the cost for storing paper forms. Frequently used documents with short (2-7 years) retention period are the best candidates for digitization.

