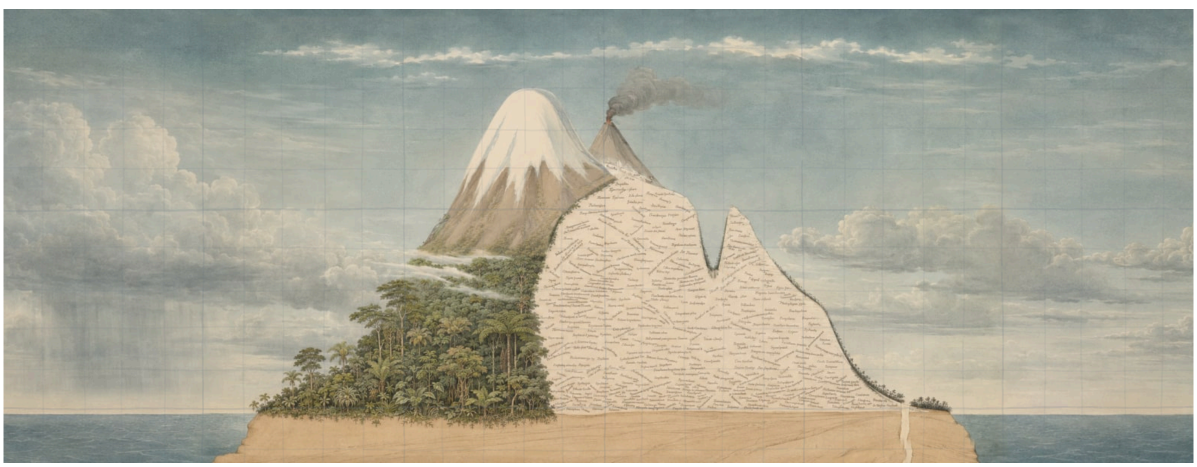
Ryan Orbuch

ryan@lowercarbon.com
x.com/orbuch

Natural General Intelligence

The case for a nature model that learns the state and dynamics of the planet itself.



September 2026

Contents

Introduction

THESE DAYS, I THINK WE COULD ALL USE A REMINDER: however powerful the geniuses in the datacenter, you and I live in the world outside it. Maybe SF people really do need to touch grass, because AI's potential to transform our relationship to that natural world has been overlooked or misunderstood by doomers and accelerationists alike.

Infinite intelligence will not eliminate our need for natural resources, nor will it supplant them, because natural resources are those not produced by human intelligence at all – they are produced only by the Earth system itself. Some are stocks accumulated over geological time, like mineral and oil deposits. Others depend on processes that continually renew them: fish reproducing, nutrients cycling through soils, water moving through a watershed. The conditions that support our farms, cities, and ecosystems emerge from the interaction of these systems.

While AI today promises abundance via *automation* of knowledge work, applying that idea here misses something basic: the Earth system is already fully automated. The processes that produce the natural resources we depend on were running long before humans evolved, and will continue to long after we're gone. Therefore, the opportunity looks less like automation and more like *stewardship*: using greater intelligence to deliberately improve what these systems can provide while sustaining the conditions that make it possible.

In the natural world, misalignment between what we intend and what our actions set in motion is not some new emerging threat but instead centuries of status quo, which no amount of panicked coordination has so far been able to pace or pause. Though we prompt and the Earth responds, we are aware of only some of nature's dials and much of the time we do not notice we are turning them. We alter the atmosphere, redirect rivers, and transform ecosystems, often discovering the consequences much later. We fertilize a field to grow food and end up feeding an algal bloom downstream.

It should be no surprise that incidents abound: in August, ten thousand years' worth of glacier collapsed in an afternoon killing hundreds in Nepal. Throughout the summer, unprecedented European heat killed at least thirty-five thousand people and forced France to take nuclear power plants offline as its rivers ran too warm to safely cool them. Meanwhile in Texas, a flesh-eating livestock parasite we'd eradicated came back with a vengeance and cattle still remain in quarantine across the state. Right now, warm water is building up in the Pacific and this winter it will rearrange drought, fire, and harvests across four continents in perhaps the biggest El Niño since records began. Learning to better anticipate and mitigate all these consequences would be an enormous achievement, but just a small part of the opportunity in stewardship.

What is to be done? I'd like to propose a new framework: **Natural General Intelligence**, a foundation model grounded in the state and dynamics of the planet itself.

Its core would be a general-purpose nature model connecting the biosphere, atmosphere, oceans, land, ice, subsurface, and all of their complex dependencies, capable of anticipating how the planet will respond to intervention and updating itself as it happens.

There is room to become much more ambitious about what the Earth can provide. NGI could understand the structure of this system and enable us to steward the relationships we haven't yet noticed, toward possibilities we haven't thought to ask for. We could stabilize the climate, help depleted fisheries flourish, or discover how to bring more life back to an exhausted landscape. After all, we live in the biggest RL environment there is, albeit one with a single episode and no reset button: the actual environment outside.

The time is now. We have much of the foundation for this project already. Long before LLMs, weather and climate modeling were already pushing the limits of supercomputing. To support that effort, generations of scientists assembled a gigantic observational data archive that's still growing every day.

Though Earth modeling today feels a decade behind the state of the art in LLMs, they suggest a clear path forward: heed the bitter lesson, move beyond hand-tuned features to learn latent structure from vast multimodal data, and expect the magic of generalization. But while LLMs train on books and the internet to attempt to generalize the output of humans, the data needed to train a nature model is not contained in the record of human output. It can only be gathered from direct observations of the real Earth.

I wrote this essay to show you how we could build NGI and why we should. In the following sections, we'll cover:

- **Motivation**: The physical foundations for stewardship, outlining a nature model, and what we might use it for
- **History**: The origins of earth observation and human understanding of its structure
- **Data**: The observing system that forms the foundation of our training data (*interactive!*)
- **Model**: Principles, constraints, and existing efforts
- **Conclusion**: Contextualizing the scientific lineage

As you read through this, get in touch by email or DM if you have a question, a suggestion, or want to work on it. With that said, let's dig in.

Superintelligence Will Be Post-Nature but We Will Not

IN THE 1960S, JAMES LOVELOCK was working at NASA's JPL on what we would today call astrobiology, developing spectrometers to analyze the atmospheric composition of faraway worlds. Asking how life might reveal itself from another planet, he realized that Earth's environment is "a single, tightly coupled process, with the self-regulation of climate and chemistry as an emergent property." This was the foundation of the Gaia hypothesis, published and popularized in the 1970s.

The core idea is that life does not passively inhabit Earth, but **actively maintains the planetary conditions that enable life to continue** – that respiration and recursive regulation of the biosphere is a necessary condition for the emergence of human intelligence in the first place.

> *I think the zone of habitability idea is flawed because it ignores the possibility that a planet bearing life will tend to modify its environment and climate in a way that favours the life upon it, as ours does.*
>
> *... The truth is that the Earth's environment has been massively adapted to sustain habitability. It is life that has controlled the heat from the Sun. If you wiped out life entirely from the Earth, it would be impossible to inhabit because it would become far too hot.*
>
> JAMES LOVELOCK, *Novacene*

Even if we set aside Lovelock's stronger framings of Gaia as woo, the foundational insight is one we have to take seriously: habitability itself is established through feedback. The Earth already runs its own loop, with life continuously self-regulating the conditions that it requires.

The sun has brightened by roughly a quarter since life began, but the oceans never boiled away. Why? Life drew down CO_2 in step via phytoplankton, land plants, and fungus. Oxygen levels have stayed between 15% and 35% for hundreds of millions of years because if it gets too high forests burn faster than they regrow, but too low and fire can't start at all (today's atmosphere is 22%). Organisms change the chemistry of the atmosphere, oceans, and soils, which in turn shapes what can survive.

We participate in these feedbacks continuously, taking in our environment, metabolizing it, and exhaling it changed. This relationship has nothing to do with how smart we are. We get it for free by being alive. And while much of the safety and alignment discussion is about imbuing AI with human values, we cannot overlook the fact that *our values arise in part from the requirements of the human organism*.

AI does not share this dependence. It's the first intelligence on earth which *doesn't breathe*. Though its intellect has sped past that of a bird, a monkey, a dumb human, and even some of the smartest humans without so much as a glance in the rearview, it has yet to surpass a humble plankton when it comes to the task of *participating in the biosphere.*

This is weird! It's weird in a way that typical AI risk fears about bioweapons or cyberattacks don't properly encapsulate, and probably should be considered on an entirely different axis. On that axis, humans and plankton stand united by something that greater intelligence doesn't supply.

The soft squishy coils that carry our consciousness through space and time will forever rely on breathable air, drinkable water, a protective ozone layer, our homes and forests not burning down, tickborne diseases not wiping us out, on those plankton making oxygen and holding up the food web, the insects still showing up to pollinate, mountain snowpack melting slowly enough to keep rivers running when the fields go dry, and temperature and humidity within the envelope that allows us to sweat our bodies cool.

Biological dependence has hardly given humans a complete understanding of the system we inhabit, but it makes the consequences of misunderstanding unusually personal. As AI expands our ability to act on that system, we need to expand its access to evidence about how the system responds. That means deliberately connecting the intelligence we're building to the observing system we've spent generations assembling and developing a nature model that can learn from what those instruments reveal.

The World Model We Need Is a Nature Model

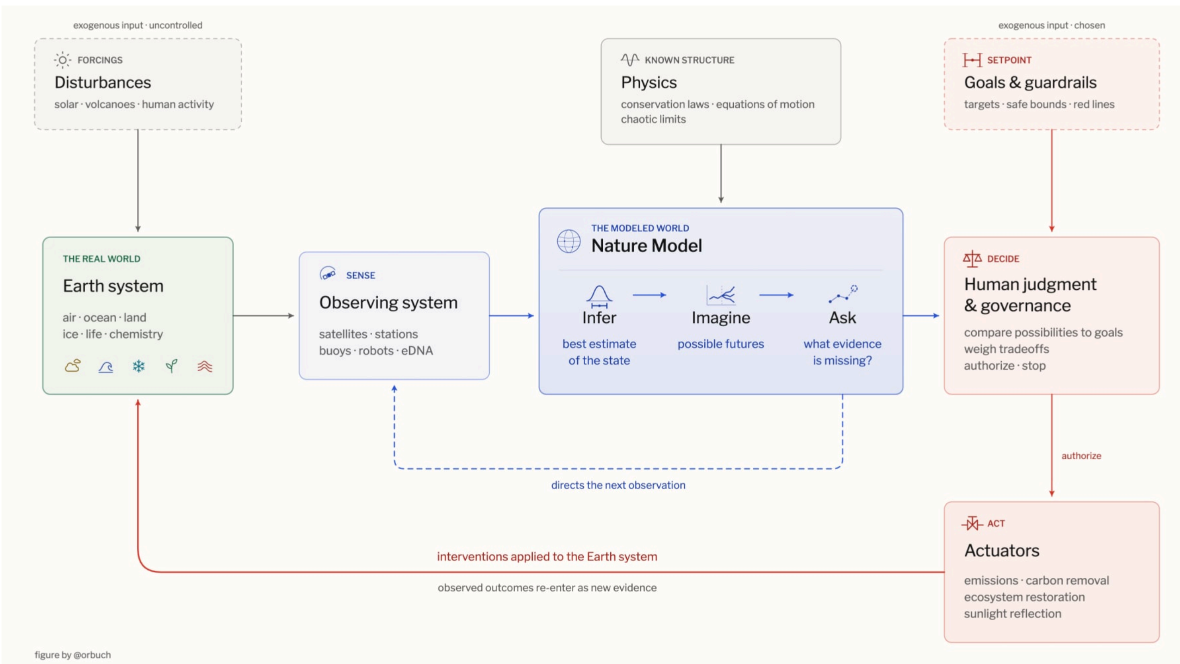
FROM BIOLOGY TO MATERIALS SCIENCE, today's models are being connected into increasingly automated labs where they can execute an experiment and learn from the result. While there is surely some low-hanging advancement in these fields that can be achieved by a language model consuming more papers and ingesting more context than any human ever could, that overhang will inevitably be exhausted because it is a compression of things that are already known. Further discovery is enabled only by a loop where the physical world can respond to the model's hypothesis to expand the scope of knowledge.

The lab enables the model to contact reality, but in doing so, shrinks the world. Select a sample, control the conditions, isolate a few variables, and point an instrument at the result. If the experiment fails, you can usually run it again. If something contaminates the sample, you can start over from scratch.

Nature is not like this. You cannot bring the ocean, the atmosphere, or a forest into the lab in its full causal context. While many important discoveries have been made by bringing back a vial of seawater or observing a tree seedling in a growth chamber, that is not enough. The lab can help to reveal mechanisms, but our understanding of nature is increasingly bottlenecked by state: what is happening in a real forest where heat is interacting with drought, fire, soil microbes, insects, land use, and thousands of other species?

Before LLMs, weather and climate modeling were among humanity's largest sustained computational projects. We have since assembled a gigantic record of the Earth, but until now it wasn't possible to **build an intelligence connected to the real state of the Earth, capable of anticipating how it will respond to intervention and updating itself from what happens.**

Here's an outline of what that might look like:



A nature model is a world model in the most literal sense: a learned model of how an environment evolves, of the kind an agent uses to predict what happens next. What sets it apart from the world models being built for games, video, and robots is the environment itself. It is the actual Earth, we only ever observe a sliver of its state, and there is no reset button. Rather than being organized around any one agent's goal, it would need to begin by representing the Earth's underlying dynamics and leave the objectives open, so that many tasks could later be posed to it.

Because a first instance of this model is necessarily incomplete, there would be three parallel lines of effort:

- One chunk of work would be to ingest observations and estimate the current biogeochemical state of the Earth system, including all that goes unobserved.

- A second would simulate possible futures using learned dynamics constrained by physics.
- A third would identify the direct Earth observations most likely to resolve an important uncertainty and direct sensors to collect it—or flag that we don't yet have an instrument that can.

And to get it out of the way, AI's recent leaps in pure math don't change our approach because the goal is not to characterize the limits of abstract systems – we want to know what state the real system is actually in! No Lean proof will tell you what the ocean is doing right now outside your window, what that means for the fisheries we depend on, or how its currents are shaping conditions on the other side of the planet. You still have to measure it. And if you intervene, you have to figure out what actually changed.

The ultimate vision is a closed loop: the nature model helps us decide what to observe, human judgment informs how to act, and the Earth's actual response returns as new evidence. The same system that informs an intervention can then measure what happened, compare it with what was expected, and update its understanding of the world while improving our ability to shape it.

In control-theory terms, the nature model is the estimator: it infers the state of the Earth from sparse sensors and learns how the system responds to inputs. It would expose which inputs the planet is actually sensitive to: the levers we have been pulling without knowing it, and how to be more intentional about those levers and others.

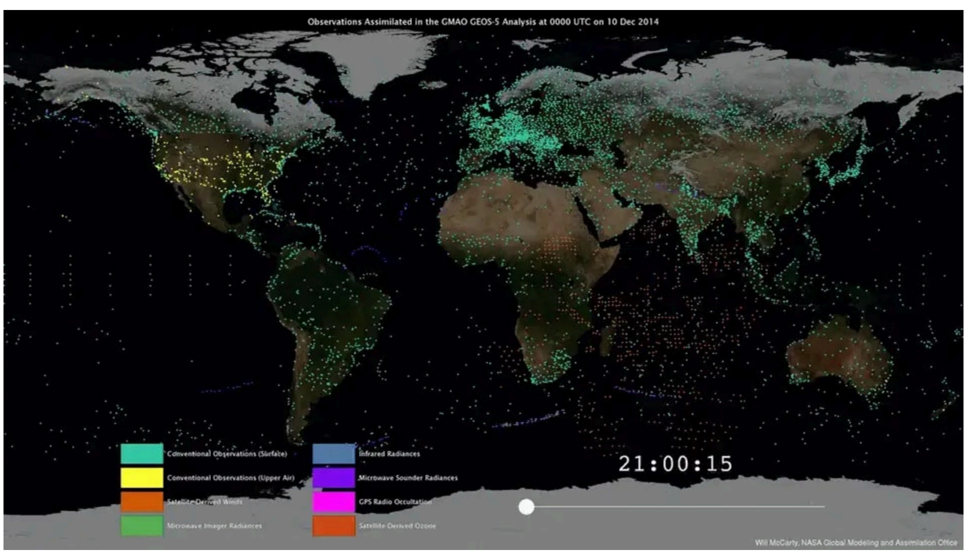
In A Stargate for Data, Will DePue describes the public internet as a "one-time civilizational subsidy" to LLMs: decades of human knowledge accumulated before anyone knew it would become training data. As that subsidy runs out, he argues, the limiting factor shifts from compute to the expensive collection of information that is private, tacit, undigitized, or simply nonexistent online. Nature is all of those.

A nature model's training data will need to comprise a record of the real, actual earth system—not only satellite images of the surface, but the chemical and physical state of the biosphere, oceans, atmosphere, ice caps, rocks in the crust, circulation, energy budget, and how it all couples and changes.

The good news: **We are not starting from scratch**. Over the last half-century, governments and academics have gradually assembled an impressive Earth-observing system:

- Satellites overhead
- Ground weather stations
- Balloons and sensors on aircraft across the atmosphere
- Floats, moorings and ships in the ocean
- Gauges and sensors on land
- Samples of ice cores, sediments, tree rings and corals that extend the record backward in time

To give you some intuition of what this system looks like in practice, here's a *small* subset of the streaming data visualized, courtesy of NASA. You can see flightpaths crisscrossing the US instrumented via the AMDAR program, infrared and microwave satellites and their orbits, inferred atmospheric processes, and more:



From Observations to Models — NASA/Goddard Space Flight Center, Global Modeling and Assimilation Office. svs.gsfc.nasa.gov/30590

All the data sources behind this preview and dozens more are detailed later on in the (*interactive*!) Today's Observing System section.

Just a year or two ago, this would have been impossible. But what excites me most is that the pieces needed to build a serious nature model have been developing independently across disciplines and institutions for decades if not centuries. Now, finally, they look ready to combine:

- **AI:** Models are becoming capable of learning shared representations from enormous, messy, multimodal datasets rather than treating every natural process as a separate problem. Essentially every paper I'll point to in later sections has come out within the last few months.
- **Compute:** We can now train and run models at a scale capable of absorbing decades of planetary observations and simulating many possible futures.
- **Sensing:** Cheaper satellites, autonomous balloons and ocean vehicles, environmental DNA, bioacoustics, smaller sensors, better batteries, and advances in robotics are making observation denser, more continuous, and more programmable.
- **Data:** Petabytes of direct observations stream in annually across dozens of sensing networks while nearly an exabyte of earth data sits in free and open archive repositories, accumulated over half a century of publicly funded observation.
- **Networking:** Starlink and other satellite networks have put the whole surface of the planet within reach of the internet. Sensors in the open ocean, on the ice sheets, and adrift in the atmosphere can now stream data home in real time instead of waiting months for a ship or an expedition to physically retrieve it.
- **Intervention:** We are gaining practical tools to restore ecosystems, modify precipitation, remove carbon, manage invasive species, and alter the planet's energy balance to negate the effects of climate change.

Done correctly, a nature model could provide the world-knowledge layer that any aligned AI would still need and close the loop between what we observe, what we predict, how we act, and how the Earth responds—whether our interventions are intentional or not. (Some principles for how we might build it are discussed later in *Designing A Nature Model*.)

What We Could Do with a Nature Model

INDUSTRIAL CIVILIZATION CONDUCTS and witnesses experiments on the earth system all the time, mostly by accident. When ships reduce sulfur emissions, a volcano erupts, oceans circulate, or an industrial region cleans up its air, it immediately alters our planet's energy balance via changes in heat distribution, clouds, and rainfall. The consequences can affect public health, food production, water supplies, and infrastructure—and yet, the observations arrive fragmented across institutions and are interpreted by models that capture only pieces of the response.

Lovelock's frame is useful because it also motivates our role as stewards. Humans are not engineers standing outside a planetary machine. We are one component of a living system, newly capable of perceiving and deliberately influencing some of its responses.

Stewardship does not imply a hubris of complete dominion, nor a fantasy that nature is a machine whose every lever we will eventually control. And yet, it's becoming clear that far more of those levers may be available to us than we ever imagined. Our task, then, is to do on purpose what we've so far done only by accident or negligence.

Climate stability matters **because it is a physical precondition for stewardship.** Our ability to perform this role depends on keeping the planet within an envelope of conditions close enough to those in which human civilization—and the biosphere it rests on—can thrive. Cross enough fundamental tipping points, and the consequences may outrun our emerging ability to repair or reverse them.

What a nature model adds is a control loop: detect a change, estimate where it leads, intervene, observe whether it worked, repeat. Some of these loops we're already trying to manage. Some we know about but struggle to intervene in. And some we don't even know exist.

Loops we're already trying to manage. A good example here is *ecological biosecurity.* The New World screwworm demonstrates a control loop, which in fact we were handling quite capably… until we weren't. Surveillance detects the organism, models estimate where it may spread, sterile flies interrupt reproduction, and new observations reveal whether containment is working. USDA analysis estimates that a renewed outbreak could cost Texas livestock producers roughly $733 million a year and the wider state economy $1.8 billion annually. Invasive species more broadly impose more than $423 billion in annual global costs, in large part because they are detected too late. USGS is already combining eDNA with automated samplers; a nature model would elevate their capabilities by combining that with port traffic, currents, winds, climate, and reproductive biology to find an invader while eradication remains possible, then direct confirmatory sampling and compare containment strategies. The same principle could apply to crop fungi, mosquito-borne disease, and zoonotic spillover.

Loops we know about but struggle to intervene in. We often pay for pollution twice: first for fertilizer or chemicals that escape their intended use, and then for water treatment, damaged fisheries, and cleanup downstream. The EPA estimates that nitrogen and phosphorus pollution in U.S. freshwaters costs at least $2.4 billion annually. We know exactly where it comes from. But no model follows these materials reliably from crops and soils through groundwater, rivers, coastal dead zones, ocean chemistry, and the atmosphere. A nature model could compare fertilizer timing, wastewater treatment, phosphorus recovery, and wetland restoration within one connected material budget—showing which intervention prevents the most damage per dollar and whether an apparent solution merely moves the problem elsewhere. It could do the same for pesticides, pharmaceuticals, PFAS, and other chemicals moving through watersheds and food webs.

Watersheds are the same story. We can observe rainfall, rivers, and reservoirs reasonably well, but the true state of an aquifer remains largely hidden. A nature model could combine well pumping data and gravity measurements with snowpack, soil moisture, river flows, plant transpiration, and land deformation into a living estimate of an entire watershed. It could help us balance the parameters we can and can't control directly in order to provide enough water without risking saltwater intrusion or irreversible compaction. In shared systems like the Colorado, Nile, Mekong, and Indus, it could give water rights negotiations a common factual foundation, and motivate cloud seeding operations or agricultural restrictions in context.

El Niño belongs here too. It's the most famous example of a global climate phenomenon: a change in Pacific Ocean temperatures reorganizes weather around the world, altering patterns of rainfall and therefore drought, wildfire, and crop productivity. We've known this for decades, and mostly we just brace for it.

Loops we don't even know exist. El Niño is simply one connection we happen to have a name for. There are many other natural oscillations or correlated climatic zones, where a small change in one region has far-reaching effects. These are discussed further in principle 4 in *Designing a Nature Model*. Doing this right would enable us to turn an early signal in one part of the planet into a plan elsewhere: adjusting reservoir releases, planting different crops, moving fishing fleets, or adjusting surveillance for insect-borne disease ahead of time.

Today, many of these interventions are made by separate institutions using separate models, each with disparate and locally defined goals. With a nature model, we could determine whether they're all responding to the same planetary event, predict long-term causal effects, and coordinate efforts accordingly. And as was the case with language models, it might turn out that the most valuable use cases are ones we can't even imagine today.

How We Began to Understand the Earth

IN 1802, GERMAN NATURALIST ALEXANDER VON HUMBOLDT (yes, the same guy with all the stuff named after him) set off on an expedition up the volcano Chimborazo, the highest peak in what is now Ecuador, with a whole turn-of-the-nineteenth-century-lab's worth of measurement instruments sloshing chemicals like mercury on his back. From the tropical rainforest at the base to the frigid glaciers at the peak he took readings of altitude, humidity, temperature, and pressure, diligently cataloging and sketching the plants and animals he encountered along the way.

Nothing about the individual measurements was particularly remarkable, except for the fact that never before had the same person observed nature in detail from the temperate alps to the full vertical span of elevation at the equator. It turned out that the plants and animals that exist when you go up in elevation at the equator are pretty similar to what you see when you go much farther north. **The climate at high altitude mirrored that at high latitude, and the life which could be found there followed suit.** There was a structure in all of it that made sense, could be measured and inferred.

As soon as he got back home, Humboldt drew the *Naturgemälde*: "painting of nature." He showed Chimborazo in cross-section, with bands of plant and animal life mapped against altitude, temperature, pressure, and geology. It was the first documented attempt to study the Earth as one interconnected system, drawn in a single frame, built up from real data collected on the ground.

Alexander von Humboldt's *Naturgemälde* (1807): Ecuador's highest peak Chimborazo, in cross-section, with bands of plant life mapped against altitude, temperature, and pressure. The first attempt to see the Earth as one interconnected system, drawn in a single frame.

It is, in my opinion, one of the most significant artifacts of human understanding we've ever produced. Humboldt showed that life follows patterns across geography: similar conditions produce lookalike environments, even on opposite sides of the world. More than a century before the term "ecosystem" was coined, von Humboldt realized that the interplay of temperature, altitude, humidity, geology and other physical conditions shape the course of life. (His story is beautifully told in Andrea Wulf's *The Invention of Nature*.)

Humboldt's most important contribution, however, **is raising the aspirations of an entire lineage of scientists** who were driven to understand the structure of the natural world.

In 1831, an enthusiastic 22-year old named Charles Darwin found Humboldt's writings at Cambridge, anointed him *"the greatest scientific traveler who ever lived,"* and said that his work *"stirred up in me a burning zeal to add even the most humble contribution to the noble structure of Natural Science."* Soon after, Darwin set sail — to Ecuador, no less! — with trunkfuls of Humboldt's journals on board.

When he reached the Galapagos, Darwin observed that an animal's traits reflected their surroundings. Species were not preordained; they descended from common ancestors and changed as environments favored some inherited variations over others. His insight of evolution by natural selection made nature's structure legible across time, as Humboldt had across space.

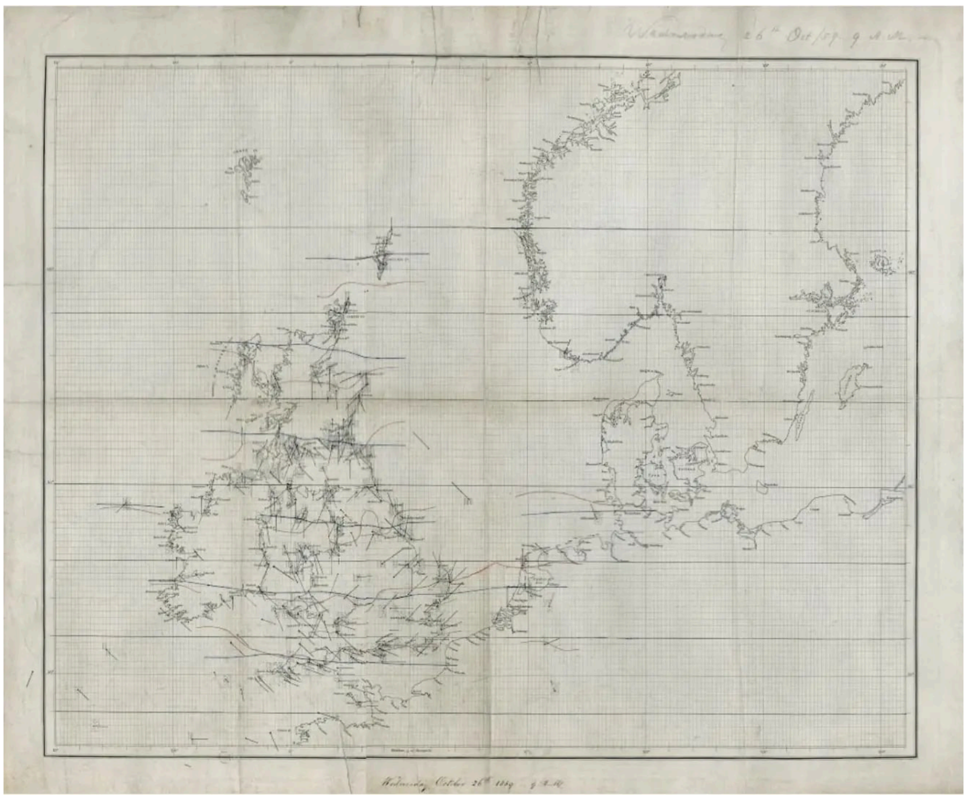
Next in line to be Humboldt-pilled was the captain of Darwin's *Beagle*, Robert FitzRoy. As a sailor, he was less concerned with plants and animals, and instead turned his eyes to the sky. After five years at sea, reading barometers and observing storms as a matter of navigational life and death, he began to wonder if there might be some as-yet-inscrutable structure in the atmospheric processes, too. There was indeed, but understanding that structure required simultaneous visibility into weather conditions on the ground over too large an area for a single observer to collect.

Click clack: now there's the telegraph. By the late 1850's, FitzRoy had wired a network of coastal stations to a central office in London. Each morning, the readings clattered in over the wire, were plotted by clerks onto one big map — and for the first time, the weather over an entire region was **viewed as a single connected system**. Modern weather models still use FitzRoy's term "synoptic", or *seen together*, to describe large-scale processes.

In 1859, a huge storm struck Britain, killing nearly a thousand people and destroying over a hundred ships. FitzRoy went back through his charts and found that the storm had signaled itself in the data days before it hit, and that anyone watching the whole map could have seen it coming in time to keep the ships safely in harbor.

This was a new kind of claim. Humboldt and Darwin had made nature's structure legible after the fact; FitzRoy proposed that we could model it forward, and then use it to inform our actions in advance.

The idea was so heretical at the time that FitzRoy was accused of false prophecy. Compelled to defend himself, he explained: *"Prophecies and predictions they are not. The term forecast is strictly applicable to such an opinion as is the result of scientific combination and calculation."* We still call them weather forecasts today, because apparently any other term would be too magical.



Robert FitzRoy's reconstruction of the Royal Charter storm, 1859. Drawn after the disaster from observations gathered across Britain, it demonstrated something radical: the storm had made itself visible in the data before it struck.

FitzRoy continued issuing forecasts under relentless criticism. In 1865, amid worsening depression, he died by suicide. After his death, an official inquiry shut public forecasting down. Storm warnings were restored only after tremendous protest from fishermen and sailors; public forecasts eventually followed, and FitzRoy is now recognized as the founder of the UK Met Office.

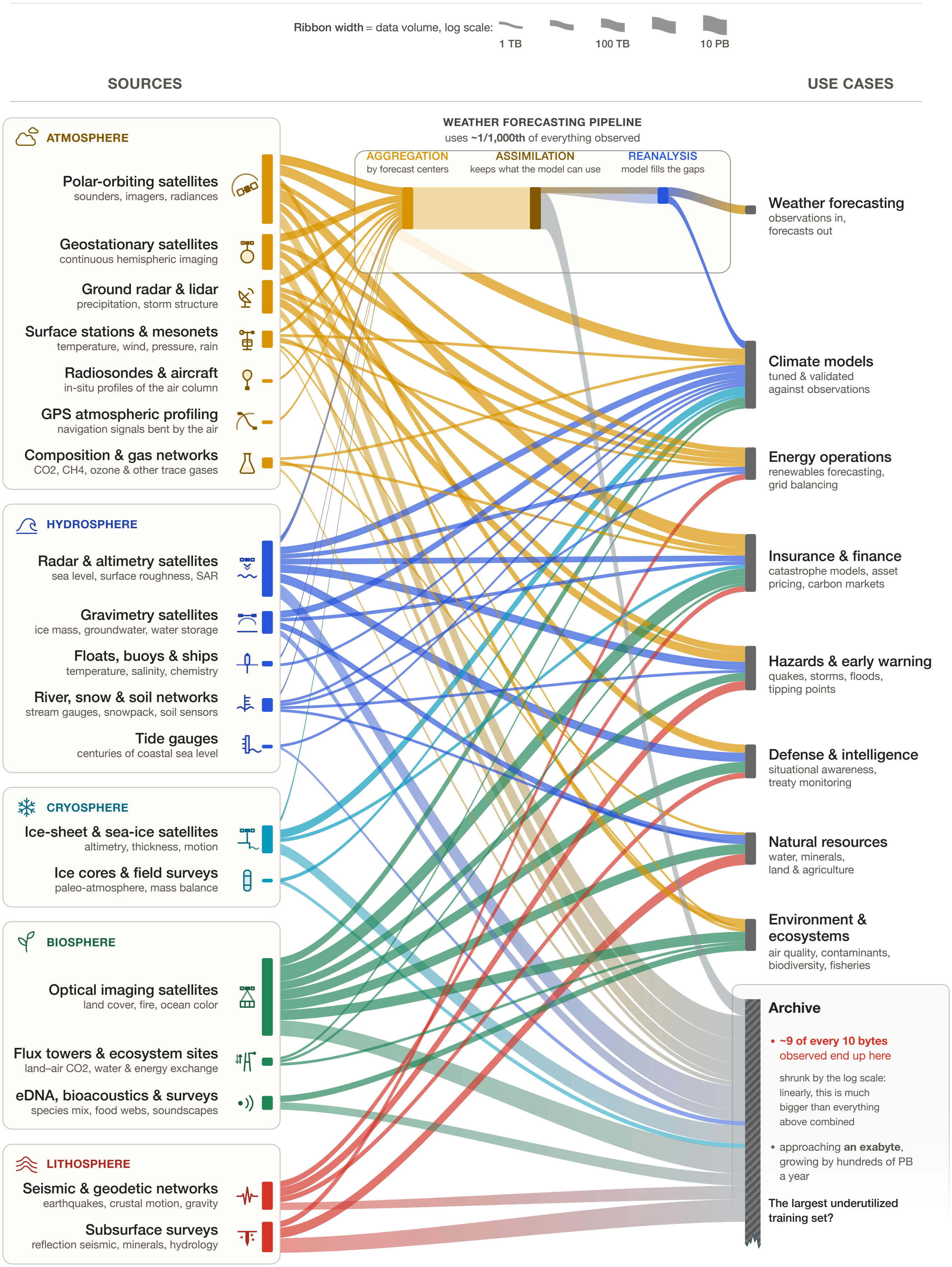
For everything we know about the natural world, there are still vast lacuna in our understanding of life on Earth. The concept of plate tectonics — which fundamentally governs the movement of the ground beneath our feet — only became widely accepted around the time of the Moon landing. Today we know how plates move, but we still don't know why our planet has them when other rocky planets in the solar system don't. Filling in our understanding of Earth is a constant effort. We have to scour the physical world, turning over actual stones and sequencing the DNA of rare species for every small crumb of knowledge.

While some of these discoveries have been pure accidents, the most tried and true way to advance the scientific frontier has been to broaden our field of view. Insights are unlocked by seeing a little more of the system at once: over more space, across longer stretches of time, widening the aperture with each advance. Today, we can observe the Earth in extraordinary detail, but when we perturb the Earth system, intentionally or otherwise, we struggle to disentangle cause and effect as they ripple across the planet.

Today's Observing System

SIGNALS CONTINUOUSLY ARISE FROM THE EARTH, even as the planet overwrites the state that produced them: the dynamics of an ocean current this season or the composition of a dust plume as it blows off the Sahara are lost forever if they aren't measured while they occur. Every missed observation is a page torn out, never to be read.

Hundreds of observing methods combined will form the foundation of a nature model's training data. I put together many of those which make up the bulk of data collected from the Earth today, shown below and interactive. A modern *Naturgemälde,* two centuries since the original.


Ribbon width = data volume, log scale:
1 TB
100 TB
10 PB
SOURCES
USE CASES
ATMOSPHERE
Polar-orbiting satellites
sounders, imagers, radiances
Geostationary satellites
continuous hemispheric imaging
Ground radar & lidar
precipitation, storm structure
Surface stations & mesonets
temperature, wind, pressure, rain
Radiosondes & aircraft
in-situ profiles of the air column
GPS atmospheric profiling
navigation signals bent by the air
Composition & gas networks
CO2, CH4, ozone & other trace gases
HYDROSPHERE
Radar & altimetry satellites
sea level, surface roughness, SAR
Gravimetry satellites
ice mass, groundwater, water storage
Floats, buoys & ships
temperature, salinity, chemistry
River, snow & soil networks
stream gauges, snowpack, soil sensors
Tide gauges
centuries of coastal sea level
CRYOSPHERE
Ice-sheet & sea-ice satellites
altimetry, thickness, motion
Ice cores & field surveys
paleo-atmosphere, mass balance
BIOSPHERE
Optical imaging satellites
land cover, fire, ocean color
Flux towers & ecosystem sites
land–air CO2, water & energy exchange
eDNA, bioacoustics & surveys
species mix, food webs, soundscapes
LITHOSPHERE
Seismic & geodetic networks
earthquakes, crustal motion, gravity
Subsurface surveys
reflection seismic, minerals, hydrology
WEATHER FORECASTING PIPELINE
uses ~1/1,000th of everything observed
AGGREGATION
by forecast centers
ASSIMILATION
keeps what the model can use
REANALYSIS
model fills the gaps
Weather forecasting
observations in,
forecasts out
Climate models
tuned & validated
against observations
Energy operations
renewables forecasting,
grid balancing
Insurance & finance
catastrophe models, asset
pricing, carbon markets
Hazards & early warning
quakes, storms, floods,
tipping points
Defense & intelligence
situational awareness,
treaty monitoring
Natural resources
water, minerals,
land & agriculture
Environment &
ecosystems
air quality, contaminants,
biodiversity, fisheries
Archive
• ~9 of every 10 bytes
observed end up here
shrunk by the log scale:
linearly, this is much
bigger than everything
above combined
• approaching an exabyte,
growing by hundreds of PB
a year
The largest underutilized
training set?

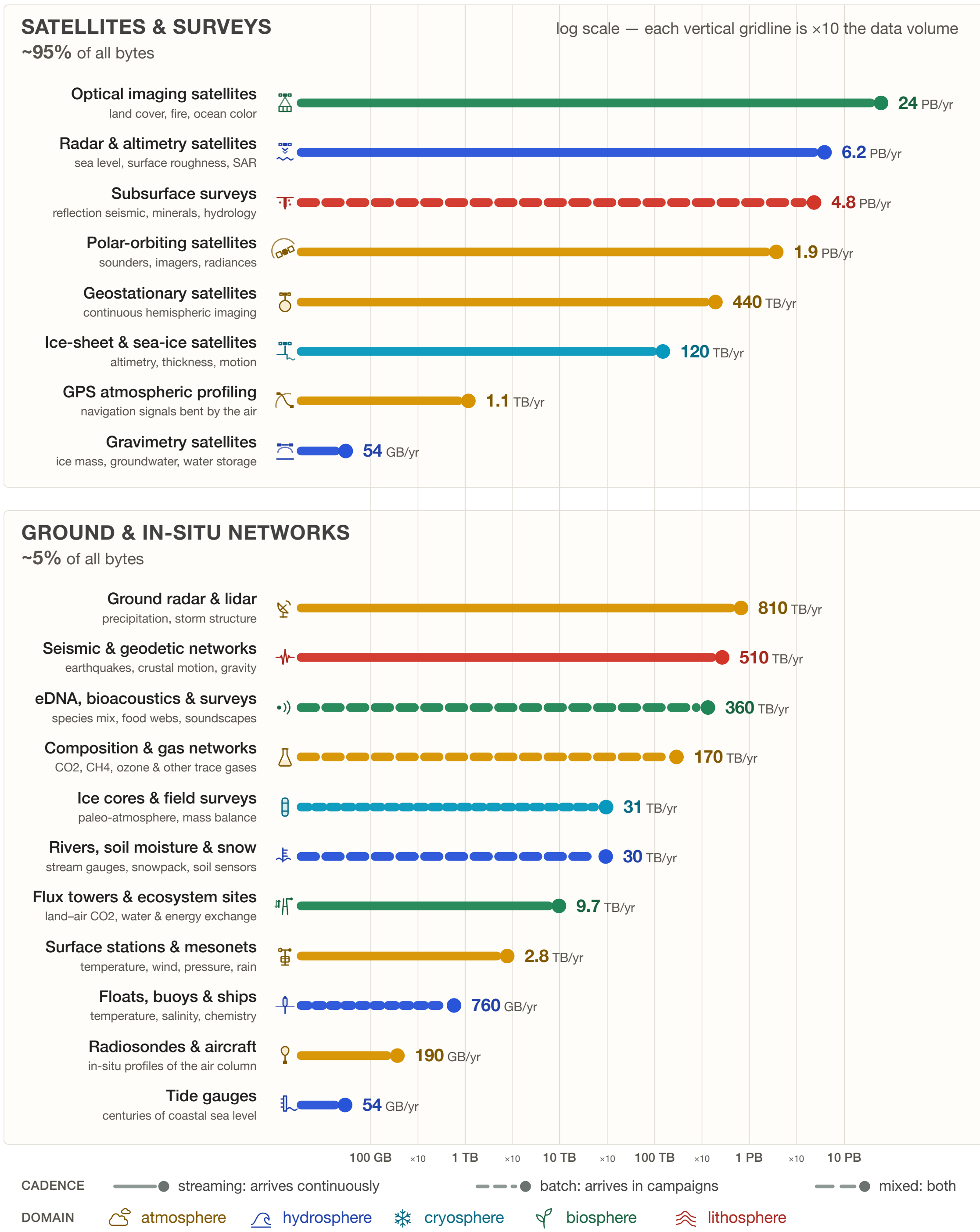
The system is simultaneously a scientific triumph and a total mess; some parts are modern, while others are almost comically vintage. It was not built from first principles, but is better understood as a many-decade cumulative shipping of the org chart. Not a single company's org chart, but a geopolitically entangled one comprising the United Nations, various countries meteorological offices, and an alphabet soup of national, international, and academic organizations.

Sifting through the mind-numbing acronyms makes clear that nearly everything in this system is paid for by governments as a public good. The flip side, of course, is that the system is more brittle than ever, with substantial parts at risk of being arbitrarily cut or, even if spared, remain extremely outdated with little to no modernization plan, let alone a holistic strategy.

While satellites are the largest data source, many necessary observations can't be made remotely. They require literally touching the physical processes, from the soil to the ice sheets. Some of these sensors stream live data over the internet while others are batch processed, requiring e.g. a ship to sail out to mid-ocean moored arrays and plug in to download the data every 6 months, or a scientific expedition to trek across the ice sheet and physically record the latest findings.

Combine these diverse systems and the scale is startling. The Earth-observing network produces an estimated **40 petabytes of data each year**. Much of this arrives through satellite downlinks and automated networks; other observations arrive through field campaigns or when scientists retrieve data from a remote ocean or ice sheet sensor. Below, I've assembled an overview of the sensing modalities and how much data each collects.

Today's Earth Data Supply



Expanding the Observing System

ON TOP OF THE LEGACY BACKBONE, the toolset for tracking the major activity across our planet is rapidly expanding and modernizing. It's driven by powerful tailwinds that hardly need to get listed out anymore. The cost to launch satellites, capture solar power, store energy in batteries, and sequence DNA are falling precipitously. Sensors are miniaturizing while becoming more power-efficient. Drones can operate autonomously and at long range not just in the air, but in the ocean, too. Starlink means we no longer need treacherous expeditions to physically recover data from faraway in-situ sensors.

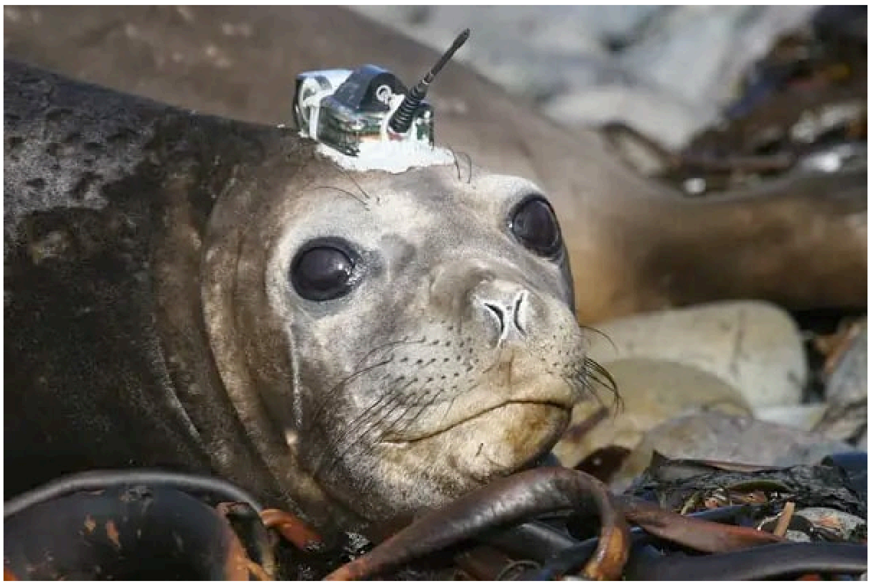
The upshot is that we can be far more ambitious about the type and volume of data about the planet that we seek out.

Extending Our Reach

The ocean remains unsurprisingly the largest gap in today's observational system. Satellites can see its surface, but observing the depths requires instruments physically in the water. The Argo program provides a backbone of roughly 4,000 robotic floats that dive as far down as 2,000 meters and transmit profiles of temperature and salinity. However, the deep ocean and polar regions remain extremely difficult to access.

A critical area is the Southern Ocean, which absorbs enormous amounts of heat and carbon and interacts directly with Antarctica's melting ice. Research ships rarely operate there in winter, while autonomous floats struggle beneath sea ice.

There is at least one way around this, and it points to the creativity that will be required to meaningfully expand our scope of data collection. Since 2004, the MEOP program (*Marine Mammals Exploring the Oceans Pole to Pole*) has coordinated an international consortium of scientists **gluing sensors to the heads of seals** (and to a lesser degree, whales and sea turtles). As a seal dives beneath the ice in search of food, the tag measures temperature, pressure, and conductivity, creating a vertical profile of ocean temperature and salinity at different depths.

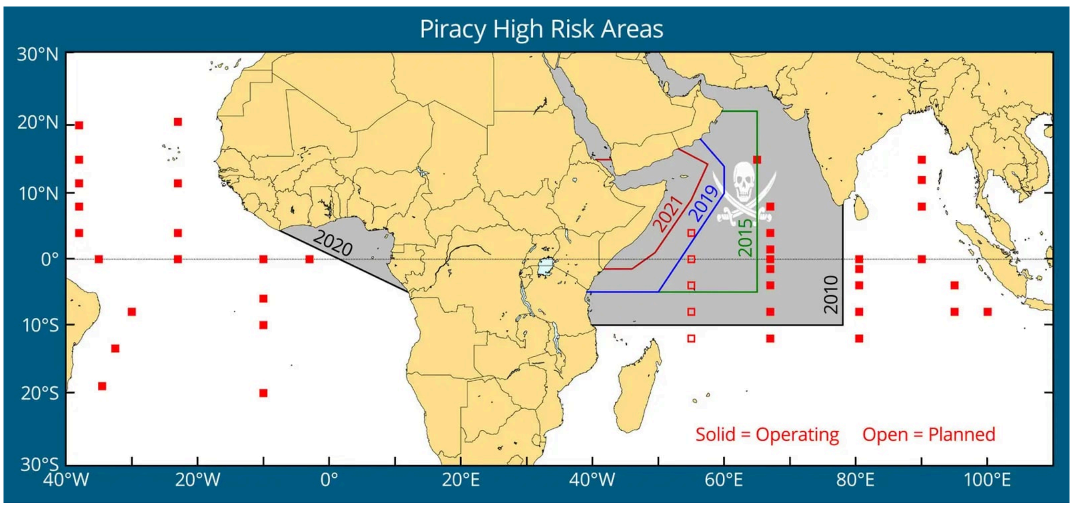


Yes this is 100% real!! It's crazy!! More pictures are here!!
Photo: Christophe Guinet

When a be-sensored seal surfaces, a compressed profile is transmitted by satellite and enters global ocean and weather forecasting systems in near-real time. They've so far recorded **800,000 vertical profiles**, filling data gaps that ships, satellites, and floats physically cannot. These little guys are both a sensor platform and part of the system being sensed: the measurements describe the ocean, while their movements reveal how life is responding.

Not all sensing programs work as smoothly as the seals. A particularly remote stretch of the Indian Ocean is home to the Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction (RAMA), one of a handful of moored arrays collecting key data on ocean circulation that underpins our ability to forecast the monsoon. Like many parts of the patchwork observing system, these arrays are exceedingly old-school: they transmit no data, but instead rely on research ships sailing up to them, docking, and downloading it directly. This is supposed to happen a couple of times a year, but RAMA's operations are regularly derailed by... **Somali pirates**.

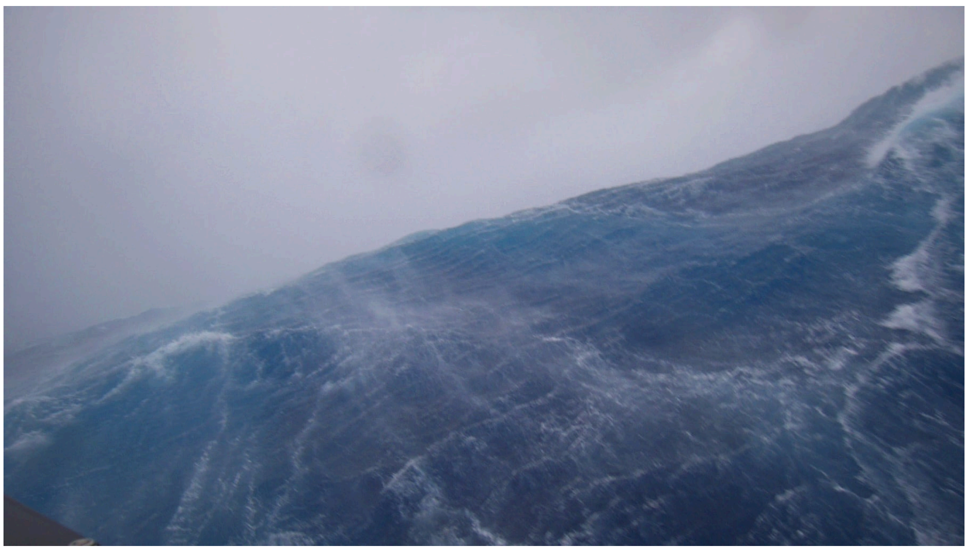
The buildout of the RAMA network faced a variety of challenges, from fishing vandalism (moored buoys attract schools of fish and therefore fishermen) to literal piracy (installations were halted or scaled back multiple times as research vessels couldn't transit the area). Look at this crazy figure from the [official magazine of the Oceanography Society]().



RAMA is the red squares in the western Indian ocean (and you can see some of the other arrays elsewhere too).

If it wasn't clear by now, Earth data collection is a lot more complicated than scraping the internet.

What we need are more platforms that can fill these gaps by moving toward events as they unfold. Saildrone builds surface vehicles that can remain at sea in harsh conditions for months while measuring the ocean, atmosphere, and the mixing between them. In 2021, NOAA directed one into Hurricane Sam, where it captured the world's first ocean surface video from inside a major hurricane while measuring winds above 120 miles per hour and waves as high as 90 feet; a feat impossible with traditional manned ships.



The view from Saildrone 1045 inside Hurricane Sam. Image: NOAA/Saildrone.

Filling In the Sky

Modern meteorology depends on a public, international patchwork of ground stations, radar, aircraft measurements, satellites, and balloons. These represent a substantial portion of the data volume in the observing system, but their coverage remains extremely uneven. The basic challenge is geometry. The atmosphere is enormous and three-dimensional, and weather systems move through the entire volume: an observational gap over an ocean or remote landmass can degrade forecasts far downstream.

Long-duration balloons offer a new way to sample that space. WindBorne builds and operates balloons that can change altitude and use wind patterns to stay afloat for weeks, producing repeated vertical profiles along wide trajectories instead of the traditional single ascent. You can think of them like seals for the sky.

A long-duration balloon capable of changing altitude and collecting repeated atmospheric profiles. Image: WindBorne.

Ground-based weather radar produces high-resolution observations of precipitation, but each station can only see immediate surroundings. The curvature of the Earth limits their range, and mountains and other terrain can block the radar beam. Stations are expensive to build and maintain so global coverage is quite spotty, particularly in lower-income countries.

Tomorrow.io is trying to fill part of that gap from orbit. It is developing precipitation-radar satellites and microwave sounders designed to observe temperature, water vapor, and precipitation across large areas poorly covered by ground infrastructure. They're a good example of how these tailwinds combine: cheaper sensing hardware itself helps, but only somewhat. But combine that with rapidly falling launch costs, and new systems are possible.

Understanding Ecosystems

The biosphere is its own beast. Figuring out what's actually going on requires understanding the complex dynamics between plants, animals, microbes, and fungi, many too small to see or too rare to keep tabs on. The current system is a mix of satellites and human reports.

For human reporting, The Audubon Society invented the concept of a large-scale citizen-science network in 1900, when it replaced a Christmas bird hunt with a Christmas Bird Count, but citizen science has come a long way from those handwritten journals. Modern platforms like eBird and Adventure Scientists turn millions of people into a distributed observing network collaboratively producing a living record of where species occur and how their ranges and seasons are changing.

We also have this:

Good luck getting one of these by labeling stuff for Mercor.

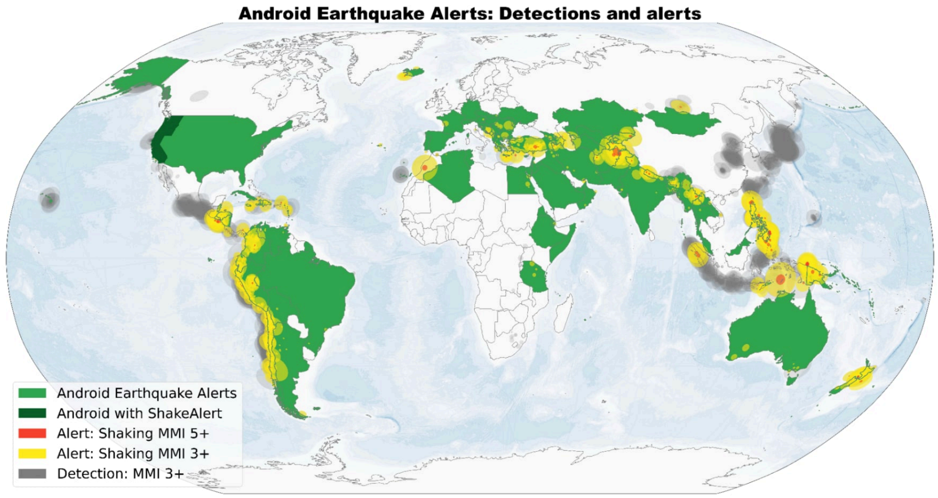
For many animals, hearing them is much easier than snapping a photo. Cheap microphones can listen continuously, but they also create far more audio than people can review. HARK builds solar-powered listening stations that run bioacoustic AI on the device, identifying species in real time.

Sound only travels so far, though, and lots of organisms are too quiet or too small to be heard with bioacoustics. Environmental DNA, “eDNA”, goes further: organisms leave genetic traces in water, soil, and air, samples can be sequenced to index species that were never seen nor heard. NatureMetrics is scaling up eDNA sequencing, leveraging the exponential drop in sequencing cost to turn this method from a niche technique to a viable survey mechanism. And while cataloging more species is the baseline we’ll need, it only tells you what's there, not whether the system is healthy or about to tip. That’s a harder problem Echo Labs is chasing: turning signals like these into a continuous measure of ecosystem condition so ecology can become less about autopsies and more about stewardship.

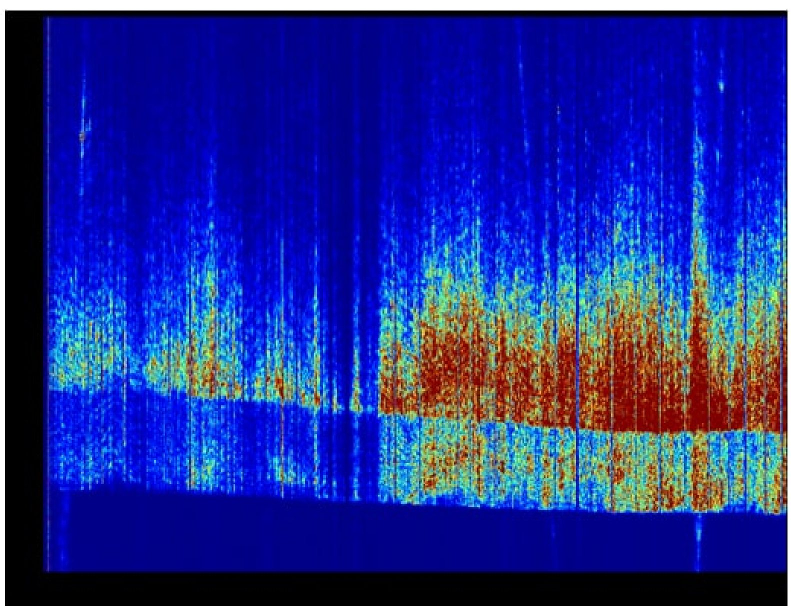
Repurposing Existing Infrastructure

Permanent seismic stations are another fixture of the existing observing system, but their limitations rhyme with those of the weather radar stations: high cost to build and maintain, so they remain unevenly distributed around the globe. That sparsity costs lives. An earthquake alert can only be sent after an earthquake has begun, in the brief interval between detecting its faster initial waves and the arrival of more damaging shaking. If the observing system could detect it fast enough, it could warn people in its path to take cover.

Google realized that smartphones have become so widely distributed that the accelerometers in each phone can provide the foundation for a very dense seismic monitoring system. Now, billions of Android phones act as a sensing network. A single phone's motion is so noisy that it's useless, but many nearby phones detect the same motion at nearly the same time: the system can confirm an earthquake, estimate its location and magnitude, and alert phones farther away before the strongest shaking arrives. It's a perfect example of how all the tailwinds in cheap sensors, batteries, and connectivity combine to make a previously unimaginable system, essentially as a by-product.



Smartphones aren't the only emergent sensing network. The fiber-optic cables powering the internet have helped give rise to the new field of Distributed Acoustic Sensing. Here, instruments send laser pulses down existing fiber-optic cable and measure tiny phase shifts in the light scattered back as the glass stretches and contracts–literally a magnifying glass for tiny subsurface changes. Every few meters of fiber represents a distinct sensor; just a few kilometers of cable is enough to provide high resolution.



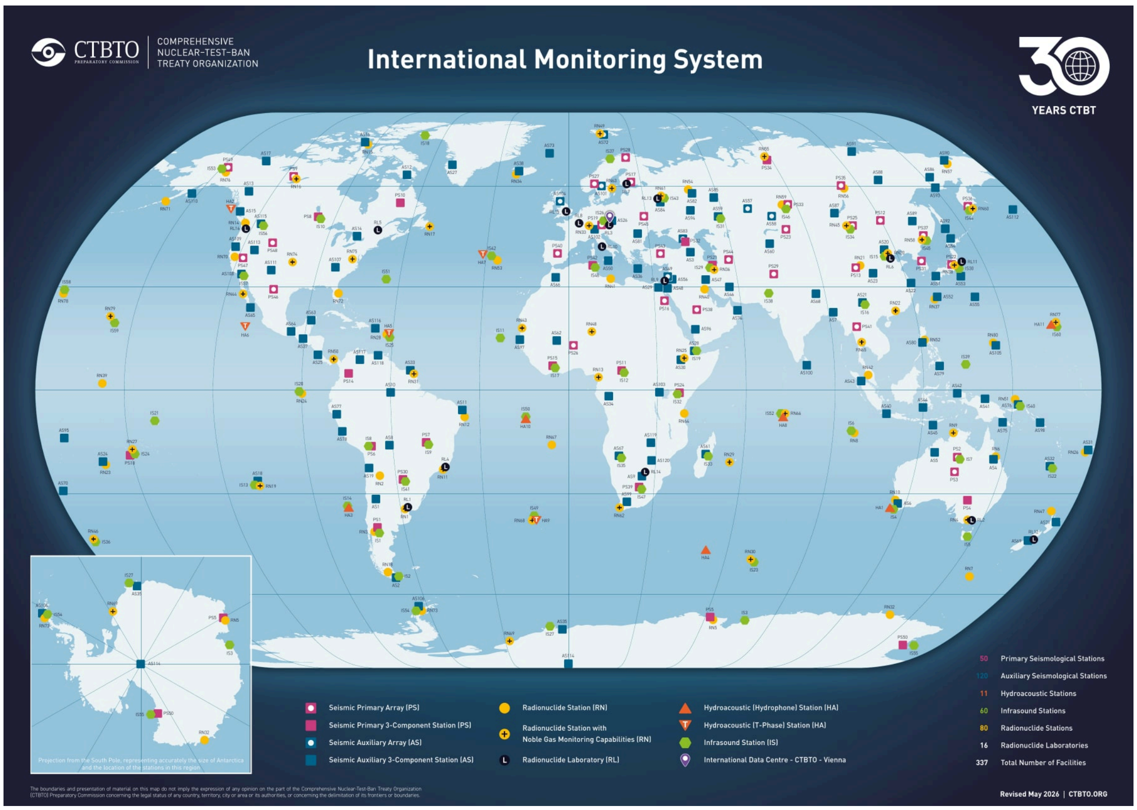
A magnitude 2.9 earthquake, seen by 15 kilometers of buried fiber. Across the length of the cable, the colors show how the ground stretched and compressed as seismic waves passed through. Image: USGS.

Fiber cables can also watch magma move. Since 2023, Iceland's Reykjanes Peninsula has experienced nine eruptions of lava, some large enough to threaten entire towns. Caltech researchers and Icelandic scientists worked with a local telecom company to connect a laser instrument to 100 kilometers of unused fiber lying beneath the surface. It took just 10 days to get the system up and running. As magma pushed through cracks underground, the cable measured millimeter-scale stretching and compression in real time, producing warnings between 30 minutes and several hours before eruptions — enough to evacuate safely.

As with Google's cell phone seismology, none of these cables were laid with earthquakes or volcanoes in mind. **We simply realized that the cheapest way to sense more of the planet is to teach infrastructure already paid for and deployed to notice something new.** There's an underexplored opportunity for much more of this across domains.

One final example, to illustrate both the scale of the sensing networks and the breadth of institutions governing them. The CTBTO was created in the wake of the Comprehensive Nuclear-Test-Ban Treaty, which since 1996 prohibits nuclear explosions. To manage compliance with the treaty, they built the International Monitoring System, with 337 sensing facilities spread across 89 countries.

Seismic stations feel vibrations, hydroacoustic stations listen through the oceans, infrasound stations detect pressure waves below the range of human hearing, and radionuclide stations sample the atmosphere for radioactive particles and gases. Their measurements flow continuously to an international data center in Vienna, where algorithms and human analysts distinguish possible explosions from the planet's constant background activity.

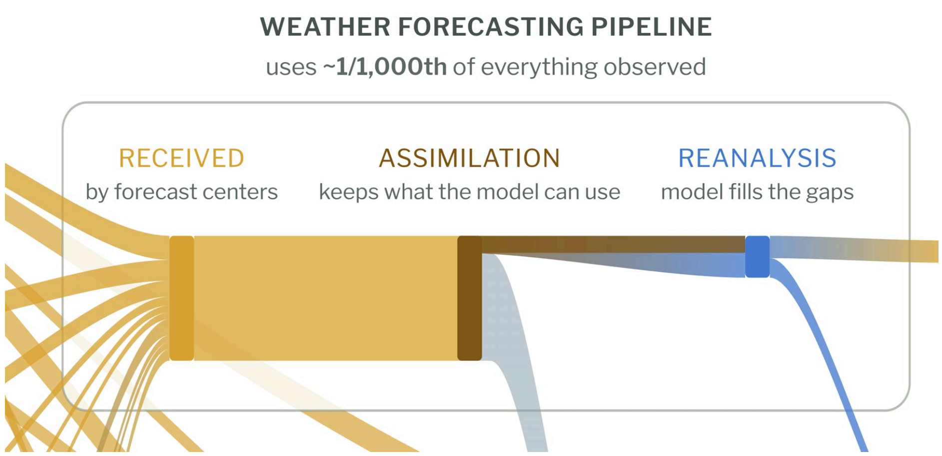


As a secondary capability, the network records far more than (hopefully the absence of) nuclear tests. It supports tsunami warnings and hears volcanic eruptions, the crackling of icebergs, and bioacoustics via hydrophone detects whale songs (it literally discovered a new whale pod).

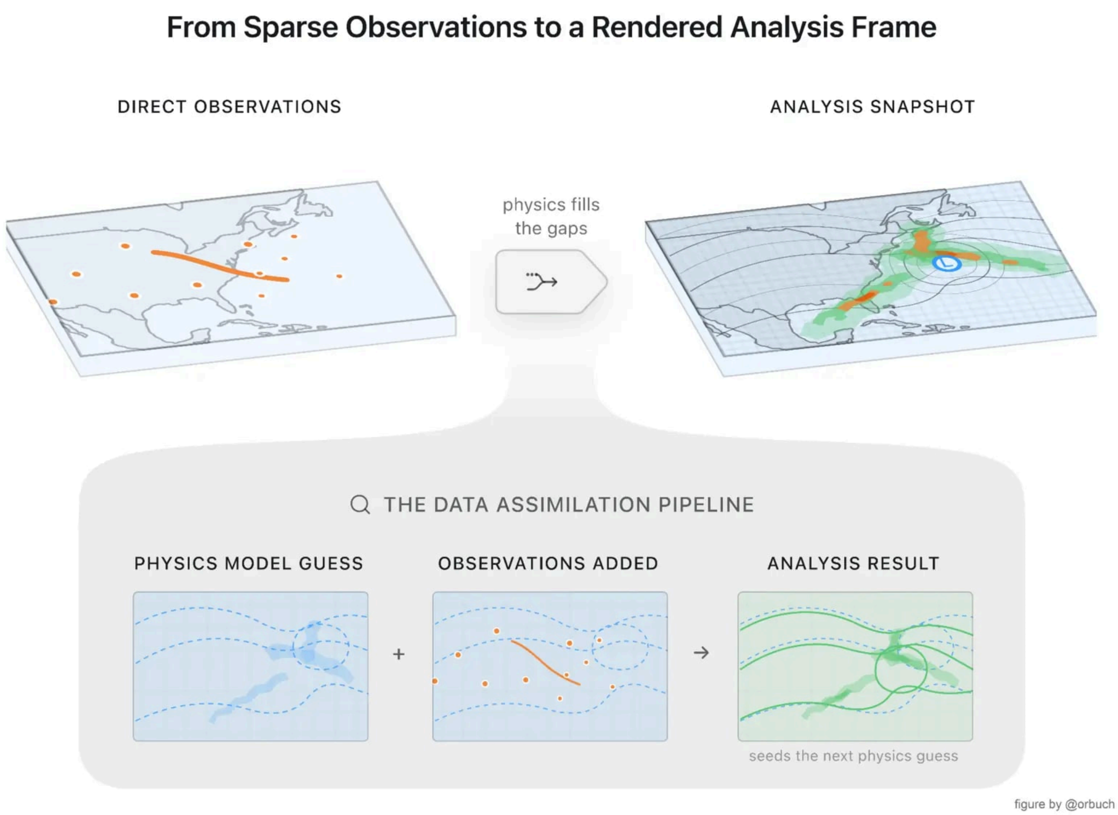
Where All This Data Goes

BY FAR THE LARGEST CONSUMER OF ALL THIS DATA is the weather forecasting complex. Weather is much narrower in scope than the nature model I'm proposing, but it's the most mature corner of Earth modeling and the first place AI has gone head-to-head with the physics-based incumbent approaches. So, it's worth understanding what happens to an observation after it's collected.

Remember this guy from the big sankey?



What you see when you zoom in is that it's really just FitzRoy's system at modern industrial scale. Observations still arrive from points around the world to be assembled into a single picture of the atmosphere. The difference is that his clerks plotting them onto the map have been replaced by a physics model that runs as a loop:



1. **Start from the last analysis and run the physics forward.** The cycle starts with a short forecast, six to twelve hours long, launched from the previous analysis. This is the background: a complete, physically consistent guess for the whole atmosphere, every grid cell, including the middle of the Pacific and the gaps between satellite passes. It is not filling holes between measurements. It is the starting point for everything, measured or not.

2. **Collect everything that arrived in the window.** While that forecast runs, observations pour in from satellites, weather balloons, aircraft, ships, buoys, and ground stations, each stamped with where and when it was taken. Everything that lands inside the assimilation window (twelve hours at ECMWF) is gathered up, quality-controlled, and, for most satellite data, thinned. Hundreds of millions of readings a day still leave most of the atmosphere unmeasured at any given moment.

3. **Ask what each instrument should have seen.** The system does not compare the model's temperatures to the satellite's temperatures, because satellites don't measure temperature. Instead, an observation operator takes the

background and computes what each instrument would have recorded if the background were true, down to simulating the infrared and microwave radiances a satellite would see. The gap between what was observed and what the background predicted is called the innovation, and it is the only thing the observations contribute. Hold onto that idea; it comes back when we get to models that skip the grid entirely.

4. **Correct the whole state, everywhere, weighted by trust.** This is **data assimilation** proper. The system finds the state that best fits both the background and the observations, given the uncertainty in each. More reliable instruments get more weight, and so does the background where the model has a good track record. Each innovation nudges not only the grid cell it landed in but its surroundings, other altitudes, and related variables, according to what the system knows about how the atmosphere fits together. Measurements never simply overwrite the model, even where they exist; the result is a weighted blend everywhere.

5. **Save the analysis, launch the forecast, repeat.** The result is an **analysis**: one complete picture of the atmosphere, made of neither pure measurement nor pure simulation. Some parts are pinned down by observations; others lean heavily on the model's reconstruction of what must lie between them. Physics helps connect the dots, but it cannot provide independent evidence about something no instrument ever detected. The analysis initializes the forecast you see and, in the same run, the short one that becomes the next cycle's background. Every six to twelve hours, forever. That loop is what keeps modern weather prediction tethered to the actual sky.

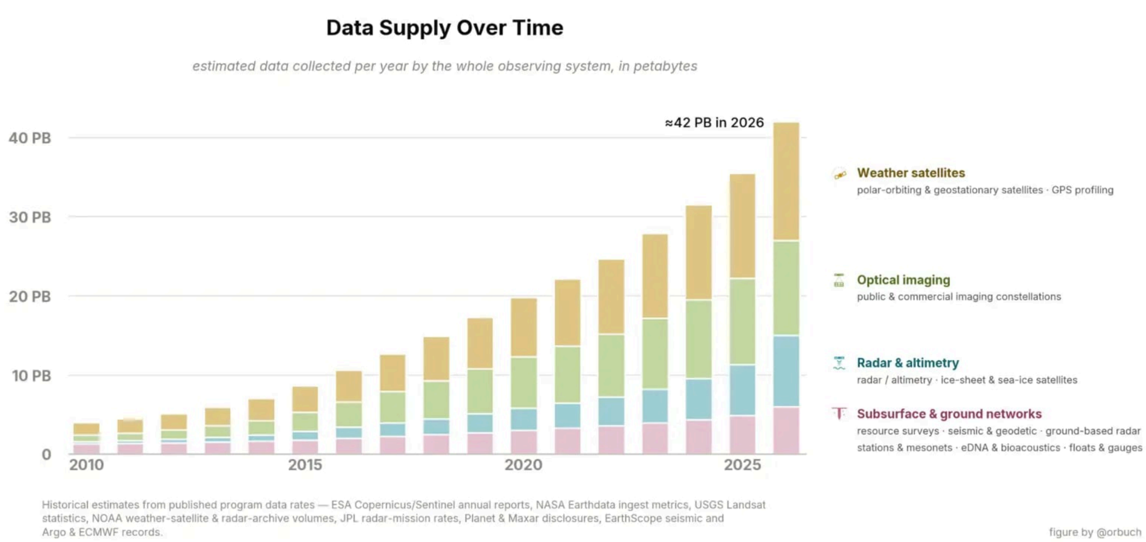
This loop is what produces the ten-day forecast that populates when you pull up the Weather app on your phone.

It is also what drives reanalysis, which aims the same machinery at past data. Hold the physics model and assimilation system constant, feed in decades of archived observations, and march forward from 1940. The result is ERA5: an hourly reconstruction of the atmosphere running to the present. It looks seamless, but it is not equally certain in every frame. Where and when observations are thin, the physics model does more of the heavy lifting. In the final output, all of this is mashed together to create essentially a pre-rendered movie of past atmospheric state.

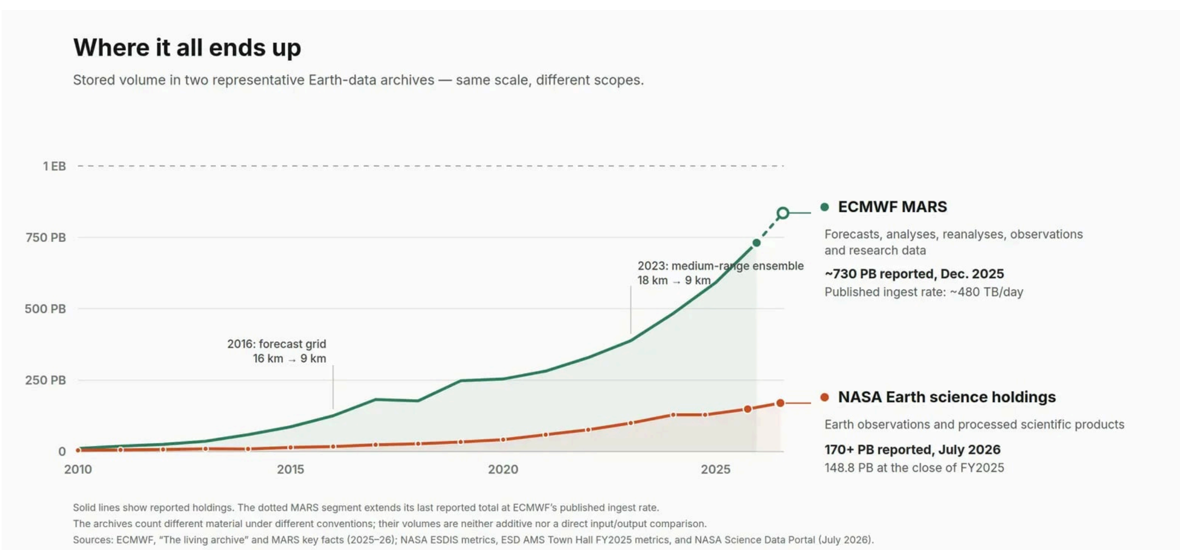
This type of reanalysis is one of the most heavily used datasets in all of Earth science, and we'll come back to it, and its fundamental limitations, in the *Designing a Nature Model* section later.

What should stand out to AI engineers is how little data goes into generating these outputs. I don't just mean compared with what a frontier LLM trains on. The pipeline that produces it is the single largest consumer of Earth observations we have, and yet it ingests only around a thousandth of what the constellation of Earth observation sensors collects.

In fact, **well over 99% of all the data that's generated isn't really used anywhere** today. In totality it recently **surpassed an exabyte** across raw observations, reanalysis runs, model outputs, and more. Where does all that stuff actually end up?



Two main places. The weather forecasting pipeline, highlighted above, dumps all output to something called the MARS archive, stored permanently on, you guessed it, physical tape, handled by robots, in a datacenter in Italy. Data volumes are growing fast:



Other data that isn't picked up by the weather forecasting pipeline mostly lands in the EOSDIS archive, a formerly distributed system across a dozen datacenters from NASA to NOAA to JPL and more. Over the last few years, this archive has been going through a massive migration of 100+ petabytes from tape to cloud, scheduled to complete at the end of this year, after which the data will be all comfy cozy in an AWS data center in us-west-2.

The existing system is pretty extraordinary in its breadth and the collaboration required to maintain it, but taken together it is still closer to a collection of sensory organs than a true central nervous system.

Designing a Nature Model

IN THE PRECEDING SECTIONS, we went into how traditional Earth system modeling is treated like a simulation problem: chop up the planet into a grid, specify the initial conditions, roll the grid cells forward in time based on physics equations, and see what happens. A nature model like what I'm proposing would still include plenty of simulation, but the more useful way to think about training one is as a task of **distillation**.

In machine learning, knowledge distillation uses a larger, more expensive "teacher" model to train a smaller "student" model that is faster and cheaper to run. Typically, the teacher's forward pass is expensive, but its outputs (and often its internal activations) are directly accessible for the student to train on.

In our case, the teacher already exists and has been running continuously for billions of years: the Earth itself. The nature model is the student. The analogy is imperfect in one important way, which we'll get to: the Earth doesn't take queries. But it flips the usual economics. **The Earth's forward pass is free, but sampling the internal state is costly because samples are physical observations of the real world.**

Beyond the sampling cost, another key difference is that the Earth follows one realized, irreversible trajectory through the enormous space of physically possible states. It never perfectly repeats itself.

We cannot pause it, checkpoint it, fork it into counterfactual timelines, or rerun the twentieth century with one variable changed. If we miss an observation, the physical state that would have produced it is overwritten and forever lost. The challenge arises because Earth observations are **sparse**. Sensors sample particular pieces of the Earth's state at different places, times, and resolutions, but never the whole state all at once.

What does that mean for the system design? While not meant to be a detailed technical proposal, I'll outline five core principles that follow from the analogy and can shape our approach. I am not the first to suggest that these are good rules of the road. However, most Earth modeling approaches today only really follow one or two of these constraints, partially, if at all.

A nature model should:

01 Learn Directly from Earth Observations

02 Keep the Full Resolution of Each Observation

03 Obey the Physics We Already Know

04 Discover Couplings across the Earth System

05 Fetch New Observations

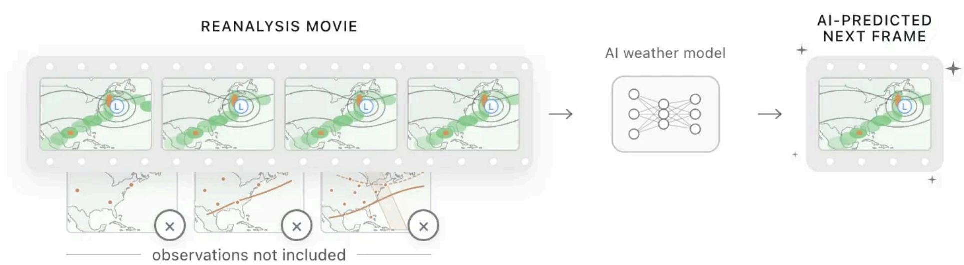
Taken together, these let the model learn as much as possible from limited (albeit growing) observations without ever losing track of what was measured, what was inferred, and what's still unknown. That's why distillation here isn't a one-time training step but an ongoing loop between the model and the planet. Let's dive into these principles one at a time.

PRINCIPLE 1

Learn Directly from Earth Observations

If Earth is the teacher, the student model should learn **directly from the Earth.** While this might sound like a truism, this is not how most AI weather models are actually trained today.

Recall the reanalysis pipeline and generated movie described back in the *Where All This Data Goes* section. You may be surprised to learn (I was!) that the models in NVIDIA Earth-2, Microsoft Aurora, and Google WeatherNext **don't primarily learn from direct Earth observations**. They **train on the reanalysis movie** (reconstructions produced by an upstream physics simulation) and try to predict the next frame. The frames they train on don't contain the underlying observations, and they generally don't preserve the balance of measurement and inference that produced any given pixel.



Whether an important detail was missing because it was too small to observe or because reanalysis smoothed it away, these models can't tell the difference. Headline scores won't tell you the difference either. **It turns out small disturbances matter a lot.** *Can Artificial Intelligence-Based Weather Prediction Models Simulate the Butterfly Effect?* found that AI weather models could produce seemingly good forecasts while learning the wrong underlying dynamics, failing to reproduce the rapid growth of the tiny disturbances that limit atmospheric predictability. *Follow-up work* found similar problems across several models. This isn't to say these models are useless! They're faster, cheaper, and better than the physics baselines. But that makes them an optimization layer bolted onto the existing forecasting system, not a foundation for what we actually need.

A step in the right direction is *AIFS-DOP*, released by the European Centre for Medium-Range Weather Forecasts (ECMWF) in June. They trained it on **forty years of direct (though extremely sparse) observations**, not reanalysis, using 16 H100s for four days. It learned to forecast the atmosphere straight from observations, with no analysis step in between, becoming competitive with the operational physics forecast on several headline scores out to ten days when scored against observations. It's a very cool result that points to how we might skip the pre-rendered movie.

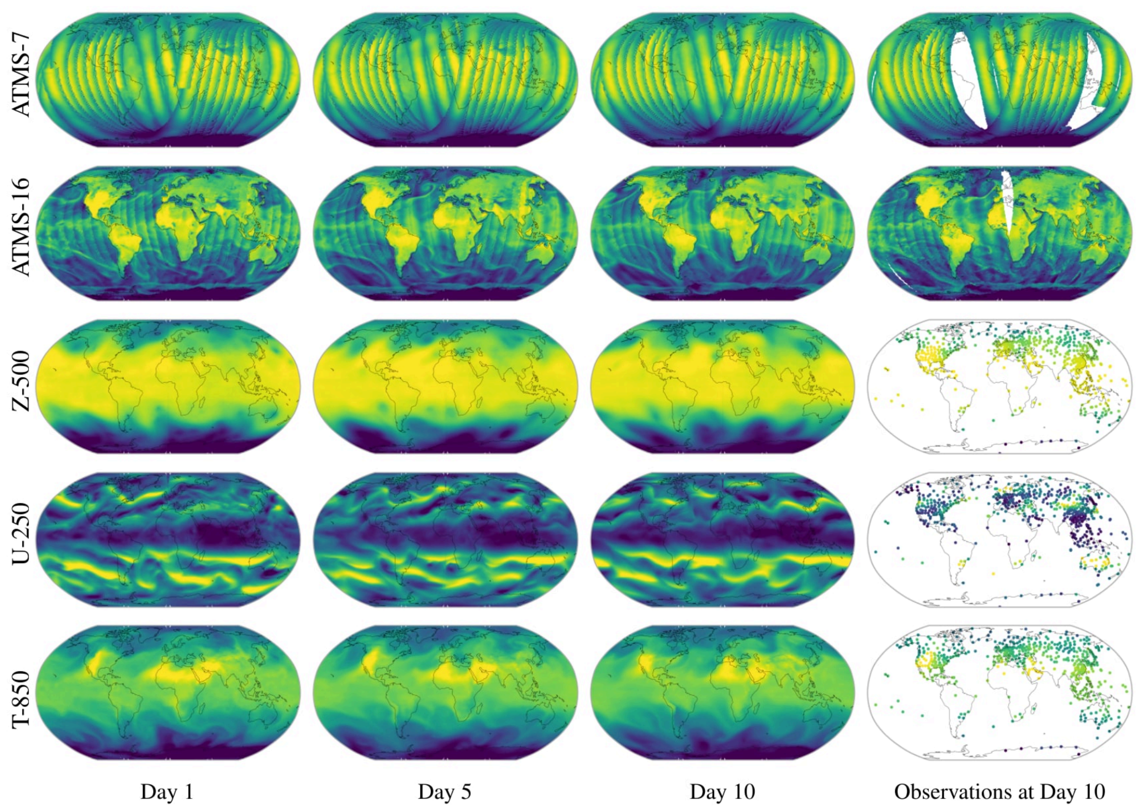


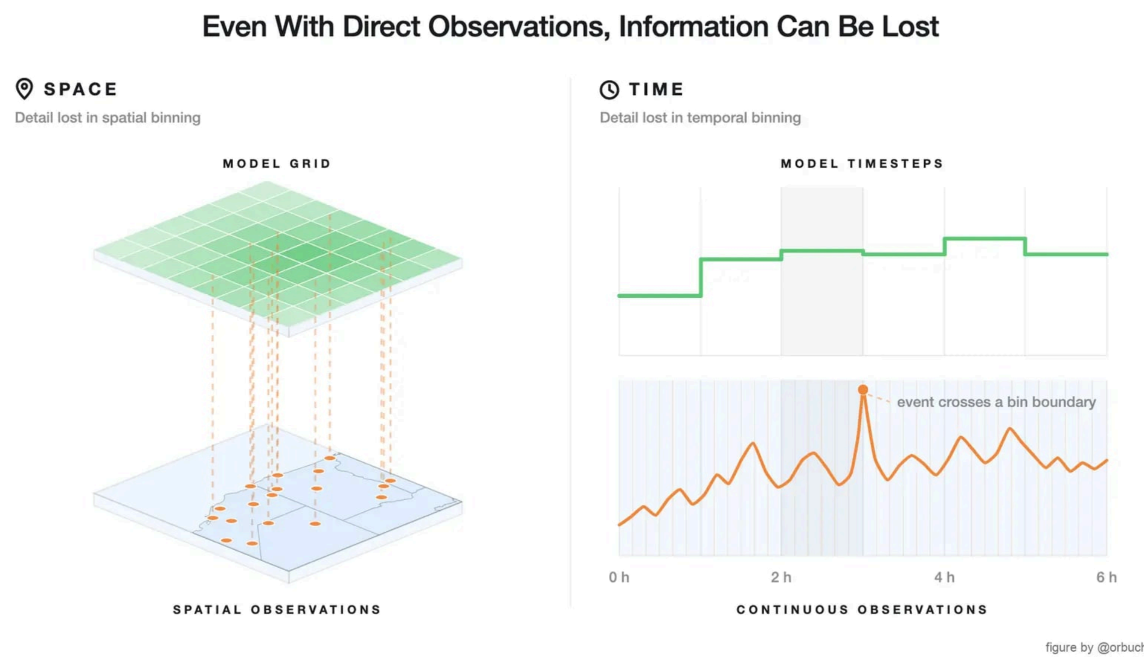
Figure 3: AIFS-DOP predictions at different forecast lead times compared to observations. Forecasts initialised on June 6th, 2021 00 UTC, observations valid at June 16th 00 UTC.

The goal with all of this is to internalize data assimilation, not to get rid of it. A nature model has to learn from the observations directly, because that's the only way to learn the explanatory value of each observation.

PRINCIPLE 2

Keep the Full Resolution of Each Observation

It isn't just enough to capture direct observations, however. We need to preserve as much fidelity as possible.



AIFS-DOP, while a promising direction, doesn't go far enough on this principle. Its observations were gridded upstream into six-hour windows and roughly 100 km cells, losing much of their original resolution in space and time, detail that could turn out to matter a great deal. A recent paper called *Earth-01* describes a model which gets closer: it learns a continuous three-dimensional atmospheric field directly from ungridded observations, so each measurement stays attached to exactly where and when it was made.

There are a handful of architectures designed to represent real-world processes continuously. Neural operators learn mappings between functions rather than between arrays, so in principle they can be queried at any resolution; the Fourier Neural Operator and its spherical variant are the best-known examples in weather. Graph-based models like GraphCast internally relax the grid to an arbitrary mesh, which is a step toward irregular inputs but still not the raw observation itself.

Neural processes go further and treat the data as an unordered set of (location, time, value) points, which is closer to what an observing system actually produces and is the family Aardvark drew on for its end-to-end forecasts.

This principle is about the pipeline, not the architecture. Any averaging, gridding, or time-windowing applied before the model sees the data is information lost for good, and no amount of downstream cleverness gets it back (information theorists call this the data-processing inequality). Which architecture best consumes raw point observations is an open question. The families above are candidates, but standard transformers have a habit of absorbing the advantages of bespoke architectures once given enough data, and the right answer may turn out to be a fairly ordinary backbone.

But we'd know it's right if it can actually utilize the full information content of the observations: does it get better if you hand it finer-grained observations or does it stop improving once the observations are finer than the model's internal resolution? How much better can it do against gridded representations on the same historical data without the grid?

PRINCIPLE 3

Obey the Physics We Already Know

In ordinary distillation the student gets to pose the teacher any arbitrary query and see how it responds. The Earth doesn't take queries. It runs one input, the trajectory that actually happened, and it runs it a single time. For any question of the form *what if the Gulf had been two degrees warmer that August*, there's simply no teacher output to learn from. Physics helps in two ways.

First, it provides constraints. Letting the model learn the couplings doesn't mean letting it invent whatever dynamics fit the training record. We're not building a language model or an image model; we're trying to depict the real Earth, which obeys real physics. **Some things we know for sure**: energy doesn't appear from nowhere. Carbon and water have to go somewhere. Matter and momentum are conserved. Fluids obey equations of motion. The model can learn how Earth's processes fit together, but it must be constrained to the same rules governing them, no matter how nicely an apparent shortcut happens to explain the data. Otherwise it is just a sign that we don't have enough data.

Second, as a substitute teacher. We can't rerun a hurricane over several alternative ocean temperatures or repeat the twentieth century with one variable changed. A numerical simulation will happily do both. The catch is provenance. What's wrong with training on reanalysis isn't that simulation is involved, it's that simulated and observed content get blended so the model can't tell which is which, and therefore can't know what each observation is worth. Simulated data is fine so long as it is labeled as simulated. Weather agencies already use simulations this way to estimate what a hypothetical new instrument would be worth before anyone builds it (they call these observing system simulation experiments), which is exactly the question principle 5 asks.

The tradeoff shows up most clearly in rare events. In Physics-based models outperform AI weather forecasts of record-breaking extremes, researchers found that GraphCast, Pangu-Weather, and FuXi all made worse forecasts as events got more extreme (and therefore, definitionally, further outside the historical distribution), while ECMWF's physics-based models performed comparatively better. The study looked at record heat, cold, and wind, and the AI models' errors grew almost linearly with how far an event exceeded the previous record, while the physics model's did not. This makes sense: a record-breaking heatwave is outside anything in the training data, but it still obeys the same physics as every heatwave before it.

There are already "hybrid models" working to strike this balance. NeuralGCM (trained on reanalysis) uses a physics solver for the large-scale flow and learns the parameterizations for what the grid can't resolve, like clouds and precipitation. CliMA and LEAP are the academic efforts closest in breadth of vision to what I'm describing, via different routes.

In each case physics supplies the skeleton and learning fills in the rest. It helps to be precise about which physics is which. Conservation laws (energy, mass, momentum) are exact, and no amount of data should ever relax them. Parameterizations, the empirical approximations for processes a model doesn't resolve, like clouds, turbulence, or soil microbes, are the human-authored, approximate part, and that is what learning should take over as observations get denser. Where observations are sparse, physics of both kinds stops the model from extrapolating arbitrarily. The long-term balance between these two can be reasoned about as a scaling law but remains an open research question.

PRINCIPLE 4

Discover Couplings across the Earth System

Traditional Earth-system models are assembled from modules. Researchers decide in advance which variables and processes to represent, divide the planet into a grid, write equations for each domain, and then "couple" separate models of the atmosphere, oceans, land, ice, and biosphere by passing predetermined variables across their boundaries. At the ocean surface, for example, the atmosphere supplies wind, rain, and sunlight, while the ocean's surface temperature helps determine how much heat and moisture return to the atmosphere.

This is the prevailing approach because it makes planetary simulation computationally manageable, lets researchers tailor each component to its particular physics, and allows specialists to develop and test those components independently.

But accurately modeling the parts does not guarantee an accurate account of their interactions. The difficulty is capturing how processes operating at vastly different scales influence one another. Clouds can form in minutes, while deep-ocean circulation redistributes heat over centuries. Microbes alter soil chemistry, vegetation changes rainfall, melting ice redirects currents. Processes too small or fast to resolve directly must be approximated, and interactions between modules can only occur through the exchanges researchers have specified. These choices determine which feedbacks a model can capture—and errors in those feedbacks can propagate through the system. A model can reproduce the behavior of each component in isolation yet misrepresent how the planet behaves as a whole.

For example: almost no existing model is capable of handling the Antarctic ice sheet. Meltwater that exists in the real world is invisible to the model's ocean module. Therefore, these models expected the Southern Ocean surface to warm over the past forty years, but the real one cooled. Hardcoding in the missing freshwater explains some but not all of the error. This isn't just an academic curiosity. Cold water from the Southern Ocean flows north into the eastern Pacific, and that pattern determines the strength of cloud feedbacks worldwide.

Those clouds turn out to be a big deal. Recent warming has accelerated largely because of a decline in low-cloud cover, which makes the Earth less reflective overall. The result is that Earth's energy imbalance – how much solar energy we absorb versus what we reflect and radiate back into space – has more than doubled since the mid-2000s. This increase in the planet's energy imbalance is roughly twice as strong as the models predict it should be. Since 2003, the loss of low clouds has added about as much to that growing imbalance as greenhouse gases have.

There are countless such interactions that traditional models simply cannot capture. In 2019, Australian bushfires sent iron-rich smoke across the Southern Ocean and fed plankton blooms thousands of miles downwind. Recognizing this cause-and-effect spans modeling of land fire, global air circulation, atmospheric chemistry, and ocean biology. Understanding what those blooms do requires us to cast an even wider net: beyond impacts on fisheries and economic productivity, those plankton release gases that further modify cloud formation and reflectivity.

Here's one final example that is hard to wrap your head around. Over the last century, we've pumped so much groundwater for irrigation, mostly in northwest India and the American West, that the planet's rotational pole has tilted by nearly a meter. Scientists only figured this out because changes to ice sheets and glaciers alone couldn't account for the drift that was observed in the real world.

New couplings continually emerge as we alter the planet and it responds. No scientist would have thought to write an equation connecting groundwater pumping by Punjabi farmers to the motion of the North Pole. It's obviously wrong for a nature model to inherit the boundaries of human scientific disciplines. Trying to enumerate every consequential interaction in advance is a dead end — in fact, it reminds me a bit of the early days of symbolic AI before deep learning took over.

This is why the archive mentioned earlier matters. So much of what the observing system has collected so far is underutilized because its pipeline was built to feed a narrow set of modules. A nature model could learn from all of it, taking advantage of much more of the information embedded in the archive.

In 2019, one of the founders of reinforcement learning wrote a now famous essay titled "The Bitter Lesson." Its core claim is that in the long run, research strategies that leverage more compute and more data will always win out against ones that seek to encode specialist human knowledge. **Well, the Earth is not exempt from the bitter lesson.** The latent structure of all these couplings is, in principle, recoverable from a large enough set of observations over enough time, so long as the relevant quantities are being observed at all (which is where principle 5 comes in). That means we need to extend the idea of data assimilation one level deeper, from separate systems for the atmosphere, the ocean, and the land to a single estimate of the full state that binds the planet together. The nature model should learn these through whatever internal representation it needs.

That freedom of representation creates a new problem, especially as the historical record of observations drifts further from the planet we live on now: what keeps it grounded in reality? The nature model must gather new observations, in addition to leveraging the physics we already know for sure.

PRINCIPLE 5

Fetch New Observations

Physical observation is the expensive step in our version of distillation, **so a nature model needs to reason about which new observations are worth collecting.** If it learned the assimilation process correctly, and if it is built to express uncertainty rather than spit out a single best guess, it already knows where it's guessing. The next step is to rank those guesses: which are least certain, which of those actually matter for a decision someone is trying to make, and which measurement would be expected to teach it the most (what experimental designers call expected information gain).

The nature model must weigh all that against what the measurement costs, and costs vary by orders of magnitude: a dedicated satellite pass, a research vessel steaming to a mooring, a balloon launch, a seal that was going there anyway, a phone accelerometer that already exists.

That should inform a running list in response to this question: **where should the next marginal dollar of the observational budget go?** Often the answer won't be one expensive instrument but instead a pile of cheap ones whose noisy readings, aggregated, add up to the same information, the way billions of Android phones can become a seismic sensing network.

For any of that to work, the observing system needs two upgrades.

1. **From preplanned to adaptive**. Plenty of instruments will stay put. Once you've drilled a sensor into the ice sheet or anchored a moored array in pirate territory, you are not moving it. But satellites, drones, balloons, floats, autonomous vessels, and seals can increasingly be pointed at the places, times, and phenomena where the model expects a measurement to matter most. NOAA steering a Saildrone into a hurricane is a preview of what this looks like when it's routine. Forecasters have flown targeted dropsonde missions into hurricanes for decades for the same reason. The difference now will be doing it for the whole Earth system, continuously.

2. **From scarce to abundant**. Sparse to dense, episodic to continuous, batch to streaming, bespoke to cheap and modular. A model that can ask "what do I most need to know, and where should I look?" only helps if there's a sensing network with enough capacity to answer.

Both trends are already underway, but serving a nature model could accelerate them.

The End of Natural History

AGI WON'T BE ABLE TO STABILIZE THE CLIMATE or steward agriculture, forests, or fisheries without being directly coupled to the Earth system through observation and feedback. This is not optional, and it is not a matter of me being insufficiently AGI-pilled: in fact, it's the opposite.

The smarter the brain in the data center gets — and the more power it has to act on the world — the more dangerous any gap between its model and physical reality becomes. We cannot escape the fact that its interventions must ultimately be implemented on the real Earth, of which there is only one. Natural General Intelligence would not itself solve alignment, but it could supply a critical prerequisite: **the empirical feedback loop that turns advances in AI capability into stewardship of the world humans physically inhabit.**

This raises the obvious question: why not wait? A superintelligent AI could design a much better ocean sensor in a few years, but it can't send that sensor back in time to the present day and give us the accumulated state change over the intervening four years that would enable a nature model to learn how cycles of natural variability play out.

While some components can be constrained with denser instantaneous observations, much of what we need to know about can only be learned from observations accumulated over time. El Niño returns irregularly every two to seven years. We have so much uncertainty about AMOC in part because we've been directly measuring the Atlantic overturning circulation only since the RAPID array began operating in 2004.

Remember, the planet overwrites its own state as it goes. An observation skipped this year is not replaceable; the physical conditions that would have produced it are continuously destroyed and rewritten. Infinite compute in 2030 cannot buy back the physical conditions of 2026 that were never measured in the first place. No matter how good AI gets, **nothing other than the real Earth can manufacture longitudinal observations**.

One final analogy. A recurring idea on the frontier of AI is some form of "merge." Brain-computer interfaces have been theorized for decades and are now in active development. The logic is simple: valuable information is locked inside a physical system that does not export it cleanly, so we need to build a channel dense enough to read its state and, eventually, to write back carefully.

The Earth poses a similar interface problem. Though the planet doesn't "think", it shares some of the characteristics that make a brain difficult to interface with:

- Enormous internal state
- Most of which is unobservable from the outside
- Tightly-coupled feedback across scales
- No second copy to experiment on
- No reset button
- A state that changes in response to intervention

Reading the brain requires tiny electrodes deep in it, or perhaps ultrasound. The Earth requires an observing system that can similarly capture enough state.

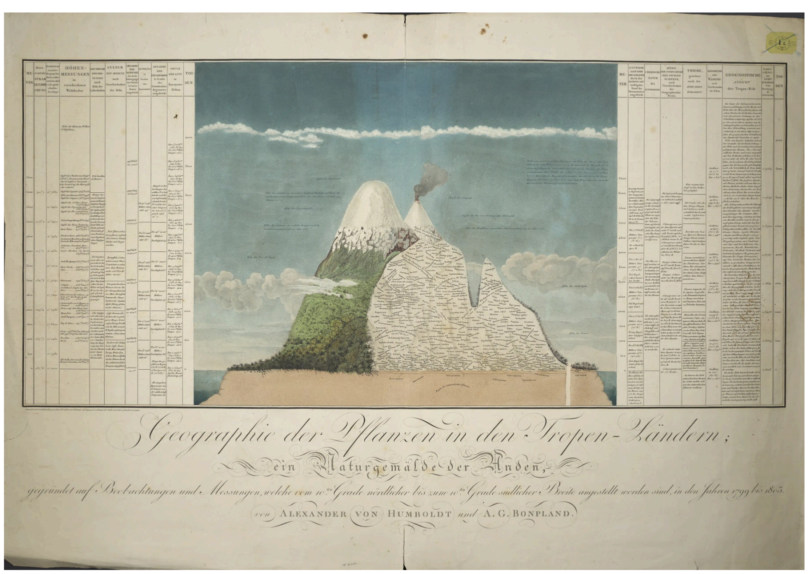
Natural General Intelligence represents the missing interface between machine intelligence and the Earth system, to do at planetary scale what a brain-computer interface could do for one organism. Lovelock was controversial in part because he viewed the Earth itself as an organism. I don't know if he was right about that, or if it matters, but the parallels are striking.

A nature model would not decide which interventions society should pursue, how to weigh the risks, or who has the authority to act. For the time being, those remain questions of human judgment and governance. Its purpose would be to make the causal chain more legible so that, just as our ability to shape the planet becomes vastly more powerful, we become less blind to the consequences.

IT CAN FEEL ALMOST OVERWHELMING to consider the full breadth of all the scientific solutions AI will unlock. Where should we focus? What should we prioritize? To me, the best way to think about how AI could accelerate scientific discovery is to ask which lineages of inquiry have yet to be fully realized.

Some scientific lineages are already fully unfolded. Take computer science: the line of inquiry that Turing and von Neumann began has been iterated to the very limit. Moore's law and the incredible diffusion of computing power throughout society fully realized their lineage of thought. Claude Shannon's lineage has been realized, beyond his wildest dreams, with modern AI tools and their unbelievably sophisticated pattern recognition.

I first learned about Alexander von Humboldt in 2020. At the time, I was a recovering software guy, learning Earth science in the context of my work on carbon removal. I was visiting some friends in DC and they brought me to the Smithsonian, which just so happened to be doing a brief special exhibition on Humboldt's life and work. The advertised centerpiece of the exhibit was a giant mastodon skeleton, which, of course, was very cool. Off in the corner, though, I found his original drawing of the Naturgemälde. I was entranced.



I realize now why I was so fascinated. To me, he represents the beginning of a lineage of inquiry about what's actually going on outside, the conditions of which humans now shape to a degree he never could have imagined. That lineage flows from him through Darwin, FitzRoy, Lovelock, and so many others. But in my view, they're all trying to answer the same fundamental question: why is the natural world around us what it is? Why does it do what it does? Progressively better instruments and wider observations have enabled better answers, though a complete understanding has remained just out of reach.

Humboldt inspired me to realize something quite simple: we are not done. **Superintelligence, such as we can conceive of it, won't mark the end of natural history.** It will instead become the bridge to its next chapter.

James Lovelock died in 2022. I'm quite sad I never got to meet him. But his death maybe clarified something for me: If he, Humboldt, or Darwin were alive today, gripped by the same questions but armed with today's instruments, what would they be working on? I'd like to think it'd be a nature model like the one I'm describing, and that they'd see it as the full realization of the lineage they began.

Thanks to Alisha Kumar, Amyad Spector, Jackson Dahl, Jared Rosner, Jessica Seddon, Kristin Ellis, Mackenzie Burnett, Maddie Hall, Nadia Asparouhova, Nan Ransohoff, Quico Torro, Will O'Brien, and members of the Lowercarbon Capital team for feedback on this post, and to Kristin for years of conversation on these topics.