The Reaggregation of Dispersed Knowledge: A Computational Epistemology of Capital

On disclosure, the join, and the physical limits of what capital has already announced

Tan Dhingra

Meridian Deal Partners · Meridian Intelligence

Working Paper · October 2026

Second edition. Supersedes the September 2026 working paper of the same title.

Abstract

Friedrich Hayek argued in 1945 that the central economic problem is not the allocation of known resources but the use of knowledge existing only in dispersed, fragmentary and largely tacit form across millions of minds. The price system, in his account, is a communication mechanism — in a vocabulary that did not yet exist, a distributed computer — that no central authority could replicate. This paper advances a narrower and, I will argue, historically specific claim: that Hayek's argument bundled together two kinds of knowledge which have since come apart. The genuinely tacit — the perishable judgement of time and place — remains exactly as uncentralisable as he said. But the itemised substructure of the capital economy, and above all its physical substructure, has been progressively externalised by four decades of disclosure law into the public record. For that stratum, and only that stratum, the binding constraint has moved from acquisition, which Hayek thought impossible, to integration, which is merely hard. Integration is a join, and a join is a computation.

What the join shows is physical. Of 202.2 gigawatts of generating capacity announced to the United States federal government in 2022, 46.5 per cent was delivering power three years later. That three-year outcome is observed for the 2022 cohort alone; the replication is at two years, where the 2022 and 2023 cohorts converted at 36.9 and 39.1 per cent, a difference of 2.2 percentage points — which is what makes this a rate and not a bad year. Over the same window the announced pipeline grew 61 per cent and the operating fleet 9.8. Conversion runs from 98.9 per cent for small plant sited beside existing wires to 1.7 for offshore wind and zero for pumped storage, and varies fourfold across jurisdictions sharing the same physics — Florida 84.2 per cent, New York 19.0 — while large-load interconnection requests in New York grew twentyfold in three years and now carry named data-centre and semiconductor requesters. What converts is what can be built where interconnection already exists, which makes an existing high-voltage connection the scarce asset in the build-out: a screen of two public federal files returns 285 retired thermal sites, 136.9 gigawatts, in counties carrying extra-high-voltage substations and with no announced reuse — a screening population for diligence, not a verified inventory of rights. A second register agrees from an earlier stage: across four system operators, 1,066 gigawatts of requested interconnection has already been withdrawn.

I describe the instrument these measurements required — 5,058,265,273 records held, of which 1,864,734,940 form one resolved estate drawn from regulators in thirty jurisdictions, held bitemporally under an explicit discipline of refusing uncertain merges — and I name the one layer beneath the whole build-out, the order books and lead times of heavy electrical apparatus, for which no public record exists anywhere. I argue that pattern detection across dispersed registers is computation rather than perception; that the instrument must model its own blindness as a first-class object; that it cannot claim causation it has not earned, except at the physical layer, where the causal model arrives from engineering rather than from inference; and that, because it consumes the same electricity whose scarcity it measures, it is an observer inescapably inside the system it observes. Forty dated predictions close the argument, resolvable against the 2028 vintage of the same file. I close on attention — the one faculty that scale does not manufacture — where Simon's economics and the contemplative traditions arrive, by different roads, at the same austere conclusion.

Keywords: disclosure · entity resolution · economics of information · computational epistemology · physical constraint · electrical infrastructure. JEL: D80, D83, L86, Q40.

In summary

Four decades of disclosure law pushed the itemised substructure of the capital economy into the public record. For that stratum the binding constraint moved from acquisition, which Hayek thought impossible, to integration, which is merely hard. This paper tests that claim where it can be checked against physics: the electrification build-out of the United States.

46.5%

OF ANNOUNCED CAPACITY ARRIVED

Of 202.2 GW proposed to the federal government in 2022, 94.0 GW was operating three years later. Two cohorts agree within 2.2pp at two years, so it is a rate and not a bad year.

1,066 GW

WITHDRAWN FROM THE QUEUE

Across four system operators, 1,066 of 1,672 GW ever requested is already flagged withdrawn — a cumulative sum across years, not simultaneous capacity.

285 sites

WHERE THE CONNECTION ALREADY EXISTS

136.9 GW of retired thermal capacity in counties carrying in-service substations at 345 kV or above, with no announced reuse. A screening population for diligence, not a verified inventory of rights.

2 of 4

SIGNALS SURVIVED A PLACEBO

Distress-flagged 8-K filers return +16.1% excess over 126 days against a placebo of −20.5%. Two intuitive insider-buying signals did not survive and are shown failing.

What would prove this wrong

1. Forty sites are registered as dated predictions, published as a companion sheet. If Set A (20 sites, 5,904 MW) does not substantially outperform Set B (20 sites, 9,089 MW) in the 2028 vintage, the constraint map describes 2022–25 and is not a property of the system.
2. If the disclosed stratum proved economically inert. An insider-transaction strategy in this estate collapsed from a large apparent premium to insignificance once dated on the filing date rather than the transaction date, and its placebo failed.
3. If the reconstructed state proved not to be decision-relevant, because allocation here is dominated by relationship and by the tacit residue Polanyi describes.

Every figure above is carried through the body and sourced there. Two measurements were withdrawn from this edition after audit and are reported as withdrawn rather than removed: a private-credit aggregate that disagreed with the filer's own published total, and a county-level attention corpus whose geocoding proved empty. Three of those errors were caught before publication and one after, by an outside reader — a ratio stated rather than rounded up, and the instrument's only current track record.

1 The problem of seeing

> *The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess.*
>
> — Hayek, 1945

Hayek's essay is remembered for its conclusion, and the premise is the part that still works. The data from which an economic calculation would proceed, he argued, are never given to a single mind. They exist dispersed, fragmentary and perishable — knowledge of time and place, held by the person on the spot, expiring shortly after. The price system solves this without anyone solving it: a scalar, broadcast to all, letting each actor take account of circumstances they will never learn. Read now, the passage describes a distributed computer in a vocabulary that did not yet exist.

Something has happened since that the argument had no occasion to anticipate. Four decades of disclosure regulation have pushed an enormous quantity of economic particulars into the public record: who holds what, who crossed which threshold on what date, which generating unit exists, where, at what capacity, in what state of construction, owned by whom, and which project has asked a system operator for permission to draw power it does not yet have. That is not tacit knowledge. It is filed, dated, itemised and machine-readable. And the cost of assembling it has fallen far enough that a small, disciplined team can attempt it.

None of this refutes him, and the careful version of the claim is worth separating from the loose one. A filing is not the knowing; it is a trace a private state left behind because the law required it. Polanyi's point stands undamaged: the operator who knows which of two identical-looking sites will clear its permits knows something no register contains and could not fully articulate if he tried. That knowledge is exactly as uncentralisable as Hayek said.

What has changed is narrower, and it is the whole of this paper. The knowledge Hayek said could never be given to a single mind is now, in significant part, given — deposited in filings that anyone may read. What is not given is the join. And a join is a computation.

So the opacity is not secrecy. It is fragmentation — and fragmentation, unlike secrecy, is an engineering condition rather than a fact of nature. The registers are fragmented by construction: each was built by an institution for its own purpose and none to be read against another.

Here is what that costs, in one instance. In the 2024 vintage of the federal generator inventory, ten units appear at a plant in Indiana County, Pennsylvania: 3,324 megawatts of proposed gas, no operating date, one of several thousand rows in a file almost nobody reads end to end. Hold the 2022 vintage beside it and the same plant code is a coal station — 2,012 megawatts, commissioned in 1969 and 1977, all operating. Open a third register, kept by another agency for an unrelated purpose, and a 345-kilovolt substation called Homer City stands in that county, in service, with nine transmission lines into it. The scarce asset at that site is not the land, the turbines or the gas. It is the right to connect. Three files, all public, all available for years, and no agency with any reason to publish them together. The fact was not hidden. It was unassembled.

It is a striking feature of the present moment that an argument about artificial intelligence, conducted almost entirely in the language of models and capital, terminates within three steps in a four-figure inventory of extra-high-voltage substations and the lead time on a class of transformer. That chain, and what it costs to assemble it, is the subject of this paper. Where a financial representation is carried far enough downward it terminates in something that cannot be argued with, and the interesting information is the discrepancy between what capital says it will do and what the physical system is willing to absorb.

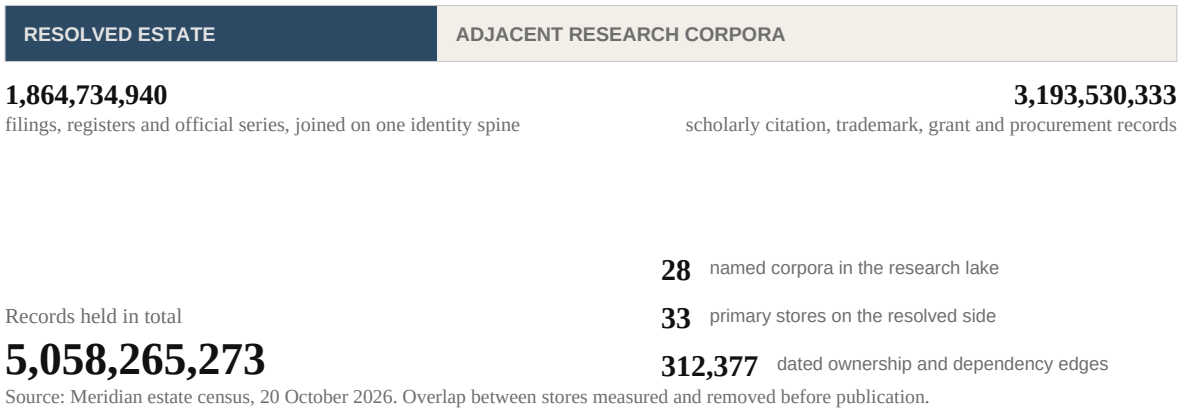


Figure 1. The estate, and what it resolves to. The two sides are not interchangeable and are never added together in an analytical claim: the left is joined on an identity spine and carries every measurement in this paper; the right is held for reference and is reported separately.

2 Prices summarise; they do not itemise

A price tells you that capital moved and at what marginal valuation. It does not tell you who moved it, into what, controlled by whom, under what mandate, or against which physical constraint. Compression is the virtue of the price system and it is also its limit: a scalar summary of a high-dimensional state is an extraordinarily efficient way to transmit what to do next and an extraordinarily poor way to transmit what is the case.

Prices have a second property that matters more here. They exist only where there is a market. There is no price for the interconnection right at Homer City. There is no price for a position in the MISO queue. There is no price for a transformer delivery slot in 2029 that an investor can observe. These are among the most consequential objects in the electrification build-out, and for every one of them the price mechanism is silent.

Where prices are silent the itemised record often is not, and the two compress complementary things. The price is dense, current and anonymous. The filing is identified, structured, sparse and lagged. Neither alone reconstructs the moving picture. Joined — the anonymous flow of prices read against the identified structure of filings — they begin to.

The strongest objection to all of this should be stated rather than waited for. A critic says: disclosure codifies precisely the part that was never tacit, so the codified stratum was always the easy part and the hard part remains where it always was. There is force in that, and Hayek's later argument sharpens it — competition is a discovery procedure, and what it discovers is not in the record until after it has been discovered. I do not think this can be argued away and I do not try. But it can be measured: the lag between action and disclosure is directly observable, and that lag is the width of the window in which the discovery argument still owns the field entirely.

Grossman and Stiglitz are usually invoked here and usually overread: what they establish is that gathering information is not self-extinguishing, not that it is profitable. That is the whole of the economic claim made here.

One question organises everything that follows: who is doing what, through which entity, with whose capital, where, when, dependent on what, constrained by what — and what evidence would show that this is wrong?

A paper that moves between a federal inventory, a thermodynamic bound and a commercial proposition owes the reader a way to tell which kind of claim is in front of them. Four kinds appear here and they do not carry equal weight. Table 1 fixes the vocabulary; every later section is written to sit in exactly one row of it, and the weakest rows are where the next edition has work to do.

Table 1. The evidentiary ladder used throughout. This paper is strongest in the first two rows. The third is argued from grid topology rather than estimated from the data, which is the licence discussed in §4; the fourth is a claim about decision advantage, registered as a prediction in §12 rather than asserted here.

Kind of claim	Instance in this paper	What it establishes
Direct observation	capacity proposed in one vintage, operating in a later one	a measured outcome in the record
Association	conversion differs by technology and jurisdiction	a pattern in the observed population
Mechanism	existing interconnection explains which projects arrive	argued from engineering, not identified here
Commercial hypothesis	the screen finds sites worth diligence before others look	untested; no outcome observed yet

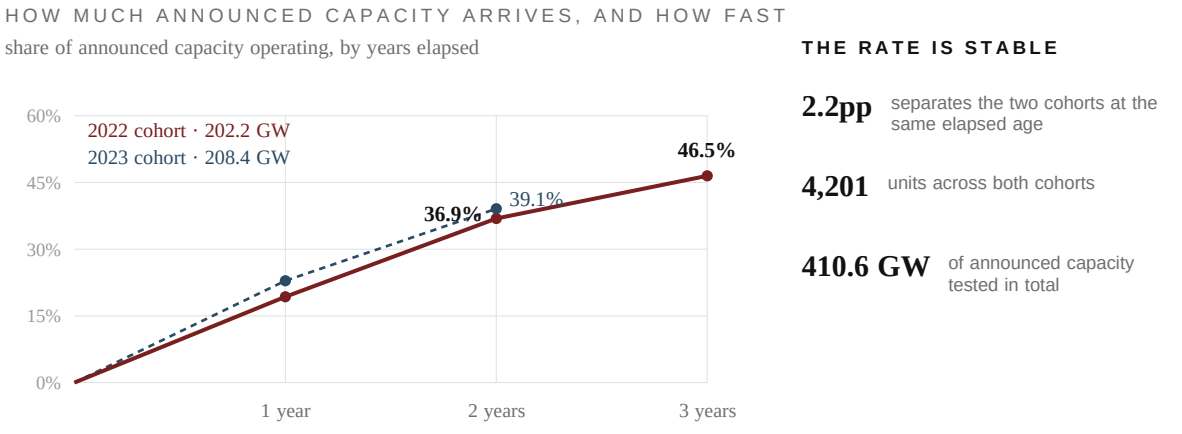
Note: a physical constraint that is real is not automatically a constraint that is mispriced. Nothing in this paper establishes mispricing, and that distinction is kept live throughout rather than conceded once in the closing section.

3 The capital–reality gap

The Energy Information Administration publishes Form 860 annually and each vintage restates the entire US generating fleet. The agency does not publish a longitudinal outcome file, and the old vintages are not generally kept. Hold four and match on plant code and generator identifier, and the inventory becomes a panel: a unit recorded as proposed in 2022 can be followed into 2025 and its fate observed rather than assumed.

The 2022 cohort is 2,023 proposed units, 202.2 gigawatts — capital's statement of intent at a stage well past a press release, reported to a federal agency as forthcoming. Three years later 94.0 gigawatts, 46.5 per cent, was operating. Another 73.6 was still proposed, 26.1 had been retired or cancelled, and 8.5 had left the file altogether.

One cohort is an anecdote about one year. The measurement holds because it replicates: the 2023 cohort, a separate population of 208.4 gigawatts, converts at 39.1 per cent at the same elapsed age against the 2022 cohort's 36.9. Two independent cohorts, 4,201 units and 410.6 gigawatts, agree within 2.2 percentage points. Figure 2 plots both against years elapsed rather than calendar time, which is what makes them comparable. Every one of these figures re-derived unchanged from the same four vintages held bitemporally. The dates are worth separating, because a re-measurement is evidence only if its interval is stated: the vintages were loaded 17 September 2026, the figures first published 22 September 2026, and re-derived on 19–20 October 2026 — an interval of one month. The first edition of this paper claimed thirteen months. That was wrong.



Two cohorts, measured independently, agree to within 2.2 percentage points at the same elapsed age. The curve also flattens: the 2022 cohort added 39.0 GW in its first year, 35.7 in its second and 19.3 in its third.

Source: EIA Form 860, annual vintages 2022–2025, matched on plant code and generator identifier.

Figure 2. The conversion curve. The solid line is the 2022 cohort and the dashed line the 2023 cohort. The vertical distance between them at two years is the replication error.

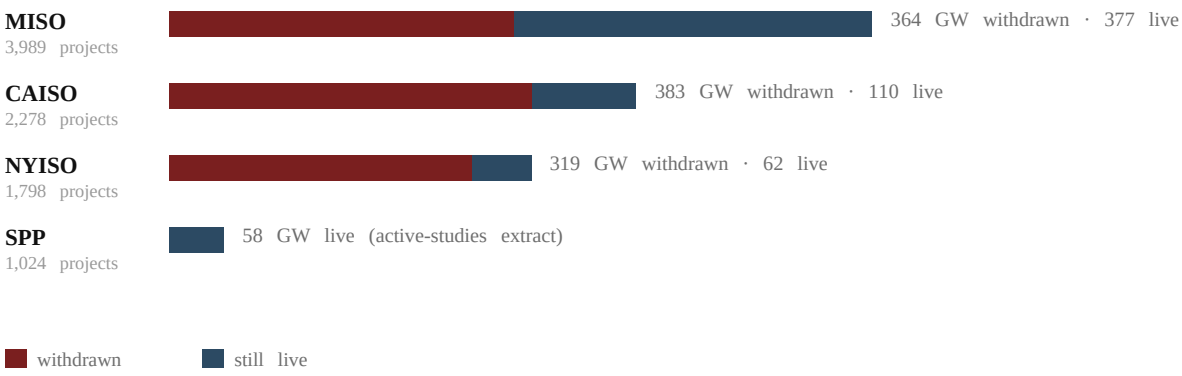
Three things follow. The divergence is widening: across the same four vintages the announced pipeline grew

from 202.2 to 326.5 gigawatts, up 61 per cent, while the operating fleet grew from 1,253.7 to 1,376.7, up 9.8. Announcement is accelerating roughly six times faster than delivery. A third of the pipeline is not moving: of the 326.5 gigawatts proposed in 2025, 103.7 have carried the status proposed for three or more consecutive vintages and 74.8 for all four, announced in 2022 and neither arrived nor dead.

And a second register agrees, from an earlier stage of the same process. Across the four system operators whose queues are published in usable form — MISO, CAISO, NYISO and SPP, which between them exclude PJM, the largest US market — 9,089 projects have requested interconnection, totalling 1,672 gigawatts. These are counted on the latest observation of each project. The source table holds seven dated snapshots of every project, and the first edition of this paper counted those rows instead, publishing figures roughly sixfold too large and describing the inflation as growth in the collection. It was not growth; it was double counting, and it is corrected here. Of the stated total, 1,066 gigawatts is already flagged withdrawn in the operators' own files — roughly two thirds of everything ever requested. Two cautions travel with that number. It is a cumulative sum of requests recorded across many years, not capacity that could ever have existed at one time, and so it is not comparable to the operating fleet, which is a stock; projects may also overlap geographically or represent alternative plans for one site. And a withdrawal records that a project stopped, not why: financing, offtake, land, policy, queue costs and technology choice all end projects, and the file does not say which. Lawrence Berkeley National Laboratory, measuring a broader population, finds 13 per cent of capacity entering US queues between 2000 and 2020 had reached commercial operation by end-2025 against 75 per cent withdrawn; the direction agrees, the definitions differ, and neither validates the other.

THE QUEUE IS MOSTLY A GRAVEYARD

interconnection requests by system operator: withdrawn against still live, gigawatts



1,066 GW withdrawn of 1,672 GW ever requested, across 9,089 projects

A cancelled project is conventionally read as absence. It is an experiment the economy ran on itself, in which a proposed state met a constraint and lost. The successful projects show what is possible; the abandoned ones show where the boundary is. SPP's published file is an active-studies extract and carries no withdrawn rows, so the withdrawal share is a floor.

Source: published interconnection queues of MISO, CAISO, NYISO and SPP, collected to 17 October 2026.

Figure 3. The interconnection graveyard. Withdrawn capacity against capacity still live, by system operator. PJM is absent because its interface requires a subscription, and ISO-NE because it publishes no machine-readable queue; both absences are recorded rather than interpolated.

A cancelled project is conventionally read as absence, a row that failed to become interesting. It is an experiment the economy ran on itself, in which a proposed state met a constraint and lost. Taleb's drowned worshippers do not leave testimonials; the failed venture does not file; the cancelled project is, in most datasets, simply not there. The successful projects show what is possible and the abandoned ones show where the boundary is.

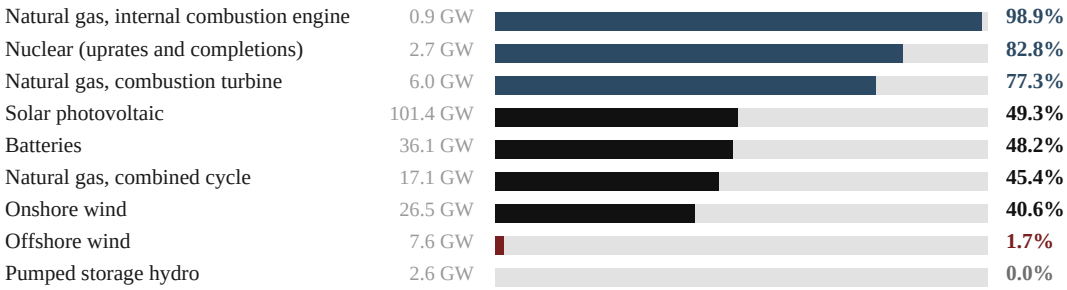
The gap between where capital has expressed a preference and where the physical world can accommodate it is not noise around the thesis. It is the thesis.

4 THE CONSTRAINT MAP

If attrition were mostly financing or mostly demand it would fall evenly across technologies. It does not.

WHAT CONVERTS IS WHAT CAN CONNECT

share of 2022-announced capacity operating in 2025, by technology



The ordering is not a merit ranking; it measures how hard each technology is to connect. Small, dispatchable plant sited beside existing wires arrives. What is large, remote and dependent on new transmission does not: offshore wind and pumped storage delivered 130 MW between them, of 10.2 GW announced.

Source: EIA Form 860, vintages 2022 and 2025.

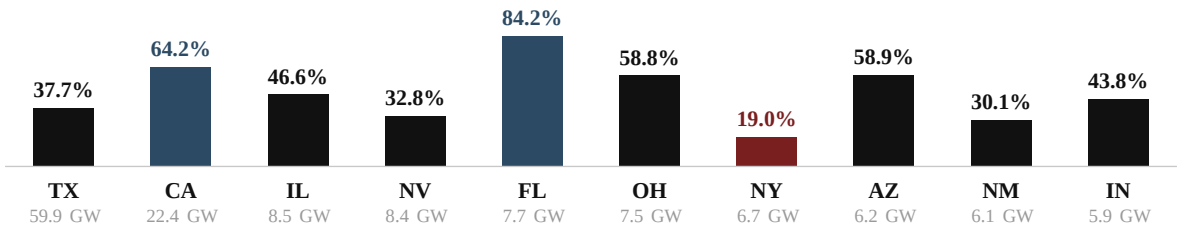
Figure 4. Conversion by technology. Share of each technology's 2022-announced capacity operating in 2025.

The comparison is descriptive, and the cohorts differ in more than technology — in project size, construction duration, whether a unit is an uprate at an existing site or a greenfield build, and in how mature a project was when first recorded. With that stated, what converts is largely what can be built where interconnection already exists. A gas reciprocating engine goes in beside an existing substation and runs. Nuclear at 82.8 per cent is not new construction — those units are uprates and completions at existing sites, capacity added behind interconnection built decades ago. At the other end, offshore wind and pumped storage need new transmission, new points of interconnection and construction sequences measured in half-decades, and between them they delivered 130 megawatts of 10.2 announced gigawatts. What is small, dispatchable and sited near existing wires arrives; what is large, remote and dependent on new transmission does not.

Disaggregated by jurisdiction the result is institutional rather than physical.

SAME PHYSICS, DIFFERENT INSTITUTIONS

ten states announcing the most capacity: announced 2022 against share operating 2025



Florida converts at 84.2 per cent and New York at 19.0. Texas announces three times more than any other state and converts at 37.7. These jurisdictions share the same physics and differ in their interconnection regimes, their permitting, and how far the party that wants the plant is vertically integrated with the party that owns the wires.

Source: EIA Form 860, vintages 2022 and 2025.

Figure 5. Conversion by jurisdiction. The ten states announcing the most capacity in 2022, ordered by announced volume, against the share of that capacity operating in 2025.

These are associations rather than identified effects: jurisdictions differ in project mix as well as in

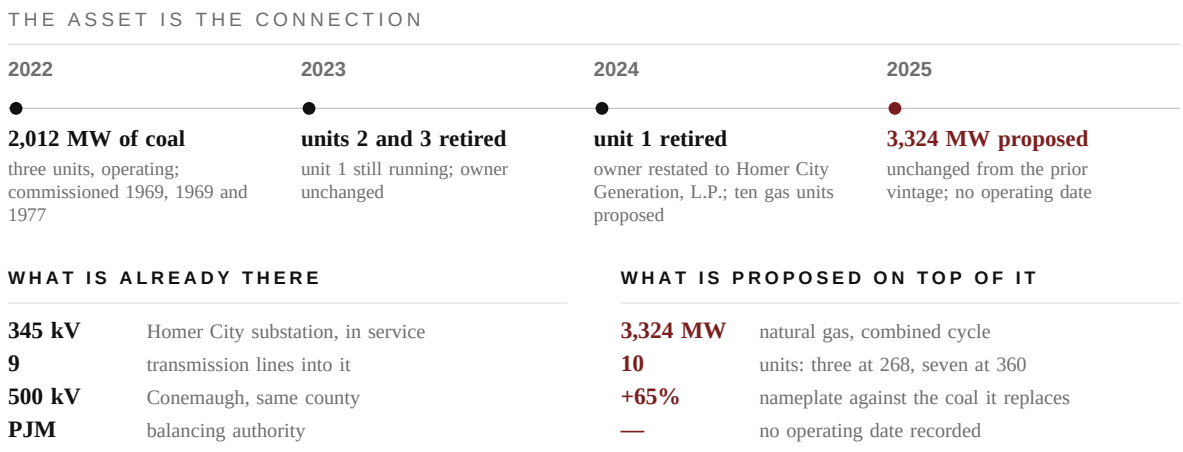
institutions, and nothing here isolates the contribution of the interconnection regime from the rest. Florida delivers 84.2 per cent of what it announces and New York 19.0. Texas, which announces three times more than any other state, delivers 37.7; California 64.2, Ohio 58.8, Arizona 58.9, Illinois 46.6, Nevada 32.8, New Mexico 30.1, Virginia 26.6. These jurisdictions share the same physics and differ in their interconnection regimes, their permitting, and the degree of vertical integration between the party that wants the plant and the party that owns the wires. This is North's argument appearing in a federal engineering file. An investor holding a portfolio of announced projects holds a portfolio of positions on interconnection regimes, whether they know it or not.

There is a reason this layer, and not the capital layer, carries the causal weight in this paper. Pearl's barrier is not a barrier against deriving a causal model from observational data; it is a barrier against deriving one from the data alone. Where a causal model arrives from outside the data — from physics rather than from inference — the ascent from association to intervention is licensed. Grid topology is such a model. The physical layer is therefore not a gesture at tangibility. It is the only place in the estate where Pearl's second rung is available, and that is precisely why it was built.

Held at its real strength: headroom is itself a modelled quantity — contingency analysis, seasonal ratings, N-1 criteria — and 'settled by the substation' is cleaner rhetorically than the engineering deserves. What survives is that the model comes from the world rather than from the correlations.

5 The scarce asset, and 285 more of it

If the interconnection right is the asset, then every retired thermal plant sitting on one is the same asset, and the two files that found Homer City will find the rest. Figure 6 shows the transaction as the record holds it: 2,012 megawatts of coal retiring across two vintages, the owner restated, and 3,324 megawatts of gas proposed at the same plant code — a 65 per cent increase in nameplate on a connection built for the smaller number.



3,324 MW at one site is half of everything New York announced in 2022, and it is going where a 345 kV substation with nine lines is already in service. The scarce asset is not the land, the turbines or the gas. It is the right to connect, which the retired plant already holds and which a new site would spend the better part of a decade attempting to obtain.

Sources: EIA Form 860 vintages 2022–2025, plant code 3122, Indiana County, Pennsylvania; HIFLD substation registry.

Figure 6. Homer City, from three public registers. A system holding only the current vintage sees a dead coal plant and a large gas proposal, with no relationship between them.

The screen generalises it, and the definition matters more than the number, so it is given in full. Take every plant carrying 50 megawatts or more of thermal capacity — coal, gas steam, combined cycle, combustion turbine, petroleum liquids, petroleum coke, integrated gasification, other gases and nuclear — recorded as retired in the 2025 vintage: 463. Exclude any plant code carrying a proposed unit, because there the transaction has already been announced: 445. Keep those in counties holding an in-service substation at 230

kilovolts or above: 374 sites, 166.9 gigawatts. Raise the threshold to 345 kilovolts and the screen returns 285 sites and 136.9 gigawatts, roughly 9.9 per cent of the entire US operating fleet, at locations whose grid connection was built and paid for and is no longer doing what it was built to do.

The concentration is high. Texas holds 41 sites and 21.1 gigawatts, Illinois 23 and 13.4, Florida 17 and 13.3. By recorded owner, the Tennessee Valley Authority holds 8 sites and 10.1 gigawatts, Luminant 12 and 8.9, Midwest Generations 12 and 8.7.

Table 2. The head of the screen. Retired thermal capacity at plants with no announced reuse, in counties carrying in-service substations at 345 kV or above. The full screen is 285 sites.

Plant	State	Retired MW	kV	Lines	Recorded owner
Martin	FL	3,623	500	11	Florida Power & Light
Monticello	TX	2,780	345	10	Luminant Generation
Collins	IL	2,650	765	10	Midwest Generations EME
Paradise	KY	2,558	500	13	Tennessee Valley Authority
Navajo	AZ	2,409	500	6	Salt River Project
P H Robinson	TX	2,314	345	16	NRG Texas Power
San Onofre	CA	2,254	500	14	Southern California Edison
Willow Glen	LA	2,178	500	5	Entergy Louisiana
Polk	FL	1,992	500	4	Tampa Electric

Note: voltage and line count are those of the highest-rated in-service substation in the plant's county. The join is by county rather than by verified electrical connection, so this is a candidate list and not an inventory of confirmed rights; confirming a site requires the interconnection agreement or the operator's one-line diagram. What the screen establishes is where to look and in what order.

The scarcity is structural. The national registry holds 75,327 substation records, 56,894 with a recorded voltage, of which 2,123 operate at 345 kilovolts or above, 518 at 500, and 36 at 765. Those few thousand nodes are the points at which a large new load can physically be connected, and there is no process by which wanting more of them produces more of them inside a decade.

Homer City does not appear on the screen, because it no longer qualifies: it has proposed units, which is the condition the screen excludes. The list returns the sites where the same transaction has not yet been announced. A claim about capital can be argued with. A claim about whether a 400-megawatt load can connect to a particular substation is settled by the substation.

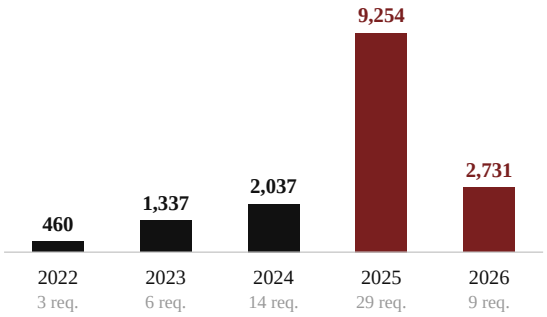
One caution belongs here rather than in a footnote, because it bears on how such a number should be read at all. The September edition of this paper reported this screen at 281 sites and 161.0 gigawatts under a narrower and now unrecoverable technology list, and the substation registry has since been re-collected. The screen is therefore restated under the definition given above rather than compared across editions, and the two numbers should not be differenced. A measurement whose definition is not retained alongside it is not a measurement; it is a number that was once true.

6 Demand arriving

The New York system operator publishes large-load interconnection requests — parties asking permission to draw substantial power, with a named requester, a megawatt figure and a queue date. It is the only operator in this collection publishing them in this form, which makes the series a floor for the United States and silent about PJM and ERCOT.

DEMAND ARRIVING, WITH NAMES ATTACHED

large-load interconnection requests in New York, by year of queue entry



megawatts requested; 2026 to 28 August

WHO IS ASKING

1,935	St. Lawrence Infrastructure 2	2025-09-02
1,000	1 Gig Data Center East Fishkill	2025-07-17
1,000	NY-Route 9W Load Center	2026-01-07
860	St. Lawrence Infrastructure 1	2025-09-02
606	Micron Fab 3	2026-08-28
576	Micron Fab 2	2023-10-31
500	North East Data LLC Data Center	2025-08-29

Requests for firm power in New York grew from 460 MW across three requests in 2022 to 9,254 MW across 29 in 2025 — a twentyfold increase in three years, in the state that converts announced generation into operating generation at a fifth of Florida's rate. The requesters are named: data-centre vehicles, a semiconductor fabricator, an industrial forge.

Source: NYISO large-load interconnection queue, collected to 17 October 2026; 2026 to 28 August.

Figure 7. Large-load interconnection requests in New York. The right-hand column is the head of the request list by size, with the requester as the file records it.

Each series is unremarkable alone. A rising queue indicates interest; a low conversion rate is a familiar complaint of New York developers. Set beside one another they describe a jurisdiction accumulating committed industrial load against a supply process that is not clearing, and the accumulation is visible years before it can reach an operating dataset, an earnings report or a price. All years together run to 74 requests and 17,642 megawatts.

The named requesters are the part that makes this concrete, and they are what the first edition of this paper could only gesture at. The largest single request on the list is an infrastructure vehicle at 1,935 megawatts. Below it sit a data centre at East Fishkill at 1,000, a load centre on Route 9W at another 1,000, and two Micron semiconductor fabrication phases at 606 and 576 — a manufacturer whose capital programme is discussed constantly in the financial press, here visible instead as a specific quantity of power requested from a specific operator on a specific date. This is the whole argument in one row. The fact is public, it is dated, it carries a name, and it is not in any dataset an allocator is likely to be reading.

7 Why this is not already visible

Everything above came from files that are public and years old. Why they are not already read this way is structural. Coase asked why firms exist when markets could co-ordinate every transaction, and answered that transacting has a cost. Williamson located the boundary at assets specific enough to make opportunism a live risk. North observed that the rules governing all of it are institutional artefacts, evolved and costly to change. Together they give the shape of the public record: an institution generates a record when a rule requires it, and the rule exists because some legislature judged that disclosure worth its cost. The record is therefore a map of institutional obligation and not a map of economic importance. The consequence for §3 is exact — the Energy Information Administration restates each vintage because its users want a current picture, nobody is obliged to keep the old ones, and so that outcome data, which the agency does not publish, does not exist until somebody holds four files instead of one.

Two timestamps attach to every fact. The event time is when the thing happened; the knowledge time is when any outside observer could first have known it. They are routinely months apart, and collapsing them produces a fictional past that always flatters. Nearly every seductive backtest in finance is an unrecorded violation of this principle. The estate has made the error itself and reports it: an insider-transaction strategy that appeared to earn a substantial premium when dated on the transaction date collapsed to insignificance when dated on the filing date, the only date on which the information is public, and its placebo leg failed outright.

Commercial vendors overwhelmingly restate, which is correct if the product is a current view of the world.

The discipline of not restating is itself long established in one domain — S&P's Compustat Point-in-Time has supplied as-of company financials to quantitative investors for decades, and no novelty is claimed here. What has no equivalent is the physical layer: no vendor sells an as-of view of the generating fleet, the interconnection queue or the substation network, which is what makes §3 unavailable to anyone buying data rather than keeping it. A firm beginning this discipline in 2029 cannot buy the 2026 view of the physical record at any price.

The second requirement is identity, and its errors are asymmetric. In a system whose output is inference over joins, a false identity is not a data-quality issue; it is an epistemic catastrophe that announces itself nowhere. Deciding that a name in one register and a differently-spelled name in another denote one firm is not a lookup but a judgement, drawn from a tacit background of how companies name and rename themselves, and performed at the margin by a person. The tacit knowledge Hayek said could not be centralised has not left this system. It has migrated — out of the fact and into the join. Figure 8 is what it costs when it is not done.

A related failure is quieter: an inference made on Monday is displayed on Tuesday without its provenance and quoted on Wednesday as a fact. The remedy is not accuracy but tiering. Of the $39.53 trillion of US-listed equity reported by 8,758 institutional managers in the most recent quarter, 44.7 per cent resolves to a jurisdiction of ultimate control. The correct statement about the remaining 55.3 per cent is that it is unresolved. It is not American. Reporting it as American would make every chart look complete and every chart wrong.

ONE DEVELOPER, FIVE RECORDED NAMES, FOUR VINTAGES

		2022	2023	2024	2025
Ocean Wind Ocean Wind, LLC · NJ	1,215 MW	proposed	**cancelled**	**cancelled**	**cancelled**
Ocean Wind II Orsted North America Inc. · NJ	1,148 MW	proposed	**cancelled**	**cancelled**	**cancelled**
Sunrise Wind Orsted Wind Power NA LLC · NY	924 MW	·	proposed	proposed	proposed
Skipjack Phase 2 Skipjack Offshore Energy, LLC · MD	846 MW	proposed	proposed	**cancelled**	**cancelled**
Revolution Wind Eversource → Orsted · RI	715 MW	proposed	proposed	proposed	proposed
Western Trail Wind Orsted Onshore North America · TX	367 MW	operating	operating	operating	operating
South Fork Wind Eversource → Orsted · NY	130 MW	·	proposed	operating	operating

5,345 MW announced · 3,209 cancelled · 1,639 still proposed · 497 operating **a 9.3% delivery record**

No single row shows this. Assembling it requires resolving five recorded owner names to one balance sheet, and retaining the 2022 vintage: two of these assets are recorded under Eversource in 2022 and under an Orsted entity thereafter, which a restating system would have overwritten.

Source: EIA Form 860, vintages 2022–2025, resolved against the Global Legal Entity Identifier register.

Figure 8. One developer, five recorded names, four vintages. No row is wrong: each is the correct legal name of the vehicle holding that asset. But the file cannot answer what the developer's delivery record is, because it does not know the five names are one balance sheet.

One consequence deserves to be stated plainly, because it is the first question a regulator or an allocator asks. This instrument does not create private information and does not want to. It industrialises the assembly of public information until the asymmetry that survives is not between insider and outsider, which is a regulated and dangerous asymmetry, but between the party that has done the join and the party that has not. That is a defensible asymmetry. It is the only kind worth building on, and it has the useful property that it can be described in full to a regulator without discomfort.

8 The arrival process

A disclosure is an event in time, and events in time have structure that a cross-section cannot see. A firm that raised capital twice in the last ninety days is not in the same state as a firm that raised the same total once, three years ago, and a feature set that counts prior filings without weighting them by recency cannot tell the

two apart.

The natural formalism is the self-exciting point process Hawkes introduced in 1971 for earthquake aftershocks, in which the arrival of an event raises the intensity of further events and that elevation decays. Written for a filing history, the intensity at time t is a background rate plus a sum of decaying kernels over prior events — one term for each thing the entity has already done, each fading on a characteristic timescale. Three timescales were wired, at thirty, ninety and three hundred and sixty-five days, together with a burstiness ratio comparing recent intensity against the long-run rate. The features sit on all 412,154 feature snapshots in the estate, written alongside the incumbent columns rather than over them.

The hypothesis was registered before the test, with its falsifier fixed in advance: reject the null if a paired bootstrap of the skill difference on a strict temporal split excludes zero. The incumbent shape summarises recency with the log of the prior count and the log of days since the last filing. The challenger adds the kernels and changes nothing else.

Table 3. Self-excitation against the incumbent recency shape, scored on identical rows at a base rate of 10.63 per cent. The paired interval excludes zero, so the pre-registered null is rejected.

Feature shape	n	Brier skill	95% interval
Incumbent recency summaries	20,057	27.7%	25.1 – 30.23
With Hawkes kernels	20,057	30.71%	28.35 – 32.99
Paired difference	—	+3.01 pp	2.64 – 3.4

Note: the comparison is paired on the same rows, which is what makes it readable. Both models see the same population, so the coverage asymmetry described below cannot account for the difference between them. It could account for the level of either, which is why the paired difference and not the level is the reported result.

So the arrival process carries information the incumbent shape was discarding, and the gain is three points of Brier skill rather than a rounding error. What the result does not do is leave the disclosed record: the outcome is still a filing, and §12 holds that boundary.

One further caution belongs here, and it is the reason the level is reported separately from the difference. A second registered hypothesis in the same lab asked whether the live evaluation population was trustworthy, and found that it was not: the shadow universe held every matured negative and roughly one matured positive in twenty. A skill figure computed on that population describes the population's shape at least as much as the model's quality. The lab established this about its own headline number and recorded it, which is why the Hawkes result is stated as a paired difference on identical rows — the one form the coverage problem cannot contaminate.

9 A stratum that became observable, and a measurement withdrawn

One stratum moved from unobserved to partially observed since the first edition, and it is worth describing for two reasons: the structural point stands, and the measurement I built on it did not survive audit. Both halves belong in the paper.

Business development companies are registered vehicles that lend to private borrowers, and they are obliged to publish a schedule of investments: borrower, instrument, reference rate, spread, maturity, principal, cost and fair value, position by position. That obligation makes a slice of private credit — an asset class whose defining property is that it is not marked in public — visible at position level, and it is a clean instance of this paper's thesis: the disclosure exists, it was created for an unrelated supervisory purpose, and nobody assembles it across filers.

The first edition reported position counts, an aggregate fair value and a dispersion statistic computed from that corpus. Those figures are withdrawn. Adversarial review asked whether the aggregate could be reconciled, and it cannot: the extraction is defective in three compounding ways. Borrower names are split across the borrower and instrument fields, so a single obligor appears as a stub name with the remainder of its own name recorded as the instrument. A large share of value carries no instrument classification at all. And one equity value is replicated across each of a borrower's share-class rows, so a single stake is counted repeatedly.

The test that settles it is external. Reconstructed from this corpus, one large listed lender's portfolio reads between three and four times the total it states in its own public filings. Deduplication does not repair the gap,

because the defect is in the parse rather than in repetition. A figure that disagrees with the filer's own statement of the same quantity is not a finding about private credit; it is a finding about the parser.

So the domain is scored partially observed in the ledger of §11 on the strength of the obligation rather than of my numbers, and no private-credit quantity appears anywhere in this edition. The stratum is real and remains the most promising unexploited disclosure in the estate. What it needs is a rebuilt extractor reconciled line by line against filed totals, which is work, not a caveat.

I report this at length rather than quietly dropping a section, because the quiet option was available and worse. The withdrawn figures were plausible, they were specific, and nothing inside the estate contradicted them. Only a comparison against the filer's own published total exposed them. An instrument whose errors are visible only from outside has not yet earned the reader's trust on the figures nobody checked — which is the argument of §11, arriving a section early and at my own expense.

10 The observer inside the system

Landauer established that information processing carries an irreducible thermodynamic cost: erasing one bit dissipates at least kT ln 2 as heat, about three zeptojoules at room temperature. Bennett completed the argument by resolving Maxwell's demon with it — the demon appears to violate the second law only until one accounts for the cost of clearing the memory it needs to keep sorting.

The bound is irreducible in principle and says nothing about magnitude; real hardware dissipates some five to six orders of magnitude more per operation. The relevant figure is therefore engineering rather than thermodynamic, and it should be stated rather than waved at. The instrument described here presently draws on the order of a few megawatt-hours a year. Setting that against the capacity it inventories requires a common basis, since gigawatts measure power and megawatt-hours measure energy: the 326.5 gigawatts announced in the 2025 vintage would, at a forty per cent capacity factor, generate on the order of a billion megawatt-hours a year. The ratio is thus around one part in a billion. It is not fixed. The measurement programme it supports is bounded by statistical power rather than by insight, and closing that bound means ingesting several orders of magnitude more of the record — which raises the numerator and leaves the denominator alone. A paper that asks others to quantify their blind spots should quantify its own, including the one that grows.

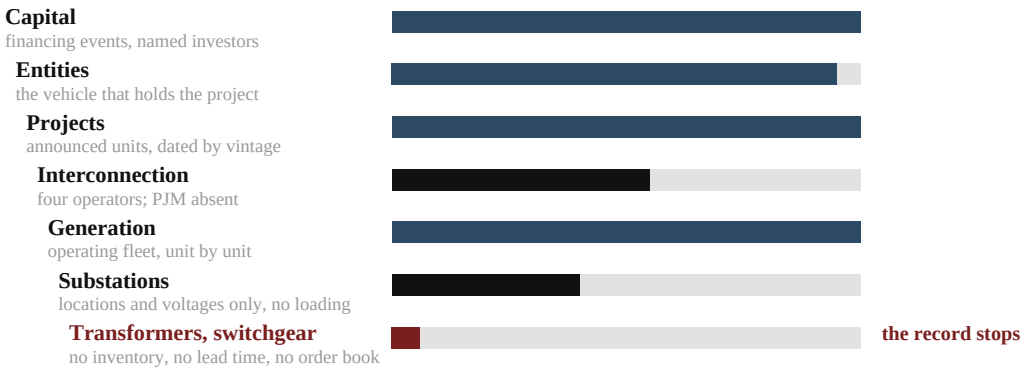
Now observe the position that instrument occupies. It models the interconnection queue — applications from data centres requesting permission to draw power from a constrained grid — by performing sustained computation, which draws power from that same grid. The demon that sorts the molecules is standing in the room whose temperature it is measuring. This is not metaphor and not paradox; it is arithmetic, currently negligible in magnitude and exact in kind. The discipline it implies — to treat one's own dissemination as an intervention in the data rather than as a neutral act of observation — is cheap to adopt now and impossible to reconstruct later.

An instrument that cannot represent its own presence is systematically wrong in a direction it cannot detect. The remedy is the one applied everywhere else here: make the limitation a first-class object and let it be queried.

11 Where the record stops

Follow the chain. Capital finances a project; the project is held through an entity; the entity acquires land and permits; it requests interconnection; interconnection requires a substation with capacity; the substation requires transformers and switchgear; the generation requires turbines. Only at the end does an operating asset exist. Resolution falls as the chain descends.

RESOLUTION FALLS AS THE CHAIN DESCENDS



Procurement silhouette, all that is public

46,434 awards · $12.11bn · 4,231 named firms across eight classifications

An award is not a delivery and a contract is not a lead time. The negative claim was searched rather than assumed: across federal statistics, operator planning documents, manufacturers' segment disclosures, trade associations and customs classifications, no public series gives order backlog in megavolt-amperes, quoted or realised lead time, plant utilisation, or allocation between named customers.

Source: Meridian observability ledger and federal procurement records, October 2026.

Figure 9. Resolution down the chain. Bar length is the share of the layer the estate can observe at entity level. The last row is not a low number; it is the absence of a series.

Federal procurement gives a silhouette and nothing more: 46,434 awards worth $12.11 billion to 4,231 named firms across eight classifications, of which transformer manufacturing is 8,936 awards and $1.845 billion, and switchgear 8,988 and $0.677 billion. It records an award, not a delivery; a contract, not a lead time.

Table 4. The procurement silhouette. Federal awards under the classifications on which the build-out depends. This is the whole of what the public record offers about the supply of heavy electrical apparatus.

Industrial classification	Awards	Value	Named firms
Power, distribution and specialty transformer mfg	8,936	$1.845bn	1,403
Switchgear and switchboard apparatus mfg	8,988	$0.677bn	1,106
Noncurrent-carrying wiring device mfg	8,999	$0.091bn	753
Current-carrying wiring device mfg	8,893	$1.228bn	1,216
Other communication and energy wire mfg	4,316	$0.504bn	575
Power and communication line construction	2,397	$6.105bn	937
Fiber optic cable mfg	2,357	$0.184bn	491
Turbine and turbine generator set mfg	1,548	$1.478bn	292

Note: an award is a contract signed, not an item delivered, and a firm may contract under one classification and manufacture under another. No column here is a lead time, because no public source carries one.

The negative claim behind this section was searched rather than assumed: any public series giving order backlog in megavolt-amperes, quoted or realised lead time, plant utilisation, or allocation between named customers, across federal statistics, operator planning documents, manufacturers' segment disclosures, trade associations and customs classifications. Found: trade volumes without lead times, backlog in currency without delivery dates. Not found: any of the four.

The layer is decisive rather than merely opaque. A constraint here does not slow one project; it sets the clock for a class of capital formation. If a transformer cannot be sourced inside two years, the interconnection request, the financing, the fund's deployment schedule and the hyperscaler's capacity plan are all subordinate to a manufacturing queue that appears in none of their documents.

It is structurally dark rather than accidentally dark: privately held, industrially unglamorous, and therefore unreachable by any foreseeable extension of disclosure law. Every other gap in this estate has a dataset behind it somewhere; this one does not. And the properties that make it dark make it defensible, which inverts the

assumption most data enterprises are built on. Where knowledge is public, assembly is a race and the advantage decays as tooling improves. Where it is structurally absent it can only be obtained the old way — expensive, slow, unscalable and therefore durable.

THE IGNORANCE LEDGER

OBSERVED	PARTIAL	BLOCKED	UNKNOWN
4 domains	10 domains	1 domain	5 domains
— UK beneficial ownership	— Ultimate control of US equity	— Gulf private ownership	— Transmission and substations
— US institutional holdings	— Swiss beneficial ownership		— Transformer and HV supply
— US private fundraising	— US registered-fund holdings		— Physical assets outside the US
— US operating generation	— Private-credit terms and marks		— Cross-border capital flows
	— OTC and dark-pool volume		— Performance and returns
	— Prices, including delisted		
	— Interconnection queues		
	— Large-load requests		
	— Supply chains and dependencies		
	— People across registers		

Twenty domains graded daily, each unknown carrying the route by which it might become observable or the finding that none exists. Transformer and high-voltage apparatus supply is the second kind of unknown: no public inventory, lead-time series or order book for it exists anywhere. The Gulf registers are scored blocked rather than unknown — the response to a robots-disallowed interface is a written request, not a circumvention.

Source: Meridian observability ledger, recomputed nightly; state as at 20 October 2026.

Figure 10. The ignorance ledger. Twenty domains graded daily, each unknown carrying the route by which it might become observable or the finding that none exists.

The discipline the ledger enforces is against a specific and almost invisible failure. A null renders as a blank; a blank reads as a zero; a zero enters an average. The path from we did not observe this to this did not happen is three careless steps long and entirely invisible once traversed. Borges imagined a cartography so complete that it coincided with the territory and was, for that very reason, useless. The discipline here is the opposite of that ambition: not a map the size of the world, but a map with its own holes drawn on it.

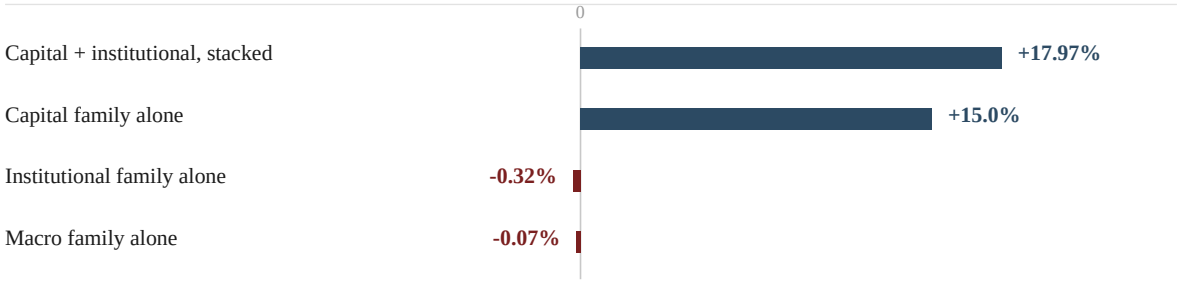
An instrument that publishes this about itself can be relied on for the domains it scores observed. The account of its own blindness turns out to be a map of where to go — and to build an epistemology of capital is, in the end, to be handed an epistemology of the builder.

12 What the instrument predicts, and what would falsify it

The measurements above are direct rather than modelled, which is their strength. The instrument also carries a prediction layer, and its limits bound the commercial claim.

WHAT PREDICTS, AND WHAT DOES NOT

Brier skill against the realised base rate, 90-day horizon, out-of-sample holdout



531,416 out-of-sample holdout predictions with outcomes attached

7.25% realised base rate on the scored population

0.6870 area under the curve, capital and institutional stacked

60,250 scored rows at the 90-day horizon

The outcome is an event in the disclosed record, not a return. Two of the four families sit at or below chance on their own and are reported at their measured value rather than withheld. The estate holds no survivorship-free price series and therefore cannot test whether any of these events are tradeable.

Source: Meridian one-model lab, walk-forward protocol with purged folds, scored 19 October 2026.

Figure 11. The prediction layer, reported at measured value. Brier skill against the realised base rate on the out-of-sample holdout, 90-day horizon. Two families sit at or below chance alone.

The estate holds 531,416 out-of-sample holdout predictions with outcomes attached across all horizons; the 90-day horizon carrying the figures below holds 60,250 scored rows, the remainder sitting at the 180- and 365-day horizons. Predictions are produced under a walk-forward protocol with purged folds. On the 90-day horizon the base rate is 7.25 per cent; the capital and institutional families stacked reach 0.6870 area under the curve with 17.97 per cent Brier skill. The institutional and macro families alone return -0.32 and -0.07 per cent, at and below chance. At the 180-day horizon the base rate is 11.13 per cent and the stack returns 15.93 per cent.

Two boundaries belong immediately beside those numbers. The outcome is an event in the disclosed record — whether an entity files again inside the horizon — and not a return; and predicting that a filer files again is a far weaker thing than predicting that a project is built, which is why the physical measurements in §3 to §6 and not this layer carry the paper.

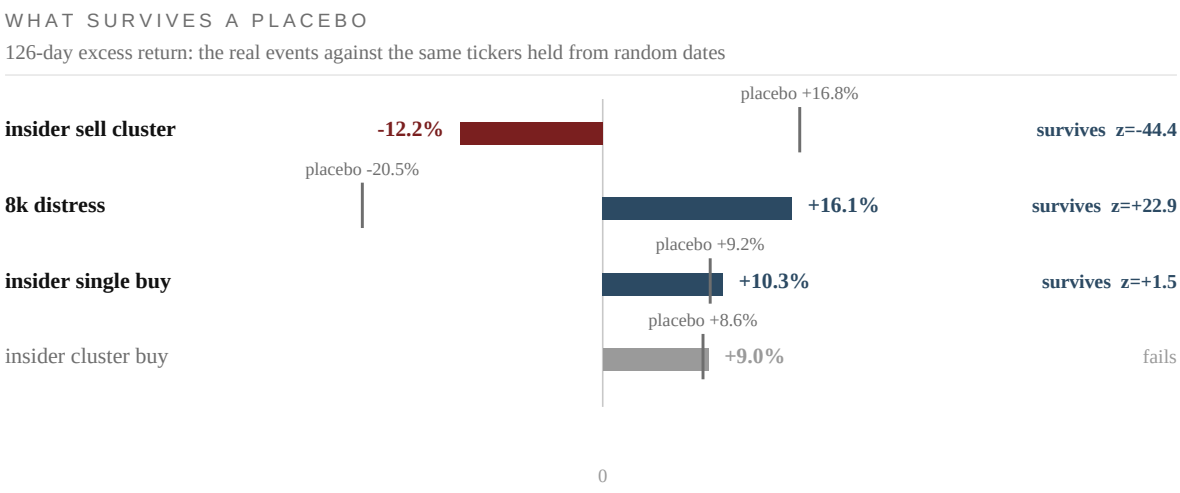
On the first of those boundaries the position has changed since the previous edition, and the way it changed is the most instructive result in this paper. A returns layer now exists: a price store of 10,422 securities and a signal ledger measuring forward return against a winsorised equal-weight universe index, for filing-dated events entered on the first trading day after the filing. Four signals carry enough observations to test, and on their face all four look strong. Before reporting any of them I ran the test that the last study in this estate failed.

The placebo holds everything constant except the one thing under test. For each real observation the event date is redrawn uniformly from the same calendar window, keeping the ticker and the per-ticker count fixed, and excess return is recomputed with identical entry, exit and benchmark machinery. Two hundred draws per cell. If a signal carries no information, its real mean excess should sit inside that placebo distribution.

Table 5. The placebo test. Real mean excess return against the distribution of the same tickers held from random dates in the same window. Two of the four signals do not survive it.

Signal	Horizon	Real excess	Placebo	z	Verdict
8k distress	21d	+2.35%	-4.59%	+6.9	Survives
8k distress	63d	+9.23%	-12.51%	+15.6	Survives
8k distress	126d	+16.14%	-20.51%	+22.9	Survives
insider cluster buy	21d	+1.24%	+1.28%	-0.1	Fails
insider cluster buy	63d	+4.38%	+3.83%	+0.7	Fails
insider cluster buy	126d	+9.01%	+8.56%	+0.5	Fails
insider sell cluster	21d	-1.97%	+2.99%	-17.5	Survives
insider sell cluster	63d	-5.53%	+8.55%	-30.2	Survives
insider sell cluster	126d	-12.17%	+16.81%	-44.4	Survives
insider single buy	21d	+1.27%	+1.50%	-0.8	Fails
insider single buy	63d	+5.04%	+4.37%	+1.3	Fails
insider single buy	126d	+10.25%	+9.17%	+1.5	Survives

Note: one residual asymmetry is not corrected. A ticker that delists loses its forward return, so real events close to a delisting drop out while placebo dates earlier in the window survive. This biases the real sample toward survivors — overstating the distress signal and understating the selling signal. Direction is known; magnitude is not.



Each bar is the mean excess return of the real events. The marker is the mean of two hundred placebo draws that hold the ticker fixed and move only the date. Where the two coincide, the signal is a property of which companies appear in the sample rather than of the event.

Source: Meridian signal ledger and price store, 200 placebo draws per cell, seed fixed, 25 October 2026.

Figure 12. The placebo separation at 126 days. Bars are the real mean excess return; the vertical marker is the placebo mean for the same tickers held from random dates in the same window. Two signals clear their own null by more than twenty standard deviations. Two sit on top of it.

Read the two columns together and a second correction becomes unavoidable, which is reported here rather than quietly fixed. The ledger stores each signal's excess return already multiplied by the direction of its thesis — it records what the trade earned, not what the stock did. The first edition of this paper read those figures as the stock's own excess return, which inverts the economic meaning of both short-thesis signals. Distress-flagged 8-K filers do not return plus sixteen per cent; they underperform by roughly that amount, and the positive number was the return to shorting them. Clustered insider selling is the mirror image: those stocks outperform, and the negative number was a short thesis losing money.

Correcting the sign makes the placebo result clearer rather than weaker, and it sets up the harder test. A placebo establishes that an event carries information its ticker does not. It does not establish that the information is not simply risk. Distressed filers and firms with insider activity are more volatile and carry different momentum than the universe average, so part of any measured excess is exposure rather than event.

The remedy is Barber and Lyon's: abandon the index benchmark and match each event to control firms on the characteristics that drive the exposure.

Each event is therefore matched to five control firms on pre-event realised volatility and momentum measured over the window from sixty to twenty-one trading days before the filing, and the abnormal return is taken against that matched portfolio. Significance uses the skewness-adjusted statistic of Lyon, Barber and Tsai, with p-values from a block bootstrap that resamples whole event dates rather than observations, so same-day events cannot be counted as independent.

Table 6. Every return signal re-measured against characteristic-matched controls, with a skewness-adjusted statistic and date-clustered bootstrap. The index column is the stock's own excess return, with the direction adjustment removed.

Signal	Thesis	vs index	vs matched	t	p	Verdict
8k distress	short	-17.65%	-15.15%	-6.03	0.0005	Survives
insider single buy	long	+9.26%	+2.33%	+2.67	0.0115	Survives
insider sell cluster	short	+11.19%	+2.01%	+2.57	0.0195	Survives As A Long
insider cluster buy	long	+8.06%	+1.85%	+1.41	0.1569	Fails

Note: insider sell cluster was carried as a short thesis. Matched, those stocks outperform, so the signal survives with its sign reversed — reported as it measured rather than as it was framed.

The matching is what matters. Three of the four edges shrink by between seventy-five and eighty-two per cent once volatility and momentum are controlled for: insider single buying falls from 9.26 per cent to 2.33, cluster buying from 8.06 to 1.85 and out of significance entirely. Most of what looked like insider information was the characteristics of the firms insiders buy.

One result holds. Shorting distress-flagged 8-K filers returns 15.15 per cent against matched controls, t = −6.03, bootstrap p = 0.0005, on 847 events across 326 distinct dates — a shrinkage of only fourteen per cent from the index measure. It survives the placebo, the matching, the skewness adjustment and the clustering correction. After four tests designed to kill it, it is the one thing in this paper I would call an edge.

That one result was then put through the questions a desk would ask before trading it. The dating objection this paper raises elsewhere turns out not to apply here: the distress events are dated at each 8-K filing, one event per filing, not at a month end. What does apply is whether the edge survives overlapping repeat filings, entry delay, and a universe you could actually borrow.

Table 7. The distress short under implementation constraints. Half of all events are repeat filings by the same issuer; removing the overlap costs 1.7 points and the result holds.

Test	n	BHAR	t	p
baseline (every filing)	756	-14.52%	-5.62	0.0005
first-only, no overlap	478	-12.82%	-4.35	0.0005
entry T+1 (as measured)	478	-12.82%	-4.35	0.0005
entry T+2	478	-13.67%	-5.03	0.0005
entry T+5	474	-14.26%	-5.02	0.0005
entry T+10	473	-13.15%	-4.42	0.0005
first-only, price >= $5	367	-11.76%	-3.58	0.0005

Note: survivorship runs in favour of this measurement rather than against it. The price store carries no delisting flag, so issuers that went to zero drop out of the sample — which understates how badly distressed filers do, and therefore understates the short.

The entry-lag rows are the ones that matter. If this were information the market absorbs on the day, delaying entry would destroy it. It does not: the abnormal return is −12.82 per cent entering the day after the filing, −14.26 entering five days later, and −13.15 at ten days. The underperformance accrues over the following months regardless of when you arrive in the first fortnight. That is not a microstructure artefact, and it means being first is not the edge.

On a borrowable universe — excluding names under five dollars, where a short is frequently unavailable at any price — the result is −11.76 per cent at t = −3.58 on 367 events. What this paper does not model, and what would decide the question in practice, is the cost of borrow in precisely this population. Distressed small capitalisations are where borrow is most expensive and least available, and no measurement here establishes

that the spread survives it.

I report this because the failures are the point. Had the placebo not been run, this paper would have carried four signals with t-statistics between 3.1 and 11.9, and two of them would have been wrong — which is precisely what happened to the insider-transaction study described in §7.

The honest scoreboard is three caught before publication and one caught after. The private-credit figures withdrawn in §9 went out in the first edition of this paper, and were found not by the instrument but by an outside reader asking whether the aggregate could be reconciled against the filers' own published totals. It could not. An error caught by a stranger is not evidence of a control; it is evidence of a missing one, and the reconciliation check it implies has since been added. I state the ratio rather than the flattering version because a paper that spends twenty pages on the ways an instrument deceives itself cannot round its own record up.

That qualitative claim is also measured. A standing audit scores each challenger against the capital stack alone on the same rows, with paired bootstrap intervals. The result is uncomfortable and belongs in the paper rather than in a drawer.

Table 8. Incremental value of each additional evidence family over the capital family alone. Correlations are between the challenger's predictions and the benchmark's, on identical rows.

Horizon	Added to the capital stack	Prediction corr.	Incremental Brier	P(improve)
90d	capital+institutional	0.999995	+0.0000126	0.984
90d	capital+macro	1.000000	-0.0000008	0.092
90d	meta:all	0.999995	+0.0000118	0.981
180d	capital+institutional	0.998787	+0.0001097	0.678
180d	capital+macro	0.999995	-0.0000386	0.038
180d	meta:all	0.999124	+0.0000922	0.783

Note: each audit is stored five times in the source table; the rows above are deduplicated. A negative incremental Brier means the addition made the forecast worse.

The challengers correlate with the benchmark at between 0.9988 and 1.000000. The capital family carries essentially all of the signal; the institutional family adds an amount that is statistically detectable at ninety days and economically meaningless; and the macro family is negative at both horizons, with a probability of improvement of 0.09 and 0.04. More data made the forecast worse, and the audit says so without being asked.

The same finding appears directly in a controlled experiment, and it bears on the claim this paper opened with. Against a within-corpus baseline of 0.709, a fourteen-column cross-corpus block added +0.0018, and a single binary indicator of whether the entity appears anywhere else recovered all of it. On the converged sub-population the estimate is negative with an interval of roughly ±0.021, which is to say that this test cannot tell rather than that the join fails there. So the claim that the binding constraint has moved from acquisition to integration does not hold generally, and I do not assert it generally. Assembly does not pay on questions one corpus already answers. It pays on questions no single corpus holds the state for — which is what §3 to §6 demonstrate, because no register contains both what was announced and what physically arrived.

A rate that cannot predict anything is a description, so the conversion curve is registered as predictions rather than promised as capability. The rule was fixed before the output was read: rank every unit still proposed in the 2025 vintage at 100 megawatts or more by the product of its technology and state conversion rates, take the twenty highest and twenty lowest distinct sites, and predict the first set operating and the second not. Set A is twenty sites and 5,904 megawatts, dominated by combustion turbines in Wisconsin, Georgia, Oklahoma and Missouri. Set B is twenty sites and 9,089 megawatts — offshore wind at Empire Wind 1, Sunrise Wind, Revolution Wind, Vineyard Wind 1 and Coastal Virginia, pumped storage in California and Utah, and New York onshore wind, solar and batteries. At the aggregate the same rate implies that of the 326.5 gigawatts proposed in 2025, approximately 152 will be operating by 2028 and approximately 175 will not. Resolution is mechanical: the 2028 vintage either records a unit as operable or it does not. These are the predictions the paper can lose, which is what makes registering them worth doing.

A claim that cannot specify what would refute it is not yet a scientific claim, whatever its sophistication. Three would refute this one: if Set A does not substantially outperform Set B in 2028, the constraint map describes 2022–25 and is not a property of the system; if the disclosed stratum proved economically inert, of which the insider-transaction collapse is the cautionary instance; and if the reconstructed state proved not to be

decision-relevant, because allocation here is dominated by relationship and by the tacit residue Polanyi describes. A fourth candidate must be relabelled rather than kept — if general-purpose systems made the join trivial for everyone at once, that would confirm the claim that integration was binding and then relieve it. It threatens the enterprise, not the thesis.

None of this prices anything. The estate holds no return series, so nothing here establishes that a constraint is mispriced rather than merely real. What changes is the search space: an allocator who can enumerate the 285 sites, the fourfold jurisdictional spread, the 1,066 gigawatts already withdrawn and the twentyfold demand growth is asking different questions from one who cannot, and those are answerable from public data in an afternoon rather than a quarter. The loop from hypothesis to capital decision to observed outcome is instrumented and has not run. That is the next paper.

13 Coda: attention as the scarce faculty

Wiener founded cybernetics on a single idea: that control is the communication of information in a feedback loop. The last loop any instrument of this kind must close — signal to action to outcome and back into the model that produced the signal — is not merely a missing dataset. It is the instrument's conscience. A perception that never learns whether it was right is not intelligence; it is observation with good posture.

There are three collapses to avoid: of data into knowledge, of knowledge into understanding, and of understanding into wisdom. A corpus is not knowledge until it is joined; a join is not understanding until it is set against outcome; and understanding is not wisdom until it is held with humility about what it cannot see.

Simon saw the real scarcity sixty years ago: a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. Attention is the single faculty that scale does not manufacture. Five billion records do not produce a gram of it, and the growth of this estate from half a billion to 5,058,265,273 in a year is an argument for Simon rather than against him. The machine's advantage is not intelligence in any deep sense; it is tirelessness at scale, and tirelessness is what buys the hours back.

It is tempting to say that Simon and the contemplative traditions are the same idea arriving by two roads. They are not, and the difference is the point. Simon's attention is a resource to be allocated efficiently. Weil's is the opposite of a resource: it is defined by wanting nothing from its object, and it is destroyed the moment it is spent for a return. One can be optimised. The other can only be made room for. The instrument cannot attend. It can only make attending affordable — and whether the cleared hours are then spent on attention in Weil's sense, or merely on more transactions, is not a question the instrument can answer. It is the only question that finally matters about it.

I have tried to admit a claim of this kind only where it names something implemented. The refusal to over-merge is a rule in the resolver; the ledger of ignorance is a table recomputed nightly; the 285 sites are a query against two public files. Where I could not point at the thing, the sentence is not here — and where a figure in the first edition could not be reproduced, it was restated under a stated definition or removed. Hayek was right that no mind holds the whole, and the machine does not hold it either: it holds a shadow, labelled as a shadow, with the shape of the missing parts recorded beside it.

> *Attention is the rarest and purest form of generosity.*
>
> — Simone Weil, 1942

Weil wrote that about something other than capital. But an instrument is an argument about where attention should go, and to see truly is to be changed by what one sees, and to be accountable to it. That is the discipline, and it is the only part of this that does not scale.

— T.D., Meridian Deal Partners, October 2026

Method, sources and interest

Every empirical figure in this paper was verified against the live estate on 19–20 October 2026 and is stated as of that date; the corpus grows daily. Where a figure could not be reproduced from the underlying store it was re-derived under a definition stated in the text or removed rather than softened. The conversion, technology and jurisdiction figures of §3 and §4 re-derived identically to the September edition one month after first measurement, from the same four vintages. The screen of §5 did not, and is restated there under an explicit

definition rather than compared.

The estate holds 5,058,265,273 records. Of these, 1,864,734,940 form the resolved estate — statutory registers, regulatory filings and official series from regulators in thirty jurisdictions, across 33 primary stores, joined on one identity spine and carrying every measurement in this paper. The remaining 3,193,530,333 are adjacent research corpora — scholarly citation, trademark, grant and procurement records — held for reference, not joined to the spine, and never added to the estate figure in an analytical claim. An overlap of 152,081,152 rows between stores is measured and removed before either figure is published, so the two components are already net of it and sum to the total rather than exceeding it. A record here is one row as its source register publishes it — a filing, a holding, a citation — not a resolved entity and not a deduplicated fact. Scale is not the argument: the strata carrying §3 to §6 are a small fraction of one per cent of the whole.

The conversion figures come from EIA Form 860, vintages 2022 to 2025, matched on plant code and generator identifier; substations from the HIFLD registry, re-collected 6 October 2026; queue and large-load figures from MISO, SPP, CAISO and NYISO, which exclude PJM because its interface requires a subscription and ISO-NE because it publishes no machine-readable queue. Lawrence Berkeley National Laboratory publishes an annual study measuring a different and earlier transition, from queue entry to commercial operation; it is the right external check on §3. Private-credit figures come from the schedules of investments filed by business development companies, at the reporting date stated in §9.

Identity is resolved against the Global Legal Entity Identifier register — 3,465,300 legal entities and 493,535 declared parent-child relationships across 235 jurisdictions — and the resolver asserts a link only where a normalised name matches exactly one candidate, holding the entity as ambiguous otherwise. Recall is 70.4 per cent on 4,000 sampled labelled pairs, which is a lower bound. Precision is 90 per cent on sixty hand-adjudicated links, 95 per cent interval 79 to 96, adjudicated by the author with no second rater and therefore no agreement statistic. All six failures were systematic, fusing either a US entity to a foreign namesake or an operating company to its own holding company; the normaliser had been discarding holdings, group and partners as legal furniture, and they are not furniture, they are precisely what distinguishes an operating company from the entity that owns it. After the correction 688 links were withdrawn and 588 asserted, and ambiguous holdings fell from 1,252 to 902.

Several commercial systems assemble overlapping material — Orbis, Capital IQ, PitchBook, Preqin, FactSet, Sayari and the compliance-screening industry — and several do it better, at greater coverage, with larger teams. Four differences are claimed and none is a claim of superiority: bitemporality on the physical record, noting that point-in-time company financials are already an established product and that the distinction claimed is one of domain rather than of technique; the physical layer, which no commercial buyer has asked them to build; the published ledger of ignorance, which runs against the incentive to present completeness; and the refusal to over-merge, which costs coverage and is reported rather than hidden. The forty named units of §12 are published as a companion sheet, with the machine-readable version beside it.

Interest. The author operates the instrument described and intends to commercialise research derived from it. Nothing here is investment advice, no security is recommended, and no claim is made that any constraint identified is mispriced.

References

Bennett, C. H. (1982). The Thermodynamics of Computation — a Review. International Journal of Theoretical Physics, 21(12), 905–940.

Borges, J. L. (1946). On Exactitude in Science. In Collected Fictions.

Coase, R. H. (1937). The Nature of the Firm. Economica, 4(16), 386–405.

Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. American Economic Review, 70(3), 393–408.

Hawkes, A. G. (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. Biometrika, 58(1), 83–90.

Hayek, F. A. (1945). The Use of Knowledge in Society. American Economic Review, 35(4), 519–530.

Hayek, F. A. (1968). Competition as a Discovery Procedure. In New Studies in Philosophy, Politics, Economics and the History of Ideas (1978).

Landauer, R. (1961). Irreversibility and Heat Generation in the Computing Process. IBM Journal of Research and Development, 5(3), 183–191.

Lawrence Berkeley National Laboratory (2025). Queued Up: Characteristics of Power Plants Seeking Transmission Interconnection.

North, D. C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge University Press.

Pearl, J. (2009). Causality: Models, Reasoning, and Inference (2nd ed.). Cambridge University Press.

Polanyi, M. (1966). The Tacit Dimension. University of Chicago Press.

Popper, K. R. (1959). The Logic of Scientific Discovery. Hutchinson.

Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27, 379–423.

Simon, H. A. (1971). Designing Organizations for an Information-Rich World. In Computers, Communications, and the Public Interest. Johns Hopkins University Press.

Snodgrass, R. T. (1999). Developing Time-Oriented Database Applications in SQL. Morgan Kaufmann.

Taleb, N. N. (2007). The Black Swan: The Impact of the Highly Improbable. Random House.

Weil, S. (1942). Letter to Joë Bousquet, 13 April 1942. In Seventy Letters, trans. R. Rees (1965). Oxford University Press.

Wiener, N. (1948). Cybernetics: or Control and Communication in the Animal and the Machine. MIT Press.

Williamson, O. E. (1985). The Economic Institutions of Capitalism. Free Press.

Primary sources. US Energy Information Administration, Form EIA-860, vintages 2022–2025. US Department of Homeland Security, HIFLD electric substations. Interconnection and large-load queue publications of MISO, SPP, CAISO and NYISO. US Securities and Exchange Commission EDGAR, including Form D, Forms 3/4/5, Schedules 13D and 13G, Form 13F, Form N-PORT and the schedules of investments filed by business development companies. USAspending.gov federal award records. Global Legal Entity Identifier Foundation golden copy.