COUNTING DATA CENTERS

The OpenStreetMap Inventory Behind a National County-Level Comparison

Series: A research note. Companion to the Virginia data center case study series
Evidence snapshot: September 17, 2026
Version: 1.1
Author: Tommy N. Turner
Location: Williamsburg and Christiansburg, Virginia
Contact: tommy@gofast.pro · tommynturner.com
ORCID: 0009-0005-7422-9881
DOI (this version, v1.1): 10.5281/zenodo.23441602
Cite all versions: 10.5281/zenodo.22819653
Published: October 2026

Version note: A research note examining the facility inventory behind a national county-level comparison of housing, employment and electricity outcomes. Section 5.3 reports a classification of the comparison register this note measures against, and states what that register contains that is not a data center.

© 2026 Tommy N. Turner

Evidence snapshot: 17 September 2026. All figures in this note were computed on that date from files held locally and listed in Methods. The underlying registers are live and revisable.

Disclosure. The author serves on the advisory board of *Data Centers and the Southern Energy Buildout*, a Sloan Foundation-funded study at Emory University. The study had no role in this note.

Evidence grades. Claims here carry a mark saying what stands behind them, so a reader can tell which sentences rest on a document and which rest on the author's reading of one. [P] is a primary record: a file with a stated hash, a named page of a published report, or a figure one of the distributed scripts computes. [S] is a secondary source corroborated elsewhere, [S¹] a single source with no corroboration sought, and [I] an inference by the author, where the step beyond the record is stated in the sentence that makes it. The marks follow the convention used in this author's other volumes.

Abstract

The *2026 Data Center Impact Report*, published by the National Association of REALTORS® on 9 September 2026, sorts more than 3,200 United States counties into four groups by the number of data centers each contains, and compares housing, employment and electricity outcomes across the groups. The grouping comes from the IM3 Open-Source Data Center Atlas, a facility file that Pacific Northwest National Laboratory derives from OpenStreetMap. This note audits that file as an exposure measure.

The report's counts reproduce exactly from the source it names. What the file counts is an object a volunteer contributor drew, and 1,237 of its 1,474 records are individual buildings rather than facilities or campuses. Between the two versions of the file the report cites, eight months apart, 55 of 252 counties fall into a different one of its four groups. The February 2026 version records no data centers in 58 of 163 counties where the Environmental Protection Agency held air-program records under the corresponding industry code, and in nine Virginia localities where the state had issued air permits to named data center sites. The report's own county appendix places all nine of those localities in its "No Data Centers" group.

Nothing here establishes that the report's conclusions are wrong. What it bears on is what a county-level data center count measures, and how much weight a comparison across such counts can carry.

Why this note exists

The report is written for people making decisions with it. Its seventh section is addressed to "REALTORS® working in markets where data centers are already or new data centers are being proposed", and its closing passage sets out what such a professional should want to know about a market [P]. The first question on that list is "How many data centers are there?" [P]

The report answers that question county by county, in a published appendix, and every comparison it makes runs on the answer. A county's group decides which median it contributes to. The count is not a descriptive detail sitting beside the analysis. It is the variable the analysis turns on.

That count came from somewhere, and the report says where, which is more than most reports do. It came from a file assembled out of OpenStreetMap, the volunteer-edited map anyone can contribute to. Whether such a file can carry the weight of a national county comparison is a question nobody appears to have asked, on either side of the argument about data centers and property values, and it is answerable in an afternoon from public records.

This note asks it.

Everything needed to check the answer is deposited with this note. The archive carries both versions of the Atlas file, the federal and state permit registers they are compared against, and eight scripts that reproduce every figure here. They run on Python 3 with the standard library and nothing else, in seconds, with nothing to download first. Section 8 lists them and says what each one feeds. A reader who believes a number here is wrong can settle it rather than argue about it.

What this note finds

In September 2026 the National Association of REALTORS® published a county-level analysis of data centers and local housing, employment and electricity markets. It sorts more than 3,200 U.S. counties into four groups by the number of data centers each contains, and most comparisons in the report run across those four groups. Its methodology states: "For most comparisons, we use the median for each group" [P].

This note examines the inventory that produces the grouping. Three findings.

The report's counts reproduce exactly from its cited source, in four separate checks [P]. Its arithmetic is correct and its source is public, named and openly licensed.

The unit counted is an object tagged in OpenStreetMap, and about five of every six such objects is a single building rather than a facility or a campus [P]. Whether one data center contributes one record or twenty depends on how a volunteer contributor chose to draw it.

The February 2026 version of the inventory records no data centers in 58 of the 163 U.S. counties where the Environmental Protection Agency held air-program records under the corresponding industry code as of 17 September 2026 and where at least one of those records is not affirmatively something other than a data center, nor in nine Virginia localities where the state had issued air permits to named data center facilities as of 17 August 2026 [P]. Under the report's grouping, those counties sit in the comparison group, and the report's own county appendix confirms it: all nine are listed with a facility count of zero under the label "No Data Centers" [P]. Section 5.3 reports the classification behind that figure.

A fourth observation concerns stability. The report cites two versions of the inventory eight months apart. Between them, 55 of 252 counties fall into a different one of the report's four groups [P].

None of this establishes that the report's conclusions are wrong. The report states its own limitations at length, and this note quotes them. What the findings bear on is narrower and prior: what a county-level data center count measures, and how much weight a comparison across such counts can carry.

---

1. The report and the file its counts come from

The National Association of REALTORS® announced the *2026 Data Center Impact Report* in a newsroom release dated 9 September 2026, and presented it in a member webinar on 15 September [P]. The report file examined here carries a creation date of 14 September 2026 in its own metadata, so the document in hand is not necessarily the file distributed on the announcement date [P]. It has two parts: a county-level analysis of more than 3,200 counties, and a survey of 2,357 members [P].

The report names its inventory in its methodology section, verbatim: "Data center locations and footprints come from the IM3 Open-Source Data Center Atlas, developed by Pacific Northwest National Laboratory (PNNL), a U.S. Department of Energy national laboratory. We use the June 2025 and February 2026 versions of the Atlas" [P].

The Atlas describes itself in its own README, verbatim: "This dataset contains locations of existing data center facilities in the United States (U.S.) derived from OpenStreetMap (OSM), a crowd-sourced database" [P]. It is distributed under the Open Database License through the MSD-LIVE repository, DOI 10.57931/3017294 for the February 2026 version [P].

Naming a public, openly licensed federal dataset is what makes every check below possible. A reader can obtain the same file and recompute the same numbers.

---

2. The counts are correct

Deduplicating the February 2026 Atlas by facility identifier yields **1,474 facilities across 251 counties** [P]. The report states, verbatim: "The latest data identify 1,474 data centers across 251 U.S. counties" [P]. Exact agreement on both figures.

Three further checks agree [P]:

Virginia totals match the webinar deck exactly for the three counties the deck names: 319 statewide, Loudoun County 213 and Prince William County 63 [P]. Fairfax County returns 27 from the Atlas and that figure appears in the report rather than the deck [P]. Fairfax does not appear anywhere in the deck's text.

The report's Virginia state profile states that its 319 facilities "are located across 11 counties" [P]. The Atlas returns exactly 11 Virginia localities.

The deck's "Net Change of Facilities 2025-2026" column reproduces in all eight cases tested against the difference between the two Atlas versions: Virginia +27, Loudoun +14, Prince William +5, Santa Clara +14, Maricopa +9, Grant +1, Texas +35, Ohio +35.

Five facility identifiers appear twice because a facility straddles a county boundary, which the Atlas README documents [P]. The report's figure of 1,474 indicates it deduplicated by identifier, which is correct.

This section should be read as stated. The report's arithmetic is right, its source is identified, and its published counts can be reconstructed from a free file by anyone who wants to. Nothing that follows is a claim that the report miscounted.

---

3. The unit counted is a drawn shape

Of the 1,474 deduplicated facilities, **1,237 are building polygons, 132 are campus polygons and 105 are points** [P]. No identifier appears in more than one of the three layers, so the count is dominated by individually tagged buildings.

The Atlas README states the consequence in its own words: "In some instances, both the data center area (e.g., campus) and individual data center buildings are included as overlapping areas in the database. Both values are retained" [P].

Loudoun County's 213 records are 189 buildings, 21 campuses and 3 points [P]. Amazon Web Services appears as the operator on 79 of the 213, and 15 records carry the operator and name fields as identical strings reading "Amazon Web Services" [P]. Operator is blank on 519 of the 1,474 records, 35.2 percent, and name is blank on 275, 18.7 percent [P].

The inference, stated as one. A contributor who traces each building on a campus produces more records than one who draws a single outline around the same campus. The Atlas retains both forms and says so. It therefore follows that a county's record count reflects tagging practice as well as facility presence, and that two counties with identical physical development can land in different groups `[I]: this is an inference from the layer composition and the README's statement about retention, not a measured decomposition of how much of the variation is tagging practice. No such decomposition has been attempted here.`

---

4. One county in five changes band between versions

The report cites two Atlas versions. Both were obtained and compared [P].

June 2025: 1,240 facilities in 209 counties. February 2026: 1,474 facilities in 251 counties. Matching on coordinates rather than identifier, 1,151 facilities appear in both versions, 323 are new in February and 89 present in June are absent in February [P].

A trap worth recording, because this note nearly reported its output as a finding. The identifier field is zero-padded differently between versions. June identifiers run 10 or 11 characters, February identifiers are all 11, so `0002744301` becomes `00002744301`. Matching on identifier reports 1,197 of 1,240 facilities dropped, a 96 percent churn, and that figure is an artifact. It was caught in review and is stated here so no one repeats it [P].

The Atlas README anticipates movement in both directions, verbatim: "some data may be missing where it has not yet been reported. As we collect information on additional data center locations and as OSM receives additional contributions, the database will be updated to capture additional data points not yet shown." And: "updated versions of this database may not contain all data center locations included in previous versions" [P].

The webinar deck presents the difference between versions as development, reporting that Ohio "added the most facilities of any state (+35, tied with Texas)" [P]. Eight months of contributor activity separates the two files. A county's increase can reflect new construction, or a contributor tracing buildings that already stood, and the two versions alone do not distinguish them `[I]: the distinction cannot be made from the two files alone.` The deck presents the column as buildout, and its own sources do not support that reading.

The distinction is testable against outside records, and the test was run for Virginia. `atlas_delta_vs_permits.py` compares the February-minus-June increment against Virginia DEQ permits issued between the two versions, taking the window as 1 June 2025 to 9 February 2026 [P].

Forty-two Virginia facilities appear in the February file and not the June file, matched on coordinates: Loudoun 19, Prince William 13, Fairfax 6, Fauquier 2, Henrico 1, Manassas 1. Thirty-eight of the forty-two are building polygons, and operator is blank on twenty-one [P]. All forty-two sit in localities holding a permit issued inside the window, which on its own is weak evidence because those are the localities that permit continuously.

The converse is the informative direction. Eight Virginia localities had a data center permit issued inside the window. Six gained Atlas records; Frederick and Mecklenburg, with one permit each, gained none. **And of the six localities that gained records, all six had permitting activity in the same window. None gained records without it** [P].

That result cuts against the strong reading of this section and is reported as such. In Virginia, the increment did not appear anywhere lacking contemporaneous permitting, which is evidence against the increment being purely catch-up mapping in this state.

What it does not show: magnitudes do not track. Prince William had two permits in the window and gained thirteen records; Fairfax had one permit and gained six; Loudoun had seventeen and gained nineteen [P]. Permitting activity and OpenStreetMap contributor attention concentrate in the same few localities, so a locality-level association cannot separate them [I]. One state, one window, eight localities.

The national version of the test, and it goes further. `atlas_vs_icis_national.py` repeats the comparison for every state using EPA's ICIS-AIR system, which records the date a facility entered an air program. Thirty-two counties had a NAICS 518210 air program begin inside the window between the two Atlas versions; 110 counties gained Atlas records; thirteen counties are in both [P].

Stated as "97 of 110 counties that gained records, 88 percent, had no program start," the same result would read as showing the increment was unrelated to real activity. That reading is wrong, and wrong in the specific way this note exists to describe: a share reported without its denominator. Program starts are rare, thirty-two counties out of roughly 3,144. Against that base rate the expected overlap if the two were independent is 1.12 counties, and thirteen were observed: an enrichment of **11.6 times**, with a hypergeometric probability of 1.15e-11 [P]. Put the other way, 41 percent of counties with a program start gained Atlas records, against a national rate of 3.5 percent.

So the increment is strongly associated with regulatory activity nationally, and the Virginia result is unlikely to be an artifact of Virginia's density alone [I]. **The confound stated for the Virginia test applies here unchanged.** Permitting activity and OpenStreetMap contributor attention concentrate in the same counties nationally as they do within Virginia, and a county-level test cannot separate them. A facility-level test, asking whether the specific new geometry sits at the specific permitted facility, would separate them and is not attempted here [I].

The surviving claim is narrow. A reader of the deck still cannot tell what share of a state's net change is construction and what share is mapping, and neither the deck nor the Atlas supplies the outside record that would answer it. What the outside record does show, in both tests, is that the increment is not mapping noise. The deck's presentation of it as buildout is closer to right than a reading of the two versions alone would support.

Limits on both tests: a program begin date is not a permit issuance date and not a construction date, a facility built under an existing permit generates no new program record, county matching is by normalized name because ICIS-AIR carries county names rather than FIPS codes, and association is not attribution [I].

On grouping stability. Applying the report's four bands to both versions, **55 of 252 counties fall into a different band depending on which version is used** [P]. Ellis County, Texas moves from 7 to 15 and Du Page County, Illinois from 8 to 14, both crossing from the 3-to-9 band into the 10-plus band. Rockland County, New York moves from 0 to 4 and Ada County, Idaho from 0 to 3, both crossing out of the no-data-centers band. Fauquier County, Virginia moves from 1 to 3.

---

5. Counties with permitted facilities that the file records as empty

Two comparisons, one state and one national. Both compare the Atlas against registers of facilities holding air permits, which is a different instrument measuring a different thing. The limits of the comparison are stated in section 6 and should be read with the figures.

5.1 Virginia

The Virginia Department of Environmental Quality publishes its issued air permits for data centers. The register held locally covers 201 permits across 185 distinct site-name strings, dated 2000 to 2026 [P]. DEQ's page carries an as-of date of 17 August 2026; the copy held locally was retrieved 23 August 2026, and every row records that retrieval date. The two dates are not the same thing and the distinction matters for any claim bounded to the register's state.

Nine Virginia localities hold DEQ-permitted data center sites and show zero facilities in the February 2026 Atlas [P]: Caroline, Chesterfield, Hanover, Louisa, Russell, Spotsylvania, Stafford, Virginia Beach and Wise.

The report's own appendix places all nine in its "No Data Centers" group [P]. `NAR_2026_Appendix_Concentration_by_county.pdf` lists each with a facility count of zero and that group label: Caroline 51033, Chesterfield 51041, Hanover 51085, Louisa 51109, Russell 51167, Spotsylvania 51177, Stafford 51179, Wise 51195 and Virginia Beach 51810. Each carries a full row in the companion data appendix, Stafford's reading a median home value of $485,100, so each contributes to the medians the report compares [P].

Those nine localities carry twelve permits between them. All twelve are listed, rather than a selection, because the selection would otherwise do argumentative work [P]:

Permit issued	Locality	Facility as named in the register
23 March 2007	Chesterfield	11751 Meadowville Owner LLC c/o GFH Capital Ltd
16 May 2008	Russell	Northrop Grumman Systems Corporation
7 August 2015	Virginia Beach	Sentara Healthcare
23 August 2016	Wise	DP Facilities Inc South LLC
27 March 2025	Stafford	Amazon Data Services Inc IAD-350

Permit issued	Locality	Facility as named in the register
20 February 2026	Louisa	Amazon Data Services Inc - IAD - 550 Louisa REB
26 March 2026	Spotsylvania	Amazon Data Services IAD-870
9 April 2026	Chesterfield	Chirisa Richmond LLC
12 May 2026	Stafford	Vantage Data Centers VA41 LLC
1 June 2026	Spotsylvania	Amazon Data Services Inc. IAD450 Cosner Tech Campus
2 July 2026	Caroline	Clean Arc Data Centers - VA-1 Campus
21 July 2026	Hanover	CLOP Ashland VA LLC

Five of the twelve predate the February 2026 Atlas version, which the Atlas's own repository record fixes at 9 February 2026 [P]. Four of those five are old: 2007, 2008, 2015 and 2016. An older permit may cover a facility that has since closed, and the Virginia Beach permittee is a healthcare system, which is a different kind of object from a hyperscale campus [I].

The fifth is the clearest single case in this note. Virginia DEQ issued an air permit to Amazon Data Services Inc for IAD-350 in Stafford County on 27 March 2025 [P]. The February 2026 Atlas version, published ten and a half months later, records no data center in Stafford County [P].

The remaining seven permits postdate the Atlas version and cannot bear on whether that version should have recorded anything. They are shown because they establish that these localities are not marginal cases: six of the nine are the sites of named, permitted hyperscale or colocation projects as of mid-2026 [P].

What is not established. Why OpenStreetMap holds no record in these localities is not determined by either register. A permit is not a building, and neither the DEQ register nor the Atlas establishes whether a structure existed to be mapped on any given date. The finding is that one inventory records zero where a state permitting authority records named, permitted facilities.

5.2 National

EPA's ECHO facility file was downloaded 17 September 2026. Scanning all 3,180,415 facility records returns 734 carrying NAICS 518210, Data Processing, Hosting and Related Services, of which 482 carry an air-program flag [P].

Method validation. Virginia returns 133 such facilities. The author's Virginia crosswalk, built 23 August 2026, reports 134 ECHO air facility records at the same industry code [P]. One record apart, three weeks apart, on a live register.

Joining on five-digit county FIPS, 64 of the 169 counties holding EPA air-program facilities in this industry code as of 17 September 2026, 38 percent, record zero facilities in the February 2026 Atlas version [P]. Those 64 counties hold 79 such facilities and span 27 states, led by Virginia with 8, Georgia with 7 and Indiana with 7 [P].

That figure should not be read before section 5.3. Six of the 64 counties hold nothing but facilities that are affirmatively not data centers. Excluding them from both sides, **58 of 163 counties, 36 percent, record zero** [P].

The FIPS join convention moves that figure across a range of 34 to 37 percent, which `fips_fallback_sensitivity.py` reports and Methods states.

The two largest entries in that count do not stand on equal footing. New Castle County, Delaware holds six records and survives inspection: three JPMorgan Chase sites, two Hosting.com sites and Computer Sciences Corporation [P]. Jackson County, Colorado holds five, and every one is a flare-gas crypto mining rig on an oil or gas well pad [P]. A coverage count that cannot tell those two counties apart is not yet a finding about data centers, which is what section 5.3 exists to address.

Restricting to facilities EPA records as operating. ECHO's air-program flag is an inference about status; ICIS-AIR carries an explicit operating-status field, and `atlas_vs_icis_national.py` repeats the comparison on it. 134 counties hold at least one NAICS 518210 facility EPA marks Operating, and **44 of those 134, 33 percent, record zero facilities in the February 2026 Atlas**, holding 59 operating facilities across 22 states [P]. On the section 5.3 classification the same figure is 42 of 132, 32 percent.

Two join conventions in that script change the result. ICIS records a literal county name of "Undetermined" for seven Ohio facilities and one in New Jersey; treated as a county, "Undetermined, Ohio" would enter the comparison holding seven facilities the Atlas records none of, and the Ohio seven are real campuses. And ICIS writes Virginia's independent cities as "Manassas (city)" where the Atlas writes "Manassas city"; a name normalizer that strips only the second form counts Manassas and Harrisonburg as coverage misses while the Atlas holds a record for each. The script handles both. Without those two conventions the test reads 48 of 136, 35 percent, 71 facilities, 23 states. The second national test in section 4 is unaffected in every figure [P].

In the other direction, 146 of the 251 counties in the February 2026 Atlas hold no EPA air-program facility in this code [P].

How well the two registers agree where both see a county, reported five ways. Among the 105 counties both record, the Pearson correlation between the raw counts is **0.837**. That figure is driven by a small number of very large counties and does not survive transformation [P]:

Statistic	Value
Pearson, raw counts, n = 105	0.837
Spearman rank correlation	0.406
Pearson on log(1+x)	0.530
Pearson excluding Loudoun County alone, n = 104	0.519
Pearson excluding the ten largest Atlas counties, n = 95	0.470

The median ratio of Atlas count to EPA count across those 105 counties is 1.67, and 38 of the 105 have an Atlas count at least three times the EPA count [P].

The raw coefficient alone reads as reassurance, and it should not. On a rank measure the two registers agree weakly. The raw coefficient is a statement about a handful of large counties rather than about the registers in general.

Those high-count cases: Loudoun 213 against 58, Santa Clara 75 against 1, Prince William 63 against 21, Maricopa 63 against 17, Grant 39 against 1, Umatilla 31 against 1, Cook 29 against 16 [P].

5.3 What the comparison instrument actually contains

Sections 5.1 and 5.2 compare an inventory against a permit register. Auditing the inventory closely, by layer, by operator string and by blank-field rate, while taking the register on trust, would repeat the failure this note describes. The register is audited here on the same terms.

The audit still runs in one direction and that is a limit of this note. Every facility behind the coverage counts is classified below. No equivalent check is made on the Atlas's own 1,474 records, which would ask whether an OpenStreetMap contributor has tagged a substation, a telecom hut or an unrelated warehouse as a data center. That check would move the figures in the opposite direction and it is not attempted here [I].

NAICS 518210 is "Data Processing, Hosting and Related Services." A facility carries the code through how its operator is classified, not through what stands on the site. `echo_icis_facility_classification.py` names every facility behind both national counts and sorts it on a rule stated in the script: a facility is called **not a data center** only where its name affirmatively identifies it as something else, and **uncertain** where the name gives an operator but no use. Everything else is left as a data center. The rule is conservative in one direction, so the first group is a floor [P].

Of the 79 facilities in the 64 ECHO counties: **10 are affirmatively not data centers, 20 cannot be classified from the name, and 49 are recognizable data centers** [P].

Six counties hold nothing but the first group, and drop out of both sides of the comparison [P]:

County	What the records are
Jackson County, Colorado	five NYDIG DFM flare-gas crypto mining rigs on oil and gas well pads
Moffat County, Colorado	GRMR Oil and Gas, Deal Gulch production
Coshocton County, Ohio	American Electric Power Conesville Plant, a coal-fired power station
Columbiana County, Ohio	GeoPetro Rosebud 16-10-2 well pad
King George County, Virginia	Birchwood Power Facility, a merchant power station
McDonald County, Missouri	a Walmart parking lot rehabilitation project

The headline moves from 64 of 169, 38 percent, to 58 of 163, 36 percent. On the ICIS operating set the same exclusion moves 44 of 134 to 42 of 132. The finding survives; the figure was two points high and one of its named examples was wrong.

What this does not settle. A facility name is weak evidence. A manufacturer or a hospital system can operate a substantial data center on its own site, and some in the uncertain group probably do. Twenty facilities cannot be classified from the name at all, which is a quarter of the count. The instrument that would settle it is a site-by-site check against permit descriptions, which this note does not build [I]. What is established is narrower: the

comparison register contains facilities that are affirmatively not data centers, the share is not negligible, and no coverage figure drawn from NAICS 518210 should be reported without saying so.

---

6. What this note does not establish

Neither register is a census, and the two measure different objects. ECHO records facilities in air programs; a data center without qualifying backup generation need not appear, which is the likeliest reason 146 Atlas counties have no ECHO match [I]. The Atlas records objects tagged in OpenStreetMap. "Loudoun 213 against 58" is not an error of 155.

NAICS 518210 is broader than data centers. Section 5.3 classifies the facilities individually and the classification changes a headline figure. The residual limit is stated there: twenty of the seventy-nine facilities cannot be classified from a name, and the classification that can be done from names is a floor rather than a measurement [I].

Air permitting practice varies by state and by local air district, but that does not explain Santa Clara. Santa Clara County's ratio of 75 Atlas records to 1 EPA facility might be read as reflecting California permitting practice rather than Atlas over-counting. **That reading does not hold.** Backup generators in the Bay Area Air Quality Management District require district permits, and the district's permitting process for generators is documented publicly [S¹]: a single petition document surfaced in review, describing the district's practice of permitting generators individually; the district's own rule text has not been pulled. Pull it before print. If those facilities are permitted locally and do not appear in the federal file, the gap sits on the federal side, in how district permits reach EPA's national systems, rather than in whether the facilities were permitted. **Which of those it is has not been established here**, and nothing in this note resolves it.

The report states its own limitations, repeatedly and in its own text. On electricity, verbatim: "The rates are summarized at the state level and matched to counties within each state. This means that they do not capture differences within a state or a specific utility area" [P]. On causation: "This does not mean that data centers caused home values to rise. Counties with more data centers also tend to have higher incomes, younger populations and more college educated residents" [P]. And: "it cannot tell us that a specific facility caused a change in home prices, jobs or electricity rates" [P]. The report also publishes the non-monotonic pattern in its own electricity results, noting that the largest residential rate increase was in counties with one or two data centers rather than those with ten or more [P].

The general property is documented elsewhere, for OpenStreetMap as a whole. Herfort, Lautenbach, Porto de Albuquerque, Anderson and Zipf, "A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap," *Nature Communications* 14:3985 (2023), DOI 10.1038/s41467-023-39698-6, finds that OSM building completeness varies sharply and systematically, exceeding 80 percent for some urban centres while remaining below 20 percent for cities holding 48 percent of the world's urban population, and that the distribution of completeness across urban centres is, in the paper's words, "characterized by spatially clustered patterns at various scales" [S]. That study concerns urban building footprints worldwide rather than U.S. data centers, so it establishes that OSM coverage varies non-randomly in general and does not establish the specific pattern reported in section 5 [I].

Scope of the negatives in this note. The 58-county finding is bounded to counties holding EPA air-program records under NAICS 518210 in the file downloaded 17 September 2026, after the section 5.3 exclusion, compared against the February 2026 Atlas version, and it moves by about two points either way on the FIPS join convention. The 44-county finding is bounded to facilities EPA marked Operating in ICIS-AIR on the same date, joined by county name. The nine-locality Virginia finding is bounded to the DEQ issued-permit register as that register stood on 17 August 2026, retrieved 23 August 2026, compared against the same Atlas version; of its twelve permits, five predate the Atlas version and one of those five carries no caveat this note could raise against it. **None of the three is a statement about how many data centers exist in any county**, including the nine-locality finding, which says only that one inventory records zero where a state permitting authority records named, permitted sites.

---

7. What follows for anyone using a county count

A county-level comparison of any outcome across data center concentration bands inherits whatever properties the concentration measure has. Three properties are documented above, each at the scope established in its own section: the February 2026 measure counts tagged objects rather than facilities, with 1,237 of 1,474 being individual buildings; it records zero in nine Virginia localities holding DEQ-issued permits as that register stood on 17 August 2026, and in 58 of the 163 counties holding EPA air-program records under NAICS 518210 as of 17 September 2026 after the facilities that are affirmatively not data centers are removed; and 55 of 252 counties fall in a different band depending on which of the two cited versions is used.

The practical consequence for a reader is that the band a county occupies is a measurement carrying error of a size these comparisons can bound but not correct. Whether that error is systematic in a statistical sense has not been tested here, and testing it would require a coverage model this note does not build `[I]`. What the comparisons do show is that the error is not small: it moves 55 counties between bands across two versions of the same dataset, and it leaves 58 counties at zero that EPA's air-program file, as downloaded 17 September 2026 and net of the facilities section 5.3 removes, records as hosting facilities in the corresponding industry code.

The practical consequence for anyone conducting such a comparison is that the inventory choice is a research design decision requiring the same justification as any other, and that a public, crowd-sourced inventory of existing facilities carries coverage properties that should be reported alongside results.

---

8. Methods and reproducibility

Every figure in this note can be recomputed from files that are free to obtain.

Sources.

IM3 Open-Source Data Center Atlas, Pacific Northwest National Laboratory, U.S. Department of Energy. February 2026 version, DOI 10.57931/3017294, MSD-LIVE record p147s-4h760, downloaded 17 September 2026, md5 `dfe532e258ab233dfe86a92340607098`. June 2025 version, same repository, md5 `0b34754b5cc41b644d90c9fcc5aa72f8`. Open Database License.

National Association of REALTORS® , *2026 Data Center Impact Report*, 55 pages, md5 `8c9aaf70acfb764522d879ff97a83258`. Webinar deck, "The Data Behind The Data Centers: A County-Level Analysis," Nadia Evangelou, 15 September 2026, 11 slides.

Virginia Department of Environmental Quality, issued air permits for data centers, as of 17 August 2026 on the agency's page, retrieved 23 August 2026. 201 permits, 185 distinct site-name strings.

U.S. EPA, ECHO Exporter national facility file, downloaded 17 September 2026. 3,180,415 facility records; extract of 734 records carrying NAICS 518210, md5 `f6aec4f0f2b90ce9aac430b5011ad88f`.

Scripts. Eight Python scripts accompany this note, each carrying its results and caveats in its own header: `atlas_count_audit.py` (sections 2, 3 and 5.1), `atlas_version_delta.py` (section 4), `atlas_delta_vs_permits.py` (section 4), `atlas_vs_icis_national.py` (sections 4 and 5.2), `atlas_vs_echo_national.py` (section 5.2), `correlation_robustness.py` (the section 5.2 table), `fips_fallback_sensitivity.py` (the join convention, section 5.2 and Methods), and `echo_icis_facility_classification.py` (section 5.3). ICIS-AIR extracts are in `icis_extract/`, and every input file is distributed alongside. Each runs in seconds on the source files, on Python 3 with the standard library and nothing else.

Every figure in this note is produced by a script distributed with it. The section 5.2 correlation table beyond the raw Pearson coefficient comes from `correlation_robustness.py`, and the join-convention band from `fips_fallback_sensitivity.py`. A number that travels without the record that lets a reader check it is the thing this note is about, and that applies to its own numbers.

Conventions. County identifiers are constructed from the Atlas's own `state_id` and `county_id` fields as a five-digit FIPS code. On the EPA side, the county identifier is taken from `FAC_DERIVED_STCTY_FIPS` where present and from `FAC_FIPS_CODE` otherwise; three records carry neither and are excluded. **That fallback order is a choice and it matters.** The three other defensible conventions return 39.2 percent twice and 35.1 percent once, against the 37.9 percent this convention gives; the 35.1 percent convention discards 76 of 482 records, which is why it is not used, and it moves the figure in the direction that would flatter this note. After the section 5.3 exclusion the same four conventions run from 33.8 to 37.0 percent. The headline should be read as approximately 34 to 37 percent `[P]`. `fips_fallback_sensitivity.py` prints all eight. On the ICIS side the county join is by normalized name, which has two known failure modes recorded in `atlas_vs_icis_national.py`: a literal county name of "Undetermined", and parenthesized independent city names. Both are handled. The five county-straddling facilities are resolved by first occurrence; resolving them by county-identifier pair instead moves one county between the 1-to-2 and 3-to-9 bands, and both results are reported in `atlas_count_audit.py`. Version matching uses coordinates rounded to five decimal places, approximately one metre, so a facility whose polygon centroid shifted between versions reads as one removal and one addition; the 89 and 323 figures in section 4 are approximate for that reason.

Verification. This note was put through independent review passes on 17 September 2026. One pass recomputed every quantitative claim directly from the raw files without access to the scripts. One reconciled quotations, evidence grades and the scope of every negative against the sources. Three argued adversarial positions against the note's conclusions, one of them from the position of the report's own research staff. One tested whether every promise the front matter makes is kept by the body, and whether every bound travels with every restatement of a negative.

Findings were incorporated where they held. Every file hash, every repository identifier, every verbatim quotation from the report, the deck and the Atlas README, and every count produced by the distributed scripts was re-verified independently and stands.

The review and source-verification passes were run by language models working cold against the raw files, without access to the drafting context and without the scripts in the recomputation pass. That is stronger than a single reading and weaker than direct verification by the author, and it is marked as its own thing rather than promoted `[P]`. The raw files and the scripts are deposited so that no part of the chain has to be taken on trust.

---

9. Data availability and licensing

Everything needed to reproduce this note is deposited with it, including the two IM3 Atlas CSV files. Unpack the archive and every figure recomputes on Python 3 with the standard library and nothing else.

The deposit carries more than one licence and the distinction is real.

The note, the eight scripts, the build and assembly scripts, the Virginia DEQ transcription and its column structure are the author's work and are released under Creative Commons Attribution 4.0 International (CC BY 4.0). The version of record is 10.5281/zenodo.23441602; 10.5281/zenodo.22819653 resolves to the most recent version.

`IM3_Atlas_v2026.02.09.csv`, `IM3_Atlas_v2025.06_JuneVersion.csv` and `IM3_Atlas_README.txt` are **not** CC BY. The IM3 Open-Source Data Center Atlas is published by Pacific Northwest National Laboratory for the U.S. Department of Energy under the Open Database License 1.0 and is derived from OpenStreetMap, © OpenStreetMap contributors. Anyone receiving those three files from this deposit receives them under ODbL and under no other terms. The source of record is MSD-LIVE, DOI 10.57931/3017294, record p147s-4h760. The two versions here carry md5 `dfe532e258ab233dfe86a92340607098` and `0b34754b5cc41b644d90c9fcc5aa72f8`; the Atlas is versioned and revised, so check the hashes before concluding you hold what this note analysed.

The EPA ECHO extract and the ICIS-AIR extracts derive from United States federal government records and are not subject to copyright in the United States.

Nothing in this deposit is confidential or embargoed, and no unpublished work by any other author is included or quoted.

ABOUT THE AUTHOR

Tommy N. Turner is an independent, self-funded researcher and writer who documents the local-government and energy dimensions of the data center build-out in Virginia. His series of jurisdiction-level case studies covers the counties and the fall-line corridor where Virginia's data center decisions have been made, refused, or contested, and the analytical volumes extend that work into the systems those decisions run on, among them the transmission network, the cost-allocation machinery, and generation. His work spans data center fiscal and environmental policy, local public finance, and the governance questions the two raise, and he writes opinion and analysis on Virginia public affairs for outlets including the *Richmond Times-Dispatch*, the *Virginia Mercury*, the *Virginian-Pilot*, the *Roanoke Times*, *Cardinal News*, and *RTO Insider*. The case studies and analytical volumes are

published as open, citable living editions, revised as the record moves. He works from Williamsburg and Christiansburg, Virginia.

Contact: tommy@gofast.pro · tommynturner.com · ORCID 0009-0005-7422-9881

VERSION AND CITATION

Version 1.1, October 2026. Evidence snapshot September 17, 2026.

Cite this version as: Turner, Tommy N. *Counting Data Centers: The OpenStreetMap Inventory Behind a National County-Level Comparison.* Version 1.1. Zenodo, October 2026. DOI 10.5281/zenodo.23441602.

To cite the work across all versions, use the concept DOI 10.5281/zenodo.22819653, which resolves to the most recent version.

This is a living edition. The registers it examines are live and revisable, and the note says so at the point each is used: the Atlas is versioned and its publisher removes records as well as adding them, EPA's files are refreshed continuously, and the Virginia permit register moves as permits issue. Corrections and additions will be issued as new versions under the same concept DOI.

On errors. Section 6 states the limits the author knows about. For the errors he does not know about, the contact address above is the right route, and a correction that arrives with a document citation will be acted on faster than one that does not. The scripts are included so that a disagreement can be settled by running them.