

UNIVERSITY OF WATERLOO

Standards for Imaging (Scanning) Paper Documents

For Capture into University Systems

Version 1: Effective August 1, 2011

Approved by the Document Management Steering Committee July 6, 2011

Table of Contents

Introduction.....	3
Scope.....	3
Unit Specific Procedures	3
Outsourcing of Imaging.....	4
Document Imaging Set-Up and Procedures	4
Scanning Room.....	4
Scanner Maintenance.....	4
Document Preparation.....	5
Document Batching.....	5
Imaging Procedures	6
Metadata and Indexing.....	6
Quality Control of Imaging and Indexing	6
Sample Set.....	7
Checking Scanner Performance	8
Quality Control of Production Scanning and Indexing.....	9
Quality Control Logs	10
Document Storage, Retention, and Disposal	10
APPENDIX 1: SCANNER TEST TARGET PM-189	11
APPENDIX 2: (UNIT) SCANNER PERFORMANCE QUALITY CONTROL LOG	12
APPENDIX 3: (UNIT) PRODUCTION QUALITY CONTROL LOG.....	13

Introduction

These standards are issued in support of the University of Waterloo Policy 12: Records Management to assist the university in meeting accepted standards for managing its records. They are designed to ensure that electronic images of paper source documents, where the image is intended to replace the original as the official record, are captured in a way that ensures their reliability, authenticity, and usability over their entire retention period and that allows the images, or copies produced from the images, to be authenticated as true copies of the originals and to be admissible as evidence in legal proceedings. Following these standards will facilitate the university's compliance with the Canadian standard CAN/CGSB 72.11-93: *Microfilm and Electronic Images as Documentary Evidence* and CAN/CGSB-72.34-2005: *Electronic Records as Documentary Evidence*, and will allow paper source documents to be disposed of once the images have been captured into a university system.

Scope

These standards apply to paper source documents approved for capture into university systems. They do not cover electronic object capture. Nor do they cover specialized imaging processes related to, for example, engineering drawings and geographical maps. In addition, they do not cover processes used to evaluate the authenticity of the source documents prior to their being scanned and imported into a system. However, compliance with these standards can be used to demonstrate that output from the system is a true reproduction of the paper source document.

Unit Specific Procedures

Any university department using document imaging, where the electronic images are intended to replace the paper source documents as the official record, must use these standards as the basis for detailed procedures for its particular documents. Each unit's detailed procedures must include the following:

- Types of documents to be scanned and the business processes they are part of;
- Procedures for document preparation, batching, and scanning, including procedures for particular document types where applicable;
- Procedures for indexing the images;
- Procedures for reviewing image quality and accuracy of indexing;
- Procedures for committal of images to the system; and
- Procedures for interim storage and disposal of paper source records.

Outsourcing of Imaging

If a unit contracts out all or part of its imaging program, the service provider must first demonstrate compliance with CAN/CGSB 72.11-93: *Microfilm and Electronic Images as Documentary Evidence*.

Document Imaging Set-Up and Procedures

The set-up and procedures below apply to scanning operations centralized within a unit. These procedures do not include distributed capture or the use of multi-function devices.

Scanning Room

It is recommended that the scanning room include the following features and equipment:

- Room dust free and preferably without carpets (to minimize static electricity);
- Temperature between 21° and 25° C and humidity between 30% and 40%;
- Dedicated electrical outlet based on the electrical specification for the scanning equipment;
- Table or desktop for document preparation;
- Paper jogger to prevent paper sticking together;
- Computer workstation and scanner placed so that the operator can easily feed documents, remove jams, and clean the equipment; and
- Container(s) in which to place the scanned documents, labelled with the scanning date, batch number, and documents types, as appropriate.

Scanner Maintenance

Regular maintenance of scanning equipment is critical to ensure that it is operating properly and achieving optimal performance. The maintenance procedures provided by the manufacturer in the equipment manual for each type and model must be followed.

Before beginning daily scanning operations, routine scanner maintenance must be carried out by the scanner operator. These procedures include, for example, cleaning the rollers and the glass plate. Refer to the scanning manual for details.

It is recommended that a set of spare parts, such as rollers and lamps, be acquired for each scanner, so that replacements can be made as necessary without undue interruption of scanning operations.

Regular servicing of scanners must be scheduled, including the regular replacement of parts that must be performed by the manufacturer. Refer to the scanning manual for details.

Document Preparation¹

Each unit must list the types of documents which will be imaged and document the business process(es) they are part of, for example, how they are received, sorted, and incorporated into the imaging process. This will serve to demonstrate the integrity of the documents from the time they are received to their storage in the system.

Routine document preparation includes the careful removal of staples, paper clips, and other bindings as well as unfolding and flattening folded documents.

As part of planning for document imaging, each unit must examine the types of documents that will be scanned and look for special features that might present scanning challenges, such as the following:

- Multi-page documents or documents that must be kept together (e.g., a letter with an attachment);
- Fragile documents;
- Poor quality or faint print, or poor colour contrast due to, for example, coloured paper or light characters on a dark background;
- Seals and / or watermarks;
- Physical attachments such as photographs or sticky notes;
- Extremely thin or thick paper, for example, onionskin or card stock; or
- Unusual size or orientation of paper.

Procedures for handling the specific issues found must be documented in the unit specific procedures. (For more information on methods of addressing scanning challenges, including photocopying, see the standard referenced in footnote¹).

Document Batching

Documents should normally be grouped into batches for scanning. The size and content of a batch should be determined by each unit based on the nature of the documents being scanned and the unit's workflow. For example, documents might be batched by type or the contents of a file folder might be considered a batch. Procedures for batching must be documented in the unit specific procedures and must include, where applicable, the use of barcodes and the use of batch sheets (with e.g., batch number, scanner operator ID, and date scanned) for subsequent traces and storage.

¹ The following procedures are largely based on ISO/TR15801: *Document Management: Information Stored Electronically, Recommendations for Trustworthiness and Reliability*, sections 5.4 – 5.7. A copy of this standard is available for consultation in the office of the University Records Manager.

Imaging Procedures

Each unit must document the specific scanning procedures for its particular documents. The procedures must include the following:

- Scanner settings, including settings for particular batches or document types, where applicable;
- Indexing requirements for use with particular batches or document types;
- Variations in the process for different document types (e.g., single-sided vs. double-sided; colour vs. black & white);
- Procedures to ensure that all documents in a batch are in fact scanned (e.g., comparing the number of scanned images with the number of documents in a batch; the use of a document jogger to prevent documents accidentally feeding together);
- Procedures or techniques used to improve the quality of images (Note: image enhancements should be used with extreme caution; de-speckling, for example, is not recommended since it may remove important isolated characters); and
- Procedures for retakes and disposal of rejected images.

Metadata and Indexing

Metadata is structured information that describes and enables finding, managing, understanding, preserving and disposing of records. Units must ensure that the minimum recordkeeping metadata elements as required by the *University of Waterloo Electronic Recordkeeping Metadata Standard*² are captured as part of any imaging process. Additional metadata elements may be required to ensure efficient retrieval and processing of images. The appropriate metadata elements for each type of document must be determined as part of the project planning process.

Indexing is the capture of appropriate values for metadata elements (e.g., 2011-06-24 as the value for the created date element of a document). Indexing may be manual or automated. Image indexing information must be accurate, securely stored, and persistently linked with the image. If the indexing information is incorrect or lost, then the image may effectively be lost also.

The procedures for indexing must be documented in the unit's procedures. If image quality evaluation is part of the indexing process, this should be made explicit in the procedures.

Quality Control of Imaging and Indexing

Quality control is an essential component of the imaging process when the image is to replace the paper source document, since the source document will not be available later to supply missing or illegible information. Procedures must be put in place to minimize the likelihood of

² In draft form only on July 6, 2011

poor quality scanned images and to ensure that problems are discovered while the original documents are still available.

Since accurate indexing is vital for the retrieval of stored information, procedures for checking the accuracy of indexing and making corrections as necessary must be established.

Each unit must document the quality control procedures for its particular documents, including the procedures for rescanning and purging the original images from the system, re-indexing, and for keeping records of quality control checks.

Sample Set

Each unit must assemble a sample set of documents representative of the types of documents that will be scanned. The sample should include examples of documents which are poor in quality compared to the majority of documents to be scanned. When new document types are added or when documents change, they need to be included in the sample set.

The sample set will be used initially to determine the scanner settings which produce images of acceptable quality and periodically thereafter to check that the scanner is working to the acceptable standard. For most documents, a scanner resolution of at least 200 dpi will be necessary. For optical character recognition, a scanner resolution of at least 300 dpi will be necessary.

The quality criteria for the document set must be agreed on by the users and documented. The needs of the users should be balanced against the capabilities of the scanner and the characteristics of the documents. ISO/TR 15801 (Section 5.4.6.1) suggests addressing the following issues:

- Overall legibility;
- Legibility of small details (e.g., punctuation, decimal points);
- Completeness of detail (e.g., acceptability of broken characters);
- Dimensional accuracy compared with the original;
- Acceptability of scanner-generated speckle;
- Acceptability of missing information at the edges of the image area; and
- Density of solid black areas and colour fidelity.

The criteria must include specifics for particular document types where applicable (for example university seals and logos on transcripts, legibility of hand-written signatures and text).

Images captured from the sample documents should be examined using a high resolution display screen. They should also be printed using a printer similar in quality to what will be used for producing printed copies from the system, to ensure that printouts are of acceptable quality for the uses which will be made of them.

Once an acceptable set of images has been produced from the sample set, the images and the printouts of the images must be saved for future reference. The scanner settings and the characteristics of the monitor and printer used to examine the images must also be documented.

The sample set of documents must be rescanned and the results compared to the original scanned images when a scanner is first installed, recalibrated, serviced, moved to another location, or when new scanning software is downloaded. Results must be recorded in the Scanner Performance Quality Control Log (See below and Appendix 2).

Checking Scanner Performance

It is recommended that a scanner test target (Scanner Test Target PM-189, IEEE Std 167A – see Appendix 1) be used regularly to determine whether the scanner is performing consistently to the required standard.

The frequency of tests runs using the test target is subject to volume, time period, and the scanner maintenance schedule. A test run should be carried out after continuous use of 8 hours or after approximately 10,000 documents have been scanned. If a scanner is in constant use, using the test target routinely after the daily scanner maintenance would be good practice. Test runs, including the sample set, must be carried out when a scanner is first installed, recalibrated, serviced, moved to another location, or when new scanning software is downloaded. The results of test scans must be recorded in the Scanner Performance Run Quality Control Log (See below and Appendix 2).

Place the test target on the scanner in the same way as production documents. If duplex mode is required, ensure that the test target is placed to test both sides. Test images should be placed in temporary storage while the results are examined.

The scanner settings for test runs must be the same as those determined for production documents. Normally, the following criteria should be met:

- The registration markings at the four corners of the test target should be visible;
- The diagonal line across the test target should be straight with no breaks or distortion;
- The grey shading level marking shown under line 85 to Line 150 should exceed the level of 15 and above;
- Resolution scale, both vertical and horizontal, should exceed 200 and above, with clear breaks of the line at the the markings at 200;
- Character readings of 6 pt should be clearly visible; and
- All isolated characters should be clearly visible.

Where specific document types require criteria different from the above, these variations must be documented.

If the scanner does not consistently produce the required quality, it must be serviced.

Quality Control of Production Scanning and Indexing

Each production batch of images must be examined initially by the scanner operator. Base the examination on the accepted criteria for the sample set. Images found to be of poor quality may be deleted and the documents rescanned or the entire batch deleted and rescanned. Unless a unit's business workflow includes an inspection of each image, a sample of images must be subject to a second inspection by a person other than the scanner operator.

Quality checks on scanned images may be performed before they are indexed or image quality and indexing accuracy checks may be performed together as suits the needs of the particular unit.

During the initial stages of implementation of an imaging process or when a scanner operator or indexer is inexperienced, a large sample of documents (80% or even 100%) must be reviewed. Once users are confident in the quality and accuracy of the scanning and indexing processes, the sample size may be reduced so that 5% of images are reviewed.

The source paper document must be compared to the image. The sample should be selected from different parts of the batch if applicable. If a poor quality image is found, the images immediately preceding and following must also be examined.

When poor quality images are found, the reviewer must determine whether the error is major (requiring rescanning) or minor (the image is acceptable as is); an isolated incident (meaning only that image needs to be rescanned) or something which might affect the entire batch (meaning the entire batch should be reviewed and perhaps rescanned). If an entire batch is to be rescanned, the reviewer should determine whether further batches should be reviewed as well.

Anything which omits or renders illegible any of the substantive content of the document must be considered major and require rescanning. The following are examples of major quality issues:

- Missing characters;
- Illegible small characters or handwriting;
- Missing lines or sections;
- Missing portions of edge of image; and
- Consistent spot or "noise" which obscures part of image.

Minor quality issues, which do not require rescanning unless extreme and/or render document or parts of it illegible, include the following:

- Image skew;
- Background noise or speckle;
- Incorrect image orientation;
- Poor contrast; and
- Loss of certain colours.

Quality Control Logs

The results of all scanner performance tests must be recorded in the Scanner Performance Quality Control Log (Appendix 2), which includes the date of inspection, the scanner operator's name, scanner settings, documents tested (test target or sample set), and the results. This log provides evidence that quality control procedures are being followed and allows a scanner operator to spot developing problems.

The results of all quality checks of production batches must be recorded in the Production Batch Quality Control Log (Appendix 3), which includes the date, the inspector's name, the batch number, acceptance or rejection, and comments. Comments should indicate the nature of the problem if any, including whether the problem is due to operator error, scanner malfunction, or the quality of the source document, and steps taken to correct it (e.g., operator re-training, scanner re-calibration, etc.). These logs must be retained for inspection/audit purposes.

Document Storage, Retention, and Disposal

Once image quality and indexing accuracy checks have been completed, the electronic images must be saved to the system.

Electronic images must be retained and disposed of in accordance with university's approved retention schedules.

The paper source documents must be retained at least until all quality reviews and any necessary retakes or corrections have been completed and the images have been saved in the system. In most cases, retaining paper source documents from 2-6 months after scanning will be adequate. Paper source documents should be stored in files or storage boxes with the batch number (as applicable), scan date, and any other information required for the retrieval of the documents if necessary.

Since the electronic image is not considered to be the official record while the paper source document exists, it is important that the source documents be destroyed once quality control procedures have been successfully completed, the images have been saved in the system, and appropriate back up procedures have been completed. Destruction of paper source documents must be authorized by the unit head and a record must be kept of the destruction.

In some cases, it may be necessary to retain paper source documents, for example, if a paper source document was of such poor quality that a legible image could not be made, or the source document is particularly valuable (e.g., a transcript which could not be replaced). In other cases, source documents may be returned to their owner (e.g., immigration documents).

Each unit must document procedures relating to the storage and destruction of paper source documents, including instructions for managing source documents which will be kept or returned rather than destroyed..

APPENDIX 1: SCANNER TEST TARGET PM-189

Scanner Test Target PM-189
Precision™ (Ritter) Target

Resolution Patterns (LINES PER INCH): 81, 71, 64, 91, 102, 112, 127, 142, 56, 51

Text Samples:
 ABCDEFGHIJKLMNOPQRSTUVWXYZ
 abcdefghijklmnopqrstuvwxyz
 1234567890 Megaron Medium 12 pt
 ABCDEFGHIJKLMNOPQRSTUVWXYZ
 abcdefghijklmnopqrstuvwxyz
 1234567890 Megaron Medium 10 pt

Color Calibration:
 WHITE
 BLACK

Vertical Scale (LINES PER INCH): 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 800, 1000

Horizontal Scale (CYCLES PER MM): 1.0, 1.1, 1.25, 1.4, 1.6, 1.8, 2.0, 2.2, 2.5

Isolated Characters:
 e n m p q r s t u v w x y z
 4 5 6 7 8 9 0
 8 9 0 1 2 3 4 5 6 7 8 9 0

Handwriting Sample:
 This PM-189 Target is your true "bench mark"!

Resolution Chart:

150 LINE	5	10	15	20	25	30	40	50	60	70	80	90
133 LINE	5	10	15	20	25	30	40	50	60	70	80	90
120 LINE	5	10	15	20	25	30	40	50	60	70	80	90
110 LINE	5	10	15	20	25	30	40	50	60	70	80	90
100 LINE	5	10	15	20	25	30	40	50	60	70	80	90
85 LINE	5	10	15	20	25	30	40	50	60	70	80	90
65 LINE	5	10	15	20	25	30	40	50	60	70	80	90

CERTIFICATION

Signatures of the printer, the printer's supervisor, and the printer's manager.

