

The most interesting thing about recent tooling around Wikidata is not that it helps a model find an entity faster. Search speed is useful, but it is not the hard part. The hard part is trust. If an agent links a person, company, place, or work to the wrong Wikidata item, the mistake does not stay local. It tends to spread into downstream records, summaries, and decisions. That is why inspectable evidence matters.

A good MCP server for knowledge work should do more than produce an answer that looks plausible. It should show what it saw, how it narrowed the field, and where uncertainty remains. That is the real value in the Wikidata + Google Knowledge Graph MCP project: it frames entity resolution as an evidence-driven process rather than a guessing game.

This project, published as an open-source MCP server and CLI, is designed to let AI agents search Wikidata, read selected facts, and link local records to Wikidata QIDs. The notable part is the way it handles evidence. It does not simply return a long pile of candidates and leave the model to improvise. It aims for bounded search, explicit outcomes, and inspectable data that a human or an agent can review. In practice, that makes the difference between a system you can audit and one you merely hope is right.

## **Why inspectable evidence changes the conversation**

Anyone who has spent time cleaning metadata or reconciling records knows the familiar failure mode. A name appears to match. A date sort of lines up. A description sounds close enough. The software picks a record, and later someone discovers that the “John Williams” in the catalog is not the composer, not the guitarist, but a third person entirely. The original match might have looked harmless, yet the fix can take hours because no one can tell why the system made the choice in the first place.

Inspectable evidence solves that problem at its root. Instead of burying the reasoning inside a model’s private chain of guesses, the system exposes concrete support: candidate entities, selected facts, qualifiers, references when requested, and deterministic resolution states. That gives the operator something specific to validate.

This matters even more in an MCP setting because the agent is not merely answering trivia. It is acting as a bridge between a model and an external knowledge source. If that bridge is opaque, the whole workflow becomes brittle. If it is transparent, you can set better guardrails and recover more gracefully from ambiguity.

The project’s read-only posture helps here too. It does not edit Wikidata, Google, or user data. That sounds modest, but it is a smart boundary. Read-only tools are easier to trust in operational environments because they separate retrieval from curation. A retrieval mistake is still a problem, of course, but it is not silently writing bad data back into the source.

## **What this MCP server is actually built to do**

The server focuses on a clear set of tasks. It lets an MCP client search Wikidata, inspect entity facts, explore related items, resolve local records against Wikidata, and check status. It also comes with a CLI that supports batch work and evidence export. Those capabilities are practical rather than flashy. They line up closely with the routine jobs people actually need in data enrichment and record linking.

For teams evaluating MCP for wikidata, that scope is worth noticing. Some tools promise broad intelligence but are vague about mechanics. This one is relatively direct. It is meant for search, inspection, and

resolution. It does not present itself as official Wikimedia or Google software, and it is not an export of the Google Knowledge Graph. That restraint is healthy. It reduces the risk of users over-interpreting what the tool can prove.

The supported MCP tools are straightforward:

1. `kg_search` for finding candidate entities
2. `kg_entity` for reading selected facts on an entity
3. `kg_related` for exploring connected items
4. `kg_resolve` for linking a local record to a Wikidata QID
5. `kg_status` for checking service status

Even that small tool surface tells a story. The sequence moves from discovery to inspection to resolution, with status checking as an operational layer. In other words, the server is not just exposing a search endpoint under an MCP wrapper. It is presenting a workflow.

## **Bounded search is a feature, not a limitation**

One of the easiest ways to overwhelm an agent is to return too much. Large raw result sets feel generous, but they often create noise rather than clarity. This project takes the opposite approach. By default, it returns three candidates, with up to five. That is a surprisingly important design choice.

Bounded search forces discipline. If the candidate pool is too wide, the model may start rationalizing weak matches. If the pool is too narrow, it may miss the right entity entirely. A default of three, expandable to five, is a practical middle ground for many entity-resolution tasks. It keeps the evidence set small enough to inspect and compare. It also encourages the caller to formulate better search inputs instead of relying on brute force.

I have seen this play out repeatedly in metadata work. When operators receive twenty or thirty possible matches, they often skim descriptions and pick the first familiar label. The probability of error rises fast. With three careful candidates, the review is slower in a good way. People read. They compare dates, occupations, places, and distinguishing relationships. They make decisions with context rather than instinct.

That bounded approach is especially useful in MCP for google knowledge graph and wikidata scenarios, where users may be tempted to treat two large knowledge sources as interchangeable lookup pools. They are not. Candidate selection needs to stay tight if the evidence is going to remain legible.

## **Selected facts beat indiscriminate dumps**

The server supports selected-fact retrieval, including ranks, qualifiers, and references on request. That phrase, selected facts, deserves attention. In practice, many errors in knowledge workflows come from indiscriminate extraction. A model grabs every available statement, gives each one equal weight, and loses the distinctions that matter.

Wikidata statements are not all equivalent. Rank matters. Qualifiers matter. References matter. If you are checking whether a person and a record truly align, a bare property value may not be enough. You may need the qualifier that narrows time, role, or context. You may need to know whether a statement is preferred or normal. You may need references before you are willing to trust the match in a production setting.

That is where inspectable evidence becomes more than a slogan. If an agent can request the specific facts needed for a resolution decision, and can also retrieve qualifiers or references where appropriate, then the human reviewer gets something meaningful to inspect. The evidence is not a blob. It is structured support.

A practical example makes this clearer. Imagine a local archive record for an author with a common name. A simple label match is weak evidence. Add a birth year and occupation, and confidence improves. Add qualifiers that distinguish a pseudonym, time period, or role in a particular work, and the picture sharpens further. The point is not to gather more data for its own sake. The point is to gather the right data, in enough detail that the match can be explained.

## **Deterministic outcomes are quietly powerful**

One of the most valuable aspects of this project is its resolution logic. The documented outcomes include `AUTO_MATCH`, `HOLD`, `AMBIGUOUS`, and `NO_CANDIDATE`. This is the kind of detail that experienced operators immediately appreciate because it affects governance, queue design, and quality control.

Many systems collapse all uncertainty into a soft confidence score. That can work, but confidence scores are slippery. Different models produce them differently, and users tend to treat them as more objective than they really are. Explicit states are often more useful operationally.

A deterministic outcome like `AUTO_MATCH` tells you the system found enough support, under its rules, to link confidently. `HOLD` suggests the evidence is not sufficient for automatic action. `AMBIGUOUS` signals that multiple plausible candidates remain. `NO_CANDIDATE` says the search did not surface a viable entity. Each state leads naturally to a different next step.

Here, the inspectable evidence matters again. An ambiguous result is not a dead end if the candidates and supporting facts are visible. It becomes a review task. A no-candidate outcome is also informative. It may indicate a bad input string, a missing entity in Wikidata, or a need for additional metadata from the local record. Those are very different scenarios, and deterministic states help separate them cleanly.

In real pipelines, this distinction saves time. Automatic matches can move forward. Holds can wait for enrichment. Ambiguous records can route to specialist review. No-candidate cases can trigger a separate workflow for missing or under-described entities. That is a mature pattern, and it is much easier to implement when the MCP layer exposes resolution as named outcomes rather than as an unstructured narrative.

## **The Google cross-check is useful, but the project is careful about what it means**

The optional Google cross-check is one of the most interesting parts of the design because it is both practical and restrained. The project documents exact ID joins using `/m/` for Wikidata property P646 and `/g/` for P2671. Just as important, it explicitly treats agreement between Google and Wikidata as provider concordance, not proof of identity.

That last point shows good judgment.

People often overestimate what cross-source agreement proves. If two providers align on an identifier mapping, that can strengthen confidence that you are looking at the same concept. It does not magically resolve every ambiguity. Providers can inherit the same upstream mistake. They can encode similar

assumptions. They can agree on a broad topic while differing [Knowledge Graph MCP data](#) on granularity or scope.

So the Google layer is best understood as a cross-check, not a verdict. It can be a useful additional signal in MCP for google knowledge graph workflows, especially when exact identifier joins are available. But it should stay in its lane. The project's language on this is refreshingly careful.

There is a broader lesson here for anyone building MCP for google knowledge graph and wikidata use cases. Provider concordance is valuable, but evidence quality still depends on context. A shared identifier can support a resolution decision. It should not replace reasoning about the local record, the candidate entity, and the specific facts that connect them.

## Where this fits within the wider Wikidata MCP landscape

Wikidata itself documents an MCP approach that provides standardized tools for LLMs to explore and query Wikidata programmatically via the Wikidata API and Wikidata Query Service. That broader context matters because it places this project in a family of tools rather than in isolation.

What sets this particular server apart, based on the available facts, is its emphasis on record linking with inspectable evidence and explicit uncertainty. In other words, it is not only about querying Wikidata. It is about making linking decisions legible.

That distinction matters in practice. General-purpose query tooling is excellent when the task is exploratory research or data retrieval. But when the task is to connect a local CRM record, catalog entry, or inventory object to a Wikidata QID, the critical issue becomes evidence management. Search and query are inputs. Resolution is the workflow. Auditability is the requirement.

For teams exploring MCP for wikidata, this creates a useful choice. If the goal is broad discovery or programmatic querying, a general Wikidata MCP setup may be enough. If the goal is disciplined record linkage with bounded candidates and clear resolution states, this project's design is more specialized.

## Clients, deployment, and the practical side of adoption

The server is documented for use in MCP clients such as Claude Code, Cursor, and Codex. That makes it easier to picture where it belongs. This is not a theoretical bridge. It is intended to be dropped into agent-facing environments where a model can call tools as part of a working session.

The authentication story is also sensible. Wikidata requires no account or API key for the documented use. The Google Knowledge Graph Search API is optional. That lowers the barrier for trying the Wikidata side immediately while preserving the option of adding a second provider when the use case justifies it.

There is a pragmatic benefit here for pilots. Teams can start with the read-only Wikidata flow and evaluate how much value inspectable evidence adds before deciding whether the optional Google cross-check is worth the extra setup. That is a safer adoption path than forcing every deployment to configure both systems at once.

The CLI is worth a brief note as well. Batch operations and evidence export are not glamorous features, but they are exactly what turn a promising tool into something operational. Once you have more than a handful of records, you need repeatability. You need exports that can be reviewed, compared, and archived. An evidence-based resolver without export capability would still be useful, but it would be harder to fold into real review cycles.

# What inspectable evidence looks like during actual resolution work

The phrase can sound abstract until you picture how a reviewer uses it. In a typical resolution task, the agent starts with a local record. That record might contain a name, a date, perhaps a place or category, and some free text. The resolver searches for a small set of candidate Wikidata entities. Then it exposes the facts needed to compare them.

At that point, the reviewer is not reading a model's polished summary. They are looking at a compact evidence packet. Candidate A may share the label but differ on occupation. Candidate B may match the occupation but conflict on dates. Candidate C may align on dates and role, but the local record lacks enough detail to distinguish it from a near peer. A deterministic outcome like AMBIGUOUS is not frustrating in that situation. It is honest.

That honesty is where trust starts. Good tooling does not eliminate uncertainty. It names it, constrains it, and shows the operator what remains unresolved.

A useful review pattern often comes down to a few checks:

1. Does the candidate match the local record on stable identifiers or exact joins when available?
2. Do the selected facts align on the details that actually distinguish entities, not just on labels?
3. Are qualifiers or ranks needed to interpret the statement correctly?
4. Is cross-provider agreement present, and if so, is it supporting evidence rather than the sole basis for the link?
5. If uncertainty remains, does the case belong in HOLD or AMBIGUOUS rather than being forced into a match?

That may sound strict, but strictness is usually cheaper than cleanup.

## What this project does not claim, and why that restraint matters

It is easy to oversell knowledge graph tooling. This project does not. It explicitly says it is not official Wikimedia or Google software. It is not an export of the Google Knowledge Graph. It is read-only. It does not edit Wikidata, Google, or user data.

Every one of those boundaries improves clarity.

When teams adopt tools too quickly, they often infer capabilities that were never promised. They assume synchronization where there is only lookup, authority where there is only convenience, or proof where there is only corroboration. Clear limits prevent that drift. They also make governance discussions easier because stakeholders can understand what the system touches and what it does not.

That is especially relevant when the users are not only engineers. Librarians, data stewards, researchers, and operations staff often need to sign off on workflows that involve entity linking. A read-only evidence-oriented service is easier to explain to that audience than a system that appears to rewrite records or blend providers into a single undifferentiated source.

## The bigger lesson for MCP design

Stepping back, the most instructive aspect of this project is not any single tool call. It is the philosophy underneath. Search is bounded. Facts are selected. Resolution is deterministic. Uncertainty is explicit. Cross-provider agreement is helpful but not over-claimed. The system remains read-only.

That combination reflects a mature understanding of how data work actually fails. Most bad links do not happen because a server could not fetch enough text. They happen because too much weak evidence gets treated as decisive, and because the system cannot explain itself after the fact.

Inspectable evidence is the antidote to that pattern. It does not guarantee correctness. No resolver can. But it changes the terms of review. It gives both agents and humans a shared artifact to reason over. It makes ambiguity visible early. It keeps matching decisions close to the facts that support them.

For anyone evaluating MCP for google knowledge graph, or comparing a general MCP for wikidata setup with a more specialized resolver, that is the key point to keep in view. The real test is not whether the tool can find a QID. The test is whether you can inspect the path it took, understand the strength of the match, and stop the workflow when the evidence does not warrant a decision.

That is where this project earns attention. It treats evidence as a first-class output rather than as a byproduct. In entity resolution, that is not a minor implementation detail. It is the difference between automation you can govern and automation you merely tolerate.