VENIO

Defensibility

The Secret Facts of eDiscovery

Inside the Guide

Prelude

You've just gotten that dreaded call from the eDiscovery Administrator that it's going to take another two days before the data is ready for review. They mention something about "container extraction" or "corrupt signatures." However, the explanation doesn't make much sense. You hang up with more questions than answers. It's 10 p.m. and you have no idea where your eDiscovery stands.

The Challenge

In litigation, the producing parties and their attorney of record are ultimately responsible for managing eDiscovery. However, many traditional solutions were built from a patchwork of outdated applications. The result? Error-prone manual steps, poor data filtering, and inefficient review cycles that inflate costs while wasting the time of the most expensive resources - attorneys.

For years, the industry has relied on these legacy tools, requiring both deep technical knowledge and constant project oversight. Instead of streamlined, end-to-end platforms, most organizations still use a mix of software that forces repetitive manual tasks and creates duplicate data.

These processing issues rarely make headlines, but they cause serious risks - what we call the "Secret Facts of eDiscovery." Left unaddressed, they can lead to data spoliation between collection and review and jeopardize the chain of custody.

Layers of Data Extraction

Secret Fact #1: The Depths of eDiscovery Processing

Processing is the stage where collected data, say a Microsoft Outlook email file (.pst), is broken down into its individual parts, or "items." Each item needs to be searchable, filterable, reviewable, and producible. But items often contain their own layers: attachments, embedded files, zipped folders, and so on. A single email can balloon into dozens of items across multiple levels.

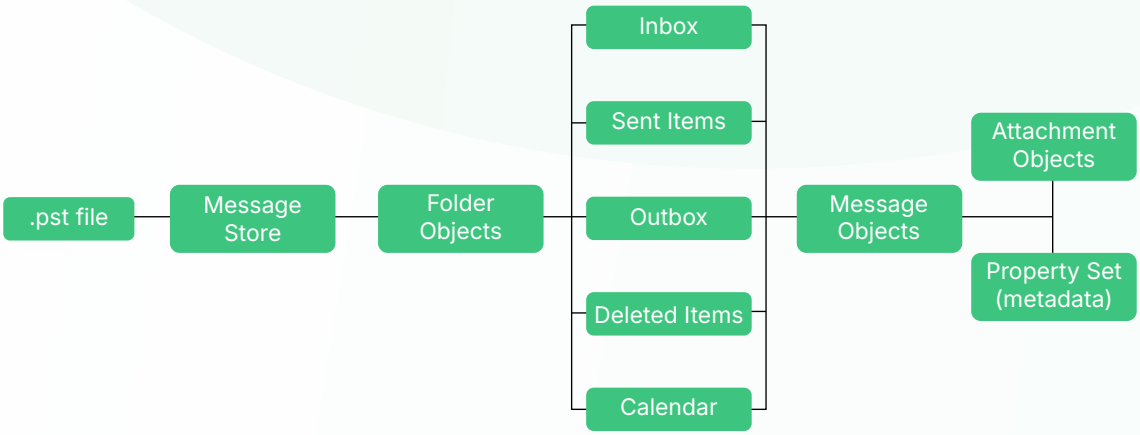

Image 1 : Structure of a Microsoft Outlook email file (.pst)

Flawed Search and Privilege Review

If the processing software sets a limit on how many levels it extracts, not all attachments or embedded files will be indexed. That means searches for responsive data won't surface them, and unless they're connected to another "hit," those files will never make it into review.
Even worse, if embedded files aren't extracted properly, reviewers may not see them at all unless they manually open each one. This creates a dangerous blind spot - potentially privileged, irrelevant, or even harmful documents could slip through and be produced without ever being reviewed.

The Solution to Secret #1

When evaluating an eDiscovery provider or platform, ask directly whether there are any level limits on extraction. Don't accept vague reassurances like "we've never had a problem" or numbers that sound large but still set a ceiling.

The only acceptable answer is simple: "none."

Without unlimited extraction, you risk missing critical files, undermining both defensibility and accuracy.

Multiple Applications

Secret Fact #2: When Too Many Tools Break the Chain

Many eDiscovery workflows are built on outdated processes that rely on a collection of separate tools. Each time data moves from one application to another - say from Tool A to Tool B, it has to be exported, converted, and then imported again. In many environments, this cycle repeats multiple times before data is finally produced.

Why does this happen? Older applications often can't handle every task effectively. As a result, service providers may use cheaper or specialized tools early on, especially when volumes are high, then switch to others later.

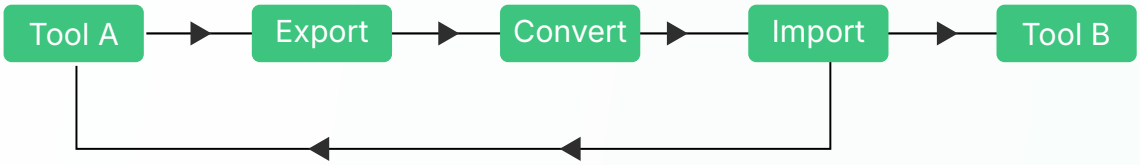

Image 2 : A loop of exporting, converting and importing while transfering data from Tool A to Tool B

Data Loss

Every export/import introduces opportunities for errors. Metadata like dates or authorship may get altered or dropped. Settings may not align between tools. And because human operators often oversee these transfers, even small mistakes can snowball into serious issues.
Many vendors market their systems as "integrated," but behind the scenes, they're still stitching together multiple applications. Middleware or cloud-based transfers may make it sound seamless, but if the process relies on moving data between systems, it's not truly integrated.

The Solution to Secret Fact #2: Technology Selection

When choosing a provider, look under the hood of their processing environment. How many separate tools or modules are actually involved? Do they rely on automated exports and imports to pass data along?

The safer choice is an end-to-end solution: a platform that performs all processing and review functions in one system, without moving or modifying data along the way.

For cloud-based offerings, apply the same scrutiny. “In the cloud” doesn’t always mean integrated. Understand exactly how data flows across components, and confirm that chain of custody and defensibility are maintained throughout the process.

Multiple Hardware Platforms

Secret Fact #3: The Cost of Moving Data Around

Remember the story from the prelude about the two-day delay? That wasn't fiction. The real cause was a server crash while moving large amounts of data between systems.

Here's the issue: when multiple eDiscovery applications run on different hardware platforms, even if hosted by the same vendor, data often has to move between servers. For example, hundreds of gigabytes may be transferred from Server A on one floor to Server B on another. If one server fails during the transfer, the entire process has to restart, creating delays and raising concerns about whether data can be fully recovered.

Time and Money

- **Time** → Every data transfer adds latency. Even in the cloud, moving data between different instances consumes processing cycles and slows deadlines.
- **Money** → More movement means more manual oversight and infrastructure costs. Each failure or restart increases labor hours and risks of overruns.

The bottom line: Multiple hardware platforms magnify inefficiency, risk, and cost.

The Solution to Secret Fact #3

When evaluating technology, don't just look at software features but also scrutinize the hardware requirements. A workflow that spans multiple servers or platforms requires multiplying everything:

- **More expertise to configure and maintain**
- **More infrastructure to support**
- **More complexity in disaster recovery**

For cloud services, ask specifically how data enters the system. The most efficient platforms allow for direct transfer into processing, without first staging it on an intermediate server. The fewer handoffs, the lower the risk of delay, cost, or data loss.

Human Participation

Secret Fact #4: People Are the Weak Link in eDiscovery

Despite advances in technology, human interaction remains one of the biggest risk factors in eDiscovery processing and review. Every manual step introduces opportunities for inconsistency, error, and inefficiency.

Continuity

Projects often span months or even years, with multiple batches of electronically stored information (ESI) arriving from different sources. Without standardized workflows or templates, each batch may be processed differently. This inconsistency can:

- **Disrupt the chain of custody**
- **Create gaps in the privilege or responsiveness review**
- **Require duplicative work, such as re-running search terms or reprocessing files for OCR, tiffing, or slipsheets**

Metrics

Many eDiscovery tools were designed primarily for technicians, not legal teams. They lack intuitive reporting features, making it difficult for project managers or litigators to track costs, timelines, or progress. Limited or hard-to-generate reports compromise the ability to manage budgets, supervise the process, and defend results in court.

The result? Increased risk of data errors, higher costs, and weaker defensibility.

The Solution to Secret Fact #4

The answer lies in reducing unnecessary human intervention and increasing automation, visibility, and control.

Automated Workflow

Modern applications should include templates that store settings and automatically apply them across all batches. Tasks like OCR, indexing, and privilege searches should run proactively - ensuring consistency, reducing errors, and making the process easier to document and defend.

Real-time Reporting: DASHBOARDS

Document review is both the biggest cost driver and the highest risk vector in eDiscovery. The only sustainable path is leveraging AI, analytics, and defensible automation to contain costs without sacrificing accuracy.

Status Reports

Dashboards don't replace reports, they complement them. The best solutions provide a full suite of customizable reports, including exception reports, so project managers and litigators can get the information they need without relying on technicians.

In short, automating repetitive tasks and giving stakeholders clear visibility not only saves time and cost, it also strengthens defensibility.

Keyword Searching

Secret Fact #5: Risks Behind Relying on Keywords Alone

While predictive coding and AI-driven review are gaining ground, keyword searching remains a central part of many eDiscovery workflows. It's used to identify responsive items, likely non-responsive materials, and potentially privileged documents. But when handled poorly, keyword searches can compromise both efficiency and defensibility.

Risks of Ad Hoc Searches

Too often, teams rely on haphazard or overly broad keyword lists, agreed upon hastily during meet-and-confer sessions. This can lead to:

- **Missing critical documents due to vague or incomplete terms**
- **Exposure to irrelevant or privileged material from overly broad searches**
- **Data sprawl that slows down review and drives up costs**

Courts have cautioned against this. In Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251 (D. Md. 2008), Magistrate Judge Paul Grimm emphasized the need for defensible, carefully planned search strategies.

Search Applications

Magistrate Judge John M. Facciola cautioned in U.S. v. Michael John O'Keefe, 537 F. Supp. 2d 14 (2008), that "this topic is clearly beyond the ken of a layman," advising litigants to seek expert guidance when handling eDiscovery issues such as search.

Multiple Applications

When keyword searches are run in different tools, search terms from Tool A must be translated into the correct syntax for Tool B. Because each application uses different search logic, results rarely match exactly. If both tools are being used to locate potentially responsive data, the number of results from the same query, such as a patent number, will likely differ.

Metadata Searches

A common mistake is assuming that a keyword search automatically covers both the text of documents and their metadata. While some tools search both by default, others do not. To avoid missing key information, users must confirm which fields are included in each search.

Level Limits

Accurate searches require full indexing of all items, including attachments and embedded files. Systems that cannot index every layer of stored content or impose limits on how deep indexing can go are not reliable search tools for discovery.

Reporting and Documentation

Strong search practices require transparent reporting that tracks every search iteration. Without this, searches risk overlooking items due to typos or incorrect terms, or pulling in irrelevant results from overly broad queries. Both scenarios increase the likelihood of incomplete results or unnecessary data sprawl.

The Solution to Secret Fact #5

Search Expertise

Courts and experts agree: effective keyword searching requires skill. If in-house teams lack this expertise, hiring a data analytics consultant or eDiscovery search expert can reduce review volumes and strengthen defensibility.

Search Engine

Whenever possible, teams should use an end-to-end platform where searches run consistently across the lifecycle of data: from ingestion through production. This avoids mismatched results and ensures a clear chain of custody.

Reports and Documentation

Defensibility depends on documentation. Strong platforms generate detailed reports for every search term, including hit counts, variations, and exceptions. This kind of reporting proved critical in Clark County v. Jacobs Facilities, Inc., No. 2:10-cv-00194-LRH-PAL, 2012 WL 4609427 (D. Nev. Oct. 1, 2012), where privilege was preserved because the responding party documented their search methodology.

That is: keyword searching isn't going away, but without expertise, unified tools, and clear documentation, it creates risks that no legal team can afford.

Conclusion

Your Responsibility: Defensible eDiscovery

The “Secret Facts” of eDiscovery may not always make headlines, but they directly impact your responsibility as a producing party. Courts and clients expect that electronically stored information (ESI) will be handled with accuracy, consistency, and defensibility.

The eDiscovery industry has advanced significantly, with modern platforms now offering end-to-end workflows that automate best practices, reducing reliance on fragmented tools, manual tasks, and human error. At the same time, the legal system continues to update rules and standards to reflect the growing importance of ESI in twenty-first-century discovery.

To conclude, you cannot delegate away responsibility for eDiscovery. Whether you are a litigator, technologist, or project manager, ensuring proper processes, technology selection, and documentation is essential. By addressing these “secret” risks head-on through automation, expertise, and defensible workflows, you not only mitigate risk but also strengthen your ability to deliver efficient, cost-effective, and compliant discovery.

About Venio Systems

At Venio Systems, we are dedicated to working with our trusted partners to bring the latest legal technology innovations to law firms, agencies, and corporations. We combine advanced technology with practical design to deliver smarter eDiscovery. Our all-in-one platform helps organizations streamline workflows, reduce costs, and maintain defensibility at every stage of the EDRM.

Why Venio is Different

Traditional Tools	Venio Systems
Multiple tools, fragmented workflows	Unified end-to-end platform
Error-prone manual processes	Automated, AI-driven workflows
Limited reporting visibility	Real-time dashboards & analytics
High per-user licensing costs	Flexible, cost-efficient pricing
Difficult to scale with modern data types	Scalable, future-ready architecture

PROTEUS DISCOVERY GROUP · Consilio · NIXON PEABODY · MERLIN · HAUG PARTNERS · REPARIO · Array · CDS · amentum

Ready to Rethink eDiscovery?

Book a Demo Today!

VENIO


Venio
Review
Venio
Legal Hold
Venio ECA
AI
Venio
AI Review
Venio
Production