Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

Wayback Machine Due Diligence: A Website Evidence 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 · 2026-10-15 · wayback machine due diligence · archived website evidence · internet archive · website history · private equity research · capture integrity · historical company claims · evidence preservation

Original article: https://docket.capital/articles/wayback-machine-due-diligence-evidence-protocol



In this article

1. Executive Summary
2. Introduction and Background
3. Key Changes
4. Implementation Considerations and Process Changes
5. Capture Integrity and Snapshot Selection
6. Corroboration and Claim Decisions
7. Data Analysis and Evidence
8. Case Studies and Real-World Examples
9. Implications and Future Directions
10. Conclusion

Executive Summary

Wayback Machine due diligence is a process for deciding whether an archived website supports a specific historical statement about an acquisition target. The archive timestamp identifies a crawl, rather than the date an author wrote a statement. An analyst should therefore distinguish the company's asserted event date, the page's publication date, the archive capture date, and the research retrieval date. Official archive help explains the timestamp; institutional citation guidance separately identifies publication, archive, and access dates. (archivesupport.zendesk.com) (www.nationalarchives.gov.uk)

The recommended decision is narrower than accepting or rejecting an entire website. Define the claim, resolve the company and URL, bracket the relevant period, inspect the material resources, and preserve the consulted capture. Assign **archive-supported historical statement**, **archive-observed but uncorroborated**, **conflicting**, **capture incomplete**, or **not found**. These are proposed research statuses, not Internet Archive classifications. A company statement can be faithfully preserved while its underlying business assertion remains unverified. Evidence-quality principles support considering independence and relevance alongside quantity, rather than using snapshot counts as confidence scores. (pcaobus.org)

Historical research explains why preservation matters without supplying a current archive-accuracy rate. Pew Research Center reported that **25%** of sampled pages from **2013 through 2023** were inaccessible by its 2023 assessment, rising to **38%** for the 2013 pages. (www.pewresearch.org) Its methodology used approximately **one million pages**. (www.pewresearch.org) A separate historical study reported that **86%** of studied website links to the archive targeted timestamped pages, and **83%** of those originals no longer existed live. These are results from different populations and methods, not probabilities that a target claim is true. (arxiv.org)

The practical deliverable is an evidence packet: original and archived URLs, literal wording, separate dates, integrity observations, corroborating records, and an explicit disposition. Citation guidance favors identifying the actual archived page consulted. (www.chicagomanualofstyle.org) A retained-file hash can help detect subsequent changes, but cannot establish the truth of the original business statement. (csrc.nist.gov) Authentication also leaves other evidentiary questions unresolved, so material legal reliance requires jurisdiction-specific counsel. (www.law.cornell.edu) For origination, the useful outcome is a dated, reviewable finding with its limits intact.

Introduction and Background

A target's present website may no longer contain an earlier location, service description, leadership biography, or ownership statement. The research question is often precise: did the company's website display a particular claim during the period relevant to a screening decision? A historical source can help answer that question even when it cannot establish whether the company actually performed the service or operated the location.

Two distinct problems motivate the work. **Link rot** concerns unavailable resources; **content drift** concerns changed resources at an existing address. Research on scholarly references distinguishes these phenomena, so finding a functioning live URL does not by itself recover the material originally consulted. (journals.plos.org) The Library of Congress likewise explains that an archived website does not update with the live site and directs readers to the live website for current information. (www.loc.gov)

This report proposes a **capture-to-claim protocol** for private equity (PE) origination research, current as a methodology on the publication date. It addresses historical company claims rather than a general website-history tour. For Docket's PE readership, the relevant outcome is a reviewable decision about a target fact, with unresolved questions visible to the next researcher.

The framework combines archive-specific checks with deliberately bounded evidence reasoning. Arora and colleagues' archived-web methodology includes sampling, crawl boundaries, collection, operationalization, integration, and analysis. (asistdl.onlinelibrary.wiley.com) The Public Company Accounting Oversight Board (PCAOB) distinguishes evidence quality from quantity; this report borrows that distinction as an analytical analogy, not as an auditing requirement for origination teams. (pcaobus.org)

The governing question is always: **what does this particular capture support?** A dated website statement may justify a historical research note while leaving current operations, legal ownership, revenue, employee count, and willingness to sell unanswered. Those fields need evidence suited to the proposition being tested.

Key Changes

The proposed changes apply to an origination team's research workflow.

Define the claim before opening the calendar

Begin with a sentence that could be accepted or rejected without interpretation. Replace “verify company history” with “determine whether the identified company's service page displayed a named service during the relevant period.” The following fields are a proposed research specification:

- **Proposition:** The exact factual statement under review.
- **Entity:** Legal name, jurisdiction, known brand, and identity uncertainty.
- **Source scope:** The page, domain, language, and material linked documents.
- **Relevant period:** The event window that matters to the decision.
- **Evidence threshold:** Website display alone, or independent corroboration.
- **Permitted wording:** The sentence a reviewer may approve if the threshold is met.

Set the threshold before seeing an attractive snapshot. Otherwise, an analyst may silently substitute evidence that the company advertised a capability for evidence that it delivered that capability. The proposed separation follows the broader principle that an independent, knowledgeable source offers a different evidentiary basis from the company's own representation. (pcaobus.org)

Separate the dates

Table 1 defines the proposed date fields. The National Archives' citation guidance distinguishes original publication, archived capture, and access dates; the asserted event date is an additional analytical field needed for company claims. (www.nationalarchives.gov.uk)

Date field	What it records	Decision rule
Asserted event date	When the source says a service, opening, appointment, or other event occurred.	Attribute it to the source unless independently corroborated.
Publication or update date	A date displayed on the page or document.	Preserve its wording and location; leave unknown when absent.
Capture timestamp	When the archive's crawler retrieved the resource.	Read the timestamp of the actual consulted capture. (archivesupport.zendesk.com)
Research retrieval date	When the analyst accessed the archived material.	Record separately, including the time zone in the working file. (www.nationalarchives.gov.uk)

The dates answer different questions. A page captured after an asserted opening can preserve a retrospective claim without demonstrating that the same wording was available on the opening date. Similarly, a first observed capture establishes an observation in the searched archive record, not the company's founding date or the page's original launch date. That inference is especially unsafe where crawl coverage has known boundaries. (www.bl.uk)

Normalize the URL set without erasing distinctions

Preserve the entered URL and the final destination. Hypertext Transfer Protocol (HTTP) origins can differ by scheme, host, or port; related addresses should not be presumed equivalent simply because they appear to represent the same business. (www.rfc-editor.org) URL parameters also require care because servers can interpret them differently. (developer.mozilla.org)

- **Protocol:** Try documented HTTP and HTTPS variants.
- **Host:** Check known www and non-www addresses.
- **Path:** Retain meaningful page names, extensions, and trailing-slash variants.
- **Redirect:** Record both the source address and destination.
- **History:** Include evidenced previous domains and localized pages.
- **Parameters:** Preserve query strings until their irrelevance is established.

URL structure distinguishes the domain, path, and parameters, so retain them as separately inspectable parts of the address. (developer.mozilla.org) This is a search map, not permission to merge distinct entities. A domain transition and a legal-entity transition should remain separate hypotheses until corroboration connects

them.

Implementation Considerations and Process Changes

Select observations deliberately

Search the archive using the recorded original page address, inspect the available capture dates, and open the selected timestamped page. Record the actual returned timestamp. (archivesupport.zendesk.com)

Choose captures around the event window rather than selecting the most visually complete page. The proposed **snapshot bracket** consists of the last usable observation before the relevant boundary and the first usable observation after it. Add intermediate captures when they could narrow a material wording change. If the earlier capture lacks a component, the bracket applies only to what was actually observed.

For each selected capture, write an observational sentence before an inference: “the service text is present,” “the linked document could not be opened,” or “the destination timestamp differs.” CalConnect's bibliographic guidance says to cite the archival copy actually used and verify the information before substituting another copy. (standards.calconnect.org)

Keep the observation window explicit. Where a later page includes a service absent from an earlier complete service list, the analyst can report a change in observed website content between the captures. The exact introduction date remains unresolved unless additional evidence supplies it. Archive exclusions and inaccessible content make a broader claim about what never appeared inappropriate. (archivesupport.zendesk.com)

Build the retained evidence packet

The following packet is a proposed internal schema rather than a universal citation standard. Oxford's guidance notes that archive and access dates can be added, while formal citation conventions differ. (libguides.bodleian.ox.ac.uk)

- **Claim identifier:** Connect the observation to the target and screening question.
- **Original address:** Preserve the complete source URL and any observed redirect.
- **Capture address:** Retain the exact timestamped archive URL consulted.
- **Date fields:** Separate event, publication, capture, and retrieval dates.
- **Literal excerpt:** Copy the relevant wording without silently correcting it.
- **Retained representation:** Keep permitted screenshots, documents, or other copies.
- **Integrity notes:** Record missing resources, substitutions, and inspection limits.
- **Review record:** Preserve the decision, rationale, reviewer, and open questions.

For an external citation, identify the original page and the archive that delivers the consulted version. Modern Language Association guidance separates the original page's information from the Internet Archive container, while Chicago guidance recommends crediting the archive and linking to the archived page. (style.mla.org) (www.chicagomanualofstyle.org) The National Archives additionally recommends preserving the complete archived address with its timestamp. (www.nationalarchives.gov.uk)

A screenshot is useful for showing the rendered wording, but the packet should explain which material resources the reviewer inspected. A visually persuasive page is not an adequate substitute for a readable supporting document. Perma.cc's documentation also asks users to review a newly created record for correct display, demonstrating that preservation services themselves expect inspection after capture. (perma.cc)

Check file stability separately

Where retention policy permits, add an optional **SHA-256 hash**, using Secure Hash Algorithm 256, to identify a retained file. NIST's Secure Hash Standard describes digests as a means of detecting changes after generation. (csrc.nist.gov) Record the algorithm, file name, hash, and collection event together. W3C guidance also recommends an independent checksum record as protection against accidental file corruption. (www.w3.org)

A matching hash supports file stability relative to the recorded digest. It does not show that the company statement was accurate, establish authorship, or recover the original publication time. Keep that interpretation separate from the claim decision. The Digital Preservation Coalition recommends regular checksum checks and allows checksum records in metadata, databases, or manifests. (www.dpconline.org)

Capture Integrity and Snapshot Selection

Inspect the material resource, not just the calendar

Calendar colors describe response classes: Internet Archive documents blue for successful responses, green for redirects, orange for client errors, and red for server errors. They are not ratings of completeness or factual accuracy. (archivesupport.zendesk.com) HTTP success likewise concerns successful receipt, understanding, and acceptance of the request, not validation of the page's business claims. (www.rfc-editor.org)

Table 2 is the proposed **capture-integrity checklist**. Each row produces an observation and a disposition for the material supporting the claim.

Check	What to record	Disposition rule
Original URL	Requested address, destination, and identity match.	Escalate unexplained domain or entity changes.
Timestamp	Actual page capture and relevant resource timestamps.	Preserve any mismatch material to the claim.
Status color	Calendar color and available response information.	Open the capture; a colored dot alone is insufficient. (archivesupport.zendesk.com)
Redirect	Starting capture and destination capture.	Inspect the destination rather than accepting the redirect as page content. (www.rfc-editor.org)
Date substitution	Requested date versus returned date.	Use the returned capture date in the observation. (archivesupport.zendesk.com)
Material assets	Images, PDFs, and other documents carrying the assertion.	Mark incomplete when necessary evidence cannot be read.

Check	What to record	Disposition rule
Interactive content	Forms, search, scripts, and content dependent on interaction.	Do not infer an absent service from a nonworking interface. (archivesupport.zendesk.com)
Live-resource exposure	Any material content retrieved outside the archive.	Separate it from the historical observation. (archivesupport.zendesk.com)

The checklist is claim-specific. Missing decorative imagery may not affect a plain-text service statement. A missing certification image, location map, or linked ownership document can be decisive if that asset supplies the claim. Edinburgh's library recommends opening the capture and following links to check capture quality, while warning that some crawler-dependent functionality may not survive. (library.ed.ac.uk)

Handle mixed dates and incomplete navigation

An archived page can be a composite of resources. Historical temporal-coherence research examined whether such recomposed pages could have existed together as presented; its findings concern the tested recompositions, not the truth of statements on every current capture. (hvdsomp.info) For research, the important response is to inspect the dates and provenance of the resources carrying the assertion.

Use the following checks for material content:

- **Main text:** Read the sentence in context and retain its qualifiers.
- **Image text:** Open the underlying image when the assertion depends on it.
- **PDF text:** Inspect the document itself and its own visible dates.
- **Navigation:** Follow the specific link that supposedly supports the claim.
- **Returned date:** Recheck the timestamp after each material navigation.
- **Live content:** Record current material separately if encountered.
- **Missing content:** State what was unavailable and why it mattered.
- **Final scope:** Limit the accepted statement to the resources inspected.

Search boxes and forms require particular caution. The Library of Congress warns that some archived functions do not work, and the British Library describes both crawl-depth limits and barriers behind login facilities. (www.loc.gov) (www.bl.uk) A homepage capture therefore does not establish that every historical service or location page was collected.

An error also needs precise interpretation. HTTP 404 does not itself establish whether the absence is temporary or permanent. (www.rfc-editor.org) Record the observed failure rather than converting it into a business conclusion such as a closure or service withdrawal.

Corroboration and Claim Decisions

Match the corroborating source to the proposition

A preserved company statement remains company-originated information. Repetition across snapshots improves the observation history but does not create source independence. PCAOB guidance considers knowledgeable independent evidence generally more reliable; its relevance here is the distinction between

additional copies and additional evidentiary bases. (pcaobus.org)

The proposed corroboration plan is field-specific:

- **Legal identity:** Compare the domain's named entity with jurisdiction-specific records.
- **Historical service:** Seek a dated operating document or independent contemporaneous source.
- **Location:** Distinguish an advertised service area, mailing address, and operating facility.
- **Leadership:** Separate a biography's displayed role from appointment and departure dates.
- **Ownership:** Seek entity-matched transaction or ownership records for the relevant period.
- **Current activity:** Obtain current evidence before carrying a historical finding forward.

For United States filers, the Securities and Exchange Commission's Central Index Key (CIK) identifies entities and individuals with disclosure records; its lookup instructions warn that recorded names can differ from expected names. (www.sec.gov) Its filing search provides registration statements, periodic reports, and other forms. (www.sec.gov) Use a filing only if its entity, period, and contents address the proposition.

In the United Kingdom, Companies House permits viewing company information without registration, but its current guidance says examinations of accounts involve basic checks. Treat the filing as a dated filed representation with the appropriate limits. (www.gov.uk) Where available, a Legal Entity Identifier (LEI) can assist matching: GLEIF describes it as a unique **20-character alphanumeric code** representing a single entity. (www.gleif.org) The identifier does not independently resolve a historical service claim.

Apply an explicit decision rule

Table 3 maps observations to proposed statuses. These are internal research decisions, not labels supplied by an archive or determinations about admissibility.

Claim status	Evidence condition	Safe wording and next action
Archive-supported historical statement	The identified capture is sufficiently intact for the exact display claim.	“The archived company page displayed the statement at the recorded capture.” Seek operational corroboration if needed.
Archive-observed but uncorroborated	The wording is visible, but the mandate asks whether the underlying business assertion was true.	Attribute the statement to the company and keep the target fact unresolved.
Conflicting	Relevant sources give incompatible answers after entity, period, and definitions are aligned.	Preserve both observations and escalate the discrepancy.
Capture incomplete	Missing or temporally unsuitable material prevents assessment.	Record what can be read and the missing dependency; seek another capture or source.
Not found	The documented URL set and period search yielded no usable support.	State the search scope and retrieval date; do not assert that the event never occurred.

The same capture can support a display claim while leaving an operating claim uncorroborated. The claim record therefore needs both the literal observation and the proposed business conclusion. A reviewer should be able to reject the inference while retaining the source observation.

Accept a historical display statement when identity, date, and material content meet the specified threshold. **Escalate** an underlying operating assertion when corroboration is required or sources conflict. **Reject** the proposed inference when it depends on the wrong entity, an unsuitable date, or an unreadable material asset. The emphasis on evidence quality follows the analytical analogy to PCAOB guidance. (pcaobus.org)

Resolve apparent contradictions in a fixed order: identity, time period, definition, source scope, then reliability. A service-area list and an office list answer different questions. If the conflict survives these checks, retain it rather than selecting the wording that best fits the investment thesis.

Data Analysis and Evidence

What the published numbers establish

The research base supports structured collection rather than a universal accuracy score. Arora and colleagues' methodology was illustrated with **300 United States small and medium-sized enterprises** in green goods industries. (vpinstitute.org) Its abstract emphasizes both machine and human effort. (asistdl.onlinelibrary.wiley.com) The sample is a methodological illustration, not a PE-target benchmark.

AlNoamany and colleagues' **2013** paper reported that **86%** of studied website links to the archive targeted timestamped archived pages, and **83%** of those originals were no longer live. (arxiv.org) The underlying access-log dataset contained **six million records** from sampled intervals on **February 2, 2012**. (www.cs.odu.edu) Those records were not six million independent company-page observations. The results describe historical linking behavior, not present-day claim validity.

Pew's **2024** report found **25%** of sampled pages from **2013 through 2023** inaccessible in its **2023** assessment, including **38%** of pages collected from **2013**. (www.pewresearch.org) Its methodology describes approximately **one million pages**, with availability checks from **October 12 through November 6, 2023**. (www.pewresearch.org) It treated several ambiguous responses as accessible, so accessibility should not be equated with unchanged, useful content. (www.pewresearch.org)

A **2025** longitudinal-sampling preprint describes **27.3 million URLs** and **3.8 billion archived pages** spanning **1996 through 2021**. (arxiv.org) Its authors note that archiving speed and capacity increased over time. (arxiv.org) Raw counts across years therefore require methodological interpretation; they cannot be read as uniform opportunities to observe a company.

These studies address different units: firms, access-log records, live URLs, and archived pages. Pooling their percentages would create a denominator with no coherent meaning. None supplies a probability that a particular ownership, location, or service statement is true.

Measure workflow coverage separately (Hypothetical Example)

The following invented batch illustrates arithmetic, not observed performance. Suppose a researcher attempts **12 URLs**, finds any capture for **9**, opens **18 captures**, judges **11** complete enough for their assigned claims, and corroborates **4 of 7 business claims**.

- **URL capture coverage:** 9 divided by 12 equals **75%**.
- **Inspected-capture usability:** 11 divided by 18 is approximately **61.1%**.

- **Claim corroboration coverage:** 4 divided by 7 is approximately **57.1%**.
- **Outstanding claims:** 7 minus 4 leaves **3** requiring a documented disposition.
- **Denominator control:** URL, capture, and claim totals remain separate.

The fractions describe work performed and evidence obtained under the proposed definitions. They are not factual-confidence estimates. Record whether redirects count as separate attempts, whether repeated captures are deduplicated, and what “complete enough” means before comparing batches. The proposed dashboard should display the actual counts alongside each rate.

Case Studies and Real-World Examples

Service and location change (Hypothetical Example)

Everything in this example, including the company, dates, and quotations, is fictional. An analyst is researching **Northbank Technical Services**, an invented business, and asks whether it offered laboratory-equipment maintenance from a regional operating location during **2021**. The invented mandate requires independent support for operations, while allowing an attributed website statement as a discovery lead.

The analyst defines a URL set containing the homepage, services page, locations page, and a linked brochure. An invented capture on **March 10, 2021** displays “industrial equipment maintenance.” The complete service list does not mention laboratory equipment. The locations page lists “Central office,” without clarifying whether it is an operating facility or mailing address.

An invented capture on **September 20, 2021** displays “laboratory equipment maintenance.” However, the linked brochure resolves to a later archived resource dated **February 2, 2022**. The September page supports a statement about the visible service text. The later brochure cannot establish that its contents were available in September. The analyst preserves both addresses and records the date mismatch.

An invented capture on **November 5, 2021** lists “East service center.” A separate fictional operating document dated **December 15, 2021** identifies the same legal entity and facility address. A hypothetical registry record supplies an entity match but says nothing about equipment maintenance. The analyst avoids treating that identity match as service corroboration.

The proposed timeline is explicit: **March observation, September changed wording, November location wording, December operating document**. The service wording changed between the observed March and September pages. The precise launch date remains unknown. The regional location has later corroboration within the invented period, but the September brochure has unsuitable timing for the September assertion.

The reviewer records separate outcomes:

- **Display claim:** Archive-supported, because the September service wording is readable.
- **Service performance:** Archive-observed but uncorroborated, because no independent operating evidence addresses that service.
- **Location activity:** Corroborated only for the scope and date covered by the fictional operating document.
- **Brochure dependency:** Capture incomplete for any claim requiring its contents as of September.

- **Current operations:** Unanswered by the historical packet.

This example shows why a single company-level confidence label loses information. The evidence can justify one sentence, qualify another, and leave a third open. An origination reviewer can preserve the useful historical signal without converting it into a stronger operating claim.

Implications and Future Directions

Preserve new observations without inventing a past

Use prospective capture when a live page matters to an active research question. Internet Archive's help page describes saving a single page rather than a whole site. (archivesupport.zendesk.com) Its October 23, 2019 announcement also documents optional “save outlinks” capture: distinguish a basic page save from a request to capture linked pages. (Internet Archive) The interface's current behavior was not independently tested for this report. Inspect every material linked page or resource after capture; do not treat the selected capture options as proof of completeness.

Another preservation service can supplement the record, provided the researcher inspects the retained representation. Perma.cc documents an interactive archive, a screenshot, and a route to the live page. These representations serve different review needs; the analyst should name which one supports the excerpt. (perma.cc)

Future automation should prioritize URL inventories, timestamp comparisons, duplicate detection, and explicit missing-resource flags. Keep the claim inference reviewable. This allocation is consistent with archived-web methodology emphasizing combined machine and human effort. (asistdl.onlinelibrary.wiley.com) An automated change detector can propose that wording differs, while the approved research record should explain the meaning of that difference for the mandate.

Know when the decision exceeds research scope

Internet Archive's help confirms that a certified-record and affidavit request process exists. Its detailed legal-procedure and terms pages could not be retrieved in this research session after the required fallback. No fees, delivery times, procedural conditions, or current permission terms are represented as verified here. (archivesupport.zendesk.com)

For legal reliance, preservation obligations, or contested authentication, consult counsel in the relevant jurisdiction and obtain the current archive procedure. United States Federal Rule of Evidence 901 concerns showing that an item is what its proponent claims; its accompanying notes make clear that authentication does not assure admission. This is a boundary explanation, not advice about a particular record. (www.law.cornell.edu)

Retention and sharing also require rights and confidentiality decisions. The United States Copyright Office explains that fair use depends on circumstances rather than a fixed permitted word count or percentage. Apply organizational policy and seek appropriate advice before retaining or distributing material beyond an approved research purpose. (www.copyright.gov)

Conclusion

Wayback Machine due diligence is most useful when the analyst starts with a precisely bounded proposition and ends with a precisely bounded statement. The archive can preserve what a company page displayed at an observed capture. The analyst still has to establish the entity, distinguish the dates, inspect the relevant resources, and decide whether the business assertion needs corroboration.

The proposed protocol turns those obligations into a repeatable review. Normalize the URL set without erasing meaningful differences. Select captures that bracket the relevant period. Inspect the material text and linked assets. Preserve the consulted address, excerpt, retrieval record, integrity notes, and decision. Keep contradictory observations and unsuccessful searches in the file.

The result should be a claim-level disposition, not an unsupported confidence score for the company. **Archive-supported** can be appropriate for the historical display statement while **uncorroborated** remains appropriate for the underlying service or operating claim. Missing captures justify a scoped not-found result; they do not establish that an event never happened.

For a PE origination team, the practical standard is whether another researcher can reproduce the observation, understand its limits, and see what evidence would change the decision. That standard preserves the value of historical website research while keeping current operations, ownership, financial performance, and legal reliance subject to their own evidence requirements.

Source links

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

1. https://archivesupport.zendesk.com/hc/en-us/articles/360004651732-Using-The-Wayback-Machine
2. https://www.nationalarchives.gov.uk/help-with-your-research/discovery-help/citing-records-national-archives
3. https://pcaobus.org/oversight/standards/auditing-standards/details/AS1105
4. https://www.pewresearch.org/data-labs/2024/05/17/when-online-content-disappears
5. https://www.pewresearch.org/data-labs/2024/05/17/methodology-link-rot
6. https://arxiv.org/abs/1309.4016
7. https://www.chicagomanualofstyle.org/qanda/data/faq/topics/Documentation/faq0434.html
8. https://csrc.nist.gov/pubs/fips/180-4/upd1/final
9. https://www.law.cornell.edu/rules/fre/rule_901
10. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0167475
11. https://www.loc.gov/programs/web-archiving/for-site-owners/frequently-asked-questions
12. https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.23503
13. https://www.bl.uk/services/legal-deposit/web-archiving
14. https://www.rfc-editor.org/rfc/rfc9110.html
15. https://developer.mozilla.org/en-US/docs/Learn_web_development/Howto/Web_mechanics/What_is_a_URL
16. https://standards.calconnect.org/docs/cc-6900-2019.html
17. https://archivesupport.zendesk.com/hc/en-us/articles/360004716091-Wayback-Machine-General-Information

18. https://libguides.bodleian.ox.ac.uk/web-archives
19. https://style.mla.org/citing-work-wayback-machine
20. https://perma.cc/docs
21. https://www.w3.org/Provider/Style/URI.html
22. https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums
23. https://library.ed.ac.uk/heritage-collections/collections-and-search/archives/digital-archives-and-preservation/web-archiving/frequently-asked-questions
24. https://hvdsomp.info/papers/Papers/2015/ht15-ainsworth-submission.pdf
25. https://www.sec.gov/search-filings/cik-lookup
26. https://www.sec.gov/search-filings
27. https://www.gov.uk/guidance/searching-the-companies-house-register
28. https://www.gleif.org/en/organizational-identity/lei-vlei/the-legal-entity-identifier-lei
29. https://vpinstitute.org/using-the-wayback-machine-to-mine-websites-in-the-social-sciences-a-methodological-resource
30. https://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_149.pdf
31. https://arxiv.org/abs/2507.14752
32. https://archivesupport.zendesk.com/hc/en-us/articles/360001513491-Save-Pages-in-the-Wayback-Machine
33. https://blog.archive.org/2019/10/23/the-wayback-machines-save-page-now-is-new-and-improved
34. https://www.copyright.gov/help/faq/faq-fairuse.html

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.