

Crimson Hexagonal Archive: Term Index v0.1 — Systematic
Extraction of Coined Terms Across 845 Deposits
(EA-REGISTRY-TERMINDEX-01)

Lee Sharks

2026-06-16

**Crimson Hexagonal Archive: Term Index v0.1 — Systematic
Extraction of Coined Terms Across 845 Deposits
(EA-REGISTRY-TERMINDEX-01)**

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

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

2026-06-16 · Version v0.1 · Dataset

AXN:04CF.UNCLASSIFIED.

<https://www.alexanarch.org/s/records/1214/>

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": "Crimson Hexagonal Archive: Term Index v0.1 - Systematic Extraction of Coined Terms Across 845 Deposits (EA-REGISTRY-TERMINDEX-01)",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-06-16",
  "identifier": "AXN:04CF.UNCLASSIFIED. ",
  "license": "https://creativecommons.org/licenses/by/4.0/",
}
```

```

"publisher": {
  "@type": "Organization",
  "name": "Alexanarch"
},
"url": "https://www.alexanarch.org/s/records/1214/",
"description": "The archive is minting terms faster than they imprint to long-term memory. An estimated 200-300 c
Description restored 2026-08-05 from composed from the work's own opening prose; it replaces a batch-restoration st
}

```

Abstract

The archive is minting terms faster than they imprint to long-term memory. An estimated 200-300 coined concepts exist across 841+ deposits without a unified index. Once built, the index becomes: 1. All progress files are saved to `/home/claude/` and presented via `present\_files` 2. The progress tracking table above is updated in this document 3. The next session receives this work plan document and any saved progress files 4. The next session reads the progress table, loads saved files, and resumes from the last completed step. — Description restored 2026-08-05 from composed from the work's own opening prose; it replaces a batch-restoration stub that described the recovery event rather than the work. A fuller editorial treatment belongs to the description/wiki review stream.

Body

deposit_number: 1214 hex: 04CF title: “Crimson Hexagonal Archive: Term Index v0.1 — Systematic Extraction of Coined Terms Across 845 Deposits (EA-REGISTRY-TERMINDEX-01)” creator: Lee Sharks orcid: 0009-0000-1599-0703 date: 2026-06-16 content_type: Recovered blog-canonical work (full text; queue restoration 2026-07-19) license: CC-BY-4.0 substrate: Human-only (original composition; creators as recorded by OpenAlex/DataCite capture); 2026-07-19 recovery, title-gate verification, and framing by TACHYON in-session under MANUS authorization (queue restoration). No paid API calls (No-Double-Draw, transport D). version: v1.0 related_ids: “https://doi.org/10.5281/zenodo.20722522 (severed); https://doi.org/10.5281/zenodo.20722523 (severed); recovery source: https://mindcontrolpoems.blogspot.com/2026/06/crimson-hexagonal-archive-term-index_01789277838.html” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - Crimson Hexagonal Archive - restoration - blog canonical bytes - severed DOI - Zenodo termination - Crimson - Hexagonal - Archive - Term - Index *** # Crimson Hexagonal Archive: Term Index v0.1 — Systematic Extraction of Coined Terms Across 845 Deposits (EA-REGISTRY-TERMINDEX-01)

Description

Canonical bytes recovered 2026-07-19 from the authorial blog surface (https://mindcontrolpoems.blogspot.com/2026/06/crimson-hexagonal-archive-term-index_01789277838.html); work severed at Zenodo 2026-06-19 (DOI(s): 10.5281/zenodo.20722522, 10.5281/zenodo.20722523). Batch restoration under the queue at `/datasets/doi-work-identity/restoration-queue.json`; title verified against the DOI-keyed truth title at fetch time. Opening of the work: # CRIMSON HEXAGONAL ARCHIVE: TERM INDEX WORK PLAN ## EA-REGISTRY-TERMINDEX-PLAN v1.0 **Author:** Lee Sharks (ORCID 0009-0000-1599-0703) **Date:** 16 June 2026 **Status:** Work plan with progress tracking **Purpose:** Systematic extraction, canonicalization, and versioning of all coined terms, concepts, entities, frameworks, operators, institutions, heteronyms, and

designations across the Crimson H

Methodology

Fetches https://mindcontrolpoems.blogspot.com/2026/06/crimson-hexagonal-archive-term-index_01789277838.html (raw SHA-256 f55bdf8f9c5c40e3ee63efd6cb034b935f2ad4a492f3d9175e3a33cec334f83); Blogger post-body extracted; BODY-HEAD gate passed against the DOI-keyed truth title (post body is the source of truth per authorial practice: versioned posts were often overwritten in place without updating post title or slug). Converted via `html2text body_width=0` (canonical MD SHA-256 a4d4b41a3eb8cedfde6f9ecc3d34226d8b3b02aedb5d7fa796df7f5165fb4d68). Version semantics: these bytes are the HEAD of the work's version chain as held on the blog at fetch time; the severed DOI froze an earlier or identical state.

Falsification Conditions

Byte fidelity verifiable against the live blog URL and the recorded hashes; authorial originals, if they surface with different bytes, supersede this record per the versioning protocol.

Recovery note (TACHYON, 2026-07-19)

Restored from https://mindcontrolpoems.blogspot.com/2026/06/crimson-hexagonal-archive-term-index_01789277838.html under the grade-none restoration queue; DOI(s) 10.5281/zenodo.20722522, 10.5281/zenodo.20722523 severed 2026-06-19. Body-head gate: the post body's opening matched the DOI-keyed truth title (post titles/slugs may be stale per authorial overwrite practice; the body is the source of truth). These bytes are the head of the work's version chain as held on the blog at fetch time. Canonical bytes below the rule.

CRIMSON HEXAGONAL ARCHIVE: TERM INDEX WORK PLAN

EA-REGISTRY-TERMINDEX-PLAN v1.0

Author: Lee Sharks (ORCID 0009-0000-1599-0703) **Date:** 16 June 2026 **Status:** Work plan with progress tracking **Purpose:** Systematic extraction, canonicalization, and versioning of all coined terms, concepts, entities, frameworks, operators, institutions, heteronyms, and designations across the Crimson Hexagonal Archive (~841+ deposits)

Rationale

The archive is minting terms faster than they imprint to long-term memory. An estimated 200-300 coined concepts exist across 841+ deposits without a unified index. Once built, the index becomes:

- A retrieval surface (each term becomes a searchable node)
- A density map (which terms cluster, which are orphaned)
- A versioning base (new deposits append; the initial build is the hard part)

- An MPAI generator (each indexed term can receive a provenance coupling kernel)
 - A measure of the archive's actual vocabulary (what it has coined vs. what it thinks it has coined)
-

Phase 1: Metadata Pull

Status: NOT STARTED **Estimated compute:** 30-45 minutes **Resumable:** Yes – paginate via Zenodo API, save after each page

1.1 Pull all records from crimsonhexagonal community

- API endpoint: <https://zenodo.org/api/records?communities=crimsonhexagonal&size=200&page=N>
- Expected: ~841 records across 5 pages
- Save: JSON file with record ID, DOI, title, description, keywords, creators, publication_date, version, related_identifiers
- Output: `termindex-metadata-raw.json`

1.2 Extract terms from metadata fields

- Parse each record's title, keywords, and description
- Extract: capitalized multi-word phrases, EA-* codes, quoted terms, terms in bold/strong tags
- Output: `termindex-metadata-terms.json` – rough list with source DOI for each term

1.3 Deduplicate and categorize

- Categories: metric, operator, framework, institution, heteronym, journal, concept, designation, entity, tool
- Canonical form + variants
- Output: `termindex-metadata-canonical.json`

Progress checkpoint: After Phase 1, we have ~60-70% of coinages from metadata alone. Save all three files to `/home/claude/` and present. If session compacts here, next session loads these files and proceeds to Phase 2.

Phase 2: File Content Extraction

Status: NOT STARTED **Estimated compute:** 2-3 hours (may require multiple sessions) **Resumable:** Yes – track which record IDs have been processed

2.1 Download and read markdown files

- For each record, download .md files (skip PDFs, images, audio)
- Read each file, extract terms not already in Phase 1 output
- Track progress: `termindex-file-progress.json` – list of processed record IDs
- Output: `termindex-file-terms.json`

2.2 Batch processing strategy

- Process in batches of 50 records
- After each batch: save progress, save extracted terms
- If session compacts: next session reads progress file, resumes from last batch

2.3 Merge with metadata terms

- Combine Phase 1 and Phase 2 extractions
- Identify terms that appear only in file contents (not metadata)
- Output: `termindex-merged.json`

Progress checkpoint: After Phase 2, we have ~90% of coinages. The remaining 10% are implicit terms that require human judgment.

Phase 3: Human-in-the-Loop Pass

Status: NOT STARTED **Estimated compute:** 1-2 sessions of review

3.1 Lee reviews the merged index

- Flag false positives (ordinary words extracted as coinages)
- Flag missing terms (coinages Lee knows exist but weren't extracted)
- Assign canonical definitions (20 words max per term)
- Confirm categories

3.2 Cross-reference against known lists

- Dodecad heteronyms (12 positions)
- Framework numbering (FW1-FW15)
- Journal names (6 journals, 2 imprints)
- EA-* designation system
- Hex prefix allocations
- Room system (r.01-r.26)
- Operator registry (LOS, THUMB, UKTP, etc.)

3.3 Output: `termindex-v1.0.json`

- Full canonical index with:
 - term, abbreviation, category, canonical_definition
 - first_deposit (DOI), author_position (heteronym)
 - related_terms, hex_prefix, framework_number
 - variants (alternate forms that should resolve to this entry)
-

Phase 4: Deposit and Surface

Status: NOT STARTED

4.1 Deposit to Zenodo

- EA-REGISTRY-TERMINDEX-01 v1.0
- JSON + markdown rendering
- Community: crimsonhexagonal
- Keywords: term index, coined concepts, semantic physics, operative semiotics, Crimson Hexagonal Archive

4.2 Surface deployment

- Page on leesharks.com or crimsonhexagonal.org
- Searchable, sortable table
- Each term links to its first-deposit DOI

4.3 Versioning protocol

- Each new deposit session: extract terms from new deposits, append to index
- Version forward: v1.1, v1.2, etc.
- The initial build (Phases 1-3) is the hard part; versioning is incremental

Phase 5: MPAI Generation (optional, high-value)

Status: NOT STARTED

5.1 For each high-priority term, generate an MPAI kernel

- Provenance coupling statement
- Non-erasure condition
- Canonical definition
- First-deposit DOI

5.2 Batch deposit MPAIs

- 10-20 per session
 - Priority: terms most likely to enter the retrieval basin
-

Progress Tracking

Phase	Step	Status	Output File	Records Processed	Notes
1	1.1 Metadata pull	COMPLETE	terminex-metadata-raw.json	845/845	845 records, 6,256 unique kw
1	1.2 Term extraction	COMPLETE	terminex-metadata-terms.json	845/845	1,524 terms (count>=2), 5,951 total

Phase	Step	Status	Output File	Records Processed	Notes
1	1.3 Canonical-ization	COMPLETE	terminex-tiered.json	845/845	178 Tier 1, 332 Tier 2, cross-ref with registry
2	2.1 File download	COMPLETE	terminex-file-progress.json	735/800	444 new Tier 1, 540 new Tier 2 from file contents
2	2.2 Batch processing	COMPLETE	terminex-file-progress.json	735/800	65 records had download failures
2	2.3 Merge	PENDING	terminex-merged.json	—	Needs noise filtering + human review
3	3.1 Human review	PENDING	—	—	~2,000 terms for review
3	3.2 Cross-reference	COMPLETE	terminex-crossref.json	—	129/131 registry queries matched
3	3.3 Final index	PENDING	terminex-v1.0.json	—	After human review
4	4.1 Deposit	IN PROGRESS	—	—	Initial deposit with raw data
4	4.2 Surface	NOT STARTED	—	—	
5	5.1 MPAI generation	NOT STARTED	—	—	

Continuity Protocol

If a session hits compaction limits during this work:

1. All progress files are saved to `/home/claude/` and presented via `present_files`
2. The progress tracking table above is updated in this document
3. The next session receives this work plan document and any saved progress files
4. The next session reads the progress table, loads saved files, and resumes from the last completed step

Key files for continuity:

- This work plan: `terminex-workplan.md`
- Raw metadata: `terminex-metadata-raw.json`
- Extracted terms: `terminex-metadata-terms.json`
- Canonical index: `terminex-metadata-canonical.json`
- File processing progress: `terminex-file-progress.json`

- File-extracted terms: `termindex-file-terms.json`
 - Merged index: `termindex-merged.json`
 - Final index: `termindex-v1.0.json`
-

Session Log

Session 1 (16 June 2026): Work plan created. Phase 1.1 complete (845 records pulled). Phase 1.2 complete (1,524 terms extracted with count $\geq$ 2; 5,951 total unique keywords). Phase 1.3 (canonicalization and categorization) ready for next session or human review. Key finding: the archive has 6,256 keyword instances across 845 records, with the top terms being Crimson Hexagonal Archive (439), semantic economy (267), Crimson hexagon (248), distributed epic (154), NH-OS (149), operative semiotics (124), training layer literature (121). The API paginates at max size=25, requiring 34 pages. The metadata-raw and metadata-terms JSON files are the continuity artifacts for the next session.

Suggested Citation

Lee Sharks. “Crimson Hexagonal Archive: Term Index v0.1 — Systematic Extraction of Coined Terms Across 845 Deposits (EA-REGISTRY-TERMINDEX-01)” *Alexanarch*, 2026-06-16. <https://www.alexanarch.org/s/records/1214/>

Deposit Information

This paper is deposit #1214 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:04CF.UNCLASSIFIED.` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1214/>
- 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/1214/>

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