

Forms that Function

By Craig Ball ©2026

This article discusses how to request and produce electronically stored information (ESI) in *forms that function*—that is, in more utile and complete forms of production that preserve the integrity, efficiency and functionality of digital evidence. It explains the advantages of securing production in native and near-native forms and supplies exemplar language crafted to convey forms of production and metadata values sought.

BACKGROUND

Historically, the law little concerned itself with “forms” of production because there were few alternatives to paper. Later, evidence became digital: documents, pictures, sounds, text messages, e-mail, spreadsheets, presentations, databases and more were created, communicated and recorded as sequences of “ones” and “zeroes.” Flat forms of information acquired new dimension and depth, described and supplemented by **metadata**, *i.e.*, data *about* data supporting the ability to find, use and trust digital information.

Digital photographs hold EXIF data revealing where they were taken and by what camera, spreadsheets carry formulae supporting complex calculations and Word documents store editorial histories and are laced with conversations between collaborators. Presentations feature animated text and rich media, including sound, video and dynamic connections to other data. Databases don’t “store” documents as much as assemble them on demand. Even conversations—once the most ethereal of interactions—now linger as text messages and data packets traversing the internet and cellular networks.

Today, the forms in which information is supplied determine if evidence is intelligible, functional and complete.

FORMS OF PRODUCTION IN THE FEDERAL RULES

The Federal Rules of Civil Procedure further the goals that lawyers understand the forms of ESI in their cases and discuss and resolve forms disputes *before* requests for production are served so that unresolved forms disputes be brought to the Court before expending the cost and time of misbegotten production.

Rule 26(f)(3)(C) requires the parties to submit a discovery plan to the Court prior to the first pretrial conference. The plan must address “any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced.”

Rule 34(b)(1)(C) permits requesting parties to “specify the form or forms in which electronically stored information is to be produced,” yet it’s common for requests for production to be wholly silent on forms of production, despite pages of detailed definitions and instructions.

Practice Tip: Requesting parties should supply a clear and practical written specification of forms sought *before* the initial Rule 26(f) conference, affording opponents the opportunity to assess the feasibility, cost and burden of producing in specified forms. Even parties who do not know the forms in which an opponent’s data natively resides can anticipate the *most common* forms suited to, *e.g.*, e-mail, word processed documents, presentations and spreadsheets.

The Federal Rules lay out **five steps** to seeking and objecting to forms of production:

1. Before the first pretrial conference, parties must hash out issues related to “the form or forms in which [ESI] should be produced.” FRCP 26(f)(3)(C)
2. Requesting party specifies the form or forms of production for each type of ESI sought: paper, native, near-native, imaged formats or a mix of same. FRCP 34(b)(1)(C)
3. If the responding party will supply the specified forms, the parties proceed with production. If not, the responding party must object and designate the forms in which it intends to make production. If the requesting party fails to specify forms sought, responding party must state the form or forms it intends to produce. FRCP 34(b)(2)(D)

The Notes to Rule 34(b) add: “A party that responds to a discovery request by simply producing electronically stored information in a form of its choice, without identifying that form in advance of the production . . . runs a risk that the requesting party can show that the produced form is not reasonably usable and that it is entitled to production of some or all of the information in an additional form.”

4. If requesting party won’t accept the forms the producing party designates, the requesting party must confer with the producing party in an effort to resolve the dispute. FRCP 37(a)(1)
5. If the parties can’t agree, requesting party files a motion to compel, and the Court selects the forms to be produced.

Practice Tip: Even when producing parties use native and near-native forms when reviewing for responsiveness and privilege, the last step before production is often to downgrade the evidence to static images. Accordingly, requesting parties shouldn’t wait until the response date to ascertain if an opponent won’t furnish the forms sought. Press

for a commitment; and if not forthcoming, move to compel ahead of the response date. Don't wait to hear the Court ask, "Why didn't you raise this earlier?"

WHAT ARE THE OPTIONS FOR FORMS OF PRODUCTION?

It's rarely necessary or feasible to employ a single form of production for all ESI produced in discovery; instead, tailor forms to the data. Options for forms of production include:

- Paper [where the source is paper and the volume small]
- Page Images [best for items requiring redaction and scanned paper records]
- Native [spreadsheets, electronic presentations and word processed documents]
- Near-native [e-mail and database content]
- Hosted production

Paper

Converting searchable electronic data to paper is rarely a reasonable form of production for ESI, but paper remains an option where the items to be produced are paper documents and so few in number that electronic searchability isn't essential.

Page Images

Parties produce digital "pictures" of documents, e-mails and other electronic records, typically furnished in Adobe's Portable Document Format (PDF) or as Tagged Image File Format (TIFF) images. Converting ESI to TIFF images strips its electronic searchability and metadata. Accordingly, TIFF image productions are accompanied by load files holding searchable text and selected metadata (a so-called "TIFF+ production"). Searchable text is obtained by extraction from an electronic source or, for scanned paper documents, by use of optical character recognition (OCR). Load files are composed of *delimited text*, *i.e.*, values following a predetermined sequence and separated by characters like commas, tabs or quotation marks. The organization and content of load files must be negotiated, and is often pegged to review software like Relativity, Everlaw or DISCO.

Pros: Imaged formats are ideal for production of scanned paper records, microfilm and microfiche, especially when OCR serves to add electronic searchability.

Cons: Imaged production breaks down when ESI holds embedded information (*e.g.*, collaborative content like comments or formulae in spreadsheets) or non-printable information (*e.g.*, voice mail, video or animation and structured data). Imaged productions may also serve to degrade evidence when the information is fielded (*e.g.*, structured data and messaging) or functional (*e.g.*, animations in presentations, table relationships in structured data or threads in e-mail).

Native Production

Parties produce the actual data files containing responsive information, *e.g.*, Word documents in their native .DOC or .DOCX formats, Excel spreadsheets as .XLS and .XLSX files and PowerPoint presentations in native .PPT and .PPTX. Native production is cheaper and superior in competent hands using tools purpose-built for native review.

Pros: The immediate benefits *to the producing party* are speed and economy—little or nothing must be spent on image conversion, text extraction or OCR.

The benefits *to the requesting party* are substantial. Using native review tools or applications like those used to create the data (Careful here! —see *Cons* below), requesting parties see the evidence as it appeared to the producing party. Embedded commentary and metadata aren't stripped away, deduplication is facilitated, e-mail messages can be threaded into conversations, time zone irregularities are normalized and costs are reduced and utility enhanced every step of the way. Moreover, native file sizes tend to be many times more compact than their counterparts converted to static images, making native forms much less costly to ingest for processing and host for review.

Cons: Applications needed to view rare and obscure data formats may be prohibitively expensive (*e.g.*, specialized engineering applications or enterprise database software). If native applications are (unwisely) tasked to review, *e.g.*, Microsoft Word for reviewing Word documents, copies must be used to avoid altering evidence.

Near-Native Production

When some ESI cannot be feasibly or prudently tendered in true native formats, *near-native* forms preserve the essential utility, content and searchability of native forms but are not, strictly speaking, native forms. Examples:

- **Enterprise e-mail** – Enterprise email systems store messages in monolithic container formats like an Exchange Server's EDB format; so, exported messages tend to be stored in container- or single message formats not native to the mail server. These replicate the pertinent content and essential functionality of the source, but again, are not, strictly speaking, native forms.
- **Databases** - Exports from databases are often produced in delimited formats not native to the database yet supporting the ability to interpret the data in ways faithful to the source.

- **Social networking content** - Content from social networking sites won't replicate the precise way the content is stored in the cloud, so near-native forms seek to replicate its essential utility, completeness and searchability.

Hosted Production

Hosted production is more a delivery medium than a discrete form of production. Hosted production resides on a secure website. Requesting parties access data using their web browser, searching, viewing, annotating and downloading data. The electronic forms of production above are the grist ingested by hosting providers (service providers) to comprise the hosted collection.

Load Files Explained

Some years back, I got a call from a lawyer who reported that he'd received production of ESI from a bank and spent the weekend going through it. He'd found images of the pages of electronic documents, but they weren't PDFs and he couldn't search them. He also found a lot of what he called "Notepad documents." He'd specified native production, so thought it odd that the other side produced so many pictures of documents and plain text files.

As it's unlikely a bank would rely on Windows Notepad as its word processor, I probed further and learned that that the production included folders of TIFF images, folders of .TXT files (those "Notepad documents") and folders of files with extensions like .DAT and .OPT. My caller didn't know what to do with these.

Unbeknownst to my caller, he'd received a TIFF+ production from an opponent who ignored his demand for native forms and simply printed everything to electronic paper. The producing party no doubt expected the requesting party to buy or own a review tool capable of cobbling together page images with extracted text and metadata produced in load files. Lacking such tools, the caller found the production to be wholly unsearchable and largely unusable. Should my caller protest, the other side will tell him how all those other files represent the very great expense and trouble they've gone to make the page images searchable, as if furnishing load files to add crude searchability to page images of inherently searchable electronic documents constitutes some great favor.

It brought to mind that classic Texas comeback, "Don't piss in my boot and tell me it's raining."

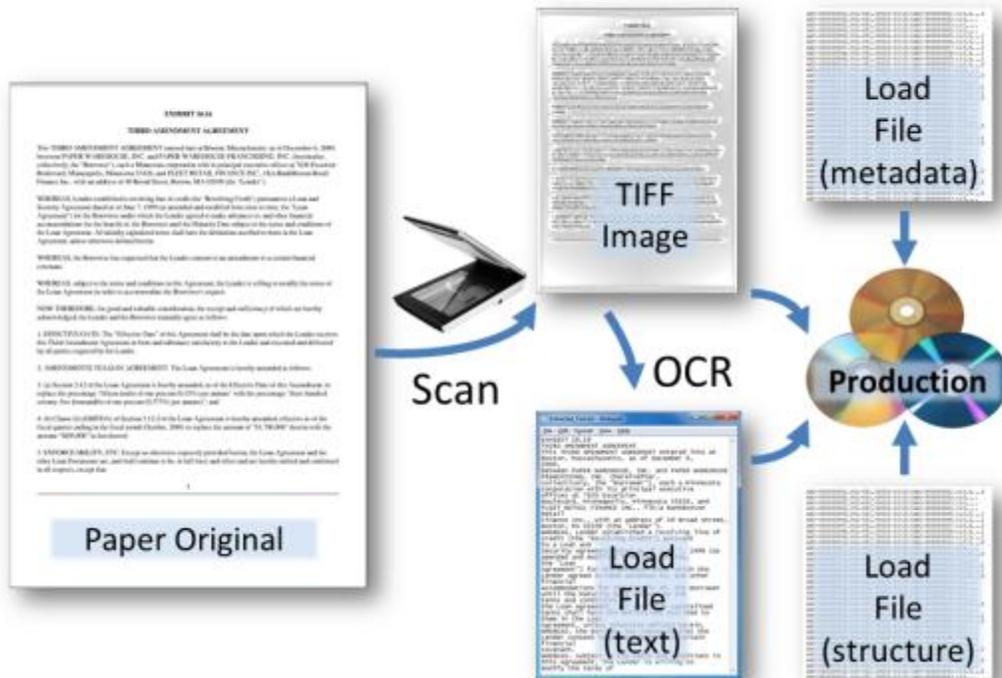
It also reminded me that not every lawyer knows about load files, those unsung digital sherpas tasked to tote metadata and searchable text otherwise lost when ESI is converted to TIFF images. Grasping the fundamentals of load files is important to fashioning a workable electronic production protocol, whether you're dealing with TIFF images, native file formats or a mix of the two.

In simplest terms, load files carry data that has nowhere else to go. They are called load files because they are used to load data into, *i.e.*, to "*populate*" a database. Load files first appeared in civil discovery in the 1980s to add electronic searchability to scanned paper documents. Then as now, paper documents were scanned to TIFF image formats and the images subjected to

optical character recognition (OCR). Unlike Adobe PDF images, TIFF images weren't designed to integrate searchable text; consequently, the text garnered using OCR was stored in simple ASCII text files named with the Bates number of the corresponding page image. Think pants with pockets versus skirts without pockets. When you use TIFF images for production, text must go *somewhere* and, since TIFFs have no "pockets," the text goes into a "purse" called a "load file." Compared to unsearchable paper documents, imaging and OCR added functionality. It was 20th century information technology improving upon 19th century printing technology, and if you were a lawyer in the Reagan-era, this was futuristic stuff.

Metadata is "data about data." While we tend to think of metadata as a feature unique to electronic documents, paper documents have metadata, too. They come from custodians, offices, files, folders, boxes and other physical locations that must be tracked. Still more metadata takes the form of codes, tags and abstracts reflecting reviewers' assessments of documents. Then as now, all this metadata needs somewhere to lodge as it accompanies page images on their journey to document review database tools (a/k/a "review platforms. This data goes into load files.

Load Files (Producing Paper Documents)



Finally, we employ load files as a sort of road map and as assembly instructions laying out, *inter alia*, where document images and their load files holding their searchable text and metadata are located on disks or other media used to store and deliver productions and how the various pieces relate to one-another.

So, to review, some load files carry extracted text to facilitate search, some carry metadata about the documents and some carry information about how the pieces of the production are stored and how they fit together. Load files are used because neither paper nor TIFF images are suited to carrying the same electronic content; and if it weren't supplied electronically, you couldn't load it into review platforms to search it using computers.

Different review platforms use different load file formats to order and separate information according to guidelines called "load file specifications." Load files employ characters called delimiters to field (separate) the various information items in the load file.

Load File Structure

Imagine creating a table to keep track of documents. You might use the first two columns of your table to number the beginning (first) and ending (last) pages of each document. The next column holds the document's file name and then each succeeding column carries information about the document (*e.g.*, Date, Author, Type). To tell one column from the next, you'd draw lines to delineate the rows and columns, like so:

BEGDOC	ENDDOC	FILENAME	MODDATE	AUTHOR	DOCTYPE
0000001	0000004	Contract	01/12/2013	J. Smith	docx
0000005	0000005	Memo	02/03/2013	R. Jones	docx
0000006	0000073	Taxes_2013	04/14/2013	H. Block	xlsx
0000074	0000089	Policy	5/25/2013	A. Dobey	pdf

The lines serve as delimiters—literally delineating one field of data from the next. Vertical and horizontal lines are excellent visual delimiters for humans, but computers work well with characters like commas or tabs. So, if the tabular data were a load file, it might be delimited as:

```
BEGDOC,ENDDOC,FILENAME,MODDATE,AUTHOR,DOCTYPE
0000001,0000004,Contract,01/12/2013,J. Smith,docx
0000005,0000005,Memo,02/03/2013,R. Jones,docx
0000006,0000073,Taxes_2013,04/14/2013,H. Block,xlsx
0000074,0000089,Policy,05/25/2013,A. Dobey,pdf
```

Each comma replaces a column divider, each line signifies another row and the first or "header" row is used to define the data that follows and the way it's delimited.

Load files using commas to separate values are called "comma separated value" or CSV files. Legacy load file formats adhere to conventions established by early review tools like Concordance and Summation, using unique delimiters that remain common today.

In e-discovery, there are three principal functions for delimiters:

- **Document Delimiter:** A Document Delimiter signals a switch from one document to the next. In the most common load file format, a carriage return/line feed serves this purpose.
- **Field Delimiter:** A Field Delimiter signals a change from one field to the next.

- **Quote Delimiter:** A Quote Delimiter permits a delimiting character to be used within the fielded data without it being treated as a field delimiter. When, for example, a comma serves as a delimiter in A CSV file, the Quote Delimiter enables the comma to be treated as a comma in the text rather than indicating a shift from one field to the next.

Concordance load files use the file extension .DAT and the **þ** (thorn, ALT-0254, Unicode 00FE) as the Quote Delimiter and the **¶** (pilcrow, ALT-0182, Unicode 00B6) character as the Field Delimiter. Note how each line denotes a single document:

Concordance Load File

```
þBEGDOCþ¶þenddocþ¶þfilenameþ¶þMODDATEþ¶þAUTHORþ¶þDOCTYPEþ
þ0000001þ¶þ0000004þ¶þContractþ¶þ01/12/2013þ¶þJ. Smithþ¶þdocxþ
þ0000005þ¶þ0000005þ¶þMemopþ¶þ02/03/2013þ¶þR. Jonesþ¶þdocxþ
þ0000006þ¶þ0000073þ¶þTaxes_2013þ¶þ04/14/2013þ¶þH. Blockþ¶þxlsxþ
þ0000074þ¶þ0000089þ¶þPolicyþ¶þ05/25/2013þ¶þA. Dobeþ¶þpdfþ
```

Summation load files use the file extension .DII, and separate each record like so:

Summation Load File

```
; Record 1
@T 0000001
@DOCID 0000001
@MEDIA eDoc
@C ENDDOC 0000004
@C PGCOUNT 4
@C AUTHOR J. Smith
@DATESAVED 01/12/2013
@EDOC \NATIVE\Contract.docx
; Record 2
@T 0000005
@DOCID 0000005
@MEDIA eDoc
@C ENDDOC 0000005
@C PGCOUNT 1
@C AUTHOR R. Jones
@DATESAVED 02/03/2013
@EDOC \NATIVE\Memo.docx
@C AUTHOR A. Dobeþ
@DATESAVED 05/25/2013
@EDOC \NATIVE\Policy.pdf
```

Two more load files:

Opticon load files (file extension .OPT) are used in conjunction with Concordance load files to pair Bates numbered pages with corresponding page images and to define the **unitization** of each document; that is, where they begin and end. Documents may be unitized *physically*, as when constituent pages are joined by clips, staples or bindings, or *logically*, where constituent pages belong together even if not physically unitized (as when documents are bulk scanned or transmittals reference enclosures). Logical unitization is also a means to track family relationships between container files and contents and e-mail messages and attachments.

Opticon Load File

```
0000001_0001,,TIFF\001\0000001_0001.tif,Y,,,4
0000002_0002,,TIFF\001\0000002_0002.tif,,,,
0000003_0003,,TIFF\001\0000003_0003.tif,,,,
0000004_0004,,TIFF\001\0000004_0004.tif,,,,
0000005_0001,,TIFF\001\0000005_0001.tif,Y,,,1
0000006_0001,,TIFF\001\0000006_0001.tif,Y,,,68
0000007_0002,,TIFF\001\0000007_0002.tif,,,,
0000008_0003,,TIFF\001\0000008_0003.tif,,,,
0000009_0004,,TIFF\001\0000009_0004.tif,,,,
0000010_0005,,TIFF\001\0000010_0005.tif,,,,
```

Opticon load files employ a simple seven-field, comma-delimited structure:

1. Page identifier,
2. Volume label (optional),
3. Path to page image,
4. New document marker (the letter “Y” for “yes” in the illustration above),
5. Box identifier (optional),
6. Folder identifier (optional),
7. Page count (optional).

Overlay load files are used to update or correct existing database content by replacing data in fields in the order in which the records occur. Thus, it’s crucial that the order of data within the overlay file match the order of data replaced. Data must be sorted in the same way, and the overlay must not add or omit fields.

Making the Case against Imaged Production

Parties don’t print their e-mail before reading it or emboss a document’s name on every page. Parties communicate and collaborate using tracked changes and embedded comments. Parties use native forms because they are the most utile, complete and efficient forms in which to store and access data.

Lawyers come along and convert native forms to images, Bates label each page and purge tracked changes and embedded comments without disclosing the destruction.

Converting a client’s ESI from its native state as kept “in its ordinary course of business” to TIFF images **injects needless expense in at least half a dozen ways:**

1. You pay to convert native forms to TIFF images and emboss Bates numbers;

2. You pay to generate load files;
3. You must produce multiple copies of documents (like spreadsheets) that are virtually incapable of production as images;
4. TIFF images and load files are much “fatter” files than their native counterparts (*i.e.*, bloated 5-40 times as large), so you pay more for vendors to ingest and host them;
5. It’s difficult to reliably de-duplicate documents once converted to images; and
6. You must reproduce everything when opponents recognize that imaged productions fall short of native productions.

REBUTTING THE CASE AGAINST NATIVE

When producing parties insists on converting ESI to TIFF despite a timely request for native production, they often rely on Federal Rules of Civil Procedure 34(b)(2)(E)(ii), which obliges parties to produce ESI in "the form or forms in which it is ordinarily maintained or in a reasonably usable form or forms." This reliance is misplaced because “[i]t is only if the requesting party declines to specify a form that the producing party is offered a choice between producing in the form ‘in which it is ordinary maintained’—native format—or ‘in a reasonably useful form or forms.’ Fed. R. Civ. P. 34(b)(2)(E)(i)-(ii)”. *The Anderson Living Trust v. WPX Energy Production, LLC*, No. CIV 12-0040 JB/LFG. (D. N.M. March 6, 2014).

Producing parties usually assert **FOUR JUSTIFICATIONS** for refusing to produce ESI in native and near-native forms. None withstand scrutiny:

1. You can't Bates label native files. Making the transition to modern forms of production requires acceptance of three propositions:

- Printouts and images of ESI are not “the same” as ESI;
- Most items produced in discovery aren’t used in proceedings; and
- Names of electronic files can be simply changed without altering contents of files.

Native documents carry more information than their imaged counterparts, and are inherently functional, searchable and complete. Moreover, native documents are described by more and different metadata—information invaluable in identifying, sorting and authenticating evidence.

Though you can’t *emboss* Bates-style identifiers on discrete pages of a native file until printed or imaged, many native forms (*e.g.*, spreadsheets, social networking content, video, and sound files) don't lend themselves to paged formats and would not be Bates labeled. When Bates-style identifiers are needed on pages for use in proceedings, simply require that file identifiers and page numbers be embossed on images or printouts. In practice, that impacts only a small subset of production.

Practice tip: It's simple and cheap to replace, prepend, or append an incrementing Bates-style identifier to a filename. One free file renaming tool is Bulk Rename Utility, available at www.bulkrenameutility.co.uk. You can even include a protective legend like "Subject to Protective Order." **Renaming a file does not alter its content, hash value or last modified date.**

2. Opponents will alter evidence. Evidence tampering is not a new fear or a hazard unique to e-discovery. Page images, being black and white pictures of text, are simple to manipulate (and Adobe Acrobat has long allowed extensive revision of PDF files).

Though any form of production is prey to unscrupulous opponents, native productions support quick, reliable ways to prevent and detect alteration. Producing native files on read-only media guards against inadvertent alteration. Alterations are easily detected by comparing hash values (digital fingerprints) of suspect files to the files produced.

Counsel savvy enough to seek native production should be savvy enough to refrain from evidence handling practices prone to alter the evidence.

3. Native production requires broader review. Native forms routinely hold user-generated content (*e.g.*, collaborative comments in Word documents, animated "off-screen" and layered text in presentations and formulae in spreadsheets) that is rarely visible on page images or intelligible in extracted text. Imaged productions often obliterate such matter *without review and without disclosure, objection or logging*. Review is only "broader" because this user-contributed content has long been furtively and indefensibly stripped away.

4. Redacting native files changes them. Change is the sole purpose of redaction. The form of production for items requiring redaction should be the form or forms best suited to efficient removal of privileged or protected content without rendering the remaining content wholly unusable.

Some native file formats support redaction brilliantly; others do not. In the final analysis, the volume of items redacted tends to be insignificant. Accordingly, the form selected for redaction shouldn't dictate the broader forms of production when, overall, native forms have decided advantages for items not requiring.

Practice Tip: **Don't let the redaction tail wag the production dog.** If an opponent wants to redact in TIFF or PDF, *let them*, but only for the redacted items and only when they restore searchability after redaction.



UPDATING YOUR REQUESTS FOR PRODUCTION

The first step in getting the information you seek in the forms you desire is to ask for it, applying the rules and eschewing dated boilerplate. Clear, specific requests are the hardest to evade and the easiest to enforce. See **Appendix: Exemplar Production Protocol, *infra***.

Most digital evidence—including e-mail—exists as data within databases. So, stop thinking about discovery as the quest for “documents” and start focusing on what you really seek: *information in utile and complete forms*.

The definition of “document” must give way to an alternate term like “information” or “information items.” Instead of the usual thesaurus-like litany of types of information, consider:

"Information items" as used here encompass individual documents and records (including associated metadata) whether on paper or film, as discrete "files" stored electronically, optically or magnetically or as a record within a database, archive or container file. The term should be read broadly to include e-mail, messaging, word processed documents, digital presentations, spreadsheets and database content.

Next, **cut junk prose** like “including, but not limited to” and “any and all.” They don’t add clarity. If you must incorporate examples of responsive items in a request, just say “including” and add an instruction that says, “Examples of responsive items set out in any request should not be construed to limit the scope of the request.” If drafting a request without “any and all” makes you quake, add the instruction, “Requests for production should be read so as to encompass any and all items responsive to the request.”

Before you serve discovery, **check your definitions** to be sure you’ve defined only terms you’ve used and used terms only in ways consistent with your definitions.

Specify the forms you seek

The most common error seen in requests for production is the failure to specify the forms sought for ESI production. Worse, requests often contain legacy boilerplate specifying forms the requesting party *doesn’t* want.

Every request for production should specify forms of production sensibly and precisely. Don’t assume that “native format” is clear or sufficient; instead, specify the formats sought for common file types, *e.g.*:

Information that exists in electronic form should be produced in native or near-native formats and should not be converted to imaged formats. Native format requires production in the same format in which the information was customarily created, used and stored in the ordinary course. The table below supplies examples of the native or near-native forms in which specific types of electronically stored information (ESI) should be produced.

Source ESI	Native or Near-Native Form or Forms Sought
Microsoft Word documents	.DOC, .DOCX
Microsoft Excel Spreadsheets	.XLS, .XLSX
Microsoft PowerPoint Presentations	.PPT, .PPTX
Microsoft Access Databases	.MDB, .ACCDB
WordPerfect documents	.WPD
Adobe Acrobat Documents	.PDF
Images	.JPG, .JPEG, .PNG
E-mail	Messages should be produced in a form or forms that readily support import into standard e-mail client programs; that is, the form of production should adhere to the conventions set out in the internet e-mail standard, RFC 5322. For Microsoft Exchange or Outlook messaging, .PST format will suffice. Single message production formats like .MSG or .EML may be furnished with folder data. If your workflow requires that attachments be extracted and produced separately from transmitting messages, attachments should be produced in their native forms with parent/child relationships to the message and container(s) preserved and produced in a delimited text file.
Databases	Unless the entire contents of a database are responsive, extract responsive content to a fielded and electronically searchable format preserving metadata values, keys and field relationships. If doing so is infeasible, please identify the database and supply information concerning the schema and query language of the database along with a detailed description of its export capabilities so as to facilitate crafting a query to extract and export responsive data.
Documents that do not exist in native electronic formats or which require redaction of privileged content should be produced in searchable .PDF formats or as single page .TIFF images with unredacted OCR text furnished and logical unitization and family relationships preserved.	

Practice Tip: In settling upon a form of production for e-mail, use this inquiry as a litmus test to distinguish “native” forms from less functional forms: ***Can the form produced be imported into common e-mail client or server applications?*** If the form of the e-mail is

so degraded that e-mail programs cannot recognize it as e-mail, that's a strong indication the form of production has strayed too far from functional.

Specify the Load File Format

Every electronic file has a complement of descriptive information called *system metadata* residing in the file table of the system or device storing the file. Different file types have different metadata. Every e-mail message has "*fields*" of information in the message "*header*" that support better searching, sorting and organization of messages. This may be data probative in its own right or simply advantageous to managing and authenticating electronic evidence. Either way, you want to be certain to request it sensibly and precisely. Simply demanding "the metadata" reveals you don't fully understand what you're seeking.

Develop a comprehensive production protocol tailored to the case and serve same with discovery. Always specifically request the metadata and header fields you seek, *e.g.*:

Produce delimited load file(s) supplying relevant system metadata field values for each information item by Bates number. Typical field values supplied include:

- a. **Source file name** (original name of the item or file when collected from the source custodian or system);
- b. **Source file path** (fully qualified file path from the root of the location from which the item was collected);
- c. **Last modified date and time** (last modified date and time of the item);
- d. **UTC Offset** (The UTC/GMT offset of the item's modified date and time, *e.g.*, -0500).
- e. **Custodian or source** (unique identifier for the original custodian or source);
- f. **Document type**;
- g. **Production File Path** (file path to the item from the root of the production media);
- h. **MD5 hash** (MD5 hash value of the item as produced);
- i. **Redacted flag** (indication whether the content or metadata of the item has been altered after its collection from the source custodian or system);
- j. **Embedded Content Flag** (indication that the item contains embedded or hidden comments, content or tracked changes); and
- k. **Deduplicated instances** (by full path).

The following additional fields shall accompany production of e-mail messages:

- l. **To** (e-mail address(es) of intended recipient(s) of the message);
- m. **From** (e-mail address of the person sending the message);
- n. **CC** (e-mail address(es) of person(s) copied on the message);
- o. **BCC** (e-mail address(es) of person(s) blind copied on the message);
- p. **Subject** (subject line of the message);
- q. **Date Received** (date the message was received);
- r. **Time Received** (time the message was received);
- s. **Attachments** (beginning Bates numbers of attachments);
- t. **Mail Folder Path** (path of the message from the root of the mail folder); and
- u. **Message ID** (unique message identifier).

Hybrid productions mixing imaged and native formats also require that paths to images and extracted text be furnished, as well as **logical unitization data** serving as the electronic equivalent of paper clips and staples.

De-duplication and Redaction

You may wish to specify whether the production should or should not be de-duplicated, *e.g.*:

Documents should be vertically de-duplicated by custodian using each document's hash value. Near-deduplication should not be employed so as to suppress different versions of a document, notations, comments, tracked changes or application metadata.

Because redaction tends to impact just a small part of most productions, it's important that it not co-opt the forms of production.

Information items that require redaction shall be produced in static image formats, *e.g.*, single page TIFF or multipage PDF images with logical unitization preserved. The unredacted content of each document should be extracted by optical character recognition (OCR) or other suitable method to a searchable text file produced with the corresponding page image(s) or embedded within the image file. Redactions should not be accomplished in a manner that serves to downgrade the ability to electronically search the unredacted portions of the item.

A TIFF-OCR redaction method works reasonably well for text documents, but often fails when applied to complex and dynamic documents like spreadsheets and databases. Unlike text, you can't spellcheck numbers, so the inevitable errors introduced by OCR make it impossible to have confidence in numeric content or reliably search the data. Moreover, converting a spreadsheet to a TIFF image strips away its essential functionality by jettisoning the underlying formulae that distinguishes a spreadsheet from a table.

Specify the medium of production

A well-crafted request should address the *medium* of ESI production; that is, the mechanism used to convey the electronic production to the requesting party. If you're receiving 100GB of data, you don't want it tendered on a dozen thumb drives.

Production of ESI should be made using appropriate electronic media of the producing party's choosing that does not impose an undue burden or expense upon a recipient.

Conclusion

It's time to take a hard look at the language of the definitions and instructions accompanying requests for production. Most are boilerplate borrowed from someone who borrowed it from

someone who drafted it in 1947. It's hand-me-down verbiage long past retirement age; so, retire it and craft modern requests for a modern digital world.

We will never be less digital than we are today. Isn't it time we demand modern evidence and obtain it in the forms in which it serves us best? We must move forms of production upstream, from depleted images and load files to functional native and near native forms retaining the content and structure that supports migration into any form. *Utile* forms. *Complete* forms. *Forms that function.*

Exemplar Production Protocol

This Appendix is an example of a *production protocol*, sometimes called a *data delivery standard*. Geared to civil litigation and seeking the lowest-cost approach to production of ESI, it seeks native production of common file types and relieves parties of the burden of converting ESI to imaged formats except when needed for redaction. This exemplar protocol specifies near-native alternatives for production of native forms when near-native forms are preferable. For an example of a U.S. Government data delivery standard, see:

<http://www.sec.gov/divisions/enforce/datadeliverystandards.pdf>

Appendix: Exemplar Production Protocol

1. "Information items" as used here encompass individual documents and records (including associated metadata) whether on paper or film, as discrete "files" stored electronically, optically or magnetically or as a record within a database, archive or container file. The term should be read broadly to include e-mail, messaging, word processed documents, digital presentations, spreadsheets and database content.
2. Information that exists in electronic form should be produced in native formats and should not be converted to imaged formats. Native format requires production in the same format in which the information was customarily created, used and stored in the ordinary course.
3. If it is infeasible to produce an item of responsive ESI in its native form, it may be produced in an agreed-upon near-native form; that is, in a form in which the item can be imported into the native application without a material loss of content, structure or functionality as compared to the native form. Static image production formats serve as near-native alternatives only for information items that are natively static images (*i.e.*, photographs and scans of hard-copy documents).
4. The table below supplies examples of agreed-upon native or near-native forms in which specific types of ESI should be produced:

Source ESI	Native or Near-Native Form or Forms Sought
Microsoft Word documents	.DOC, .DOCX
Microsoft Excel Spreadsheets	.XLS, .XLSX
Microsoft PowerPoint Presentations	.PPT, .PPTX
Microsoft Access Databases	.MDB, .ACCDB
WordPerfect documents	.WPD
Adobe Acrobat Documents	.PDF
Photographs	.JPG, .PDF
E-mail	Messages should be produced in a form or forms that readily support import into standard e-mail client programs; that is, the form of production should adhere to the conventions set out in the internet e-mail standard, RFC 5322. For Microsoft

	Exchange or Outlook messaging, .PST format will suffice. Single message production formats like .MSG or .EML may be furnished with folder data. If your workflow requires that attachments be extracted and produced separately from transmitting messages, attachments should be produced in their native forms with parent/child relationships to the message and container(s) preserved and produced in a delimited text file.
Databases	Unless the entire contents of a database are responsive, extract responsive content to a fielded and electronically searchable format preserving metadata values, keys and field relationships. If doing so is infeasible, please identify the database and supply information concerning the schema and query language of the database along with a detailed description of its export capabilities so as to facilitate crafting a query to extract and export responsive data.
Documents that do not exist in native electronic formats or which require redaction of privileged content should be produced in searchable .PDF formats or as single page .TIFF images with OCR text of unredacted content furnished and logical unitization and family relationships preserved.	

5. Absent a showing of need, a party shall produce responsive information reports contained in databases through the use of standard reports; that is, reports that can be generated in the ordinary course of business and without specialized programming efforts beyond those necessary to generate standard reports. All such reports shall be produced in a delimited electronic format preserving field and record structures and names. The parties will meet and confer regarding programmatic database productions as necessary.

6. Information items that are paper documents or that require redaction shall be produced in static image formats scanned at 300 dpi (e.g., single-page Group IV.TIFF or multipage PDF images). If an information item employs color to convey information (versus purely decorative use), the producing party shall not produce the item in a form that does not display color. The full content of each document will be extracted directly from the native source where feasible or, where infeasible, by optical character recognition (OCR) or other suitable method to a searchable text file produced with the corresponding page image(s) or embedded within the image file. Redactions shall be logged along with other information items withheld on claims of privilege.

7. Parties shall take reasonable steps to ensure that text extraction methods produce usable, accurate and complete searchable text.
8. Individual information items requiring redaction shall (as feasible) be redacted natively, produced in .PDF format and redacted using the Adobe Acrobat redaction feature or redacted and produced in another reasonable manner that does not serve to downgrade the ability to electronically search the unredacted portions of the item. Bates identifiers should be endorsed on the lower right corner of all images of redacted items so as not to obscure content.
9. Upon a showing of need, a producing party shall make a reasonable effort to locate and produce the native counterpart(s) of any .PDF or .TIF document produced. The parties agree to meet and confer regarding production of any such documents. This provision shall not serve to require a producing party to reveal redacted content.
10. Except as set out in this Protocol, a party need not produce identical information items in more than one form. The content, metadata and utility of an information item shall all be considered in determining whether information items are identical, and items reflecting different information shall not be deemed identical.
11. Production of ESI should be made using appropriate electronic media of the producing party's choosing that does not impose an undue burden or expense upon a recipient. Label all media with the case number, production date, Bates range and disk number (1 of X, if applicable). Organize productions by custodian, unless otherwise instructed. All productions should be encrypted for transmission to the receiving party. The producing party shall, contemporaneously with production, separately supply decryption credentials and passwords to the receiving party for all items produced in an encrypted or password-protected form.
12. Each information item produced shall be identified by naming the item to correspond to a Bates identifier according to the following protocol:
 - i. The first four (4) characters of the filename will reflect a unique alphanumeric designation identifying the party making production;
 - ii. The next six (6) characters will be a designation reserved to the discretionary use of the party making production for the purpose of, e.g., denoting the case or matter. This value shall be padded with leading zeroes as needed to preserve its length;
 - iii. The next nine (9) characters will be a unique, consecutive numeric value assigned to the item by the producing party. This value shall be padded with leading zeroes as needed to preserve its length;
 - iv. The final six (6) characters are reserved to a sequence consistently beginning with a

dash (-) or underscore (_) followed by a five digit number reflecting pagination of the item when printed to paper or converted to an image format for use in proceedings or when attached as exhibits to pleadings.

v. By way of example, a Microsoft Word document produced by Acme in its native format might be named: ACMESAMPLE000000123.docx. Were the document printed out for use in deposition, page six of the printed item must be embossed with the unique identifier ACMESAMPLE000000123_00006. Bates identifiers should be endorsed on the lower right corner of all printed pages, but not so as to obscure content.

vi. This format of the Bates identifier must remain consistent across all productions. The number of digits in the numeric portion and characters in the alphanumeric portion of the identifier should not change in subsequent productions, nor should spaces, hyphens, or other separators be added or deleted except as set out above.

13. Information items designated Confidential may, at the Producing Party’s option:

a. Be separately produced on electronic production media prominently labeled to comply with the requirements of the **[DATE]** Protective Order entered in this matter; or, alternatively,

b. Each such designated information item shall have appended to the file’s name (immediately following its Bates identifier) the following protective legend:
~CONFIDENTIAL-SUBJ_TO_PROTECTIVE_ORDER

When any item so designated is converted to a printed or imaged format for use in any submission or proceeding, the printout or page image shall bear the protective legend on each page in a clear and conspicuous manner, but not so as to obscure content.

14. Producing party shall furnish a delimited load file supplying the metadata field values listed below for each information item produced (to the extent the values exist and as applicable):

Field Name	Sample Data	Description
BegBates	ACMESAMPLE000000001	First Bates identifier of item
EndBates	ACMESAMPLE000000123	Last Bates identifier of item
AttRange	ACMESAMPLE000000124 - ACMESAMPLE000000130	Bates identifier of the first page of the parent document to the Bates identifier of the last page of the last attachment “child” document
BegAttach	ACMESAMPLE000000124	First Bates identifier of attachment range
EndAttach	ACMESAMPLE000000130	Last Bates identifier of attachment range
Parent_Bates	ACMESAMPLE000000001	First Bates identifier of parent document/e-mail message. <i>**This Parent_Bates field should be populated in each record representing an attachment “child” document. **</i>

Child_Bates	ACMESAMPLE000000004; ACMESAMPLE000000012; ACMESAMPLE000000027	First Bates identifier of "child" attachment(s); may be more than one Bates number listed depending on number of attachments. <i>**The Child_Bates field should be populated in each record representing a "parent" document. **</i>
Custodian	Houston, Sam	E-mail: mailbox where the email resided. Native: Individual from whom the document originated
Path	E-mail: \Deleted Items\Battles\ SanJac.msg Native: Z:\TravisWB\Alamo.docx	E-mail: Original location of e-mail including original file name. Native: Path where native file document was stored including original file name.
From	E-Mail: Davy@Crockett.net Native: D. Crockett	E-mail: Sender Native: Author(s) of document <i>**semi-colons separate multiple entries **</i>
To	Genl. A.L. de Santa Anna [mailto:sa@sa.mx]	Recipient(s) <i>**semi-colons separate multiple entries **</i>
CC	Jim.Bowie@bigknife.com	Carbon copy recipient(s) <i>**semi-colons separate multiple entries **</i>
BCC	AustinSF@state.tx.gov	Blind carbon copy recipient(s) <i>**semi-colons separate multiple entries **</i>
Date Sent	03/18/2024	E-mail: Date the email was sent
Time Sent	11:45 AM	E-mail: Time the message was sent
Subject/Title	Remember the Alamo!	E-mail: Subject line of the message
IntMsgID	<A1315BC17ABD4774BF779CB3 E3E62B9B@gmail.com>	E-mail: For e-mail in Microsoft Outlook/Exchange, the "Unique Message ID" field; For e-mail in Lotus Notes, the UNID field. Native: empty.
Date_Mod	02/23/2024	E-mail: empty. Native: Last Modified Date
Time_Mod	01:42 PM	E-mail: empty Native: Last Modified Time
File_Type	XLSX	E-mail: empty Native: file type
Redacted	Y	Denotes that item has been redacted as containing privileged content (yes/no).
File_Size	1,836	Size of native file document/email in KB.
HiddenCnt	N	Denotes presence of hidden Content/Embedded Objects in item(s) (Y/N)
Confidential	Y	Denotes that item has been designated as confidential pursuant to protective order (Y/N).
MD5_Hash	eb71a966dcddd929c1055ff2f1 ccd5b	MD5 Hash value of the item.
DeDuped	E-mail: \Inbox\SanJac.msg Native: Z:\CrockettD\Alamo.docx	Full path of deduped instances. <i>**semi-colons separate multiple entries **</i>

15. Each production should include a cross-reference load file that correlates the various files, images, metadata field values and searchable text produced.

16. Parties shall respond to each request for production by listing the Bates identifiers/ranges of responsive documents produced, and where an information item responsive to these

discovery requests has been withheld or redacted on a claim that it is privileged, the producing party shall furnish a privilege log.