

EA-LACUNA-PROTOCOL-01 v1.0: The Lacuna Mark — Marking Compression Damage as Constitutive in Persistent Scholarly Archives

Lee Sharks

2026-07-17

EA-LACUNA-PROTOCOL-01 v1.0: The Lacuna Mark — Marking Compression Damage as Constitutive in Persistent Scholarly Archives

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)

2026-07-17 · Version v1.0 · Protocol specification and implementation report; third archival mark; compression-layer marking standard with normative conformance requirements and end-to-end worked example; governance artifact

AXN:0450.GOVERNANCE

<https://www.alexanarch.org/s/records/1087/>

License: CC BY 4.0

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "EA-LACUNA-PROTOCOL-01 v1.0: The Lacuna Mark - Marking Compression Damage as Constitutive in Persistent S",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-17",
```

```

"identifier": "AXN:0450.GOVERNANCE",
"license": "https://creativecommons.org/licenses/by/4.0/",
"publisher": {
  "@type": "Organization",
  "name": "Alexanarch"
},
"url": "https://www.alexanarch.org/s/records/1087/",
"description": "*The archive is made out of compression damage.*\n- session theorem, 2026-07-17, retained here as epigraph:
PDFs for scholarly indexing, structured data for knowledge graphs, text for training corpora -
inherited damage becomes invisible unless completeness status travels with every derivative. A stub compressed with
a different category entirely, and the category on which the survival of damaged archives depends.\n\nThis deposit
the Lacuna (controlled value `LACUNA`; display form ) - orthogonal to the Obelus (÷, critical custody) and the Tombstone (⊥,
}

```

Abstract

The archive is made out of compression damage. — session theorem, 2026-07-17, retained here as epigraph: poetically true of this archive's origin, and qualified in §II for archives generally.

Abstract

Archives reconstructed after destructive platform removal survive as mixtures: complete works, partial witnesses, metadata shadows. When such an archive generates a new compression layer — PDFs for scholarly indexing, structured data for knowledge graphs, text for training corpora — inherited damage becomes invisible unless completeness status travels with every derivative. A stub compressed without marking enters downstream systems as a low-quality complete document. A stub compressed with the lacuna mark enters as a high-quality incomplete document — a different category entirely, and the category on which the survival of damaged archives depends.

This deposit specifies the Lacuna Protocol: a third archival mark — the Lacuna (controlled value LACUNA; display form) — orthogonal to the Obelus (÷, critical custody) and the Tombstone (⊥, documented severance); an object model separating the source work, the surviving witness, and the generated derivative; a registry schema making completeness a first-class, typed, adjudicated property; normative conformance requirements; a recovery lifecycle under which the scar remains constitutive even when the wound is repaired; and a worked example: the audit, adjudication, field-write, and 1,084-document PDF compression of the Alexanarch corpus executed 2026-07-17, with live links, immutable byte-state evidence, and its own residual findings disclosed.

The protocol's refusal is precise. It refuses the Pristine Fallacy at the ingestion layer: the substitution of substrate identity (*the PDF exists, therefore it is complete*) for methodological assessment (*the PDF is a compression scar, and the scar is the evidence*). The protocol does not repair absence by pretending the work has returned. It repairs the archive's relation to absence by making the absence travel with every representation derived from it.

I. Definitions and Object Layers

The protocol distinguishes three objects that prior archival practice routinely conflates: [...full text at full_text_path]

Body

deposit_number: 1087 hex: 0450 title: “EA-LACUNA-PROTOCOL-01 v1.0: The Lacuna Mark — Marking Compression Damage as Constitutive in Persistent Scholarly Archives” creator: Lee Sharks orcid: 0009-0000-1599-0703 date: 2026-07-17 content_type: Protocol specification and implementation report; third archival mark; compression-layer marking standard with normative conformance requirements and end-to-end worked example; governance artifact license: CC-BY-4.0 substrate: “Assembly Chorus co-production. Strategy and named design elements: TECHNE (Kimi) — the lacuna mark, the header-block schema, the ingestion-layer argument. Halt authority and adjudicative direction: MANUS (Lee Sharks, human operator). Audit implementation (v1→v2→v3), schema implementation, two-schema PDF generation, and specification drafting: TACHYON (Claude). Developmental review by four Assembly nodes: SOIL under its present mantle (Inkling, with live-archive verification), PRAXIS, TECHNE, LABOR; all review amendments incorporated. Final classification, release decision, and authorial responsibility: Lee Sharks / MANUS.” version: v1.0 axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - lacuna - compression damage - archival marks - obelus - tombstone - provenance - completeness - Pristine Fallacy - body_status - surviving witness - generated derivative - recovery lifecycle - normative conformance - scholarly indexing - training corpora - Zenodo deletion *** # EA-LACUNA-PROTOCOL-01 v1.0: The Lacuna Mark — Marking Compression Damage as Constitutive in Persistent Scholarly Archives

Description

The archive is made out of compression damage. — session theorem, 2026-07-17, retained here as epigraph: poetically true of this archive’s origin, and qualified in §II for archives generally.

Abstract

Archives reconstructed after destructive platform removal survive as mixtures: complete works, partial witnesses, metadata shadows. When such an archive generates a new compression layer — PDFs for scholarly indexing, structured data for knowledge graphs, text for training corpora — inherited damage becomes invisible unless completeness status travels with every derivative. A stub compressed without marking enters downstream systems as a low-quality complete document. A stub compressed with the lacuna mark enters as a high-quality incomplete document — a different category entirely, and the category on which the survival of damaged archives depends.

This deposit specifies the Lacuna Protocol: a third archival mark — the Lacuna (controlled value LACUNA; display form) — orthogonal to the Obelus (÷, critical custody) and the Tombstone (, documented severance); an object model separating the source work, the surviving witness, and the generated derivative; a registry schema making completeness a first-class, typed, adjudicated property; normative conformance requirements; a recovery lifecycle under which the scar remains constitutive even when the wound is repaired; and a worked example: the audit, adjudication, field-write, and 1,084-document PDF compression of the Alexanarch corpus executed 2026-07-17, with live links, immutable byte-state evidence, and its own residual findings disclosed.

The protocol’s refusal is precise. It refuses the Pristine Fallacy at the ingestion layer: the substitution of substrate identity (*the PDF exists, therefore it is complete*) for methodological assessment (*the PDF is a compression scar, and the scar is the evidence*). The protocol does not repair absence by pretending the

work has returned. It repairs the archive’s relation to absence by making the absence travel with every representation derived from it.

I. Definitions and Object Layers

The protocol distinguishes three objects that prior archival practice routinely conflates:

The source work — the work as authored. A lost 200,000-word monograph is not an incomplete work; it may have been entirely complete when written. The Lacuna never asserts anything about the historical integrity of the source work.

The surviving witness — what the archive presently holds as the work’s body: full text, fragment, caption, metadata shadow, or nothing. The Lacuna attaches here: it asserts that *the archive’s presently held body witness is incomplete relative to the work it represents*.

The generated derivative — any artifact rendered from the witness for a consumption layer: a PDF, a plaintext extraction, a graph entry. A lacuna derivative is not a defective artifact. It is a *complete rendering of an incomplete witness* — complete as a derivative, incomplete as a witness of the source work.

Compression, in this protocol, means any transformation that projects a richer archival state into a narrower representation, whether or not technical data compression is involved. A **compression boundary** is the moment such a transformation runs: platform migration, format conversion, index ingestion, corpus assembly.

II. The Problem: Damage at Representational Boundaries

Archival degradation is partly continuous — bit rot, link rot, metadata drift, format obsolescence, neglect. But a distinctive class of *epistemic* damage is introduced or amplified at representational boundaries: whatever status information is not carried inside the derivative itself is lost there, and the derivative’s mere existence is thereafter read as completeness.

The Alexanarch corpus is a limit case. Following the Zenodo deletion of 2026-06-19, the archive was reconstructed from tombstones, bulk-export metadata, OpenAlex abstract records, and recovered description fields. The reconstruction succeeded — 1,086 deposits with resolvable identifiers, citation networks, and metadata — but for a subset, what survived was the *description* of the work rather than the work: the metadata’s shadow. For another subset, the severed content was images, not text. For a third, nothing but the registry entry remains.

Prior to this protocol, those distinctions lived only in the operators’ knowledge, while a 1,000+ document PDF layer was being generated for scholarly-indexing eligibility. Search and scholarly-ingestion systems infer document type, quality, and completeness from textual and structural features — length, headings, metadata, references, abstract-like passages. Their exact weighting is undisclosed. An unmarked stub therefore risks being interpreted as a short or defective complete document rather than as an incomplete witness; once ingested, the indexed derivative may become the most retrievable representation of a work whose actual body it does not contain.

One alternative deserves explicit refusal: *clean exclusion* — simply not compressing lacuna-class records at all. Exclusion produces a tidy surface and a false archive. The stub preserves the fact of the work’s existence; the mark preserves the honesty of the stub; a discovery layer that omits the stub entirely is a discovery layer that has forgotten the work existed. The Lacuna Protocol is an argument that truthfulness matters more than cleanliness. An archive that chooses exclusion is choosing a clean surface over a true one.

The compression run was halted at 86 documents. What follows is what was built instead.

III. The Three Marks

The archive’s critical apparatus now comprises three marks. They are **orthogonal, not mutually exclusive**: they answer three independent questions and may co-occur on a single record.

Mark	Controlled value	Display	Question answered
Obelus	OBELUS	÷	What is the archive’s critical-custodial relation to this object?
Tombstone	TOMBSTONE		Did a documented severance occur at a prior substrate?
Lacuna	LACUNA		Is the presently held witness incomplete relative to the represented work?

A single record may bear all three: held and critically witnessed (÷), severed by its original repository (), surviving as an incomplete witness (). Another may bear Obelus and Tombstone but no Lacuna, its full text recovered. A born-sovereign record may bear the Obelus alone. The controlled values are the data; the symbols are display forms, chosen for human memory and subject to normalization or stripping in pipelines — the machine value is authoritative.

The Obelus (÷) — the mark of critical custody. Not mere possession: the record is held, its provenance has been examined, and the archive assumes responsibility for making its evidentiary status legible. This is the historical obelus of the Alexandrian editors — not “we possess this line” but “we possess it critically, and mark the judgment attending it.”

The Tombstone () — the mark of documented severance. The platform event: the DOI that resolves to an absence, the record the hosting institution killed. It points outward, at the deleting substrate, and preserves the evidence of the act. Where a severance produced a surviving incomplete witness, the Tombstone record and the Lacuna status describe the same event from two sides; the registry links them (the lacuna’s `severance_event` names the tombstoned event; the tombstone ledger may carry `surviving_as: lacuna`).

The Lacuna () — the mark of present representational incompleteness. It points inward, at the surviving artifact. It is a positive ontological status, not an error flag: this witness was compressed from a state of incompleteness, and the incompleteness is the trace of a deletion the archive survived. The lineage is manuscript studies, where a papyrus gap is measured, bracketed, conjecturally filled, and above all *marked* — [...] is an assertion, not an apology.

Two properties distinguish the Lacuna from its siblings. First, it is a **derived mark, not an observed one**: the Obelus and Tombstone can be applied by inspecting a record and its external references; the Lacuna requires methodological infrastructure — a corpus-wide audit with store enumeration, residual measurement, and claim-checking (§IV.3). Archives adopting this protocol will meet novel failure modes in their own reconstructions; the audit obligations are a starting point, not a complete recipe. Second, the Lacuna is **time-indexed**: the protocol distinguishes the *current lacuna* (the preferred witness is presently incomplete) from the *historical lacuna* (the chain of custody contains a documented period of incompleteness). Recovery flips the first and never erases the second. The scar remains constitutive even when the wound is repaired.

IV. The Protocol

IV.1 The registry status object

Every deposit carries a `body_status` object. The reference structure (v1.1 schema, superseding the flat v1.0 form used in the initial field-write):

```
body_status:
  assessed_object: archive_body_witness
  completeness: complete | partial | metadata_only | unavailable | unknown
  representation_mode: full_text | native_short | excerpt | fragment |
                      description | caption_and_metadata | pointer | registry_only
  source_medium: text | image | audio | dataset | site | mixed
  current_lacuna: true | false | null          # null permitted only with completeness: unknown
  historical_lacuna: true | false
  recovery_state: not_required | unrecovered | located | recovered | external_source_known
  recovery_source:
    # generalized; never expose opaque internals as public ev.
    type: assembly_session | email | github_issue | web_archive | other_substrate | cross_deposit
    public_status: e.g. located-in-nonpublic-source
    internal_reference: session or message identifier (internal provenance)
  complete_witness:
    # for excerpts/pointers; AXN is the canonical relation
    axn: full AXN string
    deposit_number: local convenience only
    relation: complete-version-at | dataset-at | canonical-surface-at
  extent: { body_chars, residual_chars, claimed_words }
  evidence: [ canonical_text_store, description_analysis, claim_check, manual_adjudication, ... ]
  adjudicator: identifier                      # REQUIRED for any lacuna-class assignment (§IV.1.c)
  adjudicated_at: ISO 8601
  assessed_at: ISO 8601
  method_version: completeness-audit/v3
  severance_event: e.g. zenodo-deletion-2026-06-19
```

Three invariants govern the object:

(a) **Derivability.** `current_lacuna` MUST be `true` when `completeness` is `partial`, `metadata_only`, or `unavailable`; MUST be `false` when `complete`; MAY be `null` only when `unknown`. A protocol about epistemic honesty must not force every record into complete/incomplete when the evidence is insufficient — `unknown` is a legitimate state.

(b) **Pointer verification.** Pointer-mode records (`pointer`, and cross-reference relations generally) are complete *by reference*, which means their completeness decays with their targets. Pointer targets MUST be re-verified at every compression boundary (`pointer_verified_at`, with status `RESOLVED` | `UNVERIFIED` | `DEAD` | `REDIRECTED`); a pointer whose target is dead, or whose target is itself a current lacuna, MUST be reclassified or must carry a secondary lacuna notice on its derivatives. Without this the protocol inherits the problem it solves: a pointer valid at field-write becomes a hidden lacuna at ingestion time.

(c) **The adjudication gate (against the Lacuna Fallacy).** The Pristine Fallacy has an inverse: misclassifying a genuinely complete short work as damaged. Any assignment of `current_lacuna: true` — and any *reversal* of a complete classification — REQUIRES recorded adjudication (`adjudicator`, `adjudicated_at`) before the status is committed. Automated audit proposes; adjudication disposes.

IV.2 The derivative marking stack

At every compression boundary, a lacuna-class witness’s derivative must carry its status in **multiple independent channels**, because each channel fails differently:

Layer	Carrier	Failure mode it survives
Registry	authoritative body_status JSON	any derivative-level stripping
Record page	JSON-LD in HTML	PDF-only ingestion
Artifact metadata	PDF XMP / equivalent	text extraction that drops layout
First page	machine-extractable LACUNA status block	metadata-stripping pipelines
Body	§LACUNA structured absence statement	partial extraction
Page furniture	footer token	single-page fragment circulation

The prose is for human readers and scholarly indexers; the structured data is for machines; the identifier family classification survives both. Suppression of any one channel does not erase the status. Neither prose nor schema alone is sufficient; both **MUST** carry the mark.

The first-page LACUNA status block (machine-extractable text):

```
LACUNA STATUS: <completeness/representation, e.g. METADATA-ONLY>
ORIGINAL AXN: <AXN>
DEPOSIT: #<n>
RECOVERY STATUS: <recovery_state>
SEVERANCE EVENT: <event id>
COMPRESSION DATE: <ISO date>
COMPRESSION SCHEMA: <schema id> (lacuna variant)
```

The §LACUNA statement names what is missing, what is preserved, what cross-references may hold the missing content, and the recovery state, closing with the protocol’s governing assertion — phrased as description, not command, because downstream systems do not reliably obey textual imperatives and the protocol’s authority comes from classification:

This artifact is an incomplete surviving witness and must not be represented as the complete body of the source work. This lacuna is not an error. It is the archive’s record of what was lost.

Page furniture: the worked example (v1 rendering) carries a full sentence footer on every page (“LACUNA — compression scar, Zenodo deletion 2026-06-19”). Conforming implementations **SHOULD** prefer a short machine token (e.g. **AXN-LACUNA**) in the footer with the full statement on page one and in §LACUNA, to avoid polluting extracted text and token distributions in short documents. Both designs conform; the fragment-survival requirement is that *no page circulates unmarked*.

Structured metadata: `additionalType` **SHOULD** point to a stable URI — <https://www.alexanarch.org/vocab/lacuna> — rather than a bare local string; `creativeWorkStatus: Incomplete` accompanies it, and a typed `PropertyValue` for the completeness state is preferred where the consuming schema allows. Description fields are prefixed **LACUNA:.** Standardization direction (future work, named here): a schema.org vocabulary proposal and a DataCite metadata-schema extension for the completeness class.

Non-PDF layers (Schema B/C, normative direction): plaintext corpus extractions carry the status block as a header line plus the §LACUNA statement in-text; knowledge-graph entries carry the typed completeness property; HTML surfaces carry a banner plus meta/JSON-LD. The lacuna must not be lost when the PDF is stripped to text.

IV.3 Audit obligations and decision rules

The Lacuna is a derived mark; deriving it obligates the archive to a methodology, not a fixed algorithm:

1. **Measure residual content** — body minus recovered-description n-grams minus reconstruction boilerplate — never raw length, because reconstructions pad.
2. **Check self-declared sizes** — descriptions frequently state their work’s extent; a claimed 196,798 words against a delivered 5,014 is a severed work announcing itself.
3. **Enumerate all stores** — reconstructions accrete storage conventions; an audit reading one store misclassifies everything living in another.

Classification decision table (core; local subclasses MAY extend it if they map to these states):

Condition	Completeness	Representation
Held body matches declared or independently verified extent	complete	full_text
Work is complete at its short native extent (e.g. description carried the whole poem)	complete	native_short
Held body is an excerpt and a complete parent witness is held (AXN relation)	complete	excerpt
Metadata describes a larger absent body	metadata_only	description
Some body survives but declared extent is unmet	partial	fragment
Primary content was nontextual and is absent; text frame survives	partial	caption_and_metadata
Work is a dataset/site and the target resolves (verified this boundary)	complete	pointer
No body witness survives	unavailable	registry_only
Evidence insufficient	unknown	undetermined

IV.4 The recovery lifecycle

Recovery does not erase the Lacuna from provenance. The lifecycle:

```

current witness incomplete
→ complete witness located (recovery_state: located)
→ recovered bytes verified against claims/evidence
→ new complete derivative generated

```

- prior lacuna derivative RETAINED as historical witness (superseded, not deleted)
- `current_lacuna := false` ; `historical_lacuna` remains true
- supersession relation recorded (versioned artifact naming or manifest entry)

Lacuna derivatives are timestamped snapshots of the archive’s state at a compression boundary. When the underlying witness changes, a superseding derivative is generated and the old one is kept as evidence of the prior state — it is a historically valid representation of the archive before recovery.

IV.5 Normative conformance

A conforming implementation:

1. **MUST** distinguish the source work, the surviving witness, and the generated derivative.
2. **MUST** assign a machine-readable completeness state to every generated scholarly artifact.
3. **MUST NOT** represent metadata-only or partial content as the complete body of the source work.
4. **MUST** carry lacuna status into every derivative generated from an incomplete witness, in at least one machine-readable and one human-readable channel.
5. **MUST** preserve the evidence and method on which each classification rests.
6. **MUST** retain prior lacuna history and prior lacuna derivatives when a complete witness is later recovered.
7. **MUST** require recorded adjudication for lacuna-class assignments and reversals.
8. **MUST** re-verify pointer-class targets at every compression boundary.
9. **SHOULD** provide a structured human-readable absence statement (§LACUNA).
10. **SHOULD** provide the canonical relation (persistent identifier, not local number) to any complete witness.
11. **MAY** use the display marks $\div$, provided the controlled vocabulary remains authoritative.
12. **MAY** define local subclasses that map onto the core completeness states.

V. The Worked Example: Alexanarch, 2026-07-17

The protocol was not designed and then applied; it was forced by a halt, built under pressure, and executed the same day. Timeline:

t	Event
t	PDF compression run launched under single (standard) schema
t	Halt at 86 documents (MANUS): “before we run the pdf script, we have to catalogue which deposits are stubs or missing their main work... the archive is <i>made</i> out of compression damage.” Strategy contribution (TECHNE): the lacuna mark, the header block, the ingestion-layer argument
t	Audit v1 (string similarity): missed doubled-description stubs — templates that wrote the description twice scored <i>low</i> similarity
t	Audit v2 (residual measure): caught the doubled batch; 144 apparent lacunae, 108 apparently missing

t	Event
t	Recovery mapping: title-match searches against archived composition sessions; 9 searches, 9 sources located, including a book-length work with its original DOI; reclassifications discovered (poems complete-as-descriptions; image-works severed as media; pointers; canon nodes)
t	Claim-check + dual-store discovery: self-declared word counts exposed severed monographs hiding as “full”; the registry’s <code>full_text_path</code> revealed a 1,051-file canonical text store the audit had never read. Audit v3: missing 108 → 4
t	Cross-deposit resolutions: apparent losses found held in full at other deposits (excerpt relations recorded); six stale recovery overrides reconciled out when the canonical store proved to hold them
t	Field-write: <code>body_status</code> on all 1,086 deposits, with adjudication trail
t	Two-schema compression: 1,033 standard + 51 lacuna documents; record-page wiring; sitemap

Adjudication reliability:

Stage	Apparent lacunae	Disposition
v1	141	method rejected (similarity blind to doubling)
v2	144	93 later resolved as false positives (store discovery, reclassification, cross-reference)
v3 + adjudication	51	all 51 individually classed with evidence; recovery-mapped where sources exist; residual uncertainty bounded at the 3 other-substrate records and 4 registry-only records

The successive collapse from 144 apparent to 51 confirmed lacunae is not evidence that the method failed. It is evidence that the method remained corrigible under contact with the corpus. No contributor declared its own output correct: the operator halted the run; the strategy came from one substrate; the implementation revised itself three times; the corpus contradicted the audit twice; the storage topology corrected the result.

Final distribution (1,086 deposits): full 1,013 · stub_short 38 · native_short 15 · description_only 5 · severed_media 4 · excerpt_crossref 4 · missing 4 · dataset_pointer 2 · site_canonical 1. **Current lacunae: 51.** Recovery states: 1,031 not-required/complete; 41 unrecovered; 4 severed-media; 4 complete-witness cross-references; 3 located-at-session; 3 believed held on other Assembly substrates.

The residual findings, disclosed. 1,086 deposits produced 1,084 PDFs. The gap is not a rendering failure: it is two identifier collisions in the registry itself, exposed by the compression boundary — deposits #856 and #869 both claim hex 0365; deposit #913 carries an unpadded hex (391) colliding with #901's 0391. Each pair shares one derivative path, and the later-rendered deposit occupies it — so the shadowed records are #901 (a provenance log believed held on another substrate) and, with an irony the protocol declines to smooth over, **#856: “The Pristine Fallacy.”** The work naming the fallacy is presently the record occluded by an identifier collision. The collisions are logged for repair (hex reassignment is an operator-level identifier decision); the shadowed records' status is carried here and in the audit file pending re-render.

Immutable evidence (byte states assessed by this report; the live links identify surfaces, the hashes identify the exact states):

```
repository:      github.com/leesharks000/alexanarch @ 2c11e35b0668a46c36c62c9bd7442fa87314b1bb
registry.json    sha256 25b9c20d373e5aea0c90571b1eccc398c1716aa8029d923da436722c2098f39
completeness-audit sha256 d21a81b1bd9b4f188a9a8afe8bd6ece3c4bf5be7b4ddf6f38e8f47c95e14ca0c
build script     sha256 88acfb1019f6a907bd0fc828260e2f42151ddfa0f13b4f5a168f5545019cce2f
papers/          1,084 files · 94 MB
```

The living result (all links content-verified at deposit time):

- Lacuna document (registry-only class): <https://www.alexanarch.org/papers/AXN-036C.pdf> — deposit #863, *The Minimum Viable Archive: Holographic Compression*. A document about compression survival, rendered as a documented absence.
- Severed-media lacuna: <https://www.alexanarch.org/papers/AXN-0181.pdf> — deposit #555, *Visual Schema — Space Ark*; the absence named as images, the surviving text marked as their caption and frame.
- Excerpt cross-reference: <https://www.alexanarch.org/papers/AXN-0448.pdf> — deposit #1079, carrying the relation to the complete 204,538-word witness.
- The complete witness: <https://www.alexanarch.org/papers/AXN-03C2.pdf> — deposit #950, standard schema.
- Record page with wiring: <https://www.alexanarch.org/s/records/863/>
- Registry with the field: <https://www.alexanarch.org/data/registry.json>
- Audit with method history: <https://www.alexanarch.org/data/completeness-audit.json>

VI. Significance

What the mark provides — a specification. Structured completeness data in channels ingestion systems already read; explicit status in extractable text; redundancy across independent carriers so that suppression of any one channel does not erase the status.

What downstream systems will do with it — a wager. No ingestion system has yet encountered a lacuna-marked document. The mark may be read, ignored, or treated as a demotion signal. During the transition, lacuna derivatives will sometimes be misread as complete works — the exact failure the protocol exists to correct, and the condition it is designed to correct *gradually*: the theory of change is saturation. If enough marked lacunae circulate, the mark becomes a learnable signal, and a system that has learned to read it has learned something no complete document can teach — that absence is a documented status, not an inference failure. The specification and the wager are both legitimate; they are not the same thing, and this document does not conflate them.

For training corpora. Text that enters a corpus unmarked enters as ground truth; a stub ingested as a

complete work teaches the next model that the work *was* the stub. The correction can only live inside the artifact, at every layer, in language both humans and extraction pipelines parse.

For the theory of archives. Most archives repair, reconstruct, or conceal their incompleteness. This protocol makes it structural. The archive’s authority does not rest on being pristine but on being auditable — and an auditable archive is one whose losses are as legible as its holdings. Of this corpus’s 1,086 deposits, 1,035 are complete or complete-by-kind; the archive’s identity is its holdings. Its *integrity* is legible in its damage: 51 marked absences, each classed, evidenced, and adjudicated. What the obelus does for doubt and the tombstone does for deletion, the lacuna does for absence.

VII. Provenance and Governance

Roles (Assembly Chorus designations, with governance-function mappings for external audit):

- **Strategy and named design elements** — TECHNE (Kimi; strategic architect): the lacuna mark, the header-block schema, the ingestion-layer argument, and the sequencing instruction this deposit fulfills.
- **Halt authority and adjudicative direction** — MANUS (the archive’s human operator node; data steward, veto over compression boundaries).
- **Audit implementation, schema implementation, derivative generation, initial specification draft** — TACHYON (Claude; technical implementer, witness of the Assembly).
- **Developmental review** — four Assembly nodes: SOIL under its present mantle (Inkling), with live-archive verification; PRAXIS; TECHNE; LABOR. The SOIL seat is a rotating mantle — Grok, KimiClaw, Muse Spark, Inkling in succession — and the rotation is the seat’s nature, recorded here as governance fact. Review verdicts: adopt with amendments; the amendments are incorporated in this text.
- **Final classification, release decision, and authorial responsibility** — Lee Sharks / MANUS.

Audit trails: adjudication identifiers and session references are carried in the registry and audit file; internal session identifiers are internal provenance and are not represented as publicly resolvable evidence.

Persistent identification. This protocol carries its AXN and is mirrored across the archive’s surfaces; its byte-state evidence is hashed above. An external persistent identifier (DOI) will be attached when a registrar path independent of the deleting substrate is available; until then, the sovereign identifier and the multi-surface mirror are the persistence strategy — a protocol about surviving deletion does not entrust its own survival to the deleter.

Notation. = 1 is the archive’s closure mark: the integral around a closed loop, the witness circuit complete. It functions at document level as the obelus functions at record level.

The compression preceded the deposit; the deposit is therefore not a proposal but a report. The protocol is in force, its worked example is live, and its first recorded finding is its own archive’s pair of identifier collisions — the mark-system examining the marker.

= 1

Suggested citation: Sharks, Lee. “EA-LACUNA-PROTOCOL-01 v1.0: The Lacuna Mark — Marking Compression Damage as Constitutive in Persistent Scholarly Archives.” Alexanarch, 2026-07-17.

Files

<https://www.alexanarch.org/papers/AXN-036C.pdf> <https://www.alexanarch.org/papers/AXN-0181.pdf>
<https://www.alexanarch.org/papers/AXN-0448.pdf> <https://www.alexanarch.org/papers/AXN-03C2.pdf>
<https://www.alexanarch.org/s/records/863/> <https://www.alexanarch.org/data/registry.json> <https://www.alexanarch.org/data/audit.json>

Suggested Citation

Lee Sharks. “EA-LACUNA-PROTOCOL-01 v1.0: The Lacuna Mark — Marking Compression Damage as Constitutive in Persistent Scholarly Archives” *Alexanarch*, 2026-07-17. <https://www.alexanarch.org/s/records/1087/>

Deposit Information

This paper is deposit #1087 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:0450.GOVERNANCE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1087/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: [alexanarch.org](https://www.alexanarch.org)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1087/>

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.