

Measuring the Physical Side of Creative Destruction: A Point-in-Time Dataset of AI Data-Center Sites, Air Permits, Substations and Grid Demand

Dated physical records for research on how new technology turns into installed capacity

Jennie Chu

Gemral Edge · support@gemral.com

ABSTRACT

The 2025 Prize in Economic Sciences recognised growth through technological progress and creative destruction. For artificial intelligence, the step from new technology to installed capacity runs through buildings, generators, substations and the electricity grid, and most of those steps leave public records well before they appear in company accounts. We describe a point-in-time dataset that collects those records for 49 U.S.-listed AI-infrastructure companies: data-center sites and owners, dated air permits with permitted backup-generator capacity, nearest high-voltage substations, daily regional grid demand, power-plant status changes, equipment imports, quarterly capital expenditure and monthly satellite measurements. Every record carries the date it became public and the owner that held the site on that date. We show what this changes: a join through today's owners mislinks 24% of permit events, a literal reading of permit listings overstates new permits, and new generating capacity is visible a median 16 months before it operates. Twelve record layers pass accuracy checks at 100%.

Keywords: creative destruction; general purpose technology; data centers; electricity grid; physical capacity; point-in-time data; remote sensing

JEL codes: O33, L94, Q41, C81, G14

1. Introduction

In October 2025 the Royal Swedish Academy of Sciences awarded the Prize in Economic Sciences to Joel Mokyr "for having identified the prerequisites for sustained growth through technological progress" and to Philippe Aghion and Peter Howitt "for the theory of sustained growth through creative destruction" (Royal Swedish Academy of Sciences, 2025). In the Aghion and Howitt (1992) model, growth comes from innovations that replace older technologies, and the pace of replacement depends on investment by entrants and incumbents. Mokyr (2002, 2016) emphasises that new knowledge becomes growth only when it can be turned into working machines and installed at scale.

Artificial intelligence is a sharp test of both ideas. It has the hallmarks of a general purpose technology (Bresnahan and Trajtenberg, 1995): it spreads across sectors and drives complementary investment. Brynjolfsson, Rock and Syverson (2021) show that such technologies require large complementary investments that standard measures capture late or not at all. For AI, a large part of that complementary investment is physical. Models run in data centers, data centers need electricity, and electricity requires generation, substations and transmission. Each of these steps takes time, and each is subject to permitting.

The 2026 Prize, awarded to Robert E. Hall and Michael Woodford for the microeconomic foundations of macroeconomic stabilisation, adds a further reason to measure this side of the economy directly. In Woodford's (2003) framework, the output an economy can sustain without inflationary pressure, its natural output, depends on productive capacity, and sudden changes in capacity act as supply shocks (Royal Swedish Academy of Sciences, 2026). Physical capacity for computing and for electricity is part of that productive base.

Financial statements see this investment late and in dollars. Quarterly capital expenditure is reported after the quarter ends, mixes equipment prices with quantities, and does not say where capacity is being built or what constrains it. Public physical records answer different questions: where a site is, what the owner has been permitted to install, how close the site is to high-voltage infrastructure, and how demand on the local grid is moving. This is the same logic that led researchers to measure growth with night-time lights (Henderson, Storeygard and Weil, 2012) and to use remote sensing more broadly in economics (Donaldson and Storeygard, 2016). Katona, Painter, Patatoukas and Zeng (2025) show that once satellite data on retailers became available, it changed who was informed in the market, a reminder that physical measurements carry information that accounts do not.

This paper describes a dataset built to put those physical records on a single point-in-time timeline linked to listed companies. Section 2 sets out what the theory asks of a measurement. Section 3 describes errors in naive construction. Section 4 describes the architecture. Section 5 reports three measurements of what the construction changes. Section 6 reports accuracy checks, Section 7 outlines research designs, Section 8 gives scope and definitions, and Section 9 covers access.

2. What the theory asks of the data

R1. Physical units, not only dollars. Creative destruction is about capacity replacing capacity. A measurement should record capacity in physical terms, such as permitted generator megawatts, distance to a high-voltage substation and grid demand, alongside dollar spending.

R2. Public dates. Each record must carry the date it became public. A permit issued in March but posted in May belongs to May in any study of what was knowable.

R3. Owner at the time. A site must be linked to the company that controlled it on the record's publication date. Data-center campuses change hands, and many are held through subsidiaries or

joint ventures whose names do not reveal the listed parent.

R4. Coverage of the bottleneck. The constraint on installed AI capacity is often outside the building: generation, substations, grid demand and equipment supply. A useful dataset links the site to those constraints, not just to its owner.

3. Errors in naive construction

3.1 Reading permit listings literally

Public permit listings often show only the most recent version of a facility's permit. When that version is higher than the first, the facility already held an earlier permit that the listing no longer shows. Treating the first version one happens to see as a new permit counts modifications as new capacity. In our Virginia data, correcting this relabelled 132 of 183 events that a literal reading would have marked as new; the count of new permits over twelve months fell from 48 to 14 (measured 19 October 2026). Without the correction, a series of new permits overstates new build activity by more than three times.

3.2 Treating backup capacity as compute capacity

Air permits for data centers record the permitted capacity of emergency backup generators. That figure is an upper bound on on-site backup capacity. It is not the electrical load of the computing equipment, and converting one into the other requires assumptions that the permit does not support. The dataset keeps permitted backup capacity under its own label and keeps non-emergency generation separately, so that the two are not added together.

3.3 Linking events through today's owner

The simplest way to attach a permit to a company is to look up who owns the site today. For sites that changed hands, or whose link to a listed parent was established by a later public document, this attaches the event to a company too early or to the wrong one. Section 5.1 measures how often this happens.

3.4 Reading satellite indices without context

Satellite built-up indices are seasonal and respond to land clearing as well as construction. In arid regions, bare ground can already score like built land. A single index read without seasonal comparison can miss a campus that is growing or flag cleared land as new buildings. The dataset adds a bright-roof share to confirm finished roofs. In hand checks, the bright-roof share at an Arizona campus rose from 3.5% to 17% of the measurement disc between summer 2023 and summer 2025 as new buildings were completed.

4. Point-in-time architecture

4.1 Coverage

The universe is 49 U.S.-listed companies in seven groups: hyperscalers, AI clouds, data-center operators, chips and servers, optical networking, power, and cooling. For these companies the dataset provides the layers listed in Table 1. Sources are described by category: U.S. state environmental permit records, U.S. county environmental permit records, U.S. county land-use and building records, U.S. federal environmental compliance records, U.S. federal energy and electricity statistics, U.S. federal infrastructure reference data, U.S. federal international trade statistics, U.S. securities-regulator public filings, public satellite imagery and open map data.

Table 1. Layers and coverage (measured 18–20 October 2026).

Layer	Coverage
Data-center sites and owners	554 sites; owner identified or owner type stated for 489 (88.3%)
Dated air permits	Virginia permits with permitted backup-generator MW for 107 of 116 primary sites (92.2%); 171 of 194 permits carry MW. Dated state permit records for Georgia (46), Arizona (15) and Texas (34); permitted MW for 8 Georgia sites
Federal air-program registrations	79 of 81 primary sites in Texas, Georgia, Ohio and Arizona
Nearest substation of 115 kV or more	629 of 676 sites within 5 km; median distance 1.02 km
Daily regional grid demand	250,000+ daily rows, 2019 onward
Power plants and monthly status changes	2021 onward, owner linked point in time
Equipment imports	73,165 rows: 17 six-digit product codes by origin country, January 2018 to August 2026
Quarterly capital expenditure	2,332 filings for 44 of 49 tickers, with filing dates
Monthly satellite measurements	2023 onward, per site
County land-use and building records	239 permitted buildings, 139 data-center parcels and 85 planned project areas, with version history

4.2 Timestamps

Every table carries `source_published_at`, `first_seen_at`, `is_backfill` and `record_hash`. Where a source gives no publication time, the dataset assigns one that can only be late, never early: a grid day becomes available at 00:00 UTC on the following day, an import month at month end plus 40 days, and a satellite period at the first day of the next period plus three days. Capital

expenditure is keyed by ticker, period and filing date, and every amendment is kept as a separate row, so a point-in-time query uses only filings published on or before the evaluation date.

4.3 Owners over time

Each link from a site or power plant to a ticker carries a `valid_from` date taken from the public document that establishes it: a subsidiary list attached to an annual report, a permit issued in the owner's name, or a press release. Point-in-time links for site events are built in the resolved event table, where each event carries the ticker valid on its publication date. An event published before the owner link was established carries no ticker.

4.4 Linking sites to their constraints

Each site is linked to its nearest substation of 115 kV or more, with distance, and to the grid zone whose daily demand it draws on. Power plants carry monthly status changes (planned, under construction, operating, retired) on the same timeline. These links let a researcher ask not only who is building, but whether the grid near the site has the capacity to serve it.

5. Three measurements of what the construction changes

5.1 Table 2: Events misattributed by a today's-owner join

We compared the ticker each site event carries at its publication date with the ticker obtained by joining the same event to today's site owner.

Table 2. Today's-owner join versus owner at publication (measured 24 October 2026).

Event type	Events at sites with a listed owner today	Ticker attached too early or to the wrong owner
Air-permit events	169	40 (23.7%)
Power-plant status changes	687	70 (10.2%)

For nearly one permit event in four, the company that owns the site today was not the listed owner on the day the permit became public. A study that credits those permits to today's owner credits capacity to a company before it controlled the site.

5.2 Table 3: How early new capacity is visible

Power-plant status changes show when new generating capacity enters the planning pipeline and how long it takes to operate. We measured, for every plant in the monthly records, the time from its first appearance in each status to its first appearance as operating.

Table 3. Time from first public status to operation, U.S. power plants (records from February 2021; measured 24 October 2026).

Transition	Plants	Median	Interquartile range
Planned → operating	1,118	16 months	9–25 months
Under construction → operating	2,884	8 months	4–13 months
Planned, not yet operating	1,398	—	—

New generating capacity is therefore visible in public records more than a year before it runs, and 1,398 planned plants are in the pipeline today. Because the records begin in February 2021, plants already planned by then enter at that date, so the measured durations are if anything shorter than the true ones.

5.3 Table 4: Depth of linked public records per site

The value of a physical dataset lies in joining several record families to the same site. Table 4 counts how many primary data-center sites carry each layer, and how many carry all five at once.

Table 4. Linked layers per primary data-center site (549 sites, excluding off-grid crypto-mining sites; measured 24 October 2026).

Layer	Sites
Nearest substation of 115 kV or more	487
Grid zone with daily demand	484
Owner entity resolved	284
Dated air permit or federal air-program registration	264
Monthly satellite measurements	183
All five layers	107 (92 in Virginia)

Each layer comes from a different public record family with its own identifiers and formats. Joining them requires geocoding, owner resolution through subsidiaries and permit names, and dating each record to its publication. The 107 sites that carry all five layers are where a researcher can follow capacity from permit to grid in one place.

6. Accuracy checks

Each linked layer is sampled at random and checked by hand against the original public document. Table 5 reports the samples.

Table 5. Precision of linked records (samples dated 18–20 October 2026).

Layer	Correct / checked
Site owner	173 / 173
Power-plant owner	200 / 200
Site permit	101 / 101
Site substation	157 / 157
Site grid zone	163 / 163
Grid-zone daily demand	200 / 200
Imports by product code	96 / 96
Imports by product code and origin country	60 / 60
Site satellite measurement	200 / 200
Permitted backup MW	58 / 58 (2 not checkable)
Quarterly capital expenditure	60 / 60
County data-center records	60 / 60

All twelve layers scored 100%. The check confirms that the dataset reproduces the public record faithfully.

Expanded accuracy check (24 October 2026). Every record type is checked against the original public documents before it ships. In the latest graded batch, all 13 record types scored between 95.6% and 100% precision, with at least 500 graded rows per type across batches (site permits: the full set), Wilson 95% intervals, and a blind second grader on at least 10% of rows. Cumulative figures are lower for four types (contracts 91.2%, pre-award notices 76.4%, subcontractor links 94.0%, bid protests 92.5%) because they include batches graded before the entity resolver was fixed.

Attribution coverage. Edge attributes 822 power-plant status events and 807 foreign arms-sale notices to the listed parent company valid on the day each record became public. Exact matching on the raw operator or contractor name attributes 0 and 6 of them. This measures coverage, not return predictability.

7. Research designs the data supports

- 1. Diffusion speed and bottlenecks.** Measure how long the step from permit to operating capacity takes, by region and by owner, and whether proximity to high-voltage substations or growth in local grid demand shortens or lengthens it. This puts numbers on the physical prerequisites that Mokyr's account of growth emphasises.
- 2. Creative destruction in the power sector.** Follow plant retirements and new plants on the same timeline as data-center permits to study how new electricity demand interacts with

the replacement of older generation, in the spirit of Aghion and Howitt (1992) and Aghion, Antonin and Bunel (2021).

3. **Physical versus financial investment.** Compare the timing of permits and site events with the timing of reported capital expenditure for the same companies. Because both are dated by publication, the comparison uses only information available at each date.
4. **Capacity and supply.** In Woodford's (2003) framework, changes in productive capacity act as supply shocks. Regional grid demand, new generation and permitted backup capacity give researchers direct, dated measures of the capacity side of the AI build-out.
5. **Equipment supply chains.** Use monthly imports of key equipment by product code and origin country as industry context for timing and supply constraints.

In each case the dataset supplies dated, owner-linked physical records. Outcome measures come from the researcher's own sources.

A pre-registered validation (October 2026). Edge maps U.S. federal international trade statistics, by product code, to the listed companies that sell those goods, using only data published before each company files. Across 12 AI hardware suppliers (NVDA, AMD, AVGO, MU, DELL, HPE, ANET and others), import growth in each company's product codes moved with its same-quarter revenue growth: $r = 0.41$ over 174 company-quarters from 2023 to 2026 (95% CI 0.21–0.57, permutation $p = 0.001$). The link held in both halves of the AI era and for all 12 names. The test was pre-registered and run once. Imports are industry-level context by product code and origin country, not company-level shipments. What the series delivers is an early, independent read on the revenue trend weeks before reported numbers. It does not replace company guidance: most of the signal overlaps with each company's prior-quarter growth.

8. Scope and definitions

- **Permitted backup megawatts** are an upper bound on on-site backup generator capacity. They are not the electrical load of computing equipment, and the dataset does not convert them.
- **Imports** are industry context by product code and origin country. They are not shipments to any named company.
- **Permit versions.** Where a listing shows only the latest permit version, earlier versions and their issue dates are recorded as modifications, as described in Section 3.1.
- **Satellite measurements** should be compared season with season, and are best read together with the bright-roof share in arid regions (Section 3.4).
- **County land-use and building records** cover one county from 20 October 2026 and reproduce every field as the issuing county publishes it, with version history. Floor area is not converted to power.

9. Data availability, licensing and conflict of interest

The dataset is available to research teams as Parquet files with a manifest (row counts and checksums per file), a data dictionary, a REST API and a server for AI assistants, on the same point-in-time timeline as the companion dataset of federal procurement and disclosure records (Chu, 2026). Sample packs are available on request at support@gemral.com. Methods are documented in the public due-diligence questionnaire at gemral.com/edge-ddq.md. All records are drawn from public sources, and the dataset contains no material non-public information. Detailed source documentation is available to a licensee's compliance team under a non-disclosure agreement.

Contains modified Copernicus Sentinel data 2023-2026. Map data © OpenStreetMap contributors.

The author is the founder of Gemral Edge, which produces and licenses the dataset described here. This paper reports no return, performance or forecasting results.

References

- Aghion, P., and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2), 323–351.
- Aghion, P., Antonin, C., and Bunel, S. (2021). *The Power of Creative Destruction: Economic Upheaval and the Wealth of Nations*. Cambridge, MA: Harvard University Press.
- Bresnahan, T. F., and Trajtenberg, M. (1995). General purpose technologies: "Engines of growth"? *Journal of Econometrics*, 65(1), 83–108.
- Brynjolfsson, E., Rock, D., and Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1), 333–372.
- Chu, J. (2026). Dating government demand: A point-in-time, entity-resolved dataset of U.S. federal procurement and disclosure records. Gemral Edge working paper GE-WP-2026-01.
- Donaldson, D., and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4), 171–198.
- Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2), 994–1028.
- Katona, Z., Painter, M. O., Patatoukas, P. N., and Zeng, J. (2025). On the capital market consequences of big data: Evidence from outer space. *Journal of Financial and Quantitative Analysis*, 60(2), 551–579.
- Mokyr, J. (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton, NJ: Princeton University Press.
- Mokyr, J. (2016). *A Culture of Growth: The Origins of the Modern Economy*. Princeton, NJ: Princeton University Press.
- Royal Swedish Academy of Sciences (2025). *The Prize in Economic Sciences 2025*. Press release, 13 October 2025. Stockholm.
- Royal Swedish Academy of Sciences (2026). *Microfoundations deliver macro insights*. Popular science background to the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2026.

Stockholm.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.