Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

How to evaluate data sources for a knowledge graph

> 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 · 2026-10-13 · knowledge graph · data source evaluation · source admission · data source reliability · data quality assessment · data provenance · dcat · dqv · context engineering

Original article: https://reliablecontext.com/articles/knowledge-graph-data-source-evaluation



In this article

1. Executive Summary
2. Introduction and Background
3. Define the object and the decision
4. Source admission gates and the evidence card
5. Measure source quality with explicit denominators
6. Represent versions and measurements with DCAT and DQV
7. Decide, retain and re-admit
8. Data Analysis and Evidence
9. Implications and Future Directions
10. Conclusion

Executive Summary

To evaluate data sources for a knowledge graph, define **the proposition, intended decision and required evidence** before inspecting the dataset. Then test permission, identity, provenance, dates, coverage, corrections and known quality issues against that use. This report proposes a **source-admission evidence card** that records an **accept, defer or reject** decision, the supporting evidence, the unresolved gap and a review trigger. Quality is purpose dependent: the Australian Bureau of Statistics (ABS) explicitly allows different users to assess the same product differently. (www.abs.gov.au) The National Institute of Standards and Technology (NIST) links data quality to fitness for purpose and describes its research data framework as nonprescriptive. (nvlpubs.nist.gov)

Standards supply useful representations, with different statuses and scopes. **Data on the Web Best Practices** is a World Wide Web Consortium (W3C) Recommendation dated **31 January 2017**; **Data Catalog Vocabulary version 3**, or DCAT 3, is a Recommendation dated **22 August 2024**. (www.w3.org) The **Data Quality Vocabulary**, or DQV, is a **Working Group Note**, dated **15 December 2016**, whose publication does not imply W3C Membership endorsement. (www.w3.org) DCAT describes catalog resources; DQV lets users describe quality and judge fitness for purpose. Neither representation, by itself, establishes truth or grants permission to reuse data. (www.gov.uk) (www.w3.org)

Measurement requires an explicit denominator and clock. The proposed observed-coverage calculation is documented applicable items divided by sampled applicable items. Freshness lag is the evaluation date minus the underlying source observation date. These are operational proposals, not measured industry benchmarks. Distinguish observation age from statistical timeliness and punctuality, which concern reference periods and availability schedules. (www150.statcan.gc.ca) Report **unmeasurable** when the necessary denominator or date is missing. Under the open-world assumption, an absent graph statement is unknown rather than false, so sampled documentation coverage cannot establish exhaustive real-world coverage. (arxiv.org)

Admission remains conditional after ingestion. A persistent identifier may span updates: DataCite permits minor content changes under the same Digital Object Identifier, while Zenodo recommends citing a particular version for reproducibility. (support.datacite.org) (zenodo.org) Re-evaluate permission changes, corrected values, changed coverage, identity mappings and measurement scope. A successful metadata assessment

supplies evidence about metadata; the European Data Portal documentation expressly excludes intrinsic correctness, completeness and legal validity from its assessment. (dataeuropa.gitlab.io) The defensible output is a reproducible decision with declared limits, supported values and a named next review.

Introduction and Background

A context engineer combining private records with public datasets must decide what a candidate source is allowed to support. “Include this dataset” leaves that question unresolved. A usable admission decision instead names the entity class, property, population, relevant period and action that depends on the evidence. The distinction matters because knowledge graph engineering involves schema, identity and context as well as creation and quality assessment, as Hogan and colleagues (2021) discuss in their survey. (arxiv.org)

This report examines **source selection before graph ingestion**. It proposes an operational method for recording why a dataset may supply particular values, what was actually tested and what remains unknown. The method draws on standards, official quality frameworks and original research. It is a synthesis, with no claim that its card is an existing standard or that the proposed calculations are universal thresholds. The NIST framework similarly permits stakeholders to select relevant components rather than prescribing one procedure for every organization. (nvlpubs.nist.gov)

Dataset documentation offers a useful precedent. Gebru and colleagues' *Datasheets for Datasets* (2021) proposes recording motivation, composition, collection process and recommended uses. (www.microsoft.com) An admission card adds a decision boundary: which proposition this organization will rely on, under which conditions, with which missing evidence. A publisher's description belongs on the card, but a statement that a dataset is “complete” should remain an attributed assertion until its scope and denominator have been examined.

The report uses three evidence states throughout:

- **Publisher claim:** What the producer or distributor says, with its source and date.
- **Measured sample:** What an evaluator checked, with the sampled scope, procedure and calculation.
- **Unavailable evidence:** What could not be established, with the reason and its effect on admission.

These states are proposed reporting conventions. They operationalize the broader need for provenance: NIST explains that missing provenance documentation makes quality and reliability difficult to assess. (nvlpubs.nist.gov) They also keep an organizational decision legible when a dataset's accessibility, apparent detail and suitability point in different directions.

Define the object and the decision

A **source** is the accountable origin or channel from which evidence is obtained. A **dataset** is the collection being evaluated. A **distribution** is a particular representation used to access it. DCAT distinguishes its conceptual Dataset class from a Distribution, defined as a specific representation of a dataset. (www.w3.org) Consequently, admitting a conceptual collection does not automatically admit every export, endpoint or transformed file bearing its name.

An **observation** is an evidence-bearing assertion with a subject, property or variable, value and relevant time. This report uses that definition operationally. In the statistical setting, Data Commons defines an observation as a measured variable value associated with an entity and time period. (docs.datacommons.org) A source publication date, an import date and the date to which an observation refers should occupy separate fields whenever they differ.

An **intended decision** is the action or conclusion for which the graph value will be used. Its specification should make the following boundaries explicit:

- **Question:** State the proposition that requires evidence.
- **Population:** Name the entities and exclusions to which it applies.
- **Property:** Define the relationship, attribute, units and interpretation.
- **Time:** Specify whether the question concerns a historical state or a current one.
- **Consequence:** Identify how missing, stale or incorrect evidence affects the action.
- **Evidence burden:** State what direct documentation, corroboration or coverage is required.

These fields prevent the same dataset from receiving a blanket reliability label. ABS quality guidance explicitly recognizes that purpose changes a user's assessment of the same statistical product. (www.abs.gov.au) A source might document that a publisher reported a relationship while providing insufficient evidence that the relationship remains current. Both assessments can be coherent because they concern different propositions.

Identity and version are separate tests. DataCite distinguishes a DOI representing all versions from version-specific identifiers, and leaves major versus minor version distinctions to individual stewards. (support.datacite.org) Zenodo's Concept DOI resolves to its latest version, while its guidance recommends a specific version DOI when the exact artifact must be reproducible. (zenodo.org) Record what the identifier identifies, rather than treating persistence as immutability.

Discovery also differs from evidence admission. Data.gov describes its role as a metadata catalog and states that it does not control the underlying data's quality, accuracy or availability. (data.gov) A catalog result is therefore a candidate to inspect. It is not an admission verdict.

Finally, a graph's missing relationship should not become a negative assertion merely because a search returned no match. Razniewski and colleagues' 2023 arXiv survey explains that under the open-world assumption an absent statement is unknown. (arxiv.org) Permission to infer an exhaustive negative requires its own coverage argument, independent of successful retrieval.

Source admission gates and the evidence card

The proposed **admission gates** identify evidence that must be resolved for the intended use. They should be applied individually. An aggregate score can obscure a failed permission test or an unknown observation date beneath strengths elsewhere. The gates below synthesize practices rather than claiming a standardized ranking.

Permission and licence come first. DWBP asks publishers to provide the agreement controlling data use, and Dublin Core defines a licence as a legal permission document. (www.w3.org) (www.dublincore.org) Record the actual terms covering the acquired distribution and intended operations, including transformation,

retention, access by readers and export. Public availability alone does not answer those questions. Data.gov specifically directs users to different terms and licences for non-federal datasets. (data.gov)

Terms must be read at their documented scope. **Creative Commons Attribution 4.0 International** requires indicating modifications when sharing licensed material and does not license privacy or similar personality rights. (creativecommons.org) The **Open Database License** does not independently cover database contents individually. (opendatacommons.org) CC BY 4.0 also contains a qualified disclaimer of warranties, including accuracy and fitness for purpose. (creativecommons.org) These are examples of why “has an open licence” is an incomplete admission finding. They are not conclusions about an unspecified candidate dataset's rights.

Provenance, coverage and limitations follow. DWBP calls for origins and explicit metadata about known quality issues. (www.w3.org) The USGS metadata policy describes spatial and temporal extents and calls for relevant information about provenance, quality and meaning beyond mandatory metadata elements. (www.usgs.gov) Use those practices to ask for the source's population, collection method, transformations and exclusions.

Table 1 supplies a reusable **source-admission evidence card**. Its entries are instructions for completing a card, not observations about a real candidate.

Field	Test	Observed value to record	Evidence URL and date	Decision effect	Review trigger
Intended decision	Can the proposition be evaluated?	Named question, entity scope and period	Decision specification URI and approval date	Defer if use is undefined	Question or consequence changes
Permission and licence	Do documented terms cover the use?	Terms, restrictions or not found	Licence URL, retained terms and access date	Defer pending terms; reject prohibited use	Terms or access rights change
Stable identity	What does the identifier identify?	Dataset ID, distribution ID and entity namespace	Identifier documentation and inspection date	Defer unresolved identity	Namespace or mapping changes
Version	Can the evaluated state be recovered?	Version ID, file fingerprint and retained reference	Snapshot locator and retrieval date	Defer irreproducible state	Content or version changes
Provenance	Who produced and transformed the evidence?	Producer, publisher and transformation chain	Origin documentation and verification date	Scope use to supported assertions	Producer or method changes
Dates	Which clock supports the question?	Observation, publication and retrieval dates, or unknown with reason	Exact supporting passage and date	Defer date-dependent use if missing	New observation or revision
Coverage	What population and exclusions apply?	Documented population and sample denominator	Methodology URL and evaluation date	Reject unsupported exhaustive inference	Population or exclusions change

Field	Test	Observed value to record	Evidence URL and date	Decision effect	Review trigger
Corrections	How are revisions communicated?	Correction contact, policy or not found	Policy URL and last review date	Set review plan or defer unstable use	Correction notice appears
Quality issues	Which limitations affect this proposition?	Known issues and sample findings separately	Issue documentation and sample date	Accept with explicit limits or defer	New limitation or test result
Evidence gap	What remains unavailable?	Not found, blocked or not applicable, with reason	Search or access record and date	State effect on this decision	Missing evidence becomes available

The table deliberately separates an observed value from its evidence locator. “Not found” means the evaluator did not locate the required evidence within the recorded search scope. “Blocked” means it could not be accessed. “Not applicable” requires a reason. None should be an empty cell, and none substitutes for permission to make the intended inference.

For private information, document access authority and the readers permitted to use derived evidence. FAIR, meaning **Findable, Accessible, Interoperable and Reusable**, explicitly allows authentication and authorisation and calls for a clear usage licence. (www.gofair.foundation) FAIR reusability also calls for detailed provenance. (www.gofair.foundation) Restricted access can therefore be documented without describing the source as freely reusable.

The minimum operational checks are:

- **Licence scope:** Distinguish the dataset, individual contents and intended output.
- **Identity scope:** Distinguish the dataset identifier from identifiers for real entities.
- **Temporal scope:** Require evidence for the relevant period, not merely a retrieval timestamp.
- **Population scope:** Preserve inclusion rules, exclusions and missing segments.
- **Correction scope:** Record who can revise the data and how users learn about it.
- **Evidence gap:** Name the unresolved item and the decision it prevents.

Measure source quality with explicit denominators

Begin measurement by declaring **the unit of evaluation**. It might be a record, a field, a relationship, a source passage or an observation. A ratio without that unit cannot be interpreted. Official frameworks illustrate why quality cannot be collapsed to one dimension: Statistics Canada defines six dimensions, while ABS advises considering seven. (www150.statcan.gc.ca) (www.abs.gov.au) These are different framework structures.

Scope the sample before counting

Specify the sampling frame, extraction filters, source version, selection method and exclusions before inspecting outcomes. Statistics Canada recommends unbiased or minimally biased sampling procedures and checking representativeness. (www150.statcan.gc.ca) If the pilot deliberately targets difficult cases, state that

design. Its observed error proportion describes those cases; it should not be presented as an unbiased population estimate.

Administrative records also require interpretation of their original business context. Statistics Canada recommends consulting providers about concepts and processing. (www150.statcan.gc.ca) A relationship collected to route internal work may not mean the contractual relationship a later analyst wishes to infer. This is a semantic question that field completeness cannot resolve.

The sample record should include:

- **Frame:** The eligible records and how they were enumerated.
- **Selection:** Random, stratified, convenience or targeted selection, with the actual procedure.
- **Applicability:** Rules separating required items from justified exclusions.
- **Numerator:** The items meeting the stated test.
- **Denominator:** Every sampled applicable item eligible for that test.
- **Missingness:** Missing, inaccessible and uninterpretable evidence counted separately.
- **Uncertainty:** Limits on extrapolation, including selection bias and reference uncertainty.

The Census Bureau distinguishes nonresponse, coverage, processing and measurement errors. (www.census.gov) Preserve those causes in sample findings rather than labeling every missing value a retrieval failure.

Calculate documented coverage

The proposed calculation is **observed coverage = documented applicable items / sampled applicable items**. “Documented” means the evidence meets the declared documentation test. It does not mean the assertion is independently correct.

Exclude an item only through a predeclared applicability rule, and retain the exclusion count. Changing the denominator after discovering missing evidence would change the question being measured. A related completeness example in DQV compares represented objects with expected objects under a declared dataset scope; its worked value is illustrative, not a benchmark. (www.w3.org)

Separate this measure from **precision** and **recall**. Manning, Raghavan and Schütze's information retrieval text defines precision over retrieved documents and recall over relevant documents. (nlp.stanford.edu) Transposing those terms to a graph requires a declared relevance or correctness criterion and a reference set. Razniewski and colleagues note inconsistent use of completeness and recall terminology in the literature. (arxiv.org)

Calculate freshness without changing clocks

The proposed **freshness lag = evaluation date minus source observation date**, reported with a unit. Also record the sample method and the meaning of “observation.” Dublin Core's issuance date concerns formal publication; it cannot simply be substituted for a date when the underlying state was observed. (www.dublincore.org)

If an observation date is absent, report freshness as **unmeasurable**, with the dated-item denominator separately stated. A web page retrieved today may contain old observations. Statistical punctuality, which Statistics Canada defines through planned versus actual availability, answers a different question. (www150.statcan.gc.ca)

Population inference requires a suitable sampling design. Webber's 2012 arXiv paper on recall estimation uses random samples and emphasizes the reliability of the estimate. (arxiv.org) A targeted pilot should report outcomes and limits on extrapolation.

Represent versions and measurements with DCAT and DQV

DCAT supplies catalog metadata; **DQV** supplies quality descriptions and measurements. Their representations can be combined without treating either as an admission authority. The UK government's guidance uses DCAT to describe datasets and services in a catalog and links complementary open standards. (www.gov.uk) The organizational decision still requires the evidence card's permission, scope and missing-evidence findings.

Record the evaluated dataset state

Represent the conceptual dataset and the particular distribution evaluated. In DCAT 3, **dcat:version** is a resource's version indicator; version relations include **dcat:previousVersion**, **dcat:hasVersion** and **dcat:hasCurrentVersion**. (www.w3.org) **adms:versionNotes** can describe changes from the preceding version. (www.w3.org) Use a dataset version identifier to bind the decision to a state, not merely a title or moving endpoint.

A **dcat:DatasetSeries** groups separately published datasets with shared characteristics. (www.w3.org) A monthly series and a corrected version of one monthly dataset need different descriptions. Otherwise, the card may mistake a new period for a correction, or evaluate a latest-version link while retaining evidence from an older distribution.

Alongside catalog metadata, this report proposes retaining:

- **Resource binding:** The dataset, distribution and version evaluated.
- **Snapshot binding:** The authorized retained artifact or reproducible access reference.
- **Acquisition record:** Retrieval time, endpoint and access conditions.
- **Decision binding:** The named use and evaluator responsible for the admission.
- **Change description:** Differences that matter to the accepted proposition.

These additions are workflow guidance, not a claim that DCAT prescribes a particular storage implementation.

Record the metric and how it was produced

DQV's **dqv:QualityMeasurement** represents a quantitative or qualitative metric value. Its properties connect the measured resource through **dqv:computedOn**, identify the metric through **dqv:isMeasurementOf**, and express the result through **dqv:value**. (www.w3.org) The metric definition must explain its unit, numerator, denominator, exclusions and evaluation procedure.

For the proposed observed-coverage metric, bind the result to the actual sampled resource or assessment artifact and retain the sample-selection description. If the result concerns a subset, document that scope explicitly; identify a partial assessment as such. DQV includes an example of provenance for a single quality measurement, including the generating activity and source distribution. (www.w3.org)

A practical measurement package should retain:

- **Definition:** The metric and the proposition it can inform.
- **Inputs:** The evaluated distribution, version and sample artifact.
- **Execution:** The method, assessment date and responsible agent.
- **Result:** The observed value and unit, or unmeasurable with reason.
- **Limits:** The population and inference boundaries.

Published implementations reinforce the importance of measurement origin. NDE distinguishes distribution-health enrichment produced by its register from publisher metadata. (docs.nde.nl) It also documents that shallow register validation and full-distribution validation can use the same measurement shape, with the serving endpoint distinguishing their scope. (docs.nde.nl) The vocabulary shape alone consequently does not reveal how much evidence was tested.

Treat an admission verdict as a separately attributable assessment. A numeric result may inform it, but **accept** is a decision about a specified use, not a number that certifies the dataset universally.

Decide, retain and re-admit

An admission decision should state **what may be relied on**, the limits of reliance and the event that will force reconsideration. ABS recommends quality statements describing both strengths and limitations. (www.abs.gov.au) The proposed decision meanings are:

- **Accept:** Required evidence is present for the named proposition and documented use.
- **Defer:** A recoverable evidence gap prevents the decision; name the missing item and next action.
- **Reject:** The source cannot support that proposition under the stated scope or permission conditions.
- **Review trigger:** Name the change or scheduled event requiring a new assessment.
- **Owner:** Assign responsibility for reviewing the trigger and updating the verdict.
- **Retained rationale:** Preserve the evidence and prior decision so a later reviewer can reconstruct it.

Acceptance of a publisher-reported value should preserve that attribution. It should not silently become acceptance of an independently verified real-world fact. Similarly, an accepted historical snapshot does not necessarily support a current-state decision.

Source changes that force re-admission

Re-admission is needed when a change invalidates an accepted condition. Evaluate changed licence terms, entity identifiers, collection methods, schemas, geographic or temporal coverage, correction notices and observation dates. USGS policy requires metadata updates reflecting data changes, while Crossref's

Crossmark documentation requires an update-policy page explaining post-publication changes. (www.usgs.gov) (www.crossref.org) Crossmark also documents different DOI records for editorially significant updates. (www.crossref.org) These supply concrete documentary practices for a change-review plan.

File and metadata changes can be distinct. Zenodo's version-management guidance uses versioning for updated published files. (help.zenodo.org) DataCite allows minor content changes under an existing DOI, which is why identifier equality alone is insufficient evidence that an evaluated artifact has remained unchanged. (support.datacite.org)

For web sources, **Hypertext Transfer Protocol**, or HTTP, validators can help detect representation changes. RFC 9110 defines **Last-Modified** through when the origin server believes the selected representation was modified; an **entity tag**, or ETag, is an opaque representation validator. (www.rfc-editor.org) Neither field establishes the observation date or accuracy of the underlying assertion. A changed validator triggers inspection; an unchanged one is not a universal factual-freshness guarantee.

Choose the supporting tool by the missing capability

An evidence card is a decision artifact. Discovery, measurement and record stewardship require different capabilities. The European Data Portal's metadata-quality documentation explicitly excludes intrinsic data correctness and legal validity from its assessment. (dataeuropa.gitlab.io)

Table 2 compares those roles, including the publisher's directly related system-of-record offering. It reports documented scope rather than a product benchmark.

Need	Appropriate category or option	Evidence to demand	Boundary
Find candidate datasets and services	Data catalog using DCAT metadata	Catalog scope, distribution links and metadata dates	Discovery leaves admission unresolved. (data.gov)
Execute repeatable profiling and validation	Data-quality suite, with measurements represented using DQV where useful	Metric definitions, sample scope, results and assessment provenance	A quality description supports a judgment about fitness. (www.w3.org)
Preserve defensible values after admission	**Reliable Context**, system of record	Its published guarantees: canonical records, cited dated values, declared unknowns and reproducible past states	These are publisher claims, not measurements of candidate-source reliability.
Exchange retained value evidence	System of record with evidence export; Reliable Context describes this capability	Source, excerpt, locator and dates accompanying an exported value	Exportability does not supply missing source evidence.

The distinction gives readers a concrete routing decision. Use a catalog when discovery is missing and a data-quality suite when automated checks are missing. When the requirement is defending retained values, assess system-of-record guarantees and test them against the evidence card. No tool category removes the obligation to establish permission or fitness for the named use.

Data Analysis and Evidence

The quantitative foundation concerns **methods and denominators**, not a universal dataset reliability score. Statistical frameworks demonstrate documented variation: Statistics Canada names six dimensions, ABS discusses seven, and the United Nations Statistical Commission adopted its National Quality Assurance Frameworks Manual at its fiftieth session in March 2019. (www150.statcan.gc.ca) (www.abs.gov.au) (unstats.un.org) NIST describes developing its framework through three plenary and fifteen stakeholder workshops from October 2021 to September 2023. (nvlpubs.nist.gov) Framework dimensions and workshop counts do not establish dataset reliability or validated admission thresholds.

Worked source-admission calculation (Hypothetical Example)

The following inputs are constructed solely to illustrate the proposed method. They are **not observed data**, a benchmark or a result attributed to any publisher. The calculation separates coverage and missingness, consistent with the Census Bureau's treatment of distinct nonsampling-error sources. (www.census.gov)

Assume an evaluator inspects **120 records** selected through a declared, deterministic pilot procedure. Under an applicability rule set before inspection, **20 records** do not require the relationship being evaluated. The remaining **100 applicable records** form the documentation denominator. **90** contain the required supporting evidence, so observed coverage is **90 / 100 = 90%**. The ten undocumented applicable records remain missing; the twenty justified exclusions are recorded separately. These are authored arithmetic results, not population estimates.

Among the ninety documented relationships, **70 have usable observation dates** and **20 do not**. Date ascertainability is **70 / 90 = 77.8%**, rounded to one decimal place. A documentation test and a date test therefore produce different findings on the same pilot. Data.gov's distinction between metadata check or modification dates and agency-supplied dataset information reinforces why a catalog timestamp should not fill the missing observation-date field. (data.gov)

Set the hypothetical evaluation date to **13 October 2026**. Of the seventy dated items, **50 were observed on 30 September 2026**, giving **13 calendar days** of freshness lag. The other **20 were observed on 1 September 2026**, giving **42 calendar days**. Dates are constructed inputs and lags are date-subtraction results. They describe age since observation, not elapsed time between reference period and publication.

Suppose this hypothetical decision specification requires observations no older than **30 days**. That is an authored policy choice for this example, not an industry recommendation. **50 / 70 = 71.4%** of dated items meet it, while **50 / 90 = 55.6%** of all documented items have demonstrated qualifying dates. Reporting only the first ratio would hide the undated evidence. The remainder has either an older date or no measurable lag.

Table 3 turns those findings into different decisions about the same hypothetical source.

Intended use	Observed evidence	Verdict and reason	Explicit evidence gap	Review trigger
Report what the publisher documented	Required support for 90 of 100 applicable sampled records	**Accept** extraction of the supported, attributed values within the pilot scope	Ten applicable records lack the required evidence; correctness is not independently tested	Source text or attribution changes

Intended use	Observed evidence	Verdict and reason	Explicit evidence gap	Review trigger
Treat every documented relationship as current under the example's policy	Qualifying dates for 50 of 90 documented items; 20 older and 20 undated	**Defer** the blanket current-state use	Updated observations or usable dates are needed for the remaining forty	New dated evidence or policy change
Infer that every relationship absent from the source does not exist	No exhaustive population denominator or closed-world coverage evidence	**Reject** that negative inference from this source	Missing statements are insufficient evidence of nonexistence. (arxiv.org)	Documented exhaustive scope becomes available

The accept verdict applies only to supported extractions and the specified reporting question. The defer verdict identifies a recoverable temporal gap. The reject verdict concerns an unsupported inference, without declaring the dataset generally unreliable.

Accuracy remains unmeasurable in this example because no independent reference or adjudication procedure was supplied. Precision and recall likewise require their respective relevance denominators; the information retrieval definitions do not permit treating documentation coverage as either metric automatically. (nlp.stanford.edu) There is also no claim of statistical confidence from this deterministic pilot. Population estimates would require a suitable sampling design and uncertainty analysis. (arxiv.org)

Implications and Future Directions

Source admission should remain connected to the questions a graph is expected to answer. A change in intended decision can alter the relevant evidence even if the files remain identical. Statistical quality guidance supports that contextual treatment, and Eurostat explicitly separates assurance, assessment and quality documentation. (ec.europa.eu) Organizations can therefore retain one source while recording different verdicts for historical attribution, current-state inference and exhaustive negative claims.

The evidence card also makes **semantic uncertainty** visible. An internally collected field and a public description may share a label while referring to different concepts. Statistics Canada's administrative-data guidance recommends validating concepts with providers and understanding processing. (www150.statcan.gc.ca) A mapping should document that interpretation, its supporting evidence and the boundary beyond which the equivalence has not been established.

Machine-readable metadata offers an opportunity to automate checks while preserving those limits. Wilkinson and colleagues' FAIR principles (2016) emphasize machines finding and using data, and extend beyond datasets to tools and workflows. (www.nature.com) That ambition concerns accessible, reusable descriptions. This report's inference is that machine-readable evidence cards can improve reviewability; no cited source establishes that they independently guarantee truth.

A practical development agenda would test whether cards improve decision reconstruction, expose missing evidence earlier and preserve different judgments across versions. Such tests should specify tasks, evaluator criteria and denominators before reporting results. **Citation potential**, adoption, search demand and

operational benefit for this particular card remain hypotheses; this report supplies no measured estimate for them.

Readers should also retain **uncertainty about source availability**. Data Commons documents that not every statistical variable has observations for every place or entity. (docs.datacommons.org) A successful query or a rich catalog entry does not remove those coverage differences. They belong in the admitted-use scope rather than being treated as a temporary inconvenience during ingestion.

The resulting process is deliberately revisable. A source may become admissible after its licence is clarified, its observation dates are supplied or its coverage is documented. A previously accepted use may require deferral after a relevant change. The card preserves the reason for both states, allowing reviewers to assess the evidence available when each decision was made.

Conclusion

Evaluating data sources for a knowledge graph begins with a specified question. It ends with a scoped admission decision that another reviewer can reconstruct. The proposed evidence card records permission, identity, version, provenance, dates, coverage, corrections and quality issues alongside the unresolved gap and review trigger.

A useful source can support a publisher-attributed historical statement while leaving a current-state or exhaustive negative inference unsupported. The card should preserve those distinctions. **Accept** means required evidence is present for the named use. **Defer** means a stated, recoverable gap prevents it. **Reject** means the proposed inference or use is unsupported under the documented conditions.

Measurements should state their unit, numerator, denominator, dates and selection method. Documentation coverage measures documented evidence within a sample. Freshness lag measures age since a defined observation. Missing denominators, absent dates and unavailable references should produce declared unmeasurable results rather than invented scores.

DCAT and DQV help represent datasets, versions and quality findings. Catalogs help discover candidates, data-quality suites execute checks, and systems of record preserve evidence behind retained values. Their roles should be selected according to the missing capability.

The operational outcome is a reviewable chain: a proposition, an evaluated source state, documented tests, measured or unavailable evidence, a decision and a reason to revisit it. Retaining that chain makes later changes assessable without implying that the original source was suitable for every possible question.

Source links

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

1. https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1520.0Main%20Features10May%202009
2. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/1500-18/NIST.SP.1500-18r2.html
3. https://www.w3.org/TR/vocab-dcat-3
4. https://www.w3.org/TR/dwbp

5. https://www.gov.uk/government/publications/recommended-open-standards-for-government/using-metadata-to-describe-data-assets-in-a-data-catalogue
6. https://www.w3.org/TR/vocab-dqv
7. https://www150.statcan.gc.ca/n1/pub/12-539-x/2019001/ensuring-assurer-eng.htm
8. https://arxiv.org/html/2305.05403v2
9. https://support.datacite.org/docs/versioning
10. https://zenodo.org/help/versioning
11. https://dataeuropa.gitlab.io/data-provider-manual/metadata-quality
12. https://arxiv.org/abs/2003.02320
13. https://www.microsoft.com/en-us/research/publication/datasheets-for-datasets
14. https://docs.datacommons.org/data_model.html
15. https://data.gov/user-guide
16. https://www.dublincore.org/specifications/dublin-core/dcmi-terms
17. https://creativecommons.org/licenses/by/4.0/legalcode.en
18. https://opendatacommons.org/licenses/odbl/1-0
19. https://www.usgs.gov/survey-manual/5027-fundamental-science-practices-metadata-usgs-scientific-data
20. https://www.gofair.foundation/fair-principles
21. https://www.census.gov/about/policies/quality/standards/standardd3.html
22. https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-unranked-retrieval-sets-1.html
23. https://arxiv.org/abs/1202.2880
24. https://docs.nde.nl/nl/services/dataset-register/data-model
25. https://www.crossref.org/documentation/crossmark/participating-in-crossmark
26. https://help.zenodo.org/docs/deposit/manage-versions
27. https://www.rfc-editor.org/rfc/rfc9110.html
28. https://unstats.un.org/unsd/methodology/dataquality
29. https://ec.europa.eu/eurostat/web/quality/quality-monitoring/quality-management
30. https://www.nature.com/articles/sdata201618

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.