



Ariadne

Desktop forensic intelligence workstation for evidence ingestion, entity graphs, geo/heat mapping, face intelligence, contradictions, and reporting.

User Manual

February 19, 2026



Important notes and limitations

Ariadne accelerates triage and investigative navigation, but it does not replace human verification. All extracted entities, relations, dates, locations, face matches, and “contradictions” are heuristic outputs that must be reviewed against the original evidence.

The built-in Evidence viewer currently renders PDF pages only. Non-PDF files can be ingested (text extraction), but opening a page as evidence may fail unless the source is a PDF.

Entity Resolution (merge/split) and Ontology migrations are data-changing actions. Back up the database before applying them in a live investigation.



Table of contents

This document uses Word heading styles. In Microsoft Word, you can generate an automatic table of contents via References → Table of Contents.

- 1. Getting started
- 2. Building the database (Ingestion and Indexing)
- 3. Exploring the case (tabs overview)
- 4. Data quality and schema management (Resolution and Ontology)
- 5. Exporting and reporting
- 6. Keyboard and mouse reference
- 7. Troubleshooting
- Appendices



1. Getting started

1.1 What Ariadne does

- Ingest evidence from the **files/** folder (PDFs and common office/text formats).
- Extract text (including OCR for image-based PDFs when Tesseract is available).
- Extract entities (people, organizations, locations, events, accounts) and build a relationship graph.
- Extract identifiers (email, URL, domain, IPv4, IBAN, Portuguese NIF, phone, hashes) and link them to entities.
- Detect faces in PDF pages (optional) and provide a Face Gallery with thumbnails per person.
- Generate investigation views: Events Timeline, Geo Map, Heat Map, Contradiction Radar, Alerts and Patterns.
- Produce exports: alerts (CSV/XLSX), dossiers (Markdown/HTML/CSV), intelligence report (DOCX).

1.2 Typical investigative workflow

1. Place evidence into the correct subfolders under files/ (PDFs, Emails, Finance, etc.).
2. Open Settings → choose AI mode (Local or OpenAI) and configure the OpenAI key/model if needed.
3. Run Ingestion and Indexing → Build Database (first run) or Reindex New/Changed Only (updates).
4. Start with Alerts and Patterns to surface obvious anomalies, then explore the Relationships graph.
5. Use Smart Search to ask investigative questions and jump directly to cited evidence pages.
6. Use Events Timeline / Geo Map / Heat Map / Contradiction Radar to build a coherent narrative and identify inconsistencies.
7. Optionally clean up identity duplication using Resolution; keep the schema consistent using Ontology.
8. Export a Dossier or an Intelligence Report for reporting and handoff.

1.3 System requirements

- Windows 10/11 (portable folder deployment).
- CPU: modern 4-core+ recommended; RAM: 16 GB+ recommended for large corpora.



- Disk: allow space for the database and face thumbnail cache (can grow with large PDF sets).
- Optional: Tesseract OCR installed for scanned PDFs (otherwise OCR is skipped).
- Optional: OpenAI API key if using OpenAI mode.

1.4 Folder layout (inputs)

Ariadne expects a folder named files/ next to the executable (or next to ui_main.py in dev mode).

Path (relative to application folder)	Purpose / notes
files/PDFs/	PDF evidence. Best supported format: evidence viewer, page references, faces, OCR.
files/Logs/	System logs, exports, CSV/JSON logs. Useful for timeline/log extraction.
files/CRM/	Contacts, CRM exports, notes.
files/ERP/	ERP exports (orders, invoices, inventory).
files/Finance/	Statements, payments, invoices, bank evidence.
files/WhatsApp/	Chat exports.
files/Email/	Email exports.
files/OSINT/	Web captures, OSINT reports, intelligence notes.
files/_tmp/	Ignored by ingestion (use for staging).
files/_ignore/	Ignored by ingestion (use for non-case materials).
files/ (root)	Legacy: PDFs placed directly under files/ are still ingested.

1.5 Supported file formats

Ariadne ingests PDFs and several common text/office formats. Non-PDF files are treated as a single text page (no rendering in the Evidence viewer).

Format	Ingestion	Notes
PDF	Full: page text + optional OCR + faces + page evidence	Recommended for primary evidence.
TXT / MD	Text extraction	Single “page”.
CSV / JSON / XML	Text extraction	Single “page”. Large files may increase ingestion time.
HTML	Text extraction	Basic tag stripping.
DOCX	Text extraction	Extracts paragraph text.



XLSX	Text extraction	Extracts cell text across sheets.
-------------	-----------------	-----------------------------------

1.6 Where Ariadne stores its data

By default, Ariadne stores state under your Desktop:

Path	Contents
Desktop/ariadne-bd/ariadne-storage/ariadne.db	Main SQLite database (entities, relations, passages, events, alerts, etc.).
Desktop/ariadne-bd/ariadne-storage/settings.json	UI and runtime settings.
Desktop/ariadne-bd/ariadne-storage/.secrets/openai_key.secure.json	Encrypted OpenAI API key (if configured).
Desktop/ariadne-bd/ariadne-output/	Logs and exports (alerts CSV/XLSX, dossiers, intelligence report DOCX, etc.).
Desktop/ariadne-faces/	Face cache DB (faces.db) and optional face crops/thumbnails.

1.7 Selecting AI mode (Local vs OpenAI)

Ariadne can run with a local model (no network calls) or with OpenAI models. AI mode affects extraction quality and speed.

Mode / model	When to use	Trade-offs
Local AI	Air-gapped workflows; predictable costs; privacy sensitive evidence.	Usually slower and lower extraction quality than OpenAI for messy sources.
OpenAI: gpt-4.1-mini	Fast, low cost triage; quick passes over large corpora.	Lower extraction quality; more misses and weaker disambiguation.
OpenAI: gpt-5-mini	Default recommendation for most cases.	Balanced cost/quality; still may miss edge cases.
OpenAI: gpt-5.2	When quality matters more than cost/latency (deep investigations).	Higher latency and cost; requires stricter budget control.

If you select OpenAI mode in Settings, Ariadne will prompt you to choose the OpenAI model. Changes are applied immediately after Save (no restart required).



2. Building the database (Ingestion and Indexing)

2.1 What “Build Database” does

- Scans files/ (all subfolders except `_tmp` and `_ignore`).
- For PDFs: reads per-page text; runs OCR on pages with no text if Tesseract is available; stores per-page snippets for citations.
- For non-PDF files: extracts text and stores it as a single “page”.
- Runs identifier extraction and links identifiers to entities.
- Runs AI extraction (Local/OpenAI) to propose entities, relations, and events.
- Optionally runs face detection/embedding on PDF pages and stores face occurrences.
- Updates indexes used by Smart Search and other tabs.

2.2 Controls in the Ingestion and Indexing tab

Buttons and options:

- Build Database — full rebuild pass over all input files.
- Reindex New/Changed Only — incremental update for new or modified files.
- Cancel — stops ingestion as soon as the current unit of work is safe to stop.
- Enable face search during ingestion — enables PDF face detection and thumbnail generation.
- Resume interrupted build from checkpoint — resumes from the last checkpoint after a crash/cancel.

Progress bars show file-level and page-level progress. The detailed log explains what is being processed, including OCR decisions, AI calls, and any parsing errors.

2.3 OCR for scanned PDFs

If a PDF page has little or no embedded text, Ariadne can run OCR using Tesseract. If Tesseract is not installed or not configured, OCR is skipped and that page may remain effectively “silent” for search.

Configuration options:

- Install Tesseract OCR on the machine.
- If Ariadne cannot find it, set the environment variable `ARIADNE_TESSERACT_CMD` to the full path of `tesseract.exe`.

2.4 Face detection during ingestion

When “Enable face search during ingestion” is enabled, Ariadne scans PDF page images for faces, generates embeddings, and stores face occurrences. Face thumbnails (crops) may be written under `Desktop/ariadne-faces/`.



If the system prompts for labels during ingestion, provide consistent naming (e.g., “John Smith”, “Unknown Male 01”). You can refine identities later using Resolution (merge/split).

2.5 Incremental updates and evidence hygiene

- Prefer immutable evidence: do not edit PDFs in place after ingestion. If you must, reindex afterwards.
- Renaming/moving files may cause them to be treated as new evidence; keep stable paths during an investigation.
- Use files/_tmp or files/_ignore for staging materials you do not want ingested.
- Back up Desktop/ariadne-bd/ariadne-storage/ariadne.db before major operations (merges, migrations).



3. Exploring the case (tabs overview)

3.1 Tabs at a glance

Tab	Primary purpose
Relationships	Interactive entity graph + evidence drill-down.
Smart Search	Ask questions; Ariadne answers with citations you can open as evidence.
Faces Intelligence	Known people list, occurrences, and Face Gallery thumbnails.
Events Timeline	Chronological timeline plus an interactive event graph view.
Geo Map	Map view of extracted locations with evidence links.
Heat Map	Heat/intensity map (People/Locations/Events) over the basemap.
Contradiction Radar	Heuristic detection of conflicting values/amounts/dates for the same identity.
Alerts and Patterns	Rule-based alerts, triage workflow (open/closed), exports.
Investigator Dossier	Curated case bundle; export Markdown/HTML/CSV.
Intelligence Report	Iterative AI narrative + exports to DOCX.
Settings	AI mode/key/model + diagnostics and paths.
Resolution	Merge/split entities to clean duplicates and conflations.
Ontology	Active ontology version + migrations + dataset stats + validation.

3.2 Relationships

Use Relationships to explore how entities connect across the corpus. The graph is built from entity co-occurrence and extracted relations.

Typical use:

- Select a focus entity (person/organization/location/etc.).
- Choose the depth (how many hops away from the focus entity).
- Build/refresh the graph.
- Click nodes/edges to view supporting evidence. Use “Open Evidence” to render the source PDF page.



If the graph feels noisy, reduce depth and/or use Resolution to merge duplicates before expanding.

3.3 Smart Search

Smart Search is a question-answer interface over the ingested corpus. Ariadne retrieves relevant passages and uses the configured AI mode to answer, always returning citations (file + page) for auditability.

- Enter your question in the text box (English or Portuguese).
- Choose Language: auto / en / pt (controls the answer language).
- Click Ask.
- Review the Answer section, then select a row in Citations and click Open Evidence.

Caveat: Smart Search answers can be wrong or incomplete. Treat them as navigation aids; verify against cited pages.

3.4 Faces Intelligence

Faces Intelligence is available when face detection is enabled during ingestion and the evidence is PDF-based.

- Select a known person from the list to view face occurrences across evidence pages.
- Use Open Evidence to open the PDF page for a selected occurrence.
- Use Open Face Gallery to browse thumbnails for the selected person and jump to the evidence page behind each thumbnail.

If thumbnails are not available, it usually means face crops were not generated (face search disabled) or the source pages were not renderable.

3.5 Events Timeline

Events Timeline aggregates extracted timeline items (e.g., flight, check-in/check-out, contacts, generic events) and provides both a chronological list and an interactive graph view.

Filters and actions:

- Set From/To dates to constrain the timeline window.
- Filter by Entity (ALL or a specific identity).
- Filter by Type: ALL / flight / checkin / checkout / contact / event / other.
- Click Refresh to rebuild the view.
- Use Reset Graph View if panning/zooming leaves the graph off-screen.
- Select an event row and use Open Selected Evidence to open the supporting PDF page.



The timeline graph supports panning/zooming and is intended for fast navigation between clustered events. Always open and verify the underlying evidence.

3.6 Geo Map

Geo Map plots extracted locations over a world basemap and links them back to evidence. It relies on a basemap image (world.jpg/png/etc.) and a locations catalogue (world_locations.json).

Prerequisites:

- Basemap image: place world.jpg (or world.png/world.webp) next to the executable if it is not bundled.
- Locations catalogue: world_locations.json can be placed next to the executable, under files/geo/, under files/, or under resources/.
- Optional: set ARIADNE_GEO_CATALOG to override the catalogue path explicitly.

Using Geo Map:

- Choose Identity = ALL or a specific identity (limits markers to that identity's evidence).
- Choose Type = ALL / PERSON / ORGANIZATION / LOCATION (controls which extracted mentions are mapped).
- Set Min confidence and Min score to reduce noise.
- Optionally enable Use date range, then set From/To.
- Optionally enable Person↔Location links to draw lines between people and locations found together.
- Click Apply Filters or Refresh Geo Cache when inputs change significantly.
- Navigation: pan with middle mouse (or Alt+left drag); zoom with mouse wheel or +/-; use Fit to Data / Reset View as needed.

3.7 Heat Map

Heat Map visualizes intensity clusters across the corpus (People / Locations / Events), helpful for finding hotspots at a glance.

Using Heat Map:

- Select the layer: People, Locations, or Events.
- Set From/To date range (optional, depending on your corpus).
- Tune Min confidence and Min score to reduce noise.
- Click Refresh to recompute the heat layer.
- Navigation: click-drag to pan; zoom with mouse wheel or +/-; press 0 to reset zoom.



- Select a row in the evidence table (if shown) and open evidence to validate.

3.8 Contradiction Radar

Contradiction Radar scans extracted events/log-like facts and highlights potential contradictions for the same identity. This is heuristic: it flags conflicts in values, amounts, or dates that appear inconsistent across sources.

Using Contradiction Radar:

- Set From/To date range.
- Select an entity (ALL or a specific identity).
- Set Min confidence to ignore weak signals.
- Set Limit to cap the number of rendered items (performance control).
- Click Scan to compute the radar graph.
- Click items in the radar/table to view details; use Open Selected Evidence to validate in the original PDF page.
- Use Reset View if the radar is off-screen after navigation.

Do not treat a radar item as proof. It is an investigator prompt for where to look, not a conclusion.

3.9 Alerts and Patterns

Alerts and Patterns provides a triage workflow: rule-based alerts can be generated, reviewed, and exported.

- Use Search and Status filters (ALL/open/under_review/closed) to triage.
- Run Pattern Rules to scan recent passages and create new alerts (current rule set is intentionally small).
- Select rows and set status using Set Open / Set Closed.
- Export alerts as CSV or XLSX (written to Desktop/ariadne-bd/ariadne-output/).

3.10 Investigator Dossier

Investigator Dossier helps you package a case narrative and key findings into a portable bundle.

- Enter a dossier title/topic and click Create / Open.
- Curate notes and key items as you navigate other tabs (entities, evidence pages, alerts).
- Export as Markdown, HTML, or CSV for sharing and archiving.



3.11 Intelligence Report

Intelligence Report runs an iterative analysis loop over evidence passages and produces a structured narrative with recurring patterns and key points.

Inputs:

- Topic — what you want the report to focus on (recommended: one clear sentence).
- Language — auto / en / pt.
- Passages per step — controls how much evidence is processed per loop (effective range is bounded).
- Steps per loop — how many retrieval/extraction steps per loop.
- Total loops — how many loops to run (more loops = deeper synthesis, more cost if OpenAI mode).

Workflow:

- Fill the parameters and click Run.
- Monitor the log output while the report is being generated.
- When complete, review the report text.
- Use Export DOCX to generate a Word document in the output folder.

3.12 Settings

Settings controls AI mode and shows diagnostic paths.

- AI mode: Local or OpenAI.
- OpenAI API key: stored encrypted (displayed as ***** once saved).
- OpenAI model: click the prompt when switching to OpenAI mode, then Save.
- Paths/diagnostics: verify where inputs, outputs, and databases live.
- Save: applies changes immediately (no restart required).



4. Data quality and schema management

4.1 Resolution: merging duplicates

Resolution is for cleaning identity duplication (e.g., “John Smith”, “J. Smith”, “Jonathan Smith”) so that the graph and analytics are less fragmented.

How to merge:

- Open the Resolution tab.
- Click Generate Candidates to compute likely duplicates.
- Select a candidate pair from the list.
- Click Preview Merge to see what will be combined (mentions, identifiers, evidence links).
- Choose which side is Primary (A is primary / B is primary).
- Click Merge to apply.

Rules: merges are allowed only for the same entity type. After merging, re-check Relationships and Timeline for improved consolidation.

4.2 Resolution: splitting conflated entities

Split is used when one entity incorrectly combines two different real-world identities (e.g., two people with the same name).

How to split:

- Select the entity to split (entity id) in the Split section.
- Click Load Split Preview to see the entity’s mentions and identifiers.
- Select which mentions and/or identifiers should move to the new entity.
- Provide a label for the new entity (be explicit: “John Smith (pilot)”, “John Smith (lawyer)”).
- Click Split to create the new entity and move the selected items.

After a split, review Relationships and Events Timeline: some relations may need manual interpretation until the corpus is re-indexed.

4.3 Ontology: active version, migrations, stats, validation

Ontology defines the formal schema used to interpret entities and relations (types, allowed relations, domain/range). Ariadne keeps the ontology versioned (semantic versioning) and can apply migrations.

Key actions:



- Reload — reload the active ontology from storage.
- Stats — compute dataset statistics by ontology class/relation.
- Validate — check whether stored relations violate domain/range constraints.
- Apply migration — upgrade the ontology to a newer version when available.

Current default ontology is intentionally minimal (PERSON, ORG, LOCATION, EVENT, DOCUMENT, ACCOUNT, IDENTIFIER). Migrations may add scaffolding for future “Claim/Fact” projection; this is schema-level only unless explicitly implemented elsewhere.



5. Exporting and reporting

5.1 Alerts exports (CSV / XLSX)

In Alerts and Patterns, use Export CSV or Export XLSX to write the current filtered alert list to the output folder.

Notes:

- Exports include severity, status, rule name, description, and evidence reference (if available).
- Default output path: Desktop/ariadne-bd/ariadne-output/alerts_export.csv or alerts_export.xlsx.

5.2 Investigator Dossier exports (Markdown / HTML / CSV)

In Investigator Dossier, export curated content for sharing or archiving.

- Export Markdown — produces a portable, diff-friendly dossier (good for Git or audit logs).
- Export HTML — produces a viewable report for non-technical recipients.
- Export CSV — produces a structured extract (useful for downstream analysis).

All dossier exports are written under Desktop/ariadne-bd/ariadne-output/.

5.3 Intelligence Report export (DOCX)

In Intelligence Report, use Export DOCX to write the final report as a Word document.

Good practice:

- Include a clear topic statement (one sentence) and keep loops small when using OpenAI mode to manage cost.
- Treat the report as a synthesis draft: verify key claims by opening cited evidence elsewhere in Ariadne.

5.4 Output folder

Default output folder is Desktop/ariadne-bd/ariadne-output/. It also contains ariadne.log.



6. Keyboard and mouse reference

The exact interaction depends on the view type. The following is the current default mapping.

View	Pan	Zoom / reset
Geo Map	Middle mouse drag OR Alt + left drag. Buttons: Fit to Data, Reset View.	Mouse wheel OR +/- buttons.
Heat Map	Left mouse drag (hand pan).	Mouse wheel OR +/- buttons. Press 0 to reset zoom.
Events Timeline graph	Middle mouse drag OR Alt + left drag. Button: Reset Graph View.	Mouse wheel. Reset Graph View resets.
Contradiction Radar graph	Middle mouse drag OR Alt + left drag. Button: Reset View.	Mouse wheel. Reset View resets.
Evidence viewer (PDF)	Scroll to move page; use open controls in the viewer dialog.	Zoom depends on viewer implementation; prefer native PDF zoom controls if available.



7. Troubleshooting

7.1 Evidence file not found

If Open Evidence reports that a file is missing, Ariadne could not resolve the original PDF path.

- Confirm the PDF still exists under files/ (same relative location as during ingestion).
- Avoid renaming/moving evidence after ingestion; if you did, reindex and keep paths stable afterwards.
- If the evidence is not a PDF, remember the evidence viewer renders PDFs only.

7.2 Geo Map shows “unresolved_or_ambiguous” frequently

- Increase Min score / Min confidence to reduce noise.
- Provide preferred countries via `ARIADNE_GEO_PREFERRED_COUNTRIES` (comma-separated) to bias ambiguous cities.
- Ensure `world_locations.json` is present and reasonably complete.
- Treat locations as hypotheses: open evidence and confirm context (e.g., “Braga” can appear in many non-location contexts).

7.3 OCR not running

- Install Tesseract OCR.
- Set `ARIADNE_TESSERACT_CMD` to the correct `tesseract.exe` path if autodetection fails.
- OCR is used only when the PDF page has no usable embedded text.

7.4 Ingestion is slow or appears stuck

- Large PDFs with OCR and face detection are expensive. Disable face search during ingestion if you only need text first.
- Prefer Reindex New/Changed Only for incremental updates.
- If you Cancel, keep “Resume interrupted build from checkpoint” enabled and re-run ingestion.

7.5 OpenAI errors / quota issues

- Verify the API key in Settings.
- Use `gpt-4.1-mini` or `gpt-5-mini` for lower cost/latency when iterating.
- Smart Search and Intelligence Report can be the highest-cost features in OpenAI mode; reduce loops/steps when needed.



7.6 Backups and resets

Main data lives in Desktop/ariadne-bd/ariadne-storage/ariadne.db. Back it up before merges/migrations or before running large ingestion passes.

To hard reset the case state, close Ariadne and move the ariadne.db file aside (do not delete without a backup). Then re-open Ariadne and rebuild the database.



Appendix A — Glossary

Entity: A node representing a person, organization, location, event, account, document, or identifier.

Mention: A text span in evidence that refers to an entity (e.g., a name occurrence in a PDF page).

Identifier: A structured token extracted from text (email, IBAN, phone, URL, domain, IPv4, NIF, hash).

Evidence page: A specific PDF page linked to a passage, mention, alert, timeline event, or face occurrence.

Relation: A link between entities (co_occurs / has_identifier / related_to).

Ontology: The formal schema describing allowed entity types and relations.

Migration: A versioned upgrade to the ontology schema.

Alert: A rule-generated item flagged for investigator review.

Appendix B — Default ontology summary

Default ontology (v1.0.0) is intentionally minimal and is designed to be extended via migrations.

Class	Description
PERSON	Person (individual).
ORG	Organization / company / collective entity.
LOCATION	Location (city, country, address, etc.).
EVENT	Event (a node when meaningful).
DOCUMENT	Document (file) when represented as a node.
ACCOUNT	Account (bank, user, etc.).
IDENTIFIER	Identifier (base) – subtypes via IDENTIFIER * prefix.

Relation	Directed	Domain	Range
co_occurs	No	PERSON, ORG, LOCATION, EVENT, ACCOUNT, IDENTIFIER	PERSON, ORG, LOCATION, EVENT, ACCOUNT, IDENTIFIER



has_identifier	Yes	PERSON, ACCOUNT, LOCATION, EVENT	ORG,	IDENTIFIER
related_to	No	PERSON, LOCATION, EVENT, ACCOUNT, IDENTIFIER	ORG,	PERSON, ORG, LOCATION, EVENT, ACCOUNT, IDENTIFIER

Appendix C — Environment variables (advanced)

Most users will not need these. They exist for controlled deployments, debugging, or data-location overrides.

Variable	Effect
ARIADNE_DESKTOP_DIR	Override the Desktop directory used for ariadne-bd/ and ariadne-output/.
ARIADNE_FACES_ROOT	Override the root directory for the face cache (faces.db, crops).
ARIADNE_GEO_CATALOG	Override path to world_locations.json used by Geo Map.
ARIADNE_GEO_PREFERRED_COUNTRIES	Comma-separated country codes/names to bias ambiguous location resolution.
ARIADNE_AI_RUNTIME_PROFILE	Select the AI runtime profile (implementation-specific).
ARIADNE_INGEST_DISABLE_AI_EXTRACT	If set, disables AI extraction during ingestion (identifiers + basic indexing only).
ARIADNE_INGEST_DISABLE_FACE	If set, disables face detection even if the UI toggle is enabled.
ARIADNE_INGEST_MAX_PAGES_PER_FILE	Cap PDF pages processed per file (useful for fast triage).
ARIADNE_TESSERACT_CMD	Full path to tesseract.exe for OCR.

End of manual.