Look Back In Paper

...And then one day you find  
Ten years have got behind you…

-- Pink Floyd, “Time”

This article wraps ten years of Tip Sheets. I started back in October 2005 with my first Concordance Tip Sheet, as it was known at the time, because I felt that the Synonyms feature was too good to remain a largely undocumented secret. Subsequent experiences and questions provided more grist that I thought worth milling and sharing.

So in this reflective state of mind, I want to talk about an item from the past that is still with us: scanning. Construction and engineering cases with drawings, receipts and change orders, contract cases with handwritten notes on draft and signatures on final, all take us away from the relatively clean and densely populated commons of electronically stored information (a/k/a ESI).

Recent projects have reminded me that paper has not entirely gone away, but rust can gather about the skill set required to run and manage this ancient discipline. Bad scanning and ingestion can cause problems that cascade through the remainder of the discovery process, driving up cost and agita. It’s worth stepping through the procedure to recall the drudgery essential to properly digitizing paper files.

Digesting Paper

Before one can even consider unjamming the automatic document feeder, there is the matter of the medium to be scanned. Think about your paper files: there are pages stapled, clipped and crinkled together. There are Post-IT notes, signature flag stickers,
and sometimes plain old tape. There are letter, legal, photo, divider, folder and larger formats. There may be binders or bound volumes. All have to be accounted for.

Staging required disassembling documents by hand and inserting coded (and generally different-colored) slipsheets to represent the various divisions:

- first page in a box (typically about 2500 pages in a nicely packed banker’s box).
- first page in a folder or binder.
- first page of a physical document (clipped, stapled, bound, etc.).
- first page of a logical document within the physical document (letter, memo, fax cover sheet, etc.).
- page for each post-it, if these are to be captured as well.
- Any specially-sized item, book or photo may require manual glasswork.

The batch images may be assigned sequenced page numbers, and overall generally described with certain “metadata” in the scanning software, such as:

- Batch ID number
- Custodian or Site
- Box Title and/or Number
- Folder Name and/or Number
- Date Scanned
- Operator

**Bad Scans Lead to Bad Days**

Obviously, if you are scanning fourth generation copies of third generation faxes, the results won’t be pretty or very well handled by the optical character recognition (OCR) process. If the contrast is poor, papers are crinkled (and thus skew in the automatic document feeder) or, better yet, upside-down, the end-product image won’t be any better, and probably just a bit worse. The operator can tweak a few controls, but does not have much control over poor originals.

The slipsheeting process, though, is also vitally important: if you start with a binder-clipped physical set of papers containing a fax cover sheet, a two-page cover letter, and a sixteen page memo, your attorneys will want to see the distinct identity of each logical document (which are of different types, have their own page numbering,
and may have different dates, authors and recipients), and that they were clipped into a single physical set.

If not done correctly, the review and production can go awesomely wrong. Think about having a highly confidential memo with technical drawings of intellectual property. If this document is not unitized, that is, classified as a distinct “logical” document within a physical set, all we have is individual page images which would then have to be split into component documents through a coding process (which is not always done on cost grounds, putting aside the fact that reviewer eyeballs carry the biggest price tag in this process). If the slipsheet is quickly dropped in after a page with a lot of white space after the text, you might have a new “logical” document (starting at page 12) that contains confidential drawings. An attorney under time pressure might assume that those first 11 pages contain all the privileged and confidential material, and inadvertently produce this privileged caboose.

Sometimes, we see an entire box scanned into a single, non-searchable PDF file, unimpeded by slipsheets of any sort. This basically creates a huge, ugly thing that has to be processed (TIFFed, unitized, OCRed, and minimally coded) to be of any use at all. It’s even worse if the PDF is formatted with Acrobat organizational tools, which tend to interfere with processing software.

Output to Review (What’s Old Is New Again)

Once everything is scanned, the generally accepted output formats are:

- Single-page images (TIFF, or JPG for photo or color document capture).
- Document-level OCR Text with Begin Bates Number as filename.
- Load files for images (Concordance OPT/LOG is a common default, or Ipro LFP, sometimes still Summation DII).
- An alternative to the above is single-document searchable PDFs, particularly when producing to counsel with no access to review software.
- Numbering and data load file, which may include:
  - Begin Bates Number (Control number)
  - End Bates Number (last page of document)
  - Begin Attach Number (Begin Bates of first document in a physical set)
  - End Attach Number (End Bates of last document in a physical set)
  - Batch ID number
  - Custodian or Site
Any further information would require bibliographic (or objective) coding, where attributes of the document such as Document type, Document Date, Estimated Date (Y/N for a guess), Document Title, Author, Recipient, Copyee(s), and perhaps Has Handwritten Notes might be captured.

If the above looks familiar to anyone, your vintage is showing. Back in the late 1980s and through the 1990s, Dataflight with Concordance and Summation with LG and iBlaze pioneered the formatted, searchable display of document content and images. It was a novelty that took years to grab the mainstream, but more and more attorneys, faced with more and more paper, turned to these document review products as an equalizer.

Disclosure: back in the day I worked for a large law firm that engaged in what we lightly called “forklift production.” This referenced the practice of copying enormous volumes of physical files and sending a tractor-trailer filled with banker’s boxes to opposing counsel. The goal was to send an amount of boxes that exceeded the square footage of said counsel’s office.

Some years later, under the banner of our litigation support consulting shop, I was asked to write an index database for plaintiffs’ counsel in a large international civil action, to replace manual indexes and make it easy to look up and lay hands on vital documents during deposition. The app I wrote was used successfully, and I adapted it for another case before deciding, after seeing Summation, that I was not in the software development business. But in the meantime, I found that technology could be an equalizer in litigation. If you don’t agree, re-read the previous paragraph.

-- Andy Kass
akass@uslegalsupport.com
917-512-7503

The views expressed in this E-Discovery Tip Sheet are solely the views of the author, and do not necessarily represent the opinion of U.S. Legal Support, Inc.
U.S. LEGAL SUPPORT, INC.
ESI & Litigation Services

PROVIDING EXPERT SOLUTIONS FROM DISCOVERY TO VERDICT

- e-Discovery
- Document Collection & Review
- Litigation Management
- Litigation Software Training
- Meet & Confer Advice
- Court Reporting Services
- At Trial Electronic Evidence Presentation
- Trial Consulting
- Demonstrative Graphics
- Courtroom & War Room Equipment
- Deposition & Case Management Services
- Record Retrieval

www.uslegalsupport.com

Copyright © 2015 U.S. Legal Support, Inc., 90 Broad Street, New York NY 10004 (800) 824-9055. All rights reserved. To update your e-mail address or unsubscribe from these mailings, please reply to this email with CANCEL in the subject line.