Docket

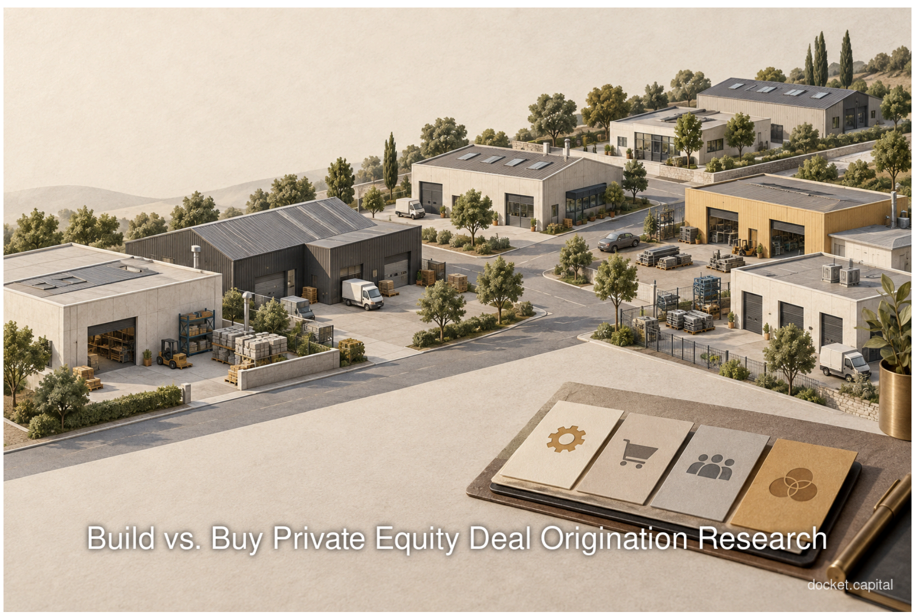
DOCKET / RESEARCH & PRACTICAL GUIDANCE

Build vs. Buy Private Equity Deal Origination Research

> Build a more reviewable acquisition pipeline with Docket. Our deal-origination software and managed research help private equity teams screen companies against a mandate and understand the evidence behind each finding.

Published by Docket · 2026-09-23 · build vs buy private equity deal origination research · private equity deal origination · deal sourcing software · in-house target research · managed research · total cost of ownership · evidence portability · vendor evaluation

Original article: https://docket.capital/articles/build-vs-buy-pe-origination-research



In this article

Executive Summary

Private equity (PE) teams choosing how to run acquisition target research face **four operating models**: build an internal workflow, buy software, commission managed research, or combine internal and external components. The comparison should begin with the work to be done, not the advertised size of a company database. A useful target file must connect a company identity to mandate questions, source material, review decisions, and a controlled handoff to a customer relationship management (CRM) system. The Institutional Limited Partners Association (ILPA) asks managers how sourcing is staffed, conducted, and documented, and how third parties participate; those are useful prompts for the operating-model design, though its questionnaire is not a performance benchmark. (ilpa.org) (ilpa.org)

Five-year total cost of ownership should include implementation, subscriptions or data rights, engineering and analyst time, review and cleanup, integrations and changes, a stated risk reserve, and exit. The Government Accountability Office (GAO) recommends a technical baseline, work breakdown structure, explicit assumptions, sensitivity analysis, and updates with actual costs; its guide says the practices also apply to internal development. (www.gao.gov) (www.gao.gov) A price is only one input. For a concrete first-party example, **Docket** currently lists **$1,500 per month** for self-serve and **$12,000 per month** for managed research, with different scope and term conditions. Those amounts are vendor prices, not estimates of a firm's fully loaded cost. (docket.capital)

Control is more than ownership of code. It includes the right to use underlying data, the fit of the research schema to an investment mandate, access controls, version history, exportable evidence, and a credible exit route. OpenCorporates explicitly warns that an empty source-license field does not mean the underlying data has no license; an internal build therefore does not automatically clear data rights. (api.opencorporates.com) The practical test is a common, time-boxed pilot: give each option the same target identities, mandate questions, permitted sources, review rubric, and export requirements. Count **accepted field answers**, not deals, and record reviewer minutes and unresolved questions. The FinOps Foundation calls for consistent unit definitions and fully loaded cost inclusions. (www.finops.org) (www.finops.org)

There is no universal winner. Build is defensible when mandate logic and evidence controls are distinctive and the team can maintain them. Buy can compress implementation work when configurable workflow meets the acceptance gates. Managed delivery can transfer day-to-day research work while retaining client review obligations. Hybrid is often the explicit design when research execution, evidence storage, and CRM relationship history have different owners. The decision should be revisited after the pilot and at planned contract or thesis changes. NIST's purchaser guidance supports requesting concrete secure-development information from a software producer, while cloud portability guidance calls for termination terms covering access to data. (www.nist.gov) (cloudsecurityalliance.org)

Introduction and Background

Deal origination research turns an acquisition thesis into a repeatable sequence: accept or construct a target list, resolve each company's identity, ask the same mandate questions, retain the supporting material, have a reviewer accept or correct findings, and pass approved records into the firm's relationship workflow. The decision about delivery is therefore a decision about **process ownership**. A database may provide useful company attributes without supplying review history; a CRM may preserve interactions without retaining the page that supports an ownership answer. Affinity's own guide describes moving sourced companies into a CRM for deeper qualification and says CRM records include interactions and notes. (www.affinity.co) (www.affinity.co)

The research object should be specified before a request for proposals or internal build estimate. At minimum, record a stable company identifier, jurisdiction and legal entity, mandate version, question and accepted-answer vocabulary, source URL or document, excerpt, retrieval time, reviewer decision, conflict status, and permitted next action. The World Wide Web Consortium (W3C) provenance model treats entities, activities, and agents as distinct objects and represents a revision as a relation to an earlier entity. That gives a useful model for keeping the research trail inspectable as an answer changes. (www.w3.org) (www.w3.org)

A minimum capability specification should name the fields and decisions at each step:

- **Intake:** who can add a candidate and which identifiers are mandatory.
- **Identity:** how legal entities, brands, duplicates and uncertain matches are handled.
- **Mandate:** which questions, accepted answers and disqualifiers apply.
- **Sources:** what may be collected, retained, quoted and reused.
- **Review:** who accepts, corrects or leaves a field unresolved.
- **Handoff:** what reaches the CRM and who can approve outreach.

These distinctions matter because a URL alone is a fragile evidence store. In a Pew Research Center study using a random sample of just under one million archived webpages, **38% of pages collected in 2013** were inaccessible in 2023. Pew measured disappearance, not every change in page content, so this is a warning about link persistence rather than a measured error rate for acquisition research. (www.pewresearch.org) (www.pewresearch.org) (www.pewresearch.org) Retaining a permissible copy or excerpt, plus its collection date, reduces dependence on a live link; the scope of retention still depends on source terms and applicable rules.

This report evaluates the **operating model** for sourcing and checking target research. It does not rank data providers or infer deal outcomes from workflow features. Docket is a direct provider of both software-led and managed target research, so its published terms appear in the comparison as identifiable examples rather than as a recommendation. Its evidence guide describes a research file that lets a reviewer trace an answer to supporting material. (docket.capital)

Build an Internal Research Workflow

Capabilities

An internal build can encode the firm's own identity rules, screening questions, source hierarchy, review states, and CRM event contract. It can also choose a data model that keeps an observation separate from the current accepted answer. The W3C model's derivation relation is a useful way to distinguish a raw source, an extracted statement, and a later conclusion; it does not prescribe a PE application. (www.w3.org) A build specification should include intake, deduplication, researcher queue, evidence capture, reviewer queue, change history, role permissions, exports, and monitoring. It also needs a way to change those components when the mandate changes.

The internal estimate should assign an owner and cost line to each continuing obligation:

- **Ingestion:** source connectors, rate limits and schema changes.
- **Storage:** records, attachments, backups and retention.
- **Review:** analyst queues, sampling and correction work.
- **Security:** access reviews, logging and incident response process.
- **Integration:** CRM sync, mapping and reconciliation.
- **Maintenance:** releases, dependency updates and support.

The build still depends on third-party inputs. Registry data, websites, commercial company data, CRM records, and cloud infrastructure can each have separate contracts and technical limits. OpenCorporates documents provenance in its detailed API responses and distinguishes its paid accounts from share-alike API terms. Those statements illustrate why the data entitlement must be inspected at the field and use-case level, even when the firm owns its application code. (api.opencorporates.com) (api.opencorporates.com) A UK registry API, for example, documents a **600-request limit per five minutes** and instructs clients to tolerate additional JSON members; those details belong in ingestion and change estimates, not in a one-time build budget. (developer.company-information.service.gov.uk) (developer.company-information.service.gov.uk)

Adoption

The internal route needs named owners for product decisions, engineering, security, data licenses, analyst operations, and reviewer policy. Start with a technical baseline: volumes, required fields, source classes, review standard, integrations, uptime need, and retention period. GAO's guide explicitly includes software cost estimating and analysis of alternatives, a stronger starting point than a developer-hours guess. (www.gao.gov) The acceptance owner should be an origination or research leader, because the central output is a defensible company answer, not merely a working interface.

Strengths and Limitations

The strength is **schema and process fit**. A firm can make mandate-specific questions and provenance mandatory, expose exactly the states its analysts use, and decide when changes ship. The limit is the ongoing obligation: dependency upgrades, source changes, reviewer training, monitoring, security work, and migration are the firm's work. Thoughtworks characterizes third-party purchase as a way to obtain proven capability quickly; that is practitioner guidance on the tradeoff, not comparative evidence about PE deal results. (www.thoughtworks.com) Build is strongest when the workflow is genuinely distinctive and the firm has a durable owner for its evolution.

Buy a Research Platform

Capabilities

A purchased platform may supply data, screening, task routing, integration, or relationship management in different proportions. SourceScrub says its PE workflow includes custom scoring and an API for embedding data in existing applications. Intapp says DealCloud supports rule-based tasks and approvals. These are documented vendor capabilities, not evidence that either will satisfy a particular mandate or field-level citation standard. (www.sourcescrub.com) (www.sourcescrub.com) (www.intapp.com) The buyer must map each required object and decision state to a configured field, API object, export, or documented gap.

A useful distinction is **research system versus CRM**. The research system stores question, answer, source, reviewer, and freshness. The CRM stores people, conversations, relationships, ownership of outreach, and pipeline stages. They may be one product or linked products, but the target identity and event semantics must match. A vendor's ability to send profiles to a CRM is a starting point, not proof that source excerpts, correction history, and not-found records travel with them.

Adoption

Buying shifts much implementation effort toward configuration, integration, security review, and vendor management. Run a field map before procurement: required object, field definition, source, access role, retention, export representation, and update frequency. Ask the provider to demonstrate the complete path from a newly imported company to an accepted answer and CRM handoff. Intapp documents import and export controls and audit logs; a buyer still has to test which records and attachments its own configuration can actually retrieve. (www.intapp.com)

A short demonstration script can keep vendor evaluations comparable:

- **Import:** accept the same frozen target sample.
- **Configure:** load the exact mandate questions and answer rules.
- **Research:** show one supported, one conflicting and one unanswered field.
- **Review:** correct an answer without erasing its prior state.
- **Integrate:** transfer an accepted record to the CRM.
- **Export:** deliver IDs, evidence, timestamps and permissions.

NIST's purchaser material recommends using Secure Software Development Framework (SSDF) terminology to organize security requests and asking for artifacts traceable to supporting detail. A PE procurement team can adapt that questioning without implying a particular certification or outcome. (www.nist.gov) (www.nist.gov) The Cloud Security Alliance's Cloud Controls Matrix offers **197 control objectives in 17 domains** as a broad checklist, but a buyer should select requirements relevant to this workload instead of treating the count as a quality score. (cloudsecurityalliance.org)

Strengths and Limitations

The main advantage is access to an already maintained application and a shorter route to a pilot when the documented workflow fits. The limits are the provider's schema, release schedule, commercial terms, and data portability. A purchase can still require extensive internal reviewer time. A licensing line item is therefore only part of buy-side total cost.

Commission Managed Research

Capabilities

Managed research assigns day-to-day target research to an external team or service while the PE firm specifies the mandate and accepts the resulting file. The deliverable should be explicit: company identity, each answer and status, source and date, reviewer decision, unresolved conflicts, and handoff format. A service may also support target-list construction or outreach, but outreach should remain a separate approved boundary. ILPA's due diligence questionnaire asks managers to identify third-party roles in sourcing, screening, and diligence, which makes responsibility mapping a useful governance exercise. (ilpa.org)

Docket's managed offer is a current first-party example of this model: it lists a dedicated analyst, describes a source, excerpt, and collection date for each researched answer, and states a three-month minimum. These are Docket's descriptions and terms, not independently measured accuracy or productivity. (docket.capital)

Adoption

The firm still needs a research owner to define questions, permitted data, rejection conditions, review cadence, and escalation decisions. Specify whether a provider may augment the target list, which fields require human review, and which outputs may enter the CRM. For any processing of personal data within the scope of UK data-protection rules, Information Commissioner's Office (ICO) guidance says controller-processor contracts specify the subject matter and duration of processing and include audit-related terms. This is jurisdiction-specific guidance; contract review must follow the relevant geography and data flow. (ico.org.uk) (ico.org.uk)

The service specification should make the following measurable:

- **Scope:** targets, fields, permitted sources and territories.
- **Turnaround:** start and acceptance clocks for each task.
- **Evidence:** the exact supporting material delivered per answer.
- **Review:** who checks a field before the client sees it.
- **Rework:** how rejected or changed answers are corrected.
- **Exit:** delivery format, rights and deletion steps.

Strengths and Limitations

A managed service can reduce the firm's need to operate the collection queue and may be useful when the volume is variable or internal research management capacity is scarce. It does not eliminate the need to test source licensing, evidence quality, confidentiality, or review burden. Service-level language should define the **accepted answer**, not merely the number of companies processed. SourceScrub, for instance, describes custom-research requests and says data requests are handled discreetly and remain available to the requesting team. That is one documented delivery feature, not an industry-wide rule. (www.sourcescrub.com)

Exit is especially important when people and software together create the output. The contract should identify the export package, time to deliver it, rights in client-provided and provider-derived material, and the treatment of retained copies. The ICO's UK guidance says a processor contract must provide for return or deletion of personal data at the controller's choice when the service ends. The exact obligation for a PE engagement depends on the parties' roles and applicable law. (ico.org.uk)

Combine Internal and External Components

Capabilities

A **hybrid** model assigns different parts of the workflow to different operators. The firm might own mandate logic, canonical company identity, accepted-answer criteria, and CRM relationships while using external data, a research platform, or managed review for collection. Another firm might buy the application but build its own evidence store and quality dashboard. These are design patterns, not empirical claims that hybrid delivery is cheaper. The split should be written as a responsibility map with one accountable owner for each handoff.

A hybrid model needs precise data contracts. A stable identity key must connect the source record, research file, and CRM object; every update should say whether it is a new observation, correction, or accepted answer. HubSpot documents an exports API for records and properties and says an export can contain associated record IDs. That makes identifiers an explicit export test, not an assumption about portability of the whole evidence package. (developers.hubspot.com) (developers.hubspot.com)

Adoption

Start with the **authoritative system** for each field. The CRM can own contact history; a research store can own sources and review decisions; the platform or service can own processing state. Then define the event that promotes a researched company to an approved outreach candidate. The route must include a person or team empowered to accept the answer and a way to reverse a mistaken merge. Salesforce documents that backup exports can provide organization data as CSV, yet its documentation also says derived formula fields are excluded from those exports. A hybrid architecture should test exact export semantics before promising reversibility. (help.salesforce.com) (help.salesforce.com)

Strengths and Limitations

Hybrid delivery preserves internal control over differentiating rules while outsourcing components that can be specified and tested. Its main limitation is interface complexity: multiple owners can duplicate records, diverge on definitions, or lose the link between answer and evidence. The controls are versioned schemas, contract tests, reconciliation reports, and a named owner for exceptions. The FinOps Foundation recommends combining billing data with operational telemetry for unit models; a hybrid PE workflow similarly needs cost and throughput data across each component to compare it honestly with the other options. (www.finops.org)

Feature Comparison

Table 1 compares the four operating models against the same decision criteria. Docket appears as two example rows because it directly offers self-serve and managed target research; the generic rows remain category descriptions. A listed capability is a starting claim for a pilot, not a measured score.

Model	Who controls mandate and review	Research and evidence path	Commercial or operating commitment	Exit test
Internal build	Firm defines logic, reviewers, releases and permissions.	Firm specifies source capture, observations, corrections and retention.	Engineering, data rights, operations and support are budgeted internally.	Rebuild a company file from database, attachments and event history.
Purchased platform	Firm configures within vendor-supported schema and roles.	Test field answers, source excerpts, reviewer decisions and exports.	License plus integration, review, security and change costs.	Export IDs, evidence and history into a neutral schema.
Docket self-serve, buy example	Firm manages criteria and results in the platform.	Vendor describes field-level sources and CSV/Excel export; published **$1,500/month**. (docket.capital)	Subscription price excludes the buyer's own review and integration.	Test the advertised export against the buyer's evidence fields.
Managed research	Firm defines acceptance; provider operates collection and often review.	Contract for source-level deliverables and rejection handling.	Service fee plus client oversight, integration and exit work.	Obtain complete deliverables and rights inventory before termination.
Docket managed, service example	Firm defines criteria; vendor describes a dedicated analyst.	Vendor describes source, excerpt and collection date per answer.	Published **$12,000/month** with a three-month minimum. (docket.capital)	Test delivery files, evidence and correction history.
Hybrid	Firm allocates each decision and handoff by contract.	Evidence store, provider and CRM may be separate.	Combined licenses, internal labor and integration maintenance.	Reconcile exports from every component by stable company ID.

The matrix shows why a single price comparison would mislead. A self-serve subscription transfers different work than a managed service, while an internal build's recurring staff and data costs may dominate its initial development. Docket's two rows document its published models and do not imply their prices are comparable to a fully loaded internal estimate. Each row needs the same acceptance tests before a cost ranking is meaningful.

Table 2 is a responsibility map to complete before signing a contract or approving a build. **A** means one accountable owner; **R** means the party that performs the task. The suggested allocations are a template for negotiation, not a claim about any named vendor's actual contract.

Work item	Build	Buy	Managed	Hybrid
Mandate and accepted-answer definitions	Firm A/R	Firm A/R	Firm A, provider R on agreed setup	Firm A/R
Data licenses and permitted uses	Firm A/R	Firm A, vendor R for its inputs	Firm A, provider R for its inputs	Firm A, each supplier R for its inputs
Collection and identity resolution	Firm A/R	Vendor R, firm A for acceptance	Provider R, firm A for acceptance	Assigned component R, firm A
Evidence retention and access rules	Firm A/R	Vendor R, firm A for policy	Provider R, firm A for policy	Firm A, named store operator R
Review, exceptions and corrections	Firm A/R	Firm A/R or documented vendor R	Firm A for acceptance, provider R for first review	Firm A, assigned reviewer R
CRM handoff and outreach approval	Firm A/R	Firm A/R	Firm A/R	Firm A/R
Export, deletion and exit rehearsal	Firm A/R	Firm A, vendor R for delivery	Firm A, provider R for delivery	Firm A, all component owners R

The **accountable firm** cannot delegate its investment mandate or approval boundary merely by buying capacity. The map also surfaces where two parties believe the other owns a missing field. ILPA asks managers to identify both technology tools and whether functions are performed internally or by third parties, a useful prompt to keep this map current. (ilpa.org)

Performance and Benchmarks

The relevant benchmark is a controlled pilot, not a vendor's database-universe claim. Create a frozen sample of target identities across obvious matches, ambiguous legal entities, duplicate candidates, and unanswered mandate questions. Give every option the same permitted source access or document the access difference. Freeze the question definitions, reviewer rubric, export schema, and time window. A pilot that lets one model use a paid data source while another cannot is a comparison of entitlements as well as operating models, and the results should be labeled accordingly.

Measure **accepted field answers**: answers a designated reviewer finds both supported by the retained material and responsive to the question. Also record attempted fields, unresolved fields, correction rate, reviewer minutes, elapsed time, research cost, and export completeness. The Public Company Accounting Oversight Board (PCAOB) defines appropriateness in audit evidence through quality and distinguishes directly obtained from indirectly obtained evidence. That standard is not a PE research regulation, but its concepts make a defensible review rubric: does the excerpt support this field, for this entity, at this date? (pcaobus.org) (pcaobus.org)

An automated citation or a confident answer is not an acceptance result. Research on attribution in generated text frames the test as whether a claim can be checked against an independent provided source. (aclanthology.org) A second reviewer should record whether an answer is accepted, corrected from retained material, or left unresolved. When sources conflict, keep both observations and record the resolution rule; PCAOB guidance treats inconsistencies among evidence sources as something to examine, not silently overwrite. (pcaobus.org)

The pilot should have **predeclared gates**. One reasonable gate is that every accepted field has a stable company ID, mandate version, source locator, retrieved excerpt, collection date, and reviewer decision. Another is that export reproduces these fields plus attachments and correction history. A third is that access roles, deletion, and an exit file are demonstrated in a test environment. These are proposed acceptance criteria, not measured industry norms. In one CRM example, Salesforce says Data Loader does not download the actual file content through the described route; that is why an attachment test must use the specific export method being purchased. (help.salesforce.com)

Finally, rerun part of the pilot after a changed mandate question or source-page update. The evaluation should include the cost of change, not only first-pass throughput. Record which party revises the rule, which records are queued for review, how previous accepted answers remain visible, and what the CRM receives. A system that is fast on a frozen list may be expensive to operate when a thesis changes each quarter; only a repeat run can reveal that in the buyer's environment.

Keep the pilot log in one shared format across all options:

- **Sample:** frozen company IDs and source entitlements.
- **Attempt:** each field researched, including no-answer outcomes.
- **Decision:** accepted, corrected, rejected or unresolved.
- **Effort:** analyst and reviewer minutes by task.
- **Cost:** fees, internal time and one-time setup separately.
- **Export:** counts and hashes reconciled after reimport.

Data Analysis and Evidence

Five-year scenario workbook

The following is an **editable planning template**, not a survey of PE costs. Copy *Table 3* into a spreadsheet and replace every bracketed input with a quote, internal loaded cost, measured pilot rate, or approved assumption. Use the same five-year horizon and target volume across models. GAO calls for explicit assumptions, sensitivity analysis, and later replacement with actual cost; the FinOps Foundation likewise asks users to document unit definitions and inclusions. (www.gao.gov) (www.finops.org)

Table 3 lists the workbook inputs and calculation rules. A separate row for each model and each year keeps license growth, staffing changes, and volume changes visible. All cash figures are nominal unless the firm supplies a discount rate; this template does not insert one.

Workbook line	Year 0	Years 1 to 5, separately	Editable calculation or source
One-time discovery and implementation	[input]	[new project input, else 0]	Include scoping, migration setup, configuration and training.
License or managed fee	[setup input]	[annual fee input]	Record contracted tiers, volume limits, renewals and growth.
Data rights and source access	[initial rights input]	[annual rights input]	Include permitted storage, derived use and export rights.
Internal engineering and operations	[setup hours × loaded rate]	[hours × loaded rate]	Include support, security and source changes.
Analyst review and cleanup	[pilot hours × loaded rate]	[field volume × review minutes ÷ 60 × loaded rate]	Use observed review minutes per attempted and accepted field.
Integration and change requests	[build input]	[annual change input]	Include CRM mapping, schema changes and reconciliation.
Downtime and risk reserve	[explicit reserve]	[explicit reserve]	State probability or reserve policy; avoid double counting.
Exit and migration	[none unless paid up front]	[forecast in intended exit year]	Include export, rebuild, deletion and parallel run.
Five-year total	**Year 0 subtotal**	**Sum of Years 1 to 5**	**Year 0 + Years 1 to 5; no present value without a supplied rate.**

To make the template operational, create one spreadsheet block for each model with **Year 0** and **Years 1 to 5** as separate rows. The following compact layout is copyable; every bracketed value is an input, and the last column is a formula copied down the rows. The review formula uses field volume, so record whether the field is attempted or accepted consistently.

```
model,year,one_time,license_or_service,data_rights,engineer_hours,engineer_rate,field_volume,review_minutes,
reviewer_rate,integration_change,risk_reserve,exit_cost,year_total
[model],0,[input],[input],[input],[input],[input],[input],[input],[input],[input],[input],
[input],=C2+D2+E2+F2*G2+H2*I2/60*J2+K2+L2+M2
[model],1,[input],[input],[input],[input],[input],[input],[input],[input],[input],[input],
[input],=C3+D3+E3+F3*G3+H3*I3/60*J3+K3+L3+M3
[model],2,[input],[input],[input],[input],[input],[input],[input],[input],[input],[input],
[input],=C4+D4+E4+F4*G4+H4*I4/60*J4+K4+L4+M4
[model],3,[input],[input],[input],[input],[input],[input],[input],[input],[input],[input],
[input],=C5+D5+E5+F5*G5+H5*I5/60*J5+K5+L5+M5
[model],4,[input],[input],[input],[input],[input],[input],[input],[input],[input],[input],
[input],=C6+D6+E6+F6*G6+H6*I6/60*J6+K6+L6+M6
[model],5,[input],[input],[input],[input],[input],[input],[input],[input],[input],[input],
[input],=C7+D7+E7+F7*G7+H7*I7/60*J7+K7+L7+M7
five_year_total,=SUM(N2:N7)
```

For each model, replace the year-zero and annual placeholders, enter a license-growth assumption in each later license cell, and move the exit cost to the year in which exit is planned. Keep base, low and high input versions for the sensitivity chart.

The same workbook can hold a quoted self-serve or managed fee without pretending the price includes internal review or data-use rights. As of September 2026, Docket's published monthly prices imply a **$90,000** self-serve subscription subtotal or **$720,000** managed-service subtotal over **60 months**, assuming unchanged monthly pricing and uninterrupted service. That multiplication is an illustration of recurring-fee arithmetic, not a five-year quotation or total cost of ownership. Its published managed minimum is shorter than the model horizon. (docket.capital)

Sensitivity tornado and unit economics

Run the model at a low and high value for each uncertain input while holding the others at the base case. The text chart below shows **illustrative mechanics only**; its bar lengths are placeholders until a firm's own inputs populate the workbook. Sort bars by the absolute change in five-year cost. Volume must flow into source charges, compute and review work where applicable; review minutes must flow into labor; engineering load into maintenance labor; license growth into contracted recurring fees.

```
Illustrative sensitivity tornado, replace bars with calculated five-year deltas
Lower total                 Base                  Higher total
Volume               [========] | [========]
Review minutes       [======]   | [======]
Engineering load     [====]     | [====]
License growth       [==]       | [==]
Delta for each side = recalculated five-year total minus base five-year total
```

For the pilot, **cost per accepted field answer = fully loaded pilot cost ÷ accepted field answers**. State the denominator and exclusions. A field answer that is later rejected belongs in attempted throughput and rework, not in the accepted denominator. FinOps guidance favors fully loaded unit metrics and explicitly identifies build-versus-buy decisions as a use for unit economics. (www.finops.org) (www.finops.org) Do not substitute cost per deal: a closed acquisition depends on many factors outside target-research production, and a short pilot cannot estimate that outcome.

Implications and Future Directions

Portability should be proved before it is needed. Ask for a bulk export with stable company IDs, field definitions, source URLs and excerpts, attachment files, collection and decision timestamps, rejected or superseded values, mandate versions, and relationship keys. Reimport a sample into a neutral schema and compare counts and hashes. HubSpot's exports documentation says a completed download URL expires after five minutes; a practical exit runbook should therefore include the timing and authentication needed to collect files, not just a statement that an API exists. (developers.hubspot.com)

For PE origination, **data control** means usable rights and a reproducible record, not merely access to an export button. Terms matter as much as technology. Distinguish client-supplied records, public records, licensed third-party fields, provider-derived judgments, and personal data. OpenCorporates' API documentation explicitly separates source provenance from licensing and warns against inferring unrestricted rights from a null license field. (api.opencorporates.com) (api.opencorporates.com) For applicable European Union cloud-switching arrangements, the European Commission describes machine-readable export obligations under the Data Act, but jurisdiction, service scope, intellectual-property exclusions, and contract terms require specific review. (digital-strategy.ec.europa.eu) A useful exit package is one that the firm can lawfully retain and operationally reuse.

Security review should be evidence-seeking. NIST recommends asking producers what secure-development information they can provide and connecting high-level claims to lower-level artifacts. (www.nist.gov) (www.nist.gov) Cloud Security Alliance portability guidance calls for termination provisions covering data access, format, retention, and deletion. (cloudsecurityalliance.org) A firm should test role access, administrator actions, audit logs, backups, retention, and export in the same pilot as field accuracy, with a named owner for each finding.

The choice is not permanent. Review the operating model when a mandate changes, a license renews, data rights change, a provider updates its export route, or internal volume moves outside the initial scenario. Compare actual spend and accepted-answer output with the baseline, then update the cost model. The ICO also advises controllers to monitor processor compliance on an ongoing basis where its UK guidance applies. (ico.org.uk) The decision record should preserve the rejected options and the assumptions that would make them attractive later.

Frequently Asked Questions (FAQs)

Is private equity deal origination software the same as an in-house research team?

No. Software can route tasks, hold data, score targets, and export records. A team still decides which questions matter, checks difficult evidence, and owns acceptance. The buy-versus-build calculation should price those retained responsibilities in both cases. ILPA's sourcing questions address staffing and technology separately. (ilpa.org)

When should a firm outsource acquisition target research?

Consider managed delivery when the work can be specified as field-level deliverables and the firm has capacity to review them but not to run every research queue. Test the provider on the same target sample and export format used for internal or software-led options. Contract for unanswered fields and corrections as well as completed profiles.

Which vendor-evaluation questions matter most?

Ask to demonstrate one difficult target end to end: identity, mandate-specific answer, source retention, reviewer action, CRM handoff, bulk export, and deletion. Request current security artifacts and exact data-license terms. NIST's purchaser guidance provides a structured way to ask software producers about development practices. (www.nist.gov)

Conclusion

A PE origination team should choose its research operating model by **required output, lifecycle cost, and control**. First define the target record and acceptance rule. Then obtain a technical baseline and comparable five-year cost inputs for build, buy, managed, and hybrid designs. Finally, run a common pilot that measures accepted field answers, review effort, export completeness, and the cost of a changed mandate.

Internal build gives the firm the most room to shape its process but also assigns it maintenance and data-rights work. Buying can make a pilot faster when a platform's schema and controls fit. Managed delivery can move collection work outside the firm while preserving the firm's need to define and accept the research. Hybrid delivery can allocate each component to the party best equipped to run it, provided the interfaces and responsibilities are explicit.

The result is a **reviewable decision**, not a permanent label. Keep the assumptions, pilot evidence, responsibility map, and exit test together. Recalculate with actual costs and accepted-answer output at agreed review dates. A model that wins only because labor, licensed data, or migration was omitted has not yet passed the comparison.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://ilpa.org/wp-content/uploads/2021/11/ILPA-DDQ-2.0.pdf
2. https://www.gao.gov/products/gao-20-195g
3. https://docket.capital/
4. https://api.opencorporates.com/documentation/API-Reference
5. https://www.finops.org/framework/capabilities/unit-economics/
6. https://www.nist.gov/itl/executive-order-14028-improving-nations-cybersecurity/software-supply-chain-security-guidance-7
7. https://cloudsecurityalliance.org/blog/2025/06/13/implementing-ccm-interoperability-portability-controls
8. https://www.affinity.co/guides/affinitys-best-practice-guide-for-deal-sourcing
9. https://docket.capital/articles/pe-target-record-entity-identifiers
10. https://www.w3.org/TR/prov-o/
11. https://www.pewresearch.org/data-labs/2024/05/17/when-online-content-disappears/
12. https://docket.capital/articles/private-company-data-providers-pe
13. https://docket.capital/evidence-standards
14. https://developer.company-information.service.gov.uk/developer-guidelines
15. https://www.thoughtworks.com/insights/e-books/build-versus-buy-strategic-framework-for-evaluating-third-party-solutions
16. https://www.sourcescrub.com/solutions/private-equity
17. https://www.sourcescrub.com/product/integration-and-delivery
18. https://www.intapp.com/dealcloud/

19. https://docket.capital/articles/crm-research-database-evidence-store
20. https://docket.capital/articles/acquisition-target-research-freshness
21. https://www.intapp.com/dealcloud/data-management/?dc=1
22. https://www.nist.gov/itl/executive-order-14028-improving-nations-cybersecurity/software-supply-chain-security-guidance-11
23. https://cloudsecurityalliance.org/research/cloud-controls-matrix
24. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/contracts-and-liabilities-between-controllers-and-processors-multi/what-needs-to-be-included-in-the-contract/
25. https://www.sourcescrub.com/product/custom-research
26. https://developers.hubspot.com/docs/api-reference/legacy/crm/exports/guide
27. https://help.salesforce.com/s/articleView?id=sf.admin_exportdata.htm&language=en_US&type=5
28. https://www.finops.org/framework/technology-categories/public-cloud/
29. https://pcaobus.org/oversight/standards/auditing-standards/details/AS1105
30. https://aclanthology.org/2023.cl-4.2/
31. https://help.salesforce.com/s/articleView?id=000382669&language=en_US&type=1
32. https://digital-strategy.ec.europa.eu/en/factpages/data-act-explained
33. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/contracts-and-liabilities-between-controllers-and-processors-multi/responsibilities-and-liabilities-for-controllers-using-a-processor/

THE PUBLISHER

About Docket

Docket provides deal-origination research software and managed research for private equity firms. We help investment teams investigate acquisition targets using structured screening criteria, retained sources and reviewable company evidence. Teams can work through a self-serve platform or use managed research, depending on how they want research delivered.

Research against a defined mandate

A useful target list needs more than company names. Docket focuses on the questions that determine whether a company fits an investment mandate, including the evidence needed to support or qualify each answer. Our research approach makes the connection between screening criteria, source material and conclusions visible to the team reviewing the work.

Triage, Scout and Audit

Docket's three named research agents perform complementary tasks. Triage resolves company identity and screens fit. Scout collects sourced answers against the mandate. Audit checks retained evidence, addresses contradictions and leaves unsupported answers visibly unresolved. This structure helps reviewers distinguish established findings from missing information and questions requiring further investigation.

Evidence that supports investment-team judgment

Our research library covers market mapping, screening criteria, private-company data, succession and ownership, source evaluation and evidence standards. These resources explain the methods and limitations behind origination research. Findings support a team's judgment; they do not establish that a company is for sale or guarantee a transaction or investment outcome.

Work with Docket

Visit Docket to explore the platform, managed research and the current contact path. Read about screening criteria, evidence standards, AI research agents and market mapping.

Public examples are illustrative unless explicitly identified otherwise. Research preparation and authorized outreach are separate activities; confidential target lists and customer outcomes should never be inferred from an educational example.

Website: https://docket.capital