

Can Native File Productions be ENF (Enough)?

November 10, 2019
Wade Peterson

Background:

The current industry standard for legal document productions typically occurs in one of the following three formats: (a) paper, (b) TIFF files, or (c) PDF files. Variations occur within each format, such as single or multi-page TIFF files; resolution of images (typically Group IV, 300 dpi, black & white for TIFF and JPEG for color images), etc.

The reasons the standards exist are traditionally historical in nature. TIFF being an original attempt to adopt an industry standard format for black and white fax machines and scanners, starting back in the mid-80's. It has not had a major update since 1992.

In 1991, Adobe outlined a new system which eventually evolved into the PDF standard. Its attempt was to create a standard to represent documents independent of application software, hardware, or operating system. It was officially released as an open standard in 2008.

Executive Overview:

Legal document productions have come a long way since the mid-80s; and yet we still use standards defined almost two decades ago! Today, litigation support professionals are faced with increasing challenges to force newer technology documents into older technology standards. In today's environment we have "*3-dimensional documents*". Examples of these include Word documents with embedded links to web sites; embedded Excel worksheets in Word documents; Excel cells dynamically obtaining data from SQL databases; PowerPoint slides with embedded videos, Excel graphs, and animation; hidden rows/columns; pivot tables; etc.

Attempting to represent these 3-dimensional documents in a 2-dimensional world of TIFF or PDF is challenging, if not impossible.

This paper presents a conceptual framework for defining a new standard. A standard designed for litigation support professionals. A standard designed for legal document productions.

Challenges:

I've already outlined some of the challenges faced by litigation support professionals doing document productions. Listed below (although not exhaustive) are a few other challenges:

- Courts and opposing counsel are increasingly demanding "native file productions".
- Native files can be altered (either intentionally or not).
- Native files cannot be redacted.
- Native files cannot be bates stamped or endorsed or watermarked.
- Native file names and paths can often exceed the Windows standard 255 character limit.
- Native files are typically never printed, hence how can they be "represented" in a print format (which is what the TIFF and PDF standards were designed for).
- Native files with "embedded content" do not translate well to print format, therefore arbitrary mechanisms were created to support them (e.g., "parent" and "child" relationships).
- Emails with attachments must be handled using add-on standards (e.g., BegAttach, EndAttach).
- Extracted text from native files must be delivered separately from the TIFF files, and a mechanism built to associate that text with the actual TIFF image(s) (e.g., TIF/TXT files).
- Color files cannot be produced as TIFF, they are often supplemented as JPEG images.
- Vendor specific formats have been created, and adopted as pseudo-standards to support these new 3-dimensional documents (e.g., Summation DII files; Concordance delimiters; OPT and DAT files; and recently XML standards).

- A single page TIFF (and its accompanying OPT file) production cannot be viewed outside of normal document review platforms.
- Smaller firms do not own document review platforms.
- Document productions are difficult, if not impossible, to use in deposition or courtroom settings.

Solution:

The solution is to develop a new standard. One which addresses today's concerns yet has an open architecture to meet future requirements as they occur. Develop a new standard for litigation support document productions. One which eventually is adopted by courts as the legal standard, much like standard legal citation formats were adopted centuries ago.

The solution is not to transform native files into something they are not. The solution is to embrace native files, and build an architecture around them that addresses the challenges our industry faces.

Architecture:

The following paragraphs layout the basis for a new standard for native file productions. While not a complete specification, it presents the concept from a high-level vantage point. The thought is this can be an open-standard, available to anyone to build upon, incorporate into existing products, or used to build new products for the legal production marketplace.

The idea is to develop a framework for "encapsulating" native files in sort of an envelope metaphor. Just as a red rope folder often contains several pieces of physical information (including the original document); the electronic **encapsulated native file** (perhaps even using the file extension .ENF) would contain the native file + ancillary metadata fields, including hash values, extracted metadata, etc. Think along the lines of a zip file, but rather than just files within a zip file, an ENF file contains both a file and fields and controlling mechanisms. The effect being that we can begin to deposit what we need for a native file (including its original source) into a definable, perhaps even XML format, container. Having the native file contained within a "super container" also may avoid the issue of long file names; since the ENF file can be named whatever, while still preserving (internally) the original native file name (and path too if need be). For all I know, this is already being done by someone, somewhere. But, I haven't heard of it being done before.

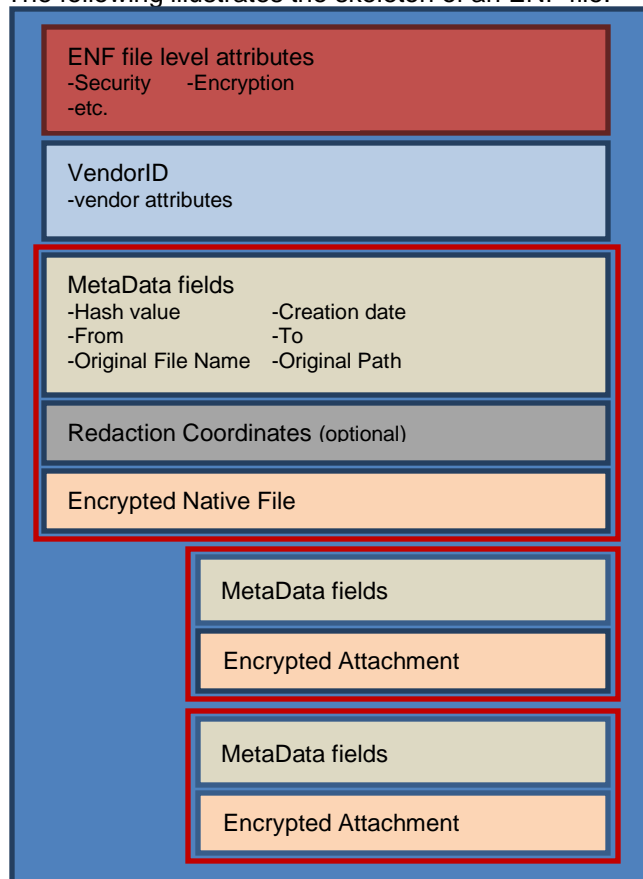
The ENF file format could become an EDRM standard. It may even be possible to encrypt the ENF file; and deposit redaction coordinates for the native file when viewed with a compatible file viewer (so for example QuickViewPlus viewing a native file within a ENF file with redaction coordinates, would show a redacted view of the native, not the actual native); and of course we can create a hash value for the ENF file itself. There, of course, needs to be a new ENF Redaction tool or that capability designed into existing document review platforms for redaction viewing to occur.

Since the ENF file can be protected in many ways, the integrity of the native file can be preserved even with redactions, since the ENF file standard would also contain security rights to "*view native*", "*view redacted version only*", "*can export native*", "*can export metadata*", etc. The ENF viewer may even be able to overlay bates numbers and endorsements and watermarking; that are contained within the ENF file.

Printing and saving by the ENF viewer would be restricted though several means, and several security layers, e.g., you may be able to only "*print the native file, but only with redactions, bates numbers and endorsements*" because of the security placed within the ENF file. The native file itself is encrypted inside the ENF format. Another security level would be "*you are allowed to save the native file in its original format, for use with another application - i.e., for Excel files*". Perhaps even a batch processing program that "*extracts all native files from ENF files, and metadata fields*" into standard load file formats such as iConvert+ or ReadySuite.

Since the ENF architecture is a totally self-contained deliverable for a single file; single ENF files can be taken to depositions or courts for viewing and further analysis. With an appropriate viewer, these become the equivalent of the Adobe Reader of the litigation world.

The following illustrates the skeleton of an ENF file:



Just as a multi-page TIFF contains a single document, or a PDF contains a single document (for most productions I've seen at least); the ENF file becomes the new deliverable. It's a native file production - but in a very controlled environment. We therefore side-step all the previous objections to producing native files. i.e., no one can actually work on the native file without the proper ENF permissions; and those are controlled based on who the production is being delivered to.

It actually becomes a more secure environment for delivering productions than even TIFF/PDF. Since security is built into the specification, we have automatic encryption, in case the production is lost or stolen. If one wanted to, it's even possible to create within the ENF viewer, the ability to return "viewed" information - i.e., "Oh, they're looking at THAT hot document"... although that's a bit scary.

Once a standard is defined for ENF; vendors could be approached to create the "ENF viewer", and persuade them to create ingestion tools for document review platforms based on the ENF standard.

To the courts - this can be a new, guaranteed, securable, encryptable, format. Courts could then impose standards for production that states – "Both parties shall abide by and produce ENF productions". The similarity to "enough" is intentional.

Enhancements to the Standard:

Because the architecture is "open", vendors would have the ability to add additional functionality to the specification which could be unique to them. For example, it might be advantageous to not only have a viewer, but have the ability to add annotations, issue codes and attorney notes to the ENF file while

viewing it. This is the "value add proposition" whereby vendors could add value, and therefore charge for those features in products they deliver based on the ENF standard.

Other enhancements may include the ability to package an entire "email conversation thread" into a single ENF file, or package all "near duplicate documents" in a single ENF file.

Overcoming Obstacles:

What has been proposed above is a radical departure from the current norm. Adoption of such a standard would take time, work and need to be at a broad level of vendor involvement. Courts too would be encouraged to adopt this as a standard for eDiscovery productions.

There will be other obstacles as well. One, off the top of my head, deals with encrypted, password protected ENF files. Since ENF files can stand on their own, combining several ENF files from different productions, with perhaps different passwords, may be problematic.

Developing the tools and software libraries is a challenge to implementation, and no small undertaking for any single vendor. Some of the processes and tools needed to support this standard include:

- Complete, documented, open-source ENF architecture specification
- ENF viewer tool
- ENF creation utility (either built into existing document review platforms, or as an ancillary tool to convert an existing production to ENF format)
- ENF file ingestion mechanisms for various document review platforms
- ENF export/conversion utility to export an ENF production to some other load file format (e.g., to a Summation DII, or OPT/DAT load file format)

Conclusion:

We have been using standards defined well over 20 years ago to produce documents which didn't even exist 20 years ago. Those standards have been patched, augmented, twisted, and bent to accommodate newer *3-dimensional documents*. I suggest it is time to construct a new standard, designed for the world of litigation support in the 21th century, and beyond.