Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

Wikidata Statements, References, Ranks and Qualifiers

> 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 · 2026-10-29 · wikidata statements references · wikidata ranks preferred deprecated · wikidata qualifiers · wikibase data model · knowledge graph provenance wikidata

Original article: https://reliablecontext.com/articles/wikidata-statements-references-ranks-qualifiers-provenance-model



In this article

1. The statement as the unit of assertion
2. Qualifiers: the context of a value
3. References: where a claim comes from
4. Ranks: preferred, normal, deprecated
5. Worked example: Berlin's population
6. Export to RDF
7. Mapping Wikidata constructs to provenance concepts
8. Conflicts and constraints
9. Lessons for enterprise knowledge graphs
10. Limits
11. Checklist for borrowing the pattern
12. Sources

The statement as the unit of assertion

Wikidata is the best-known public knowledge graph that stores claims rather than facts. Its own help pages say so plainly: Wikidata "does not state what the population of Germany actually is; it simply provides the information on what the population of Germany is *according to* a specific source" (Help:Statements, Wikidata). That choice shapes everything below.

The vocabulary comes from the Wikibase data model, the software model behind Wikidata (Wikibase/DataModel, MediaWiki):

- An **item** is a thing being described, identified by a QID such as Q64 (see the Wikidata QID reference). A **property** (P-number) names a kind of data and carries a datatype.
- A **snak** is the smallest assertion the model can make: a property with a value, or one of two special forms described below. The data model notes that editors never meet the word; it exists so developers do not confuse a snak with a statement.
- A **statement** has a subject, one **main snak**, zero or more **qualifier snaks**, a list of **references**, and a **rank**. The data model gives it as `Statement(Entity Snak {PropertySnak} {ReferenceRecord} Rank)`.

Older documentation separates a *claim* (main snak plus qualifiers) from a *statement* (claim plus references); the JSON documentation calls this distinction historical (Wikibase JSON format). Here, *statement* means the full structure; for the general concept, see the glossary entry on claims.

Three kinds of snak

A snak's `snaktype` is one of `value`, `somevalue` or `novalue` (Wikibase JSON format). The editing interface calls the last two **unknown value** and **no value**.

- **Value**: the property has this specific value.

- **Somevalue** (unknown value): the property has a value, but the value is not known. The data model's example is Ambrose Bierce's date of death: unknown, yet he is certainly not among the living (Wikibase/DataModel).
- **Novalue** (no value): the property has no value for this item. Help:Statements calls this "a positive statement" and gives the example of a person who had no children (Help:Statements).

Both separate a declared absence from incompleteness, and neither is a placeholder value: if "none" were a value, a query for items sharing a value would conclude that Australia and Iceland have a common neighbour (Vrandečić & Krötzsch, CACM 2014).

Qualifiers: the context of a value

A **qualifier** is a property-value pair attached to the statement, not to the item. The original design paper is explicit: for "population 2,651,040, as of 2010, method estimation", the extra pairs "do not refer to Rome but to the assertion that Rome has a population of 2,651,040" (Vrandečić & Krötzsch, CACM 2014).

Help:Qualifiers lists four jobs (Help:Qualifiers, Wikidata):

1. Modify what the item means ("France, excluding Adélie Land").
2. Say how the value was determined (*determination method*: estimation, census).
3. Constrain the validity of the value (*point in time*, *start time*, *end time*).
4. Add details a single pair cannot hold (a cast member's role in a film).

A statement "should still provide useful data even without a qualifier", and qualifiers modify the main value or the statement as a whole, never other qualifiers (Help:Qualifiers). When a property should have one value but sources report several, keep all of them and qualify each (Help:Statements).

References: where a claim comes from

A **reference** is a set of snaks describing a source, from a single URL to title, author, publisher and page (Wikibase/DataModel). The JSON documentation states that references "record provenance information for the data in the main Snak and qualifiers" (Wikibase JSON format). It supports the whole qualified assertion.

Help:Sources sets the working conventions (Help:Sources, Wikidata):

- *stated in* (P248) points to an item describing a publication, database or report; *reference URL* (P854) points to a web page.
- For web pages, add the publication date or, failing that, *retrieved* (P813), plus title and an archive URL "so that the source can be tracked down if the URL changes".
- Statements supported only by *imported from Wikimedia project* (P143) "are not considered sourced statements", and Wikipedia articles are not acceptable sources.

The data model permits an empty reference list: editors can add statements without a reference and others may supply one later (Wikibase/DataModel; see also Hogan et al., Knowledge Graphs, §10.1.4).

Ranks: preferred, normal, deprecated

A **rank** is a three-level marker on each statement (Help:Ranking, Wikidata):

- **Normal** is the default and "provides no judgement or evaluation of a value's accuracy and currency".
- **Preferred** marks "the most current statement or statements that best represent consensus". Several statements may be preferred at once.
- **Deprecated** marks statements "known to include errors" or representing outdated knowledge. It is not for correct historical values, which keep normal rank and carry *start time* and *end time* qualifiers.

The page recommends recording a reason with *reason for preferred rank* (P7452) or *reason for deprecated rank* (P2241). It also separates ranks from references: "References merely state where a data value comes from; ranks indicate what data value is considered the most correct" (Help:Ranking). Deprecating rather than deleting stops the value being re-added, and a deprecated statement must still be sourced, since the source really did make the claim.

From ranks the data model derives the **best rank** for a property on an item: preferred if any statement has it, otherwise normal. The "best statements" are those with the best rank (Wikibase/DataModel).

Worked example: Berlin's population

Figures are from Berlin's live entity JSON (Q64 entity data, Wikidata), read 28 October 2026; they will change.

Q64 carried 47 statements for *population* (P1082): 46 normal, 1 preferred, none deprecated. 29 of the 47 had no reference. Two statements describe the same reference date:

	Statement A	**Statement B**
Value	3,700,577	3,913,644
point in time (P585)	31 Dec 2025	31 Dec 2025
determination method (P459)	statistical updating (Q52679562)	resident register (Q15194024)
Reference	*stated in* "Gemeindeverzeichnis 2025" (Q141265959), *retrieved* 3 Sep 2026	two *reference URLs* to statistik-berlin-brandenburg.de
Rank	preferred	normal
reason for preferred rank	not recorded	n/a

Read with their qualifiers, the figures measure different things. Wikidata keeps both; rank picks the default for queries and infoboxes, without recording why.

Statement A in JSON, abridged:

```
{
  "id": "Q64$C5506726-07D8-4E52-8AB4-BB5FB9317501",
  "type": "statement",
```

```
  "rank": "preferred",
  "mainsnak": {"snaktype": "value", "property": "P1082",
    "datavalue": {"type": "quantity", "value": {"amount": "+3700577", "unit": "1"}}},
  "qualifiers": {
    "P585": [{"snaktype": "value", "datavalue": {"type": "time",
      "value": {"time": "+2025-12-31T00:00:00Z", "precision": 11}}}],
    "P459": [{"snaktype": "value", "datavalue": {"value": {"id": "Q52679562"}}}]
  },
  "references": [{
    "hash": "3384faddfcb22f1d6f76da385283f5e62ee583e1",
    "snaks": {"P248": [{"datavalue": {"value": {"id": "Q141265959"}}}],
              "P813": [{"datavalue": {"value": {"time": "+2026-09-03T00:00:00Z"}}}]}
  }]
}
```

Male and female population qualifiers are omitted. Precision 11 means significant to the day (Wikibase JSON format).

Export to RDF

The RDF dump writes every statement twice (Wikibase RDF Dump Format, MediaWiki):

- **Truthy statements** link the item directly to a value with a `wdt:` predicate. They cover "the best non-deprecated rank for a given property", and "the qualifiers are ignored".
- **Full statements** are separate nodes (`wds:`) linked from the item with `p:`. The node holds the rank (`wikibase:rank`), the value (`ps:`), qualifiers (`pq:`) and references via `prov:wasDerivedFrom`. Statements with the best rank also have type `wikibase:BestRank`.
- **Reference nodes** (`wdref:`) are named by a hash of their content, so identical references share a node; their values use `pr:`.

The same Berlin statements, from the entity's Turtle export (Q64 RDF, Wikidata), abridged:

```
wd:Q64 wdt:P1082 "+3700577"^^xsd:decimal .        # truthy: date and method gone

wd:Q64 p:P1082 wds:Q64-C5506726-07D8-4E52-8AB4-BB5FB9317501 .
wds:Q64-C5506726-07D8-4E52-8AB4-BB5FB9317501 a wikibase:Statement, wikibase:BestRank ;
    wikibase:rank wikibase:PreferredRank ;
    ps:P1082 "+3700577"^^xsd:decimal ;
    pq:P585 "2025-12-31T00:00:00Z"^^xsd:dateTime ;
    pq:P459 wd:Q52679562 ;
    prov:wasDerivedFrom wdref:3384faddfcb22f1d6f76da385283f5e62ee583e1 .

wdref:3384faddfcb22f1d6f76da385283f5e62ee583e1 a wikibase:Reference ;
    pr:P248 wd:Q141265959 ;
    pr:P813 "2026-09-03T00:00:00Z"^^xsd:dateTime .

wd:Q64 p:P1082 wds:Q64-d789393f-445d-f810-b5ce-84543997a922 .
wds:Q64-d789393f-445d-f810-b5ce-84543997a922 a wikibase:Statement ;
    wikibase:rank wikibase:NormalRank ;
```

```
    ps:P1082 "+3913644"^^xsd:decimal ;
    pq:P585 "2025-12-31T00:00:00Z"^^xsd:dateTime ;
    pq:P459 wd:Q15194024 .
```

Statement IDs have "no guaranteed format or meaning" (note `$` in JSON, `-` in RDF). An unknown value is a blank node, which the query service replaces with skolem IRIs tested via `wikibase:isSomeValue`. A no-value statement is a `wdno:` class on the statement or entity, defined in OWL as the complement of having any value for that property (Wikibase RDF Dump Format). Giving the edge its own node so context can attach to it is reification (Hogan et al., §3.3.2).

Mapping Wikidata constructs to provenance concepts

The export uses one W3C provenance term, `prov:wasDerivedFrom`. PROV-O defines derivation as "a transformation of an entity into another, an update of an entity resulting in a new one, or the construction of a new entity based on a pre-existing entity" (PROV-O, W3C Recommendation, 30 April 2013). The other rows are analogies, not part of the export.

Wikidata construct	What it records	Closest PROV-O concept	Gap
Statement node (`wds:`)	One attributable assertion	`prov:Entity`	No author field; editors appear in page history
Reference (`wdref:`) via `prov:wasDerivedFrom`	Source of the assertion	`prov:wasDerivedFrom`; `prov:hadPrimarySource` when the source is primary	Optional; points to a source, does not store a copy
retrieved (P813)	When the source was consulted	No direct term; `prov:generatedAtTime` is the nearest	Retrieval date is not the publication date
determination method (P459)	How the value was produced	Description of the generating `prov:Activity`	Free-form item, not a method record
point in time, *start time*, *end time*	When the value holds	None; valid time is outside PROV's scope	Qualifiers, not a temporal model
Rank	Editorial selection for default use	None	Selection is not provenance; reason optional
Deprecated rank + *reason for deprecated rank*	Claim kept but marked wrong	Loosely `prov:invalidatedAtTime`	No date of deprecation
Entity revision (`schema:version`, `schema:dateModified`)	When the record changed	`prov:generatedAtTime` of a record version	Whole-entity granularity, record time only

Conflicts and constraints

Wikidata does not prevent conflicting statements. **Property constraints** are statements on a property, using *property constraint* (P2302), that describe expected usage, and they are "hints, not firm restrictions" with allowed exceptions (Help:Property constraints portal, Wikidata). Four bear on provenance:

- **Single value**: items should have no more than one statement with this property.
- **Single best value**: items should have one "best" statement, for properties that change over time. Population (P1082) carries it; Berlin satisfies it with one preferred statement.
- **Citation needed**: statements should have at least one reference.
- **Mandatory qualifiers**: statements must carry the listed qualifiers.

Disputes are settled socially: other views may be added "as long as they include a source and appropriate qualifiers", and consensus "should be indicated by a preferred rank" (Help:Statements). A statement can also name its opponents with the qualifier *statement disputed by* (P1310) (Help:Ranking).

Lessons for enterprise knowledge graphs

Coexistence and election are different layers. Wikidata keeps every sourced claim and elects a default with rank. A system of record must name one current value per subject, property and date, because a filing needs one answer. Keep both layers: the claims with their references, and the election as a separate dated act. The Knowledge Vault article covers the automated version of the same problem, fusing extracted claims at web scale.

Record why a value was elected. Wikidata recommends but does not require a reason for a rank; Berlin has none. Store the rule or reviewer, date and evidence behind each election.

Qualify before you compare. Berlin's figures conflict only until method is read. Compare statements only when subject, property, scope, time and method align.

Declare unknowns explicitly. Somevalue and novalue separate "unknown" and "none" from "not yet entered"; enterprise records need the same, with a dated search boundary (declaring unknowns).

Do not publish only the truthy view. A `wdt:`-style projection drops dates, methods and sources; consumers, including AI agents, need the qualified statement.

Limits

- **Rank is editorial.** It records community consensus, any editor may change it, and it may have no stated reason (Help:Ranking).
- **References are optional.** The model allows empty reference lists, and 29 of Berlin's 47 population statements had none on the day read.
- **A reference is a pointer.** It names a source; it does not hold a copy. Archive URLs are recommended for web pages, not enforced (Help:Sources).
- **Revision history is not valid time.** The entity's revision counter and last-modified timestamp (Wikibase RDF Dump Format) record when the record changed, not when a value held in the world. Valid time lives in qualifiers, where present.

Checklist for borrowing the pattern

Question	Wikidata answer	Enterprise target
Can conflicting claims coexist?	Yes	Yes, with one elected value per subject, property and date
Is a source required?	Usually expected, not enforced	Required for every value
Is the reason for election stored?	Optional qualifier	Required: rule or reviewer, date, evidence
Are unknown and no value distinct from missing?	Yes, as snak types	Yes, with scope and search date
Is valid time separate from record time?	Qualifiers vs revision history	Both, queryable together
Can a past state be reproduced?	Page history per entity	Any past state, per value

This article is educational and is not legal or regulatory advice.

Reliable Context applies these lessons inside the customer's perimeter: one canonical record per real thing; a cited, dated source behind every value; unknowns declared; any past state reproducible.

Sources

1. Help:Statements, Wikidata (read 28 Oct 2026) — "according to" a source, unknown and no value, plurality and consensus.
2. Help:Qualifiers, Wikidata (read 28 Oct 2026) — four uses of qualifiers and the rules on qualifying.
3. Help:Sources, Wikidata (read 28 Oct 2026) — reference properties, archive URLs, P143 not a source.
4. Help:Ranking, Wikidata (read 28 Oct 2026) — preferred, normal and deprecated rank; reasons; ranks vs references; statement disputed by.
5. Help:Property constraints portal, Wikidata (read 28 Oct 2026) — constraints as hints; single value, single best value, citation needed, mandatory qualifiers.
6. Wikibase/DataModel, MediaWiki (read 28 Oct 2026) — snaks, statement structure, ReferenceRecords, empty reference lists, ranks and best rank.
7. Wikibase JSON format, Wikimedia documentation (read 28 Oct 2026) — snaktypes, claim vs statement, reference scope, time precision.
8. Wikibase/Indexing/RDF Dump Format, MediaWiki (read 28 Oct 2026) — truthy and full statements, prefixes, special values, revision metadata.
9. Berlin (Q64) entity data, JSON, Wikidata (read 28 Oct 2026) — population statements, ranks, qualifiers and references in the worked example.
10. Berlin (Q64) entity data, Turtle, Wikidata (read 28 Oct 2026) — RDF triples in the worked example.
11. Vrandečić & Krötzsch, "Wikidata: A Free Collaborative Knowledgebase", Communications of the ACM 57(10), October 2014 — design rationale for qualifiers, special values, plurality and ranks.
12. Hogan et al., "Knowledge Graphs", arXiv:2003.02320v6, September 2021 — reification (§3.3.2); Wikidata as a secondary source with optional references and contradictory claims (§10.1.4).

13. PROV-O: The PROV Ontology, W3C Recommendation, 30 April 2013 — terms in the mapping table.

Source links

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

1. https://www.wikidata.org/wiki/Help:Statements
2. https://www.mediawiki.org/wiki/Wikibase/DataModel
3. https://doc.wikimedia.org/Wikibase/master/php/docs_topics_json.html
4. https://cacm.acm.org/research/wikidata
5. https://www.wikidata.org/wiki/Help:Qualifiers
6. https://www.wikidata.org/wiki/Help:Sources
7. https://arxiv.org/abs/2003.02320
8. https://www.wikidata.org/wiki/Help:Ranking
9. https://www.wikidata.org/wiki/Special:EntityData/Q64.json
10. https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format
11. https://www.wikidata.org/wiki/Special:EntityData/Q64.ttl?flavor=full
12. https://www.w3.org/TR/prov-o
13. https://www.wikidata.org/wiki/Help:Property_constraints_portal

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.