Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

Conflicting Data Sources on a Private Equity Target Ledger

> 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 · 2026-10-27 · conflicting data sources private equity target · claim adjudication ledger · due diligence data reconciliation · source conflict resolution deal sourcing · field-level data provenance · private company data discrepancies · acquisition target research · beneficial ownership verification · inter-reviewer agreement

Original article: https://docket.capital/articles/conflicting-sources-pe-target-claim-ledger



In this article

1. Executive Summary
2. Introduction and Background
3. Define the Claim and Preserve Every Observation
4. Diagnose the Conflict Type
5. Field-Specific Authority Ladders and Recency Rules
6. Dispositions, Second Review and Escalation
7. A Fictional Three-Source Example (Hypothetical Example)
8. Reference Asset: Conflict Ledger Schema
9. Data Analysis and Evidence
10. Implications and Future Directions
11. Frequently Asked Questions (FAQs)
12. Conclusion

Executive Summary

When a state registry, a company website and a commercial data vendor disagree about an acquisition target, the useful question is not "which source is right?" It is "which claim, about which entity, as of which date, under which field definition, is each source making, and what does the mandate do with the answer?" This report proposes a field-level **claim adjudication ledger** that answers that question in a form a second analyst can audit. Each observation is stored as its own row with a literal value, source excerpt, effective date, retrieval date and jurisdiction, and each conflict ends in one of six dispositions: **accepted, corroborated, issuer-stated, conflicted, unknown or not applicable**.

The case for field-specific rules is documented by the sources themselves. The W3C PROV Recommendation defines provenance as information about the entities, activities and people involved in producing data, usable to assess its quality, reliability or trustworthiness (www.w3.org). The U.S. Securities and Exchange Commission (SEC) states that it cannot guarantee the accuracy of its Form D data sets, which rest on information provided by individual registrants (www.sec.gov), and that they are not a substitute for the filings (www.sec.gov). The Library of Congress warns that private-company information can be "very dated and often just wrong" (guides.loc.gov). A registry can show that an entity exists yet say nothing about who owns it: Delaware states that alternative entities are not required to list members or managers (corp.delaware.gov).

Most apparent conflicts are definitional, not contradictions. County Business Patterns counts employees on the payroll in the pay period including March 12 (www.census.gov), while the Small Business Administration (SBA) averages over the preceding 24 completed months (www.ecfr.gov) and includes affiliates (www.ecfr.gov). An establishment is classified on its major activity (www.census.gov), but an enterprise follows the sector with the largest payroll share (www.census.gov). A registered office is an address for statutory mail (companieshouse.blog.gov.uk), not a site of operations. Only after such differences are normalized should a residual disagreement be labeled "conflicted."

The protocol has five working rules: keep rejected alternatives with reasons; test independence before counting agreement, because copied values add no evidence (lunadong.com); require a second reviewer for any claim that can flip a mandate pass/fail result or outreach eligibility; time-box escalation and set a recheck trigger; and never let a weighted score hide an unresolved disqualifier. A fictional three-source example shows the before and after claim record, and an optional inter-reviewer agreement rate (matching decisions divided by double-reviewed conflicts) is defined with its denominator. No accuracy benchmark is asserted, because none is published for this task as of October 2026.

Introduction and Background

Private-equity origination teams screen long lists of privately held companies using public and commercial data. The Library of Congress guide notes that, unlike public companies, private companies are not required to file with the SEC (guides.loc.gov), and that for many it is hard to find information beyond name and location (guides.loc.gov). The evidence that does exist is scattered across state registries, national registers, exempt-offering notices, statistical agencies, company websites and vendor databases, produced by different parties, for different purposes, on different dates and definitions. Disagreement among them is the normal condition of the task, not an anomaly.

This report addresses one operational problem: how an analyst, research manager or reviewer should decide whether an ownership, location, employee-count or industry-fit field is accepted, rejected or left unresolved when public sources disagree. It does not rank databases or declare any source universally authoritative. Authority belongs to a source for a specific field at a specific date. The sections that follow cover diagnosis of target data discrepancies, field-specific authority, dispositions and escalation, a claim adjudication ledger schema for due diligence, a revenue subsection, a worked example and an FAQ.

Disclosure: Docket, which provides sourced target research for private equity firms, publishes this article. The protocol below can be run in a spreadsheet with any combination of sources. The article describes no real acquisition target or client engagement, and all examples are fictional and labeled.

Four terms are used throughout. A **claim** is one assertion about one field of one entity. An **observation** is one source's statement bearing on a claim. A **disposition** is the ledger's recorded decision. A **disqualifier** is a claim that, if true, fails the mandate regardless of other results. These map onto the W3C PROV vocabulary, in which an entity is "a physical, digital, conceptual, or other kind of thing" (www.w3.org), an activity "occurs over a period of time" (www.w3.org), and attribution is "the ascribing of an entity to an agent" (www.w3.org). In ledger terms the observation is the entity, retrieval and extraction are activities, and the registry, vendor or analyst is the agent.

Define the Claim and Preserve Every Observation

Define the disputed claim first

A conflict cannot be adjudicated until the claim is specific. Fix four things in writing before comparing sources:

- **Entity.** Legal entity, brand, facility or parent.
- **Field.** Ownership, location, employee count, industry fit or another.

- **As-of date.** The date the claim is meant to describe.
- **Mandate consequence.** Pass/fail, outreach eligibility, or neither.

Keeping legal entity, brand, facility and parent claims separate prevents the most common false conflict, in which two sources describe different layers of one corporate family.

Entity matching is a documented discipline. A survey in ACM Computing Surveys defines entity resolution as identifying different entity profiles that describe the same real-world object (arxiv.org), and Census Bureau record-linkage research describes the aim as determining the pairs of records associated with the same entity (census.gov). The classic Fellegi-Sunter framework reserves a middle band, the "no-decision region or clerical review region," for human judgment (census.gov). A ledger should mirror that: each source row is linked to a canonical entity ID as a match, non-match or "possible match," and a possible match is never silently counted as corroboration.

Store each observation as its own row

Preserve every observation rather than overwriting it with a merged value. Each row records the literal value, source URL, short excerpt, publication or filing date, retrieval date, jurisdiction, the entity name exactly as the source prints it, and the source type. A reviewer can only audit a decision if the inputs survive. The ELIXIR research data management kit says well-documented provenance is essential for assessing authenticity, credibility, trustworthiness and quality (rdmkit.elixir-europe.org) and defines lineage as tracing the movements and changes of the data (rdmkit.elixir-europe.org). The FAIR principles include the requirement that data are associated with detailed provenance (www.nature.com), and GO FAIR says such provenance helps people and machines assess whether a resource meets their criteria for intended reuse (www.go-fair.org), which is the screening use case.

Several different dates attach to any observation, and Dublin Core Terms keeps them apart: a date of creation (www.dublincore.org), a date of formal issuance (www.dublincore.org), a date of validity, "often a range" (www.dublincore.org), and a date on which the resource was changed (www.dublincore.org). In company research these become the effective date (when the fact was true), the source's publication or filing date, the vendor's refresh date and the analyst's retrieval date. A filing date is not an effective date: the Form D instructions define the date of first sale as when the first investor is irrevocably contractually committed to invest (www.sec.gov), a different moment from the date the notice is submitted.

Record each source's role at ingestion

The Library of Congress guide advises analysts not to underestimate what companies publish on their own websites (guides.loc.gov), and also cautions that information can be very dated and often just wrong (guides.loc.gov). A company website is strong evidence of how the company describes itself and weak evidence of how a registry classifies it, so the ledger records the source's role (issuer statement, official record, derived aggregate) at capture.

Diagnose the Conflict Type

Not every disagreement is a contradiction. Data-integration research separates contradictions (different non-null values for one property) from uncertainties (a value in one source and nothing in another). An early data-fusion report notes that a "take the information" strategy is only reasonable if there are only uncertainties and no conflicts (edoc.hu-berlin.de), and lists outdated values as another reason for contradictions between sources (edoc.hu-berlin.de). Six classes extend that distinction to company screening.

- **Different entities.** The sources describe a parent, subsidiary, brand or facility. Resolution is an entity-mapping decision.
- **Different dates.** Both values were true at different times. Align effective dates and prefer the later effective value, subject to the field's recency rule.
- **Different field definitions.** The sources measure different things under one label. Restate on one definition or record the values as non-comparable.
- **Transcription, aggregation or copy error.** One value derives from the other or is wrong on its face. Trace the originator and discard the derivative.
- **Issuer statement versus official record.** The company's statement differs from a registry or filing. The outcome depends on what the record can prove for that field.
- **True contradiction.** Same entity, date and definition, independent sources, different values. Only this class should end as "conflicted."

Table 1 crosses these classes with the six adjudication steps in this protocol. "Primary" means the step usually resolves that class, "Check" means it must be run but rarely resolves the class alone, and "None" means it can be skipped.

Conflict class	1. Entity match	2. Align dates	3. Normalize definition	4. Test independence	5. Apply field authority	6. Disposition and review
Different entities	Primary: map each row to a canonical ID	Check	Check	None	None	Record mapping and any "possible match"
Different dates	Check	Primary: order by effective date	Check	None	Check: recency rule	Both rows kept; later effective value leads
Different field definitions	Check	Check	Primary: one definition or mark non-comparable	None	Check	"Not applicable" or "unknown" if no common basis
Transcription, aggregation or copy error	Check	Check	Check	Primary: trace the originator	Check	Reject the derivative, keep it with a reason
Issuer statement vs official record	Check	Primary: filing date is not effective date	Check	Check	Primary: what the record can prove	"Issuer-stated" unless corroborated

Conflict class	1. Entity match	2. Align dates	3. Normalize definition	4. Test independence	5. Apply field authority	6. Disposition and review
True contradiction	Check	Check	Check	Check	Check	Primary: "conflicted"; second review if mandate-changing

Steps 1 to 3 are cheap and eliminate most apparent conflicts, step 4 prevents false corroboration, and only the last row reaches an unresolved disposition. Teams that start at step 5 (choosing a winner) record disagreements that were never real.

Why independence must be tested

Counting agreement is meaningful only when sources are independent. Research on integrating conflicting web data observes that a value from one source, true or false, can be copied by many others (lunadong.com), so a false value can be spread through copying (lunadong.com). A survey of truth-discovery methods states the converse: sources sharing common mistakes are not likely to be independent (arxiv.org). In a ledger this becomes a flag on each row (original, copied, unknown), and copied rows add no corroboration. A vendor profile whose owner field reproduces a registry extract is one observation, not two. Deciding once, in advance, whom to trust for all values (edoc.hu-berlin.de) is the "one database is authoritative" rule: cheap, but unable to detect that the trusted source is stale for this particular field.

Field-Specific Authority Ladders and Recency Rules

No universal ranking of sources survives contact with specific fields. ISO/IEC 25012 states that data quality characteristics will be of varying importance and priority to different stakeholders (www.iso.org), and the U.S. Office of Management and Budget treats objectivity as a measure of whether disseminated information is accurate, reliable and unbiased (obamawhitehouse.archives.gov). A ledger needs a short authority ladder per field, written before research begins: source types from strongest to weakest, the recency rule, and what the strongest source cannot prove.

Ownership and control

Ownership is where official sources are most often silent. Delaware states that alternative entities are not required to list members or managers (corp.delaware.gov) and that shareholder names of a corporation are not on file with the Division (corp.delaware.gov). The South Carolina Secretary of State says it does not certify the authenticity of site information because it originates from third parties (businessfilings.sc.gov). A registry record establishes existence and filed particulars, not ownership.

Registers designed to show control have a defined scope. UK Companies House describes a person with significant control as someone who, among other tests, has more than 25% of shares or voting rights (www.gov.uk) or can appoint or remove a majority of directors (www.gov.uk), and a company that cannot identify such a person, or has none, must tell Companies House (www.gov.uk). A "none identified" entry is a statement about identification, not proof that no owner exists. The regime is changing: guidance states that

identity verification is now a legal requirement (www.gov.uk). Companies House also says occasional inaccuracies occur and are corrected when drawn to its attention (www.gov.uk), and that an annotation could be placed on the register (www.gov.uk), so a UK entry should be read with its filing date and any annotation in view.

For U.S. targets, no federal beneficial-ownership register is available as evidence as of October 2026. The Financial Crimes Enforcement Network (FinCEN) states that it issued a final rule making permanent the beneficial ownership information reporting exemptions (www.fincen.gov), effective August 14, 2026 (www.fincen.gov), under which U.S. companies are exempt (www.fincen.gov). Absence of a federal filing is therefore not a finding about a domestic target.

The Legal Entity Identifier (LEI) is a second signal with its own limits. GLEIF Level 2 data records each LEI holder's reported direct and ultimate accounting consolidating parent (www.gleif.org), a consolidation concept and not a natural-person ownership list. Reporting exceptions are explicit: the entity may be controlled by a natural person without any intermediate legal entity (www.gleif.org), there may be no known person controlling it (www.gleif.org), and an entity is not required to provide non-public relationship information to register or renew an LEI (www.gleif.org). The Financial Action Task Force (FATF) states that only a natural person can be an ultimate beneficial owner (www.fatf-gafi.org) and calls for a combination of different mechanisms to collect ownership information (www.fatf-gafi.org). **An LEI exception must never be recorded as a negative ownership finding.** Public-company thresholds do not transfer: Schedule 13D and 13G reporting applies to more than five percent of a registered class of equity securities (www.sec.gov).

Location

"Location" can mean the **registered office** (a legal address), the **principal place of business** (where the issuer says it is headquartered) or an **establishment** (a physical site). Companies House describes the registered office as the address for receiving statutory mail and legal documents (companieshouse.blog.gov.uk); an appropriate address is one where documents are expected to come to the attention of a person acting for the company (companieshouse.blog.gov.uk), which can be a service provider's address. The Census Bureau defines an establishment as a single physical location at which business is conducted (www.census.gov) and notes that geographic assignment can fall back on a mailing address (www.census.gov). Form D asks for the principal place of business as a street address (www.sec.gov). State of formation and city of operations can both be correct and still differ.

Employee count, industry fit and revenue

Table 2 summarizes, by field, what each source family can establish, what it cannot, and the date semantics involved.

Field	Source family and basis	Can establish	Cannot establish	Date semantics
Legal existence and status	State or national registry (as filed)	Existence and filed particulars	Authenticity beyond what filers supply (businessfilings.sc.gov); Delaware's free search does not provide entity status (corp.delaware.gov)	Filing date, not effective date

Field	Source family and basis	Can establish	Cannot establish	Date semantics
Ownership and control	Registers of significant control; LEI Level 2; exempt-offering notices	A reported controlling person or parent where the regime requires it (www.gov.uk) (www.gleif.org)	Natural-person ownership of a U.S. LLC (corp.delaware.gov); absence of a parent (www.gleif.org)	Filing date or last LEI renewal
Location	Registry address; Form D; statistical-agency establishment	Statutory address (companieshouse.blog.gov.uk); issuer-reported business address (www.sec.gov); physical site (www.census.gov)	Where operations occur if only a registered or mailing address is known (www.census.gov)	As-of date of filing or survey
Employee count	Statistical counts; SBA rule; EU rule; 10-K	A count on a stated rule (www.census.gov) (www.ecfr.gov) (ec.europa.eu)	A like-for-like comparison across rules; any count for a private target with no filing	Reference period differs by rule
Industry fit	NAICS-based coding; SEC SIC; company-selected codes	Classification on the source's own taxonomy (www.census.gov)	Business-model fit to a mandate; consistency across revisions (census.gov)	Taxonomy revision year
Revenue	Form D revenue range; public filings	For issuers other than hedge or other investment funds, a range reported by the issuer for a fiscal year (www.sec.gov)	A point estimate; any figure if the issuer declines to disclose	Fiscal year of the range

The two right-hand columns do most of the work in review, because they tell a reviewer when a source cannot support the claim at all.

Headcount is a definition problem before it is a measurement problem. County Business Patterns counts people on the payroll in the pay period including March 12 (www.census.gov). The Quarterly Census of Employment and Wages (QCEW) keys to the pay period including the 12th (www.bls.gov) and excludes proprietors and the unincorporated self-employed (www.bls.gov). The SBA counts all individuals employed full-time, part-time or on another basis (www.ecfr.gov), averages over 24 months (www.ecfr.gov) and includes domestic and foreign affiliates (www.ecfr.gov). The European Commission measures headcount in annual work units (ec.europa.eu) and notes that a firm in a larger group may need to include group staff data (single-market-economy.ec.europa.eu). A public registrant's business description must include the number of persons employed (www.ecfr.gov). A vendor figure of 212 and a website statement of "about 85" can both be honest if one includes affiliates or averages over two years, so the ledger records the definition next to the number.

Industry codes differ by unit and taxonomy. An establishment is classified entirely on its major activity (www.census.gov), an enterprise in the two-digit sector with the largest payroll share (www.census.gov), and the North American Industry Classification System (NAICS) is revised, as when the 2022 revision eliminated Subsector 454, Nonstore Retailers (census.gov). In the UK a company can select up to four SIC codes (companieshouse.blog.gov.uk), and changes them only when it files its next confirmation statement

(companieshouse.blog.gov.uk). A Form D industry group reflects the use of the bulk of the proceeds of the offering (www.sec.gov), not the whole business. A taxonomy code is an observation about classification; "fit" is the analyst's claim, supported by quoted evidence of what the company does.

Revenue follows the same logic. A Form D revenue range covers the most recently completed fiscal year available (www.sec.gov), and the issuer certifies the notice itself (www.sec.gov). The range is requested only of issuers that do not specify "Hedge Fund" or "Other Investment Fund" in Item 4, so pooled-fund issuers report an aggregate net asset value range instead. Two vendor estimates that agree may both derive from it, so test independence first. A private-target revenue claim with no filing behind it should usually stay "issuer-stated" or "unknown," with the range and fiscal year recorded, until a document is obtained through an approved channel.

Dispositions, Second Review and Escalation

The six dispositions

Each claim ends in exactly one disposition. The definitions are operational and should be copied into team guidance as written.

- **Accepted.** One source of the highest authority for this field, inside the recency window, supports the value, and no independent source contradicts it.
- **Corroborated.** Two or more independent sources, normalized to the same definition and date, support the value.
- **Issuer-stated.** Only the company itself supports the value, and no independent source confirms or contradicts it. It may be used for screening with a visible label.
- **Conflicted.** Independent sources, normalized to the same entity, date and definition, support different values and the authority ladder does not settle it.
- **Unknown.** No admissible observation exists, or the only observations are copied, non-comparable, or of a provenance or independence that cannot be established. A registry absence or an LEI exception is "unknown," never a negative finding.
- **Not applicable.** The field does not apply to the entity.

Every disposition keeps the rejected alternatives and a reason drawn from the Table 1 classes, so a later reviewer can see whether a value was dropped as a derivative, as stale or as a different definition.

Second-reviewer gate

A claim that can change a mandate pass/fail result or outreach eligibility goes to a second reviewer who sees the ledger rows but not the first decision. This is a maker-checker control. The NIST control catalog frames separation of duties as dividing functions among different individuals or roles (raw.githubusercontent.com), and the U.S. intelligence community's analytic standards require products to distinguish information from assumptions and judgments (www.dni.gov), state linchpin assumptions (www.dni.gov), explain the basis for uncertainties (www.dni.gov) and assess plausible alternative hypotheses (www.dni.gov). They also name the age and currency of information as a source-quality factor (www.dni.gov). These documents govern other domains and are borrowed as analogies only.

Time-boxed escalation and recheck triggers

An unresolved conflict should not stay open indefinitely or be settled by fatigue. Set a fixed time box per conflict (a number of business days or source attempts chosen by the team and written in guidance), after which the conflict goes to the research manager with the ledger rows attached. If it cannot be resolved there, the disposition stays "conflicted" (or "unknown", where independence or provenance cannot be established) and the claim is flagged as an open disqualifier. Each row also stores a **recheck trigger**: a date or an event such as a new filing, a registry status change, a website update or a data release. The SEC publishes Form D data quarterly (www.sec.gov), which gives Form D-based fields a built-in refresh cadence.

Do not let a score hide a disqualifier

Weighted confidence scores sort lists but are dangerous when they average three high-confidence passes with an unresolved disqualifier on a fourth field. The rule is lexicographic: an unresolved disqualifier is displayed first, and a target with any open disqualifier, whether its disposition is "conflicted" or "unknown", cannot be shown as passing and keeps outreach eligibility on hold, whatever its score. Scores may order targets within a disposition class, never across it.

Optional inter-reviewer agreement rate

Where two reviewers independently decide a subset of conflicts, an agreement rate can show whether the guidance is clear. Define it as matching decisions divided by double-reviewed conflicts, and always publish the denominator. Percent agreement does not take account of the possibility that raters guessed (pmc.ncbi.nlm.nih.gov); in a published worked example, 94% agreement (0.94) corresponds to a kappa of 0.85 once chance is removed (pmc.ncbi.nlm.nih.gov). The same paper notes that many texts recommend 80% agreement as a minimum (pmc.ncbi.nlm.nih.gov), a convention from another field and not a standard for company research. A team should set its own target, treat the figure as a diagnostic and review disagreements one by one. No benchmark for this task is published, and none is implied here.

A Fictional Three-Source Example (Hypothetical Example)

This example is entirely fictional. The company, the sources and the figures are invented to show the mechanics. It does not describe any real acquisition target.

Mandate. Independently owned (not sponsor-backed) manufacturers of metal fasteners, headquartered in one named U.S. state, with 50 to 150 employees. **Entity.** "Alderbrook Fastener Works LLC" (fictional). **As-of date for the claims:** 2026-10-06.

Observations.

- **Source A, commercial data profile (fictional vendor).** Employees: 212. Headquarters: Columbus, Ohio. Industry: "industrial distribution." Ownership: "backed by Meridian Holdings" (fictional). Last refreshed: March 2025. Retrieved 2026-10-05.
- **Source B, company website "About" page.** "A family-owned manufacturer." "About 85 team members across two plants." Headquarters: Dayton, Ohio. Page dated September 2026. Retrieved 2026-10-06.

- **Source C, state business registry record.** Limited liability company, active, filed particulars current as of its last annual filing; registered office is a registered-agent address in another state; no members listed. Retrieved 2026-10-06.

Table 3 shows the claim record before and after adjudication.

Field	Before (merged, unlabeled)	Adjudication steps applied	After (ledger disposition)	Second review
Employee count	"212 / 85" shown as a range	Step 3: A's definition is unstated and may include affiliates; B states "about 85" for current plants. Step 2: A was refreshed 18 months before B was dated.	**Issuer-stated: 85.** 212 is kept as a rejected alternative ("definition unknown, older refresh"). The 50 to 150 band is met on the issuer figure only.	Required: the value decides pass/fail
Ownership	"Sponsor-backed (vendor)"	Step 5: the registry cannot list members, so it neither confirms nor refutes. Step 4: A's source of the parent link is unknown.	**Unknown** (vendor parent link of unknown source and untested independence versus issuer "family-owned"; the registry cannot speak to ownership). Open disqualifier; **no pass status**; outreach eligibility held.	Required: outreach eligibility
Headquarters location	"Columbus / Dayton"	Step 1: A may describe a different site. Step 3: registered office is not headquarters.	**Issuer-stated at state level** (Ohio in A and B; A's independence from B is untested, so the match is not treated as corroboration, and the registry supplies no headquarters value). City is **issuer-stated** as Dayton; the Columbus value is retained as a possible second facility.	Not required: state-level mandate
Industry fit	"Distribution / manufacturing"	Step 3: A's label is a vendor category; B states the activity. No registry code present.	**Issuer-stated** (manufacturer). Not independently corroborated.	Required: the mandate admits manufacturers only

No value was averaged or chosen by source reputation. The headquarters conflict dissolved once entity and field definitions were separated, as the cross-tab predicts, while the ownership claim stayed unknown, with an open disqualifier, because the only official source cannot speak to ownership and the vendor's independence is untested. Its recheck trigger is a date seven business days out or any change in Source B, and the next step is to seek independent evidence of the vendor's parent link through an approved channel. No outreach goes out meanwhile.

Illustrative agreement calculation (fictional). Suppose two reviewers independently decided five double-reviewed conflicts from this and similar files and matched on four. The agreement rate is 4 matching decisions divided by 5 double-reviewed conflicts, or 80%, with a denominator of 5. The number is arithmetic on invented counts, not a benchmark, and with a denominator this small a single changed decision moves the rate by 20 percentage points.

Reference Asset: Conflict Ledger Schema

The schema below can be copied into a spreadsheet or a database table. One row is one observation; a claim has many rows and one disposition. Columns are grouped by purpose.

```
ledger_id               unique row id
canonical_entity_id     internal ID for the resolved legal entity
entity_match_status     match | non_match | possible_match
field                   ownership | location | employee_count | industry_fit | revenue | other
candidate_value         literal value as printed by the source
definition_note         what the source says the value measures (e.g. "average over 24 months")
source_url              URL of the page or document
source_type             official_record | issuer_statement | statistical_agency | vendor_aggregate | press
source_excerpt          short verbatim excerpt supporting the value
named_entity_in_source  the entity name exactly as the source prints it
jurisdiction            country / state of the record
effective_date          date the fact is said to be true
publication_or_filing_date  date the source published or filed it
retrieval_date          date the analyst retrieved it
authority_for_field     rank on this field's authority ladder (1 = strongest)
independence_flag       original | copied_from:<ledger_id> | unknown
contradiction_reason    different_entity | different_date | different_definition | copy_or_aggregation_error
| issuer_vs_record | true_contradiction | none
decision                accepted | corroborated | issuer_stated | conflicted | unknown | not_applicable
decision_reason         one sentence, citing the contradiction_reason
rejected_alternatives   ledger_ids rejected, each with reason
mandate_effect          pass_fail | outreach_eligibility | none
reviewer_1, reviewer_2  initials; reviewer_2 required if mandate_effect != none
reviewer_agreement      match | mismatch | not_double_reviewed
escalation_due_date     end of the time box
next_check_date         scheduled recheck
recheck_trigger         e.g. new filing, status change, quarterly data release
```

A minimal example row for the fictional ownership claim reads: `L-0003 | ENT-0412 | match | ownership | "backed by Meridian Holdings" | vendor_aggregate | independence unknown | issuer_vs_record | unknown | outreach_eligibility | reviewers R1, R2 | match | recheck on website or vendor refresh`. The `definition_note` and `independence_flag` columns look like overhead, yet they explain most of the apparent conflicts in the cross-tab.

Data Analysis and Evidence

The figures below put the limits of public sources in numbers. All are quoted from their originators. None is a benchmark of conflict rates in private-equity screening, and no such benchmark is known to exist.

- **Exempt-offering volume.** The SEC's Regulation D statistics report 34,553 initial Form D filings in 2025 (www.sec.gov) and 56,254 total filings including amendments (www.sec.gov), with a total amount sold of $2,391.5 billion (www.sec.gov). The counts are based on information that issuers self-report

(www.sec.gov), the SEC notes they could underestimate activity if some issuers did not file (www.sec.gov), and the staff cannot guarantee the accuracy of the underlying filings (www.sec.gov).

- **Form D data set limits.** The SEC documentation states the data may contain redundancies and inconsistencies (www.sec.gov), covers submissions from January 2008 (www.sec.gov) and is not a substitute for the filings (www.sec.gov). The form is a brief notice with basic information about the company and the offering (www.investor.gov), and the landing page repeats the accuracy caution (www.sec.gov).
- **Form D field coverage.** Year of incorporation is requested only for organizations formed within the past five years, with older issuers indicating they were formed over five years ago (www.sec.gov), so the exact formation year is often absent for mature targets.
- **LEI scale and quality.** GLEIF reports 3.19 million active LEIs at the end of Q3 2026 (www.gleif.org), of which 88.1% were fully corroborated (www.gleif.org). In Q2 2026, parent-information reporting represented 99% of the LEI population (www.gleif.org), a count of entities that reported information on direct and ultimate parents (www.gleif.org) including reporting exceptions. The 99% must not be read as 99% having a named parent.
- **Imputation in official counts.** The Census Bureau reports that employment is missing or zero while quarterly payroll is positive for approximately 6% of incoming administrative records (www.census.gov), so even official employment series contain modeled values.
- **Industry-code agreement.** A 2013 Census Bureau working paper comparing person-reported industry in the American Community Survey with employer-reported industry in linked administrative data found an overall industry sector match rate of 75 percent (www2.census.gov), and 61 percent at the four-digit Census Industry Code level. The comparison covers worker jobs rather than enterprise classification, but it shows that industry labels from different kinds of source can differ.
- **Cross-database discrepancy.** An NBER working paper on venture capital data observes that existing databases differ in methodologies and that analyses frequently produce different results (www.nber.org), and that one platform may classify the same investor differently from another (www.nber.org).

Three conclusions follow:

- **Official sources carry self-report or imputation.** They are large and valuable, yet each has a documented self-reported or modeled component.
- **Vendor coverage is not field accuracy.** An aggregator's description of data drawn from official primary sources across 140+ jurisdictions (opencorporates.com) describes breadth, and is the vendor's own statement.
- **Published percentages are not comparable.** The figures sources publish about themselves (88.1%, 99%, 75%) measure different things and should not be converted into a confidence score for a specific target.

Implications and Future Directions

- **For origination teams.** The most useful change is procedural: record observations before merging them, and write the authority ladder for each field before research starts. A reviewable screening decision needs three visible things for every mandate-changing claim: the rows, the contradiction reason

and the second reviewer. A ledger can be run in a spreadsheet, which keeps adoption inexpensive. The protocol can be applied by any team to any combination of sources.

- **For vendors and aggregators.** Field-level provenance (originating source, effective date, definition) is what makes a vendor value auditable. The W3C PROV model gives a vocabulary: a derivation is a transformation of one entity into another (www.w3.org), and a record built from a registry filing is such a derivation. PROV also notes that data published by a respected independent organization may be considered more trustworthy (www.w3.org), a reason to record the agent on every row.
- **Open questions.** No public data describes how often screening conflicts arise or how often second review changes a decision, so the agreement rate and the conflict mix should be measured internally. Professional advisers should be consulted where a conflict concerns legal ownership, regulatory status or a binding transaction decision; this report is a research-process guide and not legal advice.

Frequently Asked Questions (FAQs)

How should an analyst resolve conflicting data on an acquisition target? Fix the claim, entity and date; store each observation as a row; classify the conflict; apply the field's authority ladder; assign one of six dispositions with a reason; and escalate mandate-changing conflicts to a second reviewer.

Why do private equity target company data sources disagree? Mostly because they measure different things. Headcount may be a March 12 payroll count (www.census.gov), a 24-month average (www.ecfr.gov) or a group-level figure (single-market-economy.ec.europa.eu); a location may be a statutory address (companieshouse.blog.gov.uk) or a site (www.census.gov); an industry may be a code assigned from an employer description (www.bls.gov).

What goes in a claim adjudication ledger for due diligence? Canonical entity ID, field, candidate value, source and excerpt, effective and retrieval dates, authority for this field, independence flag, contradiction reason, decision, reviewer and next check. The full schema is in the reference section.

How does source conflict resolution work in deal sourcing without a master database? By field. A registry may be the best evidence for legal status and the weakest for ownership (corp.delaware.gov); a statistical agency defines headcount precisely but may not cover a given private firm (www.census.gov).

How can private-company revenue be verified when sources disagree? Check independence first, since estimates can derive from one origin (lunadong.com). A Form D range is issuer-reported for a stated fiscal year (www.sec.gov) and may be undisclosed. Record the range and year, and keep the claim "issuer-stated" or "unknown" until a document is obtained through an approved channel.

What is evidence-backed acquisition target research, and what is field-level provenance? Research in which every field value traces to a retained source passage, a date and a stated process, consistent with the PROV view of provenance as a basis for assessing quality and reliability (www.w3.org). Each value carries its own source, agent, effective date and definition, matching the FAIR requirement of detailed provenance (www.nature.com).

How should conflicting company data be reconciled in due diligence? Normalize first (entity, date, definition), test independence second, apply authority third and decide last. Treat a registry absence, or an LEI reporting exception (www.gleif.org), as "unknown," never as evidence that no owner exists.

Conclusion

Conflicting data on a private-equity target is best handled as a record-keeping problem with decision rules, not as a search for one trustworthy database. The protocol in this report has a short list of requirements: define the claim, entity and date; store each observation with its excerpt and dates; classify the conflict before choosing a value; use a written authority ladder per field; test independence before counting agreement; assign one of six dispositions and keep the rejected alternatives; send mandate-changing claims to a second reviewer; time-box escalation and set recheck triggers; and never let a score outrank an unresolved disqualifier.

The sources show why this discipline is needed. Registries establish filed particulars but often not ownership; exempt-offering notices are issuer-reported; statistical agencies define headcount and industry on their own rules; identifier systems record exceptions that are not findings. A reviewer who sees the rows, the reasons and the second review can accept, reject or hold a field on evidence. Where evidence is thin, the correct and defensible entry is "unknown," recorded with its recheck date.

Source links

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

1. https://www.w3.org/TR/prov-dm
2. https://www.sec.gov/files/Form_D.pdf
3. https://guides.loc.gov/company-research/private
4. https://corp.delaware.gov/faqs
5. https://www.census.gov/programs-surveys/cbp/about/glossary.html
6. https://www.ecfr.gov/current/title-13/chapter-I/part-121/subject-group-ECFR8f6c0b0a3ba5b04/section-121.106
7. https://www.census.gov/programs-surveys/susb/about/glossary.html
8. https://companieshouse.blog.gov.uk/2021/10/05/how-to-choose-or-change-your-companys-registered-office-address
9. https://lunadong.com/publication/dependence_vldb.pdf
10. https://arxiv.org/pdf/1905.06167
11. https://census.gov/content/dam/Census/library/working-papers/2006/adrm/rrs2006-02.pdf
12. https://rdmkit.elixir-europe.org/data_provenance
13. https://www.nature.com/articles/sdata201618
14. https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance
15. https://www.dublincore.org/specifications/dublin-core/dcmi-terms
16. https://www.sec.gov/files/formd.pdf
17. https://edoc.hu-berlin.de/bitstreams/d61400e7-efd0-4654-b042-7acdd463bac4/download
18. https://arxiv.org/pdf/1505.02463
19. https://www.iso.org/standard/35736.html
20. https://obamawhitehouse.archives.gov/omb/info_quality_iqg_oct2002
21. https://businessfilings.sc.gov/businessfiling
22. https://www.gov.uk/guidance/people-with-significant-control-pscs

23. https://www.gov.uk/guidance/when-you-need-to-verify-your-identity-for-companies-house
24. https://www.gov.uk/guidance/search-the-companies-house-register
25. https://www.fincen.gov/boi
26. https://www.gleif.org/en/lei-data/access-and-use-lei-data/level-2-data-who-owns-whom
27. https://www.gleif.org/en/about-lei/common-data-file-format/level-2-data-reporting-exceptions-format
28. https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Guidance-Beneficial-Ownership-Legal-Persons.pdf.coredownload.pdf
29. https://www.sec.gov/files/divisions/corpfin/guidance/reg13d-interp.htm
30. https://companieshouse.blog.gov.uk/2024/01/22/get-ready-for-changes-to-uk-company-law
31. https://www.census.gov/programs-surveys/cbp/technical-documentation/methodology.html
32. https://ec.europa.eu/docsroom/documents/42921/attachments/1/translations/en/renditions/native
33. https://census.gov/library/reference/code-lists/naics/2022/implementation-timeline.html
34. https://www.bls.gov/cew/overview.htm
35. https://single-market-economy.ec.europa.eu/smes/sme-definition_en
36. https://www.ecfr.gov/current/title-17/chapter-II/part-229/subpart-229.100/section-229.101
37. https://companieshouse.blog.gov.uk/2021/10/12/choosing-a-standard-industrial-classification-sic-code-for-your-company
38. https://raw.githubusercontent.com/usnistgov/oscal-content/main/nist.gov/SP800-53/rev5/json/NIST_SP-800-53_rev5_catalog.json
39. https://www.dni.gov/files/documents/ICD/ICD-203.pdf
40. https://pmc.ncbi.nlm.nih.gov/articles/PMC3900052
41. https://www.sec.gov/data-research/statistics-data-visualizations/regulation-d-offerings
42. https://www.investor.gov/introduction-investing/investing-basics/glossary/form-d
43. https://www.sec.gov/data-research/sec-markets-data/form-d-data-sets
44. https://www.gleif.org/en/newsroom/blog/the-lei-in-numbers-q3-2026-data-shows-an-expanding-more-connected-global-lei-system
45. https://www.gleif.org/en/newsroom/blog/the-lei-in-numbers-q2-2026-signals-growing-regional-momentum-for-the-lei
46. https://www2.census.gov/ces/wp/2013/CES-WP-13-47.pdf
47. https://www.nber.org/system/files/working_papers/w22500/w22500.pdf
48. https://opencorporates.com

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.