

THE MOST **EXPENSIVE** **DATA IS THE DATA YOU** **FORGOT YOU HAD**



Proactive information governance helps organizations reduce eDiscovery costs, minimize legal risk, and improve defensibility by understanding, organizing, and managing their data before litigation begins.

How Proactive Information Governance Transforms eDiscovery from a Cost Center into a Competitive Advantage

Four million paper documents. Hand-coded by a team of people. One record at a time. Over two years. For a cost that approached \$10 million before a single attorney ever touched the database.

That was the state of the art in 1993. And while no one would argue that legal discovery hasn't come a long way since the days of hand-stamping Bates numbers on every page, a surprising amount of the cost and risk baked into today's eDiscovery process stems from the same root problem: organizations don't know what data they have, where it lives, or how long they've been holding on to things they no longer need.

The tools have changed dramatically. The underlying challenge has not. And for companies that treat eDiscovery as a reactive problem something to scramble around only after litigation hits the cost of that mindset is very real.

This white paper makes the case for a different approach: one that starts left of the litigation event, with the data itself, and one in which a trusted eDiscovery partner helps organizations build the information governance foundation that makes every future matter faster, cheaper, and more defensible.

The Data Problem Nobody Talks About

The Explosion of Digital Information

More than 98% of documents created today never exist as paper. They are born digital, live in digital environments, and, unless actively managed, die digital, sitting in storage indefinitely, long after any practical reason to retain them has passed.

This was not always the case. A generation ago, the challenge of discovery was physical: how do you get paper into a system that attorneys can actually search? That problem required armies of coders, weeks of scanning, and millions of dollars in infrastructure just to stand up a single searchable database. Today, that same four million records could be processed in 72 hours and hosted online.

But the data volumes we now face would have been unimaginable to those earlier practitioners. And the types of data have multiplied far beyond email and documents. Every organization now generates a sprawling ecosystem of electronically stored information:

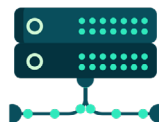
- **Email and documents** - still the foundation, but increasingly only a fraction
- **Text messages, SMS, and MMS** - now primary business communication channels
- **Over-the-top messaging platforms** - Slack, Teams, WhatsApp, Signal
- **Call center audio recordings** - transcribed, indexed, and legally discoverable under federal rules
- **Security camera footage and badge access logs** - increasingly captured in corporate investigations
- **Video and audio meeting recordings** - stored automatically across dozens of collaboration platforms
- **Cloud-based documents** - OneDrive, iCloud, Google Drive, often outside traditional IT oversight

Context is no longer simple. A thumbs-up emoji in a Slack thread carries different meaning depending on who sent it, to whom, and in what decade they were born. A one-line reply “OK by me” is either an innocuous sign-off on a birthday cake order or a green light on something far more damaging. The volume is massive. The nuance is critical. And most organizations have no map of where any of it actually lives.



98%

**OF DOCUMENTS
LIVE ENTIRELY
DIGITALLY**



2.5Q

**BYTES OF NEW
DATA CREATED
DAILY**



20%

**OF UNSTRUCTURED
DATA IS
MISMANAGED**



\$25B+

**GLOBAL
EDISCOVERY
MARKET BY 2029**

The Real Cost of Waiting

Why eDiscovery Shouldn't Start at the Triggering Event

Most companies approach eDiscovery reactively. A lawsuit is filed. A regulatory inquiry arrives. An internal investigation is triggered. And the scramble begins: collect everything, process everything, review everything.

The problem is that by the time litigation is on the table, the cost of that approach is already largely locked in. And the organizations that feel that cost most acutely are often those whose data practices have been quietly accumulating risk for years.

The Retention Problem

Storage is extraordinarily cheap. A four-terabyte external drive that once cost hundreds of dollars per gigabyte now costs roughly \$100. And because storage is cheap, organizations let data accumulate indefinitely, well past any legal, regulatory, or business reason to retain it.

The practical consequence: a company that thought it was facing a 100-gigabyte collection discovers, when litigation arrives, that it actually has terabytes sitting in archived servers, backup environments, and legacy systems. Data with no legal hold obligation. Data that should have been disposed of years ago under the organization's own retention policies. Data that is now, potentially, all discoverable.

THE IT CLOSET PROBLEM: In a documented real-world scenario, a company's official policy was to recycle backup tapes every 90 days. When deposed, the IT administrator revealed he had been pulling tapes from the drive, putting them in a closet, and replacing them with new ones for over ten years. A decade of potentially discoverable data, off the official record.

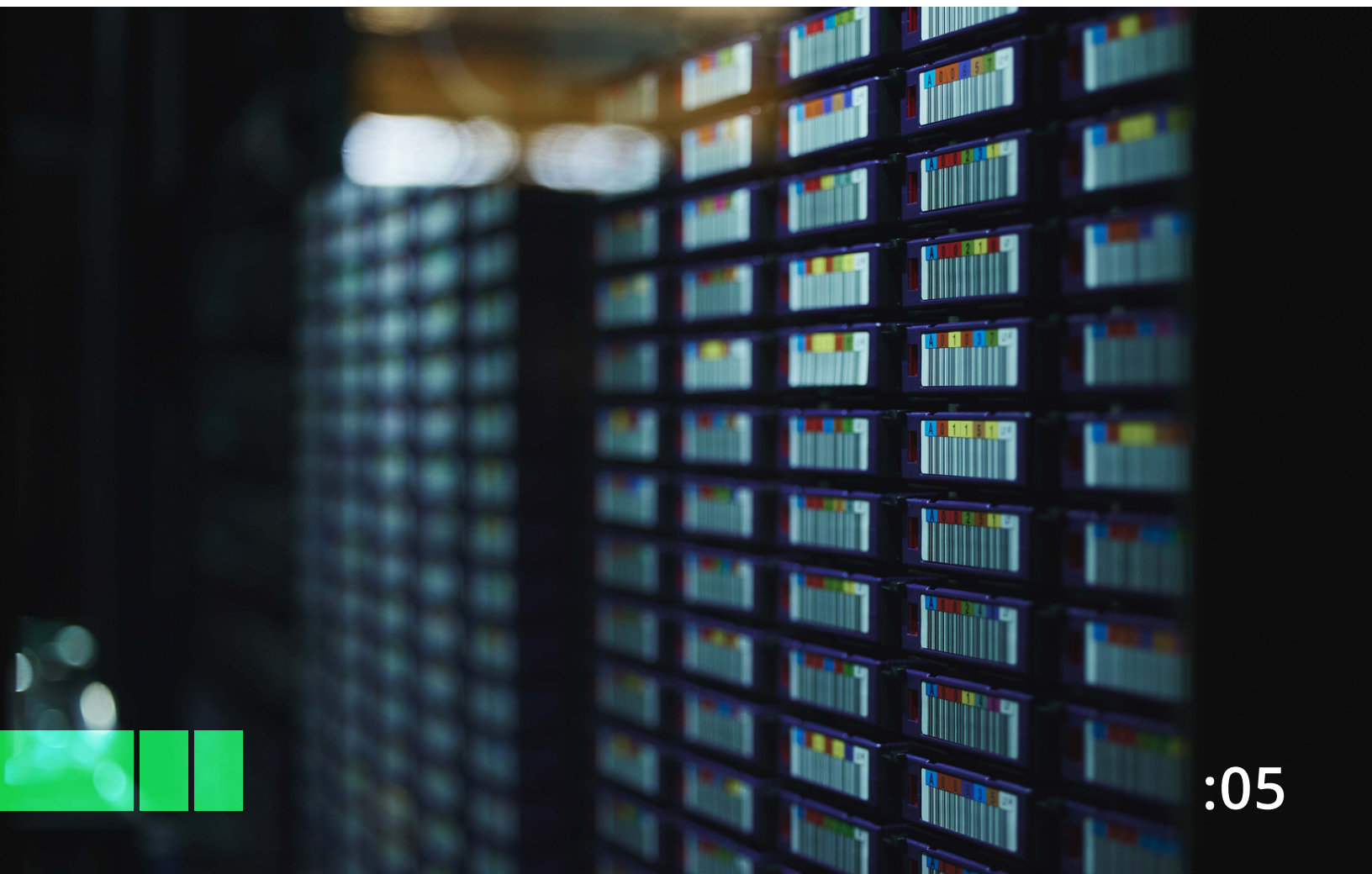
THE LESSON: Policies without enforcement are just paper. IST's consultative approach helps organizations not just write the policies, but audit whether they're actually being followed.

The BYOD Blind Spot

Bring-your-own-device policies are among the most commonly misunderstood data governance risks in corporate environments. Organizations frequently believe they have managed this exposure because they issue company devices and prohibit business communications on personal phones.

The gap is in the verification. When a single employee is using both a company-issued device and a personal device for business communications, whether in violation of policy or simply because the policy was never communicated clearly, the eDiscovery cost for that one custodian can significantly increase collection scope, review volume, and risk exposure.

These are not hypotheticals. They are patterns IST's team sees across engagements, in organizations of every size and sector. And they are entirely preventable with proactive information governance support.



Information Governance: Starting Left of the Problem

What a Data Map Actually Means

The single most powerful question an organization can answer before litigation arrives is deceptively simple: where is your data?

Not “what data do you have” though that matters too. Where is it stored? What platforms generate it? Who owns it? What are the applicable retention policies? Which data has outlived its legal and business purpose and should be deleted?

This is the function of an enterprise data map: a structured inventory of an organization’s information assets, the systems that generate them, the people responsible for them, and the policies that govern them. A good data map is the foundation of defensible information governance. And most organizations, even those with written policies, don’t have one.

An information governance engagement often begins with stakeholder interviews across legal, IT, compliance, records management, and business units. The goal is to understand where information is created, where it is stored, how it moves throughout the organization, and whether existing policies reflect actual practice.

In one engagement, a client believed the majority of its discoverable information resided within email archives. Through stakeholder interviews and data mapping exercises, additional repositories were identified across collaboration platforms, shared drives, archived storage, and employee devices. The result was a more defensible governance framework and a clearer understanding of future discovery obligations.

The EDRM Starts Earlier Than You Think

The Electronic Discovery Reference Model, the industry framework that most legal professionals use to describe the eDiscovery process, begins with information governance, not collection. That is not an accident. The decisions made at the information governance stage determine what is available, what is relevant, what is defensible, and what the total cost exposure will be when any triggering event occurs.

The organizations that invest in the left side of that model — governance, identification, and preservation, are the ones that arrive at litigation with confidence. They know their data. They have auditable policies. They can make targeted, strategic collection decisions rather than reflexive over-collection.

Those that skip the left side pay for it on the right, in hosting costs, in review volume, in sanctions exposure, and in time.

IST'S CONSULTATIVE APPROACH: When engaging a new corporate client, IST's first questions are not about the matter. They're about the data. Do you have a data map? What are your retention schedules? Are they being followed? Where are your employees actually communicating? The answers to those questions shape every downstream decision — and often reveal opportunities to save significant time and money before a single document is processed.

AI in eDiscovery: The Reality Behind the Hype

What AI Actually Does... and What It Doesn't

The term "AI" is now used so broadly in legal technology marketing that it can be difficult to separate meaningful capabilities from marketing claims. Vendors promise that AI will handle everything. Legal teams are told to just run AI on their data and move on.

The reality is more nuanced, and more useful, when understood correctly.

At its core, AI in the eDiscovery context is a pattern matcher. It does not replace legal judgment. It is not exercising judgment. It is taking a prompt, identifying data points, searching across an indexed universe of information, and returning the best match. For high-volume data reduction tasks, TAR, Continuous Active Learning, semantic clustering, image recognition, it does this with extraordinary efficiency.

But the pattern matcher has limits. The most important of these is the nuanced document. Consider a one-line email: “OK by me.” Was the prior message a birthday party announcement or a proposal to steal trade secrets? The AI sees a pattern. It does not see context. It cannot distinguish the innocuous from the incriminating based on three words alone.

This is why IST’s approach to AI has always been to treat it as a tool, not a replacement for judgment. AI handles what AI does best. Human expertise handles what requires human expertise. The results are faster, more accurate, and fully defensible.

Where AI Delivers Real Value

Deployed correctly, AI in eDiscovery creates genuine and measurable cost savings:

- **Technology-Assisted Review (TAR 2.0) and Continuous Active Learning:** AI trained on attorney-reviewed samples continuously reprioritizes the document set, surfacing the most relevant material first and pushing clearly irrelevant content to the bottom. Review is 30–50% faster without sacrificing accuracy.
- **Concept search and semantic clustering:** Rather than relying solely on keyword hits, AI identifies thematic connections across documents, finding relevant content that keyword searches miss entirely.
- **Privilege log support:** Once a document is identified as privileged, AI can surface all semantically similar documents across the entire dataset, dramatically reducing the risk of inadvertent privilege waiver.
- **Summarization and tone analysis:** AI can synthesize large document collections to surface key themes and flag sentiment shifts, insights that would previously have required hundreds of attorney hours to develop.
- **Image and multimedia recognition:** AI scans photographs, handwritten notes, and audio transcriptions to make non-text evidence fully searchable and reviewable.

THE HALLUCINATION CAUTION: Generative AI tools are increasingly being used by legal professionals to draft briefs, summarize cases, and research precedent. IST's guidance to clients is clear: AI is a powerful drafting accelerator, but it is not a verifier. The well-documented issue of AI "hallucinating" citations — confidently producing case names that do not exist — is a direct consequence of how pattern matching works. Generative AI can accelerate drafting and research, but it should always be validated by legal professionals before being relied upon in a filing, investigation, or legal strategy. It does not know whether that case is *Marbury v. Madison* or a fiction. Human verification is not optional.

The Role of the Project Manager: Where Cost Savings Actually Happen

Beyond the Platform

Organizations evaluating eDiscovery partners tend to focus on platform capabilities, pricing per gigabyte, and turnaround timelines. These are legitimate considerations. But the single most important variable in determining the total cost of an eDiscovery engagement is neither the platform nor the price sheet.

It is the project manager.

IST's project managers are recruited from Am Law 100 firms, with an average of 13+ years of hands-on eDiscovery experience. They are in the data every day. They have seen the trips and traps that accumulate across hundreds of matters. And when they sit down with a new client, they are not waiting to be asked, they are proactively counseling on where the money is being spent unnecessarily.

Targeted Collection: The Single Biggest Lever

The instinct when litigation arrives is to collect everything and sort it out later. It is understandable. It feels safe. It is also one of the most expensive decisions a legal team can make.

A more strategic approach: identify the top five to ten custodians most central to the matter. Collect and process those data sets first. Get into the matter within the first 30 to 60 days. In most cases, those key custodians will tell you everything you need to know about the case. And if the picture that emerges suggests a settlement is appropriate, at a cost significantly lower than the cost of full collection, processing, hosting, and trial, that intelligence has real financial value.

The difference between a \$2 million settlement and a \$20 million one, reached earlier in the process because the right data surfaced first, is not a marginal cost savings. It is a strategic advantage.

Preservation vs. Processing: Not the Same Decision

One of the most important distinctions IST's project managers help clients understand is the difference between preservation and processing. These are often conflated, but they carry very different cost implications.

- **Preserve everything** subject to a litigation hold or reasonable anticipation of litigation. This is a legal obligation and a cheap insurance policy against spoliation arguments.
- **Process strategically.** Not every preserved data set needs to be loaded into a review platform, run through analytics, and reviewed by attorneys. That decision should be driven by what the matter actually requires.
- **Host only what you need.** Hosting costs accumulate over the life of a matter. A dataset that does not need to be actively reviewed does not need to be actively hosted.

IST's project managers help clients make these distinctions with clarity and confidence, ensuring that every dollar spent is spent on what the matter actually needs, not on default behaviors that benefit the vendor more than the client.

Addressing the Real Questions

“WE ALREADY HAVE RETENTION POLICIES. ISN'T THAT ENOUGH?”

Only if they're being followed — and verified. IST regularly encounters organizations whose written policies are solid but whose actual practices have drifted significantly from what's on paper. A policy that isn't audited and enforced is, in practical terms, no policy at all. IST's consultative team helps clients close the gap between what the policy says and what is actually happening at the IT and user level.

“HOW DO WE KNOW WHICH CUSTODIANS TO PRIORITIZE?”

This is one of the most valuable conversations IST has with new clients. Using case facts, organizational structure, and data mapping, IST's project managers help counsel build a prioritized custodian list that focuses early collection efforts on the people and data most central to the matter. In most cases, this smaller targeted set surfaces the most critical evidence quickly — and informs whether a broader collection is actually warranted.

“IS AI-ASSISTED REVIEW DEFENSIBLE?”

Yes — when it is deployed with proper human oversight, validation protocols, and documented workflows. IST's approach combines AI efficiency with human QC at every stage. False positives and false negatives are tested for. Privilege decisions are validated. Every step in the process generates an audit trail that supports defensibility in court.

“WE'RE WORRIED ABOUT DATA VOLUME GROWING. WHEN IS THE RIGHT TIME TO ENGAGE?”

Before you need to. The organizations that manage eDiscovery costs most effectively are those that engage a partner during periods of stability — to build a data map, audit their governance practices, and put scalable processes in place. When litigation arrives, they are ready. Those who wait until the triggering event find themselves reactive, over-collecting, and overpaying.

The Path Forward: From Reactive to Ready

The eDiscovery landscape has transformed beyond recognition since the days of scanning paper documents onto floppy disks. The technology is faster, smarter, and more capable than anything the industry could have imagined in 1993. The global eDiscovery market is projected to exceed \$25 billion by 2029.

But the fundamental challenge has not changed: too much data, unclear where it lives, and a legal event that requires making sense of it, quickly, accurately, and at a defensible cost.

The organizations that will navigate that challenge most successfully are the ones that get ahead of it. That means building and maintaining a data map. That means ensuring policies are not just written but followed. That means making smart, targeted collection decisions guided by experienced project managers who have seen what works and what doesn't.

And it means treating AI for what it is: a powerful, cost-saving tool that delivers extraordinary results within its capabilities, and that still requires expert human judgment to deploy correctly.

IST's approach brings all of these elements together. From information governance consulting to early case assessment, from targeted forensic collection to AI-powered analytics and managed review, IST combines information governance consulting, experienced project management, and proven eDiscovery workflows to help organizations make better decisions about their data before, during, and after litigation.

The most expensive data is rarely the data an organization actively uses. It is the forgotten data sitting in legacy systems, archived repositories, personal devices, and unmanaged platforms.

The organizations that understand their data before litigation arrives are the ones that spend less, move faster, and respond with confidence when it does.

That is the value of information governance. And that is where IST helps clients start. ■ ■ ■

IST Discover-E provides end-to-end eDiscovery and information governance support — from data mapping and retention consulting through AI-powered analytics, managed review, and production.

IST Discover-E

800.636.8015 | info@istdiscover-e.com

www.istdiscover-e.com

People + Technology + Agility = Results



Greg Caraway
Vice President, Legal Solutions
IST Management Services

404.394.4166
gcaraway@istmanagement.com