Athenica AI Document Review System – Plain-English Guide (v2 draft)

Oct 21, 2026 · @Someone

Document status

This is a revised draft of the October 15, 2026 guide by @Shubh. It is not yet approved for build.

Item	Detail
Version	2 (draft)
Replaces	v1, October 15, 2026
Original author	@Shubh
Decision owner	Open: who signs off the build (suggest the lead Environmental Professional plus IT)
Audience	Part 1 (sections up to "Decisions to make"): staff and leadership. Technical appendix: the build team.

The big picture

The system helps Athenica produce Phase I Environmental Site Assessments (ESAs) faster. It reads every document gathered for a site, flags each finding with the exact sentence and page it came from, and drafts the report for an Environmental Professional (EP) to review and sign.

Think of it as a very fast junior analyst with a highlighter. Claude does the reading, flagging and first drafting. Athenica staff stay in charge: reviewers approve or edit every finding, and the EP signs the report and remains responsible for its conclusions.

Two rules hold throughout:

- No finding is accepted unless it links back to the original words or image in a source document.

- The system speeds up review and drafting. It does not replace the site visit, interviews or the EP's judgment.

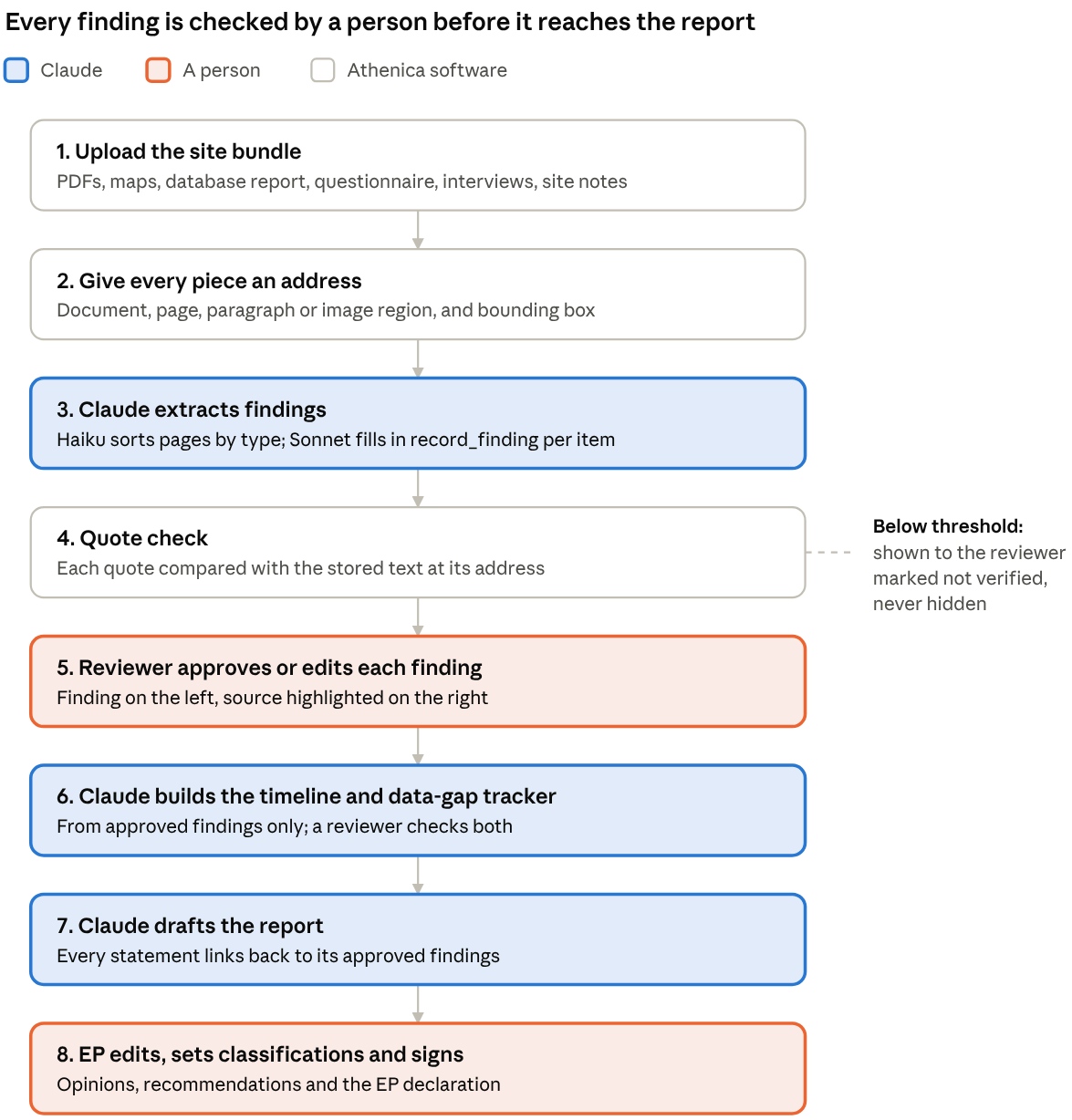
What goes in

One job is one site, not one PDF. A site job holds a bundle of sources, often 1,000+ pages in total, and each needs its own handling.

Source	Examples	How the system reads it
Text reports	Prior Phase I/II ESAs, regulatory files, permits	Text extraction, paragraph by paragraph
Scanned text	Old letters, city directory pages	OCR first, then as text
Maps and images	Sanborn maps, aerial photographs, topographic maps	Sent to Claude as images; findings point to a marked area of the image
Data tables	Lab results, tank registrations	Read as tables, kept row by row, compared against screening levels
Regulatory database report	EDR, ERIS or similar	Read as a structured listing of sites, distances and directions
User Questionnaire	Environmental liens, activity and use limitations, specialized knowledge, purchase-price questions	Entered as a form; answers become findings
Site reconnaissance	Field notes and photos	Entered as notes and uploaded photos, each tied to a location
Interviews	Owner, occupants, local officials	Entered as notes with name, role and date

How it works

Eight steps take a site from upload to a signed report. Claude reads and drafts; people approve every finding, and the EP signs.



system flow · 8 steps from upload to signed report

A quote that fails the check still reaches the reviewer, marked "not verified". Nothing enters the report without a reviewer's approval.

Example (illustrative): Claude flags a former dry cleaner on the adjoining lot. The reviewer clicks the flag, page 412 opens with the sentence highlighted, and they confirm in seconds what the document says.

What Claude returns for each finding

Claude answers by filling in a fixed form called `record_finding`, never with free text. Each field below is required.

The address. Every paragraph, table row and image region gets one address when the document is processed. It has four parts: document ID, page number, paragraph or region ID, and bounding box (the rectangle it occupies on the page). Claude copies the address it was given and never works out page numbers itself. This is why highlights land in the right place.

Field	What it holds	Example (illustrative)
Conclusion	The finding in one plain sentence	Former dry cleaner on the adjoining lot to the north
Evidence	The exact quote, or the image region for maps and photos	The sentence as printed in the report
Address	Copied from the input, as defined above	Document 102, page 412, paragraph 412-08
Location	Subject property, adjoining property, or nearby (with distance and direction)	Adjoining, north
Classification	REC, CREC, HREC, de minimis, or not a concern	REC (proposed; EP confirms)
Reason	One sentence on why that classification fits	Solvent use; no record of closure or cleanup
Rule matched	The checklist item it relates to	An ASTM E1527-21 historical-use item
Date of use	The years the evidence covers, if stated	1958–1972

Classifications are proposals. Only the EP sets the final classification.

From findings to the Phase I report

Approved findings feed three outputs, each reviewed by a person before it reaches the report.

1. Property-use timeline. A second Claude pass reads the approved findings, not the raw pages, and builds one chronology for the subject and adjoining properties. It connects sources that each tell part of a story, such as a 1960s city directory listing, a Sanborn map showing the building, and a database hit for the same address.

2. Data-gap tracker. Every required ASTM E1527-21 element is marked found, not found, or data gap. Examples of data gaps:

- Historical sources that don't reach back to 1940 or the property's first developed use
- Gaps of more than about five years between historical sources
- Interviews that could not be completed
- Areas not accessible during the site visit

The EP decides whether each data gap is significant. Extraction alone cannot find these, because they are about what is missing.

3. Draft report. Claude drafts the standard report sections from the approved findings, the timeline and the tracker. Every statement in the draft links back to its findings. The EP edits the draft, writes the opinions and recommendations, and signs the EP declaration.

The report records two dates for viability: the earliest component date, which starts the 180-day and one-year clocks, and the report date.

Keeping it private and accurate

Client files stay under Athenica's control, and no AI finding is trusted until a person has checked it against the source.

Privacy

- **Files stay in-house.** Original PDFs and photos are stored on Athenica's own servers.
- **Only staff get in.** Login uses Microsoft Entra ID, the same single sign-on as work email.
- **The AI provider's handling is set by contract.** Text and images sent to Claude are handled under the provider's written data terms for Athenica's account (Zero Data Retention with Anthropic, or the equivalent terms on AWS Bedrock). Confirm those terms in writing before go-live (see Decisions).

- **Retention.** Client files and findings are kept and deleted on a set schedule (to be decided; see Decisions).

Accuracy

- **Every quote is checked.** A script compares each quote Claude returns with the stored text at that address. A quote below the match threshold is shown to the reviewer with a clear "not verified" warning. It is never hidden, because a small OCR difference should not make a real finding disappear.
- **Wrong conclusions are a reviewer's job.** A real quote can still support a wrong conclusion. The classic case is negation: "no evidence of USTs" read as a UST finding. Reviewers see the quote next to the conclusion so this is quick to catch.
- **Missed findings are measured.** Before go-live, the system is tested against past Athenica reports with EP-approved findings (see "How we will know it works"). The test is rerun whenever the model or the rulebook changes.
- **Everything is traceable.** For each finding, the audit log records who approved or edited it, when, and which model version, rulebook version and prompt version produced it. A report can be defended years later.

How we will know it works

The system goes live only if it matches EP-approved results on past reports and saves reviewer time. Targets marked "set" are for the decision owner to fill in.

Measure	How it is measured	Target
Missed RECs and CRECs	Run on 10–20 past Athenica Phase I files; compare with the EP-approved findings	Zero missed on the test set
Missed findings overall	Same test set, all finding types	Set (e.g. at least 95% found)
Wrong classifications	Share of proposed classifications the EP changes	Set
Unverified quotes	Share of quotes flagged "not verified"	Set; track by source type
Reviewer time per report	Hours from upload to EP sign-off, before and after	Set (baseline needed first)

Measure	How it is measured	Target
Cost per site job	Model and server cost for a typical 1,000-page bundle	Set (estimate during pilot)

Decisions to make before building

Eleven open items could change the plan. Owners are suggestions.

Decision	Options or question	Suggested owner
AI provider	Anthropic API directly, or Claude on AWS Bedrock. Choose one; it sets the data terms and which models are available.	IT + leadership
Data terms in writing	Written confirmation that Zero Data Retention (or the Bedrock equivalent) covers Athenica's account and every feature used	IT + legal
Model availability	Confirm Claude Sonnet 5.5 and Haiku 5.5 are enabled for the chosen account or AWS region	IT
PDF viewer	PDF.js alone, or react-pdf-highlighter on top of it	Build team
OCR tool	Tesseract or Marker; test both on real scanned files	Build team
Quote-match threshold	How close a quote must be to count as verified	Build team + lead EP
Rulebook	Turn ASTM E1527-21 and the NYC OER guidelines into the checklist Claude receives. Keep them as separate rulebooks, since they serve different purposes. Include vapor migration (ASTM E2600) as its own item.	Lead EP
Disclosure and liability	Whether and how AI use is disclosed to clients; any requirements from the professional liability insurer	Leadership + legal

Decision	Options or question	Suggested owner
Retention	How long client files and findings are kept, and how they are deleted	Leadership + legal
Server size	How many site jobs must run at once; size the server to that	IT
Maps and images	Test Claude's image reading on a handful of real Sanborn maps and aerials before relying on it	Build team + lead EP

Technical appendix

This section is for the build team. Everything runs on one internal server, packaged with Docker Compose, except the calls to Claude.

Ten parts of the system

Part	Tools	Job
Web app	Next.js (React) + Tailwind CSS	The screens staff use: upload, review, timeline, data gaps, report draft
PDF viewer	PDF.js, possibly with react-pdf-highlighter (see Decisions)	Shows the source and highlights the exact box when a finding is clicked
Backend	Python + FastAPI	Receives uploads and coordinates every other part; the only part that talks to Claude
Document processing	PyMuPDF, plus Tesseract or Marker for OCR (see Decisions)	Extracts text, tables, images and bounding boxes; assigns addresses
Background jobs	Celery + Redis	Runs long jobs and parallel Claude calls; tracks progress; retries on rate limits

Part	Tools	Job
Database	PostgreSQL	Stores addresses, findings, reviewer edits, rulebooks, jobs and the audit log
File storage	MinIO	Keeps original files on Athenica's servers; works like Amazon S3
AI	Claude Sonnet 5.5 and Haiku 5.5, via the provider chosen in Decisions	Reads text and images, fills in findings, builds the timeline, drafts the report
Login	NextAuth.js + Microsoft Entra ID	Single sign-on with work accounts; roles for analyst, reviewer and EP
Packaging	Docker + Docker Compose	Starts all parts together with one command

pgvector (search by meaning) is left out of v2 because no step uses it yet. Add it if staff need to search past reports.

How the backend uses Claude

1. **Access.** The API key or AWS IAM role lives in the server's secrets store, never in the code or the browser. Connect with Anthropic's official Python SDK (the `anthropic` package, or its Bedrock client).
2. **Sorting pass.** Haiku 5.5 sorts pages by type (report text, map, table, directory page) so each goes to the right extraction prompt.
3. **Extraction pass.** Sonnet 5.5 receives the rulebook (marked for Prompt Caching), a batch of addressed paragraphs or images, and the `record_finding` tool. Batches follow section and table boundaries rather than a fixed paragraph count, so a table that spans pages stays together.
4. **Quote check.** A fuzzy string match compares each quote with the stored text at its address. Results below the threshold are saved as "not verified".
5. **Synthesis pass.** After review, Sonnet 5.5 reads the approved findings for the whole site and builds the timeline and data-gap tracker.
6. **Drafting pass.** Sonnet 5.5 drafts report sections from approved findings only, citing finding IDs.

7. **Logging.** Every call records model version, rulebook version, prompt version and token cost against the job.

Build stages

1. **Server.** Start PostgreSQL, Redis, MinIO and FastAPI with Docker Compose.
2. **Document processing.** Extract text, tables, images and bounding boxes; assign addresses; run OCR on scans. Test on real files of every source type.
3. **Extraction.** Connect Claude with Tool Use and Prompt Caching; add the quote check.
4. **Review screen.** Findings on the left, source on the right; a click opens the page and box (for example `#page=412&bbox=...`).
5. **Field inputs.** Forms for the User Questionnaire, interviews and site visit notes and photos.
6. **Synthesis and drafting.** Timeline, data-gap tracker and draft report.
7. **Test and lock down.** Run the accuracy test set; confirm data terms; turn on retention rules and the audit log.

Change log

Version	Date	Changes
2 (draft)	October 21, 2026	Scope widened from document review to supporting the full Phase I: site bundles, field inputs, images, classification, timeline, data-gap tracker, report drafting and EP sign-off. Mismatched quotes now shown with a warning instead of hidden. One address definition used throughout. Three overlapping step lists merged into one flow plus a technical appendix. Open choices moved to Decisions. Privacy wording tied to written provider terms. Success measures added. Copy fixes.
1	October 15, 2026	First guide by @Shubh. Updated model names from Claude 3.5 Sonnet and Haiku to Claude Sonnet 5.5 and Haiku 5.5.