

A newly asserted permission that authorized its own transaction would collapse the distinction between describing power and legitimately installing it. Repairing such counterexamples makes the proposal more coherent. It supplies no evidence that the underlying correction policy, institutional mandate, or distribution of power is justified.

Formalization also exposes the limits of a result. Four-valued support retains positive and negative attributed support under specified coordinates; it does not measure evidentiary quality. Deterministic projection reconstructs accepted receipts given retained dependencies; it does not establish the fairness of the decisions recorded in those receipts. These limitations should remain visible in service documentation and interfaces. Evaluation should test whether users understand them, since a mathematically precise backend can still support a misleading public presentation.

15.3 Comparative Design and Institutional Baselines

A useful comparison must configure alternatives to meet the same substantive needs. One baseline should combine an existing repository with relational assertions, versioned schemas, and temporal queries. A second should combine RDF, provenance, validation, a query service, and an event ledger. A third should use a typed rule engine for the finite signed-support fragment. The comparison should document extensions, integration work, maintenance, interoperability assumptions, and the consequences of unavailable inputs. Existing tools may satisfy a domain's requirements at lower cost. Such a finding would appropriately narrow the rationale for a dedicated language or additional infrastructure layer.

Institutional comparison is equally necessary. A well-supported archive with accessible staff, documented editorial decisions, and an effective appeals route may preserve consequential context more successfully than a richer technical system with weak stewardship. Conversely, a manual process may work within one institution while losing correction histories when records circulate. The relevant comparison therefore includes combinations of technology and institutional practice. It should assess the marginal benefit of additional structure relative to the labor and risk that structure creates. Claims of superiority based solely on feature counts would neglect this relationship.

Ablation studies can identify the contribution of individual distinctions. Removing attribution separation, explicit valid time, or immutable schema identity should change answers to specified competency questions. Removing review independence or participation support should be studied through institutional exercises rather than simulated as a missing database column. The methods differ because the causal mechanisms differ. Both forms of comparison should preserve adequate baseline resources and record the conditions under which results would generalize.

15.4 Participatory Evaluation and Consequential Outcomes

The first institutional study should involve a bounded inquiry with willing custodians, contributors, and affected persons, including people whose views could be represented without their having submitted records. Participants should help define the consequential questions and the costs they regard as acceptable. Domain expertise is needed to assess evidence; experience of exclusion is needed to assess whether the categories and procedures make that evidence accessible. Recruitment through the system's existing administrators alone would provide a narrow view of participation.

An initial exercise can follow a disputed account through collection, translation, assessment, publication, challenge, correction, and downstream reuse. Participants should examine both ordinary use and cases in which a steward's decision is itself challenged. Appropriate consent, confidentiality, and independent review of the research protocol are needed before undertaking research involving people or sensitive records. The constructed cases in this paper provide materials for planning such a study; they supply no observations about how any actual community would respond.

Table 5 relates the normative commitments to evidence that could support or weaken them. The indicators belong to an interpretive evaluation; they are unsuitable as a single score of institutional justice. Low appeal volume, for example, may reflect satisfaction, inaccessible procedures, fear, or lack

Table 5: Evaluation dimensions, evidence, and adverse interpretations

Dimension	Evidence to examine	Potential failure
Representational adequacy	Participant review of categories, translations, and consequential omissions	Valid encoding suppresses the substance of an account
Epistemic participation	Access to contribution, discoverability, and response to criticism	Dissent is stored but practically inaccessible
Procedural fairness	Reasons, assistance, time to remedy, independence, and implementation of decisions	Formal completion substitutes for effective hearing
Distribution of burdens	Time, cost, exposure, maintenance labor, and access to assistance across groups	Benefits accrue to consumers while contributors bear risks
Rights and disclosure	Scoped legal assessment, authorized access, retention actions, and protected review	Preserved metadata recreates an exposure
Institutional continuity	Budget, succession exercise, export usability, and ability to challenge rule changes	Dependence on a funder or custodian defeats accountable exit

of awareness. Short processing time may reflect efficient remedy or superficial dismissal. Interviews, observation, decision records, and participant review are needed to interpret the measurements.

Evaluation should follow consequences beyond the database. A corrected record may leave an earlier public report, recommendation, or institutional decision unchanged. Researchers should examine whether notices reach relevant recipients, whether the correction changes later use, and who bears the work of obtaining those changes. The same applies to recognition: adding a contribution edge may have little value if attribution disappears in the interfaces and publications through which credit is distributed. Meaningful assessment therefore requires a specified follow-up period and an account of the downstream contexts the study can actually observe.

15.5 Generative Use and Further Inquiry

The proposal also requires an evaluation of the activity that gives generative preservation its purpose: developing further knowledge from a retained trajectory. A comparative task could ask participants to explain a change in an assessment, recover a discarded interpretation worth reconsidering, or formulate a qualified research question from a disagreement. Another task could ask for a synthesis that identifies both shared grounds and unresolved differences. Participants would receive the same permitted source materials, with different arrangements for accessing their contextual relations and histories. Baselines should retain the contextual information that their actual workflows would ordinarily provide.

The study should assess the grounds, qualifications, and traceability of the resulting interpretations as well as time and workload. Domain reviewers and relevant participants may disagree about whether an interpretation is useful; those disagreements should remain part of the findings. Producing more hypotheses or a longer account would not alone demonstrate improved knowledge. A persuasive result would show that the retained relationships helped participants identify consequential evidence, avoid an unsupported inference, or develop an inquiry with clearer grounds and limits.

AI-assisted conditions require additional attention to the relation between apparent fluency and evidentiary adequacy. A synthesis may become easier to read while losing an uncertainty or minority interpretation that mattered to the task. Comparing the adopted result with its available sources and subsequent criticism can expose that loss. Such a study would assess whether the infrastructure supports responsible knowledge development under the specified conditions. It would provide no automatic measure of originality or guarantee of progress across other domains.