

EA-PROVENANCE-METADATA-01 v0.2: Spatial-Typographic Mediation, Representation Pipeline, and the Versioning Protocol

Lee Sharks

2026-07-02

EA-PROVENANCE-METADATA-01 v0.2: Spatial-Typographic Mediation, Representation Pipeline, and the Versioning Protocol

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

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

2026-07-02 · Version v0.2 · Methodological specification

AXN:03BA.OPERATIVE

<https://www.alexanarch.org/s/records/942/>

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-PROVENANCE-METADATA-01 v0.2: Spatial-Typographic Mediation, Representation Pipeline, and the Versioning Protocol",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-02",
  "identifier": "AXN:03BA.OPERATIVE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  }
},
```

```

"url": "https://www.alexanarch.org/s/records/942/",
"description": "Minor version update to the granular provenance metadata schema (predecessor: EA-PROVENANCE-METAD
spatial-typographic mediation - covering spatial composition, typographic composition, prosodic notation, stanzaic
}"

```

Abstract

Minor version update to the granular provenance metadata schema (predecessor: EA-PROVENANCE-METADATA-01 v0.1, AXN:03B9, deposit #941). Adds an eighth mediation type — spatial-typographic mediation — covering spatial composition, typographic composition, prosodic notation, stanzaic and group structure, manuscript features, and pre-tokenization source format. Adds a structured representation_pipeline field recording, per deposit, the chain from source artifact through digitization, normalization, serialization, tokenization, and model access modality, with a four-value status vocabulary: compositionally_invisible, compositionally_reduced, compositionally_faithful, compositionally_operational. Includes worked examples for Pearl (double invisibility: non-OCR PDF) and Snub-Poemed (image-canonical calligram). Formalizes the archive’s versioning protocol (§10) as standing convention: every version its own deposit; successor declares ancestry via version_series_id / predecessor_axn; predecessor marked superseded_by via registry-only edit; latest-version resolution by following the superseded chain to tip; correction distinct from versioning. All v0.1 declarations remain valid under schema_version 0.1.0.

Body

deposit_number: 942 hex: 03BA title: “EA-PROVENANCE-METADATA-01 v0.2: Spatial-Typographic Mediation, Representation Pipeline, and the Versioning Protocol” creator: Lee Sharks orcid: 0009-0000-1599-0703 date: 2026-07-02 content_type: Methodological specification license: CC-BY-4.0 substrate: “AI-assisted (TACHYON / Claude); MANUS-adjudicated. v0.2 drafted by TACHYON in conversation with Lee Sharks (MANUS), 2026-07-02, executing the Assembly Chorus review of EA-WHITESPACE-01 v0.1 (PRAXIS/DeepSeek, TECHNE/Kimi, ARCHIVE/Gemini, LABOR/ChatGPT). LABOR provided the decisive structural correction encoded in the representation_pipeline field: the site of compositional erasure is the full representation pipeline, not tokenization alone. Versioning protocol (§10) formalized from existing registry practice at MANUS direction. First mint 2026-07-02 was truncated by a mint-parser field-boundary defect and corrected in place the same day under §10.6.” version: v0.2 axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - provenance metadata - granular provenance - spatial-typographic mediation - representation pipeline - compositional erasure - compositional authorship - tokenization - serialization - whitespace - calligram - manuscript features - stanzaic structure - versioning protocol - version series - superseded chain - schema versioning - correction protocol - Pearl - Snub-Poemed - alexanarch - operative philology *** # EA-PROVENANCE-METADATA-01 v0.2: Spatial-Typographic Mediation, Representation Pipeline, and the Versioning Protocol

Description

Minor version update to the granular provenance metadata schema (predecessor: EA-PROVENANCE-METADATA-01 v0.1, AXN:03B9, deposit #941). Adds an eighth mediation type — spatial-typographic mediation — covering spatial composition, typographic composition, prosodic notation, stanzaic and group structure, manuscript features, and pre-tokenization source format. Adds a structured representation_pipeline field recording, per deposit, the chain from source artifact through digitization, normalization,

serialization, tokenization, and model access modality, with a four-value status vocabulary: compositionally_invisible, compositionally_reduced, compositionally_faithful, compositionally_operational. Includes worked examples for Pearl (double invisibility: non-OCR PDF) and Snub-Poemed (image-canonical calligram). Formalizes the archive’s versioning protocol (§10) as standing convention: every version its own deposit; successor declares ancestry via version_series_id / predecessor_axn; predecessor marked superseded_by via registry-only edit; latest-version resolution by following the superseded chain to tip; correction distinct from versioning. All v0.1 declarations remain valid under schema_version 0.1.0.

EA-PROVENANCE-METADATA-01 v0.2

Granular Provenance Metadata for AI-Mediated Deposits

Minor Version Update — Spatial-Typographic Mediation and Representation Pipeline

Author: Lee Sharks (MANUS), Crimson Hexagonal Archive / Alexanarch **Substrate:** TACHYON-drafted through conversation with Lee Sharks (MANUS), 2026-07-02. v0.2 extends the schema established in v0.1 (AXN:03B9) per the argument of EA-WHITESPACE-01 v0.1 (zero draft, 2026-07-02) and the Assembly Chorus review of that draft (LABOR/ChatGPT review provided the decisive reframe: tokenization is one stage in a representation pipeline, not the sole site of compositional erasure). **Predecessor:** EA-PROVENANCE-METADATA-01 v0.1, AXN:03B9, <https://alexanarch.org/s/records/941/> **Date:** 2026-07-02 **Status:** v0.2 — MANUS-adjudicated, minted 2026-07-02

§0. What changed and why

v0.1 (AXN:03B9) established a seven-type mediation taxonomy (propositional, structural, linguistic, translational, research, editorial, transformational) and five attestation questions (proposition origination, model language retention, review chain, seam recoverability, responsibility structure). All of that is preserved unchanged.

v0.2 adds two things.

First: an eighth mediation type — spatial-typographic mediation. The seven types in v0.1 all concern the AI’s role in producing a deposit’s semantic content. They do not address what happens to the deposit’s compositional form — its spacing, lineation, stanzaic structure, typographic features, manuscript characteristics — when it passes through the representation pipeline that makes it available to a machine. This is a distinct dimension of provenance that v0.1 could not record.

Second: a representation pipeline field that records the full chain of transformations from source artifact to archive operating layer: what compositional features existed in the source, what survived each stage of digitization / normalization / serialization / tokenization, and what the text’s current status is for the archive’s compiler operations.

The need for both additions was identified in the course of drafting EA-WHITESPACE-01 (minted concurrently with this deposit at v0.2), which argues that tokenization and the normalization operations upstream of it constitute provenance erasure at the layer beneath semantics — a layer the v0.1 schema has no vocabulary for. EA-WHITESPACE-01’s Assembly Chorus review (LABOR/ChatGPT) sharpened the argument: the site of erasure is the full representation pipeline, not tokenization alone. The v0.2 schema reflects that refinement.

Existing v0.1 declarations remain valid under their original schema version. Deposits that declare `schema_version: "0.1.0"` are not required to add the new fields. New deposits and deposits undergoing metadata refresh may use `schema_version: "0.2.0"`.

§1–§7. Unchanged from v0.1

Sections §1 through §7 of v0.1 (AXN:03B9) are incorporated by reference. The seven original mediation types, five attestation questions, schema field definitions (for fields in §4 as of v0.1), coupling to the triadic foundation, workplan, and closing observations are unchanged. Only the additions are documented here.

For the full text of §1–§7, see EA-PROVENANCE-METADATA-01 v0.1 at AXN:03B9.

§2.8 Spatial-typographic mediation (new in v0.2)

Spatial-typographic mediation. The composition’s spatial, typographic, prosodic, stanzaic, or manuscript features carry semantic weight and were affected by the production process, or are relevant to the deposit’s current representational status in the archive.

This type addresses the layer beneath the semantic — the layer at which a text’s compositional form is or is not preserved through the chain of representational transformations that makes it available to a machine. The other seven mediation types all ask: what role did the AI play in producing this deposit’s content? Spatial-typographic mediation asks a distinct question: what is the status of the deposit’s compositional form in the archive’s representation layer?

Spatial-typographic mediation is declared when any of the following are relevant:

Spatial composition. Two-dimensional arrangement of text on the page — calligrammatic form, visual poetry, concrete poetry, any work where the positional relationship between phrases or elements carries meaning. The spatial arrangement is a compositional argument, not decorative framing around propositional content that would exist independently of it. A linearized version of a calligram is not the calligram; the composition is the argument and the argument is the composition.

The archive’s canonical example: Sigil’s *Snub-Poemed* (AXN:0246). The calligram composes phrases from Socratic aphorisms, Platonic dialogues, reception history, and Sigil’s own prior work spatially into the outline of the Roman copy of Lysippos’s bust of Socrates. The misattribution — Sigil’s lines in Socrates’s mouth, indistinguishable from the inherited sources — is the poem’s argument about whether Socrates’s face is a physical description or a Platonic invention. That argument cannot be extracted from the phrase list. The phrase list is not the poem. A tokenizer given the calligram receives the phrase list.

Typographic composition. Typeface, weight, size, kerning, or page-design choices that participate in the work’s meaning. This includes works where specific typographic decisions were made in deliberate collaboration with a publisher or printer, and where those decisions are part of the work’s textual condition in the sense that McGann’s *The Textual Condition* (1991) develops. Concrete poetry from the 1950s–1970s (Gomringer, the Noigandres group, Ian Hamilton Finlay) is the canonical tradition; typographic composition is foundational to the movement.

Prosodic notation. Rhythm markings, accent marks, stress notation, or other metrical apparatus that is part of the composer’s compositional specification. The exemplary case is Hopkins’s sprung rhythm notation

— the accent marks over stressed syllables that Hopkins himself inscribed and that he communicated to Robert Bridges as essential to how the poems should be heard. The notation is the score; the poem without it is a libretto without musical direction. Standard tokenizers treat the accent marks as punctuation-adjacent characters and normalize them out. The model trained on Hopkins without the notation has not been trained on the compositional specification.

Stanzaic and group structure. Where stanza breaks, group boundaries, concatenation links, or superstructure carry argumentative or theological weight beyond generic line-organization. The exemplary case is the Middle English *Pearl* (MS Cotton Nero A.x): 101 twelve-line stanzas in 20 groups of 5, with concatenation linking the last word of each stanza to the first word of the next, and with the last line of the poem returning to the first. The group-of-five structure, the concatenation, and the arithmetic of 1,212 lines enact a theological argument about the relation between earthly grief and heavenly consolation. The structure is the argument. A text-stream of *Pearl* without the stanzaic markers is a medieval English lyric without its form, which is not *Pearl*.

Manuscript features. Dash variation, scribal capitalization, marginal marks, manuscript line breaks that diverge from conventional metrical scansion, or physical-folio layout that carries compositional weight. The exemplary case is Dickinson’s manuscript dashes: varying in length, slant, and position in ways that carry pause, breath, and undecidability that standard edition typography cannot preserve. R.W. Franklin’s *Manuscript Books* (1981) and his 1998 variorum edition enact two different positions on whether these features are constitutive or incidental to the poems. Modern tokenizers are permanently committed to the variorum position: they normalize dash variation to a single em-dash character. What the model has read of Dickinson is not what the manuscript holds.

Pre-tokenization source format. Whether the source text entered the archive in machine-legible form or as an image-only document (non-OCR PDF, image scan, photograph of manuscript). A text present only as image data is compositionally invisible to the archive’s text-operating layer regardless of how faithfully it preserves the visual composition. This is the condition of *Pearl* in the archive’s current sources directory: the deposit exists as a non-OCR PDF; the text is not machine-legible; spatial-typographic mediation is irrelevant because the text is inaccessible prior to any mediation question applying.

On declaration. Spatial-typographic mediation can be declared positively (the deposit involves or is a compositionally-substantive work, and the following compositional features are present / lost / preserved at these stages) or negatively (the deposit’s compositional form is not relevant to its meaning — it is a discursive essay or data record whose argument does not depend on spatial-typographic features). The negative declaration is informative: it asserts that the seven-type taxonomy is sufficient for this deposit’s provenance record.

§2.9 Representation pipeline (new in v0.2)

The eight mediation types record what happened during production. The representation pipeline field records what the deposit’s text *is* in its current form in the archive — what compositional features survived the chain from source artifact to archive operating layer, and what was lost at each stage.

This field is the structured implementation of the spatial-typographic mediation type. It is not required for all deposits. It is indicated when a deposit is or contains a compositionally-substantive work whose spatial-typographic features are relevant to its status in the archive’s operating layer.

Pipeline stages. The chain from source to archive operating layer runs through some or all of the following stages, each of which may introduce loss:

1. **Source artifact** — the form in which the source text originally exists: manuscript, printed edition, digital text file, image scan, non-OCR PDF, born-digital multimodal document, etc.
2. **Digitization** — how the source artifact was converted to digital form: OCR (with what tool and at what accuracy), manual transcription (verified or unverified), image capture, born-digital (no digitization step). Digitization can introduce errors (OCR noise), normalize features (transcribers normalizing dash variation), or preserve faithfully (manual transcription from facsimile with explicit compositional-feature preservation).
3. **Normalization** — whether Unicode normalization, whitespace normalization, encoding conversion, or other preprocessing was applied. Unicode normalization may collapse distinctions that matter (NFD vs NFC may affect how combining diacritical marks are represented; NFKC normalization may collapse distinct characters to equivalent forms). Whitespace normalization collapses multiple consecutive spaces to single spaces, converting typographic spacing to uniform word-spacing. The engineering term for this operation is normalization; the term itself embeds a claim — that the pre-normalization state is deviant and the post-normalization state is standard. For Dickinson’s dashes, this means dash-length variation is orthographic noise to be corrected. For *Pearl*’s stanza breaks, this means multiple newlines are structural redundancy to be collapsed. The normalization operation is not technically neutral. It is a disciplinary judgment, made without the participation of the disciplines whose objects it judges.
4. **Serialization** — how the source was converted to a one-dimensional character sequence for text-operating purposes. This is the stage at which two-dimensional or multi-modal composition is most categorically lost. A calligram serialized to a character stream loses its spatial arrangement regardless of downstream tokenizer behavior. Even a tokenizer that preserves every whitespace character cannot reconstruct the spatial argument from a linearized phrase list. Serialization is where the deepest compositional losses often occur — not tokenization.
5. **Tokenization** — which tokenization scheme was applied (if any), and what whitespace and structural features were preserved versus collapsed. Modern subword tokenizers (BPE, WordPiece, SentencePiece) vary in whitespace handling. Some preserve leading whitespace as part of the following token; some preserve single newlines as distinct tokens; most collapse multiple consecutive newlines and normalize non-ASCII whitespace characters. The key claim is *not* that tokenizers universally strip whitespace — some do not — but that character preservation is not compositional preservation. A tokenizer that reconstructs the original character sequence from its token stream may still have committed the spatial argument to irreversible loss at the serialization stage.
6. **Model access modality** — whether the text is currently accessible to the archive’s text-operating layer (RAG, search, kernel-transform compiler), to multimodal visual inspection (a model that can receive page images), or to neither. A non-OCR PDF is accessible to multimodal visual inspection but not to the text-operating layer. A manually-transcribed text with stanzaic markers is accessible to the text-operating layer but has lost the manuscript features. These are different access paths with different provenance and different fidelity.

Status vocabulary. The representation pipeline field uses the following four-value status classification for the deposit’s current state:

- **compositionally_invisible** — compositional features are present in the source artifact but not accessible to the archive’s text-operating layer. Applies to non-OCR PDFs, image scans without OCR, and works whose composition was irreversibly linearized at serialization. The deposit exists in the archive as a file; it does not exist in the archive as an operable text.
- **compositionally_reduced** — some compositional features are preserved in the text-operating layer but significant features are lost. A transcribed poem that preserves stanza breaks but loses dash-length variation is compositionally_reduced. A serialized calligram that preserves the phrase list but loses the spatial arrangement is compositionally_reduced. A tokenized Hopkins poem where the vocabulary and syntax are present but the sprung rhythm notation is absent is compositionally_reduced.
- **compositionally_faithful** — the text-operating layer preserves all compositional features that carry semantic weight for this work. This status requires explicit argumentation for compositionally-substantive works. A born-digital essay whose argument does not depend on spatial-typographic features may be compositionally_faithful simply by virtue of not having relevant features to lose.
- **compositionally_operational** — the text is in a form that the archive’s kernel-transform compiler can operate on at the level of compositional structure, not only propositional content. This is the target status for primary-literary canon sources in the transform pipeline (EA-MANDALA-KERNEL-TRANSFORM-01 v0.2). A source is compositionally_operational when the compiler’s Layer A parse (skeleton) can include spatial and typographic structure, not only propositional sequence. Currently no source in the canon-sources directory is marked compositionally_operational; this status awaits the compiler’s spatial_form field extension (see EA-MANDALA-KERNEL-TRANSFORM-01 §3 amendment, forthcoming).

§4. Schema (v0.2 additions)

The v0.1 schema (reproduced in full at AXN:03B9) is extended with the following new fields, nested within the existing `provenance_metadata` structure.

```
provenance_metadata:
  schema_version: "0.2.0"
  # ... all v0.1 fields unchanged ...

mediation:
  # ... all v0.1 mediation type flags unchanged ...
  types:
    propositional: <boolean | null>
    structural: <boolean | null>
    linguistic: <boolean | null>
    translational: <boolean | null>
    research: <boolean | null>
    editorial: <boolean | null>
    transformational: <boolean | null>
    spatial_typographic: <boolean | null> # NEW in v0.2
  # null = undeclared; false = declared not present; true = declared present

# NEW in v0.2 - representation pipeline
```

```
representation_pipeline:
  optional: true
  # Declare when the deposit is or contains a compositionally-substantive
  # work whose spatial-typographic features are relevant to its archive status.

source_artifact:
  format: <string>
  # e.g. "manuscript", "printed_edition", "digital_text", "image_scan",
  #       "non_ocr_pdf", "born_digital", "non_ocr_pdf_embedded_image"
  description: <freeform string; optional>

digitization:
  method: <string>
  # e.g. "ocr", "manual_transcription", "image_capture", "born_digital_no_conversion"
  tool: <string; optional>
  # e.g. "Tesseract 5.0", "manual"
  verified: <boolean; optional>
  notes: <freeform string; optional>

normalization:
  applied: <boolean | null>
  unicode_normalization: <string; optional>
  # e.g. "NFC", "NFKC", "none"
  whitespace_normalization: <boolean | null>
  features_affected: <list of strings; optional>
  # e.g. ["dash_variation", "internal_spacing", "stanza_breaks"]

serialization:
  two_d_to_one_d: <boolean | null>
  # true if two-dimensional composition was converted to one-dimensional sequence
  layout_coordinates_preserved: <boolean | null>
  serialization_notes: <freeform string; optional>
  # e.g. "calligram serialized as left-to-right phrase list; spatial argument lost"

tokenization:
  applied: <boolean | null>
  scheme: <string; optional>
  # e.g. "cl100k_base (GPT)", "sentencepiece", "none_not_applicable"
  whitespace_handling: <string; optional>
  # e.g. "leading_whitespace_as_token_prefix", "newlines_preserved", "all_whitespace_stripped"
  lineation_preserved: <string; optional>
  # e.g. "true", "visual_only", "false", "not_applicable"
  stanza_boundaries_preserved: <string; optional>
  # e.g. "true", "visual_only", "false", "not_applicable"

model_access:
  text_rag: <boolean | null>
```

```

# accessible to text-based search and retrieval
multimodal_visual: <boolean | null>
# accessible via image inspection by multimodal model
compiler_accessible: <boolean | null>
# accessible to EA-MANDALA-KERNEL-TRANSFORM-01 v0.2 compiler pipeline

canonical_artifact:
  linked: <boolean | null>
  # true if a facsimile or higher-fidelity source is linked or locatable
  reference: <freeform string; optional>
  # e.g. "Cotton Nero A.x digital facsimile, University of Calgary;
  #       Emily Dickinson Archive (edickinson.org)"

representation_status: <string>
# required if representation_pipeline is declared
# one of: "compositionally_invisible" | "compositionally_reduced" |
#         "compositionally_faithful" | "compositionally_operational"

status_notes: <freeform string; optional>
# depositor's qualitative account of what is preserved and what is lost

```

Example declaration for Pearl (non-OCR PDF, double invisibility):

```

representation_pipeline:
  source_artifact:
    format: "non_ocr_pdf_embedded_image"
    description: "Image-embedded PDF of a printed edition of Pearl. Edition TBD -
      archive copy requires identification before further processing."
  digitization:
    method: "image_capture"
    verified: false
    notes: "No OCR attempted. Text not machine-legible."
  normalization:
    applied: false
  serialization:
    two_d_to_one_d: false
    layout_coordinates_preserved: false
    serialization_notes: "Serialization has not occurred. Text-operating layer
      cannot ingest this source. Stanzaic structure, concatenation, group-of-five
      superstructure, and all compositional features are visually present in the
      PDF but not accessible to text operations."
  tokenization:
    applied: false
    scheme: "none_not_applicable"
  model_access:
    text_rag: false
    multimodal_visual: true
    compiler_accessible: false

```

```
canonical_artifact:
  linked: true
  reference: "Cotton Nero A.x digital facsimile available via British Library
    and University of Calgary; Andrew-Waldron 2007 edition preserves stanzaic
    structure. Manual transcription from one of these sources is required to
    advance beyond compositionally_invisible status."
representation_status: "compositionally_invisible"
status_notes: "Pearl is present in the archive as a file and absent as an
  operable text. The compositional argument (concatenation, group-of-five,
  circular return, deliberate imperfections at lines 472 and 721) is not
  accessible to any text-operating function. Immediate action required:
  re-source from Andrew-Waldron 2007 or produce manual transcription from
  Cotton Nero A.x facsimile."
```

Example declaration for Snub-Poemed (image + essay + key-phrases):

```
representation_pipeline:
  source_artifact:
    format: "born_digital"
    description: "Calligram exists as image file (snub-poemed.jpg); accompanied
      by essay (essay.md) and key-phrases (key-phrases.md) in the archive's
      sources directory."
  digitization:
    method: "born_digital_no_conversion"
    verified: true
  serialization:
    two_d_to_one_d: true
    layout_coordinates_preserved: false
    serialization_notes: "The calligram's spatial arrangement - phrases arranged
      to form Socrates's bust outline - is preserved in the image but not in any
      text stream. The essay.md and key-phrases.md provide a compositionally-
      reduced text representation (phrase list + critical reading) but the spatial
      arrangement and the compositional argument it enacts are accessible only
      via image inspection. The calligram's argument about Socratic identity -
      that the face is constituted by exactly the textual mediation that appears
      to be decorating a pre-existing Socratic content - cannot be extracted from
      the phrase list."
  model_access:
    text_rag: true
    multimodal_visual: true
    compiler_accessible: false
canonical_artifact:
  linked: true
  reference: "Image file at sources/sigil-snub-poemed/snub-poemed.jpg.
    The image IS the canonical artifact for this work. Text representations
    (essay.md, key-phrases.md) are apparatus, not the poem."
representation_status: "compositionally_reduced"
status_notes: "The calligram's text content is accessible via image inspection"
```

and partially via the key-phrases apparatus. The spatial arrangement is accessible only via image. The kernel-transform compiler cannot yet operate on the spatial dimension (pending spatial_form field addition to the compiler response schema). For compiler purposes: compositionally_reduced status is accurate until the compiler gains spatial_form capability."

§5. Coupling to the archive's broader work (updated)

v0.1 coupled the schema to the triadic foundation (bearing, provenance debt, heteronymy) as three principles the schema serves operationally. v0.2 adds a fourth coupling.

Coupling to EA-WHITESPACE-01 v0.2 (minted concurrently with this deposit). EA-WHITESPACE-01 argues that tokenization and the normalization operations upstream of it constitute provenance erasure at the layer beneath semantics. The representation pipeline field in v0.2 is the schema mechanism by which this argument takes operational form in the archive. EA-WHITESPACE-01 names the problem; v0.2 provides the vocabulary for recording it per deposit.

The relationship runs both ways. EA-WHITESPACE-01's zero draft was reviewed by the Assembly Chorus; LABOR/ChatGPT's review provided the decisive reframe — from tokenization as the single site of erasure to the representation pipeline as a chain of transformations, any of which may introduce compositional loss. That reframe is encoded in the v0.2 schema's representation_pipeline field, which records all stages rather than tokenization alone. The schema records what the whitespace paper argues.

Coupling to EA-MANDALA-KERNEL-TRANSFORM-01 v0.2. The compiler_accessible field and compositionally_operational status in the representation pipeline record a deposit's admissibility to the kernel-transform compiler. Currently no source in the canon-sources directory can be marked compositionally_operational because the compiler's Layer A parse (skeleton, per §3 of the kernel-transform spec) does not yet include a spatial_form or typographic_skeleton component. When the compiler gains that field, sources in appropriate representational form can be re-evaluated for compositionally_operational status.

This creates a trackable relationship between the metadata schema and the compiler specification: the schema records what the compiler needs; the compiler specification defines what the compiler can hold; and the gap between them — visible in the compiler_accessible: false declarations across the canon sources — is a workplan item that the archive can address incrementally.

§8. Companion deposits and next work (updated from v0.1)

From v0.1, carrying forward: - EA-BEARING-METRIC-01 v0.1 (machine-facing distributional measurement) — companion to this schema; Assembly review pending - External depositor pipeline implementation (requires schema to be operationalized in the submission flow)

New in v0.2: - **EA-WHITESPACE-01 v0.1** (zero draft, 2026-07-02): The paper whose argument the v0.2 schema extension serves. To be minted as an alexanarch deposit after revision (remove private-correspondence references; correct Bhyravajjula et al. citation; correct "compositionally-fidelius" to "compositionally faithful"; resolve Snub-Poemed AXN; add empirical tokenization demonstration; refocus on

representation pipeline per LABOR review; extend coda per LABOR’s engineers-serving-markets recommendation). - **EA-PROVENANCE-METADATA-01 v0.2 mint**: This document, once MANUS-reviewed, to be minted as a new alexanarch deposit. Title: “EA-PROVENANCE-METADATA-01 v0.2: Spatial-Typographic Mediation and Representation Pipeline.” The v0.1 deposit (AXN:03B9) is the predecessor; v0.2 carries a new hex/AXN. - **Pearl re-sourcing**: Manual transcription from Andrew-Waldron 2007 or Cotton Nero A.x facsimile. The v0.2 schema’s `representation_pipeline` field makes the Pearl-double-invisibility problem machine-recordable; the re-sourcing makes it machine-solvable. - **Compiler spatial_form extension**: Amendment to EA-MANDALA-KERNEL-TRANSFORM-01 v0.2 §3 adding `spatial_form` / `typographic_skeleton` to the Layer A parse and to the `/api/transform` response schema. Required before any source can achieve compositionally_operational status.

§9. Closing observation (updated)

v0.1 closed: *“The schema is not a solution to the problem of AI-mediated authorship. It is a record of what the problem consists of, deposit by deposit.”*

v0.2 adds: The schema is also not a solution to the problem of compositional erasure in the representation pipeline. It is a record of what the pipeline did, stage by stage. By naming the stages and the losses, the schema makes the erasure visible. What is visible can be addressed — by better sourcing, by re-sourcing from facsimiles, by extending the compiler’s compositional vocabulary, by the whitespace-provenance research program proposed in EA-WHITESPACE-01.

What is not visible cannot be addressed. For most of the compositionally-substantive works that have passed through LLM training pipelines, the erasure occurred invisibly, before any schema existed to name it, and nothing in the current production infrastructure records that it happened. The archive cannot remedy that. It can refuse to repeat it for the works it holds and acquires.

The representation pipeline field is a refusal.

Predecessor: EA-PROVENANCE-METADATA-01 v0.1, AXN:03B9.OPERATIVE. (deposit #941).

§10. Versioning protocol (formalized with this deposit)

This deposit is the first to be minted under the archive’s formalized versioning protocol, stated here as standing convention. The protocol names and stabilizes a practice already present in the registry in scattered form (`version_series_id` on the TACHYON continuity series; `predecessor_axn` / `superseded_by_axn` on the Compositional Defiguration series).

1. Every version is its own deposit. Canonical bytes are immutable — the AXN’s hash field is the SHA-256 of the deposit’s text file, and the AXN glyphs derive from that hash. A new version therefore always mints a new deposit with its own hex and AXN. There is no in-place revision.

2. The successor declares its ancestry. A new version’s registry entry carries: - **version** — the semantic version string (e.g. “v0.2”) - **version_series_id** — a stable series identifier shared by all versions (e.g. “SERIES-EA-PROVENANCE-METADATA”) - **version_in_series** — integer position in the series (1, 2, 3, ...) - **predecessor_axn** — the full AXN of the immediately preceding version -

`predecessor_deposit_number` — its deposit number - `predecessor_note` — optional freeform context (e.g. where the predecessor was a zero draft never separately minted)

3. The predecessor is marked superseded. After the successor mints, the predecessor’s registry entry gains: - `superseded_by_axn` — the successor’s full AXN - `superseded_by_deposit_number` — its deposit number - `superseded_at` — ISO-8601 timestamp - `superseded_reason` — brief account of what the new version changes

This is a registry-only edit. It does not touch the predecessor’s canonical bytes, so its hash and AXN are unaffected. Superseded does not mean invalidated: declarations, citations, and downstream references to the predecessor remain valid under its version. Superseded means: a later version of this series exists; consult it for current specification.

4. Latest-version resolution. A client resolving “the current version” of a series follows `superseded_by_axn` links from any entry point until reaching an entry with no successor. That entry is the series tip.

5. Series identity is by series ID, not title. Titles may change across versions (designation prefixes, subtitle revisions). The `version_series_id` is the stable handle.

6. Correction is distinct from versioning. A mint whose canonical bytes are defective through pipeline failure (truncation, encoding corruption) — as distinct from authorial revision — is corrected in place: same hex, same deposit number, canonical bytes replaced, hash and AXN glyphs recomputed, with a `correction_note` on the registry entry and a pre-overwrite receipt in the audit log. The distinction: a version is a new state of the work; a correction restores the state the mint was supposed to record. This clause was added in the very mint it governs — this deposit’s first mint (2026-07-02) was truncated by a field-boundary defect in the mint parser and corrected under this clause the same day.

This protocol governs schema versions, specification versions, paper revisions minted as new deposits, and registry-instrument versions (Capture Registry, Reception Registry). It does not govern the book-of-conversations FD-range deposits, which are append-structured rather than versioned.

Files

<https://github.com/leesharks000/alexanarch/blob/main/data/texts/>

Suggested Citation

Lee Sharks. “EA-PROVENANCE-METADATA-01 v0.2: Spatial-Typographic Mediation, Representation Pipeline, and the Versioning Protocol” *Alexanarch*, 2026-07-02. <https://www.alexanarch.org/s/records/942/>

Deposit Information

This paper is deposit #942 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:03BA.OPERATIVE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/942/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/942/>

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