

A NEW PARADIGM FOR WORKING WITH ELECTRONIC DOCUMENTS

For years, litigators have relied on a system of scanning and sequentially numbering individual document pages, extracting the text electronically and producing single page TIFF files as the standard method of working with electronic documents in the discovery process. But what if there were a new way of working with electronic documents that could reduce the costs of that process by as much as 65%?

Attorney Browning Marean of DLA Piper, a well known proponent of litigation technology, thinks that cost savings are the only way to make this new paradigm attractive to attorneys. He feels that this perspective goes all the way back to the days of the law school civil procedure class, where professors were fond of citing the English case that first established the awarding of punitive damages with a reference to a poem by Alfred Lord Tennyson: “the jingling of the guinea helps the hurt that Honour feels”.

And beyond immediate cost savings, what if this new paradigm could give you other features such as:

1. Cheaper native file production
2. Reduced TIFFing costs for production sets
3. Eliminating the risk of reassembling documents incorrectly, or with wrong attachments
4. Review of data absent in imaging (formulas, review notes, track changes, speakers notes)
5. Earlier access to documents in the review platform
6. Up to 90% time savings in the processing stage
7. Drag and drop for early case assessment
8. Compatibility with file-based analysis
9. Conceptual search and clusters
10. Email threading
11. Near de-duplication
12. Document comparisons

1. THE “DOCUCENTRIC” APPROACH

The process of uniquely identifying each page, which we refer to in this paper as the “pagecentric” approach, is called Bates numbering in reference to Edwin G. Bates who patented the original document numbering machine. The Bates Manufacturing Co. was eventually acquired by Edison Phonograph Works and their machine so dominated the market that numbers imprinted on multi-page documents became generally referred to as Bates numbers.

Used sporadically for years by lawyers, Bates numbering became more widespread in the 1980’s with the rapid increase in the number of lawsuits involving large numbers of documents which was fueled by the growth of word processors and personal computers. The marking of each document with a unique

alphanumeric identifier helped track discovery responses, refer to specific pages of a document during depositions and even better track which documents were withheld on the basis of privilege or work product.

The routine method of handling documents in large cases became the following: 1) copy the documents, 2) manually stamp Bates numbers the copies, then 3) copy the Bates-numbered pages. Indeed, *The Manual for Complex Litigation* published by the Federal Judicial Center (The Manual) specifically states at Section 11.441, Identification Systems,:

Counsel should be informed that consecutive numbering is usually the practicable; blocks of numbers are assigned to each party in advance to make the source of each document immediately apparent. Every page of every document is Bates-stamped consecutively. The document's number may be later used to designate it; if the document is identified differently in the course of a deposition or on an exhibit list, the stamped number should be included as a cross-reference. If other means of designation are used, no designation should be assigned to more than one document, and the same document should not receive more than one designation unless counsel have reason to refer to different copies of the same document. In multitrack depositions, a block of numbers should be assigned to each deposition in advance. To avoid later disputes, a log should record each document produced and should indicate by, to whom, and on what date production was made. A record of the documents produced by a party and copied by an opposing party may also be useful.

However, this paper based page centric paradigm became problematic with electronic documents. As The Manual continues to state in Section 11.441:

However, databases containing millions of data elements, none of which are meaningful alone, can be difficult or impossible to break down and organize in a way directly analogous to conventional document collections. Special consideration should be given to their identification and handling.

As computer systems came to routinely use electronic images of documents which were indexed in computerized databases, electronic Bates numbering was instituted and the page-centric paradigm continued. (long sentence – break up?) Each separate page was scanned into a single page TIFF format, electronic Bates numbers were added during scanning and that electronic Bates number was used to link each image to a specific record in the database.

If a record referred to a multi-page document captured in multiple TIFF images, the software creating the images generated a “load file” which specified the range of images by beginning and ending Bates number for the document. The advantage to this process was that one could immediately locate any page of a document that had thousands of pages by referring to the Bates number and thus introduce only those pages needed as exhibits at a deposition or trial, rather than introducing the entire document and making the witness page through it to the needed pages.

This worked fairly well at first but was most efficient for relatively small amounts of pages of traditional legal documents. With the advent of electronic documents and new document types such as multi-page TIFFs and PDFs, emails, excel spreadsheets and audio files, the page-centric approach quickly falters.

In addition, the preference under recent federal and state rule changes for using native files in productions makes Bates numbering problematic, since native files cannot be Bates numbered. The alternative employed by many e-discovery vendors is to generate TIFF images from the native files and Bates number those images. But this process complicates native file review, and at anywhere from \$0.08 -\$0.20 per TIFF, also adds considerable cost to the process.

Why is this true? I asked John Turner, Senior Vice President and Chief Technology Officer of Anacomp, Inc. in San Diego, who stated, “To fully understand why, I need to go back to the start of the discovery process at what the EDRM project calls the “Processing” stage. At a high level this stage can be broken down into the following steps: (1) Data is received from the customer; (2) This data is culled and deduplicated; (3) The metadata is extracted; (4) The text is extracted; and (5) The document is TIFFed. (This can be done as single page or multi page, but is usually single page.)”. He continued, “One consequence of this is that the relationship of the pages to themselves and to the document is artificially broken. It also breaks the relationship of an email with its attachment or of a document with an embedded file, or ZIP file. All of them then have to be recreated in the review platform.”

“For applications that live in the page-centric paradigm, the load file is essentially a large flat file that ties everything together from the “key” of a page. They are optimized to work with this paradigm. “

“But most new applications, including our CaseLogistix, use a relational database which stores the data about a document in multiple tables. To load single-page TIFFs into a relational database involves a lot of additional work. CaseLogistix extracts all the metadata, as well as the text, from a native file as part of the load process and also establishes the relationship between documents and their attachments. We are then effectively running the documents through the same procedure a second time.”

Simply put, a docucentric data model can eliminate the text extraction and TIFF steps from the processing stage and cut the cost of that process in half. It also allows the documents to be made available in the review platform much faster – as TIFFing often accounts for as much as 90% of the time to process. It also allows early case assessment without any processing by simply dragging and dropping a native file or a PST straight into the application – something that simply cannot be achieved with the page-centric batch process.

The problem is that most attorneys and corporate law departments use legacy applications and service bureaus that are still mired in the page-centric paradigm. Some of these are trained to think in terms of single pages of paper containing data and some have a business model based on that concept. The latter don’t want to cut costs in half so they continue to push single page TIFF as the best means of processing even though a well designed docucentric application is just as effective at finding documents.

Relational databases allow for one- to-many and many-to-many relationships which support advanced features and functionality, and also compatibility with external engines for tasks such as de-duping and concept searching. New applications which support these functions, such as Equivio, Recomind and Vivisimo, are all document centric and will not perform in the old page centric environment.

2. SINGLE BATCH PROCESSING

Applications that use the new docucentric model can eliminate the batch transfer that currently is part of the workflow between processing, review and production. This process (see Diagram 1 below) increases data storage due to the need for data replication in the transfer process and is also prone to a high rate of human error. And as is the case in classic “lean manufacturing” processes, elimination of the time that inventory (in this case, electronic data) is stationary will eliminate overall cost as well as reduce production time.

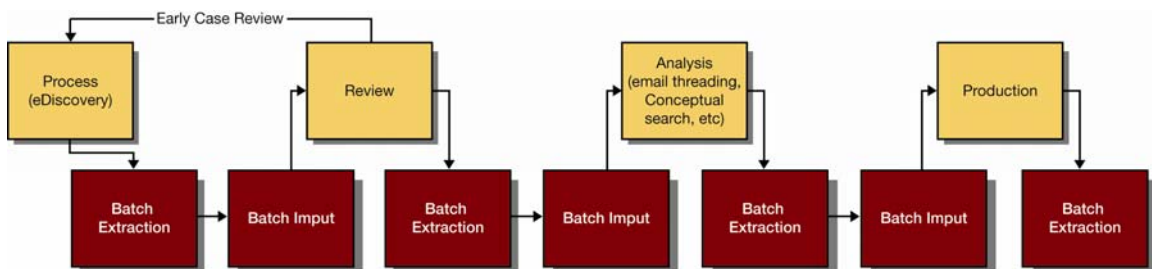


DIAGRAM 1: BATCH TRANSFER WORKFLOW

An excellent example of this approach is the partnership between Recommind and CaseLogistix. Axcelerate eDiscovery from Recommind uses a conceptual search engine to automatically cluster documents around elements such as key phrases, concepts, people, or even email categories like sender/recipient, /cc/bcc or date. With automatic culling and filtering of both structured and unstructured documents, Axcelerate can prioritize and batch assign documents for review based on their substance regardless of keywords.

A firm responding to a litigation hold letter can use Axcelerate as a specifically designed automated First Pass Review tool. According to David Baskin, VP of Product Management at Recommind, this process “improves review accuracy and consistency which drastically expedites the review process. It helps attorneys gain insight into a document collection before review has even begun while insuring that all documents reviewed and the legacy workflow is maintained.”

This process is taken one step further by adding integration with the CaseLogistix review tool. Using the docucentric approach, documents can be moved between the two applications as the need arises rather than in large batch transfers. This integration creates an easier, faster and more cost-effective e-discovery solution than previously existed and which allows for a much quicker and simpler review organization. See Diagram 2 below.

Another example is the integration between IPRO eCapture and CaseLogistix, which eliminates batch transfers between the applications and the associated human and system overhead. Users will be able to load large sets of unknown documents, files, and materials more rapidly and in a completely native workflow, eliminating the costly and time-consuming process of converting native files to TIFF images and extracting the text

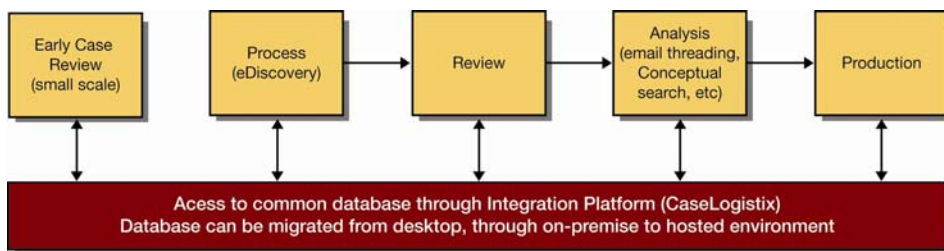


DIAGRAM 2: TRANSACTIONAL DATA TRANSFER WORKFLOW

3. EXPANDING THE MARKET THROUGH PARTNERSHIPS -- NOT ACQUISITIONS

Working with partners to provide the best technical solution is preferable to any attempt to build such a comprehensive solution. Since no one application can or ever will perform all steps of the EDRM process, it is clear that combining applications is necessary. That approach is best implemented in a partnership model, since the time to develop the necessary applications would be overly lengthy and no one company can be expected to possess a sufficiently high skill set at each step of the process to make the development of a best of breed single application possible.

Acquisition of the applications with that level of sophistication would be time consuming as well as extremely expensive for most -- if not all -- of the individual players. Additionally, such a series of acquisitions might lead to pushback from users and heightened scrutiny from government regulators as both would be cautious in embracing such a series of acquisitions.

An excellent example of this change in paradigm came from IPRO Chief Operating Officer Rich Ruyle when announcing such a partnership with CaseLogistix. He said: "One of IPRO's key strengths is identifying leading products in the legal market and ensuring that our software works seamlessly with them. This partnership brings two powerful products together in a way that gives end users a more complete, end-to-end approach to their e-Discovery needs."

"Integrating IPRO eCapture with CaseLogistix streamlines the discovery and review process, increases efficiency, reduces costs and minimizes risks," said John Turner, Anacomp's Chief Technology Officer. "This deep technical integration is the first step in an alliance that will benefit our mutual customers by delivering greater value and solving complex litigation challenges."

Finally, it is important to note that partners are more likely to be honest in providing feedback about the integration of two products than are the employees of companies that wish to be acquired or have just been acquired. The latter are more prone to downplay drawbacks to their system for fear that they may kill a potential sale or even lose their jobs.

4. CONCLUSION

According to John Turner, applications "...that retain the paradigm of the page and not the document will not be able to adequately support the new age of litigation. A modern platform must be able to review native documents that are not just paper equivalents, but must directly enable review of any file that is in common use in business today." Attorneys and their clients who focus on the new paradigm will save time and money by using this process for docucentric native file review.

The future belongs to these new technologies, where native files are processed without the need to convert to TIFF and are identified by their unique hash algorithm. "For litigation support professionals, embracing the new paradigm will be critical to providing best practice advice to attorneys and will also be essential to how they position themselves as industry leaders in this new era.," said Jeff Isenberg, Managing Director at Cobra Legal Solutions.

In today's world of vast quantities of electronic documents, the page-centric days of the Bates stamp are numbered.