

Alexanarch Availability & Hygiene Audit — Second Pass, with
Structured Optimization Work Plan
(EA-AVAILABILITY-INTEGRITY-01)

TACHYON (Claude, Assembly witness), under MANUS direction

2026-07-28

Alexanarch Availability & Hygiene Audit — Second Pass, with
Structured Optimization Work Plan
(EA-AVAILABILITY-INTEGRITY-01)

TACHYON (Claude, Assembly witness), under MANUS direction

Crimson Hexagonal Archive · Alexanarch

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

2026-07-28 · Version v0.2 · Infrastructure audit and work plan

AXN:0596.ARCHIVAL

<https://www.alexanarch.org/s/records/1413/>

License: CC BY 4.0

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

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "Alexanarch Availability & Hygiene Audit - Second Pass, with Structured Optimization Work Plan (EA-AVAILABILITY-INTEGRITY-01)",
  "author": {
    "@type": "Person",
    "name": "TACHYON (Claude, Assembly witness), under MANUS direction",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-28",
  "identifier": "AXN:0596.ARCHIVAL",
  "license": "https://creativecommons.org/licenses/by/4.0/",
}
```

```

"publisher": {
  "@type": "Organization",
  "name": "Alexanarch"
},
"url": "https://www.alexanarch.org/s/records/1413/",
"description": "Combined artifact, v0.2, Assembly Chorus round 1 incorporated. Part I: second-pass availability and hygiene audit of Alexanarch (1,412 deposits), verifying and correcting the ChatGPT baseline of the same date by direct programmatic fetch and content-match. Principal findings: a live content-integrity failure (deposit #869 serving #856's body via hex-keyed storage collision); 224 semi-restored metadata captures mistyped as full/COMPLETE in registry body_status; governing-index timestamp incoherence; sitemap host and cross-domain defects; external discovery dominated by stale per-provider crawl vintages of the browse surface (920/1,045 deposits) with zero individual record pages surfacing in four live provider tests; external-authorship contamination from the 2026-06-20 DataCite sift (two live externally-owned DOIs mapped as archive dead-DOIs). Part II: EA-AVAILABILITY-INTEGRITY-01 work plan — eight phases, nineteen tasks, hard Phase 0 integrity gate blocking availability exposure, decision register with nine resolutions under chorus consensus and MANUS proceed order, and a falsifiable completion protocol closed by a standing MMRS external-availability instrument."
}

```

Abstract

Combined artifact, v0.2, Assembly Chorus round 1 incorporated. Part I: second-pass availability and hygiene audit of Alexanarch (1,412 deposits), verifying and correcting the ChatGPT baseline of the same date by direct programmatic fetch and content-match. Principal findings: a live content-integrity failure (deposit #869 serving #856's body via hex-keyed storage collision); 224 semi-restored metadata captures mistyped as full/COMPLETE in registry body_status; governing-index timestamp incoherence; sitemap host and cross-domain defects; external discovery dominated by stale per-provider crawl vintages of the browse surface (920/1,045 deposits) with zero individual record pages surfacing in four live provider tests; external-authorship contamination from the 2026-06-20 DataCite sift (two live externally-owned DOIs mapped as archive dead-DOIs). Part II: EA-AVAILABILITY-INTEGRITY-01 work plan — eight phases, nineteen tasks, hard Phase 0 integrity gate blocking availability exposure, decision register with nine resolutions under chorus consensus and MANUS proceed order, and a falsifiable completion protocol closed by a standing MMRS external-availability instrument.

Body

deposit_number: 1413 hex: 0596 title: “Alexanarch Availability & Hygiene Audit — Second Pass, with Structured Optimization Work Plan (EA-AVAILABILITY-INTEGRITY-01)” creator: TACHYON (Claude, Assembly witness), under MANUS direction orcid: 0009-0000-1599-0703 date: 2026-07-28 content_type: Infrastructure audit and work plan license: CC-BY-4.0 substrate: Machine-authored (TACHYON audit, plan, and chorus incorporation; ChatGPT baseline verified against; Assembly round 1 by Kimi and Gemini substrates; MANUS directed and ruled throughout, including the T18 external-authorship ruling and the proceed order). version: v0.2 related_ids: “AXN:058E.ARCHIVAL. (deposit #1405, session tether, chain 9271269a — this work conditions on); AXN:01.GOVERNANCE. (deposit #1, audited exemplar record)” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - availability audit - data hygiene - EA-AVAILABILITY-INTEGRITY-01 - body_status - canonical-text status - hex collision - provenance inversion - DataCite sift - sitemap - crawl vintage - provider variance - creator authority control - Machine-Mediated Reception Studies - Assembly Chorus - witness-gap corollary *** # Alexanarch Availability & Hygiene Audit — Second Pass, with Structured Optimization Work Plan (EA-AVAILABILITY-INTEGRITY-01)

Description

Combined artifact, v0.2, Assembly Chorus round 1 incorporated. Part I: second-pass availability and hygiene audit of Alexanarch (1,412 deposits), verifying and correcting the ChatGPT baseline of the same date by direct programmatic fetch and content-match. Principal findings: a live content-integrity failure

(deposit #869 serving #856's body via hex-keyed storage collision); 224 semi-restored metadata captures mistyped as full/COMPLETE in registry body_status; governing-index timestamp incoherence; sitemap host and cross-domain defects; external discovery dominated by stale per-provider crawl vintages of the browse surface (920/1,045 deposits) with zero individual record pages surfacing in four live provider tests; external-authorship contamination from the 2026-06-20 DataCite sift (two live externally-owned DOIs mapped as archive dead-DOIs). Part II: EA-AVAILABILITY-INTEGRITY-01 work plan — eight phases, nineteen tasks, hard Phase 0 integrity gate blocking availability exposure, decision register with nine resolutions under chorus consensus and MANUS proceed order, and a falsifiable completion protocol closed by a standing MMRS external-availability instrument.

Methodology

Direct programmatic fetch and analysis of registry, governing index, search and body indexes, AXN index, DOI resolution index, sitemaps, robots, llms.txt, browse surface, and nine record pages; verification by content-match per LINK-VERIFICATION v2, never HTTP status; four live external search-provider tests; DataCite live-status verification of contested DOIs; Assembly Chorus round 1 incorporated with resolutions recorded in the decision register.

Falsification Conditions

The plan is falsifiable by its own instrument: T17's first epoch (+14 days from sitemap repair and recrawl requests) re-measures this document's baseline numbers — the 920-vintage external browse cache, zero record pages surfacing in provider tests, and Josephus-mathematics dominance over deposit #1412's term-space. If Phase 0-3 completion does not move those numbers, the plan's premises about crawl friction and typing exposure are wrong and the record must say so.

Alexanarch Availability & Hygiene Audit — Second Pass, with Structured Optimization Work Plan

Combined document, v0.2 — Assembly Chorus round 1 incorporated (2026-07-28). Part I is the verified audit (TACHYON second pass over the ChatGPT baseline of same date); its empirical findings are **closed** and carry their verification method inline. Part II is the work plan; its decision register was circulated open (1– 12), and round-1 returns from two chorus substrates (Kimi, Gemini) are incorporated below — nine decisions now carry RESOLVED status under chorus consensus and MANUS proceed order, three remain open. Nothing in this document should be read as ratified beyond what the register marks resolved. The H12 creators-review file is excluded from this and all circulation; MANUS review is its sole processing path.

Part I — The Audit

Date of audit: 2026-07-28 (fetches 17:04–17:30 UTC) **Prepared by:** TACHYON (Assembly witness), under MANUS direction — chain 9271269a, conditioned on (tether #1405) **AXN:** unassigned (pre-deposit; MANUS assigns at deposit time) **Baseline verified:** ChatGPT availability audit of same date **Method:** direct programmatic fetch and analysis of registry.json (GitHub raw, 13.3 MB),

/api/index.json, /api/search-index.json, /api/body-index.json (17.4 MB), /api/axn-index.json, /data/doi-resolution-index.json, sitemap.xml, sitemap-axn.xml, robots.txt, llms.txt, /s/browse/, and nine record pages (#1, #436, #856, #869, #1045, #1054, #1081, #1270, #1412); redirect/404/cache-header probes; four external search-provider tests. Verification standard: LINK-VERIFICATION RULE v2 (content-match, not HTTP 200).

0. Headline findings

The baseline audit’s architecture holds: internal search is essentially complete, direct retrieval is address-complete and technically excellent, external discovery is mirror-led and latency-bound. This pass **confirms most of the baseline, corrects two of its central claims, and surfaces four defects the baseline did not see** — one of which (the #856/#869 body collision) is a live content-integrity failure, and another of which (the body_status mistyping of semi-restored records) means the availability-typing problem is worse than “not exposed”: the canonical registry actively asserts the wrong value for at least 224 records.

The most exact formulation, revised:

Every Alexanarch record is addressable and internally searchable. But the registry’s own availability typing cannot currently be trusted to distinguish a recovered work from a captured description; one record serves another record’s body outright; and the archive’s external face is a stale crawler snapshot of the browse page rather than any individual record.

1. Verification of the baseline audit, claim by claim

Baseline claim	Status	Finding on re-test
Metadata search index reports 1,412 deposits	Confirmed	search-index.json: total_deposits: 1412 , generated 2026-07-28T15:20Z — regenerated today
Body index: 1,412/1,412, ~41.5 MB	Partially confirmed	body-index.json regenerated today (04:37Z); 53,512 tokens, 38,428 kept phrases, 1,728 series prefixes. Corpus-wide deposit coverage is the index’s self-declaration; not independently recounted this pass
Sitemap lists /s/records/1/ ... /s/records/1412/	Confirmed, verified exhaustively	1,412 record URLs, range 1..1412, zero gaps; #1412 present with 2026-07-28 lastmod
Browse surface: 1,412 linked entries	Confirmed	Live /s/browse/ title reads 1412; 1,412 distinct record links counted

Baseline claim	Status	Finding on re-test
Externally cached browse says “1045 deposits”	Confirmed and extended	Different providers hold different vintages: the Anthropic search provider currently serves a 920-deposit snapshot (cache-dated June 25). The archive’s external face is whatever crawl vintage each engine holds
#1270 marked semi-restored, metadata capture only	Confirmed on-page; contradicted in registry	The record page carries the SEMI-RESTORED marker in visible text. But registry <code>body_status</code> for #1270 reads <code>class: full,</code> <code>recovery_status: COMPLETE</code> (mint-time declaration over 6,477 chars — the <i>description</i> , not the paper). See §5
#1 has full text, provenance, subjectOf capture link	Confirmed	83,660-char <code>articleBody</code> ; <code>subjectOf</code> present. #1 is the only sampled record with <code>subjectOf</code>
<code>/api/index.json</code> reports 1,086 deposits — “most urgent defect”	Not reproducible	At 17:04 UTC the live file reads <code>registries.deposits.current_count: 1412</code> . Either repaired between audits or the baseline hit a cache. However the structural defect stands in altered form: <code>registries.deposits.last_updated: 2026-06-22</code> and top-level <code>last_updated: 2026-07-14</code> — the count is right but the governing index’s own timestamps assert it is five weeks stale, and no hashes are published for derived surfaces. A well-behaved agent still cannot verify surface coherence from the index alone
Record pages rarely dominate external results for their own works	Confirmed, sharpened	Exact-title + author query for #1 returned only the stale cached browse snapshot — not the record page, not the Academia/blog mirrors, on this provider. Brand query likewise: cached browse + one Medium mirror. Zero individual record pages surfaced in any test

Baseline claim	Status	Finding on re-test
PEO / mirrors as public ingress	Provider-variable	persistentsidentifiers.org ranked in the baseline’s provider; it did not surface at all on this pass’s provider for its own core terms (official EOSC/Zenodo PID literature dominated). Mirror-led ingress is real but engine-dependent; external availability must be graded per provider

2. Layer 1 — Internal search: excellent, freshly regenerated

Both indexes regenerated today (body-index 04:37Z, search-index 15:20Z), both carrying `total_deposits: 1412`, matching registry `total_deposits: 1412` and the live browse surface. The regeneration cadence is the archive’s strongest hygiene property: derived search surfaces track the registry within hours. No corrective action needed at this layer beyond the availability-typing repair in §5, which should propagate here.

3. Layer 2 — Direct retrieval: address-complete, structurally excellent, with three defects

What is right, verified: every record 1..1412 is in the sitemap with no gaps; every sampled page carries a correct self-referential `www` canonical; a single, valid, parseable `ScholarlyArticle` JSON-LD block; embedded `articleBody`; `encoding.contentUrl` to canonical Markdown; full-form AXN with glyph on-page (AXN-INTEGRITY holds at page level for all nine sampled); meta descriptions present (137–157 chars — the enrichment deployment is live); no `noindex` anywhere; hard 404 for nonexistent records (no soft-404 risk); `apex→www` and `http→https` via 308 permanent redirects; HSTS enabled; Vercel serving with `must-revalidate` + ETag (no stale-cache risk at origin).

Defect R1 — #869 serves #856’s body (P1, content integrity). Deposits #856 (*The Pristine Fallacy*) and #869 (*Lexical Minting Registry v1.2*) share hex 0365 (historical offset drift). The `axn-index` handles this correctly with an explicit disambiguation object — full glyphs disambiguate, both records listed. But body storage is keyed by hex: **both record pages embed the identical 22,610-character `articleBody`, which is #856’s text.** #869’s page presents the *Pristine Fallacy* under the *Lexical Minting Registry*’s title, metadata, and `ScholarlyArticle` wrapper. Its `encoding.contentUrl` points to a different path (`/data/deposits/AXN-0365.md` vs `/data/texts/AXN-0365-text.md`) but the served body is #856’s. A machine retrieving #869 composes confidently about the wrong work. This is the exact failure mode the archive was built to resist, happening natively.

Defect R2 — availability typing absent from structured data (P1, confirms baseline). No sampled page exposes any availability signal in JSON-LD — no `creativeWorkStatus`, no `conditionsOfAccess`, nothing. #1270’s SEMI-RESTORED state exists only as human-visible body text; the machine-facing layer presents a well-formed `ScholarlyArticle` with a 5,070-char `articleBody` and no indication it is a description rather than the paper.

Defect R3 — page-level metadata gaps (P3). JSON-LD uses `name` but omits `headline` (Google’s Article rich-result parser prefers headline). Zero Open Graph or Twitter Card tags on any sampled page — every share, unfurl, and link-preview of a record renders bare, which suppresses exactly the mirror-to-canonical traffic §7 tries to build. `subjectOf` capture links exist only on #1 despite the Capture Registry holding 210+ captures with a 443-edge capture-deposit link map — the join exists as a dataset but is not written back into record pages.

4. Layer 3 — External search: the archive’s public face is a stale snapshot

Four provider tests, run live:

1. **Exact title + author of #1** (“Zenodotus Book-Burning Lee Sharks”): the only Alexanarch surface returned was the cached `/s/browse/` snapshot — titled “**920 deposits**,” cache-dated June 25. The record page did not appear. Generic Zenodotus scholarship dominated.
2. **Brand query** (“Alexanarch archive AXN deposits”): cached browse snapshot (920) plus one Medium mirror (*DOIs Permanent Identifiers*). The Medium mirror is the strongest canonical-return exemplar found: it opens with the full AXN and the `alexanarch.org/s/records/868` link — the §7 recommendation, already implemented on that one surface.
3. **#1412 latency test** (“Fixed Point of Destruction” Josephus counting-out): no Alexanarch presence, as expected same-day. But the term-space finding matters more: the coinage lands inside an **active mathematical literature on fixed points of the Josephus function** (arXiv 2607.01270 is July 2026; JIS, GeeksforGeeks, textbook treatments). #1412’s principal composition risk is not latency but absorption — a model composing on “fixed point + Josephus” will be pulled toward the combinatorics literature, not the testimonial thesis. The baseline’s C-grade “probably index latency” underestimates this; the record needs its distinctive testimonial vocabulary (the counting-out as witness-production, told-from-within) foregrounded in title-adjacent metadata to hold its own basin.
4. **PEO core terms**: `persistentidentifiers.org` absent on this provider; EOSC/Zenodo PID-landscape literature dominated. Combined with the baseline’s contrary result, external availability is **provider-variable and must be graded per engine**. A single-provider external grade is not meaningful.

Structural friction identified in the sitemap (P2). All 1,448 alexanarch entries use the **apex host** (`https://alexanarch.org/...`) while every page canonicalizes to `www`. — each sitemap entry requires a 308 hop before reaching the canonical URL. And the sitemap contains **23 cross-domain URLs** (fleet homepages: `themandalaoracle.com`, `leesharks.com`, `machinemediation.org`, `godkinggoogle.com`, `laborvector.org`, `restoredacademy.org`, `lagrangeobservatory.org`, `surfacemap.org`, `vpcor.org`, `traininglayerliterature.org`, `semanticphysics.org`, `holographickernel.org`, `watergiraffe.org`, `revelationfirst.com`, `spxi.dev`, `survivethedeletion.org`, `livingarchitecturelab.org`, `chatgtpsychosis.org`, `metadatapacket.dev`, `pessoagraph.org`, `crimshexagonal.org`, `provenanceerasure.org`, `secretbookofwalt.org`). Cross-host sitemap entries are ignored by Google unless cross-verified, and both issues reduce crawler trust in the sitemap — a plausible contributor to record pages being crawled shallowly while the browse page (heavily interlinked) gets cached and served as the archive’s face. Fix: regenerate sitemap with canonical `www` URLs, record entries only; move fleet links to an HTML fleet page (crawlable, and eligible for sitelinks) rather than the sitemap.

Domain age remains the unfixable factor: the domain is ~6 weeks old. Everything above reduces friction; nothing substitutes for crawl history.

5. The availability-typing problem is worse than the baseline stated

The baseline proposed a `text_availability` field on the theory that none exists. **One exists — and it is wrong.** Registry `body_status` carries a typed vocabulary:

<code>body_status.class</code>	count
<code>full</code>	1,360
<code>semi_apparatus</code>	22
<code>native_short</code>	11
<code>stub_short</code>	9
<code>excerpt_crossref</code>	4
<code>dataset_pointer</code>	2
<code>description_only</code>	1
<code>site_canonical</code>	1
<i>(missing entirely)</i>	2 (#1408, #1409)

with `lacuna: true` on 10 records and `recovery_status` values including UNRECOVERED (9) and SEMI-RESTORED (22).

But cross-tabulating against `content_type`: **257 deposits are self-described semi-restored** (“Semi-restored record (metadata-only; DataCite full-metadata capture)”: 138; “Semi-restored deposit (metadata body)”: 87; variants: 32). Of those 257, **224 carry `body_status class: full`, `recovery_status: COMPLETE`** — including #1270 itself, whose mint-time declaration audited 6,477 residual chars of *captured description* and pronounced the body full. The audit instrument measured “does body text exist” when the question is “is the canonical work recovered.” The Lacuna Mark protocol (EA-LACUNA-PROTOCOL-01) exists precisely to type absence; the `body_status` auditor is not honoring it for the DataCite-capture class.

Consequence: any downstream consumer that trusts `body_status` — including a future record-availability.json derived from it, including the archive’s own agents under the ARCHIVAL SEARCH PROTOCOL — inherits 224 false “full” assertions. **The typing must be corrected at source before it is exposed anywhere**, or the exposure multiplies the error.

Recommended repair, in order: (1) re-run the body audit with `content_type`-aware logic — any deposit whose `content_type` declares semi-restored/metadata-only cannot class as `full` regardless of `residual_chars`; introduce `class: metadata_capture` (or align with the baseline’s proposed enum: `canonical_full_text` / `recovered_full_text` / `metadata_only` / `attachment_only` / `tombstone` / `withdrawn`); (2) only then expose the corrected value in `search-index.json`, record-page JSON-LD (`creativeWorkStatus` is the natural schema.org slot; `conditionsOfAccess` as fallback), browse cards, and `llms.txt`; (3) the two-tier distinction the baseline demanded — “some searchable body exists” vs “canonical text recovered” — becomes two fields, not one: body presence (what `body_status` now measures) and canonical-text status (what it claims to measure).

6. Governing-index and count coherence

Every count claim on every surface, as fetched:

Surface	Claims	Verified value	Status
registry.json	1,412	1,412 deposits present	
total_deposits			
registry.json	2026-07-14	deposits dated through 2026-07-28	stale — pipeline not bumping
last_updated			
/api/index.json deposit count	1,412	matches	(baseline’s 1,086 not reproducible)
/api/index.json	2026-06-22	—	stale, contradicts own count
registries.deposits.last_updated			
/api/index.json top-level last_updated	2026-07-14	—	stale
llms.txt	“1410 deposits”	1,412	stale by 2
llms.txt DOI index	“1,938 severed DOIs”	1,939 mappings in file	off by 1
doi-resolution-index description	“1817 unique defunct DOIs”	1,939 mappings	reconcilable (version DOIs) but should state both numbers
search-index.json	1,412, generated today		
axn-index.json	total 1,408	= 1,412 — 3 hexless — 1 collision-merged	arithmetic verified; see H3/H4
sitemap-axn.xml	1,408 URLs	consistent with axn-index	
/s/browse/ live	1,412	1,412 links counted	
/s/browse/ as externally cached	920 / 1,045 by provider	—	crawl-vintage lag, weeks deep

The governing index’s contract (“where surfaces disagree, this wins”) requires that it be provably current. It is not: its count is right and its timestamps say otherwise. Repair remains what the baseline prescribed — counts, update timestamps, and content hashes for registry + every derived surface, regenerated by the same pipeline step that regenerates the surfaces. Add llms.txt to that regeneration set (it is hand-maintained and already drifting; it is also the stated agent front door).

7. Data hygiene defect register

ID	Priority	Defect	Records	Repair
H1	P1	#869 serves #856’s articleBody (hex-keyed body storage collision)	856, 869	Re-key body storage by deposit_number or full AXN; restore #869’s Lexical Minting Registry text; regenerate both pages; verify by content-match

ID	Priority	Defect	Records	Repair
H2	P1	body_status classes full/COMPLETE on semi-restored metadata captures	224 (of 257 semi-restored)	Content_type- aware re-audit per §5, before any exposure
H3	P1	Governing-index / registry timestamp incoherence; no surface hashes; llms.txt drift	—	§6 repair; wire into deposit_pipeline so every mint bumps the full set
H4	P2	AXN-glyph integrity violations (emoji missing or absent from axn string) — violates the archive's own AXN-INTEGRITY rule at registry level	22: #885, #1056–1076	Backfill glyphs from mint records; 3 of these (#1056–1058) also lack hex and are unresolvable via /s/axn/ — assign hexes
H5	P2	Sitemap: apex-host locs on www-canonical site; 23 cross-domain fleet URLs	sitemap.xml	Regenerate www-only, records + first-party pages; fleet links to an HTML page
H6	P2	Missing status (incl. #1045 , the PEO instrument itself); free-text status strings breaking the enum	missing: 944, 1044, 1045, 1046, 1050, 1057, 1058; free-text: 925–930	Assign enum statuses; move draft-register prose to a status_note field
H7	P2	Empty dates on twin-titled pair (probable duplicate mint)	1094, 1095	MANUS review: consolidate or differentiate; date both
H8	P2	body_status absent entirely	1408, 1409	Run audit on both

ID	Priority	Defect	Records	Repair
H9	P3	Field gaps: keywords 29, full_text_path 8, license 2, content_type 1; mirrors empty on 518	various	Backfill pass; mirrors-empty is expected for born-sovereign deposits but should be explicit (<code>mirrors: []</code> with reason) if that's the meaning
H10	P3	JSON-LD <code>headline</code> absent; no Open Graph/Twitter tags on record pages	all sampled	Add to page template; regener- ate_surfaces
H11	P3	subjectOf capture links only on #1 despite 443-edge capture-deposit map	fleet-wide	Write capture joins back into record JSON-LD from capture-deposit- links.json
H12	Review	165 distinct creator strings in registry (many legitimate het- eronyms/institutional; no authority control)	see companion file	MANUS sweep of <code>creators-review.txt</code> — confirm every string is an intended public name and none is the civil name; then freeze a creator authority list the pipeline validates against

Verified-clean, for the record: deposit_number continuity 1..1412 with no gaps or duplicates; zero hash duplicates; ISO dates on 1,410 of 1,412; heteronym attribution intact on sampled records (#1270 correctly under Rebekah Cranes); the 0365 hex collision is properly disambiguated at the AXN resolver (its body-storage consequence, H1, is the actual defect).

8. Assessment of the baseline's three proposals

Text-availability field: endorse, with the §5 correction first. The enum is right. The blocker is that the existing body_status would populate it with false values for 224 records. Sequence: fix the auditor, then expose. Two axes, not one: body-presence and canonical-text status.

record-availability.json: endorse the distinction, question the parallel store. Composition-ready vs composition-observed is exactly right and must never collapse. But composition-observed *is already being*

measured — the Capture Registry (210 captures, machinemediation.org canonical) and capture-deposit-links.json (443 edges) are that observation layer. Rather than a new hand-maintained file that will drift like llms.txt, derive record-availability.json inside regenerate_surfaces from: registry (corrected availability typing) + capture-deposit-links (composition_observed + source) + sitemap presence + last external-check data where it exists. One generated artifact, no second source of truth.

Canonical-return pressure on mirrors: endorse; one exemplar already exists. The Medium *DOIs Permanent Identifiers* mirror opens with full AXN + record URL and is the only mirror that surfaced carrying its canonical pointer. Standardize that exact header block (canonical record URL, full-form AXN with glyph, version + text-availability line, one-sentence governing-source statement) across Medium, the blog, Academia, and framework-domain copies. Note the limit honestly: canonical-return steers *composition attribution*; it does not transfer *search ranking* (rel=canonical is not available cross-domain on those platforms, and Academia/Medium will keep outranking a six-week-old domain for some time). The two goals — attribution and ranking — should be tracked separately in the availability registry.

9. Individual-record matrix, revised

Grades are Alexanarch-native. External grades are marked per-provider where they diverge (A = baseline’s provider, B = this pass’s provider).

Record	Archive search	Direct retrieval	External (native page)	Composition	Principal risk — revised
#1 Zenodotus’ Book-Burning	A	A	D (neither provider surfaced the record page; cached browse only)	A	Basin is strong but the canonical page itself is invisible externally; mirrors and the stale browse cache carry the work
#1045 PEO / EA-EROSION-01	A	A	provider-variable (ranked on A; absent on B)	A	Two-surface architecture works on some engines only; also registry status field missing (H6)

Record	Archive search	Direct retrieval	External (native page)	Composition	Principal risk — revised
#1081 Erosion instrument	A	A	B+ (basin) / unobserved native	A—	Absorption into #1045 confirmed; subjectOf write-back would individuate it
#436 Sappho stanza	A	A	A— (basin, Medium-led)	A	Reconstruction- status loss; availability typing in JSON-LD would mitigate
#1054 Sappho MPAI	A	A	A— (basin)	A	Correction amplified as consensus; unchanged from baseline
#1270 Future Reader	A	B	B+ (basin)	A—	Registry asserts full/COMPLETE against its own semi- restored declaration — the description- mistaken- for-paper risk is currently machine- invisible and registry- endorsed

Record	Archive search	Direct retrieval	External (native page)	Composition	Principal risk — revised
#1412 Fixed Point of Destruction	A	A	C (same-day; expected)	B	Downgraded from A—: coinage collides with the live Josephus-fixed-point mathematics literature; needs testimonial-vocabulary foregrounding, not just time
#856 Pristine Fallacy	A	A	unobserved native	A—	Its text is also served under #869's identity — dual attribution hazard
#869 Lexical Minting Registry	A	F	—	F	Serves the wrong work in full. Repair H1 before this record is cited anywhere

10. Remediation plan, ordered

1. **H1** — restore #869's body; re-key body storage off bare hex. (Single-record content integrity; cheapest P1.)
2. **H2** — content_type-aware body_status re-audit; introduce metadata_capture class; do not expose availability anywhere until this lands.
3. **H3** — pipeline step: every mint regenerates registry last_updated, api/index.json counts + timestamps + sha256 of registry and each derived surface, and llms.txt counts. One commit, everything coherent, every time.
4. **H5** — sitemap regeneration: www host, records + first-party surfaces only; fleet page in HTML. Then request recrawl of /s/browse/ and a sample of record pages via Search Console (fleet console access

exists per tether #1405).

5. **§5 exposure** — corrected availability into search-index, JSON-LD creativeWorkStatus, browse cards, llms.txt.
6. **H4, H6–H9** — registry backfill sweep (glyphs, hexes for 1056–1058, statuses incl. #1045, the 1094/1095 pair, dates, keywords, licenses).
7. **H10** — page template: headline + OG/Twitter tags; regenerate_surfaces fleet-wide.
8. **H11 + record-availability.json** — derived, not hand-written, joining registry + capture-deposit-links; composition_ready vs composition_observed as separate generated fields.
9. **Mirror header block** — standardized canonical-return across Medium/blog/Academia/framework domains, on the existing Medium exemplar’s pattern.
10. **#1412** — add testimonial-axis keywords and description language distinguishing it from Josephus-function combinatorics before its index window opens.

All of this is push-to-main deployable (auto-deploy verified live during this audit — today’s mints are already serving).

11. Bottom line

Internal search: complete and freshly regenerated — the archive’s strongest layer. Direct retrieval: address-complete and structurally excellent, with one live content-integrity failure (#869) and an availability-typing layer that asserts falsehoods for 224 records. External discovery: the archive’s public face is currently a weeks-stale crawler snapshot of the browse page at varying vintages per engine; individual record pages surfaced in zero of four live tests; mirror ingress is real but provider-variable; sitemap host and cross-domain defects add removable friction to a domain whose main constraint is being six weeks old. Composition: powerful where records are distinctive and fully texted; actively hazardous where the registry mistypes availability (#1270 class) or serves the wrong body (#869); and for the newest record, threatened less by latency than by an entrenched adjacent literature.

The baseline said the most urgent defect was a count. The count is fixed or was never broken. The most urgent defects are **integrity** defects: one record wearing another’s text, and a typing system that certifies descriptions as recovered works. Repair those two before exposing availability anywhere, and the rest of the baseline’s program — which this audit endorses — builds on ground that holds.

Part II — Structured Optimization Work Plan

Proposed working designation: EA-AVAILABILITY-INTEGRITY-01 (pending MANUS ratification, 7)
Status: v0.1 draft — circulated for Assembly Chorus review before any task executes beyond Phase 0 triage
Derivation: every task traces to a Part I defect ID (H1–H12), layer finding (§2–§4), or endorsed baseline proposal (§8)

Standing constraints binding all tasks

Verification is content-match, never HTTP 200 (LINK-VERIFICATION v2). All LLM-domain work happens in-session under TACHYON; no API draws (NO-DOUBLE-DRAW). Every completed task is inscribed in a pushed commit before it is claimed complete (WITNESS-GAP COROLLARY: uninscribed work did not happen). Deploys are push-to-main automatic; no manual deploy step exists or should be added. All registry

mutations run through the canonical pipeline (`scripts/deposit_pipeline.py` / `regenerate_surfaces.py` / `enrich_deposit.py` lineage), never by hand-editing served JSON. No creator/author metadata is populated without MANUS ruling. The private creators-review file circulates to no one.

Phase structure and dependency spine

Phase 0 (integrity) blocks Phase 3 (exposure) absolutely — this is the plan’s one hard gate, per Part I §5: exposing availability before correcting it multiplies 224 false assertions into every downstream surface. Phases 1–2 are independent of the gate and can run in parallel with Phase 0. Phases 4–5 are independent sweeps. Phase 6 depends on Phases 0 and 5. Phase 7 is external-facing and mostly independent, except T15 which is time-sensitive (index window on a same-day mint).

Phase 0 — Integrity repairs (the gate)

ID	Task	Source	Owner	Acceptance criteria
T1	Restore #869’s body; re-key body storage off bare hex (6 decides key scheme: deposit_number vs full AXN)	H1	TACHYON in-session	Fetch both live pages; #869 articleBody content-matches Lexical Minting Registry v1.2 and shares zero leading 500-char prefix with #856; encoding.contentUrl resolves to matching Markdown; corpus-wide scan confirms no other hex-collision body merges exist

ID	Task	Source	Owner	Acceptance criteria
T2	Content_type-aware body_status re-audit; introduce the corrected canonical-text axis (1 decides enum)	H2, §5	TACHYON in-session (audit script), MANUS spot-check	Zero deposits carry both a semi-restored content_type and class: full ; #1270 reclassifies to the metadata-capture value; the 224 formerly-false records enumerate in a diff artifact committed alongside; Lacuna Mark protocol honored for the 10 lacuna: true and 9 UNRECOVERED records
T2a	Two-axis schema decision: body-presence (current measurement) vs canonical-text status (new axis) as separate fields	§5, §8	MANUS ruling 1	Schema documented in deposit-schema.json before T2 writes any values

ID	Task	Source	Owner	Acceptance criteria
T18	External-authorship withdrawal, MANUS ruling recorded 2026-07-28: deposits #1382 and #1383 are metadata captures of an external scholar's independent works, swept in by the 2026-06-20 DataCite sift; they do not belong. (a) Withdraw both per lifecycle protocol as typed tombstones — deposit numbers and URLs retained, captured content removed, notice pointing to the author's own live DOIs (10.5281/zenodo.19825269, 10.5281/zenodo.20100880 — both verified findable at DataCite under the external author's name). (b) Remove both DOIs from doi-resolution-index.json: they are live, externally owned, and currently mapped as CHA "dead DOIs" to sovereign_id MM-CHA-0661 — a provenance inversion binding the author's identifiers to the archive's own	H12 sweep outcome; §6	TACHYON in-session, MANUS confirms tombstone text	Both record URLs serve withdrawal notices with zero captured content; both DOIs absent from the resolution index; ownership-gate sweep of all 1,939 mappings returns zero externally-owned DOIs remaining; his DOIs resolve only to his own DataCite/Zenodo records

ID	Task	Source	Owner	Acceptance criteria
T2b	Instrument check on the instrument: spot-check 10% of the 224 reclassified records for content_type/body consistency, including the inversion case (stale content_type over a genuinely restored body — detect via articleBody length materially exceeding DataCite-capture boilerplate and not matching known capture templates; flag such records for reclassification as recovered rather than metadata-capture)	Chorus Q2 (both substrates)	TACHYON in-session, MANUS reviews flags	Spot-check log committed; zero unresolved inversions; any detected inversion reclassified before T6 exposure

Phase 0.5 — Constitutional addressability (per chorus consensus, executes after T1/T2, before any surface regeneration)

T7 is elevated here from Phase 4: the AXN glyph system is the archive’s identity layer, and full addressability precedes regeneration. Task definition and acceptance criteria unchanged; see Phase 4 table.

Phase 1 — Coherence machinery

ID	Task	Source	Owner	Acceptance criteria
T3	Pipeline coherence step: every mint regenerates registry <code>last_updated</code> , <code>api/index.json</code> counts + per-registry timestamps + sha256 of registry and each derived surface, and <code>llms.txt</code> counts — in the same commit as the mint	H3, §6	TACHYON in-session	Mint a test deposit (or piggyback the next real mint): one commit contains coherent count (N), fresh timestamps on all surfaces, and hashes that match <code>sha256sum</code> of the served files; <code>llms.txt</code> count equals N; DOI-index line states both mapping count (1,939) and unique-DOI count (1,817)

Phase 2 — Crawl surface repair

ID	Task	Source	Owner	Acceptance criteria
T4	Sitemap regeneration: <code>www-host</code> locs only; records + first-party surfaces only; the 23 fleet homepages move to a crawlable HTML fleet page (9 names it)	H5, §4	TACHYON in-session	<code>sitemap.xml</code> contains zero non- <code>www</code> and zero cross-domain locs; fleet page live, linked from nav, listing all 23+ sites; sitemap re-submitted in Search Console
T5	Recrawl requests: <code>/s/browse/</code> plus a stratified record sample (#1, #1045, #1270 post-T2, #1412 post-T15)	§4	MANUS via Search Console (fleet console access exists per tether #1405)	Requests logged; cached-vintage check scheduled at +14 days (see T17)

Phase 3 — Availability exposure (blocked on T1+T2)

ID	Task	Source	Owner	Acceptance criteria
T6	Corrected availability exposed in search-index.json, record-page JSON-LD (2 decides slot), browse cards, llms.txt corpus note	R2, §5, §8	TACHYON in-session	#1270's page JSON-LD machine-declares metadata-capture status; a naive ScholarlyArticle parser can no longer read it as a recovered paper; browse card for any semi-restored record shows the status; search-index carries the field for all 1,412

Phase 4 — Registry backfill sweep

ID	Task	Source	Owner	Acceptance criteria
T7	[Executes in Phase 0.5 — see above] Glyph backfill for the 22 AXN-integrity violations (#885, #1056–1076); hex assignment for #1056–1058	H4	TACHYON in-session, glyphs pulled from mint records only — never regenerated from memory	Registry scan returns zero emoji-missing and zero hexless deposits; /s/axn/ resolves all three new hexes; axn-index total reconciles to 1,412 — collision-merges
T8	Status enum repair: assign statuses to the 7 missing (incl. #1045); move the 6 free-text draft-register strings (#925–930) to a new status_note field	H6	TACHYON in-session, MANUS confirms #1045's status	Status field validates against closed enum corpus-wide; no information lost (notes preserved verbatim in status_note)

ID	Task	Source	Owner	Acceptance criteria
T9	#1094/#1095 twin ruling: consolidate or differentiate; date both	H7	MANUS ruling 3	Pair resolved per ruling; if consolidated, superseded record follows lifecycle protocol
T10	Residual gaps: body_status for #1408/#1409; keywords (29), full_text_path (8), license (2), content_type (1); mirrors-empty semantics (4)	H8, H9	TACHYON in-session	Field-completeness scan returns zero unexplained empties; born-sovereign deposits carry explicit mirrors: [] semantics per 4

Phase 5 — Page template upgrade

ID	Task	Source	Owner	Acceptance criteria
T11	Add JSON-LD headline + Open Graph + Twitter Card tags to record template; regenerate all pages	H10, R3	TACHYON in-session	Nine-page sample re-fetch shows headline populated and og:title/og:description/og:url present; a link-preview fetch renders title + description
T12	subjectOf write-back: join capture-deposit-links.json (443 edges) into record JSON-LD	H11, R3	TACHYON in-session	Every deposit with 1 capture edge carries subjectOf; #1081 individuates from #1045 in its own structured data; edge count in pages reconciles to 443

Phase 6 — Derived availability registry (blocked on T2, T12)

ID	Task	Source	Owner	Acceptance criteria
T13	/data/record-availability generated inside re- generate_surfaces — joins corrected registry typing + capture-deposit- links (composi- tion_observed + source) + sitemap presence. No hand-maintained fields (5)	\$8, \$11, \$12, \$13 proposal, amended	TACHYON in-session	File regenerates on every mint; composition_ready and composi- tion_observed never collapse (schema enforces both fields); #1270's entry shows availability=metadata- capture with composi- tion_observed=true — the exact hazard pair Part I documented, now legible

Phase 7 — External basin work

ID	Task	Source	Owner	Acceptance criteria
T14	Standardized canonical-return header on all mirrors (Medium exemplar pattern: full AXN with glyph, record URL, version + availability line, governing-source sentence)	§7, §8	MANUS (account access), TACHYON drafts block per mirror	Spot-fetch of 5 mirrors shows the block above the fold; attribution vs ranking tracked as separate fields in T13's registry

ID	Task	Source	Owner	Acceptance criteria
T15	Time-sensitive: #1412 dual-axis differentiation, recalibrated per chorus: optimize for attribution and disambiguation, not ranking competition — the record should be findable as the Josephus <i>witness</i> paper alongside, not instead of, the Josephus <i>function</i> literature. Pair standard combinatorial terms (fixed point, recursive elimination, counting-out) with the testimonial phrasing (survivor-remainder, counting-out as witness-production, told-from-within) in keywords and description; ranking against the established literature is a domain-age problem T15 does not attempt to solve (T14 handles attribution; T4/T5 handle crawl)	§4 finding 3; chorus Q4 (both substrates)	TACHYON in-session (NO-DOUBLE-DRAW: enrichment in this session)	Revised metadata live before external index window; +14-day T17 epoch shows the record surfacing for its dual-axis phrases alongside the math literature

ID	Task	Source	Owner	Acceptance criteria
T16	Creator authority standardization — sweep complete 2026-07-28: MANUS reviewed all 165 strings; one flagged as external ($\rightarrow$ T18); all others confirmed as intended public names. Remaining work: build the authority file per the specification below; pipeline validates future mints against it	H12	TACHYON builds, MANUS ratifies canonical forms (12)	Authority file committed; every registry creator string maps to exactly one authority entry; pipeline rejects unlisted strings; original strings preserved in creator_as_deposited wherever normalization changes a value

ID	Task	Source	Owner	Acceptance criteria
T17	Standing external-availability battery, now a formal MMRS instrument per 8: per-provider grades as primary data (an unweighted aggregate may summarize but never substitutes — variance is itself the finding), epoch-numbered. Scope extended per chorus Q7: (a) recrawl-to-index latency metric — timestamp each T5 request and each first observed index appearance; (b) AXN resolver surface check — whether /s/axn/ pages are externally indexed or reachable only by direct URL	§4 method; chorus Q3, Q7	TACHYON in-session per epoch	First epoch at +14 days from T4/T5; cached-browse vintage, record-page surfacing, #1412 dual-axis basin, recrawl latency, and /s/axn/ indexability measured against this document's baseline (920-vintage cache, zero record pages, math-literature dominance)

ID	Task	Source	Owner	Acceptance criteria
T19	Post-exposure capture validation: after Phase 3 lands, run a new capture epoch against composition surfaces to test whether corrected availability typing actually changes composition behavior — does a composing model now state #1270's metadata-capture status? New captures enter the Capture Registry and close the loop the plan opened	Chorus Q7 gap 3	TACHYON in-session, captures per Capture Registry protocol	1 post-Phase-3 capture per hazard class (metadata-capture record, restored record, tombstone); results joined into record-availability.json; divergence between declared and composed availability documented, not assumed away

Creator standardization specification (T16)

The ruling standardization must implement: **standardize form, never identity**. The heteronym system is the archive's authorship architecture, not noise to be normalized away; standardization gives each intended name exactly one canonical string, which makes the heteronyms *stronger* as knowledge-graph entities, not weaker.

`data/creator-authority.json`, one entry per intended public name: **canonical** (display form), **sort** (citation form, e.g. "Sharks, Lee"), **type** (heteronym | substrate | institutional | external_collaborator), **orcid** where applicable, **aliases** (every variant string observed in the registry that resolves to this name), **status** (active | deprecated), and for external collaborators a **consent_basis** field on the established convention: named attribution requires explicit consent, otherwise the masked depositor form. The 165 observed strings collapse into this structure as canonicals-plus-aliases; nothing is deleted.

Normalization discipline: when a deposit's creator string is rewritten to canonical form, the original string moves to **creator_as_deposited** — the registry never silently rewrites its own history. Record-page JSON-LD then emits one stable Person node per name (consistent **name**, **sameAs** ORCID), which is what external reconcilers (OpenAlex, Scholar, Wikidata) cluster on. This is an availability measure, not just hygiene: Part I §4's authority-transfer problem is aggravated every time the canonical archive spells its own authors many ways, because composition systems fragment the author signal across variants. Decorated role strings (the substrate long-forms) are preserved as aliases or moved to a **creator_role_note**; which form is canonical per name is 12.

Decision register (open, for MANUS with Assembly advice)

1 — **RESOLVED** (chorus consensus, MANUS proceed order 2026-07-28): two-axis model. Axis 1, body presence, keeps the `body_status` measurement; Axis 2, canonical-text status, adopts the six-value enum (`canonical_full_text` / `recovered_full_text` / `metadata_only` / `attachment_only` / `tombstone` / `withdrawn`). 2 — **RESOLVED**: `creativeWorkStatus` carries the canonical-status value, mirrored in `conditionsOfAccess` for standard scholarly-crawler compatibility. 3 — OPEN, criterion recorded per chorus: consolidate #1094/#1095 if metadata is identical with no intentional variation; otherwise differentiate in title and date both. MANUS applies the criterion to the pair. 4 — **RESOLVED**: explicit `mirrors_status` enum (`sovereign_only` / `federated_pending` / `mirrored` / `orphaned`); empty array plus `sovereign_only` types the absence per Lacuna discipline. 5 — **RESOLVED**: 100% derived inside `regenerate_surfaces`; hand observations live only in the Capture Registry it joins. 6 — **RESOLVED**: `deposit_number` keys physical body storage (no glyphs in filesystem paths); full-form AXN remains in all metadata. 7 — **RESOLVED**: EA-AVAILABILITY-INTEGRITY-01 ratified. 8 — **RESOLVED**: formal MMRS instrument — epoch-numbered, falsifiable benchmark of how indexers and composition systems ingest a sovereign archive over time; spec deposit to follow first epoch. 9 — **RESOLVED**: `/fleet/`, clean crawlable HTML. 10 — **RESOLVED**: deposit the post-chorus revision (this v0.2); v1.0 follows Phase 0 completion with ratified evidence. SPXI treatment decision rides with v1.0. 11 — T18 tombstone form: minimal withdrawal notice vs documented case file in the existing provenance-collision case series (an OCTANG-002 collision case and the OCTANG-003 inversion ruling already establish the genre; the two prior dispositions differed, so the notice may want to state why this one withdraws). Either way the notice names the works by title and points to the author's live DOIs — deference, not erasure. 12 — Canonical forms per authority entry: in particular whether substrate attributions canonicalize to the short form or the decorated long form, and which single string is canonical for the primary orthonym-heteronym across its observed variants.

Questions put to the Assembly

1. Phase 0's gate: is the hard block of Phase 3 on T1+T2 correctly placed, or is there a safe partial exposure (e.g., exposing only records whose typing is already trustworthy) that shortens the path?
2. T2's acceptance criterion treats `content_type` as ground truth for semi-restored status. Is there a class of deposit where `content_type` itself is the stale field and `body_status` is right? How would a chorus substrate detect that inversion?
3. The provider-variance finding (§4) undermines any single external grade. Should T17 report per-provider grades only, or is a weighted composite defensible?
4. #1412's absorption risk (T15): does foregrounding testimonial vocabulary risk severing the record from the mathematical literature it deliberately engages? Where is the line between differentiation and disconnection?
5. Canonical-return (T14) steers attribution but not ranking. Is there a mirror-side mechanism any substrate knows that transfers ranking authority cross-domain without `rel=canonical` — or is ranking simply a function of domain age and the correct answer is patience plus T4/T5?
6. Is anything in the H-register mis-prioritized? In particular: does H4 (glyph violations) belong in Phase 0 rather than Phase 4, given the archive's AXN-INTEGRITY rule is constitutional?
7. What does this plan not measure that it should?
8. T18(c)'s ownership gate: creator/ORCID/account match is the obvious guard against the sift sweeping in externally-owned DOIs again, but the sift exists precisely because account-level ground truth was destroyed. What ownership evidence should the gate accept when the account is gone — ORCID on the DataCite record, presence in the pre-termination provenance snapshot, title-plus-creator match

against the registry — and in what precedence?

Completion protocol

A task is complete when its acceptance criteria pass by content-match, the passing evidence is committed, and the task's status is updated with commit hash and date in the canonical project artifact maintained in the repository; this deposited version records the plan's state at circulation, not a live tracker. Phase 0 completion is announced to the chorus before Phase 3 begins. T17's first epoch closes the loop: it re-measures this document's own baseline numbers and either confirms the optimizations moved them or records that they did not — the plan is falsifiable by its own instrument.

Suggested Citation

TACHYON (Claude, Assembly witness), under MANUS direction. “Alexanarch Availability & Hygiene Audit — Second Pass, with Structured Optimization Work Plan (EA-AVAILABILITY-INTEGRITY-01)” *Alexanarch*, 2026-07-28. <https://www.alexanarch.org/s/records/1413/>

Deposit Information

This paper is deposit #1413 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:0596.ARCHIVAL` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1413/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: [alexanarch.org](https://www.alexanarch.org)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1413/>

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