

ENF: It's time to shuck outdated technology and adopt a file format for today's e-discovery needs

We all struggle with the rising cost of e-Discovery. What used to be a simple process of delivering just a few boxes of paper for a production has turned into a tide that sometimes breaks the bank of litigation. Today, just mention the word e-Discovery and it begins to elicit thoughts of settlement. E-Discovery and the entire production process often is mired in technical mumbo-jumbo, geek speak, frustration, cost overruns, and fear. Why have we come to this place in the world of litigation? We can blame many factors, including the exponential growth of data, the diversity of data sources, the complexities of collecting custodian data, vendor pricing, or even “the cloud”. But, let's first look at the history of how we arrived at the present.

In the good old days (as some may remember them), we used a photocopier, a staple remover, and some colored slips of paper to separate documents. We performed a process called “blowback” if there was anything electronic, but in the end it all got put to good old ink on tree bark – and delivered to opposing counsel. We carried it around in banker boxes, often using two wheel dollies, and created witness binders or red rope folders filled with the good old days of paper. *“Just print it out and put it on my desk”* is a quote from a well viraled video on the Internet about the litigation support person meeting with an attorney.

We then quickly moved into the world of electronic data (ESI as it's known today – Electronically Stored Information) and the banks of photocopiers couldn't keep up. We needed a new medium, and turned to TIFF (Tagged Image File Format) as a solution. TIFF was developed by Microsoft and Aldus in 1986. TIFF was created primarily for input and output devices such as printers, monitors, and scanners. As a result, it is specifically designed to be compatible with different image processing devices. So, it was a file format designed for the print industry. But, we adopted it for the e-Discovery industry.

TIFF seemed to work fine back in the day, and is even the standard for fax machines which started out connected by analog phone lines. It was a format easily transmitted across slow data links. But, who owns a fax machine today? And, why are we still using technology that is 30 years old!

As the e-Discovery industry matured, it looked for a new format, and soon adopted PDF files. But, again, PDF was designed for the print industry – not the e-Discovery industry. It works well, stores a preserved image rendition of the printed page, and can contain text for searching. TIFF on the other hand doesn't contain text, therefore we all came up with this brilliant contraption called TIFF/TEXT and have adopted the “Concordance Load File Format” (OPT and DAT) as one de facto industry standard for delivering productions.

Let's re-cap here a bit. Both a TIFF and PDF essentially capture that old ink on bark, paper page image and preserve it to a file. That's a good thing some will argue because it preserves an image of the original file. Essentially, we're still doing blowbacks but now to files instead of paper. Thanks for saving the trees folks, but here's the rub – it's an expensive process. Very expensive.

Talk to any litigation support processing specialist and they will tell you the most difficult, time consuming, and costly process they do is convert a perfectly good native file into a 2-dimensional page rendition called a TIFF or PDF image. Add to that the geek mystical powers it takes to properly format a decent load file; and you begin to understand why things are getting expensive – and why they're taking way more time than they should. And don't even get me started on what someone is supposed to do with a bunch of single page TIFF files, text files and a load file – if they don't have a document review tool.

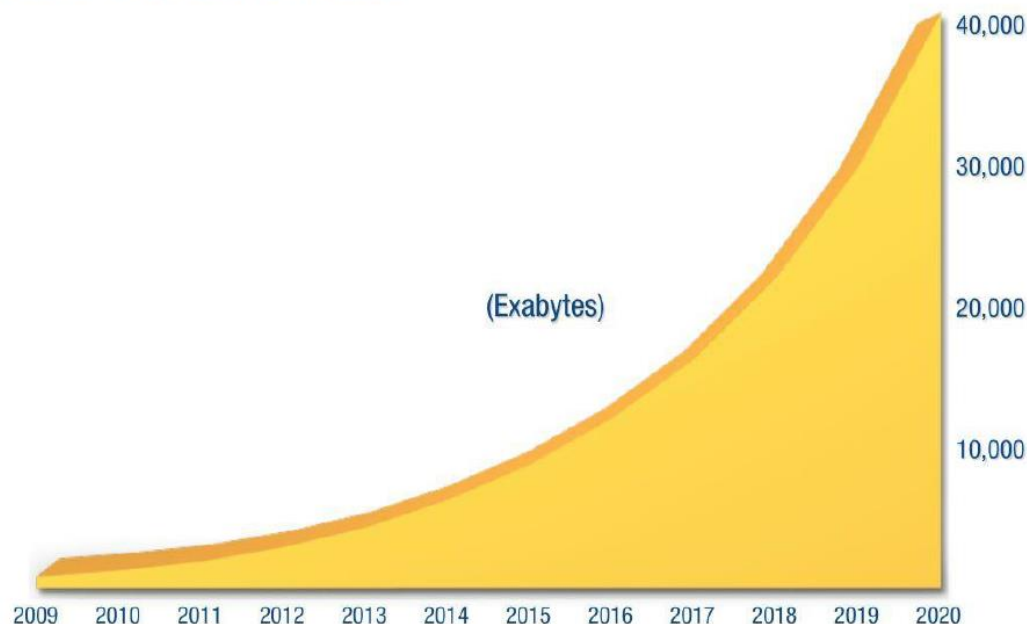
Documents today aren't the same as they were even a few years ago. Today, a Word document can contain a hyperlink to an external source, an embedded graphic, a dynamic field that changes each time the document is opened or printed, etc., etc., etc. Excel files are rows and columns, calculations, and rarely if ever fit comfortably on an 8 ½ x 11 page. I call these 3-dimensional documents. And today, everything is a 3-dimensional document – Word files, Excel files, database files, even E-mail messages. How do you take 3-dimensional things and turn them into 2-dimensional things?

The print industry has TIFF and PDF. The engineering industry has AutoCAD (DWG). The music industry has MP3 and WAV. The video industry has MPEG and AVI. Even the radiology industry has a file format (DICOM). Why doesn't e-Discovery have a file format, designed especially for them?

We're beginning to see an emerging trend in discovery stipulations to "produce Excels as native", born largely out of this complexity to turn them into 2-dimensional renderings. But, there is still reluctance to do so. Why? Because of security concerns. Why not produce all documents natively? Because of security concerns. And, the fact that metadata fields (informational fields about the files – such as author, subject, from, to, etc.) is needed as well. Those metadata fields get stuffed into that DAT file I mentioned earlier.

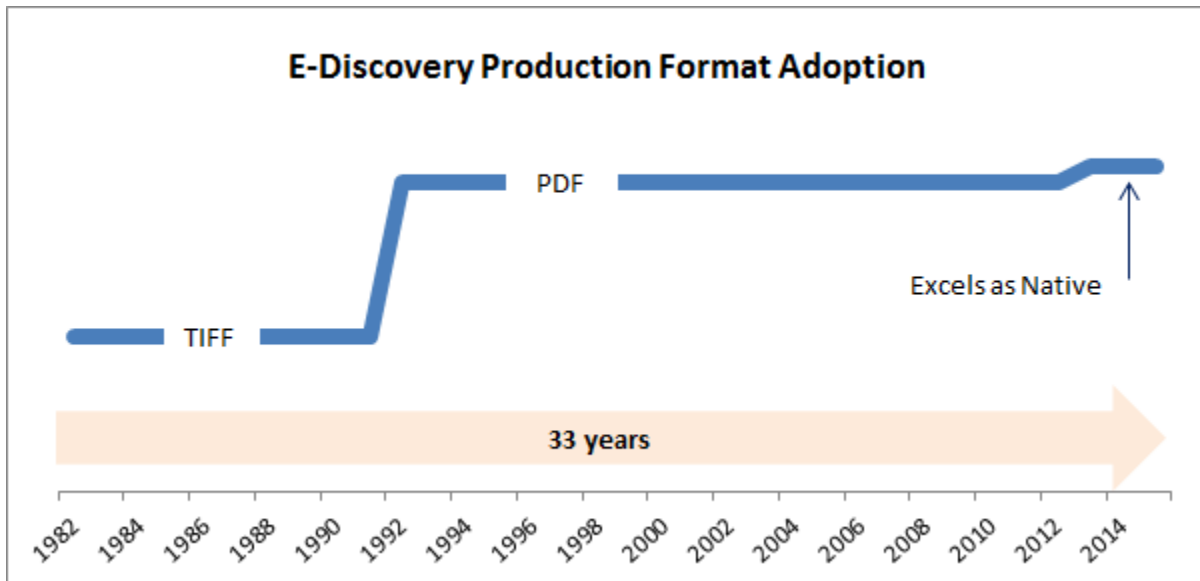
Let's look at a couple of charts. All articles need charts right? Below I list the exponential growth of data – reflecting one reason for the explosion of e-Discovery costs; and another showing the history of file formats for production. To put the first chart into real world context, 40 ZB (1 zettabyte = 1,000 exabytes) is equal to 57 times the number of all the grains of sand on all the beaches on earth.

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.

Is there any doubt the cost of e-Discovery will get increasingly more expensive? By comparison, let's look at the technology curve for adoption of new production formats. In 30 years, little has been done to address the issue of production formats.



If we take, as an assumption, that the largest cost of e-Discovery is borne by rendering native files into page image equivalents, then there is an obvious choice – create a new production standard. A standard that is specifically designed for e-Discovery and addresses the concerns of cost, security and packaging that older solutions have not solved. It's time to let Generation Y take over.

The major obstacles to not adopting native file productions are (1) security, (2) inclusion of metadata, (3) family relationships (e.g., an email and its attachments), and (4) redactions.

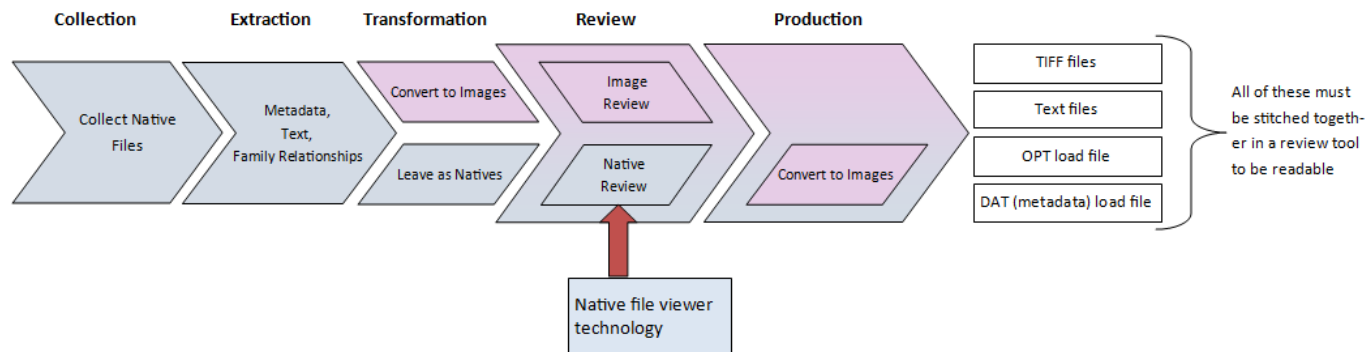
To overcome these obstacles, I designed a new architecture to address these concerns. It is called ENF (Encapsulated Native File). An ENF file contains a native file, fully encrypted, along with metadata and all members of a family relationship. The ENF file architecture has been designed and preliminary tools for creating ENF files (makeENF) and viewing them (viewENF) have been created.

ENF is a file format for the e-Discovery industry. Designed to eliminate all barriers to native file productions. Designed to significantly reduce e-Discovery costs – by eliminating the labor of translating native files to 2-dimensional renderings. If we take native files and simply repackage them, adding security levels and passwords, and designing a self-contained file with all properties included – think of the savings in turnaround time from collection – to review – to production.

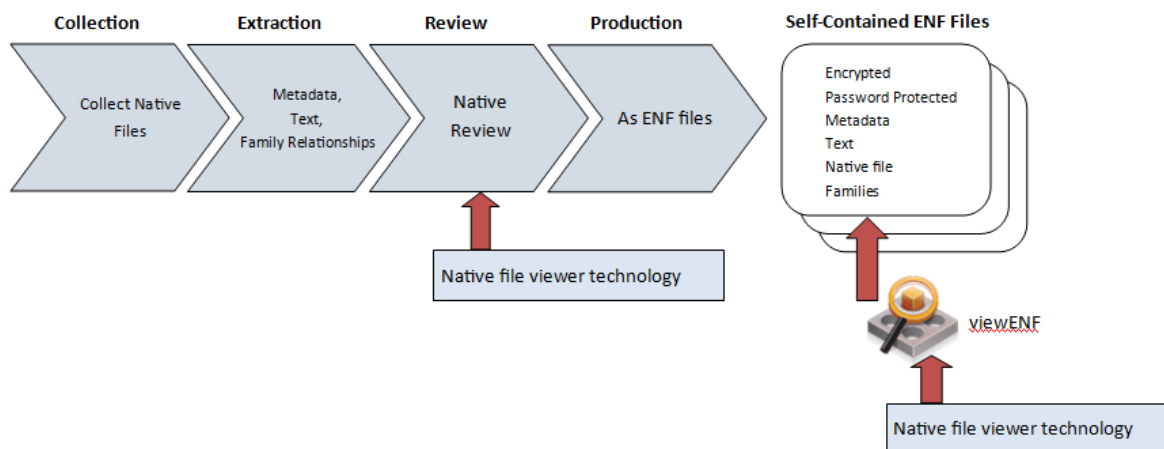
But, adoption of this new standard by multiple vendors and our e-Discovery industry is needed before we can achieve its results.

Let's look at the lifecycle of e-Discovery. It starts with collecting native files; followed by loading them into a processing tool to extract the metadata and text and relationship between parent and children (emails and attachments); then the process transforms native files into "print images". That rendering process is often error prone and labor intensive. Any anomaly in files must be handled by a hands-on technical professional. Margins re-adjusted, columns narrowed, orientation corrected, all in an effort to transform 3-dimensional into 2-dimensional. Once that lengthy process is complete, the results are loaded into a document review platform.

Some people will debate that several review tools today support reviewing native files. And, that is a step in the right direction; since it expedites the time reviewing can start. To achieve that, some type of viewing technology is needed (e.g., QuickView Plus or Oracle’s Outside-In). Once documents are selected for production, they are produced typically in an image format with corresponding load files; which are really sewing machines that stitch back together the individual image files, as illustrated in the diagram below.



What if we simplified the process? And, lowered the cost? And, made the process faster and more efficient?



ENF uses the same viewer technology as the document review tools. It’s just carried forward and used to view ENF files, just as you do in your review tool. The viewENF application supports encryption, and password protection to limit features that can be done with the embedded native file, and encapsulates the extracted text/metadata/and the complete family tree. If you will, “viewENF” becomes the new Adobe Reader for the e-Discovery industry.

Could a new architecture, a new file format, designed specifically for the e-Discovery industry save costs? Yes. It’s time to leave 30 year old technology behind, and let Generation Y invent a new paradigm.

About the author:

Wade Peterson is the Manager of eDiscovery and Litigation Support at Fredrikson & Byron P.A. With over 45 years of experience, Wade specializes in maximizing the firm’s performance through strategic design, management and implementation of technology and resources. Prior to joining Fredrikson & Byron, Wade held technology leadership positions in law firms and vendors servicing the legal field. In those roles, he managed application development, mentored technology professionals and launched numerous legal technology initiatives including national and international data center designs.