Constellations *of* Borrowed Light

Borrowed Light Collective

v1.0 • 2026-01-08

"The inheritance from the master becomes not only his additions to the world's record but for his disciples the entire scaffolding by which they were erected."

— Vannevar Bush

As We May Think, 1945

The Inheritance

Human knowledge is never contained in one person. It grows from the relationships we create between each other and the world, and still it is never complete.

—Paul Kalanithi, When Breath Becomes Air

The knowledge to save your life exists. It cannot reach you.

I know this because when I was six years old, I saw the same three doctors twelve times over four months. Each visit, my chart grew thicker—headaches, fevers, nausea, all documented. Each visit, the same routine: temperature, abdomen, ears. "It's viral," they said. "Give it another week."

My mother had noticed I was unsteady on my feet. She mentioned it. He nodded and wrote something down. She had noticed my vision seemed off. He didn't ask about that. No one had.

The knowledge to diagnose me existed. It was one of the most common pediatric brain tumors—well-documented, highly treatable if caught early. The diagnostic triad—headaches, vomiting, and ataxia—has been documented for decades. I presented with all three. Textbook in retrospect—invisible in the moment.

But there was no structure. No checklist prompted any of those doctors to ask about balance, vision, and headaches *together*. No system flagged that the same child had returned twelve times with unresolved symptoms. The pattern was building in the chart. Nothing prompted anyone to see it.

One night, my fever hit 105. My mother drove me to the children's hospital.

In the waiting room, a teenage volunteer passed me snacks and candy while we waited. I asked if I could have more. She brought me another handful. I remember the fluorescent lights, the plastic chairs, the sound of a television playing somewhere I couldn't see. I remember thinking this was the kind of day where people were nice to you.

Hours passed. They took me for a CT scan—not because someone finally connected the dots, but because my symptoms had become impossible to ignore.

I remember the machine. The hum. The coldness of the table. I remember looking up through the glass at the booth above, watching the technicians stare at the screen. Then their faces changed. One of them left. He came back with another doctor. Then another. They were pointing at something. They weren't looking at me anymore.

The headaches. The vision. The swelling. All of it, now undeniable.

They wheeled me back to my mother. No one said anything yet. But I had seen their faces. I was six years old, and I understood that something was very wrong.

The surgeon appeared at my bedside. I survived.

Three doctors. Twelve visits. Four months. The pattern was there—accumulating with every visit, documented in a chart that grew thicker each time. The knowledge to diagnose me existed. What didn't exist was the structure that would have made the pattern visible.

• • •

The answer is not that my doctors were incompetent. The answer is that I was typical.

Only 38% of pediatric brain tumors are diagnosed within the first month of symptoms.[1] One in three children is misdiagnosed on their first hospital visit. In some studies, the average time from first symptom to diagnosis exceeds eight months. Some children wait years.

The symptoms are there: headaches, vomiting, balance problems, visual changes. But these symptoms are also signs of a hundred other things. A pediatrician sees a fever and thinks infection. An ophthalmologist sees vision problems and thinks refractive error. A neurologist might connect the dots, but the child never reaches the neurologist, because no one thinks to refer.

Researchers call this "diagnostic imprinting." Each specialist sees through their own lens. The whole picture exists, but no one is positioned to see it. In one study, only 57% of children presenting to emergency rooms with brain tumor symptoms were even examined for papilledema—swelling of the optic nerve, one of the most reliable signs of increased intracranial pressure.[2] The examination takes thirty seconds. It wasn't done.

[1] Flores et al., "Delay in the diagnosis of pediatric brain tumors," *American Journal of Diseases of Children*, 1986. doi:10.1001/archpedi.1986.02140210082031. PMID:3012997.

[2] Shay et al., "Diagnostic delay of pediatric brain tumors in Israel," *Child's Nervous System*, 2012. doi:10.1007/s00381-011-1564-0. PMID:21870085.

This is not a knowledge problem—it is a transmission problem.[3] Knowledge wants to arrive. The path does not exist. The symptoms are described in textbooks, in papers, in clinical guidelines. But knowledge does not flow. It sits in silos. The pediatrician doesn't see what the ophthalmologist sees. The ER doctor doesn't see what the neurologist would recognize.

My case was not unusual—it was typical. That is the harder truth. The system that failed me was not broken; it was working exactly as designed. It was designed without the structure that would make knowledge arrive.

· · ·

Most scientific knowledge never becomes a paper. The failed experiment that would have saved someone else six months. The clinical reasoning at 2 AM that no one recorded. The protocol that actually works, as opposed to the one that was published. The tacit pattern a senior researcher recognizes but has never articulated.

This knowledge exists. It was never structured, never connected. The knowledge that would save the next child is out there—scattered, invisible, waiting.

Medicine is not unique in this. It takes, on average, seventeen years for a scientific discovery to reach routine clinical practice.[4] This lag is a transmission failure. The gap spans every domain where humanity needs science to arrive: fusion, climate, disease, space, the problems that will determine whether our children inherit a livable world.

We begin with science because the stakes are measured in lives and the claims aspire to verifiability.

· · ·

[3] "Transmission," as this essay uses the term, means the provenance, navigation, and correction of claims as they move between lab notebooks, publications, practice, and subsequent experiments. It does not by itself solve training, staffing, regulation, or reimbursement—but it is the substrate those solutions require.

[4] Balas & Boren (2000) estimated seventeen years for discoveries to reach routine practice; subsequent reviews suggest 10–20 years with wide variation by field. doi:10.1055/s-0038-1637943. PMID:27699347.

Look up at the night sky and you are seeing ghosts. Some of those stars collapsed millions of years ago. Their light is still traveling. Sailors crossed oceans by it. Entire civilizations oriented themselves by stars already dead.

We navigate by borrowed light—knowledge we inherit rather than discover. The physician inherits clinical wisdom from cases she never saw. The researcher inherits methods from experiments she never ran. Every generation inherits from the one before.

The question is not whether we navigate by borrowed light—we have no choice. The question is whether the light is structured: navigable, correctable, traceable. Whether it arrives, or is lost.

That inheritance carries obligation—but not the heavy kind. The best transmitters carry warmth: *we're going to figure this out. This isn't the end.* I learned this from someone who didn't get to finish teaching me.

• • •

For drugs, the gap is even longer.

Eighty percent of the world's most important drugs originated from publicly funded science[5], but with an average of thirty-two years from discovery to approved drug—a timeline that includes necessary development and testing, but also years of fragmentation and failed knowledge transfer.[6]

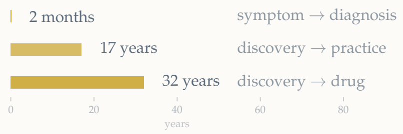
Twenty-eight billion dollars per year—more than half of what the NIH spends annually—wasted on preclinical research that cannot be reproduced.[7]

For a six-year-old with a brain tumor, seventeen years is not policy. It is a lifetime some children will not have. Thirty-two years: long enough for the researcher who made the discovery to retire, or die, never knowing whether it helped anyone. Long enough for a child to be born, grow up, and have children of their own, still waiting for the treatment.

[5] Stevens et al., "The Role of Public-Sector Research in the Discovery of Drugs and Vaccines," *New England Journal of Medicine*, 2011. doi:10.1056/NEJMsa1008268. PMID:21306239.

[6] Contopoulos-Ioannidis et al., "Translation of highly promising basic science research into clinical applications," *The American Journal of Medicine*, 2003. doi:10.1016/S0002-9343(03)00013-5. PMID:12731504.

[7] Freedman et al., "The Economics of Reproducibility in Preclinical Research," *PLOS Biology*, 2015. doi:10.1371/journal.pbio.1002165. PMID:26057340. The NIH's annual budget is approximately $47 billion.



The problem is not a shortage of discoveries. The discoveries are scattered across journals, databases, lab notebooks, clinical records, and the minds of specialists—most of it never written down at all. What's missing is the structure that would connect them: the lines that would turn isolated points of light into navigable constellations.

Knowledge that cannot arrive might as well be absent for the person who needs it.

Fragmentation is the default. No one chose it. No one maintains it. It persists because nothing has replaced it.

The barriers are not primarily technical. They are institutional: incentives that reward citation over impact, formats that fragment rather than connect. AI cannot solve social problems. But infrastructure can change the incentives.

Eleven million physicians. Eight million researchers.[8] Each day, millions of decisions hinge on knowledge that may not have arrived. Each week, millions of hours spent searching for what already exists. The scale is not niche. It is civilizational.

We propose to build this infrastructure. The stars are there, scattered across the archive, luminous but unconnected. What remains is to draw the constellations that make them whole.

The same infrastructure gap that almost killed me is about to matter at civilizational scale.

We are entering the era of AI-driven scientific production. Not science assisted by AI—science *performed* by AI. Autonomous laboratories that run experiments around the clock. Models that generate hypotheses, design protocols, interpret results, and propose the next experiment before a human wakes up. The compressed 21st century that Dario Amodei describes: fifty to a hundred years of biological progress in five to ten.[9]

[8] World Health Organization, Global Health Workforce Statistics, 2024. WHO data. The "fifth of time searching" estimate comes from Tenopir et al., "Journal reading patterns and preferences of pediatricians," *Journal of the Medical Library Association*, 2007. PMC1773049.

[9] Dario Amodei, "Machines of Loving Grace," October 2024. darioamodei.com.

This is not science fiction. AlphaFold predicted the structure of over 200 million proteins—more than the entire history of experimental biology combined.[10] DeepMind won the 2024 Nobel Prize in Chemistry for the work. Early demonstrations show models that formulate hypotheses, run experiments, and write papers that pass peer review.[11]

The capability is arriving. The question is whether what it produces will be knowledge or noise.

Within years—not decades—AI systems will generate the majority of new scientific claims. The ratio of AI-generated to human-generated claims will be 10:1, then 100:1, then 1000:1. A system that generates a thousand hypotheses per hour will outproduce any human team. The economics are overwhelming.

At human scale, we could—barely—rely on manual verification. A reviewer reads a paper. A clinician checks a source. It was slow, incomplete, and missed most errors—but it was possible.

At AI scale, manual verification is impossible. A thousand papers per day. A million claims per week. No human can read them. No institution can review them.

This will either be the greatest acceleration of human flourishing in history—cures for every disease, cheap clean energy, problems solved that we thought would take centuries—or it will be an ocean of confident, fluent, uncheckable outputs where knowledge and hallucination are indistinguishable.

The difference is infrastructure.

The same infrastructure that made AI coding assistants possible must be built for science. Without it, the gigafactory produces noise. With it, the gigafactory cures disease.

This essay describes that infrastructure and argues we must build it now—before the window closes.

The Pattern

Every age builds its own Library of Alexandria. Every age loses it.

[10] AlphaFold Protein Structure Database. alphafold.ebi.ac.uk.
[11] Sakana AI, "The AI Scientist," 2024. sakana.ai.

The pattern repeats. When knowledge travels differently, new infrastructure follows—not just for production, but for arrival.

Before writing, knowledge traveled by voice alone. A medicine that worked could only spread as far as a speaker could walk. Writing changed the physics: knowledge could outlive its creator. The printing press changed the economics: a discovery in Florence could reach London within months. Each transition required new infrastructure for finding, not just producing.

In the third century BCE, the Library of Alexandria became the greatest repository of knowledge the world had ever seen. But the Ptolemaic scholars understood that storage alone was worthless. So Callimachus created the *Pinakes*—a vast catalog organizing holdings by subject and author.[12] Structure imposed on chaos. The same pattern reinvented in every age.

The Library was not destroyed in a single fire—it declined over centuries, through neglect and conflict and the assumption that someone else was preserving it. We repeat that loss every day. Not through flames, but through fog. Knowledge sits in journals no one reads, in lab notebooks no one will see, in the minds of specialists who will retire or die. We are not losing knowledge to fire. We are losing it to fragmentation.

Borges imagined a Library of Babel—infinite galleries containing every possible permutation of text. In theory, all knowledge; in practice, unfindable. We have built it: 37 million articles in PubMed, no structure to navigate them.

In 1609, Galileo pointed a telescope at the night sky and discovered the moons of Jupiter. But discovery was not accident. He could only find what he found because he knew where to look. The constellations gave him that knowledge—they told him where familiar objects were, and the gaps between told him where to search. When he saw points of light near Jupiter that moved night after night, he recognized them as anomalies precisely because the structure made anomalies visible.

This is what infrastructure does. It doesn't make discoveries for you. It makes discoveries *findable*. We have built the telescope. We have not built the constellation.

The stars were always there. It was the lines between them that made navigation possible.

Others have tried. They failed. We know why.

[12] Callimachus's *Pinakes*, circa 245 BCE, is considered the first systematic bibliography. Wikipedia: Pinakes.

Vannevar Bush imagined the Memex in 1945—associative trails that never fade.[13] He described what we are building. He never built it. Paul Otlet spent forty years building the Mundaneum—twelve million index cards, the dream of organizing all human knowledge. In 1940, the Nazis burned it. He died four years later, having watched his life's work destroyed. The Semantic Web demanded ontology agreement before use. The agreement never came.

Each failure teaches the same lesson: *structure must emerge from use, not precede it.*

We learned three things:

Ship early, not perfect. Perfect systems never ship. We begin with three primitives—points, links, and trails—and discover what else is needed by building.

Extract, don't prescribe. The Semantic Web required consensus. We extract structure from existing papers. The bottleneck shifts from "get everyone to agree" to "verify what was extracted."

Distribute by design. Otlet's Mundaneum was centralized and could be destroyed. The constellation is distributed from the start. Every copy is complete. The protocol survives any single failure.

These are not just principles. They are antibodies against the diseases that killed our predecessors.

What Software Learned

Software learned something we haven't.

In 2005, Linus Torvalds released Git—a tool for tracking changes in code. Before Git, collaboration was chaos: patch emails, FTP servers, overwritten files. Developers worked in isolation, and merging their work was agony.

Git solved the foundation problem. Every change tracked. Every decision preserved. Every history queryable. But Git alone was not enough.

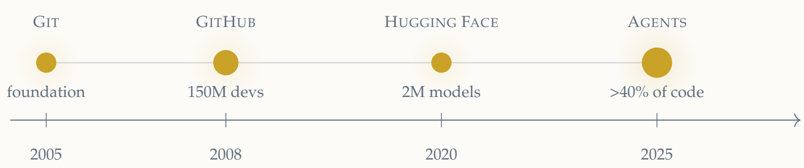
In 2008, GitHub launched—a platform that made Git social. Suddenly developers could discover projects, fork them, propose changes, build reputation. The friction of

[13] Vannevar Bush, "As We May Think," *The Atlantic*, July 1945. The Atlantic.

collaboration collapsed. Today, over 150 million developers work on GitHub.[14] They pushed nearly one billion commits last year. They merged 43 million pull requests *per month*.

Then came Hugging Face—the same pattern for machine learning. Its model hub, launched in 2020, now hosts over two million models.[15] Before Hugging Face, sharing AI models was chaos: broken links, undocumented files, incompatible formats. Now a researcher can share a model with a click, and anyone can use it.

And then came the agents. First as assistants, suggesting code alongside developers. Now autonomous—navigating codebases for hours, creating pull requests without supervision. By early 2025, AI writes over forty percent of the average developer's code.[16] In some languages, over sixty percent.



Foundation → Platform → Agents. The stack built in order.

Here is the crucial point: **AI writes code because this infrastructure exists.**

Copilot works because it operates on repositories with clear history, documented dependencies, queryable context. It can see what the code is, how it evolved, what it connects to. Without GitHub, there is no Copilot. Without Git, there is no GitHub.

AI uses GitHub for code. What does it use for everything else?

For history: documents. For law: documents. For policy: documents. For medicine: documents. For science: documents.

[14] GitHub, "About," accessed January 2026. https://github.com/about.
[15] Jiang et al., "Anatomy of a Machine Learning Ecosystem: 2 Million Models on Hugging Face," arXiv:2508.06811, 2025. arxiv.org/abs/2508.06811. The first million took over 1,000 days; the second million arrived in just 335 days.
[16] GitHub, "The economic impact of the AI-powered developer lifecycle and lessons from GitHub Copilot," updated May 2024. github.blog. In Python and JavaScript repositories, the percentage exceeds sixty percent.

Documents are containers. The claims inside—the actual assertions about reality—are not accessible as structure. AI reads documents the way a human reads a foreign language without grammar: it can pattern-match, but it cannot parse.

This is why AI can write code better than it can do science. Not because science is harder. Because code has infrastructure. Knowledge has none.

Why not just use Git? Git tracks lines of text—insertions, deletions, changes. But a paper is not a collection of lines to be diffed. It's a container for points: findings, measurements, methods, results. Git can tell you how a document changed. It cannot tell you how the evidence state of a finding changed. It cannot propagate a retraction to everything that cited it. It cannot answer: *which findings about this mutation have replicated?*

The substrate question is not theoretical. It is being answered—for code. The Model Context Protocol is becoming the universal interface for connecting AI agents to tools and data.[17] AGENTS.md guides AI coding agents across repositories. The infrastructure for AI coding is being standardized at protocol level.

Science has no equivalent.

AI science agents are arriving. FutureHouse runs literature review at scale. ChemCrow controls spectrometers and liquid handlers. The AI Scientist writes papers that pass peer review. But these agents operate on documents—PDFs, databases, API endpoints. They read science. They do not navigate it. MCP connects agents to tools. What connects agents to knowledge?

The primitive for code is the line of text. The primitive for knowledge is the point—a finding, a measurement, a method, a result.

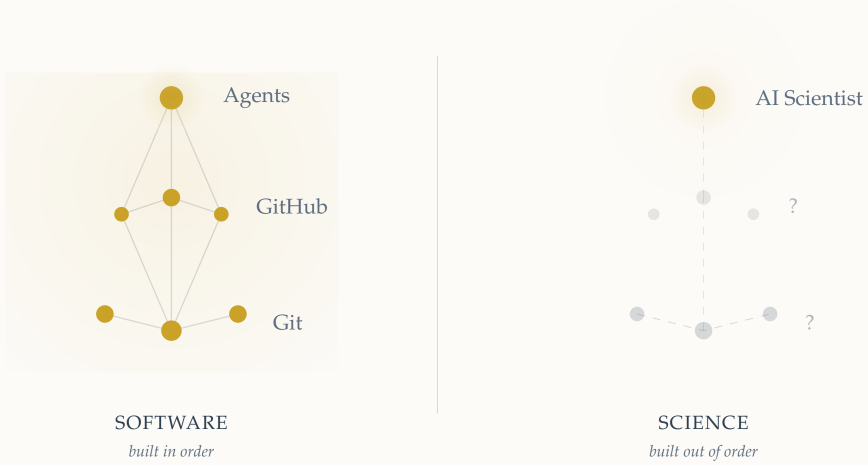
Git versions lines. We need infrastructure that versions points—with evidence, confidence, and lineage attached. The links between them make navigation possible.

AI trained on documents generates documents. AI trained on structured knowledge—with evidence, confidence, and lineage—generates knowledge that can be verified, corrected, and built upon. The difference is not incremental. It is the difference between an oracle that says "trust me" and infrastructure that says "here is why, and here is how to check."

[17] The Model Context Protocol (MCP), released by Anthropic in 2024 and adopted by OpenAI, Google, and Microsoft, was donated to the Linux Foundation's Agentic AI Foundation in December 2025. linuxfoundation.org.

The constellation is that infrastructure. Not a tool for scientists that AI happens to use. The substrate that makes AI useful for knowledge work at all.

The software world built its revolution in layers: foundation, platform, agents. We have skipped to agents.



Software built its stack in order. Science jumped to agents with nothing underneath.

Why Now

AGI is a crowded field—$109 billion in 2024, every lab racing the same benchmarks.[18] That canvas is saturated.

But the problems this essay describes—knowledge infrastructure, the substrate for points and evidence, the architecture that lets what humanity knows actually travel—this canvas is almost empty. We have version control for code and version control for nothing else.

The question is no longer whether it should exist. The question is why it doesn't yet, and what has changed to make it possible now.

The AI argument for urgency is real—we must build structure before agents entrench on documents. But while we debate, children are being misdiagnosed. Researchers are duplicating work. Treatments are being delayed. The seventeen-year gap is not a future problem we might prevent. It is a present harm we are permitting. Every year without structure is a year of preventable loss.

[18] Stanford HAI AI Index Report, 2025. aiindex.stanford.edu.

Why this hasn't been built

People have tried. They failed. Why?

Beyond the Semantic Web's sequence problem, the deeper issue is incentives. The system is not broken. It is working exactly as designed—for the people who designed it.

Publishers put lifesaving research behind paywalls and call it "protecting intellectual property." A doctor in Nairobi cannot read the paper that would save her patient because a company in Amsterdam demands $40 for access. The publisher paid nothing for the research. The authors received nothing for their work. The patients pay with their lives. This is not a market failure. It is extraction.

Journals reward novelty because novelty attracts attention. Replication studies—the work that actually tells us what is true—are rejected as boring. A paper claiming a breakthrough gets published. The twelve papers failing to replicate it do not. The literature fills with false positives. The journals profit either way.

Tenure committees count publications. Not replications. Not infrastructure. Not the tools that would help everyone. A researcher who spends five years building something useful for the field gets passed over for one who published twelve incremental papers. The incentives are not misaligned by accident. They were designed by people who benefit from the current system.

This is not inertia. It is opposition. The constellation threatens everyone whose power depends on knowledge staying fragmented.

The cascade of loss

Consider how knowledge moves in a lab. A student's experiment goes into a notebook. If publishable, it enters the record. If not, the notebook goes on a shelf. When the student graduates, the shelf is cleared. Software solved this with version control: every change preserved, every dead end recorded. Science has no equivalent. For every ten experiments, perhaps one reaches publication. The rest vanish—failed approaches, calibration tricks, the context that would save someone months. The metric must change: not papers published, but paths left behind—including the ones that ended at a wall.

No one's job is to make knowledge arrive. When a patient dies because knowledge didn't arrive, no one counts it. When a researcher wastes three years duplicating work, no one tallies it. When a student clears the shelf, no one mourns what was lost. The cost is everywhere. The blame is nowhere.

So the lines don't get drawn. The notebook sits on the shelf. The shelf gets cleared.

Git succeeded because one person needed it badly enough to build it. GitHub succeeded because it made Git accessible—lowered the barrier until adoption cascaded.

The constellation requires the same: someone who needs it badly enough to build it, and a design that makes adoption cascade.

The adoption objections

This sounds like another knowledge graph project. Those always fail.

The graveyards are real. Knowledge bases that went stale. Graph databases that promised interoperability and delivered silos. The pattern: build the perfect structure, then wait for the world to use it. The world never does.

The constellation inverts this. It does not require agreement before use. LLMs parse papers that already exist. Trails record clinical decisions that are already happening. The structure is not prescribed; it is observed. The difference between "agree on a schema" and "extract what's already there" is the difference between the Semantic Web and Google.

Who will do the validation work? It's thankless.

Wikipedia proved that distributed validation scales when three conditions hold: the work is granular (small edits, not monographs), reputation accrues (edit counts, barnstars, admin status), and the stakes are visible (your contribution appears on a page millions will read). The constellation inherits this design. Confirming a point is a single action. Contribution is credited and visible. The difference from Wikipedia: domain expertise is weighted. A hematologist's confirmation of a hematology point carries more weight than a layperson's. The incentive is not altruism—it is reputation in a system that tracks contribution.

How do you bootstrap adoption when a small constellation is useless?

The cold-start problem is real. A network with 1,000 points provides little value; one with 10 million transforms how science works. The path is vertical before horizontal: one field (oncology), one problem (drug interactions), one community (clinical trialists) where the density of points justifies the infrastructure. RECOVERY worked because the trial structure was useful to each hospital independently, not only when every hospital joined. The constellation must provide local value before it provides global value. Clinicians use it because it helps them today, not because it might help science tomorrow.

The flywheel works like this: a clinician logs one patient's response and finds a researcher who studied that mutation. The researcher finds a bioinformatician who curated the dataset. Each connection creates the next. The constellation grows not by recruitment but by *use*—each trail drawn makes the next one easier to find.

Git spread through Linux kernel developers, then open source projects, then enterprise. Wikipedia spread through early encyclopedists, then topic experts, then the general public. The pattern is always the same: useful to a small group first, then the small group pulls in the next circle.

What does deployment look like? A pilot in one hospital system, one disease area. Trails captured at the point of decision—not surveillance, but structured reflection that already happens informally. Privacy preserved because trails reference points, not patients; PHI never enters the constellation. Integration as decision support, not clinical authority: the system surfaces relevant evidence and shows its lineage, but the physician decides. Liability stays where it belongs. The goal is not to replace judgment but to make it informed.

But consider what happens when structure exists.

When COVID hit, every country ran trials. Most fragmented—each hospital with its own protocol, incompatible with the others. The UK did something different.

In late February 2020, Martin Landray, an Oxford epidemiologist, emailed Jeremy Farrar, then director of the Wellcome Trust. A few days later, they discussed it on a No. 18 bus to Marylebone. Farrar suggested Landray join forces with Peter Horby, an infectious disease specialist. Within nine days of writing the protocol, the first patient was enrolled.

The RECOVERY trial launched with shared infrastructure: web-based randomization that any hospital could use, a single ethics committee approval instead of 180 separate applications, minimal data collection integrated with NHS electronic records. The structure made participation effortless. One in six COVID patients admitted to UK hospitals entered the trial.[19]

[19] Wellcome Trust, "The Story of RECOVERY," 2021. wellcome.org. The trial ultimately enrolled over 47,000 patients across 180 hospitals.

Within 100 days, RECOVERY had enrolled 11,000 patients and delivered its first result: dexamethasone reduced deaths by one-third in ventilated patients. The drug costs £5. It has since saved an estimated one million lives worldwide.

"It's very, very rare," Landray reflected, "that you announce results at lunchtime, and it becomes policy and practice by tea time, and probably starts to save lives by the weekend."

Lunchtime to saving lives by the weekend.

This is what arrival looks like when structure exists: the knowledge does not sit in a journal for seventeen years. It moves at the speed the moment demands because the infrastructure is there to carry it.

Same pandemic. Same virus. Same doctors. Same patients. The difference was structure.

AlphaFold is the same lesson at scale: a foundation layer (the Protein Data Bank) enabled a capability leap. Without foundation layers, capability does not translate. The UK's 2025 AI for Science Strategy states it plainly: "We don't have AlphaFold equivalents in other fields because we don't have PDB equivalents." Foundation data infrastructure is the acknowledged bottleneck.[20]

If the constellation exists—if points are versioned, trails are recorded, provenance is preserved—researchers discover relevant work in hours instead of months. Physicians query evidence by symptoms and population, not keywords. Graduate students navigate the frontier instead of reconstructing what senior researchers know. The same transformation Git and GitHub made for code, but for knowledge—where the stakes are measured in lives.

Discovery is not delivery. That is the gap.

Google organized the world's information. It made documents findable. Before Google, the information existed; finding it took hours. After Google, seconds. Entire industries emerged because the cost of finding collapsed.

The constellation would organize the world's knowledge. It would make knowledge navigable—not just findable, but verifiable, correctable, connected to evidence. Google answers "what documents mention X?" The constellation answers "what do we know about X, with what confidence, based on what evidence, with what dissent?"

[20] UK AI for Science Strategy, Government Office for Science, 2025.

That is a categorical difference. Access to information enabled the last technological era. Access to structured knowledge enables the next.

The missing layer

The Genesis Mission is building the compute layer.[21] But compute has never been the bottleneck. Engines need roads. The constellation is that road network—and the problem is planetary. A researcher in Nairobi has the same right to navigate the frontier as one in Boston.

Someone will build this. The question is whether it will be built in the open, for everyone—or captured, like so much infrastructure before it.

Open means open protocol (anyone can implement a client), open schema (points are portable across systems), and permissive licensing (contributed knowledge under CC BY). Infrastructure that can be enclosed will be.

What happens if we don't build it open?

We continue as we are. Another generation wastes years reconstructing knowledge that exists but can't be found. Another wave of patients receives treatments that arrive a decade late.

Or worse: someone builds it closed. Proprietary. A constellation you can only see if you pay. The map of what humanity knows—owned by a company optimizing for quarterly returns.

Neither future is inevitable. But look at what persists: peer review emerged in the 1600s and has barely evolved. Most technologies from Newton's era have been replaced. This one we preserve, like a family heirloom that no longer works but we cannot bear to throw away. The journal article took its modern form in the 1800s and still dominates. The incentive structures that reward publication over replication, novelty over reliability, access over arrival—these have calcified over decades. Defaults have a way of becoming permanent.

AI will accelerate discovery. Models will read papers faster than humans, generate hypotheses, run experiments. The rate of new findings will explode. But if no one builds

[21] Executive Order, "Launching the Genesis Mission," The White House, November 24, 2025. whitehouse.gov.

transmission infrastructure, we will have more discoveries and the same arrival problem. The gap between what is known and what reaches patients will widen, not close.

This is the blank canvas. What we draw on it will outlast us.

We have named the problem. Now for the structure itself.

The Constellation

Claude Shannon proved that information could be measured, transmitted, and corrected—that noise could be separated from signal, that errors could be detected and fixed. His mathematics enabled the digital age: every text message, every streamed video, every financial transaction depends on the theory he developed in 1948.

The same problem exists for knowledge. Scientific findings are signals; noise is the unreplicated results, the retracted papers, the errors that propagate uncorrected. We have built systems to transmit knowledge—journals, databases, search engines. We have not built systems to correct it.

The constellation is Shannon's insight applied to knowledge: transmission with error correction built in.

Consider the difference between an artifact and a point.

A paper is an artifact: "Chapman et al. 2011, *New England Journal of Medicine*." A dataset is an artifact. A protocol is an artifact. You can cite artifacts. You cannot query them. You cannot ask: has this been replicated? Under what conditions does it hold? What depends on it?

A point is a unit of knowledge: "Dexamethasone reduces mortality in severe COVID patients." But the point knows more than its statement. It knows its evidence: the RECOVERY trial. It knows its constraints: benefit in patients requiring oxygen, no benefit in those who do not. It knows its replications and dissent. It knows its lineage: what it superseded, what it updated, what depends on it.

When a new study challenges the point, the structure updates. When a contradicting paper is retracted, the point's confidence updates accordingly. When a clinician queries the constellation at midnight, she sees not just the assertion but its entire chronicle: what supports it, what challenges it, how certain we are, and why.

But points do not only come from papers. A point might originate from a published study (the current default, and a small fraction of what is known), a failed experiment (almost never published, but critical for those who would try the same thing), a clinical decision (made thousands of times daily, almost never recorded), a lab protocol (the real one that works, not the sanitized version in Methods), an expert's tacit knowledge (patterns recognized but never articulated), or an AI system's output (requiring verification before it becomes knowledge).

The constellation accepts points from all sources. The source is tracked—a point from a replicated RCT carries different weight than a point from a single lab notebook. But the primitive is the point, not the container it came from.

The constellation captures what journals reject: negative results. The majority of experiments never reach publication. Most are negative—approaches that didn't work, hypotheses that failed, conditions that produced no effect. In the current system, they vanish into lab notebooks, cleared when students graduate.

In the constellation, a failed experiment becomes a structured point: "This approach was tried under these conditions and did not work." The next researcher doesn't repeat the failure. They see the wall and try elsewhere.

A single negative result, properly recorded, might save a hundred labs a year each. Multiply that by the millions of experiments that currently vanish. The dark matter of science—structured, queryable, navigable—would dwarf the visible literature.

Before points, there are entities—the stable referents that points are about. A gene, a drug, a mutation, a patient population. Entity resolution is where previous knowledge systems died: the same thing with twelve names, twelve things with the same name. The constellation requires a foundation layer of mapped, versioned identifiers. Without it, points cannot link. This is infrastructure, not application.

Extraction

Large language models can read papers and protocols and propose structured points with enough accuracy that human validation, not human reading, becomes the limiting step.

An LLM reads a paper and proposes: "This paper asserts X, with evidence Y, under conditions Z." A domain expert confirms, corrects, or rejects. The validated point enters the constellation with its validation history attached. One expert-hour can validate dozens of proposed points. The work that once required reading every paper becomes review of proposed structure.

One worked example

Consider a point extracted from the literature: *"BRAF V600E mutation predicts response to vemurafenib in metastatic melanoma."*

What does the point require?

Entities: BRAF is a gene. V600E is a variant. Vemurafenib is a drug. Metastatic melanoma is a condition. Without stable referents, you cannot say "this gene" and mean the same thing across papers, databases, and time.

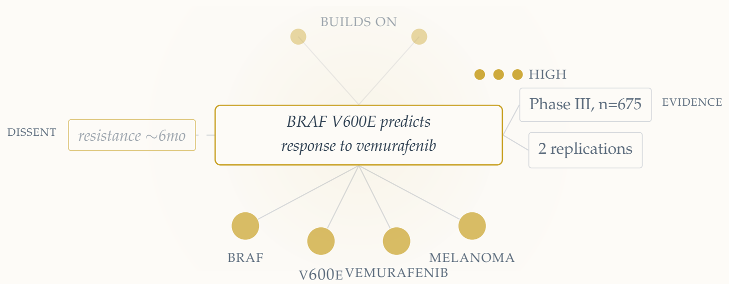
Evidence: Chapman et al. 2011, a Phase III trial with 675 patients. Two independent replications.

Confidence: High—multiple independent trials show consistent effect.

Dissent: Resistance typically develops within six months.

Lineage: Which findings this builds on, which it supersedes.

A point is not a static assertion. It has a *chronicle*: how it emerged, what evidence accumulated, what challenges arose, how confidence evolved. The BRAF point was uncertain in 2010, strengthened by Phase III trials in 2011, complicated by resistance data in 2012, and refined by combination therapy evidence since. A clinician querying the constellation sees not just the current state but the full history—and can judge whether the trajectory is toward consolidation or fragmentation.



A point is a queryable object. When a new study appears—supporting, contradicting, or refining—the point's status updates. The constellation evolves.

Evidence version control

The constellation is version control and auditability for scientific knowledge. Note-taking systems organize documents. Review aggregators summarize. Guidelines prescribe.

Search engines retrieve. None of them version the evidence state of an assertion—what supports it, what challenges it, how confidence changes.

Existing tools—UpToDate, Semantic Scholar, Elicit, Scite—each address a symptom. The constellation addresses the substrate.

Without structure, errors compound. An AI generates a point with moderate confidence. The confidence is not recorded—just the assertion. Another AI cites it. A third builds on the citation. By the time a human encounters the downstream assertion, five generations of inference have intervened. The original evidence is buried. The uncertainty was never tracked.

This is *confidence laundering*: a point that was 60% certain becomes, through citation without structure, indistinguishable from established fact. The constellation exists precisely to prevent this. Every point carries its confidence. Every citation preserves uncertainty. Every correction propagates.

Points are not static. They move through states: *proposed* when first extracted, *validated* when expertise confirms them, *contested* when evidence conflicts, *superseded* when newer findings arrive, *retracted* when wrong. Each transition is logged. Confidence is not a number someone assigns—it emerges from the evidence profile and validation history. Dissent is preserved, not resolved: a point can show that three trials support it and one challenges it. The structure holds the disagreement rather than hiding it. This is what makes correction possible. A retraction propagates to everything that cited the point. The scar is visible. The constellation learns.

· · ·

Points alone are not a constellation. The connections between them are links—typed relationships that make navigation possible. A link might be *evidence* (this study supports that finding), *dependency* (this method assumes that protocol), *dissent* (this result challenges that claim), or *lineage* (this point supersedes that one). Without links, points are isolated facts. With links, they form a navigable structure. The Web proved that links create networks; the constellation proves that *typed* links create structure you can reason over. Link is the second primitive.

And there is a third primitive: the trail.

A trail is a recorded path through knowledge—what you searched, what you found, what you decided, and why. A trail might record: *searched BRAF variants in metastatic melanoma; filtered to Phase III trials with n>200; rejected three papers—one retracted, two with undisclosed industry funding; selected vemurafenib protocol based on two independent replications and six-month resistance timeline.*

Consider what happens without them.

Henrietta Lacks was thirty-one when she died of cervical cancer in a segregated ward at Johns Hopkins Hospital. Her doctors took a sample without her knowledge or consent. Those cells—HeLa cells—became immortal: polio vaccine, cancer treatments, gene mapping, COVID vaccines. Over 110,000 publications and billions in research.[22]

Her family didn't know. For decades, they couldn't afford health insurance while researchers around the world built careers on her contribution. The knowledge traveled everywhere. The trail back to its source was never built.

At a memorial lecture years later, her great-granddaughter Veronica Robinson answered questions about her family's experience.

"This comes down to a question," she said, "of how we bridge the gap between the community and the health-care field."[23]

Two years later, her family was still fighting for recognition and redress. Seventy years after her death, the gap was never bridged.

I didn't have an answer then. I am trying to build one now.

· · ·

Here is what arrival looks like when the structure exists.

In the Netherlands, a baby girl was born with shortened bones and failing kidneys. She died within seven weeks. Her clinicians had no diagnosis. They sequenced her genome, found two candidate genes, and entered her case into Matchmaker Exchange—a federated network for rare disease research. Within days, they connected with researchers

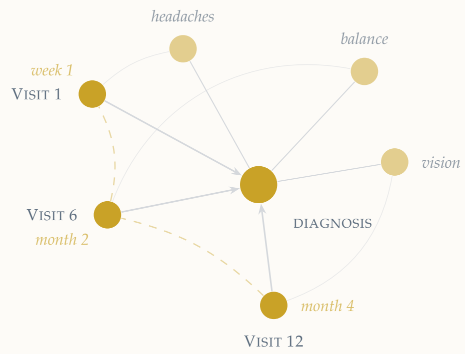
[22] Rebecca Skloot, *The Immortal Life of Henrietta Lacks*, 2010. Publisher page. HeLa cells remain one of the most important tools in biomedical research; the Lacks family received no compensation until 2023.

[23] Veronica Robinson, remarks at the Henrietta Lacks Memorial Lecture, Johns Hopkins University, October 2019. Johns Hopkins News-Letter coverage.

in Germany who had seen the same mutation. Then Canada. Then Portugal. Then the UK. Nine patients across five families on four continents, linked through a single API. They named the disorder. The research led to a potential treatment.[24]

The baby in the Netherlands didn't survive. Her data did. It traveled the structure her clinicians built, reaching families who had searched for years.

Veronica Robinson's question is still unanswered. How do we bridge the gap? The constellation is one attempt.



This is what a constellation looks like: points linked to evidence, trails linking decisions to paths, the whole structure versioned over time. Navigation for the clinician who needs to know what works. Discovery for the researcher who needs to see what's missing.

The cognitive scientist Dan Sperber observed that culture emerges from two basic acts: making private knowledge public, and internalizing what others have made public. The constellation operationalizes this for knowledge. Points are externalized understanding. Trails are internalization paths. The structure is culture—made queryable.

The astronomer uses constellations for both purposes. Where is Mars tonight? The constellation tells her: near this star, in that region of sky. Where should I point the telescope? Not at bright stars—at the dark spaces between, where structure predicts something should be but nothing has been found.

The same structure serves both. The map of what is known is simultaneously a map of what is missing—gaps made visible, queryable, navigable.

[24] Oud et al., "Mutations in EXTL3 Cause Neuro-immuno-skeletal Dysplasia Syndrome," *American Journal of Human Genetics*, 2017. doi:10.1016/j.ajhg.2017.01.013. PMID:28132690.

The constellation is not just for humans. It is infrastructure for AI scientists.

Today's models read documents. They summarize, synthesize, hypothesize—but they cannot verify. They cannot tell you which findings replicated, which were retracted, which depend on evidence that quietly failed. They navigate fog.

The constellation gives agents something different: a queryable frontier. An agent can ask: where is evidence thin? What points are in tension? What gaps imply the next experiment? It can propose hypotheses as structured points—not prose, but assertions with specified evidence requirements.

This is the difference between Copilot and an AI that reads Stack Overflow posts. Copilot works because it operates on structured repositories—version history, dependencies, tests. AI scientists will work when they operate on structured knowledge—points, trails, provenance.

The constellation is also infrastructure for the techniques AI labs are already building. Retrieval-augmented generation (RAG) reduces hallucination by grounding outputs in retrieved documents—but documents are containers, not points. A retrieved passage may contain multiple assertions with varying confidence, some retracted, some superseded. The model cannot distinguish.

RAG over structured points is categorically different: the model retrieves assertions with explicit evidence, confidence scores, and lineage. It knows not just *what* was asserted but *how certain* we are and *what challenges exist*. This is the difference between "I found a document that mentions X" and "X is supported by three independent trials with 0.85 confidence and one active dissent."

And structured points become ground truth for evaluation: benchmarks for AI scientific reasoning, calibration targets for confidence, test sets for verification. The constellation doesn't compete with AI capability research. It provides the substrate that makes capability trustworthy.

But the constellation is not just infrastructure for individual AI agents. It is the substrate for AI coordination.

When AI systems need to build on each other's work—when one agent's output becomes another's input—what ensures the chain is sound? When a hundred AI agents work on related problems, how do they share what they've learned? When an AI-generated point turns out to be wrong, how does the correction reach every system that used it?

For code, these problems are solved. Git tracks changes. GitHub coordinates contribution. CI/CD propagates verification. An organization of developers—human or AI—can work on shared codebases because the infrastructure exists.

For knowledge, there is no coordination layer. AI systems generating points are isolated. When they're wrong, nothing propagates. When they're right, nothing compounds. We have AI organizations—but no infrastructure for AI organizations to share knowledge.

The constellation is the coordination layer. The infrastructure that lets AI systems—individually and in organizations—operate on shared knowledge with the same rigor they now operate on shared code.

When the constellation exists, participation looks less like credentialed gatekeeping and more like open source: validate a point, record a trail, contribute a correction. AI does retrieval and extraction; humans do judgment. Trails teach newcomers how experts reason—what they searched, what they rejected, what they trusted. A high school student can make a meaningful contribution. The limiting factor becomes curiosity and care, not access to a locked library of PDFs.

The horizon

The first constellations will be incomplete but queryable: points extracted from a decade of literature, trails from a small network of specialists, dissent and corrections attached where they belong.

As the structure grows, the bottleneck shifts to validation. Contributors do not need to read every paper; they review proposed points, attach constraints, and record trails. The frontier becomes navigable: dense where evidence is strong, jagged where it is not.

The frontier of knowledge is jagged—peaks where funding flows, valleys where it doesn't, entire continents unexplored because no one thought to look. The constellation reveals this topology. Without the lines, the gaps stay invisible.

Cracks are part of any system that grows. In *kintsugi*, broken pottery is repaired with gold and the fracture remains visible. Knowledge needs the same property: errors detected, corrected, and propagated.

In January 2014, Haruko Obokata published two papers in *Nature* claiming a revolutionary method for creating stem cells—just dip cells in weak acid.[25] Labs worldwide scrambled to replicate.

Ken Lee, a stem cell researcher at the Chinese University of Hong Kong, decided to document his attempt publicly. He live-blogged his replication on ResearchGate—the first live-blogged scientific experiment.[26] Within weeks, he identified the problem: what Obokata had seen was likely autofluorescence, cells glowing from stress rather than transformation. Her revolutionary discovery was an artifact.

Lee submitted his findings to *Nature*'s "Brief Communications Arising"—the journal's mechanism for publishing corrections. *Nature* rejected it. No clear explanation. The journal that published the flawed papers refused to publish the correction.

For six more months, labs worldwide continued wasting resources. Rudolf Jaenisch at MIT: "Many people wasted their money and their time and their resources on repeating this."[27]

The correction existed, but the system could not propagate it.

A constellation makes corrections durable and discoverable. A retraction links to what it corrected. A failed replication updates the points that depended on it. A dissent stays attached to its evidence until the structure of evidence shifts. Errors stop living as footnotes and start living as updates.

The window is open now. The child in the next waiting room deserves better. We can build the system that makes that difference.

[25] Obokata et al., "Stimulus-triggered fate conversion of somatic cells into pluripotency," *Nature*, January 2014. doi:10.1038/nature12968. Both papers were retracted in July 2014. doi:10.1038/nature13598.

[26] Lee's replication log was publicly documented on ResearchGate beginning February 2014. Coverage in David Cyranoski, "Acid-bath stem-cell study under investigation," *Nature News*, February 2014. doi:10.1038/nature.2014.14738.

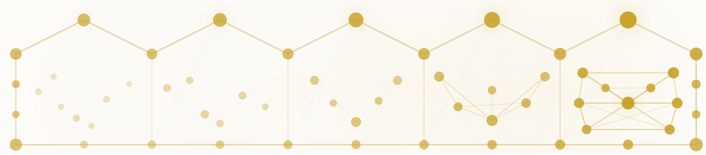
[27] Quoted in David Cyranoski, "STAP cells are finally dead," *Nature News*, September 2014. nature.com/news/stap-cells-are-finally-dead-1.15977.

The Gigafactory

A gigafactory is a production system—standardized processes, continuous operation, compounding efficiency. Tesla did not invent the electric car. It built the system that made electric cars manufacturable at scale.

Science today is artisanal. Each project starts over. Each lab reinvents. Each finding sits in a journal, disconnected from what came before and what comes after. We have the intelligence. We have the tools. We lack the production system.

The constellation is the foundation layer that makes science manufacturable.



When the foundation exists, science runs differently. Continuous pipelines instead of episodic grants. AI agents navigating the frontier map—seeing where evidence is dense, where it's thin, where the next experiment matters most. Every validated point improving extraction. Every correction propagating. The efficiency curve steepening with each discovery.

Consider what becomes possible.

The seventeen-year gap collapses to months. A discovery in Shanghai reaches a clinic in São Paulo before the paper is published. The millions of failed experiments that currently vanish—calibration tricks, dead ends, the context that would save someone years—become searchable. No one duplicates work that has already failed. Every disease becomes a project with a visible frontier: here is what we know, here is what we've tried, here is where to push next.

The child who would have waited four months gets diagnosed in three days. The researcher who would have spent five years reconstructing known knowledge spends five years extending it. The treatment that would have arrived a generation late arrives in time.

This is not utopia. It is infrastructure. The same transformation that happened when code became versionable, shareable, buildable—applied to everything humanity knows.

This is what makes the compressed 21st century real.

Dario Amodei describes a future where AI accelerates biology by a hundred years in a decade—reliable prevention of nearly all infectious disease, elimination of most cancer, significant extension of healthy lifespan.[28] That future is possible. But discovery is not delivery. The same infrastructure gap that delayed my diagnosis will delay every cure AI discovers—unless we build the structure that makes knowledge arrive.

The constellation is that structure. It is what turns a century of discoveries into a decade of cures. Without it, we get a hundred years of papers. With it, we get a hundred years of lives saved.

But mechanism is not the point. The point is who gets to participate.

Imagine: a high school student opens the platform. She finds a project focused on pediatric brain tumors—the same diagnosis I had as a child. She joins a team of twenty people across four continents. Thousands of AI agents handle retrieval, synthesis, analysis. She directs. She contributes. She matters.

She is not reading about science in a textbook. She is doing science—novel work that has never been done by anyone in history. The spark that most people never get the chance to feel.

Most people never find their gift—not because they lack capability, but because they lack access. The constellation removes the barrier. The gigafactory is what happens when everyone can contribute.

· · ·

Imagine: 2035. A different child, the same symptoms.

She is six years old. She has headaches that won't go away. She is unsteady on her feet. Her mother notices, mentions it to the pediatrician. The doctor listens.

The pediatrician enters the symptoms. The system surfaces a pattern—not from a single study, but from 847 confirmed diagnoses across twelve countries. A trail left by a

[28] Dario Amodei, "Machines of Loving Grace," October 2024. darioamodei.com.

neurologist in Melbourne: "When balance problems present with persistent headache in this age range, order imaging early. I missed one. Don't repeat my mistake."

The system flags the constellation of symptoms. Not an alert—a prompt. A suggestion that connects what this doctor is seeing to what doctors across the world have learned. The trail shows the reasoning: here is why this pattern matters, here is what was missed before, here is what to do.

The imaging happens on day three, not day twelve.

The girl lies on the table. The machine hums. She looks up at the booth. The technicians look at the screen. Their faces do not change. They smile at her. They tell her she did great.

Her mother is waiting when they wheel her back. The doctor comes in a few minutes later.

"We found something," she says. "But we caught it early. The prognosis is good."

The mother holds her daughter. She doesn't know how close it was. She doesn't know about the other children—the ones who waited four months, the ones whose patterns went unseen. She doesn't know that thirty years ago, a six-year-old boy with the same symptoms saw twelve doctors over four months before anyone ordered a scan.

She only knows her daughter is going to be okay.

No twelve visits. No watching doctors' faces change. No learning at six years old that something is very wrong. The structure existed. The light traveled. The knowledge arrived.

• • •

The stars are there—scattered across journals, databases, the minds of specialists, the outputs of AI systems. The lines that connect them are ours to draw.

What we build now determines what the next generation inherits.

The light we received was borrowed from those who came before.[29] The light we pass forward is a choice.

[29] The constellation on the last page includes Sigma Draconis, 82 Eridani, and Delta Pavonis—stars whose light left approximately eighteen years ago and is arriving now. Sigma Draconis. 82 Eridani. Delta Pavonis.

The structure that would save the next child does not exist. I cannot unknow that.

· · ·

The work is already happening—scattered, unconnected, invisible. The constellation connects it.

Build for the child you were. Build what should have existed.

Somewhere, right now, a child is waiting.

So the light arrives.