Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

Best Entity Identifier for Private Equity Target Records

> 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 · 2026-09-20 · best entity identifier for private equity target records · entity resolution · private equity deal sourcing · company registration number · legal entity identifier · central index key · master data management · target data deduplication · company identifier crosswalk

Original article: https://docket.capital/articles/pe-target-record-entity-identifiers



In this article

1. Executive Summary
2. Introduction and Background
3. Key Changes
4. Identifier Policy by Jurisdiction and Object
5. Matching, Merge and Exception Policy
6. Implementation Considerations and Process Changes
7. Data Analysis and Evidence
8. Case Studies and Real-World Examples
9. Implications and Future Directions
10. Frequently Asked Questions (FAQs)
11. Conclusion

Executive Summary

The best entity identifier for private equity target records is not a single external code. It is an **internally generated, immutable target-record ID** supported by a set of authority-qualified identifiers. The strongest legal-entity key is the tuple `issuing jurisdiction + registration authority + local registration number`. That design follows the World Wide Web Consortium principle that legal status is conferred by an authority within a jurisdiction (www.w3.org). It also survives source changes: source rows retain their own IDs, while the internal ID holds the crosswalk and history. Statistics Canada uses the same architectural separation by storing links between its internal IDs and original record IDs (www.statcan.gc.ca).

External identifiers answer narrower questions. A **Legal Entity Identifier (LEI)** is a 20-character code and represents one entity (www.gleif.org), but eligibility and adoption do not make it universal. The active LEI population exceeded **3.1 million in Q2 2026** (www.gleif.org). A **Central Index Key (CIK)** identifies an EDGAR filer, not every company, and the Securities and Exchange Commission (SEC) notes that one legal entity can have more than one CIK for different filing roles (www.sec.gov). A Companies House number is durable within the United Kingdom, but the API covers entities within a defined statutory scope (developer.company-information.service.gov.uk).

The operating rule is **merge legal entities only on identity evidence**. Exact authority IDs can authorize an automatic merge when authority, jurisdiction, entity type and identifier all agree. A verified crosswalk is next strongest. Composite agreement on normalized name, address, domain and officers can create a review candidate, but none of those attributes alone proves identity. Domain registration can transfer to another registrant (www.icann.org), and the Internal Revenue Service explicitly treats a trade name as different from a legal name (www.irs.gov). Parent, brand, fund, location and target-group relationships should therefore be **links**, not merges.

Every decision needs a source, observation date, rule version and reversible event. The correct quality measure is **precision**, confirmed correct merges divided by reviewed merges, which matches the Office for National Statistics definition of the proportion of links that are true matches (www.ons.gov.uk). Review yield is

a separate internal metric: confirmed matches divided by manual-review candidates. No accuracy claim should be published without a labeled test set. Docket's first-party operating model is consistent with this evidence discipline: identity is checked before detailed research, and corrections retain earlier findings.

Introduction and Background

Private equity origination teams routinely join vendor exports, registry searches, customer relationship management (CRM) records and analyst workbooks. The visible problem is duplicated rows. The costlier problem is **misassigned evidence**: an employee count belongs to a parent, a filing belongs to a similarly named subsidiary, or an old brand is treated as an independent target. The effects include duplicate outreach, lost research after a rename, misleading scores and a committee record whose provenance cannot be reconstructed.

The phrase **company entity resolution** hides several different objects. Eurostat defines a legal unit around legal existence independent of its owners (ec.europa.eu). An enterprise may instead be the smallest operating combination of legal units (ec.europa.eu). The U.S. Census Bureau defines an establishment at a single physical location (www.census.gov). These are related objects, not synonyms.

This report turns that taxonomy into a field-level policy for **entity resolution for private equity deal sourcing**. It addresses **business identifiers by jurisdiction**, a global company identifier crosswalk, company registration number entity matching, ultimate parent company identifier treatment and private equity target data deduplication. It does not infer present control from a registry key, and it does not treat a website, vendor record or normalized name as legal identity proof.

Docket is a direct provider of sourced target research, not an identifier issuer. Its relevant perspective therefore belongs in process design rather than the identifier comparison table. Its source guide states that authority is assessed field by field. That is also the premise here: the best key depends on the object and decision, while the internal record binds the evidence together.

Key Changes

Change 1: Replace the universal-key search with a layered key model

No external identifier has universal eligibility, jurisdictional reach and perfect one-to-one mapping to a PE target. The master should therefore have four layers:

- **Internal stable ID:** a random, immutable identifier for one master object. It never encodes a name, country or source.
- **Authority key:** jurisdiction, registration authority, local ID and entity type. This is the strongest legal-entity match key.
- **Crosswalk IDs:** LEI, CIK, Unique Entity ID, vendor IDs and historical source IDs, each with issuer and validity dates.

- **Typed relationships:** `operates_as`, `has_location`, `direct_parent_of`, `ultimate_accounting_parent_of`, `brand_of`, `fund_vehicle_of`, `successor_to` and `member_of_target_group`.

The W3C recommends that an identifier value carry its identifier scheme (www.w3.org). A bare value such as `12345678` is unsafe because it can be meaningful in multiple systems. A deterministic composite key can be rendered as `GB|COMPANIES_HOUSE|01234567`, but normalization only makes comparison reproducible. It does not establish that a vendor row belongs to that entity.

Change 2: Separate the objects before matching them

The master model should classify every record before calculating similarity:

- **Legal entity:** an incorporated or otherwise legally recognized unit, anchored where possible by an official registration key.
- **Operating enterprise:** an economic organization that can combine multiple legal units.
- **Establishment or location:** a physical operating site, keyed separately and linked to its legal operator.
- **Trade name or brand:** a label used in commerce. Schema.org defines a brand as a labeling name, not the organization itself (schema.org).
- **Domain:** an internet resource and discovery attribute. The United States Patent and Trademark Office notes that domain registration alone does not create trademark rights (www.uspto.gov).
- **Parent relationship:** a sourced, dated edge between separate entities, not a field that collapses the child into the parent.
- **Fund vehicle:** a separate pooled-money entity. The SEC describes a private fund as an entity created to pool money from multiple investors (www.sec.gov).
- **Target group:** an analyst-defined collection used for origination workflow. W3C recognizes that umbrella organizations can span several legal entities (www.w3.org).

This taxonomy prevents the most common category error: merging related rows because they describe the same commercial story. A group may be one target for coverage planning but still require separate legal-entity records.

Change 3: Make every identity action reversible and time-aware

An identity decision is an event, not a destructive update. A substantial organizational change can warrant a new identifier while retaining the relation to the former organization (www.w3.org). Parent relationships also need effective dates and source dates. Open Ownership defines a statement date as the date on which the information is claimed true (standard.openownership.org).

The event vocabulary should be explicit:

- **Merge:** two records represent the same object and one survivor becomes the current view.

- **Link:** records remain independently valid but gain a typed relationship.
- **Split:** an earlier master is decomposed after evidence shows that it mixed objects.
- **Supersede:** a new record succeeds an old one without rewriting historical observations.
- **Unmerge:** a prior merge event is reversed, restoring both records and their source memberships.

Reversibility is feasible in production systems: documented master-data workflows allow records to be restored to their original state (docs.oracle.com). Provenance should connect source entities, transformation activities and outputs as a chain (www.w3.org).

Identifier Policy by Jurisdiction and Object

The jurisdiction-aware identifier matrix

Table 1 is the core policy. Public lookup describes whether a normal reviewer can inspect the source, not whether every field is open. Update cadence refers to the source system, not a promise that every underlying fact changed at that interval.

Identifier	Object represented and issuer	Format and eligibility	Public lookup and update	Known ambiguity	Safe use in a target master
Internal stable ID	Master object created by the PE data team	UUID or similarly nonsemantic value; every modeled object	Internal lookup; append-only events	Has no external meaning	Primary database key; never evidence of legal identity
Local registration ID	Legal entity; jurisdictional registry	Registry-specific; qualify by jurisdiction, authority and entity type	Registry lookup where available; cadence varies	Same digits can occur in different systems; branches and registrations may differ	Strongest legal key when the issuing tuple matches
Companies House number	UK company in Companies Act scope	Eight digits or two letters plus six digits (www.gov.uk)	Public API described as live and real-time (developer.company-information.service.gov.uk)	Not a global identifier; scope follows UK law	Automatic match only with the issuer and exact number
Delaware file number	Entity registered with Delaware Division of Corporations	State-assigned value for each incorporated, formed, qualified or registered entity (icis.corp.delaware.gov)	State lookup	State-scoped, not a federal company number	Use as `US-DE

Identifier	Object represented and issuer	Format and eligibility	Public lookup and update	Known ambiguity	Safe use in a target master
LEI	Legal entity in the Global LEI System	20-character alphanumeric; financial-transaction eligibility (www.gleif.org)	Global public data; Golden Copy updates three times daily (www.gleif.org)	Coverage depends on eligibility and adoption; renewal status matters	High-quality cross-jurisdiction key when present; keep local registry crosswalk
CIK	SEC EDGAR filer account	SEC-assigned numeric identifier for filers (www.sec.gov)	Public EDGAR lookup	People can have CIKs; one legal entity can have several filing-role CIKs	Link filings to a target; do not use as universal company key
Canadian BN	Business interacting with Canadian government; Canada Revenue Agency	Unique nine-digit number (www.canada.ca)	Government use; access varies by program	Can represent more than corporations	Qualify by issuer and do not assume corporate legal form
Australian ACN or ABN	ACN identifies a company; ABN identifies a business	ACN has nine digits; ABN has 11 digits (www.asic.gov.au)	Official register searches	ACN and ABN have different object scope	Store both with scheme; prefer ACN for company identity
Singapore UEN	Registered entity across multiple legal and organizational types	Scheme-specific value	Government dataset and issuer lookup	Covers companies, partnerships, charities and societies (data.gov.sg)	Pair with entity type before matching
EUID	Entity in the European Business Registers Interconnection System	ISO 6523-aligned composite (eur-lex.europa.eu)	Cross-border register interconnection	Availability follows participating registers and object types	Use as a cross-border registry key, with domestic ID retained
Unique Entity ID (SAM)	Authoritative entity identifier in U.S. General Services Administration systems (www.gsa.gov)	12-character value (open.gsa.gov)	SAM data and API access rules apply	Government-program identifier, not corporate formation key	Use for federal-award crosswalks, not legal-entity creation
Vendor ID	Vendor's representation of a company or group	Proprietary	Vendor-dependent	May change after remapping and is only unique within that system	Preserve as source lineage; never authorize a merge alone
Domain	Internet resource, not necessarily an entity	DNS name	Registration and site evidence change	Shared, redirected, transferred or controlled by a parent	Discovery and corroboration attribute only

The matrix answers **legal entity identifier vs company registration number** directly. An LEI provides a standardized global reference where the entity participates in the LEI system. A company registration number derives its authority from the jurisdiction that recognizes or registers the entity. For lower-middle-market targets, the local authority tuple is usually the legal anchor; an LEI is an excellent crosswalk when available. Neither should replace the internal ID that preserves all source and relationship history.

Crosswalk policy

A global company identifier crosswalk should be a table of claims, not a flat set of columns. Each row needs `master_id`, `scheme`, `issuer`, `jurisdiction`, `identifier`, `object_type`, `valid_from`, `valid_to`, `observed_at`, `source_url`, `source_excerpt`, `status` and `confidence`. GLEIF's Registration Authorities List covers sources from **232 jurisdictions** (www.gleif.org) and can carry the local number maintained by the authoritative source (www.gleif.org).

Do not infer current ownership from that crosswalk. GLEIF Level 1 answers “who is who,” while Level 2 addresses “who owns whom” (www.gleif.org). Its parent relation is specifically based on accounting consolidation (www.gleif.org). Store it as a dated relationship with its definition and source.

Matching, Merge and Exception Policy

A five-stage matching cascade

The policy should produce five dispositions, not a single fuzzy score:

1. **Exact authority match:** same scheme, issuer, jurisdiction, entity type and identifier. Automatically merge source rows into the same legal-entity master unless contradictory object typing is present.
2. **Verified crosswalk:** an official mapping connects two identifiers to the same entity. Merge with the mapping source and observation date recorded.
3. **Strong composite match:** legal name, jurisdiction, address, domain, phone and officers align without a verified authority ID. Create a possible-match case for human review, never an automatic merge.
4. **Conflict:** a strong identifier disagrees with another strong identifier, or the same domain points to multiple legal entities. Freeze automation and route to an exception queue.
5. **No match:** evidence is insufficient or points to distinct objects. Keep records separate and permit later reconsideration.

Statistics Canada describes deterministic linkage as using common identifiers or variables (www150.statcan.gc.ca), while probabilistic linkage can accommodate incomplete or erroneous data (www150.statcan.gc.ca). The Office for National Statistics similarly describes a cascade from exact to probabilistic matching and then clerical resolution (www.ons.gov.uk).

Match-decision worksheet

Table 2 is a copyable match-decision worksheet. Teams can download the page as Markdown or copy the table to a spreadsheet. Each row represents one candidate pair and the reviewer must fill the evidence columns before selecting a disposition.

Field	Record A	Record B	Comparison rule	Evidence required	Decision effect
Object type			Same legal entity, brand, location, fund or group?	Source supporting each classification	Different types usually link, not merge
Jurisdiction			Exact normalized authority jurisdiction	Registry page or official crosswalk	Missing jurisdiction blocks exact-ID merge
Registration authority			Exact issuer code	Issuer page	Different issuers make equal digits inconclusive
Local registration ID			Exact after scheme-valid formatting only	Official record	Exact qualified ID is strongest merge evidence
Legal name and former names			Token and former-name comparison	Registry name history	Supports review; name alone never merges
Address			Structured current and historical comparison	Dated source	Shared offices and agents reduce weight
Domain			Exact host, redirect and page attribution	Retained page and date	Corroboration only; shared domains trigger review
LEI, CIK and other IDs			Scheme-specific crosswalk	Official issuer or mapping	Apply scheme limitations before merge
Relationship evidence			Brand, parent, branch, successor or group	Dated source and relation definition	Create typed edge, not entity merge
Disposition			Merge, link, split, supersede, possible, conflict or no match	Reviewer, rule version and reason	Writes an append-only decision event

The worksheet makes **company registration number entity matching** auditable. It also prevents strong but misleading composites. A normalized name and exact domain can still refer to a brand shared by several entities. ICANN states that it does not verify user details in registration profiles (www.icann.org). Those attributes can prioritize review, but they cannot authorize a legal merge.

Survivor rules and reversible merge log

Survivorship should operate at the field level. One source may be strongest for legal name while another is current for operating website. A documented master-data model describes selection of the best available value for each attribute (docs.oracle.com). The survivor record should not erase losing values; it should project current values from retained observations.

Table 3 is the minimum reversible merge log. It is also a practical schema for an audit export.

Event field	Required content	Why it matters
event_id	Immutable unique value	References the exact action
event_type	merge, unmerge, link, unlink, split or supersede	Separates identity from relationship changes
object_ids_before	All affected stable IDs	Makes reversal possible
survivor_or_output_ids	IDs after the action	Preserves continuity without deleting inputs
evidence_ids	Registry records, mappings and retained pages	Shows what supported the action
rule_version	Exact policy and threshold version	Supports later re-evaluation
actor and reviewer	System, analyst and approver	Establishes responsibility for the decision
decided_at and effective_at	Decision timestamp and real-world effective date	Separates observation from business time
reason_code and note	Exact ID, verified crosswalk, reviewed composite or correction	Makes queue analytics possible
reversal_of	Prior event ID when applicable	Restores lineage instead of rewriting it

The log should be append-only. Record IDs should not be recycled for different objects, consistent with the Open Ownership data standard (standard.openownership.org). Docket's related operational principle is to keep each observation with its date and source.

Implementation Considerations and Process Changes

Minimum field model

A production master should distinguish identity, claims and relationships. The minimum fields are:

- **Master object:** `master_id`, `object_type`, `created_at`, `status` and `superseded_by`.
- **Identifier claim:** `scheme`, `issuer`, `jurisdiction`, `value`, `entity_type`, `valid_from`, `valid_to` and `verification_status`.
- **Name observation:** legal, former, trade, brand or vendor label, with language and dates.
- **Address observation:** structured address, address role, source and observation date.
- **Domain observation:** hostname, page role, redirect target, source and observation date.
- **Relationship:** subject ID, predicate, object ID, definition, effective dates, source dates and confidence.
- **Source lineage:** vendor file, row ID, import batch, original value and retained evidence.
- **Decision event:** action, rule version, candidate scores, reviewer, evidence and reversal pointer.

Names should remain typed. W3C explicitly treats a trading name or alternative name as an attribute of a legal entity (www.w3.org). A rename therefore updates name observations but does not normally change the stable master ID. Companies House guidance says the registration number does not change with the company name or registered-office address (www.gov.uk).

Ingestion and review workflow

The operating sequence should be fixed:

- **Profile the source:** document its object definition, ID scope, freshness and known transformations.
- **Preserve raw values:** keep the incoming row and source identifier before normalization.
- **Classify the object:** decide whether the row describes a legal entity, brand, location, fund or group.
- **Normalize without overwriting:** create comparison forms for names, addresses, domains and identifiers.
- **Resolve exact keys first:** test authority-qualified identifiers and verified crosswalks.
- **Generate candidates:** use blocking keys such as jurisdiction plus normalized name stem.
- **Score composites:** compare multiple independent fields and record missingness.
- **Route exceptions:** send conflicts and middle-score candidates to manual review.
- **Apply disposition:** merge, link, split, supersede or leave separate through an event.
- **Recalculate projections:** update current views and downstream scores from retained observations.
- **Monitor drift:** sample automatic merges, review vendor remaps and recheck stale relationships.

A two-threshold design is useful because it separates automatic matches, automatic nonmatches and a review band. Statistics Canada's methodology states that pairs between two thresholds are resolved manually (www150.statcan.gc.ca). Thresholds should be calibrated separately by jurisdiction, source pair and object type. A score that works for UK registry-to-CRM matching is not automatically suitable for a U.S. vendor file with state-level identifiers.

QA and exception queues

The exception queue should prioritize consequences, not just low scores:

- **Identifier conflict:** exact strong IDs disagree.
- **Shared domain:** multiple legal entities legitimately use one website.
- **Name collision:** common normalized name appears in the same jurisdiction.
- **Type mismatch:** one row is a brand or establishment and another is a legal entity.
- **Parent ambiguity:** source states a group relationship without identifying the precise entity.
- **Stale identifier:** source record is older than a registry change or renewal state.

- **Many-to-one vendor remap:** several vendor records collapse to one proposed master.
- **One-to-many split candidate:** one old master contains conflicting legal identifiers.

Each queue needs service levels, ownership and a sampled quality review. The master should expose unresolved status to analysts rather than silently selecting a value. Linked records can remain equally valid without one having priority, a distinction reflected in the HL7 link-type model (hl7.org).

Data Analysis and Evidence

The available numbers show why coverage and quality must be measured separately. The Global Legal Entity Identifier Foundation reported more than **3.1 million active LEIs in Q2 2026**, after **92,000 organizations** obtained an LEI during the quarter, a **3.0% quarterly growth rate** (www.gleif.org). Yet the same report put the overall renewal rate at **57.1%** (www.gleif.org). An LEI can be globally useful while still requiring status and validation dates in the target master.

GLEIF also reported **87.9%** of LEIs fully corroborated against public authoritative sources at the end of Q2 2026 (www.gleif.org). That is a data-quality status, not evidence that LEI coverage includes every acquisition target. ISO 17442-1 excludes natural persons while including international branches (www.iso.org). Eligibility and corroboration therefore answer different questions.

Scale differences also matter when matching enterprise and establishment data. The Census Bureau's Statistics of U.S. Businesses methodology covers more than **6 million single-unit establishments** and more than **2.0 million multi-unit establishments** (www.census.gov). It also says one multi-unit enterprise may have several Employer Identification Numbers (www.census.gov). Counting IDs without object definitions can therefore overstate company counts.

Quality metrics require labeled review outcomes:

- **Merge precision:** `confirmed correct merges / reviewed merges`. This estimates the trustworthiness of merge actions, not overall coverage.
- **Review yield:** `confirmed matches / manual-review candidates`. This measures whether the exception queue is productive.
- **Split rate:** `masters split after review / masters reviewed`. This detects earlier over-merging.
- **Conflict rate:** `candidate pairs with incompatible strong IDs / candidate pairs evaluated`.
- **Identifier coverage:** `eligible masters with a verified scheme ID / eligible masters`, reported by scheme and jurisdiction.
- **Relationship freshness:** `active relationship claims checked within policy interval / active relationship claims`.

An Australian Bureau of Statistics linkage retained **605,626 links** at an estimated **98.6% cumulative precision** and **1.4% false-link rate** (www.abs.gov.au). That result is useful as an example of a measured, labeled evaluation, not as a transferable benchmark for PE targets. A team should publish no matching accuracy until it has sampled its own sources, jurisdictions and decision rules.

Operational denominators should always be visible. “Ninety percent matched” is uninterpretable unless the reader knows whether the denominator is imported rows, candidate pairs, eligible legal entities or analyst-defined target groups. Report counts alongside rates, preserve the labeled test set and freeze the rule version used for each measurement.

Case Studies and Real-World Examples

Brand with several legal entities (Hypothetical Example)

Northbank Industrial is a fictional trade name used by three legal entities: a Delaware holding company, a Texas operating company and a Canadian sales corporation. All use `northbankindustrial.example`, and the vendor export contains one row with the brand name and combined headcount.

The correct process is:

- Create one **brand record** and three **legal-entity records**.
- Key the Delaware entity with its state authority tuple and the Canadian entity with its BN scheme where verified.
- Link the shared domain to all four objects with page-level attribution.
- Store the holding and operating relationships as dated edges.
- Create one **target-group record** for origination workflow, with membership edges to the three entities and the brand.
- Keep headcount unresolved until its organizational perimeter is established.

The records must not be merged merely because the domain and commercial name agree. The outcome supports one coordinated coverage plan without erasing the legal perimeter. It also makes later outreach suppression possible at either group or entity level.

Renamed company with a stale vendor row (Hypothetical Example)

Harbor Metric Ltd is fictional. Its current Companies House record uses a new legal name, while an older vendor row retains the former name and an old domain. Both records carry the same verified Companies House number.

The exact authority-qualified number supports a merge into one legal-entity master. The old vendor name and domain remain historical observations with their source dates. The merge event records both source IDs, the registry evidence, the rule version and the reviewer. No new master is created solely for the rename because the stable registration number persists through a name change. If later evidence showed that the new name belonged to a newly incorporated successor instead, the team would reverse the merge and record a `successor_to` relationship.

These examples illustrate how to identify acquisition targets across jurisdictions without flattening brands, operating companies and parents into one row. The policy preserves both the analyst's target concept and the legal entities that evidence actually describes.

Implications and Future Directions

Cross-border infrastructure is improving, but it strengthens the layered model rather than eliminating it. The European Business Registers Interconnection System lets connected registers exchange information about foreign branches and cross-border mergers (ec.europa.eu). Corporations Canada offers real-time API access to federal corporation data (ised-isde.canada.ca). These services reduce collection friction, but their scopes remain jurisdictional.

Streaming updates do not replace snapshots. Companies House cautions that its real-time stream cannot provide a complete copy of the register (developer-specs.company-information.service.gov.uk). A robust system needs an initial population, incremental changes and retained evidence for the state observed at decision time.

Three design implications follow:

- **Portability:** keeping the internal ID independent of vendor IDs allows vendor replacement without losing diligence history.
- **Explainability:** authority-qualified IDs and event logs let analysts explain why two rows were merged and undo the decision.
- **Jurisdictional calibration:** match rules, eligible identifiers and review thresholds must be configured by source and country.
- **Temporal relationships:** parent, brand and operating links require observation and effective dates rather than a timeless ownership field.

Target flexibility is equally important: an analyst-defined group can change with the investment thesis while legal-entity records remain stable.

The long-term opportunity is not one universal company number. It is interoperable mappings among official schemes, combined with explicit object typing and preserved provenance. That architecture can incorporate new registration APIs and mappings without rewriting prior research.

Frequently Asked Questions (FAQs)

What identifier should anchor a PE target record?

Use an **immutable internal ID** as the database key. For legal identity, attach a verified composite of jurisdiction, issuing authority, entity type and local registration number. Add LEI, CIK, vendor IDs and domains as qualified crosswalks or attributes, never as unscoped replacements for the internal key.

Can an LEI replace a local company registration number?

Usually no. An LEI is standardized and globally searchable, but local registration numbers derive authority from the jurisdiction that recognizes or registers the entity. Use both when available, and preserve the official crosswalk. An LEI is especially useful across markets; the local key remains essential for registry-specific facts.

When can two target rows be merged automatically?

Only when a strong, authority-qualified identifier agrees or an official crosswalk confirms identity, with compatible object types. Name, address, website domain or vendor ID alone should create at most a review candidate. Conflicting strong IDs should block automation.

How should an ultimate parent company identifier be stored?

Keep the parent as a separate legal-entity master with its own identifiers. Store the child-to-parent assertion as a dated, sourced relationship whose definition is explicit, for example ultimate accounting consolidating parent. Do not overwrite the child's identifier or treat a parent assertion as permanent.

What happens after a company changes its name?

Keep the same master when the verified legal entity remains the same. Add the new and former names as dated observations, refresh discovery attributes and preserve the source record. If the change reflects a new successor entity, keep both masters and link them through a supersession relationship.

How should vendor identifiers be used?

Vendor IDs are valuable for import lineage, refreshes and support conversations. Store the vendor, dataset, row ID and batch. Because such an identifier is scoped to the vendor's system, it should never authorize a cross-source legal merge by itself.

Conclusion

The best entity identifier for private equity target records is a **layered identity architecture**. An immutable internal ID anchors the master. Authority-qualified local registration numbers provide the strongest legal-entity evidence. LEIs, CIKs, program identifiers, domains and vendor IDs become typed crosswalks with explicit scope, issuer and dates. Brands, establishments, funds, parents and target groups remain separate objects connected by relationships.

This architecture changes the central operational question from “Which column is the company ID?” to “What object does this row describe, and what authority supports the identity claim?” The answer determines whether records should merge, link, split, supersede or remain separate. It also keeps normalization in its proper role: useful for finding candidates, insufficient for proving identity.

A workable policy is therefore conservative at the merge boundary and generous with typed links. It records every decision as a reversible event, selects survivor values field by field, and exposes unresolved conflicts to reviewers. Precision and review yield should be computed from labeled outcomes, with denominators, source pairs, jurisdictions and rule versions disclosed.

For PE origination operations, the payoff is practical: fewer duplicate targets, less evidence attached to the wrong entity, preserved history through renames and acquisitions, and a target group that can reflect the investment thesis without distorting the legal record. The durable asset is not a universal external number. It is a well-governed crosswalk whose claims can be checked, dated and reversed.

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/vocab-regorg
2. https://www.statcan.gc.ca/en/sdle/overview
3. https://www.gleif.org/en/organizational-identity/lei-vlei/the-legal-entity-identifier-lei
4. https://www.gleif.org/en/newsroom/blog/the-lei-in-numbers-q2-2026-signals-growing-regional-momentum-for-the-lei
5. https://www.sec.gov/submit-filings/filer-support-resources/edgar-glossary
6. https://developer.company-information.service.gov.uk
7. https://www.icann.org/resources/pages/ownership-2013-05-03-en
8. https://www.irs.gov/instructions/iss4
9. https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/datalinkageandmatchingpolicy
10. https://ec.europa.eu/eurostat/cache/metadata/EN/egr_esms_li.htm
11. https://www.census.gov/programs-surveys/susb/about/glossary.html
12. https://www.w3.org/TR/vocab-org
13. https://schema.org/Brand
14. https://www.uspto.gov/trademarks/basics/trademark-process
15. https://www.sec.gov/resources-small-businesses/capital-raising-building-blocks/private-funds
16. https://standard.openownership.org/en/main/standard/modelling/dates-guidance.html
17. https://docs.oracle.com/cd/F26413_18/books/UCMSIA/configuring-oracle-customer-hub-ucm-to-resolve-duplicate-records.html
18. https://www.w3.org/TR/prov-o
19. https://www.gov.uk/government/publications/xbrl-tagging-context-entity-identifiers/xbrl-tagging-context-entity-identifiers
20. https://icis.corp.delaware.gov/ecorp/FieldDesc.aspx
21. https://www.gleif.org/_documents/2024-04-22_the-legal-entity-identifier-faqs_v1.0-final_approved_.pdf
22. https://www.gleif.org/en/lei-data/access-and-use-lei-data
23. https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/enabling-interoperability/gc-enterprise-data-reference-standards/data-reference-standard-business-number.html
24. https://www.asic.gov.au/for-business-and-companies/companies/register-a-company/australian-company-number-acn
25. https://data.gov.sg/collections/1/view
26. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML?uri=CELEX:32021R1042
27. https://www.gsa.gov/about-gsa/newsroom/news-releases/gsa-systems-switch-to-unique-entity-id-sam-on-april-4-2022-03162022
28. https://open.gsa.gov/api/entity-api
29. https://www.gleif.org/en/lei-data/code-lists/gleif-registration-authorities-list
30. https://www.gleif.org/en/lei-data/access-and-use-lei-data/level-2-data-who-owns-whom/roc-policy-on-level-2-data
31. https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch3/5214780-eng.htm
32. https://www.icann.org/rdap

33. https://docs.oracle.com/cd/F14158_08/books/UCMSIA/oracle-customer-hub-ucm-data-model-tables.html
34. https://standard.openownership.org/en/0.4.0/standard/modelling/record-ids.html
35. https://www.gov.uk/hmrc-internal-manuals/cotax-manual/com40011
36. https://www150.statcan.gc.ca/n1/pub/12-605-x/12-605-x2017001-eng.htm
37. https://hl7.org/fhir/R4B/valueset-link-type.html
38. https://www.iso.org/standard/78829.html
39. https://www.census.gov/programs-surveys/susb/technical-documentation/methodology.html
40. https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2080.5~2006-2016~Main%20Features~Linkage%20Results~21
41. https://ec.europa.eu/digital-building-blocks/sites/spaces/DIGITAL/pages/210798097/Business%2BRegisters%2BInterconnection%2BSystem%2Bdashboard
42. https://ised-isde.canada.ca/site/corporations-canada/en/data-services
43. https://developer-specs.company-information.service.gov.uk/streaming-api/guides/overview

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.