VENIO

EBOOK

The Cost Curve Nobody Budgeted For

Why eDiscovery costs keep climbing even when your legal risk does not, and what actually bends the curve

4.5x
Enterprise-held data growth projected from 2025 to 2030

~35%
Annual data growth driving volume-based discovery cost

62%
Estimated 2025 share of spend tied to review work

The number that should concern every data-rich company

Most companies budget for eDiscovery as if its cost tracks their legal exposure: more lawsuits, more spend. For data-intensive businesses, that assumption is quietly wrong. The cost increasingly tracks something else entirely: how much data you generate.

And data is compounding. According to ComplexDiscovery's reconciled market estimates, enterprise-held data, the slice most relevant to discovery, is on track to grow from roughly 54 zettabytes in 2025 to about 243 zettabytes by 2030. That is a 4.5x increase in five years, close to 35% a year. None of that growth is litigation. It is product, headcount, customers, and the ordinary exhaust of a company doing well: messages, documents, code, tickets, and logs.

If your eDiscovery is priced by volume, the same growth that signals a healthy business also raises the floor on what discovery costs. Your spend climbs in years you never set foot in a courtroom. The better the company does, the worse this line item looks.

> **THE UNCOMFORTABLE LINK**
>
> Under a volume-based model, your discovery cost line starts to mirror your data growth line. It is no longer a litigation-driven expense. It becomes an operating cost of being a successful, data-producing company.

Where volume quietly becomes cost

A volume-based model rarely charges you once. It meters you at three stages, and every one scales with data.

Processing

Processing comes first, ingesting and preparing data, typically billed per gigabyte, before a single document has been looked at. Larger custodians and data-heavy matters cost more by default, regardless of what's actually relevant.

Hosting

Hosting is the charge that compounds. Data sits in the platform, usually billed per gigabyte per month, for as long as the matter is open. Matters that drag on cost more for reasons that have nothing to do with their merits, and resolved matters often keep accruing rent indefinitely.

Review

Review is the highest cost in most matters and the one most sensitive to volume. RAND placed review at 73% of total task spend, and reconciled 2025 modeling still places it around 62%. When volume rises, review rises with it, and review is where the real money is.

The collaboration-data multiplier

For technology and IT-driven companies, there is a fourth force that makes all three stages worse: the data is not documents anymore. It is a conversation.

Slack threads, Teams channels, Jira comments, and chat are now among the fastest-growing categories of discoverable data, and they behave nothing like email. Tools built for email treat each message as a standalone record. A single threaded conversation between four engineers does not enter your review set as one item. It enters as dozens, sometimes hundreds. You process that inflation, you host it, and you pay attorneys to review it.

Collection has become its own rising expense as well. Its share of total discovery spend has roughly tripled over the past decade as data sources multiplied and modern collection grew more specialized. Native Slack exports, for instance, often arrive as raw JSON that is nearly unreadable without dedicated tooling, so simply making the data reviewable becomes a billable project of its own.

The cost curve

Put it together, and a clear shape emerges. Under volume-based pricing, your discovery cost line is, in effect, the same line as your data growth. The curve below follows the attached indexed comparison: volume-priced cost rises from 100 to 246 while flat-rate cost stays at 100.

Indexed Cost Curve Comparison

Two lines, two completely different futures. Under volume pricing, cost rises lockstep with the data your business naturally creates. Under a flat-rate model, the two decouple. Data can grow 35% a year while the discovery number stays something you set once and plan against.

Why this is a forecasting problem, not just a budget one

The deepest issue with volume pricing is not simply that it is expensive. It is that it is unpredictable. Every large matter and every data-heavy custodian becomes a fresh negotiation with your own budget, and the eDiscovery line becomes the one finance can never count on.

And if you do not run eDiscovery at all yet, the curve is still there. You just cannot see it. The data is accumulating now, in tools that auto-delete on their own schedules, and the first time you will meet the full cost is under deadline, when a matter forces you to find, hold, and produce all of it at once. The curve does not wait for you to be ready for it.

What actually bends the curve

Four things change this math, and none of them require accepting that discovery cost must rise with data growth.

LEVER 1

AI-assisted review

Because review is the largest line in most matters and scales with the number of items, the highest-leverage move is reviewing fewer items. TAR, CAL, and concept clustering surface what is relevant early and set aside what is not, so attorneys spend their hours on what matters instead of reading noise.

LEVER 2

Flat-rate pricing

Flat-rate pricing decouples cost from data entirely: the number is known before the matter starts, so a 200%-growth year or a single enormous custodian does not reopen the budget.

LEVER 3

Native handling of collaboration data

Reconstructing Slack and Teams conversations in context, rather than exploding them into thousands of standalone items, cuts volume before review begins and compounds the savings from AI review.

LEVER 4

Deployment control

For global and security-conscious organizations, running discovery in the cloud, on-premises, or fully air-gapped on a matter-by-matter basis turns data residency from a constraint into a choice.

This is the approach Venio is built around: defensible AI that shrinks the review set with humans in control, flat-rate pricing that holds the line as data grows, native handling of collaboration data, and deployment on your terms. The result is not a cheaper version of the same curve. It is a flat line where there used to be a rising one.

The bottom line

Your data will keep growing. That is the point of a healthy business. The real question is whether you are ready for what that means: that the cost and the risk of finding, holding, and producing it are scaling right alongside it.

Whether you already run discovery or have not had to yet, the curve is the same. The only difference is whether you meet it on your own terms or under a deadline. The fastest way to see yours is to walk through a recent matter, or your likely first one, with someone who can map it.

See your cost curve in a 30-minute walkthrough

For legal ops, discovery counsel, and finance leaders

Bring one recent matter, or one likely upcoming one, and we will map where processing, hosting, collaboration data, and review costs start to compound. You leave with a clearer forecast and a practical strategy to flatten the curve.

- Forecast cost exposure
- Spot hidden cost multipliers
- Leave with a practical plan

Schedule the walkthrough

Sources

1 Enterprise-held data ~54 ZB (2025) to ~243 ZB (2030), about 4.5x / ~35% per year — ComplexDiscovery OÜ reconciled market model (Rob Robinson), “Market Intelligence: eDiscovery market growth from 2012 to 2030” (May 1, 2026; republished by EDRM May 19, 2026).

2 Review share of task spend: 73% (2012) to about 62% (2025); collection's share roughly tripled over the decade — same ComplexDiscovery analysis, with 2012 baseline drawn from RAND Corporation, Where the Money Goes (Pace & Zakaras).

3 AI features including TAR, CAL, concept clustering, and defensible human-in-control workflows — Venio Systems product and analytics materials.

4 Cost-curve table/graph in this document is illustrative and indexed to show directional shape rather than quoted pricing figures.

veniosystems.com