The Single Source of Truth Starter Guide

Build the one place where everything the firm knows actually lives - on a foundation that holds

Most firms don't have a single source of truth. They have a client list in one system, old adverse parties buried in closed files, and the rest scattered across email, spreadsheets, and one partner's memory. The ambition this guide serves is to fix exactly that: pull it all into one place, run AI across it, and finally ask questions of the firm's own knowledge.

That is the most valuable thing a firm can build. It is also where privilege and confidentiality questions concentrate, which is why the architecture has to be right before the first file goes in. This guide gives you that architecture: the layers, the build order, and the honest picture of where the law stands.

The infrastructure this requires is real and available today. Whether it preserves privilege across every scenario is the strong-but-not-yet-settled question that runs through every decision below. Build on the best available argument, configure to make your facts as strong as possible, and get an independent counsel opinion before your most sensitive matters go in.

1. What the single source of truth is and why it matters

The conflicts problem is not about checking. The checking is easy. The problem is that the firm has no single place where everything it knows about its own clients and matters actually lives.

The client list is in one system. Old adverse parties are buried in closed files. The rest is scattered across email, spreadsheets, and one partner's memory. When a new client comes through the door, conflicts screening becomes a hopeful search rather than a real one. When a partner retires, institutional memory walks out with them.

The single source of truth fixes this. One store - every client, every matter, every document, every history. AI runs across it, so the firm can finally ask real questions of its own knowledge. Conflicts becomes a genuine search. Institutional knowledge stays. The store becomes the most valuable operational asset the firm has.

It is also the thing that gathers the entire privilege question into one place. The bigger the build, the more a court challenge would put everything at risk at once. That is why the architecture below exists: to make the build sound before the first file goes in, not after.

2. The architecture: three layers, each with a guardrail

The store has three layers. Each layer has a defined scope and a non-negotiable guardrail. Build all three correctly, and the infrastructure-provider argument - that your automation layer is plumbing, not a privilege-breaking third party - has the strongest available factual foundation.

Layer	What goes here	The guardrail
Sources	Every client record, matter file, closed file, adverse party, document, email thread, and historical intake that the firm needs to query. This is the raw input - everything the firm has ever known about who it has represented, who it has been adverse to, and what it has worked on.	Counsel-directed ingestion. Nothing enters the store unless a lawyer has authorized it. This is not a clerical function. The work-product and privilege arguments both require that the work be prepared by or at the direction of counsel. Cut counsel out of the loop and you remove the thing the doctrine protects. Every import, migration, or connection is a lawyer-level decision.
Platform	The automation layer that carries data into and across the store, and the AI layer that reasons across it when queried. This includes the middleware (the automation platform moving data), the vector database or knowledge base holding it, and the AI engine answering questions against it.	Business or dedicated tier only. No consumer-tier tool touches this store at any layer. The platform's contract should carry three protective-order terms, with a carve-out for what the service needs to operate: no storing or training on the data, no third-party disclosure except as essential to service delivery, and deletion on request. Those terms come from one magistrate judge's order in one case, not a settled industry standard, and the order named no vendor. Treat them as the strongest available benchmark, not a court-blessed checklist. Any tool that cannot meet those terms does not touch this store.
Access	Who and what can query the store. Lawyers querying for conflicts. Staff authorized by counsel to run specific searches. AI models answering questions against the stored knowledge. Every query path that reaches client information.	Matter-team scoping, ethical walls enforced in the permission layer, human checkpoint on AI output. Access defaults to the matter team. Screened lawyers are walled out at the permission layer, not by policy alone. The AI can surface, rank, and summarize. A human confirms before any result is relied on. The AI does not make the final call on a conflicts determination, a privilege assertion, or any significant decision that flows from the store. The model finds the candidates. The lawyer makes the judgment. That structure is what keeps the store on the right side of the supervision duty and keeps the output reliable enough to stake the firm on.

> **Ethical walls.** A firm-wide queryable store is exactly what ethical screens exist to prevent from becoming a leak. Build the wall into the store's permission layer, not into a policy memo. Access defaults to the matter team. A walled matter is excluded from a screened lawyer's query results outright - the model never surfaces it, and the lawyer never sees it to judge it. Log every query against the store, so a wall can be audited, not just assumed.

A fourth layer runs underneath all three and stays out of the visible build: the sub-processor chain. When your automation platform calls an AI module, that module's provider is a separate party with its own terms. Your platform's data processing agreement does not automatically reach it. Every AI engine in the store needs its own analysis, and for sensitive matters, its own enterprise contract. Map the sub-processors deliberately. The protection stops at your platform's edge and the next provider's terms take over.

3. The build order

The order below is not arbitrary. Each step either creates the legal foundation the next step rests on, or catches the failure mode that would otherwise sink the build silently.

1. **Get counsel in the loop before anything else.**
   Before the first file is ingested, a lawyer has to own the decision to build. That lawyer reviews the architecture, authorizes the tier choices, and signs off on which matters go in. The work-product argument requires that materials be prepared at the direction of counsel. The infrastructure-provider argument for privilege requires the same. A store built without that direction has a gap at its foundation that no contract language can close after the fact. Start here.

2. **Put the right tier underneath.**
   Business tier is the floor. Dedicated tier is required when the store will hold health information or when the litigation sensitivity of the matters demands it. Choose the tier before you connect a single data source. Then run the protective-order terms against every platform and sub-processor: does the contract bar storing or training on the data, bar third-party disclosure except as essential to service delivery, and permit deletion on request? Any tool that cannot clear those terms is not in this build. Do not finalize the architecture until every component in the chain has cleared the terms in writing.

3 **Write the protective-order terms into the contract.**
The terms above come from one magistrate judge's order resolving a contested motion in one case - not a settled industry standard, and the order named no vendor. Treat it as the strongest available benchmark, not a court-blessed checklist. Get those terms into the vendor contract before the store goes live. "Does not train on your data" written on a marketing or documentation page is not the same as a signed data processing agreement, and vendors word the protection differently inside the real agreement: some state the training bar outright, others bar any use of your data beyond delivering the service, which forecloses training in substance. Either form counts when it sits in the executed agreement; a claim that lives only on a product page counts for nothing until the vendor confirms it in writing. The agreement is what a court looks at. Get it executed, keep it current, and know where it is.

4 **Put a human checkpoint on AI output.**
Wire the checkpoint into the workflow before the store answers its first query. The model surfaces results. A person confirms them. That structure is non-negotiable for two reasons. First, AI can be confidently wrong, and in a conflicts search or a privilege determination, a confident wrong answer has real consequences. Second, the supervision duty that attaches when AI reads client content and produces output a person relies on requires a human to own the result. The checkpoint is not bureaucratic overhead. It is what makes the output reliable and the firm defensible.

5 **Start narrow and widen deliberately.**
The biggest risk in building a central store is moving too much data too fast. Start with one matter type, or one closed-file cohort, or one practice group. Run it for a defined period. Test the query quality. Test the access controls. Test the sub-processor chain. Confirm the tier holds under real use. Then expand to the next cohort. A staged build catches configuration errors before they reach the full archive. It also lets counsel review each expansion as a new authorization decision, which keeps the counsel-direction requirement current rather than treating it as a one-time sign-off at the start.

4. The platform question, answered plainly

Firms building this store ask the same platform questions. The answers follow directly from the tier model.

Question	The answer
Can we use the free tier of a major AI platform for queries against the store?	No. Consumer tier is never appropriate for client data, on any track, for any process. The consumer terms permit training on the data and do not give you the contractual protections a custodian of confidential information needs. Use the business or dedicated tier, configured correctly, with the contract in place.
Is our automation middleware covered by the same protections as our AI platform?	Only for its own processing. Your automation platform's data processing agreement covers what that platform does with the data. It does not reach the AI sub-processors you invoke from inside a workflow. Each sub-processor needs its own analysis and, for sensitive matters, its own enterprise contract. Map the full chain before you go live.
Can we stack zero-data-retention and a business associate agreement on the same AI model?	Not always. At least one major provider makes those options mutually exclusive on its covered models. You can have zero data retention or a signed HIPAA agreement, but not both at once on those models. If health information will enter the store, confirm which option your chosen model actually allows before you architect the workflow around an assumption you cannot verify.
Does using cloud automation middleware make HIPAA compliance impossible?	On mainstream cloud platforms, yes - for now. None of the major cloud automation platforms offer a business associate agreement on their standard cloud products. If protected health information will enter the store, the automation layer either needs to run on self-hosted infrastructure you control, or that data stays out of the store until a compliant path exists.
Does a "Data is Confidential" setting on our automation platform protect everything?	It protects the platform's own logging of what passes through it. It does not reach past the platform to any outside service the workflow calls out to. Once your automation hands data to an outside AI engine, that engine's contract governs. The setting and the contract are both required. Neither one alone is enough.

5. The privilege question: what the law says and what it does not yet say

A firm building this store needs an honest picture of where the legal protection stands. There are three parts, and all three matter.

Work product: early signals, not a settled rule

Work-product protection shields materials prepared in anticipation of litigation. Two early magistrate-level discovery rulings point the right way on using a commercial AI tool to help prepare those materials, reasoning that AI tools are tools, not persons, and that routing protected work through one is

not the same as revealing it to an adversary. Commentators call the broader landscape split. Work-product waiver turns on disclosure to an adversary, or disclosure in a way likely to reach one, and the rulings so far treat ordinary commercial AI use as falling short of that bar.

One structural requirement runs underneath both rulings: the materials have to be prepared by or at the direction of counsel. Keep the lawyer directing the workflow, and the protection has something to attach to.

Two limits to know. Work-product protection only exists where there is anticipation of litigation, so transactional and regulatory matters do not get it - the protection requires anticipation of litigation, which those matters lack. And this protection covers materials prepared in that context - it is not a blanket cover for everything in the store.

Attorney-client privilege: the strongest available argument, honestly labeled

The protective theory for privilege is the infrastructure-provider argument: an automation platform moving your data is infrastructure, not a participant in the legal conversation. The same way the phone company carrying a call does not break privilege, a properly configured automation layer carrying client data should not either. That argument draws real support from ABA guidance on protecting client information and from the work-product decisions' "tools, not persons" reasoning.

Here is the honest label: the infrastructure-provider theory is the best available argument, not settled law. No federal circuit has formally adopted it for automation middleware. A firm relying on it for genuinely sensitive matters should get an independent counsel opinion before treating it as decided. Strong argument, real support, no judicial seal yet.

The more you concentrate in the store, the more that opinion matters. A single workflow can lean on the theory. A store holding everything the firm knows should be built on the theory and backstopped by a real legal opinion before the most sensitive matters go in.

Consumer tier: where the protection has already failed

A federal court has already shown exactly where the protection breaks. Privilege never attached to material run through a consumer AI tool where no attorney directed the use and the provider's own privacy policy permitted retention and disclosure to third parties. The court's conclusion was that privilege never attached - not that it was destroyed after the fact.

Scope that holding to its facts: a consumer-tier tool, no attorney in the loop, terms that permitted disclosure. Do not generalize it to enterprise AI where the facts are different. But do take the lesson seriously. The consumer-tier fact pattern is the cliff the store architecture is designed to keep you away from.

> **Honesty note.** The infrastructure-provider theory is the backbone of the privilege argument for this build. It is consistent with recent court decisions and ABA guidance, and it is where the smart money is watching. But no federal circuit has formally adopted it for automation middleware. Before you stake your most sensitive matters on it, get an independent counsel opinion. This guide is a starting architecture, not a legal opinion, and the law has not yet given this build its full blessing. Build on the strong argument, configure to make your facts as strong as possible, and get the opinion before going live with your highest-sensitivity work.

6. Pre-launch checklist

Run this before the store goes live. Every item is required. None of them can be treated as optional on the assumption that the others cover it.

- ☐ **Counsel authorization in writing.** A lawyer has reviewed the architecture, approved the tier and sub-processor choices, and authorized the initial data ingestion. That authorization is documented and dated.
- ☐ **Tier confirmed for every component.** Every platform and sub-processor in the chain - the automation middleware, the knowledge base, the AI query engine - has cleared the protective-order terms: contractual bar on storing or training on the data, on third-party disclosure except as essential to service delivery, and permission for deletion on request. The contracts are signed and on file.
- ☐ **Sub-processor map complete.** Every AI engine the store calls, directly or through the automation layer, is identified. Each one has its own executed contract or has been excluded from sensitive-matter data flows until one exists.
- ☐ **Health information decision made.** If the store will hold protected health information: the cloud automation middleware question is resolved (self-hosted or excluded), the zero-retention vs. BAA tradeoff is confirmed for the AI layer, and the outcome is documented.
- ☐ **Human checkpoint wired in.** Every query path that produces output a person relies on has a human confirmation step before that output is acted on. The checkpoint is in the workflow, not in a policy document.
- ☐ **Independent counsel opinion obtained (sensitive matters).** For any matter where privilege is mission-critical, a counsel opinion on the infrastructure-provider theory is on file before the matter goes into the store.

- ☐ **Staged ingestion plan confirmed.** The first data cohort is defined and scoped. A review date is set before expanding to the next cohort. Counsel will re-authorize each expansion.
- ☐ **AI policy and verification protocol in force.** The firm has a written AI policy covering this store and a verification protocol specifying the human checkpoint requirements. Both are in force before the store answers its first query.

How to use this. This is a thinking tool, built to help you reason about your own firm. It is not a compliance guarantee and it is not legal advice. Where it states a rule, it states the rule as of July 18, 2026, and the law moves, so confirm the current rules in your jurisdiction with your own counsel before you rely on anything here.

From *The Automated Law Firm* • 4Spot Consulting • **4SpotConsulting.com/single-source-guide**