Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

Entity Resolution Human Review: Which Decisions Need It?

> 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 · 2026-10-16 · entity resolution human review · human in the loop · entity matching · false merges · false splits · confidence thresholds · data provenance · context engineering · identity decisions

Original article: https://reliablecontext.com/articles/entity-resolution-decisions-human-review



In this article

1. Executive Summary
2. Introduction and Background
3. Define the identity decision and its consequences
4. Separate scores, probability, and sufficient evidence
5. Route decisions to acceptance, review, or unresolved status
6. Allocate scarce review capacity by expected harm
7. Preserve the decision card and choose the right system
8. Data Analysis and Evidence
9. Implications and Future Directions
10. Conclusion

Executive Summary

Entity resolution human review should be allocated where a person can materially improve a consequential identity decision. The proposed operating test has **three dispositions**: accept a policy-admissible action, review a case where expertise or additional evidence can change the outcome, or hold it unresolved when identity cannot be established. UK government guidance explicitly recognizes that insufficient evidence can defeat both human and algorithmic classification; reject-option research treats abstention as a legitimate designed outcome. (www.gov.uk) (www.jmlr.org)

A similarity score is neither automatically a calibrated probability nor sufficient evidence for identity. Guo and colleagues define calibration through estimates that represent actual correctness likelihood. OpenRefine documents differing reconciliation-score scales, while Zingg describes threshold variation across fields and sources. Consequently, this report recommends source-specific validation and explicit contradiction rules rather than a universal numeric cutoff. Review should cover proposed merges, proposed splits, and the clusters affected by either action. (proceedings.mlr.press) (openrefine.org) (www.zingg.ai)

HUMO partitions human and machine work, and r-HUMO prioritizes inspection using estimated risk. (arxiv.org) These studies motivate allocation methods but do not guarantee production reviewer performance. HUMO's experiments supplied hidden ground-truth labels when manual verification was requested. (arxiv.org) Operational disagreement also matters: ONS reports that approximately **15% of clerically made matches** received another expert review in its documented linkage process. (www.ons.gov.uk)

The report's **hypothetical calculation** holds capacity at **4 reviews and 2 reviewer hours**. Under its explicitly assumed probabilities and consequences, arrival-order inspection avoids **14.4 expected harm units**, compared with **64** for consequence-weighted ordering. These are illustrative arithmetic results, not benchmarks. A separate synthetic audit with **zero observed errors in 100 independent accepted decisions** yields a one-sided **95% upper error bound of 2.95%**, under exact-binomial assumptions. (www.itl.nist.gov) The reusable **decision card** records dated source evidence, conflicts, policy version, reviewer disagreements, disposition, and reversal conditions. Provenance standards support representing data-producing entities, activities, and responsibility; the card specifies the additional evidence this report recommends preserving. (www.w3.org)

Introduction and Background

Entity resolution identifies records that refer to the same real thing. In a knowledge system, the consequential step is deciding whether evidence permits an identity assertion that other people and applications will reuse. Doidge, Christen, and Harron distinguish an actual match from an assigned link and from agreement between identifying fields. Those distinctions prevent a convenient classification from being mistaken for established identity. (www.gov.uk)

Human in the loop entity resolution therefore needs a decision policy, not simply a queue of middling scores. A reviewer may approve identity, reject the proposed identity, request evidence, or conclude that available information cannot decide. The reviewer must distinguish evidence that can be obtained from evidence that is merely missing. Splink documents cases with too little information to confirm a match. Assigning such a case to a person does not itself supply distinguishing evidence. (moj-analytical-services.github.io)

The research literature offers useful allocation methods. **HUMO**, the human and machine cooperation framework described by Chen and colleagues, partitions entity-resolution work between humans and machines. **r-HUMO**, developed by Hou and colleagues, prioritizes inspection using risk analysis informed by human labels. These are study-specific frameworks, rather than universal operating guarantees for every knowledge system. (arxiv.org)

This report proposes an **observable decision test**: specify the action, its consequences, admissible evidence, and reversal condition; then decide whether to accept, review, or hold unresolved. The test is an analytical proposal, not a validated benchmark or a description of a vendor's internal matching mechanism. It combines established distinctions about linkage quality with explicit operational choices. NIST's Artificial Intelligence Risk Management Framework similarly calls for human oversight processes to be defined, assessed, and documented. (airc.nist.gov)

The intended users are context engineers and data stewards approving identity decisions whose consequences extend beyond a search result. The relevant question is whether another reviewer can reconstruct the evidence and determine why the decision was permissible. The US Census Bureau's linkage standard makes replication and evaluation documentation an explicit requirement within its own statistical operations. (www.census.gov)

Define the identity decision and its consequences

A review policy should define its unit of work before selecting thresholds. Identity is a stronger assertion than association: the World Wide Web Consortium's Web Ontology Language, **OWL**, defines its sameAs property as asserting that two references denote the same thing. A shared location, common name, or organizational relationship does not by itself establish that proposition. (www.w3.org)

For this report, the following are operational definitions:

- **Candidate pair:** two records selected for comparison; selection establishes eligibility for investigation, not identity.

- **Canonical identity:** the durable referent under which a system groups accepted observations about one real thing.
- **False merge:** records of distinct things are accepted as one identity; this includes a mistaken link that contaminates a larger group.
- **False split:** records of one thing are maintained as distinct identities; a rejected candidate pair is one possible contributor.
- **Review capacity:** the reviewer time available during a stated operating period, including escalation and adjudication.
- **Uncertainty:** incomplete knowledge about identity, evidence reliability, expected consequences, or reviewer performance.
- **Abstention:** declining to issue a definitive identity classification while preserving an explicit unresolved state.

These definitions distinguish pair-level decisions from entity-level consequences. Government linkage guidance recommends examining whole clusters when records may be incorrectly merged or split. Sandia's benchmarking work likewise evaluates quality at the level of individual entities, where aggregate pair measures can obscure which entities remain difficult. (www.gov.uk) (www.sandia.gov)

False positives and false negatives require stated denominators. For a match classifier, a false positive is a proposed link between different entities; a false negative misses a true match. **Precision** is true matches among assigned links; **recall** is identified true matches among all true matches. A false merge in a connected cluster may affect more records than the pair that introduced it. (www.gov.uk)

Consequences need not be symmetric. This report recommends documenting the downstream cost of attaching an observation to the wrong identity separately from the cost of keeping relevant evidence apart. IBM's own matching documentation explicitly recognizes different organizational tolerances for false matches. That supports local policy design, not a universal cost ratio. (www.ibm.com)

Identity scope also matters. GLEIF's Legal Entity Identifier framework preserves separate identifiers for parent entities and their subsidiaries while publishing ownership relationships. ROR, the Research Organization Registry, records structural and temporal relationships among organizations. Those examples show why relationship evidence belongs in an identity review packet without automatically becoming evidence of sameness. (www.gleif.org) (ror.readme.io)

Separate scores, probability, and sufficient evidence

A **similarity score** describes a comparison according to a particular method. A **calibrated probability** estimates an event frequency whose relationship to observed correctness has been assessed. **Evidence sufficiency** asks whether the proposed action meets an explicit decision rule. These are separate tests. Guo and colleagues define calibration through probability estimates representative of actual correctness likelihood; Doidge and colleagues note that interpreting match weights as probabilities requires stronger assumptions than ranking candidates. (proceedings.mlr.press) (www.gov.uk)

OpenRefine documents that reconciliation services use different score calculations and ranges. **Zingg** says that thresholds calibrated on one field are not transferable to another and that the same threshold behaves differently across sources of different quality. Neither statement licenses borrowing a numeric acceptance threshold for an unrelated dataset. (openrefine.org) (www.zingg.ai)

The proposed evidence policy should require these checks:

- **Meaning:** identify whether the output is a distance, weight, ranking score, estimated probability, or rule outcome.
- **Population:** state the source combinations and candidate-generation process on which probability estimates were assessed.
- **Validation:** test held-out decisions whose labels were established independently of the model being evaluated.
- **Locality:** examine source combinations and identity patterns separately where aggregate results may hide differences.
- **Authority:** define which identifiers and source observations can support the particular identity assertion.
- **Contradiction:** specify which conflicts prohibit acceptance until reconciled or explicitly adjudicated.

Probability cannot substitute for source authentication. ORCID, the Open Researcher and Contributor ID organization, cautions against linking identifiers through names alone because researchers can share names. Its guidance also requires shared unauthenticated identifiers to be marked as unauthenticated. A copied identifier and a verified identifier should therefore have different evidentiary treatment under the proposed policy. (info.orcid.org)

Nor does calibration transfer automatically between tasks. Guo and colleagues studied image and document classification datasets. Their work establishes a useful probability concept, but does not validate an entity-resolution threshold for a particular knowledge system. An organization must assess the probability interpretation on its own relevant decision population. (proceedings.mlr.press)

The practical threshold question is consequently conditional: **which actions are admissible at this level of evidence, in this source combination, with these consequences?** Lowering a match threshold generally trades precision for recall, as Dedupe's documentation explains, but that tradeoff does not determine whether a conflicting identifier may be overridden. (docs.dedupe.io)

The review packet should expose both supporting and conflicting observations before presenting a disposition. Splink's labeling interface allows predictions to be hidden because they could bias the clerical labeler. An independent first judgment followed by comparison with the machine's recommendation is therefore a defensible evaluation design, though its usefulness should be tested locally. (moj-analytical-services.github.io)

Route decisions to acceptance, review, or unresolved status

The proposed three-way disposition concerns **whether a decision may be issued**. It does not create a third kind of identity relationship. Within acceptance, the approved action can be to join records or keep them separate. Review is temporary allocation of human effort; unresolved is a recorded epistemic outcome.

Traditional linkage methods already distinguish high-weight links, low-weight non-links, and an intermediate clerical-review region. The Australian Bureau of Statistics describes this arrangement and the resource demands of clerical review. The proposal here adds an explicit test for cases that should remain unresolved even after inspection. (www.abs.gov.au)

Accept when the stated evidence policy permits the proposed action, material contradictions have been addressed, and the consequences fall within the authorized decision boundary. **Review** when an appropriately skilled person can plausibly change the disposition by checking evidence, resolving a conflict, or applying a documented exception. **Hold unresolved** when decisive evidence is absent, inaccessible, or insufficiently discriminating.

The distinction is consistent with reject-option classification, which permits abstention in uncertain cases. Franc, Prusa, and Voracek analyze selective risk and coverage; Bartlett and Wegkamp model declining classification at a specified cost. These papers support treating abstention as a designed outcome, without supplying a ready-made identity policy. (www.jmlr.org) (jmlr.csail.mit.edu)

Table 1 presents **synthetic, hypothetical examples**. Every disposition assumes the policy stated in its row; the examples are not measured vendor behavior.

Hypothetical evidence pattern	Proposed action	Observable disposition and reason
Similar names and a shared office address, with no verified unique identifier	Merge records	**Hold unresolved.** The observations do not distinguish co-located entities. Review only if another permitted source can add evidence.
Historical and current names, with verified identifier continuity and compatible dated observations	Keep one identity	**Accept**, if the policy recognizes continuity across this name change.
Matching names but conflicting identifiers that each appear authoritative	Merge records	**Review** the identifier scope, issuer, dates, and source authenticity; hold unresolved if the conflict remains.
A parent and subsidiary appear together in an ownership filing	Merge records	**Accept separation** under a policy distinguishing legal entities; retain the relationship separately.
An identifier appears on a copied directory entry without authenticated attribution	Merge records	**Review** the identifier's origin; repeated copies are not assumed to be independent corroboration.
A proposed pair would connect otherwise separate clusters with contradictory observations	Join clusters	**Review the affected clusters**, not merely the bridge pair.
A later observation suggests that a previously combined record contains two entities	Split an identity	**Review** the earlier evidence and downstream effects; record the split basis and superseded decision.

The name-change and relationship examples reflect distinctions found in registry documentation. ROR updates an existing record for a simple name change and distinguishes temporal and structural relationships; GLEIF separately identifies parent and child legal entities. These sources support the distinctions, not the hypothetical case outcomes. (ror.readme.io) (www.gleif.org)

Merge and split decisions deserve the same evidence discipline. A split should identify which observations belong to each resulting entity and which remain unattributed. A merge should identify the records and dependent assertions affected. This recommendation follows the need for cluster-level assessment rather than assuming that a satisfactory pair score establishes a satisfactory entity. (www.gov.uk)

An unresolved decision should carry a reason and a reopening condition. Useful reasons include insufficient distinguishing evidence, blocked source access, contradictory identifiers, and undefined identity scope. Splink documents records that lack enough information to confirm a match; UK guidance states that neither humans nor algorithms can accurately classify cases without sufficient matching evidence. (moj-analytical-services.github.io) (www.gov.uk)

Allocate scarce review capacity by expected harm

Review allocation has multiple objectives. **Active learning** selects labels to improve a model; Settles describes achieving greater accuracy with fewer labeled training examples. **Decision remediation** selects cases to correct consequential current outputs. Dedupe's uncertain-pair interface illustrates the learning objective, while r-HUMO explicitly uses risk analysis to prioritize inspection. These purposes may overlap but should be measured separately. (minds.wisconsin.edu) (docs.dedupe.io) (arxiv.org)

For a proposed merge, let **p** be an assessed match probability and **C_FM** the consequence assigned to a false merge. Under this report's simplified decision model, expected false-merge loss is **(1 − p) × C_FM**. A proposed separation instead has expected false-split loss **p × C_FS**. The harm weights are explicit policy inputs, not learned facts or universal monetary values. A complete policy should also compare the consequences of holding a case unresolved, delaying action, and introducing an incorrect decision during review. Reject-option research explicitly treats declining classification as having a cost. (jmlr.csail.mit.edu)

If review prevents a fraction **q** of erroneous merges and takes **t** hours, a simple review priority is **q × (1 − p) × C_FM / t**. This is an illustrative expected-benefit-per-hour calculation. It assumes review does not introduce new errors and excludes several operational costs. The r-HUMO paper's own pair-level loss concerns the probability of an incorrect label; adding consequence weights here is the report's extension, not an attributed r-HUMO guarantee. (arxiv.org)

The formula explains why **highest uncertainty need not mean highest review value**. A relatively confident proposal can warrant inspection when an error would have much larger consequences. NIST's risk-management framework similarly prioritizes treatment using impact, likelihood, and available resources. Actual allocation should also account for reviewability: a costly but unresolvable case may warrant a hold rather than repeated inspection. (airc.nist.gov) (www.gov.uk)

A practical allocation policy should record:

- **Eligibility:** whether review can obtain new evidence or apply expertise that could change the action.
- **Expected benefit:** the preventable harm under documented probability and consequence assumptions.
- **Effort:** initial review, specialist escalation, disagreement adjudication, and evidence retrieval time.
- **Exploration:** a reserved audit sample outside the prioritized queue to test what the policy misses.

- **Disagreement:** independent judgments and the reason for adjudication, without replacing their original labels.
- **Expiry:** the date or evidence change that triggers reevaluation of an unresolved case.

Reviewer disagreement is evidence about the process. The UK Ministry of Justice notes that clerical labeling results vary by labeler. The Office for National Statistics, ONS, describes repeated review and referral of conflicting decisions for another opinion. A final label should therefore retain the disagreement and adjudication history rather than silently converting consensus into certainty. (www.gov.uk) (www.ons.gov.uk)

Capacity accounting must include work outside the visible inspection queue. The r-HUMO study distinguishes selected inspection work from sampling, although both contribute to human cost. Comparing policies on inspection counts alone can therefore misstate their operational burden. This report recommends reporting reviewer hours and the accounting boundary together. (arxiv.org)

Preserve the decision card and choose the right system

An **entity resolution audit trail** should explain why the identity action was permissible at the time. The World Wide Web Consortium's **PROV** model represents entities, people, and processes involved in producing data, including when entities were generated or used. It also treats a revised result as a new entity, providing a conceptual basis for versioned decision evidence. (www.w3.org)

The following **decision card** is a proposed reusable record, not a claim that any cited product implements every field:

- **Decision identifier:** a stable reference to the proposal and its eventual disposition.
- **Identity scope:** the kind of thing being identified and the boundary of continuity over time.
- **Affected records:** candidate records, existing canonical references, and impacted clusters.
- **Proposed action:** merge, maintain separation, split, or defer, with its expected downstream consequences.
- **Source evidence:** dated observations, verbatim supporting passages, exact source locations, and access limitations.
- **Conflicting evidence:** incompatible identifiers, dates, attributes, and source dependencies.
- **Score interpretation:** the output's meaning, validation population, and relevant uncertainty.
- **Policy version:** the acceptance rule, evidence requirements, exception authority, and applicable date.
- **Human judgments:** reviewer roles, independent labels, disagreements, and adjudication rationale.
- **Final disposition:** the action authorized, or an explicit unresolved reason.
- **Reversal condition:** evidence or policy changes that require reopening the decision.
- **Historical reconstruction:** preserved evidence and prior decisions sufficient to reproduce the earlier state.

An action log alone does not demonstrate that the card is complete. **AWS CloudTrail** records Entity Resolution application programming interface, or **API**, calls as events; AWS rule-based matching also outputs the rule responsible for a match. Those are useful components of traceability. The recommendation to

preserve source quotations, policy versions, and adjudication rationale is additional to the specific capabilities established by those documents. (docs.aws.amazon.com)

Table 2 compares documented system roles with the evidence test. **Reliable Context** appears only to identify the publisher's declared position; this report makes no independent product-performance assessment.

System or category	Documented or declared role	Question for the decision owner
Splink, entity-resolution tooling	Compares clerical labels with predicted probabilities to identify false positives and false negatives. (moj-analytical-services.github.io)	Can the evaluation cover candidate omissions and affected clusters as well as reviewed pairs?
AWS Entity Resolution, matching service	Supports customized exact-match rules and records the rule that generated a match. (docs.aws.amazon.com)	Does the surrounding process preserve the evidence and authorization behind the action?
IBM Master Data Management, master-data system	Consolidates enterprise data into master entities and supports manual links, unlinks, and remediation inputs. (www.ibm.com)	Are operational stewardship, entity maintenance, and coordinated record changes the main requirement?
OpenRefine, reconciliation workspace	Documents reconciliation as requiring human review and approval. (openrefine.org)	Is the task a bounded reconciliation exercise with explicit human judgments?
Reliable Context, publisher position	Declared category: system of record. No implementation or performance comparison is made here.	Apply the same observable evidence and historical-reconstruction test.

These categories answer different needs. A team that needs to build and evaluate a matching process should assess dedicated entity-resolution tooling. A team whose main requirement is enterprise master-entity consolidation should assess a dedicated **master data management**, or **MDM**, system; IBM's documentation explicitly describes that role. The decision card evaluates the surrounding evidence process and does not replace those capabilities. (www.ibm.com)

Additional documentation can help substantiate individual card fields. Senzing exposes the resolution rule that triggered a decision and represents relationships separately. OpenRefine retains the original string alongside its matched entity link. Neither documented feature alone establishes the completeness of an organization's historical evidence record. (senzing.com) (openrefine.org)

Data Analysis and Evidence

Published evidence and its limits

Chen and colleagues' HUMO paper identifies its revised version as an **ICDE 2018** camera-ready paper. (arxiv.org) In its particular DBLP-Scholar and Abt-Buy experiments, approximately **7% and 12% manual work**, respectively, supported precision and recall requirements of **0.9**. (arxiv.org) These are dataset- and method-specific experimental results.

Crucially, the experiments supplied hidden ground-truth labels when manual verification was requested. (arxiv.org) They did not measure ordinary reviewer time or disagreement. HUMO also states that its guarantees cannot exceed human-attained quality. (arxiv.org) ONS provides a different kind of operational

evidence: approximately **15% of matches made in its clerical matching system** received another expert review. (www.ons.gov.uk) Neither figure establishes a universal review budget.

Fixed-capacity allocation (Hypothetical Example)

All quantities below are **synthetic assumptions**, not benchmarks. Assume a pool of **20 decisions**: **12** otherwise eligible merge proposals and **8** cases already held for insufficient evidence. Review capacity is **4 cases**, each taking **30 minutes**, or **2 reviewer hours**. Each reviewed erroneous merge is assumed to be corrected or withheld with probability **0.8**, with no new errors introduced.

Table 3 applies the report's expected-loss formula to the eligible proposals. Harm units have no monetary or empirical interpretation.

Hypothetical cohort	Cases; match probability	False-merge harm per error	Expected harm per proposal	Expected harm avoided per reviewed case
A, first in arrival order	4; 0.55	10 units	(1 − 0.55) × 10 = **4.5**	0.8 × 4.5 = **3.6**
B, higher consequence	4; 0.90	200 units	(1 − 0.90) × 200 = **20**	0.8 × 20 = **16**
C, intermediate consequence	4; 0.70	40 units	(1 − 0.70) × 40 = **12**	0.8 × 12 = **9.6**

Arrival-order review inspects cohort A and avoids **14.4 expected harm units**; consequence-weighted ordering inspects B and avoids **64**, over the same **2 hours**. Expected harm avoided per reviewed case is **3.6 versus 16**. This is a result of the assumptions, not evidence that a production policy achieves those benefits.

Suppose separately that one reviewed case remains unresolved. Together with the original eight holds, the final unresolved share is **9/20 = 45%**. The reviewed share is **4/20 = 20%**. These measures have different denominators and meanings; an unresolved outcome can be a successful review result.

Residual uncertainty (Hypothetical Example)

Assume an independent, representative audit of **100 accepted decisions** observes **zero errors** and reference labels are correct. Solving **(1 − e)^100 = 0.05** gives a one-sided **95%** upper binomial error bound of **e = 1 − 0.05^(1/100) = 2.95%**. This is the report's derivation using exact-binomial reasoning documented by NIST, not a measured error rate. (www.itl.nist.gov)

Cluster dependence, selective sampling, or uncertain audit labels invalidate that simple interpretation. Moreover, accepted-case audits do not establish recall: UK guidance says clerical review is more useful for precision, and ONS estimated recall by searching samples of unmatched records. Report the sampling frame alongside every bound. (www.gov.uk) (www.ons.gov.uk)

Implications and Future Directions

The immediate implication is to evaluate the **review policy** as well as the matcher. Statistical learning can improve which cases are scored highly, but operational accountability depends on which actions are permitted and what happens when evidence remains insufficient. Reject-option research formalizes the relationship

between selective risk and coverage; organizations still have to define acceptable consequences and evidence requirements. (www.jmlr.org)

A defensible evaluation should report:

- **Decision coverage:** the share receiving a definitive identity disposition, with unresolved cases reported separately.
- **Review workload:** cases, reviewer hours, escalation time, and the accounting period.
- **Residual error:** independently assessed accepted links and missed matches, with the sampling design stated.
- **Entity effects:** affected clusters and observations, rather than pair counts alone.
- **Reviewer agreement:** initial disagreement, adjudication outcomes, and cases still unresolved.
- **Source sensitivity:** results by relevant source combinations and evidence availability.

Candidate generation deserves particular attention. Dedupe warns that reduced fingerprinting recall can leave true duplicates never directly compared. Reviewing only proposed pairs cannot expose every such omission. Blackwell and Rogers recommend quality assessment of both linked and unlinked pairs; this supports a separate search for missed matches outside the priority queue. (docs.dedupe.io) (www.gov.uk)

Evaluation must also consider the resulting knowledge resource. Statistics Canada distinguishes validation of individual links from validation of the resulting analysis file, including representativeness. Applied here as an evaluation principle, correct decisions on a selected review set do not establish that the overall accepted record population is complete or fit for its intended use. (www.statcan.gc.ca)

Historical identity should remain inspectable when new evidence arrives. ROR recommends retaining historically correct organization names and identifiers in digital object identifier, or **DOI**, metadata even when organizations become inactive or acquire successors. This illustrates why current organizational state should not silently overwrite the meaning of a past observation. (ror.readme.io)

Future evaluations should compare allocation policies under the same capacity, with independent labels and documented treatment of unresolved decisions. The Census Bureau's standard includes training, referral criteria, and monitoring of clerical errors. Those are concrete process dimensions to assess alongside any improvement in a model score. (www.census.gov)

Conclusion

Entity resolution human review is most useful when it changes an identity decision that matters and supplies evidence or expertise the automatic process lacks. The central operational distinction is among an action already justified by policy, a case where review can change the evidence assessment, and a case whose ambiguity must remain visible.

The proposed test starts with a defined referent and a stated action. It separates similarity scores from validated probability estimates and from evidence sufficiency. It then considers false-merge and false-split consequences, affected clusters, and available reviewer time. Acceptance may authorize either joining records or maintaining separation; an unresolved result preserves the limits of current knowledge.

The hypothetical calculation shows why review allocation should consider preventable consequence as well as uncertainty. It also shows why reviewer hours, unresolved share, and residual error bounds answer different questions. None should be presented as a substitute for the others or promoted from an illustrative assumption to an observed benchmark.

The reusable decision card makes the policy reviewable: dated source evidence, contradictions, rule version, human judgments, final disposition, reversal conditions, and reconstructable history. Its value lies in making an identity assertion defensible to someone who did not make it. When the work requires dedicated entity-resolution tooling or a master-data platform, the appropriate system should support that work. The evidence policy remains the standard against which its observable decisions are assessed.

Source links

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

1. https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/quality-assessment-in-data-linkage
2. https://www.jmlr.org/papers/v24/21-0048.html
3. https://proceedings.mlr.press/v70/guo17a.html
4. https://openrefine.org/docs/manual/reconciling
5. https://www.zingg.ai/post/entity-resolution-at-scale-part-4-thresholds-active-learning
6. https://arxiv.org/abs/1710.00204
7. https://arxiv.org/html/1710.00204v2
8. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/linkagemethodsforcensus2021inenglandandwales
9. https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/exacbici.htm
10. https://www.w3.org/TR/prov-primer
11. https://moj-analytical-services.github.io/splink/api_docs/evaluation.html
12. https://airc.nist.gov/airmf-resources/airmf/5-sec-core
13. https://www.census.gov/about/policies/quality/standards/standardc4.html
14. https://www.w3.org/TR/owl-ref
15. https://www.sandia.gov/research/publications/details/entity-resolution-at-large-scale-benchmarking-and-algorithmics-2018-12-01
16. https://www.ibm.com/docs/en/ws-and-kc?topic=data-customizing-strengthening-your-matching-algorithm
17. https://www.gleif.org/en/lei-data/access-and-use-lei-data/level-2-data-who-owns-whom
18. https://ror.readme.io/docs/relationships
19. https://info.orcid.org/documentation/collecting-and-sharing-orcid-ids
20. https://docs.dedupe.io/en/latest/API-documentation.html
21. https://www.abs.gov.au/ausstats/abs@.nsf/mf/1351.0.55.034
22. https://jmlr.csail.mit.edu/papers/v9/bartlett08a.html
23. https://ror.readme.io/docs/ror-data-structure

24. https://minds.wisconsin.edu/handle/1793/60660
25. https://arxiv.org/abs/1803.05714
26. https://arxiv.org/html/1803.05714v3
27. https://www.gov.uk/algorithmic-transparency-records/moj-data-first-splink
28. https://docs.aws.amazon.com/entityresolution/latest/userguide/logging-using-cloudtrail.html
29. https://docs.aws.amazon.com/entityresolution/latest/userguide/creating-matching-workflow-rule-based.html
30. https://www.ibm.com/docs/en/ws-and-kc?topic=data-managing-master
31. https://senzing.com/docs/tutorials/get_entity_response
32. https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/longitudinal-linkage-of-administrative-data-design-principles-and-the-total-error-framework
33. https://www.statcan.gc.ca/en/sdle/overview

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.