The Ultimate Guide to eDiscovery

A PRIMER ON GATHERING ELECTRONIC EVIDENCE
Contents

Pages 3-28  
**CHAPTER 1**  
An Introduction to Electronic Discovery

Pages 29-53  
**CHAPTER TWO**  
Data Preservation and Legal Holds

Pages 54-64  
**CHAPTER THREE**  
Information Governance

Pages 65-82  
**CHAPTER FOUR**  
Early Case Assessment

Pages 83-98  
**CHAPTER FIVE**  
Document Review

Pages 99-114  
**CHAPTER SIX**  
Predictive Coding

Pages 115-131  
**CHAPTER SEVEN**  
Discovery Software
CHAPTER 1

An Introduction to Electronic Discovery

This chapter describes and defines electronic discovery, or “eDiscovery,” for anyone involved in the processes of litigation or investigations. It should provide an introduction and overview of the basic concepts and terms you will need to begin any conversation around the practice of electronic discovery. Following chapters will dive into more nuanced and specific issues within the process of managing electronic evidence in litigation.
The purpose of this guide is to describe the fundamentals of eDiscovery. If you are involved in modern litigation or investigations, you will almost certainly encounter eDiscovery. However, we should note that while the legal profession now recognizes eDiscovery as a distinct practice area, the word is really an artificial construct.

eDiscovery is really just another way to describe discovery, the process of obtaining and exchanging evidence, or information that might potentially become evidence, in litigation. But as the world has moved almost exclusively to electronic communications, some legal professionals began adding the “e” to make it clear that electronic records were now involved. So at a basic level, eDiscovery describes the process of discovery, updated to address the challenges and complications of collecting, reviewing, and producing evidence in the modern, digital world.
Discovery is the legal process governing the right to obtain and the obligation to produce non-privileged matter relevant to any party’s claims or defenses in litigation. In other words, it is the legal procedure by which parties are required to exchange information and evidence with one another in state and federal courts. While we primarily describe discovery in the context of litigation, the process is also essential to investigations, arbitrations and, generally, other forms of dispute resolution whereby parties must gather facts.

Most discovery still comes in the form of testimony or recorded interrogations. Discovery can also involve physical items, like medical exams or a defective product. But increasingly, discovery is focused on electronically stored information (ESI), which is why lawyers often use the term eDiscovery to distinguish the discovery of electronic records from other forms of discovery.
In years past, parties exchanged paper documents—often hundreds and hundreds of boxes of them. Over time, paper documents have been largely replaced with computer generated content, and the process of discovery was forever fundamentally changed. ESI can now be email, social media, cell phone data, digital audio or video recordings, global databases, apps, global positioning data, data stored in a household appliance, onboard computers in a car, or any of the thousands of digital records produced by an average person on an average day.
WHY IS EDISCOVERY SO HARD?

Until recently, discovery was a relatively pro forma affair, involving written requests for production in which a party to a lawsuit asked another party or a third party to furnish information. To do this, they described the documents or records of interest or particular types of information they needed to support the claims of a case. The party responding to the request was expected to locate responsive and produce the evidence or copies of the evidence.

The responding party could withhold or redact items containing privileged information, such as confidential communications between lawyer and client, but was obliged to furnish a log describing what had been withheld. The court served as a referee, assessing when a request was unduly burdensome or compelling production when responses proved insufficient.

But with the mass adoption of personal computing and the internet in the 1990’s, the processes and practices that had worked well for paper documents began to break down.
Today, what once was ink on paper are now pixels on screens, or are made up of database records which are incomprehensible masses of numbers and symbols except to trained computer experts.

Another important fact to understand is that the United States has a long tradition of broad discovery, which allows litigants to request a wide range of information and document types for almost any matter. Attorneys in other countries are likely shocked and horrified at the amount of information discoverable in U.S. courts, as many countries have much more restrictive rules and regulations regarding the discovery of personal or private information.

To help you understand the scope of the problem, consider the volume of evidence now available for litigation. For example, in one recent patent dispute, Samsung collected and processed about 3.6 terabytes of data, or 11,108,653 documents in a case against Apple Computers. The cost to process that evidence during a 20-month period was over $13 million dollars.

Of course, this case is an extreme example, but eDiscovery is a fact in matters of all sizes. In fact, digital evidence is playing an increasingly important role in even family law and criminal cases. As we all live increasingly digital lives, we are all leaving a daily trail of electronic evidence that often tracks our every move. And it is all available for discovery.
Understanding eDiscovery begins with understanding the Federal Rules of Civil Procedure. Don’t worry- we’re not going to make you take a Civil Procedure class. However, it is important to understand that the Federal Rules of Civil Procedure describe and define the discovery of electronic information. In fact, the rules have been updated twice already to address the specific concerns of electronic discovery.

At the federal level, the Federal Rules of Civil Procedure (FRCP) govern the procedure for civil lawsuits in United States district courts. At the state level, each state has their own set of statutes and rules that govern the procedure for civil lawsuits in state court. However, we mainly refer to the FRCP, which most jurisdictions model their own rules after.

In December 2006, significant amendments to the FRCP shaped the landscape for how eDiscovery works today. Among other changes, the 2006 amendments redefined discoverable material; encouraged early attention to issues relating to eDiscovery; introduced the concept of “reasonably accessible”; provided a procedure for asserting claims of privilege and
work product after production; and provide a mechanism for “safe harbor”
limits on sanctions related to loss of ESI as a result of routine operation of
computer systems.

Arguably, the most important 2006 amendment was simply to include the
words “electronically stored information” on the list of information that is
discoverable throughout the discovery process. Specifically, FRCP Rule
34 dictated disclosure and discovery related to “producing documents,
electronically stored information, and tangible things or entering onto
land, for inspection and other purposes.” The phrase “electronically stored
information” is broad enough to cover all current types of computer-based
information and intended to be flexible enough to encompass future changes
and technological developments.

The Rules were amended again in December of 2015. These more recent
changes were less sweeping but no less important. The amendments were
aimed primarily at addressing a few outstanding issues not adequately
addressed in the 2006 changes.

The rule change that got the most attention was Rule 37. The amended Rule
37(e) allows sanctions for failure to preserve ESI, but limits sanctions for
failure to preserve so that negligence, even gross negligence, will not be
sufficient for imposition of most severe penalties.
Perhaps the most sweeping change was to FRCP 26(b)(1) regarding proportionality and the scope of discovery. Rule 26(b)(1) was rewritten to limit discovery to that which is “proportional to the needs of the case” and provided five factors for courts to consider. This Amendment redefined the scope of discovery so that parties must address concerns about whether the amount of discovery is reasonably necessary to resolve the case fairly.
CHAPTER 1
An Introduction to Electronic Discovery

EDISCOVERY IN HISTORY

Many people may claim not to have heard of eDiscovery, but in fact, eDiscovery has played a starring role in many of the most famous legal battles of the past few decades.

Lawyers have known for almost 30 years that digital documents—especially e-mail—are a key to winning many types of cases. eDiscovery is the art and science of collecting, preserving, reviewing, and presenting digital evidence for litigation. Though it may seem like an exotic practice area, eDiscovery has a long and important history.

eDiscovery’s history arguably begins in February of 1989, when Oliver North stood trial on twelve counts related to lying to Congress about his role in the Iran-Contra Affair. One of the most damning pieces of evidence that led to North’s trial was his own email. Specifically, emails that North had deleted from his computers at the National Security Council. The email server in the White House kept archives of all sent and received email and the deleted emails became evidence for the committee investigating the Iran-Contra affair and subsequent trial.

eDiscovery again played a starring role in the 1998 Microsoft monopoly trial. In that case, Bill Gates became whiny, insolent, and defensive on the stand as his own e-mails were read back to him. In these e-mails, recovered from Microsoft’s email servers, Bill Gates asked his subordinates to think of creative ways to sabotage the company’s rivals.
eDiscovery has played a starring role in the world of sports as well. In 2015, New England Patriot’s quarterback Tom Brady instructed his assistant to destroy the cellphone that he had been using. Brady’s act was judged to be willful obstruction of justice and, “a deliberate effort to ensure that investigators would never have access to information that he had been asked to produce,” leading to a four-game suspension for the star.

More recently, former Secretary of State Hillary Clinton’s legal team deleted emails from her personal email server, most of which were later recovered and reviewed by Congressional investigators. The handling of the Clinton’s emails and controversy around the review process became a central point of contention in the 2016 Presidential Election.
Rule 26(f): Setting the Ground Rules
The actual negotiation will begin with what is known as a 26(f) Conference, which happens before any discovery can occur. The courts have made it clear these conferences should happen as early as possible and parties should agree on foundational principles like the forms of production. If parties can’t agree within two weeks, they face the judge for what is known as the Rule 16 Conference. Once an actual discovery request is issued, the responding party may object, under Rules 26(c) and 37(a).

Rule 26(g): A Reasonable Inquiry
The Federal Rules of Civil Procedure mandates a standard of care, including “that to the best of the person’s knowledge, information, and belief formed after a reasonable inquiry: with respect to a disclosure, [the response] is complete and correct as of the time it is made.” However, “reasonable” is a matter for the court to decide on the totality of the circumstances.

If parties fail to cooperate and judges or magistrate judges become involved, parties lose much of their freedom to negotiate. Also note Rule 26(b)(3), which says discovery of opposing counsel’s work-product is admissible
only upon showing the party seeking it has substantial need and is unable to obtain equivalent materials by other means.

**FRCP Rule 26(d)(2): There’s More Than One Way to Do eDiscovery**

Don’t let opposing counsel or judges impose a set pattern on the process. Just because one approach has worked for them in the past, doesn’t mean it’s always the right way to proceed. In addition, don’t let the other party’s timeline get in the way of what you need to do. The rules clearly state that, “methods of discovery may be used in any sequence,” and “discovery by one party does not require any other party to delay its discovery.”
Rule 34(b): Get Data How You Want It

Rule 34(b) allows the requesting party to decide how it wants information to be produced and lets the responding party object if impractical. Note that if the requesting party fails to specify the form for producing data, the producing party has the option to either produce the information in a form in which it is ordinarily maintained, or in an electronically search-able form. Courts have rebuked parties that produce data in printed or other non-native formats. (When in doubt, get the native format. It retains potentially useful metadata and is usually easier to access.)

Rule 37: Still a Safe Harbor?

Rule 37(f) used to be known as the “safe harbor” provision. However, few legal teams used this provision. The amended Rule 37(e) allows sanctions for failure to preserve ESI, but limits sanctions for failure to preserve so that negligence, even gross negligence, will not be sufficient for imposition of most severe penalties. Under the amended rule, courts may only sanction parties if one party is found to have “intent to deprive another party” of information and if that information cannot be recovered or produced from another source. If electronically stored information that should have been preserved in the anticipation or conduct of litigation is lost because a party failed to take reasonable steps to preserve it, and it cannot be restored or replaced through additional discovery, the court may impose sanctions.
Rule 26(b)(1): Keep it in Proportion

The old Rule 26(b)(2)(C)(iii) was clear that a court could limit discovery when the burden outweighed benefit. However, new Rule 26(b)(1), implemented by the December 1, 2015 amendments, takes the factors in these old requirements and puts them at the heart of any discussion about the scope of discovery.
In 2005, two consultants, George Socha and Tom Gelbmann, created the Electronic Discovery Reference Model. The name is clunky, and the model is by no means perfect, but it is the best and most commonly accepted description of the eDiscovery process. Not all litigation will follow all of the steps described (and new steps have been added over the years), but it remains a useful guide.

The EDRM consists of nine stages, which we will describe in more detail throughout this book. The process begins with information governance, identification, preservation, and collection. The data management functions include processing, review, analysis, production, and presentation.

Information Governance
Information governance is a more recent addition to the EDRM. In recent year, large organizations have begun looking for ways to reduce eDiscovery costs before litigation happens, which means managing ESI from its initial creation through its final disposition.
CHAPTER 1
An Introduction to Electronic Discovery

Identification
Locating potential sources of ESI, the volume of data that might be discoverable, the custodians and locations of discoverable evidence. The key is not only identifying the evidence but addressing the potential scope and technical issues of the project at hand.

Preservation
Parties must ensure that ESI that is discoverable for litigation is not altered or destroyed. ESI is often deleted in the course of routine business, but when potentially discoverable information is deleted, that can be considered spoliation, a sanctionable offense in some cases.

Collection
Collecting ESI is a significant challenge. Data must be collected in a forensically sound manner so that evidence is not altered or changed.

Processing
In order to review evidence in a forensically secure manner, ESI is often converted to forms more suitable for review and analysis, often an image file. The original, native document is preserved as well for more detailed, forensic analysis.
Review
The heart of the process. Attorneys must review documents and evidence for relevant information while protecting privileged information from being accidentally produced to opposing counsel.

Analysis
Attorneys must review ESI for content and context, identifying key custodians, subjects, patterns, and discussions.

Production
Delivering ESI to others in appropriate forms. Parties still often produce evidence on hard drives or disks, although electronic production is also employed.

Presentation
Once ESI has been reviewed for relevance, a few key pieces or passages may actually be presented at a deposition, hearing, or trial. Evidence is presented to help witness testimony, demonstrate key facts, or persuade a jury or audience.
KEY CASES TO KNOW

Zubulake v. UBS Warburg L.L.C. (“Zubulake V”)
U.S. District Court, Southern District of New York
May 2004

Zubulake is the mother of eDiscovery case law. This was not the first case to mandate parties make a reasonable inquiry to identify likely custodians or sources of information. However, Judge Shira Scheindlin wrote a series of long, detailed opinions with detailed balancing tests that were so thoughtful, complete, and persuasive that the issue was no longer possible to ignore. The defendants initially argued that recovering and reviewing ESI from backup tapes would be too expensive. Judge Scheindlin’s seven-factor test compelled the defendants to produce the evidence, although the costs of recovery and review of the emails were shared by both parties.

Victor Stanley Inc. v. Creative Pipe Inc.
U.S. District Court, District of Maryland
May 2008

Victor Stanley sued Creative Pipe for copyright and patent infringement over the design of an end frame for a park bench. U.S. Magistrate Judge Paul Grimm ruled Creative Pipe had waived attorney-client privilege
when it inadvertently produced several electronic documents. Grimm found Creative Pipe waived the privilege because it had failed to work out a privilege search protocol with the opposition, it had declined to use a “clawback” agreement, and it had not proven that its search process for privileged documents was reasonable. As such, Grimm deemed the documents to have been voluntarily disclosed. Victor Stanley prevailed at trial in September 2011, winning more than $2 million in damages and more than $1 million in a monetary sanction against Creative Pipe for destroying electronic evidence. The verdict was upheld on appeal in February 2013.

**Mancia v. Mayflower Textile Services Co.**

**U.S. District Court, District of Maryland**

**October 2008**

Once again, Grimm made his mark in eDiscovery in this wage-and-hour case brought by several Mayflower employees. Grimm held that the failure of opposing counsel to cooperate and work out disputes on their own was the biggest reason why eDiscovery costs were skyrocketing. He strongly advised counsel from both sides to work together and make the eDiscovery phase move more smoothly. The parties settled in November 2009. Many judges have followed Grimm by advising attorneys before them to cooperate and hold regular conferences before even setting foot into court. The proposed eDiscovery amendments to the Federal Rules of Civil Procedure also contain a duty to cooperate.
The Pension Committee of the University of Montreal Pension Plan, et al. v. Banc of America Securities
U.S. District Court, Southern District of New York
January 2010
Judge Scheindlin dealt with the question of spoliation of electronic evidence and found herself reiterating many of the points she had made in Zubulake. (She even called the case “Zubulake Revisited: Six Years Later.”) Investors in the pension plan had filed suit to recover lost funds. Several investors were accused of failing to preserve evidence, and Scheindlin focused heavily on the duties of counsel. She ruled that attorneys were obligated to issue written litigation holds in order to preserve electronic and paper records, and that failure to do so would be considered gross negligence. Scheindlin also reinforced the idea that sanctions should be mild, but that parties behaving recklessly, negligently and knowingly could be sanctioned, depending on the level of harm experienced by the other party. In this case, Scheindlin handed down monetary sanctions and the dreaded adverse inference. The parties eventually settled out of court.

Rimkus Consulting Group Inc. v. Cammarata
U.S. District Court, Southern District of Texas
February 2010
One month after Pension Committee, U.S. District Judge Lee Rosenthal refused to follow Scheindlin’s jurisprudence in a case where a consulting
group was trying to enforce a noncompete agreement with a group of former employees. Rimkus accused its ex-employees of intentionally deleting relevant emails and asked for sanctions. Rosenthal ruled sanctions were only appropriate in instances like this where a party behaves in bad faith. Rosenthal’s opinion also highlighted a major split among federal circuits as to when sanctions were appropriate—a split that could only be resolved by the U.S. Supreme Court or the proposed new federal rules. In June 2010, the case was resolved when Rosenthal imposed a permanent injunction preventing the ex-employees from using certain facts and methods they had obtained while working at Rimkus.

**Da Silva Moore v. Publicis Groupe & MSL Group**  
**U.S. District Court, Southern District of New York**  
**February 2012**  
U.S. Magistrate Judge Andrew Peck approved protocols to allow for both sides to use predictive coding in conducting eDiscovery. The decision—believed to be the first to endorse the use of technology-assisted review—clearly states that “while some lawyers still consider manual review to be the ‘gold standard,’ that is a myth, as statistics clearly show that computerized searches are at least as accurate, if not more so, than manual review.” However, the case, which remains open, got bogged down as attorneys representing Da Silva Moore filed a petition for Peck to recuse himself from
the case, accusing him of being biased as a result of his previous public comments in favor of using predictive coding. The appellate courts backed Peck, and the U.S. Supreme Court declined to intervene in October 2013.

**Arthur Andersen, LLP v. United States**

**United States Supreme Court**

**May 2005**

In the conclusion to the famous criminal action against the Arthur Andersen accounting firm stemming from the collapse of Enron, the Supreme Court overturned Arthur Andersen’s conviction for criminal obstruction of justice for shredding documents prior to receiving a subpoena from the SEC. The Court determined that the trial court’s instruction to the jury omitted the essential element of scienter—actual knowledge of a proceeding (in contrast to the “reasonable anticipation of litigation” standard in civil actions) and the intent to obstruct that proceeding.

**Coleman (Parent) Holdings, Inc. v. Morgan Stanley & Co.,**

**Florida Circuit Court**

**March 2005**

In a lawsuit alleging accounting fraud and misrepresentation in the sale of stock, the plaintiff filed a motion for sanctions, including an adverse
inference jury instruction for the defendant’s destruction of e-mails. The defendant had a practice of overwriting e-mails after twelve months, although it was required by the SEC to retain e-mails for two years. The court had ordered the defendant to review backup tapes, conduct searches, produce e-mails and a privilege log, and certify compliance with discovery obligations. The defendant certified discovery as complete despite having failed to review more than 1,400 backup tapes. In its order on March 23, 2005, the court revoked the pro hoc vice license of the defendant’s trial lawyer and disqualified the law firm, forcing the defendant to seek substitute counsel two weeks before trial. Sanctions in the case were ultimately reversed, but the threat of punishment in this suit scared many lawyers straight when it comes to eDiscovery.
TERMS TO KNOW

As you move through this book, there will be some terminology that is specific to the eDiscovery practice. Here are some of the essential terms to know:

**Custodian:**
The individual recognized to have created or controlled an electronic file.

**De-Duplication/DeNISTing:**
Techniques for removing duplicate files from a document collection.

**Forensic Image:**
An electronic or digital format for capturing and storing data without corruption or alteration.

**Keyword Search:**
The most common approach for searching document collections including keywords and Boolean strings.

**Load file:**
The file used to import data (coded, captured or extracted data from processing) into a database; or the file used to link specific files.

**Metadata:**
Data about data; hidden from direct view, including information such as, author, recipient, creation date, modified date, and other potentially important data.

**Native File:**
A file in its original file format that has not been converted to a digital image or other file format such as TIFF, JPEG, or PDF.

**OCR Text:**
Optical Character Recognition is software that scans paper and creates searchable text.

**Processing:**
Data must be narrowed down, converted, and prepared for analysis and relevance review. Data must be imported into a software platform for analysis and production, which is where volumes can expand greatly.
TERMS TO KNOW

Production:
Data can be produced to opposing parties in a number of formats, including images like TIFF, file formats like PDF, or native formats. Images are often easy to manage and Bates stamp, but do not retain metadata. Determine which format works best for your needs.

Personal Storage Table (PST):
A common file format used to store messages, calendar events, and other items within Microsoft software.

Predictive Coding or Machine Learning:
Refers to a process, not a search technology. Machine learning makes it possible for computers to assist in the relevancy review process by recognizing responsive documents in eDiscovery.
CHAPTER 2

Data Preservation and Legal Hold

This chapter will help you understand the technology environment that must be navigated in the context of preserving information potentially relevant to litigation, investigations or other disputes, as well as the steps practitioners must take to ensure that data maintains its evidentiary integrity.
WHAT IS DATA PRESERVATION?

Data preservation is defined as the processes that must occur to ensure that information potentially relevant to anticipated, pending or active litigation, investigations or other legal disputes retains its evidentiary integrity. In U.S. courts, legal precedent requires that potentially relevant information must be preserved at the instant a party “reasonably anticipates” litigation or another type of formal dispute. The event or occurrence that precipitates the party to begin preserving information is referred to as the “trigger” or “triggering event.”

The goal of preserving information is to ensure that the information’s evidentiary integrity is maintained for potential use in the case. This means that the information should not be altered from its original form or the form in which it exists at the time of the triggering event. Not only does this apply to the appearance of the information, such as the way a document looks, but to the hidden metadata as well. The slightest modification -- such as opening an email that is potentially relevant -- can alter the information’s metadata, and thus compromise its evidentiary integrity and potentially draw penalties.
Data preservation is defined as the processes that must occur to ensure that information potentially relevant to anticipated, pending or active litigation, investigations or other legal disputes retains its evidentiary integrity. In U.S. courts, legal precedent requires that potentially relevant information must be preserved at the instant a party “reasonably anticipates” litigation or another type of formal dispute. The event or occurrence that precipitates the party to begin preserving information is referred to as the “trigger” or “triggering event.”

The goal of preserving information is to ensure that the information’s evidentiary integrity is maintained for potential use in the case. This means that the information should not be altered from its original form or the form in which it exists at the time of the triggering event. Not only does this apply to the appearance of the information, such as the way a document looks, but to the hidden metadata as well. The slightest modification -- such as opening an email that is potentially relevant -- can alter the information’s metadata, and thus compromise its evidentiary integrity and potentially draw penalties.
Generally, the first step to preserving data involves the application of a legal hold. Immediately after a triggering event, it is also imperative that the party take steps to suspend activities that could potentially alter information that must be preserved, including those that occur in the normal course of business. For example, companies should halt deletion of email archives according to automated retention schedules.

Below, we will dive more thoroughly into perhaps the most important step of preserving information: the legal hold.
WHAT IS A LITIGATION HOLD?

A legal hold, also known as a litigation hold, is an essential early step in the eDiscovery process, and crucial to showing defensible and good faith efforts to preserve evidence. It is the mechanism by which parties that must preserve information potentially relevant to a dispute tell “custodians” of that data it must be preserved, and ensure compliance until the obligation no longer applies. The initial document outlining the scope of the preservation requirement is called the legal hold notification or notice.

The legal hold notice is where the rubber hits the road: the point at which all parties are aware of pending litigation and the need to preserve ESI for that litigation. It is also the point at which you need to have identified the potential sources of information, the custodians of that information, and the technical and practical challenges to be faced in the coming dispute.
LEGAL HOLDS IN PRACTICE

Legal holds are a relatively recent phenomenon. In *Zubulake v. UBS Warburg*, which we discussed in Chapter 1, Judge Shira Scheindlin made it clear that parties have an obligation to preserve relevant information as soon as litigation is imminent. But when is litigation imminent? And what is potentially relevant information?

A litigation hold notice is designed to answer those questions and eliminate confusion. The best definition for legal hold is that it is a way to communicate to a party or custodian that a) litigation is imminent, or reasonably anticipated, b) potentially relevant information must be preserved, and c) that electronically stored information (ESI) and paper documents that may be relevant to this pending litigation must not be deleted or altered.

Unfortunately, this third part of the litigation hold process is often the most problematic, either because parties continue with the routine destruction of data — such as deleting emails after a set period of time — or, carelessly spoil data because they don’t know better, or, less often, through deliberate efforts to thwart justice.
But as Judge Scheindlin wrote in *Zubulake*, “…once a party reasonably anticipates litigation, it must suspend its routine document retention/destruction policy and put in place a ‘litigation hold’ to ensure the preservation of relevant documents.”

If the requesting party fails to offer a prompt, detailed, and understandable litigation hold notice to its adversary, it may not receive the information necessary to conduct litigation. On the other hand, when a producing party fails to take a litigation hold letter seriously, *sanctions, adverse inferences or even worse penalties may be imposed*. Ultimately, the party that possesses potentially relevant ESI is under an obligation to preserve that information when litigation becomes reasonably anticipated, regardless of whether its adversary has alerted it of its preservation obligations.
DEFINING THE LEGAL HOLD PROCESS IN EDISCOVERY

A legal hold notice is usually sent to an organization and distributed to legal counsel, custodians, and IT staff. It is also common for an organization that anticipates litigation to draft and circulate its own legal hold. The notice must be broad enough to ensure that all potentially relevant information is preserved, but narrow enough so as not to be overburdensome for the party that possesses the information.

Unfortunately, too many lawyers rely on checklists or boilerplate language when issuing a legal hold. However, an effective hold notice can only be drafted if a legal team takes the time to consider the relevant issues of a case, the potential sources of evidence, and the questions that must be answered via discovery.
DRAFTING A LEGAL HOLD NOTICE

How do lawyers avoid sending formulaic, boilerplate notices? A litigation hold must be a thoughtful document tailored to the needs of the case. Certainly, there are elements of the legal hold process that can be repurposed over and over again, but there must also be an attention to detail in each new case as well. It’s a balancing act: minimizing cost, complexity and business disruption while ensuring that all potentially discoverable information related to the matter at hand is preserved. A sample notice can be found here.
It is important that once the initial notice is issued, reminders are issued early and often to ensure compliance. Cloud-based discovery automation software, like Logikcull, automates notices and reminders, and keeps track of which custodians confirm receipt of the notice and when.

Key custodians must receive email and messaging hold notices, and IT and HR must receive notices to preserve specific machines, devices, and backend systems. While issuing a legal hold notice that demands backup drives and media be preserved will help ensure you can recover evidence even if the original records are lost, it is important to be aware that courts have said that, often, producing legacy media such as tapes can be overly burdensome to the producing party. Parties should seek further counsel if they are unsure of how far their preservation efforts should extend or where original sources of evidence reside. At a minimum, a sound legal hold notice will instruct parties to suspend data destruction efforts if appropriate so that evidence will be easier to find later.
HOW DO I MAKE MY LEGAL HOLD DEFENSIBLE?

In the context of litigation holds, defensibility means understanding how your opposition, judge, jury, or regulator will view your legal hold process in light of applicable legal requirements. While perfection is not required, your actions will need to be reasonable given the information available at the time of the action.

In other words, acquire as much information as possible once litigation is anticipated and build it into a legal hold notice. As the process continues and you learn more about the opposing party’s processes, data collections and/or the information it seeks, update and refresh the notice accordingly.

Though a litigation can apply to nearly everyone in an organization, it is a good idea to limit and refine its scope. Interview people identified as key players and ask them who else is likely to have potentially relevant ESI. You should also interview both the personnel who manage the content, and those who manage the electronic implementation environment or network infrastructure.
HOW DO YOU WORK WITH DOCUMENT RETENTION POLICIES?

As noted above, email systems can be set to automatically delete messages after a certain elapsed time according to the organization’s email retention policy. When preserving and collecting data from a company’s communication systems (email, phone, messaging, etc.), consider the company’s ESI retention policy and whether it must be suspended.

Two groups to consider interviewing and involving are the users of the data and the IT department. The IT department may have to modify their practices in order to comply with the company’s preservation obligation. So be sure to interview users and custodians who can help identify data you need, but also the IT group, which can help you to preserve and collect it.
Courts have criticized the practice of allowing individuals who are stakeholders in the outcome of the case to identify, preserve or collect their own data, since they arguably lack the necessary objectivity to collect all of what is relevant and responsive. This practice of “self-collection” is often described as the “fox guarding the hen house.” Moreover, custodians simply cannot be expected to identify and preserve all the ESI that could be relevant to a case.

Judge Scheindlin touched upon the risk of relying solely on custodial holds in her decision in the Nat’l Day Laborer Org. Network v. U.S. Immigration & Customs Enforcement Agency (S.D.N.Y. July 13, 2012). Judge Scheindlin made it clear that if custodial holds fail, it is the party and/or counsel’s fault for not enforcing the hold notice, not the custodian who fails to preserve evidence. So it is important for legal teams to ensure that the legal hold is received and that its instructions are followed.
Legal teams must take the following steps to help ensure a sound legal hold.

1. Build a defensible, current and relevant list of the people and departments who will be receiving a notice. Distributing the notice to overly broad groups (e.g., the entire organization) can lead to non-compliance or misunderstanding since many employees may ignore a notice.

2. Once a notice is distributed, require recipients to confirm receipt and understanding of the message. This can be a written email confirmation or legal software that allows the recipient to fill out an acknowledgement that they understand and will comply with the message.

3. Send regular reminders to recipients of legal holds reiterating that the hold and its instructions are still in place, and that potentially relevant information should not be altered.
WHERE WILL YOU FIND ESI?

It’s always a good idea to instruct the recipient of a legal hold that the data to be preserved may be located in a wide variety of storage locations such as: email, desktop drives, portables devices, shared drives, home computers, tablets, smartphones, internet storage locations, and document management systems.

The most common source of ESI is a custodian. However, non-custodian ESI is a critical source of ESI as well. Examples of custodian data are: email, personal storage on hardware devices or cloud accounts, allocated storage, private data storage, data associated with social networking sites used by the custodian, tablets, smartphones or even private web-based email accounts.

Examples of non-custodian data include: databases, cloud storage databases hosted by third parties, and shared network storage locations.

Below we will go into more depth on each of these potential sources of evidence.
Email

Users of corporate computers will have email stored on one or more email servers. These servers may be physical hardware managed by IT staff or virtual machines leased from a cloud provider, either running mail server software, most likely applications like Microsoft Exchange. A third potential source is a Software as a Service (SaaS) offering from a cloud provider, which are ubiquitous. Webmail may be as simple as a single user’s Gmail account or, like the Microsoft Office 365 product, a complete replication of an enterprise email environment.

Users often have a different, but overlapping complement of email stored on desktops, laptops and handheld devices they’ve regularly used. On desktops and laptops, email is found locally (on the user’s hard drive) in container files with the file extensions .pst and .ost for Microsoft Outlook users or .nsf for Lotus Notes users, for example. Finally, each user may be expected to have a substantial volume of archived email spread across several on- and offline sources.

Network Shares

Apart from email, custodians generate content in the form of productivity documents like Microsoft Word documents, Excel spreadsheets, PowerPoint presentations and the like. These may be stored locally, i.e., in a folder on the C: or D: drive of the user’s computer. But more often, corporate custodians
store work product in an area reserved to them on a network file server and mapped to a drive letter on the user’s local machine. The user sees a lettered drive indistinguishable from a local drive, except that all data resides on the server, where it can be regularly backed up. This is called the user’s network share or file share.

**Mobile Devices: Phones, Tablets, IoT**

According to the U.S. Center for Disease Control, more than 41 percent of American households have no landline phone and instead rely solely on wireless service. For those between the ages of 25 and 29, two-thirds are wireless-only. A recent survey (sponsored by Facebook) found that four out of five people start using their smartphones within 15 minutes of waking up and, for most, it’s the very first thing they do.

The Apple App Store supplies over 1.5 million apps accounting for more than 100 billion downloads. All of them push, pull or store some data, and many of them surely contain data relevant to litigation. More people access the internet via phones than all other devices combined. The bottom line is if you’re not including the data on phones and tablets, you’re surely missing relevant, unique and often highly probative information.

Relatedly, the so-called Internet of Things (IoT) -- a vast, growing network of interconnected devices, applications, appliances and objects connected to
the internet -- poses challenges for legal practitioners, who increasingly must glean data from it for discovery purposes.

Local Storage
Organizations deploy network shares to ensure work is backed up routinely. Nevertheless, despite the best efforts of companies to try to keep all work in a single location, users will store data on local, physical media, including local laptop and desktop hard drives, external hard drives, thumb drives and other devices.

Social Networks
Given the prevalence and public nature of social media, today it is possible for anyone to draw a detailed profile and gain intimate personal information of an individual in just minutes. Needless to say, social media has become increasingly important in an evidentiary context.

It is a rich source of information for litigators in many types of matters, and can provide crucial context to employee investigations and other types of company disputes.

Recently, advisory notes to the updated Federal Rules of Civil Procedure put a fine point on this, stating, “It is important that counsel become familiar with their clients’ information systems and digital data — including social media — to address these issues.”
Databases (server, local and cloud)

From Microsoft Access databases on desktop machines to enterprise databases running multinational operations (think UPS or Amazon.com), databases of many kinds are embedded in company systems. Other databases are leased or subscribed to from third-parties via the cloud (like Salesforce.com or Westlaw). Databases hold so-called structured data, meaning it is unreadable outside of the database in which it is created.

The Cloud

Corporate applications and IT infrastructure increasingly operate in cloud environments like Amazon Web Services and Microsoft Azure while individuals increasingly store data in tools like Box, Dropbox, Google Drive, Microsoft OneDrive, Apple’s iCloud and others. The cloud must be considered alone as adjunct to the other sources when seeking to identify and preserve potentially responsive ESI.

Related reading: Debunking the Myths of Cloud with Esteban Kolsky

Now that we’ve addressed some of the most common types of information that must be preserved, we’ll revisit the legal hold itself, the individuals you should consider involving when it must be deployed, and considerations for maintaining the integrity of data that must be preserved and collected.
WHO CAN HELP WITH LEGAL HOLDS?

In order to create a reliable and consistent legal holds process, you will need reliable human guides. Some custodians may know a lot about the data itself, such as fields in a database, what information may be relevant to the matter, and even how to create valuable reports. Find people in the IT department who understand the relationship of the data to the business. These people understand the mechanics and technology of the system and know where important information may be stored. However, they are not experts on the content of the system.
How do you make sure people don’t ignore your legal hold notice?

It is important to perform regular follow-up interviews and send reminders to custodians and IT personnel subject to the hold. Remember, responding to litigation is not a usual function of most people’s jobs — they are busy and will not necessarily prioritize preservation duties unless they are regularly reminded. Case law requires litigating attorneys to take affirmative action in monitoring the hold. Just as importantly, when a hold is no longer necessary, make sure to release the custodians from their obligations. This lets companies return to their normal document retention schedule, and allows individual parties to use the device or data without fear of spoliation.
Courts are increasingly sensitive to the costs of eDiscovery and the concept of proportionality, which should be taken into account when assessing the scope of the collection. In most cases, the use of software will aid in validating the collection of ESI. Failure to use commonly accepted methods and technologies may expose the client to additional risk.

**WHAT IF SOMETHING HAS BEEN DELETED?**

Collecting ESI can be accomplished by the client, custodians, or a third-party gathering ESI for further use in the eDiscovery process.

It is important to understand that it is possible to recover evidence that may have been deleted as well. In most computer systems, deleted files still live in the unallocated space on a hard drive, which is where many applications store temporary files while the application is in use. However, preserving data in unallocated space requires forensic collection using special software.
For most data, there are two types of metadata, system and file metadata, and each have different jobs. System metadata is data involving the architecture of a computer system, and offers information like timestamps of when file was last modified, accessed, and created as well as where the file is physically located in the volume.

File metadata is data about a file and is stored within the file. File metadata includes printed date and time and other user-supplied data that can be altered. It also includes timestamps, author information, and edits. Metadata will help you authenticate documents and verify important information when conducting eDiscovery. Unfortunately, many lawyers are unaware that metadata can be altered or corrupted if not properly handled.
WHAT IS A FORENSIC COLLECTION?

In order to collect data without corrupting it, legal teams may need to engage in a forensic collection. Forensic collections are most useful if you suspect someone has altered or deleted data and you need to investigate the metadata and unallocated space from a computer or device. A forensic collection makes a forensic copy of a hard drive that includes every bit of data on that drive, including data in unallocated space. This collection type often uses a “write-blocker” to prevent alteration of the content when a device is attached to retrieve the data.

In addition to using specialized tools, legal teams need to keep documentation of the decisions and actions made during the collection process. This includes answering whether or not a computer should be forensically collected. More often than not, collecting the active data and relevant network shares is sufficient.
DEFENDING YOUR PRESERVATION

If an opposing party or presiding judge questions the preservation methods employed, you will need to defend and show your work. Here are some important steps that should be taken along the way, regardless of whether you expect a challenge:

- Create documentation (the chain-of-custody log) or testimony by the ESI collector about the steps taken.

- Always consider and take into account the cost, accessibility, and needs of the case.

- Take steps to ensure that nothing about the data is altered or degraded.

- A forensically sound collection will preserve all potentially relevant metadata that may be used by the trial team in its claims.

- A collection by a third-party vendor or with appropriate Legal Intelligence software will often be the best method.
CHAPTER 3

Information Governance

Information governance pertains to any program or system designed to get your electronic house in order for the purposes of mitigating risk and cost, and making wiser business decisions. Our focus in this chapter is to summarize the eDiscovery implications of information governance. Information governance is not a litigation tactic, but a program that should be in place from the initial creation of ESI through its final disposition. While it sounds like a straightforward program, there are many complications that make it difficult to actually make an information governance plan work in the face of litigation.
WHAT IS INFORMATION GOVERNANCE?

It is a fact of modern life in our digital world that companies generally have a poor understanding of the nature, quantity and form of the electronically stored information (ESI) in their possession. Information governance is any set of policies, procedures, processes, and controls designed to manage an organization’s data in order to mitigate the risk of future regulatory, legal, and other risks associated with keeping data for long periods of time. More recently, information governance guidance has focused on how it can provide business intelligence such that corporate teams can be in a better position to make bottom line-driven decisions.

The Gartner consulting firm defines information governance as “the specification of decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, storage, use, archival and deletion of information. It includes the processes, roles, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals.”
In 2002, Arthur Andersen partner David Duncan pled guilty to destruction of records related to Enron, leading to a criminal indictment of his firm. In that case, Duncan led an effort to shred documents even after he had learned of a pending federal investigation and received warnings that destroying documents violated the firm’s policies.

Despite that abject lesson in the downside of data destruction in the face of litigation, spoliation continues to happen. If a company or organization fails to suspend an information governance program in the face of a litigation hold notice, it can be sanctioned or penalized in court.
An information governance regime is in part designed to set standards for when and how records and documents should be disposed of within the company. The policies are in place to designate when the organization no longer needs to be retained or no longer needs to be collected for compliance and regulatory purposes. When litigation arises, if the organization has faithfully deleted or destroyed data within the guidelines of the program, the organization in most cases cannot be forced to produce that data for discovery. This ability to purge records according to corporate policies set with respect to legal and compliance obligations is generally referred to as “safe harbor.”

This does not necessarily mean that all documents or records can be deleted when the retention period is over. In highly regulated industries such as banking, securities, and pharmaceuticals, there is often a need to retain information long after it has served its business purpose in order to comply with laws and industry-specific regulations.
In banking, for instance, some information will often need to be retained for 7 to 10 years. Life insurers need to retain policy documents and information for as much as 100 years. Business units may work with a Records Management department to determine how long different classes of information should be retained before deletion. Some companies are establishing information governance committees, drawing representation from departments such as risk management, compliance, legal, I.T. and various executives to oversee the process.

In some companies, information is retained indefinitely and never deleted. This increases costs and risks in the face of litigation, as more information is caught up in the discovery dragnet and must be collected, analyzed and reviewed.
HOW DO I HANDLE DEPARTING EMPLOYEES?

One of the biggest sources of risk and headaches is how to handle the records belonging to employees who have left a company. In order to make an IG program work, organizations must track departing employees to ensure ESI under litigation hold is not lost or taken when an employee leaves the company.

To begin, efforts must be taken to ensure that potentially relevant information is not improperly deleted or that the departing employee’s hard-drive is not re-purposed and given to another employee. Regular retention and deletion policies which govern ESI of employees must be suspended as soon as it is known that an employee has left the company. Any policies in place to preserve ESI of departed employees should be implemented immediately.
Federal Rule of Civil Procedure 26(a)(1)(A) demands that an organization be responsible for all data within its “care, custody or control.” Courts have often punished parties or counsel for eDiscovery failures when they fail to know what ESI they had and in what forms, what custodians held it, where they stored it and what risks of alteration or disposal affected the ESI.

Federal Rule of Civil Procedure 30(b)(6) allows a party to have a corporate employee give testimony about their legal hold and all the decisions around it. Having a well-documented hold will make preparing for such a deposition much easier and reduce the opposing party’s chances of finding the hold insufficient.
When a litigation hold is issued, an organization must respond and suspend deletions accordingly. In order to show that an organization has acknowledged and responded to a litigation hold order, it should be able to address these issues:

1. The legal hold notice and any updates to it

2. Evidence of distribution of the legal hold notice, including the list of recipients

3. Confirmations of receipt and follow-up communications

4. Notes of witness interviews

5. Copies of company documents, such as organizational charts

6. An up-to-date data map (see below)
Before you can preserve, review or produce ESI, you must first know what you have, where you have it, the forms it takes and how much of it you’ve got. The process by which a litigant and counsel build an inventory of potentially relevant ESI is called data mapping.

Data mapping is one of those eDiscovery buzz phrases — like technology-assisted review or information governance itself — that takes on any meaning the fertile minds of marketers can think up. But at its most basic, a data map is likely a list, table, spreadsheet, database or simple Excel spreadsheet. In fact, a “data map” might be better termed an “Information Inventory.” It’s very much like the inventories that retail merchants undertake to know what’s on their shelves by description, quantity, location and value.
As a rule of thumb, as the business value of information decreases, the relative costs and risks to maintain the information increases. Information governance and litigation readiness is also referred to as proactive eDiscovery because it encompasses the policies, procedures and steps an organization has in place to follow when and if a litigation event occurs.

There are a few questions that need to be asked about an organization’s ESI and its environment. These questions fall within the scope of information governance which addresses a need larger than litigation readiness. They are: What is it? Where is it? How safe and secure is it? Who controls it? Why do we need it? How long do we need it? What does it cost to keep it?

Making an IG program work in the digital world is not easy. Many businesses have adopted a BYOD (Bring Your Own Device) model when they allow employees to connect their personal phones and tablets to the corporate network. Many other companies don’t have a BYOD policy but their employees use personal devices for work purposes regardless. Securing the
ability to access these devices for eDiscovery requires employers obtaining consent in employment agreements, complicating the work of enforcing an information governance program. In order to make a plan work in the real world, consider the following steps:

1. Create a data map to account for business record locations and document which custodian/department is in control of that source.

2. Centralize email storage and do not allow users to archive emails on multiple machines.

3. Make sure deletion policies can be suspended if a litigation hold notice arrives.

4. Review legal holds and delete data that is no longer under a hold per your records retention policy.

5. Archive or delete legacy document review databases for closed matters so they are no longer under active storage. Ask outside counsel or vendors for a data certificate of destruction when a case is closed out.

6. Set-up rules and stick to them. If a policy says e-mails should be deleted after six months, they must be deleted after six months. It is impossible to defend an irregularly enforced program in court.
CHAPTER 4

Early Case Assessment

Early case assessment (ECA) is defined as the practice of estimating the risk to prosecute or defend a legal case.
WHAT IS EARLY CASE ASSESSMENT?

Organizations will often spend significant cost, time, and money on a case only to find they need to settle the case unfavorably after the cost or exposure becomes too burdensome. Much of that cost is directly related to eDiscovery. In a broader sense, ECA can be thought of as the process of gaining insights about the strength of legal positions, and potentially relevant issues, witnesses, custodians and evidence as soon as a legal matter surfaces. In some instances, it has been described as “conducting discovery prior to a formal discovery context.”
HOW DOES EARLY CASE ASSESSMENT WORK?

More than 90 percent of all cases settle prior to trial. In federal cases, this number is closer to 98 percent. Therefore, discovery has become a de facto form of dispute resolution in U.S. courts. That is why the early phases of discovery are arguably the most important of any legal battle or dispute. This is when you identify whether or not your side will likely be able to proceed with an action or if it will be more cost-effective to pursue a settlement or alternative resolution.

The initial phase of discovery is the time when you will have to determine the evidence that will need to be collected, the custodians who will be interviewed, and the cost, complexity and challenges your team will face. Early case assessment provides a foundation for creating overall case strategies. Considerations include identifying the trigger event of litigation, the facts of the case, value of the damages, the capability of your opponent, your judge’s sophistication in eDiscovery and history of imposition of sanctions, date ranges, keywords, case merit, risk analysis and many other considerations.
We will address the basics of early case assessment in this chapter, but first, let’s discuss two important recent trends which are changing the practice of ECA.
FROM LINEAR REVIEW TO ANALYTICS

Traditional linear review is giving way to an analytics-first workflow. That means legal teams can increasingly rely on software to understand case data early through analysis and intelligent culling, rather than old-fashioned manual review, where documents had to be reviewed one at a time.

For example, the use of visual analytic technology is helping lawyers find relevant information by illustrating relationships and communications patterns. Clustering technology helps group data into categories so that legal teams can identify where relevant information might be found. In Logikcull, culling filters that are akin to those used by Amazon.com are used to whittle down large amounts of data, so, for example, users can hone in on just emails to and from certain domains, or just files above a certain size. This is similar to how online shoppers search for products by identifying a size, a make, a brand and so forth, so they are not looking at thousands of available options one at a time, but a targeted subset instead.
In terms of visual analytics, Logikcull automatically provides “infographics” that create a high-level timeline showing when important communications have been created and sent.

In addition, analytic tools can identify how custodians are talking about their work, which words are important, and what unexpected terms might be useful search topics. This information can help create a simple word list from a collection of emails and documents that can be used to find smoking gun documents and evidence later.
The next factor that is changing how ECA is employed is changes to the Federal Rules of Civil Procedure on proportionality. As mentioned in Chapter 1, Federal Rule of Civil Procedure 26(b)(1) was rewritten to limit discovery to that which is “proportional to the needs of the case” and provided five factors for courts to consider.

The old Rule 26(b)(2)(C)(iii) was clear that a court could limit discovery when the burden outweighed benefit. However, new Rule 26(b)(1), implemented by the December 1, 2015 amendments, takes the factors in these old requirements and puts them at the heart of any discussion about the scope of discovery.

In addition, Rule 26(b)(1) eliminates the phrase “reasonably calculated to lead to the discovery of admissible evidence.” This means that requesting parties can no longer assume they can obtain discovery of virtually anything that’s “reasonably calculated” to be helpful in litigation. According to the rules, the
factors parties must address as they relate to proportionality include the:

- Importance of the issues at stake in the action,
- Amount in controversy
- Parties’ relative access to relevant information
- Parties’ resources
- Importance of the discovery in resolving the issues, and,
- Whether the burden or expense of the proposed discovery outweighs its likely benefit.

In sum, under the new rules, lawyers now have to provide more details and more information in order to make a claim that discovery is proportional to the needs of a case. This means lawyers must dispense with vague, broad claims or objections and rigorously vet proportionality factors in order to control the scope of discovery.
In practical terms, this may entail providing:

- Estimates outlining the proposed steps and the associated expenses with conducting certain aspects of discovery.
- Suggestions for alternative forms of production, if preserving evidence poses a significant burden.
- Itemized expenses, including time and cost to help bolster arguments and proposals.
CHOOSING THE RIGHT ECA TOOLS

Given the exploding volumes of information subject to discovery, ECA tools are essential for providing insight into the data. Technology allows legal teams and investigators to filter, search and perform different types of analysis prior to fully processing data.

Some ECA tools ingest native files and the metadata associated with the native files and thus do not require processing to permit analysis. Other “full service” solutions, like Logikcull, have the ECA function built into their core framework.

Choose technology that works with your data collections. For example, a project that contains numerous spreadsheets would likely drive a decision to use a review platform appropriate for reviewing spreadsheets, or a tool that can natively ingest and analyze all types of common files.

One important trend to be aware of is the rise of in-place ECA applications.
In traditional ECA, litigators must first collect all potentially relevant ESI at the outset of a legal matter and try to assess its potential relevance. This is disruptive to business operations and captures a lot of duplicative and useless data. In-place ECA applications assess the available data as it is on the network, leaving it in place during the initial review. This allows for more targeted and thoughtful collections if data is deemed relevant, and less business disruption.
DEFENSIBLE DELETION

Careful analysis and review will allow parties to identify and eliminate inessential data so that it doesn’t become a burden later on. Tactics for eliminating useless data, collectively termed “defensible deletion,” include:

- **Culling** - the process of removing low-value materials from higher value information using a variety of analytical tools.
- **De-duplication** - the process of removing exact copies of files, and sometimes near matches, from a particular data set.
- **De-NISTing** - the process of removing all so-called system files that are deemed to have no evidentiary value.

Of course, your team should always document and record the use of these technologies so that if opposing counsel questions why potential evidence was deleted, documentation exists which can support your actions. Legal Intelligence tools like Logikcull automatically create audit logs of all user actions that have been taken within the system so they are defensibly
Early data assessment is similar to ECA, and the two terms are often used interchangeably. Early data assessment allows you to know what your data looks like before you process it, and gives insights into the scope of the project and its costs.

Often, the most important part of this process is the selection and testing of keywords. Keywords are the words or phrases you will use to search data collections and identify potentially relevant documents, whether they are applied as basic searches or in a more predictive or analytical context.

In its most basic application, early data assessment may mean searching your available data collections to see if potentially relevant documents hit on selected keywords. If no documents are found, refine your search until better results are found. However, if the results continue to be disappointing, it may also indicate that there is no evidence to support the claims in a case, or that rudimentary keyword searches are insufficient to properly assess the available data.
Once accurate keywords have been developed and you have a bearing on the scope of the potentially relevant information, take the time to count your potential custodians and estimate the volume of evidence for each. This is also the best point to evaluate how much time and effort a discovery project will take, as time periods can be different for each custodian and source.
The early case assessment process is iterative, meaning actions must be repeated until your team feels it has a complete picture of the digital landscape. The total universe of potentially collectible ESI will usually be defined during the process of formulating the internal preservation directive/litigation hold. The universe usually consists of four main categories of data locations: Individual employee files, department/group files, enterprise databases, and, potentially, backup media. Not all data identified for preservation needs to be collected right away. Some data may never need to be collected. Collecting all data that has been preserved may unnecessarily inflate costs and overwhelm the legal team with irrelevant data. Once the timing of collection from a data location has been decided, the legal team must assess what level of forensic defensibility should be employed for the collection.
Important steps include:

- Taking notes of witness and IT interviews
- Conducting custodial surveys
- Retaining copies of documents reviewed to determine scope of legal hold (e.g., organizational charts)
- Drafting and tracking legal hold communications
- Identifying and tracking hold recipients, confirmations, and follow-ups
- Reporting on each step along the way
WHAT DOES MY POLICY SAY?

The ability to defensibly remove immaterial information from business and IT systems is largely dependent on corporate policy. Most organizations have record management policies which describe actions to be taken by the client to preserve data, especially in the case of a legal event. However, even the best record management policies fail if no one enforces or tracks the progress. For example, companies often fail to ensure that potentially relevant information is not improperly deleted or re-purposed when employees leave the organization.

Consult records management policies which determine how long different classes of information should be retained before deletion. In some cases, you can meet with information governance committees, risk management, compliance, legal, I.T. and various individuals who oversee data management.
When identifying potentially relevant ESI, use data maps or data surveys of IT systems, conduct periodic surveys of an organization's digital landscape and document who has access to which records and where the records reside. Data maps list the IT systems used by employees, onsite and offsite, and paper document storage locations, including backup media that may contain responsive documents or data.

When surveying the organization’s IT systems and data, it is essential to determine:

- Type of documents stored,
- Security and access controls,
- Format and reporting capabilities,
- Backup, archival, and purging processes and schedules, and
- Audit-trail capabilities.

But always remember that while organization charts and data maps are valuable resources, data maps can get out of date very quickly. Ensure a data map is verified with the IT staff you interview.
CHAPTER 5

Document Review

When you think about eDiscovery, you are probably thinking about review. Review is the phase of eDiscovery where your team will actually discover potentially relevant documents and evidence needed for a matter. A document review distinguishes relevant data from irrelevant data and uncovers information and evidence about the facts and issues of the matter. However, there are important pitfalls to address in review which, if not properly managed, can jeopardize your case, especially around protecting privileged material.
In today’s digital world, billions of emails, text messages, presentations, database files, electronic documents, and the like are sent and received every day, creating eDiscovery dilemmas for lawyers and judges. Any one of them could be the pivotal document that either builds a case or destroys one.

If there is one case that demonstrates the awesome power of eDiscovery, it is United States v. Microsoft, the infamous antitrust trial that began in the late 1990’s. A key contention was that Microsoft was conspiring against Sun Microsystems. Thanks to the review process, an internal email from Bill Gates surfaced. In it, he asked: “Do we have a clear plan on what we want Apple to do to undermine Sun?” It was a modern-day smoking gun.

A more recent case demonstrates the danger inherent in the review process. The Samsung v. Apple litigation pitted two tech giants against each other in an epic court battle between the world’s dominant smartphone makers. But the case also showed that even the biggest companies can stumble in the face of eDiscovery review.
During document review, Samsung’s outside counsel improperly redacted a sensitive Apple contract it had obtained during the discovery process, a contract that was under a protective order, limiting its use to the litigation alone. The associate then uploaded the contract to Samsung’s internal intranet, where it “went viral.” When Apple found out, sanctions soon followed.

Another example comes from Oracle v. Google, which pitted two more tech companies against each other in a high-stakes copyright dispute. In that case, Google’s famous search abilities didn’t seem to transfer to its document review process. During its privilege review, the company screened out an email that was labeled “Attorney Work Product.” But Google’s review did not catch nine drafts of the same email, which were then produced to Oracle. Worse yet, that email became a key piece of evidence against the company.

As you can see, document review is not a matter to be taken lightly. A well-executed review process can win a matter for your client. But even a well-intentioned review team can make mistakes that can sink a case or expose litigants to a negative inference or even incur sanctions.
WHAT IS DOCUMENT REVIEW?

Document review is the act of identifying responsive documents to produce and privileged documents to withhold from opposing counsel. Review is an iterative, learning process, meaning you will need to repeat and refine your work over and over again. Through this process, your team will uncover factual issues in a case and develop legal strategies to pursue based on information that is found in a collection of documents.

Due to the exploding volumes of evidence available for eDiscovery, your team should limit the scope of what needs to be reviewed whenever possible using targeted collection, filtering, and culling techniques prior to a manual review. Other advanced technologies can be used prior to or in combination with a human review to reduce the amount of documents the human reviewers must examine.
The review process is the most complex and expensive part of eDiscovery. In order to begin, many teams create a review guide to define the process and help reviewers make informed decisions. The guide should explain the facts in the case, the search terms being used, and the codes to be used to mark responsive, non-responsive, privileged, and other determinations. Consider how you will:

- Find out how information was collected—by a client, vendor or responding party—and determine how it is formatted and organized before attempting to review it.
- Document your search and review procedure. Provide details of the software used in review, how you allocated sets of data, and any use of predictive coding or TAR methods.
- Develop a quality control program and stick to it.
- Create consistent coding or tagging sets which fit the needs of the case. Coding screens typically contain data fields for: responsive or non-responsive, privileged, confidentiality levels, and “key” documents.
- Evaluate the search capabilities of the review database.
- Determine the final output formats and production to outside parties.
• Make sure that a privilege log can be generated for documents coded as privileged and supporting privilege assertion is captured.

• Ensure confidential and redacted documents are specially recognized and properly handled.
As discussed in Chapter 1, Federal Rule of Civil Procedure 26(f) requires parties to meet and confer about preserving discoverable information and to develop a proposed discovery plan addressing discovery of electronically stored information and the form or forms in which it should be produced.

According to Rule 26(f), eDiscovery topics covered must include preservation, the form of production, and privilege and work-product protection claims. Other important issues to be discussed include the scope of the review, the issues in dispute, the search protocols, and the mechanics of the production of materials between parties after review. Parties should also facilitate communication channels between team leaders in order to resolve issues quickly.
Important Issues to Consider

Start by determining what form your information will come in. Will the dataset in question be all ESI or mixture of ESI and paper? Converting paper to an electronic format can simplify the process, keeping the entire review in one collection.

Consider how the ESI will be searched. Parties often redact ESI in the way they once redacted paper documents, by blacking out text. To make that possible, ESI is converted to an image file, such as TIFF, in a process that makes it no longer searchable. Files you can’t search aren’t much help, however, so after redaction, searchability needs to be restored by using “optical character recognition.” OCR technology can extract the remaining text from the TIFF image file, restoring searchability to the non-redacted sections.

This TIFF-OCR method is most appropriate for documents that are largely text based. It is much less fitting for complex, dynamic documents, such as spreadsheets or databases. Consider the types of files likely to be reviewed before agreeing to a conversion method.
Similarly, parties should address Bates numbering early in the process. The Bates system is still the standard tool for tracking reviewed documents, but native files are not easily Bates numbered. However, Bates numbers can serve as filenames for native files. If you expect to handle a large number of native files, negotiate the Bates numbering conventions to be used at the Meet and Confer.

It is possible to begin privilege review before human eyes even see the first document. Most digital review systems can search and tag documents that have privileged names, email addresses, law firm domains or legal terms before reviewers examine those documents. Reviewers are aware that a document is potentially privileged and can review it to confirm it is in fact a privileged communication or attorney work product.
Perhaps the most important consideration in the review process is the search terms to be employed. As mentioned earlier, this process will be iterative. You will continually refine your search as the review process begins. After negotiating an initial list of search terms, you will begin reviewing the documents returned using these terms.

If too many irrelevant documents are being returned you are experiencing “noise,” also commonly referred to as “false positive” search results. If that is the case, search criteria may be refined, optimized, and tweaked to get the desired results. You will have to narrow search terms if too many irrelevant documents are found and broaden the search if too few relevant documents are found. This may also involve testing different search technologies.
SEARCH TIPS AND TRICKS

Use Misspellings: This may be counter-intuitive, but even in the age of spell check, people get words wrong. Include common misspellings in your list of search terms.

Look for code words and jargon: Consider terms that only the players in a matter might know. Use language real people use in their emails, not lawyerly terminology.

Learn Boolean logic: Boolean searches use connectors such as “AND”, “OR”, and “NOT” to provide more refined searches. Proximity searches can identify when specific words are used in the same sentence or paragraph.

Play Your Wild Cards: Symbols can replace one or more characters (i.e., e-disc*very)

Stemming: Reduces a word to its root form, as in “eating,” “eat,” and “eater.”

Stop Words Stop Pointless Searches: Stop words tell the search engine to ignore common words that will not help find important information, like the, and, or with.
CHECK YOUR WORK!

Quality control and due diligence should be performed at every stage of the review process to ensure consistent and accurate document designation. One important consideration is to allow the review team to tag documents for “further review” when unsure of coding. The review platform should also be able to apply quality control restrictions such as automatically identifying duplicate documents, or grouping families of documents such as threaded emails and attachments for easy identification.

In order to QC your work, include a second level review of all designated production documents (or a sample of them) by senior attorneys and randomly review the team’s coding to check for inconsistencies. Provide feedback to the reviewers as they work in order to improve quality and spot issues as they arise.
As we briefly highlighted above in the *Oracle v. Google* example, a failure to protect privileged documents can ruin a case. When privileged documents are produced to opposing counsel, privilege could be waived. In some cases, it is possible to demand the documents be returned and stricken from the record, but, as attorneys often say, you can’t unring a bell. That is, opposing counsel knows what the document says and will certainly use that knowledge to the best of their ability.

The work product doctrine protects records such as documents which were prepared by the client, the attorney, or an agent for either the client or the attorney. Attorney-client privilege pertains to communications by and between the attorney (or his/her agent) and a client that were made for the purpose of giving or receiving legal advice.
In order to review for privilege, search the review data set using a list of search terms that will likely identify privileged documents. Some possible search terms include:

- Names and internet domains of in-house and outside counsel, paralegals, and other legal staff.
- Email addresses for in-house and outside counsel, paralegals, and other legal staff.
- Terms such and legal jargon like deposition, subpoena, and interrogatory.

Be sure to provide all reviewers with a list of people and topics that may involve privilege. Parties should have a discovery agreement in place that allows inadvertently produced privileged documents to be returned. A Rule 502(d) order may also be sought, in order to protect against waiving privilege through inadvertent disclosure. Otherwise, the court may deem the privilege to be waived.
FORMS OF PRODUCTION

Remember that what you see on screen as an email or a document is actually a partial representation of all of the data and information in an electronic document. Meeting the obligation to produce data means choosing suitable forms of production which deliver the content without destroying the hidden and associated information in a file, such as its metadata.

Requesting parties often demand native production of reviewed documents, but native production is complicated and often involves a form of production that closely approximates the contents and usability of the source. When negotiating the production of reviewed materials to outside parties, determine if native production is the best approach or if there is a more appropriate form of production.

Important questions to consider include:

- Can the production maintain the integrity of header fields, change logs, and metadata? Will it support sorting and searching by these data?
- Does the chosen format preserve family relationships between messages and attachments?
- Will it facilitate redaction of privileged and confidential content and identification and sequencing with Bates numbering or a similar system?
The document review process is one of the most critical phases of an eDiscovery project. In order to properly manage this process, make sure to:

- Understand how information was collected.
- Document your search and review procedure.
- Cull and reduce your data collection to minimize review costs.
- Search, assess the results, and refine the search terms.
- Verify your results with quality control tests.
- Control attorney-client privilege.
CHAPTER 6

Predictive Coding

With the growing volumes of evidence and the expense of manual review, many legal teams are considering computer assisted review to help weed through digital records faster and with less human intervention. Predictive coding is the common name for this process which is emerging as an accepted practice in some extremely large cases.
The typical professional spends about a third of their work day reading, organizing, and responding to email. Those emails quickly add up, to more than 100 billion business emails sent every day. These messages don’t just crowd our inboxes—they may become subject to litigation. When that happens, emails and other ESI will need to be collected, reviewed, and, if responsive and unprivileged, produced. Reviewing even a tiny fraction of the emails sent every day can be, to put it mildly, a daunting task.

Getting through the ever-growing volumes of emails and other ESI is one of the main challenges in litigation today. To make this process more efficient, some lawyers and technologists are turning to emerging technologies like predictive coding and computer-assisted review, through which they hope to reduce the time and expense associated with eyes-on review.

But while predictive coding solves some problems, it introduces new complications. And despite a spate of recent cases employing the technology, questions remain as to how effective it can be.
WHAT IS PREDICTIVE CODING?

Predictive coding is the automation of document review. In other words, instead of manually reading every single document in a collection, document reviewers use computer-categorizing software that classifies documents according to how they match concepts in sample documents.

This typically works by taking information gained from manual coding and automating that logic to a larger group of documents. Reviewers use a set of documents to identify potentially responsive documents and then train the computer to identify similar ones. Technology is used to predict how certain documents would be coded, based on how they were coded manually. Hence, the name predictive coding.

But how do you know if your system’s predictions are accurate? To guide the process and measure effectiveness, these processes generally incorporates statistical models and/or sampling techniques alongside human review.
To avoid confusion, it is important to note that not all technology-assisted review, or TAR, involves predictive coding, although the terms are sometimes used interchangeably. Predictive coding is just one form of TAR, which is itself a broader category that encompasses many uses of technology in the document review process. It is also important to note that predictive coding does not replace culling or early case assessment in the review process. For example, some TAR methods, like culling irrelevant data from a document collection, are used before applying predictive coding technology.
WILL COMPUTERS REPLACE DOCUMENT REVIEW ATTORNEYS?

Since computers can analyze millions of documents in minutes, some imagine that they may displace many of the lawyers, paralegals, and other professionals that are traditionally needed for document review. Yet, even the most automated review workflows can’t completely replace human reviewers. They simply change their role in the process.

Even with TAR, the reviewer’s role is still critical and is at least as important as during a traditional linear review. Rather than reviewing each and every document, the reviewer is now responsible for coding documents within the sample set. Those documents, in turn, refine the computer’s understanding of either a specific issue or the concept of responsiveness. Reviewers are also responsible for validating the review result.
In the traditional review model, many reviewers who are typically distant from the actual facts of the case complete a rote, manual review. In the predictive coding model, fewer reviewers with expert understanding of a case do the initial coding, setting the standard for the review and managing its quality. The role is altered, but not eliminated.
After culling, most predictive coding workflows call for selecting and coding a sample from the collection. The purpose of the initial sample is to see the how prevalent responsive documents may be in the collection. The two types of samples are:

Control Set: A random sample of documents coded by human reviewers at the start of a search or review process that is separate from and independent of the training set. Control sets are used in to measure the effectiveness of the machine learning algorithm.

Training Set: A sample of documents coded by one or more subject matter experts as relevant or non-relevant, from which a machine learning algorithm then infers how to distinguish future documents.

In TAR that uses machine learning, the predictive algorithms are continuously updated, based on the judgement of the review team. This “continuous active learning” deemphasizes the role of control and training sets, allowing the predictive coding to change alongside human reviewers.
As with the manual review process, predictive coding workflows are iterative. The iterative process involves repeatedly updating the training set with additional examples of coded documents to improve results. It is critical to have a lawyer review all documents and processed data, including reviewing documents coded by the machine to verify quality. Attorneys should continually review the results and refine their search method in order to train the system.
CONCERNS ABOUT PREDICTIVE CODING

Because predictive coding relies on highly complex algorithms to determine responsiveness, rather than individual human judgement, some critics describe it as a “black box.” That is, information goes in and information comes out, but the internal processes are opaque.

How the technology work, and thus how coding decisions are made, is understood by only a few experts—and very few of them are lawyers. Whereas an associate in a traditional eyes-on review can explain his or her individual coding decisions, understanding why predictive coding treated a document a specific way takes highly specialized technological knowledge.

There is also debate about the level of scrutiny courts should apply to TAR claims. Magistrate Judge David J. Waxse, for example, has argued that courts must act as gatekeepers when evaluating the appropriateness of TAR. Evidence about TAR’s capabilities would have to meet the standards established in the Supreme Court’s *Daubert v. Merrell Dow Pharmaceuticals*. 
decision and Federal Rule of Evidence 702. On the other hand, Magistrate Judge Andrew J. Peck has said that those standards are “not applicable to how documents are searched for and found in discovery.”

Concerns over TAR, along with the complexity and cost of predictive coding products, may have contributed to the legal industry’s slow adoption to predictive coding review.
HOW DO I DEFEND THE USE OF PREDICTIVE CODING?

Because predictive coding is still a nascent practice, it may be necessary to defend any use of the technology in litigation. A defensible process will look something like this:

- Assembling a data sample.
- Coding the sample in a linear fashion.
- Using the sample to assemble a training set.
- Coding the training set.
- Running the algorithm against the training set and comparing the machine coding with the human coding.
- Repeating the process until the human reviewer is satisfied with the machine’s understanding of the coding criteria.
- Running the algorithm so that the coding is applied to the rest of the collection.
- Assessing the machine-generated coding.
- Manually reviewing the production set for a final quality check.
The first court case to embrace predictive coding as an allowable review strategy was 2012 *Da Silva Moore v. Publicis Groupe.* “What the bar should take away from this opinion is that computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review,” U.S. Magistrate Judge Andrew Peck of the Southern District of New York wrote at the time. “Counsel no longer have to worry about being the ‘first’ or ‘guinea pig’ for judicial acceptance of computer-assisted review.”

In the years since, a handful of subsequent opinions have continued to address the issue. Here are some of the most notable:

**In re Actos (Pioglitazone) Products Liability Litigation, MDL No. 6:11-md-2299 (W.D. La. July 27, 2012).**
This product liability action included a case management order allowing the parties to use a “search methodology proof of concept to evaluate the potential utility of advanced analytics as a document identification mechanism for the review and production” of ESI. The search protocol provided for the use of a TAR tool on the emails of four key custodians.
Global Aerospace, Inc. v. Landow Aviation, L.P., No. CL 61040
(Vir. Cir. Ct. Apr. 23, 2012).
Virginia Circuit Court Judge James H. Chamblin ruled that the defendants may use predictive coding for the purposes of processing and producing ESI.

U.S. District Court Judge Shira Scheindlin held that the federal government’s searches for responsive documents, requested pursuant to FOIA, were inadequate because of the government’s failure to properly employ modern search technologies. Judge Scheindlin urged the government to “learn to use twenty-first century technologies,” including predictive coding as opposed to simple keyword search.

U.S. District Court Judge Robert L. Miller Jr. ruled that defendants need not identify which of the documents, from among those they had already produced, were used in the training of the defendants’ TAR algorithm. However, the court said that it was troubled by the defendants' lack of cooperation.
Hyles v. NYC, No. 110-cv-03119 (S.D.N.Y. Aug. 1, 2016)

In this case, Judge Peck again addressed TAR, ruling that plaintiffs could not force the producing party to use TAR in its review. “It certainly is fair to say that I am a judicial advocate for the use of TAR in appropriate cases,” Judge Peck wrote, but he found no reason to compel its use in this case. “Responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information.”
TAR has had an impact across the pond as well. In the UK, courts have issued several opinions on predictive coding in litigation. One key decision in Merry Old England is Pyrrho Ltd. v MWB Property Investments, evaluating the suitability of TAR in document review.

**How to Speak TAR**

TAR involves complex algorithms, based on sophisticated mathematical and linguistic models. As in the law, the expert lingo can seem impenetrable to outsiders. Here are some of the common concepts and terms you might need to know to navigate this world.

**Algorithm:** A specified set of computations used to accomplish a particular goal. The algorithms used in eDiscovery are implemented through computer software.

**Artificial Intelligence:** A general term for computer programs that are designed to simulate human judgement. Artificial intelligence includes machine learning, which allows computers to change when exposed to new data, without needing manual programming.

**Boolean Search:** A search methodology that uses keywords to pull results through using connecting words like “and” or “or” to find specific combinations. In more complex litigation, more sophisticated Boolean strings are often used with a fuzzy search technique, designed to account for variations in spelling and word choice.

**Concept and Categorization Tool Search Systems:** These techniques rely on a thesaurus to capture documents that use different words to express the same thought.

**Clustering:** A grouping method in which documents are organized into categories so that those in one category are more similar to each other than to those in another category. Clustering is an automated process and the categories made may or may not be valuable for review.
HOW TO SPEAK TAR

**Fuzzy Search Models:** A method to refine a search beyond specific words, recognizing that words can have multiple forms. In fuzzy search models, even if search terms don’t use the exact words in a relevant document, the document might still be found.

**Natural Language Search:** A non-Boolean search method, whereby search commands are input as one would speak naturally. This is the type of search associated with search engines like Bing and Google.

**Probabilistic Search:** Search based on language models, including Bayesian belief networks, which make inferences about the relevance of documents based on how concepts are communicated in a collection.

**Subjective Coding:** The classification of documents based on subjective judgement about their responsiveness, privilege, or other categories.
CHAPTER 7

Discovery Software

The proliferation of PCs, laptops, smartphones, and even the Internet of Things have created exploding volumes of discoverable, electronic evidence. That growth in electronic documents threatens to break the discovery process. The only solution is to use equally disruptive technology to tame the challenges of modern discovery.
A SHORT HISTORY OF EDISCOVERY SOFTWARE

It may be hard to remember now, but discovery was once an analog affair. A large case would require mountains of cardboard banker’s boxes. Teams of contract attorneys would hole up in a conference room or stifling warehouse where they would read, one by one, every document, memo, receipt, accounting ledger, and record that was dumped on their desk.

In the late 1970’s and early 1980’s, the emergence of corporate computing began to slowly change this system. Volume wasn’t the problem yet—complexity was. No contract reviewer could be expected to make heads or tails of a computer punch card.

But then, in the 1980’s the desktop PC came into the law firm and lawyers realized they could use these machines to organize hundreds, or maybe thousands, of pieces of evidence.

In short, lawyers began to see the challenge and the opportunity afforded them by the desktop computer.

Indeed, the database technology that is still used by many litigators dates
back to the early 80’s. One such platform, CT Summation, was first thought up more than 35 years ago. Jon Sigerman, founder CT Summation, writes:

In 1982, David Rotman and I conceived the idea for what became CT Summation. At the time, David was a megafirm litigation partner and I was a solo practitioner. Really, we combined our ideas for applying personal computers to litigation support. I had a docket of smaller cases, some of which at any given time were hot, while others were not. My desire was to have a system that made it easy for me to winnow each case down to a critical mass of evidence by clipping and assimilating excerpts of the key portions of documents or deposition transcripts.

David on the other hand, wanted a platform on his and his paralegal’s desktop IBM PC’s to index and archive potentially important documents and transcripts from a case.

Another discovery platform, Concordance, is almost as old. Concordance was launched in 1984, as a system that could offer full-text information retrieval on a personal computer.

These once ground-breaking tools still remain in use by some firms today. After all, Summation and Concordance may be approaching middle age, but the problems they were designed to solve still exist. In 2017, as in 1982,
attorneys still need to narrow down mounds of evidence to the most critical documents and recall those documents as needed. But the technology was designed for managing documents, not strictly for eDiscovery.

By the turn of the millennium a few entrepreneurs and software vendors had recognized that the volume of evidence was a challenge. Reviewing it was a nearly impossible task with existing software.

To remedy the situation, software vendors began introducing new search tools designed to help make sense of it all. Not all of it worked. The most recognized member of this next generation is kCura's Relativity, a review platform designed to help ingest, process, and review large amounts of electronic evidence.

At the same time, eDiscovery vendors began to proliferate. Vendors, making use of platforms such as Relativity, could help guide a firm’s discovery process, from data ingestion to production—often for a hefty fee.

Technology from this era allowed parties to deal with growing electronic evidence, but often in a way that was highly complex and, in many cases, prohibitively expensive.

Years later, these tools are reaching the limits of what is needed for modern litigation.
Since the start of the Computer Age, discoverable data has grown at incredible rates. The amount of digital information that attorneys needed to review in 1984, or even 2004, pales in comparison to the volumes of data available today.

There’s a reason for this. Moore’s Law posits that computing power doubles every 18 months. Since it was coined by Intel founder Gordon Moore in 1965, it has remained a relatively accurate predictor of processing power growth. As computational resources increase, so too does the amount of data that is produced. Indeed, according to Ion Stoica, a computer sciences professor at the University of California, Berkeley, data growth even outpaces Moore’s Law.

It’s an endless feedback loop: More processing power means more data, in turn requiring more processing power, allowing for more data. Consider the backlog of unread emails in your Gmail account. Gmail’s inbox size has made it possible to hoard millions of emails, which has in turn resulted in ever greater electronic document creation.
That’s just the tip of the iceberg. There are many places that ESI can be stored within an organization, either on-premise, in a data center or cloud service. It may reside in unstructured systems or structured systems, and may exist as active data, archived, legacy, or deleted data. Almost all of this data is discoverable in litigation.

As data grows at ever-increasing rates, the tools used to process and review that data need to be able to keep up.
The evolution of eDiscovery technology has led to the cloud. Cloud-based services can scale, or grow to meet any sized challenge. Scalability allows a system to access resources needed to accommodate larger workloads or to free up resources when they are not needed.

But scale alone is not the solution. A true eDiscovery software solution must also be intelligent and flexible. In the cloud, there is enough computing power and flexibility so that legal teams can address their discovery challenges intelligently and strategically. This next stage in the evolution of discovery software uses process automation and “intelligent” data processing to manage data in the cloud.

Automating many of the labor and time-intensive processes typically associated with eDiscovery is now possible, from early case assessment through production. That means formerly slow or costly processes can be handled automatically and in a matter of minutes. The technology automates processes such as data ingestion, indexing, OCRing, virus scanning, and so on. It takes advantage of user-friendly interfaces, real-time collaboration, 24/7 accessibility, cost-predictability, and fast deployment via the cloud.
This automated, cloud-based process also allows for use cases well beyond traditional document review. In some organizations, compliance teams are using such systems to conduct internal investigations. General counsel’s offices are using it to move eDiscovery in-house, while finance departments utilize the technology to conduct tax audits. It’s not a matter of simply using the scalability of the cloud, but also a maturing of legal technology more generally, where user-friendly technology accessible by cloud and infinitely scalable can be deployed to address a wide range of different projects concurrently.
That law firms are bullseyes for cybercriminals should go without saying. Through litigation, investigations, M&A, and other delicate legal activities, law firms and other service providers gather reams of highly valuable data into one place. For hackers, breaching a law firm’s security means hitting the figurative jackpot.

There’s a common adage among cybersecurity professionals: data is at its most vulnerable when it’s in motion. Unfortunately, the traditional eDiscovery workflow is a process of motion. And it’s also a process of replication. Under the EDRM-driven model of eDiscovery, data is transferred, shared and copied repeatedly, often with no log of who touched it or where it resides. Really, it’s an ideal workflow—for hackers, that is.

The traditional eDiscovery process is perhaps the most at risk to that kind of intrusion. First, valuable data is collected by clients, often without confidential and sensitive information removed. That data is then transferred to the law firm and, from there, sent onwards to the firm’s tech team and/or external vendors. The information is loaded onto a review platform and reviewed for responsiveness and privilege, after which it’s produced to the
other side—sometimes through a secure online transfer, sometimes through hard drives sent through the mail. Every time that information is in transit, it’s put at risk.

Cloud-based discovery solutions can help protect against these risks. A secure, cloud-based platform that encrypts data in motion and at rest can provide important safeguards against hackers. When data is stored on the cloud, permission-based access can be used to make sure information is retrievable instantly, to those who should have it, without requiring endless reproductions. Productions can be shared with outside parties via a secured download link that will automatically expire after a specified time.

Finally, by hosting information in one centralized hub, you can entrust much of the security infrastructure to the experts, instead of having to build and maintain your own systems. All in all, secure, cloud-based discover means fewer opportunities for data to be put at risk.
The growth of eDiscovery is a major contributor to expanding litigation costs. One of the most expensive part of the discovery process is the attorney and paralegal hours required to review documents. But another major cost center is eDiscovery vendors.

eDiscovery pricing is notoriously opaque. Vendors often treat their pricing as trade secrets, leaving buyers in the dark as to how their cost structure compares to the rest of the market. As a result, many pricing schemes in the eDiscovery services industry are based not on the value provided, but the max clients are willing to pay. Hidden costs—$500 for rushing an upload here, extra fees for customer support there—can quickly push projects over budget. Meanwhile, the lack of market transparency leaves many clients unfamiliar with what’s a fair rate—and slow to recognize when they’re getting a bad deal.

eDiscovery pricing should be transparent and predictable. Unfortunately, that’s rarely the case. Here’s a quick overview of the hidden costs that can arise during discovery.
These are the old school eDiscovery vendors, the companies that use Relativity and similar tools to manage the discovery process. Hidden costs include:

- Data loading fees
- Database setup and training
- Advanced analytics fees
- Seat license fees
- Storage fees
- 3rd-party production fees
- Database restoration
- Native processing
- Project management
- Minimum invoicing

**Total cost: $300-$500 per gigabyte.**
VENDORS-IN-THE-CLOUD

The cloud is revolutionizing discovery, but cloud-based vendors still have many of the hidden costs of traditional vendors. These include:

- Per-project minimums
- Extra costs for admin seats
- Extra costs for “data reduction”
- Project management fees
- Downtime for manual ingestion
- Hosting overage fees
- Minimum invoicing

Total cost: $75-$350 per gigabyte.
You may think that legacy systems are “already paid for,” but maintaining aging infrastructure while putting up with outdated software means your costs can quickly add up. Costs such as:

- Server maintenance & security
- Storage infrastructure
- Slow time-to-review
- Write-offs for poor quality
- Inability to bill back to client
- Lawyers “going rogue”
- Operating system maintenance
- Downtime for upgrades,
- installation

**Total cost: $30-$75+ per gigabyte.**
With Logikcull’s Discovery Automation, there are no hidden costs. For $40 a gigabyte, you get a fully automated discovery solution with world-class support. That means:

- NO user limits
- NO matter limits
- NO production limits
- NO onboarding fees
- NO training fees
- NO in-app chat fees
- NO support fees
- NO technology-assisted review fees
- NO system maintenance
- NO minimum invoices

**Total cost: $40 per gigabyte per month.**
EDISCOVERY SOFTWARE CHECKLIST

When evaluating eDiscovery solutions, ask tough questions to ensure that the software will fit your demands for the long-term, whether it is ongoing litigation, or preparing your firm, company, or organization for the legal and regulatory challenges your digital data collections create. Here are some important questions to ask any potential technology partner for litigation:

- Is the workflow design and methodology flexible and adaptable to your needs?
- What processes are automated? When would you need to rely on a third-party or vendor?
- Will the vendor be around for years? Remember your case might be around for a long time.
- Is the application easy to use?
- Can projects be started quickly or will you be stuck waiting for someone else to process and ingest your data?
- Is the software defensible in court, if necessary, or is it a mysterious “blackbox” algorithm?
- Does the platform protect your data when in motion and at rest?
- Can the software scale easily to handle growing case volumes?
- How fast is the system?
• Is the workflow flexible enough to handle different case and matter types?

• What is the pricing structure? What is included? What are the hidden costs?
ABOUT LOGIKCULL

Logikcull.com is instant discovery for modern legal teams. Its secure, cloud-based solution helps law firms and organizations of all sizes solve the expensive, complex, and risky challenges associated with eDiscovery, internal investigations, and open records response. With Logikcull, you can start a discovery project in five seconds, from anywhere at any time on any device. Reviewing data is as easy as performing a Google search. And in Logikcull, your data is always secure. That’s why it’s trusted by the Fortune 500, Am Law 200, and governments of the biggest cities in the world. Founded in 2004 by CEO Andy Wilson and CTO Sheng Yang, Logikcull builds powerfully simple software that democratizes discovery.

Call 1-855-986-5444
Email hi@logikcull.com

Watch a demo at http://logikcull.com/demo