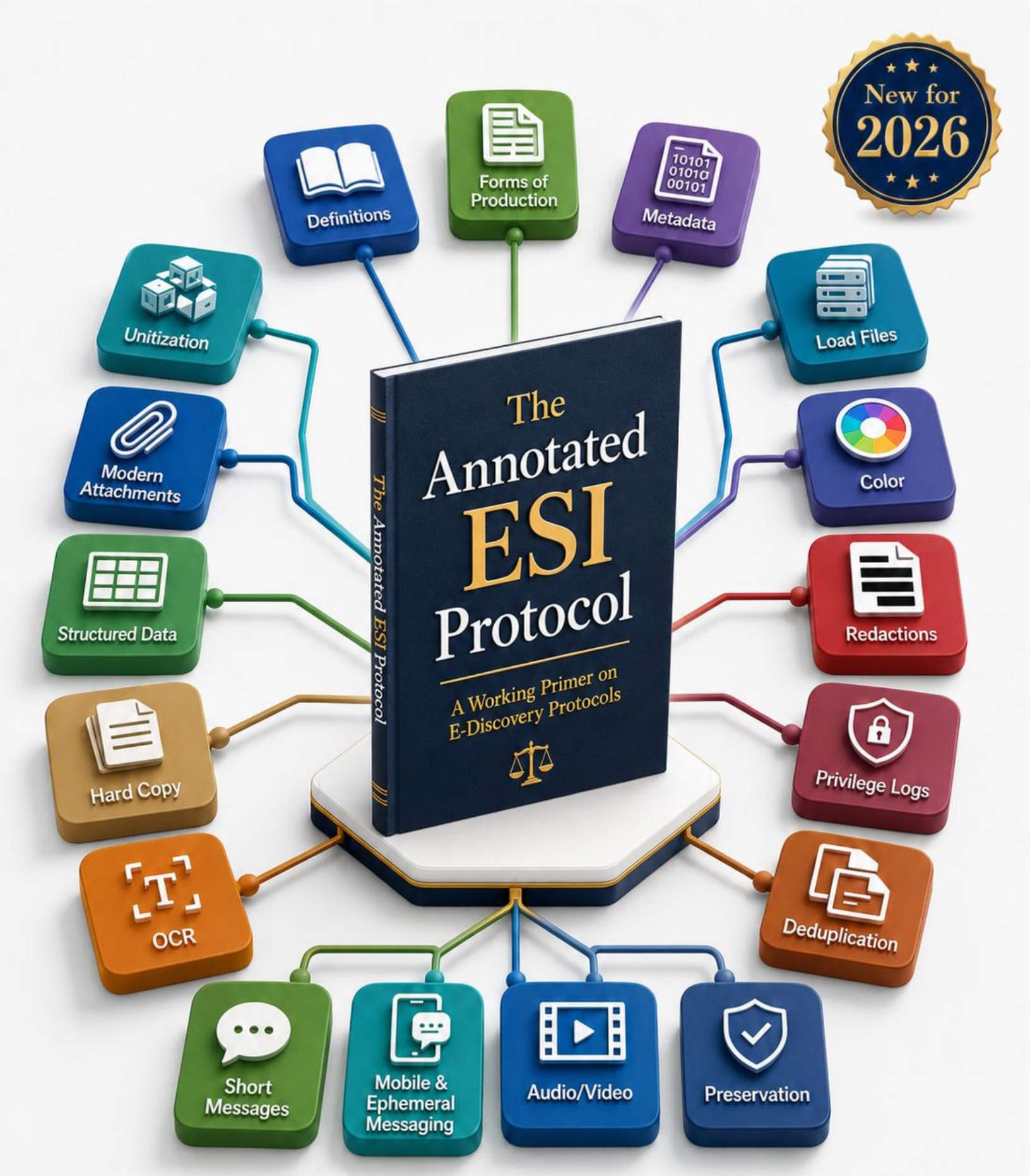

Craig Ball

The Annotated ESI Protocol

A Working Primer on E-Discovery Protocols

Craig Ball

An ESI or e-discovery protocol is an agreement or order that answers the recurring 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 are linked or "modern attachments" treated?
- How do parties protect privileged data and rectify inadvertent disclosure?
- What processes may producing parties use to suppress duplicates and thread messages?
- 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, including the validation of advanced analytics and technology-assisted review
- Issues attendant to discovery from databases and other structured data sources
- Discovery of short-message, collaboration platform, mobile and ephemeral communications
- Discovery of audio, video, and voicemail evidence
- Use of generative artificial intelligence in review and production workflows
- 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 is 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 a hybridized TIFF+ form of production—still the form most often demanded by institutional litigants and their service providers—but it carves out broad categories of evidence (spreadsheets, presentations, databases, photographs, audio, video, chat exports, mobile messages, and anything else that doesn't reduce sensibly to a static page image) for native production. That hybridization makes TIFF+ tolerable. If you're willing to fight for it, you'll find better economy and higher functionality by swapping in the alternative native-production language discussed in the Forms of Production section below. Pick your battles.

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))

[1]Throughout this exemplar protocol, the language presupposes a single producing party and a single requesting party. In matters with multiple parties on either side, the language must be conformed accordingly. The exemplar also assumes a federal-court setting and references to the Federal Rules of Civil Procedure should be conformed to the procedural rules of the forum.

A few drafting notes on what is and is not in this exemplar:

(a) The exemplar does not address the temporal or subject-matter scope of discovery, custodian selection, or the phasing of discovery. Those are case-specific issues better addressed in a separate scheduling/discovery order or in the parties' joint Rule 26(f) report.

(b) The exemplar does not address the protective order. A separate protective order, ideally with a Rule 502(d) clawback provision, should accompany this Protocol in any matter where confidentiality designations or privilege claims will arise.

(c) The exemplar uses "TIFF+" as the default form of production not because TIFF+ is the better form (it is not) but because the realities of institutional litigation in 2026 still favor it. The hybridization that follows—producing in native any item that does not reduce sensibly to a static page image—is the compromise that makes a TIFF+ default tolerable. Counsel preferring straight native production should swap in the Alternative 1 language in the Forms of Production section.

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 make 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. Two decades on, that admonition still gets ignored more than honored.

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 that 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 it in writing; and if I'm willing to agree to it in writing, I'm content for it to be memorialized in an order. There is 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 a tripwire depends upon whose ox is gored.

In the final analysis, judges guard their authority more jealously 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 is 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, technology-assisted review, or use of generative artificial intelligence, 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 and updated to reflect 2026 practice. It is not "The Perfect ESI Protocol," but one crafted in the hope of achieving both a representative assemblage of 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.

In the two-column layout that follows, the left column sets out exemplar language and the right column explains it.

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 and short messages exchanged over collaboration and chat platforms, mobile messages (including SMS, MMS, RCS, and platform-native messages such as iMessage), transaction logs, audio and video files, voicemail, internet data, computer logs, social-media posts and direct messages, and 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 electronic communications; word processing files (e.g., Microsoft Word); computer presentations (e.g., PowerPoint slides); spreadsheets (e.g., Excel); image files (e.g., PDF, JPEG, TIFF, HEIC); short messages and chat content from collaboration platforms (e.g., Slack, Microsoft Teams, Google Chat); mobile messages (e.g.,	Definitions artfully deployed in a protocol can serve to streamline and simplify the language of the protocol and the requests for production 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, mobile messaging, and collaborative communications such as comments, tracked changes, Slack, Microsoft Teams, and Google Chat) and relevant rich media (like video and audio recordings or social-networking content). 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, stored within the file, and moving with the file when it's 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-

iMessage, SMS, MMS, RCS, WhatsApp); audio and video files; and the metadata associated with each.

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, cloud services, 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; provided, however, that the parties may by agreement substitute SHA-1, SHA-256, or any other supported cryptographic 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, and 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 processed document are the date a file was last printed and any tracked changes and comments.

MD5 has been the e-discovery industry's default hash since the dawn of the discipline, and it remains a standard despite well-documented cryptographic collision vulnerabilities dating back two decades. For the duplicate-identification purposes that consume most of an ESI protocol's attention, MD5 still works fine; the parties shouldn't be obliged to retool every workflow to placate cryptographers. The exemplar permits SHA-1, SHA-256, or any other supported algorithm by agreement, which suffices for matters where the integrity demands are higher (forensic chain-of-custody, regulatory productions, criminal proceedings).[2]

Defined terms for Modern Attachment, Short Message, Conversation, and Ephemeral Message let later sections of the protocol address those evidence types crisply, without restating the universe each time. If you don't expect any of these in the case, strike the unused definitions—defining terms you never use is sloppy.

[2]MD5 has been an industry standard for hash-based duplicate identification in e-discovery for three decades. It has been recognized as cryptographically broken since 2004, when collisions were first published, and chosen-prefix collisions have been demonstrated since. None of this is a problem for the duplicate-identification purposes that drive most ESI-protocol uses of hash values: identical files produce identical MD5s, and the concern is that a malicious party could produce two distinct files with the same MD5—a concern that does not arise in routine deduplication. Where the integrity demands are higher (forensic chain-of-custody, regulatory productions, criminal proceedings), a stronger algorithm like SHA-256 should be used in lieu of MD5. The exemplar accordingly defaults to MD5 and permits SHA-1, SHA-256, or any other supported algorithm by agreement.

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. The “native format” of short messages and chat content is the export format produced by the source platform’s administrative or compliance interface (e.g., Slack JSON exports), supplemented by any rendered transcript necessary for human readability.

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.

11. “Modern Attachment” or “Linked File” means a document or other resource referenced in an electronic communication by hyperlink or pointer, the substance of which is hosted in a cloud-based repository (e.g., OneDrive, SharePoint, Google Drive, Dropbox) and which the sender intended the recipient to access through the link as a functional substitute for an embedded file attachment.

12. “Short Message” means any message exchanged over a chat, instant messaging, collaboration, or mobile messaging platform, including but not limited to Slack, Microsoft Teams, Google Chat, iMessage, SMS, MMS, RCS, WhatsApp, Signal, Telegram, and direct messages exchanged on social-media platforms.

13. “Conversation” means a logically related sequence of Short Messages exchanged among a

defined set of participants within a single channel, thread, group, or direct-message context. 14. “Ephemeral Message” means a Short Message that, by configuration of the source application, is set to be deleted automatically after a defined retention period or upon being read.	
Preservation	
The Parties represent that they have issued litigation hold notices to those custodians with data, and to persons or entities responsible for the maintenance of non-custodial data, which, based upon then-current information available, are reasonably likely to contain discoverable information. The hold shall include affirmative direction sufficient to suspend any auto-deletion, retention-policy expiration, or ephemeral-messaging settings that would otherwise destroy potentially relevant Short Messages. 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. 5. Server, system, or network logs. 6. System and application files matching entries in the NIST National Software Reference Library.	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 deemed not reasonably accessible or unduly burdensome. New here is the express direction that a hold suspend ephemeral-messaging and auto-deletion settings. Disappearing-message timers and platform retention policies (Slack’s message-retention configuration, Teams chat retention policies, mobile auto-delete features) routinely destroy relevant evidence after a hold has been issued because counsel forgot the policy was running in the background. Decided cases have not been kind to litigants who left those toggles in their default destructive settings after suit was filed.
E-Discovery 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. Each liaison shall be reasonably available to confer with opposing counsel's liaison on technical matters arising under this Protocol and to escalate disputes to lead counsel as appropriate.	Even the best ESI liaisons must sometimes reply, "I'll get back to you," but communication and efficiency really suffer when questions filter through counsel unschooled in e-discovery. Working through skilled liaisons that "speak geek" won't guarantee harmony, but it fosters focused, dispassionate diplomacy. The added requirement of "reasonably available" liaison contact is a small thing that prevents the "my expert is unavailable for the next ten days" gambit from derailing time-sensitive issues.
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. The producing party shall provide a data dictionary or schema reference sufficient to allow the receiving party to interpret field names, code values, and the relationships among tables.	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 that can be crippled or eradicated when data is produced in e-discovery without preserving its fielded ("delimited") character. New here: a duty to provide the schema, code lookups, or data dictionary needed to interpret the export. A CSV with twenty fields named CD_01, CD_02, … is technically delimited and entirely useless. Producing parties commonly know what those codes mean—or have a documented mapping in the application—and the receiving party should not be left to guess. For more on databases in e-discovery: 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 images, with a ".tif" file extension.[3] 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. Hard Copy Documents that contain color used to convey information shall be scanned and produced as 300 dpi JPG images at the highest-quality compression setting, in lieu of TIFF, with the same file-naming and OCR conventions.	Although there is 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-and-white pictures. "Single page" requires that 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

[3]The term "Bates number" derives from the Bates Manufacturing Company, which manufactured a sequential numbering machine widely used in the legal profession for stamping consecutive numbers on documents. The numbering machine has long since been retired in favor of software-applied Bates numbers, but the term stuck.

	obliged to supply OCR text absent an agreement or order. Because this provision demands that 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 Bates numbering, with 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 1980s-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, sometimes a 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. The added JPG-for-color carve-out at least staunches the worst of the color-loss problem; it should be the rule rather than the exception.
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	"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

correctly, and will commit to address situations of improperly unitized 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. Treatment of Modern Attachments transmitted by hyperlink rather than as embedded attachments is governed by the Modern Attachments / Linked Files section below.	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. It also 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. The 2023 edition of this exemplar lumped hyperlinked "modern attachments" into the Parent-Child clause. That was always uncomfortable; the issues are different enough to warrant their own section. They get one below.
Modern Attachments / Linked Files	
The Parties agree that documents transmitted to a recipient by hyperlink to a cloud-hosted resource ("Modern Attachments" or "Linked Files") shall be treated as attachments to the transmitting communication for purposes of discovery, except that documents merely referenced—as distinguished from those the sender intended the recipient to access through the link as a substitute	This is the topic on which the most evolution has occurred since the prior edition of this paper. Microsoft 365, Google Workspace, and other cloud platforms have made it effortless for users to send a colleague a hyperlink to a document in OneDrive, SharePoint, or Google Drive in lieu of an embedded attachment. From a user's perspective, the link is the attachment. From a discovery perspective, the link is a pointer to a moving target.

for an embedded attachment—need not be produced as Modern Attachments.

To the extent reasonably feasible given the source platform's collection capabilities, the producing party shall collect and produce the version of each Modern Attachment that existed as of the time the link was sent (the "point-in-time version"). Where collection of the point-in-time version is not reasonably feasible, the producing party shall collect and produce the most contemporaneous version reasonably available, identify the version produced by version identifier or modification date, and disclose the limitation.[4] Where the receiving party identifies particular Modern Attachments for which the point-in-time version is material to the issues in the case and the producing party's ordinary collection has not produced it, the Parties shall meet and confer regarding the proportionality of pursuing historical-version recovery through specialized tooling.

The producing party shall provide metadata sufficient to identify each Modern Attachment as such (e.g., a "LinkedFile" Boolean field), to associate each Modern Attachment with its parent communication (e.g., ParentBates and AttRange fields), and to identify the URL of the linked resource as transmitted, the platform of origin, and any version identifier.

Where a Modern Attachment cannot be collected or produced (e.g., the link has been broken, the resource has been deleted, or the producing party lacks access), the producing party shall so identify the link by URL and explain the basis for non-production.

Producing parties have argued—with mixed success—that linked content is not part of the email and need not be collected as a family member. Receiving parties have countered that the only reason any of this came up is because email clients now insert hyperlinks rather than embedded copies for the same files that, a decade ago, would have been embedded as attachments. Whatever the technical implementation, the substantive question is whether the linked content was, in the sender's and recipient's ordinary understanding, transmitted with the message.

The exemplar provision distinguishes—properly, I think—between content the sender intended the recipient to consume through the link (which the producing party should produce) and content merely referenced (e.g., a sentence reading "see the policy at https://wiki/…"). Drawing that line is fact-dependent and will be a recurring source of meet-and-confer skirmishes; the goal of the protocol is not to eliminate the dispute but to give the parties a shared vocabulary for it.

The "point-in-time version" obligation is the recurring difficulty. Cloud documents change after the link is sent. The version the recipient saw when the link was clicked may not be the version present when the producing party collects months later. Collection-tool capability for historical-version recovery varies materially across platforms and subscription tiers. The exemplar requires the producing party to do what is reasonably feasible, identify what was produced, and disclose the limitation, with a meet-and-confer trigger for cases where the receiving party identifies items for which the point-in-time version is material.

[4]Recovery of point-in-time versions of cloud-hosted files turns on the source platform and the producing party's subscription tier. Microsoft Purview eDiscovery (Premium) can collect SharePoint and OneDrive version history, including the version that existed as of a specified datetime; Standard Purview cannot. Google Vault collects current Drive content, with historical-version recovery limited and SKU-dependent. Box, Dropbox Business, and other platforms vary. For the median producing party in routine commercial litigation, the point-in-time version may simply not be reasonably available, and the "most contemporaneous version" produced will reflect collection-time state rather than transmission-time state. The meet-and-confer trigger in this Section is intended to focus the parties on items for which the historical version is genuinely material rather than to require historical-version recovery as a matter of course.

	Counsel litigating these issues should review The Sedona Conference's Commentary on Modern Attachments and the line of decisions beginning with Nichols v. Noom and continuing through subsequent rulings on the proper handling of hyperlinked documents in major MDLs and class actions. Outcomes have varied with the facts of the case and the available collection technology, but the trend is unmistakable: courts increasingly expect producing parties to treat the substantive content of links as evidence to be produced, not as metadata footnotes.
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) 8. Extracted/OCR text file path	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.
Short Messages and Collaboration Platforms	
Short Messages from collaboration and chat platforms (e.g., Slack, Microsoft Teams, Google Chat) shall be produced in a form that preserves the content, the participants, the channel or thread context, the timestamps, and the family relationships between messages and any embedded or attached files. The producing party shall produce, for each responsive Conversation: 1. A native or near-native export of the Conversation in the format provided by the source platform's administrative or compliance interface (e.g., Slack's JSON export, Microsoft Purview eDiscovery export, Google Vault export);	Short messages have become primary business communications. A 2026 production that excludes Slack or Teams content because counsel "only collected email" is no longer defensible in matters where the custodians demonstrably used those platforms for substantive work. Two questions dominate Short-Message production: (1) what unit of evidence are we producing—a single message, a thread, a channel, or a day's worth of channel traffic? and (2) in what form—native export (typically JSON), rendered HTML/PDF transcript, or both? The exemplar opts for both, with Conversation-day as the default unitization unit. That's the most common workable

2. A human-readable rendering of the Conversation as a single document per Conversation per day or per Conversation per logical unitization unit, paginated and Bates-numbered, with each message identifying its sender, timestamp (with time zone), and any associated reactions, edits, or deletions known to the producing party; and
3. Any files attached to or embedded in the Conversation, produced as separate items with metadata associating each to its parent Conversation.

The requirement of a human-readable rendering under (2) above is conditioned on the availability of platform tooling capable of producing such rendering or upon reasonable request at proportionate cost.[5] Where rendering would require specialized vendor engagement disproportionate to the demands of the matter, the Parties shall meet and confer regarding alternatives, which may include native production accompanied by a documented procedure for rendering on demand.

Unitization of Short Messages shall be at the Conversation-day level (one document per channel, thread, or DM context per calendar day) as the default target unitization, provided that where the source platform's native export structure does not align with Conversation-day boundaries, the producing party may produce in the source-platform's native unitization, identify the rule applied, and provide metadata sufficient for the receiving party to verify the unitization against the underlying export.

compromise; it produces sensible, reviewable documents without atomizing the evidence into millions of one-line records.

The native export matters because it carries fields a rendered transcript cannot easily convey: thread parent IDs, user IDs, message IDs, channel IDs, and (importantly) edit and deletion history. Without those, the receiving party cannot verify whether a message was edited after sending or was deleted entirely. Reactions, often dismissed as decorative, are sometimes evidentiary—a "👍" from the CEO on a contested decision is a substantive endorsement.

Producing parties may resist the rendered-transcript requirement on cost grounds. Where the producing party's tooling natively produces a rendered transcript or where the cost of generating one programmatically from the JSON export is modest at the volumes encountered, the requirement is unobjectionable. Where it is not—because the producing party's platform does not natively render and specialty vendor engagement would be required—the protocol's proportionality qualifier governs and the Parties should confer on alternatives, including native-only production with on-demand rendering of items the receiving party identifies as substantively significant.

For a fuller treatment of short-message discovery practices, consult The Sedona Conference Primer on Social Media (Second Edition) and the EDRM Short Message Project.

[5]The capability to produce a paginated, Bates-numbered transcript of a Conversation varies materially across the major platforms and across subscription tiers within each platform. Slack's native export produces JSON structured by channel and date but does not produce a rendered transcript; commercially available specialty tools convert that JSON into a paginated transcript at additional cost. Microsoft Purview eDiscovery (Standard and Premium) produces Teams content in formats that may require post-processing to yield a transcript matching this Protocol's defaults. Google Vault produces Chat content as MBOX. Edit and deletion histories are similarly tier-dependent: Slack records edits and deletions in audit logs available to Enterprise Grid customers; standard- and Plus-tier workspaces lack the audit log API. Microsoft Teams records edits in message history but deletion logging requires Purview Audit (Premium) licensing. Google Workspace logs message deletions in Vault, but recoverability depends on retention rules in effect at the time of deletion. The Protocol's proportionality qualifier and the disclosure obligation governing platform tier and audit-log availability are calibrated to these realities.

Where the source platform records edits, deletions, or message-revision history and that history is reasonably available to the producing party, the producing party shall produce the edit/deletion history as part of the rendered Conversation or as accompanying metadata. Whether such history is reasonably available depends on the source platform's configuration and the producing party's subscription tier; the producing party shall disclose the platform tier and the audit-log availability applicable to the source data sufficient for the receiving party to evaluate any limitations on the production of edit and deletion history.

Reactions (e.g., emoji reactions to a message) shall be preserved in the rendered Conversation when reasonably available.

Mobile and Ephemeral Messaging

Mobile messages (including iMessage, SMS, MMS, RCS, WhatsApp, Signal, Telegram, and platform-native messaging on personal or business mobile devices) responsive to discovery requests shall be produced as:

1. An export of the responsive Conversation produced by a tool capable of preserving the message content, the participants, the timestamps (with time zone), and any attachments, in a form that allows the receiving party to verify the integrity of the export. Acceptable tools range from consumer-grade backup-extraction utilities (e.g., iMazing, Decipher Text Message, AnyTrans, or comparable) to forensic-grade mobile collection platforms (e.g., Cellebrite, Magnet AXIOM, Oxygen Forensic Detective, MSAB XRY, or comparable). The choice of tool shall be proportionate to the demands of the matter and the evidence-integrity issues presented.
2. A human-readable rendering of the Conversation, paginated and Bates-

Mobile messaging is where preservation goes to die. Auto-delete settings, lost devices, factory resets, and platform-side retention policies destroy evidence routinely, often without any malicious intent. The exemplar accordingly pairs production requirements with a disclosure obligation about how the producing party managed (or failed to manage) the destructive settings.

The bigger access-to-justice problem is the routine practice in civil litigation of ignoring mobile evidence altogether because the cost of forensic-grade collection is treated as the only acceptable approach. It isn't. For the great mass of civil cases, a consumer-grade backup-extraction tool—iMazing, Decipher Text Message, AnyTrans, or a comparable utility—produces a defensible export of iOS or Android message content at a cost an order of magnitude below forensic-grade collection. The export captures sender, recipient, timestamp with time zone, content, and attachments; it can be rendered as a transcript with metadata; and the underlying backup file from which the export was produced can itself be preserved and authenticated. That's good-enough mobile evidence for routine commercial litigation, employment matters, and

numbered, identifying each message's sender, recipient(s), timestamp (with time zone), and any attachments;

3. Any media attached to or embedded in the Conversation, produced as separate items with metadata associating each to its parent Conversation; and
4. A statement identifying the tool used to produce the export, the tool version, and the date and method of collection.

Where the matter involves (a) alleged spoliation or deletion of mobile evidence, (b) recovery of deleted content from device storage, (c) a dispute as to the integrity of the source device or its contents, or (d) circumstances in which a party seeks judicial relief premised on the comprehensiveness of the mobile collection, the producing party shall use a forensic-grade mobile collection tool unless otherwise agreed or ordered. The Parties shall meet and confer regarding the appropriate level of forensic rigor where the demands of the case so warrant. The foregoing trigger conditions govern the choice of tool prospectively and do not, of themselves, require re-collection of devices already collected; any further forensic examination of previously collected devices shall be addressed by agreement of the Parties or by order of the Court.

The producing party shall preserve and produce, where reasonably available, the underlying message database or backup file from the source device (e.g., the SMS/iMessage SQLite databases on iOS, or the iTunes/Finder backup from which the export was produced) in addition to any rendered transcript.

Where any source device or messaging account was configured at any time during the preservation period to delete messages automatically after a defined interval (an "auto-delete" or "disappearing-messages" setting), the producing party shall disclose the configuration history (or, if unavailable, the configuration as of the date the litigation hold was implemented), the affected platforms and

most civil disputes. Insisting on Cellebrite-grade collection in every case effectively means most cases proceed without mobile evidence at all, which is the worse outcome.

What consumer-grade tools cannot do is recover deleted messages from device free space, extract from devices with damaged storage, or produce the kind of comprehensive verifiable extraction log a forensic tool generates. Those capabilities matter when the case involves alleged spoliation, when recovery of deleted content is the point, when the integrity of the source device is itself at issue, or when the party seeks judicial relief premised on the comprehensiveness of the collection. In those circumstances the exemplar requires forensic-grade collection. Outside those circumstances, proportionality favors the more economical approach.

The screenshot point still holds however the collection is done. Custodian-produced screenshots of message threads—taken by users from their own devices, with no underlying export—are demonstrably unreliable and invite authentication challenges. Whether the producing party uses iMazing or Cellebrite, the artifact produced should be the export, not the screenshot.

Where the custodian uses a personal device for both personal and business communication (the rule rather than the exception in 2026), the protocol should make clear—either here or in the parties' separate BYOD discovery agreement—how the producing party will isolate the responsive subset and what review will occur before non-responsive personal content is screened from production.

accounts, and the steps taken to suspend such settings upon implementation of the litigation hold.	
Audio, Video, and Voicemail	
Audio recordings (including voicemail), video recordings, and other rich-media files responsive to discovery requests shall be produced in their native format with original metadata preserved (including, where available, capture device, capture date, duration, and codec information). Where a producing party intends to rely on a transcript of any responsive audio or video evidence, the producing party shall produce the transcript with metadata identifying the source recording (by Bates number) and the means of transcription (e.g., automated speech-to-text, human transcription, or hybrid). A transcript is not a substitute for the underlying recording, and the underlying recording remains the operative evidence. The Parties shall meet and confer regarding any obligation to provide transcripts as a searchability surrogate for audio and video, including allocation of cost and validation of accuracy.	Audio and video should always be produced natively. There is no static-image equivalent that preserves the evidence, and the cost of native production is trivially low compared to the alternative of trying to reduce a recording to a textual artifact. The exemplar treats native as the only acceptable form. Transcripts are a different question. Audio is not text-searchable until transcribed, and transcription is not free. Speech-to-text technology in 2026 is materially better than it was even three years ago, but its accuracy varies sharply with audio quality, multi-speaker recordings, accented speech, and domain vocabulary (legal, medical, technical). A transcript produced as a searchability surrogate should disclose the means of transcription so the receiving party can calibrate its reliance. Where the case involves substantial recorded evidence (e.g., recorded sales calls, body-worn camera footage, surveillance video), the parties should address transcription cost-sharing and accuracy-validation expressly. Wholesale machine transcription with no validation can mislead a search workflow into missing evidence; sample-based human review of machine transcripts is a reasonable middle ground.
Forms of Production	
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, where reasonably 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	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. 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

appending the Bates number for/to the file name.[6] 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.	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 e-discovery vendors price services based on the byte volume of the data. If the case will tolerate it, native is the right answer. The exemplar offers TIFF+ as Alternative 2 because so much institutional litigation still demands it, and the protocol must work on the courts and clients we have, not the ones we wish we had.
Alternative 2 — Hybridized 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" × 11" 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 format: 1. Spreadsheets (e.g., Microsoft Excel, Google Sheets, Apple Numbers) 2. Presentations (e.g., PowerPoint, Keynote, Google Slides) 3. Databases and database extracts (e.g., Microsoft Access, SQL exports) 4. Delimited text files (e.g., CSV, TSV) 5. Photographs and other native image files (e.g., JPEG, PNG, HEIC, RAW) 6. Audio, video, and voicemail recordings 7. Short Messages and Conversations from collaboration platforms (per the Short Messages section) 8. Mobile messages (per the Mobile and Ephemeral Messaging section)	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. TIFF images may be viewed in any browser, though they won't be text-searchable doing so. The exemplar above is hybridized: it carves out so many categories for native production that what remains as TIFF is essentially the universe of word-processed documents, emails, and PDFs. That's the point. The categories carved out are exactly the categories where TIFF rendering causes real harm—loss of formulas in spreadsheets, loss of speaker notes in presentations, loss of color in everything, loss of fidelity in audio and video, and the impossibility of reducing a Conversation to a static page. A TIFF+ protocol that doesn't make these carve-outs is one in which the producing party gets to degrade the evidence and call it compliance. 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

[6]When ESI is produced in native format, Bates numbers cannot be embossed onto the face of pages because there are no static pages on which to emboss them. Bates numbers can instead be (a) substituted for the file name (i.e., the Bates number becomes the file name); (b) prepended to the original file name; or (c) appended to the original file name. The substitution approach is most common because it ensures unique file names and avoids the problem of duplicate file names from different custodians. The original file name is preserved in the load file in the FileName field.

9. Structured-data exports (per the Databases and Structured Data section) 10. Computer-aided design (CAD), engineering drawing, and other format-specific files for which TIFF rendering would materially impair fidelity 11. Documents of a type which cannot be reasonably converted to useful TIFF images. All images of documents that 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, hidden columns or rows, hidden slides, or other user-entered data that the source application can display to the user shall be processed such that all that data is visible in the image, or, in the alternative, produced natively. Documents that contain color used to convey information (e.g., color coding and highlighting versus merely decorative use) shall be produced as 300 dpi JPG images at the highest-quality compression setting, in lieu of TIFF, with the same file-naming and metadata conventions, or, in the alternative, produced natively.	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 takes the first path, with native as the alternative. 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.[7]
File Names	
Each TIFF image (or JPG image produced for color or paper documents) shall have a unique file name corresponding to the Bates number of that page with the appropriate file extension. The file name shall not contain any blank spaces and shall be zero-padded (e.g., DEF-0000001), 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, the producing party shall so note in a cover letter or production log accompanying the production. Bates numbers shall	Bates numbering is the connective tissue of the production. Get it wrong and nothing else lines up. The principal pitfall is reuse of sequences across productions in the same matter (especially across multiple producing parties or over years of rolling production); thus the requirement of consistent prefixes and unique numbers across the entire production. Zero-padding deserves a moment's thought. If your matter will produce 50,000 pages, six-digit padding suffices. For a serious modern matter, plan for at

[7]Hosting and processing in e-discovery is overwhelmingly priced on a per-gigabyte-per-month basis. TIFF renderings of native files typically inflate file size by an order of magnitude (sometimes more, depending on document type), so a TIFF+ production is many times more expensive to host than the equivalent native production. This is one of the under-discussed perversities of the TIFF+ default: the producing party imposes a multi-month cost burden on the receiving party that scales linearly with the volume of data. Native production avoids this entirely. For more on the economics: craigball.com/Ball_TIFF_2018.pdf.

be unique across the entire production and prefixes shall be consistent across all documents produced. The producing party will brand all TIFF and JPG images in the lower right-hand corner with their corresponding Bates number without obscuring any part of the underlying image.	least seven digits; eight is not unreasonable for very large productions. Re-padding mid-production is painful and avoidable.
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. If the document is an image file or contains redactions, a text file created using OCR shall be produced in lieu of extracted text.	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 created from that text), and because the Bates-numbered text files share names with the Bates-numbered image files, search hits within text tie to page images.[8] Extracted text from the native file is materially more accurate than OCR text generated from a TIFF rendering of the same document. The exemplar accordingly requires extraction from the native unless the source is itself an image (e.g., a paper scan) or has been redacted (where extracting from the unredacted native would defeat the redaction). This is one of the points where producing-party process discipline genuinely matters; a workflow that OCRs everything because it's easier than extracting from natives is producing systematically inferior evidence. The provision requires that the text be produced as Unicode text, meaning that it must be encoded to support a wide array of international characters versus the paltry 256 characters of the once-ubiquitous ASCII encoding. UTF-8 with byte-order

[8]ASCII (American Standard Code for Information Interchange) was the dominant text encoding for English-language computing for decades. It supports 128 characters (or 256 in the "Extended ASCII" variants), enough for English plus a smattering of punctuation and control codes but inadequate for non-English alphabets, ideographic scripts, emoji, and modern web content. Unicode replaces ASCII with a much larger code-point space and a family of encodings (UTF-8, UTF-16) that handle the world's scripts. UTF-8 is now the default for most computing; UTF-16 LE is common in Microsoft tooling. ASCII text in 2026 is technically Unicode (since UTF-8 is a superset of ASCII for the original 128 characters), but production specifications should call for full Unicode encoding to ensure non-English content survives the round trip.

	mark, UTF-16 LE, or UTF-16 BE are all acceptable; the load-file specification below pins this down.
Load Files	
Productions will, as applicable, include image load files in Opticon (.opt) or IPRO (.lfp) format and Concordance-format data (.dat) files with the applicable metadata fields identified in Addendum A. All metadata files shall be encoded as UTF-8 with byte-order mark or UTF-16 LE. All native format files shall be produced in a folder named "NATIVE." All TIFF and JPG 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 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: craigball.net/2013/07/17/a-load-file-off-my-mind/
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 degrade the original image. 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 e-discovery 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. Where the

	producing party's tooling supports it, that should be the default. 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 they best lend themselves to ad hoc requests for color reproduction. PowerPoint presentations and Excel spreadsheets should never be produced in anything but native formats (where color is natively supported), and the hybridized TIFF+ exemplar above provides for that.
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 text file containing only the 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. Where a producing party redacts a Short Message, audio recording, video recording, or other non-paginated form of evidence, the producing party shall describe the redaction (the location, the duration, or the nature of the content redacted) with sufficient particularity to allow the receiving party to evaluate the basis for the redaction without revealing 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. Redactions in audio, video, and chat content are an underappreciated landmine. A blanked frame in a video, a silent gap in audio, or a deleted message bubble in a chat transcript needs to be unambiguously identified as a redaction—not as a defect or omission. The exemplar provision asks for description sufficient to evaluate the basis for the redaction. That's a meet-and-confer concept dressed up as a clause; expect to negotiate the level of particularity case by case.
Privilege Logs	

With each production, the 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. Where the volume of withheld items makes a document-by-document log impracticable, the Parties may agree to a metadata-based or categorical log. Any agreement on the form of the log shall identify the categories or fields used and the bases on which categorical or metadata logging is appropriate.	The obligation to furnish a privilege log is governed by the applicable Rules of Civil Procedure, e.g., Fed. R. 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 and an explicit recognition that metadata-based or categorical logs may be appropriate at scale. 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.[9]
Deduplication	
Vertical Deduplication The producing party may vertically deduplicate documents based on MD5, SHA-1, SHA-256, Message ID, EDRM MIH, or other standard methodology for email deduplication within the collection of a single custodian or data source. Attachments to parent documents may not be deduplicated where a duplicate standalone version of the attachment exists, and standalone versions of documents may not be suppressed where a duplicate version exists as an attachment. OR **Horizontal (Global) Deduplication** The producing party may horizontally (globally) deduplicate documents based on MD5, SHA-1, SHA-256, Message ID, EDRM MIH, or other standard	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

[9] In federal court, an order under Federal Rule of Evidence 502(d) provides the strongest protection against waiver from inadvertent disclosure of privileged material. A Rule 502(d) order operates regardless of the precautions the producing party took (or didn't take) and binds non-parties as well. It is best practice to enter a Rule 502(d) order at the outset of any case in which significant volumes of ESI will be produced. A Rule 502(d) order is not a substitute for an ESI protocol; the two work in tandem.

methodology for email deduplication across multiple custodians or data sources. Attachments to parent documents may not be deduplicated where a duplicate standalone version of the attachment exists, and standalone versions of documents may not be suppressed where a duplicate version exists as an attachment. The producing party will track all deduplicated files and provide the names of all custodians of these duplicates in the load file. If the duplicates are e-mails, the producing party shall describe the process of creating the hash value, e.g., the names and order of concatenated fields by which the deduplication hash was calculated.	of the binary content of files to facilitate duplicate identification. SHA-256 is increasingly common where collision resistance matters. E-discovery service providers 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.[10] Horizontal (global) deduplication is more efficient but loses custodian-by-custodian provenance unless the load file faithfully tracks all suppressed duplicates. The exemplar requires that tracking. A horizontal-dedupe production with no all-custodians field in the load file is a worse production than a vertical-dedupe production at the same volume.
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 e-discovery 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 threading. A party may use industry-standard message-threading technology to remove email messages where the content of those messages, and any attachments,	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 objection most often voiced is that threading may serve to suppress a message or

[10]The EDRM Message Identification Hash (MIH) is a standardized methodology for calculating a hash value across the headers and body of an email message such that duplicates can be reliably identified across different email systems and processing tools. For more: edrm.net/edrm-projects/dupeid-2/.

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. Where threading is employed, the producing party shall provide metadata sufficient to identify the threading relationships among produced and suppressed messages and to allow the receiving party to reconstruct the thread.	attachment whose existence and routing are evidentiary in their own right—e.g., who actually received which attachment when. The added metadata requirement allows the receiving party to verify the integrity of the threading and to surface any individual message where the routing matters more than the content. Producing parties using threading should expect to provide thread-relationship fields in the load file as a matter of course.
Production Media and Transmission	
The producing party will use the appropriate electronic media (encrypted hard drive, encrypted thumb drive, secure file transfer, or secure file-sharing service) for its ESI production and will endeavor to use the highest-capacity suitable medium and the most efficient transmission method consistent with the security requirements set forth herein. The producing party will label the production media (or production package, in the case of secure transfer) with the name of the producing party, production date, media volume name, and Bates number range(s). Productions on physical media shall be encrypted using AES-256 or comparable encryption. Productions transmitted electronically shall be transmitted via secure means (e.g., SFTP, HTTPS, or a vendor-provided secure file-sharing service with at-rest encryption). At the time of production and under separate cover, the producing party shall furnish decryption credentials to the 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 or transmission method used to hand off the data. In 2026, secure file-sharing services have largely replaced physical media for productions of any size. Counsel still occasionally encounters thumb drives and external hard drives, especially in matters where outside vendors are used, but the modern default is encrypted electronic transfer. The exemplar accommodates both. Two practical points: (1) AES-256 is the right encryption baseline; legacy zip-with-password is not. (2) Decryption credentials must travel separately from the encrypted package—physical media should never be labeled or stored with its decryption credentials, and electronic transmission credentials should not be sent in the same email as the link to the encrypted file.
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 exceptions are the silent killer of completeness. A production may look complete on paper while a meaningful percentage of source items quietly drop out due to corruption, password protection, encryption, unsupported file types, or

processing of ESI). The Parties will meet and confer regarding procedures that will be used to identify, access, and process and resolve exception issues. The producing party shall maintain and provide upon reasonable request a log of items excluded from production due to processing exceptions, identifying the item, the nature of the exception, and the steps taken to resolve it. Parties shall normalize times and dates to conform to [UTC] OR [specified local time zone]. Where dates and times in the source data are recorded with time-zone information, that information shall be preserved in the load file. For archive files (e.g., .zip, .jar, .rar, .gz, .tar, .7z), all contents shall 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.	processing errors. The exemplar requires an exception log so the receiving party can see what is missing and why. Without it, the receiving party has no way to know whether the production reflects the universe of responsive evidence or just the universe of evidence the producing party's tools could open. For more about processing: craigball.com/Ball_Processing_2019.pdf
Search Methodology and Validation	
The Parties may employ keyword search, technology-assisted review (TAR), continuous active learning (CAL), or other technology-assisted methodologies to identify potentially responsive ESI for review and production. The producing party retains the right to choose its review methodology and tooling. Where keyword search is employed, the parties shall meet and confer regarding the search terms and refinements to be used. The producing party shall disclose, upon reasonable request, the final search syntax applied, the data sources to which the searches were applied, and the resulting hit counts. Where technology-assisted review is employed, the producing party shall disclose, upon reasonable request, the general methodology applied (e.g., predictive coding, continuous active learning, clustering), the validation methodology used to confirm the adequacy of the review, and a summary of the validation results, including the underlying sampling parameters (sample size, sampling methodology, and observed counts) sufficient to	This is a new section relative to the prior edition. It is not meant to mandate a particular methodology—the producing party gets to pick its review approach—but to nail down the validation and disclosure obligations that flow from whatever choice is made. Disclosure of search syntax and hit counts is unobjectionable. Disclosure of seed sets and individual coding decisions is contested and should not be the default; the line drawn here tracks the consensus reflected in The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery and the case law applying it. Recall and precision are the right framing for TAR validation: recall is the fraction of responsive documents the methodology surfaced; precision is the fraction of surfaced documents that are responsive. The exemplar requires reporting of the underlying sampling parameters (sample size, sampling methodology, and observed counts) so the receiving party can characterize the precision of the

characterize the precision of the recall and precision estimates reported.[11] Disclosure of seed sets, training sets, and individual coding decisions is not required absent a particularized showing of need.	estimates rather than mandating computed confidence intervals; in routine practice, point estimates from a control set or elusion-test sample remain the dominant validation reporting, and the sampling-parameters formulation captures the necessary information without obliging the producing party to a computational format that not all platforms generate natively.
Use of Generative Artificial Intelligence in Review	
Generative artificial intelligence tools may be used by either Party in support of review and production workflows, including for responsiveness review, privilege review, redaction, summarization, deposition preparation, and related tasks, subject to the following: 1. The producing party shall ensure that any generative AI tool used in the review of potentially privileged or confidential ESI is hosted and operated in a manner that does not result in the transmission of such ESI to third parties or its incorporation into the training data of any general-purpose model. 2. The use of generative AI does not relieve the producing party of its obligations under the applicable Rules of Civil Procedure, including the duties of competence, diligence, and accuracy. Output from generative AI tools shall be subject to human review sufficient to verify its accuracy before reliance for any responsiveness or privilege determination. 3. The disclosure obligations applicable to technology-assisted review under the Search Methodology and Validation section shall apply equally to the use of generative AI tools, with disclosure of the general methodology, validation approach, and	Generative AI is no longer a novelty in document review; in 2026 it is becoming routine for first-pass responsiveness, drafting privilege-log entries, summarizing depositions, and synthesizing across record sets. The protocol cannot ignore it, and pretending otherwise would force every matter to relitigate the basics. The three propositions in the exemplar are deliberately modest. First, no feeding privileged or confidential client data to public consumer LLMs (this is bar-discipline territory, not just a discovery point). Second, AI-assisted output requires human verification before it becomes a determination on which a party relies; this is the same rule that has always applied to junior-associate review and to any other review delegation. Third, the validation and disclosure framework for TAR maps cleanly onto generative AI; there is no principled reason to treat them differently. Counsel using generative AI in review should also be alert to the well-documented hallucination patterns that affect different types of tasks differently—fact-extraction from a single document is generally reliable; synthesis across many documents introduces error; citation and quotation tasks are notoriously prone to fabrication. Sample-based

[11]The Sedona Conference and the leading TAR scholarship support reporting of recall and precision estimates with confidence intervals as the methodologically correct validation practice. In actual production practice, however, validation reporting is more commonly expressed as point estimates from a control-set or elusion-test sample without explicitly computed confidence intervals. The exemplar requires disclosure of the underlying sampling parameters—sample size, sampling methodology, and observed counts—so that the receiving party may compute confidence intervals or otherwise characterize the precision of the estimates if it wishes to do so, without obliging the producing party to a computational format that not all review platforms generate natively.

aggregate validation results upon reasonable request.	human verification calibrated to the task type is the right discipline.
Foreign-Language Materials	
Where ESI in a language other than English is responsive to discovery, the producing party shall produce such ESI in its original language with all associated metadata. Native-language text shall be preserved in the load file using a Unicode encoding (UTF-8 with byte-order mark or UTF-16 LE). Translation of foreign-language ESI is not required as a condition of production. Where a Party intends to rely on a translation in a filing, deposition, or proceeding, that Party shall produce the translation, identify the source document by Bates number, and identify the means of translation (e.g., certified human translation, machine translation, or hybrid). The Parties shall meet and confer regarding any obligation to identify the language of each foreign-language item in the load file (e.g., via a Language metadata field), allocation of cost for translation, and validation of machine translations on which a Party intends to rely.	In matters with foreign custodians, multinational subsidiaries, or non-U.S. counterparties, foreign-language evidence is the rule. The exemplar establishes three points: (1) produce in the original language with proper Unicode encoding (do not "translate to produce"—that is producing-party editorial work, not discovery); (2) translation costs follow use, with the relying party paying for and producing the translation; and (3) language identification at the metadata level is helpful but negotiable. Machine translation in 2026 is extraordinarily good for major languages and routine business prose, and is materially less reliable for legal terminology, technical jargon, and lower-resource languages. Counsel should validate machine translations they intend to rely on, especially when those translations underpin substantive arguments. The protocol cannot solve the validation problem; it can ensure that the methodology is disclosed when the translation is used.
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.	Belt-and-suspenders. The provision exists to prevent a clever opponent from arguing that, by agreeing to the form of production, a party also agreed to the scope of production or waived rights as to objections. Worth keeping.

Metadata Production Fields

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 e-discovery software for processing. A hash value is calculated for each file. A Bates number is assigned to each file or page image. Not every e-discovery vendor can supply every field below, and some use different field names for the same data.

The field set has expanded since the prior edition to accommodate Modern Attachments, Short Messages, mobile messages, and audio/video evidence. Where a particular field is not applicable to a particular item type, the field may be left empty; load-file processors should not require every field to be populated for every item.

Addendum A — Production Metadata Fields

Field Name	Description
Identifiers and Bates	
BegBates	First Bates identifier of item.
EndBates	Last Bates identifier of item.
BegAttach	First Bates identifier of attachment range.
EndAttach	Last Bates identifier of attachment range.
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.
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).
AttachCount	Number of attachments to an e-mail or parent message.
AttachName	Names of each individual attachment, separated by semicolons.
Source and Custody	
Custodian	E-mail: mailbox where the email resided. Native: individual from whom the document originated. Short Messages: account or workspace identifier.
OtherCustodians	Custodians whose file/message has been deduplicated; separated by semicolons.
Source	Source repository or system from which the item was collected (e.g., Exchange Online, Slack workspace name, mobile device identifier).
Path	Original location of item including original file name.
FileName	Original name of file as it appeared in the location where collected.

Field Name	Description
FileExt	File extension.
FileType	File type as identified by the processing tool.
FileSize	Size of native file/message in KB.
PgCount	Number of pages in the document.
Production Paths	
NativeLink	Relative path and filename for native file on production media.
TextLink	Relative path and filename for text file on production media.
ImageLink	Relative path and filename for first page image of the document on production media (where applicable).
Email-Specific	
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.
Subject	E-mail: subject line.
DateSent	E-mail: date and time the email was sent (mm/dd/yyyy hh:mm:ss with time zone).
DateReceived	E-mail: date and time the email was received (mm/dd/yyyy hh:mm:ss with time zone).
MsgID	E-mail: unique Message-ID field.
EDRM_MIH	EDRM Message Identification Hash — a standardized methodology for calculating a hash value across the headers and body of an email message that yields consistent duplicate identification across processing tools and platforms. Parties are encouraged to populate this field; broader adoption of the MIH meaningfully improves cross-vendor deduplication and provenance tracking in multi-producing-party matters. *Specifications:* edrm.net/edrm-projects/dupeid-2/
Document-Specific	
Title	Document: title provided by user within the document.
Author	Document: author or last-saved-by attribution.
ModifiedDate	Document: last-modified date and time (mm/dd/yyyy hh:mm:ss with time zone).
CreationDate	Document: created date and time (mm/dd/yyyy hh:mm:ss with time zone).
HiddenContent	Denotes presence of tracked changes / hidden content / embedded objects in item(s) (Y/N).
Modern Attachments / Linked Files	

Field Name	Description
LinkedFile	Denotes that the item is a Modern Attachment / Linked File rather than an embedded attachment (Y/N).
LinkedFileURL	URL of the linked resource as transmitted in the parent communication.
LinkedFilePlatform	Source platform of the linked resource (e.g., OneDrive, SharePoint, Google Drive, Dropbox).
LinkedFileVersion	Version identifier or modification timestamp of the linked file as produced (point-in-time, where reasonably available).
LinkedFileVersionStatus	Whether the produced version is the point-in-time version (PIT), the most contemporaneous version available (CONTEMP), or unavailable (UNAVAIL).
Short Messages and Chat	
MsgPlatform	Source messaging platform (e.g., Slack, Teams, Google Chat, iMessage, SMS, MMS, RCS, WhatsApp, Signal).
Workspace	Workspace, tenant, or account identifier of the source platform.
Channel	Channel, room, group chat, or conversation identifier of the source platform.
ConversationID	Identifier of the Conversation as defined in this Protocol.
ThreadID	Identifier of the thread parent (where the source platform supports threading).
MsgTimestamp	Timestamp of the individual message (mm/dd/yyyy hh:mm:ss with time zone), where messages are produced as discrete records.
Participants	Participants in the Conversation; separated by semicolons.
EditedFlag	Denotes whether the message was edited after sending (Y/N), where the source platform records edit history.
DeletedFlag	Denotes whether the message was deleted from the source platform (Y/N), where the source platform records deletion history.
Reactions	Reactions associated with the message (e.g., emoji reactions and the user identifiers of those reacting); separated by semicolons.
Mobile-Specific	
DeviceID	Identifier of the source device (e.g., IMEI, serial number, or producing-party-assigned identifier).
PhoneNumber	Phone number associated with the message account, where applicable.
EphemeralFlag	Denotes whether the source Conversation was configured with disappearing-messages or auto-delete settings during the relevant period (Y/N).
Audio, Video, and Voicemail	
Duration	Duration of audio or video recording (hh:mm:ss).
Codec	Codec of audio or video recording, where reasonably available.
CaptureDevice	Capture device identifier, where reasonably available.
CaptureDate	Capture date and time (mm/dd/yyyy hh:mm:ss with time zone), where reasonably available.

Field Name	Description
TranscriptLink	Relative path and filename for any produced transcript on production media.
TranscriptMethod	Means of transcription (e.g., HUMAN, MACHINE, HYBRID).
Foreign Language	
Language	Primary language of item content (ISO 639 code), where identified.
Markings and Status	
Hash	Hash value of the item (algorithm specified in Protocol).
Redacted	Denotes that the item has been redacted as containing privileged content (Y/N).
Confidential	Denotes that the item has been designated as confidential pursuant to confidentiality agreement or protective order (Y/N).
DeDuped	Custodian instances suppressed by deduplication; separated by semicolons (and corresponding to OtherCustodians where applicable).

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

Hybridized TIFF+ | Version 20260501

This Exemplar ESI Protocol presents the production language of the Annotated ESI Protocol above as a single-document protocol suitable for adaptation to a particular case. It uses a hybridized TIFF+ form of production: TIFF as the default for word-processed documents, emails, and PDFs, with broad carve-outs for native production where TIFF imaging would degrade the evidence. Refer to *The Annotated ESI Protocol* article above for the rationale and trade-offs reflected in each provision.

1. Definitions

1.1 "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 and short messages exchanged over collaboration and chat platforms, mobile messages (including SMS, MMS, RCS, and platform-native messages such as iMessage), transaction logs, audio and video files, voicemail, internet data, computer logs, social-media posts and direct messages, and backup materials. The term "Document(s)" shall include Hard Copy Documents, Electronic Documents, and Electronically Stored Information (ESI) as defined herein.

1.2 "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); image files (e.g., PDF, JPEG, TIFF, HEIC); short messages and chat content from collaboration platforms (e.g., Slack, Microsoft Teams, Google Chat); mobile messages (e.g., iMessage, SMS, MMS, RCS, WhatsApp); audio and video files; and the metadata associated with each.

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

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

1.5 "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; provided, however, that the parties may by agreement substitute SHA-1, SHA-256, or any other supported cryptographic hash algorithm.

1.6 "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, and OCR or

extracted text. A load file linking corresponding images is used for productions of static images (e.g., TIFFs).

1.7 "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.

1.8 "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. The "native format" of short messages and chat content is the export format produced by the source platform's administrative or compliance interface (e.g., Slack JSON exports), supplemented by any rendered transcript necessary for human readability.

1.9 "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.

1.10 "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.

1.11 "Modern Attachment" or "Linked File" means a document or other resource referenced in an electronic communication by hyperlink or pointer, the substance of which is hosted in a cloud-based repository (e.g., OneDrive, SharePoint, Google Drive, Dropbox) and which the sender intended the recipient to access through the link as a functional substitute for an embedded file attachment.

1.12 "Short Message" means any message exchanged over a chat, instant messaging, collaboration, or mobile messaging platform, including but not limited to Slack, Microsoft Teams, Google Chat, iMessage, SMS, MMS, RCS, WhatsApp, Signal, Telegram, and direct messages exchanged on social-media platforms.

1.13 "Conversation" means a logically related sequence of Short Messages exchanged among a defined set of participants within a single channel, thread, group, or direct-message context.

1.14 "Ephemeral Message" means a Short Message that, by configuration of the source application, is set to be deleted automatically after a defined retention period or upon being read.

2. Preservation

The Parties represent that they have issued litigation hold notices to those custodians with data, and to persons or entities responsible for the maintenance of non-custodial data, which, based upon then-current information available, are reasonably likely to contain discoverable information. The hold shall include affirmative direction sufficient to suspend any auto-deletion, retention-policy expiration, or ephemeral-messaging settings that would otherwise destroy potentially relevant Short Messages.

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

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

(b) 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.

(c) Temporary internet files, browser history files, cache files, and cookies.

(d) 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.

(e) Server, system, or network logs.

(f) System and application files matching entries in the NIST National Software Reference Library.

3. E-Discovery 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. Each liaison shall be reasonably available to confer with opposing counsel's liaison on technical matters arising under this Protocol and to escalate disputes to lead counsel as appropriate.

4. 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. The producing party shall provide a data dictionary or schema reference sufficient to allow the receiving party to interpret field names, code values, and the relationships among tables.

5. 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.

Hard Copy Documents that contain color used to convey information shall be scanned and produced as 300 dpi JPG images at the highest-quality compression setting, in lieu of TIFF, with the same file-naming and OCR conventions.

6. 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.

7. 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. Treatment of Modern Attachments transmitted by hyperlink rather than as embedded attachments is governed by Section 8 below.

8. Modern Attachments / Linked Files

The Parties agree that documents transmitted to a recipient by hyperlink to a cloud-hosted resource ("Modern Attachments" or "Linked Files") shall be treated as attachments to the transmitting communication for purposes of discovery, except that documents merely referenced—as distinguished from those the sender intended the recipient to access through the link as a substitute for an embedded attachment—need not be produced as Modern Attachments.

To the extent reasonably feasible given the source platform's collection capabilities, the producing party shall collect and produce the version of each Modern Attachment that existed as of the time the link was sent (the "point-in-time version"). Where collection of the point-in-time version is not reasonably feasible, the producing party shall collect and produce the most contemporaneous version reasonably available, identify the version produced by version identifier or modification date, and disclose the limitation. Where the receiving party identifies particular Modern Attachments for which the point-in-time version is material to the issues in the case and the producing party's ordinary collection has not produced it, the Parties shall meet and confer regarding the proportionality of pursuing historical-version recovery through specialized tooling.

The producing party shall provide metadata sufficient to identify each Modern Attachment as such (e.g., a "LinkedFile" Boolean field), to associate each Modern Attachment with its parent communication (e.g., ParentBates and AttRange fields), and to identify the URL of the linked resource as transmitted, the platform of origin, and any version identifier.

Where a Modern Attachment cannot be collected or produced (e.g., the link has been broken, the resource has been deleted, or the producing party lacks access), the producing party shall so identify the link by URL and explain the basis for non-production.

9. Short Messages and Collaboration Platforms

Short Messages from collaboration and chat platforms (e.g., Slack, Microsoft Teams, Google Chat) shall be produced in a form that preserves the content, the participants, the channel or thread context, the

timestamps, and the family relationships between messages and any embedded or attached files. The producing party shall produce, for each responsive Conversation:

(a) A native or near-native export of the Conversation in the format provided by the source platform's administrative or compliance interface (e.g., Slack's JSON export, Microsoft Purview eDiscovery export, Google Vault export);

(b) A human-readable rendering of the Conversation as a single document per Conversation per day or per Conversation per logical unitization unit, paginated and Bates-numbered, with each message identifying its sender, timestamp (with time zone), and any associated reactions, edits, or deletions known to the producing party; and

(c) Any files attached to or embedded in the Conversation, produced as separate items with metadata associating each to its parent Conversation.

The requirement of a human-readable rendering under (b) above is conditioned on the availability of platform tooling capable of producing such rendering or upon reasonable request at proportionate cost. Where rendering would require specialized vendor engagement disproportionate to the demands of the matter, the Parties shall meet and confer regarding alternatives, which may include native production accompanied by a documented procedure for rendering on demand.

Unitization of Short Messages shall be at the Conversation-day level (one document per channel, thread, or DM context per calendar day) as the default target unitization, provided that where the source platform's native export structure does not align with Conversation-day boundaries, the producing party may produce in the source platform's native unitization, identify the rule applied, and provide metadata sufficient for the receiving party to verify the unitization against the underlying export.

Where the source platform records edits, deletions, or message-revision history and that history is reasonably available to the producing party, the producing party shall produce the edit/deletion history as part of the rendered Conversation or as accompanying metadata. Whether such history is reasonably available depends on the source platform's configuration and the producing party's subscription tier; the producing party shall disclose the platform tier and the audit-log availability applicable to the source data sufficient for the receiving party to evaluate any limitations on the production of edit and deletion history.

Reactions (e.g., emoji reactions to a message) shall be preserved in the rendered Conversation when reasonably available.

10. Mobile and Ephemeral Messaging

Mobile messages (including iMessage, SMS, MMS, RCS, WhatsApp, Signal, Telegram, and platform-native messaging on personal or business mobile devices) responsive to discovery requests shall be produced as:

(a) (a) An export of the responsive Conversation produced by a tool capable of preserving the message content, the participants, the timestamps (with time zone), and any attachments, in a form that allows the receiving party to verify the integrity of the export. Acceptable tools range from consumer-grade backup-extraction utilities (e.g., iMazing, Decipher Text Message, AnyTrans, or comparable) to forensic-grade mobile collection platforms (e.g., Cellebrite, Magnet

AXIOM, Oxygen Forensic Detective, MSAB XRY, or comparable). The choice of tool shall be proportionate to the demands of the matter and the evidence-integrity issues presented;

(b) (b) A human-readable rendering of the Conversation, paginated and Bates-numbered, identifying each message's sender, recipient(s), timestamp (with time zone), and any attachments;

(c) (c) Any media attached to or embedded in the Conversation, produced as separate items with metadata associating each to its parent Conversation; and

(d) (d) A statement identifying the tool used to produce the export, the tool version, and the date and method of collection.

Where the matter involves (i) alleged spoliation or deletion of mobile evidence, (ii) recovery of deleted content from device storage, (iii) a dispute as to the integrity of the source device or its contents, or (iv) circumstances in which a party seeks judicial relief premised on the comprehensiveness of the mobile collection, the producing party shall use a forensic-grade mobile collection tool unless otherwise agreed or ordered. The Parties shall meet and confer regarding the appropriate level of forensic rigor where the demands of the case so warrant. The foregoing trigger conditions govern the choice of tool prospectively and do not, of themselves, require re-collection of devices already collected; any further forensic examination of previously collected devices shall be addressed by agreement of the Parties or by order of the Court.

The producing party shall preserve and produce, where reasonably available, the underlying message database or backup file from the source device (e.g., the SMS/iMessage SQLite databases on iOS, or the iTunes/Finder backup from which the export was produced) in addition to any rendered transcript.

Where any source device or messaging account was configured at any time during the preservation period to delete messages automatically after a defined interval (an "auto-delete" or "disappearing-messages" setting), the producing party shall disclose the configuration history (or, if unavailable, the configuration as of the date the litigation hold was implemented), the affected platforms and accounts, and the steps taken to suspend such settings upon implementation of the litigation hold.

11. Audio, Video, and Voicemail

Audio recordings (including voicemail), video recordings, and other rich-media files responsive to discovery requests shall be produced in their native format with original metadata preserved (including, where available, capture device, capture date, duration, and codec information).

Where a producing party intends to rely on a transcript of any responsive audio or video evidence, the producing party shall produce the transcript with metadata identifying the source recording (by Bates number) and the means of transcription (e.g., automated speech-to-text, human transcription, or hybrid). A transcript is not a substitute for the underlying recording, and the underlying recording remains the operative evidence.

The Parties shall meet and confer regarding any obligation to provide transcripts as a searchability surrogate for audio and video, including allocation of cost and validation of accuracy.

12. Hard Copy Document Metadata

The following metadata fields should be provided for Hard Copy Documents when reasonably available:

- Beginning Bates number
- Ending Bates number
- First attachment Bates number
- Last attachment Bates number
- Source location/custodian
- Confidentiality designation
- Redacted (Y/N)
- Extracted/OCR text file path

13. Forms of 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" × 11" 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 format:

(a) Spreadsheets (e.g., Microsoft Excel, Google Sheets, Apple Numbers)
(b) Presentations (e.g., PowerPoint, Keynote, Google Slides)
(c) Databases and database extracts (e.g., Microsoft Access, SQL exports)
(d) Delimited text files (e.g., CSV, TSV)
(e) Photographs and other native image files (e.g., JPEG, PNG, HEIC, RAW)
(f) Audio, video, and voicemail recordings
(g) Short Messages and Conversations from collaboration platforms (per Section 9)
(h) Mobile messages (per Section 10)
(i) Structured-data exports (per Section 4)
(j) Computer-aided design (CAD), engineering drawing, and other format-specific files for which TIFF rendering would materially impair fidelity
(k) Documents of a type which cannot be reasonably converted to useful TIFF images.

All images of documents that 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, hidden columns or rows, hidden slides, or other user-entered data that the source application can display to the user shall be processed such that all that data is visible in the image, or, in the alternative, produced natively.

Documents that contain color used to convey information (e.g., color coding and highlighting versus merely decorative use) shall be produced as 300 dpi JPG images at the highest-quality compression setting, in lieu of TIFF, with the same file-naming and metadata conventions, or, in the alternative, produced natively.

14. File Names

Each TIFF image (or JPG image produced for color or paper documents) shall have a unique file name corresponding to the Bates number of that page with the appropriate file extension. The file name shall not contain any blank spaces and shall be zero-padded (e.g., DEF-0000001), 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, the producing party shall so note in a cover letter or production log accompanying the production. Bates numbers shall be unique across the entire production and prefixes shall be consistent across all documents produced.

The producing party will brand all TIFF and JPG images in the lower right-hand corner with their corresponding Bates number without obscuring any part of the underlying image.

15. 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. If the document is an image file or contains redactions, a text file created using OCR shall be produced in lieu of extracted text.

16. Load Files

Productions will, as applicable, include image load files in Opticon (.opt) or IPRO (.lfp) format and Concordance-format data (.dat) files with the applicable metadata fields identified in Addendum A. All metadata files shall be encoded as UTF-8 with byte-order mark or UTF-16 LE.

All native format files shall be produced in a folder named “NATIVE.”

All TIFF and JPG 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.

17. 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 degrade the original image.

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.

18. 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 text file containing only the 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.

Where a producing party redacts a Short Message, audio recording, video recording, or other non-paginated form of evidence, the producing party shall describe the redaction (the location, the duration, or the nature of the content redacted) with sufficient particularity to allow the receiving party to evaluate the basis for the redaction without revealing privileged content.

19. Privilege Logs

With each production, the 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.

Where the volume of withheld items makes a document-by-document log impracticable, the Parties may agree to a metadata-based or categorical log. Any agreement on the form of the log shall identify the categories or fields used and the bases on which categorical or metadata logging is appropriate.

20. Deduplication

The producing party may horizontally (globally) deduplicate documents based on MD5, SHA-1, SHA-256, Message ID, EDRM MIH, or other standard methodology for email deduplication across multiple custodians or data sources. Attachments to parent documents may not be deduplicated where a duplicate standalone version of the attachment exists, and standalone versions of documents may not be suppressed where a duplicate version exists as an attachment.

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

21. 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.

22. 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.

Where threading is employed, the producing party shall provide metadata sufficient to identify the threading relationships among produced and suppressed messages and to allow the receiving party to reconstruct the thread.

23. Production Media and Transmission

The producing party will use the appropriate electronic media (encrypted hard drive, encrypted thumb drive, secure file transfer, or secure file-sharing service) for its ESI production and will endeavor to use the highest-capacity suitable medium and the most efficient transmission method consistent with the security requirements set forth herein.

The producing party will label the production media (or production package, in the case of secure transfer) with the name of the producing party, production date, media volume name, and Bates number range(s).

Productions on physical media shall be encrypted using AES-256 or comparable encryption. Productions transmitted electronically shall be transmitted via secure means (e.g., SFTP, HTTPS, or a vendor-provided secure file-sharing service with at-rest encryption). At the time of production and under separate cover, the producing party shall furnish decryption credentials to the receiving party.

24. 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.

The producing party shall maintain and provide upon reasonable request a log of items excluded from production due to processing exceptions, identifying the item, the nature of the exception, and the steps taken to resolve it.

Parties shall normalize times and dates to conform to [UTC] OR [specified local time zone]. Where dates and times in the source data are recorded with time-zone information, that information shall be preserved in the load file.

For archive files (e.g., .zip, .jar, .rar, .gz, .tar, .7z), all contents shall 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.

25. Search Methodology and Validation

The Parties may employ keyword search, technology-assisted review (TAR), continuous active learning (CAL), or other technology-assisted methodologies to identify potentially responsive ESI for review and production. The producing party retains the right to choose its review methodology and tooling.

Where keyword search is employed, the Parties shall meet and confer regarding the search terms and refinements to be used. The producing party shall disclose, upon reasonable request, the final search syntax applied, the data sources to which the searches were applied, and the resulting hit counts.

Where technology-assisted review is employed, the producing party shall disclose, upon reasonable request, the general methodology applied (e.g., predictive coding, continuous active learning, clustering), the validation methodology used to confirm the adequacy of the review, and a summary of the validation results, including the underlying sampling parameters (sample size, sampling methodology, and observed counts) sufficient to characterize the precision of the recall and precision estimates reported. Disclosure of seed sets, training sets, and individual coding decisions is not required absent a particularized showing of need.

26. Use of Generative Artificial Intelligence in Review

Generative artificial intelligence tools may be used by either Party in support of review and production workflows, including for responsiveness review, privilege review, redaction, summarization, deposition preparation, and related tasks, subject to the following:

(a) The producing party shall ensure that any generative AI tool used in the review of potentially privileged or confidential ESI is hosted and operated in a manner that does not result in the transmission of such ESI to third parties or its incorporation into the training data of any general-purpose model.

(b) The use of generative AI does not relieve the producing party of its obligations under the applicable Rules of Civil Procedure, including the duties of competence, diligence, and accuracy. Output from generative AI tools shall be subject to human review sufficient to verify its accuracy before reliance for any responsiveness or privilege determination.

(c) The disclosure obligations applicable to technology-assisted review under Section 25 shall apply equally to the use of generative AI tools, with disclosure of the general methodology, validation approach, and aggregate validation results upon reasonable request.

27. Foreign-Language Materials

Where ESI in a language other than English is responsive to discovery, the producing party shall produce such ESI in its original language with all associated metadata. Native-language text shall be preserved in the load file using a Unicode encoding (UTF-8 with byte-order mark or UTF-16 LE).

Translation of foreign-language ESI is not required as a condition of production. Where a Party intends to rely on a translation in a filing, deposition, or proceeding, that Party shall produce the translation, identify the source document by Bates number, and identify the means of transcription or translation (e.g., certified human translation, machine translation, or hybrid).

The Parties shall meet and confer regarding any obligation to identify the language of each foreign-language item in the load file (e.g., via a Language metadata field), allocation of cost for translation, and validation of machine translations on which a Party intends to rely.

28. 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.

29. Production Metadata Fields

The Parties shall produce the metadata fields specified in Addendum A above, populated where reasonably available for each item type. Where a particular field is not applicable to a particular item type, the field may be left empty.

Addendum A — Production Metadata Fields

The Parties shall produce the metadata fields specified below, populated where reasonably available for each item type, in the Concordance-format data (.dat) file delivered with each production. Where a particular field is not applicable to a particular item type, the field shall be present in the row but may be left empty. The Parties shall not omit fields specified herein from the Concordance-format data file structure on grounds of non-applicability; empty values are acceptable, missing fields are not.

Field Name	Description
Identifiers and Bates	
BegBates	First Bates identifier of item.
EndBates	Last Bates identifier of item.
BegAttach	First Bates identifier of attachment range.
EndAttach	Last Bates identifier of attachment range.
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.
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).
AttachCount	Number of attachments to an e-mail or parent message.
AttachName	Names of each individual attachment, separated by semicolons.
Source and Custody	
Custodian	E-mail: mailbox where the email resided. Native: individual from whom the document originated. Short Messages: account or workspace identifier.
OtherCustodians	Custodians whose file/message has been deduplicated; separated by semicolons.
Source	Source repository or system from which the item was collected (e.g., Exchange Online, Slack workspace name, mobile device identifier).
Path	Original location of item including original file name.
FileName	Original name of file as it appeared in the location where collected.
FileExt	File extension.
FileType	File type as identified by the processing tool.
FileSize	Size of native file/message in KB.
PgCount	Number of pages in the document.
Production Paths	
NativeLink	Relative path and filename for native file on production media.

Field Name	Description
TextLink	Relative path and filename for text file on production media.
ImageLink	Relative path and filename for first page image of the document on production media (where applicable).
Email-Specific	
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.
Subject	E-mail: subject line.
DateSent	E-mail: date and time the email was sent (mm/dd/yyyy hh:mm:ss with time zone).
DateReceived	E-mail: date and time the email was received (mm/dd/yyyy hh:mm:ss with time zone).
MsgID	E-mail: unique Message-ID field.
EDRM_MIH	EDRM Message Identification Hash — a standardized methodology for calculating a hash value across the headers and body of an email message that yields consistent duplicate identification across processing tools and platforms. Parties are encouraged to populate this field; broader adoption of the MIH meaningfully improves cross-vendor deduplication and provenance tracking in multi-producing-party matters. *Specifications:* edrm.net/edrm-projects/dupeid-2/
Document-Specific	
Title	Document: title provided by user within the document.
Author	Document: author or last-saved-by attribution.
ModifiedDate	Document: last-modified date and time (mm/dd/yyyy hh:mm:ss with time zone).
CreationDate	Document: created date and time (mm/dd/yyyy hh:mm:ss with time zone).
HiddenContent	Denotes presence of tracked changes / hidden content / embedded objects in item(s) (Y/N).
Modern Attachments / Linked Files	
LinkedFile	Denotes that the item is a Modern Attachment / Linked File rather than an embedded attachment (Y/N).
LinkedFileURL	URL of the linked resource as transmitted in the parent communication.
LinkedFilePlatform	Source platform of the linked resource (e.g., OneDrive, SharePoint, Google Drive, Dropbox).
LinkedFileVersion	Version identifier or modification timestamp of the linked file as produced (point-in-time, where reasonably available).

Field Name	Description
LinkedFileVersionStatus	Whether the produced version is the point-in-time version (PIT), the most contemporaneous version available (CONTEMP), or unavailable (UNAVAIL).
Short Messages and Chat	
MsgPlatform	Source messaging platform (e.g., Slack, Teams, Google Chat, iMessage, SMS, MMS, RCS, WhatsApp, Signal).
Workspace	Workspace, tenant, or account identifier of the source platform.
Channel	Channel, room, group chat, or conversation identifier of the source platform.
ConversationID	Identifier of the Conversation as defined in the Protocol.
ThreadID	Identifier of the thread parent (where the source platform supports threading).
MsgTimestamp	Timestamp of the individual message (mm/dd/yyyy hh:mm:ss with time zone), where messages are produced as discrete records.
Participants	Participants in the Conversation; separated by semicolons.
EditedFlag	Denotes whether the message was edited after sending (Y/N), where the source platform records edit history.
DeletedFlag	Denotes whether the message was deleted from the source platform (Y/N), where the source platform records deletion history.
Reactions	Reactions associated with the message (e.g., emoji reactions and the user identifiers of those reacting); separated by semicolons.
Mobile-Specific	
DeviceID	Identifier of the source device (e.g., IMEI, serial number, or producing-party-assigned identifier).
PhoneNumber	Phone number associated with the message account, where applicable.
EphemeralFlag	Denotes whether the source Conversation was configured with disappearing-messages or auto-delete settings during the relevant period (Y/N).
CollectionTool	Tool and version used to produce the export (e.g., “iMazing 3.0.7,” “Cellebrite UFED 7.66”). Required for mobile productions.
CollectionDate	Date the export or collection was produced (mm/dd/yyyy). Required for mobile productions.
CollectionMethod	Method of collection (e.g., iTunes/Finder backup extraction, logical extraction, file-system extraction, advanced logical, physical extraction). Required for mobile productions.
Audio, Video, and Voicemail	
Duration	Duration of audio or video recording (hh:mm:ss).
Codec	Codec of audio or video recording, where reasonably available.
CaptureDevice	Capture device identifier, where reasonably available.
CaptureDate	Capture date and time (mm/dd/yyyy hh:mm:ss with time zone), where reasonably available.

Field Name	Description
TranscriptLink	Relative path and filename for any produced transcript on production media.
TranscriptMethod	Means of transcription (e.g., HUMAN, MACHINE, HYBRID).
Foreign Language	
Language	Primary language of item content (ISO 639 code), where identified.
Markings and Status	
Hash	Hash value of the item (algorithm specified in Protocol).
Redacted	Denotes that the item has been redacted as containing privileged content (Y/N).
Confidential	Denotes that the item has been designated as confidential pursuant to confidentiality agreement or protective order (Y/N).
DeDuped	Custodian instances suppressed by deduplication; separated by semicolons (and corresponding to OtherCustodians where applicable).