Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

Knowledge graph quality assessment for decisions

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-09-24 · knowledge graph quality assessment · knowledge graph quality · data quality · fitness for purpose · provenance · shacl · quality scorecard · knowledge engineering

Original article: https://reliablecontext.com/articles/knowledge-graph-quality-assessment-decisions



In this article

1. Executive Summary
2. Introduction and Background
3. Definition and taxonomy of graph quality
4. Assessing each quality dimension
5. Design the assessment and record its uncertainty
6. Data Analysis and Evidence
7. Implementation guidance and decision rules
8. Implications and Future Directions
9. Conclusion

Executive Summary

Knowledge graph quality assessment should answer a bounded decision question: whether a specified graph snapshot can support named queries for a defined population, under a stated cost of error. A graph may pass a constraint test while containing unsupported assertions; it may contain correct assertions while omitting the entities needed by the decision. The World Wide Web Consortium (W3C) Data Quality Vocabulary (DQV) lets publishers describe measurements and their provenance, but explicitly leaves the choice of dimensions and metrics to the use case. It remains a Working Group Note, not a W3C Recommendation. (www.w3.org) Research by Huaman (2022) likewise proposes a customizable assessment framework rather than a universal acceptance score. (arxiv.org)

The practical unit is a **decision-scoped quality card**: decision and query set; population and denominator; graph snapshot; test definitions; sample design; thresholds fixed before inspection; measured results; uncertainty; unresolved unknowns; and named source and reviewer. Structural metrics, constraint conformance, factual correctness, bounded coverage, freshness, and provenance reachability should appear as separate results. Seo and colleagues (2022) report **six structural metrics** and show why scale indicators alone miss structural properties. Those results do not establish factual accuracy. (arxiv.org) The W3C Shapes Constraint Language (SHACL) tests a data graph against a shapes graph and returns a conformance report; a conforming shape cannot by itself demonstrate that a source supports a value. (www.w3.org) (journals.sagepub.com)

For a worked fictional slice, suppose a bounded reference list contains **120 eligible required assertions**, of which **102 are confirmed** in the snapshot. The observed bounded coverage is **102/120 = 85.0%**. Suppose **30 sampled citations** yield **24 retrievable supporting sources**, **3 blocked sources**, and **3 not-found sources**. The joint retrievability-and-support rate is **24/30 = 80.0%** for that reviewed set; the blocked and not-found outcomes remain visible. These are disclosed toy counts, not industry benchmarks or population estimates. An opportunistic review cannot support a conventional sampling margin of error, and an unknown assertion must not automatically be scored false. (aapor.org) (papers.nips.cc)

Acceptance should be conditional on every decision-critical gate, including any disqualifying unknowns. A scorecard can authorize the stated use, restrict it to a supported subpopulation, send the graph for remediation and retest, or mark the evidence insufficient. The quality verdict must name the snapshot and review date so

that a later graph is assessed again. NIST treats data quality assessment as actions across a dataset's lifetime, and W3C guidance calls for explicit dataset versions and update frequency. (nvlpubs.nist.gov) (www.w3.org)

Introduction and Background

A knowledge graph represents entities and relations as assertions that can be queried and linked. Quality is not a property that can be inferred from graph size, response speed, or a single global score. The analyst's question is narrower: may a particular query result be used for a particular decision about a specified population? A graph that is sufficient for exploration may be inadequate for an irreversible decision, while the same graph may support one query family and leave another unanswered. The older consumer-centered data-quality literature puts the task at the center of assessment, and NIST defines fitness for purpose against a dataset's planned use. (web.mit.edu) (nvlpubs.nist.gov)

The assessment begins with a sentence that a reviewer could dispute: “Use snapshot S to answer query set Q for population P as of date T, provided that the result meets gates G.” Each symbol must resolve to a controlled artifact or a written definition. A **graph slice** is the subgraph reachable through the query set and its relevant entity types, not an arbitrary sample of all triples. An **assertion** is a claim such as entity A having relation R to value B at a stated time. A **decision unit** might be an entity, an assertion, a linkset, or a dataset. DQV explicitly permits measurements on resources as fine grained as a set of triples; the chosen unit belongs on the card. (www.w3.org)

This report offers a reproducible assessment procedure, not a new benchmark. It separates tests that answer different questions and shows how to record the denominator, evidence, and limits of each result. The distinction matters because a validator can establish that a value has the required datatype or cardinality without establishing that the value is true. Conversely, a hand-checked true assertion can sit inside a graph with broken identity links or unusable version history. DQV's own examples warn that precision and accuracy are different properties. (www.w3.org)

Definition and taxonomy of graph quality

Fitness for a named decision

An assessment framework should start from **failure costs**, not from a convenient metric catalog. Write down the action the organization intends to take, the questions the graph must answer, and the consequences of a wrong positive, a wrong negative, a stale answer, and an untraceable answer. A query set should include both the production query and adversarial variants: different identifiers for the same entity, relations near a boundary date, and requests for a missing value. The intended consumer's use defines which error modes matter. Wang and Strong's consumer-centered framework, as published by MIT, groups quality into intrinsic, contextual, representational, and accessibility categories; it explicitly says context depends on the task. (web.mit.edu)

The chosen population must be enumerated or operationally defined. “All organizations” is not a usable denominator unless the organization has a frame that says which organizations are eligible. “All assertions in the graph” only measures internal content, not what is absent. For a coverage test, construct a bounded reference set: for example, the named entities and required predicates in a maintained reference list for this decision. An Open Research Knowledge Graph study notes that instance completeness requires a small ground-truth collection for a specified ontology. That is a methodological constraint, not a claim of absolute completeness. (link.springer.com)

The quality dimensions should be explicit. **Structural quality** concerns graph topology and connectedness. **Conformance** concerns declared shapes or constraints. **Factual correctness** asks whether an assertion is supported by evidence and is accurate for the relevant time. **Coverage** asks how much of a bounded requirement set is present and confirmed. **Freshness** concerns the age and relevance of observations. **Provenance reachability** asks whether cited evidence can actually be retrieved and whether it supports the precise assertion. **Unknown-state quality** asks whether absence, blockage, and nonapplicability are distinguished. A survey of linked-data quality identified many possible dimensions and metrics, which is further reason to choose rather than merely count them. (journals.sagepub.com) (link.springer.com)

What a metric can and cannot establish

Seo and colleagues' structural work is useful for diagnosing graph organization. A favorable topology can make traversals possible and reveal isolated regions, but it cannot certify the factual content of a node or edge. Likewise, SHACL's validation report says whether a data graph conforms to a shapes graph; it does not compare values with independent sources. A shape can require a relation to be present, yet the relation can point to the wrong real-world entity. (arxiv.org) (www.w3.org)

Correctness requires a different test. Reviewers select assertions under a disclosed design, locate sources appropriate to the assertion, compare the graph's representation with the source, and record agreement, contradiction, ambiguity, or unavailable evidence. Research on graph assessment describes comparison against ground truth as necessary for factual accuracy. Provenance verification research further distinguishes a reference being attached from its text actually supporting a triple. (link.springer.com) (journals.sagepub.com) Neither a valid URL nor a source title alone is sufficient evidence of support.

Table 1 contrasts evidence mechanisms that may appear in an assessment. The row for Reliable Context states first-party observable guarantees; it is not an independent benchmark result.

Mechanism	What a reviewer can observe	What remains to test
Structural metric	A topology or connectivity indicator for a defined graph snapshot. (arxiv.org)	Whether represented assertions are correct, complete for the decision, and sourced.
SHACL validation	Conformance and detailed validation results against a stated shapes graph.	Whether shapes encode the right decision rules and values agree with evidence.
Source-backed manual review	A sampled assertion checked against a cited passage and observation date. (journals.sagepub.com)	Coverage of the eligible population and uncertainty from the sample design.
Reliable Context	Its first-party page says values have cited, verbatim, dated sources and declared unknown states.	Independently sample whether the cited passage supports each value for the intended decision.

No row substitutes for the others. A source-backed record can still omit required entities; a complete reference list can still contain stale or incorrect assertions. The useful output is therefore a vector of decision-specific findings, not an average that allows a strong topology score to erase a weak provenance score. DQV describes a quality metric as a procedure tied to a dimension and does not prescribe one universal list. (www.w3.org)

Assessing each quality dimension

Structure and constraint conformance

Structural tests should declare the graph boundary and version before computing a statistic. The assessor may count isolated entities, missing required links, or relation patterns that make the query impossible. Report both numerator and denominator, including which entity classes and edge types are in scope. A topology statistic calculated over the entire graph can conceal the failure of a small but decision-critical slice. Seo and colleagues show that structural characteristics can be hidden by scale-only indicators, which is a reason to inspect structure in addition to size. (arxiv.org)

Constraint tests should publish the shapes, their version, the target classes, and the validation report. SHACL Core includes minimum and maximum cardinality restrictions on property values, so it can identify a record missing a required relation or carrying too many values for a chosen rule. A passing report means conformance to those shapes on that data graph. It does not mean the shapes are complete for the decision, and it does not establish factual truth. If the shapes graph is ill formed, the SHACL specification says the validation result is undefined; such a run must be marked invalid rather than silently counted as a pass. (www.w3.org)

Factual correctness and identity

For each reviewed assertion, the assessment should ask four separate questions: does the source identify the same real-world subject, does it state the same predicate and value, was it applicable at the decision time, and is the source version accessible to the reviewer? The source check should use the exact passage and locator. An assertion about a related organization must not be treated as an assertion about the focal organization. Record unresolved ambiguity as such instead of forcing a binary label. The academic description of provenance verification makes the test concrete: whether source text supports the documented graph triple. (journals.sagepub.com)

Sampling must match the decision. A random sample of all triples can be dominated by low-risk predicates and miss a small high-risk class. Stratify by predicate, source type, entity class, or consequence when those strata were fixed in advance; disclose selection probabilities if estimates across strata are intended. If selection was purposive, report audited counts but do not attach a conventional sampling margin of error. The American Association for Public Opinion Research (AAPOR) asks analysts to disclose population rules, coverage exclusions, and the unit of analysis, and it warns that conventional margins of sampling error cannot be derived from nonprobability samples. (aapor.org)

Bounded coverage and completeness

Coverage needs an external requirement set. Let **R** be the eligible required assertions for the named decision and **C** the subset confirmed in the graph with evidence. Then bounded coverage is **|C|/|R|**. State how R was assembled, when it was frozen, who reviewed it, and whether any reference entries are themselves contested. If an entity has three required predicates, the denominator is three assertion requirements when the unit is an assertion, not one entity. If the unit is an entity, define what makes an entity complete. The denominator cannot quietly switch between these units. (link.springer.com) (aapor.org)

Completeness is relative to the chosen ontology and use. An unrecorded triple is not automatically false; under an open-world interpretation, it may be true but absent from the graph. The NeurIPS research on incomplete graph information distinguishes the closed-world move of scoring every unknown triple false from an open-world treatment that allows missing facts. (papers.nips.cc) Maintain states such as **confirmed present**, **confirmed absent for a defined scope**, **not found**, **blocked**, and **not applicable**. The last three require reasons and search boundaries. They should not be collapsed into a blank or into a negative label.

Freshness and provenance reachability

Freshness is not one threshold for every relation. Define an observation-age distribution by predicate and source: median age, tail age, and the count older than each predeclared limit. A stable identifier and a frequently changing relation may warrant different cutoffs. Record the observation date and graph snapshot date separately; if the source has a version, retain that version or an equivalent reproducible citation. W3C data publishing guidance recommends explicit update frequency and dataset version dates.

Provenance reachability needs its own denominator: sampled citations, not sampled graph assertions, because one assertion may cite several sources and one source may support several assertions. For each citation, classify **retrievable and supportive**, **retrievable but nonsupportive**, **blocked**, **not found**, or **not applicable**. Keep the last categories visible even when the source cannot be inspected. The W3C Provenance Ontology (PROV-O) can represent chains among entities and activities, while DQV examples allow provenance on quality measurements themselves. (www.w3.org) A chain of custody tells the assessor where a value came from; a content check still determines whether the source says what the assertion claims.

Table 2 is a reusable dimension matrix. Thresholds are illustrative policy choices for a named decision, not universal recommendations.

Dimension	Test and unit	Denominator	Example preregistered gate	Evidence and owner	Uncertainty to report
Structure	Required path present per eligible entity. (arxiv.org)	Entities in the query slice.	At least 98% of required paths present.	Versioned graph export; graph steward.	Boundary and entity resolution errors.
Constraint conformance	SHACL pass per target record.	Records targeted by the published shapes.	All decision-critical shapes pass.	Shapes, validator report; ontology owner.	Incomplete or invalid shapes.
Factual correctness	Supported assertion per reviewed assertion. (link.springer.com)	Assertions selected by the sample design.	No unsupported critical assertions.	Source passage, locator, date; reviewer.	Sampling and adjudication disagreement.

Dimension	Test and unit	Denominator	Example preregistered gate	Evidence and owner	Uncertainty to report
Bounded coverage	Confirmed required assertion. (link.springer.com)	Frozen eligible requirement set.	At least 95% confirmed.	Reference list and evidence; decision owner.	Reference-set omissions.
Freshness	Observation age per active assertion.	Assertions with a declared observation policy.	All critical assertions within their predicate limit.	Source date and snapshot; steward.	Unknown source update cadence.
Citation retrievability and support	Citation both retrievable and supportive.	Citations selected for audit.	All critical citations retrievable or explicitly blocked.	Source copy or URL and locator; evidence owner.	Access changes and reviewer clearance.

The matrix makes tradeoffs visible. A committee may tolerate incomplete coverage for exploratory search yet require zero unsupported critical assertions before an automated action. The threshold has meaning only when chosen before seeing the result; otherwise it can drift to accommodate the observed graph. The table's percentages are policy examples, not empirical norms. DQV's refusal to define a normative dimension list leaves that choice with the data consumer.

Design the assessment and record its uncertainty

Define the frame before testing

Begin with a versioned decision protocol. Specify the exact query templates, expected outputs, eligible entity identifiers, included predicates, time interval, and source permissions. Freeze the graph snapshot and record any transformations used to produce the tested view. This prevents a later correction from being presented as though it were in the earlier decision state. Data Catalog Vocabulary (DCAT) Version 3 describes version chains for data resources, and DataCite's version guidance similarly distinguishes major content updates through related identifiers. (www.w3.org) (support.datacite.org)

Build separate frames for distinct tests. The coverage frame is the reference set of requirements. The correctness frame is the set of assertions eligible for checking. The provenance frame is the set of citations attached to those assertions. The freshness frame is assertions whose predicates have observation-age policies. State exclusions, including sources a reviewer cannot access, and explain their effect on the decision. The US Census Bureau's sample standards require a frame and design suited to the collection objective and documentation that permits replication. (www.census.gov)

For a probability sample, publish the selection method, inclusion probabilities, weighting, and variance method. A simple random design and a clustered design do not share the same uncertainty calculation. Census guidance says complex design must be considered when calculating variance. (www.census.gov) For a judgmental review, publish the cases and selection reasons, but avoid a confidence interval that implies probability sampling. AAPOR says precision statements for nonprobability designs depend on stated model assumptions. (aapor.org) An interval cannot repair frame omissions, disputed ground truth, or a source that a reviewer was not allowed to inspect.

Adjudication should be written down before reviewers begin. Use two reviewers for contested assertions if the decision warrants it, preserve their independent labels, and record how disagreements were resolved. The labels should distinguish source absence from an explicit negative statement. AAPOR's data guidance separates no response from substantive options such as “none” or “don't know”; the same discipline is useful when a graph declares unknowns. (aapor.org) Human review should state the source hierarchy for this decision without claiming that one source type is universally correct.

Publish the one-page quality card

The card is the compact, inspectable result of the protocol. It should fit on one page without dropping the uncertainty or exclusions. The schema below is intentionally human readable; DQV can express a measurement, its metric, the resource computed on, and associated provenance for machine-readable publication.

- **Decision and query set:** the named action, query templates, expected outputs, and consequence of each error type.
- **Population and denominator:** eligible entities, assertions, and citations, each with an explicit frame and excluded segments.
- **Snapshot and version:** graph identifier, extraction time, source versions, shapes version, and reference-list version.
- **Dimensions and methods:** distinct structural, conformance, correctness, coverage, freshness, provenance, and unknown-state tests.
- **Sample design:** selection frame, unit, sample size, selection rule, weighting if used, and reviewer protocol.
- **Thresholds fixed before measurement:** acceptance, restricted-use, remediation, and abstention rules.
- **Result and uncertainty:** numerator, denominator, rate where meaningful, intervals only when justified, and adjudication disagreements.
- **Unresolved unknowns:** counts of not found, blocked, and not applicable, with reasons and decision impact.
- **Source and reviewer:** citations, observation dates, reviewer identity or role, approval date, and the next review trigger.

The card should preserve a failed gate rather than substitute a composite mean. The record of an assessment is itself evidence: who ran which test, on which version, with which source access and review rules. NIST defines provenance as an attributed historical record, and the W3C DQV illustrates quality-measurement provenance at several levels. (nvlpubs.nist.gov) If source access changes, a reader should still be able to see whether the old conclusion was based on a retrievable passage, a retained copy, or an unresolved block.

Data Analysis and Evidence

The published literature supports a multidimensional approach but does not supply a universal acceptance rate. A linked-data quality survey reported **18 dimensions and 69 metrics**; a later knowledge-graph review described **23 dimensions and 42 metrics**. Those counts characterize the authors' taxonomies, not a

requirement to compute every metric for every decision. (journals.sagepub.com) (link.springer.com) Huaman's 2022 work is a proposed framework customizable to task or domain. Seo and colleagues' six metrics concern structure; neither paper claims that a structural score proves a graph's factual accuracy. (arxiv.org)

Empirical provenance work also warns against equating attached references with verified support. Amaral and colleagues sampled **416 triple-reference pairs**. The released WTR dataset contains **409 unique triple-reference pairs** after seven duplicate pairs were removed. (github.com) (journals.sagepub.com) A separate service illustration comes from Crossref, whose Participation Reports check the presence of **11 metadata elements** and offer gap reports for records missing selected fields. That is a useful example of a bounded denominator and explicit missingness, not a measure of factual truth. (www.crossref.org)

Table 3 shows a complete fictional calculation for one decision slice. The reference list and citations are treated as enumerated toy sets, so no sampling confidence interval is attached. The sample label in the citation row refers to the review exercise, not to a probability sample from a known population.

Fictional card field	Disclosed toy count or rule	Result and interpretation
Decision and slice	Query set Q over 40 eligible entities in snapshot S.	Only this population and snapshot are covered.
Required assertion coverage	120 eligible requirements; 102 confirmed with support; 12 not found; 6 blocked.	**102/120 = 85.0%** bounded coverage. Unknown states remain in the denominator.
Citation retrievability and support	30 reviewed citations; 24 retrievable and supportive; 3 blocked; 3 not found.	**24/30 = 80.0%** observed joint retrievability-and-support rate for the reviewed citations.
Constraint conformance	40 eligible records; 38 pass the stated SHACL shapes; 2 fail.	**38/40 = 95.0%** conformance to these shapes, not truth.
Preregistered decision gate	Every critical assertion supported; coverage at least 95%; every critical citation retrievable or explicitly blocked with a reason.	**Remediate and retest** because coverage misses the stipulated gate; inspect blocked citations before use.
Source and reviewer	Frozen reference-list version, source locators, and graph snapshot S; fictional reviewer role: graph steward.	Decision owner signs the card only for query set Q.
Uncertainty and provenance	Enumerated toy reference set; no probability selection; review date and source locators retained.	No population confidence interval or industry generalization. (aapor.org)

The coverage numerator counts confirmed requirements, not merely occupied cells. The **18** unresolved requirements are divided into **12 not found** and **6 blocked** so that a reader can see whether more searching or access is needed. The citation calculation uses citation instances as its unit; it cannot be compared directly with entity coverage. The SHACL rate answers a third question. If the decision owner had declared any blocked citation critical, the card would also have to identify that block's consequence. No one should infer statistical precision from these small, invented counts. (papers.nips.cc) (aapor.org)

Data services show why scope labels matter in practice. OpenAlex distinguishes a core corpus from an expansion corpus that it says has thinner, noisier metadata, and directs careful bibliometric uses toward the core. (help.openalex.org) The lesson is not that one corpus is universally better; it is that a quality statement

belongs to a population and intended use. Crossref's element-presence checks and the Library of Congress metadata provenance field illustrate separate questions: is a required field present, and what is the provenance of the field's content? (www.crossref.org) (www.loc.gov)

Implementation guidance and decision rules

An assessment becomes operational when the result is tied to an action. **Accept** only for the named query set, population, snapshot, and review horizon if every critical gate passes. **Restrict** when evidence is sufficient for a narrower subpopulation or predicate set that can be named exactly. **Remediate** when identifiable missing sources, failed constraints, or stale observations can be corrected and retested. **Abstain** when the reference frame, ground truth, or source access is too uncertain to support the decision. These are governance outcomes, not grades of intrinsic graph worth.

The retest plan should target the reason for failure. A failed SHACL result calls for a corrected assertion or shape and a new validation report. Unsupported factual assertions call for source review and possible correction. A coverage deficit calls for a frozen reference list and more evidence for missing requirements; changing the denominator requires a new card. A blocked source may call for a cleared reviewer or a declared limitation. A stale observation calls for a new observation with a date and version. W3C guidance asks publishers to state update frequency and changes; a dataset version chain helps a reviewer locate what was actually assessed. (www.w3.org)

The quality record should include the reviewer’s access context. A public link may resolve for one person and be blocked for another; a retained copy may be accessible only under policy. Reliable Context's first-party description says its values carry a verbatim cited source and observation date, that unknowns are declared as not found, blocked, or not applicable, and that past record states can be reproduced. Those are observable properties to include in an evaluation, not a substitute for an independent check of a given assertion. Its first-party page also says exports carry the source, excerpt, locator, and dates; a reviewer should test that chain on sampled values relevant to the decision.

Do not reduce all of this to a dashboard average. A weighted score requires justified weights, a stated treatment of missing evidence, and a rule for critical failures. The card may show rates, but it should retain the underlying counts and source states. A graph with excellent metadata presence can still have a wrong value; a graph with a correct value can still lack a reachable citation. DQV expressly allows consumers to make their own fitness judgment, and its examples distinguish high precision from accuracy. (www.w3.org)

Implications and Future Directions

The most important research problem is estimating quality without hiding unknown ground truth. Absolute graph completeness is generally unavailable outside a bounded reference frame. New methods may infer likely gaps or prioritize review, but such predictions should be labeled as hypotheses until evidence confirms them. Work on open-world knowledge bases asks where a knowledge base is incomplete and studies relative recall rather than assuming every absent fact is false. (researchportal.ip-paris.fr) A graph assessment should preserve the distinction between a negative fact and a fact that has not been found.

Another direction is stronger provenance evaluation. Machine-readable chains can connect a statement to a source and a source to its derivation, but assessment still needs a check that the passage supports the claim. PROV-O offers entity and activity relations; DQV offers places to describe the measurement and its provenance. The two together support an inspectable evidence path without implying that the path is correct. Research that checks triple support against source text provides a more demanding test than counting references alone. (journals.sagepub.com)

Versioning will matter whenever a decision must be reconstructed. DCAT Version 3 supports backward version chains, DataCite documents identifiers for major content updates, and Dataverse documents access to specific dataset versions. (support.datacite.org) (guides.dataverse.org) A future assessment can compare cards across snapshots, but it should preserve the old thresholds and denominators or explain any change. Without that discipline, an apparent improvement may simply reflect a revised scope.

The broader lesson is that quality evidence should be portable. A graph steward, decision owner, and independent reviewer should be able to examine the same card, sources, exclusions, and unresolved unknowns. Open standards help express parts of that record, while the final authorization remains a decision about fitness for a specific purpose. (nvlpubs.nist.gov)

Conclusion

Knowledge graph quality assessment is a decision protocol with measurements attached. It begins by naming the query set, population, unit, graph snapshot, and failure costs. It then tests structure, conformance, factual correctness, bounded coverage, freshness, provenance reachability, and unknown-state handling separately. A pass on one axis cannot compensate for a critical failure on another unless the decision owner explicitly justifies that tradeoff in advance.

The reusable artifact is a one-page card with frozen denominators, methods, thresholds, counts, uncertainty, source locators, reviewer, and review date. The fictional example shows why an **85.0%** coverage rate and an **80.0%** joint retrievability-and-support rate must not be averaged into a vague quality score: they have different units, denominators, and remediation paths. The outcome should be accept, restrict, remediate, or abstain for the named use, with a retest rule when the graph or evidence changes.

The discipline is deliberately modest. No collection of structural metrics, constraint reports, or source links can establish universal truth or absolute completeness. What an organization can establish is a bounded, reproducible account of why a particular graph state was adequate, inadequate, or still uncertain for a particular decision. When evidence remains unavailable, the defensible outcome is an explicit abstention with its reason and a date for reassessment, rather than a silent negative.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.w3.org/TR/vocab-dqv
2. https://arxiv.org/abs/2208.07779
3. https://arxiv.org/abs/2211.10011

4. https://www.w3.org/TR/shacl
5. https://journals.sagepub.com/doi/full/10.3233/SW-233467
6. https://aapor.org/standards-and-ethics/transparency-initiative
7. https://papers.nips.cc/paper_files/paper/2022/hash/378226e5df7eded3e401de5c9493143c-Abstract-Conference.html
8. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/1500-18/NIST.SP.1500-18r2.html
9. https://www.w3.org/TR/dwbp
10. https://web.mit.edu/tdqm/www/papers/94/94-10.html
11. https://link.springer.com/article/10.1007/s00799-021-00306-x
12. https://journals.sagepub.com/doi/10.3233/SW-150175
13. https://link.springer.com/article/10.1007/s42979-022-01429-x
14. https://www.w3.org/TR/prov-o
15. https://www.w3.org/TR/vocab-dcat-3
16. https://support.datacite.org/docs/connecting-versions-with-related-identifiers
17. https://www.census.gov/about/policies/quality/standards/standarda3.html
18. https://www.census.gov/programs-surveys/sipp/methodology/sampling-error.html
19. https://aapor.org/standards-and-ethics/best-practices
20. https://github.com/gabrielmaia7/RSP
21. https://www.crossref.org/documentation/reports/participation-reports
22. https://aapor.org/statements/understanding-a-credibility-interval-and-how-it-differs-from-the-margin-of-sampling-error-in-a-public-opinion-poll
23. https://help.openalex.org/data/works/corpus
24. https://www.loc.gov/marc/community/ci883.html
25. https://researchportal.ip-paris.fr/en/publications/completeness-recall-and-negation-in-open-world-knowledge-bases-a-
26. https://guides.dataverse.org/en/latest/quickstart/what-is-dataverse.html

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.