



Preparing Your Documents for Imaging: Increasing Productivity and Accuracy in Paper Conversion

By: Sergio Quiroz and Dave Steimling

Quality Associates and their subsidiary DocPoint Solutions have been engineering document conversion solutions for clients for over 20 years. Although we can almost always find room for improvement in many areas of the process, we recognize that the document preparation methodology ('doc prep' or simply 'prep') is an area where accuracy is of the highest importance. We have also found that this is a section of the process in which improvements can often be made with little problem regarding training and adoption.

Our experience has shown that proper planning and training in the separation and preparation of documents is the single most important factor in producing accurate end-results in a document conversion endeavor. Obviously the other various steps in the document conversion process are also important, but if the documents are not properly prepared for the process, delivering accurate work to those subsequent processing steps becomes less likely as one gets further down the processing line. Although in any imaging job a small percentage of documents will almost always need rescanning to correct for very poor image quality, re-working masses of poorly prepared documents can do more to skew a job's P/L statement than any other factor. For this reason, it is also the area in which the best recoveries can be made, and, even better: this is where the greatest losses can be prevented.

Documents not being separated properly will result in one of two scenarios: either the documents are "over-separated" (i.e.: the documents were split apart more than was necessary, incorrectly subdividing individual documents) or they are under-separated (i.e.: the documents are not separated enough, and multiple documents are stacked together as if they were a single document). Either of these cases results in decreased accuracy in the retrieval of documents from the final repository. If one had to choose the lesser of two evils, the former would likely be better, in that in the latter example documents that are combined with (behind) other documents will likely not be recognized as such and therefore will not be indexed as a separate document. The result of this are lost documents.

Correcting such errors past the point of scan can be time-consuming and costly. If the errors become recognized decisions have to be made on how to handle the problems, which would be based on where in the process the errors were found and on the process itself. Some processes may lend themselves to correcting such errors more easily than others, and some modules of a process may lend themselves to assisting with the correction more than others. Once the documents have been released into a final repository, things can become more difficult. This is especially true dependent on the access to the digital repository, the access to the original documents, the file types of the errant documents, their storage format, etc.

Let's explore how we can maintain a high rate of accuracy throughout our process by looking at how to best prepare the documents for conversion.

There are generally three portions of a document conversion process: the backfile conversion, the day-forward process and on demand scanning. Often the process for these portions will be different, as (for example) the conversion of the backfile may be best suited with a broader, less specific scope depending on how and how often the documents are used, what the retention for those documents will be, etc. On the other hand, the day-forward scanning might be best served with tighter specifications and further document subdivisions, depending on their use. On demand scanning has its own challenges, getting specific documents to the client out of the backfile conversion as they need to access them. The differences in these methodologies should be calculated in advance of beginning any document conversion so that all of the scopes are well defined. This reduces 'surprises' for the client, and also for the contractor.

The Purge:

The first step in processing a backfile conversion is the removal of documents and pages that do not need scanned. This is referred to as 'purging'. More often than not, this process is completed by the client as they will always have the greater institutional knowledge about their own documents. It is always most efficient to have only the documents and pages that are to be scanned brought into document preparation. Sometimes the purge will be part of the prep process. If purging is part of the prep, and to be considered the contractor's responsibility, detailed instructions must be given to ensure that the contractor has as much knowledge as possible of the document types, duplication schemes, etc, so that the purge occurs accurately: insuring no unintentional document losses or inclusion of excess scans.

The Prep:

Document Preparation (prep) has many steps that generally occur relatively concurrently. Not every step will be necessary for every document set in the conversion (i.e.: some may have staples and some may not), but the steps that are necessary will be applied more or less 'on the fly' by the prep processor (prepper). Because a prepper works fairly independently, it is of the utmost importance that they be trained properly so they may make the decisions regarding the document prep with minimal supervision. In order to insure the level of training of the prepper (as well as any operator that is involved in a conversion process) their work should be checked regularly and thoroughly after their initial training, and then spot checked on regular occasion thereafter. Here I list the steps of document prep, and include notes about each:

1. Initial orientation –
 - a. Prepping and processing should start and continue in a manner logical to the document collection. If the document collection is sorted alphabetically, the batches should start with "A" and work through to "Z". If they are numeric, start at the file cabinet, etc, with the lowest number and work toward the highest number. In most cases, the preferred order will be fairly obvious. If it is not, the specifications should be included in, and followed from, the statement of work.
 - b. Start with a stack of 'raw' documents from the first folder or drawer, oriented face up.
2. Staple and paper clip removal –
 - a. After all staples and paper clips are removed, the operator can "fan" the pages to check for staples that may have been missed and to insure good paper separation upon induction into the scanner.
 - b. If a staple remains, it will pull into the scanner all of the documents that are in that stapled group. Assuming multi-feed detect is turned on, the document feed will stop, but the attached documents will be at least wrinkled and more likely damaged by being torn as they are brought into the scanner.
 - c. Care should be taken when removing staples in order not to tear the page while removing the staples, but to remove them carefully to maintain the page integrity.
 - d. There is an "industrial" staple used to bind large piles of papers. There is a special tool to remove such large, heavy staples.
3. File and page details –
 - a. Keep pages in same order as they were in the file.
 - b. Keep the material arranged in such a way so that the documents can be re-filed into their respective files and/or folders. Sometimes it is easier to use two different color separators: one at the beginning of the folder (let's say: pink) and one in between each individual document from that folder (let's say: orange). This will allow for a more efficient un-prepping process (see below).
 - c. Flatten dog-eared corners of pages.
 - d. Unfold any folded pages.
 - e. Sticky notes will stop the scanning process, as they will be detected as a double-feed. Often they can be put to one side of the page, which will not be detected by the sensor (as the sensor is often in the middle). If there is no space to put the note on the side of the paper without covering up something on the page, it might be easiest to make a copy of the sticky note and scan the full-paged duplicate in front of the page it was attached to. (Client may prefer it behind. Always get clarification in writing.)

- f. If some pages are too large to be scanned in the conventional batch mode, insert them into the batch in a predetermined method so that the scan operator can see them in order to stop his batch scan, and then scan these documents on the flat bed (or whatever method is necessary).
 - g. If some of the pages need scanned in color or grayscale (when the norm for this task is B&W), insert them into the batch in a predetermined method so that the scan operator can see them in order to stop his batch scan, and then scan these documents in color / grayscale mode.
 - h. The prepper can also mark pages that need special attention with paper clips or binder clips that will force the scan operator to stop and handle the issue.
 - i. Spiral bound reports should have the binding removed.
 - j. Magazines and books may require hardware and training in order to cut the binding off of the material to facilitate scanning. (Care must be taken to insure that *only* the binding is removed and that the data is not removed.
 - k. Pages should be stacked up so that
 - l. Small pages (anything smaller than the norm in your task) may require copying to “standard” size in order to insure that they will be pulled in to the scanner un-skewed.
4. Prepped Orientation –
- a. In creating a batch, all pages are to be pulled from a pile that is oriented face-up. Remove staples, etc, as listed above, and then restack the pile face down in the same order. Because the paper gets placed into the scanner face-down, the page 1 that was on top of the un-prepped pile, will be page one in the prepped stack in the scanner.
 - b. Documents should all have their tops to the same side (i.e.: they should not be upside down to one another).
 - c. All prepped batches of documents should be arranged in such a way so that the tops of all pages are flush. Smaller pages may exist in the batch, but they must be aligned with the pull-edge of the standard size documents to avoid double-feeds, document damage, paper jams, etc.
5. Separation –
- a. All pages immediately following a separator sheet will be prepped and scanned as a unique document.
 - b. A new unique document will begin at the next separator sheet.
6. Notes on Paper Types –
- a. Fold outs – if an operator comes across a multipage fold out, it will likely need cut into separate pages, and the scan operator will have to rearrange the pages, as the back of page 1 will not be page 2.
 - b. Onion skin – very thin paper, prone to tearing and jamming in a scanner. Staples are difficult to remove without damaging.
 - c. Carbonless copies
 - d. Card Stock – very thick. Depending on the scanner it may require hardware setting adjustments. May also detect as multi-feed in some scanners.
7. Un-prepping (reconstituting the files)
- a. Take a batch for which the scanning is complete. They should be taken in order, as they should have been scanned (A-Z, 1-9999, etc). It should come into the un-prepper face up.
 - b. An un-prepper should only have one batch at a time and should un-prep it completely before starting work on another batch.
 - c. From the completed, scanned batch, separate the documents back into their folders/files by inserting the respective set of documents “as-is” back into its respective folder from the front (top) of the batch to the back (bottom), re-stacking the filled folders face down. (Again, pulling from the top of a face up stack and replacing as the bottom of a face down stack maintains the order.) Separating the batch into the respective folders prior to actually un-prepping the documents ensures that there are no excess folders in, or missing folders from, the batch.

- d. Starting from the top of an upside down pile, the operator would take the first folder and unprep the documents, in order, by starting at the back of the top folder and work to the front, stacking the un-prepped documents face up, laying the others face up on top of them. This will ensure that the documents end up back in the folder in the same order from which they started. This should be done accurately regardless of what the final disposition of the documents is to be.

Un-prepare by:

- i. Stapling back together the documents as they originally were prior to prepping.
- ii. Pulling out the separator sheets
- iii. Rebinding books
- iv. Stamping the outside of the folder "scanned". The stamp should always be in the same place on the folder: either at the front upper left or the front upper right. If something is printed on the folder on the right, stamp the left and vice versa. Placing the stamp in the same place every time ensures that when someone goes to check the folder, they'll see the stamp and not have to look all over the folder to find whether or not it's stamped.
- v. Re-file into drawers, archives, etc.

Note: In re-filing, a methodology must exist to ensure that new documents are not filed into folders that have already been scanned, nor that folders of unscanned documents are mixed with folders of scanned documents, as the unscanned folders may end up being overlooked.

Notes:

Most scanners have duplex capability, allowing both sides of a page to be captured simultaneously. Also, most scanning software packages allow for the configuration for deletion of blank page backs.

There are, of course, many methods that might work for prepping documents for inclusion into a scanning process, and this is but an overview. Being that there may be many possibilities, a manager must be vigilant in making sure that the prep and scan operators are not taking unacceptable shortcuts. As they may not always be aware of the previous or subsequent steps, they might not understand why a certain instruction should be carried out the way they were trained as opposed to the way they may see it. Training operators in previous and subsequent steps can help avoid errors made by taking unacceptable short cuts. It is also recommended that the operators be instructed to follow the instructions that they were given and ask if they believe they may have an alternative method of processing.

It may not be necessary to un-prepare documents, and there are various levels of un-prepare. It is recommended that at least the separator sheets are removed so that they can be re-used, but it may not always be necessary to re-staple the documents, re-bind the books or to re-file them into the cabinets (they may be heading for destruction). This should be specified in the statement of work. If it is not, inquiry should be made to see a) if it is necessary, b) to see if un-prepare was calculated into the overall processing budget (as it may take as much, or even more, time to un-prepare as it does to prepare).

Glossary:

Backfile conversion: Converting masses of existing legacy documents.

Batching: Processing large groups of documents with minimal intervention. (i.e.: 50 documents of 20 pages each, separated with bar codes, scanned in as one group)

Day Forward Conversion: The method that governs the processing of new documents coming into the system.

Duplex: Describes a double-sided document.

On Demand Scanning: Scanning documents from a backfile as needed, one or a small group at a time – as needed.

Simplex: Describes a single-sided document.