METHOD / INSTITUTIONAL OBSERVABILITY / WORKING PAPER

Institutional Reality as a Correctible Evidence Chain

A methodological specification of temporal claims, witnessed execution and institutional self-observation

Juan Antonio Marín Liera
System Friction Institute
ORCID 0009-0002-3714-1467

OBJECT: SFI-RC-2026-METHOD-001
16 NOVEMBER 2026
STATE: AUTHOR APPROVED
PEER REVIEW: NOT OBSERVED
DOI: NOT ASSIGNED

PUBLIC CLAIM BOUNDARY

A falsifiable reporting framework and bounded documentary illustrations. No independent comparative trial, general predictive validity, or external institutional RETURN is claimed.

TEXT AND ORIGINAL EDITORIAL FIGURES: CC BY 4.0 UPON RELEASE
THIRD-PARTY SOURCES AND THE SFI MARK ARE EXCLUDED

READER'S MAP / EVIDENCE CONTRACT

How to read this research object

The diagram below defines reportable states, not a completed experimental result. An absent step remains absent. Observation and authorization cannot be replaced by inference.

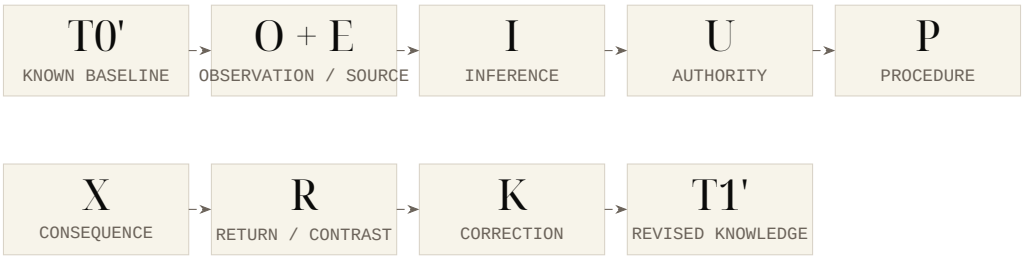


FIG. 01 / PROPOSED REPORTING SCHEMA / NOT A CAUSAL MODEL / DASHED ARROWS = DECLARED SEQUENCE ONLY

EPISTEMIC READOUT / THIS MANUSCRIPT

OBSERVED / REPORTED	Public source vintages and bounded SFI register snapshots; documented as such.
INFERRED	Feasibility of identity/state separation within explicit identifiers and retained lineage.
HYPOTHESIZED	H1–H4 comparative improvements; no matched prospective test.
NOT OBSERVED	Independent comparative gain, external RETURN, scientific validation.
AUTHORIZED	The author requested publication of a bounded methodological working paper; independent review not implied.

What this paper does and does not establish

The manuscript supplies a reproducible reporting specification and a protocol for later independent evaluation. References to NIST, W3C PROV and SEC historical records are antecedents or documentary sources, not endorsements of SFI. The publication's supporting inventory is supplied separately; restricted Method Lab records are not redistributed.

PROVENANCE / SEE THE TECHNICAL RECORD AND MANIFEST IN THE COMPANION PACKAGE.

Abstract

Institutional systems can preserve a record of a claim without preserving what was knowable when the claim was made, who was permitted to act on it, whether execution occurred, or what the external world subsequently showed. This methodological article specifies a correctible, provenance-oriented evidence chain for distinguishing observations, inferences, attestations, authorization, execution, externally observed consequences and later revisions. It defines temporal and authority constraints and four hypotheses for independent evaluation against information-equivalent conventional logs and audit workflows. Two bounded institutional continuity episodes and one public documentary package provide feasibility illustrations; none constitutes an independent test of predictive superiority or a universal identity function. The contribution is a falsifiable reporting and evaluation protocol, a compact Reality Passport, and an explicit account of which empirical claims remain open.

Keywords: institutional observability; provenance; temporal integrity; human–AI systems; authorization; audit trails; falsifiable methodology; scientific reporting.

1 Introduction and research question

Organizational records routinely conflate the existence of a statement with its evidentiary validity, permission to perform an operation with the operation itself, and completion of an internal workflow with an external consequence. These distinctions matter when evaluating human–AI processes that generate large volumes of apparently coherent explanations. The central research question is: *Under which observable conditions can an institution reconstruct what was known, inferred, authorized, executed and subsequently observed, without allowing later information to rewrite its earlier knowledge state?*

This paper is a methodological specification and bounded documentary report. It does not claim that a comparative randomized trial has been conducted. Its secondary question—whether continuity-bearing human–AI–institutional configurations improve reconstruction or correction after participant or model substitution—remains an untested hypothesis. A method's stated intent is not a demonstrated effect.

2 Background and contribution

Pragmatic accounts of inquiry emphasize corrigibility under experience and criticism [1]. Early work on purposeful behavior examined feedback as an explanatory relation [2]. Resilience engineering studies organizational performance, adaptation and failure [3]. The NIST AI Risk Management Framework provides a baseline for mapping, measuring, managing and governing AI risks [4]; W3C PROV supplies a standard vocabulary for entities, activities, agents and derivations [5]. These literatures provide antecedents, not evidence that the SFI implementation is novel in every respect or already superior.

The proposed contribution is operational: (i) preserve the time of an event, its observation and its later institutional recognition separately; (ii) make an authority claim inspectable independently of the actor's identity; (iii) require a link from any claimed consequence to suitable evidence; and (iv) carry missingness, contradictory evidence and correction forward without silently rewriting the historical baseline.

3 Reality Chain and state model

For a bounded case c, let the record be an ordered, typed sequence

$$\mathcal{C}_c = \left\{T_0', O, E, I, U, P, X, R, K, T_1'\right\}, \tag{1}$$

where T'_0 is the frozen initial knowledge state; O is observation; E records evidence and provenance; I is inference; U describes authority and scope; P records the chosen procedure or attempted perturbation; X is an externally observable consequence, if one exists; R is the evidence-bound contrast or RETURN; K is subsequent correction/integration; and T'_1 is the updated, bounded knowledge state. The sequence is a reporting schema, not a claim that all stages necessarily occur or that they have a single universal causal order. An observational study may have no deliberate intervention, and a study without an external consequence must not invent one.

Each assertion should be linked to an identifier, source, version, actor/instrument, relevant time, epistemic class, asserted authority, supporting and contradictory evidence, and correction lineage. The following times are stored independently when available:

$$t_{\text{event}}, \quad t_{\text{observed}}, \quad t_{\text{recognized}}. \tag{2}$$

They are *not presumed equal*; the notation is a warning against conflation, not a universal empirical inequality. A frozen T'_0 must be reconstructible without drawing on later sources.

Table 1: Compact reporting contract for an evidence-bearing transition.

Field	Minimum reportable content
T'_0	Source cut-off, available facts, unresolved questions, permitted knowledge.
Observation	Instrument/observer, source identity, event time, acquisition time.
Inference	Method, assumptions, alternative explanations, purpose, uncertainty.
Authority	Actor, role, scope, basis of permission, review/approval boundary.
Procedure	Proposed versus executed operations, inputs, code, transformations and exclusions.
RETURN	External outcome evidence, observation window, or explicit NOT OBSERVED.
Contrast	Expected versus observed, residuals, adverse evidence, rival causes.
T'_1	Supported revision, unchanged claims, missing facts and next test.

3.1 Attestation is not certification

An attestation associates a responsible actor with a bounded statement and a defined evidence scope. A cryptographic signature or a successful software receipt may strengthen provenance but does not, by itself, prove the statement true, establish external effects, or grant broader authority. The ability to access a system is distinct from a policy-grounded right to make or approve a particular claim.

4 Hypotheses, rivals and failure conditions

Four hypotheses are specified as prospective candidates. The absence of a registered, independently evaluated prospective test is stated explicitly.

H1: Reconstructibility. Given information-equivalent evidence, a typed Reality Chain improves evaluators' accuracy in identifying facts knowable at T'_0 relative to flat chronological logs. *Rival:* conventional logs with matched provenance metadata perform equivalently.

H2: Temporal integrity. Source-vintage enforcement reduces retrospective leakage. *Rival:* explicit instructions and ordinary version control provide the same protection.

H3: Correction. Linking later independently observed consequences to earlier decisions reduces unsupported case closure. *Rival:* standard audit/review procedures perform equally well.

H_4: Distributed continuity. Historical linkage improves reconstructibility after operator or model substitution. *Rival:* context volume, retrieval quality or ordinary project management explains the difference.

Predefined falsifiers include equivalent or worse performance, leakage across the frozen time boundary, irreproducible provenance, unexplained operator dependence, and failures to distinguish authorization from execution. There is no favorable interpretation of results stipulated in advance.

5 Proposed independent evaluation

5.1 Study design

A future evaluator would select eligible publicly documented institutional cases, freeze source vintages and reconstruct each case under three matched conditions: A, a flat chronological log; B, the typed chain; and C, a conventional audit/project-management system supplied with equivalent information. Condition labels would be concealed from scorers where practical. Case selection, excluded information, evaluator training, outcome definitions, randomization scheme, stopping criteria and scoring rubric must be preregistered *before* examining comparative outcomes. No such comparative run is claimed here.

5.2 Outcome measurements

Suggested outcomes are: accuracy of T_0' factual reconstruction, rate of false retrospective attribution, rate of confusion between authorized and observed action, missingness recognition, time to reconstruct, agreement between independent raters, correction quality and the remaining error after operator/model substitution. Appropriate denominators, uncertainty intervals and adverse cases must accompany any reported statistic. Internal test pass rates are not substitutes for these outcomes.

5.3 Privacy and independence

Public derivatives must exclude private relational communications, raw user data and proprietary materials without a documented basis for sharing. The investigator preparing a case should not be its sole independent evaluator. Claims about individual motives, mental states or institutional wrongdoing require evidence appropriate to those claims. An automated assessor cannot confer its own publication authority.

6 Bounded documentary observations

6.1 Case A: institutional identity and certification status

The SFI-A laboratory object records a bounded case in which a completion-controller's historical records and requirement identity remained traceable while present certification admissibility changed. This supports *identity/state separability under explicit identifiers and lineage*. It does not establish persistence of the underlying functional capability, a universal criterion of identity, or robustness when identifiers are missing or manipulated.

6.2 Case B: research-object content revision

A second SFI-A episode reports that a Method Lab research object retained its identifier, object class, objective and method across a version change from 0.3 to 0.3.1 while a finding and lineage record were added. Again, traceability is aided by explicit identifiers; independent individuation under ambiguous identity conditions has not been tested. These are bounded process observations, not a representative sample from which to derive a population effect.

6.3 Case C: public retrospective package

The October 2026 candidate manuscript describes a bounded Knight Capital (2012) public derivative transported to the SFI Research Hub. Its contemporaneous package manifest reported 15 documentary findings, five official source records, eight relations and zero scientifically validated tracks. Those numbers are reported as a historical package snapshot, not refreshed measurements. Different source-publication vintages must not be projected backward into the T_0' before the incident. The transport demonstrates an inspectable publishing operation; it is not evidence of prospective predictive accuracy, MIHM calibration or cognitive individuation.

7 Temporal inventory and limitations

The source manuscript recorded 27 Method Lab research objects on 12 October 2026. The subsequent read on 16 November 2026 returned 32 objects, including SFI-A (working, findings registered), SFI-B (working, findings registered), SFI-C0 (frozen candidate), SFI-C1 (withheld), a private relational field object with no linked primary communications, and several simulated or documentary research objects. Counts are inventory snapshots from different dates and are *not* experimental results; they are not evidence of scientific growth or quality. The SFI-A findings were last marked updated on 2 October 2026 in the retrieved projection. The current method lab also described synthetic MIHM diagnostic runs and an unobserved external RETURN, which must remain distinct from operational execution.

The evidence base is small, internally instrumented and heterogeneous. There is no blinded A/B outcome, no calibrated general identity threshold, no independent external replication and no demonstrated causal effect of the reporting grammar. The comparison protocol, pass/fail rubric and statistical power considerations are future work. Internal metadata may be incomplete or reclassified in later reports, so publication must version the consulted evidence snapshot.

8 Discussion: what can legitimately be concluded

The documented cases establish feasibility of recording selected identity, version, source and authority distinctions in an operational institutional setting. They also illustrate a general reporting hazard: later publication or recollection can make earlier knowledge appear more complete than it was. This is a methodological concern, not proof that the proposed schema improves decisions compared with established alternatives. The distinction between institutional observability and empirical validity is therefore not cosmetic: it limits the central claim.

The Reality Passport provides a compact way to navigate these distinctions in an institutionally formatted PDF. The passport is a *projection* of source-level structured records and must not become their sole repository. Where a journal prohibits marginal annotations, the scientific content is retained in ordinary

text, tables and supplementary material. Typography and color are editorial encodings, not evidence grades or publication authorization.

9 Conclusion

A correctible record must preserve what was previously knowable, what was observed, inferred, authorized and attempted, and what evidence later justified correction. This manuscript offers an explicit and falsifiable specification with bounded documentary illustrations. Independent evaluation is necessary before claiming that it outperforms equally provisioned conventional audit systems. A DOI, when deposited, identifies a scholarly object; it does not certify scientific validity.

Declarations

Authorship. Juan Antonio Marín Liera (author-approved release candidate). AI-assisted editorial drafting, restructuring and compilation were used and should be disclosed in the deposited version.
Affiliation. System Friction Institute (author-declared affiliation; no external institutional accreditation implied).
Ethical scope. No new human-participant intervention or personal relational records were analyzed for this article. This documentary/methodological manuscript makes no human-subject research approval claim.
Data availability. The accompanying review package contains a structured index of the documentary observations and sources used here, a claims/status register and a methodological README. It does not include private raw operational records or a full independent replication dataset.
Funding and competing interests. Not declared in the materials inspected; absence of funding or conflicts is not presumed.
Publication state. Author-approved working paper; not peer reviewed; institutional promotion and DOI not claimed. **License.** Author-originated article text and newly prepared editorial diagrams offered under CC BY 4.0 upon authorized public release; third-party references and SFI marks are excluded.

References

[1] C. S. Peirce, "The Fixation of Belief," *Popular Science Monthly*, vol. 12, pp. 1–15, 1877. https://www.peirce.org/writings/p107.html.
[2] A. Rosenblueth, N. Wiener and J. Bigelow, "Behavior, Purpose and Teleology," *Philosophy of Science*, 10(1), 18–24, 1943. https://doi.org/10.1086/286788.
[3] E. Hollnagel, D. D. Woods and N. Leveson, eds., *Resilience Engineering: Concepts and Precepts*. Ashgate, 2006.
[4] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, 2023. https://doi.org/10.6028/NIST.AI.100-1.
[5] W3C, *PROV-O: The PROV Ontology*, W3C Recommendation, 2013. https://www.w3.org/TR/prov-o/.
[6] System Friction Institute, "Individuation as a Measurement Problem," Method Lab research object SFI-A, v0.4, consulted 16 November 2026. Internal object metadata; primary case records not redistributed.
[7] System Friction Institute, "From AI Governance to AI Observability," Method Lab research object SFI-B, v0.4, consulted 16 November 2026. Internal object metadata; primary case records not redistributed.
[8] System Friction Institute, "Knight Capital 2012 — real case with MIHM-T, LCI and retrospective reconstruction," research object SFI-KNIGHT-2012-TRIPLE-RESEARCH-001. https://github.com/Aptymok/SFI-RESEARCH-HUB/tree/main/05_RESEARCH_OBJECTS/SFI-KNIGHT-2012-TRIPLE-RESEARCH-001 (historical documentary snapshot, as described in the 12 October manuscript).