

# Public records as evidence

A reproducible method for Virginia data-center permit indexing, ambient monitoring transcription, and facility identity resolution.

| Version | Date          | Status                                           |
|---------|---------------|--------------------------------------------------|
| 1.0     | July 22, 2026 | Public methods release - open for defined review |

Author: Public Evidence Project

Contact: [review@publicevidence.org](mailto:review@publicevidence.org)

License: CC BY 4.0 for the project's original arrangement and documentation. Agency records retain their own public-record status and controlling authority.

## No claim of peer review

This paper is a public methods release and an invitation to review defined questions. Cited agencies and researchers are sources, not endorsers or reviewers.

# Executive summary

The Public Evidence Project converts source-linked government records into narrow, reusable evidence products. This paper documents three connected layers: a Virginia DEQ issued air-permit index, a transcription of DEQ's weekly Loudoun County air-monitoring tables, and a candidate queue for linking state records to EPA facility identities.

**The governing rule is simple: each output must say what its unit of observation is, which source controls it, how it was transformed, and what it cannot establish.**

## Principal outputs

- A versioned permit-row index with direct Virginia DEQ document links.
- Forty monitoring statistics transcribed from four numbered DEQ tables dated July 17, 2026.
- Nine sensor-deployment records documenting starts, stops, moves, and maintenance.
- A 25-record federal identity candidate queue that contains no forced EPA matches.
- A five-sheet workbook and public CSV distributions with explicit interpretation boundaries.

## Core findings

- A permit row is a regulatory record, not evidence of current operation or actual emissions.
- APEX/Kunak sensors provide concentration data and trends but are not regulatory instruments for determining NAAQS attainment.
- Name-only facility matching is unsafe because one permit may cover multiple buildings and federal searches may return several similarly named records.
- A documented blank is more useful than a plausible but unverified identifier.

# 1. Research question and evidence boundaries

The immediate research question is not whether Virginia data centers are good or bad. It is whether public records can be made easier to find, cite, compare, and correct without converting administrative records into unsupported claims.

## Units of observation

| Layer          | Unit                             | What it supports                           | What it does not support                                    |
|----------------|----------------------------------|--------------------------------------------|-------------------------------------------------------------|
| Issued permits | One DEQ table row                | Existence of a listed issued permit record | Construction, operation, actual emissions, compliance, harm |
| Monitoring     | One site-pollutant-statistic row | A value printed in a named DEQ table       | Source attribution, attainment, facility-level emissions    |
| Identity       | One permit-to-federal candidate  | A review status and evidence gap           | Identity until address and program linkage agree            |

## Source hierarchy

- Controlling source: the agency page, permit document, monitoring report, or EPA program record.
- Project distribution: a normalized table with provenance and transformation notes.
- Editorial interpretation: a bounded explanation that must remain separable from the underlying record.
- External context: reporting, advocacy, or research used only for clearly attributed context.

**The project does not infer ownership from facility names. It does not add emissions across equipment units unless an official source explicitly provides the total and scope. It does not treat an absence from a search result as evidence of regulatory absence.**

## 2. Permit index pipeline

The issued-permit index begins with Virginia DEQ's table of issued air permits for data centers. The current website release contains the source fields facility, permit, issue date, program type, locality, regional office, and source URL.

### Acquisition and validation

- Retrieve the agency table and preserve the stated source date.
- Reject structurally incomplete responses rather than replacing the collection with an empty or partial result.
- Require all seven fields, a valid date, a direct DEQ document URL, and a unique facility-permit pair.
- Fingerprint the verified table and record added, removed, or changed rows only when the source changes.

### Controlled document-enrichment pilot

A closed 25-document pilot manually reviewed permittee identity, facility address, permit action, equipment references, operating-limit wording, emissions-table scope, and page provenance. It met internal completeness thresholds, but it is not published as a completed equipment or emissions dataset. Ambiguous fields remained blank. Combined capacity, operating status, actual hours, actual emissions, compliance, ownership, and harm were not derived.

| QA check                             | Result      |
|--------------------------------------|-------------|
| Permittee present                    | 25 of 25    |
| Address present                      | 24 of 25    |
| Equipment count present              | 25 of 25    |
| Operating or emissions limit present | 25 of 25    |
| Page provenance for populated rows   | 100 percent |

**These results justify a controlled scale decision, not immediate publication of all extracted conditions.**

# 3. Monitoring table transcription

Virginia DEQ operates a special air-monitoring project in Loudoun and Prince William counties. The July 17, 2026 report presents CO, NO2, and PM2.5 data from APEX/Kunak sensors, with collocated or nearby regulatory monitors used for context.

## Transcription rule

- Accept only values printed in numbered report tables.
- Do not digitize chart traces or estimate points from axes.
- Carry report date, analysis end date, pollutant, averaging basis, site, instrument class, unit, statistic, table number, and source URL.
- Record deployment changes separately from concentration statistics.

| Table | Pollutant and basis | Rows | Unit  |
|-------|---------------------|------|-------|
| 1     | CO hourly           | 10   | ppm   |
| 2     | NO2 hourly          | 10   | ppb   |
| 3     | PM2.5 hourly        | 10   | ug/m3 |
| 4     | PM2.5 24-hour       | 10   | ug/m3 |

## Required qualifications

- DEQ states the sensors are not Federal Reference or Equivalent Method instruments and cannot determine NAAQS attainment.
- The report documents voided periods, sensor stabilization, maintenance, an airflow failure, and regulatory-monitor data-quality issues.
- DEQ identifies July 4 fireworks and wildfire smoke as important period effects in the PM2.5 series.
- A short study-period statistic is not a three-year NAAQS design value.

# 4. Facility identity resolution

Facility identity matters because a state permit record, an EPA FRS registry entity, and an ICIS-Air program record can use different names, addresses, or campus boundaries. The pilot treats federal identifiers as claims that require supporting evidence.

## Acceptance gates

| Gate            | Evidence required                                                | Failure result           |
|-----------------|------------------------------------------------------------------|--------------------------|
| Candidate       | Usable address, coordinate, or parcel description                | Blocked or manual review |
| Physical place  | Street number, street, locality, ZIP or a reviewed spatial match | No identifier            |
| Program linkage | FRS program record supports the same facility                    | No ICIS-Air assertion    |
| Conflict review | Multi-building and similarly named results resolved              | Documented blank         |
| Publication     | Dated EPA URL and short match-basis note                         | Remain in queue          |

## Pilot result

**The first public candidate register contains 25 records, zero forced FRS or ICIS-Air matches, one blocked missing-address case, one out-of-scope record, and a set of manual-review candidates. This is a useful negative result: current evidence is insufficient for a defensible published crosswalk.**

The queue is designed for correction. A reviewer can resolve one facility without accepting the rest of the dataset or endorsing the project.

## 5. Quality assurance and reproducibility

Each public distribution is accompanied by a data dictionary or explicit field descriptions. Files are versioned, downloadable, and linked to controlling sources. The website exposes interpretation boundaries beside downloads rather than burying them in a general disclaimer.

### Release checklist

- Validate row counts, required fields, types, uniqueness, and source domains.
- Review every transformed field against the source definition.
- Confirm that all URLs are direct and current at release time.
- Render and visually inspect the workbook and PDF outputs.
- Run website build, lint, rendered-page tests, and diff checks.
- Publish checksums and archive stable release pages when deployment is complete.

### Correction model

Routine agency refreshes belong in the source-change history. Material factual or methodological corrections belong in a dated corrections log. Private correspondence is not published as external review. A named reviewer appears only with permission, a defined scope, reviewed item, date, outcome, and disclosure boundary.

### Reproducibility boundary

The public outputs make the source trail and transformations inspectable. The project does not claim that every processing component is currently open source. Reproduction means another researcher can retrieve the cited source, understand each field, verify the transcription, and identify where judgment remains.

## 6. Limitations and next tests

- The Virginia permit index is not a national data-center registry and does not cover every environmental approval.
- Agency pages can change, move, or replace documents without notice.
- Permit rows and facilities are not necessarily one-to-one.
- The monitoring layer currently transcribes one weekly report, not raw hourly observations.
- The federal identity queue is intentionally incomplete.
- The monitoring report does not apportion observations to data centers or any other source category.

### Priority next tests

- Add a new monitoring snapshot only when DEQ publishes a changed report; preserve prior releases.
- Resolve a small number of single-address EPA identity candidates using official FRS program links, then have a second reviewer reproduce them.
- Test whether public-notice records can be linked to issued permits without implying that proposals have been approved.
- Request expert review of the unit-of-observation and identity-resolution rules, not a general endorsement.

**The method should be judged by how easy it is to find and correct an error, not by how many institutions appear beside it.**

# References and distributions

Virginia DEQ. Issued Air Permits for Data Centers. <https://www.deq.virginia.gov/news-info/shortcuts/permits/air/issued-air-permits-for-data-centers>

Virginia DEQ. Data Center Air Monitoring Project. <https://www.deq.virginia.gov/news-info/shortcuts/topics-of-interest/data-center-air-monitoring-project>

Virginia DEQ, Office of Air Quality Monitoring. Data Center Air Quality Analysis: Loudoun County Study Area. July 17, 2026. <https://www.deq.virginia.gov/home/showpublisheddocument/37251/639198901999830000>

U.S. EPA. Facility Registry Service API. <https://www.epa.gov/frs/frs-api>

U.S. EPA. ECHO Web Services. <https://echo.epa.gov/tools/web-services>

Public Evidence Project. Virginia Data Center Air Permit Dataset. <https://publicevidence.org/data>

Public Evidence Project. Virginia Data Center Air Monitoring Data. <https://publicevidence.org/virginia/data-centers/air-monitoring>

Public Evidence Project. Facility Identity Crosswalk Pilot. <https://publicevidence.org/research/facility-identity-crosswalk>

Persistent permit-dataset identifier: <https://doi.org/10.7910/DVN/PDK5U4>

## Suggested citation

Public Evidence Project. 2026. Public records as evidence: A reproducible method for Virginia data-center permit indexing, ambient monitoring transcription, and facility identity resolution. Version 1.0. July 22, 2026.

**Defined review is invited at [review@publicevidence.org](mailto:review@publicevidence.org). Please identify the section, disputed rule or field, and the strongest available primary source.**