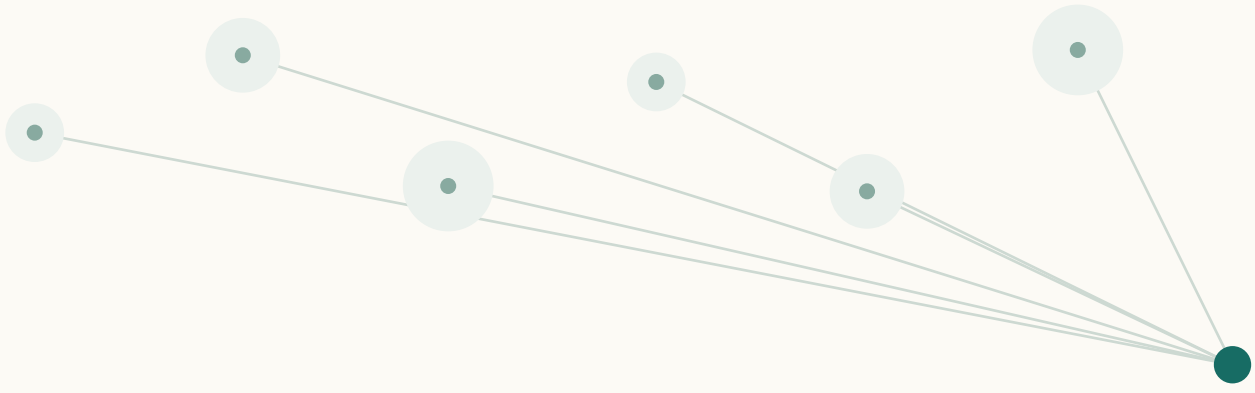
OpenHealth Atlas

Connecting personal health data to AI reasoning

Records, calculations and explanations I can investigate.

KAJEESAN JEEVENDRA

Project creator

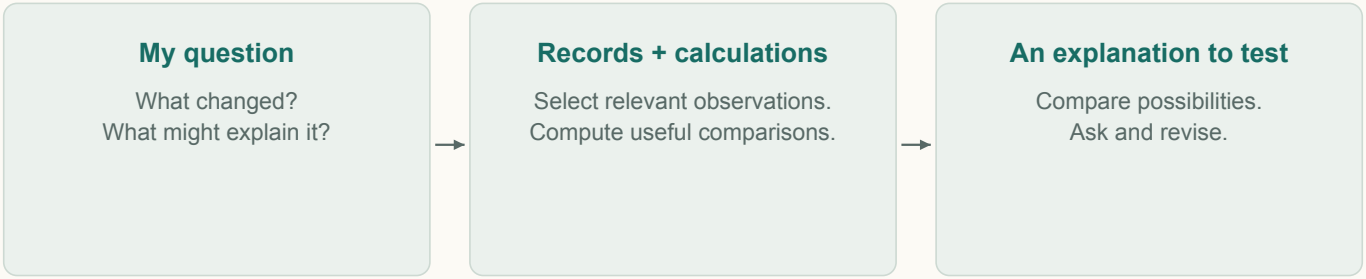


From scattered records to questions I can investigate.

21 SEPTEMBER 2026 / SYSTEM EXPLANATION

A clearer picture of my health

I built OpenHealthAtlas to help an AI agent investigate questions using my recorded history.



QUESTION → EVIDENCE → INTERPRETATION

Figure 1. The question guides the investigation. OpenHealthAtlas supplies records and calculations; the external agent develops an explanation that can be questioned and revised. [1,2]

Abstract

My workouts, sleep, meals, symptoms and test results describe different parts of the same life. When they remain in separate applications, it is difficult to see what changed together or to investigate a question across several months.

I designed OpenHealthAtlas to bring those records together, calculate useful comparisons and retain the evidence behind each result. Hermes, or another connected AI client, uses that evidence to ask questions and develop possible explanations.

The aim

I want to move from a general question such as “Why have I been feeling worse?” to a specific investigation: what changed, when it changed, what supports an explanation and what would weaken it.

I have used and tested OpenHealthAtlas on myself. The worked numerical examples in this report use fictional values to explain the calculations without publishing my personal health records.

How the pieces connect

The data tools, AI agent and conversation channel each have a distinct job.

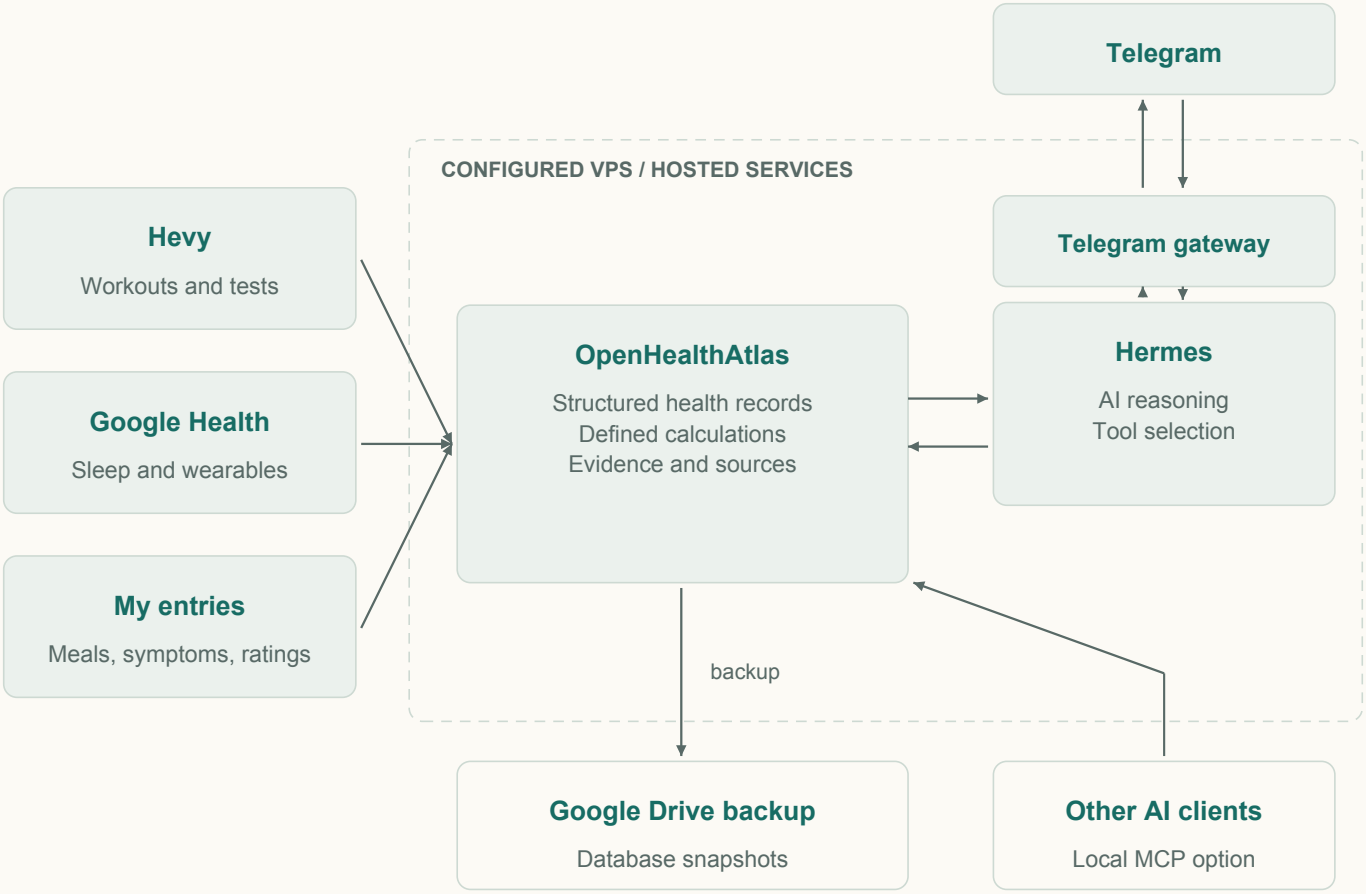


Figure 2. A configured Hermes deployment. Collectors bring external records into the database; the agent requests evidence through tools. The VPS hosts services. A separate local MCP connection also supports other AI clients. [1-3]

Records and calculations

Hevy supplies training records. Google Health supplies wearable measurements such as sleep and heart-rate data. I can also record meals, symptoms, day ratings and quarterly tests.

OpenHealthAtlas keeps structured records and performs the calculations. A VPS is the server that can keep the configured services available between conversations.

Conversation and storage

Telegram gives me a way to speak with Hermes and enter supported check-ins. Hermes chooses which tools to call and interprets their results; it is separate from the calculation engine.

The original setup includes database backups copied to Google Drive. Backup coverage and successful restoration are operational checks; they are separate from the agent's reasoning.

Six pillars, one timeline

A health question often crosses the boundaries between applications and categories.

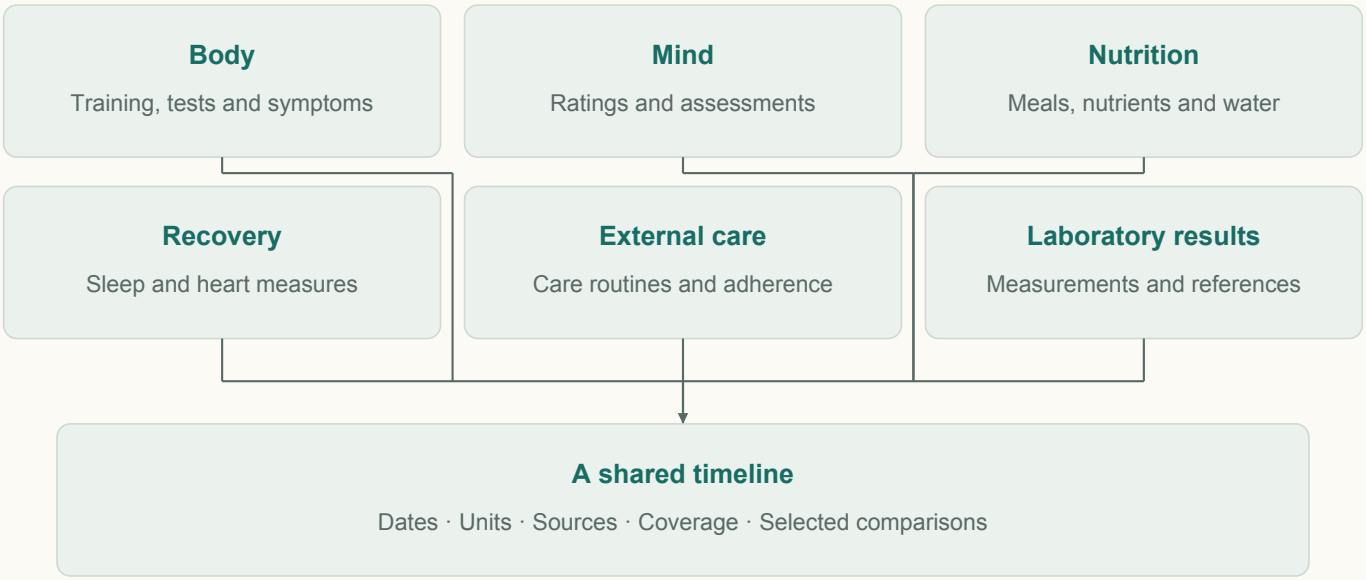


Figure 3. The pillars describe available kinds of information. Dates, units and sources make selected records comparable. The connecting lines show data flow, not an assumption that every pair is correlated. [1,3,5]

Why I connect the pillars

A change in training can coincide with changes in sleep, energy or pain. Food records can add another explanation. The shared timeline lets the agent inspect these possibilities together instead of treating each application as a separate story.

The current care area mainly supports care-product routines and adherence. The laboratory area stores results with units and reference information.

Meaning travels with a value

A sleep value needs a date and a source. A lifting result needs an exercise, load, repetitions and protocol. A symptom needs timing and the description I provide.

I also preserve the difference between “not recorded” and “recorded as absent.” No meal entry does not establish that I ate nothing. Source changes and missing observations affect the comparisons.

Give the agent the right evidence

The model brings learned knowledge. OpenHealthAtlas brings relevant facts about my recorded history.

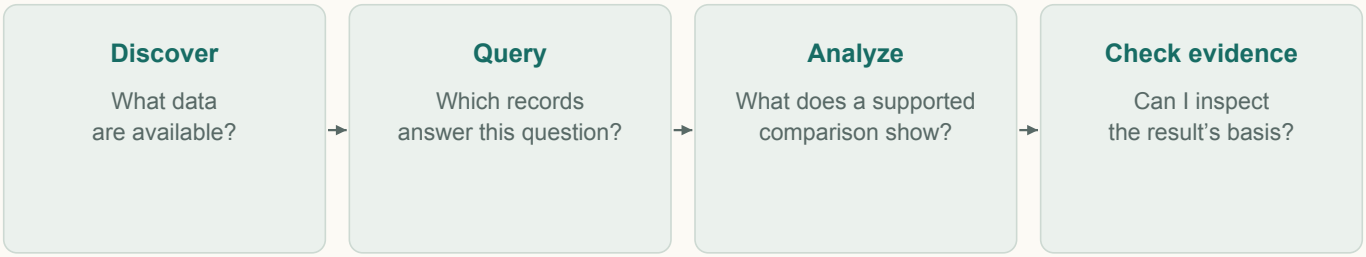


Figure 4. Four core evidence operations support a question-led investigation. The agent can request more detail as its explanation changes. MCP is the standard connection used by compatible AI clients. [2]

A calculator and control room

The deterministic layer acts like a calculator: it applies specified rules to a particular set of records and returns reproducible results. The agent decides which available calculations and observations are useful for the question.

Instead of sending every record into every conversation, it can select a period, inspect relevant measurements and retrieve the evidence behind a finding.

Room for interpretation

The agent can combine those results with its learned knowledge and my replies. A missing predefined formula does not prevent it from proposing a tentative explanation or asking a useful follow-up question.

Selected context can make the question more specific within the model's context limit. It does not guarantee that every important factor was selected or that the explanation is correct. I keep the calculated result and the model's interpretation distinguishable.

Investigating what changed

“Why might my lower back have started hurting?” is the beginning of an investigation.

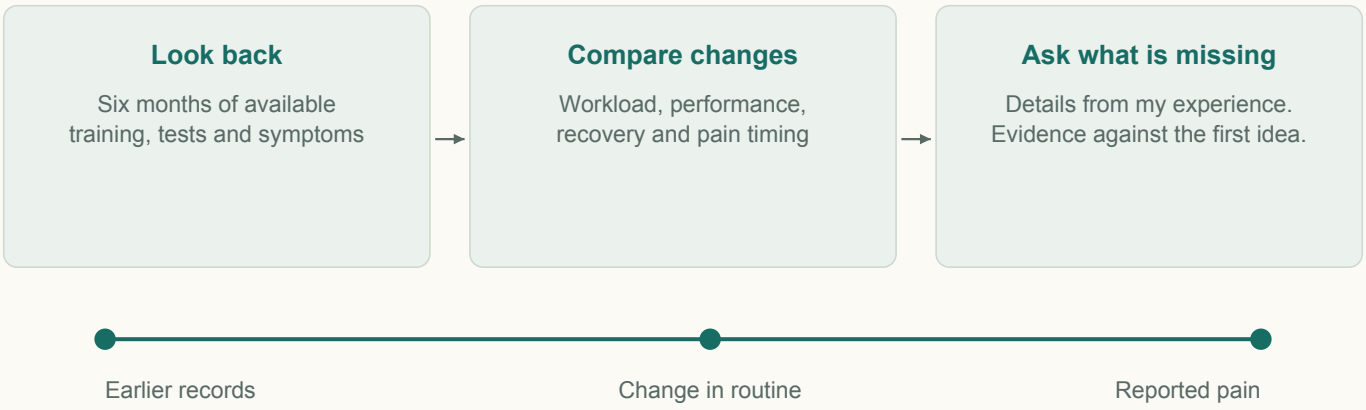


Figure 5. An illustrative investigation across six months. Existing records, date-selected queries and the agent’s follow-up questions support the process. This particular pain explanation is not a demonstrated causal finding. [2,4,5]

Start with the history

I want Hermes to examine symptom timing alongside training, tests and recovery. It can look for changes in exercises, frequency, load or recorded discomfort, then ask about details the records do not contain.

Suppose I logged 30 chest working sets and 3 back working sets in one period. That is ten times as many sets. It describes training exposure; the strength tests provide a separate comparison of performance.

Develop competing explanations

The agent can consider whether an imbalance, a workload change or another factor is relevant. It should also look for observations that do not fit its first explanation.

It can ask when the discomfort occurs and compare my answer with the recorded timeline. Even without a validated ideal ratio for a particular muscle pair, it can use the available evidence to form a hypothesis that I can investigate or discuss with a clinician.

My baseline starts with repeated tests

A personal ratio is useful when I understand exactly which measurements it compares.

ILLUSTRATIVE QUARTERLY TESTS

Chest exercise

40 kg × 8 repetitions
40 × (1 + 8 ÷ 30) = 50.7 kg

Row exercise

30 kg × 8 repetitions
30 × (1 + 8 ÷ 30) = 38.0 kg

PERSONAL EXERCISE-PERFORMANCE RATIO

50.7 ÷ 38.0 = 1.33

Repeat comparable protocols to see how this relationship changes.

Figure 6. Fictional eight-repetition tests using the implemented estimated-one-repetition-maximum formula. The 1.33 result is an exercise-performance ratio, not a universal ideal or a direct measurement of all chest versus back strength. [4]

What the numbers mean

I repeat defined tests quarterly so that I can compare my own measurements over time. Direct maximum tests and repetition-based estimates need clear protocol labels; equipment and technique affect the comparison.

For the illustrated estimates, the formula is load × (1 + repetitions ÷ 30). The same repetition count makes the two estimated results straightforward to compare.

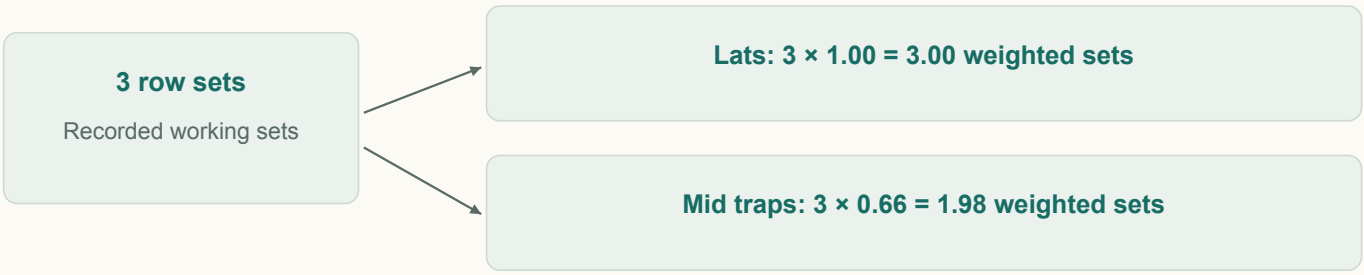
Compare with my history

The system stores personal tracking pairs even when no suitable published target band exists. If a ratio changes from 1.33 to 1.20 in a later comparable quarter, the relationship between those recorded performances has changed.

That change gives the agent something specific to investigate. My starting baseline describes me; it does not automatically define an ideal balance. Research references are useful where their measurement protocol fits the test.

From an exercise to a useful estimate

The system combines measured performance with clearly identified estimates of muscle involvement.



A SEPARATE SUBREGION ESTIMATE / ALL SIGNALS AVAILABLE



$$0.60 \times 73 \;+\; 0.25 \times 60 \;+\; 0.15 \times 80 \;=\; 70.8$$

Rounded score: 71. An authored estimate, not a direct strength measurement.

Figure 7. Three row sets become weighted exposure for mapped regions. The lower example shows the authored combination used when all three signals exist. These weights describe an inference method, not measured activation percentages. [4]

Mapping the work

An exercise can involve several muscles and subregions. The mapping distinguishes research-informed entries from biomechanical estimates. Three sets at a weight of 0.66 contribute 1.98 weighted sets to that region.

My original map includes quarterly-test exercises beyond the smaller public starter map. The active mapping determines which direct tests contribute.

Interpreting the score

The subregion estimate combines progress relative to an exercise's own history, mapped direct tests and exposure. Missing signals change the effective weighting; sufficient repeated observations are required before a ranking appears.

A stable performance history can score below an improving one. I therefore treat this score as an indicator to inspect alongside the actual tests and symptoms, rather than as a direct measure of one region's strength relative to another.

What do my better days have in common?

I provide the day rating. The system investigates the observations associated with it.

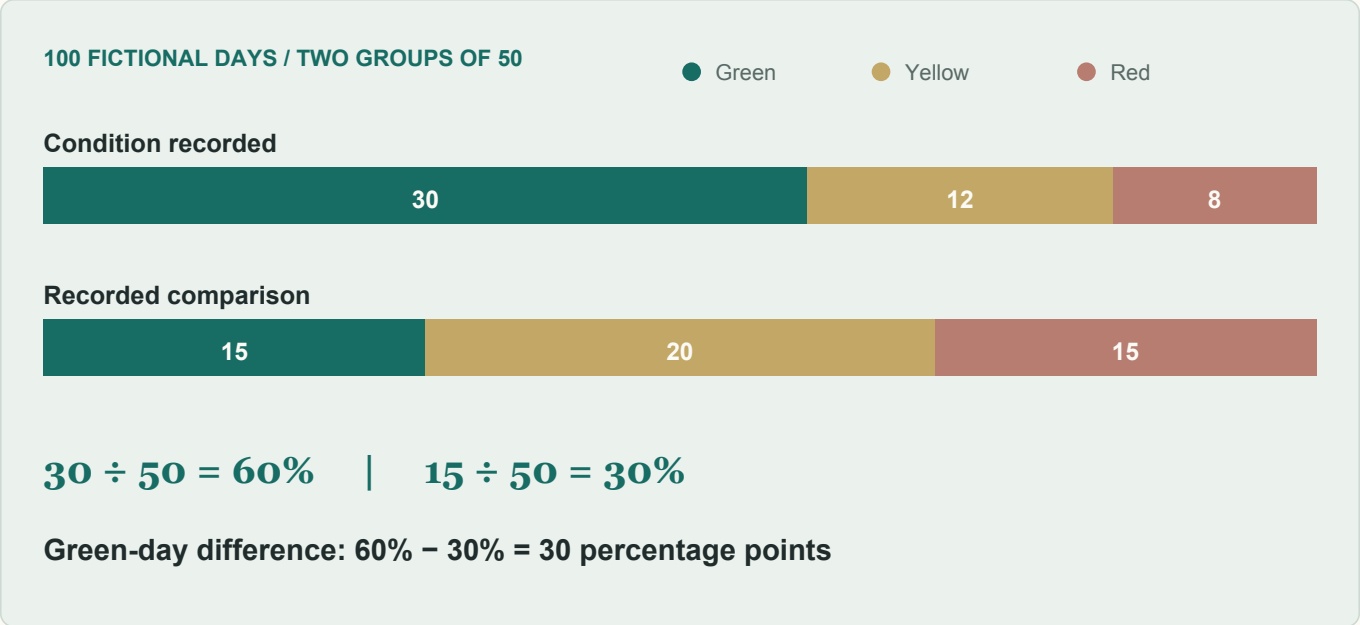


Figure 8. Fictional groups of 50 days: 30 and 15 green ratings give rates of 60% and 30%. Their difference is 30 percentage points. Yellow and red remain recorded outcomes, not missing observations. [5]

An outcome I recognize

Green means I rated the day good, yellow okay and red bad. These are my subjective ratings, not a health grade automatically calculated from all my devices.

The agent can investigate what accompanied those ratings: sleep, training, food, symptoms or other recorded factors. The system can compare selected periods and account for a defined delay between an observation and an outcome.

A difference to investigate

In the example, green days were 30 percentage points more common with the recorded condition. That is the calculation. Why the groups differed is a question for further investigation.

The engine also checks usable observations, timing and uncertainty. The agent can consider alternative explanations, ask for missing context and see whether the pattern appears again in a later period.

Could meal timing be relevant?

A useful question connects the last logged meal with the sleep that follows it.

SIX FICTIONAL NIGHTS / LAST LOGGED MEAL AND SUBSEQUENT SLEEP

Earlier meals

18:30 · 19:00 · 19:15

7.7 · 7.9 · 7.8 hours

7.8 h

average recorded sleep

Later meals

22:00 · 22:15 · 22:30

6.4 · 6.8 · 6.9 hours

6.7 h

average recorded sleep

Observed difference: 7.8 − 6.7 = 1.1 hours

A question to investigate; meal timing is not isolated from other changes here.

Figure 9. A constructed six-night example: average recorded sleep is 7.8 hours in one group and 6.7 in the other. The difference is 1.1 hours. These small illustrative groups are not an accepted statistical finding. [2,5]

Make the comparison explicit

The tools can retrieve meal times and sleep observations. A one-day maximum can identify the last logged meal. I then need to align that evening with the appropriate sleep record and consider incomplete logs.

The example makes the question visible: did later eating accompany less sleep, or did bedtime, meal size, training or another part of the routine change too?

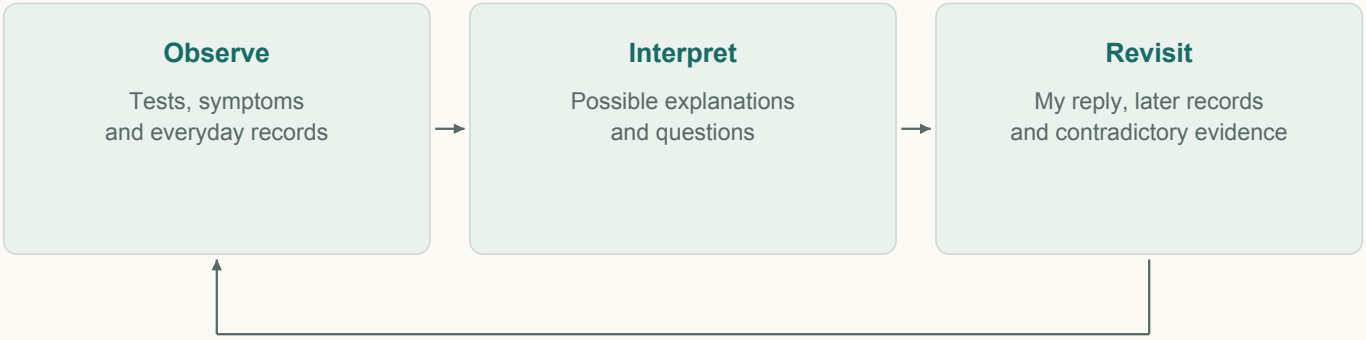
Use interpretation honestly

An agent can inspect these observations and ask whether a change in meal timing is worth exploring. Subsequent records and my feedback can support or weaken that idea.

A ready-made last-meal-versus-sleep predictor is not yet part of every analysis route. That is a specific calculation I can add or connect later; it does not prevent the agent from discussing the available evidence as a tentative pattern.

An explanation should be revisable

The value comes from returning to a question as better information becomes available.



New information can change the explanation.

Figure 10. New observations and replies can change the investigation. Conversation history, structured health records and the deterministic hypothesis ledger are separate forms of continuity. [1,5,6]

Keep the evidence available

The database stores health observations and analytical history. Registered hypotheses can accumulate supporting or weakening evidence over later periods. The existing pain workflow also organizes self-tests and recorded responses to trials.

An external Hermes session can retain conversation history, letting a later reply build on an earlier question.

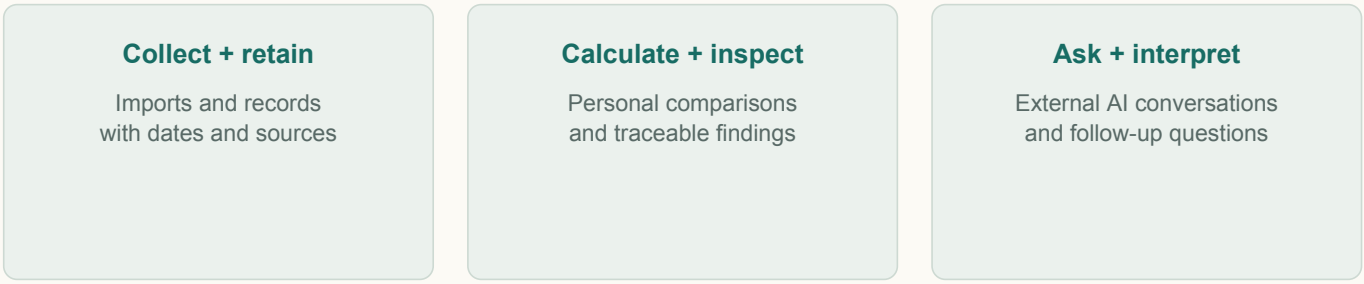
Keep the forms of learning clear

A useful conversational explanation is not automatically saved as a structured hypothesis. Saving an agent's conclusion with its evidence is optional and was switched off in the controlled test runs cited here.

More records and better feedback can improve the context for the next investigation. They do not automatically retrain the underlying model. I want the system to make revision possible, including the decision that an earlier explanation no longer fits.

Personal use and software testing

I have used and tested OpenHealthAtlas on myself, alongside controlled checks of the software and agent connections.



Existing components form an investigation workflow; the connection defines tool access.

Figure 11. Personal use and controlled software tests provide complementary evidence about the working system. The examples in this report illustrate its methods. [1-6]

What is already present

The implementation includes imports, database storage, personal tests, nutrition and recovery calculations, selected cross-pillar associations, evidence retrieval and data-driven screens.

My own use is part of the project's practical testing. Separately, controlled tests with fictional records have demonstrated model/tool conversations, continuing Telegram conversations, and the import-to-dashboard workflow.

What depends on the connection

I can connect an AI client to a selected personal database through MCP. Controlled Hermes demonstrations use an isolated test connection. My original configured Hermes workflow has broader tool access, so I distinguish what is available in each setup.

Some detailed calculations use a different tool connection from the general queries. I treat checks of the active deployment and comparisons of answer quality as distinct tasks. The aim is to establish when selected evidence leads to useful, accurate explanations.

Sources and scope

This report draws on my personal use and the reviewed implementation. Worked examples explain the calculations without publishing personal records.

[1] System architecture and responsibilities

OpenHealthAtlas public source: README, Architecture and Product authority. Describes records, calculations, external AI clients and the retained integration boundaries.

Open the implementation source

[2] Agent tools and interpretation

Local MCP guide; scripts/openhealthatlas_mcp.py; hermes_surface.py; shipped Hermes skill and delivery rules. Defines discovery, queries, analysis, evidence replay and model interpretation.

Open the implementation source

[3] Inputs and integrations

Hevy and Google Health collectors, import commands, Telegram actions and deployment templates. The original configured setup also contains the Google Drive database-backup script.

Open the implementation source

[4] Personal tests and muscle calculations

Shared calculations.py and catalogs.py; quarterly adapters; health.py muscle detail; authored-submuscle mapping documentation. Source for the equations and distinctions in Figures 6-7.

Open the implementation source

[5] Timing, associations and hypothesis history

Registry and adapters; associations.py; stats.py; ledger.py. Source for outcome definitions, bounded comparisons, evidence identity and subsequent evaluation.

Open the implementation source

[6] Storage and verification

synthesis.py; VERIFICATION.md; dated implementation checkpoints. Documents controlled software tests and genuine-agent demonstrations. My personal use is described separately in this report.

Open the implementation source

Implementation version

Public source snapshot ab3ce1c4b4194ce3ebe9e50189402f785cf49f0e, reviewed on 20 September 2026. The original authored map and backup script were examined separately as configuration and operational source. No personal health observations are reproduced in this report.

What I am looking into next

My priority is to improve the reliability and usefulness of the existing system while preserving its division of work.

1 / Reliable inputs

Validation, dates and source baselines

2 / Clear indicators

Measurements, estimates and references

3 / Verified operation

Configured integrations and restoration

4 / Useful explanations

Same questions, comparable evaluations

The next stage combines focused engineering corrections with evaluation of real user questions. These are planned priorities, not claims of completed validation.

Make the evidence more dependable

I am prioritizing stronger validation for general logging, correct date boundaries for current measurements and consistent source baselines for recovery calculations.

I also want every strength label to match the measurement method behind it. Personal trends should remain useful even where a research-derived target cannot be transferred directly.

Evaluate the whole investigation

I plan to verify the configured integrations and backup restoration, then compare how the same model answers selected questions with ordinary history, structured records and traceable calculations.

I will look at factual accuracy, relevant evidence, useful follow-up questions and whether the agent revises an explanation when new information conflicts with it. OpenHealthAtlas will remain the calculator and evidence workspace; the external agent will remain the interpreter.