

Against Retrospective Certainty: Temporal Honesty and the Governance of Institutional Memory

Ilias Konstantinidis

`ilias.konstantinidis@ou.ac.uk`

Research Article

Keywords: temporal honesty, institutional memory, epistemic reconstructibility, provenance, recordkeeping, accountability

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11207500/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Institutional retrospective accounts can present past judgments as more coherent, confident, and unified than they were when made. Once outcomes are known, uncertainty can be compressed, dissent can recede, and provisional reasoning can be reconstructed as certainty. This paper treats such retrospective certainty as one downstream governance risk of institutional memory: when the contemporaneous epistemic record is incomplete, overwritten, or stripped of lineage, later accounts become harder to test against what the institution actually recorded at the time.

The article develops temporal honesty as a *pro tanto* norm for institutional memory: prior epistemic states—including evidence, uncertainty, conditional judgments, dissent, and revision—should remain recoverable and auditable over time. Its design implications concern preservation of the contemporaneous record on which later investigation and accountability depend; decision quality and institutional responsiveness remain separate problems.

Drawing on social epistemology, organizational memory, archival and recordkeeping theory, and provenance, the paper examines two deliberately heterogeneous diagnostic cases: the Iraq WMD inquiries, which expose limits in reconstructing an institutional decision record, and public evaluation documentation across successive AI systems, which exposes a narrower limit at the public audit surface. Their common question is which aspects of epistemic lineage remain recoverable at the relevant audit layer. The paper then proposes a bounded working specification—state recoverability, evidentiary provenance, revision lineage, attribution and access lineage, and governed exposure—through which institutional memory can support retrospective reconstruction while leaving substantive judgment to the relevant domain.

1 Introduction

The Iraq WMD inquiries provide a compact example. In 2004, the United Kingdom's Intelligence and Security Committee disclosed that its previous review had not received the full set of relevant JIC papers, despite having reported that it had examined the assessment record (Intelligence and Security Committee 2003, 2004).

The episode matters because an oversight body reconstructing the institution's prior position could not initially establish whether its documentary corpus was complete. Before evaluating a past judgment, an inquiry must first be able to recover the evidence, assessments, uncertainty, dissent, access, and revision history that constituted the contemporaneous epistemic record.

This paper argues that the preservation of those conditions is an epistemic governance problem in its own right. Institutions routinely make decisions under uncertainty, but they are later judged under conditions in which the outcome is already known. Tetlock and Gardner (2015) identify the familiar cognitive mechanism of hindsight bias—the tendency to reconstruct prior uncertainty as though the eventual outcome had been more foreseeable than it was. In institutional settings, however,

retrospective certainty is not produced by memory alone. It is also shaped by documents, databases, reporting chains, model outputs, briefing practices, access rules, and systems of record. If earlier judgments are overwritten, if confidence qualifiers become detached from conclusions, if dissent is not preserved as part of the decision lineage, or if later revisions silently replace earlier states, the surviving record may no longer provide an adequate check on retrospective narrative.

I use temporal honesty as a stipulative label for a pro tanto norm of institutional memory: where an institution makes a materially significant epistemic commitment, it should preserve enough of the contemporaneous evidential and decision context for an authorized later evaluator to reconstruct that commitment substantially as it stood at the relevant time. The term is introduced here rather than borrowed from archival science, provenance research, or social epistemology. 'Honesty' is structural: it concerns the relation between successive institutional states independently of actors' sincerity or correctness. Later states should not silently substitute for earlier ones in a way that makes the institution's past position appear cleaner, more unified, or more determinate than the surviving contemporaneous record warrants.

An institutional epistemic state is the analytic object whose recoverability temporal honesty protects. The phrase is stipulative and operational: it denotes the configuration at time T of institutionally attributable evidence or source material, interpretations, uncertainty, assumptions, dissent, review and access conditions, and authorized judgments that bear on the commitment under audit. The state may be distributed across records, versions, logs, and other traces rather than contained in a single document. This usage brackets debates in collective epistemology over group belief and group knowledge, which caution against reducing institutional epistemic attribution either to a fictional collective mind or to the simple aggregation of what individual members know (Lackey 2021).

The argument intersects several established literatures. Forecasting supplies hindsight, calibration, and accountable belief revision (Tetlock and Gardner 2015); Longino treats criticism and disagreement as structural conditions of objectivity (Longino 1990); organizational-memory scholarship analyzes acquisition, retention, and retrieval (Walsh and Ungson 1991); and archival and recordkeeping theory treats records as evidential objects whose meaning depends on contextual relations, provenance, and revision histories (Upward 1996; Duranti 1997; Anderson 2024; King 2024). Organizational sociology explains why hierarchy, secrecy, incentives, and sensemaking can prevent warnings from affecting action (Vaughan 1983, 1996; Weick 1995; Roe and Schulman 2008), while secure logging and computational provenance show how durable lineage can support accountability (Schneier and Kelsey 1999; Moreau and Missier 2013; Singh et al. 2019). The narrower question here is what must be preserved for a later auditor to reconstruct the epistemic state from which a decision emerged.

Epistemic reconstructibility is the graded property of a record environment by virtue of which a later evaluator, operating at a specified audit layer, can recover the relations needed to distinguish a materially significant contemporaneous epistemic state from later reconstruction. For a given audit question, reconstructibility is sufficient when the surviving record allows the evaluator to identify the institutionally

attributable judgment, the material source or evidential basis and qualifications relevant to that judgment, and the revisions or successor states necessary to establish how the position changed. An audit layer denotes an access and disclosure position—such as the institution's internal record, an authorized oversight record, or the public audit surface. Reconstructibility is analytically separate from public transparency, completeness of capture, and decision quality. Retrospective certainty, by contrast, is a downstream representational condition in which an earlier position is later presented as more settled, unified, or continuous than the contemporaneous evidence supports. It differs from hindsight bias: hindsight bias is a cognitive tendency associated with outcome knowledge, whereas retrospective certainty concerns the relation between a later representation and the surviving record. It can arise through ordinary compression, record loss, or selective survival independently of deliberate falsification. Loss of epistemic lineage increases vulnerability to that condition without by itself demonstrating that it occurred.

The cases serve a diagnostic function: they identify reconstructibility problems rather than test whether a different architecture would have produced better substantive outcomes. Their heterogeneity is intentional. Iraq exposes an institutional process later opened to investigation; the AI case tests what an external reader can reconstruct from public system documentation. The analytic aim is to identify recurring reconstructibility questions and assess whether Section 3's record properties map onto reconstructed lineage problems, residual reconstructibility gaps, or boundaries of public reconstructibility.

The architectural component is a working audit specification: prior states are not silently overwritten; observations remain distinguishable from interpretations; uncertainty, assumptions, and dissent remain attached to judgments; revisions preserve lineage; material review and access events remain attributable; and legitimate secrecy restricts exposure while preserving relevant record structure. These are auditable constraints for epistemic reconstructibility rather than a general-purpose decision architecture.

The contribution is threefold: it identifies retrospective certainty as a record-mediated governance risk that becomes harder to contest when epistemic reconstructibility is weak; distinguishes decision quality from recoverability of the epistemic state in which a decision was made; and specifies which relations among evidence, interpretation, uncertainty, dissent, attribution, access, and revision should remain auditable when institutional epistemic positions change over time.

2 Temporal Honesty as an Institutional Epistemic Requirement

This section grounds temporal honesty as an institutional epistemic requirement through institutional responsibility, the distinction between source material and interpretation, and non-erasure memory under legitimate constraints on access and forgetting. It then states the conceptual boundaries among the norm, its object, the property it seeks to preserve, and the downstream risk it is intended to make more contestable.

2.1 Epistemic responsibility in institutional contexts

Institutional epistemic commitments differ from ordinary individual avowals because organizational judgments operate through delegated authority, routinized procedures, continuity across personnel, and often asymmetrical consequences. In this article, an institutional epistemic commitment means an institutionally attributable judgment, assessment, or model-mediated conclusion that is sufficiently authorized, relied upon, or consequential to become an appropriate object of later accountability. It need not be unanimous and it need not correspond to what every member privately believes. A court opinion, intelligence assessment, corporate risk model, or public health advisory can function as such a commitment because it is institutionally issued or relied upon in ways that shape action and accountability.

Longino's social account of objectivity supplies one foundation for treating epistemic responsibility as structurally organized rather than exhausted by individual good faith. In *Science as Social Knowledge*, scientific communities are objective to the degree that they sustain transformative criticism through four conditions: recognized avenues for criticism; shared standards that critics can invoke; community responsiveness to criticism; and equality of intellectual authority among qualified practitioners (Longino 1990, pp 76–81). Its relevance here is that objectivity depends partly on whether a community is organized so that assumptions and inferences can be challenged and criticism can matter.

Longino's account concerns the organization of critical practice. The present extension is diachronic: if later evaluators are to determine whether criticism existed, whether it received uptake, which standards governed assessment, and whose judgments carried institutional authority, some traces of those relations must survive. Temporal honesty adds that record condition to the social organization of criticism.

Three working obligations adapt that account to institutional memory. First, institutional judgment should preserve the distinction among source material, assumptions, interpretation, and uncertainty so that later criticism can address the grounds and conditionality of the claim rather than only its final wording.

Second, institutions should preserve continuity across successive epistemic states. Organizational action persists across personnel turnover and decision cycles, so institutional memory cannot depend on individual recollection alone. When a past judgment is silently overwritten rather than linked to an explicit successor, later evaluators lose evidence of how the recorded understanding changed and accountability becomes more dependent on retrospective reconstruction.

Third, materially significant epistemic commitments should remain reviewable, reconstructible, and attributable. A later evaluator should be able to establish, to the extent permitted by legitimate access restrictions, what judgment was institutionally recorded, what material supported or qualified it, how criticism or dissent related to it, and how it was subsequently revised or superseded. Institutional

epistemic responsibility is therefore partly a governance problem of preserving relations through which past judgments can remain answerable across time.

2.2 Source/observation and interpretive layers: preserving the chain from record to judgment

The distinction here is analytical and recordkeeping-based, not metaphysical. 'Source/observation' denotes an item as it was available to the institutional process before the particular interpretation under audit; it does not denote an unmediated fact or a representation-free view of the world. Archival theory and computational provenance both make relations central, but in different technical vocabularies. Duranti treats the archival bond as the relationship linking a record to other records produced in the same activity (Duranti 1997, pp 215–217). Anderson (2024) sharpens a related distinction: provenance concerns origin and trajectory, while context encompasses the wider relationships surrounding creation, use, and interpretation. PROV-DM, by contrast, models entities, activities, and agents together with generation, usage, derivation, and responsibility relations (Moreau and Missier 2013). Temporal honesty therefore treats provenance as necessary but not exhaustive: later reconstructibility also depends on preserving the interpretive and institutional context in which recorded material acquired evidential significance.

The source/observation layer therefore consists of the inputs available to the institution before the particular interpretation under audit: source reports, measurements, documents, sensor outputs, imagery, transcripts, legal instruments, recorded events, and other evidentiary artifacts. These inputs are not assumed to be neutral, complete, or representation-free. A source report can be mistaken, a measurement can embed methodological choices, and a legal document can itself be institutionally constituted. The point is only to preserve what entered the reasoning process in the form in which it was available at time T.

The interpretive layer consists of claims constructed from those inputs: hypotheses, risk scores, scenario models, policy assumptions, assessments, and official narratives. Such judgments are necessarily selective and contingent on evidence, methods, categories, and institutional purposes. Their substantive quality remains a separate question; temporal honesty requires that the interpretive state remain distinguishable from the source material on which it relied.

Layer conflation occurs through organizational sensemaking rather than through a single institutional mind. Observations and source materials are categorized, circulated, summarized, weighted, and authorized across roles, documents, and hierarchy (Weick 1995; Vaughan 1996; Roe and Schulman 2008). Across those transformations, an assumption can lose its label as an assumption; a confidence qualifier can disappear from a briefing; a model output can be treated as though it were a direct observation; or a minority interpretation can become administratively invisible. The audit question is how a particular evidentiary input became an institutional judgment, who participated in that transformation, and which intermediate traces remain recoverable.

For temporal honesty, the implication is correspondingly modest: source/observation records and interpretive judgments should remain distinguishable, traceably connected, and versioned where either changes. Section 3 translates that requirement into record properties.

2.3 Non-erasure memory and the ethics of institutional forgetting

These demands pull in opposite directions. On one side, temporal honesty creates a defeasible reason to preserve materially significant past judgments—especially judgments made under uncertainty—in a form that remains attributable and temporally anchored. On the other, modern governance frameworks—including the GDPR (with Article 17 subject to public-interest archiving exceptions), national-security protocols, and privacy-by-design principles—can impose legitimate obligations to delete, redact, or compartmentalize information that should not remain permanently accessible (European Union 2016).

The resulting question is not whether institutions should ever forget, but how restriction or deletion can occur without making the surviving epistemic history materially misleading. Erasure conflicts with temporal honesty when it removes the relations needed to establish what judgment was institutionally recorded, how it was qualified, or how it changed. Legitimate redaction—removing personally identifiable information, sealing classified segments, or limiting downstream exposure—need not destroy the temporal record of what was institutionally recorded, accepted, assumed, or contested at the time.

Temporal honesty therefore seeks a bounded structural trace compatible with legitimate restriction: enough metadata and lineage to establish the existence, timing, authority, and relation of an epistemic state even when its content cannot be disclosed. Cryptographic commitments and other mechanisms are possible implementations; the normative requirement is implementation-independent.

In his discussion of psychological or mnemonic traces, Ricoeur contrasts memories that have been definitively erased with memories that may instead be "rendered inaccessible, unavailable" (Ricoeur 2004, p 416). The article uses that distinction as a limited analogy for separating preservation from availability in institutional governance. Records-continuum and mediated-recordkeeping approaches provide the directly relevant recordkeeping foundations for distinguishing capture, organization, access, and later evidential use (Upward 1996; Gibbons 2026).

2.4 Conceptual boundaries: norm, object, property, and risk

The central vocabulary separates four functions. Temporal honesty is the normative requirement. The institutional epistemic state is the analytic object to be reconstructed. Epistemic reconstructibility is the graded property of a record environment that makes such reconstruction possible at a specified audit layer. Retrospective certainty is a downstream risk in later representation. A record can therefore be highly reconstructible even if the judgment it preserves was wrong, and a poorly reconstructible record need not prove that anyone later falsified the past.

Provenance provides an enabling family of relational concepts. In archival science, provenance and the archival bond concern origin, context, and relationships among records; in computational systems, data provenance concerns derivation and processing lineage; decision provenance extends lineage to decision pipelines and organizational context. Temporal honesty uses these resources to specify which epistemic relations should remain recoverable when an institutional judgment changes over time.

Non-overwrite and versioning are implementation-level record properties that can support the norm. Non-erasure memory remains defeasible: legitimate deletion, redaction, secrecy, and restricted access may limit what survives or who can inspect it. Temporal honesty asks whether, subject to those constraints, enough of the existence, status, provenance, qualification, and succession of a materially significant commitment remains reconstructible for the relevant form of accountability.

Materiality is the scope condition that keeps the norm from becoming a requirement to preserve every epistemic event. In this article, an epistemic commitment is materially significant when one or more features of its institutional role make loss of its contemporaneous lineage consequential for later accountability: the authority attached to the commitment, the extent to which others are expected to rely on it, the consequences of action or error, and the foreseeability or importance of later scrutiny. These factors are domain-sensitive rather than a universal numerical threshold. Different domains may operationalize them differently, but the applicable criteria and their use should themselves be reviewable rather than reconstructed only after outcomes are known.

3 From Temporal Honesty to a Recoverable Epistemic Record

This section translates the preceding norm into a bounded, implementation-agnostic set of record properties: contemporaneous source material and interpretation, uncertainty and material dissent, attribution and access, and explicit revision or supersession. Append-only or tamper-evident systems are possible means, but the governing property is recoverability rather than any particular architecture.

3.1 Non-overwrite and versioned epistemic states

The first requirement is that prior analytical states be treated as primary objects of record rather than as transient stages on the way to a final answer. A judgment at time T may include a claim, the evidence to which it refers, an uncertainty or confidence characterization, stated assumptions, and the identity or role of the actor responsible for the judgment. If that judgment changes at $T + 1$, the later state should identify what changed and link the revision to the state it supersedes rather than silently replacing it. Archival description provides a concrete analogue: King (2024) argues that evolving finding aids require meta-metadata about authorship, earlier versions, and revision history because descriptions that conceal their own development can appear final and disembodied.

Correction remains compatible with this requirement: mistaken or obsolete claims may be revised or withdrawn through linked supersession.

3.2 Preserving uncertainty, dissent, and provenance

A recoverable record must preserve more than the proposition an institution ultimately adopted. One route to retrospective certainty is compression: qualifications disappear while the headline conclusion survives. Temporal honesty therefore requires epistemic qualifications to remain attached to the judgments they qualify. Where an institution uses numerical confidence, probability ranges, or formal uncertainty measures, those values should be preserved with the relevant state. Where uncertainty is expressed qualitatively, the record should preserve the language and conditions under which the judgment was made rather than retrospectively translating it into a sharper claim.

The same principle applies to assumptions and dissent. A minority assessment is not adequately preserved if only the final majority conclusion remains in the institutional record. Nor is a model assumption preserved merely because the final model output can later be reproduced. Assumptions, caveats, alternative interpretations, and recorded objections should remain identifiable as parts of the contemporaneous epistemic state, together with their relation to the judgment that prevailed. Materiality limits durable preservation to disagreements that meet explicit criteria whose application can itself be reviewed.

Computational provenance can supply part of this evidentiary lineage by representing derivation, use, generation, and responsibility among recorded entities, activities, and agents (Moreau and Missier 2013). Here it supports traceability of captured transformations; archival provenance and substantive evaluation of the inference remain distinct tasks.

3.3 Decision lineage, attribution, and controlled access

Because institutional judgments are produced through roles and information flows, the record should preserve material review and authorization events. Across time, an auditor should be able to determine who authored, reviewed, approved, rejected, or superseded an epistemically significant state. Attribution serves an evidentiary function by identifying the institutional actors and roles connected to a recorded judgment; responsibility remains a separate evaluative question.

Access is part of that lineage. In secrecy-sensitive institutions, the question is not simply whether a document existed but who was entitled to examine it when a judgment was formed. A record architecture should therefore preserve access classifications and material distribution or review decisions: which roles were authorized to see an item, which relevant reviewers were excluded by compartmentalization, and under what rule or decision that restriction operated. Where systems reliably record actual retrieval events, those events may add further evidence, but an access log should not be mistaken for proof that information was cognitively understood or acted upon.

Controlled auditability permits redaction or delayed access where authorized oversight can still establish the existence and lineage of the relevant state.

3.4 A working audit specification

For this article, these principles form a bounded working audit specification with five elements: (1) state recoverability—for a judgment classified as epistemically material under the institution's applicable record rules, an authorized auditor should be able to recover the state as it existed at the relevant time, including its qualifications; (2) evidentiary provenance—the sources or observations supporting it should remain identifiable without being replaced by later interpretations; (3) revision lineage—changes should create linked successor states rather than silent modification; (4) attribution and access lineage—material authorship, review, authorization, and access restrictions should remain reconstructible; and (5) governed exposure—confidentiality or redaction may restrict content while preserving enough structure to establish that a state existed and how it related to successors. Materiality rules are themselves part of the governance layer addressed in Section 6.3.

Where analytical artifacts must be registered, a durable manifest can let later inquiry test whether the corpus supplied for review matches the institution's own formal record. Its reach stops at that formal corpus: unregistered documents, conversations, or observations remain outside the test, so record integrity after capture is easier to secure than completeness of capture.

These properties preserve the conditions for later evaluation; truth, inferential justification, and substantive correctness remain separate questions.

4 Evidentiary Gaps and Retrospective Reconstruction

The cases ask a retrospective question at two different audit layers: what can the record available to a relevant auditor establish about the epistemic state in which a consequential judgment was formed or assessed? Their heterogeneity is intentional: inference rests on the shared reconstructibility question rather than on sectoral or technological comparability. They test whether enough of the contemporaneous lineage survives at the relevant audit layer to compare later accounts with what was recorded at the time. An internal record layer, an authorized oversight layer, and a public audit surface can preserve different amounts of information without being equivalent.

The diagnostic target is epistemic reconstructibility. At an institutional or authorized-oversight layer, the question is whether contemporaneous lineage was directly preserved or had to be recovered later; at the public layer, the question concerns what an outside auditor can reconstruct from disclosed material. The counterfactual asks whether the record properties specified in Section 3 would have made that audit question more determinate at the relevant layer, subject to the capture boundary already identified there.

4.1 What counts as a temporal-honesty record problem?

At the institutional-record layer, this article distinguishes two outcomes. A reconstructed lineage problem exists when a material contemporaneous relation was not directly recoverable from the record initially available to oversight and had to be established through supplementary disclosure, testimony, cross-document comparison, or other later evidentiary work. A residual reconstructibility gap exists when, after the available inquiry, a material relation remains incomplete or indeterminate because the

surviving record does not preserve it. Both can reveal weakness in how contemporaneous lineage was captured, organized, or surfaced to oversight; the residual gap denotes the stronger condition in which later reconstructibility remains incomplete.

Either category is relevant here when four conditions hold. The retrospective question must be material to accountability or learning; it must concern a contemporaneous epistemic state or lineage; the difficulty must be evidenced in the surviving record or in the work later investigators had to perform; and a temporal-honesty requirement must bear directly on the relation at issue—for example, a document manifest on corpus completeness, version history on supersession, or access lineage on who could inspect or authorize a claim. A rich archive can therefore contain reconstructed lineage problems as well as residual reconstructibility gaps. The analytical question is whether the contemporaneous record preserved the relevant relation directly and, where it did not, whether later reconstruction ultimately resolved the uncertainty. The AI case marks a different category: a boundary of public reconstructibility.

4.2 Iraq WMD: incomplete records, dissent, and access

The Iraq WMD record allows the distinctions in Section 4.1 to be applied to documentary completeness, artifact integrity, dissent, access, and review lineage.

The first problem is corpus completeness. In its September 2003 report, the Intelligence and Security Committee stated that it had examined all JIC assessments produced since Iraq's invasion of Kuwait in 1990 (Intelligence and Security Committee 2003, para. 16). In its 2003–2004 Annual Report, however, the Committee reported that this statement had proved erroneous: a subsequent check identified eight additional JIC papers relating to Iraqi WMD and UN inspections that had not been provided during the original inquiry. The Committee had previously reviewed 93 assessments; five of the newly supplied papers dated from the period before September 2002 and three from October 2002 to March 2003. Importantly, the Committee concluded that the additional papers would not have changed its substantive findings. Its concern was instead institutional: it identified defective internal systems and record-keeping within the Cabinet Office's Intelligence and Security Secretariat and subsequently began receiving regular lists of JIC papers (Intelligence and Security Committee 2004, paras. 145–147).

The temporal-honesty question precedes substantive evaluation: could oversight establish that the corpus supplied was complete relative to the institution's own formal record? A durable manifest would allow the supplied corpus to be checked against the contemporaneous list of registered assessments while leaving unregistered discussion or papers outside that test. Because the omitted papers were later identified and supplied, this episode is best classified as a reconstructed lineage problem concerning corpus completeness rather than a residual reconstructibility gap. Its value is documentary determinacy rather than counterfactual policy improvement.

The later Iraq Inquiry also identified a distinct record-integrity limitation. Chilcot reported that, although the Inquiry had examined most original and unredacted documents relevant to the production of the September 2002 dossier, the Government was unable to provide original copies in a small number of

cases, primarily emails previously supplied to the Hutton Inquiry (Iraq Inquiry 2016, vol. IV, sec. 4.2, para. 3). A subsequent inquiry can compensate through copies, testimony, and cross-document comparison, while loss of an original may remove metadata, version context, or authenticity evidence that no longer survives. The Inquiry identified no material epistemic relation that remained unresolved specifically because those originals were unavailable, so the episode is best treated as a record-integrity limitation rather than a residual reconstructibility gap.

A different kind of episode concerns dissent and escalation lineage. During its 2003 investigation, the ISC learned that two members of the Defence Intelligence Staff had written to their line managers on 19 and 20 September 2002 expressing concerns about language in the draft dossier. The letters had been copied to the then Deputy Chief of Defence Intelligence, but not to the Defence Secretary or the JIC Chairman. Before a second evidence session, the Defence Secretary supplied the ISC with copies of the original letters, enabling the Committee to reconstruct how the concerns had been handled. The ISC nevertheless criticized the initial failure to disclose that written concerns existed as 'unhelpful and potentially misleading' and recommended that formally written concerns about JIC assessments be brought to the attention of the JIC Chairman (Intelligence and Security Committee 2003, paras. 101–105, 114–116).

This episode is likewise a reconstructed lineage problem rather than a residual reconstructibility gap. The surviving material and further disclosure allowed the ISC to establish that the dissent existed and where its formal route stopped. Its relevance is diagnostic because it demonstrates why existence, recording, disclosure to oversight, and escalation are distinct audit events. A temporal-honesty record would preserve those relations contemporaneously rather than requiring a later inquiry to recover them through additional disclosure and testimony.

The same category applies to access and review lineage. The Butler Review concluded that the protection of sensitive human-intelligence sources had, in one important instance, produced unduly strict compartmentalization. A report significant to the drafting of the September dossier was not shown to key DIS experts who could have assessed its validity and credibility; Butler concluded that source protection should not prevent technical reporting from being exposed to the most expert available analysis (Butler Review 2004, para. 24). Butler was able to establish this restriction, so the episode is a reconstructed lineage problem, not evidence that the access history remained unknowable. Its value for the present argument is that access itself formed part of the epistemic lineage: a later audit needed to establish who was in a position to evaluate a material source and which exclusions followed from deliberate access rules.

By contrast, Chilcot records a residual reconstructibility gap in the review lineage of the dossier's Foreword. Sir John Scarlett later wrote that he and 'some JIC members' had commented on the draft, but the Iraq Inquiry stated that it had not seen comments from any JIC members other than Sir David Omand and Simon Webb (Iraq Inquiry 2016, vol. IV, sec. 4.2, paras. 529–537). The surviving record therefore does not provide a complete documentary account of which additional JIC members commented, what

they proposed, or how any such interventions related to the final text. A versioned review history could preserve the formal comment and revision lineage through which an institutional document acquired its final form.

The Iraq material thus supports three distinct claims that should not be conflated. The omitted JIC papers, DIS dissent, and Butler compartmentalization are reconstructed lineage problems: the relevant relations were not all directly available to oversight at first, but later inquiry substantially established them. The unavailable originals are a record-integrity limitation: some contemporaneous artifacts did not survive in original form, without the Inquiry demonstrating that their absence left a material epistemic relation unresolved. The incomplete Foreword comment history is the residual reconstructibility gap: aspects of the formal review lineage remained indeterminate even after investigation. Together, the episodes raise recurring questions about corpus completeness, dissent routing, access and review, artifact integrity, and revision lineage. Temporal honesty's forensic value lies in preserving those relationships more directly, reducing the later burden of reconstruction, and making genuine residual gaps easier to identify.

4.3 AI evaluation across model generations: a public-reconstructibility boundary case

The second case concerns the public audit surface. Model reports and system cards are selective publications produced during or around deployment, and their purposes, scopes, and evaluation suites can change as systems and risks change. The question is whether, when an institution publicly reports an evaluation at one model generation, an external reader can later determine from successive public documentation whether that evaluation was repeated, modified, superseded, or retired.

OpenAI's 2023 GPT-4 Technical Report provides a useful starting point because it makes factual reliability an explicit object of evaluation. The report states that GPT-4 remained capable of hallucinating facts and making reasoning errors, and reports that GPT-4 scored 19 percentage points higher than the latest GPT-3.5 model on nine internally developed, adversarial factuality evaluations covering learning, technology, writing, history, mathematics, science, recommendation, code, and business (OpenAI 2023, sec. 5, Fig. 6). The same report separately presents performance on TruthfulQA, a public benchmark designed to test resistance to plausible but false answers (OpenAI 2023, sec. 5, Fig. 7). These provide two identifiable contemporaneous evaluation reference points: one tied to an internal nine-category factuality suite and one tied to a public benchmark.

The 2024 GPT-4o System Card documents a different evaluation portfolio. Its central emphasis is on risks introduced or altered by the model's multimodal and speech-to-speech capabilities, although it also reports text and vision evaluations (OpenAI 2024). In its section on historically underrepresented languages, the card publishes zero-shot results on translated versions of ARC-Easy and TruthfulQA, alongside English reference results, for GPT-3.5 Turbo, GPT-4o mini, GPT-4, and GPT-4o. TruthfulQA therefore has a visible successor measurement across GPT-4 and GPT-4o. By contrast, the card does not

report a result for the nine-category internal adversarial factuality evaluation highlighted in the GPT-4 Technical Report or identify that earlier suite as rerun, modified, superseded, or retired.

The later public record makes the longitudinal issue clearer rather than resolving it. The GPT-4.5 System Card evaluates hallucinations with PersonQA and compares GPT-4.5 with GPT-4o and o1 (OpenAI 2025a). The GPT-5 System Card shifts again, reporting factuality on prompts representative of ChatGPT production conversations and adding LongFact, FActScore, and SimpleQA; it explicitly describes the open-ended evaluations as new additions (OpenAI 2025b). The GPT-5.2 update continues production-style hallucination measurement (OpenAI 2025c), while the GPT-5.5 Instant and August 2026 GPT-5.6 documentation use challenging factuality-heavy, user-flagged-failure, and high-stakes prompt sets intended to provide a sensitive research signal over time (OpenAI 2026a, 2026b). This sequence shows sustained public attention to factuality while the evaluation regime itself changes.

Across the later public documentation reviewed here, however, the 2023 nine-category adversarial factuality suite is not given an explicit successor relation. The later cards do not state that it was rerun, revised into one of the newer factuality evaluations, superseded by them, or retired. The distinction matters because continuity of a risk domain is not the same as continuity of an evaluation object. The same August 2026 card also cautions, in discussing cross-card evaluation comparisons more generally, that policies, graders, datasets, evaluations, and other measurement details evolve over time and that scores from earlier system cards not included in a current comparison should generally not be treated as directly comparable (OpenAI 2026b). The public record therefore supports a bounded audit question: how did one previously reported evaluation relate to the later factuality regime?

A temporal-honesty approach can distinguish this question across audit layers. At the internal record layer, an evaluation could retain a stable identifier and version, model snapshot, prompt set, scoring procedure, and status such as repeated, revised, replaced, or retired. An authorized oversight layer could expose enough of that lineage to verify continuity without revealing sensitive test items. At the public audit surface, a system card could, where disclosure is appropriate, identify a successor relation even when methods or benchmarks change. The layers need not contain identical content. The governance value lies in preserving the distinction between an evaluation that ceased to exist, one that changed form, and one whose later status is simply not publicly disclosed.

The inference is confined to the public audit surface. Successive public documents show continued factuality evaluation without preserving a public predecessor-successor lineage for the earlier nine-category suite. Questions about internal preservation, benchmark quality, and disclosure obligations lie beyond what this comparison establishes. The AI case therefore complements Iraq at a different audit layer.

5 Intellectual Foundations and the Scope of the Contribution

The account draws on four adjacent traditions: organizational memory, archival and records-continuum theory, provenance, and forensic recording. This section locates the proposed norm within those

traditions and identifies the narrower diachronic audit object that distinguishes it.

5.1 Organizational memory and recordkeeping

The language of institutional memory has a substantial organizational antecedent. Walsh and Ungson define organizational memory as stored information from an organization's history that can be brought to bear on present decisions and organize their account around information acquisition, retention, and retrieval (Walsh and Ungson 1991, pp 61–63). Their retention model is explicitly distributed: five internal retention facilities are supplemented by external archives rather than a single organizational repository (pp 63–69). This article adopts that process orientation and uses 'institutional memory' for the narrower governance problem of whether the evidence needed to reconstruct an earlier institutionally attributable epistemic position remains available across organizational change.

Archival and recordkeeping theory supplies a more precise relational foundation. Upward presents the records continuum as a time/space model in which creation, capture, organization, and pluralization are dimensions rather than sequential stages (Upward 1996, pp 268, 280–281). Duranti's archival bond identifies the internal relations among records produced in the same activity as part of what gives a record evidential meaning (Duranti 1997, pp 215–217). Anderson (2024) complements this by distinguishing provenance from the wider contextual relationships surrounding a record's creation, use, and interpretation. Hurley et al. (2024) then broaden provenance through a records-continuum account of multiple, simultaneous, and parallel relations situated in wider social and organizational ambience. Across these approaches, the common point is not that provenance and context are interchangeable, but that what a record can evidence depends on preserved relationships rather than payload alone.

Gibbons (2026) moves closer to the present problem by linking mediated recordkeeping to epistemic accountability: archives should be able to clarify what their collections can evidence and how socio-technical conditions shape that evidentiality. King (2024) provides a complementary precedent at the level of archival description. Because finding aids are interpretive artifacts that evolve, he argues for preserving revision history, earlier versions, and authorship as meta-metadata; without those traces, descriptions can appear more final and disembodied than their production history warrants. Temporal honesty extends this archival logic to institutionally attributable epistemic commitments. Its audit object is the changing relation among source material, interpretation, uncertainty, dissent, authorization, access, and revision through which an institutional judgment becomes reconstructible across time.

5.2 From data provenance to decision provenance

The term provenance has a different technical emphasis in data systems. The W3C PROV-DM recommendation defines provenance as information about the entities, activities, and people involved in producing data or another thing, and formalizes relations including generation, usage, derivation, and responsibility (Moreau and Missier 2013). For the present argument, its important contribution is a vocabulary for represented lineage. Secure logging adds a distinct integrity problem: mechanisms for making later alteration of recorded events more detectable (Schneier and Kelsey 1999). Computational

provenance and archival provenance remain distinct relational traditions, both relevant to reconstructibility.

Singh, Cobbe, and Norval's concept of decision provenance moves still closer to the present argument. They propose using provenance methods to expose decision pipelines: chains of inputs to decisions, the nature of decisions and actions taken at design and run time, and their flow-on effects throughout interconnected systems (Singh et al. 2019, p 6562). Their aim is explicitly accountability-oriented: decision provenance is proposed as a means of facilitating oversight, audit, compliance, risk mitigation, and user empowerment across technical and organizational boundaries. It therefore already rejects the idea that accountability can be achieved by inspecting an isolated algorithm or final output alone.

Within this literature, temporal honesty functions as a normative specialization of accountability-oriented provenance logic. Decision provenance reconstructs pipelines across technical and organizational boundaries; temporal honesty identifies a narrower diachronic audit object within such a pipeline: the institutional epistemic state from which a judgment was formed, qualified, reviewed, authorized, and later revised. Provenance systems can represent much of this context (Singh et al. 2019); the proposed contribution is to specify which relations matter when the accountability question is whether an earlier institutional position remains reconstructible without silent substitution by its successor.

5.3 The black-box analogy and its limits

The black-box analogy invoked by the present argument also has a direct precedent in work on autonomous systems. Winfield and Jirotko propose an 'ethical black box' for robots and autonomous systems: the equivalent of a flight data recorder that continuously records sensor and relevant internal-status data, intended to help investigators discover why and how an accident occurred and to support the establishment of accountability and responsibility (Winfield and Jirotko 2017, pp 262–263). The present article borrows its forensic function: the survival of a contemporaneous operational trace for later investigation.

That proposal itself draws on an established logic of accident investigation. The US National Transportation Safety Board describes cockpit voice and flight data recorders as devices installed to help reconstruct the events leading to an aircraft accident (NTSB n.d.). ICAO's accident-investigation framework similarly requires effective use of flight recorders in investigations and defines the objective of safety investigation as the prevention of future accidents and incidents rather than the apportionment of blame or liability (ICAO 2024). In each case, the recorder matters because post-event testimony and surviving physical evidence may be incomplete. A durable contemporaneous trace changes what investigators can subsequently know.

Temporal honesty changes the recorded object from an operational to an epistemic trajectory: evidence, interpretations, confidence, assumptions, dissent, review, access, and revision distributed across actors and a longer decision lifecycle. The question becomes not only what a system did, but what was

institutionally recorded or attributable at the time and how that configuration became the state from which action was authorized.

The analogy stops at that forensic function. Unlike flight-data telemetry, institutional epistemic records depend on materiality rules, reporting practices, classification, and actor compliance. Temporal honesty therefore governs lineage after capture and the reviewability of capture rules, while capture completeness and substantive inferential justification remain separate questions. The useful commonality is that contemporaneous traces reduce later dependence on selective recollection and retrospective reconstruction.

Taken together, these traditions provide the components on which temporal honesty relies: relational record context, represented lineage, decision-pipeline provenance, and forensic preservation of contemporaneous traces. The narrower addition is to treat evidence, interpretation, uncertainty and dissent, authorized judgment, and revision or supersession as a time-indexed institutional epistemic trajectory whose reconstructibility is itself a governance concern.

6 Organizational Limits and Governance Conditions

This section separates record integrity, capture completeness, substantive adequacy, and organizational responsiveness. Temporal honesty addresses the first directly and provides governance hooks for the others without turning into a general theory of organizational rationality.

6.1 Record integrity, capture completeness, and substantive adequacy

Record integrity differs from completeness of capture. Once an epistemically material state enters a well-designed record, non-overwrite, provenance, and version lineage can make later alteration more detectable. That guarantee begins after capture: actors may still omit, delay, move reasoning to informal channels, or misclassify relevant observations, conversations, drafts, dissent, or judgments. Strategic under-classification therefore remains a governance risk.

Completeness therefore depends on complementary rules: what must be registered, when, who may classify material as outside the record, what reasons must accompany exclusion, and what independent oversight tests omission or systematic under-classification. A durable manifest tests the completeness of the formal corpus, while unregistered activity remains beyond that test. Temporal honesty can make capture and classification rules auditable while leaving the capture boundary visible.

Substantive adequacy is different again. A perfectly preserved judgment may rest on false evidence, invalid inference, poor calibration, or an inappropriate model. Preservation exposes the contemporaneous judgment to criticism; substantive validity still depends on domain expertise and evidentiary standards.

6.2 Power, incentives, and ignored warnings

The second limit is organizational rather than technical. Better records leave hierarchy, incentives, political commitment, and professional culture in place. The literature on organizational decision-making repeatedly shows that signals can be available and still fail to redirect action. Vaughan's analyses of corporate misconduct and the Challenger launch decision show how reporting structures, normalization, and organizational categories can render warning signals ineffective without requiring their literal disappearance (Vaughan 1983, 1996). Sensemaking and high-reliability approaches likewise emphasize the interpretive capacities through which organizations notice, classify, and respond to ambiguous information under pressure (Weick 1995; Roe and Schulman 2008).

The Iraq case illustrates the same limit: preserving dissent, access history, and revision lineage would still leave substantive weighting to decision-makers. Reconstruction and responsiveness are separate capacities; protected dissent channels, escalation, independent review, and incentives to reconsider settled positions belong to a wider governance problem. A recoverable record can expose failures of responsiveness more clearly, while prevention depends on those wider institutional arrangements.

6.3 Governing the record: materiality, access, and oversight

Temporal honesty therefore requires governance of the record itself: how the materiality factors identified in Section 2.4 are operationalized, what metadata and lineage accompany a material state, when revision creates a new state, how long elements are retained, and who may inspect them. Materiality cannot be an unreviewable label. Inclusion and exclusion criteria should be explicit, consistently applicable, versioned when changed, attributable, and contestable through appropriate oversight; otherwise changes in capture rules can silently change what future investigators can reconstruct.

Capture, materiality, and retention rules should themselves be historically recoverable. An auditor should be able to identify which rules applied at a given time, while independent or functionally separate oversight should be able to sample exclusions to test whether labels such as 'immaterial', 'informal', or 'out of scope' are used systematically to route inconvenient evidence or dissent around the record. This cannot guarantee complete capture, but it makes strategic under-classification more detectable.

Access governance poses a parallel problem. Secrecy and compartmentalization may be legitimate, but they can also determine which reviewers are able to test a source or challenge an interpretation. Temporal honesty is compatible with restricted disclosure so long as appropriate oversight can establish the existence, provenance, status, and lineage of the relevant epistemic record and, where material, the rules that governed access.

6.4 Legitimate forgetting and the costs of durable memory

As Section 2.3 argued, preservation and exposure are distinct. Privacy, confidentiality, security, and legal duties may require deletion, redaction, sealing, or restricted access; temporal honesty therefore asks what lawful structural trace, if any, can survive without making the remaining epistemic history materially misleading (Ricoeur 2004, p 416; European Union 2016).

Finally, durable epistemic memory has operational costs. Versioned records grow; richer provenance and access histories require storage and administration; registration requirements create friction; and review and oversight consume attention. Additional governance machinery can itself add complexity and new failure points. Temporal honesty therefore rests on proportional preservation: its scope should track the stakes, reversibility, and accountability demands of the decision environment rather than maximal recording.

7 Conclusion

Temporal honesty addresses a specific vulnerability of institutional accountability: substantive error can be compounded by loss of the contemporaneous epistemic record needed to reconstruct how a judgment was formed, qualified, challenged, and revised. Preserving that record keeps later narratives answerable to the evidence of the institution's earlier epistemic position.

The two cases illuminate different audit layers. The Iraq inquiries encountered reconstructed lineage problems, a record-integrity limitation, and a residual reconstructibility gap within an institutional process later opened to investigation. The AI case shows a public audit surface on which successive reports can remain individually informative while leaving the successor status of an earlier evaluation unspecified. Together, they show that epistemic reconstructibility depends on the access and disclosure position from which an audit is conducted.

The resulting scope is precise: record integrity, capture completeness, substantive correctness, organizational responsiveness, and public disclosure are related but distinct governance dimensions. Temporal honesty concerns the preservation of epistemic lineage across them, subject to privacy, secrecy, cost, and legitimate forgetting.

Against retrospective certainty, temporal honesty asks institutions to preserve the difference between what is known now and what the contemporaneous record could support then. It specifies the epistemic relations that should remain reconstructible when materially significant institutional judgments are formed, qualified, challenged, authorized, and revised, so that later accounts can be tested against the surviving epistemic history rather than silently substituted for it.

Use of generative AI

Generative AI tools were used during manuscript preparation for drafting support, structural and language revision, and source-checking assistance. All arguments, source selection, interpretations, and final text were reviewed and approved by the author, who assumes full responsibility for the manuscript.

Declarations

Author Contribution

I.K. conceived the study, developed the conceptual framework and methodology, conducted the literature and source analysis, analyzed the case materials, and wrote and revised the manuscript.

References

1. Anderson BG (2024) Kindred contexts: archives, archaeology, and the concept of provenance. *Arch Sci* 24:761–781. <https://doi.org/10.1007/s10502-024-09459-5>
2. Butler R (2004) Review of Intelligence on Weapons of Mass Destruction: Report of a Committee of Privy Counsellors. HC 898. The Stationery Office, London
3. Duranti L (1997) The archival bond. *Arch Mus Inf* 11:213–218. <https://doi.org/10.1023/A:1009025127463>
4. European Union (2016) Off J Eur Union L119:1–88. <https://eur-lex.europa.eu/eli/reg/2016/679/oj> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation)
5. Gibbons L (2026) Mediated recordkeeping and epistemic accountability in social media archiving. *Arch Sci* 26:29. <https://doi.org/10.1007/s10502-026-09550-z>
6. Hurley C, McKemmish S, Reed B, Timberly N (2024) The power of provenance in the records continuum. *Arch Sci* 24:825–845. <https://doi.org/10.1007/s10502-024-09463-9>
7. Intelligence and Security Committee (2003) Iraqi Weapons of Mass Destruction—Intelligence and Assessments. Cm 5972. The Stationery Office, London
8. Intelligence and Security Committee (2004) Annual Report 2003–2004. Cm 6240. The Stationery Office, London
9. International Civil Aviation Organization (ICAO) (2024) Annex 13 to the Convention on International Civil Aviation: Aircraft Accident and Incident Investigation, 13th edn. ICAO, Montreal
10. Iraq Inquiry (2016) The Report of the Iraq Inquiry, vol IV. HC 265-IV. The Stationery Office, London
11. King OC (2024) Archival meta-metadata: revision history and positionality of finding aids. *Arch Sci* 24:509–529. <https://doi.org/10.1007/s10502-024-09443-z>
12. Lackey J (2021) The epistemology of groups. Oxford University Press, Oxford. <https://doi.org/10.1093/oso/9780199656608.001.0001>
13. Longino HE (1990) Science as social knowledge: values and objectivity in scientific inquiry. Princeton University Press, Princeton
14. Moreau L, Missier P (2013) PROV-DM: The PROV Data Model. W3C Recommendation, 30 April 2013. <https://www.w3.org/TR/prov-dm/>

15. National Transportation Safety Board (NTSB) (n.d.) Cockpit Voice Recorders (CVR) and Flight Data Recorders (FDR). https://www.nts.gov/investigations/process/Pages/cvf_fdr.aspx. Accessed 28 Aug 2026
16. OpenAI (2023) GPT-4 technical report. arXiv:2303.08774. <https://doi.org/10.48550/arXiv.2303.08774>
17. OpenAI (2024) GPT-4o system card. arXiv:2410.21276. <https://doi.org/10.48550/arXiv.2410.21276>
18. OpenAI (2025a) OpenAI GPT-4.5 system card. 27 February 2025. <https://openai.com/index/gpt-4-5-system-card/>
19. OpenAI (2025b) GPT-5 system card. 7 August 2025. <https://openai.com/index/gpt-5-system-card/>
20. OpenAI (2025c) Update to GPT-5 system card: GPT-5.2. 11 December 2025. <https://deploymentsafety.openai.com/gpt-5-2>
21. OpenAI (2026a) GPT-5.5 Instant system card. 5 May 2026. <https://deploymentsafety.openai.com/gpt-5-5-instant>
22. OpenAI (2026b) GPT-5.6—August updates. 6 August 2026. <https://deploymentsafety.openai.com/gpt-5-6-august-update>
23. Ricoeur P (2004) *Memory, history, forgetting*. University of Chicago Press, Chicago
24. Roe E, Schulman PR (2008) *High reliability management: operating on the edge*. Stanford University Press, Stanford
25. Schneier B, Kelsey J (1999) Secure audit logs to support computer forensics. *ACM Trans Inf Syst Secur* 2(2):159–176. <https://doi.org/10.1145/317087.317089>
26. Singh J, Cobbe J, Norval C (2019) Decision provenance: harnessing data flow for accountable systems. *IEEE Access* 7:6562–6574. <https://doi.org/10.1109/ACCESS.2018.2887201>
27. Tetlock PE, Gardner D (2015) *Superforecasting: the art and science of prediction*. Crown, New York
28. Upward F (1996) Structuring the records continuum part one: post-custodial principles and properties. *Arch Manuscr* 24(2):268–285
29. Vaughan D (1983) *Controlling unlawful organizational behavior: social structure and corporate misconduct*. University of Chicago Press, Chicago
30. Vaughan D (1996) *The Challenger launch decision: risky technology, culture, and deviance at NASA*. University of Chicago Press, Chicago
31. Walsh JP, Ungson GR (1991) Organizational memory. *Acad Manag Rev* 16(1):57–91. <https://doi.org/10.5465/amr.1991.4278992>
32. Weick KE (1995) *Sensemaking in organizations*. Sage, Thousand Oaks
33. Winfield AFT, Jirotko M (2017) The case for an ethical black box. In: Gao Y, Fallah S, Jin Y, Lekakou C (eds) *Towards autonomous robotic systems*. Lecture Notes in Computer Science, vol 10454. Springer, Cham, pp 262–273. https://doi.org/10.1007/978-3-319-64107-2_21