Capacity versus Resiliency

The Hidden Air Permit Trade-Off for Data Center Backup Generators

Garrett Jansma

Picture a 300-megawatt data center campus at two in the morning. Every rack is humming. Every server is processing transactions for millions of users who have no idea the regional grid went dark six hours earlier. One hundred diesel generators throb in unison outside the facility walls carrying the full electrical load. The campus is performing exactly as designed. The load-transfer protocol worked flawlessly. Fuel reserves are ample. The facility is largely unstaffed, and compliance with its air permit is monitored remotely, calculated indirectly through fuel consumption rather than measured at a smokestack. Somewhere in that indirect calculation, though, a number is climbing toward a ceiling that almost nobody ever thinks about: the annual emissions budget written into the facility's air permit.

Last year, Larsen and Lillie warned in these pages that a large data center losing utility power for too long could out-run its backup capacity, "essentially like a car running out of gas in the fast lane of an interstate." Lainey Larsen & Andrew Lillie, *Personal Power Plants: On-Site Generation for Data Centers*, 40 Nat. Res. & Env't, no. 1, Summer 2025, at 1, 2. The analogy is vivid and deserves sharper definition for present purposes. In this context, the "gas" is not diesel in a tank. The constraint is legal, not physical: an annual emissions budget written into an air permit; a fixed and nonrenewable tonnage ceiling on how much a generator fleet may emit in a calendar year. Fuel can be resupplied by truck. The emissions ceiling cannot be refilled midyear. Once it is spent, it is spent. The facility is not running out of gas. It is running out of time.

Consider the math in concrete terms. Our hypothetical 300-megawatt campus holds a minor source air permit with an annual cap of 100 tons per year (tpy) of nitrogen oxides (NOx). NOx is the limiting pollutant because diesel engines emit it at a higher rate than other regulated pollutants. Regulators typically do not meter NOx directly. Instead, they work backward from the 100-tpy cap and the chosen engine model's emissions specifications to calculate a maximum fuel-throughput value: the total gallons of diesel the fleet may burn in a year before NOx emissions exceed 100 tpy. That fuel-throughput figure converts into a total hours-of-operation allowance across the fleet. For our hypothetical data center, that works out to roughly 700,000 gallons of fuel and 5,000 generator-hours, or 100 generators multiplied by 50 hours each. Fuel consumption, not a NOx sensor, is the compliance proxy that regulators and operators rely on. The facility tracks and reports that fuel consumption in periodic compliance reports, which the regulator uses to tally cumulative emissions against the annual cap.

Fifty hours. That's the emissions runway for our hypothetical data center campus: the total emergency runtime the permit can support before the NOx cap is exceeded. Major grid disruptions caused by hurricanes, ice storms, and wildfire-driven public safety power shutoffs have lasted days, and sometimes weeks, in recent years. A campus with a 50-hour emissions runway faces a hard ceiling well short of those durations. This article walks through why that number exists, why it shrinks as campuses grow, and two strategies that operators can pursue on their own: cleaner generator technology and larger permitting pathways. The article then turns to a third and different kind of fix: regulatory improvements the industry may be able to shape.

The Permit Behind the Generators

Emergency generators do not operate in a regulatory vacuum. They are stationary sources of air pollution, subject to permitting under the Clean Air Act's Title V and New Source Review (NSR) programs 42 U.S.C. §§ 7661–7661f and 7475, and further governed by federal engine emissions standards, including New Source Performance Standards (NSPS) and the National Emission Standards for Hazardous Air Pollutants for reciprocating internal combustion engines (the RICE NESHAP). 40 C.F.R. pt.

Published in *Natural Resources & Environment* Volume 41, Number 2, Fall 2026. © 2026 by the American Bar Association. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

60, subpts. IIII, JJJJ; *id*. pt. 63, subpt. ZZZZ. The applicable standard depends on engine type, but the mechanics are the same.

Under these frameworks, engines classified as "emergency generators" receive more lenient emissions standards in exchange for operational restrictions. No single federal definition of "emergency" applies uniformly; Virginia and other states define the term as a sudden, unforeseeable loss of primary power beyond an operator's control. *See, e.g.*, 9 Va. Admin. Code § 5-530-20. The principal federal restrictions cap emergency use at 500 hours and nonemergency use at up to 100 hours per year of testing & maintenance (T&M), of which up to 50 hours may be used for nonemergency and non-T&M purposes or go to narrowly defined, reliability-driven demand response. Memorandum from John S. Seitz, Dir., Office of Air Quality Planning & Standards, U.S. EPA, *Calculating Potential to Emit (PTE) for Emergency Generators* (Sept. 6, 1995); 40 C.F.R. pt. 60, subpt. IIII; *id*. pt. 63, subpt. ZZZZ. These are per-engine limits and do not override facility-wide emissions caps imposed by individual air permits.

The hours and the facility-emissions limits are independent constraints. While federal rules may authorize as many as 100 hours of T&M, a facility's air permit—through its NOx cap and fuel-throughput limit—often will not allow that many hours before the tonnage cap binds first. T&M hours consume part of the same annual emissions budget that emergency runtime draws from. So, if routine testing burns through even a modest share of the cap, the runway available for a real outage shrinks accordingly.

The mechanism is straightforward and applies whether a facility holds a minor or major source permit. Only the size of the ceiling changes. Many permits cap total annual emissions of regulated pollutants on a mass basis, in tons per year. Each hour a diesel generator runs, it burns a quantifiable volume of fuel, and that fuel burn converts—hour by hour—into a quantifiable mass of NOx, along with particulate matter, carbon monoxide, and volatile organic compounds. The fleet's cumulative runtime, multiplied by per-engine emission rates, produces the total annual mass. Once that mass reaches the permit cap, further operation exceeds the permit. The cap does not reset during an emergency. It does not pause during a declared grid failure. It does not flex upward because a hurricane knocked out transmission lines across three counties. It simply is.

Return to the hypothetical: one hundred 3-MW generators sharing a 100-tpy NOx budget yielding roughly 50 hours of collective runtime. Shrink the campus to a single 3-MW generator, and that lone engine could run for thousands of hours before approaching the same cap: abundant resiliency, negligible capacity. Double the fleet (and data center capacity) to 200 generators, and the same fixed budget, along with its corresponding 700,000-gallon, 5,000-hour allowance, now divides across twice the equipment. Runtime per generator collapses by 50% to 25 hours, leaving even less time for an actual emergency.

A permit's tonnage cap does not care how many generators are pulling from it. It only cares how much they burn, together, before the year runs out. This inverse relationship is not a new concept. Practitioner commentary has discussed the mechanics by which hours-based limits become easier to exceed during prolonged outages at facilities with large generator fleets. Susan Brancaccio, *Frequently Asked Air Questions (FAAQs): Understanding Data Center Emergency Generators and Their Air Permits*, Williams Mullen (June 2, 2026). The structural trade-off underlying that mechanic has received less attention: Capacity and resiliency are inversely related once a facility operates under a fixed annual cap because adding generators to grow capacity does not enlarge the permit's fixed budget. Every generator added to gain capacity is one that shortens the emergency runway.

Where Capacity Planning and Runtime Planning Diverge

Data center developers may prioritize capacity for entirely rational reasons. Land entitlements take years to secure. Utility interconnection queues now stretch out for years. Consolidating capacity at a single site carries its own logic: Finding parcels to host five separate 100-MW facilities is often difficult, permitting five sites is far more cumbersome than permitting one, and splitting capacity across locations sacrifices the economies of scale a single large campus offers. Meanwhile, surging demand for AI training and inference workloads means every megawatt of capacity brought online today earns revenue immediately. Against those incentives, the air permit's annual emissions ceiling can recede from view in project planning. Resiliency carries no comparable line item. It produces no return unless the specific event for which it exists actually occurs. And because emergency outages long enough to test the runway are rare, the ceiling that measures that runway is abstract, invisible during normal operations, and relevant only during an event everyone hopes never happens.

Capacity and resiliency are inversely related once a facility operates under a fixed annual cap because adding generators to grow capacity does not enlarge the permit's fixed budget.

Yet even the hypothetical's baseline of 50 hours already looks thin against real-world outage durations, before any capacity growth occurs. FEMA's P-1019 guidance notes that most building codes require only about 90 minutes of emergency power for basic life-safety systems, far short of what continued operation requires, though critical facilities in high-seismicity areas face a more demanding requirement of up to 96 hours of onsite fuel storage. The ANSI/TIA-942 data center standard associates

Published in *Natural Resources & Environment* Volume 41, Number 2, Fall 2026. © 2026 by the American Bar Association. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

higher resiliency tiers with progressively greater on-site fuel capacity, ranging from roughly 12 hours at a Basic tier to 96 hours at a Fault Tolerant tier. Training by Capitoline, *How Much Diesel Fuel Does a Data Centre Need?* (Apr. 7, 2020). A campus that plans for fault-tolerant resiliency but holds a permit supporting only 50 hours has a real gap between aspiration and regulatory ceiling.

A data center company may establish a minimum runtime standard (say 50 hours) and design its fleets around it. The harder problem is what happens next: steady pressure to shrink that standard as capacity targets climb since expanded capacity is directly monetizable while a wider resiliency margin is not, at least not until the outage arrives. Absent a sustained commitment to hold the line, capacity pressure erodes the buffer between a campus's permitted runway and the outage it actually may face.

Some operators already deploy Tier 4 generators (cleaner generators) voluntarily, even where regulation does not yet require it.

A runtime standard is only defensible if someone actually planned for the outage it is supposed to survive. The enforcement implications of that planning gap are real. EPA's Clean Air Act Stationary Source Civil Penalty Policy (Oct. 25, 1991) explains that penalty assessment must reflect the factors listed in Clean Air Act § 113(e) (42 U.S.C. § 7413), including the violator's full compliance history and good faith efforts to comply, the duration and seriousness of the violation, and such other factors as justice may require. Subsequent EPA civil-penalty guidance and annual influence adjustments have preserved that framework. *See* U.S. EPA, *Policy on Civil Penalties: EPA General Enforcement Policy #GM21* (Feb. 16, 1984); Memorandum from David M. Uhlmann, U.S. EPA, to Regional Admins., *Amendments to the EPA's Civil Penalty Policies to Account for Inflation* (effective Jan. 15, 2024) (Jan. 10, 2024). In practice, those factors tie penalty calculation and enforcement posture to the reasonableness of an operator's prior compliance planning and its response to reasonably foreseeable risks of noncompliance. A facility that never modeled extended-outage emissions against its permit cap is in a weaker position than one that identified the cap as a constraint, evaluated outage scenarios and alternatives, and documented its decision-making, even if both facilities ultimately exceed the cap during the same grid failure. Under EPA's penalty policies, regulators examine whether the risk was reasonably anticipatable and whether the operator's planning and response can fairly be characterized as reasonable in light of its compliance history and good-faith efforts, and the economic benefit it realized from delayed or avoided compliance.

Two Levers Within a Single Operator's Control

The hypothetical's 50-hour figure is not destiny. It is the product of two variables: a fleet's per-engine emission rate and a permit's annual emissions cap. Change either one, and a campus can spend the resulting headroom on more runway, more capacity, or some blend of both. Two ways of changing the underlying variables sit entirely within an individual operator's control: technology selection and permitting strategy. A third path—reshaping the regulatory framework itself—would likely require industry-wide engagement.

The first strategy involves choosing cleaner generators and deploying generators equipped with Tier 4–equivalent emissions controls, including in jurisdictions where Tier 4 is not yet required and Tier 2 generators would satisfy applicable requirements. This is a design decision: Operators choose Tier 2 or Tier 4 generators before construction begins, and the choice matters because of emission rates. Tier 4 units employ selective catalytic reduction (SCR) for NOx, diesel particulate filters for particulate matter, and diesel oxidation catalysts for carbon monoxide, and (for nonroad generator sets above 900 kW) emit NOx at up to roughly one-tenth of the Tier 2 unit (0.67 versus 6.4 g/kW-hr). *See* Off. of Transp. & Air Quality, *Nonroad CompressionIgnition Engines: Exhaust Emission Standards, EPA420B16022*, tbl. (Mar. 2016). Those are certification-level standards; actual field emissions depend on load profiles and maintenance, but the order-of-magnitude reduction holds.

Because Tier 4 units draw down the annual emissions budget more slowly on the whole, the same fixed 100-tpy NOx permit in the hypothetical could support a materially larger generator fleet at the same 50-hour runway, a longer runway at the same fleet size, or some blend of the two. All of this occurs without changing the permit itself, though land, fuel storage, and capital constraints will limit how much of that headroom any given site actually can use. For a developer deciding how much capacity a given permit can support, that is a capacity question as much as a resiliency question, and often the more consequential one.

Some operators already deploy Tier 4 generators (cleaner generators) voluntarily, even where regulation does not yet require it. Capacity is a leading reason: A Tier 4 fleet lets a developer bring more megawatts online within the emissions envelope a site can lawfully support. The resiliency and reputational payoffs are also important. But for a campus weighing how large it can build, capacity is usually what moves the investment decision.

Those gains come with trade-offs: Tier 4 units cost meaningfully more than Tier 2 units, can be harder to source, and require more space and maintenance attention. And while Tier 4 expands what a permit can support, in capacity and hours, it does not guarantee that expanded runway will be long enough for every outage a campus might face.

Published in *Natural Resources & Environment* Volume 41, Number 2, Fall 2026. © 2026 by the American Bar Association. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

There is regulatory momentum toward Tier 4 as a baseline. This began in California, where the Bay Area Air Quality Management District determined, in a December 21, 2020, letter, that Tier 4 emissions standards constitute best available control technology (BACT) for large (≥1,000 brake horsepower) diesel engines used for emergency standby power. Other jurisdictions have since followed suit, including Maricopa County, Arizona (establishing, as of 2021, that BACT for emergency engines of 1,000 bhp or larger requires meeting Tier 4 emissions limits); Virginia (adopting guidance in December 2025 making Tier 4–equivalent controls the presumptive BACT for new data center diesel generators); and Illinois (requiring, as of January 2026, that new diesel-fired emergency generators at data centers meet Tier 4 emissions standards under the state's Clean and Reliable Grid Affordability Act).

The second strategy addresses the other variable: the permit's annual tonnage cap itself. Pursuing a major source permit (a "bigger" permit), rather than remaining under a minor source or synthetic minor cap, authorizes a higher annual-emissions ceiling. In the hypothetical, a larger permit would lift the 100-tpy NOx ceiling that produced the 50-hour figure, which the operator could spend by adding resiliency and/or expanding capacity.

The major source threshold matters here. Sources with actual or potential emissions of around 100 tpy of a regulated pollutant generally trigger Title V major source obligations, though precise thresholds vary by pollutant and attainment status. 42 U.S.C. § 7602(j); 40 C.F.R. pts. 70, 71. The hypothetical's 100-tpy NOx cap sits at that threshold; the facility is managing emissions to remain a minor source. Releasing that constraint opens headroom.

The permitting trade-offs are real. Moving to major source status makes preconstruction review substantially more burdensome, although the specific requirements depend on whether the facility sits in an attainment or nonattainment area. In attainment areas, a source large enough to trigger Prevention of Significant Deterioration (PSD) review, which is generally 250 tons per year for most source types, must apply BACT and undergo detailed air-quality and increment modeling. 40 C.F.R. § 52.21(b)(1), (j)–(l). In nonattainment areas, a source triggering nonattainment New Source Review, at thresholds that vary by pollutant and by how severely an area exceeds air quality standards, instead must meet the lowest achievable emission rate (LAER), which for generators of this size could functionally require Tier 4–equivalent aftertreatment, and obtain emission offsets from other sources in the area. *Id.* § 51.165(a)(1)(xiii), (a)(3). That preconstruction layer, rather than Title V itself, typically drives any added construction timeline since Title V is an operating-permit program that consolidates applicable requirements rather than a separate preconstruction approval and often follows startup. Marcia L. Spink, U.S. EPA, *Clean Air Act 101: Module 5—Title V: Operating Permits* (archived; last visited Aug. 31, 2026). Title V nonetheless adds a layer of monitoring, reporting, and public-comment obligations that can invite opposition beyond what a routine minor source permit attracts. For operators accustomed to minor source permitting, the step up can present a meaningful increase in burden. But for larger campuses seeking more capacity than a minor source permit can support, that burden buys capacity emissions and runtime headroom a minor source ceiling cannot, which may determine whether the facility can ride through an extended outage without exceeding permit limits.

Practically, a larger permit paired with cleaner generators is the strongest combination available today, and it works on both axes that matter. Together, a bigger ceiling and a slower emissions draw rate let a campus build more capacity, hold a longer runway, or both.

A Seat at the Table: Shaping Runtime Rules

The third path differs from the first two. It likely is not a lever any individual operator could pull unilaterally. Instead, it is more realistically a collective undertaking the industry could pursue through trade associations and coalitions. Rather than treating the permit's fixed ceiling as a constraint to be discovered only after an outage forces the issue, the industry could engage lawmakers and regulators directly to shape workable runtime standards before a crisis arrives.

Sources with actual or potential emissions of around 100 tpy of a regulated pollutant generally trigger Title V major source obligations, though precise thresholds vary by pollutant and attainment status.

Consider an operator that has done everything right (adopting a defensible runtime standard, deploying Tier 4 generators, and obtaining an appropriately sized permit), yet still faces a regional outage running longer than that planning anticipated. At hour 51, the operator confronts a choice with no good option: keep the generators running past the permitted cap (a violation regardless of cause) or shut the facility down (potentially cutting off hospitals, emergency dispatch, and public safety systems that depend on the data center staying online). Current law offers nothing in between, and no amount of planning can eliminate that choice. While regulators may exercise enforcement discretion after the fact, relying on unwritten discretion is a fragile foundation for facilities that carry critical public-safety loads and face such significant liability exposure.

That gap is worth closing, and closing it is squarely in the public's interest as well as the industry's. Modern society now

Published in *Natural Resources & Environment* Volume 41, Number 2, Fall 2026. © 2026 by the American Bar Association. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.

relies on data centers remaining online, for better or worse, and federal and state regulators share operators' interest in having a clear framework for what happens when an outage lasts longer than any permit anticipated, rather than improvising after the fact through enforcement discretion. The industry's central ask of government should be to remove this risk and uncertainty before the next extended outage forces the question, rather than leaving well-prepared operators stuck between a rock and a hard place, facing the same binary choice as operators that never planned at all.

What that collaboration could produce is an open question this article does not pretend to answer. But consider one possibility: a narrowly tailored safe harbor for data center outages that extend beyond a documented, reasonable runtime standard. Any such idea would need to address who approves a standard in advance, what happens if a regulator later disagrees, and how transparency is preserved for nearby communities since nothing along these lines should shield concealment or delayed notice. It should be narrower and more rigorously documented than a bare emergency defense, built around a runtime standard fixed in advance, verifiable compliance steps, and real transparency to nearby communities so it cannot become an after-the-fact, undocumented excuse. The specific mechanism ultimately matters less than the underlying point: Government and industry both have reason to work out how extended outages should be handled now, rather than letting the next one settle the question by accident.

Getting Ahead of the Next Outage

Neither the hypothetical's 50-hour runway nor its 300 megawatts of capacity is a fixed law of nature. In the long run, the more important work may involve sitting down with federal and state governments to remove the risk and uncertainty that still surround extended outages so that no well-prepared operator ever has to face the choice this article describes.

Moving quickly on this front carries its own value, separate from any deadline. The longer a campus operates without a documented runtime standard, the longer it accumulates exposure to the exact enforcement risk this article describes, and the harder it becomes to argue that a foreseeable gap went unaddressed. An industry that engages now positions itself to help shape workable rules rather than simply having to comply with the rules that emerge without its input.

The enforcement stakes underscore the urgency. Knowing violations of Clean Air Act requirements (including Title V, NSPS, and NESHAP), in egregious cases, can carry criminal exposure under 42 U.S.C. § 7413(c)(1). And civil judicial enforcement under § 7413(b) carries inflation-adjusted maximum penalties of roughly $120,000 per day per violation, so even a single count can run into the tens of thousands of dollars per day, per violation. 40 C.F.R. § 19.4, tbl. 1. A campus that exceeds its permitted emissions cap during an extended outage does not automatically face penalties at that level. EPA's own civil penalty policy considers factors such as the violator's compliance history, good-faith efforts, and the duration and seriousness of the violation. But the exposure is real enough that emissions-budget planning deserves the same attention as generator sizing and fuel logistics.

Early planning changes the outcome, and arguably it should change the legal exposure too, even if the law does not yet say so explicitly. A campus that evaluated its runtime constraint during project design, deployed Tier 4 generators, and pursued a calibrated permit ceiling sits in a fundamentally different position than one that never modeled the scenario. That is foreseeability addressed rather than foreseeability ignored, a difference regulators and courts already weigh when deciding how to respond to an exceedance.

The companies that plan for the outage they hope never happens, and talk to regulators about it now, will be in a very different position than those that did not. Larsen and Lillie's analogy holds. No one wants to run out of gas in the fast lane. But the more precise goal is not just avoiding that outcome. It is planning the trip and, where possible, working with regulators to redraw the map so that scenario never has to be tested.

Garrett Jansma is senior environmental counsel at Allen Matkins in Irvine, California. He may be reached at gjansma@allenmatkins.com.

Your gift to the ABA Section of Environment, Energy, and Resources—whether $10 a month or $1,000 a year—helps sustain support for communities around the nation, advance environmental justice, and support equitable access to our profession.

You can donate today by scanning the QR code or visiting **ambar.org/DonateToSEER**.

Thank you for your support!

ABA AMERICAN BAR ASSOCIATION
Environment, Energy, and Resources Section

You can now deduct your gift even without itemizing! For the 2026 tax year, new rules let non-itemizers deduct up to $1,000 (single) or $2,000 (joint) in cash donations "above the line."

Published in *Natural Resources & Environment* Volume 41, Number 2, Fall 2026. © 2026 by the American Bar Association. Reproduced with permission. All rights reserved. This information or any portion thereof may not be copied or disseminated in any form or by any means or stored in an electronic database or retrieval system without the express written consent of the American Bar Association.