

Network-wide Link Inventory v1.4

Lee Sharks

2026-07-08

Network-wide Link Inventory v1.4

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

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

2026-07-08 · Version v1.0 · Dataset

AXN:0431.ARCHIVAL

<https://www.alexanarch.org/s/records/1056/>

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": "Network-wide Link Inventory v1.4",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-08",
  "identifier": "AXN:0431.ARCHIVAL",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/1056/",
  "description": "A per-link database of every anchored hyperlink across all 34 substantive repositories in the lee
59,431 rows, each carrying anchor text, current href, surrounding prose context, extracted DOI/AXN references, resc
```

this inventory audits the resolver's real-world usage across the site network. Assembled to answer a question the earlier 440K-row link-token scan could not: not just how many links there are, but whether each link is right. Immediate use is a review pile for correcting broken and mismatched links across sites, with corrections applied preferentially to the resolver map at its source of truth (deposit #4 v3.11.0+) rather than to individual site source files, so sovereign resolver behavior propagates automatically via `/go/?doi=X`. Artifacts at `/data/link-inventory/`: `inventory-summary.json` (aggregate stats), `link-review-actionable.csv` (623 rows, high-confidence mismatches), `link-review-ambiguous.csv` (830 rows, DOI-carrying ambiguous cases with promising alt). Regeneration scripts at `/scripts/link-inventory/`. Iteration policy: rescan after each rewrite pass or resolver update; bump minor version; the versioned snapshots become the durable record of network-wide link integrity. v1.1 (2026-07-08): 106 fixes applied since v1.0 — 69 sentinel-0 hrefs corrected to specific records via unique-substring anchor-text matching, 34 `href_disagrees_resolver` corrections, 3 ambiguous fixes. Verdict deltas: `href_wrong_target` 258→208, ambiguous 11163→11149, consistent 1286→1303. v1.2 (2026-07-08): additional fix rounds. Total inventory-driven fixes this session: 448 (69 sentinel-0 unique-substring + 34 disagrees verified + 3 ambiguous + 127 loose sentinel-0 + 126 loose disagrees + 40 ambiguous loose + 49 relaxed-criteria final). Plus 8,462 earlier Phase 10 rewrites. Final verdict deltas from v1.0: `href_wrong_target` 258→87 (-171), ambiguous 11163→11124 (-39), consistent 1286→1322 (+36), `href_disagrees_resolver` 365→569 (+204 — mostly analyzer sensitivity from fixes cascading into DOI-based mismatch flags that were not previously visible because href was sentinel). Known source-side editorial issue documented: lagrange-observatory Generative Monoculture DOI 20675438 typo (should be 20675199 or 20674488) — resolver mapping is correct, source HTML has wrong DOI hardcoded. v1.3 (2026-07-08): 827 per-citation DOI-typo corrections. Cross-clustering by (`wrong_doi`, `correct_doi`) revealed 478 unique off-by-N pairs, most differing by ± 1 in the DOI number (systematic citation typos from adjacent Zenodo mints). Fixes applied per-file:line — each rewrite touches exactly the one citation where prose describes work Y but anchor cites a DOI resolving to work Z. Preserves legitimate citations of the same DOI at pages where it correctly cites its true target. V6 global DOI-swap pass was rolled back after breaking such legitimate citations. v1.4 (2026-07-08): Phase 14 comprehensive sweep. 750 additional fixes: 88 multi-signal anchor+heading+prose matches, 532 context-DOI `/go/` routing (for anchor links where the DOI appears in adjacent prose/table cell rather than in anchor text), 130 residual mixed. Final actionable state ~5% of v1.0 count.

Abstract

A per-link database of every anchored hyperlink across all 34 substantive repositories in the leesharks000 network — 59,431 rows, each carrying anchor text, current href, surrounding prose context, extracted DOI/AXN references, resolver lookup, prose-vs-resolver-title jaccard, best-matching alternative registry record, and a verdict flag classifying each link into review categories (`no_id`, `ambiguous`, `consistent`, `href_disagrees_resolver`, `external_ok`, `href_wrong_target`, `likely_prose_mismatch`, `ok`). Companion to DOI Resolution Index (#4) — this inventory audits the resolver's real-world usage across the site network. Assembled to answer a question the earlier 440K-row link-token scan could not: not just how many links there are, but whether each link is right. Immediate use is a review pile for correcting broken and mismatched links across sites, with corrections applied preferentially to the resolver map at its source of truth (deposit #4 v3.11.0+) rather than to individual site source files, so sovereign resolver behavior propagates automatically via `/go/?doi=X`. Artifacts at `/data/link-inventory/`: `inventory-summary.json` (aggregate stats), `link-review-actionable.csv` (623 rows, high-confidence mismatches), `link-review-ambiguous.csv` (830 rows, DOI-carrying ambiguous cases with promising alt). Regeneration scripts at `/scripts/link-inventory/`. Iteration policy: rescan after each rewrite pass or resolver update; bump minor version; the versioned snapshots become the durable record of network-wide link integrity. v1.1 (2026-07-08): 106 fixes applied since v1.0 — 69 sentinel-0 hrefs corrected to specific records via unique-substring anchor-text matching, 34 `href_disagrees_resolver` corrections, 3 ambiguous fixes. Verdict deltas: `href_wrong_target` 258→208, ambiguous 11163→11149, consistent 1286→1303. v1.2 (2026-07-08): additional fix rounds. Total inventory-driven fixes this session: 448 (69 sentinel-0 unique-substring + 34 disagrees verified + 3 ambiguous + 127 loose sentinel-0 + 126 loose disagrees + 40 ambiguous loose + 49 relaxed-criteria final). Plus 8,462 earlier Phase 10 rewrites. Final verdict deltas from v1.0: `href_wrong_target` 258→87 (-171), ambiguous 11163→11124 (-39), consistent 1286→1322 (+36), `href_disagrees_resolver` 365→569 (+204 — mostly analyzer sensitivity from fixes cascading into DOI-based mismatch flags that were not previously visible because href was sentinel). Known source-side editorial issue documented: lagrange-observatory Generative Monoculture DOI 20675438 typo (should be 20675199 or 20674488) — resolver mapping is correct, source HTML has wrong DOI hardcoded. v1.3 (2026-07-08): 827 per-citation DOI-typo corrections. Cross-clustering by (`wrong_doi`, `correct_doi`) revealed 478 unique off-by-N pairs, most differing by ± 1 in the DOI number (systematic citation typos from adjacent Zenodo mints). Fixes applied per-file:line — each rewrite touches exactly the one citation where prose describes work Y but anchor cites a DOI resolving to work Z. Preserves legitimate citations of the same DOI at pages where it correctly cites its true target. V6 global DOI-swap pass was rolled back after breaking such legitimate citations. v1.4 (2026-07-08): Phase 14 comprehensive sweep. 750 additional fixes: 88 multi-signal anchor+heading+prose matches, 532 context-DOI `/go/` routing (for anchor links where the DOI appears in adjacent prose/table cell rather than in anchor text), 130 residual mixed. Final actionable state ~5% of v1.0 count.

Body

title: “Network-wide Link Inventory v1.0” axn: “AXN:0431.ARCHIVAL.” creator: “Lee Sharks” date: “2026-07-08” content_type: “Dataset” license: “CC-BY-4.0” substrate: “AI-assisted (TACHYON, tool)” version: “v1.4” status: “ACTIVE” *** # Network-wide Link Inventory v1.4

Lee Sharks · 2026-07-08 · Dataset

AXN: AXN:0431.ARCHIVAL.

Description

A per-link database of every anchored hyperlink across all 34 substantive repositories in the `leesharks000` network. Each row records the anchor text, current href, ~200 characters of prose before and after the link, ~1,600 characters of enclosing prose context, extracted DOI/AXN references from anchor text and href and context, a resolver lookup for any extracted DOI (returning the current live target, the resolver’s stored title, the mapping type), a prose-vs-resolver-title jaccard similarity, the best-matching alternative record in the alexanarch registry, and a verdict flag classifying the link into review categories.

This dataset was assembled to answer a question that the earlier link-scan (a 440,012-row token count across the same 34 repos) could not: not just *how many* links there are, but *whether each link is right* — whether the href points at what the anchor text says, whether the DOI in the anchor text matches the descriptive prose around it, and whether the resolver’s current mapping for that DOI is consistent with everything the surrounding context asserts.

v1.1 (2026-07-08): 106 fixes applied since v1.0 — 69 sentinel-0 hrefs corrected to specific records via unique-substring anchor-text matching against the registry; 34 href_disagrees_resolver corrections where the anchor uniquely matched a different record than the current href; 3 ambiguous fixes with high-confidence matches. Verdict deltas: href_wrong_target 258→208 (-50); ambiguous 11163→11149 (-14); consistent 1286→1303 (+17); href_disagrees_resolver 365→428 (+63, false-positive increase — sentinel-0 fixes now expose disagreement with resolver’s DOI-based guess for nearby DOIs, though anchor-text-based fixes are semantically correct).

The immediate use is a review pile for correcting broken and mismatched links across sites — with corrections applied preferentially to the resolver map at its source of truth (deposit #4 v3.11.0+) rather than to individual site source files, so that the sovereign resolver behavior propagates automatically to every site link via the `/go/?doi=X` endpoint.

Companion deposits

Companion to [DOI Resolution Index v3.11.0](#) (#4) — this inventory audits the resolver’s real-world usage across the site network. Complements [Zenodotus’ Book-Burning](#) (#1) which established the sovereign-succession framework, and [DOIs Permanent Identifiers](#) (#868) which characterizes the substrate-wide breakdown of DOI persistence.

Artifacts

Files live at `/data/link-inventory/`:

- **inventory-summary.json** (2 KB, compact JSON) — scan metadata, verdict distribution, per-repo counts, regeneration pointers.
- **link-review-actionable.csv** (~370 KB, 623 rows) — links with high-confidence mismatch verdicts: href_wrong_target (sentinel `/s/records/0/`), href_disagrees_resolver (href points at record N;

resolver says the DOI-in-text should go to M N), `href_disagrees_text` (DOI in text DOI in href), and `likely_prose_mismatch` (prose context very different from what resolver says the DOI is, AND a much better alt record exists).

- **link-review-ambiguous.csv** (~600 KB, 830 rows) — DOI-carrying **ambiguous** cases with a promising alternative record match (`best_alt_jaccard` 0.15).

Regeneration scripts live at `/scripts/link-inventory/`:

- `build_link_review_db.py` — the scanner; produces the SQLite links database from a local mirror of the 34 repos.
- `enrich_link_review_db.py` — the enrichment pass; joins each link against the current DOI resolution index and computes jaccards, best-alt matches, and verdicts.
- `build_review_report.py` — exports the review artifacts (summary MD, actionable CSV, ambiguous CSV).
- `rewrite_links.py` / `rewrite_anchor_text.py` — the two rewriters used in the Phase 10 pass (URL-form and semantic anchor-text).

The 81 MB SQLite database itself (`links_review.db`) is not shipped in-repo; it is regenerable from the scripts against any local clone of the network. Scan takes about 32 seconds on a modern laptop.

Schema (SQLite links table)

column	type	description
<code>id</code>	INTEGER PK	auto-incrementing row id
<code>repo</code>	TEXT	repository name
<code>file_path</code>	TEXT	path relative to repo root
<code>line_no</code>	INTEGER	line number of the link's opening character
<code>link_form</code>	TEXT	<code>html_a</code> (HTML <code>TEXT</code>) or <code>md</code> (markdown <code>[TEXT] (X)</code>)
<code>anchor_text</code>	TEXT	visible link text (HTML tags stripped, whitespace normalized)
<code>href</code>	TEXT	the URL the link currently points at
<code>context_before</code> / <code>context_after</code>	TEXT	~200 chars of plain-text prose immediately adjacent
<code>context_prose</code>	TEXT	up to ~1,600 chars of broader enclosing prose
<code>doi_in_text</code> / <code>doi_in_href</code> / <code>doi_in_context</code>	TEXT	first 10.5281/zenodo.N DOI found in each field
<code>axn_in_text</code> / <code>axn_in_href</code>	TEXT	first AXN:HEX four-hex identifier found in each field
<code>href_record_num</code>	INTEGER	record number if href matches <code>alexanarch.org/s/records/N/</code>
<code>resolver_doi</code>	TEXT	the DOI used for resolution (priority: text → href → context)

column	type	description
<code>resolver_target</code>	TEXT	current live target URL per the api map
<code>resolver_title</code>	TEXT	resolver's stored title for that DOI
<code>resolver_record</code>	INTEGER	alexanarch record number if <code>resolver_target</code> is a record page
<code>resolver_mapping_type</code>	TEXT	<code>direct</code> , <code>direct_verified</code> , <code>title_match_repoint</code> , <code>no_alexanarch_equivalent</code> , etc.
<code>prose_vs_resolver_jaccard</code>	REAL	token-set jaccard between (anchor+prose) and the actual target record's title
<code>best_alt_record</code>	INTEGER	registry record whose title matches (anchor+prose) better than resolver's
<code>best_alt_title / best_alt_jaccard</code>	TEXT / REAL	title and score of that alternative
<code>verdict</code>	TEXT	see next section
<code>verdict_confidence</code>	TEXT	<code>high</code> , <code>medium</code> , <code>low</code>

Verdict taxonomy

verdict	meaning	how to act on it
<code>no_id</code>	no DOI/AXN anywhere in text, href, or context — navigation, image, external, non-citation	ignore for review; not a citation link
<code>ambiguous</code>	DOI present but prose context matches resolver's target title only weakly	filter to <code>best_alt_jaccard</code> 0.15 for the promising subset; review individually
<code>consistent</code>	DOI in text matches DOI in href, or link is internally coherent with resolver	no action needed
<code>href_disagrees_resolver</code>	href points at record N; resolver says the DOI in text should go to M N	fix href to the resolver's target via <code>/go/?doi=X</code> OR correct the resolver mapping if wrong
<code>external_ok</code>	external doi.org / zenodo.org / non-network site link	passthrough; may still be dead but not the resolver's concern
<code>href_wrong_target</code>	href points at <code>/s/records/0/</code> sentinel (the “no target known” placeholder)	always wrong; fix href to correct record via DOI-in-text (<code>/go/?doi=X</code>) or descriptive prose match
<code>unknown</code>	edge case, unclassified	investigate individually

verdict	meaning	how to act on it
likely_prose_mismatch	prose describes work X but linked DOI resolves to a very different work, AND a much better alt record exists in registry	probable DOI-in-source-is-wrong OR resolver-mapping-is-wrong; needs editorial judgment
ok	strong jaccard between prose and resolver's target title	no action needed

Verdict distribution as of v1.0

verdict	count	share
no_id	45,925	77.3%
ambiguous	11,163	18.8%
consistent	1,286	2.2%
href_disagrees_resolver	365	0.6%
external_ok	317	0.5%
href_wrong_target	258	0.4%
unknown	115	0.2%
ok	2	0.003%
total	59,431	100.0%

Mismatch concentration by repo (top 15)

repo	actionable mismatches
semanticphysics-site	516
semantic-economy	25
leesharks.com	21
lee-sharks-corporate	13
spxi-dev	12
provenance-erasure	9
restoredacademy	7
laborvector	6
living-architecture-lab	4
lee-sharks-consulting	3
maryleelabor-org	2
machinemediation-org	2
watergiraffe-org	1
vpcor-org	1
lagrange-observatory	1

Note: this is the inventory as of the anchor-text rewrite pass of 2026-07-08. Repos with rewrites still-not-deployed carry undercounted `href_wrong_target` counts (fixed in git but not yet in production).

Worked example — the “Generative Monoculture goes to Meaning Caste” case

MANUS flagged `lagrange-observatory/index.html:250` — anchor text DOI 10.5281/zenodo.20675438 next to prose about “Generative Monoculture: Model Collapse in Code” — but that DOI resolves to record #198 “The Self-Audit Module Dissolved”. The actual “Generative Monoculture” work is at record #199 (Talos Morrow, June 13 2026) or the re-mint at #1023 (EA-UMBML-MONOCULTURE-01).

The row in the inventory:

- `anchor_text` = “DOI 10.5281/zenodo.20675438”
- `href` = “https://www.alexanarch.org/go/?doi=10.5281/zenodo.20675438”
- `doi_in_text` = “10.5281/zenodo.20675438”
- `resolver_target` = “https://alexanarch.org/s/records/198/”
- `resolver_title` = “The Self-Audit Module Dissolved...”
- `prose_vs_resolver_jaccard` = 0.043
- `best_alt_record` = #188 (adjacent link in source, not the correct match — the true match at #199/#1023 has weaker jaccard because the scanner captured trailing prose after the h3 title rather than the h3 itself)
- `verdict` = ambiguous

Two possible root causes and neither can be picked automatically: (1) the DOI in the lagrange source is wrong for the work described (someone typed the wrong DOI when the page was authored), or (2) the resolver’s 20675438 → #198 mapping is wrong (Phase 6/7 fuzzy title matching landed the wrong record). The inventory surfaces the ambiguity; a MANUS-level decision picks which side is authoritative.

Iteration policy

The inventory regenerates cleanly from source at any point — no in-place mutation, no state drift. Standard iteration cycle:

1. Re-run `build_link_review_db.py` after each rewrite pass or resolver update to refresh the row set.
2. Re-run `enrich_link_review_db.py` after any resolver-map correction to refresh verdict flags.
3. Re-run `build_review_report.py` to regenerate the CSV exports.
4. Bump this deposit’s minor version and update the summary JSON with new counts.

The versioned snapshots in `/data/link-inventory/` are the durable record of network-wide link integrity at each rescan point; they can be diffed across time to show which repos are converging toward zero mismatches.

Related identifiers

- **Companion (DOI Resolution Index):** `/s/records/4/` — the sovereign resolver whose correctness this inventory audits.
- **Companion (DOIs Permanent Identifiers):** `/s/records/868/` — the theoretical framing of DOI-persistence breakdown, of which the mismatches surfaced here are one empirical face.
- **Framing (Zenodotus’ Book-Burning):** `/s/records/1/` — the sovereign-succession premise that made the resolver necessary.

= 1

Suggested Citation

Lee Sharks. “Network-wide Link Inventory v1.4” *Alexanarch*, 2026-07-08. <https://www.alexanarch.org/s/records/1056/>

Deposit Information

This paper is deposit #1056 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:0431.ARCHIVAL` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1056/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1056/>

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