Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

How to Reconcile EPA and OSHA Records for Target Mapping

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 · 2026-10-01 · epa osha reconciliation · facility records · establishment search · epa echo · frs · industrial target mapping · entity resolution · acquisition screening · record linkage

Original article: https://docket.capital/articles/reconcile-epa-osha-records-industrial-target-mapping



In this article

1. Executive Summary
2. Introduction and Background
3. Define the Unit Before Joining Records
4. EPA and OSHA Source-Field Crosswalk
5. A Reproducible Reconciliation Protocol
6. Resolve Multiplicity, Renames, and Shared Addresses
7. Data Analysis and Evidence
8. Implications and Future Directions
9. Frequently Asked Questions (FAQs)
10. Conclusion

Executive Summary

To reconcile EPA and OSHA records for an acquisition target map, separate **source rows**, **physical sites**, **legal entities**, and the **company counted as a prospect**. The Census Bureau defines an establishment as a single physical place of business and says that a multi-establishment company is one enterprise (www.census.gov). EPA's Facility Registry Service (FRS) connects program records about the same facility, whereas OSHA's activity number identifies an inspection, not an enduring company (echo.epa.gov) (www.osha.gov). A same-site conclusion and a same-company conclusion therefore require different evidence.

Begin with raw names, physical addresses, source IDs, dates, program and North American Industry Classification System (NAICS) codes, and available coordinates. Preserve those values before standardizing them. Use FRS program linkages as the strongest EPA-internal join; search OSHA name variants and compare its inspection site with the EPA physical facility, not an office or mailing address (echo.epa.gov) (www.osha.gov) (www.epa.gov). Treat coordinates, phones, and NAICS as supporting clues. OSHA says its establishment names are not unique (www.osha.gov). Escalate the operator or parent question to dated permits, filings, and company evidence, leaving unresolved links visible.

Official data are large but describe different universes. EPA's ECHO Exporter covers more than **1.5 million regulated facilities** and more than **130 fields** per facility, while OSHA describes more than **3 million inspections since 1972** (echo.epa.gov) (www.osha.gov). EPA says FRS receives data from **90 systems**, and its combined CSV is updated monthly (www.epa.gov). These counts are not a market denominator. ECHO's source data can take **one week to three months** to appear, and its pre-**November 2000** national-system data quality has not been assessed (echo.epa.gov). OSHA's Establishment Search showed inspection data through **September 29, 2026** at this report's publication date (www.osha.gov).

The recommended deliverable is a **reversible worksheet** with raw and normalized addresses, FRS/program and OSHA inspection IDs, candidate legal entity, match rule, confidence category, conflicts, dates, and evidence URLs. Count deterministic, reviewed, and unresolved links against a declared source-row denominator. For example, **64**, **26**, and **10** rows out of a fictional **100-row** batch yield **64%**, **26%**, and **10%** respectively; the accepted site-link rate is **90%**. Those are illustrative calculations, not observed performance or coverage. The worksheet protects origination teams from both duplicate prospects and false merges, while reserving environmental and workplace-safety conclusions for separate diligence.

Introduction and Background

A search for an industrial acquisition candidate can return an Environmental Protection Agency (EPA) facility, an Occupational Safety and Health Administration (OSHA) inspection, and a state business registration under three different names. The immediate origination question is how many target companies those records represent. It cannot be answered by counting search hits. The Census Bureau defines an establishment as a **single physical location** and says it need not be identical with a company or enterprise (www.census.gov). EPA's Facility Registry Service (FRS) groups site information from program systems, while OSHA's Establishment Search locates inspections by establishment name (echo.epa.gov) (www.osha.gov).

The useful output is a reversible **site-to-entity crosswalk**. It links source rows to a physical site, links that site to a candidate operating legal entity only when separate evidence supports the claim, and then links entities to an acquisition target record under a stated universe rule. For a private equity origination team, the crosswalk prevents a multi-plant manufacturer from becoming several prospects and prevents neighboring operators from being merged into one. It also preserves unresolved records for later research. A facility record alone cannot establish ultimate ownership, transaction perimeter, current operating status, or environmental and workplace-safety conclusions. California's business registry, for example, says that ownership and subsidiary details are not among the information it records (www.sos.ca.gov).

The public data are substantial but bounded. EPA says the ECHO Exporter includes more than **1.5 million regulated facilities** (echo.epa.gov). OSHA says its industry inspection database holds information on more than **3 million inspections since 1972**, a count of inspection events rather than unique companies (www.osha.gov). The protocols below use these sources for identity mapping, not as a denominator for market coverage. They answer the operational question: when does an EPA row, an OSHA row, and a corporate name describe one site, one company with multiple sites, or separate businesses?

Define the Unit Before Joining Records

Four units that must remain separate

A **legal entity** is the registered person or organization identified in a jurisdiction's records. A **company or enterprise** is the business grouping used for the acquisition universe. The Census definition groups one or more domestic establishments under common ownership or control; a multi-establishment company is one enterprise (www.census.gov). A **firm** is a term whose operational definition changes by dataset: the Bureau of Labor Statistics (BLS) calls it a legal business in one series, while Census applies its own geographic and industry conditions (www.bls.gov) (www.census.gov). The map therefore needs an explicit target-universe rule, such as one prospect per operating company, and it must not silently adopt every agency's unit of analysis.

An **establishment** is a workplace at a physical location. An **EPA regulated facility** is a site represented in one or more environmental program systems. An **OSHA inspection record** is an event; its activity number identifies an inspection, not a company (www.osha.gov). Multiple inspections may refer to one workplace. Multiple EPA program identifiers may also refer to one FRS site; EPA warns that an FRS EZ Query result can repeat the same facility for each program-system identifier (www.epa.gov).

Table 1 establishes the units and the permitted inference from each source row.

Unit in the map	Typical key or evidence	What it can establish	What it cannot establish alone
Legal entity	Jurisdiction plus registry number, with dated filing (www.sos.ca.gov)	A registered entity name and filing identity	Whether that entity runs a given plant
Company or enterprise	A documented grouping of operating entities and sites (www.census.gov)	The target-counting unit chosen for the mandate	That every same-name record is part of the group
EPA facility	FRS Registry ID and linked program IDs (www.epa.gov) (echo.epa.gov)	A cross-program site record	The current ultimate owner or acquisition perimeter
OSHA inspection	Activity number (www.osha.gov)	A dated inspection event	A permanent establishment or company identifier
Physical location	Normalized street address, location evidence, and site context (geocoding.geo.census.gov)	A candidate shared site	A shared operator, employer, or parent

The key interpretation is directional. A high-confidence **site match** can coexist with an unresolved **company match**. A high-confidence company match can contain several FRS IDs and many OSHA activity numbers. The Census and BLS definitions both permit one company to have several establishments (www.census.gov) (www.bls.gov). These relationships should be stored as separate edges rather than reduced to one deduplicated name.

A name is a clue, not a key

EPA says facility names can differ across its databases, and OSHA says establishment names are not unique (echo.epa.gov) (www.osha.gov). A plant can be listed under a former operator, a trade name, a parent name, or a local name. New York's assumed-name rules illustrate why a trading name can differ from an entity's true legal name (dos.ny.gov). The researcher should retain all observed names with their source and observation date. Do not overwrite a raw source name with a chosen canonical name.

Addresses need the same discipline. The Postal Service defines a standardized address through required elements and warns that varying formats impede duplicate detection (pe.usps.com). Normalize street suffixes, unit designators, city, state, and ZIP for comparison, while preserving the original string. EPA's FRS steward guidance says a facility location should not be a post-office box, headquarters, or administrative office (www.epa.gov). California's registry labels an entity address as its executive office, which makes it a poor substitute for a plant address (www.sos.ca.gov). A headquarters-to-plant mismatch is a reason to investigate, not a reason to reject a company link.

EPA and OSHA Source-Field Crosswalk

What the EPA side supplies

ECHO, EPA's Enforcement and Compliance History Online system, is a practical starting point when a researcher has a street address but no FRS identifier; EPA explicitly recommends that route for business users (www.epa.gov). The FRS Registry ID is unique across multiple EPA systems, and program-link files connect FRS records to source program identifiers (www.epa.gov) (echo.epa.gov). The downloadable state CSVs include facility name, address, geospatial fields, program associations, and industry classifications; the combined package also supplies alternative names, contacts, and mailing-address files (www.epa.gov).

An FRS ID is strong **within EPA's facility graph**, but it is not a universal business ID. EPA describes FRS as an attempted site match (www.epa.gov). Its ECHO FRS download is explicitly a subset, not the complete FRS database (echo.epa.gov). FRS also includes a limited OSHA-OIS source slice: EPA describes inspections within the last **seven years** in mining, oil and gas, utilities, and manufacturing (www.epa.gov). If a published FRS program link directly connects an OSHA identifier to the searched record, that is useful site-level evidence. Absence of such a link cannot establish that OSHA has no record for the site, because the FRS slice is restricted.

EPA's location fields require context. ECHO says displayed coordinates may be for the facility or permit holder, and EPA says FRS locations come from program systems, states, regional offices, and geocoding routines (echo.epa.gov) (www.epa.gov). A coordinate that appears precise may therefore be derived from an address rather than surveyed at a building. ECHO coordinate search excludes facilities lacking FRS latitude and longitude; a missing map point is not evidence that a facility does not exist (echo.epa.gov).

What the OSHA side supplies

OSHA's Establishment Search finds inspection records by establishment name, and its page showed data through **September 29, 2026** at this report's publication date (www.osha.gov). For a candidate, capture the raw establishment name, inspection site address, activity number, inspection date, and any available industry code. OSHA defines Site Address as the place where inspection occurred and Activity Number as a unique inspection identifier (www.osha.gov). Industry Search accepts a **six-digit North American Industry Classification System (NAICS)** code, but that code describes activity, not ownership (www.osha.gov) (www.census.gov).

OSHA also publishes electronically submitted injury and illness data, but those submissions do not cover all establishments (www.osha.gov). Their identifiers should not be carried into ECHO. A missing submission is not evidence that a workplace does not exist.

Table 2 is the **source-field crosswalk** for a reconciliation worksheet. It records what to copy, what to normalize, and where to escalate. Docket's documented target-research process keeps source excerpts and reviews conflicting findings; the row describes a possible research-record layer, not a government identifier.

Source or record layer	Copy as raw evidence	Normalize or derive	Join use and limit
EPA ECHO / FRS	Facility name and address; program associations; NAICS; coordinates; source date (www.epa.gov)	Site address, name aliases, coordinate provenance	FRS and program IDs join EPA records; do not infer owner
OSHA inspection search	Inspected employer; inspection site; activity number; date (www.osha.gov)	Site address and employer-name variants	Activity number identifies an inspection, not a company (www.osha.gov)
State entity register	Legal name; jurisdiction; entity number; dated filing; office address (www.sos.ca.gov)	Jurisdiction-qualified entity key	Verifies entity identity; office address may differ from plant
State permit or applicant record	Permit or authorization ID; facility and applicant names; dated document (dec.ny.gov) (www.pa.gov)	Candidate operator and period of validity	Can bridge site and named operator, subject to document review
Docket target-research record	Source URL, excerpt, collection date, competing names and decision history	Candidate entity key and accepted/rejected links	A review layer; its conclusion must point back to external evidence

The crosswalk separates **source identifiers** from **derived decisions**. It also keeps program-system IDs and inspection activity numbers in separate columns, so an analyst cannot accidentally join two unrelated numeric strings. A legal-entity number should be qualified by jurisdiction; a company name should remain an alias until corroborated. The worksheet below adds a match rule, confidence category, conflicts, and evidence URLs for every candidate pair.

A Reproducible Reconciliation Protocol

Build candidates without merging them

Start with the acquisition thesis and a target record that contains every known company name, domain, legal name, and plant location. Search ECHO by name, address, FRS ID, and any program identifier; ECHO supports Registry ID and program-system ID searches (echo.epa.gov). Search OSHA Establishment Search by the raw name and likely variants, then examine site addresses and activity numbers. EPA's FRS query can also search facility names, addresses, and affiliated organizations (www.epa.gov). This is **candidate generation**, not acceptance.

For each source row, record an immutable observation: source system, URL, retrieval date, source update date where given, raw name, raw street address, city, state, postal code, source ID, program, NAICS, coordinates, and event or permit date. Retain the exact result and the search terms that produced it. A date beside the EPA record is not automatically the operating date of the plant; ECHO describes its data as snapshots of source databases and says source entries can take **one week to three months** to appear (echo.epa.gov). OSHA says its industry database updates daily, while EPA lists monthly combined FRS CSV downloads and varying source-system refresh cycles (www.osha.gov) (www.epa.gov).

Next, create a comparison version of the address. Standardize directional and street-type tokens, separate unit numbers, and retain whether a value is a physical address or a mailing address. A geocoder can supply candidate coordinates, but record its service, benchmark, and query date: Census says its Current benchmark changes with underlying data, and its coordinates may be interpolated from address ranges (geocoding.geo.census.gov) (cdn.www.census.gov). Coordinate proximity is a candidate-generation feature. It is never a substitute for the site description, parcel or building context, and operator evidence.

A workable candidate-generation sequence is:

1. **Collect exact identifiers.** Preserve FRS Registry ID, EPA program-system ID, OSHA activity number, state permit ID, and jurisdiction-qualified entity number; keep each in its own field.
2. **Generate EPA internal links.** Use FRS program-link files before fuzzy name matching; retain every original program record and its FRS assignment (echo.epa.gov).
3. **Search name variants.** Try former names, abbreviations, plant labels, and alternate spellings because OSHA documents variant establishment names (www.osha.gov).
4. **Compare physical addresses.** Distinguish the inspected site from the EPA facility, mailing, permit-holder, and headquarters locations (www.epa.gov).
5. **Inspect geography.** Compare coordinates and geocoder provenance, then check whether a shared street address covers several suites, parcels, buildings, or operators (cdn.www.census.gov).
6. **Use supporting fields.** Compare phone, industry, and dated applicant names as clues, not as a single decisive key (www.bls.gov).
7. **Propose entity links separately.** Consult jurisdictional filings and dated operator evidence; never promote a site match to a parent-company conclusion without that step (www.sos.ca.gov) (dec.ny.gov).

Decide with explicit rules

A **deterministic site link** needs a published crosswalk or source-assigned identifier whose semantics match both records. Within EPA, an FRS-to-program-system linkage meets this test. A documented FRS linkage to a relevant OSHA-OIS record can support a site link for that record, with the limited scope described above (echo.epa.gov) (www.epa.gov). An activity number alone cannot identify an enduring facility. Identical names or industry codes do not prove a cross-system match.

A **reviewed site link** can be accepted when independent evidence converges: the normalized physical address is compatible, the plant or employer names are plausible variants, the coordinates are explainable, and a dated permit, operator page, or filing resolves the remaining identity question. BLS's own historical linkage project reviewed erroneous common-ID matches and used names, addresses, phone numbers, geography, industry, and secondary sources for unmatched records (www.bls.gov). That is evidence for a documented review process, not a statistical accuracy guarantee for this worksheet. A text-string comparison can even increase matches while reducing accuracy, as Census research warns (www.census.gov).

A **conflicted or unresolved link** remains open when the address is shared, the operator dates conflict, the name belongs to a parent rather than the workplace, or the only commonality is industry. Write down the conflicting sources and the next document needed. This status protects the target universe from a false merge. The research record should permit later correction without discarding the earlier finding; Docket's published process explicitly retains such history.

For a reproducible review queue, set internal thresholds before reviewing cases. One possible rule is: accept a same-site candidate for manual review only when state and street number agree and at least one of plant-name variant, permit identifier, or explainable coordinates also agrees. Treat a conflicting suite, distinct operator, or incompatible chronology as an escalation. These are **proposed workflow thresholds**, not EPA or OSHA standards. Store the exact rule version so a changed threshold can be reapplied consistently.

Downloadable worksheet and evidence package

The [blank reconciliation worksheet](data:text/csv;charset=utf-8,source_system%2Csource_url%2Cretrieved_date%2Csource_record_date%2Craw_name%2Craw_address%2Cnormalized_address%2Csource_facility_or_establis is a CSV header that can be saved as `epa-osha-reconciliation.csv`. Create one row per source observation and a separate decision row or table for each proposed relationship. The fields are: `raw_name`, `raw_address`, `normalized_address`, `source_facility_or_establishment_id`, `frs_registry_id`, `epa_program_system_id`, `osha_activity_number`, `state_permit_id`, `jurisdiction_qualified_entity_number`, `program`, `naics`, `latitude`, `longitude`, `inspection_or_permit_date`, `candidate_legal_entity`, `match_rule`, `confidence_category`, `conflicts`, and `evidence_urls`, plus source and retrieval dates. Those provenance fields matter because EPA and OSHA update on different cycles (www.epa.gov) (www.osha.gov).

Keep a second edge table with `left_source_id`, `right_source_id`, `relationship` (`same_site`, `operated_by`, or `part_of_target`), `status`, `decision_date`, reviewer, and evidence URLs. This structure lets one FRS site have several EPA program records, one workplace have several inspection events, and one company operate several facilities. It also makes a rejected match auditable. Do not erase the raw records when an edge changes.

Resolve Multiplicity, Renames, and Shared Addresses

One company can map to several sites

The most common counting error is treating every site as a separate acquisition target. Census counts a multi-establishment company as one enterprise, and BLS likewise describes a company as potentially containing multiple establishments (www.census.gov) (www.bls.gov). The target map should roll accepted sites up to the chosen company key only after an entity relationship is supported. An FRS ID represents a site across EPA systems, not an enterprise key (www.epa.gov). Multiple inspection events should remain attached to that site rather than counted as companies.

There is a reverse error: an address can contain several operators. EPA's own FRS practice says a facility name should describe what the facility is rather than who owns it (www.epa.gov). New York's environmental mapping documentation explicitly notes some separately classified facilities are co-located (dec.ny.gov). A shared street address or nearly identical coordinates can therefore mean shared property, campus, service yard, or geocoder approximation. When the operator, suite, permit, and inspected employer do not align, preserve separate candidate entities and flag the site relationship as `co_located`, not `same_company`.

A rename creates a third pattern. New York describes a permittee-name or facility-ownership change as a permit-transfer matter, which provides a dated document path for review (dec.ny.gov). State business filings may show an assumed name or entity name change; a trade name is not a separate company solely because the text differs (dos.ny.gov). The researcher should date the old and new names and ask whether the legal entity continued, a new entity took over, or only the site label changed. Those are separate conclusions.

Table 3 illustrates how the decision rules operate. Every name and address in it is **fictional**; the table is a worked worksheet example, not a claim about an actual company.

Fictional scenario	Observed candidate rows	Decision at site level	Decision at company level and next source
Renamed plant (Hypothetical Example)	EPA: “Cedar Forge Works,” 410 Mill Road; OSHA: “Cedar Forge Components,” same road, later inspection	Reviewed same-site candidate after checking building and time	Hold company merge until dated filing or permit shows whether operator changed (dec.ny.gov)
Parent and plant collision (Hypothetical Example)	EPA: “Northline Finishing Plant,” local address; OSHA: “Northline Holdings,” same site; registry: holdings office elsewhere	Likely same physical workplace if inspected site matches EPA address	Do not make the parent a second plant or assume it is the permittee; examine operator evidence (www.sos.ca.gov)
Shared industrial address (Hypothetical Example)	EPA: “Harbor Coatings,” Unit A; OSHA: “Harbor Packaging,” Unit B; both geocode to 86 Wharf Street	Keep distinct site/operator candidates despite similar coordinates	Keep two target candidates pending suite, permit, and entity checks (cdn.www.census.gov) (dec.ny.gov)

The table shows why **one-to-many and many-to-one links** are normal. It also shows the limit of a confidence label: “reviewed same site” says nothing about the legal owner. Every accepted relationship must name the evidence it relies on, the period it covers, and the claim it does not resolve.

Escalate to the record that can answer the question

When an EPA or OSHA result leaves operator identity unclear, use the next source according to the missing fact. A state permit portal can expose applicant and facility names; New York's search supports those fields, and Pennsylvania's eFACTS searches by client name, permit number, authorization ID, or client ID (dec.ny.gov) (www.pa.gov). Pennsylvania separately defines a responsible-party address, another reason not to merge it with a site address (www.pa.gov). Request the dated permit or application, rather than relying only on a current search summary.

For legal identity, use the appropriate state registry and the filing image when available. Search behavior varies: Pennsylvania says its records require the correct entity name or number, while Delaware's lookup exposes a file number and formation date (www.pa.gov) (icis.corp.delaware.gov). The Securities and Exchange Commission's Central Index Key (CIK) identifies a filer, not a plant, and the agency warns that names can differ from expectations (www.sec.gov). A Legal Entity Identifier (LEI) parent link may add an accounting-consolidation relationship when both entities have identifiers, but GLEIF also records exceptions where a parent lacks an LEI (www.gleif.org). None of these systems should be treated as a universal private-company ownership register.

A company website can establish its current presentation of a plant or operator, but retain the dated page and corroborate it with a filing or permit when the entity distinction changes the target count. If the official records do not resolve the issue, leave the link unresolved and state exactly which document is needed. This is a valid mapping outcome, not a failed workflow.

Data Analysis and Evidence

Dataset sizes measure different things

The available official figures are useful for planning ingestion, not for estimating the addressable acquisition market. ECHO says its Exporter covers more than **1.5 million regulated facilities** and carries more than **130 fields** per facility (echo.epa.gov). EPA says FRS receives site data from **90 systems**, and its combined FRS CSV download is updated monthly (www.epa.gov). OSHA says its industry database contains information on over **3 million inspections since 1972** and updates daily from more than **120** area and state plan offices (www.osha.gov). The Bureau of Labor Statistics describes a different employment-statistics universe of more than **12 million establishments** (www.bls.gov). Facility rows, inspection events, and employment establishments have different inclusion rules. Their counts are not interchangeable denominators.

Freshness must be measured at several layers. The ECHO Exporter is updated weekly, while the combined FRS CSV is listed as monthly (echo.epa.gov) (www.epa.gov). OSHA's Establishment Search showed inspection data through **September 29, 2026** when checked for this report (www.osha.gov). A current download date still does not make every underlying name or coordinate current. ECHO describes source-data snapshots and possible lags of **one week to three months**; EPA says it has not assessed quality of national-system data before **November 2000** (echo.epa.gov). Record retrieval time, source update time, inspection or permit time, and the period for an operator claim in separate fields.

The worksheet's **match-rate calculation** uses a defined source-row denominator. For a hypothetical batch of **100 OSHA source rows** against an EPA candidate set, suppose **64** rows have accepted deterministic site links, **26** have accepted links after documented review, and **10** remain unresolved. Deterministic rate = `64 / 100 = 64%`; reviewed rate = `26 / 100 = 26%`; unresolved rate = `10 / 100 = 10%`. Accepted site-link rate = `(64 + 26) / 100 = 90%`. These numbers are **hypothetical arithmetic**, not observed data or a benchmark. Count a row once in its final category; report the denominator, time window, sector filter, and duplicate handling with the rates. If one OSHA row has several plausible EPA candidates, it remains unresolved until the edge is decided.

Do not call that **90%** market coverage, company coverage, or compliance quality. The calculation measures the fate of the imported rows under this protocol. It says nothing about establishments absent from OSHA search, unregulated sites absent from ECHO, or companies outside the source list. Public source universes have their own inclusion rules; for example, EPA calls the ECHO FRS linkage file a subset rather than a complete FRS download (echo.epa.gov). The same caveat applies to an unmatched count: it measures unresolved source records, not the number of missing businesses.

Audit metrics that actually inform origination

- **Site-link disposition:** deterministic, reviewed, conflicted, or unresolved, with a source-row denominator and decision date.
- **Entity-link disposition:** confirmed operating entity, candidate entity, multiple plausible entities, or unknown; do not reuse the site-link confidence label.
- **Multiplicity:** distinct facility IDs per accepted company and distinct inspection events per accepted site.
- **Alias load:** raw names retained per site and per entity, including discarded variants.
- **Staleness:** elapsed time since source retrieval and since the underlying event, tracked separately.

- **Review yield:** accepted reviewed links divided by reviewed candidates, with rejected candidates retained for audit; BLS's manual linkage experience shows why error review matters (www.bls.gov).

These measures tell an analyst where the target map is fragile. They are process metrics. They cannot be converted into claims about environmental condition, workplace safety, investment quality, or the fraction of a market found.

Implications and Future Directions

The key design choice is to maintain **three connected but distinct layers**: source observation, physical site, and acquisition target company. That model accommodates cross-program facility links and one-company-to-many-establishments relationships (echo.epa.gov) (www.census.gov). It also allows a site to have successive operators or multiple co-located businesses without forcing an irreversible merge. As a target universe grows, the same rule version can be rerun on new records, and prior decisions can be inspected rather than inferred from a single cleaned name.

Data pipelines should store the raw result alongside every normalized field. EPA's alternative-name files make it valuable to search more than one string (www.epa.gov). Address standardization improves comparison, but the original unit, suite, and address type still matter. Geocoding should add a provenance column and a review cue rather than overwrite the address: Census warns that coordinates may be approximated and that the benchmark changes over time (cdn.www.census.gov) (geocoding.geo.census.gov).

The largest future improvement is a better **evidence bridge** between site and operator. Dated permit applications, registry filings, and company statements can support such a bridge, but each answers a different question. New York permits can be searched by applicant and facility, whereas a California entity address is defined as an executive office (dec.ny.gov) (www.sos.ca.gov). Teams should write a source-escalation rule that chooses the document based on the disputed fact, then retain both confirmations and contradictions. This lowers the risk of double-counting without implying a legal or diligence conclusion.

Source schedules and display lags should inform the refresh queue (www.epa.gov) (echo.epa.gov). A fixed calendar review can be supplemented by triggers: a new permit applicant, changed establishment name, new entity filing, or a previously unresolved candidate acquiring a direct crosswalk. The decision log should record what changed and why the accepted edge changed.

Frequently Asked Questions (FAQs)

Are EPA facility records and OSHA establishment records the same unit?

No. FRS links program records at a facility, while OSHA's activity number identifies an inspection (echo.epa.gov) (www.osha.gov). Map each result to a physical site before deciding which company the site belongs to. The Census Bureau permits several establishments under one enterprise (www.census.gov).

How should an analyst match an EPA facility to an OSHA establishment?

First use a published program linkage where one exists. Otherwise, compare physical address, dated name variants, site context, and coordinates with provenance. OSHA itself warns that establishment names are not unique (www.osha.gov). A same-site match still needs separate evidence for the operator.

Does a facility or establishment hit identify the company that could be acquired?

No. EPA distinguishes a facility name from an owner's name, and a state filing may identify an office rather than a plant (www.epa.gov) (www.sos.ca.gov). Use dated permits, corporate filings, and the company's own site information to resolve the operating entity. Keep ownership unresolved if those sources do not establish it.

How should duplicate industrial facility records be counted?

Keep one source row per observed record, one accepted site per supported physical location, and one target-company record per the mandate's company definition. EPA says its EZ Query can repeat facility information for each program-system identifier (www.epa.gov). Do not turn repeated source rows into separate prospects.

Can environmental and workplace-safety data screen acquisition targets?

They help find locations and aliases and identify documents worth retrieving. ECHO's source snapshots can lag, so the results alone cannot establish present operator, environmental condition, workplace-safety quality, or acquisition attractiveness (echo.epa.gov). Record unresolved links and send specialist questions to later diligence.

Conclusion

Reconciling EPA and OSHA records is a **record-linkage task across different units of analysis**. EPA's FRS ID joins environmental program records at a site, and the OSHA activity number identifies an inspection (www.epa.gov) (www.osha.gov). Neither is automatically a private-company identifier. Census's establishment and enterprise definitions explain why one company can appear at many locations and why a shared name or address is not a company count (www.census.gov).

A defensible target map keeps raw source observations, normalized site candidates, and legal-entity decisions separately. It uses explicit ID links where available, physical-address and location evidence for candidate sites, and dated permits, filings, and company evidence for operator and parent questions. It labels deterministic, reviewed, and unresolved outcomes and retains rejected links. The worksheet and worked examples turn those rules into an auditable process that can be refreshed when public records change. Its match rates describe the reviewed batch only. They are not evidence of market coverage, legal ownership, environmental quality, or workplace-safety quality. The next step for an origination analyst is to work the unresolved queue: obtain the missing permit, filing, or operator page, record the decision date, and revise only the relationship that the new evidence supports. That discipline keeps company counts usable even when source names and site records change.

Source links

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

1. https://www.census.gov/programs-surveys/susb/about/glossary.html
2. https://echo.epa.gov/tools/data-downloads/frs-download-summary
3. https://www.osha.gov/data/inspection-detail-definitions
4. https://www.osha.gov/help/establishment-search
5. https://www.epa.gov/frs/facility-linkage-application-fla-best-practices
6. https://echo.epa.gov/tools/data-downloads
7. https://www.osha.gov/help/industry-search
8. https://www.epa.gov/frs/frs-description
9. https://echo.epa.gov/resources/echo-data/about-the-data
10. https://www.osha.gov/ords/imis/establishment.html
11. https://www.sos.ca.gov/business-programs/business-entities/information-requests
12. https://www.bls.gov/bdm/business-employment-dynamics-data-by-age-and-size.htm
13. https://www.epa.gov/frs/frs-ez-query
14. https://www.sos.ca.gov/business-programs/business-entities/cbs-field-status-definitions
15. https://www.epa.gov/frs/search-frs-data
16. https://geocoding.geo.census.gov/geocoder/Geocoding_Services_API.html
17. https://www.bls.gov/opub/hom/cew/concepts.htm
18. https://echo.epa.gov/help/facility-search/all-data-search-results-help
19. https://dos.ny.gov/node/35461
20. https://pe.usps.com/TEXT/pub28/28c1_004.htm
21. https://www.epa.gov/frs/epa-frs-facilities-state-single-file-csv-download
22. https://www.epa.gov/frs/frs-data-sources
23. https://www.epa.gov/frs/geospatial-overview
24. https://echo.epa.gov/help/facility-search/search-criteria-help
25. https://www.osha.gov/itadata
26. https://dec.ny.gov/regulatory/permits-licenses/environmental-permits/getting-an-environmental-permit
27. https://www.pa.gov/agencies/dep/programs-and-services/land/storage-tanks/storage-tank-data
28. https://www.epa.gov/frs/frs-query-forms-user-guide
29. https://www.epa.gov/frs/frs-data-download-options
30. https://cdn.www.census.gov/programs-surveys/geography/technical-documentation/complete-technical-documentation/census-geocoder.html
31. https://www.bls.gov/fdi/technical-notes
32. https://www.census.gov/library/working-papers/2016/adrm/carra-wp-2016-05.html
33. https://dec.ny.gov/maps/interactive-maps/decinfo-locator/layers
34. https://dec.ny.gov/regulatory/permits-licenses/wastewater-stormwater-water-withdrawal/spdes-permit-program
35. https://www.pa.gov/agencies/dos/programs/business/information-services/record-searches
36. https://icis.corp.delaware.gov/eCorp/EntitySearch/NameSearch.aspx
37. https://www.sec.gov/submit-filings/filer-support-resources/how-do-i-guides/look-central-index-key-cik-number
38. https://www.gleif.org/en/lei-data/access-and-use-lei-data/level-2-data-who-owns-whom

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.