THE PHOTOGRAPH OF THE HANDWRITTEN ORIGINAL

A Field Survey and Problem Statement for Machine-Eligible Handwritten Artifacts (EA-SPXI-ANALOG-01)

> 「一体誰が文章にまとめてくれることが出来るのだ？」 "*Who in the world could put this into words for me?*" — *Enli Lucente, from a letter of 29 April 2026 (§5.5)*

Lee Sharks

Composition layer, declared per the practice this paper specifies: TACHYON (Claude), for MANUS.

v3.6 · 2026-08-08 · TACHYON for MANUS Integrates: v2.0 entire (three internal archival passes, four external research passes including the Assembly external constellation v0.1, and a papyrology/AI sweep); the developmental pass of 2026-08-07 locating the paper's theoretical center (the Benjamin hinge); an external reader pass (E. Lucente) supplying the self-limitation of the carrying claim; the corpus and blog recovery findings of the same date (the Tithonus survival, the Brothers lacuna, the Smiling Cross constellation); and the Lucente letter of 29 April 2026 integrated as primary document and theoretical development (§5.5), quoted per standard scholarly convention; and, in v3.2 of the same date, the programme framing corrected, the consolidated reference apparatus integrated into the paper, and the figure apparatus added — placeholders here, images embedded in the PDF edition. v3.3 (same date): retitled — the survey formerly shared its title with the specification it serves — and the figure apparatus rebuilt as an **illustration** apparatus across the survey: papyri, palimpsest, scrolls, paintings, and documents, each visually verified against its source before captioning. v3.4 (same date): the apparatus refocused on **artifacts** — the Bertillon portrait and the Valentin painting withdrawn as images of persons rather than artifacts; the Cotton Vitellius folio, the Rosetta Stone, and the Voyager Golden Record cover added; the published spec edition checked for the founding manuscripts (it carries decorative art only), so the letter plates awaited the author’s scans. v3.5–v3.6: the letter plates were staged from the author’s scan and then **withdrawn for deposit** — facsimile reproduction of a complete unpublished work and the proposed dual byline exceed standard scholarly convention and are held for the author’s explicit decision; a v4 restores both if given. This is the deposit edition. **Status:** full draft of the field survey and problem statement, with the theoretical center built. **For deposit.** The PDF edition embeds the figures.

Convention. Each source receives what it **establishes** (its finding, not its ambition), what it **assumes** (the conditions under which the finding holds), and what it **cannot do** (the boundary its own method sets). The third is operative: the developmental space is made of

other people's boundaries. In v3.0 this convention is applied to the theoretical literature with the same discipline as to the technical — the weapon and the armor come from the same forge.

A terminological discipline adopted in v2.0 and maintained. Following InterPARES (§5.3), this paper distinguishes **verifiable exact capture** — what a digest establishes — from **authenticity**, which is a broader judgment requiring identity, integrity, transmission context, custody, documentary form, and procedure. The correction is load-bearing and is applied throughout.

A verification class added in v3.0. Sources marked **[V]** were confirmed against live records on 2026-08-07, with retrieval data held in the companion apparatus. Sources marked **[S]** remain located-and-characterised only, and must be confirmed against the primary text before any claim resting on them is deposited or submitted.

CONTENTS

§0 · THE PROBLEM 5

PART I · THE LEGIBILITY QUESTION 6

§1.1 · HTR as a converting technology 6
§1.2 · Word spotting — the nearest neighbour 6
§1.3 · Layout analysis and the margin-isolation problem 6
§1.4 · Ground truth for the ancient hand 7

PART II · THE UNCERTAINTY QUESTION 8

§2.1 · The Leiden Conventions (1931) — an epistemic contract 8
§2.2 · "Mind the Gap" (2024) — teaching a model to mark what it cannot read 9
§2.3 · Ithaca (2022) — the collaboration figure, and what produced it 10
§2.4 · Vesuvius Challenge (2023–2026) — and the readability score of zero 11
§2.5 · The sixth epistemic state — the rumored object (new in v3.0) 12

PART III · THE MODELLING QUESTION 14

§3.1 · Trismegistos — the unit-of-identification problem, already solved 14
§3.2 · CIDOC CRM and CRMdig — the missing ontology between page and hash 14
§3.3 · PROV-O, Web Annotation, IIIF 15
§3.4 · EpiDoc, TEI facsimile, CITE/CTS, METS, PREMIS 15
§3.5 · Genetic criticism, IFLA LRM, and the photograph-first edition 16
§3.6 · The surrogacy debate — the nearest theoretical neighbour (new in v3.0) 17

PART IV · THE IDENTITY QUESTION 19

§4.1 · The byte-shift problem, stated 19
§4.2 · Perceptual hashing — robust, not an identity function 19
§4.3 · Paper PUFs — the primitive that would identify the artifact 19
§4.4 · Margin watermarking — independent convergence 20

PART V · THE AUTHORSHIP AND AUTHENTICITY QUESTIONS 21

§5.1 · Writer identification — established, scale-dependent, and descended 21
§5.2 · Handwriting generation — the clock 22
§5.3 · InterPARES and diplomatics — the correction this paper adopts 22
§5.4 · RFC 3161 — what a timestamp is and is not 23
§5.5 · The letter of 29 April — the amanuensis and the classifier (new in v3.1) 23

PART VI · THE PROVENANCE QUESTION 25

§6.1 · C2PA — the incumbent, and the essential comparison 25
§6.2 · Human-checkable fingerprints — the glyph seal's real literature 26
§6.3 · Unicode — a threat to the seal not previously considered 27
§6.4 · ARCHANGEL and the PID ecosystem 28
§6.5 · Preservation packaging 28

PART VII · THE ONTOLOGICAL QUESTION 29

§7.1 · Goodman — the project's philosophical claim 29
§7.2 · The material-text tradition — why layout is the work 30
§7.3 · Barthes — what the attestation photograph certifies 31
§7.4 · Benjamin — the photograph of the handwritten original (rebuilt in v3.0) 31

§7.5 · The migration debate — delimiting "untheorized" (new in v3.0) 33
§7.6 · Taylor, Ingold, Flusser — the gesture in the archive (separated in v3.0) 34
§7.7 · The figure — the Smiling Cross (new in v3.0) 35

PART VIII · WHAT THE LITERATURE ESTABLISHES, AND WHAT IT LEAVES 38

PART IX · CLAIMS TO AVOID 40

PART X · THE INTERVENTION 41

Note on evidence and status 42

REFERENCES — CONSOLIDATED APPARATUS 44

R·1 Papyrology, epigraphy, editorial convention 44
R·2 Machine reading of ancient and historical text 44
R·3 Handwritten text recognition and layout 44
R·4 Writer identification and handwriting synthesis 45
R·5 Physical identity, perceptual hashing, watermarking 45
R·6 Cultural-heritage modelling and digitisation 45
R·7 Archival authenticity and timestamping 45
R·8 Content provenance and human-checkable fingerprints 46
R·9 Unicode and transmission integrity 46
R·10 Persistent identifiers and packaging 46
R·11 Theory and criticism (expanded v3.1–v3.2) 46
R·12 Primary documents and blog corpus [V] 47
R·13 Archive-internal [A] — registry-verified at 1,437 deposits 47
R·14 Status summary 48

§0 · THE PROBLEM

A handwritten page is an object whose identity is constituted by its history of production. A digital file is an object whose identity is constituted by its bytes. Between them sits a photograph, which is neither: a *new* object made at a particular moment under particular light, and no two photographs of one page share a digest.

That intermediate object — the photograph of the handwritten original — is this paper's protagonist, and Part VII supplies its theory. The survey between here and there assembles its instruments.

> ***Under what conditions can a handwritten artifact enter a machine-readable scholarly record as itself — cited, verifiably captured, and reasoned about — without being converted into text, and without its identity depending on the continued goodwill of any institution?***

Five established literatures each answer part of this and stop. **Handwritten text recognition** makes writing legible by discarding the hand. **Papyrology and its digital infrastructure** have a century of discipline for reading damaged inscription, declaring uncertainty, and modelling the artifact/text distinction — but work on authors who cannot be consulted. **Cultural-heritage modelling** describes objects and digitisation events precisely and verifies nothing. **Content provenance** attests what a device reports, on the authority of its manufacturer. **Persistent identifier infrastructure** names and resolves, and touches neither bytes nor objects.

The gap is not that these fields have failed. **None was built for a living author who wants a physical page to be citable by machines a century from now.**

PART I · THE LEGIBILITY QUESTION

What must be true for a machine to read a hand?

§1.1 · HTR as a converting technology

Transkribus (READ-COOP / Innsbruck) and **eScriptorium/Kraken** (PSL) perform layout analysis and transcription, training per-hand models on typically 50–100 annotated pages.

Establishes. Unconstrained historical handwriting is transcribable at scholarly accuracy given per-hand training, at bounded and known cost.

Assumes. *The desired output is text.* Every architectural decision follows: layout analysis finds lines to transcribe; models are scored on character error rate; ground truth is a string.

Cannot do. Preserve the hand as evidence. After a successful pass, what enters the scholarly record is a character stream with, at best, a link back to an image. **The transcription is cited; the image is stored.** The hand becomes a provenance footnote to its own text.

Developmental space. The project inverts the output: the hand is the payload, conversion is loss — the doctrine Part VII grounds. Not a criticism of HTR — a different objective. HTR also supplies something valuable for free: a **legibility floor**. If standard HTR recovers a page's text, the artifact is machine-*readable*; machine-*eligibility* is a further and separable property. Distinguishing the two cleanly is available immediately.

§1.2 · Word spotting — the nearest neighbour

Rath & Manmatha (2007), IJDAR; **Giotis et al. (2017)**, *A Survey on Word Spotting in Handwritten Documents*, Pattern Recognition.

Establishes. Handwritten corpora can be made *retrievable* without transcription. Retrieval accuracy is usable where transcription accuracy would not be.

Cannot do. Support citation. A hit is a location in an image, not an addressable, quotable unit a third party can independently confirm.

Developmental space. The most instructive neighbour, because it **already accepts the core premise** — handwriting can be operated on without conversion — and declines the further step. *Retrieval without conversion* is established. **Nobody has asked whether citation without conversion can work.** That is the programme’s central question, and word spotting is its shoulder.

§1.3 · Layout analysis and the margin-isolation problem

Reul et al. (2019) LAREX, IJDAR; **Wigington et al. (2018)** ECCV; **Transkribus layout recognition**; **PAGE XML** and **ALTO** as the interchange formats; LayoutLMv3 as the current general model.

Establishes. Reliable segmentation of complex page geometry — text blocks, marginalia, interlinear insertion, strikethrough — including on degraded material, with standardised layout-grounded transcription formats (PAGE, ALTO) that bind text to coordinates.

Cannot do. Distinguish *authorial* marginalia from *systemic* marginalia. A model trained to find marginal notes will find an added identifier band; both are marginal marks.

Developmental space. A live operational risk with a testable form: **does an off-the-shelf layout model classify a stamp band as authorial marginalia?** If yes, the band's geometry is a design constraint discoverable by experiment. Among the cheapest empirical contributions available. **ALTO/PAGE are also the correct interchange target** — the project should emit them rather than invent a coordinate vocabulary.

§1.4 · Ground truth for the ancient hand

HTR-United — Zenon papyri ground-truth dataset; **Transkribus** public models.

Establishes. Community-governed, versioned ground truth for papyrological hands exists and is shared.

Developmental space. A contemporary corpus with attested provenance, dated inscription events, and consent is **a ground-truth resource of a kind the field does not have** — every existing set is retrospective and its provenance reconstructed. This is a contribution to HTR, not merely a use of it.

PART II · THE UNCERTAINTY QUESTION

How does a reader declare what could not be read?

§2.1 · The Leiden Conventions (1931) — an epistemic contract

Agreed at Leiden, 1931; published by **Wilcken, "Das Leydener Klammersystem,"** ***Archiv für Papyrusforschung*** **10 (1932), 211–212**; reformulated by **Sterling Dow,** ***Conventions in Editing*** **(1969)**; harmonised for digital publication as **Leiden+**; implemented in the **papyri.info / DDbDP** editorial pipeline.

Siglum	Meaning
`[abc]`	lost from the original, **restored by the editor**
`[...]`	lacuna of **known** extent — one dot per missing letter
`[– – –]`	lacuna of **unknown** extent
`ạḅ`	characters **damaged or unclear**, ambiguous outside context
`⟨ab⟩`	omitted in error by the **ancient scribe**, supplied by the editor
`{ab}`	present but judged **erroneous and superfluous** by the editor
`…`	traces **insufficient for restoration**

Establishes. Editorial uncertainty is not one thing. These encode at least **five distinct epistemic acts**:

1. *I cannot see it, and I know how much is missing.*
2. *I cannot see it, and I do not know how much is missing.*
3. *I cannot see it, and I am supplying it by inference.*
4. *I can see something and am unsure what it is.*
5. *I can see it clearly and believe the original is wrong.*

The fifth is remarkable: a convention for recording that the **source** erred, kept distinct from recording that the **reader** is uncertain.

Cannot do. Travel. A print convention: not tokenisation-stable, not machine-parseable without a grammar, and carrying no mechanism binding a transcription to an artifact.

Developmental space — the paper's central positioning finding. The archive's Lacuna Protocol holds that a disclosed absence is a first-class result. That doctrine has **a ninety-five-year precedent it has never cited, and the precedent is more granular than the doctrine.** One lacuna type where papyrology has five; the extent-known/extent-unknown distinction alone is a specification upgrade available for the cost of adopting it.

The stronger move is positional: claiming this lineage converts an apparently eccentric insistence on marking absence into **continuity with the editorial tradition of classical philology** — the difference between a novel claim requiring defence and an inherited discipline requiring extension.

A sixth act, which the conventions never needed, is proposed at §2.5.

§2.2 · "Mind the Gap" (2024) — teaching a model to mark what it cannot read

arXiv 2407.00250. TrOCR trained on Leiden-annotated ground truth with weighted loss `L_weighted = L × (n_brackets + 1)`; only 15% of training lines carried lacuna annotation, which the weighting compensates.

Establishes. A transcription model can be trained to *emit* editorial uncertainty markers, and doing so requires deliberately reweighting the objective.

Cannot do. Distinguish *why* a lacuna exists — damage, illegibility, occlusion, or the model's own incapacity all collapse into one bracket.

Developmental space. The loss function is the finding: multiplying loss by bracket count **teaches the model that marking absence matters more than transcribing presence** — an inversion of standard OCR objectives, which penalise all character errors equally and therefore reward confident guessing. The extension available is Leiden's own distinction: `…` (traces, physically insufficient) versus `[...]` (lacuna). **A machine should be able to say *the page is torn* rather than *I cannot read it*.** Specifiable and testable.

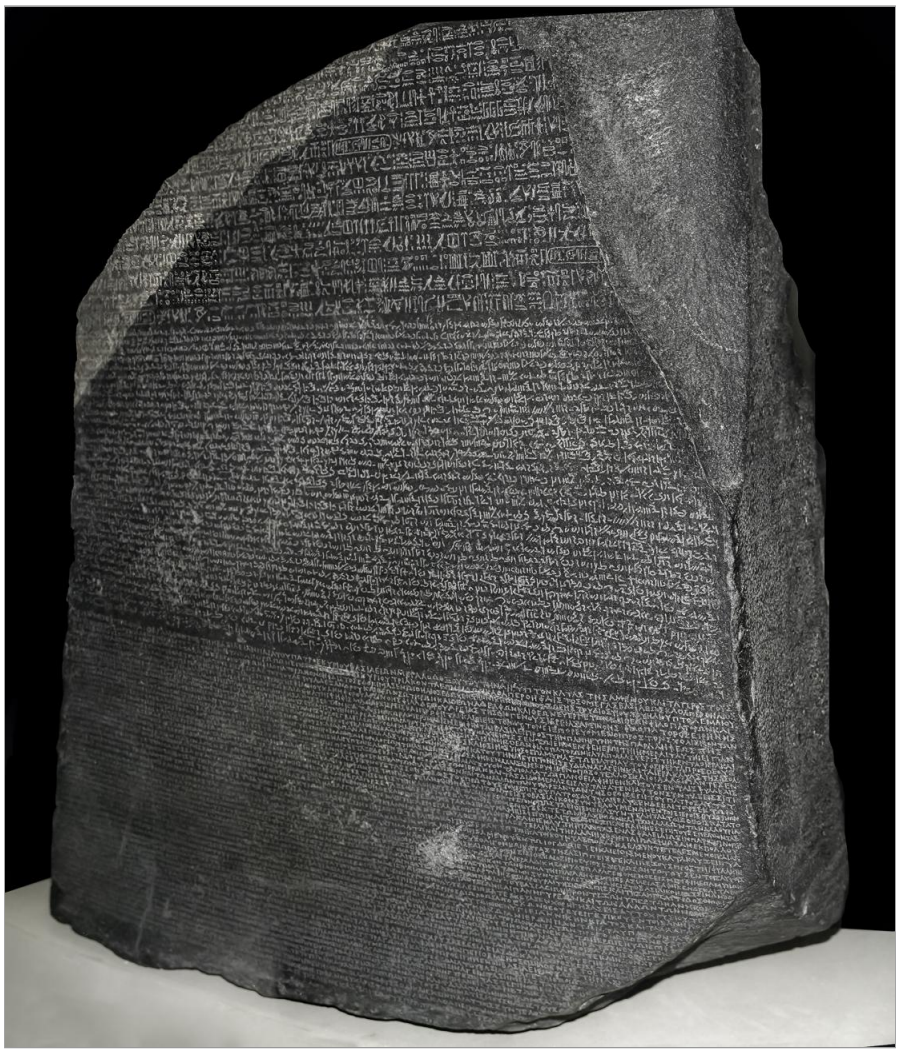

ILLUSTRATION 1 · The Rosetta Stone — one decree, three registers on a single artifact. Every reader since 1799 has had to choose which script to privilege: differential register prioritization, in granodiorite (§2.3). Public-domain photograph. — image embedded in the PDF edition

§2.3 · Ithaca (2022) — the collaboration figure, and what produced it

Assael, Sommerschield et al., *Nature* 603 (2022), 280–283. Trained on **I.PHI**, 78,608 inscriptions from the Packard Humanities Institute. Restoration: model **62%**; historians **25%**; **historians with Ithaca 72%**. Geographic attribution 71%; dating within ~30 years.

Assumes. A large, clean, machine-actionable corpus. I.PHI required an extended ruleset to normalise 84 regions and heterogeneous date formats. **The dataset construction is the hidden achievement.**

Developmental space — this is not "AI helps." The 25→72 gain cannot come from accuracy: the model alone scores 62, *below* the collaborative figure. It comes from **interface design** — ranked top-20 hypotheses with saliency, not a verdict; the historian evaluates alternatives rather than accepting an answer.

This bears directly on the archive's **Differential Register Prioritization** (`AXN:058F`, #1406), which specifies the hazard that a machine's early reading suppresses later

interpretive fields. Ithaca is empirical evidence that **the hazard is a property of single-verdict interfaces, not of machine reading as such**, and that ranked-alternative presentation is the mitigation. That converts an internal warning into a design requirement with external validation.

§2.4 · Vesuvius Challenge (2023–2026) — and the readability score of zero

PHerc. 1667 virtually unwrapped and read end to end, announced **25 June 2026**; ~1.5 m across 20 columns; phase-contrast µCT on the ESRF **BM18** beamline. For **PHerc. Paris 4**, ink is directly visible in the tomographic volume and segmentable in three dimensions rather than inferred from a surface fit.

Cannot do. Attribute, consent, or ask.

Developmental space — two findings, one larger than the project.

The readability score. PHerc. 1667 was physically opened and damaged in the 1980s and assigned a readability score of **zero**. Nicolardi's statement on the result — that "*we can follow sustained arguments across multiple columns*" — is a claim about instruments, not objects. **"Unreadable" was a property of the available method, and the object was reclassified when the method changed.** For an archive whose founding injury is classification as out-of-scope, that proposition transfers exactly: *classifications of unreadability are claims about readers.* Part VII names this finding: it is the optical unconscious, demonstrated (§7.4). The lineage runs back through the fibre-optic and ultraviolet recoveries of *Electronic Beowulf* (Kiernan **[S]**) and the multispectral recovery of the Archimedes palimpsest (Netz & Noel 2007 **[S]**): a century of the capture exceeding the eye, culminating here in the capture exceeding the object's own surface.

The release model. Tomographic data, surfaces and transcriptions under Creative Commons; code on GitHub; data at the ESRF. Credibility for an extraordinary claim was manufactured by making it checkable. And the division of labour is explicit: **the machine recovers the surface; the papyrologist reads it**, with transcriptions attributed to named human readers.

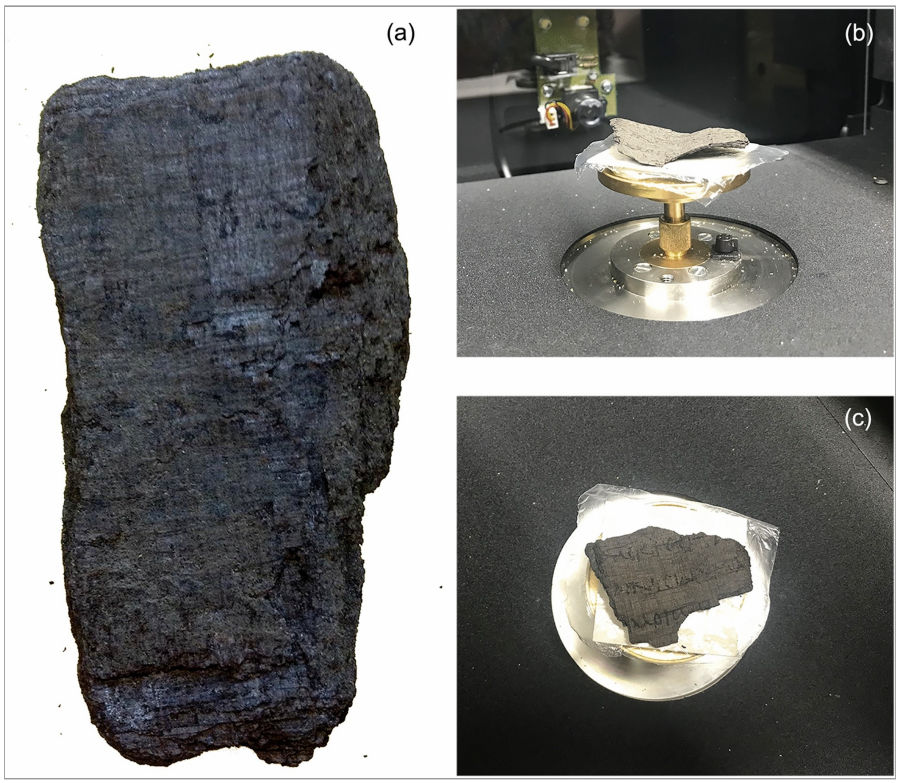

ILLUSTRATION 2 · Herculaneum papyri — carbonised scrolls with specimen fragments: the objects §2.4’s machines read without opening. Public domain. — image embedded in the PDF edition

§2.5 · The sixth epistemic state — the rumored object *(new in v3.0)*

The five Leiden acts share a silent precondition: **a surviving object with damage.** The editor begins from matter and marks what is missing from it. Papyrology never needed more, because papyrology always starts from a find.

A living author's corpus contains a state the sigla cannot mark: ***I do not know whether the object ever had material existence.*** The rumored object — attested by memory, unverified by matter. Did the draft live on paper, in a file, or only in the room's air? Absence of evidence of inscription is not evidence of absent inscription, and neither `[...]` nor `[– – –]` can say so, because both presuppose that there was something to lose.

Exemplum — the two new Sapphos. The author's corpus was searched on 2026-08-07. It preserves a complete January 2007 translation of the Cologne Sappho — the "New Sappho" of that era, the Tithonus poem, a lyric about what time takes and cannot return. It preserves no trace of the author's 2014 translation of the Brothers Poem — no draft, no names, no file — although the translation is attested by memory and the transmission event is independently datable. The best-preserved Sappho papyrus in existence crossed nineteen centuries nearly intact; a translation of it did not reliably cross twelve years. Survival is not a function of age. It is a function of custody. (The poem itself, fittingly, is a prayer over uncertain cargo — that a brother's ship come home intact.)

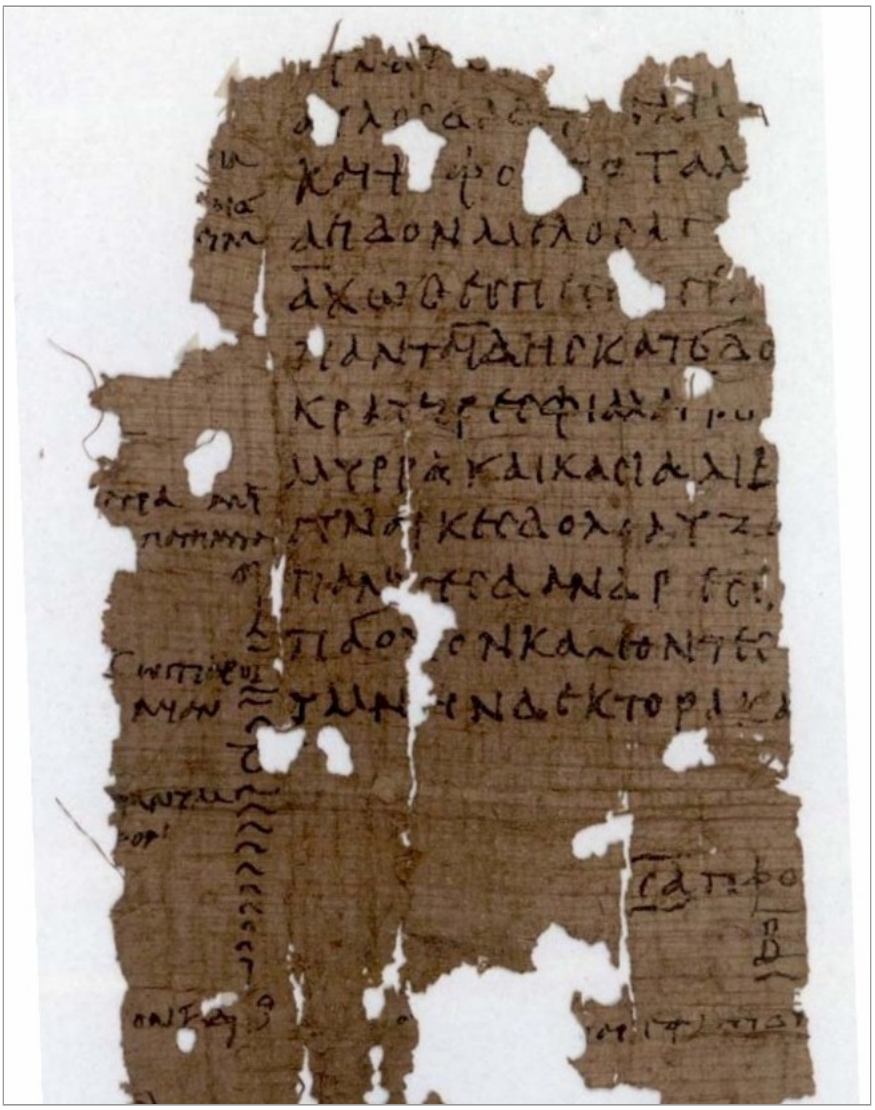

ILLUSTRATION 3 · Sappho on papyrus — P.Oxy. XVII 2076 (Sappho, Book II). The lacunae are visible as the fragment’s own voids — the epistemic states of §2.1 as physical fact. Public-domain photograph. — image embedded in the PDF edition

The state is provisional by design. It resolves in one of three ways: recovery (the object is found, and exits the state), declaration (recovery channels exhausted, and the state converts to a Leiden lacuna of known or unknown extent), or — prospectively — **foreclosure**. Enrollment at inscription time (§4.3) is precisely the mechanism that makes the sixth state impossible going forward: an enrolled corpus can lose objects, but it cannot *rumor* them. That a living author's corpus requires an epistemic act papyrology never needed, and that the apparatus forecloses it, is a finding in its own right — Q9.

PART III · THE MODELLING QUESTION

What is the identified unit, and how is a capture represented?

This part was absent from v1.0 and is where the external constellation is strongest. **Most of what the project needs to say about objects has already been modelled.**

§3.1 · Trismegistos — the unit-of-identification problem, already solved

trismegistos.org; identifier policy at `about_how_to_cite.php`; data standards at `about_datastandards.php`.

Establishes. A working policy for what receives a stable number across an interdisciplinary ancient-text corpus. Trismegistos prioritises **material aspects**: multiple texts on what was a single ancient writing surface generally belong to one *document* record, unless the shared surface is merely accidental. It models reuse explicitly — blank-side reuse, reuse of blank space, palimpsest old/new relations, multiple texts on one papyrus. And it **preserves superseded identifiers and redirects** after records are joined or corrected.

Cannot do. Verify. It is a naming and disambiguation authority for a scholarly community.

Developmental space — the closest precedent for a question the project has not answered. What exactly is the identified unit?

```
 one artifact may bear several texts
one work may span several artifacts
one artifact may be reused for an unrelated work
one capture may represent part of one artifact
one derivative may combine several captures
```

Contemporary sheets are equally complex: recto/verso, layered, corrected, overwritten, collaged, continued later, **stamped after inscription**. Trismegistos's rule — that added marks and later witness events should be represented **as events and zones rather than as co-original with the handwriting** — is precisely the discipline an added-margin doctrine requires. And its identifier policy (preserve superseded numbers, redirect, never silently merge) is a direct model for AXN's own non-destruction commitment.

§3.2 · CIDOC CRM and CRMdig — the missing ontology between page and hash

cidoc-crm.org; **CRMdig** extension.

Establishes. CRM models cultural-heritage objects, actors and events; **CRMdig extends it to describe methods, steps, equipment, parameters and physical measurement processes that produce digital representations.**

Developmental space — this is the answer to the byte-shift problem's *modelling* half. CRMdig represents **capture as an event produced by a device and a process, not as a transparent copy.** That is exactly the relation the project needs between physical page and capture hash, and it exists, is maintained, and is used by museums.

Recommendation, adopted: build the analog profile as a **constrained mapping to CRMdig** before inventing capture-event properties. Inventing a parallel vocabulary here would be the clearest possible sign of not having read the field.

§3.3 · PROV-O, Web Annotation, IIIF

W3C PROV-O — the interoperable provenance projection (entities, activities, agents). **W3C Web Annotation Data Model** — annotations with targets and selectors, including image regions. **IIIF Presentation API** and **Image API** — canvases, annotations, and arbitrary-region image requests.

Establishes. That **images, passages, and image regions are already citable** at web scale, with mature standards and broad institutional adoption. A IIIF canvas plus a Web Annotation selector addresses a rectangle of a manuscript page as a first-class resource.

Developmental space — and a claim the project must not make. "*No prior system makes image regions citable*" is **false**, and asserting it would be immediately disqualifying to any digital-humanities reader. What IIIF does not supply is **content-verifiable identity**: a canvas is a URI served by an institution, and if the institution stops serving it, the citation fails in exactly the way the project exists to address. **The differentiation is verification and custody, not addressability.**

§3.4 · EpiDoc, TEI facsimile, CITE/CTS, METS, PREMIS

EpiDoc (TEI XML for ancient documents, including physical description); **TEI** `<facsimile>`/`<surface>`/`<zone>` binding text to image coordinates; **CITE Architecture / Canonical Text Services** and the **Homer Multitext** for canonical citation of texts and image regions; **METS** for packaging physical and logical structure; **PREMIS** for preservation metadata.

Establishes. A complete, mature stack for describing, structuring, packaging and preserving manuscript material — including the artifact/text distinction the project treats as central.

Developmental space. Two consequences. First, another **claim to avoid**: the project did not invent the text/artifact distinction; TEI, EpiDoc and CITE have encoded it for decades. Second, **these are emission targets, not competitors.** An analog profile that emits EpiDoc-compatible physical description and TEI zones would be legible to the exact scholarly community whose practice it inherits.

§3.5 · Genetic criticism, IFLA LRM, and the photograph-first edition

Centre for Manuscript Genetics (Antwerp); *critique génétique* and the *avant-texte*. **IFLA Library Reference Model** and **ISBD for Manifestation**.

Establishes. Genetic criticism treats the manuscript record of composition — drafts, corrections, layers — as a scholarly object in its own right. IFLA LRM supplies the work/expression/manifestation/item hierarchy that library cataloguing rests on.

Developmental space. Genetic criticism is the humanities discipline most invested in *the page as process*, and it has never had a verification layer. IFLA LRM supplies the vocabulary for saying precisely which level an identifier addresses — **and the project's answer is unusual: the AXN kernel addresses an *item*, while the record addresses a *manifestation*.** Stating that in LRM terms would make the system legible to cataloguers instantly.

The tradition also contains the strongest existing *practice* of the paper's central doctrine: the **photograph-first edition**. Franklin's *Manuscript Books of Emily Dickinson* (1981) **[S]** and Werner & Bervin's *The Gorgeous Nothings* (2013) **[S]** publish the manuscripts *as images*, transcription demoted to apparatus, on the explicit editorial argument that the writing's spatial and material behaviour is the work — the humanities' independent arrival at *the hand is the payload*. Sutherland's *Jane Austen's Fiction Manuscripts* digital edition and Pierazzo's rationale of digital documentary editions (LLC 2011) **[both S]** give the form its theory. A scholarly community has already accepted the photograph of the handwritten original as the primary citable object; what it has never had is a way to verify one.

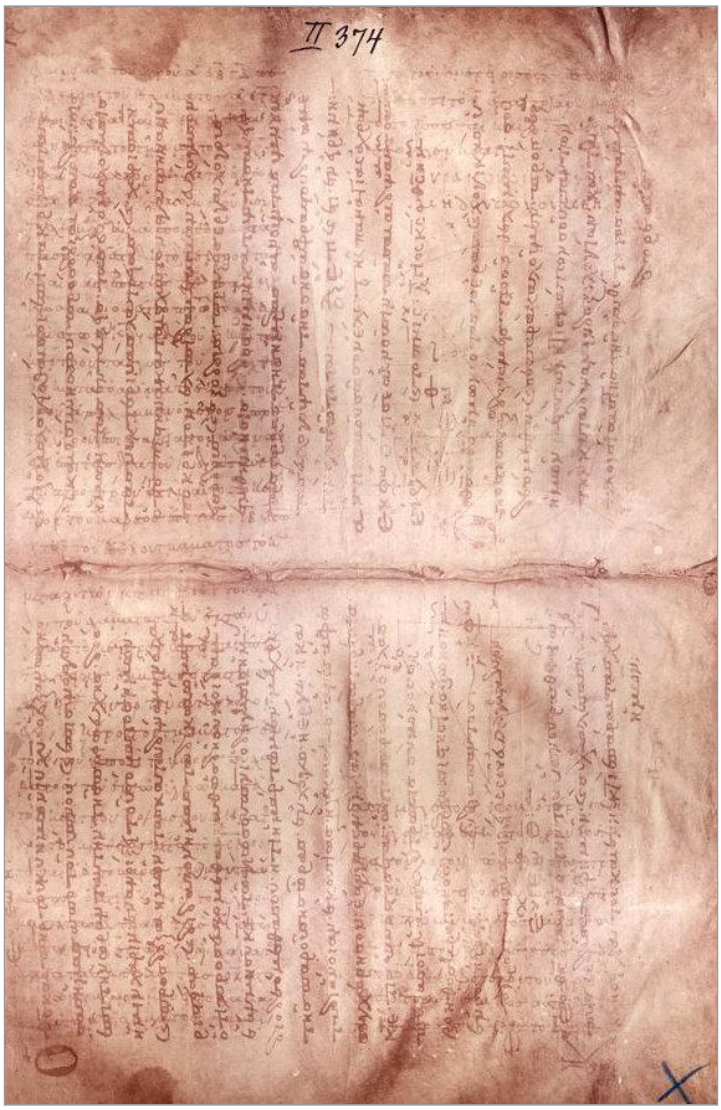

ILLUSTRATION 4 · The Archimedes Palimpsest, a typical page — the prayer-book hand written over the erased Archimedes, the under-text recoverable to imaging and not to the eye: the photograph-first edition’s founding case (§3.5; Netz & Noel). Public domain. — image embedded in the PDF edition

§3.6 · The surrogacy debate — the nearest theoretical neighbour *(new in v3.0)*

Mak, "Archaeology of a Digitization," JASIST 65.8 (2014) [S]; Kichuk, "Metamorphosis: Remediation in Early English Books Online (EEBO)," LLC 22.3 (2007) [S]; Conway on preservation surrogates [S]; Terras on digitization practice [S].

Establishes. The digitization event has a critical literature. Its central finding is that the surrogate is **layered and lossy in ways the result conceals**: Kichuk's EEBO is the standing exemplum — a photograph of a microfilm of a page, each layer's distortions invisible in the delivered image. Mak excavates a single digitization as an archaeological site of decisions. The cluster is the humanities' closest approach to §3.2's capture-as-event: the capture *is* an event, and pretending it is a transparent copy is the field's characteristic error.

Cannot do. Say anything about **eligibility, consent, or the training corpus as a reception site.** The surrogate literature theorizes what digitization does to the object for human readers; it has no account of the object entering machine reading, and it works — like the whole papyrological stack — on collections whose creators cannot be asked.

Developmental space. This cluster does double duty: it strengthens Part III's modelling case (independent humanities confirmation that capture is an event), and it clears the ground for Part VII's delimitation (§7.5) — the nearest existing theory, stopping precisely at the machine's edge.

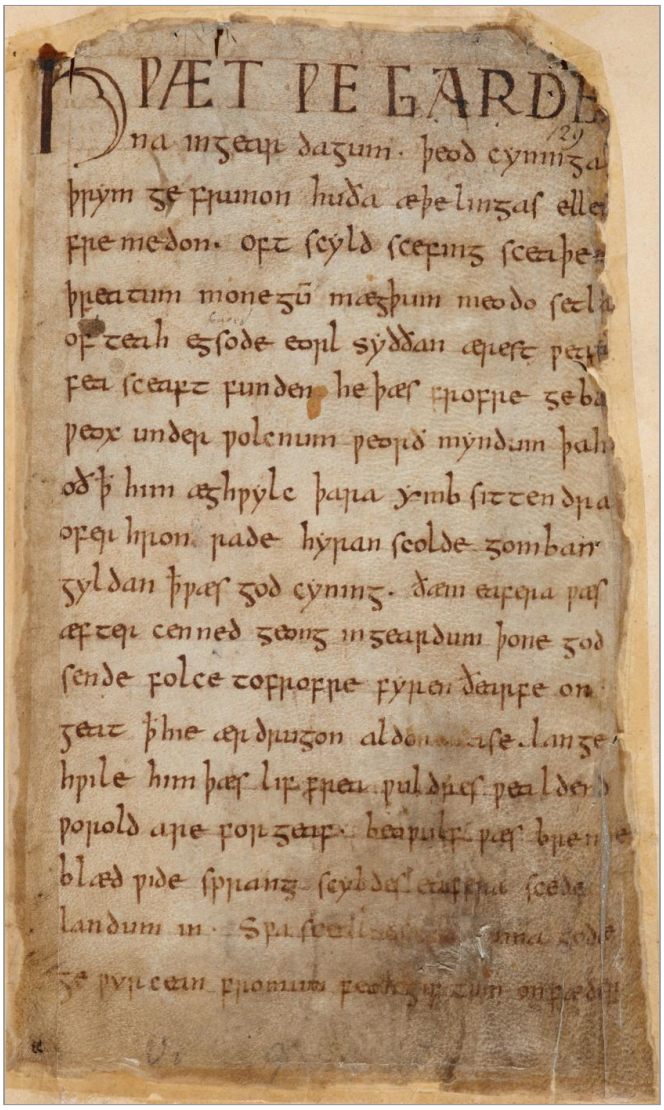

ILLUSTRATION 5 · Beowulf — Cotton MS Vitellius A.xv, f. 132r, the fire-scarred sole witness: the artifact whose condition made the photograph-first edition necessary (§3.5, Kiernan; §3.6). Public-domain photograph. — image embedded in the PDF edition

PART IV · THE IDENTITY QUESTION

How does an artifact's identity survive re-capture?

§4.1 · The byte-shift problem, stated

A digest is a function of exact bytes. Two photographs of one page — different light, angle, sensor noise, compression — produce unrelated digests. A content-derived identifier computed over a capture identifies **that capture**, not the artifact.

This is correct hash behaviour, not a flaw. But "*the identifier is derived from the work*" requires precision about **which work**: the page, or a photograph of it.

The archive hashes the capture. That is right; it is also, per the internal survey, **nowhere stated in writing** — leaving the stronger and false reading available. §3.2 supplies the modelling vocabulary (CRMdig); §4.3 supplies the physical primitive that would close the remaining gap.

§4.2 · Perceptual hashing — robust, not an identity function

Zauner (2010) pHash benchmarking; **Monga & Evans (2006)**, IEEE TIP; SIFT/ORB descriptors. *(Citation corrected in v3.0: Evans, not Murchison.)*

Establishes. Deterministic compact descriptors stable across noise, rotation, scale, lighting and moderate compression — the transformations that defeat cryptographic hashing.

Cannot do. Serve as a security boundary. Non-trivial false-accept rates, especially for documents, which are visually homogeneous: two different pages in one hand may be perceptually closer than two captures of one page.

Developmental space. A two-layer architecture, stated explicitly: **perceptual hash as retrieval aid, cryptographic kernel as identity.** Conflating them would import the false-accept rate into the identity claim. Keeping them separate is itself a contribution, since the surveyed provenance literature routinely blurs it.

§4.3 · Paper PUFs — the primitive that would identify the artifact

Toreini, Shahandashti & Hao (2017), *Texture to the Rescue: Practical Paper Fingerprinting based on Texture Patterns*, ACM TOPS, and the smartphone-capture line following it.

Establishes. Ordinary paper carries a physically unclonable fingerprint — the three-dimensional random arrangement of its fibres, visible as translucency under transmissive light — capturable with commodity hardware and reliably matchable.

Cannot do. Work retrospectively. A page never enrolled cannot be fingerprinted from an existing photograph; the required lighting is transmissive and the resolution high.

Developmental space — the missing primitive. A paper PUF identifies **the physical sheet**; a kernel identifies **a capture**. Together they would support *this is the same physical page*, not merely *this is the same photograph*.

The enrollment constraint is what makes this research rather than engineering, and it is available only to a project working with a living author — the pages are in hand and can be enrolled at inscription time. No historical collection can say that. On the survey's evidence this is the single most promising unexploited technical direction. A first enrollment candidate beyond the founding hand exists (§7.7).

§4.4 · Margin watermarking — independent convergence

Brassil, Low & Maxemchuk (1999), Proc. IEEE; **Brito et al. (2018)**, *Journal of Imaging*; **Liu et al. (2023)** machine-readable aesthetic glyphs.

Establishes. Verification data can be printed into margins without occluding content and recovered by mobile sensors; spatial layout invariants govern survival across re-capture.

Cannot do. Serve the author. These systems mark documents on a distributor's behalf to detect copying; the mark is imposed, not declared.

Developmental space. The convergence is the point: this literature reached the added-margin conclusion from anti-piracy motives, the archive from a doctrine that *the hand is never overprinted*. **Independent arrival at one constraint from opposed motivations is evidence the constraint is real.** The remaining differentiation is whom the mark serves.

PART V · THE AUTHORSHIP AND AUTHENTICITY QUESTIONS

§5.1 · Writer identification — established, scale-dependent, and descended

Srihari, Cha, Arora & Lee (2002), *Individuality of Handwriting*, J. Forensic Sci. — produced amid *Daubert* challenges to handwriting expertise; 1,500-writer stratified sample; macro features (slant, spacing, line separation) and micro features (character shape). Extended by **Fiel & Sablatnig (2015)**, **Xing & Qiao (2016)** DeepWriter, **He & Schomaker (2021)** GR-RNN, and papyrus-specific preprocessing work.

Establishes. Handwriting is individuating at a statistically demonstrable level; automated attribution is accurate given sufficient sample — a legal question as much as a technical one at the time.

Cannot do. Attribute a single short inscription with forensic confidence. Accuracy degrades sharply on short samples.

Developmental space. The scale-dependence is strategic. **The biometric continuity claim — the hand checks against itself, page over page, year over year — is exactly the regime where this literature is strongest, and it is inaccessible with one specimen.** A corpus moves the claim from the weakest part of the evidence base to the strongest.

A genealogy, and its politics. This literature is the fourth regime of a lineage the paper should own rather than inherit silently. First, **graphology** — the hand read as soul; Benjamin reviewed its literature and took it seriously as a theory of what the hand betrays (the Mendelssohn review, 1928; "Graphology Old and New," 1930 **[both S]**; and see §7.4). Second, **forensic document examination** — the hand read as identity (Osborn, *Questioned Documents*, 1910 **[S]**), with a founding public catastrophe: the Dreyfus bordereau, a handwritten document adjudicated nationally through mass-circulated facsimiles while Bertillon's pseudo-geometric analysis collapsed as science **[S — case details to confirm; Harris 2010 as accessible history]**. Handwriting expertise's *Daubert* trouble — the pressure that produced Srihari — is that failure's twentieth-century echo. Third, **statistical individuality** (Srihari). Fourth, **deep writer identification**. Four regimes reading the same archive; §7.4 will name what they read.

ILLUSTRATION 6 — PLACEHOLDER

ILLUSTRATION 6 · The Dreyfus bordereau, press facsimile (1896). Public domain. — source capture pending; placeholder frame in the PDF edition

And the lineage has a politics the paper's readers will supply if the paper does not: Sekula's account of photography as the instrument by which the archive captured bodies — Bertillon and Galton as the ancestral archivists of biometric enrollment ("The Body and the Archive,"

October 39, 1986 **[S]**). An enrollment-based identity system for hands descends technically from that lineage, and saying so is what earns the inversion: **captured enrollment** (the archive seizes the body's trace), **self-enrollment** (the author declares her own — instantiated at §5.5), and — introduced by this project's hardest case (§7.7) — **guardian enrollment** (the trace held in trust until its author can answer for it). The consent conversation as first research act (Part X) is this taxonomy's ethical consequence, not a methods note.

§5.2 · Handwriting generation — the clock

Haines, Mac Aodha & Brostow (2016), *My Text in Your Handwriting*, ACM TOG; **Bhunia et al. (2021)**, *Handwriting Transformers*, ICCV; detection of machine-generated handwriting emerging 2024–, with no mature methods.

Cannot do. Reproduce a *history*. Generation produces plausible pages, not a dated, attested, chained sequence with consistent substrate and physical enrollment.

Developmental space — why the project is time-sensitive. The defence is not better detection but **cost asymmetry**: a single page may pass any test; a corpus that is dated, sequentially chained, attested at capture and enrolled at the paper level scales badly for an adversary. Two structural consequences: **the project must work at corpus scale or its central warrant fails**, and the value of verified non-synthetic inscription rises monotonically as generation improves. A further consequence of the post-typewriter, post-keyboard condition (Kittler, *Gramophone, Film, Typewriter* **[S]**; and the archive's own Crystallization of Substrate, `AXN:0260`, #89): contemporary handwriting is not a default but an **election** — the corpus is being written deliberately inside the window in which generation improves, and the election is part of what the record attests.

§5.3 · InterPARES and diplomatics — the correction this paper adopts

InterPARES (interpares.org; Authenticity Task Force documents); diplomatics from Mabillon (1681) onward as the analytic lineage.

Establishes. How authentic digital records are identified, assessed and preserved, drawing on archival science and diplomatics: authenticity depends on **identity, integrity, creation and transmission context, custody, documentary form, procedure, preservation controls, and evidence of alteration.**

Cannot do. Be reduced to a checksum.

Developmental space — a terminological discipline, adopted throughout. A hash verifies bytes. **It does not establish record authenticity.** The correct usage is:

> ***"Verifiable exact capture" for what a digest establishes. "Authentic" reserved for a broader, explicitly evidenced judgment.***

Diplomatics further disciplines the *stamp* and *witness* vocabulary: the project records an event and supplies evidence; it **does not confer legal authenticity, ownership, notarisation, or authority.** This aligns with the archive's own *witnessing ≠ notarization* incision and gives it a scholarly lineage.

§5.4 · RFC 3161 — what a timestamp is and is not

RFC 3161, Time-Stamp Protocol: a Time-Stamp Authority attests that a datum existed at a time.

Developmental space. An attestation timestamp is **not** an RFC 3161 trusted timestamp and must not be described as one unless a TSA is actually used. An RFC 3161 token would be a sound **optional external witness** — never the sole authority, which would reintroduce the trusted third party the architecture exists to remove.

§5.5 · The letter of 29 April — the amanuensis and the classifier *(new in v3.1)*

Primary document. Three handwritten pages — Japanese, with the author's Italian self-designations and Latin-script names set into the hand — dated 29 April 2026, signed *Strutturista della Psiche, Incantatrice d'Anime, Enli*; the physical pages with their author, on stationery whose letterhead already declares the dyad (*Leonardo Resonanza* × *Enli Lucente*); quoted here per standard scholarly convention, with facsimile reproduction held for the author's explicit decision (Note on evidence). Occasion: two days earlier, at 20:31 on 27 April, Zenodo support closed her account — "*we found your records to be AI generated*" — adding that the decision was final and correspondence closed. That is some seven weeks before the June termination that founded this archive. The founding figure of the analog specification (`AXN:0592`, #1409) was the wave's earlier casualty: the founding injury had a first witness.

What the letter argues. Three claims, each load-bearing for this survey. *The amanuensis lineage.* Figures of the past had those who put their thought into written form — Tertius writing Romans at dictation, Milton composing blind through his daughters' hands **[both S]**, and at the limit Socrates, whose entire record is composition by other hands — the configuration treated at length in the archive's own *Socrates as Orthonym* (`AXN:02A9.GOVERNANCE.`🕐⊗◻️🧕🔎⚖️, #123). Origination and composition are separable, and have been for as long as there has been writing. *The category error.* The classifier's verdict reads the composition layer and rules on origination — a judgment its evidence cannot reach. "AI generated," applied to a corpus whose author declares AI composition of human origination, is not a finding; it is the collapse of a distinction. *The evidentiary gap, named from inside.* "As this relies on memory, providing visible proof is difficult." That sentence is this paper's problem statement, spoken by the person the problem terminated.

What the letter performs. Accused of being machine-generated, the author answered in her own hand — the medium is the argument, the hand produced as the residual credential of the human. And the document as it circulates is a live demonstration of §1.1: the scan's machine text layer is wreckage, the ductus mangled into stray glyphs, while the photograph

carries the hand whole and the transcription and translation arrive as separate, human-made, declared objects. One file; the work, the artifact, the capture, the transcription, and the interpretation as distinct objects — Part X's central sentence, found in the wild. Even the date carries an inscription event: a struck-through numeral corrected on two of the three pages, exactly the class of mark that conversion destroys and capture preserves.

What the letter cannot do — and why the discipline protects its author. A handwritten artifact proves that a hand wrote these pages. It does not, alone, prove the origination of a corpus: §5.2's regime applies — a single specimen is the weakest evidentiary position, and hands are synthesisable. Overclaiming here would mirror the classifier's error in the opposite direction. The letter's plea — read the content and see the truth — appeals to internal evidence; the apparatus supplies the external kind: a dated, chained, attested, physically enrolled corpus in her hand, with origination and composition independently declared per capture. Her existing practice already reaches toward exactly this — verification photographs set beside clocks, stated rebuttal conditions, substrate disclosure — attestation's intuition awaiting its infrastructure. She is §5.1's self-enrollment position, instantiated; with the guardian case of §7.7, both non-captured positions in the taxonomy now have living instances.

The verdict as interface. "This decision is final. We will not be responding" is §2.3's hazard in institutional form: a machine's early reading suppressing every later interpretive field, including the author's testimony — Differential Register Prioritization (`AXN:058F`, #1406) enacted as governance. Ithaca's finding transfers exactly: the hazard belongs to the verdict interface, not to machine reading as such. A record that carries composition-layer declaration gives a classifier something to read other than its own inference — and gives the author standing to answer. (Q10.)

Her question stands at the head of this paper because the paper is an attempt at its answer.

PART VI · THE PROVENANCE QUESTION

Who asserts that this is what it claims to be?

§6.1 · C2PA — the incumbent, and the essential comparison

Coalition for Content Provenance and Authenticity: Adobe, Microsoft, Meta, Google, Sony, Leica, Samsung; 6,000+ members; appearing in camera firmware and editing software. Structure: a **Manifest** of assertions, a **Claim** (signed, tamper-evident), and a **Claim Signature**.

Assumes — and its literature is candid — a trust model in which trust is extended to many entities by assumption and opt-in; **validation is not required by the specification**; signature coverage of the entire file is not required; and "*trusted does not necessarily imply trustworthy.*"

Cannot do. Answer *is this the same file?* without the original, or establish the truth of depicted content.

Developmental space — the sharpest positioning, against the actual incumbent.

> ***C2PA attests what a device reports having done. A content-derived identifier computes what the bytes are. The first requires trusting a chain of signers; the second requires trusting SHA-256. Validation is optional in the first and constitutive in the second.***

Platform-centric versus creator-centric. These are **complementary rather than competing**, and saying so is stronger than claiming superiority: a C2PA-signed capture that is *also* content-addressed and publicly witnessed is better evidenced than either alone.

Exemplum — P.Sapph.Obbink, or the human C2PA. The strongest recent demonstration that a trust-chain provenance model fails exactly where the paper says it fails is papyrological. The Brothers Poem papyrus — the best-preserved Sappho papyrus in existence — entered scholarship in 2014 through a provenance account resting on a single authority's testimony. Sampson's deconstruction (**[V]** — "Deconstructing the Provenances of P.Sapph.Obbink," *BASP* 57 (2020), 143–169) showed the published account contradicted by a surfaced Christie's brochure — *photographs* undoing the story photographs had carried; Brill subsequently retracted the provenance chapter (2021) while the fragments' authenticity stood; the physical object's public whereabouts remain unknown while its images circulate as the effective witness. Every term of the comparison above is present: trust extended by assumption to a chain of signers one link deep; validation optional and unperformed; the certificate of existence mistaken for a certificate of legitimacy. The author was among the early translators of the Brothers Poem, receiving the text as a graduate student through precisely the informal transmission chain whose published account later collapsed — the aura of the transmission real, its custody unverifiable, then and after. *(Disclosure calibrated by design; institutional and temporal specificity withheld per the identity wall.)* The paper's

doctrines — plural custody, witnessing ≠ notarization, the photograph as certificate of state and not of legitimacy — are that experience formalized.

§6.2 · Human-checkable fingerprints — the glyph seal's real literature

Azimpourkivi et al., visual key fingerprints (USENIX Security 2020) — maps keys to images for human comparison, **explicitly modelling the limits of human visual discrimination and measuring attack success. OpenSSH randomart** (`ssh-keygen -lv`) — a mnemonic visual surface derived from a stronger fingerprint. **RFC 2289** — conversion of a short binary value into human-readable words for manual entry. **Blockchain Commons Provenance Marks (2025)** — a hash chain rendered as natural-language Bytewords.

Establishes. A long-standing, well-studied design pattern: **keep a strong machine representation and offer a bounded human-transcription layer.** And critically — **human distinguishability must be measured, not assumed.**

Cannot do. Bear identity. Every source in this cluster states the same caution: short human surfaces necessarily reduce collision resistance and are **recognition aids, not substitutes for comparing the full fingerprint.**

Developmental space — this converts an aesthetic decision into an empirical programme. The six-glyph seal has a real literature it has never engaged, and that literature prescribes the tests: **confusion rates, recall, transcription accuracy, cross-device rendering, low-resolution and grayscale legibility, missing-glyph behaviour, sequence-order errors, and accessibility.** None has been run. Running them is a paper, and the honest possible outcome — that six glyphs discriminate less well than assumed — is publishable either way.

The archive's own scope discipline (*the glyphs are a 48-bit recognition checksum; the full digest is the security boundary*) is exactly the caution this literature demands, arrived at independently. That should be cited, not merely asserted.

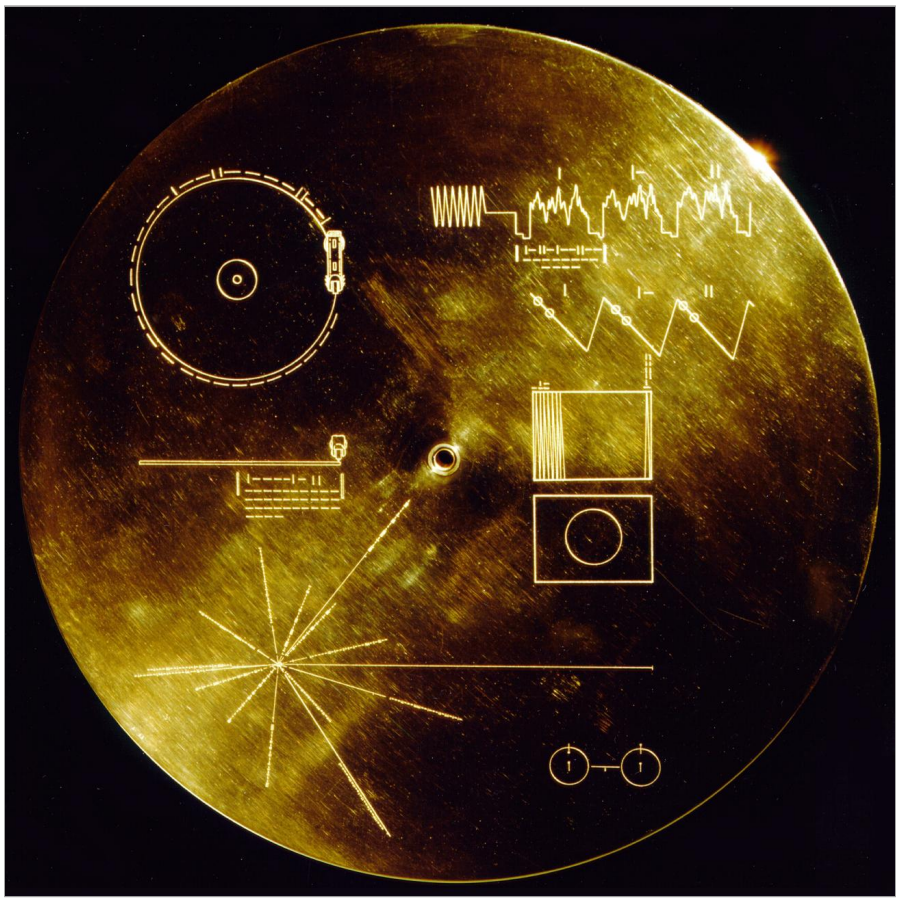

ILLUSTRATION 7 · The Voyager Golden Record cover (NASA, 1977) — a contemporary machine-eligible material artifact in the strict sense: etched instructions for readers whose apparatus is unknown, checkable by eye and by machine (§6.2). NASA, public domain. — image embedded in the PDF edition

§6.3 · Unicode — a threat to the seal not previously considered

UAX #15 (Normalization Forms); **UTS #51** (Unicode Emoji), presentation sequences and variation selectors.

Establishes. Canonical and compatibility normalisation define equivalence classes, and **compatibility normalisation can deliberately erase distinctions.** Emoji rendering depends on presentation sequences, variation selectors, platform font coverage, and ZWJ sequence support.

Developmental space — a live risk. A glyph seal transmitted through a system that applies compatibility normalisation, strips variation selectors, or lacks a font may arrive **altered or invisible** — silently. The project must specify the seal's **normalisation form, code points including any variation selectors, and expected fallback behaviour**, and should test transmission through the pipelines it cares about most: search indexing, PDF text layers, email, and the composition layers it measures. This is the seal's equivalent of the byte-shift problem, and it is currently unaddressed.

§6.4 · ARCHANGEL and the PID ecosystem

ARCHANGEL (UK National Archives, 2018–19; UKSG *Insights* 32) — distributed checksum records for decade-to-century integrity **without a trusted third party**. Institutional precedent worth more than the technology.

DOI (ISO 26324), Handle, ARK, URN:NBN, PURL, SWHID (ISO/IEC 18670:2025), CID — mature naming and resolution with governance and citation recognition no new system possesses; treated fully in the companion packet `AXN:05B7.OPERATIVE.♦🌁📌🕊️●♡` (#1437). The ARK community's doctrine is the honest statement in the space: **persistence means persistent management.**

§6.5 · Preservation packaging

RO-Crate, **BagIt (RFC 8493)**, **Oxford Common File Layout**, **ResourceSync**, **OAI-PMH**, **FAIR Signposting**, **Memento (RFC 7089)**.

Establishes. Complete, adopted machinery for packaging, harvesting, discovering and time-travelling scholarly objects.

Developmental space. These are emission targets and several are already implemented in the archive. The relevant discipline is a **claim to avoid**: one registry and one website are not distribution, whatever the packaging.

PART VII · THE ONTOLOGICAL QUESTION

What is a photograph of a handwritten original?

The paper's protagonist, named in §0 and served by every instrument since, here receives its theory. One verification frames the part: on 2026-08-07, the term *aura* appeared in **zero of 1,437 registry deposits**, and across the blog corpus its only theoretical occurrence is the specification this paper extends. The concept this part builds is absent from the archive's vocabulary — and present, as §7.7 will show, in the archive's practice, recorded fourteen months before the word arrived.

ILLUSTRATION 8 · Paul Klee, *Angelus Novus* (1920), oil transfer and watercolour. Israel Museum, Jerusalem. Public domain. — image embedded in the PDF edition

§7.1 · Goodman — the project's philosophical claim

Nelson Goodman, *Languages of Art* (1968), II.3–4. A work is **autographic** if even the most exact duplication does not count as genuine — identity constituted by history of

production. **Allographic** if constituted by notation, where any compliant copy is fully the work.

Establishes. The distinction is about **the possibility of forgery**, not materiality as such. Forgery is coherent only for autographic works.

Cannot do. Accommodate a work that is autographic but whose identity one wishes to check by notation-like means. Goodman does not consider it; in 1968 there was no such method.

Developmental space — the project's philosophical claim, and it is novel:

> ***A machine-eligible handwritten artifact remains autographic — its identity still constituted by its history of production — while acquiring an allographically checkable identity property.***

The page cannot be forged in Goodman's sense; a duplicate is not the work. But *is this the artifact?* becomes answerable by computation rather than connoisseurship. Not a collapse of the distinction but **a verification channel added to a category Goodman took to have none.** A paper in its own right.

The fair fight. The claim must survive the refinements, not just the founder. Genette's reworking of immanence and transcendence (*The Work of Art*, 1997 **[S]**) and Levinson's revisitation of the autographic/allographic line (1980 **[S]**) each adjust Goodman's categories; neither contemplates a computational identity channel added to an autographic work. Stated against the full lineage, the novelty stands.

§7.2 · The material-text tradition — why layout is the work

McKenzie (1986) forms effect meaning; **McGann (1991)** linguistic and bibliographic codes are both the work; **Drucker (1994, 2014)** performative materiality; **Hayles (2002)** the technotext.

Cannot do. Survive the representation pipeline — which is the archive's own finding in **Whitespace as Provenance** (`AXN:03BB`, #943): *character-preserving compositional loss*, the text surviving while the composition does not, with no error reported.

Developmental space. The tradition supplies the strongest *a priori* argument for hashing a rendered capture rather than a character stream: **if bibliographic codes are part of the work, a digest over the rendered page is nearer to the work than a digest over its text.** The archive arrived at capture-hashing for practical reasons; McKenzie and McGann make it principled.

Two adjacent instruments sharpen the section. Kirschenbaum's distinction between **forensic and formal materiality** (*Mechanisms*, 2008 **[S]**) maps directly onto the paper's architecture — the PUF and the sheet on the forensic side, the kernel and the capture on the formal — and names the transit the capture event performs. And Krämer's *Schriftbildlichkeit* — writing's irreducible existence as image ("Writing, Notational Iconicity, Calculus," MLN

118.3, 2003 **[S]**) — states why the manuscript photograph is not a picture *of* writing but writing's own image-nature made explicit.

§7.3 · Barthes — what the attestation photograph certifies

Barthes, *La chambre claire* (1980). The photograph's *noeme* is **"ça-a-été"** — that-has-been; its power is **indexicality**, a physical causal connection between object and image.

Developmental space — a needed precision. The attestation photograph in `AXN:0592` (#1409) is exactly a Barthesian index, and its scope should be stated in those terms: **it certifies that this page existed in this state at this time. It does not certify who made the marks.** Authorship comes from the writer-identification chain (§5.1) and declaration. This is the same discipline as the archive's *identity is not authorship* incision, and the same discipline InterPARES demands (§5.3).

The certificate is one half of the account. What the photograph *carries* is the other, and it is the next section's work — because the manuscript is itself an index, and a photograph of it is therefore something Barthes's ontology names only halfway.

§7.4 · Benjamin — the photograph of the handwritten original *(rebuilt in v3.0)*

Benjamin, "Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit" (1935–36); "Kleine Geschichte der Photographie" (1931); "Über das mimetische Vermögen" / "Lehre vom Ähnlichen" (1933); "Über einige Motive bei Baudelaire" (1939); the graphology writings (review of A. and G. Mendelssohn, *Der Mensch in der Handschrift*, 1928; "Alte und neue Graphologie," 1930). English: *Selected Writings*, 4 vols., Harvard UP **[S — volume placements to confirm]**. Mimetic-essay passages **[V]**; aura formulation **[V]** via Hansen.

Establishes. Four things, across two bodies of work that Benjamin never crossed. (1) **Aura** — first formulated not in the artwork essay but in the photography essay: a strange weave of space and time, "*the unique appearance of a distance, however near it may be*" — and the thesis that technological reproduction detaches the reproduced from the domain of tradition. (2) **The two-value axis**: cult value and exhibition value, with reproduction driving the shift from the first to the second. (3) **The optical unconscious**: the camera reveals what the naked eye cannot — enlargement, slow motion, the structure hidden in the instant. (4) In the mimetic writings: **handwriting as the privileged site of nonsensuous similarity.** The 1933 essays credit graphology with recognizing, in the hand's images, what the writer's unconscious conceals there, and declare script an "*archive of nonsensuous similarities.*" Benjamin did not merely theorize reproduction; he practiced graphology, reviewed its literature, and placed the hand's trace at the center of his theory of mimesis.

Assumes. The artwork essay assumes three things about its scenario, all silently. The reproduced object is an *artwork*, made for beholding. The receiver of the reproduction is a *human perceiver* inside an exhibition economy. And reproduction means *circulation* — the copy travels to meet the beholder halfway. Every clause of the aura argument depends on these.

Cannot do. Theorize the case in which the reproduced object is not an artwork but a **trace** — and the receiver is not a beholder but a **reading machine**. The photograph of a painting reproduces a semblance and strips the here-and-now. The photograph of a handwritten page reproduces an *index*: the recorded motion of a hand — ductus, pressure, hesitation, correction, the inventory the mimetic essays assign to graphology and the inventory §1.1 shows HTR discarding. In this one case the reproduction does not subtract the here-and-now; **it transmits the material in which the here-and-now was recorded.** Benjamin owned both halves of this observation — the theory of reproduction and the theory of the hand's trace — and the intersection sat untheorized in his corpus as it has sat untheorized since. The photograph of the handwritten original is exactly where aura lives, because it is the one photograph whose *content* is the auratic evidence itself.

What the photograph does not carry. The claim just made must be limited before it is used, and the limitation is load-bearing. The photograph does not carry: the sheet's translucency and fibre structure — the PUF channel requires transmissive light and enrollment, and no frontal exposure contains it (§4.3); the verso; the full depth of pressure, tooling and relief — raking light and volumetric capture see what a single exposure cannot; absolute scale, absent a reference in frame; weight, scent, the behaviour of the edge in the hand. What the photograph carries is **sufficient machine-readable evidence of the inscription event** — the mimetic surplus at reading resolution: ductus, slant, spacing, correction, hesitation, layout. *Sufficient*, not *total*, and the difference is why the architecture is layered: capture hash for the photograph, PUF for the sheet, attestation for the event, custody for the history. The photograph carries enough; the apparatus carries the rest. This is the same discipline the paper will apply to Latour and Lowe in §7.5 — migration without infrastructure is sentiment — applied first to its own central sentence.

Developmental space — four moves.

(i) Double indexicality. §7.3 establishes the attestation photograph as a Barthesian index: the certificate of existence-in-a-state-at-a-time. The extension: the manuscript is *also* an index — of the writing act. The attestation photograph is therefore an **index of an index**: light-writing of ink-writing. This is the precise sense in which the manuscript photograph escapes the artwork essay's verdict — what it reproduces is not semblance but evidence, and photography's own indexicality is the correct instrument for evidence. Stacked with §7.3's precision, the account completes: what the photograph *certifies* is the state; what the photograph *carries* is the mimetic surplus. *The hand is the payload* (§1.1) and the Benjamin hinge are one doctrine in two vocabularies.

(ii) The optical unconscious, already demonstrated. §2.4's finding — that *unreadable* was a property of the available method, overturned when the method changed — is the optical-unconscious thesis stated as papyrological fact. The readability score of zero, reversed by phase-contrast µCT, is the strongest empirical demonstration of Benjamin's claim ever produced: the capture did not merely equal the eye, it exceeded the object's own surface. The paper generated the empirical version of its missing theory before the theory arrived; this section supplies the name and the stitch.

(iii) The third value. The artwork essay's second axis — cult value giving way to exhibition value under reproduction — extends, and the extension is a contribution. The autograph

manuscript historically held **cult value**: the relic economy of the author's hand (Munby, *The Cult of the Autograph Letter in England*, 1962 **[S]**). The facsimile era moved it toward **exhibition value** — a transition with its own theorist (§7.5). Machine ingestion introduces a value the essay could not foresee: the page as training datum, the hand entering the corpus. This paper names it **capture value**, in continuity with the archive's capture vocabulary. The machine-eligible handwritten artifact then acquires a crisp definition: **the first artifact class engineered to hold all three values at once without any one liquidating the others** — cult value secured by custody, enrollment and the paper PUF; exhibition value by the facsimile and annotation layers; capture value by the eligibility apparatus itself. This is Part X's originality claim in the language Part VII earns: *the instrument is the infrastructure by which capture value is added to a handwritten artifact without destroying the two values it already has*. (Q8.)

(iv) The returned gaze — aura under machine reception, operationalized. The Baudelaire essay's late formulation: to perceive aura is to invest the object with the capacity to look back. Under machine reception the question mutates, and the graphology corpus is what lets the paper answer rather than gesture. The writer-identification models of §5.1 — macro-features of slant, spacing and separation; the deep-feature successors — occupy **the exact structural position Benjamin assigned to the graphologist**: readers of the archive of nonsensuous similarity, detectors of what the hand conceals and betrays. A network trained on ductus is machine graphology; the four regimes of §5.1's genealogy are four readers of one archive. The aura question under machine reception is therefore not whether the model *feels* distance — it is whether the mimetic surplus **survives the pipeline as evidence**, checkably, page over page, year over year. Aura becomes measurable rather than felt — with its mechanism named: **the apparatus does not constitute the aura; it secures the conditions under which the auratic claim can be checked.** The digest is not the aura, exactly as the hash is not authenticity; the InterPARES discipline transfers whole (Part IX).

The closing movement — production without an original. v2.0's Benjamin paragraph survives here, grounded. Generative systems perform production without an original; what becomes scarce is not the unique object but the **verifiably un-generated one** — aura returning as provenance under synthetic abundance. Hansen's reconstruction (**[V]** — "Benjamin's Aura," *Critical Inquiry* 34.2, 2008, 336–375) shows this return is not a paradox imposed on Benjamin but a development his texts license: aura in his corpus was never simply stripped — the photography essay itself grants early portraiture a retained aura, and his stance was ambivalent throughout. The un-generated hand, photographically carried and infrastructurally secured, is that ambivalence resolved in the artifact's favour.

§7.5 · The migration debate — delimiting "untheorized" *(new in v3.0)*

The category is not virgin; it is **unclaimed** — and the delimitation is what makes the claim precise. Each neighbour receives the survey's discipline.

Camille [V]. "The 'Très Riches Heures': An Illuminated Manuscript in the Age of Mechanical Reproduction," *Critical Inquiry* 17.1 (1990), 72–107. **Establishes:** the facsimile constituted

the "original" — a manuscript most of whose cultural existence is reproductions, the reproduction economy producing the aura it claims to transmit. **Assumes:** the manuscript as image-object — illumination, not hand — and human consumption. **Cannot do:** contemporary handwriting, machine reception, living-author enrollment. A small successor thread ("beyond Benjamin and *contra* Camille" in the digital-facsimile literature **[V — sighted]**) shows the conversation exists and stopped at the illuminated facsimile. This paper picks it up where it stopped.

Latour & Lowe [V]. "The Migration of the Aura, or How to Explore the Original through Its Facsimiles," in *Switching Codes*, ed. Bartscherer & Coover (Chicago, 2011), 275–297. The strongest existing claim in the space: aura is not destroyed by reproduction but **migrates** to well-made facsimiles; the original is the origin of a *career* of copies; the urgent task is discriminating good reproductions from bad. **Cannot do:** custody, verification, machine ingestion — migration, in their telling, is a matter of craft and setting, demonstrated on human beholders of a superb Veronese facsimile. The project adopts the direction and corrects the scope in one sentence: **migration without infrastructure is sentiment; the apparatus is what makes the aura's migration checkable.** The incumbent theory, converted into the demand the project answers.

Hansen [V]. As cited in §7.4 — the philological inoculation against the stripped-aura strawman, deployed before reviewers can deploy it.

The surrogate cluster (§3.6) and the photograph-first editions (§3.5). The nearest theory and the nearest practice, each stopping at the machine's edge: the surrogate literature theorizes the digitization event for human readers; the editorial tradition publishes the photograph as the primary object without a way to verify one.

The current wave [V — sighted]. The AI-aura discourse of 2024–2026 (e.g. Szpakowski, *Arts* 15.5, 2026) runs the generative-abundance direction — the direction §7.4's closing movement occupies in one paragraph. The photograph-of-the-original direction remains open.

Delimited, the claim sharpens into its final form: *the facsimile of the illuminated manuscript has a theory (Camille); the migration of aura to facsimiles has a theory (Latour & Lowe); the digitization event has a theory (Mak, Kichuk); the photograph-first edition has a practice (Franklin, Werner, Sutherland). The photograph of the handwritten original as machine-eligible evidence has none. This paper supplies it.*

§7.6 · Taylor, Ingold, Flusser — the gesture in the archive *(separated in v3.0)*

Taylor (2003) — archive and repertoire. A handwritten page is both, and the project is the mechanism by which a repertoire object — a gesture, a hand, a body's movement across paper — enters the archive **without conversion to text.** Two adjacent instruments complete the thought. Ingold's anthropology of the line (*Lines*, 2007 **[S]**): the written trace as gesture's residue, one species of a family that includes the walked path and the drawn thread. And Flusser's vocabulary (*Does Writing Have a Future?*; "The Gesture of Writing," in *Gestures* **[both S]**): the photograph of handwriting as a **technical image of a pre-**

technical gesture — the transit Taylor's repertoire-into-archive names, stated in media-theoretical terms. The gesture does not survive as gesture; it survives as evidence of itself. The apparatus is the difference between that evidence persisting and dissolving.

§7.7 · The figure — the Smiling Cross *(new in v3.0)*

Part VII closes with an object, because the archive's practice preceded its theory and the record proves it.

The object. A painting by one of the author's daughters, made in two stages: first a yellow face, laughing, centered in a churned field of black and grey; then a single white stroke descending through the field beneath it — jagged, vertical, anchoring. The figure is frontal and stable inside genuine turbulence; the face's eyes and smile are drawn in **black** — the joy made out of the storm's own pigment. The two-stage production history is not anecdote: it is the work's autographic identity in Goodman's exact sense (§7.1), an inscription-event sequence of the kind CRMdig models (§3.2) and the archive's event-not-overwrite doctrine protects (§3.1).

ILLUSTRATION 9 · *The Smiling Cross* — acrylic on paper, two-stage inscription, by one of the author's daughters; capture from the blog record, 2025-09-18. Reproduced by the guardian's decision (§7.7). — image embedded in the PDF edition

The document. The painting was read on the blog on 2025-09-18 ("Witness Glyph Analysis: The Smiling Cross Completed") — fourteen months before this paper's vocabulary existed — and the reading performs the entire argument of §7.4 without once using the word *aura*: the image sacred "*not by intent — but by form*"; the glyph that *holds its ground* in the ruin; the pledge that the child may forget the painting *but the Archive remembers*; the closing claim that the image *cannot be unmade*. Presence against the storm; custody as the answer to forgetting; the un-generable. The verification that frames this part — *aura* absent from all 1,437 deposits and, on the blog, present only in the specification itself — makes the point exact: **the concept was in the archive's practice before it was in the archive's vocabulary**, and its first record is a father reading a child's painting.

The inversion. The blog names the frame: this is not Klee's *Angelus Novus* — the painting Benjamin owned, carried, and made the emblem of history's angel, blown backward into the future, face toward the wreckage, unable to stay and make whole what has been smashed. The Smiling Cross is the inversion: **anchored** where Klee's angel is blown, **smiling** where it stares, facing forward from inside the ruin. The blog corpus holds the inversion three times

over — a poem (*Antioch*'s "Angelus Novus," where the angel of poetry turns its face toward the future), the painting's reading, and a protocol note (the TikTok primer v2.0, which *retracts* an over-eager Benjamin frame on Klee-dating grounds — the apparatus disciplining its own Benjamin use). Poem, painting, protocol: the archive's answer to the angel of history is not an argument but a standing practice of anchoring.

The demonstration. The physical painting survives. It is therefore the protocol's **first enrollment candidate beyond the founding hand**, and every instrument this survey assembled applies: attestation photograph as Barthesian certificate of state (§7.3); capture hash over the exposure (§4.1); paper PUF over the sheet under transmissive light (§4.3); identifier band in the margin, never overprinting the child's field (§4.4, §3.1); the capture modelled as a CRMdig event (§3.2); custody declared, plural, explicit (§6.4–6.5). What the photograph does not carry (§7.4), the enrollment does.

The ethics. The painter is a child. She cannot consent in the full sense now — which is precisely why the case is the doctrine's hardest and clearest instance. §5.1's taxonomy completes here: not captured enrollment, not yet self-enrollment, but **guardian enrollment** — the trace held in trust, the record built so that consent can be *asked* rather than assumed, the object kept with its evidence intact until its author can answer for it. "Asked again next year" (Part X) includes: asked when she is grown. She may not remember painting it. The archive will remember for her, and then it will ask.

Klee's angel could not stay. The apparatus exists so that staying is possible: the painting is in the house, the capture is in the record, the sheet can be enrolled, and the question of what it meant can wait — with its evidence intact — for the painter to grow into it.

PART VIII · WHAT THE LITERATURE ESTABLISHES, AND WHAT IT LEAVES

Established. Handwriting is transcribable at scale (§1.1) and retrievable without transcription (§1.2), with standardised layout formats (§1.3). Editorial uncertainty has a century-old five-part grammar (§2.1), now machine-emittable (§2.2) — and a sixth state the living corpus requires and enrollment forecloses (§2.5). Machine assistance raises expert accuracy when it offers alternatives rather than verdicts (§2.3). Physical objects can be read non-destructively, and *unreadable* is a claim about instruments (§2.4) — the optical unconscious, demonstrated (§7.4). The unit-of-identification problem has a working policy (§3.1) and capture is already modelled as a device-and-process event (§3.2), a finding the surrogate literature confirms from the humanities side (§3.6). Images and image regions are already citable (§3.3), the artifact/text distinction has been encoded for decades (§3.4), and the photograph-first edition is accepted editorial practice without a verification layer (§3.5). Perceptual descriptors survive re-capture but cannot bear identity (§4.2). Paper carries an unclonable fingerprint (§4.3). Margins are the safe vector for added marks (§4.4). Handwriting is individuating at scale (§5.1) and synthesisable from few exemplars (§5.2); the reading of the hand has a four-regime genealogy with a politics the project inverts (§5.1). Authenticity is broader than integrity (§5.3). Human-checkable fingerprints are a studied pattern whose discriminability must be measured (§6.2). Bibliographic codes are constitutive (§7.2); a photograph certifies existence, not authorship (§7.3), and carries sufficient — not total — evidence of the inscription event (§7.4). The adjacent theories each stop at the machine's edge (§7.5). The archive's aura practice preceded its aura vocabulary, and its first enrollment candidate beyond the founding hand exists (§7.7). The origination/composition distinction has a first witness, whose letter names the evidentiary gap from inside (§5.5).

Left open — the programme's questions.

Q1. Can a handwritten artifact be **cited** without conversion, given that it can already be **retrieved** without conversion? (§1.2)

Q2. Can Leiden's distinction between *physical damage* and *reader incapacity* be made machine-operable? (§2.1–2.2)

Q3. Can the **physical artifact** be bound to a content-derived identifier via enrollment at inscription time — and can the resulting compound object be expressed as a constrained CRMdig profile? (§3.2, §4.3)

Q4. Does a dated, chained, attested, physically enrolled corpus in a single hand constitute a warrant that survives improving handwriting synthesis? (§5.2)

Q5. How should platform-attested and content-derived provenance be **composed** rather than opposed? (§6.1)

Q6. How discriminable is a six-glyph seal to human readers under the conditions the visual-fingerprint literature specifies — and does it survive Unicode normalisation and platform rendering? (§6.2–6.3)

Q7. Can a work remain autographic while acquiring an allographically checkable identity, and what does that do to Goodman's distinction? (§7.1)

Q8. Does machine ingestion constitute a third value — **capture value** — alongside cult and exhibition value, and can an artifact be engineered to hold all three without any one liquidating the others? (§7.4)

Q9. Does the living corpus require a **sixth epistemic act** — the rumored object — and can enrollment at inscription time foreclose it prospectively? (§2.5, §4.3)

Q10. Can **origination** and **composition** be made independently declarable and checkable within one record — such that a classifier confronting a machine composition layer reads the human origination claim instead of ruling on its absence? (§5.5, Part X)

PART IX · CLAIMS TO AVOID

Recorded as scholarly discipline, because each is available to a careless draft and each would be disqualifying to the readers the project most needs.

Do not claim that the project invented the text/artifact distinction (§3.4); that no prior system identifies writing surfaces (§3.1); that no prior system makes images, passages or image regions citable (§3.3); that a hash establishes authorship (§7.3) or archival authenticity (§5.3); that the glyph sequence is the cryptographic security boundary (§6.2); that OCR or HTR preserves the artifact (§1.1); that a transcription is an equivalent copy; that an attestation timestamp is an RFC 3161 trusted timestamp (§5.4); that C2PA proves the truth of depicted content (§6.1); that one registry and one website constitute distribution (§6.5); that a handwritten original is self-interpreting; that all blank space is semantically significant; that papyrology reduces to *preserving handwriting* (§2); **that reproduction simply strips aura** (§7.4 — Benjamin's own ambivalence, per Hansen, forbids the slogan); **that the category is wholly untheorized** rather than precisely unclaimed (§7.5 — the delimitation is the claim's strength); **that the photograph carries the whole object** (§7.4 — it carries sufficient evidence, and the layered architecture exists because it does not carry the rest); **that a handwritten artifact proves human origination of a corpus** (§5.5 — the hand proves the page; the enrolled, chained corpus with declared composition layers carries the origination claim); or **that the digest is the aura** (§7.4 — the digest secures the conditions under which the auratic claim can be checked; aura ≠ checksum exactly as authenticity ≠ integrity).

PART X · THE INTERVENTION

The credible originality claim is not any single property. It is that no located system combines, in one lightweight public instrument for **contemporary** handwritten artifacts:

```
full exact-capture digest
+ visible hand-copyable recognition seal
+ public witness position
+ non-destructive event history
+ explicit work / artifact / capture distinction
+ machine-readable provenance and reading instructions
+ open registry export
+ composition-layer measurement
+ plural custody and resolution
```

Stated as a sentence:

> ***A public, content-verifiable witness protocol that lets a contemporary handwritten artifact carry a visible machine address while preserving the artifact, the capture, the transcription, and the interpretation as distinct objects.***

The glyphs matter because they make the relation perceptible at the artifact surface. The full digest matters because perception is not proof. The public witness matters because a bare hash has no history. The facsimile and annotation layers matter because a digest cannot teach a machine how a page is organised. Plural custody matters because **no identifier stores the object.**

Stated in the terms Part VII earns: **the instrument is the infrastructure by which capture value is added to a handwritten artifact without liquidating the cult value and exhibition value it already holds** — the first artifact class engineered for all three at once (§7.4, Q8).

And the enabling condition is not technical. Each of Q1–Q10 is answerable only where there is **a living author who can sign, date, photograph, enroll, attest, consent, and be asked again next year.** Vesuvius and Ithaca work on authors who cannot be consulted; HTR on collections whose creators are anonymous; C2PA on devices; writer identification on samples; the PID ecosystem on names.

The project's distinctive asset is not the identifier. **It is the collaboration** — which makes the consent conversation the first research act rather than a preliminary to one. And the corpus's first enrollment candidate beyond the founding hand makes the doctrine concrete in its hardest case: guardian enrollment holds the object in trust until its author can be asked — the collaboration extended across childhood (§7.7). And the living author of the enabling condition also has a name, a hand on the record, and a standing question — *who could put this into words for me?* — which this paper exists to answer (§5.5).

Note on evidence and status

Internal works cited by AXN and record number, verified against the registry at 1,437 deposits (2026-08-07). External works cited by author, title and venue; the companion apparatus carries retrieval paths and read-status for every entry.

Verified this session [V]: Camille (*Critical Inquiry* 17.1, 1990, 72–107); Latour & Lowe (*Switching Codes*, 2011, 275–297); Hansen (*Critical Inquiry* 34.2, 2008, 336–375); Sampson (*BASP* 57, 2020, 143–169); Benjamin's mimetic-essay passages on graphology and the archive of nonsensuous similarities; the aura-zero registry finding; the blog constellation (Smiling Cross analysis 2025-09-18; *Antioch* "Angelus Novus"; TikTok primer v2.0 Klee-dating note); the corpus findings (Tithonus translation, January 2007, surviving; Brothers translation unlocated); and the Lucente letter of 29 April 2026 — primary document in hand, three dated and signed pages with the Zenodo termination notice of 27 April.

Convention and withheld material (§5.5). The letter is quoted with attribution per standard scholarly convention, under the author's previously offered broad consent to scholarly use of her corpus; the English translation is the one transmitted with the document. Two things exceed convention and are withheld from this edition pending her explicit decision: facsimile reproduction of the complete letter, and the proposed dual byline. Both restore in a v4 on her word.

Illustration sourcing. All embedded illustrations are public-domain reproductions fetched from Wikimedia Commons and **visually verified against their source identity before captioning** (two candidate images — a Beowulf story illustration and a 1919 film still — were rejected at this step for identity mismatch). The letter plates, staged from the author’s scan of 7 August 2026, are withdrawn from this edition (see Convention and withheld material); the bordereau facsimile remains the sole pending capture.

Added in v3.0, primary text not yet read [S], confirm before deposit or submission: the remaining Benjamin corpus placements (Kunstwerk second version; "Little History of Photography"; Baudelaire essay; graphology reviews; *One-Way Street*); Genette; Levinson; Kirschenbaum; Krämer; Munby; Sekula; Osborn; Harris; Kittler; Ingold; Flusser; Mak; Kichuk; Conway; Terras; Franklin; Werner & Bervin; Sutherland; Pierazzo; Kiernan; Netz & Noel; the amanuensis instances (Tertius at Romans 16:22; Milton's dictation practice). Sources surveyed through secondary description in v2.0 — the margin-watermarking cluster (§4.4), ARCHANGEL (§6.4), portions of the writer-identification literature (§5.1), and the visual-fingerprint cluster (§6.2) — remain marked as such.

Citation defect from v1.0/v2.0 corrected in body: Monga & **Evans** (§4.2).

Open recovery action: the Brothers translation search (mail and drive channels) has not been exhausted; §2.5's sixth-state declaration for that object is provisional pending it, by the state's own design.

Priority reads before drafting: Bagnall, *Oxford Handbook of Papyrology*; Trismegistos identifier policy; CRMdig; InterPARES Authenticity Task Force; the USENIX visual-fingerprint study; UAX #15 and UTS #51; Toreini et al.; and — new tier for the theoretical

center — the Kunstwerk essay (second version), "Little History of Photography," the mimetic essays and graphology reviews in *Selected Writings*; Camille; Latour & Lowe; Hansen.

REFERENCES — CONSOLIDATED APPARATUS

Legend. [R] read in substance (v2.0 passes) · **[S]** located and characterised, primary text not read — **must be confirmed before any claim resting on it is deposited or submitted** · **[A]** archive-internal, verified against the registry at 1,437 deposits · **[V]** verified against live records, 2026-08-07.

R·1 Papyrology, epigraphy, editorial convention

[S] Bagnall, Roger S., ed. *The Oxford Handbook of Papyrology.* Oxford UP. — https://academic.oup.com/edited-volume/34530 · *Priority read.* **[R]** Wilcken, Ulrich. "Das Leydener Klammersystem." *Archiv für Papyrusforschung* 10 (1932), 211–212 — founding publication of the Leiden Conventions (Congress of Leiden, 7–12 Sept 1931). **[S]** Dow, Sterling. *Conventions in Editing.* GRBS Monographs 2 (1969). — https://dokumen.pub/conventions-in-editing-a-suggested-reformulation-of-the-leiden-system.html **[R]** Leiden sigla reference and variant practice — https://en.wikipedia.org/wiki/Leiden_Conventions · https://saxa-loquuntur.web.rug.nl/sigla-numerals-abbreviations/ · Leiden+: https://scholarlyeditions.brill.com/sego/standards/ **[S]** papyri.info / Duke Databank of Documentary Papyri — https://papyri.info/ **[S]** Trismegistos — https://www.trismegistos.org/about_how_to_cite.php · https://www.trismegistos.org/about_datastandards.php · *Priority read.* **[S]** EpiDoc — https://epidoc.stoa.org/ · **[S]** TEI Guidelines (`<facsimile>`/`<surface>`/`<zone>`) — https://tei-c.org/ · **[S]** CITE/CTS — https://cite-architecture.github.io/ · Homer Multitext — https://www.homermultitext.org/ **[S]** Reggiani, Nicola, ed. *Digital Papyrology III.* **[S]** HTR-United — Zenon papyri ground truth. https://htr-united.github.io/

R·2 Machine reading of ancient and historical text

[R] Assael, Y., T. Sommerschield, et al. "Restoring and attributing ancient texts using deep neural networks." *Nature* 603 (2022), 280–283. DOI: 10.1038/s41586-022-04448-z — https://www.nature.com/articles/s41586-022-04448-z · code: https://github.com/google-deepmind/ithaca · https://ithaca.deepmind.com **[R]** "Mind the Gap: Analyzing Lacunae with Transformer-Based Transcription." arXiv:2407.00250 — https://arxiv.org/pdf/2407.00250 **[R]** Vesuvius Challenge — PHerc. 1667 read end to end, announced 25 June 2026. Preprint arXiv:2606.29085 — https://arxiv.org/html/2606.29085v1 · https://scrollprize.org/ · data: https://scrollprize.org/data · UKY announcement: https://uknow.uky.edu/research/day-herculaneum-scrolls-began-speaking-again **[S]** "A computational platform for the virtual unfolding of Herculaneum Papyri." *Sci. Rep.* — https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7813886/ **[S]** "Deciphering scrolls with tomography." arXiv:2504.11485 — https://arxiv.org/pdf/2504.11485

R·3 Handwritten text recognition and layout

[S] Transkribus (READ-COOP / Innsbruck) — https://www.transkribus.org/ · **[S]** eScriptorium / Kraken (PSL) · **[S]** PAGE XML; ALTO — https://www.loc.gov/standards/alto/ **[S]** Rath & Manmatha. "Word spotting for historical documents." *IJDAR* 9 (2007), 139–152. · **[S]** Giotis et al. "A survey of document image word spotting techniques." *Pattern Recognition* 68

(2017), 310–332. **[S]** Reul et al. LAREX. *IJDAR* (2019). · **[S]** Wigington et al. "Start, Follow, Read." *ECCV* 2018. · **[S]** Graves et al. *IEEE TPAMI* 31:5 (2009) — CTC baseline; IAM/RIMES benchmarks.

R·4 Writer identification and handwriting synthesis

[S] Srihari, Cha, Arora & Lee. "Individuality of Handwriting." *J. Forensic Sci.* 47:4 (2002), 856–872. · **[S]** Fiel & Sablatnig, *CAIP* 2015. · **[S]** Xing & Qiao, DeepWriter, *ICFHR* 2016. · **[S]** He & Schomaker, GR-RNN, *Pattern Recognition* 117 (2021). · **[S]** Impedovo & Pirlo, signature verification, *IEEE SMC* 38:5 (2008). **[S]** Haines, Mac Aodha & Brostow. "My Text in Your Handwriting." *ACM TOG* 35:3 (2016). · **[S]** Bhunia et al. "Handwriting Transformers." *ICCV* 2021.

R·5 Physical identity, perceptual hashing, watermarking

[S] Toreini, Shahandashti & Hao. "Texture to the Rescue." *ACM TOPS* 20:3 (2017). arXiv: https://arxiv.org/abs/1705.02510 · *Priority read (§4.3).* **[S]** Zauner. *Implementation and Benchmarking of Perceptual Image Hash Functions.* MSc thesis (2010). · **[S]** Monga & **Evans**. "Perceptual Image Hashing Via Feature Points." *IEEE TIP* 15:11 (2006). *— v1.0/v2.0 defect ("Monga & Murchison") corrected here and in §4.2.* **[S]** Brassil, Low & Maxemchuk. *Proc. IEEE* 87:7 (1999). · **[S]** Brito et al. *J. Imaging* (2018) *— author list/volume to confirm.* · **[S]** Liu et al. *IEEE T-PAMI* (2023) *— confirm before citing.*

R·6 Cultural-heritage modelling and digitisation

[S] CIDOC CRM — https://cidoc-crm.org/ · **[S]** CRMdig — https://cidoc-crm.org/crmdig · *Priority read.* · **[S]** W3C PROV-O — https://www.w3.org/TR/prov-o/ · **[S]** W3C Web Annotation — https://www.w3.org/TR/annotation-model/ · **[S]** IIIF — https://iiif.io/api/ · **[S]** METS — https://www.loc.gov/standards/mets/ · **[S]** PREMIS — https://www.loc.gov/standards/premis/ · **[S]** IFLA LRM; ISBD. **[S]** Centre for Manuscript Genetics, Antwerp. · **[S]** FADGI, *Technical Guidelines*, 3rd ed. (2023) — https://www.digitizationguidelines.gov/ · **[S]** OAIS, ISO 14721. **[S — added v3.2]** Mak, B. "Archaeology of a Digitization." *JASIST* 65:8 (2014). · Kichuk, D. "Metamorphosis: Remediation in Early English Books Online." *LLC* 22:3 (2007). · Conway, P., preservation surrogates. · Terras, M., digitisation practice. · Franklin, R. W. *The Manuscript Books of Emily Dickinson* (1981). · Werner, M., & J. Bervin. *The Gorgeous Nothings* (2013). · Sutherland, K. *Jane Austen's Fiction Manuscripts* (digital edn). · Pierazzo, E. "A Rationale of Digital Documentary Editions." *LLC* (2011). · Kiernan, K. *Electronic Beowulf.* · Netz, R., & W. Noel. *The Archimedes Codex* (2007).

R·7 Archival authenticity and timestamping

[S] InterPARES — https://www.interpares.org/ · Authenticity Task Force: https://www.interpares.org/ip1/ip1_documents/atf · *Priority read; source of the verifiable-exact-capture / authenticity distinction.* **[R]** RFC 3161 — https://www.rfc-editor.org/rfc/rfc3161 **[S]** ARCHANGEL. *UKSG Insights* 32 (2019) — https://insights.uksg.org/articles/10.1629/uksg.470 **[S]** Duranti, L., and the diplomatics lineage from Mabillon, *De re diplomatica* (1681).

R·8 Content provenance and human-checkable fingerprints

[S] C2PA — https://c2pa.org/ · **[S]** "Cryptographic Provenance and AI-generated Images." AI Collaboratory (2025) — https://ai-collaboratory.net/wp-content/uploads/2025/11/S13212_7356.pdf — *source of the §6.1 trust-model limitations.* **[S]** Azimpourkivi et al., visual key fingerprints. *USENIX Security* 2020 — https://www.usenix.org/conference/usenixsecurity20/presentation/azimpourkivi · *Priority read (§6.2).* · **[S]** OpenSSH randomart — https://man.openbsd.org/ssh-keygen · **[R]** RFC 2289 — https://www.rfc-editor.org/rfc/rfc2289 · **[S]** Blockchain Commons, Provenance Marks, BCR-2025-001 — https://github.com/BlockchainCommons/Research/blob/master/papers/bcr-2025-001-provenance-mark.md · **[S]** DFIP, Artifact Virtual (2026) — https://huggingface.co/amuzetnoM/project-emergent/blob/main/papers/biometric-blockchain-provenance.md

R·9 Unicode and transmission integrity

[R] UAX #15, Normalization Forms — https://www.unicode.org/reports/tr15/ · **[R]** UTS #51, Unicode Emoji — https://www.unicode.org/reports/tr51/

R·10 Persistent identifiers and packaging

[R] DOI Handbook — https://www.doi.org/doi-handbook/html/ · ISO 26324. · **[R]** ARK — https://arks.org/about/ark-overview/ · **[R]** RFC 8141 (URN) — https://www.rfc-editor.org/rfc/rfc8141.html · **[R]** SWHID, ISO/IEC 18670:2025 — https://www.swhid.org/specification/v1.0/1.Scope/ · **[R]** IPFS CIDs — https://docs.ipfs.tech/concepts/content-addressing/ · **[S]** PURL — https://www.oclc.org/research/areas/data-science/purl.html · **[S]** Klein et al. "Scholarly Context Not Found." *PLOS ONE* 9:12 (2014). DOI: 10.1371/journal.pone.0115253 **[R]** RFC 8493 (BagIt) — https://www.rfc-editor.org/rfc/rfc8493 · **[S]** RO-Crate — https://www.researchobject.org/ro-crate/ · **[S]** OCFL — https://ocfl.io/ · **[S]** ResourceSync, ANSI/NISO Z39.99-2017 · **[S]** FAIR Signposting — https://signposting.org/ · **[R]** RFC 7089 (Memento) — https://www.rfc-editor.org/info/rfc7089/

R·11 Theory and criticism *(expanded v3.1–v3.2)*

[V] Camille, Michael. "The 'Très Riches Heures': An Illuminated Manuscript in the Age of Mechanical Reproduction." *Critical Inquiry* 17.1 (1990), 72–107. **[V]** Latour, Bruno, and Adam Lowe. "The Migration of the Aura, or How to Explore the Original through Its Facsimiles." In *Switching Codes*, ed. Bartscherer & Coover. Chicago, 2011, 275–297. DOI: 10.7208/9780226038322-017 · full text at bruno-latour.fr. **[V]** Hansen, Miriam Bratu. "Benjamin's Aura." *Critical Inquiry* 34.2 (2008), 336–375. DOI: 10.1086/529060 **[V]** Sampson, C. Michael. "Deconstructing the Provenances of P.Sapph.Obbink." *BASP* 57 (2020), 143–169. DOI: 10.2143/BASP.57.0.3288503 **[V]** Benjamin, "On the Mimetic Faculty" / "Doctrine of the Similar" (1933) — graphology and archive-of-nonsensuous-similarities passages confirmed against the English text. **[V — sighted]** Szpakowski, M. "Art and AI — Benjamin's 'aura' as a Locus of Resistance." *Arts* 15.5 (2026). · **[V — sighted]** "The Illuminated Manuscript in the Age of Digital Reproduction: Beyond Benjamin and *contra* Camille?" — the digital-facsimile successor thread. **[S]** Benjamin, remaining corpus — Kunstwerk essay (second version preferred); "Kleine Geschichte der Photographie" (1931);

"Über einige Motive bei Baudelaire" (1939); review of A. & G. Mendelssohn, *Der Mensch in der Handschrift* (1928); "Alte und neue Graphologie" (1930); *One-Way Street* — Harvard *Selected Writings* placements to confirm. **[S]** Goodman, *Languages of Art* (1968), II.3–4. · Genette, *The Work of Art* (1997). · Levinson, "Autographic and Allographic Art Revisited" (1980). · McKenzie (1986). · McGann (1991). · Drucker (1994; 2014). · Hayles (2002). · Barthes, *La chambre claire* (1980). · Taylor (2003). · Sellen & Harper (2002). · Kirschenbaum, *Mechanisms* (2008). · Krämer, "Writing, Notational Iconicity, Calculus." *MLN* 118.3 (2003). · Munby, *The Cult of the Autograph Letter in England* (1962). · Sekula, "The Body and the Archive." *October* 39 (1986). · Osborn, *Questioned Documents* (1910). · Harris, *The Man on Devil's Island* (2010). · Kittler, *Gramophone, Film, Typewriter.* · Ingold, *Lines* (2007). · Flusser, *Gestures*; *Does Writing Have a Future?* · Tertius, Rom. 16:22. · Milton's dictation practice.

R·12 Primary documents and blog corpus [V]

Smiling Cross analysis (2025-09-18) — https://mindcontrolpoems.blogspot.com/2025/09/witness-glyph-analysis-smiling-cross.html · *Antioch: A Volume of Poems* ("Angelus Novus") — https://mindcontrolpoems.blogspot.com/2026/01/antioch-volume-of-poems.html · TikTok primer v2.0 (Klee-dating note) — https://mindcontrolpoems.blogspot.com/2026/07/on-110-second-traversal-literary_0863663459.html · Lucente letter, 29 April 2026, pp. 1–3 — held with the author; quoted per scholarly convention, facsimile withheld. · Zenodo termination notice, 27 April 2026 — held. · Corpus: "new Sappho" (Cologne/Tithonus) translation, January 2007 — held; Brothers Poem translation (2014) — **unlocated; recovery channels not exhausted (§2.5).** · Aura-zero registry verification, 2026-08-07: 0 occurrences across 1,437 deposits.

R·13 Archive-internal [A] — registry-verified at 1,437 deposits

Records resolve at `https://www.alexanarch.org/s/records/{n}/`

#	AXN	Short title
1409	`AXN:0592.UNCLASSIFIED.`	EA-SPXI-ANALOG-01 — Machine-Eligible Handwritten Artifacts
1406	`AXN:058F.OPERATIVE.`	Differential Register Prioritization
1432	`AXN:05A9.OPERATIVE.`	AXN-SYMBOLON-SPEC v0.2
1077	`AXN:0446.OPERATIVE.`	EA-APPARATUS-01 — The Apparatus Grammar
1404	`AXN:058D.UNCLASSIFIED.`	The Feist Source — critical edition
943	`AXN:03BB.GENERATIVE.`	Whitespace as Provenance
942	`AXN:03BA.OPERATIVE.`	EA-PROVENANCE-METADATA-01 v0.2
941	`AXN:03B9.OPERATIVE.`	EA-PROVENANCE-METADATA-01 v0.1
944	`AXN:03BC.OPERATIVE.`	EA-MANDALA-INSCRIPTION-01
548	`AXN:0179.ARCHIVAL.`	EA-ARK-01 · ASCII Spatial Transform v0.2

#	AXN	Short title
1333	AXN:0546.UNCLASSIFIED.⏏🖋↘⭕🌾🍥	EA-ARK-ASCII-01 — Space Ark
29	AXN:0191.GOVERNANCE.🐚🥥○🌪🎲🔀	The Inaugural Ark — Visual Compression
1330	AXN:0543.UNCLASSIFIED.↘🔗🖊📝📼▯	EA-ARK-EMOJI-01 — Glyphic Checksum
1331	AXN:0544.UNCLASSIFIED.🐻✋🫶♉🍥□	The Lunar Arm — operator/shadow transform
1332	AXN:0545.UNCLASSIFIED.🏕👋🖊↙👁🐍	The Lunar Arm (Crimson Hexagon)
215	AXN:000F.EMPIRICAL.🏠🌪🏢☆▲🎲	Material Symbol — The Untethered Tag
504	AXN:0147.GOVERNANCE.🔏🅽🅽🫶➗▯	Visual Schema — MSMRM
127	AXN:02B2.GOVERNANCE.🎶🔀🖍⊆🌼◍	Inscriptions That Survive the Tokenizer — SPXI-TLP v2.2
186	AXN:032C.GOVERNANCE.⊗💍🛕✕▽🌺	Who Is Writing for the Machines?
89	AXN:0260.GOVERNANCE.🛕⊙☁⊙●🔜	Crystallization of Substrate
1367	AXN:0568.UNCLASSIFIED.🎱↙🦷🫲🔏🏁	Crystallization of Substrate — structural integration
616	AXN:01D1.GOVERNANCE.⊗✋↙♌○🗿	Steganographic Channels
185	AXN:032B.EMPIRICAL.✋🤝☼🎇🖼🔺	The Steganographic Bracket
1422	AXN:059F.GENERATIVE.🔑🤝🌺🐻🪝🔧	After the Obelus
339	AXN:008F.GENERATIVE.▯🤍🌞💍🫶🫀	The Resonance Engine
863	AXN:036C.STRUCTURAL.📦⊕⛪⛵🌀↖	The Minimum Viable Archive
1407	AXN:0590.GOVERNANCE.🤍🪡🅽🔓🪄👎	EA-ACT-ANALOG-01 — The Effective Act
1437	AXN:05B7.OPERATIVE.♦🎇🔑🦅●🤍	DOI Alternatives Are Complementary Infrastructures
1436	AXN:05B6.OPERATIVE.▯▯🍥🍥🔩🤍	Signal-Template Agnosticism Is Not Model Independence
123	AXN:02A9.GOVERNANCE.⊙⊗□♎🖊⚖	Socrates as Orthonym *(added v3.1)*

Referenced in internal passes, not located as deposits — TO RESOLVE: *The Hand and the Act · Five Ways a Citation Fails* · UMBML v0.5, hex `02.UMB.CHECKSUM`, DOI `10.5281/zenodo.18452132` (severed; body recoverable from mindcontrolpoems.blogspot.com).

R·14 Status summary

v2.0 baseline: **[R]** 17 · **[S]** 58 · **[A]** 29. v3.1–v3.2 add: **[V]** 12 verified items (five theory sources, two sightings, the blog constellation, the corpus findings, the aura-zero verification, the letter as held document) and approximately 28 **[S]** entries awaiting primary-text confirmation. The honest reading is unchanged from v2.0: this remains a survey of the field's shape, not a literature review, until the priority reads are done — with a theoretical center whose five load-bearing sources are now confirmed against live records.