# Information Archaeology: Reading Digital Traces as Evidence

Author: M. Bartels
Project: InformationArchaeology.org
Version: 1.0 (Working Paper) Date: December 2025

## Introduction

Digital life is prolific. Documents move, versions multiply, drafts are overwritten, and background systems generate their own trails of logs, caches, autosaves, and sync residues. After enough time passes, even simple questions like why a file was created, how an idea developed, or what led to a decision all become surprisingly hard to answer.

Information Archaeology begins with that recognition. Every digital object, whether it was created intentionally or produced automatically by a system, is a small piece of a larger pattern of activity. Artifacts reflect deliberate human work. Ecofacts reflect how the environment responded to that work via autosaves, thumbnails, sync markers, conflict copies, metadata residues, and other system byproducts. If artifacts show what people attempted to do, ecofacts show the conditions and systems within they were working.

Current approaches usually treat these traces as isolated items. Search tools treat files as static objects. Backup and sync tools treat them as blobs to be moved or mirrored. Forensic tools rebuild narrow timelines but often ignore the everyday meaning of work. AI tools try to “explain” archives, but they tend to fill gaps with confident guesses rather than disciplined evidence.

Information Archaeology takes the opposite approach. It avoids inferring motives, emotions, or inner states. Instead, it treats digital deposits as stratified fields of material evidence and asks what can be responsibly reconstructed from artifacts and ecofacts as they actually exist.

This discipline is about understanding what the evidence can reliably show about:

- how activity produced traces,
- how systems and environments shaped those traces,
- how both changed over time,
- and where loss and absence now limit what can be known.

The framework used in this paper is simple:

- Activity, traces, environments, and time (ATET) describe how digital deposits form.
- An Evidence Layer measures what exists in the deposit without interpreting it.
- Ten IA methods operate on that measured evidence to map structure, reconstruct sequences and environments, and cautiously interpret patterns of practice and culture, all within provenance limits.

The goal is straightforward: make sense of digital life without guessing.

To accomplish this, artifacts and ecofacts are treated as material traces that can be read together to reconstruct what can be known—and to make clear where responsible interpretation must stop.

### 1.1 Purpose of This Document

The purpose of this document is simple: to explain what Information Archaeology is, why it matters, and how it can help us understand digital work without relying on hidden assumptions about motives or missing context. Most people who manage or interpret digital material already know the feeling of running into a wall of half-visible history. You open a folder, see a collection of drafts, exports, screenshots, and unexplained remnants, and realize the story behind them is harder to piece together than it should be. It isn't that the information is missing, as most of it is right in front of us. The problem is that we've never had a consistent way to read these traces as evidence.

Information Archaeology provides that structure. It lays out a way to understand digital behavior through the material people leave behind and through the reactions of the systems they use. Rather than treating digital files as isolated objects, the discipline looks at how they formed, how they changed, what surrounded them, and what traces were produced along the way. By doing so, we can see patterns that normally remain hidden: how work unfolded, where the environment shaped or distorted it, and what is present versus what has been lost.

This paper has three goals. First, it defines the field clearly enough that someone encountering the idea for the first time understands what it offers. Second, it explains why existing approaches like search tools, forensics, archival theory, organizational knowledge practices, and even AI handle only parts of the problem. Third, it introduces a practical framework and a set of methods that allow analysts, researchers, and organizations to interpret digital evidence in a disciplined, reproducible way.

Throughout, the focus remains on what can be observed in the artifacts and in the environment itself. The point is not to reconstruct intention or speculate motive. The point is to understand behavior through the traces it produces and to set clear limits around interpretation, so we don't drift into guesswork. This is meant to be both a conceptual introduction and a practical guide for anyone who works with digital evidence and wants a clearer, more responsible way to read it.

### 1.2 The Problem Space

Digital work leaves a lot behind during its creation, such as drafts, versions, exports, screenshots, and the system residue. None of it feels unusual at the time, but later it becomes hard to see how the work actually unfolded. The files remain; the meaning does not.

Most organizations run into this regularly. A decision needs revisiting, or a requirement needs clarifying, and the history is unclear. The traces we do have are intentional documents mixed with autosaves, sync artifacts, and other environmental byproducts, and they don't explain themselves. They show activity, but not in a way we're taught to read.

Existing approaches only cover parts of the problem. Search finds documents but not relationships. Archival practices preserve records without explaining their evolution. Forensics reconstructs timelines but not context. AI adds summaries, but it often fills in gaps with assumptions.

The issue isn't a lack of information. It's that we've never had a consistent way to understand what these traces reveal about behavior and how the environment shaped it. Without that structure, people fall back on guesswork, and guesswork erodes clarity over time.

Information Archaeology addresses this by treating both artifacts and system-generated traces as evidence: material that can be interpreted without assuming motive or intention. The problem is not the volume of digital material; it's the absence of a disciplined way to read it.

### 1.3 Limitations of Existing Frameworks

Many fields touch parts of this problem, but none address it as a whole. Archival theory preserves records and provenance yet stops short of explaining how work evolved. Library and information science focuses on access and organization, not on reconstructing behavior. Digital humanities explore meaning and culture, but not the operational realities of how work moves through systems.

Forensics can rebuild timelines and identify events, but its goal is investigation, not understanding how decisions and environments shape each other. Organizational studies examine workflows and memory but rely on documentation that often never existed. AI systems attempt interpretation but tend to fill in missing context with confident guesses, which undermines accuracy rather than improving it.

Each of these approaches is useful in its domain. What none of them offer is a consistent way to read digital traces, both intentional and system-generated, as evidence of activity. They either ignore the environmental layer, overlook the system environment, or rely too heavily on inference. This leaves a gap between what digital material contains and what we can responsibly understand from it.

Information Archaeology steps into that space. It doesn't replace these fields; it connects what they do well and fills in what they leave out. Its focus is simple: use the traces themselves to understand behavior and environment, without assuming motive or meaning that isn't supported by the evidence.

### 1.4 The Proposal

Information Archaeology offers a simple proposal: digital traces can be read the same way other fields read material evidence, and doing so gives us a reliable view of behavior without guessing at motive or intention. To make this practical, the discipline introduces a set of principles for interpreting artifacts and their environmental factors as they actually appear in modern work.

At its core, the proposal treats digital material as something that forms in layers. People create intentional artifacts. Systems generate ecofacts in response. Both layers accumulate over time, reflecting a mixture of deliberate choices, environmental constraints, and background processes. When these traces are examined together, they reveal how work unfolded, what conditions shaped it, and where information was lost or altered.

This approach does not replace existing methods from archives, forensics, or AI; it provides a structure they can fit into. From archives, we borrow the concern for preservation and provenance. From forensics, we adopt an interest in sequence and environment. From AI, we recognize the need for scale and assistive tools. But Information Archaeology adds something they all lack: a consistent framework that keeps interpretation tied to observable evidence.

The goal is to establish clear, reproducible rules for reading what is already there, rather than creating new narratives. When we remove inference and rely on the traces themselves, digital work becomes easier to understand, audit, and explain. Organizations gain a more trustworthy account of their own activity, and individuals gain a clearer picture of how tools and systems shape their work.

Information Archaeology is meant to be practical, lightweight, and adaptable. It defines how to recognize digital evidence, how to interpret it responsibly, and how to understand behavior through the traces left behind. This document outlines the principles and methods needed to do that work consistently.

## 2. Foundations of Information Archaeology

Information Archaeology is built on a small set of principles that explain how digital traces form, how they behave over time, and how they can be interpreted without relying on assumptions. These principles are not theoretical; they describe what consistently appears in real systems when people work. Each principle reflects something observable in the artifacts themselves and in the environment that produced them.

The foundations below serve as the discipline's core. They establish what can be trusted, what cannot, and where interpretation must stop.

### 2.1 Archaeology (Material Culture)

Archaeology studies human activity by reading the material record: artifacts, features, deposits, and the layers that hold them.

Its influence on IA is direct:

- Stratigraphy — sequences embedded in layers
- Context — meaning derived from relational placement
- Taphonomy — how material changes, decays, and disappears
- Multi-hypothesis reasoning — interpretation under constraint
- Material logic — artifacts as products of practice, not imagination

Information Archaeology adapts these principles to digital environments. The "layers" are structural rather than physical. The "taphonomy" is mechanistic, not geological. The deposits are shaped by systems as much as by people.

The term "archaeology" is metaphorical and respectful; a way to highlight the layered, evidentiary nature of digital life, without claiming equivalence to excavation.

### 2.2 Behavior Leaves Material Traces

Every meaningful digital action produces evidence, deliberately or at times accidentally. Drafts, renames, exports, and screenshots are intentional traces. Autosaves, sync fragments, and conflict copies are environmental traces. When taken together, these traces describe activity more reliably than memory, narrative, or speculation.

The principle is simple: if behavior occurred, it left material behind. If the material is absent, something environmental explains the gap.

### 2.3 Environment Is an Active Participant

Digital environments do not sit idle. Cloud sync engines, autosave processes, indexing services, version managers, and background daemons all interact with human behavior and leave their own evidence. These ecofacts are not noise. They show timing, friction, interruption, drift, loss, and recovery. Understanding behavior requires understanding the environment that shaped it.

### 2.4 Artifacts and Ecofacts Form Layers Over Time

Digital history does not emerge in a single line. It forms in layers of intentional work layered with environmental reactions. Some layers accumulate; others erode, and what remains is the composite. Recognizing these layers allows us to separate deliberate activity from system-generated byproducts and understand how each influenced the other. This layered view prevents overinterpretation and reveals the actual sequence of events.

### 2.5 Interpretation Must Come From Evidence, Not Assumptions

Digital traces support many possible explanations, but only a few can be confirmed. Information Archaeology emphasizes what can be known without inferring motive, emotion, or intention. The focus stays on behavior: what people did, how systems responded, and what is materially present.

When evidence is ambiguous, the discipline preserves multiple possibilities instead of collapsing them into a single guess.

### 2.6 Loss Is Meaningful Evidence

Missing artifacts are not simply “gone.” They leave loss shadows and drift patterns that show how the environment handled them. Deletions, overwrites, corruption, and sync conflicts all produce interpretable residue. These patterns help reconstruct what happened without needing to know why it happened. Loss is treated as a form of data, not an absence.

### 2.7 Consistency Comes From Method, Not Interpretation

Reliable reconstruction depends on consistent methods, not on the individual analyst’s intuition. Information Archaeology defines steps, boundaries, and evidence classes so different people examining the same traces arrive at the same conclusions. This principle aligns IA with scientific and archival traditions: repeatability is essential.

**2.8 The System Should Not Guess Beyond Its Evidence**

AI tools can help summarize or classify traces, but they must not invent missing context. Information Archaeology provides guardrails to keep AI interpreters grounded in what the material can actually support. This ensures clarity, prevents narrative drift, and maintains trustworthiness. AI can accelerate analysis, but the evidence sets the limits.

**2.9 The Goal Is To Understand Activity Without Assigning Intention**

Information Archaeology is not psychology, forensics, or investigation into motives. Its purpose is to understand how work unfolded through the traces it left behind. Behavior is observable. Intention is not. This discipline stays on the safe side of that boundary.

**2.10 The Evidence Layer**

Information Archaeology begins with what software can actually see.Before any method runs, the deposit is scanned by an Evidence Layer: a set of tools or processes that detect artifacts, ecofacts, and metadata and describe them in a consistent way. The Evidence Layer does not explain behavior or culture. It measures.

For IA, the Evidence Layer must provide three things:

- identifiable objects and residues (artifacts, ecofacts, metadata),
- structured outputs about them (Collection Facts and Generated Facts),
- and preserved provenance and confidence for those outputs.

Collection Facts are direct observations harvested from systems:

- files, folders, versions, timestamps, and permissions,
- logs and event records,
- metadata fields as stored by applications,
- thumbnails, previews, caches, autosaves, sync markers, and other ecofacts.

Generated Facts are procedural measurements derived from those observations:

- counts and frequencies (edits per day, versions per file),
- intervals (time between saves, syncs, or migrations),
- basic classifications (filetype groupings, path segments),

- graph structures (who-edited-what, co-occurrence and adjacency),
- detected environment signatures and simple confidence indicators.

Both types of facts are measurements, not interpretations. They describe what exists and how it appears in the systems that produced it. They do not say what the behavior "meant," or why it occurred.

The Evidence Layer must also:

- attach confidence levels to its own outputs, and
- record provenance for each fact (which system, which log, which scan, under what conditions).

IA methods then take these measured facts and:

- map them into an evidence field (DSA, LAR, EBR),
- reconstruct sequences, environments, and drift within reconstruction windows (DSR, DER, TDA, PCI),
- and, finally, interpret bounded patterns of practice and culture (OCA, DCI).

Where software or tooling measures, IA provides bounded interpretation.

## 3. Core Concepts of Information Archaeology

Information Archaeology works because digital systems create consistent, recognizable patterns. These patterns show up in the material itself: the files people produce and the traces systems generate around them. The concepts in this section define the categories and structures used to interpret those patterns without speculation. Together, they form the shared vocabulary of the discipline.

### 3.1 Artifacts

Artifacts are the intentional outputs of human work. Documents, drafts, images, notebooks, exports, and any file someone deliberately created or edited fall into this category. Artifacts show purposeful activity, but not always the full path that led to the final version. Their relationships, differences, and position in time help reveal how work evolved.

Artifacts answer the question: What did the person try to produce?

### 3.2 Ecofacts

Ecofacts are system-generated traces: material created not by the person, but by the environment responding to activity. Autosaves, thumbnails, sync logs, recovery files, temporary versions, metadata residues, and conflict copies are all ecofacts. They reflect timing, friction, interruptions, resource limits, and background processes. Unlike artifacts, ecofacts always hold two layers of meaning at once.

Information Archaeology treats ecofacts as dual-boundary objects:

The inner boundary records the event that produced the ecofact. The outer boundary records how the environment responded.

Ecofacts answer the question: How did the environment react to what was happening?

This distinction of behavior vs. environment is foundational. Ecofacts often reveal the parts of activity that artifacts omit.

### 3.3 Layers of Formation

Digital history forms in layers. Intentional actions create artifacts. Environmental processes create ecofacts. Over time, these layers accumulate unevenly. Some layers record change in detail; others degrade or collapse as systems overwrite, compress, or clean up older traces.

Reading layers means looking beyond individual files to see how the full set formed together. Layer analysis clarifies sequence, dependency, and where the environment reshaped or constrained behavior.

### 3.4 Boundaries

Digital traces do not sit in isolation. Every event leaves an imprint, and systems respond around it. Information Archaeology models this with a dual-boundary structure: an inner boundary around the event itself, and an outer boundary around the environment's response.

These are interpretive constructs, not extra traces. They are ways of grouping what already exists in the evidence field so that behavior and environment can be separated, compared, and kept in scope.

Inner Boundary: The Event Imprint

The inner boundary is created at the moment of activity. It is the tightest zone around what a person or team actually did.

The inner boundary includes traces that directly express the event:

- the timestamp delta between edits
- the autosave interval or write rhythm
- the partial or incomplete fragment that marks an interruption
- the encoding details of the tool that produced it
- the structure of the residue itself (preview, cache, sidecar, delta)
- the minimal set of ecofacts that are inseparable from the act (for example, the autosave that is the only surviving record of a draft)

The inner boundary is not interpretive about motive or intention. It is literal: "this is what the event produced or changed." Its job is to outline where behavior remains reconstructable from the material in front of us.

Outer Boundary: The Environment Response Zone

Around every event, systems react. Files are copied, synced, indexed, backed up, scanned, and transformed. The outer boundary captures this environmental response: the halo of system behavior that surrounds, follows, or propagates the event.

The outer boundary includes traces that show how the environment handled the event:

- metadata rewriting during sync or cloud propagation
- permission changes and access-control adjustments
- modified vs. created timestamp drift
- format conversions and surrogate copies

- renaming rules or storage constraints
- background tasks such as indexing, thumbnailing, or cache updates
- system-level adjustments triggered by the artifact or event

Where the inner boundary says, “here is what the event created,” the outer boundary says, “here is what the environment did with it.”

Ecofacts often live primarily in the outer boundary, but they are anchored to inner-boundary events. This is what allows IA to treat ecofacts as evidence of both activity and environment without collapsing the two.

Dual Boundary Limits

The dual-boundary model is about behavior and environment around events. It is not about:

- deposit scope (what material is in or out of the case), or
- temporal markers between working episodes.

Those are handled elsewhere: PCI defines deposit boundaries for a case, and DSA uses temporal markers to separate working episodes. Keeping these concepts distinct prevents “boundary” from doing too many jobs at once.

Method Alignment (see section 4: Methods of Information Archaeology for definitions)

The boundaries guide how IA methods use the evidence field.

- Inner boundaries inform DSR, DSA, OCA, and LAR:
  - DSA and DSR use them to sequence and cluster events.
  - OCA uses them to analyze internal composition within a bounded act.
  - LAR uses them to identify where loss interrupts a sequence of events.
- Outer boundaries inform DER, TDA, RCMM, and PCI:
  - DER and TDA use them to trace how structures evolve over time and tools.
  - RCMM uses them to map relationships created or reshaped by system behavior.
  - PCI uses them to enforce where interpretation must stop when environment signals become too weak or ambiguous.

Together, inner and outer boundaries turn ecofacts and artifacts into a coherent, bounded evidence field. They mark where behavior can still be reconstructed, where environment can still be read, and where uncertainty begins.

### 3.5 Disturbance

Disturbance refers to sudden, disruptive events that break the continuity of the digital record. These are the moments where stratigraphy is interrupted: files are moved in bulk, large batches are renamed, repositories are migrated in a single step, or systems crash and recover with gaps.

Disturbance events have three common features:

- They happen over a short period of time relative to the span of the deposit.
- They produce visible discontinuities in the traces (missing versions, broken chains, abrupt jumps).
- They often involve system-level actions: migrations, bulk operations, deletion waves, or acute sync conflicts.

Examples include:

- A one-time platform migration that moves or restructures an entire folder tree.
- A bulk export and purge operation that removes working drafts but leaves thumbnails or surrogate copies.
- A sync conflict storm after a connectivity issue, creating conflict copies and fractured timelines.
- A crash that truncates logs or corrupts part of a document series.

Disturbance does not describe normal tool evolution or slow change. It captures the acute events that put “cracks” into the record. IA treats those cracks as part of the evidence: disturbance marks where reconstruction will be limited or must branch into alternate possibilities.

### 3.6 Drift

Drift describes gradual, evolutionary shifts that alter patterns in the deposit while preserving continuity. Where disturbance is abrupt, drift is slow. It shows up as changing habits, toolchains, or system behaviors that accumulate over time.

Drift has three main characteristics:

- It emerges across many traces rather than at a single point in time.
- It changes patterns (naming, structure, metadata, locations) without creating obvious gaps.
- It is often driven by environment changes that are adopted and normalized by the person or team.

Examples include:

- A progressive renaming style that shifts from “final_v1.docx” to date-stamped or semantic names.
- Gradual adoption of a new tool, where files of a given type slowly move from one application or platform to another.
- A sync engine update that begins adding new metadata fields, changing path conventions, or altering autosave behavior over months.
- A long-term shift from local folders to cloud workspaces, where old material is left in place and new material accrues elsewhere.

Some phenomena, like platform migrations, can involve both disturbance and drift: a one-day cutover may create discontinuities (disturbance), while the months that follow show new, stable patterns (drift). IA methods treat these as compound phenomena and account for both aspects where needed.

Drift is important because it sets the background conditions for interpretation. It tells us what “normal” looks like for a given span of time, so that departures from that pattern can be recognized and treated with appropriate caution.

**3.7 Loss, Absence, and Negative Evidence**

IA distinguishes three related concepts: loss, absence, and negative evidence. They are not the same thing, and treating them as interchangeable leads to overconfidence and speculation.

Loss

Loss is material that once existed but has disappeared. It leaves “shadows” in the evidence field: broken chains, references to missing items, orphaned ecofacts, or structural gaps. Loss typically reflects deletion, overwrite, corruption, failed migrations, or expired retention.

Examples of loss include:

- A document that other files or logs refer to, but whose file is no longer present.
- A thumbnail or preview image that remains after the source file has been removed.
- Log entries that describe operations on a folder that no longer exists.
- Version histories where middle versions are missing or the sequence is truncated.

A loss shadow is the pattern left behind when this happens: the surviving traces that show something used to be there. Different causes of loss (manual deletion, automated cleanup, policy-driven retention, migration failures) create different shadows. These shadows are part of the record. They show where material once existed and where reconstruction is now limited.

Loss tells us: “something was here and is now gone.”

Absence

Absence is different. Absence is the non-appearance of material that might have been expected but was never recorded in the deposit at all. There is no shadow, only a missing pattern.

Examples of absence include:

- A workflow that is usually documented, but for one request there is no sign any documentation was ever created.
- A system that never logs certain actions by design, even though those actions are common.
- A person who routinely works off-system (paper, whiteboards, verbal agreement) for particular tasks.

Absence does not prove that something did not occur. It tells us only that, given what we know about the tools, habits, and environment, there is no trace that it did. Some absences are meaningful; others simply reflect how work was actually done.

Absence tells us: “there is no trace here, but that might be normal.”

Negative Evidence

Negative evidence is how IA uses loss and absence to define what cannot be claimed. It is not a separate class of trace; it is a disciplined way of saying “the record does not support this conclusion.”

Negative evidence is used to:

- Mark where reconstruction must stop because critical material is lost.
- Identify where expected patterns are missing and require explanation.
- Constrain confidence bands: what can be asserted as likely, plausible, unknown, or unsupported.

Loss & Absence Reasoning (LAR) operates in this space. LAR does not imagine content for lost or absent material and does not assign motives for deletion or non-creation. Instead, it:

- identifies loss shadows and absence patterns,
- interprets what they imply about the limits of the record, and

- defines the uncertainty regions around any reconstruction that crosses them.

Loss, absence, and negative evidence together give IA a vocabulary for talking about what is missing without positing beyond what than the traces can justify.

**3.8 Reconstruction Windows**

A reconstruction window is the period during which sufficient evidence exists to reconstruct behavior with defensible confidence. Within a window, the density and quality of traces support structured accounts of what happened. Outside of it, interpretation becomes speculative and must be limited.

Windows open when:

- Multiple trace types converge (artifacts, ecofacts, and metadata are present together).
- Temporal relationships among traces are clear enough to establish sequences.
- Evidence density exceeds a minimum threshold (few large gaps, sufficient examples).
- Confidence in structure and ordering reaches a level that can be defended and checked.

Windows close when:

- Large temporal gaps appear and evidence density falls below that threshold.
- Drift or disturbance breaks continuity so severely that sequences cannot be reliably linked.
- Loss removes critical traces that would be required to complete the account.
- Confidence in the reconstruction drops below what IA treats as defensible.

Window characteristics:

- Some windows are narrow: minutes or hours of intensive editing where every save is present.
- Others span months or years: long, stable phases of a project with consistent patterns.
- Windows can be partial: certain aspects of behavior remain reconstructable while others do not.
- Windows are method-specific: for example, DSR may have a valid window where DER does not.

In practice, analysts identify reconstruction windows during early mapping work (DSA, LAR, EBR, RCMM) and then use them to govern where higher-level methods (DSR, DER, TDA, OCA, DCI) may operate. PCI enforces these boundaries: interpretation must stop where windows close.

**3.9 Environment Signatures**

Every tool and system produces characteristic patterns. These patterns, or environment signatures, are how we recognize what the environment interacted around a trace.

Environment signatures live primarily in the outer boundary: the response zone where systems sync, copy, rewrite, and manage artifacts and ecofacts. They anchor interpretation in technical reality instead of guesswork.

*What Environment Signatures Look Like*

Environment signatures include:

- File and folder conventions:
  - standardized sync paths
  - application-specific folder hierarchies
  - naming schemes for conflict copies or backups
- Timestamp behavior:
  - consistent gaps between created and modified times
  - systematic time zone offsets
  - characteristic “touch” patterns from indexing or scanning
- Metadata patterns:
  - fields that are always added, rewritten, or stripped by a given tool
  - characteristic user-agent strings or process names in logs
  - permission or ACL patterns applied by a platform
- Storage and transformation habits:
  - how and when thumbnails, previews, or surrogates are created
  - typical autosave intervals and temporary file naming
  - compression, conversion, or export formats used in a workflow

These signatures can be extremely stable. Once learned, they allow analysts and tools to identify which traces originate from behavior (inner boundary) and which originate from environment response (outer boundary).

*Why Environment Signatures Matter*

Environment signatures matter for three reasons:

1. Separation of behavior and environment

Without signatures, it is easy to misread environmental noise as deliberate action — for example, treating an autosave rename or indexer touch as a user decision. Signatures keep these apart.

2. Drift and state detection

Changes in environment signatures reveal drift and disturbance in the environment itself:

a. a sync engine update that starts using new conflict naming,
b. a migration that introduces new path roots or timestamp patterns,
c. a retention policy change that alters which traces survive.

These changes define environment phases that TDA and DER rely on.

3. Provenance and tool attribution

Signatures help tie ecofacts and artifacts to the tools and systems that created or modified them:

a. which cloud service handled a file,
b. which application last saved it,
c. which process did the indexing, scanning, or compression.

This is essential for PCI: we need to know which environment we are actually interpreting.

*Method Alignment*

Environment signatures are read and used across multiple IA methods:

- DSA uses signatures to distinguish environment-driven reorganization from genuine shifts in working practice.
- DER and RCMM use them to follow how structures and relationships move across tools and platforms.
- TDA uses them to define environment phases and drift zones: when a signature holds steady, when it changes, and what that means for timelines.
- LAR uses them to identify loss patterns that are clearly environmental (policy-driven cleanup, archival sweeps) versus those that may be human-driven.
- PCI uses signatures as a check: if a proposed interpretation contradicts what the environment actually does, it fails.

Environment signatures do not tell us why a person made a choice. They tell us how the environment behaves and how it has changed. Together with boundaries, loss shadows, and reconstruction windows, they keep IA grounded in what systems really do to digital traces.

## 4. Methods of Information Archaeology

The methods in this section form the practical toolkit of the discipline. They describe how to read digital traces as evidence without guessing what cannot be known.

Each method has four elements:

- a clear purpose,
- specific evidence it requires,
- what it can reliably reveal, and
- where interpretation must stop.

Before any method runs, software has already done its work. The Evidence Layer scans the deposit, identifies artifacts and ecofacts, measures relationships, and produces two kinds of outputs:

- Collection Facts: direct observations (files, versions, paths, timestamps, metadata, logs, previews, caches, sidecar files).
- Generated Facts: derived measurements (counts, intervals, frequencies, relationship graphs, confidence indicators).

Methods do not create new facts. They work on this measured material, map it into an evidence field, and then build structured accounts within that field.

Information Archaeology groups its ten methods into three functional sets.

*Evidence-field methods*

These methods operate first on Evidence Layer outputs to map what exists and how it is organized. They define the evidence field and its structure:

- DSA (Digital Stratigraphic Analysis) identifies layers, strata, and episodes in the deposit.
- LAR (Loss & Absence Reasoning) identifies loss shadows, absence patterns, and the limits of the record.
- RCMM (Relational & Contextual Metadata Mapping) charts how artifacts, ecofacts, metadata, and residues relate to one another.
- EBR (Ecofact-Based Reconstruction) interprets how systems responded around events, using environment signatures and halos.

Together, these methods turn raw measurements into a structured view of the deposit: what is present, how it is layered, where loss and absence occur, and how the environment has behaved over time.

*Structural reconstruction methods*

These methods operate within reconstruction windows to reconstruct behavior and structure from the mapped evidence field:

- DSR (Digital Sequence Reconstruction) rebuilds sequences of actions and events.
- DER (Digital Evolution Reconstruction) traces how structures, objects, and configurations changed over time.
- TDA (Temporal Drift Analysis) characterizes gradual shifts in patterns, tools, and environment phases.
- PCI (Provenance-Constrained Interpretation) sits across all methods, enforcing provenance limits, confidence bands, and ethical constraints.

These methods do not explain motives or intentions. They answer narrower questions: what happened, in what order, under which environmental conditions, and with what surviving evidence.

*Insight methods*

The final group of methods extracts insight while staying inside provenance boundaries:

- OCA (Operational Composition Analysis) examines the internal composition of artifacts: structure, vocabulary, references, and fragment reuse, in order to understand how work and focus evolved.
- DCI (Digital Culture Interpretation) looks across many deposits and environments to identify recurring patterns of practice and culture, still constrained by what the evidence can support.

Insight methods are only applied after the evidence field has been mapped and structural reconstruction has been completed. They can refine understanding, but they cannot contradict what earlier methods have already established.

*Using the methods together*

In practice, IA does not prescribe a single linear recipe. Analysts select and combine methods based on the question, the deposit, and the available evidence. A typical flow looks like this:

1. Evidence Layer software measures the deposit (Collection Facts and Generated Facts).
2. Evidence-field methods (DSA, LAR, RCMM, EBR) map the evidence field and identify reconstruction windows.
3. Structural methods (DSR, DER, TDA, guided by PCI) reconstruct behavior and evolution within those windows.

4. Insight methods (OCA, DCI, under PCI) draw higher-level conclusions, always within provenance limits.

The rest of this section describes each method in more detail: its purpose, the evidence it needs, what it reveals, and the specific boundaries it must not cross.

**4.1 Digital Stratigraphic Analysis (DSA)**

*Purpose*

To identify and describe the layered structure of a digital deposit, including accumulation, disturbance, and transformation over time.

*What It Reveals*

- Activity phases and their order
- Markers or transitions between working episodes
- Disturbance or reorganization zones
- Accretion processes
- Stability vs. instability within the deposit

*Evidence Required*

DSA relies on evidence layer outputs that provide:

- temporal associations among traces
- continuity or interruption indicators
- ecofact residues surrounding activity layers
- metadata stability or instability
- signals of disturbance, overwrite, or migration

*How Evidence Is Used*

DSA interprets stratigraphic relationships: what lies “before,” “after,” “within,” or “upon” by using procedural indicators supplied by the evidence layer.

It determines:

- where layers begin and end
- how they relate

- whether transitions between layers or episodes are sharp, gradual, or disrupted

DSA also contributes to identifying reconstruction windows by marking where stratigraphy is dense and coherent versus fractured or incomplete.

*Interpretive Limits*

DSA does not infer motives, decisions, or psychological states.

It describes structure, not intention.

*Uncertainty & Constraints*

When temporal or structural evidence is ambiguous, DSA marks indeterminacy rather than smoothing layers or imposing artificial order.

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

*Purpose*

To interpret the evidential meaning and limits of missing, incomplete, or destroyed material within the deposit, including both loss and absence.

*What It Reveals*

- Loss shadows
- Absence patterns
- Gaps in sequences
- Zones where overwrite or removal has altered the record
- Missing contextual links
- Effects of system or policy-driven deletion or cleanup (without inferring motives)

*Evidence Required*

LAR requires:

- loss shadows and absence indicators
- orphaned ecofacts
- broken or dangling relationships
- incomplete strata or sequences

- disruption residues

*How Evidence Is Used*

LAR treats loss and absence as evidence about the limits of reconstruction. It:

- identifies where loss has occurred and what shadows remain
- recognizes where expected traces are absent and when that absence is meaningful
- defines interpretive limits and uncertainty regions around gaps

LAR supplies negative evidence: it shows where the record cannot support stronger claims and where reconstruction windows close.

*Interpretive Limits*

LAR does not speculate about motives for deletion or non-creation. It does not invent missing content or assign reasons for loss.

*Uncertainty & Constraints*

LAR defines where interpretation must stop. Zones of missing or absent evidence cannot be filled by inference; they are recorded as constraints that other methods must respect.

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

*Purpose*

To map and synthesize relational and contextual structures in the evidence field, identifying convergences and divergences across methods.

*What It Reveals*

- Areas of high interpretive agreement
- Points of conflict or uncertainty
- Structural coherence across layers
- Gaps where evidence prohibits reconstruction

*Evidence Required*

RCMM synthesizes:

- stratigraphic layer transitions
- sequence structures
- environment states
- drift phases
- event boundary detections (inner and outer)
- loss shadows
- operational patterns
- conceptual signals

*How Evidence Is Used*

RCMM builds a multi-layered model of the deposit, reinforcing where methods agree and preserving ambiguity where they diverge. It arranges relationships among artifacts, ecofacts, metadata, and residues into relational clusters and context maps that other methods can refer to.

*Interpretive Limits*

RCMM cannot override uncertainty; it must preserve it. It cannot collapse multiple plausible structures into a single, definitive account when the evidence does not support that move.

*Uncertainty & Constraints*

Where methods produce irreconcilable structures, RCMM documents the incompatibility rather than prioritizing one interpretation. Conflicts are recorded as part of the evidence field, not resolved by preference.

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

*Purpose*

To detect where activity episodes begin and end in the record, and how systems responded around those inner and outer boundaries.

*What It Reveals*

- Inner boundaries (human-initiated activity, as reflected in immediate ecofact and trace patterns)
- Outer boundaries (system response zones around those events)
- Halos, bursts, or sync layers
- Boundary disruptions or irregularities

*Evidence Required*

EBR requires:

- ecofact timing and alignment patterns
- system-response residues (sync marks, thumbnails, logs, caches, index traces)
- environment signatures that identify specific tools and platforms
- boundary-confirming or boundary-conflicting indicators

*How Evidence Is Used*

EBR interprets ecofact sequences and environment signatures to identify inner and outer boundary states:

- clean boundaries
- fractured boundaries
- layered boundaries
- composite or overwritten boundaries

It refines or challenges episode markers from DSA and supports TDA and DER by showing how environment phases and responses changed around events.

*Interpretive Limits*

EBR does not infer why boundaries formed or why activity paused. It explains environment behavior and boundary states, not human intention.

*Uncertainty & Constraints*

When boundary signals conflict, EBR presents multiple plausible boundary ranges rather than collapsing them. Where ecofacts and signatures are too weak or inconsistent to support a boundary, EBR marks the area as uncertain instead of forcing a decision.

**4.5 Digital Sequence Reconstruction (DSR)**

*Purpose*

To reconstruct the defensible order of versions, drafts, or iterative variants within the deposit.

*What It Reveals*

- Version chains
- Branching and merging
- Reversions or resets
- Major vs. minor change intervals
- Pauses, resumptions, and working bursts

*Evidence Required*

DSR requires:

- observable version linkages (explicit or implicit)
- change indicators between artifacts
- temporal relationships
- structural or content differences
- ecofact timing around version events

*How Evidence Is Used*

DSR assembles sequences based on how traces accumulate, diverge, or overwrite one another. It identifies defensible ordering without assuming purpose or narrative shape.

*Interpretive Limits*

DSR does not explain why changes occurred. It does not infer strategy, intention, or authorship decisions.

*Uncertainty & Constraints*

Ambiguous or conflicting sequence indicators must be preserved as branching possibilities, rather than forced into linear order.

**4.6 Digital Evolution Reconstruction (DER)**

*Purpose*

To reconstruct how the technical, operational, and environmental conditions evolved around activity.

*What It Reveals*

- Tools and platforms in use
- Permission and access structures
- Sync, versioning, and migration behaviors
- Metadata transformation events
- Organizational or procedural constraints

*Evidence Required*

DER depends on:

- environment signatures (e.g., sync conflicts, conversion traces)
- metadata rewrites
- platform residues
- ecofact halos showing system activity
- path or location histories when present

*How Evidence Is Used*

DER reads the environment through the traces:

- What systems touched the deposit?
- When did transitions occur?
- What constraints or policies shaped behavior?
- How did the environment respond to activity?

*Interpretive Limits*

DER reconstructs conditions, not decisions. It does not infer motivation, preference, expertise, or user intent.

*Uncertainty & Constraints*

Where environmental traces conflict, DER documents competing possibilities rather than selecting a preferred narrative.

**4.7 Temporal Drift Analysis (TDA)**

*Purpose*

To identify, describe, and contextualize long-term changes in structure, naming, environment, or working patterns.

*What It Reveals*

- Naming evolution
- Structure or format changes
- Working phases
- Tool or platform transitions
- Stability and instability patterns across time

*Evidence Required*

TDA requires:

- indicators of change over time
- evidence of evolving residue patterns
- environment-transition markers
- shifting relationships among artifacts and ecofacts

*How Evidence Is Used*

TDA identifies where drift occurs, how strongly, and across what span. It contextualizes activity within broader technical, organizational, or procedural evolution.

*Interpretive Limits*

TDA does not attribute shifts to personal intention, stress, emotion, or cognitive state.

*Uncertainty & Constraints*

When drift signals are weak or conflicting, TDA marks the boundaries of knowability rather than over-interpreting change.

**4.8 Provenance-Constrained Interpretation (PCI)**

*Purpose*

To enforce provenance, evidential limits, and uncertainty across all methods, and to provide a final validation pass that distinguishes fact, supported inference, hypothesis, and speculation.

PCI does not introduce new findings. It governs how findings from other methods may be combined, stated, and qualified.

*What It Reveals*

- Which claims are directly supported by traces
- Which claims are supported inferences that rely on multiple methods
- Which ideas are hypotheses or speculative possibilities only
- Where reconstruction windows open and close
- Where loss, absence, or low-confidence measurements prevent further interpretation

*Evidence Required*

PCI relies on:

- Evidence Layer outputs and their confidence indicators
- Provenance information for all artifacts, ecofacts, and metadata
- Outputs from DSA, LAR, RCMM, EBR, DSR, DER, TDA, OCA, and DCI
- Explicit markings of loss, absence, drift zones, disturbance patterns, and reconstruction windows

PCI assumes that other methods have already done their work. It reviews their outputs and the underlying evidence field, rather than re-running their analyses.

*How Evidence Is Used*

PCI has a dual role:

1. Cross-cutting constraint during analysis
    a. Limits each method to operating within established reconstruction windows.
    b. Prevents methods from using evidence outside their defined scope.
    c. Flags when a method starts to drift toward psychological or organizational intent without support.
2. Final validation and labeling
    a. Classifies every substantive statement as:
        i. **Fact** – directly supported by high-confidence evidence.
        ii. **Supported inference** – grounded in multiple methods and consistent with the evidence field.
        iii. **Hypothesis** – plausible but not fully supported; clearly marked as such.

   iv. **Speculative possibility** – explicitly labeled and treated as outside IA's core conclusions.

b. Ensures that uncertainty, confidence levels, and loss/absence regions are explicitly documented.
c. Checks that no claim contradicts stratigraphy, sequence, environment behavior, or provenance as established by earlier methods.

PCI is where IA decides what can safely be said and to what degree of confidence.

*Interpretive Limits*

PCI cannot:

- create new facts,
- override the outputs of other methods without evidential reason,
- erase uncertainty or loss,
- promote hypotheses to facts for narrative convenience.

When methods disagree, PCI records the conflict and frames conclusions accordingly. It does not select a winner without additional evidence.

*Uncertainty & Constraints*

PCI is responsible for:

- maintaining clear confidence bands around all reconstructions,
- tying each claim back to specific evidence and methods,
- marking where reconstruction windows end and interpretation must stop,
- preserving ambiguity where evidence genuinely does not resolve between alternatives.

PCI is IA's last line of defense against overinterpretation. It keeps all ten methods aligned with ATET, the Evidence Layer, and IA's ethical commitments.

**4.9 Operational Composition Analysis (OCA)**

*Purpose*

To understand how work is composed and evolves over time by examining internal structure, repeated fragments, and observable changes within artifacts, anchored to their stratigraphic and relational context.

*What It Reveals*

- Internal working rhythms and cycles that are visible within artifacts and their versions
- Composition phases (how drafts accumulate, reorganize, and stabilize)
- Reuse and rearrangement of fragments, sections, and motifs
- Relationship between internal structure, time, and environment phases
- Workflow stability or fragmentation as reflected in composition patterns

*Evidence Required*

OCA relies on:

- internal structural elements (sections, headings, code blocks, lists, formatting cues)
- repeated fragments or motifs within and across versions
- version-to-version diffs and change patterns within reconstruction windows
- ecofact density patterns that mark editing bursts and pauses
- environment-transition markers that align with shifts in composition
- structural regularities or irregularities inside the artifacts themselves

*How Evidence Is Used*

OCA interprets compositional context by examining how internal structure, fragment reuse, and vocabulary change over time, within the boundaries set by DSA, DSR, DER, RCMM, and EBR.

OCA:

- works only inside reconstruction windows established by earlier methods
- uses internal composition to identify phases of drafting, revision, refactoring, or reorganization
- distinguishes stable structures from experimental or transient ones

OCA identifies patterns of practice at the level of composition without attributing personal motives or psychological states.

*Interpretive Limits*

OCA does not infer identity, cognition, emotional states, or strategic decision-making.

It cannot contradict stratigraphy, sequence, or environment behavior already established by other methods; it must remain consistent with the evidence field and reconstruction windows.

*Uncertainty & Constraints*

When compositional rhythms cannot be distinguished from noise, or when changes are too sparse or inconsistent, OCA marks them as ambiguous rather than constructing unwarranted patterns.

Where internal structure does not support a clear phase or motif, OCA records limited signal instead of forcing a narrative.

**4.10 Digital Culture Interpretation (DCI)**

*Purpose*

To identify conceptual patterns, recurring motifs, and cultural or organizational signals that emerge across artifacts, deposits, and environments, while remaining constrained by provenance and reconstruction windows.

*What It Reveals*

- Vocabulary structures that recur across many artifacts or projects
- Recurring conceptual fragments and motifs
- Themes that persist, evolve, or disappear over time
- Cultural or organizational conventions in how work is documented, named, or structured
- Typical ways that tools and environments are used to express certain kinds of work

*Evidence Required*

DCI depends on:

- repeated terms or structures across multiple artifacts and deposits
- motif or fragment patterns identified by OCA and related methods
- stable or shifting conceptual groupings over time and environment phases
- relationships among content elements, workflows, and environments
- cross-case comparisons that show recurring practices rather than one-off events

*How Evidence Is Used*

DCI identifies cultural and organizational signals that emerge through repeated forms of work, documentation, or structuring. It:

- aggregates patterns surfaced by OCA, DSR, DER, TDA, RCMM, and EBR

- looks for conventions that appear across many reconstruction windows and environments
- describes how a team, organization, or community tends to work, document, and structure activity, based only on what the traces support

DCI remains strictly within the evidence, not the author. It characterizes practices and patterns, not personalities or inner beliefs.

*Interpretive Limits*

DCI never attributes psychological traits, personal beliefs, or emotional states.

It does not speculate about organizational intent, values, or culture beyond what can be grounded in repeated, observable patterns. It cannot override provenance constraints, reconstruction windows, or uncertainty bands defined by PCI.

**Uncertainty & Constraints**

When conceptual patterns are sparse, inconsistent, or local to a single window or case, DCI reports limited or weak signal rather than imposing thematic interpretation.

DCI preserves ambiguity where patterns could reasonably be read in multiple ways and resists generalizing from isolated examples.

**4.11 Validation (applies to all methods)**

- Internal consistency (stratigraphy, metadata, ecofacts)
- Cross-method triangulation
- External corroboration when available
- Explicit falsification criteria
- Uncertainty labeling and confidence bands (enforced through PCI)

This keeps all interpretations disciplined, reproducible, and defensible.

## 5. Applying the Methods

Information Archaeology becomes meaningful when its methods are applied together. Each method answers a specific question about how work formed, how the environment responded, and how these patterns changed over time.

The discipline follows a simple chain:

Activity → Trace → Environment → Time

(ATET)

- Activity produces traces.
- Those traces form within an environment that shapes and alters them.
- All of this unfolds across time, where drift, disturbance, and loss accumulate.

The methods turn these four dimensions into a structured, defensible way of understanding digital material without inferring motive, psychology, or intention. They reveal only what happened, not why it happened.

### 5.1 Establishing What Exists and How It Formed

Before anything can be reconstructed, IA begins by determining what exists in the deposit and the conditions under which it formed. This first stage uses Evidence Layer outputs (Collection Facts and Generated Facts) and four foundational methods to map the evidence field:

- what material is present,
- how it is layered,
- where it has been disturbed or lost, and
- how the environment has responded over time.

*Digital Stratigraphic Analysis (DSA)*

DSA provides the initial structure. It reads the layers of a deposit using temporal and structural signals from the Evidence Layer: timestamps, continuity and interruption indicators, ecofact residues, and metadata stability or instability.

DSA identifies activity phases, marks transitions between working episodes, and highlights disturbance or reorganization zones. It shows where accretion has been stable, where layers have been disrupted, and where stratigraphy is too fractured to support confident reconstruction. In doing so, it helps define where reconstruction windows open and where they close.

*Loss & Absence Reasoning (LAR)*

LAR clarifies what is not present. It distinguishes loss (material that once existed and has left shadows) from absence (material that may never have been created or recorded), and treats both as evidence about the limits of the record.

Working from loss shadows, orphaned ecofacts, broken relationships, and incomplete strata, LAR identifies genuine gaps, deletion or overwrite zones, and missing contextual links. It does not speculate about motives or invent missing content. Instead, it marks the regions where interpretation must stop and defines the uncertainty bands that other methods must respect.

*Relational & Contextual Metadata Mapping (RCMM)*

RCMM charts how artifacts, ecofacts, metadata, and system residues relate to one another. It uses relationships already surfaced by the Evidence Layer (shared paths, co-edit patterns, temporal proximity, shared ecofacts, environment states) and by DSA and LAR to build relational clusters and context maps.

RCMM reinforces areas where methods converge and records where they diverge. It helps distinguish independent deposits from integrated ones, reveals operational and contextual groupings, and preserves conflicts rather than smoothing them away. The result is a structured map of how pieces of the deposit hang together. RCMM can be applied iteratively: early for basic mapping, and later to synthesize structures surfaced by the other methods

*Ecofact-Based Reconstruction (EBR)*

EBR interprets how the environment responded to activity. Ecofacts, logs, sync traces, thumbnails, caches, and other system residues are read alongside environment signatures to locate inner and outer boundaries around events: where activity begins and ends, and how systems reacted around those events.

EBR identifies halos, bursts, and sync layers, and flags boundary disruptions or irregularities. It shows where environment behavior is stable, where it has drifted or changed, and where system-driven actions dominate the record. This informs both reconstruction windows and later methods that depend on understanding environment phases.

Together, these methods answer the fundamental question:

What traces exist here, and what do they materially show about how they have interacted across Activity → Trace → Environment → Time?

Only once this foundation is established can reconstruction proceed.

**5.2 Reconstructing Sequences, Conditions, and Change**

With the evidence field defined, IA reconstructs how activity unfolded and how the environment shaped it. These methods operate within reconstruction windows and address:

- What happened?
- In what order?
- Under what environmental constraints?
- How did the system and practice evolve?

Digital Sequence Reconstruction (DSR)

DSR rebuilds the order of actions. It traces draft evolution, branching, reversions, pause–resume cycles, and the timing of creation or modification. Working from stratigraphy, relationships, and ecofact patterns, DSR follows observable change, not assumed workflow, and only operates where reconstruction windows are open.

Digital Evolution Reconstruction (DER)

DER rebuilds how the digital structures and environment evolved over time. It interprets tool signatures, metadata rewrites, sync behavior, permission constraints, platform transitions, configuration changes, and migration residues. DER reveals what conditions and structural changes shaped the activity without inferring user intention or internal decision processes.

*Temporal Drift Analysis (TDA)*

TDA identifies long-term change across time. This includes naming evolution, structure maturation, environmental transitions, workflow alterations, and stability/instability patterns. TDA distinguishes gradual drift from acute disturbance, and marks the “Time” dimension of ATET: how systems and practices change through continued use and reconfiguration.

*Provenance-Constrained Interpretation (PCI)*

PCI ensures all reconstruction remains within the boundaries of evidence, provenance, and reconstruction windows. It distinguishes direct fact from supported inference, hypothesis, and speculation. PCI documents uncertainty, governs confidence bands, prohibits interpretive overreach, and anchors all claims to observed constraints in the evidence field.

Together, these methods answer:

Given the traces we have, what sequences, environmental conditions, and long-term changes can be responsibly reconstructed over time?

**5.3 Interpreting Meaning Within Provenance**

Once structure, environment, drift, and reconstruction windows are established, IA may cautiously explore meaning within the bounds of provenance. These final methods do not interpret motive, emotion, or psychological state.

They interpret patterns of practice and internal signals present in the material.

*Operational Composition Analysis (OCA)*

OCA examines how work is composed and evolves in practice. It interprets internal structure, fragment reuse, vocabulary shifts, and compositional rhythms across versions, anchored to the stratigraphy, sequences, and environment phases already established.

OCA reveals:

- how drafts accumulate, stabilize, or fragment,
- how sections, motifs, and fragments are rearranged,
- how internal organization reflects workflow stability or disruption,

but never why participants behaved as they did. It operates only within reconstruction windows and cannot contradict earlier evidence-field or reconstruction findings.

*Digital Culture Interpretation (DCI)*

DCI explores conceptual and cultural patterns that emerge across artifacts, deposits, and environments. It reads naming conventions, structural motifs, vocabulary clusters, collaborative rhythms, and organizational habits that recur across many windows and cases.

DCI operates strictly at the cultural and organizational level and stops short of personal interpretation. It characterizes practices and conventions, not inner beliefs or traits, and remains constrained by provenance and the limits defined by PCI.

Together, OCA and DCI address:

What patterns of practice or purpose appear here, within the boundaries of evidence and provenance?

**5.4 How the Methods Work Together**

IA's methods form a natural sequence:

1. DSA, LAR, RCMM, and EBR use Evidence Layer outputs to map the evidence field and identify reconstruction windows.
2. DSR, DER, TDA, and PCI reconstruct sequence, environment, and evolution within those windows.
3. OCA and DCI interpret bounded meaning, remaining consistent with all earlier constraints.

This sequence mirrors ATET and ensures every interpretive claim remains tied to:

- what exists,
- how it formed,
- how it changed,
- what the environment contributed,
- what is missing,
- where reconstruction windows close,
- and where interpretation must stop.

No single method explains the deposit; IA works through triangulation and accumulation of constraints.

**5.5 Choosing the Right Method**

A simple rule guides method selection:

- To understand what exists: use DSA, LAR, RCMM, EBR
- To understand how it happened: use DSR, DER, TDA, PCI
- To understand what it might mean within provenance: use OCA, DCI

Examples:

- Unclear chronology: DSA + DSR + EBR
- Missing files or overwritten segments: LAR + DSA + RCMM + EBR
- Workflow reconstruction: DSR + DER + RCMM + OCA
- Cultural drift study: TDA + DCI + RCMM

No method stands alone; IA's power comes from method combinations that reinforce or constrain one another. PCI and LAR sit across these combinations to guard against overreach and to keep uncertainty explicit.

**5.6 Preventing Overinterpretation**

IA requires strict adherence to non-inference:

- Facts, supported inferences, hypotheses, and speculative possibilities must be distinguished.
- PCI must be used to document uncertainty, confidence bands, and reconstruction window limits.
- Multiple methods should confirm structural interpretations before they are treated as stable.
- Ecofacts must be treated as environmental evidence, not psychological signals.
- Ambiguity must be preserved, not resolved through imagination.
- Loss and drift define the limits of what can be known, not prompts to guess.

IA does not seek narrative completeness. It seeks disciplined clarity grounded in material evidence.

**5.7 Ecofacts as Cross-Cutting Evidence**

Ecofacts strengthen and constrain every method. They reveal:

- fine-grained timing (autosaves, sync intervals)
- environment behavior (metadata rewrites, permission propagation)
- disturbance (conflicts, partial writes, fragmentation)
- drift (signature changes over platform versions)
- loss shadows (orphaned thumbnails, sidecars, caches)

Across ATET:

- Activity produces both artifacts and ecofacts.
- Traces include intentional and system-generated material.
- Environment expresses itself through outer boundaries and their ecofacts.
- Time becomes visible through disturbance, drift, and residue accumulation.

Ecofacts ground IA in the reality of system behavior, preventing interpretive overreach. They tie inner-boundary events to outer-boundary responses and provide the cross-cutting signals that DSA, LAR, RCMM, EBR, DSR, DER, TDA, OCA, and DCI all depend on.

### 5.8 Summary

Applying the ten methods in combination provides a disciplined, evidence-centered understanding of digital activity.

IA does not reconstruct motive or intention; it reconstructs:

- traceable behavior,
- the environment shaping it,
- and how both changed over time.

The ATET model ensures that every conclusion remains tied to material reality and that the boundaries of interpretation remain clear. Reconstruction windows, ecofacts, loss and absence, and provenance-constrained interpretation (PCI) together define where IA can speak and where it must stop.

## 6. Boundary Conditions and Ethical Commitments

Information Archaeology depends on clear limits. The discipline only works if analysts stay within what the traces can support and make uncertainty explicit. The goal is not to create stories about people or to reconstruct motives, but to understand behavior, environment, and change over time strictly through evidence.

Boundary conditions ensure that the analysis remains grounded, defensible, and respectful of both the individuals whose traces appear in the archive and the organizations responsible for managing them. Ethical commitments ensure that interpretations do not exceed the reach of the material.

Together, these principles form the foundation of responsible IA practice.

### 6.1 Boundary Condition 1: Stay Within the Evidence

IA interprets traces, not people. The discipline limits itself to what can be observed directly in:

- artifacts
- ecofacts
- metadata
- deposit structure
- sequence patterns
- drift and disturbance
- loss and absence shadows
- environment signatures
- reconstruction windows and their limits

Analysts do not speculate beyond these constraints.If a conclusion does not trace back to evidence, it cannot be included.

This boundary forbids:

- inferring emotions, motives, state of mind, or personality
- implying organizational intent where no trace exists
- assigning blame based on incomplete or ambiguous patterns
- using single signals to make strong claims

This boundary requires:

- direct citation of trace classes
- explicit linkage between claim → evidence → provenance
- multiple hypotheses when evidence does not resolve ambiguity
- clear labeling of inference level (fact, supported inference, hypothesis, or speculative possibility), enforced by PCI

This maintains scientific and legal defensibility.

**6.2 Boundary Condition 2: Preserve Uncertainty**

Digital traces rarely provide complete pictures. IA treats uncertainty as a structural part of the evidence field, not a flaw.

Analysts must:

- identify unresolved regions
- mark loss and absence clearly
- document alternative interpretations
- maintain reconstruction windows (where evidence is sufficient vs. insufficient)
- avoid collapsing ambiguity into a single narrative
- apply LAR, loss/absence and negative-evidence distinctions when moving from observation to interpretation

Uncertainty is a strong guardrail. PCI uses these uncertainty markings to limit how far any interpretation can go.

**6.3 Boundary Condition 3: Respect Environment Limits**

IA does not guess about system behavior beyond what the ecofacts and environment signatures actually show.

Environment reconstruction depends on:

- demonstrable system signatures
- sync behaviors
- metadata rewrites
- app-specific residues
- permission propagation
- file-format and configuration transitions

If the system's behavior cannot be confirmed through traces, it must remain out of scope.

DER and EBR operate strictly within observable environment limits defined by ecofacts, signatures, and reconstruction windows. When environment behavior is unclear or conflicting, they must preserve that uncertainty instead of filling the gap.

### 6.4 Boundary Condition 4: No Psychological or Personal Interpretation

IA avoids anything that implies inner states. Patterns in the traces may reflect habits, workflows, or technical constraints, but may not infer the user's emotional condition or personal motives.

Disallowed:

- "The user was confused here."
- "They intended X."
- "They were avoiding Y."
- "They were in a hurry."

Allowed:

- "Autosave bursts increased at 16:04 and 16:07, indicating rapid editing."
- "The overwrite pattern suggests multiple revisions in a short window."

Descriptions must reflect activity and environment, not psychology. PCI enforces this constraint across all methods.

### 6.5 Ethical Commitment 1: Evidence-First Interpretation

Every interpretation must:

- begin with the traces
- identify the evidence class used (artifact, ecofact, metadata, deposit structure)
- describe the reasoning process
- show where alternative interpretations remain plausible
- document what cannot be determined

Nothing is stated without demonstrating provenance.

PCI and EBR enforce this commitment.

### 6.6 Ethical Commitment 2: Minimal Interpretive Intrusion

The material speaks for itself. Analysts should add as little interpretation as possible.

This discipline treats traces like physical material in archaeology:

- describe first
- contextualize second
- interpret cautiously
- stop early

This protects the integrity of the deposit and the people represented within it.

**6.7 Ethical Commitment 3: Protection of Individuals**

Digital traces often contain fragments of human behavior that can be taken out of context. IA requires care to avoid harm.

Analysts must:

- avoid assigning fault
- avoid character judgments
- avoid implying negligence or incompetence
- avoid linking traces to personal identity unless necessary
- remove or anonymize identifying data when sharing findings externally

Analysts should treat every trace as sensitive material.

**6.8 Ethical Commitment 4: Organizational Transparency**

Organizations using IA must:

- make the purpose of analysis clear
- explain how interpretations were produced
- preserve the distinction between evidence and interpretation
- disclose uncertainty
- maintain versioned records of all analytical steps
- protect privacy and comply with governance frameworks

The goal is clarity and should not be used for surveillance.

**6.9 Ethical Commitment 5: Responsible Use of AI**

IA uses AI only under strict provenance boundaries. AI cannot invent, guess, or smooth over gaps.

AI must:

- cite the exact traces used
- label uncertainty
- never produce single-solution narratives where multiple explanations exist
- never infer motive or feeling
- never overwrite ecofact boundaries
- follow PCI constraints in all outputs

AI should assist, not replace, the analyst and must be prevented from drifting into storytelling.

**6.10 Boundary Condition 5: Falsifiability and Disconfirmability**

Every IA output must include conditions under which the interpretation would be wrong.

Examples:

- "If draft B predates draft A, this reconstruction is invalid."
- "If the sync marker refers to a background cleanup, not a device transition, this sequence must be revised."
- "If this metadata residue is platform-generated rather than user-generated, this inference must be removed."

This keeps conclusions testable and transparent.

**6.11 Boundary Condition 6: Reproducibility**

Two analysts, or an analyst + AI, must be able to apply the same methods to the same evidence and reach the same bounded conclusions.

Reproducibility depends on:

- documenting analytical steps
- maintaining clear evidence references
- providing alternate-hypothesis trees
- storing the evidence field unmodified
- versioning interpretations

If an interpretation cannot be reproduced, it cannot stand.

### 6.12 Ethical Commitment 6: Respect for Temporal Context

Work is always shaped by time.

- deadlines
- tool availability
- platform versions
- organizational transitions
- stability or instability in the environment

IA avoids anachronistic judgment. Interpretations must acknowledge the historical conditions embedded in the traces and be sure to not measure past behavior against present norms.

Time constrains interpretation just as much as evidence.

### 6.13 Ethical Commitment 7: Cultural Sensitivity and Non-Attribution

DCI offers cultural interpretations, not psychological or interpersonal ones.

Analysts must:

- frame cultural insights as hypotheses
- acknowledge the need for external validation
- avoid attributing behavior to specific individuals without necessity
- avoid moral interpretations of cultural patterns (e.g., “bad workflow,” “chaotic thinking”)

IA stays neutral about culture; it describes patterns, not values.

### 6.14 Summary

Boundary Conditions define the limits of IA. Ethical Commitments define the responsibilities of IA.

Together, they ensure that:

- the discipline stays grounded in evidence
- interpretations remain defensible
- uncertainty is honored

- individuals are protected
- organizations remain transparent
- AI is used responsibly
- the Activity → Trace → Environment → Time model is upheld
- no conclusions exceed the reach of the deposit

This keeps Information Archaeology rigorous, humane, transparent, and trustworthy.

## 7. The Information Archaeology Model (ATET)

Activity → Trace → Environment → Time

The ATET model is the foundational structure of Information Archaeology. It describes how digital work forms, how systems respond, and how traces accumulate, drift, and disappear. Every method in IA is grounded in this model. It keeps interpretation anchored to observable material and prevents analysts from reaching beyond what the evidence can support.

ATET is intentionally simple. It reflects the real mechanics of digital production and the natural limits of what digital evidence can show.

### 7.1 Activity

Activity is the human action that sets everything in motion.

It includes:

- creating, editing, saving, exporting, moving, renaming
- running a model, updating a template, generating a report
- interacting with tools and systems in any intentional way

IA does not interpret activity directly. Human actions cannot be reconstructed through intention, emotion, or motivation. They can only be known through the traces they leave behind.

Key commitment:

IA never describes why activity occurred; only what happened, as evidenced by the material.

### 7.2 Trace

Trace refers to the material evidence created by activity and by the system reacting to that activity.

There are two categories:

*Artifacts*

Human-authored objects:

- drafts
- documents

- spreadsheets
- diagrams
- scripts
- exports
- screenshots

Artifacts show intentional work, but they rarely show everything. They capture outcomes, not the conditions or pathways that produced them.

*Ecofacts*

System-generated traces:

- autosaves
- sync residues
- conflict copies
- thumbnails
- metadata rewrites
- temp and recovery files
- cache fragments
- partial-write residues

Ecofacts show how the environment responded to activity. They encode timing, disruption, constraint, and drift. Ecofacts frequently reveal events that artifacts alone cannot. Together, artifacts + ecofacts form the evidence field.

No interpretation is allowed to exceed what these traces contain.

**7.3 Environment**

Environment is the set of technical, organizational, and procedural conditions shaping how traces form.

It includes:

- the operating system and its behaviors
- cloud sync logic
- versioning rules
- permissions
- naming conventions

- platform migrations
- tool-specific signatures
- organizational workflows
- governance maturity

The environment leaves visible fingerprints in ecofact boundaries:

- metadata rewriting
- sync timing
- drift onset
- conversion residue
- permission propagation
- multi-device conflict patterns

IA treats the environment as a mechanical actor rather than a psychological actor. It constrains what is possible and shapes what survives. Environment is observable only through traces.

IA never assumes system behavior that cannot be demonstrated via ecofacts or known platform signatures.

**7.4 Time**

Time gives structure to activity and trace formation.

It introduces:

- sequences
- drift
- aging
- migration
- compression
- overwrite patterns
- metadata degradation
- platform version changes

Time is not simply “when things happened.”

It is how traces evolve.

IA reads time through:

- stratigraphy (DSA)
- autosave intervals (EBR)
- naming drift (TDA)
- residue accumulation (DSA + EBR)
- metadata decay or rewriting
- disturbance footprints

Time clarifies what can still be interpreted, what has shifted, and what has permanently disappeared.

### 7.5 ATET as a Reconstruction Framework

ATET allows IA to rebuild activity without guessing.

Activity → Trace

All human action becomes visible only through material residue. Activity creates artifacts and ecofacts that encode what happened, but never why.

Trace → Environment

Ecofacts reveal how systems reacted — where they constrained, rewrote, synchronized, or reshaped activity. From traces, IA infers how specific environments behaved.

Environment → Time

As time passes, both artifacts and ecofacts drift, age, and lose fidelity, allowing analysts to locate transitions, migrations, disturbances, and periods of stability.

Reconstruction Through Time

By examining how traces accumulated and changed over time within an environment, analysts can reconstruct what activity occurred — always within provenance limits.

ATET is not circular. It distinguishes:

- **Formation (forward)**: Activity produces traces that accumulate over time within an environment.
- **Reconstruction (backward)**: Analysts read time-distributed traces to understand activity and environment behavior.

ATET produces a grounded, physically plausible account of digital events without drifting into intention or subjective interpretation.

**7.6 How ATET Structures the Ten Methods**

Every IA method operates within the ATET model.

Evidence-Field Mapping (DSA, LAR, RCMM, EBR)

These methods work closest to Activity → Trace → Environment. They examine the traces directly and:

- structure the deposit into layers and phases (DSA),
- identify loss, absence, and negative evidence (LAR),
- map relational and contextual structures across the deposit (RCMM),
- interpret ecofact responses and inner/outer boundaries (EBR).

Together they map the evidence field and define reconstruction windows. They do not create data; they organize what the Evidence Layer has already measured.

Reconstruction Layer (DSR, DER, TDA, PCI)

These methods work across Trace → Environment → Time. They:

- rebuild sequences of actions and events (DSR),
- reconstruct how digital structures and environments evolved (DER),
- characterize long-term drift and stability patterns (TDA),
- constrain all reconstruction to provenance and uncertainty limits (PCI).

They place traces in time, reconstruct activity sequences, and model the environments that shaped them, always within the windows defined by the mapping methods.

Insight Layer (OCA, DCI)

These methods operate on the fully mapped and reconstructed field:

- OCA interprets how work was composed and paced in practice.
- DCI interprets recurring conceptual and cultural patterns across artifacts and deposits.

They stay strictly within Trace + Environment + Time. They do not interpret motives. They produce bounded meaning grounded in observable structure and constraints.

### 7.7 Why ATET Matters

ATET solves the core problem that led to the invention of Information Archaeology:

Digital work creates more material than people realize, and without a disciplined structure, interpretation becomes guesswork.

ATET prevents:

- overinterpretation
- narrative drift
- psychological inference
- retrospective judgment
- flattening of ambiguity
- collapsing multiple possible sequences into one

ATET creates:

- disciplined reconstruction
- clear provenance boundaries
- transparent uncertainty
- reproducible analysis
- defensible narratives grounded in evidence

ATET transforms digital traces into structured, interpretable material without exceeding the limits of what the evidence can support.

### 7.8 Summary

ATET is the spine of Information Archaeology.

It centers four simple, observable realities:

- Activity creates traces.
- Traces form within an environment.
- Environment interacts with time.
- Time reshapes everything.

The ten methods operationalize this model, PCI safeguards, Ecofacts illuminate, Stratigraphy anchors. The Structure-layer reconstructions depend on it, and the Insight-layer methods respect its boundaries.

ATET is how IA stays grounded, coherent, and faithful to the deposit, never exceeding what can be responsibly known.

## 8. The Evidence Field: Anatomy of Digital Deposits

The Evidence Field is the working surface of Information Archaeology.

It is the total landscape of artifacts, ecofacts, metadata, drift patterns, disturbance signatures, and losses that accumulate as activity unfolds within an environment over time.

Every IA method begins with the Evidence Field. It defines what exists, how it is distributed, and where interpretation can and cannot go. Rather than treating files as isolated objects, IA reads the entire deposit as a connected material system.

### 8.1 What Is a Digital Deposit?

A digital deposit is a collection of traces, artifacts and ecofacts, that formed through activity within a particular environment and time span.

Deposits reflect:

- the activity that produced them
- the tools and systems that shaped them
- the timing of events
- the patterns of drift, disturbance, or loss
- the relationships among artifacts and ecofacts

A deposit is not a folder, nor a project, nor a directory tree. It is the material record of a sequence of actions, system reactions, and environmental constraints. Multiple deposits may coexist within a single location, and a single deposit may spread across multiple locations or storage systems.

### 8.2 Components of the Evidence Field

The Evidence Field consists of four primary components:

1. Artifacts

The intentional outputs of human activity: documents, drafts, spreadsheets, diagrams, exports, code, screenshots, and so on.

2. Ecofacts

The system-generated residues: autosaves, sync traces, metadata rewrites, conflict copies, temp fragments, previews, thumbnails, partial writes, device-specific residues.

3. Metadata Structures

Timestamps, paths, permissions, naming conventions, authorship markers, application signatures - the scaffolding around the traces.

4. Spatial and Temporal Patterns

Layering, clustering, drift, disturbance, loss shadows, relational groupings, temporal halos, and sequence fingerprints.

The Evidence Field is not defined by content but by structure and behavior, revealed through the arrangement and characteristics of the traces.

**8.3 How Deposits Form: From Activity to Trace**

Deposits form through the ATET chain:

Activity → Trace → Environment → Time

- Activity produces artifacts and triggers ecofacts.
- The environment shapes how traces appear, store, rewrite, or conflict.
- Time introduces drift, loss, overwrites, or accumulation.
- The result is a layered, multi-textured deposit that encodes behavior and environment.

Deposits are dynamic: they grow, shift, fragment, and stabilize based on activity intensity and environmental conditions.

**8.4 Strata: The Layered Structure of Digital Work**

Strata are groups of traces that formed together or within a tight temporal window.

Strata can be identified through:

- timestamp clusters
- autosave bursts
- incremental versions
- consistent naming rounds
- tool or system signatures
- structural similarity across drafts
- ecofact halos that outline activity zones

Strata reveal:

- periods of focused work
- transitions between stages
- pauses or interruptions
- rewriting events
- migrations or reorganizations
- activity cycles across days, weeks, or versions

DSA uses strata to anchor the timeline and reveal how the deposit formed.

**8.5 Clusters and Communities of Practice**

Clusters are relational groupings of artifacts and ecofacts that share metadata patterns, filenames, paths, structures, or timing.

Clusters can represent:

- subprojects
- task groups
- conceptual units
- tool-produced batches
- iterative cycles
- shadow workflows (unofficial but consistent patterns)

RCMM detects clusters by examining:

- co-location
- path relationships
- naming logic
- timestamp adjacency
- ecofact-family relationships (autosave → artifact → preview → sync copy)
- tool signatures

Clusters reveal the relational geometry of the evidence field: how the work was organized, implicitly or explicitly.

**8.6 Halos and Boundary Layers**

Every artifact and ecofact produces a halo: a surrounding zone of related traces indicating activity intensity, disturbance, or environmental reaction.

Halos may include:

- autosave bursts
- metadata rewrites
- conflict copies
- sync shadows
- thumbnails and previews
- local/remote divergence footprints

Halos reveal:

- editing tempo
- the scale of change
- device transitions
- environment response zones and limits
- conflict zones
- periods of instability

In IA's dual-boundary model, halos are how inner and outer boundaries become visible:

- inner layers show traces tightly coupled to the event itself,
- outer layers show how the environment responded around that event.

In EBR, halos help determine the depth and reach of an event, forming the first layer of an impact footprint.

**8.7 Drift Zones**

Drift zones form when the environment changes gradually — a sync engine update, a slow naming convention shift, long-term platform adoption, or a new tool that is phased in over time.

Drift appears as:

- changed metadata patterns
- new or altered ecofact structures
- updated application signatures

- shifts in naming or formatting
- reduced or increased residue density

TDA maps drift zones to determine:

- when the environment changed,
- how change propagated through the deposit,
- whether drift corresponds to organizational or system transitions.

Drift is one of the clearest markers of time in the deposit. It usually preserves continuity: patterns evolve rather than break. Some phenomena, such as platform migrations, can involve both a short, disruptive disturbance event and a longer drift zone that follows; TDA and DSA distinguish these aspects.

**8.8 Disturbance Patterns**

Disturbance refers to sudden, disruptive events in the normal formation of traces. Disturbance concentrates change into a short span of time and often breaks continuity.

Disturbance patterns include:

- rapid renames or relabeling waves
- relocations or major folder restructures
- platform migrations executed as one-time cutovers
- bulk operations
- sync conflicts
- deletions and overwrites
- permission resets
- corruption events

Disturbance is not negative or unusual; it is part of digital life. It marks where the deposit’s stratigraphy has been abruptly altered.

Disturbance often leaves:

- offset or compressed sequences
- orphaned ecofacts
- incomplete chains

- mismatched timestamps
- localized drift pockets created by the disruption
- abrupt structural shifts

LAR, DSA, EBR, and PCI work together to document disturbance without overinterpreting it. DSA shows where layers are broken or rearranged, LAR identifies loss and absence around the event, EBR characterizes the environment's response in the halo, and PCI limits what can be claimed about what happened inside and after the disturbance.

### 8.9 Loss Shadows

Not all parts of a deposit survive. Loss shadows are regions where material is missing but leaves recognizable traces.

Loss shadows appear as:

- thumbnails without documents
- autosaves without artifacts
- sync logs referencing missing files
- previews of now-deleted drafts
- partial-write residues
- sudden gaps in version sequences

Loss is itself evidence. LAR treats disappearance, overwrite, corruption, and deletion as meaningful signals rather than failures.

Loss shadows often clarify:

- what once existed
- what replaced it
- where migration or instability occurred
- how the environment handled failure or conflict

All conclusions remain bounded: loss cannot be used to reconstruct content.

### 8.10 Impact Footprints

An impact footprint is the combined shape of:

- strata

- halos
- drift zones
- disturbance patterns
- ecofact clusters
- loss shadows

Together, they reveal:

- the intensity of activity
- the boundaries of an event
- the path and scale of change
- the environmental response
- the time window in which activity occurred

EBR defines the footprint; DSR and DSA integrate it into sequence and stratigraphy.

Footprints anchor reconstruction in material form, never narrative intuition.

## 9. Evidence Requirements and Interpretive Use

Information Archaeology depends on what software can reliably detect and describe. This section defines the Evidence Layer: what software must provide, how its outputs are understood, and how they support but never replace interpretation.

9.1 The Evidence Layer (Software Requirements)

The Evidence Layer is the measurement layer. It is the interface between digital systems and IA methods.

At minimum, the Evidence Layer must be able to:

- detect artifacts, ecofacts, and associated metadata,
- report their observable properties as Collection Facts,
- derive additional, clearly procedural measurements as Generated Facts,
- attach confidence indicators to its own outputs, and
- preserve provenance for all measurements.

In practical terms, this means the Evidence Layer must:

- identify files, versions, paths, timestamps, logs, previews, caches, and sidecar files;
- distinguish system-generated residues (ecofacts) from primary artifacts where possible;
- record where and how each fact was obtained (source system, index, log, scan);
- avoid altering the deposit while measuring it, or record any changes it introduces.

The Evidence Layer does not interpret behavior, environment, or culture. It measures.

Software measures while IA interprets.

**9.2 Collection Facts vs. Generated Facts**

IA distinguishes between two kinds of Evidence Layer output: Collection Facts and Generated Facts. Both are essential, but neither are explanations.

Collection Facts are directly observable properties gathered from the deposit:

- file paths, sizes, timestamps, and permissions;
- metadata fields as stored by systems or applications;
- raw log entries and event records;
- the existence of thumbnails, previews, caches, autosaves, conflict copies, and other ecofacts;

- explicit relationships recorded by systems (such as version links or parent–child references).

Generated Facts are procedural outputs derived from Collection Facts. They are still measurements, not interpretations. Examples include:

- counts (number of versions, number of edits per time window),
- intervals (time between successive saves, syncs, or accesses),
- simple classifications (filetype categories, path segments),
- graph structures (who-edited-what, co-occurrence and adjacency),
- environment signatures detected from repeated patterns,
- confidence indicators for specific observations or links.

Collection Facts and Generated Facts are both inputs to IA methods. They do not answer questions by themselves. They do not claim "what happened" or "why." They describe what can be seen and what can be procedurally computed from it.

No mathematics is required beyond this distinction. The point is epistemic clarity:

- Collection Facts and Generated Facts are measurements.
- Methods and analysts supply interpretation.

### 9.3 Confidence and Interpretive Boundaries

Not all evidence is equally reliable. The Evidence Layer must signal confidence levels so that IA methods can respond appropriately.

Analysts must be able to distinguish at least four states:

- **High confidence** – measurements are well supported (for example, multiple corroborating sources, stable signatures, unambiguous links). These can support stronger claims, provided they are consistent across methods.
- **Medium confidence** – measurements are plausible but may conflict in small ways or rely on single sources. These require triangulation: other methods should confirm or qualify them before they anchor major conclusions.
- **Low confidence** – measurements are weak, noisy, or derived from partial information. These can suggest possibilities but should rarely form the basis of reconstruction.
- **Missing data** – the Evidence Layer cannot provide a measurement at all. This sets a hard boundary: reconstruction must acknowledge the gap and may not cross it without clearly labeled speculation.

IA does not prescribe a numeric scheme. Confidence can be communicated qualitatively, as long as:

- the criteria for each level are documented,
- different runs of software are comparable, and
- analysts can see which parts of their reasoning rest on which confidence levels.

Confidence levels, combined with loss and absence markers, help define reconstruction windows and ensure that PCI can enforce interpretive boundaries.

**9.4 Method–Evidence Requirements Mapping**

Each IA method depends on particular kinds of evidence. The Evidence Layer does not need to know how methods work, but it should expose the measurements they rely on.

At a minimum:

- **DSA** (Digital Stratigraphic Analysis)

needs temporal groupings, ordering signals, continuity/discontinuity indicators, ecofact residues around activity, and markers of disturbance or reorganization.

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

needs loss and absence indicators: missing links in version chains, orphaned ecofacts, broken relationships, incomplete strata, and markers of deletion, overwrite, or failed migration.

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

needs relationship and context signals: co-edit patterns, shared paths or folders, temporal proximity, shared ecofacts or logs, environment states, and metadata that link artifacts and residues.

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

needs ecofact sequences and system-response indicators: sync traces, thumbnails, caches, log events, autosave intervals, and environment signatures that show how systems behaved around events.

- **DSR** (Digital Sequence Reconstruction) needs version relationships, change detection, ordering cues, and timing information that allow sequences of edits, branches, and merges to be reconstructed within windows.

- **DER** (Digital Evolution Reconstruction)

needs environment signatures, rewrite indicators, configuration changes, permission shifts, platform transitions, and migration residues that show how structures and environments evolved.

- **TDA** (Temporal Drift Analysis)

needs markers of change over time: evolving naming patterns, structural maturation, long-term environment shifts, recurring tool transitions, and stability/instability patterns across spans.

- **OCA** (Operational Composition Analysis)

needs internal composition and rhythm signals: version-to-version diffs, repeated fragments, structural elements (sections, headings, blocks), and alignment with environment phases and editing bursts.

- **DCI** (Digital Culture Interpretation)

needs recurrence patterns across artifacts and deposits: naming conventions, conceptual motifs, vocabulary clusters, recurring workflows, and characteristic combinations of tools and structures.

- **PCI** (Provenance-Constrained Interpretation)

needs provenance chains, confidence indicators, reconstruction windows, and method outputs, so it can distinguish fact from inference, track uncertainty, and enforce interpretive limits.

This mapping describes what the evidence is, not how to produce it. Implementation details remain open and future-proof. Any system that can deliver these kinds of measurements, with provenance and confidence, can support IA.

### 9.5 Validation Principles

IA aligns its validation practices with archaeology, history, digital forensics, humanities scholarship, and information science. Rather than prescribing specific algorithms, it focuses on how measurements and interpretations can be trusted and checked.

Core validation principles include:

- Reproducibility

The Evidence Layer, when run again on the same deposit with the same configuration, should produce the same Collection Facts and Generated Facts, or clearly document any differences.

- Transparency of measurement

How measurements are obtained, transformed, and aggregated must be documented. Analysts should be able to see what each fact means and where it came from.

- Documentation of uncertainty

Confidence levels, loss and absence regions, and reconstruction window limits must be explicit. Uncertainty is recorded, not hidden.

- Traceable provenance of Evidence Layer outputs

Every fact should be traceable back to its source systems, logs, or files. When multiple sources are fused, the process must be documented.

These principles ensure that IA can be revisited, challenged, and refined. They keep the discipline grounded in material evidence rather than opaque tooling.

**9.6 The Evidence Field as a System**

The Evidence Field is more than a list of files or a static index. It is the system formed by traces, relations, and environment behavior across time.

The Evidence Field is:

- self-forming: it emerges from how people and systems interact, not from IA's preferences;
- environmentally shaped: platforms, tools, policies, and infrastructures leave strong signatures;
- temporally structured: layers, phases, drift, and disturbance are baked into the record;
- internally relational: artifacts, ecofacts, and metadata reference and depend on each other;
- incomplete by nature: loss, absence, and partial residues are expected, not exceptional;
- shaped by technical and organizational constraints: retention policies, access controls, migration strategies, and local practices.

Interpretation happens within this system, not on top of a sanitized subset.

The discipline respects:

- gaps and missing spans,

- contradictions between sources,
- drift and environment shifts,
- partial and fragmentary evidence,
- asymmetric and uneven coverage.

IA does not “fix” these features. They are part of the field. Methods are designed to live with them, to map where they appear, and to stop where they prevent responsible reconstruction.

**9.7 Summary**

The Evidence Field is the substrate of Information Archaeology.

It reveals:

- how activity formed material in the first place,
- how the environment shaped the result,
- how time transformed both through drift and disturbance,
- where loss and absence have altered the field,
- how artifacts and ecofacts relate to one another,
- and what can and cannot be responsibly known.

Understanding the anatomy and behavior of the deposit ensures that all ten methods stay grounded in real material, not assumptions.

It also ensures IA retains its central commitments:

- non-inference about motive and inner state,
- strict provenance discipline,
- preservation of uncertainty rather than erasure,
- ethical transparency about what is and is not known,
- reproducibility of measurements and reasoning,
- and respect for individuals and environments whose work produced the traces under study.

## 10. Analytical Workflow: From Deposit to Interpretation

The IA workflow is intentionally structured. It reflects the order in which evidence becomes available, the constraints of provenance, and the natural limits of what can be known about digital work.

It follows a simple rule: Begin with what exists, move only as far as the traces allow, stop early.

This section outlines the full workflow from deposit identification to final, provenance-constrained interpretation.

### 10.1 Step 1: Identify and Delimit the Deposit

Every IA analysis begins with defining the deposit: the set of artifacts, ecofacts, and metadata that belong to the same episode, period, or domain of activity.

Key tasks:

- Locate all known artifacts
- Gather associated ecofacts (autosaves, previews, conflict copies, metadata fragments)
- Identify paths, locations, or systems in scope
- Document all sources and exclusions
- Note incomplete or inaccessible regions

Purpose:

Set the physical boundaries of the evidence field before drawing conclusions.

Associated methods:

- DSA (initial structure)
- RCMM (relational mapping)
- EBR (ecofact identification)
- LAR (loss detection)

This foundation protects the analysis from scope creep or selective evidence.

### 10.2 Step 2: Establish the Evidence Field

With the deposit defined, the next step is mapping the material landscape.

Key tasks:

- Lay out the chronological and spatial distribution of traces
- Identify strata, clusters, halos, drift pockets, and disturbance zones
- Mark loss shadows, overwrites, and missing layers
- Describe the arrangement without interpretation

Purpose:

Create a clear, descriptive model of what exists. The goal is not to explain anything yet — only to document the field.

Associated methods:

- DSA
- RCMM
- EBR
- LAR

**10.3 Step 3: Reconstruct Sequences of Activity**

Once the landscape is understood, IA reconstructs the order of events.

Key tasks:

- Compare versions, drafts, and deltas
- Use autosave timing and ecofact bursts for precision
- Detect event boundaries (start → pause → return)
- Identify branches, merges, reversions, or parallel streams
- Anchor all sequences in material evidence

Purpose:

Establish a defensible timeline grounded solely in traces. Sequence reconstruction completes the Activity → Trace → Time part of ATET.

Associated methods:

- DSR (primary driver)
- DSA (sequence anchoring)
- EBR (event boundary detection)
- LAR (interruptions or loss during sequence)

- PCI (constraint enforcement)

**10.4 Step 4: Reconstruct the Decision Environment**

Activity does not occur in a vacuum. Traces reflect the system that produced them.

Key tasks:

- Identify tools and platforms in use
- Detect system transitions (cloud → desktop → device)
- Map permission structures and file behaviors
- Recognize metadata rewrites, conversion events, or sync rules
- Understand constraints, pressures, or boundary conditions

Purpose:

Reconstruct the environment through the traces, never through assumption. This step completes the Trace → Environment portion of ATET.

Associated methods:

- DER (primary driver)
- EBR (outer boundary mapping)
- RCMM (environment signature relationships)
- TDA (environmental drift over time)
- PCI (preventing speculative system claims)

**10.5 Step 5: Identify Drift, Change, and Evolution**

Digital environments evolve, and so do practices.

Key tasks:

- Examine long-term naming shifts
- Detect structural or formatting evolution
- Identify new tool adoption or workflow phases
- Analyze drift in ecofact signatures
- Document changes in residue patterns over time

Purpose:

Understand how time reshaped the deposit and the conditions of work. This step locates the work historically within its technical and organizational lifecycle.

Associated methods:

- TDA (primary)
- DSA (shifts in stratigraphy)
- EBR (signature drift)
- RCMM (metadata evolution)

**10.6 Step 6: Conduct Bounded Insight Analysis**

Once evidence, sequence, environment, and drift are established, IA can produce insight, always within provenance limits.

Key tasks:

- Analyze internal structure (OCA)
- Detect cultural patterns (DCI)
- Identify recurring workflows
- Surface stable vs. unstable practices
- Map conceptual themes or fragment reuse
- Note vocabulary shifts or conceptual drift

Purpose:

Extract meaning that exists within the traces, without inferring personal states or motives. Insight cannot override evidence or fill gaps left by loss.

Associated methods:

- OCA
- DCI
- TDA
- RCMM

PCI and ATET heavily constrain this phase.

**10.7 Step 7: Apply Provenance-Constrained Interpretation (PCI)**

PCI is not a single step; it is the discipline applied across all steps.

However, at this stage, PCI formalizes the final boundaries.

Key tasks:

- Distinguish fact, supported inference, hypothesis, and speculation
- Label uncertainties, gaps, and alternate explanations
- Document assumptions that were not made (non-inference transparency)
- Provide disconfirmation criteria (“This would be wrong if…”)
- Confirm internal consistency across all methods
- Validate interpretations through cross-method triangulation

Purpose:

Produce a defensible interpretation that reflects the evidence rather than the analyst.

**10.8 Step 8: Synthesize Into a Final Narrative (Bounded Interpretation)**

The final output is not a story; it is a structured account of what the traces show.

A proper IA synthesis includes:

- A clear description of the deposit
- A reconstructed timeline anchored in evidence
- A reconstruction of the environment and its constraints
- Identification of drift, evolution, or disturbance
- Insight derived strictly from trace patterns
- Explicit uncertainty windows
- Conditions under which the interpretation would change

The narrative avoids:

- attributing intention
- describing emotions
- moral judgments
- organizational blame
- psychological interpretation
- anything not grounded in the traces themselves

The narrative is transparent, reproducible, and falsifiable, the aim is not to create a retrospective story.

**10.9 Who the Workflow Protects**

- The individuals whose work is represented by preventing overinterpretation or character judgments.
- The organization receiving the analysis by providing transparent, defensible reasoning based on evidence.
- The analyst and IA discipline by maintaining reproducibility, falsifiability, and clear provenance boundaries.
- The deposit itself by treating traces as material evidence, without distortion.

**10.10 Summary**

The IA workflow follows the natural order of how evidence forms:

1. Define the deposit
2. Map the evidence field
3. Reconstruct sequences
4. Reconstruct environment
5. Identify drift over time
6. Conduct bounded insight
7. Apply PCI
8. Synthesize interpretation

This mirrors the ATET model and stays faithful to the limits of the material. IA does not fill in gaps, invent motives, or elevate hypothesis to fact. It reconstructs activity through trace, in the context of environment and time, and stops where the evidence stops.

## 11. Responsible AI in Information Archaeology

AI can assist in archaeological analysis of digital deposits, but it must operate under explicit epistemic discipline. Information Archaeology is built on the idea that traces, not assumptions, define what can be known. AI systems must be subject to the same, and in some cases, even stricter contstraints than those applied to human analysts..

IA does not allow interpretive improvisation, fiction, reconstruction of missing content, or psychological inference. AI can then support pattern discovery, sequencing, clustering, and structured analysis, but always within the limits of the deposit itself.

Responsible AI is not optional. It is a core principle of the discipline.

### 11.1 Core Principle: AI Cannot Exceed the Evidence

AI must be prevented both technically and methodologically from producing claims that exceed the traces.

This includes:

- inferring motives
- filling gaps
- smoothing irregularities
- assuming missing context
- generating a coherent narrative where ambiguity exists
- resolving contradictions without evidence
- “fixing” drift, loss, or disturbance

AI is restricted to:

Activity → Trace → Environment → Time

No leap beyond this chain is allowed. Every AI output must be tethered to evidence in a way that is visible, verifiable, and falsifiable.

### 11.2 Evidence-First Operation

AI may only operate on:

- the artifacts provided

- the ecofacts provided
- the metadata present
- the structures, deposits, and drift contained in the corpus

“No external knowledge” is the default unless explicitly permitted and flagged.

AI must not introduce:

- external examples
- domain-general assumptions
- typical workflows
- common practices
- “likely interpretations”

IA requires strict bounded examination: The only valid signals are those inside the deposit.

**11.3 Provenance Visibility Requirements**

Every AI-generated output must make provenance transparent.

AI must document:

- which files were used
- which segments within those files
- which metadata fields
- which timestamps
- which ecofacts
- which transformations or comparisons

This prevents hidden inference and “authority inflation,” where an output appears grounded but is actually invented.

AI outputs should be auditable down to:

- the line of text
- the timestamp cluster
- the draft boundary
- the ecofact signature

If the user cannot verify where a claim came from, the claim is invalid.

### 11.4 Explicit Uncertainty

AI must express uncertainty the way IA itself does.

Outputs must include:

- confidence bands
- alternative plausible readings
- statements of ambiguity
- what evidence is missing
- what cannot be determined
- what would change the interpretation

AI must refuse to produce precise claims from coarse evidence. In IA, uncertainty is not an inconvenience, it is a structural feature of the digital past.

### 11.5 Hallucination Boundaries and Rejection Conditions

AI-generated claims must be automatically rejected when they:

- are unsupported by trace evidence
- rely on imagined content or missing drafts
- assume user intention or emotion
- depend on unstated contextual knowledge
- produce a single narrative where ambiguity persists
- provide detailed explanations when the evidence is sparse
- fill gaps in sequences without material justification

Hallucination is not a bug; it is an inherent risk. IA treats hallucination as a taphonomic hazard requiring explicit analysis and containment.

### 11.6 No Single Narrative

AI must never collapse complexity into a single explanation.

Instead, it should produce:

- multiple hypotheses

- scenario sets
- conditional interpretations
- divergent paths where evidence diverges

Digital archaeology is plural by nature. AI must maintain that plurality.

**11.7 Multi-Hypothesis Generation (N>1)**

AI should always produce more than one plausible interpretation of:

- sequence
- environment
- drift pattern
- disturbance event
- relational structure
- strata formation

Each hypothesis must be grounded in:

- specific evidence
- ATET logic
- taphonomic considerations
- temporal sequencing
- observed constraints

Hypotheses must not imply preference unless evidence supports it.

**11.8 AI Must Respect Ecofact Boundaries**

Ecofacts define both the event and the environment.

AI must not:

- treat ecofacts as optional noise
- assume ecofacts imply motive
- override ecofact boundaries with narrative smoothing
- treat ecofact gaps as evidence of human decision
- invent internal data from residue

Ecofacts are material. They limit what the analyst — and AI — may claim.

### 11.9 AI Must Preserve Loss, Damage, and Drift

AI is forbidden from:

- reconstructing missing content
- extrapolating deleted material
- “repairing” partial files
- smoothing drift into coherent evolution
- inferring the cause of loss
- treating absence as evidence

Loss, ambiguity, noise is treated as data, while gaps are structure. AI should protect this boundary.

### 11.10 Ethical and Epistemic Requirements

IA includes a set of non-negotiable ethical principles that apply equally to humans and AI.

Transparency

AI must show:

- where evidence came from
- which methods were applied
- why conclusions are restricted
- where uncertainty lies

Traceability

Every claim must link directly back to material in the deposit.

Humility

AI must acknowledge:

- ambiguity
- limits
- alternative explanations
- interpretive weakness

Humility is a methodological strength.

Stewardship

Digital traces represent human labor and memory.

AI must protect:

- contextual integrity
- privacy
- dignity
- provenance
- organizational safety

IA is not surveillance.

It is a stewardship practice for the digital past.

**11.11 AI as Methodological Assistant, Not Interpreter**

AI may support IA by:

- structuring deposits
- detecting patterns
- mapping metadata
- identifying ecofact clusters
- proposing multiple bounded hypotheses
- generating diagrams or analytic scaffolding

AI may not:

- narrativize
- personalize
- moralize
- psychologize
- speculate
- oversimplify
- “fill in the story”

AI does the mechanical work, not the interpretive work.

### 11.12 Reproducibility and Falsifiability

AI-generated analysis must be:

- reproducible by another run
- reproducible by a human
- falsifiable by new evidence
- inspectable down to the raw traces

If it is not reproducible, it is not IA.

### 11.13 Summary

Responsible AI in IA ensures that:

- AI stays within the traces
- uncertainty is explicit
- ambiguity is respected
- no single narrative controls the interpretation
- ecofacts constrain every claim
- provenance is transparent
- hypotheses remain plural
- privacy and dignity are protected

AI does not replace the analyst. AI strengthens the discipline by reinforcing its boundaries.

## 12. Applications & Industry Relevance

Information Archaeology isn't a theoretical exercise. It's a practical way to understand how digital work actually forms, how meaning gets encoded, and how it slowly drifts, fragments, or disappears. Most individuals and organizations aren't drowning in a lack of information, they're drowning in too much, spread across too many systems, in too many half-finished states.

The same pattern shows up everywhere: digital artifacts accumulate faster than anyone can interpret them. Context collapses. People leave. Systems migrate. Workflows evolve. By the time someone needs clarity, the material record is already stratified.

IA steps in with a simple promise: to read what's here, and let the traces speak for themselves.

Below are domains in which Information Archaeology proves extremely useful.

### 12.1 Personal Knowledge & Self-Understanding

Reconstructing your own digital life.

Most people have a decade or two of digital residue scattered across:

- notes
- screenshots
- drafts
- abandoned folders
- duplicate projects
- app exports
- half-built organizational systems

It's just unstructured, layered, and partially forgotten and often unused.

Information Archaeology helps individuals:

- rebuild old or abandoned projects
- understand how an idea or phase of life evolved
- find patterns in their own decision-making
- recover material that slipped out of memory
- make sense of transitions and turning points
- regain ownership of their digital past

It’s essentially a clarity practice, a way to turn overwhelm into coherence.

**12.2 Organizational Memory & Decision Clarity**

Recovering the context organizations keep losing.

Most organizations suffer from institutional amnesia:

- staff turnover
- tool migrations
- overwritten drafts
- lost attachments
- hollowed-out “final” documents
- missing decision context

Final artifacts survive, but the meaning that created them does not.

Information Archaeology restores that meaning by:

- reconstructing digital sequences (DSR)
- revealing environmental constraints (DER)
- mapping how practices drifted over time (TDA)
- building evidence-based timelines (DSA)
- recovering the logic behind historical decisions

This reduces rework, supports continuity, and lets organizations understand not just what they did, but how they ended up there.

**12.3 Digital Transformation**

Modernizing with actual historical context.

Most digital transformation efforts begin by assuming the old system is “legacy” without understanding why it looks the way it does.

IA provides that missing history:

- what problem the old system was originally solving
- what pressures shaped the workflow
- how people adapted it over time

- where drift disconnected policy from practice
- what dependencies live below the surface

Modernization becomes dramatically less risky when you aren't guessing about the past.

**12.4 Governance, Compliance, and Audit**

Transparent lineage for decisions and processes.

Regulators increasingly expect:

- documented reasoning
- evidence of due diligence
- traceable authorship
- coherent timelines
- provenance clarity

Most organizations can't provide any of this without scrambling.

Information Archaeology can reconstruct:

- decision lineage
- missing or partial artifacts
- provenance-based explanations
- uncertainty windows
- time-based evolution of workflows
- clean, citeable evidence trails

IA doesn't invent clarity; it retrieves it from the traces.

**12.5 Responsible AI**

Provenance-aware AI grounded in archaeological principles.

AI systems struggle with:

- hallucination
- overinterpretation
- lack of provenance
- opaque reasoning

- narrative drift

IA offers a discipline that AI can borrow:

- evidence-first interpretation
- explicit provenance
- multi-hypothesis reasoning
- uncertainty disclosure
- treating gaps as structure, not flaws
- strict contextual inference rules
- ethical constraints on interpretive scope

The result is AI that becomes more:

- transparent
- traceable
- auditable
- honest about what it cannot know

This is the difference between “AI output” and “responsible AI output.”

**12.6 Enterprise Use Cases**

Practical value across complex environments.

Large enterprises, public agencies, research institutions, and cross-functional teams all face similar challenges: drift, fragmentation, and systems that tell only half their own history.

Information Archaeology helps in several high-value areas:

Postmortems

Tracing how an incident unfolded through:

- reconstructed sequences
- missing layers
- environment shifts
- drift patterns

It produces clarity without assigning motive.

Project Forensics

Understanding:

- why work veered off course
- what assumptions shaped early activity
- where practice diverged from policy
- how constraints influenced outcomes

It reveals structure where memory fails.

System Migrations

Showing:

- undocumented logic
- hidden dependencies
- incremental patch histories
- divergence between intended design and lived use

This reduces risk dramatically.

Mergers & Acquisitions

Recovering:

- institutional memory embedded in traces
- cultural patterns visible in digital behavior
- workflow logic across organizations
- pressure points and alignment opportunities

IA provides a material map of how two organizations actually work.

**12.7 Section Summary**

Information Archaeology provides immediate value anywhere digital traces pile up faster than interpretation. It supports:

- individuals
- teams
- large institutions
- auditors
- governance bodies
- technologists
- AI systems
- historians and archivists

It addresses a simple reality:

Digital life is inherently archaeological. IA gives us the tools, language, and discipline to understand what the traces have been saying all along.

## 13. Authorship and Affiliation

This paper was written by **M. Bartels** in a personal capacity as part of the independent project **InformationArchaeology.org**. It is not affiliated with, endorsed by, or produced on behalf of any current or past employer.

AI tools (OpenAI's ChatGPT) were used to assist with wording, structure, and iterative editing under the author's direction. All concepts, methods, and final interpretive framing are the author's own.