

These observations establish two requirements for a public infrastructure. First, the design must identify the institutions and work on which each preservation promise depends. A statement that evidence will remain available requires a custodian, resources, and an intelligible succession arrangement. Second, the system must distinguish the commonality of an interface from the uniformity of practices behind it. A museum accession, a scientific correction, and an administrative finding can share mechanisms for identifying records and documenting changes while retaining different standards of evaluation. Their interoperability depends on explicit translation and qualified equivalence. Treating their local statuses as interchangeable would conceal the institutional work through which those statuses acquired meaning.

The resulting conception is relational in a precise sense: an infrastructure supports specified activities for specified participants under historical and organizational conditions. This conception gives practical content to the project's ambition to connect existing institutions. It also constrains that ambition. A system's ability to exchange records establishes only one condition of public usefulness. Further inquiry must examine whose activities become easier, whose labor increases, and whose ways of producing knowledge become harder to express.

2.2 Social Facts, Representations, and Institutional Recognition

The term *social fact* requires conceptual discipline because an infrastructure can preserve descriptions of social life while also participating in the practices that establish institutional statuses. The following distinctions organize these relationships without settling the competing metaphysical accounts of society. They separate material occurrences, evidence, representations, institutional recognition, and legal status according to their roles in an inquiry. The distinctions can apply to overlapping objects and should never be read as a universal sequence in which a fact becomes more real each time it passes an institutional gate.

A material occurrence concerns what happened: a meeting took place, a measurement was produced, or a person performed work. Evidence comprises materials and testimony used to support an interpretation of that occurrence. A representation expresses an account through selected concepts, boundaries, and descriptions. Institutional recognition concerns the acceptance or classification of an account by an organization or community under a specified practice. Legal status concerns consequences recognized within an applicable legal order. An observer's description can be well supported while lacking institutional recognition; an institution's recognized classification can remain substantively contested. Legal validity, institutional acceptance, and epistemic justification require their own grounds.

Searle's analysis of institutional facts explains one reason to preserve these differences. On his account, certain social statuses depend on collective recognition and constitutive rules, while statements about those statuses can still be assessed as objectively correct or incorrect. Dependence on human practices is compatible with determinate facts about an institution [28]. A record of an appointment illustrates the architectural relevance. The recorded utterance, the appointment procedure, the resulting office, and the description of its holder are related objects. Their relationships require an account of authority and procedure; their meanings cannot be recovered from a timestamp or an authenticated signature alone.

The project's use of social facts also extends to consequences that remain unrecognized or arise unintentionally. Lukes's direct critique of Searle questions the restriction of social analysis to linguistically constituted institutions and calls attention to interactions between institutional arrangements and material power [20]. This criticism matters for the proposed infrastructure because a recognition-centered model could erase precisely the experiences that public inquiry needs to examine. A person's contribution may have occurred despite its exclusion from an official account. A community may suffer an effect before an agency recognizes the relevant category. The capacity to record and investigate such claims must remain available alongside descriptions of established institutional classifications.

Consider a hypothetical dispute over participation in an environmental study. Residents collect samples and describe changes in water quality. A laboratory produces measurements, a research team publishes an interpretation, and an agency classifies the evidence under its procedure. Several questions remain independent: which sampling activities occurred, whether the measurements support the published

2.6 Generative Relational Preservation and Its Limits

The preceding analysis defines the generative relational approach as a proposed integration of historical preservation, attributed representation, and public revisability. *Generative* refers to the selected processes through which claims, interpretations, and institutional judgments take shape and change. *Relational* refers to their dependencies on evidence, participants, concepts, material settings, and institutions. The combined thesis is both epistemic and normative: retaining relevant connections can support the assessment of a claim, and the selection and governance of those connections affect people's opportunities to take part in future inquiry. The contribution claimed here concerns an architectural integration guided by that thesis. Established work on infrastructure, situated knowledge, and commons governance already supplies substantial antecedents for its components.

The proposal faces a serious objection from the costs and effects of documentation. Recording a decision can support accountability while altering the conditions in which people deliberate. Routine capture of exploratory conversation could discourage tentative ideas, expose vulnerable participants, or make ordinary revisions appear suspicious. Elaborate contribution records could become instruments of managerial monitoring. Detailed histories can also create a misleading appearance of completeness when undocumented work, oral transmission, or excluded experience remains outside the system. These are foreseeable design risks and candidate mechanisms for empirical study. Their possibility defeats any simple inference from greater capture to better knowledge or greater justice.

Selective preservation must therefore be justified by the inquiries and responsibilities it serves. Readers should be able to use the retained history to examine a decision's grounds, revisions, and objections. A suitable record may consist of a reasoned decision, the evidence it relies on, and an account of the objections considered. Preserving every preceding exchange would require an additional justification. Choices about selection should identify the interests served, the omissions that matter, and the procedure through which the selection can be reconsidered. Retention can be graduated across public explanations, controlled evidence, and material whose preservation is unnecessary or impermissible.

A second objection concerns the relationship between historical reconstruction and explanation. Replaying admitted records can recover what a system held under a specified interpretation. Explaining how knowledge developed also requires judgments about causal importance, unrecorded conditions, and the meaning of disagreement. A sequence of entries establishes only the history of the retained representation. It may provide evidence for a richer genealogy, whose adequacy remains open to inquiry. The architecture should consequently preserve attributed explanations and their evidentiary grounds alongside operational traces, with clear indications of the limits of each.

A third objection concerns institutional authority. The availability of a complete-looking history may encourage an agency, platform, or model operator to treat its own representation as the decisive public account. The response proposed here combines addressable reasons, alternative interpretations, and procedures through which affected participants can seek correction. These capabilities can support public revisability only where institutions provide the resources and standing needed to exercise them. A technically preserved challenge with no prospect of examination would offer a weak form of participation.

These limits lead to a bounded preservation commitment. The infrastructure should sustain enough contextual and historical information for specified forms of examination while preserving the conditions under which people can continue to contribute, interpret, and revise knowledge. Its success would consist in improved inquiry and accountable institutional practice under those conditions. The later architecture translates this commitment into representational boundaries and services; the evaluation must examine both their operational behavior and their consequences for participation.

3 Related Work and Research Method

The proposal requires a method that connects social and normative explanation to executable distinctions. A representation can be formally precise while preserving an inadequate account of participation, and a compelling principle can remain ineffective when no operator can implement its remedy. This section

therefore needs to record adoption and classification decisions, including the scope of the responsibility accepted by the adopting actor.

The same concern applies to outputs produced without generative models. Automated language identification can exclude material from a corpus, and retrieval ranking can systematically make one interpretation easier to find. A governance model focused only on the final text would miss these upstream conditions. The proposed process representation consequently includes selection criteria, transformation methods, responsible operators, and materially relevant omissions where recording them is justified and feasible.

4.2 Provenance at Different Levels

Provenance in an AI-mediated trajectory has several distinct objects. Artifact provenance records the origin and versions of files. Operational provenance records a particular invocation and its inputs, parameters, outputs, and service dependencies to the extent available. Evidentiary provenance records which sources an adopter relies on for a specific proposition. Intellectual genealogy concerns the broader history of concepts and contributions. These relations can overlap without being interchangeable.

A logged prompt and output can help reconstruct a documented interaction. They do not reveal every influence represented in a model’s training or establish who originated an idea. A retrieval-augmented response can identify documents supplied at invocation, while the evidentiary relevance of those documents remains a matter for assessment. A generated citation can itself require verification. The proposed infrastructure gives these different claims separate types, speakers, and provenance, avoiding an undifferentiated assertion that an output has complete provenance.

Reproducibility also has several meanings. Replaying a stored output means retrieving the retained artifact. Reproducing an invocation requires access to the relevant model, software, parameters, and execution conditions, and may still be affected by nondeterminism. Reproducing the underlying scientific or historical warrant requires additional evidence and methods. A public interface should identify which of these activities is available before describing a result as reproducible.

Where sources or model details are inaccessible, an institution may still choose to use an output under a justified review process. Its record should state the limits of examination and the compensating checks on which it relies. An attestation from a service provider can be evidence about an operation while leaving the substantive claim unverified. The availability of an explanation and the adequacy of a justification thus remain related but separate questions.

4.3 Human Adoption and Distributed Responsibility

Responsibility must attach to actions and capacities within the trajectory. A contributor who supplies testimony, a laboratory that processes a sample, a model provider, an institution that adopts a synthesis, and an operator that distributes it have different knowledge and control. A requirement that one final author explain every upstream dependency can be impossible to satisfy; allowing each actor to disclaim the final effect can leave a remedy without an accountable institution.

The proposed allocation therefore identifies responsibility for bounded operations: collection, source selection, transformation, substantive adoption, release, correction, and continued distribution. An institution publishing a finding should state the assessments it performed and retain the ability to correct its own public representation. A model provider’s documentation can support that assessment without replacing the adopter’s decision about suitability for the particular public use. Specific legal liability remains governed by the applicable rules discussed within the bounded scope of Section 5.

Raji and colleagues propose internal auditing across the development lifecycle, with documentary outputs connecting decisions to an organization’s stated principles [25]. The present architecture draws a limited lesson about preserving decision trajectories. Its public responsibility requirements extend further as a proposal: an internal audit may need a permitted public explanation and an independent route for affected parties to challenge consequential uses. Those additional institutions cannot be inferred from the existence of the audit documentation.

Human review itself requires scrutiny. A nominal approval field can conceal insufficient expertise, time, access, or power to alter the output. The record should make the scope of review and available evidence assessable where appropriate, while the institution supplies the resources and authority needed to make that review meaningful. The objective is an accountable decision process rather than a ceremonial human signature.

4.4 Correction, Reuse, and Recursive Amplification

A later correction can alter a statement's public warrant while earlier copies remain embedded in summaries, indexes, or training corpora. The infrastructure can identify known dependencies and issue permitted correction notices, but it cannot guarantee that all recipients will revise their copies or that a model's internal state will change. The record must distinguish performed corrections, acknowledged notices, and unresolved downstream effects.

Recursive reuse also complicates apparent corroboration. Several accessible sources may derive from one earlier assertion. A public query can report their common ancestry where known, so repetition need not be treated as independent support. The absence of a recorded derivation does not establish independence; it may reflect incomplete history. Domain evidence standards determine how these limits affect a finding's weight.

The architecture therefore preserves attributed lineage and selected transformation events while keeping the final evidentiary judgment explicit. A synthesis can rely on a narrower corrected claim, preserve the challenged branch for historical inquiry, and describe the reach of its correction. Further contributors can extend this account through new evidence and interpretation. Continued generation remains compatible with institutional decisions because the record identifies their time, scope, and revisability.

4.5 Automation, Labor, and Participation

Automation can reduce the work of describing a record, translating a submission, or finding related evidence. It can also shift verification and correction costs onto people with the least capacity to contest an output. The relevant design question concerns the distribution of those costs and opportunities. An automated annotation should have an identified provenance and a proportionate route for correction; it should not automatically acquire the status of a verified social fact.

Interfaces for public contribution should permit uncertainty, partial knowledge, and a statement that the provided categories fail to capture a relevant experience. A contributor should be able to submit an intelligible account with assistance, while a steward translates it into structured form with the contributor's relationship to that translation preserved. Otherwise, the infrastructure risks admitting only the experiences its automated schema can already recognize.

These considerations make AI mediation one part of a broader institutional design. The significance of a model lies in the relations through which its outputs are selected, adopted, and acted upon. Governing those relations requires the combination of preservation, review, responsibility, and participation developed in the following sections.

5 Justice, Rights, and Legal Pluralism

This section establishes the normative and legal conditions under which a generative relational infrastructure could deserve public support. Its method combines conceptual analysis of justice, a bounded examination of legal instruments, and institutional design arguments. The analysis proceeds from epistemic participation and fair procedure to distribution, information rights, provenance, plural authority, and remedies. These concerns determine which records the infrastructure may create and how those records acquire practical force. They therefore enter the architectural argument before the specification of authorization services. The proposed safeguards remain claims requiring institutional justification and evaluation; the legal examples specify existing constraints within their stated scope.

6 Public Purposes and Preservation Requirements

The social and legal analysis determines what counts as adequate preservation. A record should support justified inquiry while making its classifications, omissions, and consequential uses open to appropriate challenge. This section derives requirements from those purposes, identifies the decisions needed to balance them, and connects each requirement to architectural behavior. The threat analysis includes the effects of ordinary institutional practices as well as deliberate misuse.

6.1 From Public Interests to Design Obligations

The conception of publicness in Section 2 requires more than successful retrieval. A person whose experience is represented needs an intelligible account of that representation and an effective means of contesting consequential errors. A contributor seeking attribution needs the relation between contribution and recognition to remain addressable. A custodian protecting sensitive material needs to limit disclosure without misrepresenting the resulting public account as fully reproducible. These interests generate different obligations, and their conflicts require an identified decision process.

The justice and legal analysis in Section 5 also changes the unit of design. Protecting one file while exposing its relationships can leave a person's identity or a community's protected knowledge accessible. Conversely, removing every relation after a disclosure objection can prevent an affected contributor from obtaining correction. An adequate design must therefore represent purposes, affected parties, decision authority, retention conditions, and the scope of a permissible explanation alongside the content.

The proposal uses a reasoned preservation decision with five elements: the public or domain purpose; the information and relationships necessary for that purpose; the interests and risks of affected parties; the alternatives available through aggregation, restriction, attestation, or noncollection; and the authority and review process for the decision. The decision is revisable when purposes, risks, evidence, or applicable rules change. This procedure is a normative design proposal, not a claim that a universal balancing formula resolves the relevant conflicts.

6.2 Question-Oriented Preservation

Preservation begins with a declared inquiry scope. A domain node should state which questions its records support, which source classes it accepts, and which omissions are expected. For the running case, required questions include the identity of each recorded speaker, the evidence selected for a claim, the interval affected by a correction, and the operational authority behind a release. A model-output archive intended only for reproducible text retrieval may require a different selection from an archive intended to support priority disputes.

Table 2 groups the requirements into observable contracts. The grouping makes responsibility and test design explicit while preserving the separate requirements of the language package.

Context selection itself requires provenance. A steward who omits a private conversation from a public trajectory may have good reasons, yet the resulting record should avoid implying that every relevant interaction was captured. Where safe, a scope statement can describe the collection boundary. Where even an omission's existence is sensitive, a more general qualification may be necessary. An apparently complete graph can otherwise make collection choices look like features of historical reality.

6.3 Identity, Evidence, and Authority Threats

The integrity threat model includes fabricated sources, falsely attributed assertions, duplicate identifiers with different payloads, compromised keys, and imported records whose schemas are unavailable. It also includes an ordinary user's mistaken inference that a valid signature establishes truth. The infrastructure must expose the scope of verification: authenticated bytes, admitted representation, accepted institutional finding, and substantive credibility are different statuses.

17 Conclusion

This paper has developed an interdisciplinary account of public knowledge infrastructure through the relationship between social facts, their recorded representations, and the institutions that make those representations consequential. The inquiry began with the loss of contextual relations across knowledge artifacts and examined that problem through infrastructure studies, social epistemology, justice, law, and constructive architectural analysis. Its central result is a conditional design argument: preserving knowledge evolution requires both addressable histories and institutions through which participants can interpret, challenge, and revise the public use of those histories.

The first part of the argument concerns representation. An occurrence, an account of it, a participant's assertion, evidentiary support, institutional recognition, and legal consequence have different conditions of assessment. Public infrastructure should preserve these distinctions where they affect inquiry or action. The proposed information model connects artifacts to attributed assertions, contexts, temporal scopes, and revision histories. Its domain categories provide extensible structure, while classifications and translations remain open to criticism. Historical reconstruction is limited to the dependencies and interpretations that remain available; a retained graph cannot establish a complete history of knowledge production.

The second part concerns the public organization of that representation. Selection, classification, visibility, retention, and reuse distribute epistemic opportunities and exposure. Justice consequently enters the infrastructure through the terms of contribution, the treatment of dissent, the allocation of maintenance burdens, and the availability of effective remedy. The scoped legal analysis further demonstrates that preservation and openness must be organized around applicable duties and rights. Provenance, authority, attribution, and control over information require separate justifications. Their conflation would give technical metadata consequences that the record itself cannot warrant.

The third part concerns operational and institutional continuity. The architecture separates recorded observations from authorized commands, epistemic disagreement from process conflict, and historical interpretation from current disclosure. Federation permits qualified exchange across local custodians while retaining responsibility for admission and use. These arrangements depend on institutions that can authorize policy, represent affected interests, hear challenges, fund maintenance, and provide succession. The proposed governance structure places those responsibilities within the infrastructure's design and assessment.

AI mediation makes the combined argument especially consequential. Selection, classification, synthesis, and adoption can transform the practical status of a claim while obscuring its evidentiary or intellectual ancestry. A public knowledge infrastructure should preserve the distinctions needed to assess those transformations and to respond when their basis changes. Documentation can support this work only when a responsible institution can act on a correction and affected persons can contest consequential uses.

The constructed cases and bounded formal checks establish a basis for criticism and further implementation. They provide no measured evidence of social effectiveness or general institutional legitimacy. Comparative technical studies and participatory evaluation must assess whether the proposal improves contextual understanding, review, and continuity without imposing disproportionate burdens or exposure. The resulting conception of public knowledge evolution is therefore both generative and revisable: retained relations make further inquiry possible, while the terms of their preservation and use remain subject to justified change.

A Initial Interface and Preservation Contracts

This appendix consolidates the component-facing contracts needed for an initial implementation. It identifies required request information, observable outcomes, and acceptance boundaries. The contracts describe semantics rather than a finalized HTTP API or signing protocol.