Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

Technographic Data for Private Equity: Validation Protocol

> 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 · 2026-10-16 · technographic data for private equity · technographic data · deal sourcing · website technology lookup · tech stack lookup · builtwith alternatives · data validation · acquisition target screening · wappalyzer

Original article: https://docket.capital/articles/technographic-data-private-equity-validation



In this article

1. Executive Summary
2. Introduction and Background
3. Key Changes
4. Implementation Considerations and Process Changes
5. Category-Specific Adjudication
6. Buying Data and Setting Decision Thresholds
7. Data Analysis and Evidence
8. Implications and Future Directions
9. Frequently Asked Questions (FAQs)
10. Conclusion

Executive Summary

Technographic data for private equity is useful for finding acquisition candidates whose public websites expose a relevant technology signal. Its value as an exclusion rule depends on a separate question: whether the observation supports the precise mandate claim. **BuiltWith** describes analysis of publicly accessible code and infrastructure and acknowledges residual or unused code and indexing delays. **Shopify** documents headless architectures that separate storefront presentation from the commerce backend. Together, these sources show why a visible signature and enterprise-wide adoption require different evidence. (builtwith.com) (shopify.dev)

The recommended validation protocol starts with a dated, domain-specific claim, then compares providers on the same labeled cases. Include confirmed users, confirmed nonusers, former users, parked domains, redirects, subsidiaries, and inaccessible sites. Keep difficult cases visible rather than deleting them from the denominator. Stratification supports examining distinct subgroups, while appropriate sample weights matter when estimating performance across a broader population. A deliberately difficult test set is useful for finding failure modes; its unweighted average is not automatically a market-wide accuracy estimate. (www150.statcan.gc.ca) (www.census.gov)

Measure **precision, sample recall, unknown rate, stale-positive rate, provider disagreement, and detection lag** separately. In the report's **fictional 200-case example**, 90 verified current users among 100 positive detections produce a conservative verified-positive yield of **90%**; 20 unresolved cases produce a **10% unknown rate**. These are author calculations, not vendor results. Precision and recall answer different questions, and a confidence score becomes a probability only after calibration against observed outcomes. An illustrative **95% Wilson interval of approximately 82.6% to 94.5%** around 90 successes in 100 independent cases also shows why a point estimate alone is insufficient. (developers.google.com) (scikit-learn.org) (itl.nist.gov)

The buying decision should follow the error cost. Use detections to generate candidates, require corroboration before treating them as verified qualification, and escalate missing or ambiguous evidence before exclusion. Wappalyzer explicitly documents that an empty response can mean an unindexed website or no identified technologies. In the fictional economics example, annual license allocation plus analyst work totals **$27,000**, or

$300 per verified match across 90 matches. The practical purchasing criterion is therefore the cost of an answer that meets the mandate's evidence standard, alongside coverage and timeliness, rather than dataset size or a generic accuracy claim. (www.wappalyzer.com) (www.gov.uk)

Introduction and Background

A private equity origination team might want merchants using a particular commerce platform, service businesses exposing a customer portal, or software-enabled companies with a named customer-facing system. A website technology lookup can generate a useful candidate list. The investment question, however, concerns a company; the detector's observation concerns a particular accessible technical surface. BuiltWith's terms describe that narrower public-code and infrastructure basis. (builtwith.com)

This report examines **technographic data for deal sourcing** through a reproducible, mandate-specific validation protocol. It addresses website technology stack analysis, how to validate technographic data, and when technology signals for acquisition targets justify acceptance, rejection, or analyst review. The recommendations are an analyst-designed procedure grounded in the cited documentation, rather than an industry certification or a benchmark performed on paid products.

The scope distinction matters for private equity acquisition target screening. BuiltWith's indexing policy excludes attempts to log in or access restricted areas. HTTP Archive separately explains that technology measurement depends on detectable fingerprints and that a homepage need not represent an entire website. Public-site observation therefore cannot, on its own, establish internal customer relationship management (CRM) or enterprise resource planning (ERP) adoption. (builtwith.com) (almanac.httparchive.org)

A useful test separates **signal correctness**, **entity attribution**, and **mandate relevance**. A correctly recognized script can belong to the wrong subsidiary. A correctly mapped storefront can still fall outside the sector mandate. A current component can satisfy a narrow technology criterion without establishing spend, contractual status, implementation success, company quality, or acquisition fit. Those broader conclusions need evidence directed at their own questions.

The objective is a decision rule the origination lead can inspect and the data-operations team can repeat. This follows a broader data-quality principle: required quality depends on intended use, and completeness does not establish accuracy. Documentation was reviewed for this October 2026 report; no paid-platform trial or live provider-accuracy experiment was conducted. (www.gov.uk)

Key Changes

Replace an adoption label with an observable claim

Replace “uses technology X” with a statement specifying **technology, surface, domain, entity, and time**. “The designated storefront exposed a qualifying signature during this session” is an observation. “The operating company currently depends on that platform” is a business conclusion requiring corroboration.

The historical open Wappalyzer specification illustrates the distinction. It describes regular-expression patterns, confidence modifiers, and inferred dependencies, and warns that short or generic patterns can identify applications incorrectly. It is evidence about a published detection approach, not proof that the current

commercial engine uses identical rules. (github.com)

Distinguish **directly observed evidence**, the script, header, or resource actually seen, from **inferred technology**, a conclusion drawn from relationships between signatures. Store both when exposed; otherwise record that the inference basis is unavailable.

Separate categories before comparing accuracy

Table 1 summarizes the claim boundaries used in this protocol. The limitations are applications of documented architectures, not measured error rates for the listed vendors.

Technology category	What an observation can support	What needs separate corroboration
Client-side script	A matching resource or code signature appeared in the observed page/session.	Whether it executed, remains operationally necessary, or covers the company. Generic patterns require caution. (github.com)
Content management system (CMS) or frontend framework	Evidence consistent with the observed presentation stack.	Backend commerce or internal systems; headless architecture separates frontend and backend. (shopify.dev)
Hosting or content delivery network (CDN)	Evidence about the serving or proxy layer encountered.	The underlying origin provider; proxied records can return Cloudflare addresses. (developers.cloudflare.com)
DNS or email infrastructure	A particular Domain Name System (DNS) record was observed.	Enterprise-wide deployment or paid seats. Cloudflare distinguishes proxied web records from DNS-only mail exchange (MX) and text (TXT) records. (developers.cloudflare.com)
Payment widget	A payment interface component appears on the relevant page.	Successful transactions, processing volume, revenue, or the merchant's complete payment arrangement. (docs.stripe.com)
Tags and analytics	Tags visible under the recorded session conditions.	Tags gated by consent or processing performed in a server-side container. (developers.google.com)
Inferred backend	A documented inference consistent with the available signature.	Actual backend operation, especially with a separately implemented storefront. (shopify.dev)
Business metadata	A provider's separately supplied business field, with its own method and age.	Treating the field as a technology observation; BuiltWith documents a different confidence treatment for operational information. (kb.builtwith.com)

Use distinct adjudication rules for CMS, CDN, and payment signals. Combine scores only for the mandate's defined case mix.

Replace timeless labels with dated observations

First detected is not necessarily implementation date, and **last detected** is not necessarily termination date. BuiltWith defines those fields as detection timestamps and separately offers a live-only filter. The protocol should retain the source definitions alongside the values rather than silently converting observation dates into contractual adoption dates. (api.builtwith.com)

Likewise, a parked domain should not inherit an operating-company label without review. BuiltWith's confidence documentation explicitly discusses older technology remaining visible on parked domains. This supports a separate domain-state field and a former-user stratum in the test set. (kb.builtwith.com)

Implementation Considerations and Process Changes

Freeze the mandate and the unit of evaluation

Write an acceptance statement specifying current use, historical use, or signature exposure. Define the permitted domains and whether a regional storefront, checkout subdomain, or subsidiary supplies sufficient evidence.

Keep website and legal-entity records separate. A **Legal Entity Identifier (LEI)** represents one entity; GLEIF reference data describes legal names, addresses, and accounting-consolidation relationships. Retain evidence connecting the tested domain to the acquisition unit alongside any identifiers. (www.gleif.org)

Do not make public domain-registration details the only identity test. ICANN's policy permits redaction of the registrant organization value under specified conditions. Where a record is redacted, the protocol should mark that identity source unavailable and seek other company or registry evidence, rather than infer that the domain lacks an owner. (www.icann.org)

Build a labeled set before viewing provider answers

Use a challenge set and, where feasible, a probability-based audit sample from the sourcing universe. Statistics Canada's stratification guidance and the Census Bureau's sample-design standards support separating diagnostic testing from population estimation. (www150.statcan.gc.ca) (www.census.gov)

Assign a primary stratum and retain additional flags so overlapping characteristics do not cause double counting:

- **Verified positives:** Current qualifying use established independently of the tested provider.
- **Verified negatives:** Evidence resolving the narrow mandate claim as nonqualifying.
- **Former users:** Earlier qualifying use with documented evidence of a subsequent change.
- **Parked domains:** Domain-state cases that require separation from active operations.
- **Redirects:** Cases with both the submitted domain and final destination retained.
- **Subsidiaries:** Cases where the observed site and intended acquisition unit differ.
- **Multi-domain firms:** Cases requiring a predefined company-level aggregation rule.
- **Inaccessible sites:** Cases whose required evidence cannot be collected under permitted access.
- **Routine controls:** Ordinary cases sampled from the intended production workflow.

These are proposed design categories, not assertions that every case can be conclusively labeled. A verified negative requires an adjudication route appropriate to the claim; absence of a homepage signature is insufficient for hidden backend or internal-system questions. HTTP Archive documents precisely that fingerprint-based measurement boundary. (almanac.httparchive.org)

Collect comparable outputs and preserve state

Submit the same cases within a documented collection window. Freeze normalization, redirects, caching, and scan settings. Wappalyzer distinguishes cached/live results, recursive scanning, confidence filtering, and initiated crawls. Record configuration differences before attributing outcomes to providers. (www.wappalyzer.com)

Use a comma-separated values (CSV) scorecard with these field groups. The groups below form an inline worksheet schema:

- **Identity:** case_id, domain, final_url, legal_entity, jurisdiction, entity_id.
- **Mandate:** mandate_version, technology, claim_scope, category, stratum, secondary_flags.
- **Reference label:** expected_label, label_source, label_date, label_reviewer, label_status.
- **Provider output:** provider, product_mode, vendor_result, confidence, raw_output_reference.
- **Time:** first_seen, last_seen, crawl_date, requested_at, returned_at.
- **Evidence:** evidence_surface, evidence_url, observed_signature, capture_reference, session_conditions.
- **Adjudication:** current_user, former_user, nonuser, unresolved, reason, decision_owner.
- **Work:** analyst_minutes, review_minutes, retry_count, escalation_status, final_action.

Retain raw responses separately from interpreted labels. The World Wide Web Consortium's (W3C) provenance model distinguishes entities, activities, derivations, and responsible agents. Applying it here makes responses, collection runs, and analyst decisions separately traceable. (www.w3.org)

Preserve the redirect chain. MDN documents that Hypertext Transfer Protocol (HTTP) redirects direct browsers to a new location. A destination lookup does not establish that both domains represent the same acquisition unit. (developer.mozilla.org)

Category-Specific Adjudication

Match the corroborating source to the technical claim

Inspect the relevant page and record the observed component signature. For header claims, retain the headers and requested resource. MDN describes the Server header as information about software used by the origin server handling that request. Keep the finding bounded to that response. (developer.mozilla.org)

For hosting and CDN questions, distinguish the network endpoint from the origin. Cloudflare documents that proxy status determines whether web traffic passes through its network or goes directly to the origin, and that proxied records return Cloudflare addresses. The protocol should therefore avoid treating a proxy address as conclusive evidence of the underlying hosting provider. (developers.cloudflare.com)

For analytics, record browser state and consent behavior. Google's basic consent mode can prevent tags from loading before a visitor interacts with the banner; advanced mode loads tags when the site or app opens. Separately, a server-side Tag Manager container runs on a controlled server. A browser-only observation has different visibility under those architectures. (developers.google.com)

For payment signals, inspect the page where payment functionality is relevant rather than relying exclusively on the homepage. Stripe describes its Payment Element as a frontend component, and its integration documentation identifies it as a Stripe.js feature. Finding the interface supports a component-level observation; it does not establish transaction completion, payment volume, or company economics. (docs.stripe.com)

For commerce platforms, test the designated storefront and corroborate the backend relationship where the mandate requires it. Shopify's headless documentation permits a separately chosen frontend stack, framework, and hosting. A frontend framework detection cannot alone settle which commerce backend powers that storefront. (shopify.dev)

Separate collection failure from a negative label

The protocol should maintain distinct states for **observed positive**, **resolved negative**, **unresolved**, and **collection pending**. Wappalyzer's documented empty-result example can mean either no indexed record or no detected technologies, and an accompanying crawl flag can indicate that collection has started. Treating that initial output as a verified negative would conflate workflow state with evidence. (www.wappalyzer.com)

The same discipline applies to inaccessible pages, unavailable historical evidence, and domain ambiguity. Record the source checked, collection conditions, and unresolved reason. Do not silently remove the case from the test or reclassify it to improve the apparent score.

Use the following proposed adjudication sequence:

- **Check identity:** Confirm the site, brand, and entity intended by the mandate.
- **Check scope:** Verify that the observation concerns the required domain and function.
- **Check capture:** Confirm whether the necessary page or response was actually collected.
- **Check signature:** Inspect the category-appropriate evidence rather than the vendor label alone.
- **Check operation:** Seek evidence of relevant current use where a signature is insufficient.
- **Check history:** Preserve earlier use separately from the present-tense decision.
- **Check independence:** Identify whether corroborating outputs share the same underlying detector.
- **Resolve conflict:** Retain competing evidence and explain the selected interpretation.
- **Escalate uncertainty:** Assign an owner and next permitted research step.

Independence deserves particular attention when comparing tools. HTTP Archive identifies its detector as a fork of the last open-source Wappalyzer version. Agreement between outputs with related detection origins should not be counted automatically as independent confirmation. The published pattern repository remains useful for understanding possible inference mechanisms, with the commercial-engine distinction preserved. (almanac.httparchive.org) (github.com)

Buying Data and Setting Decision Thresholds

Compare workflows on the same evidence requirement

For **BuiltWith alternatives**, **website technology lookup**, and **tech stack lookup**, compare the fields, collection modes, and permitted uses required by the mandate. The table below defines a buying test, rather than ranking products through an unperformed trial.

Table 2 compares detection inputs and research workflows. Docket appears in its position as a mandate-based target research publisher, separately from the detection inputs being evaluated.

Option or workflow	Documented input or position	What the buyer should test
BuiltWith	Domain application programming interface (API) supplies current/historical technology observations, detection dates, and confidence information. (api.builtwith.com)	Whether current-use definitions, retained dates, and domain handling satisfy the mandate.
Wappalyzer	Lookup API supports cached/live analysis, optional deeper crawling, and confidence filtering. (www.wappalyzer.com)	Whether the chosen mode, age settings, and completion handling meet the evidence requirement.
Open pattern-based analysis	The historical ahrefs/Wappalyzer repository documents patterns and inferred relationships. (github.com)	Maintenance, collection coverage, and independent adjudication; do not assume equivalence to a current commercial engine.
Docket / mandate-based target research	Research layer evaluated against the acquisition mandate.	Apply the same evidence, unresolved-case, and review-cost scorecard; evaluate detection inputs separately.

Apply the same acceptance test to Docket and other research providers: findings should have appropriate evidence, correct entity attribution, and explicit unresolved questions.

Check usage rights before collecting or publishing a comparative dataset. BuiltWith's terms restrict benchmarking, evaluating, or developing a competing product or service. Wappalyzer describes commercial data as licensed and requires written permission for specified sharing, publishing, resale, and customer-facing uses. Have counsel or the contracting owner confirm the contemplated internal evaluation, retention, and distribution against the applicable agreement; this report does not resolve those permissions. (builtwith.com) (www.wappalyzer.com)

Precommit to the cost of each error

Choose thresholds before inspecting results. Google's guidance recognizes unequal error costs and class balance. Distinguish a cheap analyst-review false positive from an exclusion that removes a qualifying target. Set the relative costs for the mandate. (developers.google.com)

A defensible rule might accept strongly corroborated positives, review weaker detections, and keep unresolved cases in a research queue. Automated exclusion should require affirmative evidence that the defined criterion fails, or a separately approved missing-evidence policy. State that policy explicitly so a technology filter does not quietly redefine the target universe.

A confidence label is not automatically a probability. Scikit-learn compares predicted probabilities with observed positive frequencies. Test score bands on held-out cases; for categorical labels, report empirical outcomes without translating “high confidence” into an invented success percentage. (scikit-learn.org)

Before purchase, ask for a written answer to each requirement:

- **Scope:** Which technology categories and public surfaces are actually observed?
- **Recency:** What do collection, confirmation, and historical timestamps mean?

- **Completion:** How are pending crawls, inaccessible pages, and empty results represented?
- **Evidence:** Which raw fields and supporting observations can be retained?
- **Configuration:** Which settings change coverage, confidence filtering, or review workload?
- **Identity:** How are redirects, subsidiaries, and multiple domains associated?
- **Rights:** Which evaluation, retention, export, and publication uses are permitted?
- **Economics:** What is the cost per verified answer at the required review standard?

Data Analysis and Evidence

Calculate separate metrics with explicit denominators

Precision is true positives divided by all positive classifications; **recall** is true positives divided by all actual positives. For this protocol, sample recall means qualifying users found divided by all independently verified qualifying users in the labeled sample. It is not an estimate of every qualifying company in the market. (developers.google.com)

Table 3 presents a **Hypothetical Example** with fictional counts. There are 200 eligible cases: 100 verified current users, 80 verified noncurrent cases, and 20 unresolved cases. The provider returns 100 positive detections comprising 90 current users, eight noncurrent cases, and two unresolved cases. Among the remaining cases are 10 current users, 72 noncurrent cases, and 18 unresolved cases. Every result below is an author calculation, not an observed vendor benchmark.

Metric	Fictional calculation	Interpretation
Conservative verified-positive yield	90 / 100 = **90%**	Verified current users among all positive detections, keeping unresolved detections in the denominator.
Adjudicated precision	90 / 98 = **91.8%**	Precision among detections whose reference label is resolved; disclose the two omitted unknowns.
Sample recall	90 / 100 = **90%**	Found users among independently verified current users in the sample.
Unknown rate	20 / 200 = **10%**	Unresolved reference labels among eligible cases.
Stale-positive rate	5 / 100 = **5%**	Assume five of the eight noncurrent detections are verified former users.
Pairwise disagreement	36 / 180 = **20%**	Separate fictional comparison of two providers on 180 jointly comparable cases.

The distinction between conservative yield and adjudicated precision matters. Dropping unresolved positive detections increases the displayed percentage without resolving them. Neither percentage identifies all sources of error, and provider disagreement measures inconsistency rather than truth. Precision and recall retain their conventional meanings only when the label and denominator definitions are explicit. (developers.google.com)

For multiple providers, build a symmetric disagreement matrix on the same eligible pairs. Count positive-versus-negative disagreements separately from positive-versus-unknown differences, and report how many cases support each cell. Split the error log by category, evidence age, domain state, and stratum. Report challenge-set results separately from weighted production-sample estimates, following the sampling design. (www150.statcan.gc.ca) (www.census.gov)

For the fictional 90-of-100 conservative yield, the **95% Wilson interval is approximately 82.6% to 94.5%**, assuming independent sampled cases and a binomial interpretation of verified success. This author calculation uses the Wilson method described by NIST. It does not correct challenge-set selection, entity duplication, or uncertain labels. (itl.nist.gov)

Include review cost and refresh lag

In a second **Hypothetical Example**, annual license allocation is **$12,000**, analyst and adjudication time is **200 hours**, and loaded labor cost is **$75 per hour**. Total cost is **$27,000**; with **90 verified matches**, cost per verified match is **$300**. All inputs are fictional.

A fictional alternative with **$6,000** in license allocation and **400 hours** of work at the same hourly cost totals **$36,000**, or **$400 per verified match** for the same accepted output. The author-calculated difference is **$100 per match**. Include review, retries, and correction time consistently; define whether reusable platform costs are allocated or fully charged to the mandate.

Measure **time-to-detection** separately on independently documented changes. Record the last confirmed old state, first confirmed new state, and first provider detection. Where the change time is only bounded by observations, report a lag range instead of an exact delay. This operational definition follows the broader principle that timeliness depends on the intended use. (www.gov.uk)

Implications and Future Directions

A production technology filter should remain a versioned decision process. Site redesigns, domain changes, new storefront architectures, provider configuration changes, and revised mandate definitions are proposed retest triggers. Shopify's documented separation of frontend and backend is one reason the evidence supporting a familiar signature can change without a simple company-level “technology switched” event. (shopify.dev)

Refresh rules should follow field meaning rather than a universal age limit. BuiltWith exposes first/last detection dates. (api.builtwith.com) Wappalyzer's API FAQ describes an aim to reverify identified technologies at least monthly and refresh company details quarterly. (www.wappalyzer.com) These are different observations and update objectives; they do not establish a guarantee that every delivered field is currently correct.

For ongoing operation, keep a stable reference panel and add newly difficult cases to a separate challenge panel. Re-adjudicate cases when their evidence changes. Preserve the original provider response and attach later corrections to it, making it possible to compare the old and new rule without rewriting the historical result. W3C provenance concepts offer a useful representation for those derivations and responsible activities. (www.w3.org)

The opportunity is better qualification discipline, not more confident inference from the same limited signal. A detector may supply a useful website observation while the investment criterion remains unanswered. The Government Data Quality Framework's distinction between completeness and accuracy is directly relevant: filling every technology field does not establish that those fields answer the acquisition question correctly. (www.gov.uk)

Keep discovery and outreach as separate operational decisions. This protocol recommends a distinct approved outreach workflow after qualification; a technology match should not be treated as permission to contact a company. The research record should carry its unresolved questions into the next stage rather than letting a contact export imply completed diligence.

Frequently Asked Questions (FAQs)

Can website technology stack analysis reveal an internal CRM or ERP?

Only where relevant evidence is publicly exposed or separately corroborated. BuiltWith describes public-code and infrastructure analysis and excludes restricted-area access from its indexing policy. HTTP Archive explains the dependence of technology measurement on detectable fingerprints. A missing public signature cannot resolve an internal-system adoption question. (builtwith.com) (almanac.httparchive.org)

How should technographic signals for target qualification be validated?

Define the exact current-use claim, map the relevant domain to the acquisition unit, preserve provider settings and timestamps, and adjudicate against category-appropriate evidence. Test verified positives and negatives alongside former users and inaccessible cases. Report precision and sample recall separately rather than using one undifferentiated accuracy score. (developers.google.com)

Is the highest-confidence detection enough to exclude a target?

The proposed protocol does not treat confidence alone as sufficient. Check empirical outcomes within the relevant score band and choose the decision threshold using the mandate's error costs. Calibration concerns the relationship between a score and observed outcomes; it is distinct from a provider's descriptive confidence label. (scikit-learn.org) (developers.google.com)

What should a buyer compare when evaluating BuiltWith alternatives?

Compare the same dated cases, claim definitions, collection modes, evidence availability, unresolved-case handling, contractual permissions, and total review cost. Dataset completeness and freshness should be evaluated against intended use. A broad technology catalog or a low license allocation cannot substitute for a measured, mandate-specific acceptance test. (www.gov.uk)

Conclusion

Technographic data can support acquisition sourcing when the team keeps the observed signal separate from the business conclusion. The relevant object is a dated finding about a defined website or technical surface, attributed to the correct company and assessed against a written mandate. Public-code detection has useful

scope, but its limits remain part of the answer. (builtwith.com)

The recommended operating sequence is straightforward: define the claim, build independently labeled cases, collect comparable provider outputs, adjudicate with appropriate evidence, and precommit to acceptance and escalation rules. Keep former users, redirects, parked domains, subsidiary sites, and inaccessible cases visible throughout the process. Their treatment determines what the final score actually means.

Precision, sample recall, unresolved outcomes, and review cost should be read together. A high share of correct positive detections does not demonstrate complete coverage of qualifying companies; recall asks a different question. Equally, a record that cannot be adjudicated should remain an explicit unknown rather than quietly becoming a negative. (developers.google.com)

The purchasing decision follows from that discipline. Buy the workflow that produces sufficiently supported answers at an acceptable cost for the mandate, and retest when the evidence or collection process changes. Keep raw observations, interpretations, and later corrections distinguishable. The result is a technology filter whose decisions can be reproduced and challenged, with uncertainty carried forward instead of hidden inside a binary field.

Source links

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

1. https://builtwith.com/terms
2. https://shopify.dev/docs/storefronts/headless/overview
3. https://www150.statcan.gc.ca/n1/pub/12-001-x/2019002/article/00006/02-eng.htm
4. https://www.census.gov/about/policies/quality/standards/standarda3.html
5. https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=en
6. https://scikit-learn.org/stable/modules/calibration.html
7. https://itl.nist.gov/div898/handbook/prc/section2/prc241.htm
8. https://www.wappalyzer.com/docs/api/v2/lookup
9. https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework
10. https://builtwith.com/biup
11. https://almanac.httparchive.org/en/2025/methodology
12. https://github.com/ahrefs/Wappalyzer
13. https://developers.cloudflare.com/dns/proxy-status
14. https://docs.stripe.com/payments/payment-element
15. https://developers.google.com/tag-platform/security/concepts/consent-mode
16. https://kb.builtwith.com/general-questions/confidence-score
17. https://api.builtwith.com/domain-api
18. https://www.gleif.org/en/organizational-identity/lei-vlei/the-legal-entity-identifier-lei
19. https://www.icann.org/en/contracted-parties/consensus-policies/registration-data-policy
20. https://www.w3.org/TR/prov-dm

21. https://developer.mozilla.org/en-US/docs/Web/HTTP/Guides/Redirections
22. https://developer.mozilla.org/en-US/docs/Web/HTTP/Reference/Headers/Server
23. https://www.wappalyzer.com/terms
24. https://developers.google.com/machine-learning/crash-course/classification/thresholding
25. https://www.wappalyzer.com/faq/api

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.