Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

Thesis-Driven Origination: From Thesis to Target List

> 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 · 2026-10-29 · thesis driven private equity · investment thesis private equity · thesis-driven origination · deal origination strategy

Original article: https://docket.capital/articles/thesis-driven-origination-from-thesis-to-target-list



In this article

1. From thesis statement to testable claims
2. Why specialization shapes origination
3. Mapping each criterion to a data source
4. Defining the universe with NAICS
5. Sampling and calibration
6. Logging exclusions
7. Adjacency and add-on logic
8. Worksheet: thesis-to-list template
9. Sources

From thesis statement to testable claims

An investment thesis usually arrives as a sentence: "fragmented recurring-service businesses in the Southeast, where a platform can professionalize the back office and add route density." That sentence is a reason to look, not a list. Thesis-driven origination turns it into claims a researcher can check company by company, with a record of which companies passed, failed or stayed unresolved.

First, decompose: pull out every claim the thesis makes about a company that would belong on the list. Each becomes one of three things:

- **A hard disqualifier**: a condition that removes the company regardless of anything else (wrong activity, already sponsor-owned, outside the geography).
- **A scored criterion**: a condition that ranks companies that survive the disqualifiers (share of recurring work, number of locations, owner tenure).
- **A research question**: something the thesis depends on but that public evidence cannot settle at the screening stage (margin profile, contract terms, willingness to sell). These stay open and go to the first conversation, not into the filter.

The test of a well-written claim is whether two researchers given the same company and the same sources would reach the same answer. "Strong recurring revenue" fails that test. "Publishes a scheduled maintenance or inspection plan on its own website, retained and dated" passes it, because it names the observable, the source and the evidence standard. It also implies what it does not prove: a published plan is not contracted revenue.

Thesis fragment	Claim	Type	Observable	Value when unknown
"recurring-service businesses"	Company performs a service on a repeat schedule	Scored	Service-plan or inspection-cycle language on the company's own site	Pending, not excluded

Thesis fragment	Claim	Type	Observable	Value when unknown
"fragmented"	No single operator dominates the defined market	Universe-level, not company-level	Count and size distribution of candidates in the defined universe	Thesis-level open question
"in the Southeast"	Operates from a location in the named states	Hard disqualifier	Registered office or operating address	Pending until resolved
"platform can professionalize the back office"	Target is founder-run and lightly systemized	Research question	Not reliably observable from public data	Carried to first call
"add route density"	Service area overlaps or adjoins a platform's	Scored (add-on lists only)	Service-area pages, location addresses	Pending

Some thesis claims are about the market, not any single company; they belong to universe definition, not the screen. Others cannot be screened at all; writing them down as research questions keeps them from becoming proxies that look precise and are not. Docket's thesis notes are written in this form: each states a mechanism, the evidence to seek, the reasons to stop and the questions for the target list.

Criteria design itself (weights, anchors, missing data) is covered in building consistent acquisition screening criteria; this article covers what goes into the criteria and how the list is built and documented.

Why specialization shapes origination

The most-cited evidence on how sponsors organize comes from Gompers, Kaplan and Mukharlyamov's survey of 79 private equity investors. In the 2015 NBER working-paper version, the authors report that only 37% of respondents are organized as generalists and that "more than 50% of the firms are organized by industry" (Gompers, Kaplan & Mukharlyamov, NBER w21133, 2015). They suggest three advantages: specialists are more likely to find an opportunity in their industry, to evaluate it when it appears, and to understand whether and how value can be added (NBER w21133). The paper was later published in the Journal of Financial Economics (2016); the figures here are from the openly available working paper, because the published version is closed access.

The same survey describes the funnel a thesis feeds. For every hundred opportunities considered, the average respondent deeply investigates fewer than 24, signs an agreement with fewer than 14 and closes on 6 (NBER w21133). Respondents described almost 36% of closed deals as "proactively self-generated" and about 33% as investment-bank generated, and considered almost 48% of closed deals proprietary in some way; the authors state they have no way to validate what "proprietary" means or how far the deals truly were (NBER w21133). These are 2015 self-reports: useful for shape, not a benchmark.

If roughly six in a hundred considered opportunities close, the thesis decides which hundred get considered. A vague thesis fills the funnel with companies that fail on facts knowable at the start.

There is a peer-reviewed literature on whether specialization confers an advantage in private equity, for example Cressy, Munari and Malipiero's study in the Journal of Corporate Finance (2007). We cite it here to point to that literature, not for a specific finding: no abstract or open full text was available to review for this

article.

Mapping each criterion to a data source

Each claim needs a source that can actually answer it, not merely one that is available. Build the map before searching.

Criterion type	Typical public source	What it can support	What it cannot support
Industry / activity	Census industry statistics (SUSB, by NAICS); the company's own site	Universe boundaries; whether the company describes the activity	That the activity is the company's main revenue line
Employer size band	Census SUSB (aggregate only)	How many employer establishments sit in an industry, geography and enterprise-size cell	Any individual company's size
Legal identity and status	State business registries	That a legal entity exists, its status and filing history	Revenue, ownership percentages in most states
Government customer exposure	SAM.gov registrations, USAspending awards	Registration and federal award history for a resolved entity	Commercial revenue or total revenue
Licensed activity	State licensing boards	That a licence exists for a named holder, scope and expiry	That the licence holder is the operating company you mean, without a crosswalk
Growth signals	Job postings, location openings	That the company advertised roles or opened sites	Hires, headcount or revenue growth
Ownership	Registries, press releases, sponsor portfolio pages	Positive evidence of sponsor or corporate ownership	Absence of a sponsor (no evidence is not evidence of independence)

Two rules make the table usable. First, record for each criterion the entity level the source speaks to: establishment, legal entity or brand. SUSB counts establishments with paid employees (Census SUSB); a registry counts legal entities; a website speaks for a brand. Joining them without an identity step produces duplicates and false matches. Second, record what a negative result means. "Not found in the licensing database" is a research status, not a disqualifier, until the search is shown to cover the right name and jurisdiction.

Defining the universe with NAICS

The universe is the set of companies the thesis applies to before company-level screening; defining it makes "we reviewed the market" checkable.

NAICS is the standard federal statistical agencies use to classify business establishments; it is production-oriented, grouping establishments by similarity of production process, and hierarchical, from 2-digit sectors through 3-digit subsectors, 4-digit industry groups and 5-digit industries to 6-digit national industries (Census

Bureau, Understanding NAICS). The 2022 Economic Census is published on the 2022 NAICS vintage (Census Bureau), so record the vintage alongside every code list.

The Census Bureau's Statistics of U.S. Businesses (SUSB) is the usual sizing layer. It is an annual series covering U.S. establishments with paid employees, reporting firms, establishments, employment and annual payroll by establishment industry and enterprise size, at U.S., state, metropolitan area, county and congressional-district level; the latest data year shown on the program page is 2022 (Census SUSB). Three consequences for a thesis universe:

1. **It excludes non-employers.** A thesis aimed at owner-operators with no payroll is not described by SUSB.
2. **It lags.** A 2022 data year sized in late 2026 describes the market four years ago.
3. **It is aggregate.** SUSB tells you how deep a cell is, not which companies are in it. The company list comes from other sources.

A workable sequence:

1. Start from five to ten seed companies that clearly fit the thesis.
2. Find how each is classified, and take the union of codes rather than picking the one that sounds right.
3. Set geography as a list of states, metro areas or counties, matching the SUSB geography you will size against.
4. Set a size band as a proxy you can observe (employment band, location count), not as revenue.
5. Write the universe definition as one record: code list with vintage, geography, size band, date.

NAICS codes bound the search; they rarely match a colloquial niche, and self-reported codes drift. Failure modes and mitigations are covered in mapping a fragmented industry from public data.

Sampling and calibration

Before running the full screen, calibrate on a sample.

1. **Draw a random sample from the universe**, not from the companies the screen passes. If you sample only passes, you measure precision but never see what the screen wrongly removed.
2. **Have a senior reviewer judge each sampled company against the thesis by hand**, without seeing the screen's result.
3. **Run the screen on the same sample** and cross-tabulate.

	Reviewer: fits thesis	**Reviewer: does not fit**
Screen: pass	True pass (A)	False pass (B)
Screen: exclude	False exclusion (C)	True exclusion (D)

Two ratios matter:

- **Precision of passes** = A ÷ (A + B). Low precision fills the outreach list with companies the deal team will reject.
- **Miss rate among fits** = C ÷ (A + C). A high miss rate means the screen is removing the companies the thesis is about, usually because a disqualifier keys on an observable that fits miss (a website phrase, a code, a licence type).

Illustrative, fictional numbers: a team samples 60 records from its universe. The reviewer judges 18 to fit. The screen passes 22, of which 14 are reviewer fits; it excludes 4 fits. Precision is 14 ÷ 22 = 64%; miss rate is 4 ÷ 18 = 22%. Reading the four misses shows all were excluded by a "publishes service plans" criterion because they describe recurring work in other words. The team changes that criterion from a disqualifier to a scored item and re-runs the sample.

Sample sizes for a pass/fail acceptance test of a research batch are a separate question, with exact zero-defect tables in the target audit sampling worksheet. For calibration the aim is narrower: find which criterion is doing the damage, change it once, and record the version.

Logging exclusions

Every company removed from the universe should carry a reason code, the criterion version that removed it and the evidence. When the thesis changes, a list without reasons must be re-screened from scratch; a list with reasons can be re-run on the affected codes only.

Code	Meaning	Evidence required	Reversible when
`ACT`	Activity outside the thesis	Retained page showing the actual activity	Thesis boundary widens
`GEO`	Outside defined geography	Address evidence for operating locations	Geography changes
`SIZE`	Outside the size band on the stated proxy	Proxy value, source and date	Band or proxy changes
`OWN`	Already sponsor- or corporate-owned	Positive ownership evidence	Ownership changes
`DUP`	Duplicate of another record	ID of the surviving record	Never; merge instead
`ID`	Identity unresolved	What was tried	Identity resolved
`UNK`	Criterion unknown	The open question	Evidence found

Keep `ID` and `UNK` out of the excluded pile; they are pending states. A screen that counts unknowns as failures is biased against companies with the thinnest public footprint, often the ones a proactive approach is looking for.

Minimum fields per exclusion row: record ID, legal name as resolved, reason code, criterion and version, evidence URL, date retrieved, reviewer, and a free-text note limited to what the evidence shows.

Adjacency and add-on logic

A buy-and-build thesis needs a boundary rule for what counts as "the same business". Bain defines buy-and-build as an explicit strategy of using a well-positioned platform to make at least four sequential add-on acquisitions of smaller companies, and frames add-on diligence around one question: how much overlap there is between the two companies' customers, costs, channels, capabilities and competition (Bain, Global Private Equity Report 2019). In the same piece, a roll-up assumes significant, if not total, overlap; close-in adjacencies have less, and two-step adjacencies have significant differences (Bain).

That lens converts into a boundary worksheet. Rate each candidate segment against the core on the five dimensions:

Dimension	Question to answer from evidence	Rating
Customers	Same buyer type and purchase decision?	High / partial / low
Costs	Same labour pool, equipment and cost structure?	High / partial / low
Channels	Reached through the same sales route?	High / partial / low
Capabilities	Same licences, skills, certifications?	High / partial / low
Competition	Do the same competitors show up?	High / partial / low

Segments rated high on most dimensions belong in the core universe. Segments with mixed ratings are adjacencies: list them separately, with their own universe record. Bain also notes that successful buy-and-build strategies target fragmented industries with sufficient acquisition targets of the right size (Bain), which makes universe depth in the right size band a thesis-level test to run before company-level work.

Worksheet: thesis-to-list template

The example below is entirely fictional and shows the shape of the record, not a market fact.

Thesis (illustrative): Recurring commercial inspection-and-cleaning services for restaurant kitchen ventilation in three neighbouring states, where a platform could consolidate scheduling and compliance paperwork.

Field	Entry (illustrative)
Thesis version	v1.0, dated
Mechanism	Shared scheduling and compliance reporting across route-based crews
Hard disqualifiers	Activity is not inspection/cleaning; no operating location in the three states; sponsor- or corporate-owned with positive evidence
Scored criteria	Evidence of scheduled service cycles; number of operating locations; years operating under current legal entity
Research questions (first call)	Share of revenue under recurring arrangements; owner's role in sales; customer concentration

Field	Entry (illustrative)
Universe codes	Union of codes from 8 seed companies, vintage recorded
Geography	Three named states
Size proxy	2 to 6 operating locations or 10 to 75 employees, whichever is observable
Sources per criterion	Company site (activity, cycles), state registry (identity, status), licensing boards where applicable (capability), ownership searches
Overlap rating of adjacent segment "grease-trap services"	Customers high, costs partial, channels high, capabilities low, competition partial: listed as an adjacency, not core

Fictional funnel, labelled as such:

Step	Records (fictional)	Note
Raw candidates from all sources	1,140	Before identity resolution
After de-duplication and identity resolution	870	270 `DUP`; 41 left as `ID` (pending)
After hard disqualifiers	312	Each exclusion coded and evidenced
Calibration sample changed one criterion	338	26 restored from `ACT` after criterion v1.1
Scored, top tier	95	Goes to the outreach approval step
Pending (`ID` + `UNK`)	41 + 63	Owned research questions, not exclusions

Checklist before the list leaves research:

- [ ] Thesis decomposed into disqualifiers, scored criteria and research questions, each with an observable.
- [ ] Each criterion mapped to a source, with the entity level and the meaning of a negative result.
- [ ] Universe record written: codes with vintage, geography, size proxy, date.
- [ ] Calibration sample drawn from the universe, cross-tab recorded, criterion changes versioned.
- [ ] Every exclusion carries a code, version and evidence; pending records counted separately.
- [ ] Adjacencies listed separately with their own overlap ratings.

Docket's Triage, Scout and Audit agents take a client's criteria as their questions and return sourced answers with the retained page behind each one.

Sources

1. Paul Gompers, Steven N. Kaplan & Vladimir Mukharlyamov, "What Do Private Equity Firms Say They Do?" — NBER Working Paper 21133 (2015), https://www.nber.org/papers/w21133.pdf, accessed 2026-10-28.

2. Paul Gompers, Steven N. Kaplan & Vladimir Mukharlyamov, "What do private equity firms say they do?", Journal of Financial Economics 121(3), 449-476 — Elsevier (2016), https://doi.org/10.1016/j.jfineco.2016.06.003, accessed 2026-10-28 (closed access; bibliographic record confirmed via Crossref).
3. Robert Cressy, Federico Munari & Alessandro Malipiero, "Playing to their strengths? Evidence that specialization in the private equity industry confers competitive advantage", Journal of Corporate Finance 13(4), 647-669 — Elsevier (2007), https://doi.org/10.1016/j.jcorpfin.2007.04.007, accessed 2026-10-28 (bibliographic record confirmed via Crossref and the University of Birmingham repository; cited for existence only).
4. "Understanding NAICS" (2022 Economic Census guidance) — U.S. Census Bureau (2022 vintage), https://www.census.gov/programs-surveys/economic-census/year/2022/guidance/understanding-naics.html, accessed 2026-10-28.
5. "Statistics of U.S. Businesses (SUSB)" — U.S. Census Bureau (page revised 2025), https://www.census.gov/programs-surveys/susb.html, accessed 2026-10-28.
6. "Buy-and-Build: A Powerful PE Strategy, but Hard to Pull Off", Global Private Equity Report 2019 — Bain & Company (2019), https://www.bain.com/insights/buy-and-build-global-private-equity-report-2019/, accessed 2026-10-28.

Source links

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

1. https://www.nber.org/papers/w21133.pdf
2. https://doi.org/10.1016/j.jfineco.2016.06.003
3. https://doi.org/10.1016/j.jcorpfin.2007.04.007
4. https://www.census.gov/programs-surveys/susb.html
5. https://www.census.gov/programs-surveys/economic-census/year/2022/guidance/understanding-naics.html
6. https://www.bain.com/insights/buy-and-build-global-private-equity-report-2019

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

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.