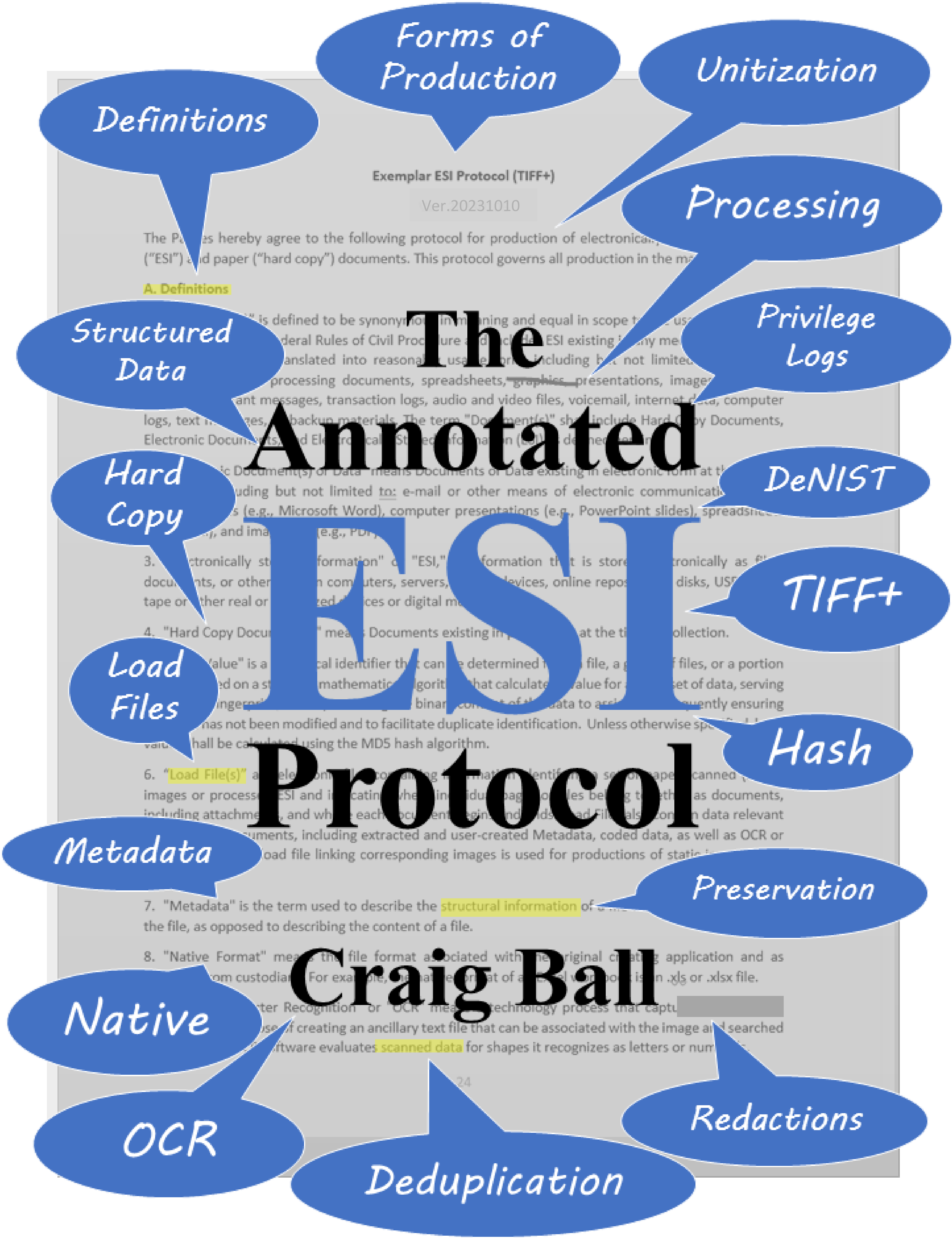

The Annotated ESI Protocol

Craig Ball

The Annotated E-Discovery Protocol: A Primer on ESI Protocols

Craig Ball ©2023

An ESI or E-Discovery Protocol is an agreement or order that answers common questions encountered when dealing with electronically stored information (ESI) in discovery, questions like:

- What forms of production should be employed?
- What metadata must be collected and produced?
- How are document "family relationships" and "unitization" handled?
- How do parties protect privileged data from and rectify inadvertent disclosure?
- What processes may producing parties use to suppress duplicates review?
- How must items produced be named and labeled?
- How is information on paper integrated with ESI production?
- How is information conveyed via color to be presented?
- How are productions efficiently transmitted and protected in transit?
- What must be made searchable by optical character recognition (OCR)?
- What must be done to resolve evidence processing exceptions and errors?
- Who serves as liaison counsel when discovery questions and disputes arise?

Ambitious ESI protocols encompass more nuanced and nettlesome issues like:

- The execution and scope of preservation duties
- Search queries and strategies
- Issues attendant to discovery from databases and other structured data sources
- Use and validation of advanced analytics
- Issues involving documents and data in foreign languages
- Confidentiality designations/legends and handling of confidential data
- The use and timing of rolling productions
- Alternative approaches to logging items withheld as privileged
- Mechanisms and timetables for dispute resolution

While it's prudent and competent to deploy an ESI protocol, anticipating consensus across too-broad a range of issues is unrealistic. *Routine ESI protocols should focus on matters of technical consistency and expediency*; that is, they should address the geeky details that ensure that what the parties exchange in discovery will be complete and utile. Yet, some parties stonewall and nitpick the most basic points of an ESI protocol in recognition that many judges—like most lawyers—are discomfited by technical disputes and retreat to solutions suited to simpler times and simpler, paper-centric discovery.

The fault for that failure lies less with Luddite judges than with advocates who can't distinguish the essential features of an ESI protocol from the merely desirable ones or articulate the "why" of either. Certainly, it's human nature to fear what we don't understand, so acceding to a different way of doing something feels risky when you don't grasp the rationale. This paper seeks to lay out the core provisions of ESI protocols, explaining their purpose and highlighting the impact of alternatives. I'll use the Federal Rules of Civil Procedure as a frame of reference, recognizing that few state courts have procedural rules entirely identical to the Federal Rules (*e.g*., not all states have a rule mirroring the FRCP's Rule 26(f) 'meet and confer' duty).[1]

A "clean" version of the exemplar protocol follows as an appendix. The example defaults to clunky TIFF+ static images as the principal form of production, so it's less efficient and economical than it could be. If you're interested in a superior protocol with lower cost and higher functionality, simply swap in the alternative native production language discussed in the Forms of Production section below.

Are ESI Protocols Compulsory?

Effectively, yes; explicitly, no. The Rules do not *expressly* require that the range of ESI-related topics on which counsel must engage be memorialized in an ESI Protocol; but where consensus exists, agreements should be memorialized as part of a discovery plan. So, *effectively* the Rules require an ESI Protocol to emerge, whether we call it that or not.

The Federal Rules of Civil Procedure require that parties confer regarding, *inter alia*:

- issues about preservation of ESI (*Rule 26(f)(3)(C)*)
- Issues about the form or forms in which ESI should be produced (*Id*.)
- Issues about claims of privilege or of protection as trial-preparation materials (*Rule 26(f)(3)(D)*)

Additionally, Rule 34(b)(1)(C) permits parties seeking production to specify the form or forms in which electronically stored information is to be produced, and it allows a party to whom the request is made to object and state the form or forms it intends to use. The 2006 Advisory Committee Comments to Rule 34 underscore that a party is not free to convert ESI to forms that

[1] You'll see this language again at the end, but I'm putting the takeaway here in case you don't get to the end: *Modern* evidence is *electronic* evidence and demands the use of electronic review tools. The *raison d'être* of an ESI Protocol is to *make productions work*, ensuring that responsive electronic evidence produced in discovery is as complete, utile and accessible as reasonably possible without exposing privileged and protected content. Modern electronic evidence resides in rich and complex information taxonomies, on systems, machines and media, in databases, accounts, folders, containers and files. Only through the meticulous management and production of data and metadata can this architecture be understood in ways essential to proving authenticity and admissibility. *These technical details matter*, and failure to attend to them thoroughly and competently prompts pernicious consequences ranging from inaccurate searches to brutally inflated review costs to losing the case because you missed probative evidence. That's the takeaway: *ESI protocols are worth fighting for, and the better both sides understand their application and purpose, the less there is to fight about.*

makes it more difficult or burdensome for the requesting party to use efficiently in the litigation or that remove or significantly degrade searchability by electronic means.

These obligations can be met by means other than an ESI Protocol, and parties are not duty bound to agree on anything. Yet, FRCP Rule 1 mandates the Rules "be construed, administered, and employed by the court and the parties to secure the just, speedy, and inexpensive determination of every action and proceeding," and judges expect lawyers to manage discovery primarily through agreement and cooperation. Isn't it just smarter that parties nail down basic discovery issues and ensure those agreements coalesce as a well-crafted ESI Protocol?

Should the Protocol be Court-Ordered?

Civil discovery was conceived as a party- and lawyer-directed process, which works well until it doesn't, at which point the Court must step in to keep discovery abuse from derailing the case. My view is, if I agree to something, I'm content to put in writing; and if I'm willing to agree to it in writing, I'm content for it to be memorialized in an order. But there's a school of thought that lawyers should afford their clients ample wiggle room in agreements, and court-ordered protocols make it difficult to adapt to the unforeseen and change direction when discovery becomes riskier, more disruptive or more costly than expected. Whether a court-ordered protocol is a guardrail or tripwire depends upon whose ox is gored.

In the final analysis, judges guard their authority more jealousy than litigants' rights; accordingly, courts tend to enforce their orders more rigorously than party agreements. If you want an ESI Protocol with teeth, get it entered as an order.

Eschew Blather and Boilerplate

Are ESI Protocols improved by stating the obvious? Many lawyers must think so because ESI Protocols can teem with blather and boilerplate. Pertinent definitions and aspirational statements defining the goals of the protocol may guide courts called on to divine the parties' intent, but paragraphs asserting that the applicable Rules apply or that discovery must be "reasonable" or "proportional" are pointless. A protocol reciting that parties must act in "good faith" or "cooperate" is no more likely to prompt salutary conduct than one silent on same. Likewise, though definitions of terms of art are helpful, defining terms never used in the protocol is sloppy. Some protocols reference e-discovery glossaries like those published periodically by The Sedona Conference. If you take that approach, be sure you can live with *all* the positions advocated by the glossary because it may contain language that will bite you in court. Also, specify the edition of the glossary agreed upon since they change over time, sometimes significantly and diametrically (*e.g.,* compare Sedona's positions on metadata across the First, Second and Third editions of The Sedona Principles). It's safer to incorporate only the definitions you need and avoid referencing materials beyond the four corners of the protocol.

Absent from the exemplar protocol language below are the customary litany of promises to meet and confer about matters left unresolved or in the face of conflicts and unforeseen complications.

Certainly, parties should seek a framework for dispute resolution short of going to court, but the obligation to confer before filing motions already exists in federal practice and most states. If the parties see a benefit to adding mandates to meet and confer respecting, *inter alia*, production of structured data, keyword search or technology-assisted review, there's no harm (albeit little benefit) to including them.

The Annotated ESI Protocol

What follows is exemplar language of the sort often seen in ESI Protocols, culled and adapted piecemeal from dozens of examples. It's certainly not "The Perfect ESI Protocol" but one crafted in the hope of achieving both a representative assemblage of protocol provisions and a measure of coherence and consistency. There are no "magic words." A suitable protocol may require tweaking to adapt to the issues and evidence in the case and, most often, to the software and capabilities of the technical staff and service providers charged to collect, process, host and produce electronic evidence.

Exemplar Protocol Language		**Explanation and Commentary**
Definitions 1. **"Document(s)" is defined to be synonymous in meaning and equal in scope to the usage of the term in Rule 34(a) of the Federal Rules of Civil Procedure and includes ESI existing in any medium from which information can be translated into reasonably usable form, including but not limited to email and attachments, word processing documents, spreadsheets, graphics, presentations, images, text files, databases, instant messages, transaction logs, audio and video files, voicemail, internet data, computer logs, text messages, and backup materials. The term "Document(s)" shall include Hard Copy Documents, Electronic Documents, and Electronically Stored Information (ESI) as defined herein.**		Definitions artfully deployed in a protocol can serve to streamline and simplify the language of the Protocol and Requests for Productions that follow. Accordingly, care should be taken to ensure that boilerplate definitions in requests conform to definitions contained in applicable protocols. Because the term "document" hearkens back to a paper-centric era of discovery, it's sensible to clarify that the term must be read expansively to include information in all its myriad forms, particularly data stored electronically, magnetically, optically and otherwise, and that "documents" encompass not only routine records (like memos, reports, presentations and ledgers) but also stored communications, like email, text messaging and collaborative communications (*e.g.,* comments as tracked change and Slack) and relevant rich media, like video and audio recordings or social networking content.

2. "Electronic Document(s) or Data" means Documents or Data existing in electronic form at the time of collection, including but not limited to: e-mail or other electronic communications, word processing files (*e.g.,* Microsoft Word), computer presentations (*e.g.,* PowerPoint slides), spreadsheets (*e.g.*, Excel), and image files (*e.g.*, PDF). **3. "Electronically stored information" or "ESI," is information that is stored electronically as files, documents, or other data on computers, servers, mobile devices, online repositories, disks, USB drives, tape or other real or virtualized devices or digital media.** **4. "Hard Copy Document(s)" means Documents existing in paper form at the time of collection.** **5. "Hash Value" is a numerical identifier that can be determined from a file, a group of files, or a portion of a file, based on a standard mathematical algorithm that calculates a value for a given set of data, serving as a digital fingerprint, and representing the binary content of the data to assist in subsequently ensuring that data has not been modified and to facilitate duplicate identification. Unless otherwise specified, hash values shall be calculated using the MD5 hash algorithm.**		

6. “Load File(s)” are electronic files containing information identifying a set of paper scanned (static) images or processed ESI and indicating where individual pages or files belong together as documents, including attachments, and where each document begins and ends. Load Files also contain data relevant to individual Documents, including extracted and user-created Metadata, coded data, as well as OCR or Extracted Text. A load file linking corresponding images is used for productions of static images (*e.g.,* TIFFs)		
7. "Metadata" is the term used to describe the structural information of a file that contains data about the file, as opposed to describing the content of a file.		Metadata remains among the most misunderstood topics in ESI discovery, encompassing not only *system metadata*, the contextual information computing devices keep about electronically stored information and stored without the file, but also *application metadata*, content about the file and stored within the file, moving with the file when copied. Examples of system metadata are a file’s name and the date the file was last modified. Examples of application metadata for a word-processed document are the date a file was last printed and tracked changes and comments.
8. "Native Format" means the file format associated with the original creating application and as collected from custodians. For example, the native format of an Excel workbook is an .xls or .xlsx file.		
9. "Optical Character Recognition" or "OCR" means a technology process that captures text from an image for the purpose of creating an ancillary text file that can be associated with the image and searched in a database. OCR software evaluates scanned data for shapes it recognizes as letters or numerals.		

10. "Searchable Text" means the native text extracted from an Electronic Document or, when extraction is infeasible, by Optical Character Recognition text ("OCR text") generated from a Hard Copy Document or electronic image.		
<u>Preservation</u> **The Parties represent that they have issued litigation hold notices to those custodians with data, and persons or entities responsible for maintenance of non-custodial data, which, based upon then-current information available, are reasonably likely to contain discoverable information.** **The Parties agree there is no need to preserve potentially relevant materials from the following sources:** **.** **1. Deleted, fragmented, or data in unallocated clusters of storage media that is only accessible by computer forensics.** **2. Volatile random-access memory (RAM), temp files, or other ephemeral data that is difficult to preserve without disabling the operating system or through the use of computer forensics.** **3. Temporary internet files, browser history files, cache files, and cookies.** **4. Back-up data that a party knows to be duplicative of ESI, documents, data or tangible things, including metadata about such information,**		ESI protocols often incorporate preservation clauses that do no more than enunciate the parties' common law duties. Unless the purpose of the provision is to narrow or expand the duty of preservation beyond the common law obligation, the provision can be dispensed with. A preservation clause may be used to identify the classes of custodians or sources that will *not* be routinely preserved, such as backup media dedicated to disaster recovery, web cache, server log files and other items that deemed not reasonably accessible or unduly burdensome.

verified to have been retained. and **5. Server, system, or network logs.**		
eDiscovery Liaison **The parties agree to designate one or more competent persons to serve as liaisons for purposes of meeting, conferring and attending court hearings regarding discovery of ESI.**		Though even the best ESI liaisons must sometimes reply, "I'll get back to you," communication and efficiency really suffer when questions filter through counsel unschooled in eDiscovery. Working through skilled liaisons that "speak geek" won't guarantee harmony but fosters focused, dispassionate diplomacy.
Databases and Structured Data **If ESI in commercial or proprietary database formats can be produced in an existing and reasonably usable, delimited report format (*e.g.,* Excel or CSV), the Parties will produce the information in such format.** **If an existing report format is not reasonably available or usable, the Parties will meet and confer to attempt to identify a mutually agreeable form of production based on the specific needs and the content and format of data within such structured data source.**		Much data sought in discovery is structured data; it resides within and is retrieved from databases. Email is a database. Social networks are databases. Financial records, health records, payroll records, customer and sales records all tend to be structured data in databases. A distinguishing feature of structured data is that it's fielded; that is, information is stored in locations dedicated to holding just that information. Fielding data serves to separate and identify information so you can search, sort and cull using just that information. It's a capability we take for granted in digital applications but can be crippled or eradicated when data is produced in e-discovery without preserving its fielded ("delimited") character. For more on databases in eDiscovery: http://www.craigball.com/Ball_DB_2010.pdf
Hard Copy Documents **Hard Copy Documents shall be scanned to single page Group IV TIFF format, 300 dpi quality or better with corresponding searchable OCR text. Image file names will be identical to the corresponding Bates numbered**		Although there's no legal duty that Hard Copy Documents be digitized, sound practice dictates that legacy paper records meld with modern digital evidence. ESI Protocols specify the form and quality of scanned items and whether and how paper records must be made text searchable. TIFF is an initialization for Tagged Image File Format, a long-used file format for storing page images as black & white pictures. "Single page" requires that

images, with a ".tif" file extension.[2] The file name of each text file should correspond to the file name of the first image file of the document with which it is associated.		each page of a document be produced as a single image file dedicated to each page. Where a 100-page file produced as a PDF would consist of a single file holding 100 pages, the same document produced in single page TIFF would consist of 100 individual files, each an image of a single page of the document. "Group IV" refers to the way the scanned image is compressed to speed transmission and optimize storage space. 300 dpi speaks to the "dots per inch," a measure of scanning and printing resolution. The higher the dots per inch, the clearer and more detailed the image; however, higher resolutions require more image data and produce larger files per page. Hard Copy Documents are inherently unsearchable electronically, so searchability may be achieved by subjecting the page images to optical character recognition (OCR). TIFF images do not store the associated text of the imaged document, so the OCR text is supplied in an accompanying file, typically a single file of text for the entire document rather than a single text file corresponding to each page. In this provision, the text file name pairs with the image file name of the first page of the document. *Note however,* Hard Copy Documents are inherently unsearchable; thus, there is no legal duty under the Rules to add searchability. The obligation to supply OCR is one the parties *choose* to take on, so apart from redacted documents, no party is *obliged* to supply OCR text absent an agreement or order. Because this provision demands an image be produced for each page, Bates numbering ensures filenames are unique and pages are produced sequentially. This requires that page images be created (or renamed) using software that supports

[2] Bates numbering has historically been employed as an organizational method to label and identify legal documents, especially those produced in discovery. "Bates" is capitalized because the name derives from the Bates Manufacturing Company, which patented and sold auto-incrementing, consecutive-numbering stamping devices. Bates numbering serves the dual function of sequencing and uniquely identifying documents.

		Bates numbering and careful attention paid to avoid reusing sequences from prior productions. **Comment:** This provision is as close to an enduring, industrywide standard as exists despite serious shortcomings. We are captive to 80's era technology when it comes to scanned hard copies. TIFF images tend to be much larger files than the same document supplied as a PDF image, making TIFF productions more expensive to host online and slower to appear onscreen. Unlike PDFs, TIFFs convert color data to black and white, a sometimes-serious downgrading of the evidence. The 300-dpi resolution works well enough for letters and reports but may be insufficient to adequately display technical drawings and fine details.
Unitizing Documents **In scanning Hard Copy Documents, distinct documents should not be merged into a single record, and single documents should not be split into multiple records (*i.e.*, paper documents should be logically unitized). For example, Hard Copy Documents stored in a binder, folder, or similar container should be produced in the same order as they appear in the container. The front cover of the container should be produced immediately before the first document in the container. The back cover of the container should be produced immediately after the last document in the container. The Parties will undertake reasonable efforts to, or have their vendors, logically unitize documents correctly, and will commit to address situations of improperly unitized documents.**		"Unitization" refers to the organization of pages into a document, chapter or volume. Paper documents are physically unitized by means of, *e.g.,* clips, staples, bindings and folders. Multiple documents may comprise a "family" unit; for example, a transmittal and its attachments or a report and its exhibits/appendices comprise a parent/child relationship. When unitized paper records are scanned, metadata supplies a logical unitization of files mirroring the physical unitization of the physical document or volume scanned. For documents that contain affixed notes, pages may be scanned once with the notes as they appear on the page and again without the notes, so all content is captured. The relationship of documents in a document collection should be maintained throughout scanning, and processing (*e.g.,* cover letter and enclosures, e-mail and attachments, binder holding multiple documents, folder and other compilations where a parent-child relationship exists between the documents). For ESI, the keys to preserving unitization lie in both the ordering of documents by Bates numbers and the metadata supplied in load files.

Parent-Child Relationships **The Parties agree that if any part of a Document or its attachments is responsive, the entire Document and attachments will be produced, except any attachments that must be withheld or redacted and logged based on privilege or work-product protection.** **The Parties shall take reasonable steps to ensure that parent-child relationships within a document family (the association between an attachment and its parent document) are preserved. The child document(s) should be consecutively produced immediately after the parent document. For further clarification, this shall not require a party to produce documents merely referenced in responsive documents; provided, however, that documents sent via a link within an email should be produced.**		Few things are as frustrating in a production review as being unable to pair a "parent" transmittal with its "child" attachments. This provision reflects the custom of extracting child attachments from the parent transmittal and supplying them *seriatim*. Too, it touches on potentially-fractious *scope* of discovery issues by requiring producing parties to treat a document family as a single item to be produced if any component is responsive (although any part may be withheld or redacted on claim of privilege). A producing party may resist, arguing that discovery allows for granular treatment of the family and does not require production of non-responsive attachments or transmittals. Note that the exemplar language obliges the parties to produce hyperlinked files or so-called "modern attachments." The parties must appreciate what this obligation entails in the context of their messaging environment. Some Cloud systems (*e.g.,* Microsoft 365) make it easy to collect documents transmitted as hyperlinked files versus embedded attachments, whereas others may demand manual collection with attendant uncertainty as to whether the item collected remains faithful to the item transmitted. As phrased, the operative distinction is whether the hyperlink in the transmittal points to a resource readily available to anyone with the link (that is, "documents merely referenced") or whether the modern attachment item is unavailable to the requesting party if not produced with the transmittal.
Hard Copy Document Metadata **The following metadata fields should be provided for Hard Copy Documents when reasonably available:** **1. Beginning Bates number** **2. Ending Bates number** **3. First attachment Bates number** **4. Last attachment Bates number** **5. Source location/custodian** **6. Confidentiality designation**		Paper documents have metadata, too, some of it essential for proper unitization and management. In the example, note that the eight data points required are not usually found within a document. Instead, these metadata values are either collected (like source location/custodian) or (like Bates numbers), assigned as part of an ESI processing and production workflow.

7. Redacted (Y/N) and **8. Extracted/OCR text file path.**		
<u>**Forms of Production**</u> ***Alternative 1: Native Production*** **The Parties will produce Electronic Documents, Data and ESI in Native Formats with the metadata specified in ADDENDUM A. Redacted ESI may be redacted natively, as feasible, or produced as redacted TIFFs with applicable, non-privileged metadata and OCR searchable text.** **Electronic Documents, Data and ESI will be Bates numbered by substituting, prepending or appending the Bates number for/to the file name. When any party prints produced ESI for use in a filing or proceeding, such party shall ensure that the Bates number of the item, any required confidentiality notices and pagination are embossed on the face of the printed item without obscuring its content.[3]**		Establishing the form or forms of production is the centerpiece of any ESI protocol, and the feature with the greatest influence on the cost of processing and hosting the data. Here, alternative clauses specify native or TIFF+ as the default form of production for ESI. Note that each approach borrows from the other in that native productions provide that redacted data be supplied in TIFF formats, and TIFF+ productions contemplate that ESI that doesn't lend itself to static imaging be produced natively. Native forms ensure a level playing field between producing and requesting parties in that a native production will faithfully mirror the ways in which the custodians view and work with evidence. Colors and functional features are preserved, along with tracked changes and comments appearing in original files. Above all, native forms are massively smaller in size versus TIFF images created from the native file. Consequently, native productions are many times less costly to load and host when eDiscovery vendors price services based on the byte volume of the data.[4]

[3] A common question is, "How do we Bates number native productions?" Because electronic files often have the same file names, the best practice is to replace the native filename with a unique Bates number and supply the original filename, paired with its Bates number, in the accompanying load file. An alternative is to ensure the filenames are unique by prepending or appending the Bates number to the filename. To facilitate page level references by Bates number when a party prints a native document for use in a deposition or proceeding, the Protocol requires that parties emboss the native file's Bates numbers and pagination on the printed document, just as with TIFF+ productions. Thus, when parties *change* the form of the evidence post-production (*e.g.,* native-to-paper), the party changing the evidence is obliged to preserve the connection between the native source and the paginated printout.

[4] Whether in native or static image format, ESI must be processed ("ingested") and hosted to be searchable and reviewable. Native forms are processed to extract their text and metadata, then indexed for search. TIFF and load file productions are indexed for search and processed to pair the page images with text and metadata. Either way, you pay a vendor to prepare the production for viewing and then pay a recurring "hosting" charge for online access to the production. The fees charged are based on the volume of data processed and/or hosted. More data costs more money. If you receive 10 times as much data, you pay a commensurate amount more to ingest and host.

OR

Alternative 2: TIFF+ Production

The Parties will produce Electronic Documents, Data and ESI as single page Group IV TIFF images, 300 dpi quality or better, and 8.5”x11” page size, except for documents requiring different resolution or page size with the metadata specified in ADDENDUM A. *However,* the Parties will produce the following forms of ESI in native formats:

1. **Spreadsheets**
2. **PowerPoint presentations**
3. **Access databases**
4. **Delimited text files**
5. **Photographs**
6. **Audio and video files**
7. **Documents of a type which cannot be reasonably converted to useful TIFF images.**

All images of documents which contain tracked changes such as comments, deletions and revision marks (including the identity of the person making the deletion or revision and the date and time thereof), speaker notes, or other user-entered data that the source application can display to the user will be processed such that all that data is visible in the image.

Parties favoring TIFF+ point to a diminished potential for fraudulent or inadvertent alteration of the evidence and the ability to emboss a Bates number on the face of a page image versus naming the produced files to their Bates numbers. Also, TIFF images may be viewed in any browser, though they won’t be text searchable doing so.

When converting electronic documents to static images, parties must consider the wealth of information users see in the native application like tracked changes and comments between collaborators in word processed documents and speaker notes in presentations. Do you require these items be made visible on the page images or leave them out of the production? The exemplar language takes the first path, but each approach has its pitfalls. Producing the document both ways doubles volume and expense. Native productions solve this issue as a native production affords requesting parties comparable access to content as the custodian of the evidence.

When parties convert evidence in native forms to static image forms like TIFF, that process strips away all electronic searchability. A monochrome screenshot replaces the source evidence. Since the Federal Rules of Civil Procedure say parties can’t remove or significantly degrade searchability, responding parties must act to restore a measure of searchability. They do this by extracting text from the native ESI and delivering it in a “load file” accompanying the page images. This (and metadata) is the “plus” when people speak of “TIFF+” productions.

To search a TIFF+ production, page images and load files must be hosted in an eDiscovery review

Vendors usually assess hosting fees as a monthly subscription, so the more data they host for you, the more you pay every month for the life of the case. More data isn’t the same thing as more information because not all electronic forms of information are equally efficient. When you convert native forms to static images and load files you explode the size of production by many multiples, and static productions come burdened by the further cost of impaired searchability, diminished functionality and lost color, animation and rich media.

File Names **Each TIFF image should have a unique file name corresponding to the Bates number of that page with a “.tif” file extension. The file name should not contain any blank spaces and should be zero-padded (e.g., DEF-000001), taking into consideration the estimated number of pages to be produced. If a Bates number or set of Bates numbers is skipped in a production, Producing Party will so note in a cover letter or production log accompanying the production. Bates numbers will be unique across the entire production and prefixes will be consistent across all documents produced.** **Producing Party will brand all TIFF images in the lower right-hand corner with its corresponding Bates number without obscuring any part of the underlying image.** **Extracted Text Files** **For each document, a single Unicode text file containing extracted text shall be provided along with the image files and metadata. The text file name shall be the same as the Bates number of the first page of the document. File names shall not have any special characters or embedded spaces. Electronic text must be extracted directly from the native electronic file to the extent reasonably feasible unless the**		“platform” capable of pairing the extracted text with the corresponding page images.[5] Once more—and unlike native files and PDFs--TIFF images are merely black-and-white pictures of pages and cannot be searched for words or phrases. They hold no text. To facilitate searchability, the text of documents must be produced in separate load files meant to be loaded into review software. Searches are then run against the text file data (more accurately, an index of created from that text) and, because the Bates numbered text files share names with the Bates numbered image files, search hits within text ties to page images. This is only possible when naming conventions are adhered to, hence attendant language of the protocol. Also, because the text of a document may include foreign languages and specialized characters, the provision requires that the text be produced as Unicode text, meaning

[5] This process can operate to materially impair accurate search as in https://craigball.net/2020/01/15/degradation-how-tiff-disrupts-search/

document is an image file or contains redactions, in which case, a text file created using OCR should be produced in lieu of extracted text.		that it must be encoded to support a wide array of international characters versus the paltry 256 characters of the once-ubiquitous ASCII encoding.[6]
Load Files **Productions will, as applicable, include image load files in Opticon or IPRO format as well as Concordance format data (.dat) files with the applicable metadata fields identified in ADDENDUM A. All metadata will be produced in UTF-16LE or UTF-8 with Byte Order Mark format.** **All native format files shall be produced in a folder named "NATIVE,"** **All TIFF images shall be produced in a folder named "IMAGE," which shall contain sub-folders named "0001," "0002," etc. Each sub-folder shall contain no more than 10,000 images. Images from a single document shall not span multiple sub-folders.** **All extracted Text and OCR files shall be produced in a folder named "TEXT."** **All load files shall be produced in a folder named "DATA" or at the root directory of the production media.**		Load files are used to import image, native, and text files and their corresponding metadata and production information into a document database or "review tool". Load files carry indispensable information, such as file names, file locations (both their origination and within a production), sources, custodians and dates. The information in load files enables search, sorting, tracing, authentication, unitization and much more. They are the Rosetta Stones of ESI production. The references to Opticon, IPRO and Concordance do not oblige a party to use a particular vendor or software; instead, those are shorthand ways to designate the *structure* of the load files and of the delimiters ("character separators") employed to distinguish one field of metadata from the next. "UTF" stands for Unicode Transformation Format, a universal way to encode alphanumeric character sets for worldwide consistency and intelligibility. For more on load files: https://craigball.net/2013/07/17/a-load-file-off-my-mind/

[6] ASCII is an acronym for American Standard Code for Information Interchange and describes one of the oldest and simplest standardized ways to use numbers—particularly binary numbers expressed as ones and zeroes–to denote a basic set of English language alphanumeric and punctuation characters.

Color **Paper documents or redacted ESI that contain color used to convey information (*e.g.*, color coding and highlighting versus merely decorative use) shall be produced as single-page, 300 DPI JPG images with JPG compression set to its highest-quality setting so as not to not degrade the original image.** **OR** **Where .TIFF images are illegible due to color content (such as colored text on a colored background) or where color is material to the interpretation of a document, JPG image files shall be provided upon reasonable request.**		JPG images and native productions show color, but TIFF images are black and white renderings, so an unsuitable form of production when color is used to convey information. Some protocols address the problem by allowing requesting parties to make *ad hoc* requests for reproduction of items in forms supporting color. The obvious problem is that it's often impossible to discern the use of color working from a black and white image. Some eDiscovery software tools offer the ability to detect the use of color in a file and can programmatically pivot the form of production between TIFF and JPG formats. As a rule, JPG images should *always* be produced when the source evidence is a JPG image (*e.g.*, a photograph). Email transmittals frequently contain decorative color (in logos), so best lend themselves to *ad ho*c requests for color reproduction. PowerPoint presentations and Excel spreadsheets should never be produced in anything but native formats (where color is natively supported).
Redactions **Any redacted material must be clearly labeled on the face of the document as having been redacted and shall be identified as such in the load file provided with the production. Each redacted document shall be produced with an OCR *.txt file containing unredacted text. A document's status as redacted does not relieve the producing party from providing all the metadata required herein unless the metadata withheld is contains privileged content.**		ESI documents can contain both apparent and non-obvious content. For example, PDFs often include an image layer and a textual layer such that altering the image won't change the searchable text. Accordingly, ESI poses unique challenges when a document contains privileged and non-privileged information. Although many forms of ESI are easy to redact reliably in their native formats and privileged content can be expurgated without impairing the searchability of non-privileged content, lawyers tend not to trust native redaction. Instead, they demand that "blacked out" TIFF images be used for redaction even when all other documents are produced natively. This requires searchability be restored for the unredacted content; and since text extraction might grab privileged content, OCR is used instead.
Privilege Logs		The obligation to furnish a privilege log is governed by the applicable Rules of Civil Procedure, *e.g.,* Fed. R.

With each production, Producing Party shall supply a log of the documents withheld or redacted under a claim of privilege and/or work product with sufficient information to allow the Receiving Party to understand the basis for the claim. **Communications involving trial counsel that post-date the filing of the complaint need not be placed on a privilege log.**		Civ. P. 26(b)(5)(A). Privilege logs don't implicate unique technical concerns except to the extent that a Producing Party seeks a "metadata privilege log" or a "categorical privilege log," to avoid the description duties required in the Rules. The exemplar language includes a categorical exemption for post-suit communications with trial counsel. Commentary: Though ESI protocols often address privilege logs, the timing and scope of privilege logs is best addressed in an agreement incorporating a liberal clawback and non-waiver provision and, in federal court, a Federal Rule of Evidence 502(d) order governing inadvertent production of privileged information.[7]
Deduplication ***Vertical Deduplication*** **Producing Party may vertically de-duplicate documents based on MD5 or SHA-I hash values at the document level, by Message ID, EDRM MIH[8] or other standard methodology for email deduplication within the collection of a custodian or a data source. Attachments to parent documents may not be deduplicated against a duplicate standalone version of the attachment exists, and standalone versions of documents may not be suppressed if a duplicate version exists as an attachment.** **OR**		Parties should endeavor to produce a single copy of each responsive document while identifying unproduced duplicates via their metadata values in load files. In this way, Receiving Parties are not burdened by production of duplicates yet can determine which custodians possessed duplicates and, *inter alia,* know the unique dates, names and locations of deduplicated instances. *Vertical deduplication* refers to deduplication within the collection of a single source or custodian, differentiated from *horizontal or global deduplication* where deduplication spans the collections of multiple sources or custodians. MD5 and SHA-1 are standard cryptographic hash algorithms, mathematical formulas that calculate a fixed length value for a given binary input of any size. These hash values serve as digital fingerprints of the

[7] A clawback provision governs what parties must do when there's been an inadvertent disclosure of privileged information: issues such as disclosure, sequestration, return, destruction, non-use and non-waiver. Such provisions are designed to minimize the harm flowing from unwitting disclosure and, crucially, to forestall the dread "subject matter waiver" whereby the release of even a narrow range of privileged material may serve to "open the door" to all privileged material touching on the subject matter of the inadvertent disclosure.

[8] The EDRM MIH (for Message Identification Hash) is a unique identifier enabling cross platform email duplicate identification. Specifications and information about the EDRM MIH are found at https://edrm.net/edrm-projects/dupeid-2/ and a white paper describing the MIH is here: https://edrm.net/download/161805

Horizontal Deduplication **Producing Party may horizontally (globally) de-duplicate documents based on MD5 or SHA-I hash values at the document level or by Message ID, EDRM MIH or other standard methodology for email deduplication within the collection of a custodian or a data source. Attachments to parent documents may not be deduplicated against a duplicate standalone version of the attachment exists, and standalone versions of documents may not be suppressed if a duplicate version exists as an attachment.** **Producing Party will track all deduplicated files and provide the names of all custodians of these duplicates of in the load file. If the duplicates are e-mails, the producing party must detail the process of creating the hash value, *e.g.*, the names and order of concatenated fields by which the deduplication hash was calculated.**		binary content of files to facilitate duplicate identification. E-Discovery service providers apply employ varying methods to calculate a hash value for email messages and attachments. The exemplar language provides that, whatever method is used won't be implemented in a way that would make it difficult to distinguish documents made attachments to email transmittals from the same documents existing as standalone files.
De-NISTing **System and application files without user created content (as identified by matching to the NIST National Software Reference Library database) need not be processed, reviewed or produced.**		The National Software Reference Library, part of the U.S. National Institute for Standards and Technology, compiles and distributes digital signatures for software, including the files comprising most operating systems and commercial applications. Because the constituents of commercial software are seldom relevant evidence in civil cases, excluding these from eDiscovery fosters efficiency.
Email Threading **To reduce the volume of entirely duplicative content within email threads, the parties may, but are not required to, use email**		When email messages are produced as static images, email threading simplifies review by presenting all messages that comprise an email conversation as a continuous, temporally-ordered "thread." The

threading. A party may use industry standard message threading technology to remove email messages where the content of those messages, and any attachments, are wholly contained within a later email message in the thread; *provided, however*, that the use of threading must not serve to obscure whether a recipient received an attachment.		objection most often voiced is that threading may serve to suppress a message or attachment
Production Media **The producing party will use the appropriate electronic media (DVD, thumb drive, hard drive or secure FTP transfer) for its ESI production and will endeavor to use the highest capacity suitable media. The producing party will label the production media with the name of the producing party, production date, media volume name, and Bates number range(s).** **Productions on physical media should be encrypted for transmission to the Receiving Party. At the time of production and under separate cover, Producing Party shall furnish decryption credentials to Receiving Party.**		ESI protocols specify both the *form* of production and the *medium* of production, the former being the file types to be supplied and the latter the type of storage device used to hand off the data. Production media should be selected to minimize the number of disks or drives required for transfer, although that's a concern tied to the era of floppy disks and optical disks and not an issue with today's huge hard drives. Parties should ensure that the contents of production media are encrypted, both to protect against loss in transit and to guard against unauthorized access. Care should be taken not to transmit encrypted data with decryption passwords and to never label or store encrypted media with its decryption credentials.
Processing **The Parties will use reasonable efforts and standard industry practices to address and resolve exception issues for items that present processing, imaging or form of production problems (including encrypted, corrupt and/or protected files identified**		For more about processing: http://www.craigball.com/Ball_Processing_2019.pdf

during the processing of ESI). The Parties will meet and confer regarding procedures that will be used to identify, access, and process and resolve exception issues. **Parties shall normalize times and dates to conform to [UTC] or [specified local time zone].** **For archive files (zip, jar, rar, gzip, TAR, etc.), all contents should be extracted from the archive with source pathing and family relationships preserved and produced. The fully unpacked archive container file does not need to be included in the production.**		
<u>**Non-Waiver**</u> **This Protocol is solely intended to address the format of document productions and does not limit the temporal or substantive scope of discovery. Nothing in this Protocol is intended to affect the right of any party to object to a request for production or to operate as a waiver of any party's right to promulgate, object to, or seek relief from a request for discovery.**		

<u>**Metadata Production Fields**</u>

The exemplar ESI protocol above contemplates that the parties will agree upon the metadata fields that will be extracted or populated and produced in the load file. Different forms of ESI hold different application metadata, and some metadata isn't collected with or extracted from the ESI but must be assigned or calculated when the data is processed. Custodians are typically determined at collection and designated when their data is ingested by eDiscovery software for processing. A hash value is calculated for each file. A Bates number is assigned to each file or page image. Not every eDiscovery vendor can supply every field below, and some use different field names for the same data.

ADDENDUM A

Field Name	**Description**
BegBates	First Bates identifier of item
EndBates	Last Bates identifier of item
PgCount	Number of pages in the document
FileSize	Size of native file document/email in KB
FileName	Original name of file as it appeared in location where collected
Path	E-mail: Original location of e-mail including original file name Native: Originating path where native file document was collected including original file name
NativeLink	Relative path and filename for native file on production media
TextLink	Relative path and filename for text file on production media
AttRange	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	First Bates identifier of attachment range
EndAttach	Last Bates identifier of attachment range
AttachCount	Number of attachments to an e-mail
AttachName	Names of each individual Attachment, separated by semicolons
ParentBates	First Bates identifier of parent document/e-mail message (will not be populated for documents that are not part of a family)
ChildBates	First Bates identifier of “child” attachment(s); may be more than one Bates number listed depending on number of attachments (will not be populated for documents that are not part of a family)
Custodian	E-mail: mailbox where the email resided Native: Individual from whom the document originated
OtherCustodians	Custodians whose file/message has been de-duplicated; separated by semicolons

From	E-mail: Sender Native: Author(s) of document; separated by semicolons
To	E-mail: Recipient(s); separated by semicolons
CC	E-mail: Carbon copy recipient(s); separated by semicolons
BCC	E-mail: Blind carbon copy recipient(s) separated by semicolons
DateSent (mm/dd/yyyy hh:mm:ss AM or PM)	E-mail: Date and time the email was sent
Subject	E-mail: Subject line of email
Title	Document: Title provided by user within the document
MsgID	E-mail: "Unique Message ID" field
EDRM_MIH	Identifier enabling cross platform email duplicate identification. Specifications and information about the EDRM MIH are found at https://edrm.net/edrm-projects/dupeid-2/
ModifiedDate (mm/dd/yyyy hh:mm:ss AM or PM)	Document: Last Modified Date and time
CreationDate (mm/dd/yyyy hh:mm:ss AM or PM)	Document: Create Date and time
FileExt	Document: file extension
FileType	Document: file type
Redacted	Denotes that item has been redacted as containing privileged content (yes/no)
Hash	MD5 Hash value of the item
HiddenContent	Denotes presence of Tracked Changes/Hidden Content/Embedded Objects in item(s) (Y/N)
Confidential	Denotes that item has been designated as confidential pursuant to confidentiality agreement or protective order (Y/N)

DeDuped	Full path of deduplicated instances; separated by semicolons

Takeaway

By now, you may be marveling at the persnickety technical details requiring precise management to enable lawyers to view and search ESI productions. Alternatively, you may be bored and irritated at having to deal with any of this…stuff. If it strikes you as fussy, then you're probably not the person responsible for making it work.

Modern evidence is *electronic* evidence and demands the use of electronic review tools. The *raison d'être* of an ESI Protocol is to *make productions work*, ensuring that responsive electronic evidence produced in discovery is as complete, utile and accessible as reasonably possible without exposing privileged and protected content. Modern electronic evidence resides in rich and complex information taxonomies, on systems, machines and media, in databases, accounts, folders, containers and files. Only through the meticulous management and production of data and metadata can this architecture be understood in ways essential to proving authenticity and admissibility. *These technical details matter*, and failure to attend to them thoroughly and competently prompts pernicious consequences ranging from inaccurate searches to brutally inflated review costs to losing the case because you missed probative evidence. That's the takeaway: *ESI protocols are worth fighting for, and the better both sides understand their application and purpose, the less there is to fight about.*

Exemplar ESI Protocol (TIFF+)

Ver. 20231010

The Parties hereby agree to the following protocol for production of electronically stored information ("ESI") and paper ("hard copy") documents. This protocol governs all production in the matter.

A. Definitions

1. "Document(s)" is defined to be synonymous in meaning and equal in scope to the usage of the term in Rule 34(a) of the Federal Rules of Civil Procedure and includes ESI existing in any medium from which information can be translated into reasonably usable form, including but not limited to email and attachments, word processing documents, spreadsheets, graphics, presentations, images, text files, databases, instant messages, transaction logs, audio and video files, voicemail, internet data, computer logs, text messages, or backup materials. The term "Document(s)" shall include Hard Copy Documents, Electronic Documents, and Electronically Stored Information (ESI) as defined herein.

2. "Electronic Document(s) or Data" means Documents or Data existing in electronic form at the time of collection, including but not limited to e-mail or other means of electronic communications, word processing files (e.g., Microsoft Word), computer presentations (e.g., PowerPoint slides), spreadsheets (e.g., Excel), and image files (e.g., PDF).

3. "Electronically stored information" or "ESI," is information that is stored electronically as files, documents, or other data on computers, servers, mobile devices, online repositories, disks, USB drives, tape or other real or virtualized devices or digital media.

4. "Hard Copy Document(s)" means Documents existing in paper form at the time of collection.

5. "Hash Value" is a numerical identifier that can be determined from a file, a group of files, or a portion of a file, based on a standard mathematical algorithm that calculates a value for a given set of data, serving as a digital fingerprint, and representing the binary content of the data to assist in subsequently ensuring that data has not been modified and to facilitate duplicate identification. Unless otherwise specified, hash values shall be calculated using the MD5 hash algorithm.

6. "Load File(s)" are electronic files containing information identifying a set of paper scanned (static) images or processed ESI and indicating where individual pages or files belong together as documents, including attachments, and where each document begins and ends. Load Files also contain data relevant to individual Documents, including extracted and user-created Metadata, coded data, as well as OCR or Extracted Text. A load file linking corresponding images is used for productions of static images (e.g., TIFFs)

7. "Metadata" is the term used to describe the structural information of a file that contains data about the file, as opposed to describing the content of a file.

8. "Native Format" means the file format associated with the original creating application and as collected from custodians. For example, the native format of an Excel workbook is an .xls or .xlsx file.

9. "Optical Character Recognition" or "OCR" means a technology process that captures text from an image for the purpose of creating an ancillary text file that can be associated with the image and searched in a database. OCR software evaluates scanned data for shapes it recognizes as letters or numerals.

10. "Searchable Text" means the native text extracted from an Electronic Document or, when extraction is infeasible, by Optical Character Recognition text ("OCR text") generated from a Hard Copy Document or electronic image.

B. Preservation

The Parties represent that they have issued litigation hold notices to those custodians with data, and persons or entities responsible for maintenance of non-custodial data, which, based upon then-current information available, are reasonably likely to contain discoverable information.

The Parties agree there is no need to preserve potentially relevant materials from the following sources:

1. Deleted, fragmented, or data in unallocated clusters of storage media that is only accessible by computer forensics.

2. Volatile random-access memory (RAM), temp files, or other ephemeral data that is difficult to preserve without disabling the operating system or through the use of computer forensics.

3. Temporary internet files, browser history files, cache files, and cookies.

4. Back-up data that a party knows to be duplicative of ESI, documents, data or tangible things, including metadata about such information, verified to have been retained; and

5. Server, system, or network logs.

C. eDiscovery Liaison

The parties agree to designate one or more competent persons to serve as liaisons for purposes of meeting, conferring and attending court hearings regarding discovery of ESI.

D. Databases and Structured Data

If ESI in commercial or proprietary database formats can be produced in an existing and reasonably usable, delimited report format (*e.g.*, Excel or CSV), the Parties will produce the information in such format.

If an existing report format is not reasonably available or usable, the Parties will meet and confer to attempt to identify a mutually agreeable form of production based on the specific needs and the content and format of data within such structured data source.

E. Hard Copy Documents

Hard copy documents shall be scanned to single page Group IV TIFF format, 300 dpi quality or better with corresponding searchable OCR text. Image file names will be identical to the corresponding Bates numbered images, with a ".tif” file extension. The file name of each text file should correspond to the file name of the first image file of the document with which it is associated.

F. Unitizing Documents

In scanning hard copy documents, distinct documents should not be merged into a single record, and single documents should not be split into multiple records (i.e., paper documents should be logically unitized). For example, hard copy documents stored in a binder, folder, or similar container should be produced in the same order as they appear in the container. The front cover of the container should be produced immediately before the first document in the container. The back cover of the container should be produced immediately after the last document in the container. The Parties will undertake reasonable efforts to, or have their vendors, logically unitize documents correctly, and will commit to address situations of improperly unitized documents.

G. Parent-Child Relationships

The Parties agree that if any part of a Document or its attachments is responsive, the entire Document and attachments will be produced, except any attachments that must be withheld or redacted and logged based on privilege or work-product protection.

The Parties shall take reasonable steps to ensure that parent-child relationships within a document family (the association between an attachment and its parent document) are preserved. The child document(s) should be consecutively produced immediately after the parent document. For further clarification, this shall not require a party to produce documents merely referenced in responsive documents; provided, however, that documents sent via a link within an email should be produced.

H. Hard Copy Document Metadata

The following metadata fields should be provided for hard copy documents when reasonably available:

1. Beginning Bates number
2. Ending Bates number
3. First attachment Bates number
4. Last attachment Bates number
5. Source location/custodian

6. Confidentiality designation
7. Redacted (Y/N) and
8. Extracted/OCR text file path.

I. Forms of Production

TIFF+ Production

The Parties will produce Electronic Documents, Data and ESI as single page Group IV TIFF images, 300 dpi quality or better, and 8.5”x11” page size, except for documents requiring different resolution or page size with the metadata specified in Addendum A. However, the Parties will produce the following forms of ESI in native formats:

1. Spreadsheets
2. PowerPoint presentations
3. Access databases
4. Delimited text files
5. Photographs
6. Audio and video files
7. Documents of a type which cannot be reasonably converted to useful TIFF images.

All images of documents which contain tracked changes such as comments, deletions and revision marks (including the identity of the person making the deletion or revision and the date and time thereof), speaker notes, or other user-entered data that the source application can display to the user will be processed such that all that data is visible in the image.

J. File Names

Each TIFF image should have a unique file name corresponding to the Bates number of that page with a “.tif” file extension. The file name should not contain any blank spaces and should be zero-padded (e.g., DEF-000001), taking into consideration the estimated number of pages to be produced. If a Bates number or set of Bates numbers is skipped in a production, Producing Party will so note in a cover letter or production log accompanying the production. Bates numbers will be unique across the entire production and prefixes will be consistent across all documents produced.

Producing Party will brand all TIFF images in the lower right-hand corner with its corresponding Bates number without obscuring any part of the underlying image.

K. Extracted Text Files

For each document, a single Unicode text file containing extracted text shall be provided along with the image files and metadata. The text file name shall be the same as the Bates number of the first page of the document. File names shall not have any special characters or embedded spaces. Electronic text must be extracted directly from the native electronic file to the extent reasonably feasible unless the document is an image file or contains redactions, in which case, a text file created using OCR should be produced in lieu of extracted text.

L. Load Files

Productions will, as applicable, include image load files in Opticon or IPRO format as well as Concordance format data (.dat) files with the applicable metadata fields identified in Attachment A. All metadata will be produced in UTF-16LE or UTF-8 with Byte Order Mark format.

All native format files shall be produced in a folder named "NATIVE,"

All TIFF images shall be produced in a folder named "IMAGE," which shall contain sub-folders named "0001," "0002," etc. Each sub-folder shall contain no more than 10,000 images. Images from a single document shall not span multiple sub-folders.

All extracted Text and OCR files shall be produced in a folder named "TEXT."

All load files shall be produced in a folder named "DATA" or at the root directory of the production media.

M. Color

Paper documents or redacted ESI that contain color used to convey information (e.g., color coding and highlighting versus merely decorative use) shall be produced as single-page, 300 DPI JPG images with JPG compression set to its highest-quality setting so as not to not degrade the original image.

N. Redactions

Any redacted material must be clearly labeled on the face of the document as having been redacted and shall be identified as such in the load file provided with the production. Each redacted document shall be produced with an OCR *.txt file containing unredacted text. A document's status as redacted does not relieve the producing party from providing all the metadata required herein unless the metadata withheld contains privileged content.

O. Privilege Logs

With each production, Producing Party shall supply a log of the documents withheld or redacted under a claim of privilege and/or work product with sufficient information to allow the Receiving Party to understand the basis for the claim.

Communications involving trial counsel that post-date the filing of the complaint need not be placed on a privilege log.

P. Deduplication

Global Deduplication

Producing Party may horizontally (globally) de-duplicate documents based on MD5 or SHA-I hash values at the document level or by Message ID, EDRM MIH or other standard methodology for email deduplication within the collection of a custodian or a data source. Attachments to parent documents may not be deduplicated against a duplicate standalone version of the attachment exists, and standalone versions of documents may not be suppressed if a duplicate version exists as an attachment.

Producing Party will track all deduplicated files and provide the names of all custodians of these duplicates of in the load file. If the duplicates are e-mails, the producing party must detail the process of creating the hash value, e.g., the names and order of concatenated fields by which the deduplication hash was calculated.

Q. De-NISTing

System and application files without user created content (as identified by matching to the NIST National Software Reference Library database) need not be processed, reviewed or produced.

R. Email Threading

To reduce the volume of entirely duplicative content within email threads, the parties may, but are not required to, use email threading. A party may use industry standard message threading technology to remove email messages where the content of those messages, and any attachments, are wholly contained within a later email message in the thread; provided however, that the use of threading must not serve to obscure whether a recipient received an attachment.

S. Production Media

The producing party will use the appropriate electronic media (DVD, thumb drive, hard drive or secure FTP transfer) for its ESI production and will endeavor to use the highest capacity suitable media. The producing party will label the production media with the name of the producing party, production date, media volume name, and Bates number range(s).

Productions on physical media should be encrypted for transmission to the Receiving Party. At the time of production and under separate cover, Producing Party shall furnish decryption credentials to Receiving Party.

T. Processing

The Parties will use reasonable efforts and standard industry practices to address and resolve exception issues for items that present processing, imaging or form of production problems (including encrypted, corrupt and/or protected files identified during the processing of ESI). The Parties will meet and confer regarding procedures that will be used to identify, access, and process and resolve exception issues.

Parties shall normalize times and dates to conform to [UTC] OR [specified local time zone].

For archive files (zip, jar, rar, gzip, TAR, etc.), all contents should be extracted from the archive with source pathing and family relationships preserved and produced. The fully unpacked archive container file does not need to be included in the production.

U. Non-Waiver

This Protocol is solely intended to address the format of document productions and does not limit the temporal or substantive scope of discovery. Nothing in this Protocol is intended to affect the right of any party to object to a request for production or to operate as a waiver of any party's right to promulgate, object to, or seek relief from a request for discovery.

ADDENDUM A

Field Name	**Description**
BegBates	First Bates identifier of item
EndBates	Last Bates identifier of item
PgCount	Number of pages in the document
FileSize	Size of native file document/email in KB.
FileName	Original name of file as appeared in location where collected
Path	E-mail: Original location of e-mail including original file name. Native: Originating path where native file document was collected including original file name.
NativeLink	Relative path and filename for native file on production media
TextLink	Relative path and filename for text file on production media
AttRange	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	First Bates identifier of attachment range
EndAttach	Last Bates identifier of attachment range
AttachCount	Number of attachments to an e-mail
AttachName	Names of each individual Attachment, separated by semicolons
ParentBates	First Bates identifier of parent document/e-mail message (will not be populated for documents that are not part of a family).

ChildBates	First Bates identifier of “child” attachment(s); may be more than one Bates number listed depending on number of attachments (will not be populated for documents that are not part of a family).
Custodian	E-mail: mailbox where the email resided. Native: Individual from whom the document originated
OtherCustodians	Custodians whose file/message has been de-duplicated; separated by semicolons
From	E-mail: Sender Native: Author(s) of document; separated by semicolons
To	E-mail: Recipient(s); separated by semicolons
CC	E-mail: Carbon copy recipient(s); separated by semicolons
BCC	E-mail: Blind carbon copy recipient(s) separated by semicolons
DateSent (mm/dd/yyyy hh:mm:ss AM)	E-mail: Date and time the email was sent
Subject	E-mail: Subject line of email.
Title	Document: Title provided by user within the document
MsgID	E-mail: “Unique Message ID” field
EDRM_MIH	Identifier enabling cross platform email duplicate identification. Specifications and information about the EDRM MIH are found at https://edrm.net/edrm-projects/dupeid-2/
ModifiedDate (mm/dd/yyyy hh:mm:ss AM)	Document: Last Modified Date and time
CreationDate (mm/dd/yyyy hh:mm:ss AM)	Document: Create Date and time
FileExt	Document: file extension
FileType	Document: file type

Redacted	Denotes that item has been redacted as containing privileged content (yes/no).
Hash	MD5 Hash value of the item
HiddenContent	Denotes presence of Tracked Changes/Hidden Content/Embedded Objects in item(s) (Y/N)
Confidential	Denotes that item has been designated as confidential pursuant to confidentiality agreement or protective order (Y/N).
DeDuped	Full path of deduplicated instances; separated by semicolons