

Research Article

Evidence Admissibility as a Requester- and Context-Relative Retrieval Constraint: A Formal Specification and Empirical Characterization

Ankur Garg¹, Jeffrey Esposito¹

1. Independent researcher

Retrieval-augmented systems select evidence by estimating topical relevance, but a topically relevant artifact may still be ineligible for use: it may have been superseded, it may apply to a different jurisdiction, its issuer may have lost standing, or the requester may not be entitled to rely on it. This paper specifies evidence admissibility as a retrieval-correctness constraint and characterizes it empirically.

We make three moves. First, we specify admissibility as a relation over (EvidenceArtifact, DecisionContext, Requester) rather than as a property of an artifact, unifying six eligibility dimensions — temporal validity, lineage validity, contextual scope, authority and provenance, access and entitlement, completeness and integrity — in a single formal decision procedure. Second, we formalize the relation in OWL 2 DL and report a structural finding: positive admissibility is a closed-world judgement while inadmissibility is open-world. A description-logic reasoner entails that an artifact is inadmissible whenever a defect is asserted, but cannot entail that an artifact is admissible without explicit local closure, because the absence of a supersession or derivation edge is not provable under the open-world assumption. Third, we characterize the constraint empirically on a controlled corpus of 1,939 artifacts and 400 query/context instances (14,911 admissibility judgements), comparing nine retrieval configurations.

The empirical results are deliberately reported including their negative findings. Relevance scores carry no admissibility signal (matched-pair AUROC 0.4981 under BM25, 0.4878 under a latent-semantic backbone), and relevance-only retrieval attains MRR 0.642 while 21.2% of returned items are relevant-but-inadmissible. Ordinary metadata filtering is close to sufficient for conditions decidable from static attributes or a simple context join (catch rates 0.838 and 0.849), but not for conditions

requiring transitive closure (0.576). Critically, an information-parity comparator — conventional flat filtering given the same materialized closure facts available to the ontology-driven gate — is statistically indistinguishable from that gate (RIER difference +0.0007, 95% CI [+0.0000, +0.0021]), and becomes bit-identical once the unknown-metadata policy is matched. Ordering effects (pre-filtering, post-filtering with over-fetch, reranking) are exactly equivalent.

Both load-bearing findings — the exposure gap and the parity result — were subsequently reproduced by a native Snowflake reimplementation, which materialized the three required closures using recursive common table expressions over native tables and matched the reference on all twelve tested system-test cells and all six tested mechanism-test cells.

We therefore do not claim a novel retrieval mechanism. Implementation of admissibility reduces to conventional filtering once the required closure facts are materialized. The contribution is a formal specification of admissibility as a requester- and context-relative relation that unifies six eligibility dimensions, makes relational-closure requirements explicit, and distinguishes closed-world admissibility from open-world inadmissibility.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

A compliance analyst asks what capital-adequacy requirements apply. The retrieval system returns a supervisory notice that addresses the question precisely, in the analyst's own terminology, from the right regulator. It was superseded eighteen months ago.

Nothing has gone wrong with retrieval. The document is topically relevant, and by any conventional measure the system performed well. What has gone wrong is that the document was not *eligible* to be used, and the retrieval abstraction had no place to express that.

This paper is about specifying that condition precisely. We call it **evidence admissibility**: whether an information artifact may legitimately support a particular decision, for a particular requester, at a particular time. Our central structural commitment is that admissibility is **not a property of an artifact**. The same superseded notice may be inadmissible for an analyst making a forward-looking compliance decision and perfectly admissible for an auditor reconstructing what rules were in force in 2023. Admissibility is a relation, and the relation's arguments include the decision context and the requester.

1.1. What this paper does and does not claim

We are explicit about scope because the individual ingredients of this work are well established and we do not wish to appear to claim them.

We do not claim that pre-filtering, access-control filtering, temporal or version-aware filtering, relational closure, or materialization of derived facts are novel. All are established prior art, and Section 2 documents this. We do not claim that eligibility evaluation ordered before relevance ranking constitutes a novel architectural mechanism; our own results show it does not. We do not claim that requester- or context-relative eligibility is new; it has substantial prior art in information-retrieval theory and in policy-aware systems. We do not claim the six dimensions are individually novel or that they constitute a complete or universal set.

Earlier work by the present authors used the term *Universal Admissibility Ontology*. **We retire that term.** Universality was never tested and the evidence presented here cannot test it. We refer to the *admissibility ontology*, or more often simply to the *admissibility relation*. Similarly, we describe admissibility as a retrieval-correctness constraint, not as *the* retrieval-correctness construct; relevance research has long recognized several constructs bearing on retrieval correctness and we do not displace them.

We claim the following, and Section 9 assesses each against the evidence:

1. A formal specification of admissibility as a relation over (EvidenceArtifact, DecisionContext, Requester) unifying six eligibility dimensions in one decision procedure.
2. An explicit statement of which conditions require transitive closure, and of the fact that positive admissibility is closed-world while inadmissibility is open-world.
3. An empirical characterization identifying where conventional filtering suffices and where it does not.
4. An information-parity result establishing that relational reasoning and conventional filtering over equivalent materialized facts are functionally equivalent.

The last of these is a negative result about our own earlier position, and we report it as a primary finding rather than a limitation. It has since been reproduced by a native Snowflake reimplementation (Sections 6.7 and 7.7). That second-engine reproduction is evidence of implementation-level robustness; it is not itself a contribution and adds no claim to the four above.

1.2. Research questions

- **RQ1.** Is admissibility empirically separable from topical relevance — that is, do relevance scores carry admissibility information?
- **RQ2.** Where is conventional metadata filtering sufficient to decide admissibility, and where is it not?
- **RQ3.** Once conventional filtering is given the same derived relational facts available to an ontology-driven decision procedure, does any difference remain?
- **RQ4.** Does the ordering of eligibility evaluation relative to ranking materially affect retrieval outcomes?
- **RQ5.** Does admissibility require closed-world assumptions that differ materially from standard open-world description-logic reasoning?

1.3. Relation to the authors' prior preprint

An earlier version of this work argued for admissibility as a novel retrieval-correctness construct and proposed *Eligibility Before Ranking* as an architectural principle, supported by a 40-document feasibility pilot. That pilot used a single coder, a small non-random sample, and, for three of four domains, documented facts about well-known cases rather than blinded corpus retrieval. **We retain it here only as historical feasibility evidence** that the dimensions are identifiable in real documents. It supports no prevalence estimate and no measurement claim, and it plays no role in any finding reported in this paper. All empirical claims below rest on the controlled study described in Section 6.

2. Related Work

We lead with the overlap. Each component of our specification has substantial prior art, and the contribution is a composition rather than a discovery. Readers should assess the composition against that background, not against a claim that the area is unexplored.

2.1. Relevance is already a relation, and already context-relative

The foundational point must be conceded immediately. Information-retrieval theory has held for five decades that relevance is relational and context-dependent. Saracevic's stratified model distinguishes system or algorithmic relevance, topical relevance, cognitive relevance or pertinence, situational relevance, and motivational relevance, each defined as a *relation* between information objects and some other term ^{[1][2]}. Situational relevance is defined as the relation between the situation, task, or problem at

hand and the retrieved texts, and is widely regarded as the most realistic form of user relevance [3]. Saracevic states the point plainly: there is always a "to" associated with relevance, and nothing can be relevant if there is no "to" involved.

It is therefore false to say that relevance is context-free while admissibility is context-relative, and we do not say it. Our distinction is narrower and must be argued rather than assumed: **situational relevance concerns an artifact's utility in solving the task; admissibility concerns its eligibility to be used at all.** These come apart in a specific way. An artifact of maximal situational utility — the document that would answer the question best — can be wholly inadmissible because it was superseded last week. Conversely, an artifact can be fully admissible and useless. The constructs are orthogonal in principle, and Section 7.1 tests whether they are separable in practice.

2.2. Formal specification of retrieval governance

The closest contemporary work is ContextNest [4], which formalizes *context governance* as six properties a governed context system must satisfy — provenance, version identity, integrity, deterministic selection, traceability, and temporal consistency — and proves that a standard retrieval-augmented architecture does not provide them as part of the retrieval abstraction. ContextNest determines which artifacts are approved, current, attributable and integrity-verified, with retrieval operating over that governed subset.

The overlap with our aims is substantial and we adopt several of its framings. The distinction is precise and stated by ContextNest itself: its eligibility is *artifact-relative*. Provenance, version identity and eligibility for AI use are framed as properties of the artifact rather than of whoever happened to prompt the model, and action governance including agent identity and runtime authorization is explicitly scoped to future work. Our specification places the requester inside the relation. This is a difference in the arity of the governed object, not a claim that we govern more properties.

Adjacent work formalizes multi-dimensional evidence governance on the retrieval side without requester-relativity. LedgerRAG maintains a claim-level evidence ledger using coverage, temporal validity, authority and conflict signals, and defines an authority–time partial order over evidence with auditable generation [5]. Relational conformance work over multi-artifact releases requires a typed artifact graph with required relation predicates, two-dimensional authority-scope orders, and blocking of affected claims and descendants [6]; that work is audience-conditioned in places, and is the closest prior art to our closure requirements, though its task is release conformance rather than retrieval.

2.3. Requester-relative eligibility

Requester-relative formal eligibility with audit is well established in the Semantic Web policy literature. In the Policy-Aware Web and Rein line of work, a guard collects the description of a resource's policy network and reasons over it to decide whether to permit access, with proofs generated on the client side and checked on the server side so that any third party can audit that the justification is valid ^{[7][8]}. Rule-based and description-logic access-control languages evaluate authorization requests against semantic information from remote knowledge bases, supporting policy composition, delegation and closure ^[9].

More recently, policy-aware data-access governance has been framed over a request triple of user, data and purpose, with a staged reasoning flow that applies non-negotiable policy gates early and follows deny-by-default when identity or policy context is missing ^[10]. That triple is materially our (artifact, requester, decision context), and the deny-by-default posture is a closed-world default, though it is not framed there as a formal asymmetry.

Access-control-aware retrieval specifically has been demonstrated for retrieval-augmented systems, including metadata-aware filtering and permission-aware retrieval over sensitive records ^[11].

What this literature does not do is unify authorization with the other five conditions. It treats entitlement as the governed dimension; temporal validity, supersession, applicability scope, issuer standing and record integrity are outside its scope.

2.4. Multi-dimensional, context-relative information quality

The information-quality literature established both multi-dimensionality and consumer-relativity long ago. Wang and Strong's framework organizes quality dimensions into intrinsic, contextual, representational and accessibility categories, defining quality as fitness for use by information consumers ^[12]. Strong, Lee and Wang argue explicitly that quality assessment must incorporate the task context of users, and that the same data may be viewed through different quality lenses depending on the decision-maker and the decision-task ^[13].

This is the strongest prior art against a claim of novelty for "multiple dimensions, unified, context-relative." We concede it. The differences we rely on are that this literature offers a conceptual and empirical taxonomy rather than a formal logical specification, does not address relational closure, and is not evaluated as a retrieval constraint.

2.5. Formal eligibility and rule-based admissibility

Formal, computable eligibility representation has a mature literature in clinical research, where eligibility criteria are represented explicitly to enable automatic classification, retrieval, search and reuse, using expression languages over relational information models supporting comparison, arbitrary conjunctions and disjunctions, and temporal constraints ^[14]. This establishes that formal eligibility evaluated for retrieval is not a new genre.

The term *evidence admissibility* itself has prior technical use. Digital-forensics work provides staged models encapsulating the technical and legal requirements that determine evidence admissibility ^[15], and Prolog-based reasoners integrate legal criteria to produce admissibility estimates over reconstructed attack scenarios ^[16]. Ontology-driven admissibility for network forensics predates our use of the phrase. Recent certification-oriented work uses "admissibility" for retrieval evidence directly, storing retrieved chunks with immutable identifiers alongside retrieval parameters and access controls ^[17]. We therefore define our use of the term explicitly in Section 3 rather than presenting it as coined, and we do not claim priority on the phrase.

2.6. Closure, materialization, and filtering

Our closure requirements are standard database theory and we claim none of it. Datalog axiomatizes recursive properties such as reachability and transitive closure, and implementations materialize the facts entailed by a program so that queries can be answered directly against the materialized facts; materialization maintenance is a special case of view maintenance ^{[18][19]}. Relational algebra has been augmented with generalized transitive-closure operators ^[20], and the expressive-power hierarchy of graph-view languages over relational data is well characterized ^[21].

This literature has a direct consequence for our study, which we state before presenting results rather than after: **if admissibility conditions are transitive-closure programs, then a non-recursive filter over the materialized closure answers exactly what live recursive reasoning answers.** Our information-parity experiment (Section 7.4) is therefore a confirmation of a theoretical prediction, and we recorded the prediction before running it.

Likewise, the choice between filtering before and after ranking is a mature systems topic. Filtered approximate-nearest-neighbour search formalizes pre-filtering, post-filtering and inline filtering, including the cardinality-estimation problem of choosing an over-fetch size. We do not claim any part of

this, and Section 7.5 reports our ordering results as a characterization of a known trade-off rather than as an architectural contribution.

2.7. Per-dimension coverage

Each of our six dimensions is separately addressed by existing work: temporal validity and version awareness in retrieval over evolving corpora; supersession and lineage in versioned and ledger-based retrieval [5]; applicability scope in constraint-satisfaction eligibility retrieval; authority and provenance in the PROV family and auditable source-verified frameworks; entitlement in permission-aware retrieval [11]; and integrity in retraction-aware filtering deployed in production research assistants. We claim the unification, not the dimensions.

3. Admissibility as a Requester- and Context-Relative Constraint

3.1. Definition

Definition 1 (Admissibility). Let A be an evidence artifact, C a decision context, and R a requester. $\text{Admissible}(A, C, R)$ holds if and only if A satisfies every one of six eligibility conditions evaluated jointly against C and R . Admissibility is a ternary relation; it is not a property of A .

The ternary form is the specification's central commitment, and it is what distinguishes this work from artifact-relative governance specifications. Three consequences follow.

The same artifact takes different verdicts. A superseded policy is inadmissible for a forward-looking decision and admissible for a retrospective audit of what was in force at a past date. A classified assessment is admissible for a cleared requester and inadmissible for an uncleared one. No property of the artifact alone decides these cases.

Admissibility is conjunctive across dimensions. Failure of any single condition defeats admissibility regardless of the others. This makes inadmissibility easy to establish and admissibility comparatively expensive, a structural asymmetry that Section 5 shows has a formal counterpart.

Admissibility is orthogonal to relevance, not a stricter form of it. An artifact can be maximally relevant and wholly inadmissible.

3.2. The six dimensions

We present the dimensions as a unification of conditions already recognized separately in the literature (Section 2.7). We do not claim them as individually novel, and we do not claim the set is complete.

Dimension	Question	Decided against
Temporal validity	Was the artifact in force at the decision time?	C: as-of date
Lineage validity	Has the artifact been superseded by something in force?	C: as-of date, plus supersession closure
Contextual scope	Does the artifact apply to this operating context?	C: jurisdiction, population, platform
Authority and provenance	Did the issuer hold standing at the decision time?	C: as-of date, plus accreditation closure
Access and entitlement	May this requester rely on the artifact?	R: clearance, role, purpose
Completeness and integrity	Is the record intact, and is its derivation sound?	Artifact, plus derivation closure

Only completeness and integrity is decidable from the artifact alone, and even then only for the non-derived case. Four dimensions require the decision context, one requires the requester, and three require transitive closure over relations between artifacts. **This distribution is the specification's substantive content:** it identifies precisely which conditions are artifact-local, which are relational, and which require the requester.

3.3. A structural taxonomy of admissibility conditions

The distribution above induces a three-way classification that organizes the empirical work and, we argue, is the practically useful output of the specification.

- **FLAT.** Decidable by a constant predicate over artifact attributes alone. Example: the record is marked retracted. Ordinary metadata filtering should suffice.

- **CONTEXT.** Decidable only with the decision context or requester as a parameter. Example: the as-of date falls outside the validity window; the requester's clearance is below the artifact's classification. Expressible in ordinary filtering, but not by a context-blind filter.
- **RELATIONAL.** Requires transitive closure over supersession, derivation, or accreditation chains of depth ≥ 2 . Example: a briefing derived from a review of a retracted study, where the briefing's own integrity field reads *intact*.

A practitioner's answer to "do I need this?" is determined by the proportion of RELATIONAL cases in their corpus. Section 7 quantifies the cost of ignoring each class.

4. Formal Admissibility Relation and Ontology

4.1. Design and axiomatization

We axiomatize the relation in OWL 2 DL. The ontology comprises 30 classes, 15 object properties and 11 data properties. Three commitments govern the design.

Admissibility is reified. `AdmissibilityDecision` is a class with functional properties `aboutArtifact` and `inContext`, and a multi-valued `violatesDimension`. Reification is required by the ternary form: a binary property on an artifact cannot carry a per-context verdict, and the decision must be able to record *which* dimensions failed for audit purposes.

Three properties are transitive. `supersedes`, `derivedFrom` and `accreditedBy` are declared `TransitiveProperty`. These are the three closures, and declaring them is what allows the reasoner to classify an artifact three revisions behind the current version, a briefing two derivation hops from a retracted source, and an issuer whose accreditor's accreditor lost standing.

Dimensions are defined classes. Each of the six is given necessary-and-sufficient conditions, so that satisfiability checking is informative. `InadmissibleArtifact` is the union of the six dimension-failure classes; `AdmissibleArtifact` is its complement within `EvidenceArtifact`. The two are declared disjoint.

4.2. What the reasoner does and does not derive

We report this division honestly rather than implying the reasoner derives everything.

OWL 2 DL cannot compare literals. Date comparisons (is the as-of date inside the validity window?) and ordinal comparisons (is clearance $\geq$ classification?) are performed by a materializer that asserts ground atoms which the reasoner then consumes. The reasoner's own contribution is confined to transitive closure and disjointness checking.

Dimension	Mechanism
Lineage validity	Reasoner — transitive supersededBy
Completeness and integrity	Reasoner — transitive derivedFrom
Authority and provenance	Reasoner — transitive accreditedBy
Temporal validity	Materializer — date comparison
Contextual scope	Materializer — set membership
Access and entitlement	Materializer — ordinal comparison

The reasoner contributes exactly on the three transitive dimensions, and those are exactly the three that Section 7 identifies as the operative ingredient empirically. We report the convergence but do not over-read it: it follows from the same underlying fact, namely that these three conditions are the relational ones.

4.3. Validation

Validation used the HermiT reasoner and comprises three distinct checks.

Consistency. The ontology together with a positive fixture set exercising every dimension, including chains deliberately longer than one hop, is consistent.

Defect detection (control). A fixture asserting an artifact to be simultaneously AdmissibleArtifact and integrity-retracted is correctly rejected as inconsistent. This control matters: an ontology that never rejects anything is not doing work.

Competency questions. Thirteen questions pass — nine requiring genuine multi-hop inference, three consuming materialized ground atoms, and one open-world control discussed in Section 5.

Two axiomatization findings are worth recording for reimplementers. First, OWL makes no unique name assumption, so without an explicit `AllDifferent` over the integrity-state individuals the reasoner cannot rule out `Intact = Retracted`, and **no artifact can ever be proven to have intact integrity**. This is silent and easy to miss. Second is the closure requirement, which is substantive enough to warrant its own section.

5. Closed-World and Relational-Closure Requirements

5.1. *The asymmetry*

Competency question CQ10 was originally a single question — is an artifact passing all six conditions classified admissible? — and it failed. The failure is not an implementation defect. It is a property of the construct.

Under the open-world assumption, HermiT:

- **entails inadmissibility** whenever a defect is asserted, and does so transitively;
- **cannot entail admissibility**, because the absence of a supersession or derivation edge is not provable.

The artifact might be superseded by something not in the graph.

We split the question accordingly. **CQ10a** (open-world control) confirms that without local closure, admissibility is correctly *not* entailed. **CQ10b** confirms that with local closure axioms — asserting `¬(supersededBy some EvidenceArtifact)`, `¬(derivedFrom some EvidenceArtifact)` and `¬(issuedBy some UnaccreditedAuthority)` for the individual — admissibility is entailed.

Finding 1. Inadmissibility is monotone and open-world. Admissibility is non-monotone and closed-world. Discovering one new supersession edge can revoke an admissibility verdict; no new edge can revoke an inadmissibility verdict.

5.2. *Why this is a construct property, not an implementation detail*

Admissibility is a licence to rely on an artifact. A licence cannot be granted on the grounds that no objection is currently known to the reasoner; it requires that the record be complete enough to assert that no objection exists. That is a closed-world commitment about the evidence base.

The practical corollary is uncomfortable and worth stating: **any system that asserts an artifact is admissible is implicitly asserting that its metadata for that artifact is complete**. Systems that cannot

warrant metadata completeness cannot soundly assert admissibility; they can only assert that no inadmissibility has been detected. These are different claims, and conflating them is a specific, nameable failure mode in evidence governance.

This also explains a result in Section 8.3 that would otherwise look anomalous: under incomplete metadata, a conservative heuristic outperforms the principled closure, because the conservative heuristic is closer to the closed-world semantics the construct actually requires.

5.3. Relational-closure requirements

Three conditions are transitive-closure programs: supersession, derivation, and accreditation. The specification's operational content is that **these three closures must be computed**, by whatever means. We claim nothing about how.

This is deliberate and follows from the prior art in Section 2.6. Because materialization of a recursive program permits direct query answering against the materialized facts, computing the closure at query time by a reasoner and computing it in advance by materialization are interchangeable. Section 7.4 confirms this empirically.

6. Experimental Method

6.1. Design rationale

The study characterizes where the constraint matters. It is not designed to show that any particular mechanism wins, and four design decisions were taken specifically to prevent that outcome from being manufactured.

Ground truth is deterministic and regenerated from frozen predicates. All 14,911 judgements regenerate bit-identically from the corpus generator. No label depends on observing system behaviour. Human annotation plays no role in any claim.

Observable metadata is not ground truth. Gold labels are computed from latent values; every configuration except the oracle sees only observable metadata, which is incomplete at a frozen rate (8% per field; 60% for the denormalized supersession pointer, reflecting the unreliability of denormalized back-references in real corpora). No configuration can be perfect.

Baselines are fitted, not strawmanned. The conventional filter’s predicate is chosen by greedy forward selection on a development split to maximize F1 against gold admissibility. It is the best context-blind filter obtainable, so any residual gap is a limit of representation rather than of engineering effort.

The corpus is not engineered to favour the constraint. RELATIONAL cases constitute 23% of inadmissible judgements; FLAT and CONTEXT cases — where ordinary filtering should and does suffice — are the majority.

6.2. *Corpus*

The corpus contains 1,939 synthetic evidence artifacts across four domains (finance, medicine, law, software engineering) and 400 query/context instances, yielding 14,911 (query, artifact) admissibility judgements. Admissible judgements are 53.6% of the total. Case classes are distributed CLEAN 7,995, FLAT 2,362, CONTEXT 2,954, RELATIONAL 1,600.

Each artifact carries observable metadata for issuer, validity interval, supersession pointer, derivation source, applicability scope, classification level and record state, plus latent true values. Each query/context carries an as-of date, a jurisdiction and a requester clearance level. Authority accreditation forms chains of depth up to two, and supersession and derivation form chains of depth three.

Marker-leakage audit. Admissibility status never appears in artifact text. A word-boundary audit over a banned vocabulary (expired, superseded, retracted, withdrawn, revoked, and related terms) reports zero violations. Without this, a lexical retriever could score admissibility by reading labels, contaminating every configuration.

Matched pairs. For the separability test, 800 matched pairs were constructed, each containing one admissible and one inadmissible artifact on the same topic, generated from the same template with only ordinary editorial variation.

Queries are split into 120 development instances (used only to fit baseline predicates) and 280 confirmatory instances (used for all reported results).

6.3. *Retrieval configurations*

Nine configurations share the same corpus, queries, relevance backbone, depth and seeds, and differ only in eligibility handling.

	Configuration
A	Relevance only
B	Conventional flat metadata pre-filter, predicate fitted on development split
C	Access-control / permission-aware pre-filter
D	Temporal and version-aware pre-filter, single-hop supersession column
D+	As D, with recursive supersession closure (isolates recursion)
E	Admissibility post-filter, no over-fetch
E5	Admissibility post-filter, 5× over-fetch
F	Admissibility as a soft reranking feature
G	Admissibility gate applied before ranking
I	Information-parity conventional filter: flat conjunction over the same materialized closure facts available to G
H	Oracle admissibility (latent truth) — upper bound

Configuration I is the decisive comparator. Parity was verified directly rather than assumed: across 1,500 artifact/context pairs, the materialized facts supplied to I are identical to G's live closure, with zero mismatches. I's predicate was fitted by the same greedy procedure as B's, over an extended candidate set including the materialized closure atoms; it was not hand-constructed to match G.

Fitted predicates (frozen):

- **B** (dev F1 = 0.878): `integrity_intact ∧ no_end_date ∧ not_derived ∧ scope_global ∧ issuer_not_directly_revoked ∧ not_superseded ∧ class_le_0`
- **I** (dev F1 = 0.965): `ctx_in_force ∧ integrity_intact ∧ mat_not_superseded_asof ∧ ctx_scope_match ∧ not_derived ∧ mat_issuer_accredited ∧ ctx_entitled`

Relevance is estimated by BM25 ($k_1 = 1.2$, $b = 0.75$) as the primary backbone, with a TF-IDF/truncated-SVD latent-semantic backbone as a robustness check. Both are lexical or latent-lexical; this is a stated limitation (Section 10).

6.4. Metrics

Conventional: Precision@k, Recall@k, MRR, nDCG@k.

Admissibility-sensitive: Admissibility Violation Rate; Admissible Precision@k (relevant *and* admissible among returned); Admissible Recall@k; and **Relevant-but-Inadmissible Exposure Rate (RIER)** — the fraction of returned items that are topically relevant but inadmissible. RIER and Admissible Recall@k are the primary endpoints. No composite score is constructed.

Primary depth is $k = 5$, with $k = 1, 3, 10$ as sensitivity analyses.

6.5. Statistics

Paired bootstrap over queries, 10,000 resamples, two-sided 95% percentile confidence intervals on paired differences, Cliff's delta for effect size, and Holm correction across a predeclared family of thirteen primary comparisons. Master seed 20260831; all derived seeds recorded.

6.6. Integrity checks

Seven checks were run before confirmatory analysis, all passing: bit-identical label regeneration from frozen predicates; absence of contradictory state combinations; absence of cycles in the three closures; development/confirmatory split separation; zero lexical leakage; direct verification of information parity; and a stratified sample audit.

Four instrument defects were identified during construction, and each affected run was declared exploratory and discarded. The most consequential: the conventional baseline was initially given a transitive accreditation predicate among its "flat" candidates, inflating its relational catch rate from 0.576 to 0.946. This is invisible in output tables and flatters the wrong configuration; we record it because any reimplementing is liable to it.

6.7. Second-engine reproduction

After the confirmatory results reported below were produced and the evidence package frozen — ontology, corpus, manifests, predicates, labels, queries, configurations, metrics, thresholds and seeds — a second-engine reproduction was carried out in Snowflake (September 2026). The reproducer received the frozen package and a specification stating the target findings, and was instructed to reproduce the frozen

science without redesigning the mechanism, corpus, predicates, configurations, metrics or interpretation.

Two reproduction runs were performed and they carry different evidential weight. The first executed the frozen Python reference code inside the Snowflake workspace. That is an **environment check** — the same code on a different machine — and it is not second-engine evidence. It is not relied upon here.

The second is a **native Snowflake reimplementation** and is the run this paper cites. It loads the frozen corpus into native Snowflake tables and materializes the three required closures using Snowflake recursive common table expressions: supersession (156 edges), derivation (144 edges), and accreditation, together with the derived predicates for ancestor-retracted integrity defects (82 artifacts) and transitively unaccredited issuers (13 issuers). Admissibility predicates for configurations A, G and I were then evaluated against these Snowflake-materialized facts, with BM25 scoring and metric computation executed in Snowpark over data read from the Snowflake tables.

This design targets the paper’s most consequential claim specifically. Configuration I is *defined* by materialized closure facts, so a second engine that materializes those closures by different means — declarative recursive SQL rather than procedural traversal — and evaluates the same predicates against them is a direct test of whether the parity result depends on one particular materialization routine. Section 7.7 reports the outcome.

7. Comparative Evaluation

7.1 Relevance carries no admissibility signal (RQ1)

On 800 matched pairs, a relevance model discriminates the admissible from the inadmissible member at **AUROC 0.4981** (BM25) and **0.4878** (latent-semantic). Per-dimension values span 0.467 to 0.603 under BM25 and 0.449 to 0.607 under the latent backbone. Access and entitlement is the highest at 0.603, modestly above chance; the remaining five dimensions fall between 0.467 and 0.550.

Because the corpus passes the marker-leakage audit at zero violations, this is not label-reading. Relevance estimation carries essentially no information about admissibility.

The consequence for evaluation is direct. Relevance-only retrieval attains:

Metric	Value	95% CI
MRR	0.642	[0.601,0.683]
nDCG@5	0.464	[0.435,0.494]
Precision@5	0.466	[0.440,0.491]
RIER	0.212	[0.189,0.235]

Finding 2. Conventional retrieval metrics are blind to a measurable failure class. A system can attain MRR 0.642 while more than one in five returned items is relevant-but-inadmissible, because gold relevance does not encode eligibility.

7.2. Where conventional filtering suffices (RQ2)

The mechanism test evaluates each eligibility predicate directly against gold admissibility, decoupled from ranking, and reports the catch rate for inadmissible items by case class.

Configuration	FLAT	CONTEXT	RELATIONAL	False exclusion (CLEAN)
B — fitted flat filter	0.838	0.849	0.576	0.061
C — access control	0.000	0.170	0.000	0.000
D — temporal, single-hop	0.074	0.493	0.122	0.000
D+ — temporal, recursive	0.170	0.493	0.270	0.000
G — admissibility gate	0.934	0.930	0.768	0.000
I — information parity	0.934	0.930	0.851	0.000
H — oracle	1.000	1.000	1.000	0.000

Finding 3. A fitted conventional filter is close to sufficient for FLAT and CONTEXT conditions (0.838 and 0.849). Its shortfall is concentrated on RELATIONAL conditions

(0.576), and it additionally over-excludes 6.1% of admissible evidence — a cost that eligibility-aware configurations do not incur.

This is the practically actionable result. **If a corpus contains no supersession, derivation or delegated-authority chains, conventional metadata filtering is an adequate implementation of admissibility and this specification offers no retrieval benefit.** The false-exclusion figure is the second half of the finding: blunt filtering is not merely incomplete, it discards usable evidence.

Access-control filtering alone catches 17.0% of CONTEXT failures and nothing else, which is expected — it governs one dimension. Single-hop temporal filtering catches 12.2% of RELATIONAL failures.

7.3. Recursion is the operative ingredient

Configurations D and D+ share identical predicates and differ only in whether supersession is closed transitively. Relational catch rises from **0.122 to 0.270**, and end-to-end RIER improves by 0.0107 (95% CI [0.0021,0.0193], Cliff's $\delta = -0.054$, significant under Holm but negligible in magnitude).

Finding 4. The gain attributable to relational treatment is attributable to *recursion*, not to ontological representation. The effect is statistically reliable and, at $k = 5$, small.

7.4. Information parity: the decisive result (RQ3)

Configuration I is a flat Boolean conjunction over materialized closure facts, fitted by the same procedure as B, and verified to receive exactly the derived facts G computes by live closure.

Comparison	Difference	95% CI	Cliff's δ	Holm
G – I, RIER	+0.0007	[+0.0000, +0.0021]	+0.004	not significant
G – I, Admissible Recall@5	–0.0006	[–0.0013, –0.0002]	–0.014	not significant

Configuration I in fact slightly *exceeds* G on relational catch (0.851 vs 0.768). A supplementary analysis isolates why. Giving the admissibility gate the same conservative treatment of unverifiable derivation edges that I's fit discovered — treating any artifact with a recorded derivation edge as integrity-unverifiable — produces a configuration **bit-identical to I** (RIER difference 0.0000, 95% CI

[0.0000,0.0000]). The gap was entirely a policy choice about unknown metadata, not a difference of representation.

Finding 5. Conventional filtering over materialized closure facts and live relational reasoning over the same facts are functionally equivalent. Once the unknown-metadata policy is matched, they are identical.

We recorded this prediction before running the experiment, on the grounds given in Section 2.6: all three closures are transitive-closure programs, and materialization of a recursive program permits direct query answering against the materialized facts. **The result is a confirmed theoretical prediction, not a surprise.** We report it as a primary finding because it establishes the boundary of the contribution: the specification tells a practitioner *what to compute*; it does not supply a new way to compute it.

The comparison that does survive is I against B, both conventional filters differing only in whether closure facts are available: RIER difference -0.0564 (95% CI $[-0.0686, -0.0450]$, $\delta = -0.268$). **The closure facts carry the value; the reasoning machinery does not.**

7.5. Ordering does not matter (RQ4)

Comparison	Difference	95% CI	Holm
G - E5 (post-filter, 5× over-fetch), AR@5	0.0000	[0.0000,0.0000]	not significant
G - F (soft rerank), RIER	0.0000	[0.0000,0.0000]	not significant
G - E (post-filter, no over-fetch), AR@5	+0.0456	[+0.0408, +0.0505]	significant, $\delta = 0.643$

Pre-filtering, post-filtering with adequate over-fetch, and soft reranking are exactly equivalent on every metric. Naive post-filtering is worse, and the reason is mechanical rather than architectural: it returns an average of 2.9 of 5 requested slots because filtering consumes the candidate window. Post-filter recall saturates once the over-fetch multiplier exceeds the reciprocal of admissibility prevalence — approximately 1.9 at the observed prevalence of 0.536 — and observed saturation falls between 2× and 3×.

Finding 6. Eligibility Before Ranking is not a novel mechanism. Ordering matters only when the candidate window is too small to over-fetch, which is a known trade-off in filtered nearest-neighbour search.

This holds at $k = 1, 3, 5$ and 10 .

7.6. Headroom and robustness

The admissibility gate reaches Admissible Recall@5 of 0.108 against the oracle's 0.118 (difference -0.0093 , 95% CI $[-0.0117, -0.0070]$), so incomplete observable metadata costs approximately 8% of achievable admissible recall. Residual RIER for G and I is 0.029 and 0.028 respectively against the oracle's 0.000.

The ordering $A > B > G \approx I > H$ on RIER holds in all four domains. Domain-level RIER for relevance-only retrieval ranges from 0.192 (law) to 0.237 (medicine), so per-domain reporting is required and pooled-only claims should be avoided. Under the latent-semantic backbone the pattern is preserved (A 0.220, B 0.060, G 0.035, I 0.035, H 0.000).

7.7. Native Snowflake reimplementations of the load-bearing findings

The native Snowflake reimplementations described in Section 6.7 evaluated configurations A, G and I. On the system test at $k = 5$, all twelve tested cells (three configurations $\times$ four metrics) matched the reference to the reported precision:

Configuration	nDCG@5	RIER	AP@5	AR@5
A — relevance only	0.4644	0.2121	0.2536	0.0627
G — admissibility gate	0.4642	0.0286	0.4371	0.1083
I — information parity	0.4656	0.0279	0.4400	0.1090

On the mechanism test, all six tested cells (configurations G and I $\times$ three case classes) likewise matched, including the $I > G$ difference on relational catch (0.851 versus 0.768) that Section 7.4 attributes to unknown-metadata policy rather than representation.

For the parity comparison, the paired bootstrap on RIER returned a difference of **+0.0007** with a 95% confidence interval of $[0.0000, 0.0021]$ — the same point estimate and interval as the reference — and a non-significant result. The bootstrap p -value differed between implementations (0.256 in the reimplementations against 0.734 in the reference) while the point estimate, interval and non-significance conclusion were unchanged. We report this difference rather than suppress it: bootstrap p -values are

sensitive to the null-construction convention, and the agreement that matters here is on the estimate and its interval.

The reproduction was assessed against the two load-bearing findings declared in the handoff specification:

Load-bearing finding	Reference	Snowflake reimplementation	Verdict
Relevance $\neq$ admissibility	A RIER 0.2121, G RIER 0.0286	A RIER 0.2121, G RIER 0.0286	Reproduced
$G \approx I$ information parity	Δ RIER +0.0007, non-significant	Δ RIER +0.0007, non-significant	Reproduced

Finding 9. Both load-bearing findings reproduce in a native Snowflake reimplementation.

Relevance-only retrieval exposes relevant-but-inadmissible evidence at 0.212 against the gate's 0.029, and the parity difference between the gate and conventional filtering over Snowflake-materialized closure facts remains +0.0007 and non-significant.

Three Snowflake-specific deviations were required and are recorded because they bear on reproducibility rather than on the findings. Snowflake does not permit multiple recursive branches in a single recursive query, so the supersession closure — which must traverse both the denormalized forward pointer and the inverse index over back-references — was expressed as one recursive branch with a join and coalesce, producing an identical edge set. Snowflake performs no cycle detection in recursive queries, so a depth guard of 20 was applied against a maximum observed chain depth of approximately four. BM25 scoring was executed in Snowpark rather than in SQL, as the relevance backbone is not the admissibility logic under test. None of the three alters the reproduced findings.

Scope of the reproduction. The Snowflake reimplementation covered the two load-bearing findings and the configurations required to establish them, and both were assessed as reproduced. It did not reimplement the remaining configurations, the ablation, the matched-pair separability test, or the ontology validation. Claims elsewhere in this paper rest on the reference implementation alone, and Section 10 treats them accordingly.

8. Ablation and Failure Analysis

8.1. Dimension ablation

Each dimension was removed from the gate in turn, at $k = 5$.

Dimension removed	RIER	Δ RIER	Δ Admissible Recall@5
Completeness and integrity	0.111	+0.083	-0.019
Contextual scope	0.076	+0.047	-0.011
Temporal validity	0.074	+0.045	-0.012
Lineage validity	0.063	+0.034	-0.007
Authority and provenance	0.058	+0.029	-0.006
Access and entitlement	0.057	+0.029	-0.005

Baseline gate RIER is 0.029.

Finding 7. All six dimensions produce a reproducible effect; none is inert. Completeness and integrity is the largest contributor, at roughly three times the smallest.

We claim independent contribution only in the sense that removing each dimension measurably degrades performance on this corpus. Relative magnitudes are corpus-dependent — they reflect our defect budget — and should not be read as general prevalence estimates.

8.2. Where each configuration fails

Exposure of relevant-but-inadmissible items at $k = 5$, decomposed by case class:

Configuration	FLAT	CONTEXT	RELATIONAL
A — relevance only	105	136	56
B — conventional filter	34	24	60
I — information parity	11	13	15
G — admissibility gate	11	13	16

The conventional filter reduces FLAT and CONTEXT exposure substantially while leaving RELATIONAL exposure essentially unchanged — indeed marginally higher than relevance-only, because filtering redistributes which items occupy the top five. This is the failure signature that motivates closure, and it is visible only when exposure is decomposed by class.

8.3. Conservative heuristics beat principled closure under incomplete metadata

The fitted parity predicate selected `not_derived` — excluding all derived artifacts — over the precise materialized predicate `no_retracted_ancestor`. The blunt predicate is empirically superior here because 8% of derivation pointers are unobservable, so the precise closure silently misses cases the blunt exclusion catches.

This is not an artefact. It follows from Section 5.2: admissibility requires closed-world semantics, and under incomplete metadata a conservative exclusion approximates closed-world semantics more faithfully than a precise open-world closure does.

Finding 8. Under incomplete metadata, conservative unknown-handling can dominate principled relational closure. Systems asserting admissibility must specify their unknown-metadata policy explicitly; it is a first-class design parameter, not an implementation detail.

9. Discussion

9.1. What the evidence supports

Admissibility is separable from topical relevance. Relevance scores carry no admissibility signal, under two backbones, on a marker-free corpus (Finding 1 of Section 7). This generalizes single-dimension

results in the temporal-retrieval literature to six dimensions jointly. It does not contradict situational-relevance theory; it distinguishes *eligibility to be used* from *utility once used*.

Conventional metrics have a measurable blind spot, quantified at 21.2% relevant-but-inadmissible exposure at MRR 0.642.

Conventional filtering is sufficient for simple conditions and insufficient for relational ones, quantified at 0.838/0.849 versus 0.576, with a 6.1% false-exclusion cost.

Positive admissibility is closed-world; inadmissibility is open-world. This is a formal property of the construct with a practical corollary: asserting admissibility implicitly asserts metadata completeness.

The specification's operational content is the closure requirement. Three transitive closures must be computed; how they are computed is immaterial.

The parity result is not an artefact of one implementation. A native Snowflake reimplementation materialized the three closures by different means — declarative recursive SQL over native tables rather than procedural traversal — and reproduced both load-bearing findings on the tested cells (Section 7.7). Because configuration I is defined by materialized closure facts, this is the specific claim most exposed to an implementation artefact, and it is the claim the second-engine reproduction was designed to test. The reproduction raises implementation-level confidence; it does not extend the scientific claims, and it creates no new claim of novelty.

9.2. *What the evidence does not support*

Relational representation as a retrieval mechanism. Materialize the closures, write a seven-term conjunction, and behaviour is identical. We withdraw the mechanism claim.

Eligibility Before Ranking as an architecture. Pre-filtering, post-filtering with over-fetch and reranking are exactly equivalent. We withdraw the architectural claim.

Universality. Untested and untestable on this corpus. The term is retired.

9.3. *The contribution, stated precisely*

We present a formal specification of admissibility as a requester- and context-relative relation that unifies six eligibility dimensions, makes relational-closure requirements explicit, and distinguishes closed-world admissibility from open-world inadmissibility.

We do not claim priority. Each constituent has substantial prior art: relevance is already relational and context-relative [\[1\]\[2\]\[3\]](#); requester-relative formal eligibility exists in policy-aware Semantic Web work [\[7\]](#) [\[8\]\[9\]\[10\]](#); multi-dimensional context-relative quality is Wang and Strong [\[12\]\[13\]](#); formal retrieval-governance properties are ContextNest [\[4\]](#); and closure and materialization are standard database theory [\[18\]\[19\]\[20\]\[21\]](#). The contribution is the composition and its formalization.

9.4. Implications for practice

For a practitioner, the specification reduces to four questions:

1. Does your corpus contain supersession, derivation, or delegated-authority chains? If not, conventional metadata filtering is sufficient and this specification offers no retrieval benefit.
2. If it does, are those three closures materialized? That is the whole implementation requirement.
3. What is your unknown-metadata policy? Under incomplete records this dominates representational choices (Finding 8).
4. Are you asserting admissibility, or merely the absence of detected inadmissibility? These differ, and only the first requires warranting metadata completeness.

Nothing here requires an ontology, a reasoner, or a change to retrieval architecture. That is a finding, not a deficiency of the specification, and the Snowflake reimplementation makes it concrete: a practitioner who materializes the three closures with ordinary recursive SQL and evaluates a conventional predicate against them obtains the behaviour reported here.

9.5. Implications for evaluation

Retrieval benchmarks encode relevance judgements but not eligibility, so a system can be rewarded for surfacing superseded, out-of-scope or unentitled evidence. RIER is straightforward to add wherever eligibility metadata exists, and Section 7.1 shows it varies independently of nDCG and MRR. We suggest it as a complement to, not a replacement for, relevance metrics.

10. Threats to Validity

Synthetic corpus with generative ground truth. The corpus rules *are* the ground truth, so the study cannot establish that these conditions are the ones that matter in real corpora, nor that our defect prevalences are realistic. Mitigations — marker-free text, incomplete observable metadata, fitted

baselines, a case mix favouring conventional filtering — reduce but do not eliminate the concern. **This is the study's principal limitation.** The specification's *formal* claims (Sections 4 and 5) are independent of it; the *empirical* characterization (Sections 7 and 8) is not.

Lexical relevance backbones only. No pretrained transformer embeddings were available in the execution environment. Results replicate across BM25 and a latent-semantic backbone but not across a dense neural retriever. The separability finding concerns whether relevance estimation carries admissibility information, which we expect to be robust to the choice of estimator, but we have not shown it.

Degenerate reranking configuration. Within-topic relevance scores are near-tied in synthetic text, so any admissibility penalty fully reorders the candidate window and soft reranking collapses onto the hard gate at every weight tested. The rerank comparison in Section 7.5 is therefore uninformative about reranking in general; the pre/post-filter comparison is not affected.

Low-prevalence regime not reached. Admissibility prevalence never fell below 0.50, so the over-fetch boundary condition in Section 7.5 is analytic rather than measured at scarcity.

Effect sizes are corpus-dependent. Ablation magnitudes reflect our defect budget. The *ordering* of dimension importance should not be generalized.

No human validation of the construct. Whether independent human annotators can apply these six conditions consistently is untested. This study makes no claim requiring it, and no claim here depends on human interpretability. It remains an open question for any application that would ask people to adjudicate admissibility.

Partially mitigated single-implementer risk. The instrument was built by one implementer, and four defects were found and corrected during construction (Section 6.6). A native Snowflake reimplementing (Section 7.7) has since reproduced the two load-bearing findings on the tested cells, which materially reduces the risk that those findings are implementation artefacts. The mitigation is bounded in three ways and should not be read more broadly. The reproduction covered configurations A, G and I only; the remaining configurations, the ablation, the matched-pair separability test and the ontology validation rest on the reference implementation alone. It used the same frozen corpus, so it cannot detect a defect in corpus construction — only in the retrieval and admissibility logic evaluated over it. And it tested the same predicate definitions, so a specification error common to both implementations would reproduce rather than surface. It is a second-engine reproduction, not a

reproduction by an independent team. The full package — ontology, corpus, predicates, manifests, configurations, metrics, seeds, code and outputs — is released to permit further reproduction by others.

11. Conclusion

Evidence admissibility is a retrieval-correctness constraint distinct from topical relevance. We have specified it as a relation over evidence artifact, decision context and requester, unified six eligibility conditions in a single formal decision procedure, identified the three transitive closures the relation requires, and established that positive admissibility is a closed-world judgement while inadmissibility is open-world.

Empirically, relevance carries no admissibility signal, and conventional metrics are blind to a failure class occurring at 21.2% of returned items under relevance-only retrieval. Conventional metadata filtering is close to sufficient where conditions are decidable from static attributes or a simple context join, and insufficient where they require transitive closure.

The information-parity result determines the boundary of the contribution, and we state it plainly: **conventional filtering over materialized closure facts is functionally equivalent to relational reasoning over the same facts, and identical once the unknown-metadata policy is matched.** Implementation of admissibility reduces to conventional filtering once the required closure facts are available. Ordering effects are equivalent after candidate-window size is controlled.

Both load-bearing findings — the exposure gap and the parity result — have been reproduced by a native Snowflake reimplementation that materialized the required closures using recursive SQL over native tables. That second-engine reproduction strengthens confidence that the parity result reflects the structure of the problem rather than a property of our code. It does not extend the claims.

What remains is a specification rather than a mechanism. It tells a practitioner which conditions constitute admissibility, which of them require closure, which require the requester, and what completeness commitment is entailed by asserting that evidence may be relied upon. We think a precise specification with a clearly marked boundary is more useful than a broader claim we could not defend.

Notes

Relation to prior version. This manuscript substantially revises an earlier preprint. The architectural claim (*Eligibility Before Ranking*) and the term *Universal Admissibility Ontology* are withdrawn. The 40-

document pilot reported previously is retained only as historical feasibility evidence and supports no claim in this paper.

Statements and Declarations

Funding

No specific funding was received for this work.

Potential Competing Interests

No potential competing interests to declare.

Data Availability

The ontology, corpus generator, frozen manifests, predicates, labels, queries, retrieval configurations, metrics, seeds, code, raw outputs and statistical outputs are released as a reproducibility package sufficient for independent implementation. The evidence package was frozen prior to the second-engine reproduction described in Sections 6.7 and 7.7; the Snowflake reproduction artefacts, including the materialized closure tables and the metric comparison table, are retained separately for audit.

Use of Generative AI

Generative AI tools were used to assist with implementation of the experimental harness, literature search, and manuscript drafting. All scientific claims, experimental design decisions, and interpretations are the authors'. All reported results were produced by executed code and verified against the released artefacts.

References

1. ^a ^bSaracevic T (1975). "Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science." *J Am Soc Inf Sci.* 26(6):321–343.
2. ^a ^bSaracevic T (2007). "Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part II: Nature and Manifestations of Relevance." *JASIST.* 58(13):1915–1933.
3. ^a ^bBorlund P (2003). "The Concept of Relevance in IR." *JASIST.* 54(10):913–925.

4. ^a ^bSulpovar M, Konsynski BR, Kanchwala Q, Goodhart G (2026). "ContextNest: Verifiable Context Governance for Autonomous AI Agents." *arXiv:2607.02116*.
5. ^a ^bLedgerRAG: Governance-Driven Agentic Chain of Retrieval for Dynamic Knowledge Scenarios (2026). "LedgerRAG: Governance-Driven Agentic Chain of Retrieval for Dynamic Knowledge Scenarios." *Electronics*. 15(7):1376.
6. ^aBeyond Object Validation: Relational Conformance in Multi-Artifact Agent Releases (2026). "Beyond Object Validation: Relational Conformance in Multi-Artifact Agent Releases." *arXiv:2607.14155*.
7. ^a ^bKolovski V, Katz Y, Hendler J, Weitzner D, Berners-Lee T (2005). "Towards a Policy-Aware Web." *Semantic Web Policy Workshop, ISWC 2005*.
8. ^a ^bKagal L, Berners-Lee T, Connolly D, Weitzner D (2006). "Using Semantic Web Technologies for Policy Management on the Web." *AAAI 2006*.
9. ^a ^bBertolissi C, Fernández M (2011). "An Access Control Language Based on Term Rewriting and Description Logic." *Springer LNCS 6702*.
10. ^a ^bPolicy-Aware Generative AI for Safe, Auditable Data Access Governance (2025). "Policy-Aware Generative AI for Safe, Auditable Data Access Governance." *arXiv:2510.23474*.
11. ^a ^bChen B, Tackman J, Setälä M, Poranen T, Zhang Z (2025). "Integrating Access Control with Retrieval-Augmented Generation: A Proof of Concept for Managing Sensitive Patient Profiles." *SAC '25*, 915–919.
12. ^a ^bWang RY, Strong DM (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers." *J Manage Inf Syst*. 12(4):5–33.
13. ^a ^bStrong DM, Lee YW, Wang RY (1997). "Data Quality in Context." *Commun ACM*. 40(5):58–65.
14. ^aWeng C, Tu SW, Sim I, Richesson R (2010). "Formal Representation of Eligibility Criteria: A Literature Review." *J Biomed Inform*. 43(3):451–467.
15. ^aAntwi-Boasiako A, Venter H (2017). "A Model for Digital Evidence Admissibility Assessment." *IFIP Advances in Digital Forensics XIII*.
16. ^aKarie NM, et al. "Legal Reasoning and Admissibility of Digital Evidence in Network Forensics." *J Digit Forensics Secur Law*.
17. ^aCompliance-by-Construction Argument Graphs (2026). "Compliance-by-Construction Argument Graphs." *arXiv:2604.04103*.
18. ^a ^bMotik B, Nenov Y, Piro R, Horrocks I (2019). "Maintenance of Datalog Materialisations Revisited." *Artif Intell*. 269:76–136.

19. ^{a, b}Abiteboul S, Hull R, Vianu V (1995). *Foundations of Databases*. Addison-Wesley.
20. ^{a, b}Dar S, Agrawal R (1993). "Extending SQL with Generalized Transitive Closure." *IEEE TKDE*. 5(5):799–812.
21. ^{a, b}Rotschild H, Peterfreund L (2025). "On the Expressiveness of Languages for Querying Property Graphs in Relational Databases." *arXiv:2510.07062*.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.