

Rule E-Discovery Within Your Company

By Bill Johnson, President and CEO, TCDI

Let's say you are CIO/head of legal technology/technical staff. You've just finished getting your company's electronically stored information (ESI) ready for discovery, and are reviewing the costs of the project. One thing's certain: the bill from the e-discovery vendor you chose is an eye-opener. As a technologist, you just don't understand why such a simple process can't be done for a lot cheaper in-house. After all, this is a litigious society, and you know this isn't the last time you're going to have to do this. Next time you swear you're going to be ready.

Now you are general counsel/director of litigation/legal staff. As visions of the impact of the new Federal Rules dance in your head, and you consider the endless stream of seminars and advertisements focused on products devoted to e-discovery, you feel the need to gain control over the process. Costs of litigation are spiraling out of control, and it burns you to write the checks paying for the full freight of processing with every single lawsuit. You've convinced yourself that if your company can bring elements of discovery in-house, you'll see major cost savings down the road. After all, there's so much software out there, each promising reduced review costs and increased control over your company's data. Bringing e-discovery in-house cannot be a bad thing, can it?

Cool, let's get started... before the next budget meeting. It all sounds so simple. Bring the processing of ESI in-house, where your internal IT staff can add it to their list of chores, and hit the cost-savings jackpot. After all, it's just a simple matter of programming (SMOP), isn't it? It's about time IT starts helping the legal department, after all, it's IT technology that started these problems. Remember how easy things were before email, when we only dealt with paper? In reality, the technology involved in the actual processing of the data is pretty simple. That doesn't mean, however, that the job is a simple one. SMOP is not the answer here. It takes a disciplined system-design approach. The process is fraught with pitfalls if you don't enter with your eyes wide open and stay aware of the dangers. The price of screwing up can go well beyond the cost of any hardware, software or implementation project. It has the real potential to endanger your company's legal standing. So

what do you need to watch out for? There are several characteristics of the process upon which you have to have a firm grasp.

Know What You Are Getting Into

There's a LOT of moving parts in this process. You'll be dealing with a number of steps, including:

Inventory. Step one of any process. Make a record of all of the ESI you're processing. Mark the easy stuff to ignore, such as system files and other known program and non-relevant files. Start the audit trail.

"Costs of litigation are spiraling out of control, and it burns you to write the checks paying for the full freight of processing with every single lawsuit."

Metadata extraction. This is the big step. This is where data about the ESI objects are extracted from the objects themselves. Data includes date and time stamps, email address information, text and other



Bill Johnson

Bill Johnson is co-founder of TCDI, and serves as the company's president and CEO. He oversees the management of TCDI's Litigation Technology Division which provides technology platforms and services for the legal community with

specific areas of focus in large volume e-discovery and case management systems. Johnson continues to play a major role in the design of software and processes used in the company's litigation support products and services. He oversees efforts in database, system architecture, and application interface design and implementation on the various active client systems in place.

data about the data. This information needs to be extracted carefully to maintain the original information intact.

Imaging. If you're doing everything in native format, including producing, this step is skipped. However, converting ESI to image files will eventually be needed if you're reviewing in image format, performing redactions or will be producing in images. Sounds simple, but this is by far the step that causes the most problems. For example, converting a spreadsheet involves lots of decisions about the formatting of the spreadsheet. Most spreadsheets these days are never meant to be printed, but are passed around and reviewed directly in Excel. Printing such a spreadsheet in a legible format becomes pretty complex... try printing a complex spreadsheet and get all of the information out without playing around with the formatting... then do it five million times. Don't forget about hidden cells and imbedded graphics.

Filtering, de-duplication, advanced metadata and more. This is where a lot of the newer data culling technologies can come into play. Using either keyword-based technology (old but understood) or newer conceptualization technologies, your ESI is marked as subject for review. In addition, duplicates or near duplicates are identified so that identical records only need to be reviewed one time. You may also be performing automated Bates numbering or doing further processing for email thread identification.

Moving the data-to-review. Having processed all of the data, you now have to make it available for review. This typically means exporting the data to a review/production system, either in-house or with a vendor.

Now compound this overall process by having millions of discrete pieces of information moving through your system. To

make things more fun, you may have a “rolling” collection, where you’re processing documents after review has started, sometimes for years at a time. And remember, there’s a distinct possibility that months down the road, you’ll be asked to explain exactly how document number 3452304 came into being, and certify that the image you created and produced is an accurate representation of the document, and the list of recipients of the email to which the document was attached is complete.

CYA...Build a Solid Foundation

So how will you sleep at night? After all, the last time you installed a system capable of tracking this much information was that ERP system (and, boy, was that install a breeze!). Here are some guiding concepts:

Track, track, track. As you see from above, there are a number of steps. The inventory step should establish a data record for each object to be processed. Every step along the way should update that record to document what was done to each object. This means this is a rather complex database project, and is where the complexity of the process overrides the simplicity of its steps. At the end of the process, you need to be able to pull up the history about any given object, and easily see what happened to it and when. Yes, you’re developing meta metadata (that is, data about the data about the record) but that’s the price of doing this job correctly. If a part of the process involves an outside vendor, then the export and import of the data to and from the vendor need to be minutely tracked. And insist on the same level of tracking on the part of your vendor.

Audit, audit, audit. Auditing has to be a full-time responsibility. The entire process needs to be periodically reviewed and evaluated. Each step needs to be audited separately against a set of standards that are developed as part of the system design. The entire system needs to be audited to make sure that the handoff of data between the steps is accurate, and that all the data is making it through the system.

Validate your system and its pieces (preferably independently). Prior to launching your system, develop a comprehensive test and validation plan that verifies each piece of data created by your system. If you have the time (and you should if you plan accordingly), have this validation done by an independent organization who doesn’t have skin in your success. Validate each step independently, as well as the system as a whole. This independent validation will give you confidence in your system, and may help if any of your data is ever questioned. Unfortunately, data processed in-house may be looked upon with less reliability by some parties without some

independent backup. And of course, document, document, document.

Be controlling. You want centralized, automated control of the entire process. Unfortunately, in many organizations where the process isn’t properly controlled, a batch of email may be inventoried by one group, metadata extracted by someone else, images created by another and so on. All the while, the flow of data is communicated by email messages. Errors may or may not be properly accounted for. At some point, the data show up ready for review and everybody is happy until...

***“If you build your
ESI processing
infrastructure
correctly, you will
be in a position
to take advantage
of the emerging
technology.”***

You want to be able to control your entire process from a central console, and be able to determine where in the process any piece of data is. Preferably, every step of the process is able to communicate with a central control, relating information about the current step, its associated status and any exceptions. Ideally, this central control will allow you to launch a set of jobs that move your ESI from start to finish with discrete, detailed auditable data.

Be prepared to go big. Your initial project may be of limited scope and with a very reasonable deadline. I’d give decent odds, however, that this comfortable set of circumstances won’t remain. Take time to design your system to know your exact throughput capabilities, and have a plan for expanding the capability quickly without losing the integrity of your process. The last thing you want is to be stuck having to resort to an outside vendor to process data beyond your capacity. This will complicate your

tracking and auditing tasks, and cause a loss of your process control and cost advantage. So, design-in systematic growth capabilities from the beginning.

If you build your ESI processing infrastructure correctly, you will be in a position to take advantage of the emerging technology. If you’re a technologist, this is the sexy part of the exploding interest in e-discovery.

Make Good Decisions

The emphasis in e-discovery software includes developments in the area of advanced search technologies, and automated extraction of concepts and categorization of documents to aid in the document review process. Included in this arena is the detection of email threads, the grouping of related documents and automated/suggestive document coding. It’s fun to be at the leading (bleeding) edge of technology if you’re a software professional, but it may not be as much fun if your job is to worry about legal risks to your company. So how can you effectively make choices among the emerging technologies?

Don’t get tied down to any one technology. Look for technology that can be built into your processes, not ones that require you to ditch your processes. You never know when something better will come along, or if the technology you pick proves less than advertised. Make sure “changing your mind” or “upgrading” is designed into your approach. Nobody has the technology nailed completely at this point in time so don’t get nailed as a consequence.

Be able to evaluate quickly, and with your data. Demos are great, but they don’t mean anything unless you can test drive with your data. Build a process where you can insert a test of a new software program without disrupting your production flow of data. This allows you to evaluate technology in the proper context. The emerging technology holds a lot of promise, but you need to be able to move quickly to make sound evaluations.

Build your ESI processing infrastructure correctly, and you’ll be better positioned to control your existing e-discovery world and help it evolve in a controlled fashion to take advantage of all the new cost-saving technologies. ■

Discovery WorkFlow™ consists of a centralized workflow definition and control module which integrates a centralized database with the ability to launch programs which process the data. Discovery WorkFlow is able to launch and control processes across multiple hardware platforms and provides a common interface for the reporting and tracking of completion codes and exceptions. Discovery WorkFlow was developed to control the internal processing of ESI for all of our clients with minimal staff intervention. The devil is in the details, and you can read the technical description at www.tcdi.com/DiscoveryWorkFlow.