

PE Manufacturing Intelligence Methodology & Data Provenance

Companion to the Q1 2026 Report (updated April 23, 2026)

Prepared by BOMForge, Inc.

Date April 23, 2026

Coverage 2020-01-02 through 2026-04-23 (2,303 days)

Transactions analyzed 3,919

Matched to manufacturer index 3,836 (97.9%)

Primary source SEC EDGAR Full-Text Search API (EFTS)

Match target BOMForge manufacturer index (1.17M US + international rows)

BOMFORGE

Why This Document Exists

The PE Manufacturing Intelligence report publishes counts, ratios, and sector verdicts. Every one of those numbers is a function of a search parameter, a filter, a regex, or a matching rule. This document makes each of those choices explicit so that:

- A reader can reproduce any headline figure from public data in under an hour.
- An analyst can evaluate whether the framing is load-bearing for the conclusion.
- A customer can cite the report in their own work and defend it downstream.
- Future refreshes converge on the same answer when the inputs have not changed.

13

EDGAR SEARCH QUERIES

2000 to 3999

SIC CODE FILTER RANGE

7

MATCH METHODS (CASCADE)

97.9%

MATCH RATE TO INDEX

I. Pipeline Overview

1

Query the SEC EDGAR Full-Text Search API

Thirteen term-sets run against forms 8-K and SC 13D. Rate limited to roughly 1.5 requests per second, well under SEC's 10 per second ceiling. Pagination caps at 500 hits per query.

2

Filter to manufacturing SIC codes

Keep only filings whose `sics` array contains at least one code in the range 2000 through 3999. Filings tagged with multiple SIC codes are attributed to the lowest-numbered manufacturing code present.

3

Deduplicate across queries by accession number

The same 8-K can match several term-sets. A Set keyed on `adsh` (EDGAR accession) prevents double-counting before any parsing happens.

4

Fetch the filing document and extract target

For the first 200 parseable filings per run, fetch the primary 8-K document, strip HTML, and apply eight regex patterns to extract the acquisition target. Bankruptcies and plant closures treat the filer as the target by construction.

5

Match target to manufacturer index

Seven-tier cascade: exact name, DBA name, fuzzy first-significant-word plus state, plus four backfill methods that populate from later enrichment runs. Stubs created for high-confidence distress filings that have no index row yet.

6

Upsert with partial unique index

Dedupe key is `(source_filing_id, event_type)`. Partial index excludes rows with null filing IDs to prevent ghost duplicates on incremental backfills.

7

Aggregate and publish

The analysis script (`scripts/pe-analysis-report.ts`) reads the full transaction table and emits the JSON payload that drives the PDF renderer.

II. Search Parameters (Exact)

Each query hits the EDGAR EFTS endpoint at <https://efts.sec.gov/LATEST/search-index> with these shared parameters:

```
forms:      8-K (12 queries) or SC 13D (1 query)
dateRange:  custom (startdt..enddt)
size:       100 per page (EDGAR max)
User-Agent: ResearchBot/1.0 build@bomforge.com
Rate limit: 600ms between requests (~1.5 req/sec)
```

The thirteen term-sets below are the complete query surface. Every headline count in the report traces back to one or more of these.

Acquisition queries (form 8-K, Item 2.01)

```
"completion of acquisition" "manufacturing"
"acquired" "manufacturing" "purchase price"
"private equity" "acquisition" "manufacturing"
```

Add-on / bolt-on queries (form 8-K)

```
"add-on acquisition" "manufacturing"
"bolt-on" "manufacturing" "acquisition"
```

Bankruptcy queries (form 8-K)

```
"chapter 11" "manufacturing"
"chapter 7" "manufacturing"
"bankruptcy" "manufacturing" "petition"
```

Plant closure queries (form 8-K)

```
"plant closure" "manufacturing"
"facility closure" "manufacturing"
"ceased operations" "manufacturing"
```

Ownership stake query (form SC 13D)

```
"beneficial ownership" "manufacturing"
```

Carve-out / divestiture query (form 8-K)

```
"divestiture" "manufacturing" "completion"
```

Coverage limitation. This query surface captures SEC registrants that file 8-Ks. It does not capture fully private PE deals that never touch EDGAR. That blind spot is what drives our roadmap item to ingest GlobeNewswire and Business Wire RSS feeds next. For 2025, Capstone Partners recorded 4,147 announced US manufacturing M&A transactions; BOMForge detected 752 via this pipeline. The 18% coverage figure should be treated as the lower bound on market visibility, not an indictment of completeness.

III. SIC Filter and Sector Mapping

Manufacturing is defined as SIC codes 2000 through 3999. A filing passes the filter if any code in its `sics` array falls in that range. Each filing is attributed to the lowest manufacturing code present, and that code's two-digit prefix drives sector classification.

Two-digit prefix to sector (exact mapping)

PREFIX	SECTOR	PREFIX	SECTOR
20	food products	30	rubber/plastics
21	tobacco	31	leather
22	textiles	32	stone/clay/glass
23	apparel	33	primary metals
24	lumber/wood	34	fabricated metals
25	furniture	35	machinery/computers
26	paper	36	electronics/electrical
27	printing/publishing	37	transportation equipment
28	chemicals	38	instruments/controls
29	petroleum/coal	39	miscellaneous manufacturing

Known conflation. SIC 28 (chemicals) is the single largest bucket by transaction count (1,170 events, 30% of the dataset). Roughly 40 to 60 percent of its bankruptcies are pharma and biotech issuers (Mallinckrodt, Acura Pharmaceuticals, Bristol-Myers Squibb subsidiaries), not traditional industrial chemistry. A roadmap item splits pharma/biotech (SIC 2834, 2836) from specialty and commodity chemicals in the next refresh.

IV. Target Extraction Regex Set

When the filer is an acquirer, we need the acquired company's name, which EDGAR metadata does not provide. The pipeline fetches the primary 8-K document, strips HTML, and runs eight regex patterns in order:

```
1. /(?:completed|completes|announces?)\s+(?:the\s+)?acquisition\s+of\s+
   ([A-Z] [A-Za-z\s&.,]+?) (?:\s*\[(,.)\]| \s+a\s+|\s+for\s+)/

2. /acquired\s+(?:substantially\s+all\s+(?:of\s+the\s+)?assets\s+of\s+)?
   ([A-Z] [A-Za-z\s&.,]+?) (?:\s*\[(,.)\]| \s+a\s+)/

3. /acquisition\s+of\s+([A-Z] [A-Za-z\s&.,]+?)
   (?:\s*\[(,.)\]| \s+a\s+|\s+for\s+)/

4. /(?:to\s+acquire|has\s+acquired|agreed\s+to\s+acquire)\s+
   ([A-Z] [A-Za-z\s&.,]+?) (?:\s*\[(,.)\]| \s+a\s+|\s+for\s+)/

5. /purchase\s+of\s+([A-Z] [A-Za-z\s&.,]+?)
   (?:\s*\[(,.)\]| \s+a\s+|\s+for\s+)/

6. /merger\s+with\s+([A-Z] [A-Za-z\s&.,]+?)
   (?:\s*\[(,.)\]| \s+a\s+|\s+for\s+)/

7. /divest(?:iture|ed)\s+(?:of\s+)?(?:its\s+)?([A-Z] [A-Za-z\s&.,]+?)
   (?:\s*\[(,.)\]| \s+a\s+|\s+division|\s+segment|\s+business|\s+unit)/

8. /(?:closed|closing)\s+(?:its\s+)?([A-Z] [A-Za-z\s&.,]+?)\s+
   (?:plant|facility|operations|manufacturing)/
```

First match wins. Candidates shorter than 3 characters or longer than 99 characters are rejected. When no pattern matches, the transaction is stored with `target_extracted=false` and the filer's name in the target slot as a placeholder.

Self-referential rows. 1,475 rows in the current dataset (38%) have `target_extracted=false`. These are transactions where the 8-K exists and the filer is manufacturing-classified, but the regex set failed to isolate a distinct target. This is the single largest source of noise in acquirer counts and is queued for a dedicated extraction pass before the next public refresh.

V. Manufacturer Matching Cascade

Once a target name is extracted (or, for distress filings, the filer is treated as the target), we cascade through seven matching methods. First hit wins. Confidence is a fixed per-method prior, not a learned score.

Method counts and confidence (current dataset)

METHOD	ROWS	SHARE	CONFIDENCE
backfill_created	1,417	36.9%	1.00
exact_name	990	25.8%	0.95
backfill_exact	772	20.1%	0.90
stub_created	437	11.4%	1.00
fuzzy_name_state	188	4.9%	0.60
name_relink	31	0.8%	0.95
backfill_dba	1	0.0%	0.85

Method semantics

- `exact_name` : case-insensitive `ILIKE` on `manufacturers.company_name` . Highest-precision match at detection time.
- `dba_name` : same match against the `dba_name` column for companies operating under trade names.
- `fuzzy_name_state` : first significant word of the target name plus a state equality filter. Only accepted when exactly one row comes back. Produces the most false positives.
- `backfill_exact` , `backfill_dba` , `backfill_created` , `name_relink` : late-bound matches that populate after the initial insert. Fire when a new manufacturer row is added or when enrichment scripts canonicalize a previously ambiguous name.
- `stub_created` : high-confidence distress filings with no index row at detection time trigger a minimal manufacturer stub. The stub carries `entity_source='edgar_pe_trace'` and is eligible for later enrichment.

Known bleed. 39 of 3,836 matched rows (1.0%) currently cite soft-deleted manufacturers. This is an artifact of matching logic that does not yet consult the lifecycle filter (`deleted_at` , `defunct` , `is_canary`). A one-line `matchToManufacturer` patch is queued for the next refresh and will drop the bleed to zero without changing any published number by more than a decimal point.

VI. Event Type Distribution

Every matched transaction carries exactly one `event_type`. The split in the current dataset:

EVENT TYPE	COUNT	SHARE
<code>acquisition</code>	1,712	43.7%
<code>plant_closure</code>	757	19.3%
<code>bankruptcy</code>	682	17.4%
<code>carve_out</code>	451	11.5%
<code>add_on</code>	317	8.1%

Event type definitions

- `acquisition`: completed or announced transfer of ownership of an operating manufacturer, detected via 8-K Item 2.01 or SC 13D.
- `add_on`: acquisition explicitly framed as an add-on or bolt-on to an existing PE platform. Narrower term-set than full acquisitions.
- `carve_out`: divestiture of a division, segment, business unit, or subsidiary from a larger filer.
- `plant_closure`: facility-level shutdown or cessation of operations disclosed by the operating entity. The plant or line stops; the company continues.
- `bankruptcy`: Chapter 11 or Chapter 7 petition disclosed via 8-K Item 1.03 or in body text. The corporate vehicle is the distressed entity; outcome can be restructure, sale, or liquidation.

The *restructure vs closure* companion report digs into the event-level outcome difference: Chapter 11 emergence, 363 sale, DIP financing, versus Chapter 7 liquidation or permanent plant shutdown: and why they are load-bearing for different investor audiences.

VII. Deduplication Strategy

- Per-run dedupe**: a `Set<string>` keyed on EDGAR accession number (`adsh`) rejects repeat hits within a single crawl before any parsing happens.
- Cross-run dedupe**: database has a partial unique index on (`source_filing_id`, `event_type`) WHERE `source_filing_id IS NOT NULL`. The partial predicate prevents a handful of legacy null-ID rows from blocking the constraint creation.
- Idempotent upsert**: inserts use `ON CONFLICT DO NOTHING`. A filing re-surfaced by a later date window never creates a duplicate row.
- Event-type split**: a single 8-K that triggers both an acquisition term-set and an add-on term-set would ordinarily produce two rows, but the per-run accession Set collapses them to one. The first matching term-set wins, which is why the `add_on` count is smaller than the actual bolt-on universe.

VIII. Deal Status Classification

- `completed`: filing's `items` array includes Item 2.01. The transaction has closed.
- `announced`: all other matches. The filing signals a transaction without having closed on the filing date. Some of these never close.
- `needs_review`: reserved for filings with ambiguous intent or failed extraction. Currently populated by backfill processes; the initial detection pipeline labels `completed` or `announced`.

In the current dataset, 6.9% are `completed`, 55.4% are `announced`, and 37.8% are `needs_review`. Ratios in the report (acquisition-to-distress, acquisition rate by state) are computed on the full table; `deal_status` is disclosed but not filtered.

IX. Geographic Attribution

Each transaction is tagged with `target_state` derived from the filing's `biz_locations` and `biz_states` fields. For bankruptcies and plant closures the filer is the target, so the filer's principal business address is attributed. For acquisitions the acquirer's state is attributed, not the acquired asset's state; this is a known limitation.

Non-US state codes. Roughly 207 rows (5.3%) carry non-US codes such as `L2`, `A6`, `V8`, `Bristol`, `Oxfordshire`, or `Dublin`. These are US-SEC filers with foreign business addresses, not US-state activity. The geographic analysis in the main report filters to two-letter codes matching `/^[A-Z]{2}$/` but does not separately expose the foreign cohort. A roadmap item splits US state from filer business location in the next refresh.

X. Reproducibility

The pipeline is a single script plus a single analysis script. Both run in under an hour on a laptop with reasonable bandwidth.

```
# Incremental (last 3 days)
bun scripts/detect-pe-transactions.ts

# Dry run: no writes
bun scripts/detect-pe-transactions.ts --dry-run

# Full backfill from a start date
bun scripts/detect-pe-transactions.ts --backfill 2020-01-01

# Regenerate aggregates
bun scripts/pe-analysis-report.ts --output data/pe-analysis-report.json

# Regenerate public PDF
bun scripts/generate-pe-report-pdf.ts \
  --data=data/pe-analysis-report.json \
  --md=docs/reports/pe-manufacturing-intelligence-q1-2026.md \
  --out=docs/reports/pe-manufacturing-intelligence-q1-2026.pdf

# Regenerate this methodology PDF
bun scripts/generate-pe-methodology-pdf.ts
```

Primary-source verification. Before publishing any refresh, we spot-check three to six marquee deals against primary sources (company IR, court dockets, SEC filings, White House publications). The April 23 refresh verified Stanley / Howmet CAM, CD&R / Sealed Air, Ascend Elements Chapter 11, Section 232 reset, and McCormick / Unilever. A "Verification status" note is pinned to the executive summary of each refresh.

XI. Known Caveats and Roadmap

Caveats (disclosed)

1. **Private deal blind spot.** SEC-filer-only surface. Roughly 82% of announced US manufacturing M&A falls outside this window.
2. **Self-referential rows.** 1,475 of 3,919 rows (38%) have the filer as target placeholder because the regex set could not extract a distinct target. Some of these are genuinely filer-is-target distress events; others are real acquisitions with unextracted targets.
3. **Soft-delete bleed.** 39 matched rows point at soft-deleted manufacturers. 1.0% bleed that does not move any published number.
4. **Pharma in chemicals.** SIC 28 includes drug manufacturers. 40 to 60 percent of the chemicals bankruptcy count is pharma/biotech, which is a different distress thesis than industrial chemistry.
5. **Foreign filer locations.** 207 rows carry non-US state codes and are currently counted in the sector totals but hidden from the US geographic tables.
6. **Acquirer-state attribution.** Acquisition rows are tagged with the acquirer's state, not the acquired asset's operating state. State-level acquisition rates are a rough proxy for capital flow, not plant location.
7. **Fuzzy match false positives.** 188 rows (4.9%) sit in the 0.60-confidence `fuzzy_name_state` bucket. These are the most likely sources of misattribution in any narrative that names a specific company.

Roadmap (queued for next refresh)

1. Lifecycle filter inside `matchToManufacturer` to eliminate the soft-delete bleed.
2. Target-extraction pass on the 1,475 self-referential rows. Expected yield: 40 to 50 percent become distinct targets.
3. US-state vs filer-business-location split in the geographic table.
4. Pharma/biotech subsector split under chemicals.
5. GlobeNewswire and Business Wire RSS ingest to close the private-PE blind spot.
6. Deal-value extraction productionized (regex already exists in `extractDealValue` ; not currently wired into the report aggregates).

XII. Data Dictionary

Every row in `pe_transactions` carries these fields; the public JSON export is a direct projection.

FIELD	TYPE	DESCRIPTION
acquirer_name	text	Filing entity for acquisitions; same as target for distress events.
target_name	text	Extracted via regex set (acquisitions) or filer name (distress).
target_extracted	bool	True when regex set isolated a distinct target, or by construction for distress events.
event_type	enum	One of acquisition, add_on, carve_out, plant_closure, bankruptcy.
deal_status	enum	completed, announced, needs_review.
sic_code	text	Four-digit SIC. Lowest manufacturing code present in the filing.
sector	text	Derived from SIC prefix via the table in section III.
target_state	text	Two-letter US code or foreign code (see section IX).
target_city	text	First segment of biz_locations .
source_type	enum	edgar_8k or edgar_13d.
source_filing_id	text	EDGAR accession number. Unique per filing.
source_cik	text	Filer's CIK (zero-padded removed).
announced_at	date	EDGAR file_date .
raw_headline	text	Generated summary for quick scan.
raw_excerpt	text	First 500 characters of stripped filing text, when parsed.
match_method	enum	One of seven values in section V.
match_confidence	numeric	Per-method prior; 0.60 to 1.00.
manufacturer_id	uuid	FK to manufacturers.id when matched.