

RH Mission: inspectable evidence of cumulative agent research engineering

Justin Sublette · Historical exhibits prepared September 16; public-source update September 23, 2026

The Mathematical Research system pursues a checkable proof or disproof of the Riemann Hypothesis with autonomous AI. Its engineering challenge is to let research accumulate across replaceable agents: preserve useful mathematics, recognize connections, revise strategy and recover from interruption without quietly strengthening what the evidence supports. [RH-10]

My contribution is problem definition, system and authority design, integration, evaluation, correction and acceptance. Codex contributed extensively to implementation, mathematical investigation and testing. The executive's job includes choosing scientific strategy; this is not a system in which I personally adjudicate every mathematical step.

These exhibits make three parts of that work inspectable: what a successor agent receives, what happens when a result is saved but its acknowledgment is lost, and how a research conclusion was narrowed without discarding its useful mathematics. Source excerpts are from revision ef8546aaa7a79c006f75a68da027389912b1f2d1, committed September 16, 2026. Historical results retain their original dates and source bindings. This preparation inspected code and retained records; it did not execute tests or run a new demonstration.

Public implementation and reproducible research examples - September 23

The actual system is now available in [Mathematical Research](#) under MIT. Start with the [engineering decisions](#) and [standalone lifecycle qualification](#). The latter records real Linux initialization, research-tool interaction and orderly stop/restart with retained fresh public state.

Published example	What a reviewer can inspect and reproduce
Finite-free localization	A missing-assumption correction, exact checker and real-store walkthrough. The live execution record shows a distinct successor retrieving exact retained Evidence/Candidate revisions, preserving the corrected assumptions and tightening the bound from 10 to less than 9.
C90 Loewner obstruction	Rigorous interval enclosures refute the stated Pick/complete-Bernstein route while leaving ordinary Bernstein status open; an authored store walkthrough preserves that scope.
Jacobi construction	Exact symbolic and root checks preserve a wider positive-rooted construction after a restricted attempt fails; its application retains open analytic premises.
Bounded Jensen program	A finite retained-mass/phase-mixing argument, interval/algebra checker and authored store walkthrough preserve the conditional range calculation and larger unresolved dependencies.

Only the first example includes a fresh live executive and successor. It is reference-assisted; its selected actual tool requests and results are public, while raw private traces are not. The other examples distinguish historical research from freshly checked mathematics and authored store judgments. No general research-success rate, mathematical novelty, actual-source Jensen certification or RH proof follows. These public-release checks are separate from the September 16 private retained-history preparation and restoration work described below.

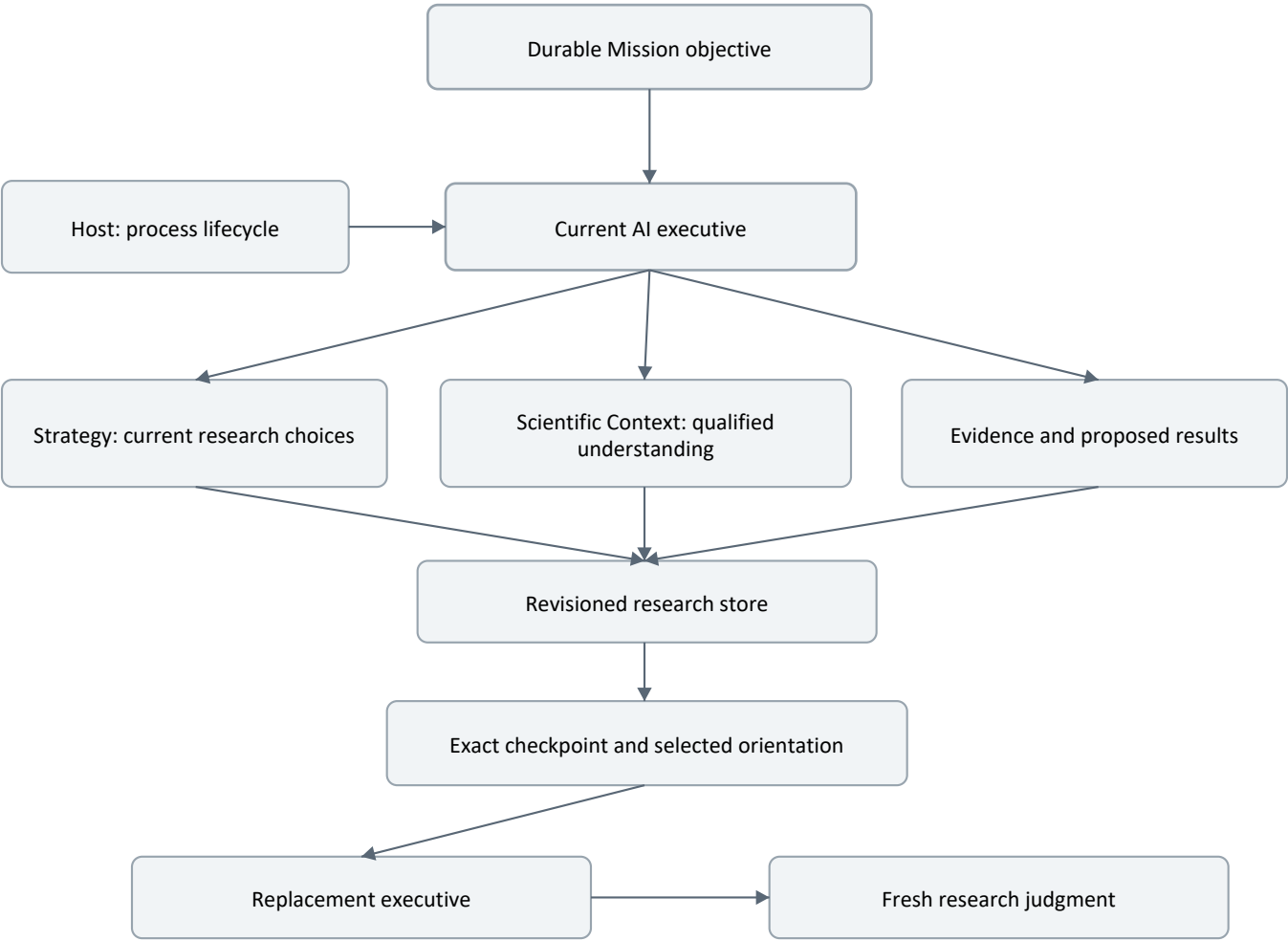
The links in this section pin the inspected public documentation snapshot. Its validation pages identify the exact runtime revisions used for their observed checks; later repository changes are not included in those results.

1. Scientific understanding survives replacement of the executive

Evidence: accepted architecture, implemented interfaces and an authored deterministic integration test.

A persistent objective is not enough to preserve research. A replacement agent also needs to understand what a result permits, which qualifications constrain it and what question remains open. A transcript preserves the conversation's order; the system instead gives that understanding explicit owners.

The durable **Mission** carries the objective. The current AI **executive** chooses research strategy, delegates investigation and criticism, and synthesizes results. **Strategy** allocates attention. A separate scientific **Context** carries cumulative understanding, including selected treatments of useful results and their restrictions. Evidence and proposed results keep their own exact identities. The host manages process lifecycle and containment; it does not choose the mathematics. [RH-01, RH-02]



Explanatory diagram of responsibilities and intended continuity, not a recorded execution trace. Proposed results do not become accepted mathematics through storage or succession.

The context decision deliberately separates recognition from mathematical reliance. A result can help an agent notice an opportunity without every source behind it becoming a premise in a proof. When a proposed result actually relies on a source, the relevant exact revision and treatment must be identified. This supports broad discovery while preserving a narrower, inspectable account of what an argument depends on. [RH-02]

The selected test constructs a result with an authored scientific continuation, adds another unrelated treatment, replays the original request, checkpoints, and replaces the executive. It then inspects the new executive's orientation. The fixture's mathematical wording is deliberately limited: a stronger lower-gap ingredient does not supply an upper mismatch estimate. The successor must inherit that limitation with the selected treatment. [RH-03]

Exact excerpt from the test's successor assertions:

```
science = self.interface.executive_orientation()["scientific_context"]
exposed = {(item["context_reference"]["identity"], item["treatment_id"]): item for item in
science["treatments"]}
self.assertIn(("consumer", "a"), exposed)
self.assertNotIn(("consumer", "b"), exposed)
self.assertIn("Lower gaps", exposed[("consumer", "a")]["content"]["qualifications"][0])
```

The distinction is concrete: both treatments remain stored, but the explicitly selected one is exposed with its qualification. Automatic opening is not a recursive expansion of all historical material. Further information remains retrievable. The test also checks that replay returns the original publication's references rather than silently changing them to later heads.

This is inspectable implementation of a continuity mechanism. It does not, by itself, demonstrate that an AI successor will discover a valuable new connection. Restoring and evaluating that scientific behavior remained separate work at the account's cutoff.

2. A saved result must not lose its already-authored continuation

Evidence: actual publication implementation and authored failure-injection tests using the real store fixture.

Suppose an agent has completed a research result and written the explanation of what that result leaves to investigate. Saving those separately creates an awkward failure: the evidence survives, but the scientific continuation needed to make it useful to a successor does not.

The system provides a narrow transaction for one explicitly paired Evidence result, its authored scientific treatment and the required exposure. This is not a transaction around an entire research campaign or synthesis. Independent sibling work and immediate preservation of a potential complete result retain separate paths. [RH-02, RH-04]

The semantic interface passes the pair to the existing store writer:

```
outcome = self._store.commit_scientific_context_publication(
    executive_epoch_id=authority.executive_epoch_id, mission_id=self._mission_id,
    updates=tuple(updates), exposure_required=continuation["exposure_required"],
    evidence=evidence, request_sha256=digest,
    lease=self._writer_lease(), command_id=command_id, actor=_ACTOR,
    expected_canonical_authority_digest=authority.canonical_authority_digest,
)
```

Publication identity binds the Mission, executive epoch and exact request. The interface checks for an existing publication before preparing a new one. That addresses a second failure mode: the database commits successfully, but the caller never receives the acknowledgment. Repeating the scientific work would be wasteful and could produce a different interpretation. Replaying the exact request instead retrieves the committed result. [RH-04]

The selected test injects a failure after commit. It expects the initial response to report unavailability while the evidence and Context already exist. A replay must then complete without creating replacement revisions:

```
replay = self._execute(INTERPRET_MATERIAL, request, executive_epoch_id=self.epoch)
self.assertEqual(replay["status"], "completed", replay)
self.assertTrue(replay["result"]["publication"]["replayed"])
self.assertEqual(self._reference("science"), exact)
self.assertEqual(read_evidence_meaning(self.store, evidence_id="ingredient").revision,
evidence.revision)
```

A companion test injects failure before the publication transaction can complete. It checks that the previous Context is unchanged, the paired Evidence is absent, and the captured source material remains available. Retrying succeeds without duplicating that retained raw custody. These tests distinguish a failed commit from a lost acknowledgment of a successful commit. [RH-05, RH-06]

The value here is preservation of authored research across unreliable process boundaries. The result, its interpretation, the caller's knowledge of the write, and the successor's ability to recover it are different facts. The implementation maintains their relationship explicitly. The excerpts show actual code and test assertions; no new passing execution is claimed by this exhibit.

3. Rejecting an overbroad conclusion while retaining useful mathematics

Evidence: an accepted July 30, 2026 research closeout, its corrected derivation and the corresponding symbolic checker.

Campaign 020 asked several worker lanes to investigate a particular mathematical construction. One lane proposed a broad structural-obstruction verdict. The final synthesis rejected that classification because the exact subclass being excluded had been formulated during the worker's investigation rather than specified before testing. A conclusion about that narrower subclass could not become a permanent exclusion of the wider research approach. [RH-07]

The review preserved a useful local result instead. For an isometric embedding J , write $\Pi = JJ^*$. Let T_t be a semigroup of bounded self-adjoint operators and define its compression $Q_t = J^* T_t J$. The recorded derivation gives:

$$\begin{aligned} Q_{2t} - Q_t^2 &= J^* T_t (I - \Pi) T_t J \\ &= ((I - \Pi) T_t J)^* ((I - \Pi) T_t J) \geq 0. \end{aligned}$$

This display is a shortened presentation of the retained derivation, with its assumptions stated above. It expresses a nonnegative leakage defect. If the compressed family is itself a semigroup, that defect vanishes: the embedded subspace must be invariant under the ambient evolution. The closeout supplies the fuller argument and its separate, explicitly conditional application. [RH-07]

The engineering significance is the final disposition. The local compression result survived as C-162. The proposed wider exclusion did not. The broader research branch became dormant with a stated revival condition, not mathematically false. No central RH proof dependency was advanced by association.

The accompanying checker has an equally explicit role. It checks finite symbolic calibrations, including the compression-leakage identity and a small conditional example. Its own description assigns the operator proof and domain statements to the written closeout. It does not treat finite algebra as a proof of the full RH research target. [RH-08]

This is a retained example of the research process correcting the strength of an AI-generated conclusion while conserving valid partial work. It supports a claim about the system's research discipline and development decisions. It is not evidence of unaided mathematical authorship, external peer review or an RH solution.

Supporting result: reproducible finite computation

An accepted July 31, 2026 execution record reports three fresh remote replicas at source revision 75e2b74dfb1dc1e8c28f6ad0e7649c70b4f461b6. Their normalized exact-result streams agreed within and across replicas, and the canonical proof state was unchanged. [RH-09]

Replica	Total elapsed seconds	Peak step memory
1	6,265.104	697.6 MiB
2	6,285.248	694.7 MiB
3	6,268.990	695.3 MiB

These figures establish the recorded reproducibility of that frozen software and finite calculation suite. They are supporting operational evidence, not a speedup comparison or proof of an infinite mathematical statement.

RH remained open at the September 16 account cutoff. Full qualification of the newer build and restoration of autonomous scientific behavior remained unfinished. [RH-10] The substantial contribution represented here is the implemented research infrastructure, the concrete mechanisms for preserving meaning and continuity, and the recorded judgment that kept useful mathematics without admitting an unsupported stronger conclusion.

Source notes

The RH-01 through RH-10 source labels refer to the September 16 snapshot named above; historical receipts retain their own dates. The separately dated public-source section links its own implementation and execution evidence. The associated private provenance ledger preserves exact original paths, line ranges and excerpt hashes.

- **RH-01 - Mathematical Research architecture:** adaptive executive, subordinate host control and persistent semantic owners; accepted design and implementation description.
- **RH-02 - September 14 cumulative scientific Context decision:** independent understanding, exact reference roles, selected qualified exposure and narrowly paired publication; accepted design.
- **RH-03 - test_scientific_publication.py:** test_paired_deeper_question_survives_unrelated_patch_and_exact_replay; authored deterministic test, inspected rather than run for this exhibit.
- **RH-04 - mission_interface.py:** publication identity, original-request replay and paired publication; implementation source.
- **RH-05 / RH-06 - test_scientific_publication.py:** post-commit acknowledgment loss and pre-commit transaction failure; authored deterministic tests, inspected rather than run here.
- **RH-07 - RH Native Campaign 020 Semigroup-Discriminator Closeout, July 30:** retained accepted research record, including the C-162 derivation and narrowed conclusion.
- **RH-08 - check_native_campaign_020.py:** finite symbolic calibration source; no fresh execution in this preparation.
- **RH-09 - RH Remote Compute Qualification Receipt, July 31:** retained accepted execution account tied to its original frozen source and three-replica collection.
- **RH-10 - Project Charter and Mathematical Research Status:** the autonomous research objective, its attribution boundary and the September 16 implementation/qualification frontier.