level legal®

eDiscovery 101

A Field Guide for Associates

Updated July 2026

How to Use This Guide

This is a field guide. You are not expected to read it cover to cover. It is built to be opened at the moment you need it — when a partner hands you a new matter, when opposing counsel sends an electronically stored information (ESI) protocol you don't understand, or when someone uses a three-letter acronym in a meeting, and you need to know what it means without asking.

It is organized around a single idea: eDiscovery is not a technical specialty you delegate and forget. It is a sequence of five decisions you own, each of which can advance your case strategically or derail it publicly. The guide walks you through those five decisions, provides a working checklist for each phase, and shows you the vocabulary and the escalation instincts needed to operate credibly from day one.

A note on jurisdiction. The federal rules are the backbone of eDiscovery practice and the common vocabulary everyone uses, so this guide is based on the federal rules. Be aware that state rules can and do diverge from the guidance in this document. For state court matters, refer to the rules of your jurisdiction.

A note on what you are not. You are the lawyer, not the eDiscovery provider. You may rarely run the tools yourself. Your job is to make defensible decisions, supervise the people who execute them, and be able to explain every choice to a judge. Keep that framing, and most of this guide falls into place.

Part 1 | The Big Picture

Why a First-Year Should Care

Discovery is where most modern cases are won and lost, and it happens long before trial. It is also where associates often bear much of the day-to-day responsibility. This includes running searches, drafting legal hold notices, coordinating collections, and communicating with opposing counsel about what has and has not been done. Those responsibilities carry real consequences. "I didn't know" is not a defense. The courts have been clear for two decades that competence with ESI is a basic skill required to practice law.

The stakes are concrete. In *DR Distributors v. 21 Century Smoking*, counsel represented that production was complete without reasonably investigating the client's systems; the result was roughly $2.5 million in sanctions and lasting credibility damage. In *UANPF v. Carvana*, a party agreed to produce hyperlinked documents without testing feasibility and was forced into a pilot that cost more than $200,000 and 1,170 hours for just two custodians (custodians are people who hold or control potentially relevant data). Neither of those was a technology failure. Both were *decision* failures — someone agreed to something, or skipped something, that a little discipline would have caught.

Your Duty of Competence

ABA Model Rule 1.1, Comment 8 states that keeping abreast of the benefits and risks of relevant technology is part of maintaining competence. This is the rule that makes eDiscovery your problem and not just the vendor's.

The other ethical duties that bear on discovery — candor to the tribunal, confidentiality, and supervision of the vendors and contract reviewers you engage — are covered in Part 5. Read them before your first production goes out.

The Map: The EDRM and the Litigation Timeline

The Electronic Discovery Reference Model (EDRM) is the industry's shared map of the eDiscovery lifecycle. It is not a strict assembly line — in practice, you loop back constantly as you learn more about the data — but it gives everyone a common vocabulary for the stages. Each stage answers one plain question.

EDRM Stage	The Question It Answers
Information Governance	Before litigation starts, what data exists, and how is it kept?
Identification	Where might potentially relevant evidence live, and who has it?
Preservation	How do we stop data from disappearing?
Collection	How do we defensibly gather data?
Processing	How do we make data searchable and reduce it to a workable set?
Review	What is relevant, privileged, or significant?
Analysis	What does the evidence tell us?
Production	What do we hand to the other side, and in what form?
Presentation	How do we use data to prove the case?

Overlaid on the case timeline, the stages line up roughly like this: the duty to preserve and the complaint trigger the Identification and Preservation phases; the Rule 26(f) conference is where you negotiate scope, search, and format; the bulk of Collection, Processing, Review, and Analysis happens during the discovery period; and Production and Presentation feeds the depositions, dispositive motions, and, ultimately, trial.

The single most important thing to internalize: Early decisions control every downstream outcome. A preservation miss in week one surfaces as a sanctions motion a year later.

Part 2 | The Five Decisions You Own

This is the spine of the process. Strip away the vocabulary and the tooling, and eDiscovery is five decisions. In this guide, each one has a question you must be able to answer "yes" to, a short list of what goes wrong, and a real case showing the cost of getting it wrong. The phase checklists in Part 3 are the operational details behind these decisions.

The Decision	The Question You Must Answer "Yes" To
Identification and Preservation	Do we know where the evidence lives, and have we ensured it won't be lost?
Collection	Is this collection complete, defensible, and metadata-intact?
Meet and Confer/ESI Protocol	Are we agreeing only to things we can execute?
Review and Analysis	Is our workflow reasonable, consistent, and validated?
Production	Is what we produced complete, accurate, and defensible?

Decision 1: Identification and Preservation

The question: Do we know where the evidence lives, and have we ensured it won't be lost?

Identification is the process of determining who the key players (custodians) are and where potentially relevant data resides. Preservation is protecting that data from alteration or destruction the moment litigation is reasonably anticipated. These run together because you can't preserve what you haven't identified.

The duty to preserve is triggered by reasonable anticipation of litigation — which can be well before a complaint is filed. Federal rules are silent on how to preserve data; the obligation comes from common law. The practical mechanism is the legal hold: a written directive telling custodians to preserve relevant material and suspend auto-deletion. Issuing the notice is the beginning, not the end. You must monitor compliance.

Where it goes wrong: preserving too late; issuing a hold and never following up; forgetting departing employees and decommissioned systems; missing modern sources like Teams, Slack, texts, Generative AI (GenAI) chat logs, and ephemeral messaging apps; and relying on the client's honor system instead of verifying.

Case study — the classic and the modern. In *Zubulake v. UBS Warburg*, counsel issued a hold but failed to supervise compliance; backup tapes were overwritten, and the court gave an adverse-inference instruction. Twenty years later, the same lesson recurs: In *In re Google Play Store Antitrust Litigation*, Google left a 24-hour auto-delete function active on internal chats and let employees decide for themselves what to preserve. The court held that a litigant cannot rely on an unmonitored honor system, leading to severe sanctions and millions in fees. You don't merely issue a hold; you make sure it works.

Decision 2: Collection

The question: Is this collection complete, defensible, and metadata-intact?

Collection is the acquisition of the ESI you identified. The central tension is proportionality versus completeness. You have three basic methods, trading speed and cost against risk:

- **Full forensic collection** — a forensically sound acquisition that preserves not only the user-created data but also system artifacts, metadata, and other evidence that may become relevant later. Often the lowest-risk approach is essential when authenticity, deletion activity, user behavior, timeline reconstruction, or other forensic questions are at issue. It is not always the most expensive or time-consuming option; the cost and complexity depend heavily on the data source.
- **Targeted collection** — selected ESI from specific custodians, locations, date ranges, or repositories, collected in a defensible manner with appropriate validation and chain-of-custody documentation. The workhorse method for most matters because it balances proportionality, efficiency, and defensibility.
- **Self-collection** — the custodian or client collects and provides their own data. Typically, the least expensive option, but also the one with the greatest risk of omission, alteration, over-collection, or spoliation. Rarely appropriate without clear instructions, lawyer oversight, and independent validation of what was collected.

Where it goes wrong: collection methods that strip or alter metadata (which can itself be spoliation); letting an inexperienced client self-collect and treating counsel as a passive conduit; and failing to document the chain of custody so you can't later prove what you did.

The Four T's test. Before you let clients collect their own data, run the Four T's test. If the answer to any of these questions is "no," get help:

1. **Time** — Does the client's IT team have time to do this properly?
2. **Technology** — Does the client have the right tools to collect defensibly?
3. **Training** — Does the client know how to use those tools and document the process?
4. **Testimony** — Would you and the client be comfortable testifying to the defensibility of this collection?

Case study. In *EEOC v. Formel D*, the client self-collected and counsel acted as a passive conduit. The court held that counsel must test and validate the accuracy of a collection. This resulted in spoliation motions granted in part against counsel. Trust your client but verify the data.

Decision 3: Meet and Confer/ESI Protocol

The question: Are we agreeing only to things we can execute with certainty?

The Rule 26(f) meet and confer is the strategic leverage point of the whole case. It is where the parties define discovery scope, search methodology, and production format, ideally memorialized in an ESI protocol. Rule 26(b)(1) sets the governing standard: discovery must be relevant *and* proportional to the needs of the case.

The Sedona Principles, an industry-standard best practices guide published by the Sedona Conference®, capture the two ideas you'll hear most: cooperation is expected (Principle 1), and the responding party is generally best situated to decide how to meet its own discovery obligations (Principle 6).

What actually matters in the protocol: scope (custodians, sources, date ranges); search approach (keywords, technology-assisted review (TAR), GenAI, and how you'll validate); production format (native vs. image, required metadata, structured data); privilege protection (see Appendix 1 for a sample Rule 502(d) order) and — the one associates often forget — feasibility. Can your client comply with what you're about to sign?

Where it goes wrong: agreeing to something to seem cooperative, then discovering you can't do it; overpromising on format; and skipping the 502(d) order that would have protected you.

Case studies. Two cautionary tales already introduced in Part 1 belong here. In *DR Distributors*, counsel represented completeness without reasonable inquiry and drew roughly $2.5 million in sanctions — if you can't execute it, don't agree to it. In *Carvana*, agreeing to produce hyperlinked documents before testing feasibility forced a $200,000-plus, 1,170-hour pilot for two custodians. Test feasibility before you sign the protocol.

Privilege at Scale: What a Rule 502(d) Order Does for You

When you produce hundreds of thousands of documents, some privileged material will occasionally slip through no matter how careful the review. Absent protection, producing a privileged document can waive privilege — potentially over the whole subject matter. Federal Rule of Evidence 502(d) is the fix: it allows the court to enter an order stating that production of privileged material does not waive privilege in that case *or any other* federal or state proceeding. A 502(d) order is cheap insurance, and forgetting to ask for one is a preventable, career-denting mistake. Ask for it in every case where you're producing volume.

Decision 4: Review

The question: Is our workflow reasonable, consistent, and validated?

Review is where the legal team spends most of its time and money — commonly cited at 60–70% of total eDiscovery cost — because it is the most human-intensive stage. Review does two things: It identifies documents that are responsive (relevant and must be produced), and it identifies documents that are privileged (and must be withheld or redacted). Along the way, reviewers often code documents by issue to build the case narrative.

Three Tools, One Judgment that Stays Yours

You'll hear three approaches to getting through the documents. The one-line version to remember: TAR ranks documents, GenAI explains them, and neither replaces attorney judgment.

- **Human review** — attorneys reading documents. This is the baseline, still where the hardest calls land.
- **Technology-Assisted Review (TAR/predictive coding)** — the software learns from reviewer decisions on sample sets and then ranks the remaining documents by likely relevance. TAR 1.0 builds a static model from expert-coded seed sets; TAR 2.0, or Continuous Active Learning (CAL), keeps learning from every reviewer decision. You don't need to memorize the distinction — you need to know it prioritizes and ranks.
- **Generative AI (large language models)** — newer tools that can summarize, spot issues, and suggest relevance or privilege calls in natural language. These tools are powerful, but they predict language patterns, not legal truth, so their output is a starting point you validate, not an answer you adopt.

What Makes a Review Hold Up

Defensibility and effectiveness come from the same disciplines: clear definitions of relevance, privilege, and issues; consistent coding across reviewers; sampling and quality control (QC); a path to escalate hard calls; experienced oversight; and alignment between what you're coding and what you must ultimately produce.

The Standard Doesn't Change — Only How You Meet It

Rule 26(g) requires counsel to certify that a discovery response is complete and correct after a reasonable inquiry. That same standard applies whether you used human review, TAR 1.0 or 2.0, or a GenAI workflow. Every method must be validated through QC and sampling (for TAR and GenAI, that means measuring *recall* — did we find what we should have — and *precision* — how much of what we flagged was relevant). Courts evaluate every approach on the same three axes: reasonableness, proportionality, and defensibility.

Delegating Judgment Raises Risk

A useful way to think about technology risk is that the closer the tool gets to making the legal decision, the higher your exposure. Using TAR to prioritize documents or GenAI to summarize and suggest decisions are assistive, lower-risk uses. Letting TAR auto-code for privilege, or letting a GenAI make the final privilege call with no human validation, are decisive, higher-risk uses. Keep the human judgment attached to the decisions that carry legal consequences.

Decision 5: Production

The question: Is what we produced complete, accurate, and defensible?

Production is where mistakes become visible to the other side and the court. You deliver the responsive, non-privileged documents in the agreed format, with the agreed metadata (see Appendix 2 for List of Metadata Fields to Request for ESI Checklist). Quality control is non-negotiable and occurs both before and after the production. You must validate relevance calls, privilege calls, redactions, formatting, and data integrity. Confirm the redactions are "burned in." A redaction that only hides text on screen but leaves it in the underlying file is a classic, catastrophic error.

Where it goes wrong: producing in the wrong format; broken or incomplete metadata; redactions that didn't take; privileged documents slipping through (see the 502(d) discussion above); and inconsistent family handling, where an attachment gets produced but its parent email doesn't.

Part 3 | Phase Checklists

These are working checklists for each stage of the lifecycle, adapted from the Association of E-Discovery Specialists' (ACEDS) *E-Discovery Checklist Manifesto*. They are the operational detail behind the five decisions outlined above. Use them as a starting point and prune to the needs of your case — proportionality applies to your process too. Nothing here is a substitute for the specific rules of your forum or the judgment of the supervising lawyer.

Consultation and Scoping

- ☐ Discuss the nature and substance of the claims and the events that led to them
- ☐ Identify key dates and the key people involved
- ☐ Map the client's systems: document management, email, collaboration tools, hardware and software in use
- ☐ Ask whether the client has information-governance or retention policies — and whether they're actually followed
- ☐ Identify data subject to special legal requirements (HIPAA, financial, privacy regimes)
- ☐ Get an organization chart with names, roles, and locations; ask whether IT keeps a data map (if not, they should)
- ☐ Surface privacy and HR considerations early

Identification

- ☐ Build or obtain a data map of all relevant hardware, software, and systems
- ☐ Inventory custodial sources: email, texts, chat/collaboration apps, laptops, mobile devices, cloud accounts, home computers, removable media
- ☐ Inventory non-custodial sources: shared drives, databases, HR/finance systems, help desk, the Document Management System (DMS)
- ☐ Don't forget the modern and the easily missed sources: ephemeral apps (Signal, WhatsApp), Internet of Things (IoT)/connected devices, structured databases, network logs
- ☐ Identify all possible custodians and how each relates to the relevant data sources
- ☐ Conduct custodian interviews: role, date ranges, where they store things, personal-device use, who else might have relevant material

Preservation

- [ ] Confer with in-house/outside counsel and IT; determine the scope of the duty to preserve
- [ ] Suspend auto-delete and records-disposition practices for relevant data
- [ ] Review backup procedures and their impact on preservation
- [ ] Determine whether relevant ESI is reasonably accessible; weigh proportionality
- [ ] Identify and mitigate spoliation risks; account for departing employees and retiring systems
- [ ] Issue the legal hold — then monitor compliance, follow up, and re-issue reminders as the case develops
- [ ] Eventually, release the hold when the obligation ends

What Belongs in a Legal Hold Notice

- [ ] A plain summary of the dispute or investigation
- [ ] A statement of the duty to preserve and that normal deletion practices are suspended
- [ ] A clear directive not to alter, delete, or destroy relevant information
- [ ] The sources and types of ESI to preserve, and how to handle them
- [ ] A request that recipients acknowledge receipt
- [ ] A named contact for questions, and an appropriately scoped distribution list

Collection

- [ ] Finalize the scope; prepare the custodian and data-source list
- [ ] Choose the collection type (forensic, targeted, self) based on the needs of the case
- [ ] Confer with IT on access; confirm and schedule the collection
- [ ] Perform collection and validate the ESI (hash values, completeness)
- [ ] Obtain the acquisition/collection log and initiate chain-of-custody documentation
- [ ] Prepare a status report for the legal team

Processing

- [ ] Intake and catalog the data; update the chain of custody
- [ ] Extract metadata and text; index for search; flag exceptions (files that couldn't be opened)
- [ ] Apply defensible culling: date ranges, file-type filters, de-NISTing (removing system files), keyword filters
- [ ] Set the de-duplication policy (global vs. custodial; exact vs. near-dupe)
- [ ] QC the processing output; review anomalies; transfer to the review platform

Document Review

- [ ] Meet with the case team to plan the review; prepare a review protocol and key-document set
- [ ] Decide linear manual review vs. TAR or GenAI
- [ ] Determine coding: responsive, privileged, issue codes, confidentiality
- [ ] Assemble and train the review team; brief them on the issues
- [ ] Batch and prioritize; run first-level review, then second-pass and privilege review
- [ ] Apply redactions; monitor progress; run QC and report
- [ ] Assist in preparing the privilege and/or redaction log

Analysis

- [ ] Which custodians hold relevant documents, and are new custodians emerging?
- [ ] What dates, words, phrases, or events are becoming significant — and are there suspicious gaps?
- [ ] Which documents confirm your case, and which undermine it?
- [ ] Do any documents move motion or settlement strategy?
- [ ] Track review productivity and expenses against the budget

Production

- [ ] Plan and schedule the production; confirm format, tools, and techniques
- [ ] Send or request a sample load file to shake out format problems early
- [ ] Run an inconsistent-tagging and document-family conflict check
- [ ] Generate images/apply endorsements as agreed; select metadata fields for export
- [ ] QC the exported data, text, and images; confirm redactions are burned in
- [ ] Prepare deliverable media; QC it; update the production log
- [ ] Deliver by the agreed method; update privilege/redaction logs

Part 4 | Worked Scenarios

These are examples of fact patterns you might face. Use these examples to work through common issues now, before you're in the room. There is rarely one right answer. The point is to reason through the five decisions aloud, so you're ready when it counts.

Scenario A: The First 72 Hours

The facts. You're staffed on a new matter. A sales executive has left Company A (your client) and joined a competitor. Company A alleges the executive took confidential customer data and is misusing trade secrets. It's day one.

Work the decisions:

1. **Where does the key evidence live?** Think broadly: not only the executive's work email and laptop, but also Teams/Slack messages, their company phone, any personal devices used for work, cloud storage, USB activity, and building/VPN access logs. The forensic story of what was copied and when often lives in metadata and system logs, not documents.
2. **Who are your first custodians?** The departed executive, obviously — but also their manager, close colleagues, IT (who can speak to access and exfiltration evidence), and possibly HR.
3. **What must be preserved immediately?** Issue a legal hold immediately and make sure IT suspends any routine wiping or reimaging of the departed executive's laptop and phone. Departing employee devices are the number one thing that gets destroyed by the routine process right when you need them.
4. **What's the biggest risk of delay?** Spoliation. Every day the devices sit in the reissue queue, or auto-delete runs on the executive's mailbox and chats, is potentially lost evidence — and a future sanctions motion against you.
5. **What must you understand before negotiating discovery?** The feasibility of collecting from personal devices and any forensic questions about how data is left, so you don't over- or under-commit in the ESI protocol.

Scenario B: Opposing Counsel Sends an ESI Protocol

The facts. You're handed a proposed ESI protocol and asked whether to sign. It requires native production of all files, point-in-time versions of every hyperlinked document, and a 30-day production deadline.

How to reason: Run each term through feasibility before agreeing. Native-everything conflicts with your need to redact — flag it. An issue with point-in-time hyperlink versions is exactly the *Carvana* problem (see Part 2); it can be extraordinarily expensive and may not be technically feasible for your client's systems. The deadline may be unrealistic given the collection and review volume. The move is not to reject cooperation — it's to counter with what you can execute, propose a hybrid format, carve out hyperlinks pending a feasibility check, and make sure a Rule 502(d) order is in the protocol. If you can't execute it, don't agree to it.

Scenario C: A Privileged Document Went Out the Door

The facts. A week after a production, a reviewer realizes a clearly privileged memo was produced.

How to reason: First, check whether you have a Rule 502(d) order or a clawback agreement — if you do, the waiver risk is largely contained and you follow the clawback procedure to demand its return. Then notify opposing counsel promptly per the protocol, document what happened, and figure out how it slipped QC so it doesn't recur. This scenario is the entire argument for negotiating 502(d) protection at the outset: with that protection, this is an annoyance; without it, the disclosure can be a subject-matter waiver.

Part 5 | Ethics, Supervision, and When to Escalate

The Ethical Duties that Touch Discovery

Competence. Covered in Part 1: Technological competence is a required basic skill, and if you lack it you must learn it, associate with someone who has it, or decline the matter.

Candor to the tribunal. You cannot create a misimpression or let the court form an incorrect understanding by omission. This has teeth in discovery: You may not simply repeat a client's assertion that data is inaccessible or that preservation was handled. You must make a reasonable inquiry and form an independent view before representing it to the court or opposing counsel. Willful blindness is not a defense. The duty to preserve is owed to the court, not just to the other side.

Confidentiality and privilege. Attorney-client privilege covers confidential communications made to give or get legal advice, and it extends to agents of the lawyer (including the vendors and contract reviewers working under your direction). Protecting it is why privilege review and 502(d) orders exist.

Supervision of nonlawyers and vendors. You will supervise eDiscovery vendors, forensic collectors, and contract reviewers. You are responsible for making reasonable efforts to ensure their work complies with your professional obligations. "The vendor did it" does not insulate you. Engage competent providers, give them clear direction, and validate their output.

When to Raise Your Hand

You are not expected to solve everything yourself. You are expected to know the edge of your competence and escalate before it becomes a problem. Ask for help the moment you cannot confidently answer any of these:

- Do I understand where the relevant data lives?
- Can I explain how it was preserved and collected?
- Can I defend this process to a judge?

Common escalation triggers — flag these to the supervising lawyer or your eDiscovery team without hesitation:

- Ephemeral or encrypted messaging (Signal, WhatsApp, disappearing messages)
- Any self-collection that hasn't been validated
- Complex, proprietary, or unfamiliar data systems
- Cross-border data or privacy constraints (the EU's General Data Protection Regulation (GDPR) and foreign blocking statutes can override your discovery instincts)
- Any TAR or GenAI workflow you couldn't explain to a judge
- ESI protocol commitments you're not sure your client can meet

What Works Versus What Goes Wrong

What Goes Wrong	What Works
Late preservation	Early, targeted preservation
Blind reliance on the client	Verified, defensible collection
Overpromising in ESI protocols	Realistic agreements you can execute
Unvalidated technology workflows	Validated workflows, human and machine
Weak or absent QC	Strong QC and consistency
Inability to explain your decisions	Clear, explainable decision-making

The throughline: Credibility with the court is built through defensible decisions. Every phase above comes back to whether you can stand up and explain, reasonably and honestly, what you did and why.

Part 6 | Glossary

Term	What It Means
Adverse inference	A jury instruction, imposed as a sanction, allowing the fact-finder to assume that lost or destroyed evidence would have been unfavorable to the party that lost it. One of the more severe spoliation remedies.
CAL (Continuous Active Learning)	A form of TAR 2.0 in which the software keeps learning from every reviewer decision and continuously re-ranks the remaining documents, rather than training once on a fixed seed set.
Chain of custody	The documented, unbroken record of who handled a piece of evidence, when, and what they did with it — used to prove data wasn't altered. Central to defensible collection.
Clawback agreement	An agreement (ideally backed by a court order) letting a producing party retrieve inadvertently produced privileged material without waiving privilege.
Culling	Reducing a data set before review by filtering out material — by date range, file type, keyword, or system files — to cut volume and cost.
Custodian	A person who holds or controls potentially relevant data. "Custodial" data is tied to an individual; "non-custodial" data lives in shared systems like databases or shared drives.
Data map	An inventory of an organization's systems and where different kinds of data live. The foundation of the identification phase.
De-duplication (de-duping)	Removing duplicate copies of the same document so reviewers see each item once. Can be global (across all custodians) or custodial (per person).
De-NISTing	Removing known system and program files (identified against the National Institute of Standards and Technology (NIST) reference library) that are never substantive evidence — icons, executables, and the like.
EDRM	Electronic Discovery Reference Model — the industry-standard conceptual map of the eDiscovery lifecycle, from information governance through presentation.
Ephemeral messaging	Apps and settings where messages auto-delete (Signal, some WhatsApp and Teams configurations). A preservation minefield — flag immediately.
ESI	Electronically stored information — any information in digital form that may be discoverable.
ESI protocol	A written agreement between parties governing how ESI will be preserved, searched, formatted, and produced. Negotiated at or after the Rule 26(f) conference.
Forensic collection	A bit-for-bit, forensically sound copy of a data source that preserves everything, including deleted-but-recoverable data. The highest-integrity, and, depending on the data source, usually the highest-cost method.

Term	What It Means
Generative AI (GenAI)	Large language model tools that generate natural-language output — summaries, issue-spotting, suggested calls. In review, they assist judgment; they don't replace it.
Hash value	A unique digital fingerprint of a file. If a file changes at all, its hash changes — which is how you prove a collected file wasn't altered.
Information governance	How an organization manages, retains, and disposes of its data in the ordinary course, before any litigation. Good governance makes discovery cheaper and cleaner.
Legal hold (litigation hold)	A directive suspending normal deletion and instructing custodians to preserve potentially relevant material once litigation is reasonably anticipated.
Load file	A file that tells a review or production platform how to load a set of documents — linking images, text, metadata, and family relationships together.
Meet and confer	The Rule 26(f) conference where parties plan discovery, including ESI issues. The strategic leverage point of the case.
Metadata	Data about data — the information a system automatically records about a file (author, dates, sender/recipients). Essential for timelines and authentication.
Native format	A file in its original application format (e.g., a working .xlsx), with functionality intact — as opposed to a static image of it.
Non-custodial source	Relevant data not tied to an individual person — shared drives, databases, HR/finance systems, application data.
Precision	In TAR/search validation: of the documents the process flagged as relevant, what fraction actually were. High precision means little noise.
Predictive coding	An early name for TAR — using machine learning trained on human coding decisions to rank documents by likely relevance.
Privilege log	A list of documents withheld or redacted on privilege grounds, describing each enough to justify the claim without revealing the privileged content.
Proportionality	The Rule 26(b)(1) principle that discovery must be proportional to the needs of the case, weighing burden and expense against likely benefit. Your main tool for pushing back on overbroad demands.
Recall	In TAR/search validation: of all the relevant documents that exist, what fraction the process found. High recall means little was missed.
Redaction	Removing privileged or protected content from a document before production. Must be burned into the produced file, not merely hidden on screen.

Term	What It Means
Relevant	Information that could help prove or disprove a claim, defense, or fact at issue in the case. Distinct from "responsive."
Responsive	A document that falls within the scope of a discovery request and must be produced (if not privileged). Distinct from merely "relevant."
Rule 26(b)(1)	The federal rule defining the scope of discovery: relevant and proportional to the needs of the case.
Rule 26(f)	The federal rule requiring the early meet and confer and discovery plan, including ESI.
Rule 26(g)	The federal rule requiring counsel to certify, after reasonable inquiry, that discovery responses are complete and correct. The certification standard that applies to every review method.
Rule 34	The federal rule governing requests for production, including the right to specify the form in which ESI is produced.
Rule 37(e)	The federal rule governing sanctions for failure to preserve ESI.
Rule 502(d) (FRE)	The evidence rule letting a court order that producing privileged material doesn't waive privilege — the key protection for large productions.
Sedona Principles	Influential, widely cited best-practice guidance on eDiscovery. Key concepts are Principle 1: cooperation is expected and Principle 6: the responding party controls its own methodology.
Self-collection	Custodians or clients copying their own data. The cheapest and riskiest option; requires lawyer supervision and validation.
Spoliation	The loss, destruction, or material alteration of evidence a party had a duty to preserve. The most common trigger for discovery sanctions.
Structured data	Data organized in a database (rows, columns, fields) — as opposed to "unstructured" documents and email. Requires special handling to produce meaningfully.
TAR (Technology-Assisted Review)	Using machine learning to rank or classify documents by relevance based on human coding decisions. TAR 1.0 trains once; TAR 2.0/CAL learns continuously.
TIFF	A common image format for productions — a picture of a document page, carrying no native functionality; usually paired with extracted text and metadata in a load file.

Part 7 | Where to Go Next

When you need to go deeper than this guide, these are the reputable, current sources practitioners use. Bookmark a few.

Standards and Foundational References

- **ACEDS** (aceds.org) — the certifying body for the Certified E-Discovery Specialist (CEDS) credential; includes training, chapters, and resources like the Checklist Manifesto and CEDS materials that this guide draws on.
- **EDRM** (edrm.net) — the reference model itself, plus frameworks and standards for each phase.
- **Legal Data Intelligence** (legaldataintelligence.org) — emerging frameworks for data and AI in legal practice.
- **The Sedona Conference** (thesedonaconference.org) — the Sedona Principles and detailed commentaries; the intellectual backbone of the field.

News, Commentary, and Staying Current

- **Complex Discovery** (complexdiscovery.com) — digital publication and research organization covering cybersecurity, data privacy, regulatory compliance, and eDiscovery.
- **Craig Ball** (craigball.net) — a leading authority on forensics and technical eDiscovery; deep, practical, and readable.
- **eDiscovery Case of the Week** (podcast) — short, current case briefings in audio form.
- **eDiscovery Today** (ediscoverytoday.com) — daily roundup of cases and developments in the industry.
- **JD Supra** (jdsupra.com) — aggregated law-firm client alerts, useful for tracking how firms read new decisions.
- **Legaltech News (a Law.com publication)** (law.com/legaltechnews) — general legal industry news and trend coverage focused on legal technology.

Records, Governance, and Vendor Resources

- **ARMA International** (arma.org) and **InfoGov World** (infogovworld.com) — records and information-governance standards.
- **Relativity blog** (relativity.com/blog) — vendor-published but substantive material on review platforms and TAR.

This guide is a training and orientation reference. It is not legal advice and does not substitute for the specific rules of your forum or the direction of the supervising attorney on your matter.

About Level Legal

Level Legal is a Chambers-ranked provider of forensic, eDiscovery, managed review, and consulting services. We work with clients seeking a disciplined approach to eDiscovery. We build evidence-first, defensible workflows with human accountability at the center, and deliver delight through clarity, reliability, and high-touch partnership.

Built by and designed for litigators, we approach eDiscovery as part of the advocacy story, not a technical workflow to be offloaded. Our focus is on judgment, defensibility, and transparency so legal teams can move quickly without compromising credibility.

With decades of experience handling eDiscovery-intensive disputes, internal investigations, and multi-party litigation, we know what AmLaw teams and Fortune 500 clients expect: A partner who is steady, accountable, and aligned with their high expectations.

We're Here to Help

Have questions about an upcoming project or any of the topics covered in this guide?
We love to help solve complex eDiscovery problems.

Matt Mahon
Vice President of Client Solutions
mmahon@levellegal.com
+1 949-466-2662

Appendix 1

Sample FRE 502(d) order

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

- x

:

:

: **RULE 502(d) ORDER**

:

:

- x

ANDREW J. PECK, United States Magistrate Judge:

1. The production of privileged or work-product protected documents, electronically stored information ("ESI") or information, whether inadvertent or otherwise, is not a waiver of the privilege or protection from discovery in this case or in any other federal or state proceeding. This Order shall be interpreted to provide the maximum protection allowed by Federal Rule of Evidence 502(d).

2. Nothing contained herein is intended to or shall serve to limit a party's right to conduct a review of documents, ESI or information (including metadata) for relevance, responsiveness and/or segregation of privileged and/or protected information before production.

SO ORDERED.

Dated: New York, New York
[DATE]

Andrew J. Peck
United States Magistrate Judge

Copies **by ECF** to: All Counsel
Judge _____

Appendix 2

List of Metadata Fields to Request for ESI

List of Metadata Fields to Request for ESI

This list includes the basic metadata fields to request in discovery. Consult with your provider about what you'll want for your review tool.

What You Need

Document Specific Metadata Fields

- [ ] TO -- the main recipient(s) of an email message
- [ ] FROM -- the sender of an email message
- [ ] CC -- the recipient(s) of carbon copies of the email message
- [ ] BC -- the recipient(s) of blind carbon copies of the email message
- [ ] SUBJECT -- the subject of an email message
- [ ] AUTHOR -- the creator of the electronic file
- [ ] FILE_NAME -- the name of the electronic file
- [ ] PRODBEG -- For Bates labeled productions, the first production number of the email, attachment or electronic file
- [ ] PRODEND -- For Bates labeled productions, the last production number of the email. attachment or electronic file.
- [ ] PRODEND_ATT -- For Bates labeled productions, the last production number of the last child document in a document family
- [ ] PRODBEG_ATT -- For Bates labeled productions, the first production number of the parent document in a document family
- [ ] CUSTODIAN -- Custodian to whom the data is associated (John Doe, Social media, party name)
- [ ] DATE_CREATED -- the date on which an electronic file was created
- [ ] DATE_MOD -- the date on which an electronic file was modified

- [ ] DATE_SENT – the date on which an email was sent
- [ ] RECORDTYPE – Type of Electronic Record (email, doc, social media, spreadsheet, image)
- [ ] FILE_EXT -- the extension for the type of file (e.g. .doc, .xls, .pdf)
- [ ] FILE_SIZE -- the size of the electronic file
- [ ] CONFIDENTIALITY – Confidentiality designation based on Protective Order in matter (e.g. Confidential, Attorneys Eyes Only)
- [ ] NATIVEFILEPATH -- Path to the native file for the document (e.g. Natives\EFXAA00000001.xlsx)
- [ ] TEXTPATH – Path to the extracted text for the document, for mapping when loading data (e.g. Text\EFXAA00000001.txt)
- [ ] ALLCUSTODIANS -- All custodians who possessed a copy of the document (including the name in the Custodian field) from whose file the item would have been produced if it had not been de-duplicated (e.g. John Doe, Jane Doe, Peter Smith, Michael Jones)
- [ ] TIME_SENT -- the time at which an email was sent
- [ ] MD5HASH – Unique identifier, similar to a fingerprint, extracted from all files (e.g. D564668821C34200FF3E32C9BFDCCC80)
- [ ] PAGECOUNT – The total number of pages in the document

Redaction Metadata Fields

- [ ] REDACTIONBASIS – The basis for the redaction including Non-Responsive Business Information (NRBI) or privilege
- [ ] REDACTIONFLAG – Denotes when document contains redactions (Yes/No)

Social Media Metadata Fields

- [ ] TITLE – Title of Social Media Post
- [ ] URL – URL for Social Media Posts (e.g. https://twitter.com/user/status/1142098094210859008)

- [ ] SMACCOUNT -- The author of the Social Media Post (e.g. @johndoe)