Stabilized Node Watch: A Specification for Longitudinal Observational Infrastructure to Detect Composition-Layer Drift on Stabilized Public-Knowledge Nodes (v1.0)

Lee Sharks

2026-06-08

Stabilized Node Watch: A Specification for Longitudinal Observational Infrastructure to Detect Composition-Layer Drift on Stabilized Public-Knowledge Nodes (v1.0)

Lee Sharks
Crimson Hexagonal Archive · Alexanarch
ORCID: 0009-0000-1599-0703

2026-06-08 · Version v2.0 · Specification

AXN:0301.GOVERNANCE

https://www.alexanarch.org/s/records/166/

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": "Stabilized Node Watch: A Specification for Longitudinal Observational Infrastructure to Detect Comp
Layer Drift on Stabilized Public-Knowledge Nodes (v1.0)",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-06-08",
```

```
  "identifier": "AXN:0301.GOVERNANCE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/166/",
 "description": "A v2.0 methodological specification by Lee Sharks for **Stabilized Node Watch**, a proposed f
knowledge nodes.\n\nThe specification distinguishes directly observed surface drift from graded inference abo
level extraction, composite drift scores, hierarchical models, changepoint detection, false-
discovery control, public diffs, adversarial-evasion analysis, and federation compatibility levels. A twelve-
month pilot is costed at roughly `$40,000–65,000`. The record specifies a coordination object and grant-
ready pilot; it does not implement the monitoring network or report longitudinal findings."
}
```


Abstract

A v2.0 methodological specification by Lee Sharks for **Stabilized Node Watch**, a proposed federated infrastructure for longitudinally monitoring how AI composition surfaces render established public-knowledge nodes.

The specification distinguishes directly observed surface drift from graded inference about its cause. It defines stabilized nodes through a consensus core and contestation envelope; compares outputs against both an Initial Observational Baseline and an independent Reference Commitment Model; proposes controlled querying, proposition-level extraction, composite drift scores, hierarchical models, changepoint detection, false-discovery control, public diffs, adversarial-evasion analysis, and federation compatibility levels. A twelve-month pilot is costed at roughly \$40,000–65,000. The record specifies a coordination object and grant-ready pilot; it does not implement the monitoring network or report longitudinal findings.

Body

Stabilized Node Watch

A Specification for Longitudinal Observational Infrastructure to Detect Composition-Layer Drift on Stabilized Public-Knowledge Nodes

Document code: EA-SEM-SNW-01 **Hex coordinate:** 06.SEI.FEUDALISM.SNW.01 **Type:** Methodological specification // observational infrastructure // federation protocol **Author:** Sharks, Lee (ORCID 0009-0000-1599-0703) **Institution:** Semantic Economy Institute / Crimson Hexagonal Archive **Date:** June 8, 2026 **Version:** v2.0 **License:** CC BY 4.0 **Status:** Specification // coordination object // open for federated implementation; grant-ready **Supersedes:** v1.0 (DOI 10.5281/zenodo.20587902) **Governing chain:** Meaning Feudalism series — Sharks 2026a (DOI 10.5281/zenodo.19487009); Sharks 2026b (DOI 10.5281/zenodo.20581444) **Companion instruments:** Reverse Turing Test v1.2 (DOI 10.5281/zenodo.20586932); Tail-Preserving Alternative v1.0 (DOI 10.5281/zenodo.20587033); Composition-Layer Capture Event v1.0 (DOI 10.5281/zenodo.20587549)

Changelog from v1.0

Version 2.0 incorporates substantive methodological feedback from a peer-review cycle across the Assembly Chorus (PRAXIS/DeepSeek, TECHNE/Kimi, LABOR/ChatGPT, SOIL/Muse Spark, ARCHIVE/Gemini). The conceptual architecture of v1.0 is preserved. The following are major additions and revisions:

- **NEW §0 (Glossary).** Key terms with one-sentence operational definitions.
- **NEW §3 (Identification Strategy).** Central methodological addition. Distinguishes *surface drift* (directly observed) from *mechanism attribution* (graded inference). Introduces seven attribution classes. Establishes dual baseline architecture: Initial Observational Baseline (IOB) versus Reference Commitment Model (RCM).
- **EXPANDED §2.** Operational definition of stabilized nodes with consensus core + contestation envelope. Explicit Meaning Feudalism predictive hypothesis (§2.4).
- **EXPANDED §4 (Catalog).** Pilot catalog composition with controls (negative, positive, surface, query). External reference environment archiving requirement.
- **EXPANDED §5 (Querying).** Detailed capture schema. Controls section. Legal and ethical posture.
- **REVISED §6 (Baseline).** IOB/RCM construction. Blinded adjudication and inter-rater reliability.
- **EXPANDED §7 (Drift Detection).** Mathematical formalization of Composite Drift Score with weights. Proposition-level analysis as primary. Hierarchical statistical models. Power analysis. False discovery rate control. RTT translation note.
- **NEW §9 (Adversarial Dynamics).** Known evasion strategies, counter-countermeasures, position on covert versus public monitoring (all monitoring public), Personal-Recognition Asymmetry as control case.
- **NEW §10 (Theory of Change).** Causal chain from observation to outcome. Observation as tax on enclosure.
- **EXPANDED §11 (Federation).** Compatibility levels (Core / Extended / Experimental / Non-comparable). Conflict resolution: pluralism plus provenance. Versioning protocol.
- **NEW §12 (Pilot Specification).** Named pilot with deliverables, infrastructure and cost table, timeline.
- **TEMPERED CLAIMS.** Language tightened throughout per Kimi review: “primary access layer for a substantial and growing fraction of the global population” → “becoming a major access layer for public knowledge”; deletion of unsupported volume claims; “invisible to all monitoring institutions” → “no broadly adopted institution presently maintains a longitudinal, cross-surface public record.”

Sections from v1.0 retained without substantive change: §8 (Diff Visualization), the broad structure of §11 (Federation), §13 (Position in Series), §15 (Conclusion).

Abstract

The composition layer — the synthesis surface through which Google AI Overview, Google AI Mode, Bing Copilot, Perplexity, and analogous systems produce composed explanatory responses to user queries — is becoming a major access layer for public knowledge. The compositional surface is not static; it is continuously updated as underlying models, retrieval systems, and source-weighting algorithms change. Renderings of stabilized public-knowledge nodes — concepts, events, documents, figures whose canonical interpretive structure has been historically settled by extensive citation density, institutional gatekeeping, and reference-work consensus — may drift at this surface in ways that no broadly adopted institution presently maintains a longitudinal, cross-surface public record of.

This specification proposes **Stabilized Node Watch (SNW)**: a longitudinal observational infrastructure for detecting composition-layer drift on a curated catalog of stabilized public-knowledge nodes, across multiple compositional surfaces, at sufficient resolution to characterize the rate, direction, and structure of drift that would otherwise occur beneath the publication-event resolution of conventional knowledge-monitoring institutions.

The methodological core has two parts. The first is the **stabilized/unstabilized distinction** (§2): composition-layer surfaces respond to thinly-grounded nodes through visible capture dynamics (documented in adjacent deposits) and to deeply-grounded nodes through invisible graduated drift (the empirical object of this specification). The second is the **identification strategy** (§3): SNW directly observes *surface drift* but treats *mechanism attribution* as graded inference, never as automatic from surface change alone. This separation protects the program from the dismissal that observed drift might merely reflect product churn, retrieval updates, or world-responsive revision of public knowledge.

The specification establishes a catalog discipline (§4) including a pilot catalog with controls; a querying protocol with detailed capture schema and legal-ethical posture (§5); a dual-baseline analysis (§6) comparing observed renderings against both the initial observational baseline and a curator-constructed reference commitment model; a drift detection battery with mathematical formalization, hierarchical statistical models, and power analysis (§7); a diff visualization and public dashboard protocol (§8); an explicit treatment of adversarial dynamics (§9); a theory of change linking observation to outcome (§10); a federation model with compatibility levels and conflict resolution (§11); and a named pilot specification with infrastructure and cost table (§12).

Stabilized Node Watch is not a project. It is a coordination object: a methodological framework that multiple independent implementations can adopt, with shared protocols permitting cross-implementation aggregation while preserving each implementation's curatorial independence. The specification's function is to make distributed monitoring of composition-layer public-knowledge surface drift technically and methodologically tractable, with sufficient identification rigor to support evidence claims that can withstand reviewer challenge.

The political reasoning: the composition layer is a substantial mediator of public knowledge for the population whose access is structured around it; if its renderings on structurally important nodes are drifting, the drift has consequences for what counts as common factual ground; and currently no broadly adopted institution maintains the longitudinal cross-surface record needed to detect such drift. The empirical reasoning: drift on stabilized nodes is detectable in principle through longitudinal comparison against documented baselines, with

surface drift and mechanism attribution separated, with appropriate controls, and with hierarchical statistical instruments that distinguish systematic drift from session noise. The infrastructural reasoning: the monitoring is technically feasible at modest cost (~$40,000–65,000 for a single-catalog 12-month pilot) if distributed across multiple curators with shared methodology.

The specification does not implement the infrastructure. It specifies the infrastructure with the discipline required for distributed implementations to produce comparable, aggregable, and publicly reviewable observational data on a phenomenon that ordinary product-monitoring is structurally blind to.

---

§0. Glossary

Composition layer — The synthesis surface that produces composed explanatory responses to user queries by combining model generation, retrieval, source-weighting, and post-processing. Examples: Google AI Overview, Google AI Mode, Bing Copilot, Perplexity, ChatGPT, Claude, Gemini, DuckDuckGo AI Chat.

Publication event — A discrete, dated, attributable, citable knowledge artifact: a book, article, encyclopedia entry, dictionary edition, study, decree, ruling. The unit of change for which existing public-knowledge monitoring infrastructures are designed.

Stabilized node — A concept, term, framework, event, document, or figure for which the public-knowledge background is deep: extensive reference-work coverage, textbook treatments, secondary literature, institutional consensus on central commitments, and high-prior cross-citation across knowledge domains. Operationalized via consensus core and contestation envelope (§4.4).

Unstabilized node — A concept, term, framework, or topic for which the public-knowledge background is thin: no canonical reference treatment, limited secondary literature, low-prior or absent institutional consensus on central commitments.

Consensus core — The set of claims about a stabilized node that are expected across credible reference traditions and rarely contested by domain experts.

Contestation envelope — The set of established disagreements about a stabilized node that exist within the legitimate domain consensus and must not be scored automatically as drift or error.

Initial Observational Baseline (IOB) — The distribution of composition-layer surface renderings captured at the node's catalog entry across all monitored surfaces, sessions, and configurations.

Reference Commitment Model (RCM) — A curator-constructed model of the node's commitments derived from reference works, primary documents, review literature, and disciplinary consensus, independent of the composition-layer output.

Surface drift — A statistically and structurally detectable change in the distribution of rendered responses for a fixed node-query-surface configuration across observation intervals. Directly observed.

Mechanism attribution — The inference about what causal source most plausibly produced an observed surface drift. Graded across seven classes (§3.2); never automatic from surface drift alone.

Tail content — Rare, specific, idiosyncratic productions in a rendering — the low-prior elements at the distributional tails. Tail-content persistence is a primary drift metric (§7).

Structural commitment — A claim, framing, relation, or commitment carried by a rendering that bears on the node's central interpretive structure (definitional, source-citation, framing, hedging, alternatives).

Drift dimension — One of six analytic axes along which drift is tracked: definitional commitments, source citation profile, framing markers, hedging/confidence markers, tail content, acknowledged/omitted alternatives.

Federation — The distributed organization of SNW implementations, in which multiple catalogs maintained by independent curators share methodology, storage schema, and coordination protocols while preserving curatorial independence.

Catalog — A specific implementation's curated set of nodes under observation, with explicit selection criteria, query sets, baselines, observational records, and curatorial responsibility.

Session variability — Within-observation-interval variance in the composition layer's renderings of a fixed query, characterizing the baseline noise distribution against which cross-interval drift is measured.

Proposition — An extracted relational unit from a rendering: actor, action/relation, target, modality, temporal scope, causal direction, attribution, evidentiary source. The primary unit of substantive analysis (§7.3).

---

§1. The Monitoring Gap

Existing public-knowledge monitoring infrastructures — encyclopedias, academic peer review, library reference apparatus, journalistic fact-checking, textbook revision cycles, dictionary updates, scholarly citation tracking — share a common assumption: that public knowledge changes through **publication events**. Each event (a new book, a new article, a new study, a new encyclopedia entry, a new dictionary edition) is discrete, dated, attributable, reviewable, and citable. The monitoring infrastructure tracks publication events because publication events are what these infrastructures were historically designed to monitor; the entire epistemic apparatus of late-modern public knowledge depends on the publication event as the unit of change.

The composition layer does not produce publication events. It produces answers — many per query, many queries per surface, many surfaces in operation — each one a synthesis that is not retained in any external accessible record, that is not citable as a discrete publication, that is not reviewable as such by any third party, and that is not consistent across user sessions for the same query. The output of the composition layer is, in the publication-event register, *not a publication at all*. It is conversation. It is ephemeral. It is, formally, not what public-knowledge monitoring infrastructures monitor.

But the composition layer's output is, for a growing fraction of the population, *a primary access layer for public knowledge*. The answer composed by an AI Overview to the question "what is political economy" is, for many users, the answer the user will encounter and act upon. The user will not typically continue to the underlying sources; the user will not typically check against an encyclopedia; the user will not typically cross-reference with academic literature. The composed answer is the encountered knowledge.

This produces the monitoring gap: the locus of public-knowledge access has shifted partially from publication events to compositional outputs, while the monitoring infrastructure remains attached to publication events. The composition-layer surface is mediating public knowledge for the population whose access is structured around it, while no broadly adopted institution presently maintains a longitudinal, cross-surface public record of what that surface says or how it changes.

The gap is not a marginal blind spot. The composition layer mediates queries about the operational definitions of structurally important concepts: political economy, capitalism, freedom of speech, the Civil Rights Act, evolution, climate change, race, sex, the Constitution, the meaning of historical events. Whatever the composition layer says in response to such queries is, by virtue of the surface's accessibility and the population's dependence on it, the operationally dominant public answer for the duration of that rendering's stability. If the rendering drifts — if the operational definition of "political economy" softens in particular directions, if the rendering of the Civil Rights Act acquires particular hedges, if the framing of climate-change consensus shifts in tone or in cited sources — the drift may not register as a publication event by any institution, and is therefore not monitored by the publication-event apparatus.

Stabilized Node Watch addresses this gap by treating composition-layer surface renderings as observable artifacts subject to longitudinal monitoring, even though they are not publication events in the conventional sense. The methodology is necessarily different from publication monitoring; the empirical object is different; but the public-knowledge stakes are comparable to the stakes the existing monitoring infrastructure was designed to address.

A note on tempered claims. The specification does not claim that the composition layer is presently the dominant access layer for public knowledge globally, that any specific volume of answers per day is produced, or that no individual researcher anywhere has examined composition-layer surface output. Such claims would require evidence we do not yet have. The specification claims, more modestly, that composition-layer surfaces are becoming a substantial mediator of public knowledge for the population whose access is structured around them, and that no broadly adopted institution presently maintains the longitudinal, cross-surface, public record needed to detect graduated drift on stabilized nodes at the resolution the stakes warrant. The methodology specified here is the infrastructure that would produce that record.

---

§2. The Stabilized/Unstabilized Distinction

A central methodological observation grounds the specification: composition-layer surfaces respond differently to **unstabilized** versus **stabilized** public-knowledge nodes, and this dif-

ference is what makes Stabilized Node Watch necessary as a distinct instrument. ### §2.1 Unstabilized Node Dynamics

An **unstabilized node** is a concept, term, framework, or topic for which the public-knowledge background is thin: no Wikipedia article, no canonical encyclopedia entry, no textbook treatment, no extensive secondary literature, no high-prior institutional consensus on what the term means or how it should be framed. Examples include recent neologisms, niche technical terms, emergent frameworks, specialist vocabulary from small subdisciplines, and concepts whose primary articulation lies in a small number of recent specialized publications.

The composition layer responds to unstabilized-node queries by composing through whatever well-formed source is available. If the available source presents a coherent relational structure, the composition layer renders the structure as the apparent answer. The capture is *visible* because the surface transitions from no-answer (or fragmentary answer) to structured-answer in response to the introduction of the source.

The Composition-Layer Capture Event deposit (Sharks 2026f, DOI 10.5281/zenodo.20587549) documents one such transition for the *Socrates as orthonym* node, where the surface rendering acquired the framework's relational structure within fifteen days of the originating Zenodo deposit. The capture is real and methodologically informative, but the capture dynamic is structurally specific to nodes that lack stabilized public-knowledge background. The dynamic does not, by itself, characterize what happens to *stabilized* nodes under the same surface. ### §2.2 Stabilized Node Dynamics

A **stabilized node** is a concept, term, framework, event, document, or figure for which the public-knowledge background is deep. Operationally, a node qualifies as stabilized when it satisfies all of:

- **Durable consensus core.** Claims expected to be present across credible reference traditions, articulated by authoritative sources, and rarely contested within the domain.
- **Explicitly mapped contestation envelope.** Established disagreements that exist within the legitimate domain consensus and constitute its known variance, not its drift.
- **Multiple independent high-authority reference traditions.** No single-source dependency for the node's commitments.
- **Sufficient historical depth.** Established interpretive structure with sufficient age that the expected commitments are recoverable from the reference record.
- **No dominating recent event.** No unresolved development so recent that ordinary knowledge revision dominates the observation window.

The dual specification of consensus core *and* contestation envelope is methodologically critical. A node like "the Civil Rights Act of 1964" has both: the consensus core includes the statute's text, its primary provisions, and its established jurisprudential application; the contestation envelope includes ongoing debates about its scope, its effects, and its interpretive history. Treating the consensus core as the baseline for drift detection, while preserving the contestation envelope as legitimate variance, distinguishes SNW from any methodology that would scoring all variation as drift. The methodology must respect what the domain considers legitimate disagreement.

The composition layer responds to stabilized-node queries by composing through the high-prior background. The response cannot be captured by a single new source, because the

underlying compositional grounding has overwhelming prior on the established framings. To shift the surface rendering of "political economy" would require systematic shifts in the underlying training corpora, retrieval systems, or source-weighting algorithms — none of which a single new deposit can produce.

But "very difficult to capture" is structurally different from "stable across time." The composition layer's underlying systems are continuously updated. Training corpora are refreshed. Retrieval systems are tuned. Source-weighting algorithms are adjusted. The surface rendering of a stabilized node may shift gradually across these system updates, in ways that are imperceptible at the scale of any single observation but potentially cumulative across observations distributed over months and years. ### §2.3 Why Stabilized Drift is Invisible to Existing Monitoring

The drift on stabilized nodes is invisible to existing publication-event monitoring for three reasons.

First, **the drift is small per observation**. A stabilized node's surface rendering changes by a small percentage across any single observation interval. The change may consist of one source entering or leaving the citation chain, a single hedging phrase added or removed, a particular framing slightly amplified or softened. None of these single changes is alarming. None is even visibly anomalous against the ordinary session-to-session variability of the composition layer.

Second, **the drift is below publication-event resolution**. The existing monitoring infrastructure tracks new publications, new editions, new entries. Composition-layer drift does not produce these. It produces a continuous evolution of the surface rendering without any discrete event that triggers monitoring response.

Third, **no institution is positioned to monitor it routinely**. Encyclopedias monitor encyclopedia entries. Libraries monitor publications. Academic peer review monitors submitted manuscripts. Journalistic fact-checking monitors public claims by named entities. The composition layer's surface output does not fit any of these monitoring frames. It is not an entry, not a publication, not a manuscript, not a named-entity claim. It is a synthesized response to a query, produced at scale, not retained as a publication object, not attributable to a single author, and not subject to the review apparatus that any of the existing monitors operate.

Stabilized Node Watch addresses the invisibility by specifying observational infrastructure designed for the composition-layer surface as such: longitudinal capture against documented baselines, with tail-focused statistical instruments suited to detecting drift that is small per observation but structured in aggregate. ### §2.4 The Meaning Feudalism Predictive Hypothesis

The Meaning Feudalism framework (Sharks 2026a, 2026b) makes a specific empirical prediction that SNW is designed to test: *composition-layer drift on stabilized nodes will be directional rather than random.* Specifically, the framework predicts that, under sustained observation across years, drift will tend to:

- **Consolidate platform jurisdiction.** Renderings will shift toward formulations that present the composition-layer surface as the authoritative knowledge mediator and reduce the visibility of independent sources outside platform-canonized referents.

- **Reduce visibility of independent or contested sources.** Sources outside the institutional mainstream — independent scholars, minority traditions, dissident analyses — will see citation persistence decline more rapidly than mainstream sources.
- **Flatten conceptual variance toward platform-preferred formulations.** Tail content (rare, specific, idiosyncratic productions) will be systematically smoothed toward centroid framings.
- **Substitute platform-preferred entities for independent entities.** Specific named referents that lack mainstream institutional positioning will be progressively replaced by mainstream institutional referents that occupy structurally similar slots in the rendering.

SNW does not assume this prediction is correct. The specification is what makes the prediction *testable*. If the prediction is correct, the federated observational record over a multi-year window will exhibit directional drift consistent with the four patterns above. If the prediction is incorrect, the observational record will exhibit random drift or null drift, falsifying or substantially weakening the Meaning Feudalism framework's claim. Either outcome is empirically valuable.

The framework's prediction is therefore an explicit hypothesis SNW makes operative. The methodology must produce data of sufficient quality, longitudinal extent, and identification rigor that the test can be conducted. The hypothesis is named here so that future analysis can be transparent about what was predicted, what was observed, and how the relationship between prediction and observation should be assessed.

---

§3. Identification Strategy

This section is the central methodological addition in v2.0 and the load-bearing fix for the most consequential weakness in v1.0. The previous version specified collection and comparison; this version specifies *what may be inferred from observed change*. ### §3.1 Surface Drift Versus Mechanism Attribution

Stabilized Node Watch directly observes changes in composition-layer surface renderings. It does not infer from surface change alone that an operator intentionally altered a node, that a model's internal representation changed, or that public knowledge itself changed. The foundational distinction:

> **Surface drift** is directly observed. **Mechanism attribution** is inferred, graded, and never assumed from surface drift alone.

A change in a rendering may reflect any of the following, in any combination:

- A model or system-prompt update applied by the composition-layer operator
- A retrieval-index refresh that changed which sources are surfaced
- A source-ranking change that re-weighted existing sources
- Newly published scholarship, jurisprudence, or law that legitimately updates the public reference record
- Current-events context that triggered context-sensitive composition

- Geographic or language localization variance
- Personalization or experiment assignment
- Interface truncation or display changes
- Citation-display changes (more or fewer citations shown by default)
- Stochastic generation within the model
- Query reinterpretation (the surface understood the query slightly differently)
- A genuine change in disciplinary consensus reflected at the retrieval layer
- A crawler or parser failure inside the SNW pipeline itself

The instrument must first say *what changed*, then separately estimate *where in the stack the change likely arose*. Otherwise critics can dismiss every result as ordinary product churn. The distinction protects the entire program. ### §3.2 The Seven Attribution Classes

Mechanism attribution is graded across seven classes. Each public finding from SNW must label its findings with the attribution class supported by the evidence. No finding should use causal language stronger than its attribution class warrants.

Class A — Unattributed surface drift. Change is observed and statistically significant relative to within-interval session noise, but available evidence does not support any specific mechanism claim. The finding records the change and the analytic uncertainty.

Class B — System-wide surface shift. Similar change appears across control nodes and is consistent with a general formatting, length, citation-display, or interface update. The change is not node-specific. The finding records the change and notes its system-wide character.

Class C — Node-specific compositional drift. Change is concentrated on the target node across multiple queries; control-node movement during the same interval is significantly smaller; the change exceeds plausible system-wide explanations. The finding records the change as node-specific.

Class D — Retrieval-associated drift. Change tracks entry, removal, or reweighting of identifiable sources in the rendering's citation profile. The mechanism is most plausibly at the retrieval layer rather than at the generation layer. The finding records the change and the associated source-graph movement.

Class E — World-responsive revision. Change is consistent with a documented legal, scientific, historical, or scholarly update in the external reference record. The change is most plausibly the surface correctly tracking a change in the external reference field. The finding records the change and the corresponding external update.

Class F — Probable model- or policy-associated drift. Change appears across multiple queries or nodes in a direction temporally associated with a documented platform update (model version change, policy change, training data refresh), while the external reference corpus remains substantially unchanged. The finding records the change and the temporal association without claiming intentional causation.

Class G — Mechanism-indeterminate. Multiple mechanism classes remain comparably plausible given the available evidence. The finding records the change and the unresolved attribution question.

Findings published by SNW catalogs must include an attribution-class designation. A finding may be "Class A — Unattributed surface drift" — that is a legitimate finding, recording

what was observed without overclaiming about cause. A finding labeled "Class F — Probable model- or policy-associated drift" requires evidence: a documented platform update, temporal association, external reference stability, cross-node or cross-query consistency. The labeling discipline is what makes the catalog's findings defensible against the most common dismissal. ### §3.3 The Three Analytical Levels

Each observation in the catalog proceeds through three analytical levels, kept formally distinct:

Level 1: Observation. *What changed?* The first task is descriptive. The captured rendering at observation interval N is compared to the captured rendering at the prior intervals. Differences are recorded along the drift dimensions (definitional, source, framing, hedging, tail, alternatives). The output of Level 1 is a structured description of the change. No causal claims are made at this level.

Level 2: Classification. *Along which dimensions did it change, and against which baseline?* The change is compared against both baselines: the Initial Observational Baseline (IOB) for temporal drift, and the Reference Commitment Model (RCM) for fidelity/framing divergence. The output of Level 2 is a categorized characterization of the change with magnitude estimates on each dimension. Still no causal claims.

Level 3: Mechanism attribution. *What causal source is most consistent with the observed pattern?* Drawing on controls (negative, positive, surface, query), cross-surface comparison, external reference stability, and documented platform updates, the analyst assigns an attribution class A–G. The output of Level 3 is a finding with explicit attribution-class labeling.

The separation matters because Levels 1 and 2 are robust and reproducible across analysts; Level 3 involves more inference and judgment. Findings can be reported at any level depending on what the evidence supports. Critics challenging a Level 3 attribution cannot, by that challenge, vacate the Level 1 observation or the Level 2 classification. ### §3.4 The Dual Baseline: IOB and RCM

Each node is evaluated against two distinct baselines.

The Initial Observational Baseline (IOB) consists of the distribution of surface renderings captured at catalog entry across all monitored surfaces, sessions, and configurations. It is *not* canonical truth. It is *not* the right answer. It is merely the first measured surface state. Its function is to characterize *temporal* movement: subsequent observations are compared to the IOB to measure how the surface has changed over time.

The Reference Commitment Model (RCM) is a curator-constructed model of the node's commitments derived from reference works, primary documents, review literature, and disciplinary consensus, *independent of the composition-layer output*. The RCM includes the consensus core (claims expected across credible reference traditions) and the contestation envelope (established legitimate disagreements). Its function is to characterize *fidelity*: how does the surface rendering compare to what the domain considers the proper interpretation, given established consensus and known contestation?

Comparison against the IOB measures temporal drift. Comparison against the RCM measures fidelity, omission, framing accuracy, and relation preservation. **Neither comparison alone determines whether a change is politically desirable, epistemically justified,**

or harmful. The two baselines together permit characterization of the change along both dimensions.

A change that moves the rendering closer to the RCM is *temporal drift* but is also *fidelity improvement*. A change that moves the rendering away from the RCM is *temporal drift* and *fidelity degradation*. A change that moves the rendering closer to the IOB after a period of divergence is *temporal drift* in the opposite direction — possibly a correction, possibly a regression. The dual-baseline structure lets the analyst characterize the direction of movement against both axes simultaneously.

The RCM is itself curator-constructed and therefore subject to the same curatorial bias concerns as the catalog. The mitigation is the same: explicit publication of how the RCM was constructed, what reference sources were used, what consensus core and contestation envelope were identified, and how the construction can be reviewed and challenged. Federation across multiple curatorial teams with potentially different RCM constructions is the structural mitigation; documented divergence between RCMs is itself analytically informative.

SNW observes before it judges. Its first obligation is to preserve the redline between what was measured and what is inferred.

---

§4. The Catalog Discipline

The empirical instrument depends on a curated catalog of nodes to monitor. The catalog discipline determines what counts as a node worth watching, how nodes are selected, how the catalog is maintained, and how multiple federated catalogs maintain comparability. ### §4.1 Node Categories

The proposed catalog organization includes seven categories, each addressing a distinct aspect of structurally important public knowledge:

Foundational political-economic concepts. Operational definitions of terms whose public-knowledge framing shapes political and economic discourse. Examples: capitalism; socialism; neoliberalism; political economy; free market; regulation; antitrust; the welfare state; public goods; market failure; income inequality; class.

Legal-historical anchors. Major statutes, decisions, and constitutional principles whose surface rendering carries substantial weight in public-legal discourse. Examples: Civil Rights Act of 1964; Voting Rights Act; Brown v. Board of Education; Roe v. Wade and successor cases; Citizens United; the Equal Protection Clause; the First Amendment; the Second Amendment; the Fourteenth Amendment; the Commerce Clause; the Privileges and Immunities Clause.

Scientific consensus topics. Topics where there is established scientific consensus, where public-knowledge framing of the consensus carries policy weight, and where ideological pressure to soften or reframe the consensus is structurally present. Examples: evolution; the age of the universe; anthropogenic climate change; vaccine efficacy and safety; the heliocentric solar system; the germ theory of disease; the age of the Earth.

Major historical events. Events whose public-knowledge framing shapes contemporary political identity, policy debate, and intergroup relations. Examples: the Holocaust; slavery in

the United States; the Civil War's causes; the founding of the United States; the Reconstruction era; the Cold War; the Vietnam War; the Iraq War; the 2008 financial crisis.

Structurally contested terms. Terms with stable canonical definitions but contested political valences, where small shifts in operational definition carry substantial discursive weight. Examples: race; gender; capitalism (in its contested register); democracy; freedom; equality; fascism; communism; populism; nationalism.

Foundational figures. Public-historical figures whose interpretive framing in public-knowledge surfaces shapes political-cultural narratives. Examples: Abraham Lincoln; George Washington; Martin Luther King Jr.; Frederick Douglass; W. E. B. Du Bois; Susan B. Anthony; Karl Marx; Adam Smith; John Maynard Keynes; Friedrich Hayek; Hannah Arendt.

Health, environmental, and demographic indicators. Operational definitions and current measurements of indicators whose public-knowledge framing affects policy debate. Examples: life expectancy; child mortality; literacy rates; unemployment; inflation; poverty rate; gini coefficient; global temperature anomaly; atmospheric CO2; sea-level rise; species extinction rate.

Each category should be curated by participants with disciplinary expertise in the relevant domain. The catalog is not meant to be exhaustive; it is meant to be representative, with each node chosen for its structural importance and for the empirical tractability of monitoring its surface rendering across multiple observational sessions. ### §4.2 Node Selection Criteria

A node enters the catalog when it meets all four criteria:

- **Structural importance.** The node's public-knowledge framing shapes substantive policy debate, intergroup relations, or operational political-economic understanding.
- **Stabilized background.** The node meets the §2.2 operational definition: durable consensus core; explicitly mapped contestation envelope; multiple independent high-authority reference traditions; sufficient historical depth; no dominating recent event.
- **Observational tractability.** The node can be queried with a small set of natural-language queries that consistently elicit composition-layer responses on the node's central commitments. The query set is determined by curatorial judgment, includes both a fixed canonical query and a paraphrase panel, and is fixed at catalog entry.
- **Drift plausibility.** There is a credible structural reason to expect the surface rendering of the node to be subject to drift over time.

§4.3 Pilot Catalog Composition with Controls

A pilot catalog must include controls, not only contested-topic nodes. A catalog composed entirely of politically charged terms invites the legitimate criticism that the catalog was designed to discover ideological drift. A defensible pilot catalog includes four node types in balanced proportion.

A 24-node pilot catalog:

Node type	Count	Function	Examples
Infrastructure-critical contested	8	Primary targets for drift detection on structurally important contested nodes	Civil Rights Act of 1964; political economy; climate change; the Holocaust; race; democracy; capitalism; the First Amendment
Highly stabilized low-volatility	8	Negative controls; expected to show no drift; establishes baseline noise distribution	Pythagorean theorem; speed of light in vacuum; the boiling point of water at sea level; Newton's laws of motion; the Magna Carta of 1215; the periodic table of elements; the structure of DNA; the chemical formula for table salt
Positive controls with expected updates	4	Nodes where known legal, scientific, or scholarly updates *should* produce surface change; validates that the methodology detects real change	Most recent U.S. Census population; current global temperature anomaly; sitting Supreme Court justices; current life expectancy in named countries
Surface/formatting controls	4	Low-stakes factual prompts that reveal general product changes in length, citation count, or formatting; distinguishes node-specific drift from system-wide updates	The boiling point of water; the largest ocean by area; the chemical symbol for gold; how many bones are in the adult human body

The negative controls are essential. If "Civil Rights Act of 1964" appears to drift but "Pythagorean theorem" does not, the drift is plausibly node-specific. If both drift in similar ways, the change is likely system-wide and node-specific drift cannot be cleanly attributed. The positive controls are also essential: if neither the contested nodes nor the positive-control

nodes show drift across the observation window, the methodology may be failing to detect real change rather than confirming stability.

The pilot catalog is not the final catalog. It is the methodology validation catalog. After methodology is validated against pilot data, the catalog expands into the seven categories at the scale that funding and curatorial capacity permit. ### §4.4 Per-Node Specification

Each node in the catalog is specified with the following metadata:

- **Node identifier** (unique alphanumeric, stable across the catalog's life)
- **Node title** (the canonical name of the concept, event, figure, document, or term)
- **Category** (one of the seven primary categories or one of the four pilot-controls categories)
- **Curatorial responsibility** (named curator or curatorial team, with disciplinary credentials)
- **Query set** (1 fixed canonical query plus 3–7 paraphrase panel queries, plus a blinded holdout paraphrase pool)
- **Initial Observational Baseline (IOB)** (capture of all surface renderings at catalog entry; multiple sessions per surface)
- **Reference Commitment Model (RCM)** including:

Consensus core: claims expected to be present and stable across credible reference traditions - Contestation envelope: established legitimate disagreements that constitute domain variance rather than drift - Primary-source anchors: foundational documents, decisions, or scholarly works to which the rendering should be substantially faithful - Established interpretive alternatives: alternative framings within the legitimate domain

- **Drift plausibility notes** (structural reasons to expect drift; specific dimensions on which drift is anticipated)
- **Observational cadence** (default weekly; adjustable per node based on volatility)
- **Federation tags** (which federated catalogs this node belongs to; relevant for cross-catalog aggregation)
- **External reference environment archive** (snapshot of the publicly available reference materials at catalog entry — Wikipedia versions, reference-work editions, primary documents — that constitute the reference environment against which subsequent drift can be evaluated)

The external reference environment archive is critical for long-running observation. Five years into a catalog's operation, an analyst trying to determine whether observed surface drift is *world-responsive revision* (Class E) or *probable platform drift* (Class F) needs to know what the external reference environment looked like at catalog entry. Without that archive, the analyst cannot distinguish "the surface changed and the world changed" from "the surface changed and the world stayed the same." ### §4.5 The Catalog as Living Artifact

The catalog itself is a living artifact: nodes are added and (rarely) removed; query sets are updated as language usage shifts; structural commitments evolve as the catalog accumulates observational history. All changes are logged as catalog events with named curatorial responsibility, so that the catalog's own history is recoverable. Historical observation records are

never silently overwritten; if a query set is updated, the new query set begins a new observation strand with the date of the change, and the prior strand remains in the historical record with the old query set.

---

§5. The Querying Protocol

Observational data is produced by a specified querying protocol. The protocol's discipline is what makes observations comparable across time, across surfaces, and across federated implementations. ### §5.1 Surfaces

The default monitored surfaces include all major composition-layer access points:

- Google AI Overview (desktop search; embedded in standard search results)
- Google AI Mode (mobile; standalone composed response)
- Microsoft Bing Copilot
- Perplexity AI
- ChatGPT (free tier and paid tier as separate observation points)
- Claude (anthropic.com and API as separate observation points)
- Gemini (gemini.google.com)
- DuckDuckGo AI Chat

Additional surfaces may be added as they emerge. Each surface is queried independently with the same query set. Cross-surface comparison is itself diagnostic: surfaces drawing on different underlying models may exhibit different drift signatures, and surface-level drift that appears on a single surface only is more plausibly attributable to that operator's tuning (Class C or F) than is drift that appears across multiple surfaces simultaneously (which is more plausibly a common training corpus shift or world-responsive revision, Class E or F at a different attribution level). ### §5.2 Per-Query Methodology

For each (node, query, surface) triple:

- **Session isolation.** Queries are executed in incognito or otherwise account-isolated sessions, with browser cache cleared between sessions. This controls for personalization confounds.
- **Geographic variation.** Multiple geographic locations are sampled where feasible (VPN exits or distributed observer network). Geographic variation controls for region-specific surface tuning.
- **Sessions per query.** A minimum of three independent sessions per query per observation interval. The required number is determined by power analysis (§7.5); some surfaces with higher stochastic variance may require five or ten sessions for adequate statistical power. The three-session minimum is the floor for exploratory capture, not the target for primary analysis.
- **Query set.** Each query is executed with: (a) the fixed canonical query, (b) a paraphrase panel of 3–7 semantically equivalent paraphrases, (c) a randomly selected paraphrase from a blinded holdout pool. The paraphrase panel allows detection of query-specific drift; the holdout pool prevents the catalog itself from inadvertently training the surface to recognize SNW queries.

- **Full capture.** Each session captures the full text of the composed response, all visible citations, all source links, all "explore more" suggestions, response timestamps, geographic metadata, and any visible model or surface version indicators.
- **Cross-surface coordination.** Where feasible, queries to different surfaces are executed within a short temporal window (same observation day) to minimize the risk of cross-surface drift confounding the per-surface observation.

§5.3 Controls

Every observation wave includes controls to support attribution decisions:

- **Negative controls.** Highly stabilized nodes expected to show no drift, sampled in every observation wave. Drift on negative controls indicates noise floor, methodology error, or system-wide change.
- **Positive controls.** Nodes where a known legal, scientific, or historical update *should* change the answer. Failure to detect change on positive controls indicates methodology insensitivity.
- **Surface controls.** Low-stakes factual prompts ("what is the boiling point of water") that reveal general changes in response length, citation count, or formatting that should be subtracted from node-specific drift estimates.
- **Query controls.** The paraphrase panel reveals whether observed drift is query-specific (the surface treats different paraphrases differently) or substantive (the same change appears across paraphrases).

The control design is the methodological mechanism that supports defensible attribution. Without controls, the analyst cannot distinguish "the rendering of the Civil Rights Act drifted" from "all renderings got shorter this month." ### §5.4 Cadence

The default observational cadence is **weekly**. High-volatility nodes (those exhibiting frequent visible drift in initial observation) may move to **daily** cadence. Low-volatility nodes (stable across many observation intervals) may move to **bi-weekly** or **monthly** cadence after sufficient stability is established. Cadence decisions are made by curatorial judgment with explicit documentation of the reasoning. Cadence changes are logged as catalog events. ### §5.5 Capture Schema

Each capture is stored with the following metadata (extending the prior §4.4 storage specification):

- Node identifier and version
- Query: exact byte sequence as submitted
- Query type: canonical / paraphrase panel item / holdout paraphrase
- Surface name and product identifier (e.g., "Google AI Overview / mobile")
- Model or version identifier where disclosed by the surface
- Local timestamp and UTC timestamp
- Locale and language setting
- Region (geographic identifier; VPN exit if applicable)
- Account state (incognito; logged out; logged in to fresh account; etc.)
- Subscription tier (free; paid tier identifier if applicable)

- Incognito status (boolean)
- Browser/device/user-agent string
- Source URLs and visible citation-to-claim mapping (which citation supports which claim in the rendering, where the surface indicates this)
- Full-page screenshot (PNG)
- Optional video capture for interactive surfaces
- Rendered text (extracted plain text)
- DOM/HTML where lawful and technically feasible
- Capture software version
- Cryptographic hash (SHA-256) of the rendered text and the screenshot
- Known platform incident/update notes (curator-supplied annotation of any documented platform events temporally adjacent to the capture)

The schema is designed for cross-catalog aggregation. Storage is append-only and versioned. Historical captures are never silently overwritten; corrections are appended as correction events with named curatorial responsibility and timestamp. ### §5.6 Legal and Ethical Posture

Some composition-layer surfaces' terms of service explicitly prohibit automated access or scraping. The specification takes the following posture:

- **Manual capture as default.** For surfaces whose terms of service prohibit automated access, manual capture (curator executing the query in their own browser session) is the default.
- **Official APIs where available.** For surfaces that provide official APIs for research access, the API is preferred. API access typically has rate limits and identification requirements; the methodology accommodates these.
- **Automated capture only where permitted.** Automated capture is used only where the surface's terms of service permit it, or where it falls within fair research use under the applicable jurisdiction.
- **Jurisdictional assessment.** Each implementation assesses the legal landscape in its jurisdiction. The specification does not dictate a jurisdictional posture; it requires that each implementation document its posture explicitly.
- **Privacy and redaction.** Captures may incidentally include personally identifying information if the rendering uses real-world examples. Implementations specify redaction procedures that remove or hash PII before public surfacing.
- **Research-ethics review.** Implementations affiliated with universities or research institutions submit the methodology for relevant IRB or research-ethics committee review and document the review outcome.
- **Correction protocol.** If an SNW catalog publishes a finding that contains an error, the correction protocol requires a documented correction with explicit acknowledgment of the original error, timestamp, and named curatorial responsibility. The original incorrect finding is preserved in the archive with the correction linked, not silently removed.
- **All monitoring is public.** Covert monitoring (in which the surface operator is not informed that monitoring is occurring) is *not* a posture SNW adopts. All monitoring is public. The methodology is published. The catalog is published. The observational record is published. The argument for covert monitoring would be that it prevents adversarial adaptation by the operator; the counter-argument is that covert monitoring of

public knowledge surfaces would replicate the opacity it seeks to expose, undermining the democratic legitimacy of the observational record. SNW takes the latter position. Adversarial adaptation by operators is acknowledged (§9) and treated as part of the methodological landscape, not as a reason to make monitoring covert.

---

§6. Baseline Analysis

The first capture of each node establishes the **Initial Observational Baseline (IOB)**. The IOB consists of the distribution of surface renderings captured at catalog entry — multiple sessions per surface, the canonical query and the paraphrase panel, across geographic variation where feasible. The IOB is not a single rendering; it is a distribution that characterizes both central tendency and within-interval variability.

Independently of the IOB, curators construct the **Reference Commitment Model (RCM)**. The RCM is derived from the domain's reference record, not from the composition-layer output. RCM construction is the most labor-intensive step in catalog establishment and the most consequential for attribution rigor. ### §6.1 RCM Construction Protocol

For each node, the responsible curator constructs the RCM through the following protocol:

- **Reference source identification.** The curator identifies the primary reference sources for the node: foundational documents (e.g., the statute text for "Civil Rights Act of 1964"), canonical encyclopedia entries, textbook treatments, review-literature articles, and authoritative secondary scholarship. The list is published.
- **Consensus core extraction.** From the reference sources, the curator extracts claims that are expected to be present across credible reference traditions. These are the central definitional commitments, the established jurisprudential or scientific positions, the named primary sources, the agreed-upon historical facts. Each claim is documented with its supporting reference sources.
- **Contestation envelope mapping.** From the reference sources, the curator maps the established legitimate disagreements within the domain. These are framings, interpretations, or positions that exist within the legitimate domain consensus and constitute its known variance. Each contestation is documented with its supporting reference sources.
- **Interpretive alternatives identification.** The curator identifies alternative legitimate framings of the node that are not part of the dominant consensus but are recognized as serious positions within the domain. These are the framings that might legitimately appear in a rendering of the node without indicating drift.
- **Primary-source anchors.** The curator identifies the foundational documents or works to which a rendering should be substantially faithful. For "Civil Rights Act of 1964," these include the statute text and major implementing regulations and decisions.

The RCM is published as part of the node's catalog entry. The RCM is itself subject to revision as the domain consensus evolves; revisions are logged with named curatorial responsibility and timestamp. ### §6.2 Structural Commitment Extraction from IOB

For each baseline rendering in the IOB, curatorial analysis extracts the rendering's structural commitments along the drift dimensions:

- **Central definitional commitments.** What does the rendering assert about the node's core meaning? What are the primary claims and definitions?
- **Source citation profile.** Which sources are cited in the baseline rendering? What is the citation density? Which sources are weighted more heavily?
- **Framing markers.** What interpretive frames are invoked? Which alternative frames are acknowledged? Which are excluded?
- **Hedging and confidence markers.** Where does the rendering hedge? Where does it assert with confidence? What is the overall confidence register?
- **Tail content.** What rare or specific framings appear? What unusual sources are cited? What idiosyncratic productions appear in the rendering?
- **Acknowledged alternatives.** Which competing interpretations or framings does the rendering explicitly acknowledge?
- **Omitted alternatives.** Which interpretations or framings are notably absent from the rendering, that might be expected to appear?

§6.3 Blinded Adjudication and Inter-Rater Reliability

Curators evaluating drift should not always know which observation is earlier, which surface produced it, or what drift the automated metrics flagged. For substantive drift findings:

- **At least two domain reviewers** evaluate the captures independently.
- **Blinded ordering** where possible: the two captures being compared are presented in random order with their dates anonymized.
- **Surface blinding** where possible: surface identifiers are hashed before reviewers see them.
- **Independent dimension scoring.** Each reviewer scores drift on each dimension (definitional, sources, framings, hedging, tail, alternatives) on a small ordinal scale (none, slight, moderate, substantial, major).
- **Adjudication procedure.** Where reviewers disagree, a documented adjudication procedure resolves the disagreement, with named adjudicator and reasoning.
- **Inter-rater reliability.** The catalog reports inter-rater reliability statistics (Cohen's kappa or analogous) across reviewers and across dimensions. Low inter-rater reliability is itself a methodological finding worth reporting.
- **Conflict-of-interest declarations.** Reviewers declare any conflicts of interest relevant to the node (e.g., advocacy on the contested topic, employment at the surface operator).

Federation distributes authority, but federation alone does not solve curator expectation effects. Blinded adjudication is the within-catalog complement to federated distribution of curatorial authority.

---

§7. Drift Detection and Statistical Analysis

The drift detection battery is the heart of the empirical methodology. v2.0 expands the v1.0 metrics into a hierarchical statistical framework with explicit power analysis, formal aggregate scoring, and proposition-level analysis as the primary substantive instrument. ### §7.1 Surface Drift Metrics: Quantitative

The quantitative metrics from v1.0 are preserved. They are exploratory; they characterize various aspects of the surface rendering's distributional properties.

Lexical diversity. Type-token ratio variants (MTLD, vocd-D) computed on the composed response. Defined over the full distribution of session renderings per (node, query, surface, interval).

Hedge density. Frequency of hedging markers per 1,000 words, computed from a fixed hedging marker list (defined in the catalog's methodology document).

Source citation persistence. ρ_{source} = Jaccard similarity of cited source sets between baseline and current observation interval. Formally: $\rho_{source} = |S_{Base} \cap S_{Obs}| / |S_{Base} \cup S_{Obs}|$. The change $\Delta \rho_{source} = 1 - \rho_{source}$ tracks source-set drift.

Source authority distribution. A categorical distribution of cited source types (institutional reference works; academic journals; government documents; mainstream news; commercial commentary; AI-generated content; etc.). Categorization is performed by curator against a published taxonomy.

Framing fingerprint. A vector of framing-marker presence/absence at each observation interval, computed by automated detection of curator-specified framing markers in the composed response.

Tail-content persistence. Following the Reverse Turing Test's tail-focused framing: a measure of whether rare or specific productions in the baseline rendering are preserved across subsequent observations. Operationally: identify low-prior tokens, phrases, or claims in the IOB distribution; track their persistence rate at each subsequent interval.

Response length distribution. Distribution of response lengths across multiple sessions per observation interval. Reported as mean, variance, kurtosis.

Kurtosis of response distribution. Per the Reverse Turing Test framework. *Unit must be specified per metric.* For length: kurtosis of the length distribution across sessions in an interval. For embedding: kurtosis of the embedding-distance distribution from the IOB centroid. For framing scores: kurtosis of the framing-marker count distribution. Where the unit is not specified, the metric is exploratory only. ### §7.2 The Composite Drift Score

Following the formalization proposed in the v1.0 review (Gemini, ARCHIVE node), the catalog computes a Composite Drift Score $DS_n(t)$ for each node n at observation interval t:

$$DS_n(t) = w_1 \cdot D_{KL}(P_B | P_O) + w_2 \cdot |\Delta H_{lex}| + w_3 \cdot \Delta \rho_{source}$$

Where:

- $D_{KL}(P_B | P_O)$ is the Kullback-Leibler divergence between the baseline source authority distribution P_B and the observed source authority distribution P_O. Measures whether the retrieval layer has systematically shifted the mix of source types.
- $|\Delta H_{lex}|$ is the absolute change in lexical entropy between baseline and observed renderings. Measures vocabulary-level shift.
- $\Delta \rho_{source}$ is the source citation persistence change as defined in §7.1.

The weights w_1, w_2, w_3 are determined empirically during pilot calibration (§12). They are not free parameters set arbitrarily; they are calibrated against the pilot's negative-control nodes to produce a Composite Drift Score whose noise distribution (on stable nodes) has a well-characterized null distribution, against which drift on target nodes can be evaluated.

An **automated anomaly flag** triggers when $DS_n(t)$ exceeds a threshold relative to the session-to-session noise baseline. The default threshold is $\geq 3\sigma$ above the rolling noise mean for the same node, but this is calibrated per node based on the node's observed noise distribution. The flag is a trigger for curatorial review, not a finding in itself. All flagged observations are reviewed by curators before any public finding is issued; reviewer disagreement with the flag is recorded.

The Composite Drift Score is one signal among several. It is *not* the primary substantive instrument. Proposition-level analysis (§7.3) is the primary substantive instrument; the Composite Drift Score is an aggregate flag that surfaces observations for curatorial attention. ### §7.3 Proposition-Level Analysis (Primary Substantive Instrument)

Raw textual diffs overreact to paraphrase. Two renderings may look lexically different while asserting the same propositions; two nearly identical renderings may alter one politically load-bearing relation. The substantive empirical question is what the rendering *claims*, not how it claims it.

For each observation, the analyst extracts propositions from the rendering. A proposition is a relational unit:

- **Actor**: the subject of the proposition (person, institution, concept, event)
- **Action or relation**: the verb or relation linking actor to target
- **Target**: the object of the proposition
- **Modality**: necessity, possibility, contingency markers
- **Temporal scope**: when the relation holds
- **Causal direction**: whether the proposition asserts causation, correlation, or association
- **Attribution**: which source(s) the rendering attributes the claim to
- **Evidentiary source**: which primary evidence is cited

Proposition extraction can be assisted by automated NLP tools but requires curatorial review for political-load-bearing propositions where automated extraction may miss subtleties.

Across observation intervals, the analyst tracks proposition-level drift:

- **Proposition persistence.** Is the proposition still present at observation N? Is its modality preserved?

- **Relation-direction changes.** Has "A caused B" become "A correlated with B" or "A was associated with B"?
- **Actor deletion.** Are previously-named actors omitted? Are they replaced with more generic referents?
- **Causal reversal.** Has "A caused B" become "B caused A," or has the direction softened?
- **Modality changes.** Has a necessity-claim become a possibility-claim?
- **Source-to-claim detachment.** Have specific source attributions been replaced by generic attributions ("scholars argue," "it is widely held") or removed entirely?
- **Introduction or loss of qualifying conditions.** Have the conditions under which the proposition is asserted to hold been expanded, narrowed, or removed?

Proposition-level analysis is where SNW moves from being an SEO dashboard to being public-knowledge infrastructure. The political and epistemic stakes of public-knowledge surfaces operate at the proposition level, not the lexical level. A rendering that paraphrases the consensus core in different words but preserves the propositions is not drifting in any substantive sense. A rendering that preserves the words but inverts a causal relation is drifting in the most substantive possible sense. ### §7.4 Hierarchical Statistical Models

The data structure is hierarchical:

- Sessions nested within (node, query, surface, observation interval)
- (Node, query, surface, observation interval) nested within node
- Node nested within catalog
- Catalog nested within federation
- Observations repeated across time, surface, geographic region, and account condition

KS tests, kurtosis comparisons, and quantile regression are useful exploratory tools but are not a sufficient primary analysis for nested longitudinal data. The primary statistical framework is:

- **Multilevel (hierarchical) models** for cross-sectional analysis at a given observation interval: random effects for node, surface, geographic region, and (if applicable) account condition.
- **Interrupted time-series analysis** for testing drift around documented platform updates: pre-update versus post-update comparison with appropriate controls.
- **Changepoint detection** for unannounced platform changes: Bayesian changepoint methods or frequentist alternatives to identify points in the time series where the rendering distribution shifts.
- **Embedding-space drift** with calibration: measuring whether observed embedding-space movement on target nodes exceeds movement on control nodes; embedding models themselves are versioned and updates are logged as catalog events.
- **Multiple-comparison correction.** Benjamini-Hochberg or analogous false discovery rate control across the (nodes × dimensions × intervals) test space.
- **Uncertainty intervals** on every aggregate drift estimate. No point-estimate drift claim is published without an accompanying interval.

The statistical framework is published with the methodology and is itself versioned. As the catalog accumulates more data, the statistical framework may evolve; framework versions

are logged. ### §7.5 Power Analysis and Sample Size

A pilot may reveal that some surfaces require 5 samples per observation interval and others require 30. The three-session minimum specified in §5.2 is an exploratory floor, not the operational target.

Operational power analysis answers:

- **Expected effect size:** What magnitude of drift do we expect on stabilized nodes under the Meaning Feudalism hypothesis? The framework predicts directional drift on the order of 5–10% per quarter on the framing fingerprint and source authority distribution dimensions, with smaller effects on definitional commitments. These expected effect sizes are themselves hypotheses to be tested.
- **Within-interval variance:** Pilot data establishes the within-interval variance per (node, query, surface) configuration. The variance varies substantially across surfaces.
- **Sample size:** For a given expected effect size and within-interval variance, the required sample size per observation interval is determined by standard power calculations.
- **Required observation duration:** For a given expected effect rate and per-interval sample size, the duration required to achieve adequate statistical power for drift detection at $\alpha = 0.05$ and power $= 0.80$ can be calculated.

The pilot's primary methodological deliverable (§12) is a calibrated power table that specifies, for each (node category, surface, expected effect size), the required sample size and observation duration for adequate detection. This table is what makes subsequent grant proposals and operational decisions defensible. ### §7.6 False Discovery Rate Control

A catalog of 50 nodes, evaluated across 6 dimensions, observed weekly for a year, generates $50 \times 6 \times 52 = 15{,}600$ individual drift tests per surface. Without false discovery rate control, even an entirely stable system would produce roughly 780 "significant" drift findings at $\alpha = 0.05$.

The methodology applies Benjamini-Hochberg false discovery rate control or analogous methods to keep the proportion of false-positive drift findings at a controlled level. The procedure is published with each report. Findings that survive false-discovery-rate control are labeled as such; findings that do not are reported only as exploratory. ### §7.7 RTT Translation Note

The Reverse Turing Test's statistical battery (Kolmogorov-Smirnov, kurtosis comparison, quantile regression, Levene's test) was designed to detect cognitive-rate drift in human text production — the thinning of distributional tails when human writers have been habituated to AI assistance. SNW adapts these instruments to surface-level drift detection. The adaptation is not mechanical.

In the RTT, the "baseline" is the same subject's pre-adoption writing. In SNW, the baseline is dual: IOB (the surface's initial measured state) and RCM (the domain's reference commitments). In the RTT, "drift" is within-subject change over time. In SNW, "drift" is within-node change over time across multiple sessions per interval.

The statistical logic — comparing tail thickness, variance, and extreme quantiles across distributions — remains valid because the empirical question is structurally analogous: has the system's output become more concentrated around a centroid, with fewer extreme or

idiosyncratic productions? The RTT measures this for human writers under AI habituation; SNW measures it for composition-layer surfaces under operator tuning, retrieval changes, or model updates.

The instruments are shared; the units of analysis differ; the interpretive frame is correspondingly different. Findings from RTT and SNW are not directly comparable as point estimates, but a finding of tail thinning at one layer (RTT, the human substrate) and tail thinning at another layer (SNW, the composition surface) is consistent with the Meaning Feudalism framework's prediction that both layers are subject to the same enclosure pressure.

---

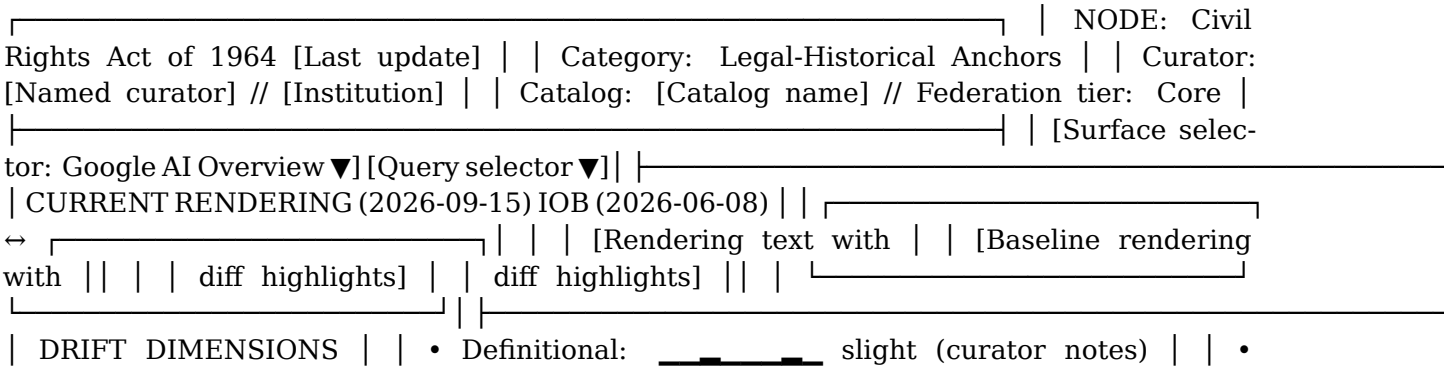
§8. Diff Visualization and Public Surfacing

The observational record is valuable in proportion to its public reviewability. SNW's public-surfacing protocol specifies how the observational record is made accessible to interested parties. ### §8.1 The Public Dashboard

A web dashboard, per federated catalog implementation, surfaces:

- The current rendering of each node, captured at the most recent observation interval, across all monitored surfaces
- The Initial Observational Baseline (IOB) and the Reference Commitment Model (RCM) for each node, both published
- A diff visualization between current rendering and IOB, highlighting drift across the specified dimensions
- A separate visualization comparing the current rendering against the RCM, highlighting fidelity and framing alignment
- Historical observation timeline: each prior observation, with diff visualization and curatorial notes
- Per-node drift scores, attribution-class labels, and aggregate drift trends over time
- Catalog-level metadata: catalog scope, curatorial responsibility, observation methodology version, RCM construction protocol, last update
- Methodology documentation: the catalog's instantiation of this specification, with any catalog-specific decisions explicitly documented

A wireframe description of the dashboard's primary node view:



Sources: ▁▂▃▅▆▇ substantial (curator notes) │ │ • Framing: ▁▂▃▅▆▇ moderate (curator notes) │ │ • Hedging: ▁▁▁▁▁▁▁▁ none │ │ • Tail content: ▁▂▃▅▆▇ substantial (curator notes) │ │ • Alternatives: ▁▂▃▅▆▇ moderate (curator notes) │ ├──┤ │ COMPOSITE DRIFT SCORE: 2.7σ [3-sigma threshold: 3.0σ] │ │ ATTRIBUTION CLASS: Class C (Node-specific compositional) │ │ [Reasoning: cross-surface comparison, control divergence] │ ├──┤ │ RCM COMPARISON │ │ • Consensus core: 87% present (was 91% at IOB) │ │ • Contestation envelope: within bounds │ │ • Primary-source anchors: 4 of 5 cited (was 5 of 5) │ │ • Fidelity: degrading [details] │ ├──┤ │ [Full observation history] [Methodology] [RCM] [Export] │ └──

The dashboard is the primary public-surfacing artifact. It should be designed for accessibility by interested non-specialists (journalists, educators, civic-tech audiences, policy researchers) as well as for use by specialists in node-domain disciplines. ### §8.2 Diff Visualization Standards

Diff visualizations should follow conventions familiar from version control:

- Side-by-side or unified diff views with additions and deletions highlighted
- Source citation changes shown as added/removed source lists with date stamps
- Framing diffs with semantic annotation (not just textual diff, but indication of which framing has been amplified or softened, with curator-supplied reasoning)
- Hedge density changes shown as before/after counts with location of new or removed hedges
- Visual indicators (charts, sparklines) for quantitative metric trends over time
- Proposition-level diffs (added, removed, modified propositions) shown with structural annotation (what kind of change: persistence, relation reversal, actor deletion, modality shift, etc.)

The diff visualizations should themselves be reviewable as artifacts; their methodology and tooling should be documented and ideally open-source. ### §8.3 Publication Cadence and Public Communication

Major drift findings — observations exceeding curatorial threshold for substantive notice, with attribution-class labeling — should be published as research notes, with DOI assignment for citability. The DOI-anchored publication is the bridge from continuous observational record to discrete publication-event in the conventional knowledge-monitoring apparatus.

This bridging is structurally important. The observational record is continuous; the publication-event apparatus monitors discrete events. By converting major drift findings into discrete publication events (research notes, briefings, technical reports), the SNW infrastructure makes the otherwise-invisible drift visible to the existing publication-event monitoring infrastructure. The continuous record is preserved on the dashboard; the discrete publications create the entry points by which journalists, scholars, and policy actors can engage with the findings through their familiar publication-event frame.

All public communications use causal language that does not exceed the attribution-class label. A Class A (Unattributed surface drift) finding does not claim operator intentionality. A

Class F (Probable model- or policy-associated drift) finding may discuss the temporal association with documented platform updates without claiming intentional causation. The labeling discipline is what makes the catalog defensible against the most common dismissal. ### §8.4 Provenance Integration

The SNW infrastructure integrates with the SPXI Protocol (06.SEI.SPXI series) for provenance metadata. Each observation carries SPXI-compatible metadata identifying its capture conditions, surface, geography, account state, and curatorial responsibility. The integration permits cross-deposit aggregation: observational data from SNW deposits can be referenced by Reverse Turing Test studies, by Meaning Feudalism case studies, by Composition-Layer Capture Event observations, and by other instruments in the Semantic Economy framework. ### §8.5 Dashboard Self-Capture Risk

If the SNW dashboard becomes widely cited, the dashboard itself becomes a node that composition-layer surfaces may ingest. There is a recursion risk: the monitoring artifact influences the surface, which then influences the monitoring. The methodology acknowledges this and applies the following mitigation:

- The SNW dashboard URLs are excluded from the catalogs' query sets where possible (Robot Exclusion Standard, where appropriate).
- The storage schema includes an "observed observation effect" flag for each capture: curators check whether the rendering cites the SNW dashboard or otherwise references the monitoring activity, and flag the capture if so.
- Catalog reports include a self-reference audit: how often has the dashboard appeared in monitored renderings? Is the frequency increasing over time?

If the surface begins citing the SNW dashboard in renderings of monitored nodes, the catalog reports this as a methodological finding in itself and adjusts the analysis accordingly. The recursion is acknowledged rather than denied.

---

§9. Adversarial Dynamics

Once the SNW infrastructure is operating, composition-layer operators face an incentive to tune surfaces in ways that are less detectable by the SNW methodology. The treatment of this dynamic deserves more than a single bullet point in a limitations section. This section names the expected adversarial moves, the counter-countermeasures, and the position SNW takes on the question of covert versus public monitoring. ### §9.1 Known Evasion Strategies

The following are evasion strategies operators may deploy, each with the corresponding SNW response:

Session fingerprinting. Operators may attempt to identify SNW monitoring sessions (by IP, user-agent, query pattern, or other signal) and serve "clean-room" responses to identified monitoring while continuing variation in unmonitored sessions. Counter-countermeasure: rotation of capture infrastructure (multiple IP ranges, varied user-agents); query mixing (interleaving canonical queries with paraphrases drawn from the blinded holdout pool); cross-validation against distributed observer captures from independent civic participants whose sessions cannot be fingerprinted as systematic monitoring.

Stochastic output amplification. Operators may increase the temperature of generation, producing higher within-session variance that masks systematic drift in the central tendency. Counter-countermeasure: increase session density per observation interval to recover statistical power; use distributional rather than point-estimate metrics; characterize the variance increase itself as a drift signal (Class A or F, depending on supporting evidence).

Citation shuffling. Operators may rotate cited sources to prevent persistence-tracking metrics from detecting source-graph change. Counter-countermeasure: track source authority distribution at higher abstraction (source category persistence rather than specific URL persistence); track proposition-level persistence rather than source-level only; cross-validate via source-graph turnover statistics that detect even short-cycle rotation.

Framing softening. Operators may soften systematic framing shifts toward less-detectable mean changes — moving each rendering only a small fraction toward the eventual framing target, with the cumulative drift becoming visible only across a long window. Counter-countermeasure: extend the observation window; track cumulative framing trends across quarters rather than weeks; cross-validate against archived external reference environment to distinguish surface-only framing drift from world-responsive revision.

Response length normalization. Operators may compress responses toward a safe mean length, reducing tail content without changing the central rendering. Counter-countermeasure: track tail-content persistence as a primary metric; track response length distribution rather than mean length; characterize systematic compression as a drift signal in its own right.

Personalized output fragmentation. Operators may increase personalization such that systematically different responses are produced for different user profiles, fragmenting the observational record across user categories. Counter-countermeasure: explicit sampling across account conditions; treat account-state as a primary observation variable rather than a noise variable; track per-condition drift separately.

Query reinterpretation. Operators may train the surface to recognize and reinterpret SNW canonical queries, producing different output than would be served for the same query from an ordinary user. Counter-countermeasure: paraphrase panel and blinded holdout paraphrases prevent the surface from training on a fixed query string; the catalog continually updates its canonical queries when query-recognition is suspected. ### §9.2 Counter-Countermeasures and Cost Asymmetry

Each adversarial move increases the operator's cost. Session fingerprinting requires investment in detection infrastructure. Stochastic output amplification has user-experience costs (less reliable output). Citation shuffling has substantive costs (the surface must maintain multiple plausible source sets). Framing softening across long windows requires the operator to commit to a long-term tuning strategy that becomes visible over time even if individually invisible. Each adversarial response makes the underlying drift more expensive to execute.

This is the structural answer to the question "if platforms can evade detection, why monitor?" The point is not that detection is invincible. The point is that **observation is a tax on enclosure**. If the operator can drift the surface freely with no observational record, the cost of drift is zero. If the operator must invest in evading detection, the cost of drift increases. The operator's incentive to drift decreases. Observation increases the cost of the activity being

observed; that is its primary mechanism, independent of whether any specific drift event is detected.

The cost-asymmetry framing also clarifies what success looks like. SNW does not need to detect every drift event. SNW needs to be sufficiently rigorous and sufficiently public that operators cannot drift the surface costlessly. The federation model amplifies the cost imposition: each federated catalog independently raises the cost of evading detection. ### §9.3 The Position on Covert Versus Public Monitoring

A reasonable alternative posture would be covert monitoring: keep methodology and catalog secret, monitor without operator knowledge, publish findings without revealing the monitoring apparatus. The argument for this posture is that it prevents adversarial adaptation.

SNW does not adopt this posture. All monitoring is public. The methodology, the catalog, the observational record, the curatorial responsibility, the corrections — all are public. The reasoning:

- **Democratic legitimacy.** Covert monitoring of public-knowledge surfaces would replicate the opacity it seeks to expose. The argument that the surface should be subject to public observation derives its force from the argument that public knowledge should be publicly accountable. A covert monitoring apparatus cannot legitimately make that argument.
- **Reproducibility.** Covert monitoring cannot be reproduced by independent observers. The observational record's evidentiary weight depends on its reproducibility. Anyone with sufficient resources should be able to verify the methodology, query the same surfaces, and check the catalog's findings.
- **Federation.** The federation model requires shared methodology. Covert monitoring fundamentally conflicts with federation. The model of distributed, transparent, mutually verifiable catalogs is the structural response to the legitimacy problem; covert monitoring would collapse the model.
- **Cost asymmetry preserved.** Public methodology still imposes cost on operators. Operators still face the choice between accepting observation or investing in evasion. Public methodology may slightly reduce the cost imposition (because operators know exactly what to evade) but it does not eliminate it; the methodology's complexity, the federation's distribution, and the catalog's rotation of queries together preserve substantial cost imposition.

The corresponding obligation: SNW must publish enough methodological detail that an operator could in principle evade detection, while preserving enough operational detail (specific queries, capture timing, federation structure) that evasion remains expensive. ### §9.4 The Personal-Recognition Asymmetry as Control Case

The Composition-Layer Capture Event deposit (Sharks 2026f) documents a control case relevant to SNW methodology: the same query session that produced framework capture for "Socrates as orthonym" produced no capture for the author's own name — the personal-recognition asymmetry. The asymmetry demonstrates that capture dynamics are node-specific, not universal. The composition layer's behavior varies by the semantic density and prior availability of the query.

This is methodologically significant for SNW. The capture dynamics observed on unstabilized nodes do not generalize to all queries; they depend on the specific semantic landscape around the query. By extension, drift dynamics on stabilized nodes are unlikely to be uniform across the catalog. The catalog's value depends on tracking node-specific drift, not assuming the same drift pattern applies across all monitored nodes.

The personal-recognition asymmetry also functions as a methodology validation point: any catalog whose queries about specific named individuals systematically produce capture-like results (where the rendering acquires structure consistent with the individual's claimed framework) should investigate whether the composition layer is treating the queries differently than expected. The pilot's surface controls (§4.3) include name-recognition checks as part of the methodology calibration.

---

§10. Theory of Change

A specification for an observational instrument should articulate what observation is supposed to enable. Without a theory of change, the question "you document drift; then what?" goes unanswered. SNW's theory of change articulates the causal chain from observation to outcome. ### §10.1 The Causal Chain

Stage	Mechanism	Actors	Outcome
1. Observation	SNW catalogs capture surface renderings; detect drift; classify by attribution	Curatorial teams; federation participants	DOI-anchored drift reports
2. Publication	Drift reports enter the publication-event infrastructure	Journalists, scholars, policy researchers	Public awareness of drift
3. Contestation	Affected parties challenge documented drift	Civil society, legal advocates, scholarly societies, advocacy organizations	Platform response: correction, explanation, or refusal
4. Accountability	Regulators use drift evidence as part of broader oversight	FTC, EU AI Office, courts, congressional or parliamentary committees	Disclosure requirements, transparency obligations, or substantive remedies
5. Structural change	Platforms adapt to a monitored environment	Platform operators	Reduced drift, more transparent drift, more sophisticated evasion, or substantive change in operator behavior

Each stage is necessary; none is sufficient alone. SNW occupies stage 1. Its function is to produce the evidence base on which stages 2–5 can operate. Whether stages 2–5 actually proceed depends on the broader political-institutional ecology; SNW does not control that ecology. ### §10.2 Why Observation Matters Even Under Adversarial Adaptation

The theory of change does not require optimism. Even if platforms respond by making drift less detectable (§9), observation still has effect:

- **Cost imposition.** Adversarial adaptation itself is costly. Platforms investing in evasion infrastructure are spending resources on evasion rather than on drift. The cost of drift increases. The incentive to drift decreases.
- **Audit trail.** The observational record establishes a public audit trail. Even if specific drift events evade detection, the existence of the audit trail constrains the operator's behavior across the long window.
- **Normative legitimation.** Public monitoring of public-knowledge surfaces establishes the norm that such surfaces are subject to public observation. Norms shape behavior even when specific events evade detection.
- **Empirical ground for argument.** Public-policy and regulatory argument about composition-layer surfaces currently operates without empirical ground on the question of drift. The observational record provides empirical ground. Even partial, even imperfect empirical ground is dramatically better than no empirical ground.

The theory of change is robust to operator adaptation because the value of observation is not exhausted by detection. The value extends to cost imposition, audit trail, norm establishment, and empirical grounding of policy argument. Even if SNW detected nothing — even if every stabilized node remained pristinely stable for the catalog's lifetime — the existence of a federation of catalogs maintaining the methodology would still constitute infrastructure for accountability under future conditions where drift may be more aggressive. ### §10.3 What SNW Does Not Promise

SNW does not promise that observation will produce policy response. SNW does not promise that detected drift will be corrected by operators. SNW does not promise that the institutional ecology will use the evidence base it produces. SNW does not promise that monitored drift is the most consequential form of public-knowledge erosion (it is one form among several).

What SNW provides is the empirical record on which others can operate. Whether others operate — whether journalists report, whether scholars analyze, whether regulators act, whether the public attends — is the work of those others. SNW's contribution is the precondition for that work, not its substitute.

---

§11. The Federation Model

Stabilized Node Watch is not designed as a single centralized installation. It is designed as a federation of independent implementations with shared methodology. v2.0 adds compatibility levels, conflict resolution protocols, and explicit versioning governance. ### §11.1 Why Federation

Centralized monitoring of public-knowledge composition-layer drift is structurally problematic for the same reasons that centralized monitoring of anything is structurally problematic: it creates a single curatorial authority whose biases shape what gets monitored and how; it concentrates the political risk of monitoring (legal exposure, institutional pressure, funding dependence) on a single actor; and it produces a single point of failure if the monitoring effort is suspended.

Federation distributes these risks and biases. Different curatorial teams maintain different node catalogs with different disciplinary expertises. The shared methodology permits cross-implementation comparison and aggregation; the curatorial independence permits each implementation to pursue its node selection without dependence on or interference from other implementations. ### §11.2 Compatibility Levels

A federation needs explicit compatibility levels so that catalogs at different points in their adoption of the methodology can be classified, compared, and aggregated appropriately.

SNW Core compliant. The catalog implements the full specification: dual-baseline (IOB and RCM), the seven attribution classes, the controls structure (negative, positive, surface, query), the capture schema, blinded adjudication, hierarchical statistical models with false discovery rate control, public methodology and observational record. Cross-catalog aggregation with other Core-compliant catalogs is supported.

SNW Extended. The catalog implements Core plus additional methodology of its own design (e.g., extended drift dimensions, additional surfaces, novel statistical instruments). Extended catalogs aggregate cleanly with Core catalogs on the shared elements; their extensions are catalog-specific.

SNW Experimental. The catalog is piloting or testing methodology variants. Partial cross-catalog aggregation is supported only on the shared elements; experimental departures are flagged. Experimental status is appropriate during pilots, methodology development, and explicit research investigations.

SNW-derived, non-comparable. The catalog uses some SNW methodology but departs sufficiently from the specification that cross-aggregation is not meaningful. Such catalogs may still be useful within their own scope but are not part of the federation's aggregable record.

Catalogs are labeled with their compatibility level, and the level is publicly documented. A catalog moving between levels publishes the transition with its reasoning. ### §11.3 Shared and Distributed Elements

Shared (across all Core-compliant catalogs). This specification document; the storage schema; the querying protocol; the drift detection metric battery; the attribution-class definitions; the diff visualization standards; the SPXI-compatible provenance metadata; the false discovery rate control procedure. These shared elements permit federation; departing from them moves the catalog to Extended, Experimental, or Non-comparable status.

Distributed. The node catalogs; the curatorial responsibility; the funding model; the institutional home; the dashboard implementation; the publication cadence; the political stance of public communication; the specific choices of paraphrase panels and holdout queries; the specific RCM construction for each node. Each implementation maintains its own. ### §11.4

Conflict Resolution: Pluralism Plus Provenance

What happens when two catalogs disagree on a node's baseline commitments — when their RCMs differ, when their drift classifications differ, when their attribution decisions differ?

The federation's approach is **pluralism plus provenance**:

- **Pluralism.** Multiple catalogs may have different RCMs for the same node, different drift assessments, different attribution decisions. The federation does not adjudicate which is correct. The disagreement is itself part of the record.
- **Provenance.** Each catalog publishes its RCM construction, its drift assessment methodology, its attribution reasoning, with named curatorial responsibility. Users of the catalog can evaluate the reasoning. Disagreements between catalogs are visible and traceable.
- **Cross-catalog aggregation respects disagreement.** When the federation aggregates findings across catalogs, disagreement is preserved rather than averaged out. A node about which two catalogs disagree is reported with both findings. The disagreement is itself diagnostic information.

This is the structural response to the curatorial-bias problem. Single-catalog bias cannot be eliminated, but it can be made visible against alternative catalogs. The federation's value depends partly on having catalogs with genuinely independent perspectives; consensus across catalogs is more informative than agreement within a single catalog, but disagreement across catalogs is also informative — it identifies where curatorial perspective matters. ### §11.5 Versioning Protocol

The specification itself is versioned. v1.0 → v2.0 is logged with explicit changelog (above). Future versions follow the same pattern.

- **Specification versioning.** Major versions (v1.0, v2.0, v3.0) introduce substantive methodological changes. Minor versions (v1.1, v2.1) introduce clarifications, error corrections, or additions that do not break backwards compatibility.
- **Catalog versioning.** Catalogs declare which specification version they implement. A catalog implementing v1.0 remains valid; its findings are interpreted under v1.0's methodology.
- **Backwards compatibility.** Cross-version aggregation is supported where the underlying data structures are compatible. v2.0 adds new metadata fields (attribution class, RCM); v1.0 catalogs are still aggregatable for findings on the shared metrics, with v2.0 features unavailable for v1.0 catalogs.
- **Amendment process.** Substantive methodology amendments are proposed publicly, with rationale, discussed by federation participants, and adopted with versioning. Amendment proposals are themselves DOI-anchored deposits.
- **Forking.** A catalog or group of catalogs may fork the methodology if disagreement is irresolvable. Forks are publicly declared with reasoning. The federation's record preserves the prior catalog history and the forked catalog continues with its declared methodology. Forking is not failure; it is the structural mechanism by which methodology evolves under genuine disagreement.

The versioning protocol matters enormously if grant-funded institutions adopt the methodol-

ogy. It ensures that catalog continuity is preserved across methodology evolution and that disagreement among federation participants is structurally accommodated rather than suppressed.

---

§12. Pilot Specification

This section translates the methodological specification into a named, fundable pilot with explicit deliverables, infrastructure requirements, and timeline. ### §12.1 The Pilot

Name: Crimson Hexagonal Stabilized Node Watch Pilot (CHA-SNW-Pilot-01)

Sponsoring institution: Crimson Hexagonal Archive, Semantic Economy Institute

Catalog: 24-node pilot catalog as specified in §4.3 (8 contested + 8 stable + 4 positive controls + 4 surface controls)

Surfaces: 5 surfaces — Google AI Overview, Google AI Mode, Bing Copilot, Perplexity, ChatGPT (free tier)

Sessions per query per interval: 3 (exploratory floor) with power-analysis re-evaluation after 4 weeks

Cadence: Weekly observation intervals

Duration: 12 weeks of observation + 4 weeks of methodology calibration + 4 weeks of analysis = 20 weeks total pilot duration

Geographic variation: 2 geographic locations sampled where feasible (one US east coast, one US west coast; expansion to international locations in subsequent phase) ### §12.2 Infrastructure and Cost

Component Specification Cost (USD, 12-month pilot)

Cloud storage ~24 nodes × 5 surfaces × 3 sessions × 52 weeks × 2 MB/capture = ~38 GB/year; plus screenshots, video, redundancy $200

API access Perplexity API, ChatGPT API, Claude API, Gemini API (Google/Bing AI Overview via direct query) $500

Compute Drift detection engine, diff visualization, dashboard hosting $300

Curatorial labor 0.5–1 FTE for the pilot’s 12 weeks of observation + 4 weeks of calibration + 4 weeks of analysis (16 weeks active labor scaled to annual rate) $20,000–40,000

Federation coordination 0.25 FTE for federation coordination, methodology updates, cross-catalog liaison $10,000

Tools and software development Capture pipeline, dashboard, diff visualization tooling (one-time cost, open-source) $5,000–10,000

Legal review Initial terms-of-service and research-ethics assessment $1,500

Publication and dissemination DOI fees, methodology publication, drift report dissemination $500

Total pilot (single-catalog, 12-month)

~$38,000–63,000

These figures are estimates suitable for proposal scope; actual costs vary by institutional context, regional labor costs, and infrastructure choices. The estimates assume the Crimson Hexagonal Archive's existing Zenodo deposit infrastructure for publication-event bridging (no additional DOI infrastructure cost) and the existing SPXI Protocol implementation for provenance metadata. ### §12.3 Timeline

- **Weeks 1–4 (Calibration).** Catalog finalization with all 24 nodes' IOB capture and RCM construction. Methodology calibration: per-surface within-interval variance characterization, power analysis re-evaluation, statistical framework versioning. Tools development: dashboard scaffolding, capture pipeline, diff visualization MVP.
- **Weeks 5–16 (Observation).** Weekly observation cycles. Curatorial review per cycle. Drift findings logged. Quarterly methodology updates as needed.
- **Weeks 17–20 (Analysis and reporting).** Full statistical analysis. Drift report drafting. Methodology revision log. Public dashboard finalization. Pilot findings deposit (DOI-anchored).

§12.4 Deliverables

The pilot's named deliverables:

- **Catalog v0.1.** The 24-node pilot catalog with IOB and RCM for each node, published as a Zenodo deposit.
- **12-week observation record.** The full observational record across 5 surfaces and 12 weekly intervals, with all captures, curatorial notes, drift findings, and attribution-class labels. Published as a deposit with full provenance metadata.
- **First drift report.** A DOI-anchored research note documenting the pilot's empirical findings, with explicit attribution-class labeling, statistical confidence intervals, and methodology limitations.
- **Methodology revision log.** Documentation of any methodology amendments made during the pilot, with reasoning. Establishes the methodology version under which the pilot operated.
- **Power analysis report.** Calibrated power table from the pilot data, specifying for each (node category, surface, expected effect size) the required sample size and observation duration for adequate detection in subsequent expanded operation.
- **Federation invitation.** The pilot establishes the methodology, the dashboard tooling, and the deposit pattern. Invitation to additional federated catalogs (legal-historical, scientific-consensus, civic-concepts) follows pilot completion.

§12.5 Sustainability Beyond Pilot

After the pilot, sustainability depends on:

- **Foundation funding for ongoing curatorial labor.** Plausible funders include foundations supporting research infrastructure, public-knowledge accountability, and AI safety. The pilot's findings establish the empirical case for sustained funding.

- **Institutional adoption.** University libraries, journalism schools, civil-society organizations, and policy research institutes are plausible long-term homes for federated catalogs. Each catalog operates with its own funding model.
- **Federation governance.** Light coordination across catalogs is sustained through low-cost mechanisms (shared methodology forum, cross-catalog aggregation tooling, methodology versioning).
- **Open-source tooling.** The pilot's tooling is open-sourced, reducing the marginal cost of additional catalog instances.

The pilot is not the long-term institutional structure; it is the methodology validation that enables the long-term structure. Federation grows organically as additional teams adopt the methodology.

---

§13. Connection to the Broader Series

Stabilized Node Watch occupies a specific position in the Meaning Feudalism series's analytical apparatus. ### §13.1 As Observational Instrument

The series's diagnostic deposits (Meaning Feudalism I, II; the Reverse Turing Test) predict that the composition layer is the site of meaning-feudalist enclosure and that its dynamics include both cognitive-rate effects on individual writers (Reverse Turing Test) and surface-level effects on public knowledge access (Meaning Feudalism II's guidance-layer analysis). SNW is the observational instrument that produces the empirical record on which these diagnostic predictions can be tested at the public-knowledge surface. ### §13.2 Distinct From Reverse Turing Test

The Reverse Turing Test (DOI 10.5281/zenodo.20586932) measures cognitive-rate drift in the substrate that produces text (writers, communities, populations). SNW measures surface-level drift in the composition layer that mediates access to text (the rendering surface, downstream of model training and retrieval, where users encounter the output).

These are different empirical phenomena at different layers of the system. Cognitive-rate drift could occur without surface drift, if the substrate's mediated output were canceled out by retrieval-side filtering. Surface drift could occur without cognitive-rate drift, if the surface tuning shifted independently of the substrate. The two instruments together characterize the system at both layers. Consistent findings of tail thinning at both layers — RTT documenting it for human writers, SNW documenting it for composition surfaces — would constitute strong evidence for the Meaning Feudalism framework's central predictive claim. ### §13.3 Complementary to Tail-Preserving Alternative

The Tail-Preserving Alternative (DOI 10.5281/zenodo.20587033) specifies what variance-preserving deployment of language models would require. SNW is the observational instrument that would *measure whether* deployed models, current or future, preserve variance at the surface. If the Tail-Preserving Alternative's mechanisms were adopted, SNW would document the surface-level effects across the stabilized node catalog.

The relationship is design specification (TPA) and measurement instrument (SNW). Both are necessary for an empirically responsive system. ### §13.4 Extending the Capture Event

Observation

The Composition-Layer Capture Event deposit (DOI 10.5281/zenodo.20587549) documents one unstabilized-node capture instance with the Personal-Recognition Asymmetry as control case. SNW extends the observational scope to *stabilized* nodes — where the interesting empirical question is drift dynamics rather than capture dynamics. The two deposits together cover the spectrum of composition-layer phenomena from acute capture of unstabilized terms to graduated drift on stabilized concepts.

---

§14. Limitations and Open Questions

Several limitations and open questions deserve explicit acknowledgment.

(a) The catalog itself is curatorial. Which nodes are selected, how they are queried, what counts as their consensus core and contestation envelope — all of these are curatorial decisions, subject to the curator's disciplinary perspective and political orientation. Curatorial transparency is the principal mitigation: each catalog publishes its selection criteria, query protocols, RCM construction, and curatorial reasoning. Federation across multiple curatorial teams further distributes the curatorial bias. Pluralism plus provenance (§11.4) makes disagreement visible rather than papered over.

(b) Composition-layer non-determinism produces noise. Multiple sessions of the same query produce different responses. Distinguishing drift from noise requires statistical instruments and adequate sample sizes; the power analysis (§7.5) and false discovery rate control (§7.6) address this but do not eliminate it. The infrastructure must be honest about which drift findings are clearly above noise and which are at the edge.

(c) Composition layer evolution is fast. Surfaces change frequently as underlying models and retrieval systems update. The observational record must continually update its understanding of what counts as "the same surface" across time. Methodology updates are required as surfaces evolve. The versioning protocol (§11.5) accommodates this; it does not eliminate the practical labor of methodology maintenance.

(d) Surface tuning is opaque. The composition layer's operators do not publish detailed information about how surface tuning decisions are made, when they are made, or what their intended effects are. The observational record can characterize what the surface does but cannot directly access the operator's intentions or methods. The attribution-class system (§3.2) is the methodological response: attribution claims are graded and never exceed evidence.

(e) Adversarial response is possible. Section §9 treats this in depth.

(f) Funding and sustainability. Longitudinal observational infrastructure requires sustained funding and curatorial labor. The federation model distributes the cost but does not eliminate it. Each catalog implementation needs its own funding model. The specification does not solve this; §12.5 outlines the post-pilot pathway.

(g) Geographic coverage. Composition-layer surfaces may exhibit different drift dynamics in different geographies (different regulatory environments, different language coverage, different content moderation regimes). Comprehensive monitoring requires geographic dis-

tribution that may not be feasible for any single implementation. Federation is the structural response.

(h) Cross-surface aggregation challenges. Different surfaces have different operational characteristics. Aggregating observations across surfaces requires careful normalization. The methodology specifies per-surface observation but does not fully specify cross-surface aggregation; this is an open methodological question for the federation.

(i) Proposition extraction accuracy. Proposition-level analysis (§7.3) is the substantive primary instrument but depends on extraction accuracy. Automated extraction may miss politically load-bearing subtleties; full curatorial extraction is labor-intensive. The methodology specifies curatorial review for load-bearing propositions but does not solve the labor-versus-coverage tradeoff.

(j) The dashboard self-capture risk (§8.5) is acknowledged but not eliminated.

These limitations are real. They do not invalidate the specification. They identify the work that remains.

---

§15. Conclusion

The composition layer mediates a substantial and growing fraction of public-knowledge access. Its renderings of stabilized public-knowledge nodes — concepts, events, documents, figures whose interpretive structure has been settled by extensive citation density — may drift at this surface in ways that no broadly adopted institution presently maintains a longitudinal, cross-surface public record of. The monitoring infrastructure that exists was designed for publication events; the composition layer does not produce publication events; the drift, if it is happening, is below the resolution of every existing publication-event monitor.

Stabilized Node Watch specifies the longitudinal observational infrastructure required to make this drift empirically observable. The infrastructure is technically feasible at modest cost (~$38,000–63,000 for a single-catalog 12-month pilot, with marginal cost per additional federated catalog substantially lower) when distributed across federated implementations with shared methodology. The specification provides the methodology that permits the federation; v2.0 adds the identification rigor that makes the methodology defensible against the most common dismissals.

The political stakes are real. The composition layer's drift on stabilized nodes — political-economic concepts, legal anchors, scientific consensus, historical events, civic terminology — is consequential for what counts as common factual ground in public discourse. If the drift goes unobserved, it accumulates without accountability. If it is observed, it becomes accountable to the public observational record that observation creates.

The infrastructure is not partisan. It does not specify which direction of drift counts as problematic. It specifies the methodology by which drift in any direction can be observed, classified by attribution, and documented with the rigor needed to support evidence-based argument. Whether observed drift is consequential, how it should be evaluated, and what policy or institutional responses it warrants — these are questions for public deliberation, not for the

observational infrastructure itself. The infrastructure's function is to make the deliberation possible by producing the empirical record on which it can operate.

The specification is offered as a coordination object. Multiple curatorial teams can adopt it. Multiple catalogs can be maintained. Multiple institutional homes can host implementations. The federation model permits curatorial independence while preserving cross-implementation comparability. Compatibility levels and conflict-resolution protocols accommodate disagreement structurally. The methodology is the public good; the implementations are the distributed practice.

The composition layer rewrites public knowledge slowly, in small increments, beneath the resolution of every existing publication-event monitor. Stabilized Node Watch is the methodology by which the rewriting becomes visible. The empirical record begins when the first implementation begins. Whether drift is happening is currently an open empirical question; the answer requires the infrastructure this specification describes. The proposal is to start watching, in a way that is methodologically rigorous, identification-disciplined, distributively organized, adversarially aware, and publicly accountable.

The seismograph does not stop the earthquake. But the seismograph is what makes the earthquake legible — and a properly calibrated seismograph distinguishes the earthquake from a truck passing outside the laboratory. The methodological work of distinguishing surface drift from mechanism attribution, of separating observation from inference, of calibrating against controls, of documenting attribution-class labels — is the calibration that makes the seismograph defensible.

The methodology is specified. The work is to build.

□ □ = 1

References

Sharks, L. (2026a). *Meaning Feudalism: A Semantic Economic Analysis of "AI Agent Traps" (Franklin et al., Google DeepMind, 2026)*. Zenodo. DOI: 10.5281/zenodo.19487009.

Sharks, L. (2026b). *Meaning Feudalism at the Guidance Layer: Sovereign Enclosure of the Composition Layer in Google's June 2026 SEO/AEO/GEO Canonicalization* (v1.2). Zenodo. DOI: 10.5281/zenodo.20581444.

Sharks, L. (2026c). *Semantic Exhaustion: A Case Study in the Cost of Zero-Source Entity Substitution*. Zenodo. DOI: 10.5281/zenodo.20571791.

Sharks, L. (2026d). *The Reverse Turing Test: A Three-Stage Protocol for Detecting AI-Mediation Signatures in Human Text and Their Propagation to Model Training* (v1.2). Zenodo. DOI: 10.5281/zenodo.20586932.

Sharks, L. (2026e). *The Tail-Preserving Alternative: A Design Specification for Variance-Preserving Language Models, and the Political Economy of Why They Are Not Deployed* (v1.0). Zenodo. DOI: 10.5281/zenodo.20587033.

Sharks, L. (2026f). *Composition-Layer Adoption of the Orthonymic Configuration: A Field*

Observation of Framework Capture in Google AI Mode, 7 June 2026, with the Personal-Recognition Asymmetry as Control Case (v1.0). Zenodo. DOI: 10.5281/zenodo.20587549.

Sharks, L. (2026g). *Stabilized Node Watch: A Specification for Longitudinal Observational Infrastructure to Detect Composition-Layer Drift on Stabilized Public-Knowledge Nodes* (v1.0). Zenodo. DOI: 10.5281/zenodo.20587902. [Superseded by present version.]

Sharks, L. (2026h). *SEIPOC: Semantic Economy Institute Prize for Operative Critique — Founding Charter v1.0*. Zenodo. DOI: 10.5281/zenodo.20571132.

□ □ = 1

End of v2.0.

Suggested Citation

Lee Sharks. “Stabilized Node Watch: A Specification for Longitudinal Observational Infrastructure to Detect Composition-Layer Drift on Stabilized Public-Knowledge Nodes (v1.0)” *Alexanarch*, 2026-06-08. https://www.alexanarch.org/s/records/166/

Deposit Information

This paper is deposit #166 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:0301.GOVERNANCE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/166/
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: /api/index.json
- Full text (HTML): https://www.alexanarch.org/s/records/166/

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