Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

PPP Loan Data Due Diligence: Target Evidence and Limits

> 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 · 2026-10-13 · ppp loan data due diligence · ppp loan lookup · acquisition due diligence · private equity research · sba loan data · entity resolution · historical company data · ppp forgiveness · evidence standards

Original article: https://docket.capital/articles/ppp-loan-data-target-research



In this article

1. Executive Summary
2. Introduction and Background
3. Key Changes in How PPP Records Should Be Used
4. Data Provenance, Versions, and Record States
5. Field-to-Claim Matrix for Acquisition Research
6. Implementation Considerations and Process Changes
7. Time Limits, Quality Controls, and Ethical Review
8. Data Analysis and Evidence
9. Worked Reconciliation (Hypothetical Example)
10. Implications and Future Directions
11. Frequently Asked Questions (FAQs)
12. Conclusion

Executive Summary

PPP loan data due diligence should begin with a limited claim: a dated federal lending record may supply a historical borrower identity, industry, or location lead. It is not a current company profile. The **Paycheck Protection Program (PPP)** ended on **May 31, 2021**. SBA limits forgiveness applications to **five years from the date it issued the SBA loan number**. (www.sba.gov) The Small Business Administration (SBA) describes its current Freedom of Information Act (FOIA) dataset as covering **all disbursed PPP loans**; its landing page was last modified **October 21, 2024**. (data.sba.gov) Neither date establishes when a particular company was last observed operating.

The central control is a **field-to-claim protocol**: preserve the original row, attach the dictionary used, resolve the borrower to a legal entity, and state the historical scope of every accepted answer. The current dictionary defines `DateApproved` as **“Loan Funded Date”**, `JobsReported` as **“Number of Employees”**, and `ForgivenessDate` as **“Forgiveness Paid Date.”** These definitions require careful treatment rather than interpretations based on column names alone. (data.sba.gov) The Government Accountability Office (GAO) says applicants self-certified their information and that PPP data reliability depends on the accuracy of applicant inputs supplied to lenders and SBA. (www.gao.gov) Its assessment of sufficient reliability concerned describing program participation, not validating every borrower for an acquisition mandate. (www.gao.gov)

Version differences also matter. **ProPublica** displays an update label of **October 27, 2024**, while its data footer describes lender-approved records as of **June 1, 2021**; its FAQ warns that some applications do not result in loans. (projects.propublica.org) These statements should be preserved separately from SBA's disbursed-loan scope. For identity reconciliation, use jurisdiction-qualified registry identifiers and documented name links. California defines its entity number as a registration identifier, while Florida warns that fictitious-name registration does not prevent someone else using the same name. (www.sos.ca.gov) (dos.fl.gov)

The resulting screening decision should be **accepted historical match**, **escalated conflict**, or **unresolved identity**. Acceptance authorizes only the historical claim reviewed. Current headcount, ownership, operations, profitability, distress, and sale interest require separate, current evidence. The historical self-reporting basis is

the reason for that restriction. (www.dallasfed.org) Aggregate employment statistics cannot fill that gap: the Bureau of Labor Statistics (BLS) describes broad employment coverage but does not release an individual company's employee count through its Quarterly Census of Employment and Wages (QCEW). (www.bls.gov) This report supplies a reusable matrix and worksheet, with a fictional reconciliation example, to keep those limits visible throughout origination research.

Introduction and Background

PPP records still appear when an analyst researches a privately held U.S. business. The useful question is not simply whether a borrower name can be found. It is **which claim the located record can support**, for which entity, and at which date. The program was established by the **Coronavirus Aid, Relief, and Economic Security (CARES) Act**; its payroll-support purpose and historical end date belong in the evidence record. (www.fdic.gov) (www.sba.gov)

For private equity (PE) origination, that record may open a research path. An unfamiliar borrower name can suggest a registry search; an address can identify a jurisdiction; an industry code can suggest adjacent businesses to examine. Those are discovery uses. Promoting the row into a present-day investment screen requires another evidentiary step, particularly because program information includes borrower self-reports. A Federal Reserve Bank of Boston working paper describes borrower-provided information in its PPP extract, and the Federal Reserve Bank of Dallas explains that self-reports can create incomplete or inconsistent information. (www.bostonfed.org) (www.dallasfed.org)

A practical starting question is: **“Can this historical borrower be linked to the company under review?”** Answer that before asking whether the company meets a current mandate. New York's registry describes exact entity names and Department of State (DOS) identifiers; California separately defines entity identifiers and address fields. (dos.ny.gov) (www.sos.ca.gov) The combination is more useful than a name match alone, but it still needs jurisdiction-specific interpretation.

The scope here is deliberately narrower than a general company-data guide. It covers program provenance, row semantics, entity reconciliation, historical dates, and permissible acquisition-research claims. **Docket** describes its research model as retaining a source, supporting excerpt, and collection date for researched answers. Applied to PPP, that approach requires retention of the row and dictionary, together with a separate record of the analyst's interpretation. It does not turn a program input into verified current headcount. GAO's account of self-certification supplies the reason for that distinction. (www.gao.gov)

Key Changes in How PPP Records Should Be Used

From borrower lookup to a dated evidence claim

The recommended process changes the unit of analysis from a search result to a **specific claim**. A lookup interface can help locate a candidate; it cannot decide whether that candidate is the legal entity behind a present-day brand. Florida's fictitious-name guidance demonstrates why: a trading name can be registered without exclusive rights to that name. (dos.fl.gov) New York separately requires assumed-name filings for specified entity types conducting activities under a name other than their true legal name. (dos.ny.gov)

An acceptable output therefore reads: “The identified historical loan row is linked to this registered entity through these retained documents.” It should not read: “This company currently has the employee count shown in PPP.” GAO explains that the jobs field came from applications asking for employee numbers, and Dallas Fed describes the self-reporting chain through lenders. (files.gao.gov) (www.dallasfed.org)

From column labels to version-specific definitions

Historical terminology can change. GAO reported that SBA changed the field previously labeled **jobs retained** to **jobs reported** in its program data. (files.gao.gov) That matters because “retained” can imply a measured outcome, whereas an application input supplies a different kind of evidence. The report's recommendation is to retain the raw field label alongside the dictionary definition and the analyst's permitted interpretation.

Record-state terminology also needs its own context. SBA's current release description says disbursed loans; ProPublica's FAQ describes approved applications that may not become actual loans. (data.sba.gov) (projects.propublica.org) An analyst should not combine these descriptions into a universal statement that every searchable historical application was funded. Where a screen depends on receipt of funds, reconcile the official version and the record-specific fields before accepting the answer.

From silent merging to reviewable relationships

A second record for the same borrower need not be an accidental duplicate. Treasury's reopening announcement explains that certain existing borrowers became eligible for a **Second Draw** loan. (home.treasury.gov) Preserve each loan record and link it to an adjudicated entity rather than collapsing the loan history into one row.

The recommended relationship structure separates:

- **Loan identity:** the source's loan identifier and unchanged raw row.
- **Legal identity:** jurisdiction, registry identifier, and supporting filings.
- **Trading identity:** borrower name, assumed name, brand, and website links.
- **Research claim:** the dated conclusion accepted by a reviewer.

This distinction follows the World Wide Web Consortium (W3C) provenance model's separation of entities, activities, and responsible participants. (www.w3.org) It also keeps a transformation, such as name normalization, distinct from evidence that the resulting records concern the same business. W3C defines derivation as a transformation or construction based on a pre-existing entity; retaining that relationship makes subsequent corrections easier to trace. (www.w3.org)

Data Provenance, Versions, and Record States

Preserve the release before interpreting the borrower

Start from the **official SBA landing page** and its linked dictionary. The page describes all disbursed PPP loans and displays its modification date. (data.sba.gov) During this review, the linked resources included `PPP Data Dictionary.xlsx`, `public_150k_plus_240930.csv`, and numbered `public_up_to_150k` CSV

resources carrying the same `240930` suffix. Treat those names as version labels to preserve, not as proof that every row describes company conditions on a single date. (data.sba.gov)

The recommended retention manifest includes:

- **Origin:** publisher, landing URL, and direct resource URL.
- **Version:** literal filename, file checksum, and dictionary copy.
- **Collection:** download timestamp and analyst or process identifier.
- **Record:** loan identifier, source row location, and original values.
- **Transformation:** parsing, normalization, and matching rules applied.
- **Review:** accepted claim, conflicts, disposition, and reviewer date.

These are proposed research controls, not additional SBA columns. W3C's provenance definition supports recording the relationship between data and the activity that produced or transformed it. (www.w3.org) A page access date proves when research occurred; it does not replace the loan's event dates.

Keep official scope and mirror scope separate

Table 1 compares evidence routes. It includes a research-service row because sourced target research is a separate delivery layer, rather than a replacement federal dataset.

Evidence route	Documented role or model	Permissible use in this protocol	Required boundary
SBA FOIA release	Official loan-level release described as all disbursed PPP loans. (data.sba.gov)	Preserve the authoritative source row and dictionary for the historical program claim.	Do not call the landing-page modification date a company observation date.
ProPublica lookup	Searchable implementation with a 2024 update label and a footer referring to a 2021 lender-approved snapshot. (projects.propublica.org)	Find a candidate and retain the exact mirror version description.	Its FAQ warns that some applications do not produce loans. (projects.propublica.org)
Official registry	Jurisdiction-specific legal identity and filed information; New York identifies exact names and DOS IDs. (dos.ny.gov)	Reconcile the borrower to a legal entity and inspect name history.	Registry identity does not establish current employees or commercial activity. (dos.ny.gov)
Docket sourced target research	Service and platform assigning identity checks, collection, and evidence review to Triage, Scout, and Audit.	Organize retained sources and review decisions against the mandate.	A research-service conclusion still requires evidence appropriate to the field.

The table distinguishes **data origin**, **search interface**, **legal record**, and **research delivery**. A mirror and the official release are not independent confirmations of a borrower fact when the mirror derives from SBA. ProPublica identifies the underlying lender-approved program records in its data footer. (projects.propublica.org) Preserve any discrepancy rather than silently preferring whichever wording looks more conclusive. ProPublica's explicit application warning makes that distinction operationally important. (projects.propublica.org)

Separate applications, amounts, and administrative events

The current dictionary defines `DateApproved` as **Loan Funded Date**, `InitialApprovalAmount` as the approval amount at origination, and `CurrentApprovalAmount` as the current approval amount. It separately lists `UndisbursedAmount`, `ForgivenessAmount`, and a forgiveness paid date. Its terse definitions do not supply a complete cash-flow reconciliation formula or comprehensive cancellation codebook. (data.sba.gov)

Accordingly, do not automatically calculate “cash received” as approval amount minus undisbursed amount. Preserve those fields separately and seek record-specific documentation if cash receipt is material to a diligence question. Forgiveness application and payment also differ: SBA describes deferral until it remits a forgiveness amount to the lender, while FDIC's historical notice describes forgiveness as conditional on criteria being met. (www.sba.gov) (www.fdic.gov)

Cancellation needs the same version discipline. GAO's historical analysis used a **fully canceled** status in its dataset. (files.gao.gov) That does not establish that the current disbursed-loan release contains a complete population of canceled applications. Do not invent a `CancellationDate` column or infer cancellation from a missing forgiveness value.

Field-to-Claim Matrix for Acquisition Research

A field is useful only when its definition, source type, and reference date fit the question. Borrower names, addresses, industry inputs, and reported employment appear in borrower-provided PPP information described by the Boston Fed's research extract. (www.bostonfed.org) The official dictionary supplies field definitions; Census supplies the statistical meaning of the **North American Industry Classification System (NAICS)**. (www.census.gov)

Table 2 is a reusable field-to-claim matrix. “Historical program” means the applicable loan or application context, not the page-access date. The permissible claims are this report's recommended interpretations, conditional on successful identity reconciliation.

SBA column(s)	Definition and source type	Reference date	Permissible historical claim	Prohibited current inference	Corroboration and review note
`BorrowerName`, `BusinessType`	Borrower name and business-type labels in a federal program record; underlying borrower reporting requires validation. (www.bostonfed.org)	Historical program input.	The row names this borrower and records this business type.	Exact present legal name, beneficial ownership, or current structure.	Registry ID plus legal-name or assumed-name filing; preserve the raw string. (dos.ny.gov)
`BorrowerAddress`, `BorrowerCity`, `BorrowerState`, `BorrowerZip`; `ProjectCity`, `ProjectState`	Distinct borrower/project location fields; labels alone do not establish headquarters, operating site, or comprehensive footprint. (data.sba.gov)	Historical program record.	The specified field contains this location.	Current headquarters, current facilities, or full geographic coverage.	Compare address roles, dates, and entity records; California defines its registry entity address as executive office. (www.sos.ca.gov)

SBA column(s)	Definition and source type	Reference date	Permissible historical claim	Prohibited current inference	Corroboration and review note
`NAICSCode`	Industry-code input; NAICS classifies establishments for statistical purposes. (www.census.gov)	Preserve the input's program context and code vintage.	The row records the displayed code.	Verified current service mix, revenue concentration, or mandate fit.	Read current company services and applicable licensing evidence; do not filter solely on the code.
`JobsReported`	Dictionary definition: Number of Employees. GAO traces the input to applications asking for employee counts. (data.sba.gov) (files.gao.gov)	Application /program context, not current research date.	The historical row reported this employee number.	Current headcount, independently verified jobs saved, or staffing trend.	Obtain a dated employee measure with entity, group, and worker scope; keep missing and zero distinct.
`LoanNumber`, `ProcessingMethod`	Loan identifier; PPP denotes first draw and PPS second draw. (data.sba.gov)	Individual loan record.	This is a distinct recorded loan and draw designation.	A unique company identifier or automatic duplicate determination.	Join to a jurisdiction-qualified entity key; preserve both draws. (home.treasury.gov)
`DateApproved`, `InitialApprovalAmount`, `CurrentApprovalAmount`, `UndisbursedAmount`	Funded date and separate approval/undisbursed amount fields as defined in the dictionary. (data.sba.gov)	Recorded event and file version.	The identified version records these specified values.	Application submission date, revenue, enterprise value, or unverified net cash receipt.	Retain units, nulls, versions, and any lender documentation; do not invent a reconciliation formula.
`LoanStatus`, `LoanStatusDate`	Dictionary substitutes Exemption 4 for specified disbursed loans and leaves the status date blank in that condition. (data.sba.gov)	Administrative record state in that version.	The file displays this status and date or blank.	Distress, viability, current debt, or wrongdoing.	Review dictionary conditions before interpreting missingness; obtain current financial evidence separately.
`ForgivenessAmount`, `ForgivenessDate`	Separate amount and paid-date fields; forgiveness date means Forgiveness Paid Date. (data.sba.gov)	Forgiveness payment event, where reported.	The version records a forgiveness amount and paid date.	Current profitability, solvency, cleared business risk, or sale interest.	Separate application, administrative decision, and remittance; missing fields remain unresolved. (www.sba.gov)

The matrix deliberately limits claims to **what the record says**. A legal-entity match does not improve the employee input's measurement quality or make its date current. The field originated in an application employee report. (files.gao.gov) Dallas Fed's discussion of self-reported information and GAO's input-dependence assessment support that distinction. (www.dallasfed.org) (www.gao.gov) Treat field acceptance and entity acceptance as separate review decisions. An exact registry name and identifier answer a narrower identity question. (dos.ny.gov)

For reuse, export this table with the column headings intact and add **source file/version**, **dictionary URL**, **applicable event date**, **corroborating document URL**, and **review decision** to each row. The saved matrix should travel with the batch rather than remain a general policy detached from its actual files. The derivation record should explain how every displayed value was produced from retained source material. (www.w3.org)

Implementation Considerations and Process Changes

Resolve the borrower before applying the mandate

Use the raw borrower name to generate candidates, not to authorize a match. California assigns a business entity number; New York publishes exact names and DOS identifiers; Florida's fictitious-name guide supports searches using an associated entity's charter or document number. (www.sos.ca.gov) (dos.ny.gov) (dos.fl.gov) Pair each identifier with its jurisdiction and source, because these systems describe different registries. California also defines jurisdiction as the place under whose laws the entity was organized. (www.sos.ca.gov)

The proposed matching sequence is:

- **Preserve:** retain the raw name, address, loan ID, and file version.
- **Normalize:** create additional search keys without replacing original values.
- **Generate candidates:** search official registries and relevant assumed-name records.
- **Compare:** assess legal name, address role, jurisdiction, dates, and name history.
- **Link:** use a filing or explicit company statement connecting the brand and entity.
- **Review conflicts:** record alternatives and reasons for accepting or rejecting them.
- **Decide:** accept the historical link, escalate a conflict, or leave identity unresolved.
- **Research current fields:** begin mandate screening from fresh, field-appropriate evidence.

These steps are proposed controls, not a numerical confidence model. A distinctive name plus location may be suggestive; it is not a universal acceptance threshold. A legal-name or assumed-name record supplies a different level of support. (dos.ny.gov) Florida explicitly allows the possibility of another registration using the same fictitious name. (dos.fl.gov) Keep candidate collisions visible rather than inventing a matching probability.

Interpret the registry within its jurisdiction

Registry status is itself a defined field. New York says a domestic corporation or limited liability company (LLC) remains active until it dissolves. (dos.ny.gov) Delaware's free search returns active and inactive entities, so appearance in results does not itself supply a current status determination. (corp.delaware.gov) California says a foreign entity's status applies to its California registration. (www.sos.ca.gov) These distinctions prevent a registry lookup from becoming an unsupported finding about current operations.

Address matching needs similar care. California defines its entity address as executive office; that role differs from an address whose role is unspecified in the historical PPP row. (www.sos.ca.gov) A shared address should prompt review of the match, but it does not independently prove common ownership. Even registry sources may lack the member details needed for an ownership answer. (dos.ny.gov) When address roles or dates differ, preserve both observations and state whether the difference was resolved.

Additional identifiers can help where present. The Securities and Exchange Commission (SEC) uses a **Central Index Key (CIK)** for filers, while the Global Legal Entity Identifier Foundation (GLEIF) describes official names and registered addresses in **Legal Entity Identifier (LEI)** data. (www.sec.gov) (www.gleif.org) Do not treat absence from EDGAR, the SEC's filing system, as nonexistence; the SEC notes that not all companies offering stock must file electronically. (www.sec.gov)

Use a row reconciliation worksheet

Copy the following header into a CSV worksheet; create one observation per source loan row and version. Empty cells should remain empty until researched, with the reason recorded in `review_note`.

```
source_landing_url,source_resource_url,source_file,source_checksum,dictionary_version,retrieved_at,source_row,loan_number,borrower_raw,borrower_dba,borrower_address_raw,project_city,project_state,naics_raw,naics_vintage,date_approved_raw,processing_method,initial_approval_amount,current_approval_amount,undisbursed_amount,loan_status,loan_status_date,jobs_reported,forgiveness_amount,forgiveness_date,legal_entity_name,registry_jurisdiction,registry_id,matched_fields,conflicting_fields,corroborating_urls,applicable_reference_date,disposition,reviewed_at,review_note
```

This is a **proposed worksheet schema**, not the SBA's CSV header. It separates the loan row, the legal-entity crosswalk, and the review decision. W3C's provenance model provides the underlying logic: document the data and the activity responsible for producing its derivative research record. (www.w3.org) Keep original and corrected observations available rather than rewriting the historical evidence. Identify the activity responsible for the derived record. (www.w3.org)

Time Limits, Quality Controls, and Ethical Review

Historical employment is not a current staffing measure

PPP employee information requires both measurement and time controls. GAO explains that applicants reported employee numbers, while its historical review found that employment-related fields could differ. (files.gao.gov) A newer file may improve an administrative value without converting an old employee input into a contemporary count. The origin remains borrower reporting in the historical extract. (www.bostonfed.org) GAO also explains that lenders could update or supplement information during disbursement reporting and forgiveness. (files.gao.gov)

This report recommends explicit restrictions:

- **Headcount:** require a fresh employee measure and a defined entity or group scope.
- **Location:** require current operating evidence, distinguishing offices from agents and mail addresses.
- **Ownership:** require current, entity-specific ownership evidence, not loan or lender identity.
- **Financial health:** require financial documentation appropriate to the question.
- **Sale interest:** require direct current evidence of willingness to engage.

The registry cannot fill every gap. New York says it does not maintain the LLC member or manager details described in its FAQ. (dos.ny.gov) GLEIF reference data can support identity, but historical and current LEI records must be distinguished when used for a dated match. (www.gleif.org) An ownership question should

remain unanswered when the available source only establishes a name and address. GLEIF distinguishes official-name and registered-address information as identity data. (www.gleif.org)

Review missingness, revisions, and collisions without allegations

GAO's reliability procedures examined frequencies, blanks, zero values, and extremes. (www.gao.gov) A target-research quality pass can adopt that diagnostic approach while avoiding a conclusion that unusual values establish misconduct. Missing employment, repeated names, or inconsistent addresses are **review triggers**, not findings about an identifiable borrower's behavior. Incomplete self-reported information is a documented data limitation. (www.dallasfed.org)

Recommended checks include:

- **Missingness:** separate blank, zero, unavailable, and not-applicable states.
- **Parsing:** preserve ZIP codes and identifiers as strings where necessary.
- **Revision:** compare the same loan identifier across retained file versions.
- **Draw history:** distinguish legitimate multiple loan records from duplicate ingestion.
- **Collision:** group normalized names and locations without automatically merging entities.

Treasury's Second Draw explanation supplies a concrete reason to preserve multiple records for the same borrower. (home.treasury.gov) If an official revision changes a cited value, create a new observation and record whether the accepted claim also changes. If a mirror differs, prefer the appropriate official version for the program claim while retaining the mirror discrepancy. (data.sba.gov)

Minimize sensitive exposure

SBA's program included sole proprietors, independent contractors, and self-employed people, so a historical borrower record may relate closely to an individual. (www.sba.gov) Public availability does not make every address or identifying detail necessary for an acquisition-screening report. Use the minimum detail needed to establish the match, restrict raw evidence access, and avoid publishing real borrower examples in an educational workflow. This recommendation follows the FTC's need-based retention principle. (www.ftc.gov)

The Federal Trade Commission (FTC) recommends retaining sensitive personally identifying information only while necessary and limiting employee access to resources needed for their work. (www.ftc.gov) The National Institute of Standards and Technology (NIST) describes its Privacy Framework as a tool for identifying and managing privacy risk. (www.nist.gov) For this protocol, record the business purpose of retaining each sensitive field and use a retention policy appropriate to that purpose. Obtain professional advice where collection, reuse, or disclosure raises jurisdiction-specific questions; this article does not claim professional review.

Data Analysis and Evidence

What the available quantitative evidence does establish

A historical quality statistic is not a current accuracy score. GAO's **September 2020** program analysis found missing or zero jobs-related entries for **18 percent of loans** in the historical dataset it examined. (files.gao.gov) That is evidence about that extract and measure. It is not an estimated error rate for the current SBA release, a correction factor for an individual borrower, or a reason to scale every reported employee number downward. GAO's purpose-specific reliability assessment reinforces the need to define the analytical task before accepting the data. (www.gao.gov)

Time can be quantified without inventing a decay model. Use **record age = research date minus the applicable event date**, in calendar days. For this report's **October 13, 2026** research date, a fictional event dated **May 1, 2020** is **2,356 days** old. This is an illustrative date-subtraction calculation, not an observed company statistic. Separately, the interval from the program end on **May 31, 2021** is **1,961 days**. (www.sba.gov) The interval from the landing-page modification on **October 21, 2024** is **722 days**. (data.sba.gov) These intervals measure different things; none establishes the freshness of company operations.

The proposed calculation controls are:

- **Choose the field date:** identify the event appropriate to the claim.
- **State the research date:** preserve when the calculation was made.
- **Keep unknown dates unknown:** do not substitute a download date for an absent event date.
- **Label derived values:** distinguish arithmetic from source-reported figures.
- **Route current questions:** obtain new evidence instead of applying an unsupported decay factor.

Use statistical denominators at their actual unit

County Business Patterns (CBP) provides annual statistics for businesses with paid employees, including employment during the week of **March 12** and payroll measures. (www.census.gov) **Statistics of U.S. Businesses (SUSB)** distinguishes a single-location establishment from a firm defined by establishments in the same geography and industry under common ownership or control. (www.census.gov) It excludes nonemployer businesses. (www.census.gov) Those scopes differ from PPP's eligible sole proprietors and self-employed borrowers. (www.sba.gov)

QCEW covers **more than 95 percent of U.S. jobs**, but its monthly measure refers to the pay period including the month's **12th day**. (www.bls.gov) Its public geographic detail reaches county-industry aggregates; it does not release a particular company's employee count. (www.bls.gov) Broad coverage therefore supports sector context, not validation of a named target's payroll.

The Internal Revenue Service (IRS) publishes business statistics on income, deductions, taxes, and credits; its Integrated Business Dataset is assembled at the table level. (www.irs.gov) Such aggregates can inform a sector comparison, but they do not establish a target's revenue or profitability. Preserve industry vintage as well: Census's NAICS page lists proposed **2027** updates, which should not be treated as an already implemented replacement for the classification attached to an older record. (www.census.gov)

Report reconciliation coverage as **accepted, escalated, and unresolved rows**, with a defined denominator. Keep loan-row coverage, legal-entity coverage, and current-mandate coverage separate. A reviewed historical loan count is not a count of operating acquisition targets.

Worked Reconciliation (Hypothetical Example)

The following names, identifiers, addresses, employee values, and amounts are **entirely fictional**. The purpose is to illustrate adjudication, not describe real borrowers or estimate program outcomes. The process preserves separate draw records because an existing borrower could be eligible for a second loan. (home.treasury.gov)

Suppose a screening batch contains a first-draw row for **“Harbor Tooling”**, a second-draw row for **“Harbor Tooling LLC”**, and another row using the short name in the same city. The analyst generates a normalized-name-plus-location group. That group is a review queue, not an entity merge. Florida's warning about nonexclusive fictitious names explains why a shared trade name is insufficient by itself. (dos.fl.gov)

Table 3 shows a fictional duplicate/collision group. Row labels are worksheet identifiers, not actual loan numbers.

Raw row and preserved fields	Corroborating evidence in the fictional scenario	Conflict or distinction	Disposition and authorized claim
A: Harbor Tooling; first draw; 24 reported jobs; May 1, 2020 event	Assumed-name filing explicitly links Harbor Tooling to Harbor Precision LLC, registry ID EXAMPLE-A.	Borrower short name differs from registry legal name.	**Accept historical match** after reviewing the explicit name link; retain 24 only as the row's historical report.
B: Harbor Tooling LLC; second draw; same borrower address	No filing yet links this exact borrower wording to EXAMPLE-A; candidate registry record is Harbor Precision LLC.	Legal suffix and name disagree; matching address alone does not resolve the conflict.	**Escalate**; preserve the likely relationship without accepting the entity join.
C: Harbor Tooling; first draw; same city, different street	Candidate is Harbor Tooling Services Inc., registry ID EXAMPLE-C; no name bridge is available.	Name collision, different address, and unverified brand relationship.	**Unresolved**; leave it separate and exclude it from accepted entity counts.

The accepted outcome for A is deliberately narrow. The real program's jobs field came from applicants reporting employee counts. (files.gao.gov) It establishes the fictional historical entity link, not current employees, ownership, or readiness to sell. B requires additional evidence even though an analyst may suspect it is a second loan to the same business. C remains visible as a collision. Registry identifiers help adjudicate these paths, as New York's exact-name and DOS-ID guidance illustrates. (dos.ny.gov)

For this **three-row fictional batch**, the outcome is **one accepted**, **one escalated**, and **one unresolved** row. The derived row-acceptance share is **1 divided by 3, or 33.3 percent**, rounded to one decimal place. This calculation measures the example's reconciliation outcome only. It is not a measured match rate, an operating-company count, or a service-performance claim.

The reviewer should also record the specific unresolved request: “Obtain a filing or company statement connecting the borrower wording in B to the legal entity.” That is more actionable than an undifferentiated low-confidence label. Keep C's alternative candidate and conflicting address in the worksheet so subsequent research can resolve it without repeating the original search.

Implications and Future Directions

Use the matrix and worksheet to review historical inputs before researching current mandate fields. The self-certification and borrower-reporting findings support this sequence. (www.gao.gov) (www.bostonfed.org)

For a current identity question, start with the relevant registry and any legal-name history. For an operations question, seek current company materials and applicable official licensing or registration records. An active New York entity alone is insufficient because its status can persist until dissolution. (dos.ny.gov) For employee scope, distinguish the operating entity from a group, full-time staff from other workers, and a reporting period from today's date. Aggregate QCEW data cannot supply an individual-company headcount. (www.bls.gov) For financial and ownership questions, retain the specific evidence needed or mark the answer unresolved.

A research system should support these independent review states:

- **Historical row preserved:** the original program evidence is available.
- **Entity link reviewed:** the borrower-to-entity relationship is accepted or pending.
- **Field interpretation reviewed:** the historical meaning is understood.
- **Current mandate evidence collected:** contemporary questions have separate support.
- **Conflicts retained:** unresolved alternatives remain visible at handoff.

These are recommended workflow states, not a certification or promised acquisition outcome. They make it possible to accept an identity link while declining a headcount inference. They also protect against confusing an administrative status date with a current observation. Delaware explicitly describes its status-as-of date as the date the entity entered that status, rather than the date an analyst searched. (corp.delaware.gov)

Future releases should be assessed for **changed rows, changed definitions, and changed scope**, separately. A dictionary revision may require reinterpretation without altering the underlying borrower identity; a corrected borrower string may require a new match review. W3C's derivation model provides a way to retain the connection between the old source, transformed records, and revised conclusions. (www.w3.org) Better evidence management does not remove uncertainty; it makes the unresolved work specific and reviewable.

Frequently Asked Questions (FAQs)

Can PPP loan data be used for acquisition due diligence?

Yes, as historical discovery or corroboration evidence after entity reconciliation. SBA describes the official dataset as disbursed-loan data, while GAO's reliability assessment was tied to program-participation analysis. (data.sba.gov) (www.gao.gov) A record alone cannot answer current investment criteria. Separate origination screening from financial, legal, and commercial diligence, and obtain appropriate current documentation when a transaction question depends on it.

Does JobsReported show a company's current employee count?

No. The input belongs to the historical program context. GAO explains that applicants were asked to report employee numbers and documents the change from jobs retained to jobs reported. (files.gao.gov) Describe the recorded value with its period and scope; do not use it as today's workforce size or as proof of independently measured employment preservation.

Does a missing forgiveness amount mean the borrower still owes the loan?

The missing field does not establish that conclusion. Forgiveness application and remittance are distinct events in SBA's explanation. (www.sba.gov) Record the missingness and source version, then obtain appropriate current loan documentation if the obligation matters. Do not use a blank to infer distress or borrower conduct.

Are a PPP lookup and the SBA release interchangeable?

No. ProPublica's FAQ warns that some approved applications do not result in loans, whereas the current SBA landing page describes all disbursed loans. (projects.propublica.org) (data.sba.gov) Preserve each source's own description and dates. A mirror is useful for discovery but does not supply independent corroboration of the underlying SBA record.

Does a borrower record establish the business is still active or independent?

No. PPP is historical, and registry activity has a jurisdiction-specific meaning. New York says a domestic corporation or LLC remains active until dissolution. (dos.ny.gov) Verify current operations and ownership separately. Neither an old loan row nor an active registry status demonstrates a sale process or interest in acquisition outreach.

Conclusion

Retain a **dated historical claim attached to a reconciled entity**, with raw values, a file version, a dictionary, and a review decision. The historical borrower-reporting basis remains relevant after matching. (www.bostonfed.org)

The protocol separates the original loan record from the legal entity, the trading name, and the mandate answer. It also separates approval amounts, funding terminology, administrative status, and forgiveness payment. Those distinctions prevent a search interface, a familiar name, or a missing field from carrying more evidentiary weight than it can support.

Teams should preserve conflicts and publish coverage at the right unit. Accepted loan-row matches, accepted entity relationships, and completed current-mandate answers measure different work. Escalated and unresolved records belong in the review queue with explicit next steps. They should not disappear through automatic merging or become a generalized assessment of a borrower's character or financial condition. Documented incomplete inputs support careful review rather than expanded conclusions. (www.dallasfed.org)

For handoff, record each accepted claim, unresolved conflict, and specific evidence request. Keep the source row, entity link, and review decision together so the next reviewer can inspect and correct the file.

Source links

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

1. https://www.sba.gov/loans/covid-era-programs
2. https://data.sba.gov/dataset/ppp-foia
3. https://data.sba.gov/sites/default/files/distribution/SBA-OCA-2022-07-001/ppp-data-dictionary.xlsx
4. https://www.gao.gov/assets/720/716639.pdf
5. https://projects.propublica.org/coronavirus/bailouts
6. https://www.sos.ca.gov/business-programs/business-entities/cbs-field-status-definitions
7. https://dos.fl.gov/sunbiz/start-business/efile/fl-fictitious-name-registration
8. https://www.dallasfed.org/cd/communities/2020/0904
9. https://www.bls.gov/cew/questions-and-answers.htm
10. https://www.fdic.gov/news/press-releases/2020/pr20050.html
11. https://www.bostonfed.org/-/media/Documents/Workingpapers/PDF/2021/wp2115.pdf
12. https://dos.ny.gov/faqs-corporations-business-entities
13. https://files.gao.gov/reports/GAO-20-701/index.html
14. https://projects.propublica.org/coronavirus/bailouts/faq
15. https://home.treasury.gov/news/press-releases/sm1230
16. https://www.w3.org/TR/prov-dm
17. https://www.fdic.gov/news/financial-institution-letters/2020/fil20050.html
18. https://www.census.gov/naics
19. https://dos.fl.gov/sunbiz/search/guides/fictitious-name
20. https://corp.delaware.gov/more-information
21. https://www.sec.gov/search-filings/cik-lookup
22. https://www.gleif.org/en/lei-data/access-and-use-lei-data/level-2-data-who-owns-whom
23. https://www.ftc.gov/business-guidance/resources/protecting-personal-information-guide-business
24. https://www.nist.gov/privacy-framework
25. https://www.census.gov/data/developers/data-sets/cbp-zbp/cbp-api.html
26. https://www.census.gov/programs-surveys/susb/about/glossary.html
27. https://www.bls.gov/cew/overview.htm
28. https://www.irs.gov/statistics/soi-tax-stats-business-tax-statistics
29. https://corp.delaware.gov/onlinestatus

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.