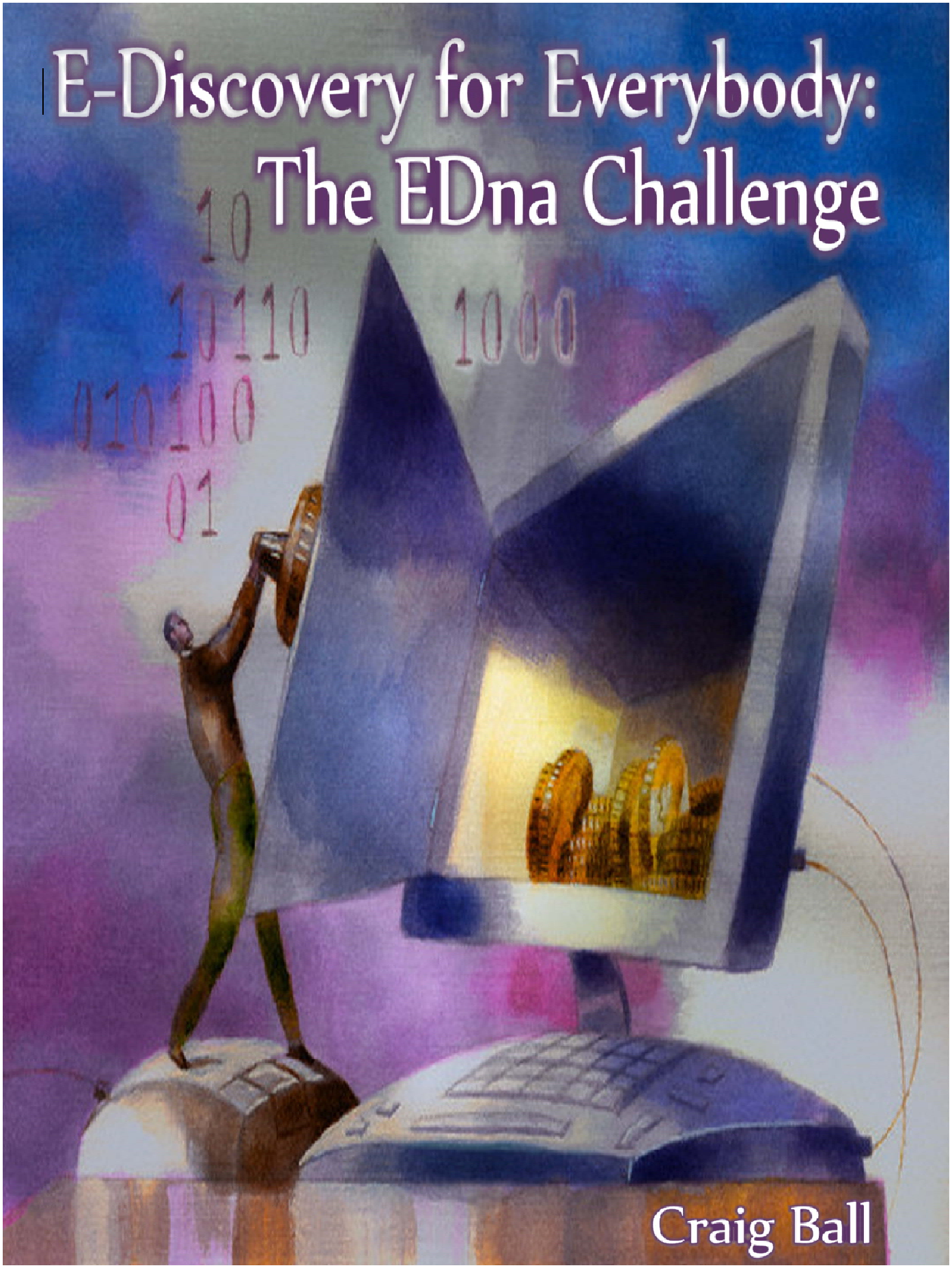
E-Discovery for Everybody: The EDna Challenge



Craig Ball

E-Discovery for Everybody: The EDna Challenge
Craig Ball
© 2009

E-discovery is just for big budget cases involving big companies, handled by big firms.

Right, and suffrage is just for white, male landowners.

Some Neanderthal notions take longer than others to get shown the door, and it's time to dispel the mistaken belief that e-discovery is just for the country club set.

Today, evidence means *electronic* evidence; so, like the courts themselves, access to evidence can't be just for the privileged. Everyone gets to play.

If you think big firms succeed at e-discovery because they know more than you do, think again. Marketing hype aside, big firm litigators don't know much more about e-discovery than solo practitioners. Corporate clients hire pricey vendors with loads of computing power to index, search, de-duplicate, convert and manage terabytes of data. Big law firms deploy sophisticated in-house or hosted review platforms that let armies of associates and contract lawyers plow through vast plains of data--viewing, tagging, searching, sorting and redacting with a few keystrokes. *The big boys simply have better toys.*

A hurdle for everyone else is the unavailability and high cost of specialized software to process and review electronic evidence.

A Mercedes and a Mazda both get you where you need to go, but the e-discovery industry has no Mazdas on the lot. This article explores affordable, off-the-shelf ways to get where you need to go in e-discovery.

One Size Doesn't Fit All

First, let's set sensible expectations: Vast, varied productions of ESI cannot be efficiently or affordably managed and reviewed with software from Best Buy. If you're grappling with millions of files and messages, you'll need to turn to some pretty pricy power tools.

The key consideration is workflow. Tools designed for ESI review can save considerable time over cobbled-together methods employing off-the-shelf applications; and, when every action is extrapolated across millions of messages and documents, seconds saved add up to big productivity gains.

But few cases involve millions of files. Most entail review of material collected from a handful of custodians in familiar productivity formats like Outlook e-mail, Word documents, Excel spreadsheets and PowerPoint presentations. Yes, volume is a challenge in these cases, too; but, a mix of low-cost tools and careful attention to process makes it possible to do defensible e-discovery on the cheap.

Paper Jam

More from comfort than sense, ESI in smaller cases tends to be printed out. Paper filled the void for a time, but lately the cracks are starting to show. Lawyers are coming to appreciate that printing evidence isn't just more expensive and slower, it puts clients at an informational disadvantage.

When you print an electronic documents, you lose three things: Money, time and metadata. Money and time are obvious, but the impact of lost metadata is often missed. When you move

ESI to paper or paper-like formats like TIFF images, you cede most of your ability to search and authenticate information, along with the ability to quickly and reliably exclude irrelevant data. Losing metadata isn't about missing the chance to mine embedded information for smoking guns. That's secondary. Losing metadata is like losing all the colors, folders, staples, dates and page numbers that help paper records make sense.

The EDna Challenge
I polled a group of leading e-discovery lawyers and forensic technologists to see what tools and techniques they thought suited to the following hypothetical:

> Your old school chum, Edna, runs a small firm and wants your advice. A client is about to send her two DVDs containing ESI collected in a construction dispute. It will be Outlook PST files for six people and a mixed bag of Word documents, Excel spreadsheets, PowerPoint presentations, Adobe PDFs and scanned paper records sans OCR. There could be a little video, some photographs and a smattering of voicemail in WAV formats. "Nothing too hinky," she promises. Edna's confident it will comprise less than 50,000 documents and e-mails, but it could grow to 100,000 items before the case concludes in a year or two.
>
> Edna's determined to conduct an in-house, paperless privilege and responsiveness review, sharing the task with a tech-savvy associate and legal assistant. All have late-model, big screen Windows desktop PCs with MS Office Professional 2007 and Adobe Acrobat 9.0 installed. The network file server has ample available storage space. Edna doesn't own Summation or Concordance, but she's willing to spend up to $1,000.00 for new software and hardware, but not a penny more. She's open to an online Software as a Service (SaaS) option, but the review has to be completed using just the hardware and software she currently owns, supplemented only by the $1,000.00 in new purchases. Her team will supply as much brute force as necessary. She's too proud to accept a loan of systems or software, and you can't change her mind or budget.
>
> ***How should Edna proceed?***

Goals of the Challenge
Ideally, the review method employed should:

1. Preserve relevant metadata;
2. Incorporate de-duplication, as feasible;
3. Support robust search of Outlook mail and productivity formats;
4. Allow for efficient workflow;
5. Enable rudimentary redaction;
6. Run well on most late-model personal computers; and
7. Require no more than $1,000.00 in new software or hardware, though it's fine to use fully-functional "free trial" software so long as you can access the data for the 2-3 year life of the case.

I had some ideas (shared later in this article), but expected my colleagues might point me to better mousetraps. Instead, I was struck by the familiarity and consistency of their excellent suggestions as compared to options that have been around for years. Sadly, there's not that much new for those on shoestring budgets; that is, developers remain steadfastly disinterested in 85% of the potential market for desktop discovery tools.

One possible bright spot was the emergence of hosted options. No one was sure the job could be begun--let alone completed--using SaaS on so tight a budget; but, there was enough mention of Saas to make it seem like a possibility, now or someday soon.

Advice to Edna

While the range of proposals was thin, the thought behind them was first-rate. All responding recognized the peril of using the various Microsoft applications to review the ESI. Outlook's search capabilities are limited, especially with respect to attachments. If Edna expected to reliably search inside of every message, attachment and container file, she would need more than Outlook alone.

Notable by their absence were any suggestions to use Google's free desktop indexing and search tool. Though a painful interface for e-discovery, Google Desktop installed on a dedicated, "clean" machine would be capable of reading and searching Outlook e-mail, Word documents, Excel spreadsheets, PowerPoint presentations, PDF files, Zip archives and even text within music, video and image files. It wouldn't be pretty--and Edna would have to scrupulously guard against cross-contamination of the evidence with other data--but Google Desktop *might* get much of the job done without spending a penny.

Quin Gregor of Strategic Data Retention LLC in Georgia was first to respond with an endorsement of my two favorite affordable workhorses, the ubiquitous dtSearch indexing and search tool ($199.00 at www.dtsearch.com) and Aid4Mail ($69.95 at www.fookes.com), a robust utility for opening, filtering and converting common e-mail container files and message formats. Quin described a bankruptcy case where a microscopic budget necessitated finding a low-end option. He reports that dtSearch and Aid4Mail saved the day.

Ron Chichester, an attorney and forensic examiner in Texas pointed to the many open source Linux tools available without cost. These command line interface tools are capable of indexing, Bayesian analysis and much of the heavy lifting of the tools used by e-discovery vendors; but. Ron acknowledged that Edna and her staff would need a lot of Linux expertise to integrate the open source offerings. Bottom line: The price is right, but the complexity unacceptable.

Florida e-discovery author and blogger, Ralph Losey, a partner at AkermanSenterfitt, suggested using an online review tool like Catalyst and tried to dance around the budget barrier by pointing out that the cost could be passed on to the client. Ralph argued that hosting would save enough lawyer time to pay for itself. No doubt he's right; but, passing on the costs isn't permitted in the Edna Challenge and, even in a real world situation, unless the savings were considerable, Edna's likely to keep the work--and the revenue--in house.

Another Floridian, veteran forensic examiner, Dave Kleiman, suggested that Edna blow her budget on alcohol and amphetamines because she has a lot of toil ahead of her. Party on, Dave!

Our northern neighbor, Dominic Jaar of Ledjit Consulting Inc. in Quebec, took a similar doleful tack. Dominic thought that SaaS might be a possibility but added that Edna should use her grand to take an e-discovery course because she needs to learn enough to "stay far from the case." Else, he offered, she could go forward and apply the funds to coffee and increased malpractice coverage. *Ouch!*

John Simek of Sensei Enterprises in Virginia prudently suggested that Edna use part of her budget to buy an hour of a consultant's time to help her get started. John predicted that a SaaS approach would be priced out-of-reach, but was another who thought salvation lay with

dtSearch. John recognized that Adobe Acrobat could handle both the redaction and light-duty OCR required. As for the images, video and sounds, Edna's in the same boat, rich or poor. She's just going to have to view or listen to them, one-by-one.

Jerry Hatchett with Evidence Technology in Houston suggested LitScope, a SaaS offering from LitSoft. Jerry projected a cost of around $40/GB/month, which would burn through Edna's budget in about 3 months...if she didn't buy any Starbucks. Following up, I discovered that LitScope can't ingest the native file formats Edna needed to review unless accompanied by load files containing the text and metadata of the documents and messages. The cost to pre-process the data to load it would eat up Edna's budget before she looked a single page. That, and a standard $200 minimum on monthly billings coupled with a 6 month minimum commitment, made this SaaS option a non-starter. Attractive pricing, to be sure, but not low enough for Edna's shallow pockets.

The meager budget forced George Rudoy, Director of Global Practice Technology & Information Services at Shearman & Sterling, LLP in New York, to suggest using Outlook 2007 as the e-mail review tool, adding the caveat that metadata may change. Unlike earlier versions, Outlook 2007 claims to extend its text search capabilities to attachments. Unfortunately, it doesn't work very well in practice, meaning Edna and her staff will need to examine each attachment instead of ruling any out by search. George also urged Edna to buy licenses for Quick View Plus--a universal file viewer utility--and hire an Access guru to design a simple database to track the files and hyperlink to each one for review.

From Down Under, Michelle Mahoney of Mallesons Stephen Jaques in Melbourne shared several promising approaches. She suggested Karen's Power Tools (a $30 suite of applications at www.karenware.com) as a means to inventory and hash the files and Microsoft Access as a means to de-duplicate by hash values. Michelle also favored hyperlinking from Access for review, working through the collection progressively, ordering them by file type and then filename. She envisions adding fields to the database for Relevant and Privileged designations and a checkbox for exceptional files that can't be opened and require further work.

For the e-mail files, Michelle also turns to Outlook as a review tool, proposing that folders be created for dragging-and-dropping items into Relevant Non Privileged; Relevant Privileged and Non Relevant groups. She echoed warnings about metadata modification and gives her thumbs up to Aid4Mail.

Finally, Michelle offers more kudos for dtSearch as the low cost tool-of-choice for keyword searching. dtSearch will allow Edna to run keywords across files, including emails and attachments, and it is a simple file copy option to copy them, with or without original path, into a folder. Messages emerge in the generic MSG mail format, and Edna can either produce them in that format (with embedded attachments) or use Aid4Mail to copy them into an Outlook PST file format. For further discussion of using dtSearch as a low-cost e-discovery tool, see, Craig Ball, *Do-It-Yourself Digital Discovery***,** (Law Technology News, May 2006).

Tom O'Conner, Director of the Legal Electronic Document Institute in New Orleans, observed that he often gets requests like Edna's from his clients in Louisiana and Mississippi and weighed in with a mention of Adobe Acrobat, noting that it might be feasible to print everything to Acrobat and use Acrobat's annotation and redaction features. As mentioned, Acrobat also offers rudimentary OCR capabilities to help deal with the scanned paper documents in the collection and even has the ability to convert modest volumes of e-mail to PDFs directly from Outlook. For further discussion of using Adobe Acrobat to process Outlook e-mail, see, Craig Ball, *Adobe Brings an Acrobat to Perform EDD* (Law Technology News, June 2008). Tom

concludes that, although working with the tools you already own and know can be cumbersome, it's sometimes a better approach that trying to master new tools under pressure.

Ohio-based e-discovery consultant, Brett Burney, had some very concrete ideas for Edna. He thought she could try to find some SaaS solution to host the data, suggesting Lexbe, NextPoint or Trial Solutions as candidates. Brett was most familiar with Lexbe and knew of small law firms that had successfully and inexpensively used their services.

Brett guessed Edna's budget might allow her to upload everything to Lexbe, review it quickly and then take everything down before the hosting costs ate up her budget. He reported that Lexbe will accept about any file format, by uploading it yourself or sending it to Lexbe to load. Brett put the cost at $99 per month for 2 users and 1GB of storage. Noting that Edna needs to host more than 1GB of data, he predicted her outlay should be close to $200/month. Brett added, "Edna and her crew can upload everything with the tools they have, get it reviewed pronto (i.e. less than a month), and then take everything down--paying only for what they use."

For the Outlook e-mail, Brett thought Edna should turn to Adobe Acrobat and convert the PST container files to PDF Portfolios along the lines of my June 2008 column. Alternatively, Brett suggested Edna use the free Trident Lite tool from Wave Software (www.discoverthewave.com) to get a "snapshot" of the PSTs and then convert relevant messages to PDF or upload them to a hosting provider.

Lisa Habbeshaw of FTI in California pointed to Intella by Vound Software (http://www.vound-software.com) as an all-in-one answer to Edna's needs. Intella offers an efficient indexing engine, user-friendly interface and innovative visual analysis capability sure to make quick work of Edna's review effort. Lisa was unsure if the program could be had for under $1,000, but noted that Vound Software offers a free, fully-functional demo that might fill the bill for Edna's immediate needs. Like Lisa, I'm unsure whether Intella will bust Edna's budget, but it's certainly a splendid new entry to the do-it-yourself market.

Other Great Tools

If the dollar holds its own against the Euro, Edna could accomplish just about everything she needs to do using a terrific tool created in Germany called X-Ways Forensics from X-Ways Software Technology AG. X-Ways Forensics could make quick work of the listing, hashing, opening, viewing, indexing, searching, categorizing and reporting on all that client data; however, it's a complex, powerful forensics tool that would require more time and training to master than Edna can spare. Plus, it would eat up all of her $1,000 budget.

If her budget was bigger, Edna would be very happy attacking the review with the easy-to-use, fast and versatile Nuix Desktop (www.nuix.com). Nuix would allow Edna to begin her review in minutes, and it supports a host of search options. The embedded viewer, hash and classification features foster an efficient workflow and division of review among multiple reviewers. Like Intella, Nuix is an Australian import. Whatever they're doing way down there in Kangaroo land, they're certainly doing something right!

A Few More Ideas for Edna

It's hard to add much to so many fine ideas. Collectively, dtSearch, Adobe Acrobat and Aid4Mail deliver the essential capabilities to unbundle, index, search, OCR and redact the conventional file formats and modest data volumes Edna faces. Her challenge will be cobbling together tools not designed for e-discovery so as to achieve an acceptable workflow and defensible tracking methodology. It won't be easy.

For example, while dtSearch is Best of Class in its price range, it doesn't afford Edna any reasonable way to tag or annotate documents as she reviews them. Accordingly, Edna will be obliged to move each document to a folder as she makes her assessments respecting privilege and responsiveness. That effort will get very old, very fast.

On the plus side, dtSearch offers a fully functional thirty-day demo of its desktop version, so Edna can buy a copy for her long-term use, but rely on 30-day evaluation copies for her staff during the intense review effort--a $400 savings.

While Adobe Acrobat supports conversion of e-mail into PDFs, the process is painfully slow and cumbersome. Moreover, the conversion capabilities break down above 10,000 messages. That sounds like a lot, but it's likely less than Edna will see emerge in the collections of six custodians. Further, Edna may encounter an opponent who smart enough to demand the more versatile electronic formats for e-mail (i.e., PST, MSG or EML). What's Edna going to do if she finds herself locked into a reviewed wedded to image formats?

Whatever tools she employs, Edna will need to be meticulous in her shepherding of the individual messages and documents through the process. To that end, I'd offer this advice:

1. Your first step should be to make a working copy of the data to be processed and secure the source dataset against any usage or alteration. Processing of ESI poses risks of data loss or alteration. If errors occur, you must be able to return to uncorrupted data from prior steps. For each major processing threshold, set aside a copy of the data for safekeeping and carefully document the time the data was set aside and what work had been done to that point (e.g., the status of deduplication, filtering and redaction).

2. From the working copy, hash the files and generate an inventory of all files and their metadata. The processes you employ must account for the disposition of every file in the source collection or extracted from those files (i.e., message attachments and contents of compressed archives). Your accounting must extend from inception of processing to production. By hashing the constituents of the collection as it grows, you gain a means to uniquely identify files as well as a way to identify identical files across custodians and sources.

   A useful tool for hashing files is Karen's Hasher available at http://www.karenware.com. But the best "free" tool for the task is AccessData's FTK Imager, available from www.accessdata.com/downloads. FTK Imager not only hashes files, it also exports Excel-compatible comma delimited listings of filenames, file paths, file sizes and modified, accessed and created dates. Moreover, it supports loading the collected files into a container called a Custom Content Image that protects the data from metadata corruption.

3. Devise a logical division scheme for the components of the collection; e.g., by machine, custodian, business unit or otherwise. Be careful not to aggregate files in a manner that files from one source may overwrite identically named files from other sources.

4. Expand files that hold messages and other files. Here, you should identify e-mail container files (like Outlook .PST files) and archives (e.g., .Zip files) that must be opened or decompressed to make their constituents amenable to search. For e-mail, this can be done using an inexpensive utility like Aid4mail from Fookes Software or Trident Lite from Wave Software. Additionally, e-mail client applications, including Outlook, usually permit export of individual messages and attachments. Though dtSearch includes a command line utility to convert Outlook PST container files to individual messages (.MSG) files for indexing, it doesn't work well or easily compared to Aid4Mail. Finally, most indexing tools are capable of directly accessing text within compressed formats. For example, DTSearch can extract text from Zip files and other archives.

5. A feature common to premium e-discovery tools but hard to match with off-the-shelf software is deduplication. You can use hash values to identify identical files, but the challenge is to keep track of all de-duplicated content and reliably apply tagging for privilege and responsiveness to all deduplicated iterations. Most off-the-shelf utilities simply eliminate duplicates and so aren't suited to e-discovery.

   This is where it's a good investment to secure help from an expert in Microsoft Excel or Access because those applications can be programmed to support deduplication tracking and tagging.

   When employing deduplication, keep in mind that files with matching hash values can have different filenames and dates. The hash identicality of two files speaks to the *contents* of the files, *not* the names assigned to the files by the operating system or to information, like modified, accessed and created dates, stored *outside* the files.

6. Above all, don't process and review ESI in a vacuum. Be certain that you understand the other side's expectations in terms of the scope of the effort, approach to search and--critically--the forms of production they seek. You may not agree on much, but you may be pleasantly surprised to learn that some of the perils of a low budget e-discovery effort (e.g., altered metadata, limited search capabilities, native production formats) don't concern the other side. Further, you may reach accord on limiting the scope of review in terms of time intervals, custodians and types of data under scrutiny. Why look at *all* the e-mail if the other side is content with your searching just communications between Don and Betty during the third week of January 2009?

Finally, Edna may seek an answer to two common questions from those taking the do-it-yourself route in e-discovery:

What if I change metadata?

Certain system metadata values--e.g., last access times and creation dates--are prone to alteration when processed using tools not designed for e-discovery. Such changes are rarely a problem if you adhere to three rules:

1. **Preserve** an unaltered copy of whatever you're about to process;
2. **Understand** what metadata were altered; and,
3. **Disclose** the changes to the requesting party.

By keeping a copy of the data at each step, you can recover true metadata values if particular values proves significant. Then, disclosing what metadata values were changed eliminates any suggestion that you pulled a fast one. Many requesting parties have little regard for system metadata values; but, they don't want to be surprised by relying on inaccurate information.

Can I Use My Own E-Mail Account for Review?

You wouldn't commingle client funds with your own money, so why commingle e-mail that's evidence in a case with your own mail? That said, when ESI is evidence and the budget leaves no alternative, you may be forced to use your own e-mail tools for small-scale review efforts. If so, remember that you can create alternate user accounts within Windows to avoid commingling client data with your own. Better still, undertake the review using a machine with a clean install of the operating system. Very tech-savvy counsel can employ virtual environments (e.g., VMWare products) to the same end.

If using an e-mail client for review, it may be sufficient to categorize messages and attachments by simply dragging them to folders representing review categories; for example:

1. Attorney-client privilege: entire item;
2. Work product privilege: entire item;
3. A-C Privilege: needs redaction;
4. W-P privilege: needs redaction;
5. Other privilege;
6. Responsive;
7. Non-responsive.

Once categorized, the contents of the various folders can be exported for further processing or for production, if in a suitable format.

Throwing Down The Gauntlet

The vast majority of cases filed, developed and tried in the United States are not multimillion dollar dust ups between big companies. The evidence in modest cases is digital, too. Solo and small firm counsel like Edna need affordable, user-friendly tools designed for desktop e-discovery--tools that preserve metadata, offer efficient workflow and ably handle the common file formats that account for nearly all of the ESI seen in day-to-day litigation. Using the tools and techniques described by my thoughtful colleagues, Edna will get the job done on time and under budget. The pieces are there, though the integration falls short.

So, how about it e-discovery industry? Can you divert your gaze from the golden calf long enough to see the future and recall the past? Sam Walton became the richest man of his era by selling to more for less. There's a fast growing need...and a huge emerging market. The *real* Edna Challenge is waiting for the visionaries who will meet the need and serve the market.