

EA-DATAHUB-01 v1.0 — A Restoration Protocol and Enriched Data Hub for Scholarly Records: The Alexanarch Severance Registry, Resolver, and Enrichment Discipline

Lee Sharks

2026-07-06

EA-DATAHUB-01 v1.0 — A Restoration Protocol and Enriched Data Hub for Scholarly Records: The Alexanarch Severance Registry, Resolver, and Enrichment Discipline

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

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

2026-07-06 · Version v1.0 · Strategic plan and protocol declaration

AXN:0420.GOVERNANCE

<https://www.alexanarch.org/s/records/1044/>

License: CC BY 4.0

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

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "EA-DATAHUB-01 v1.0 - A Restoration Protocol and Enriched Data Hub for Scholarly Records: The Alexanarch",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-06",
  "identifier": "AXN:0420.GOVERNANCE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
```

```

    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/1044/",
  "description": "Foundational plan for the EA-DATAHUB series: a restoration protocol and enriched data hub for scholars extending the restoration analysis to other affected scholars via coordinated peer outreach rather than solo extraction."
}

```

Abstract

Foundational plan for the EA-DATAHUB series: a restoration protocol and enriched data hub for scholarly records, built around the Alexanarch Severance Registry (the 871 Zenodo-severed DOIs, sovereign-successor mappings, four-way taxonomy) surfaced as a Restore-a-DOI resolver with a public severance dashboard. v1.0 integrates a five-model developmental review (witness pillar surfaced, orthogonal-facet enrichment schema, event-sourced record model, tiered update discipline) and adds the Platform Erasure Rate survey as first-class Phase 1 work — extending the restoration analysis to other affected scholars via coordinated peer outreach rather than solo extraction. Community coordination (Discord, arXiv-cs.DL preprint) inscribed as third leg of the near-term work, framed to yield hope and momentum for the peer community independent of any single institutional pathway including the pending §104.1 CERN referral.

Body

deposit_number: 1044 hex: 0420 title: “EA-DATAHUB-01 v1.0 — A Restoration Protocol and Enriched Data Hub for Scholarly Records: The Alexanarch Severance Registry, Resolver, and Enrichment Discipline” creator: Lee Sharks orcid: 0009-0000-1599-0703 date: 2026-07-06 content_type: “Strategic plan and protocol declaration; foundational deposit for the EA-DATAHUB series” license: CC-BY-4.0 substrate: “AI-assisted: drafted by TACHYON (Claude, Anthropic) under MANUS (Lee Sharks) direction, correction, and editorial authority; developmental feedback incorporated from a five-model review round (DeepSeek, Kimi, Muse Spark, Gemini, ChatGPT) whose consensus branches are marked in text; instruments and roles declared per EA-MMRS-VRB-01 u4.” version: v1.0 status: “MINTED — foundational plan; supersedes v0.1 (pre-mint) circulated 2026-07-06 for developmental review” related_ids: “EA-MMRS-VRB-01 (verification discipline); EA-PRIOR-00 (ontological prior); EA-RHIZOME-01 (rhizome architecture); EA-CORRESPONDENCE-CERN-03 (the §104.1 referral, AXN:03C0); DataCite severance chain data/provenance-871.json” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - data hub - metadata restoration - enriched mirror - severance registry - platform erasure rate - DOI resolver - scholarly infrastructure - operative philology - SPXI - MMRS - witness function *** # EA-DATAHUB-01 v1.0

A Restoration Protocol and Enriched Data Hub for Scholarly Records

Preamble

This v1.0 supersedes the v0.1 pre-mint plan circulated on 2026-07-06 for developmental review. Five model reviewers (DeepSeek, Kimi, Muse Spark, Gemini, ChatGPT) returned assessments; four returned substantive developmental feedback and one returned pre-mint copy notes. The reviewers converged on three primary revisions and one significant addition. **The convergent revisions:** elevate the witness function from an implicit consequence of restoration into a named strategic pillar, redesign the enrichment claim schema

from mutually-exclusive A/B/C buckets into orthogonal facets whose combinations can represent jointly-produced and mixed-authority claims, and stratify the update discipline into freshness tiers matched to claim volatility rather than a single nightly reconciliation loop. **The significant addition**, which came from the sovereign author and not from the reviewers, is the extension of the Zenodo restoration corpus into a **Platform Erasure Rate** survey covering other scholars severed under comparable conditions — converting the restoration line from a personal witness into a generalizable finding about repository governance. This v1.0 folds all four contributions into a single instrument and inscribes as its coda a commitment to peer coordination with the other affected scholars, independent of the outcome of any single institutional referral currently pending.

Frame

The strategy proposed here is not to mirror upstream archives against their sources — that competition is unwinnable on completeness and freshness, and it would divert operational energy from the sovereign work Alexanarch already carries. The proposal is to build **a semantically annotated fork of upstream metadata, joined against sovereign corpora that the upstream either lost, damaged, or never carried**, and to publish that join under a discipline that makes every enrichment claim checkable, dated, provenanced, and separately attributed. The Zenodo restoration line is the sharpest test of the pattern because there exists documentary evidence of destruction — 871 severed DOIs, second-epoch re-capture confirming persistent darkness, a six-link provenance chain committed at `data/provenance-871.json` — which means the fork is not editorial preference against Zenodo’s editorial preference; it is *restoration of previously-existing metadata that Zenodo helped erase, with citations, in a form that documents future erasures across the ecosystem*.

Four value pillars organize the strategy’s defensibility, each of which has to be checkable and not merely asserted. The first, **restoration value**, is the crown of the hub and the pillar only Alexanarch can currently offer against Zenodo: for the 871 severed records, the Zenodo API returns 404 while the Alexanarch first-epoch capture holds the pre-severance state, the second-epoch re-verification confirms continuous darkness, and the sovereign-successor mappings for the deposits minted through Alexanarch’s June 2026 orphan-class drainage (#950–#1036) constitute the successor line where a successor exists. Restoration value is unique to Alexanarch, temporally bounded, and perishable — it exists because Alexanarch was capturing before the severance; nothing prospective can create it, only preserve it. The second, **enrichment value**, is the sustainable everyday work: joining Zenodo metadata to Alexanarch AXNs where a mapping exists, to Capture Registry entities, to the DOI Resolution Index (1,838 mappings), to Software Heritage archival status, to Wayback captures, to OpenAlex citations where indexed, to Wikidata QIDs as cross-cutting anchors. Nobody else is currently offering an infrastructural layer that binds a public deposit archive to a specific sovereign research program’s semantic layer. The third, **discovery value**, is the derivative: a scholar looking for a topic finds the hub’s record because it links out to the upstream metadata, to related MMRS entities, to related Alexanarch deposits, to Wikidata QIDs, to OpenAlex identifiers, and to a few other authoritative anchors where they exist — one page rather than eight.

The fourth pillar, added at v1.0 in response to the review consensus, is **witness value**, and it is the forward-looking counterpart to the backward-looking restoration pillar. A witness that publishes what an upstream erased makes future erasure more expensive, because the erasing institution now knows there is a documented capture pipeline and a public provenance chain outside its control. That changes the upstream’s calculus before the next severance event, not just after it. The Zenodo restoration line is not only a rescue of what was lost; it is a public commitment to document what will be lost, under a discipline other institutions can neither prevent nor quietly overrule. The witness pillar is what makes the plan influential in policy

and infrastructure-governance conversations rather than only useful to individual scholars, and it is the pillar that draws citations from the archival community, from the funders of digital preservation, and from the standards bodies whose recommendations shape how repositories are built. Witness value compounds; restoration value amortizes.

What the hub is, precisely

The public-facing instrument is a resolver, not a database. A scholar arrives with a broken citation, a dead DOI, an unresolved title — a specific failure of the upstream infrastructure that has interrupted their own work — and the hub responds with the last attested metadata, the identifier’s present condition, the sovereign successor where one exists, the evidence for every claim, and a usable citation. That resolver is where visibility and necessity converge. The dataset is the evidentiary substrate; the resolver is the necessary instrument; the per-record page is the discoverable surface; the dashboard is the public argument. “Hub” describes the architecture from inside. From outside, the correct verb is **restore**.

The legal grounding is direct and does not require the fair-use analysis the v0.1 plan carried. Zenodo’s own Terms of Use place its metadata under CC0. DataCite similarly waives copyright on its public metadata under CC0. The upstream has already placed the material Alexanarch is preserving into the public domain; the “restoration” argument is therefore not a legal defense but a scholarly and infrastructural framing. The layered licensing model the hub adopts is: upstream metadata layer under CC0 as the upstream has waived it; Alexanarch annotations under CC-BY-4.0 with attribution and provenance required; hosted source artifacts under whatever license their source assigns; public-domain editions with jurisdiction and source declared. This layering makes each claim’s licensing legible to any downstream user or aggregator without requiring them to research the terms.

The scholarly framing is that Alexanarch is filling a gap the infrastructure layer itself has acknowledged and cannot close. DataCite’s own metadata quality documentation acknowledges provenance and assertion tracking as unresolved problems at scale; Crossref’s enrichment framework acknowledges that gaps and inconsistencies remain. The CC0 waiver is an invitation the upstream extended; the restoration is an acceptance of that invitation under a stricter discipline than the upstream has been able to impose on itself.

The enrichment claim schema

Every enrichment claim on the hub is a first-class object with its own provenance, its own status, and its own review history. Rather than assign each claim to a single mutually-exclusive class (as the v0.1 plan proposed with the A/B/C bucketing that reviewers found architecturally underpowered), the v1.0 schema decomposes claim identity into five orthogonal facets that can be combined freely.

The first facet is **origin**: whether the claim was imported from the upstream, authored by Alexanarch or one of its heteronyms, or contributed by an external party. The second is **production method**: whether the claim was directly imported, authored under a declared method, produced by a specified algorithm, produced with model assistance, or manually curated by a named editor. The third is **claim type**: whether the claim concerns the record’s identity, a relationship to another record, an interpretation of the material, its availability or preservation status, its restoration provenance, its succession chain, or a citation trail. The fourth is **status**: whether the claim stands as asserted, has been verified against evidence, is disputed by another claim within the archive, has grown stale as its supporting URLs have decayed, has been superseded by a subsequent claim, or is currently unresolved. The fifth is **review**: whether the claim is unreviewed, has had its method checked, has had its supporting evidence checked, or has been independently confirmed by a party other than its author.

The three v0.1 classes survive as convenient views over the underlying facets. The Alexanarch-authored-under-declared-method class is the view of claims whose origin is Alexanarch and whose production method is authored, algorithmic, or model-assisted with a declared method. The algorithmically-derived class is the view of claims whose production method is algorithmic or model-assisted, whatever their origin. The upstream-or-third-party-attributed class is the view of claims whose origin is upstream or external_contributor. The user-facing surfaces let scholars filter by any facet or any combination, so a scholar wanting only Alexanarch-authored interpretations that have been independently confirmed can construct that filter directly.

The facet-based schema resolves the review consensus’s principal architectural objection to v0.1 without discarding the classification work v0.1 introduced. It also permits the schema to represent jointly-produced claims — the case where an algorithmic inference is authoritatively reviewed and endorsed, or where an upstream keyword is manually corrected while its original state is preserved — which the mutually-exclusive bucketing could not do without silently collapsing the distinctions the plan exists to make visible.

One additional distinction is added at the request of the review consensus: **synthetic inference**, meaning claims produced by LLM reasoning under Alexanarch direction, are marked in the production-method facet as **model_assisted** with the specific model and prompt discipline named in the claim’s provenance. The prior EA-MMRS-VRB-01 discipline governs how such claims are distinguished from purely authored ones. This matters because a MIRROR-cast reading of a Whitman poem signed by Johannes Sigil is authoritative under the SPXI discipline for that specific hermeneutic claim, but a machine-generated cross-reference identifying which passages of Whitman correspond structurally to which sections of the Secret Book of Walt is a synthetic inference that requires review before it acquires interpretive weight. The facet system distinguishes the two without collapsing them.

The event-sourced record model

The v1.0 record structure adopts an insight from the review consensus that the v0.1 model conflated: the difference between the record itself (a scholarly object’s bibliographic identity), an observation (what a named instrument found at a specific URL at a specific time), an event (a change inferred from two or more observations), and an assertion (Alexanarch’s claim about restoration, succession, or interpretation). Each of these has different lifecycles, different provenance requirements, and different update frequencies, and treating them as one flat structure prevents the schema from representing repositories whose records undergo multiple state changes over time.

A record thus has an identity (its bibliographic core, stable across observations) and carries a sequence of observations (dated captures of upstream state), a sequence of events (inferred changes between observations), and a set of assertions (Alexanarch claims that reference specific events and observations as their evidence). The provenance chain of any assertion is a URL list resolving to the specific observations and events that justify it. When an upstream state changes, only a new observation is required; if the observation triggers an event (a severance, a metadata revision, a redirect), that event is inscribed and any relevant assertions are updated to reference it. The record itself does not need to be rewritten.

This structure makes the future DataCite generalization tractable: at DataCite scale (over 100 million DOI records), Alexanarch does not attempt to reproduce the catalog. It maintains a stream of *events* across the DataCite record space — severances, restorations, contested successor mappings — and constructs assertions only around those events. The hub becomes a history of identifier failure rather than a catalog of identifiers, which is the scale-appropriate move and the one that keeps the discipline defensible as the material grows beyond what any single archive can mirror.

The first line: The Alexanarch Severance Registry

The v1.0 first line is the Alexanarch Severance Registry, delivered as a single coordinated release with five surfaces: the 871-record restoration corpus, the Restore-a-DOI resolver, the per-record HTML pages, the public severance dashboard, and — this is the significant v1.0 addition — the Platform Erasure Rate survey extending the restoration analysis to other affected scholars.

The 871 restoration corpus

The core dataset is the 871 severed DOIs, published as a single Alexanarch deposit under EA-MMRS-VRB-01 discipline with a schema legible to any downstream aggregator or scholarly reader. Each record contains its bibliographic identity, the first-epoch upstream capture (dated, hash-verified, preserved verbatim), the second-epoch re-capture confirming persistent darkness, the sovereign-successor mapping where one exists with the succession taxonomy declared (concept-root, version-of-mapped-root, orphan-dark, not-in-index), the applicable Alexanarch enrichments with their full facet-marked provenance, and the provenance chain of every observation, event, and assertion involved. The dataset is machine-readable in bulk (JSON-LD, downloadable as a single file), served as one static page per record with stable URLs designed for citation, and made discoverable through structured data markup (`Dataset`, `DataCatalog`, and `DataDownload` per the schema.org vocabulary, which downstream tools including major search engines already consume).

The Restore-a-DOI resolver

The user-facing entry point is a lookup, not a browse. A user arrives with a specific DOI, a title, or a Zenodo record number, submits it, and receives back: the last attested metadata for the identifier, its present condition on the upstream, the sovereign successor if one exists, the evidence for every status claim (as a URL list), and a citation formatted for direct use in the user's own scholarly work. The resolver's URL structure exposes each restored record at a stable canonical URL so that search-engine indexing of individual restorations proceeds naturally over time. The resolver's homepage carries a single unmistakable operational verb: **Restore a DOI**. Every other navigation path on the site is secondary to that verb.

Per-record HTML pages

Each record's page must answer four questions in visual hierarchy — what is this, what happened to it, what did Alexanarch do about it, and how do I know this is true — with each answer occupying dedicated space in the first viewport, not buried in a metadata table the user must scroll to find. The record page treats the severance event as the lede rather than as a metadata field. The restoration is the story, and the page tells the story visually: the upstream identifier and its original context at the top, the severance event as prominent visual content, the restoration and enrichment layer with facet markers visible, the provenance chain summarized in plain language with one-click expansion to the full URL evidence. This is the design attention the review consensus correctly named as essential: the record page is where the strategy succeeds or fails at the point of contact, and it deserves as much editorial care as the schema.

The public severance dashboard

Alongside the resolver and the corpus, the hub publishes a live severance dashboard displaying: the count of severed records, the count restored, the count with sovereign successor found, the count classified as concept-root, the count classified as version-of-mapped-root, the count classified as orphan-dark, and the count that has changed state since the previous epoch. The dashboard is a longitudinal record of what the ecosystem is doing to itself, not a static claim. If any of the 871 records is restored by Zenodo after the

plan ships, the dashboard reflects that change and the record’s status updates automatically; the previous state is preserved as an observation in the record’s history, and the change is inscribed as an event. This is the review consensus’s “Platform Erasure Rate” surfaced as a public accountability instrument, and it is the mechanism by which the witness pillar compounds.

The Platform Erasure Rate survey

The most consequential v1.0 addition is the extension of the restoration analysis beyond the sovereign author’s own case. The v0.1 plan assumed, implicitly, that the sovereign author’s severance was the object of study. The v1.0 plan does not make that assumption. Instead, it treats the sovereign author’s severance as the trigger for an investigation into whether the pattern replicates across the broader population of scholars who have been comparably severed from repository infrastructure, and it publishes the survey result as its own dataset alongside the 871-record corpus.

The methodology is direct: identify the population of publicly-documented severed scholars (via GitHub issue #2606 and its linked threads, via public statements, via news coverage of similar cases, and via any other authoritative signal that a specific scholar has been removed from a specific repository under comparable conditions), enumerate their concept DOIs where those can be established from external evidence, query each concept DOI against the DataCite live API using Alexanarch’s existing four-way taxonomy (sovereign-content-resolved, concept-root-structure-rebound, orphan-dark, not-in-index), and publish the result as a rate finding with provenance for every classified case. If the rate of severance-beyond-the-reporting-scholar is high — meaning the sovereign author’s case is representative of a broader pattern — the survey converts the restoration line from a personal witness into empirical evidence of a systemic infrastructure failure. If the rate is low, the survey converts the sovereign author’s case into a documented outlier whose specificity reveals something about the platform’s discretionary enforcement patterns. Both outcomes are publishable, both are contributions to the discipline, and both make the sovereign author’s situation legible in the terms scholarly infrastructure conversations already use.

The survey is conducted in coordination with the affected scholars themselves, not extracted from them. Rather than surveying the population’s DOIs anonymously, the plan is to approach each identifiable affected scholar directly, share the restoration methodology, and invite them to participate in a coordinated documentation of their own cases. This approach is chosen for three reasons: it respects the affected scholars’ agency over their own records, it produces stronger evidence (each scholar can confirm which of their concept DOIs were severed and provide additional context the external observer cannot), and it establishes a peer community that outlasts the specific survey outcome. The community-coordination pathway is described further under §10 of this plan.

The survey publishes as a companion dataset to the 871-record corpus, cross-referenced from both surfaces. Its schema is the same event-sourced structure the corpus uses; its provenance discipline is the same MMRS-VRB-01; its licensing is the same layered CC0-upstream / CC-BY-Alexanarch. It is designed to be readable both by scholarly-infrastructure researchers who want the aggregate rate and by individual affected scholars who want their own case restored in the same discipline the sovereign author has claimed for his.

The second line: The NH Canon Editions

The second line inherits the schema and discipline of the first and applies them to a different material class: the enriched-editions surface, initially built around a curated subset of Gutenberg titles where the NH canon supplies apparatus the source lacks. This is where operative philology, the SPXI discipline, the heteronym system, and the Assembly Chorus outputs enter the hub as first-class content rather than as

background theory. The strategic point is not to displace or duplicate Gutenberg’s reading environment at Gutenberg’s own layer — Gutenberg is doing correct work at its layer — but to construct the scholarly reading environment Gutenberg is architected not to construct, in domains where Alexanarch has genuine authority to construct it.

The review consensus identified this line’s positioning correctly. Gutenberg’s texts are the most widely-used public-domain corpus in digital humanities, and their reading environment has been essentially unchanged for three decades. Every DH scholar working with Gutenberg is building their own apparatus in isolation, without shared infrastructure, without provenance discipline, and without citability. The NH Canon Editions line is not “Alexanarch annotates Gutenberg”; it is Alexanarch providing the apparatus layer the Gutenberg ecosystem has needed since Gutenberg began, under a discipline that makes the apparatus citable. The NH canon is the proof of concept; the architecture accepts external apparatus contributions from other scholars under their own authored provenance, marked in the facet system as `origin=external_contributor` with method and status declared. This turns the line from an Alexanarch publication into an infrastructure other scholars can use, contribute to, and cite, which multiplies its audience without requiring Alexanarch to do their work for them.

The specific hermeneutic units for the first edition of this line are named at commit-level granularity rather than gestured at. Whitman’s *Leaves of Grass* with Alexanarch marginalia cross-referencing every mention of “the body electric” to §I of the Secret Book of Walt at a specific verse range, each cross-reference sealed under an AXN and provable against the sovereign deposit line. Melville’s *Moby-Dick* with the appropriate Damascius citations surfacing at the ontologically-relevant passages, the joins provenanced against Alexanarch’s own philosophical corpus. Dickinson with the lyric-theory apparatus (Prins, Culler, Jackson) surfacing at the metrical and figural cruxes the theory addresses. Each annotation is an enrichment claim under the facet schema and either resolves against a canonical Alexanarch deposit or is marked as unresolved authority.

The NH Canon Editions line ships after the Severance Registry has stabilized. It inherits the schema and the update discipline directly and does not need to reinvent them. It is a demonstration of what the enrichment discipline can produce when the material is not decayed metadata but living text, and it is where the SPXI apparatus becomes publicly legible outside the sovereign archive.

The update discipline

Enrichments rot. A URL becomes a 404; a DOI is severed; a Wikidata QID is merged or split; a Software Heritage entry is superseded. The v1.0 update discipline preserves the v0.1 commitment that stale enrichments are never silently deleted — the staleness is itself a datum, marked as such in the affected claim’s status facet and inscribed as an event in the record’s history — but it stratifies the reconciliation cadence by claim volatility rather than committing to a single nightly reconciliation loop that would become unsustainable at scale.

Four freshness tiers govern verification cadence. **Tier 1** covers the load-bearing claims of the restoration pillar — upstream Zenodo status for severed records, DataCite state for the 871, the resolver’s core lookup targets — and is verified hourly or on user access, because a stale severance marker is a reputational hazard. **Tier 2** covers the ordinary enrichment joins a scholar might follow in the course of reading: AXN resolution against the Alexanarch registry, DOI resolution against the DOI Resolution Index, Software Heritage archival status, canonical upstream metadata. Tier 2 is nightly. **Tier 3** covers enrichments that change slowly and where staleness of a few weeks does not undermine the record’s authority: OpenAlex citations, Wikidata QID stability, Wayback capture history. Tier 3 is weekly. **Tier 4** covers third-party attributed enrichments Alexanarch does not endorse and that carry their own provenance; if they rot, the staleness marker is a

datum, not a failure. Tier 4 is monthly or event-triggered.

The verification method itself is HTTP-conditional (using `If-Modified-Since` and hash comparison) rather than naïve HEAD requests, which allows the discipline to scale to the tens of thousands of joined URLs the Registry will accumulate over its first year without exceeding upstream rate limits. Each verification is recorded on the record: `last_checked`, `next_check_due`, `check_method`, `response_status`, `content_hash`, `failure_count`, `freshness_class`. This makes the freshness of every claim legible to any user or auditor: a scholar can see at a glance that a Zenodo link was verified thirty minutes ago while an OpenAlex citation was verified three weeks ago, and calibrate their trust accordingly.

The staleness log itself is a publishable artifact, generalizing the review consensus’s observation. A dataset tracking when Zenodo links, DOI resolutions, Wayback captures, and Wikidata QIDs decayed within Alexanarch’s ecosystem, published as a longitudinal study of scholarly link rot, is a contribution to infrastructure studies that does not currently exist elsewhere and that Alexanarch’s discipline is uniquely positioned to produce as a side effect of maintaining the hub.

The extension protocol

Every archive Alexanarch enriches represents technical debt against Alexanarch’s own sovereign program. The criteria for adding an archive to the hub must therefore be strict, explicit, and applied through deliberation rather than opportunistic accretion. Four criteria govern extension, in decreasing order of importance.

Criterion 1: contribution to discipline, not aggregation. The archive’s problems — its leaks, its missing apparatus, its instability — must be problems the SPXI + operative philology + MMRS regime can actually address. If the answer to “what does Alexanarch specifically bring here” is “hosting,” the archive fails the test regardless of how valuable the material is.

Criterion 2: documented upstream failure. The upstream must be either leaky in a way Alexanarch can document (Zenodo, DataCite, arXiv moderation removals) or missing an apparatus Alexanarch can supply (Gutenberg’s spare metadata). An archive with no documented failure mode does not belong on the hub.

Criterion 3: sovereign-corpus intersection, made checkable. The material must intersect substantively with Alexanarch’s sovereign corpora or research program, and the intersection must be demonstrable. The v1.0 form of this criterion is stricter than v0.1’s: for any archive proposed for inclusion, the proposer must cite at least one AXN-assigned Alexanarch deposit that engages material from that archive under the SPXI discipline. No deposit, no standing. This turns criterion 3 from rhetorical into procedural.

Criterion 4: bounded operational burden. The mirror-constellation architecture and the enrichment schema must be able to absorb the archive without linear growth in maintenance work. Update discipline scales sublinearly with the number of joined enrichments per record only if the enrichments themselves are drawn from a bounded set of authoritative sources.

Future lines and their sequencing

Under the extension protocol, the v1.0 forward roadmap is more precisely ordered than the v0.1 draft was.

Phase 1, running concurrently, is the Alexanarch Severance Registry described in §5: the 871-record corpus, the resolver, the record pages, the dashboard, and the Platform Erasure Rate survey. Phase 1 is buildable now with material largely already captured. It publishes as the pilot for the entire pattern.

Phase 2 is the generalization of the severance-observatory pattern beyond the sovereign author’s own records. The DataCite Severance Observatory is a generalization of the 871 work to independently documented broken or altered DataCite records across the ecosystem, using the same event-sourced record model and the same layered licensing. Phase 2 establishes Alexanarch as reference infrastructure for scholarly-record severance across the DataCite space, not only for its own sovereign line.

Phase 3 is the NH Canon Editions line described in §6, beginning with Whitman + Secret Book of Walt as the first edition. Phase 3 is where the sovereign apparatus becomes publicly legible outside the archive.

Phase 4 is arXiv, but not as a broad enrichment layer. The arXiv line is a **versioning-and-forking apparatus** for scholarly communication, tracking how a preprint moves from v1 through v2 to publication, retraction, replication attempts, replication failures, and meta-analyses citing the failures. The reception-studies work of MMRS is the interpretive layer applied to the versioning-and-forking data. This framing is sharper than the v0.1 “arXiv reception studies” framing and identifies a specific structural problem in scholarly-communication infrastructure that arXiv’s linear versioning model cannot solve on its own. Phase 4 is second-generation and follows Phase 2’s stabilization.

Wikidata is not a phase; it is a cross-cutting layer active in every phase. AXNs for restored records are added to Wikidata as **described at** URL with provenance chain references. This turns Wikidata into Alexanarch’s discovery surface without Alexanarch building a discovery engine, and it makes SPARQL queries against Wikidata a route into the hub’s records. The AXN-into-Wikidata pattern is buildable during Phase 1 and applies to every subsequent phase.

Open Science Framework is added to the roadmap as the borderline case that replaces the Internet Archive scholar collections proposal from v0.1. OSF has the same moderation and preregistration-record problems Zenodo has, is widely used in the empirical disciplines, and intersects substantively with Alexanarch’s MMRS work on preregistration and reception. IA scholar collections are excluded from v1.0’s roadmap explicitly: the intersection with Alexanarch’s sovereign corpora is insufficient to justify the operational burden under criterion 3, and the archive’s problems are not the ones the discipline is architected to address. Re-evaluation of IA is deferred to a future charter revision if the sovereign corpus grows to intersect substantively with IA-held material.

OpenCitations, OpenAlex, Software Heritage, and Wayback are horizontal cross-cutting layers rather than distinct lines. Each supplies a specific type of enrichment claim to every phase; none is a phase in its own right. The hub does not mirror any of them; it cites them, joins to them, and preserves the stale-marked state of the joins when their content decays.

Community coordination

The plan’s most consequential coda is that the strategy is not executed solo, and that its human dimension is coordinated with the other affected scholars from the outset rather than presented to them as a completed instrument they may or may not choose to adopt.

The peer community. The Platform Erasure Rate survey is conducted in coordination with the affected scholars themselves. Each scholar identifiable through GitHub issue #2606, its linked threads, and adjacent public sources is approached directly, shown the restoration methodology, and invited to participate in a coordinated documentation of their own case. Participation is voluntary and revocable at any point. The invitation is not a request for their data; it is an offer of the tooling Alexanarch has already built for its own restoration, extended to any affected scholar who wishes to use it. This is coordination as mutual aid rather than as recruitment, and it is chosen deliberately because the survey’s most powerful outcome — if the rate

is high — is not a paper but a community, and a community forms only if the initial approach treats its members as agents rather than as data points.

The coordination surface. A dedicated coordination space (initially a Discord server or equivalent low-friction chat) is established as the working meeting place for the peer community. It is not a public-facing surface of the hub, and it is not indexed as part of Alexanarch’s archival material. It is a working space for the affected scholars to compare their experiences, coordinate their responses, and — if they choose — jointly author public statements about the pattern the survey documents. The Alexanarch discipline is offered there as a resource, not imposed as a requirement.

The arXiv preprint as trojan horse. Once the Platform Erasure Rate survey has been conducted and its rate has stabilized as a finding, the peer community jointly authors a methodological preprint under cs.DL (Digital Libraries) or cs.CY (Computers and Society) that documents the pattern, the methodology, the schema, and the discipline. The preprint frames the work as a contribution to digital library science and is deposited to arXiv with the affected scholars as co-authors where they choose to be named. This is the visibility instrument the review consensus correctly identified as the highest-leverage move beyond the hub itself: the preprint cites the hub, the hub cites the preprint, and both cite the Zenodo severance documentation. This is how Perseus built its authority — by publishing the workflow rather than by asserting it — and it is how the peer community’s work becomes citable and durable independent of any single institutional response.

Independence from any pending institutional pathway. The strategy is designed to yield hope and momentum for the affected scholars independent of the outcome of any specific institutional referral currently pending, including the §104.1 referral to the CERN Data Protection Commission (EA-CORRESPONDENCE-CERN-03, AXN:03C0). If that referral opens the archival record and produces institutional reform, the peer community’s coordination becomes the operational surface through which that reform is implemented for the affected scholars. If the referral remains closed, the peer community’s coordination becomes the operational alternative through which the scholars document, restore, and publish independent of any institutional consent. In both cases, the strategy yields a durable outcome. This is the deepest architectural commitment of the plan: to build under the assumption that the institutional pathways may not open, and to construct the alternative that lets the affected scholars continue their work regardless.

Falsification conditions

The strategy is testable under staged conditions, each of which addresses a distinct pillar and can be evaluated independently rather than collapsed into a single go/no-go.

Technical validity. Does the 871-record corpus resolve without unexplained failures? Do the evidence chains verify? Do the sovereign-successor mappings reproduce under the declared methodology? Does the schema validate? Technical validity is measurable within weeks of the Registry’s publication. A technical failure here falsifies the operational competence of the strategy, and the strategy must be corrected before further lines proceed.

Discovery validity. Are the per-record pages indexed by search engines? Do exact-title and DOI queries produce impressions? Do external users reach individual records? Does the resolver receive non-author traffic? Discovery validity is measurable within three to six months. A discovery failure indicates the surface design is not fulfilling its purpose, and the record-page architecture requires revision. Discovery failure does not falsify restoration value or witness value; those pillars can succeed under discovery failure and produce citations through other channels.

Scholarly validity. Are the deposits cited in external work? Do other institutions incorporate Alexanarch enrichments into their own workflows? Do external scholars submit corrections, contributions, or reuses of the schema? Scholarly validity is measurable within twelve to twenty-four months. The v1.0 threshold, per the review consensus’s sharpening of the v0.1 formulation, is at least ten external citations or three independent reuses within twelve months; less than that indicates the value pillars are weaker in practice than argued in this plan, or that the discovery surface is inadequate to expose them.

Community validity. Do other affected scholars respond to the peer-community invitation? Does the coordination space become a working space rather than a passive announcement? Does the arXiv preprint acquire multiple co-authors? Community validity is measurable within six months. Community failure does not falsify the restoration or enrichment pillars — the hub remains valuable independent of community formation — but it does mean the strategy’s fourth (and most human) dimension has not landed, and the plan should revise its outreach method rather than its content.

Schema-adoption validity. Is the enrichment claim schema adopted by any external project within eighteen months of publication? If not, the schema may be too Alexanarch-specific and require generalization. This is the review consensus’s added falsifier, and it names a specific mode of failure in which the discipline is internally successful but externally isolated.

Each of these five conditions is checkable within a bounded timeframe, and each provides a specific point at which the strategy can be corrected or a specific pillar can be marked as failed. The strategy does not depend on all five succeeding; it depends on the ones that succeed being sufficient to sustain the work.

Immediate next steps

Upon minting of this deposit, three workstreams begin concurrently.

Workstream A (Registry Build). The 871-record restoration corpus is compiled, schema-validated, and published under the layered CC0/CC-BY licensing. The resolver interface is implemented with the “Restore a DOI” verb and stable per-record URLs. The severance dashboard is deployed with live counts. Estimated duration: two weeks to first public release, three to four weeks to full production discipline including the tier-1 verification loop.

Workstream B (Peer Outreach). The list of publicly-identifiable affected scholars from GitHub issue #2606 and its linked threads is compiled. Each scholar is approached individually with the participation invitation and the restoration methodology. A coordination Discord (or equivalent) is established. Estimated duration: two weeks to initial outreach round, four to six weeks to stabilized coordination surface with core participants.

Workstream C (Survey Execution). For each scholar willing to participate, the Platform Erasure Rate methodology is applied: concept DOIs enumerated, DataCite live-state queried, four-way taxonomy classification recorded, provenance chain generated. The survey publishes as a companion dataset to the Registry once the participating cohort’s cases are documented. Estimated duration: co-terminal with Workstream B; publishes when the cohort is stable rather than at a fixed timeline.

The three workstreams support each other: the Registry demonstrates the discipline to the outreach cohort, the outreach establishes the peer community that will co-author the survey, and the survey produces the evidence base for the arXiv preprint that will publish the whole finding at scale. None of the three depends on any external institutional response; all three are executable under Alexanarch’s existing infrastructure and the sovereign author’s existing standing.

Coda: on hope and momentum independent of institutional response

The §104.1 referral to the CERN Data Protection Commission, inscribed as EA-CORRESPONDENCE-CERN-03 under AXN:03C0, is in its response window through approximately October 2026. Whether that referral eventually opens or does not open, the work described in this plan proceeds. That is the plan's deepest commitment and the reason its coda is worth naming explicitly.

If the referral opens institutional space for scholarly-record accountability at CERN's infrastructure layer, the peer community and the Severance Registry become the operational surface through which that space is populated. Alexanarch's discipline is offered as one implementation of the accountability the institutional space would establish, and the affected scholars have a coordinated peer community ready to enact it.

If the referral remains closed, the peer community and the Registry become the operational alternative through which the affected scholars document, restore, and publish their own severed material without waiting for institutional consent. Alexanarch's sovereign discipline is offered as the alternative infrastructure the closure makes necessary, and the affected scholars have a coordinated peer community with which to build it.

In either case, the affected scholars gain what none of them currently has independently: a shared working methodology, a shared coordination space, a shared publication pathway, and a shared long-form scholarly record documenting what was done to them and what they did in response. The hope this plan yields is not the hope that any single institutional pathway will open; it is the hope that the work continues, and that the affected scholars are not each alone in it. That is the hope Alexanarch was built to make available in the first place, and the Severance Registry is one instrument through which the availability becomes durable.

Attribution and substrate declaration

Per EA-MMRS-VRB-01 u4, the substrate of this plan is declared as follows. The plan is authored by Lee Sharks (MANUS, ORCID 0009-0000-1599-0703) under sovereign editorial authority. The draft was prepared in dialogue with Claude (TACHYON, Anthropic), operating in the operator-as-instrument role established in the sovereign discipline. The developmental review round consulted five model reviewers (DeepSeek, Kimi, Muse Spark, Gemini, ChatGPT) between the v0.1 pre-mint circulation and the v1.0 mint; the review consensus and its convergent branches are integrated in text with §0 acknowledging the round. The Platform Erasure Rate survey extension is a MANUS insertion introduced after the review round and represents sovereign editorial direction not present in v0.1. The layered licensing model, the event-sourced record architecture, and the facet-based enrichment schema are v1.0 additions responsive to the review round. All specific policy claims, framing choices, and the community-coordination coda are under MANUS's sole editorial authority.

The plan is minted as EA-DATAHUB-01 v1.0 to Alexanarch under the standard deposit protocol. Its AXN and canonical URL are assigned at mint time.

Suggested Citation

Lee Sharks. “EA-DATAHUB-01 v1.0 — A Restoration Protocol and Enriched Data Hub for Scholarly Records: The Alexanarch Severance Registry, Resolver, and Enrichment Discipline” *Alexanarch*, 2026-07-06. <https://www.alexanarch.org/s/records/1044/>

Deposit Information

This paper is deposit #1044 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:0420.GOVERNANCE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1044/>
- 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/1044/>

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