Information Archaeology

CIO / Technical Leader Brief

Primary Author: M. Bartels

Project: InformationArchaeology.org

Version**:** 1.0 (Executive / Technical Brief)

Date: December 2025

1. The Problem Information Archaeology Addresses

Modern organizations sit on years of digital residue: shared drives, cloud folders, wikis, ticket systems, chats, exports, backups, and sync ghosts from tools that may not even be in use anymore.

When someone asks:

- *"How did we get to this design?"*
- *"Why is this process the way it is?"*
- *"What changed between version A and version B?"*

the assumption is that "the answer is in the files." Usually, it isn't.

What we see instead:

- Folders that read like strata: old drafts, copied subtrees, renamed folders, abandoned experiments.
- Artifacts mixed with system residue: autosaves, conflict copies, thumbnails, sync traces, path churn.
- Long projects where the visible documentation and the actual activity don't line up.

The data is there, but the *history* is not. We have traces without a consistent way to read them.

Most existing approaches only touch slices of this:

- Search finds files, but not how they relate.
- Archival practices preserve records, but don't explain how work evolved.
- Forensics reconstructs specific incidents, not everyday project behavior.
- AI can summarize content but will happily fill in missing context with guesses.

The result is a recurring pattern:

- When it really matters, people fall back on memory, assumption, and politics.
- Critical decisions get justified with stories, not evidence grounded in how the work actually happened.

Information Archaeology exists to change that.

2. Why "Information Archaeology"?

Archaeology works with material traces of past activity:

- It doesn't interview the people who were there.
- It doesn't assume motive.
- It studies how artifacts and layers formed over time, and where the record is broken or incomplete.

Information Archaeology applies that mindset to **digital work**.

- **Artifacts**: documents, spreadsheets, models, reports, tickets, code—anything people intentionally produce.
- **Ecofacts**: autosaves, sync markers, conflict copies, thumbnails, logs, metadata rewrites, sidecars, caches—anything systems produce as a side effect of use.

Artifacts tell you what people tried to do.

Ecofacts tell you how the environment responded and changed around that work.

IA treats both as evidence inside a **digital deposit** (a project archive, shared drive, export, or system history). It asks:

- How was this deposit formed?
- What layers and transitions can we see?
- Where did systems reshape or lose information?
- What can we responsibly reconstruct about behavior and environment—*and where does the trail actually end*?

The "archaeology" is in respect to method:

- Work backward from traces.
- Treat gaps and disturbance as part of the record, not noise to be smoothed out.
- Refuse to infer motive, psychology, or intent where there is no direct evidence.

3. Core Concepts in Plain Language

3.1 Artifacts, Ecofacts, and the Evidence Field

Every digital deposit can be treated as an **evidence field**:

- **Artifacts** – intentional work products.
- **Ecofacts** – system-generated side effects and residues.
- **Metadata** – timestamps, paths, users, devices, permissions.
- **Structure** – how items are grouped, layered, renamed, moved, and copied over time.

IA assumes that:

- Activity produces traces (artifacts + ecofacts).
- Systems and environments leave signatures in those traces.
- Time introduces drift, loss, and disturbance.

Plainly speaking, the job is to read that field without pretending it says more than it does.

3.2 ATET: How Digital Deposits Form

IA uses a simple formation model:

- **Activity** → human work: writing, editing, saving, syncing, reorganizing.
- **Trace** → the material left behind: artifacts, ecofacts, metadata, structure.
- **Environment** → the systems and tools that shaped those traces: platforms, sync engines, policies, device contexts.
- **Time** → drift, accumulation, and loss; the way traces age, change, and fragment.

IA works backwards through this chain:

1. Start from what exists in the deposit.
2. Reconstruct sequences and environments from the traces.
3. Identify drift, disturbance, and loss.
4. Stop where the evidence stops.

3.3 The Evidence Layer (Measurement vs Interpretation)

In practice, you need a layer—implemented in whatever stack you already have—that can:

- detect and normalize artifacts and ecofacts,
- extract metadata and relationships,
- record quality and confidence signals,
- preserve provenance for every transform.

This is the Evidence Layer.

- It **measures**: “this file existed here at time T, with these properties.”
- It does *not* explain or infer.

Information Archaeology sits *on top* of that.

IA methods use the measured evidence to build a reconstruction that is:

- structured,
- explicit about uncertainty,
- and constrained by provenance.

4. The Ten Methods – High-Level View

IA defines ten named methods. They're not strict steps as much as disciplined ways of looking at the evidence field**.** They tend to be used in three phases.

4.1 Mapping What Exists (Evidence-Field Methods)

These methods describe the deposit before we try to explain it.

- **Digital Stratigraphic Analysis (DSA)**

Identifies layers, boundaries, accumulation, and disturbance over time—how different "episodes" of work stack up.

- **Loss & Absence Reasoning (LAR)**

Examines missing or broken traces—deleted files, partial histories, orphaned residues—and defines where reconstruction has to stop.

- **Relational & Contextual Metadata Mapping (RCMM)**

Maps relationships among artifacts, ecofacts, metadata, and environments: which items travel together, share paths or timestamps, or co-occur in the same structural contexts.

- **Ecofact-Based Reconstruction (EBR)**

Uses ecofacts—autosaves, sync bursts, conflict copies, thumbnails, environment signatures—to detect where activity begins, ends, and collides with system constraints.

Together, these four describe the shape of the evidence field: what's there, how it is layered, how it relates, and where the obvious gaps sit.

4.2 Reconstructing Sequence, Environment, and Change

These methods explain how things unfolded and evolved.

- **Digital Sequence Reconstruction (DSR)**

Rebuilds observable sequences: edit chains, branching, merges, reversions, pause–resume cycles.

- **Digital Evolution Reconstruction (DER)**

Reconstructs structural and environmental change: schema evolution, platform transitions, environment shifts that reshaped the deposit.

- **Temporal Drift Analysis (TDA)**

Tracks long-term change: naming evolution, structural maturation, new tools entering the picture, old ones fading out.

- **Pattern-Constrained Interpretation (PCI)**

Enforces discipline across all methods:

  - distinguishes facts from supported inferences and hypotheses,
  - labels uncertainty and confidence,
  - ensures conclusions stay within reconstruction windows.

These four methods convert a pile of traces into a time-aware structure: a plausible, bounded account of what changed, when, and under which constraints.

4.3 Interpreting Bounded Meaning

Only after structure and environment are understood does IA look at **patterns of practice**.

- **Operational Composition Analysis (OCA)**

Examines how work is actually composed:

  - recurring structures in documents,
  - typical edit rhythms,
  - fragment reuse,
  - how small units of work assemble into larger outputs.

- **Digital Culture Interpretation (DCI)**

Looks across deposits to find **cultural and organizational patterns**:

  - naming conventions,
  - recurring motifs,
  - shared habits of structuring and documenting work.

Crucially:

- OCA and DCI do not infer motives, personality, or emotional state.
- They stay at the level of practice and culture as expressed in traces.

5. How the Methods Work Together

In practice, IA rarely runs "one method at a time." Methods are combined in patterns.

5.1 A Typical Analysis Flow

A common pattern looks like this:

1. **Map the field**
    a. Use **DSA**, **LAR**, **RCMM**, **EBR** to:
        i. identify layers and boundaries,
        ii. mark obvious loss/absence regions,
        iii. map basic relationships,
        iv. understand where systems have left strong ecofact halos.
2. **Reconstruct structure and change**
    a. Use **DSR**, **DER**, **TDA**, with **PCI** watching the edges, to:
        i. rebuild key sequences,
        ii. identify environment states and transitions,
        iii. locate drift zones and disturbance events,
        iv. define reconstruction windows: where the picture is stable vs where it fragments.
3. **Interpret bounded meaning**
    a. Use **OCA** and **DCI**, again under **PCI**, to:
        i. identify patterns in how work is composed,
        ii. surface cultural and organizational habits,
        iii. correlate these patterns with layers, drift zones, and environment changes.

Throughout, **PCI** forces every conclusion to be tagged with:

- what evidence supports it,
- what alternative explanations exist,
- and where we simply do not know.

5.2 Example Problem Patterns

For a CIO or technical leader, these method combinations map naturally onto problems you already have:

- **Understanding how a critical process actually runs today**
    - DSA + RCMM + EBR → what exists and how it is layered.
    - DSR + DER + TDA → how the process and environment have changed.
    - OCA → how work is composed in practice vs on paper.
- **Preparing for a major system migration or consolidation**
    - DSA + TDA → legacy layers, drift, and stabilization points.
    - LAR → where history is already broken or lost.
    - DER + RCMM → which structures and relationships must be preserved.
- **Using AI on archives without losing traceability**
    - Evidence Layer defines what is visible and how it is measured.
    - PCI + IA methods set the rules for what AI is allowed to claim.
    - OCA/DCI operate only within what the Evidence Layer and reconstruction windows actually support.

You don't need to deploy all ten methods every time. The point is to have a named, reusable toolkit that keeps analysis grounded in traces rather than stories.

6. How to Start, Practically

Information Archaeology is a framework and methodology: a **lens** and a **method set** you can start applying to the systems and archives you already have.

A realistic starting path:

1. **Adopt the vocabulary in a small circle** (architecture, data, records, risk, AI leads):
    a. artifacts vs ecofacts,
    b. evidence field,
    c. drift and disturbance,
    d. loss and absence,
    e. reconstruction windows.
2. **Pick one real deposit that matters**:
    a. a long-running project folder,
    b. the history behind a key recurring report,
    c. a legacy repository that is about to be retired.
3. **Run a light IA-style pass**:
    a. DSA + RCMM + EBR to understand layers, relationships, and ecofact halos.
    b. LAR to explicitly mark where history is broken.
    c. DSR/TDA where the evidence supports it.
4. **Use the results to inform one decision**:
    a. migration scope,
    b. documentation cleanup,
    c. where AI should and should not be trusted on this material.

7. Conclusion

Information Archaeology is not another dashboard, search tool, or AI feature. It is a way of reading what your systems already record.

The discipline gives you:

- a clear vocabulary for artifacts, ecofacts, drift, disturbance, and loss,
- a formation model (ATET) that explains how digital deposits come to look the way they do,
- and ten named methods that let you map what exists, reconstruct how it changed, and interpret bounded patterns of practice—without drifting into stories about motive or intent.

For a technical leader, the value is straightforward:

- You gain a consistent way to talk about “what actually happened” in your digital estate that does not depend on memory or narrative.
- You can ask better questions of vendors and internal teams: beyond “What data do we have?” but “What traces do we have, what can they support, and where are the gaps?”
- You get a framework that sits underneath AI, governance, migration, and audit work, instead of treating each as a separate problem.

IA doesn’t promise completeness; it promises clarity about what the evidence can and cannot say.

If you adopt the vocabulary, run it once on a real deposit, and insist on separating measurement from interpretation, you will already be doing Information Archaeology in practice. The full whitepaper exists for when you’re ready to go deeper into the methods; this brief is meant to give you enough to start using the lens in real conversations and decisions.

Authorship and Affiliation

This publication was produced independently as part of the InformationArchaeology.org project.
It is not affiliated with or endorsed by any employer or organization.

AI-assisted tools were used in the preparation of this document under human direction.