

PLI Civil Justice Institute

Tools, Techniques, and Technologies for Efficient Litigation Management

March 15, 1997

Using Technology to Manage Documents

Ron Friedmann, Esq.
Director of Computer Applications
Wilmer, Cutler & Pickering
Washington, DC

Introduction and Acknowledgments

Reducing Costs and Improving Results with Technology

- I. Document management goals**
- II. Document management costs**
- III. New technology and procedures to reduce costs**
- IV. Improving results**

Document Review Software: Desirable Features

- V. Document review features**

- VI. Retrieval features**
- VII. Image printing**
- VIII. Document Coding and Scanning**
- IX. Productions**
- X. Report Writing**
- XI. Quality assurance features**
- XII. Remote access**
- XIII. General system and technical requirements**

Introduction and Acknowledgments

From 1989 through 1996, Wilmer, Cutler & Pickering developed its own software to manage discovery documents. We integrated image management, optical character recognition, full-text databases, bar code technology, and structured databases to create our own system. Parallel with creating our own software and systems, we developed new procedures for planning document management and for building and using databases.

Because of changes in technology and improvements in commercial off-the-shelf litigation support software, we changed our strategy in 1996. We stopped developing our own software and outsourced this function. Nonetheless, we continue to apply what we learned in how we manage documents.

The outline that follows on page 5 reflects some of the key lessons we learned over the last several years. Perhaps the greatest lesson is that the human element and human processes are more important than technology in achieving cost-effective results. The best technology does not help unless the litigation team plans for its use and changes the way it works. The suggestions in the outline sound easy enough but require planning, discipline, and new ways of working. Having the right technology is necessary, but by no means sufficient. Absent a commitment to change, new technology is likely to have only limited impact.

The feature checklist for document retrieval and review software that follows on page 8 reflects features that we believe are valuable in litigation support software. (A limited number of features intentionally appear under more than one heading.) Neither our own system nor any commercial system we evaluated has all these features. Although we believe that over time these features need to be incorporated, few cases require all of them at once. Thus it is always important to keep in mind the specific needs of the firm or department and the matter. Nonetheless, it may be useful to consult this list and compare how different products compare against it.

It is important to remember that technology evolves rapidly. The attached outline and checklist are a snapshot from a specific moment. We expect that new technology will lead to new tools, which will require new procedures. For example, software is emerging to code documents automatically for date, author, and recipients. Similarly, advances in full-text searching may allow credible first-passes at issue coding. Such changes may set the stage for further cost reductions and quality improvements. The promise is great, but it means that we cannot afford to become too comfortable with the way we happen to manage our work today, for surely it will change.

Acknowledgments

Our new approach to document management is the result of a collective effort. Technical staff, legal assistants, and lawyers at Wilmer, Cutler & Pickering have all played an integral role. I would like to thank the following current Wilmer, Cutler & Pickering personnel for their past and present support: John Burden, Olivia Cherry, Chris Chih, Joyce Cutlip, Peter Einaudi, Ed Macnamara, Yvonne McCoy, Yvonne Nearing, Lucy Pearson, Brian Sanfelici, Quang Tran, and Cindy Vide. Particular thanks and recognition go to my former colleagues Michael Pergola, Lisa Antel, and Nick Ryerson, now with ISIS Communication Associates, where they are pursuing many of the ideas they first developed at WC&P.

Reducing Costs and Improving Results with Technology

- I. Document management goals
 - A. Achieve the best outcome for the client
 - B. Minimize cost
 - C. Learn the facts of the case
 - D. Prepare offense or defense
 - E. Fulfill obligations to produce documents

- II. Document management costs
 - A. Build database
 1. Variable costs
 - a. Identify every page with a unique number
 - b. Define documents (which pages constitute a document; which documents are attachments)
 - c. Code objective information such as date and author
 - d. Scan and OCR (optical character recognition) each page
 - e. Quality check a statistically controlled sample of all work
 2. Semi-Variable costs
 - a. Computer hardware and software
 - b. Project management
 - B. Use database
 1. Variable costs
 - a. Lawyer or legal assistant time to search for and review documents
 - b. Legal assistant time to retrieve, print or copy, sort, and re-file documents
 - c. Printing or duplicating costs
 2. Semi-Variable costs
 - a. Computer hardware and software
 - b. Project management

III. New technology and procedures to reduce costs

A. Build database

1. Reduce the number of documents that require processing
 - a. Conduct a box or folder level review to assess significance and relevance at an aggregate level
 - b. For boxes or folders with *relevant* documents, process each document
 - c. For boxes or folders with *irrelevant* documents, build a database index of box or folder level information (which can be used subsequently to help find documents not processed at the document level)
2. Eliminate Bates stamping
 - a. Manually apply bar code labels or automatically affix a bar code during the scanning process. Either approach can include human readable numbers as well.
 - b. If every page has a bar code label, then lawyers and staff no longer have to type multi-digit numbers, a time-consuming and error prone process. Instead, they can use a bar code wand to enter the number automatically.
3. Substitute computer effort for human effort in the coding process
 - a. Limit objective coding to a minimum number of fields
 - (i) Even if OCR accuracy is as low as 60%, it is often possible to surface relevant documents with a full-text search
 - (ii) Key fields to code are date, author, recipients, and marginalia
 - (iii) Think carefully about how document type is coded. If there are more than a few choices, a surprising degree of judgement is required to distinguish among them.
 - (iv) Avoid any coding from document text (e.g., names mentioned)
 - b. Take advantage of database controls, picklists, drop down boxes, and other features to minimize coding time and decrease the likelihood of errors

B. Use database

1. Reduce lawyer review time
 - a. Full-text searches save time
 - (i) Find documents pertaining to issues without having to look at each and every document every time a new issue emerges
 - (ii) Refine searches of coded information so that lawyers review fewer irrelevant documents
 - (iii) Find documents with particular names mentioned in the text
 - b. Capture results of lawyer reviews electronically so that they can be re-used
 - (i) Lawyers should work on-screen and record information pertaining to privilege, responsiveness, issues, and comments
 - (ii) On-screen review of images facilitates this process but is not essential. Use a 21" monitor to display image on one-side and data input form on other side

- (iii) If working from paper, lawyers should nonetheless enter any notes about documents on-screen. They can use a bar code wand to retrieve the associated database record and type comments or code issues using on-screen forms.
- c. Use this information as the case progresses
 - (i) Searches can be refined based on this information, resulting in fewer documents to re-review
 - (ii) Privileged documents can be segregated and a privilege log automatically generated
 - (iii) Responsive documents can be printed or copied to cd-rom for productions

- (iv) Issue binders can be created automatically by printing only those documents coded for an issue
- (v) Documents whose comments contain particular words can be retrieved and printed
- d. As lawyers re-review documents for various purposes, as is almost always necessary, less time is required because the subsequent reviewers have ready access to the judgement of all prior reviewers
 - (i) If working on-screen, subsequent reviewers can view coding on-screen
 - (ii) If working from printed copies, both the objective coding and lawyer coding and comments can be included on a cover sheet preceding each document.
- 2. Reduce clerical time handling documents
 - a. Printing from images saves approximately three minutes per document of clerical time required to find, copy, and re-file each document
 - b. Printing from images also eliminates the need to sort documents (a time-consuming process). Documents can be printed in whatever order they need to appear (e.g., chronologically by witness)
 - c. Working with images eliminates the lost document problem and hours spent looking for missing documents
- 3. Reduce administrative effort
 - a. Generate automatic indexes of documents produced
 - b. Automatically track how each document has been used

IV. Improving results

- A. Searching on lawyer coding and using full-text searches increases the likelihood of finding relevant documents and therefore decreases the chances that an important document will be missed
- B. Maintaining strict database control and rigorous procedures helps ensure that duplicate documents are handled appropriately, reducing the chances of inconsistent productions and the sanctions that may follow
- C. Maintaining history of how documents are used (e.g., to whom proffered, to whom produced) can help understand opponent's strategy and what they may know

Document Review Software: Desirable Features

[Note: Italicized items are of particular interest to lawyers]

V. Document review features

- A. *Users should be able to code for*
 - 1. Significance (with flexibility to define degree of significance)
 - 2. Responsiveness (with flexibility to designate for a particular production request)
 - 3. Privilege (with flexibility to define multiple types and to which production the privilege applies)
 - 4. Issues
 - a. Hierarchic and collapsible issue outlines with check-mark box next to each issue for “one mouse-click” issue designation
 - b. Ability to modify issue outline quickly and easily on a global basis, for all users
 - c. Ability to customize which issues appear on-screen for each user
 - 5. Comments
 - a. Pertaining to entire documents
 - b. Pertaining to specific issues only
- B. *Document tagging (either in a hit list or when viewing a document or its coded information) for*
 - 1. Printing in a batch job
 - 2. Applying same issue code or comment to multiple documents
 - 3. Creating a new, re-usable list of documents to be used subsequently
 - 4. Including the newly tagged documents in an already existing list
 - 5. Including in a personal notebook of documents
- C. *Conduct reviews on-screen from images or from paper*
 - 1. If from paper, be able to synchronize on-screen information by using bar code wand to select appropriate database record
 - 2. Interface should have option for check-boxes, pull-down lists, or radio-buttons; should not require typing text except for comments
- D. *Create customized notebooks of selected documents*
 - 1. Database should track that a document has been placed in notebook
 - 2. Changes in document data should automatically flow through to all notebooks
 - a. In real-time for all users on the LAN
 - b. At next synchronization for remote users
- E. *Performance criteria*
 - 1. Images must appear on-screen in under one second after pressing “next page” or “next document”
 - 2. Coding should require a minimum number of keystrokes, preferably by radio buttons or checklists
 - 3. Changing a field value should not require noticeable wait time for a database update
- F. *Customize interface*
 - 1. Select and size columns/fields on hit lists
 - 2. Option to view any and all of coded data, document images, and OCR full-text; option to keep the three data types synchronized while navigating through any one of them

- G. Image viewing and use
 1. Thumbnail sketch of document (with user defined number of images)
 2. Print all or some of current or selected documents
 3. Usual rotation, inversion, and zoom features
 4. Cut and paste all or some of image into other applications
 5. Toggle between image, text, coded data OR have all three on-screen at once
 6. Features to redact and annotate documents
- H. Image annotation, highlighting, and redactions
 1. Annotations and highlighting
 - a. Annotations or highlighting anywhere on image. Specific colors for annotation flag and highlighter should be assignable to named users
 - b. Annotations should be full-text searchable
 - c. Software should be able to search for annotations and highlights and display the specific page (as opposed to the first page of the document)
 - d. It should be apparent when viewing any part of a document that there are annotations somewhere in the document and there should be a feature to navigate to the next annotation
 2. Redactions
 - a. Users should be able to redact images
 - b. System must keep audit trail of who makes each redaction and when it was made.
 - c. Option to enter a comment specific to each redaction or specific to all redactions made in a single session
- I. Exhibit management
 1. Synchronize original document numbers with exhibit numbers across depositions
 2. Jump links from exhibit references in transcripts to document images (software should automatically insert the jump links)
- J. Navigation features
 1. Scroll through all images in a hit list without needing to return to the hit list
 2. Navigate directly to any image (quick find feature)
 3. Move from hit to hit within a text search
 4. Option to view adjacent documents in database (although not on hit list) and then return to hit list
- K. Export review set
 1. By e-mail or fax
 2. To other applications
- L. Document review administration and technical features
 1. Audit trail of all coding changes (user, date, and change)
 2. All on-screen coding should update the database in real-time
 3. Feature to update globally all values of a coded item
 - a. Option to change default value of uncoded records only

4. Option to generate productivity reports showing documents reviewed by user, per day

VI. Retrieval features

A. *Search techniques*

1. Full-text searching of document contents
 - a. Boolean
 - b. Natural language
 - c. Relevance ranking
 - d. Thesaurus capabilities
 - (i) Built-in thesaurus or semantic network
 - (ii) User defined thesaurus for case-specific terms
 - e. Fuzzy searching and Soundex (techniques to search for similar sounding words and possible misspellings)
2. Structured database searching
 - a. Be able to access box-, folder-, or document-level objectively coded data (e.g. source, date, or author)
 - b. Be able to access any lawyer coding (e.g., issues, privilege, and responsiveness)
 - c. Simple front-end for beginners; more complex interface for advanced users
 - d. Fast full-text searching of one or more fields within the structured data (useful for searching comments)
 - e. Fully SQL compliant
3. Combine structured and full-text searching
 - a. Be able freely to mix and match searching across coded data and full-text of documents.
4. 'More like this'
 - a. Option to query by example, that is, submit a document as a search to find similar documents

B. *Search management*

1. Save prior searches
 - a. Annotate searches
 - b. Re-execute at will
2. Conduct searches within searches
3. Retrieve and report on all prior searches

C. *On-the-fly definition of unit of retrieval*

1. Search on Q&A pairs for depositions
2. Family, document, or page for documents
3. Transcript and/or exhibits for depositions

D. *List management features*

1. Combining lists
 - a. Intersection

- b. Outersection
- 2. Managing lists during searches
 - a. Select and size fields to display on hit list
 - (i) Be able to adjust selection while viewing hit list, not just in an administrative module
 - b. Sort hit list on any field
- 3. Manage document families
 - a. Option to gather document families

VII. Image printing

A. *Features*

- 1. Option to print with one or more sets of redactions
- 2. Option to print customized legend (e.g., Protective Order or Privileged) based on coded information
- 3. Option to print with or without families
- 4. Track each print occurrence and require user to input purpose of printing
- 5. Print document separators
 - a. On colored sheets (requires printer with two paper drawers)
 - b. With customizable selection of coded data
- 6. Print optional, customizable table of contents for print job
- 7. Option to replace original document numbering with new, sequential numbers, at print time
- 8. Automatic stapling

B. *Handling privileged documents*

- 1. Automatic on-screen warning if print request contains any privileged documents
 - a. Option to generate on-screen or printed report
 - b. Indicate type of privilege for each document
 - c. If overridden, keep audit trail
- 2. Options to include privileged documents
 - a. Apply Privilege watermark
 - b. Print from separate tray (on different color paper)
 - c. Print a page indicating the document is privileged (with custom control of which coded fields appear)

C. *Print administration*

- 1. Capture information about each print job (e.g., who ordered, purpose of printing, client and matter)
- 2. Generate monthly, system-wide reports for cost tracking

D. *Technical considerations*

- 1. Use industry-standard image format

2. Software de-compression
3. Drive printers at rated speeds
4. Running high-volume print jobs should not appreciably affect image viewing performance.
5. Support multiple printer brands

VIII. Document Coding and Scanning

- A. The software for production coding and scanning does not have to be fully integrated into the document retrieval and review system
- B. But the retrieval software or an associated module should allow easy updating of coded records and adding small batches of new documents

IX. Productions

- A. Be able to tag hit lists for productions
- B. Option to prohibit inclusion of privileged documents
- C. Generate appropriate reports
- D. Option to replace original document numbering with new, sequential numbers, at print time
- E. Option to print (or copy to cd) with one or more sets of redactions
- F. Tracking
 1. Track at a document level for each production or proffering
 2. Option to track by page level and redaction set

X. Report Writing

- A. Easy to use custom report writer
- B. Ability to add true footnotes
- C. Output to
 1. Paper
 2. MS Access or other standard format
 3. CD-ROM

XI. Quality assurance features

- A. Reports for gaps, overlaps, and duplicates
- B. Audit trail for all changes to data

XII. Remote access

- A. Features for mobile users
 1. Tag some or all documents needed while operating remotely
 2. Automatically move all relevant data to portable media

3. Re-synchronize upon return to office
- B. Network users should see flag that document is also on a remote system and who 'owns' remote copy

XIII. General system and technical requirements

- A. Type of data to handle
 1. Discovery documents on paper
 2. Deposition transcripts
 3. Marked exhibits
 4. E-mail databases
 5. Word processing documents produced in electronic format
 6. Other documents in digital format
- B. Technical features
 1. Can run on Win 95 or Win NT
 2. Run on Netware 3.x and 4.x or Windows NT server
 3. Open architecture (no proprietary image file or database formats)
 4. Relational database design
 5. Optical Character Recognition (OCR) requirements
 - a. Ability to swap OCR software as needed
 - b. OCR software should have a case-specific spelling engine to help improve accuracy
 6. Flexible data storage systems
 - a. Store images on magnetic or any optical media
 - b. Caching and staging system for images stored on optical
 - c. RAID arrays for full-text and structured data
- C. Administration of multiple cases
 1. Set up cases
 2. Control security by user and by document
 3. Run multiple cases separately
 4. Control how review fields are used (e.g., ok to add new names or not)