

On Site Scanning Projects – Helpful Tips

By

Manuel Bulwa

Integrated Scanning of America, Inc.

March, 2007

This article is a succinct summary of valuable lessons learned during large on-site capture projects. The very first lesson is to acknowledge that an end user has expertise in their core competency business, but not necessarily in scanning.

Times have changed. Transporting boxes containing documents in paper or micrographic form out of the client premises for scanning and publishing poses new kinds of risks and liabilities. Considering recent regulations such as HIPAA, Sarbanes-Oxley, and others, consequences may include financial burden, pain and suffering, death, jail time, jobs, or just unnecessary embarrassment. The war on terror and natural disasters add more reasons to get serious about it.

Staying in the dark ages is no longer an option, sooner rather than later we must digitize some or all of our records. We must be prepared to plan on who, how much, when, where and how.

Some of the options are:

All capture (backlog and day forward) happens wherever the documents reside. Nothing leaves the client premises.

The opposite extreme: an outside vendor does it all (backlog and day forward). Documents are transported back and forth.

Compromise: Backlog is outsourced, day forward stays home.

The best of both worlds: Off the many tasks involved in professional document capture, we perform some in house while a well equipped, professional service vendor performs the rest wherever appropriate. Only tasks requiring direct contact with physical documents stay home (prepping, scanning, rescanning, de-prepping and disposition). All other tasks (QC, post processing, barcode recognition, OCR, forms id, indexing, database publishing, image publishing, media publishing, hosting, retention and contingency management, backups, auditing, etc) are handled remotely. Let's call each of these respective sites: "*the client site*" and "*the plant*".

The underlying proposition of this paper is that the last option above is the most versatile and offers the most advantages. The paper also lays the foundation for certain methodologies and technologies designed to optimize performance while keeping quality standards. We will start by focusing on the backlog side of the equation first.

On site scanning presents many unique challenges, such as:

The client premises are unlikely to meet the ideal specifications for a smooth production project: Access to the facility, security, allowed shifts per day and days a week, electrical and cooling conditions, etc. The more hours per week we keep expensive, production-grade equipment idle, the more damage we inflict on the cost-effectiveness of the project

Documents are in daily use and scattered around many offices or buildings. The processes of inventory and check-in/check-out should be implemented in a way that:

Will not interfere with your daily use of the documents

Guarantees that scanned documents will make it safely back to where they belong.

Guarantees that no individual sheet or drawing is missed or scanned more than once.

Helps locating an original if needed for rescan or QC.

Provides a mechanism to control records moving around multiple phases.

Flow Control. It is imperative to capture the origin, custodian, container type (box, cabinet, book binder, stick, drawer, roll, etc) and external label of each container sent for scanning. We handwrite this data on our "Begin Box" barcode sheets while prepping.

Poor quality of certain documents. Our approach is to scan in color or grayscale. This allows special post processing techniques to turn images into bitonals more reliably and with additional flexibility.

Subs and temps. Posting our own staff may imply travel, lodging and other expenses. Supervising temps at a remote location without posting one or more of our staff on each and every shift, is a difficult and dangerous proposition. Temps know that there is no future or career on a limited time project, even if it lasts for months, and it shows in their dedication, attitude, ethics, initiative and passion (or lack thereof). It is not uncommon to have subs or temps performing on-site prepping and scanning for weeks or months. This creates a whole new

set of challenges, mostly related to inter-personal issues, supervision and control. We enforce strict on-line software control of scanning performance and quality, sometimes with video surveillance as well.

Monitoring: Delegating work to remote temps under limited supervision requires the use of special operation procedures and extraordinary auditing, QC, and error management methodologies. In one of my recent projects, all paper documents were scanned in duplex and color, so we would not be forced to trust human decision during prep or capture. At post processing we removed blank pages and turned into bitonals all images that did not contain color or grayscale photos

Error hysteresis: There will always be human and other errors, guaranteed. The longer the delay in catching these errors, the more difficult their remediation will be. For example, having to rescan a document that has already been moved to an archive storage location (or shredded). It is imperative that we implement both preventive and corrective error checks. Special preventive and corrective methodologies are needed to reduce the risk of human errors during on-site capture projects. For example:

Managing accountability and integrity:

Large drawings: Individually stick a small barcode on each and every drawing larger than B size. This offers a number of advantages besides strict integrity control, as it also serves as proof of scanning, makes possible efficient rescan and QC, and facilitates check-in/check-out. Unused or damaged stickers must be kept and accounted for at the end.

Other documents: You should use a Begin Box/End Box pair of barcoded target sheets to help assure a certain degree of batch integrity. More can be done by counting sheets or measuring linear inches.

Early error detection. We constantly monitor rates, sizes and special imaging metrics on-line. The key is to "catch" errors sooner rather than later, as they propagate very fast otherwise and become very costly to fix.

Mutilations: It is not uncommon to scan less than the whole original. Our solution is to scan everything at full size, especially drawings and aperture cards. More reliable cropping is then performed during post-processing.

Adjusting scan settings: Catching color on the fly while scanning is not as reliable as when scanning everything in color and let post-processing decide what needs to be bitonalized and what stays in color. For standard size paper documents, we capture all-color, full size, full duplex. Large format drawings are captured all-grayscale or color, full size. Microfiche and aperture cards are captured all grayscale, full size or strips

File naming, renaming, cutting/pasting, dragging/dropping, deleting: None of this should be allowed manually. It must be fully automated and an audit trail kept. Purging records to reclaim disk space is especially critical, and must be 100% under software control.

Scan retakes should not depend on operator file naming or numbering. We scan a special "Retake" barcoded target sheet followed by the rescan to eliminate human errors. Other handy barcoded target sheets include "Poor original" and "Operator Alarm" (wild card for scan operators to communicate unstructured messages to QC staff).

DVD production should not depend on people sizing, dragging and dropping. We use robotic publishers under strict software control to size, write and print DVDs. Images are stored encrypted, as DVDs are likely to be sent out for QC, post processing and indexing.

Manual tasks: Never allow operators to use Windows Explorer or similar to manually cut, paste, move or delete anything, at any time. It must be fully software driven and with audit trails.

What is meant to be sent out (to scan, QC, indexing, rescan, etc) is not necessarily what is actually received. Use redundant production logs to provide vital crosschecks. Use verified production logs to drive automated purging.

Naming folders and files after metadata is a bad idea: There are many reasons why this popular method backfires all the time. Consult with your vendor for better solutions.

Utilities: You will need software and help to ease your burden on tasks such as:

Acceptance, QC and reporting. If you want to perform basic cosmetic or other changes, make sure there is an audit trail.

Preliminary posting of selected documents to various servers and users.

Permanent posting onto your EDMS system. Be prepared to do and undo importations as the QC phase progresses. Import metadata such as container location, batch name, import date and media volume id.

Fortunately, for each of the challenges listed (and many more we did not list), a set of proven methodologies exists that makes on site capture safe and efficient. In traditional capture projects, service providers as well as end users prefer the "one-stop shop" approach, i.e. to do all that can possibly be done during capture, and let post-processing and QC determine what has to be rescanned. We have adopted exactly the opposite approach: do as little as practical during capture, while deferring as much as possible to post-processing. We capture paper and drawings in full color, full size, full duplex. We capture microfiche in raw grayscales, multiple rows, multiple frames. All sophisticated segmentation, color detection and conversion, image enhancement, etc happens at post-processing time.

We developed a number of specialized methodologies and techniques specifically for *On Site* projects. Some of these follow:

Preventive QC: Keep workorder files with vital data such as scanning resolution, duplex or simplex scanning, color, black and white or grayscale scanning, workflow inter-dependencies, etc. If a process catches a potential violation of workorder business rules, it can be preempted before mistakes turn into bad results.

Corrective QC: Create statistics for as many factors in the process as practical. Examples: Histograms, averages and standard deviations for image sizes, image widths, image heights, image densities, image colors, folders per batch, documents per batch, indexes per batch, index lengths, batches per volume, pages per hour, keystrokes per hour, rescans per batch, etc. Any suspects deviating from statistical expectations is flagged and queued for human or further automatic inspection. "Booby traps" do help keep the QC concept on its toes. Our PDA should wake us up in the middle of the night if images suddenly deviate from workorder specifications. This may be caused by an operator that at 3 AM discovers that scanning at 100 dpi is so much faster, and decides to show off his/her creative initiative. Or perhaps documents that need to be captured in color are inadvertently scanned in black and white or grayscale.

Operator behavior monitoring: This should include video surveillance cameras at the client site with IP replay at the plant, frequent surveys to client staff regarding their observations and comments on our temps behavior, and plain old surprise visits.

Productivity monitoring: Each production line has certain productivity goals set. For example, a paper scan station is expected to produce a certain amount of new images per hour. The system should keep track of all new image files produced, and if after a certain period of time this number falls below a preset threshold, a warning notification is sent to the supervisor assigned to this task category. Email notification is a good option. Monitoring and alert notification should be defined and implemented for human operators as well as automated processes. We can only hope that our PDAs will not behave like Kato in the Pink Panther movies.

Operating environment monitoring: Disks may be running out of space. DVD writers may have more and more frequent retries than usual. Scanners may be jamming too frequently or creating noise and lines. Operators may become "too creative" and use unorthodox procedures to show off how much they learned at their weekend community college classes.

Records and container management: Each and every solid object that moves through the working environment must be tracked down. Barcode stickers on boxes, roll film boxes, microfiche envelopes, etc are a very good option. Barcode guns and a strict workflow control are used to reconcile physical tracking and check in/check out with system data. Our technology includes an "external" workflow to monitor movements in the real world and an "internal" workflow of what the system perceives as happening. It feels good to know that an operator wanding a barcode in a box to prepare for returned shipment to the client will be alerted to the fact that that box has not been scanned, QC-ed and/or de-prepped yet.

Acceptance monitoring: Based on pre-established acceptance criteria, certain portions of the project may be bookmarked to produce results that can be periodically submitted for acceptance compliance.

Equipment maintenance: Preventive maintenance on critical units such as scanners must be strictly followed through and monitored. Loosely supervised temps may attempt to take shortcuts that can be very detrimental.

About Integrated Scanning of America, Inc:

Integrated Scanning of America, Inc. was established in 1993 as a highly automated, large volume scanning service bureau. With a track record of hundreds of millions of images digitized from paper documents, drawings, microfiche, microfilm, aperture cards, checks, X-Rays and bound books, ISA produced over two hundred satisfied clients including Government, Health, Nuclear, Industrial, Financial, Education and Services. ISA customizes production lines for large projects with emphasis in on-site capture. Other specialties include A/P mailroom automation and massive searchable PDF publishing.

ISA is headquartered in Newbury Park, CA and has a branch office in Las Vegas, NV dedicated to on-shore data entry and large format scanning.

MBE-WBE, 8(a). www.isausa.com