METHODOLOGY WHITE PAPER

Narrative Geo

Defensible, reproducible geospatial analytics on 106 U.S. retail and restaurant chains.

Version 1.1

Published 2026-06-09

Author Nick Gill, Founder · Nan Point Group

Contact nick@nanpointgroup.com

Coverage Trade-area · saturation · whitespace · cannibalization · demographics · weather sensitivity

This document specifies, in full, the data sources, processing pipeline, computational methods, and known limitations underlying every output the Narrative Geo platform produces.

1. Executive Summary

Narrative Geo produces defensible, board-ready geospatial analyses of U.S. retail and restaurant chain footprints. Each analysis is generated from a fixed pipeline that combines public road-network and demographic data, chain store coordinates reconciled to management-disclosed unit counts, and a documented spatial framework built on Uber's H3 hierarchical hexagonal grid.

The platform's design principle is **methodological transparency**. Every figure cited in a delivered PDF is reproducible from the data sources documented in §2, using the computational methods documented in §§3–8, with the limitations acknowledged in §9 and the audit trail provided in §10.

This document is the citation anchor for any figure the platform produces. It is intended for the sell-side analyst, corporate IR officer, CFO, board committee, and outside auditor who must defend a number sourced from the platform.

---

2. Data Sources & Provenance

The platform consumes three classes of input data, all of which are public-source or licensed-clear, and all of which are version-stamped at the time of analysis.

2.1 Demographic Data — U.S. Census ACS 2023 5-year

Source: U.S. Census Bureau American Community Survey, 5-year estimates, vintage 2023 (covering survey years 2019–2023).

Geographic grain: Block group (~600–3,000 people per BG; ~242,000 BGs nationally).

Variables consumed: Total population (B01001_001E), median household income (B19013_001E), age cohort distribution (B01001 family), educational attainment (B15003 family), household composition (B11001 family), employment status (B23025 family).

Refresh cadence: Annual, per Census release schedule (typically December for the prior vintage).

Geographic boundary file: Census TIGER/Line block-group shapefiles, vintage-matched to the ACS release. Connecticut uses the 2023 vintage to align with the state's 2022 transition from county codes to planning-region codes; the other 49 states plus DC and Puerto Rico use the 2020 vintage.

2.2 Road Network & Drive-Time — OpenStreetMap / OSRM

Source: OpenStreetMap planet extract (United States subset), processed through OSRM (Open Source Routing Machine) version 5.27 with the Contraction Hierarchies (CH) algorithm.

Speed model: The OSRM `car.lua` default profile, which assigns class-of-road speed limits based on OSM `highway=*` tagging (motorway 90 km/h, primary 65 km/h, residential 25 km/h, etc.). Where OSM provides an explicit `maxspeed=*` tag, that value is used; otherwise the profile's class default applies.

Coverage: Contiguous United States (CONUS). Alaska, Hawaii, and U.S. territories are excluded from the trade-area matrix; chains operating exclusively or substantially in those regions are flagged as out-of-scope.

Refresh cadence: Quarterly from the OpenStreetMap weekly planet extract.

Profile note: Drive times are computed under OSRM `car.lua` and reflect free-flow off-peak conditions. They are not directly comparable to figures from routing engines that apply congestion factors by default; quoted reach and ceiling figures must be interpreted under this profile.

2.3 Employment Data — LODES 8

Source: U.S. Census Bureau Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics, version 8 (LODES8), 2022 vintage.

Use: Employment density per H3 cell, used as a co-variate in demographic similarity scoring and in the customer-mix profile output.

2.4 Chain Store Coordinates

Sources: 1. Primary: company-published store locator (scraped quarterly per chain). 2. Secondary: Foursquare/Factual POI extracts cross-referenced against company sources. 3. Tertiary reconciliation: chain 10-K filings, 10-Q updates, and shareholder-letter unit counts.

Reconciliation rule: Every chain's store count is reconciled to the most recent company-disclosed unit count from a 10-K, 10-Q, or 8-K filing. The full reconciliation tolerance and disclosure policy is specified in §6.

Format inclusion: Each chain is documented as company-operated only or company-operated plus licensed/franchised. For company-operated chains, only company-operated stores are counted. For franchised systems (McDonald's, Subway, Domino's, and similar), all operating locations are counted regardless of ownership, with the breakdown disclosed in the PDF appendix. Airport, in-store concession, and seasonal pop-up locations are included or excluded based on

operational substance: full-menu, full-hours formats are included; cart, kiosk, or limited-menu formats are excluded. The rule applied is disclosed per analysis.

Coordinate vintage (as of this version's publication date). Every store record carries a `snapshot_date` field, exported in every store-list CSV. Current distribution:

Vintage cohort	Chains	Store records
May 2026 refresh (direct first-party scrape)	66	176,993
Feb–Mar 2025 baseline (aggregator extract)	40	138,729

Chains on the 2025 baseline are refreshed via the platform's per-chain scraper framework on a rolling basis, prioritized by customer coverage requests; a chain-level refresh takes 2–15 minutes once its scraper exists. The per-chain vintage is disclosed in the methodology footer of any analysis covering that chain, and the §6 reconciliation against company-disclosed unit counts bounds the staleness impact: a chain whose POI count has drifted beyond the §6.1 tolerance is flagged rather than silently reported.

---

3. Spatial Framework: H3 Hierarchy

The platform uses Uber's H3 hierarchical hexagonal grid as the primary spatial index. Two resolutions are used.

3.1 H3 Resolution 8 — The Calculation Grain

All demographic, employment, and trade-area calculations are performed at H3 resolution 8. A res-8 cell covers approximately 0.74 km² (~282 acres, ~180-meter radius). There are roughly 12.9 million res-8 cells in CONUS.

Every input dataset — Census block groups, OSM road network, chain store coordinates, LODES employment — is apportioned to the res-8 grid before any cross-dataset operation is performed. This eliminates the entire class of join-key mismatch issues that arise when polygon boundaries differ across source datasets.

3.2 H3 Resolution 7 — The Map and Report Grain

User-facing maps and report tables aggregate the res-8 calculation grid to H3 resolution 7. A res-7 cell covers approximately 5.16 km² (seven child res-8 cells) and is the resolution at which trade-

area tiles render in the web interface and at which whitespace tables and saturation curves are reported.

Aggregation from res-8 to res-7 is a simple sum over child cells for extensive variables (population, jobs, household counts) and a population-weighted average for intensive variables (median household income, demographic percentages).

3.3 Why H3, Not ZIP Codes or Census Tracts

ZIP codes are a mail-delivery construct with irregular shapes that change yearly and do not nest cleanly. Census tracts are stable but their irregular polygons make set operations — the underlying primitive for cannibalization analysis — computationally expensive and visually noisy. An equirectangular lat/lon grid distorts cell area significantly across latitudes.

H3 provides uniform cell area across CONUS, hierarchical containment for clean cross-resolution aggregation, native polygon arithmetic for trade-area set operations, and a stable cell identifier that does not change between data vintages.

---

4. Drive-Time Computation

4.1 The Matrix

The platform's foundational artifact is a U.S. drive-time matrix: for every (store-anchored origin H3 cell, destination H3 cell) pair within 60 minutes' drive at the OSRM `car.lua` speed model, the matrix stores the integer minute drive-time.

Origins are constrained to H3 res-8 cells that contain at least one store from any of the 106 covered chains (approximately 50,000 origin cells nationally). A full cell-to-cell matrix at res-8 would be 12.9 M × 12.9 M = 165 trillion pairs — infeasible on any hardware. By anchoring origins to store-containing cells, the matrix is reduced to ~100 million stored pairs while still answering every trade-area question the platform supports.

4.2 Validation Gates

Every published matrix release passes four mandatory validation gates:

1. **Non-empty and finite.** Every (origin, destination) entry has a finite integer drive-time or is omitted entirely. Null and infinite values are forbidden in the published matrix.

2. **Monotone reach.** For any drive-time tier T, the set of cells reachable in T minutes is a subset of the set reachable in T+5 minutes.
3. **Plausible per-store reach.** Maximum physical distance reached per drive-time tier implies an effective speed between 20 and 90 mph.
4. **Materialization parity.** Pre-materialized chain trade-area parquets agree with on-the-fly matrix queries on the same chain and cut within 0.5% on cell count and reached population.

4.3 Pre-Materialized Trade Areas

For each covered chain, trade areas are pre-materialized at 5-, 10-, 15-, 30-, and 60-minute drive-time tiers. A trade-area cell is any H3 res-8 cell whose centroid is reachable within T minutes from at least one store in the chain's footprint. The pre-materialized parquets are the input to all white-space, cannibalization, and demographic-profile outputs.

---

5. Demographic Apportionment

5.1 The Core Operation

ACS block-group data is apportioned to H3 res-8 cells before any chain analysis is performed. The apportionment is area-overlap-weighted: for each (block group, H3 cell) pair where the BG polygon intersects the cell hexagon, the cell receives a share of the BG's value equal to the intersection's area as a fraction of the BG's total area.

For extensive variables (population, jobs, household counts), this is mathematically exact under the standard assumption of uniform-within-BG distribution. For intensive variables (median household income, demographic percentages), the platform stores the population-weighted value, which is the appropriate aggregation when later summing across cells.

5.2 National Reconciliation

The platform's national population total reconciles against the official Census ACS 2023 5-year B01001_001E API call within 0.02%. State-level totals reconcile within 0.1%. These reconciliation tolerances are checked on every quarterly demographic refresh.

5.3 Income

The canonical income measure is the population-weighted median household income (`median_hh_income` , ACS variable B19013_001E):

> `SUM(median_hh_income × pop_total) / SUM(pop_total)` over the relevant cell set, filtered to cells where `median_hh_income` is non-null and positive.

National pop-weighted result reconciles to $93,120 against the Census 2023 5-year publication. Any income figure quoted in a delivered PDF is calculated by this method.

---

6. Store-Count Provenance & 10-K Reconciliation

6.1 Reconciliation Rule

Before any analysis is generated, the platform's store count for the subject chain is compared to the most recent company-disclosed unit count from the chain's 10-K, 10-Q, or 8-K filing:

Discrepancy	Action
< 0.5%	Proceed; report the platform figure with a footnote citing the source.
0.5%–2%	Proceed; print both figures in the PDF appendix with a one-line reconciliation note.
> 2%	Halt. Refresh the store-coordinate scrape. If still > 2%, manual reconciliation by Narrative Geo staff before PDF generation.

6.2 Treatment of Licensed and Non-Traditional Locations

For each covered chain the platform documents which formats are included (full-service company-operated, full-service franchised, licensed in-store concessions, airport locations, kiosks) and the most recent date of company-disclosed unit count by format.

When a chain reports unit counts under multiple format definitions (e.g., Starbucks separates "company-operated" and "licensed"), the platform's analyses are run under the format definition that matches the customer's question. Alternative definitions are available on request as an additional PDF deliverable.

7. Saturation Curve Methodology

7.1 The Question

A long-term unit guide — "we can grow to 7,000 stores" — is implicitly a geographic claim. Somewhere in the United States there must be a sufficient number of distinct catchment areas that resemble the chain's existing markets, and each next store must open in one of them without simply cannibalizing the trade area of a store the chain already operates. The methodology in this section answers this question directly:

> Starting from the chain's existing footprint, in what order would the best next stores actually be opened — and at what store count does the next store reach so few additional people that it cannibalizes more than it adds?

The figure reported as the "geographic ceiling" is the answer to that question. **No parametric curve is fitted at any stage.** The platform simulates opening stores one at a time, in the best possible order, and stops when the next store no longer earns its way against the chain's demonstrated economics.

7.2 The Algorithm

The model produces the ceiling through a greedy max-marginal-coverage simulation with explicit cannibalization accounting.

1. Candidate identification. Every H3 cell outside the chain's existing trade area is scored on demographic similarity to the markets where the chain operates today. Scoring uses five axes: total population, median household income, age cohort distribution, an opportunity-size factor, and a broader demographic look-alike index built from housing, education, ethnicity, and daytime-employment features. Cells passing a minimum-population floor and a minimum-similarity floor become the candidate pool (typically ~1,500 cells per chain).

2. Ranking. Candidates are sorted by similarity score, highest to lowest.

3. Sequential selection. The algorithm walks the ranked list and, for each candidate, computes the population inside the candidate's drive-time trade area that is **not** already covered by either the

chain's existing footprint or any previously selected candidate. This is the candidate's *marginal population* — the people it would reach who are not already served. If the marginal population meets the threshold defined in §7.3, the candidate is selected; otherwise it is rejected.

4. Cannibalization is computed per candidate, not estimated in aggregate. A high-scoring candidate three miles from one already selected may be rejected because its trade area is already largely covered by the prior selection. The model knows exactly which population cells have been claimed and subtracts them.

5. The ceiling. The geographic ceiling is the count of candidates whose marginal population at selection met the threshold. Candidates accepted but flagged "below saturation" are reported separately — the simulation includes them, but they fall below the per-store reach threshold the chain's existing footprint demonstrates.

The reported output is a discrete cumulative-coverage curve: store count on the x-axis, cumulative population reached on the y-axis, with a vertical line at the ceiling count.

7.3 The Threshold

A store is counted "above saturation" if its marginal population is at least **30% of the chain's average existing per-store reached population**. The 30% threshold is configurable; 30% is the platform default and is the figure reported in every delivered PDF unless explicitly noted.

The intuition matches how an operator already evaluates a marginal new opening:

> A chain whose 500 existing stores each reach an average of 200,000 people in their trade area is operating at a baseline of 200K per store. A hypothetical 501st store that would reach only 50,000 incremental people (25% of the baseline) is opening at roughly a quarter of the current portfolio's reach per location. The 30% threshold is the level at which most operators we have spoken to begin to underwrite a candidate as a "fill-in" location requiring different standards than a greenfield open.

Lower thresholds produce higher ceiling figures; higher thresholds produce lower, more conservative ceilings. Customers can request the simulation re-run at alternative thresholds; the reported figure changes monotonically and predictably with the threshold.

Worked sensitivity example. The table below is the actual platform output for Cava (472 stores, 30-minute drive-time) re-run at four thresholds on 2026-06-09 — the sensitivity any reviewer can request for any chain:

Marginal-reach threshold	Geographic-expansion stores selected	Implied geographic ceiling (total units)
20% of avg per-store reach	350	822
30% (platform default)	**307**	**779**
40%	269	741
50%	243	715

The ceiling moves −13% as the threshold doubles from 20% to 50% — monotone and bounded, with no parametric cliff. A chain near its geographic ceiling (e.g., McDonald's at ~13.6K units) shows near-zero threshold sensitivity because the candidate pool, not the stop rule, is the binding constraint. The intensification component (§7.6) is threshold-invariant by construction.

7.4 Why This Approach Is More Defensible Than Curve-Fitting

The two alternative methods most commonly used in retail saturation analysis — linear extrapolation and three-parameter logistic-curve fitting — both have structural weaknesses that the greedy approach avoids.

Linear extrapolation assumes the marginal store reaches the same incremental population as the average existing store. This assumption is false for any chain with overlapping trade areas. The method ignores cannibalization entirely, ignores demographic fit, and systematically overstates ceilings for urban-concentrated chains while understating them for chains that have already saturated their core demographic.

Logistic-curve fitting estimates three free parameters (asymptote, growth rate, inflection point) from the chain's historical reach trajectory. The sigmoid shape is an assumption rather than an observation; the asymptote is weakly identified when the chain's footprint is small relative to its theoretical ceiling; and the method cannot answer *where* the next stores would go, whether they would be in look-alike markets, or how they would cannibalize each other.

The greedy approach answers each of these on the only basis that is empirically grounded — the chain's own demonstrated geographic and demographic footprint:

Question	Linear	Logistic	Greedy set-cover
Accounts for cannibalization between candidate stores?	No	Implicit only	**Yes — per candidate**
Identifies *where* the marginal stores would be?	No	No	**Yes — specific H3 cells**
Uses the chain's demonstrated economics as the stop rule?	No	Fitted asymptote	**Yes — threshold relative to per-store reach**
Auditable step-by-step?	One formula	One regression	**Yes — full per-candidate ledger**
Robust to chains whose history doesn't fit a sigmoid?	n/a	No	**Yes — no fitting**

A reviewer who disagrees with the 30% threshold can request a re-run at any other threshold and see how the ceiling shifts. A reviewer who disagrees with a logistic fit can only argue about whether the fit is correct.

7.5 What This Algorithm Does Not Estimate

The ceiling answers a supply-side question: how many geographically distinct trade areas exist that look like the chain's existing markets and have not been claimed by an earlier store in the same simulation. It does not answer:

- **Whether the marginal store would be profitable.** The threshold uses reached population as a proxy for unit-economics fit; real unit economics depend on rent, labor, daypart mix, and competitive density that the platform does not measure. The reported ceiling is a geographic upper bound, not a profitability ceiling.
- **Whether the chain's actual opening order would resemble the algorithm's order.** Real-world site selection is constrained by lease availability, capital allocation, and regional manager bandwidth. The algorithm produces a best-possible-ordering trajectory; the chain's realized trajectory will lie at or below it.
- **Whether competitor moves would erode the marginal economics.** Cells remain in the candidate pool regardless of subsequent competitor expansion.
- **What the parametric shape of the curve is.** No parametric curve is fitted; the simulation's discrete output is the model output.

These limitations are stated explicitly in the methodology footer of every PDF that includes a saturation curve or ceiling figure.

7.6 Intensification: A Parallel View

In parallel with the geographic-expansion model above, the platform reports a second figure addressing a different question:

> Holding the chain's set of operating counties fixed, how many more stores can fit in counties the chain already operates in, before each county's uncovered population is exhausted?

For each county where the chain has at least one store, the platform computes the population inside the county that lies outside the chain's existing drive-time footprint (the uncovered population), divides by the chain's national average reached population per store, and reports the result, capped at 2× the chain's current store count in that county. The 2× cap is an operational-realism guardrail reflecting the practical constraints of densifying an already-dense metro.

Intensification and geographic expansion answer different questions and are reported as additive components of a combined ceiling, with a chain-level realism cap applied to the sum.

---

8. Cannibalization Methodology

8.1 Pairwise Trade-Area Set Operations

For two chains A and B, the pairwise cannibalization analysis computes the following set operations on their respective trade-area cell sets at each drive-time tier T:

- **|A ∩ B|** — cells reachable by both A and B within T minutes
- **|A − B|** — cells reachable by A but not B
- **|B − A|** — cells reachable by B but not A
- **|A ∪ B|** — cells reachable by either

The reported "overlap percentage" is **|A ∩ B| / |A|** — the fraction of A's trade area also covered by B. This is the operationally useful figure for the question "what fraction of A's footprint is contested by B?"

8.2 Population-Weighted vs Cell-Count-Weighted

The platform reports both population-weighted and cell-count-weighted overlap. The population-weighted version answers "how many people live in the contested zone." The cell-count-weighted version answers "how much geographic area is contested." For chains operating across dense urban metros and sparsely populated rural areas, the two figures can diverge significantly; both are reported in every cannibalization PDF.

8.3 Drive-Time Tier Reporting — All Five Tiers, Always

Every cannibalization analysis reports the full five-tier overlap curve (T = 5, 10, 15, 30, 60 minutes) — in the platform UI, in delivered PDFs, and via the CSV audit trail. A single-tier overlap number is not a deliverable of this platform: 30-minute overlap is structurally high for any pair of national chains (most metros are less than 30 minutes wide), and a number quoted without its curve invites the legitimate objection that the tier was chosen for effect. The 10- or 15-minute tier is typically the most discriminating for retail competitive analysis.

Worked example (actual platform output, 2026-06-09). Cava × Chipotle Mexican Grill, population-weighted union overlap by tier:

Drive-time tier	Union overlap (|A∩B| / |A∪B|)	Share of Cava's footprint contested (|A∩B| / |A|)
5 min	13.8%	87.8%
10 min	26.5%	97.2%
15 min	40.0%	98.4%
30 min	55.0%	98.8%
60 min	62.5%	98.7%

The two columns answer different questions. The union overlap grows with the tier as both footprints expand into shared metros. The directional column shows that even at a 5-minute drive — the strictest tier — 87.8% of Cava's footprint already sits inside Chipotle's reach: a statement about Cava's siting strategy, not an artifact of a generous drive-time assumption.

---

9. Known Limitations

Every PDF citing platform output includes the relevant subset of these limitations in its methodology appendix.

1. **OSRM `car.lua` speed model.** Does not apply rush-hour congestion factors. Reach figures should be interpreted as off-peak free-flow estimates.
2. **ACS 5-year smoothing.** The 2023 5-year covers survey years 2019–2023 and reflects an average across that period. Rapidly changing markets (e.g., Sun Belt growth corridors) are systematically understated in their current-year demographics by this smoothing.

3. **Closed and relocated stores.** Store-coordinate scrapes capture currently operating locations. Stores closed in the past quarter may persist until the next quarterly refresh.
4. **Acquisition and renaming events.** When chains acquire or are acquired (e.g., a regional concept rolled into a national brand), the platform's chain definition follows the company's current reporting entity. Historical analyses use the entity that was operating at the time, not the current entity.
5. **Licensed and franchised location ambiguity.** As described in §6.2, the inclusion or exclusion of licensed/concession formats is documented per chain but may differ from a customer's preferred definition. Alternative definitions are available on request.
6. **Saturation curve as a supply-side bound.** As described in §7.5, the curve characterizes geographic reach, not unit economics. Reported ceiling figures are not profitability ceilings.
7. **Cannibalization is geographic, not behavioral.** The platform measures overlap of geographic trade areas, not observed customer cross-shopping. Two chains with high geographic overlap may have low behavioral overlap (different customer demographics, dayparts, or occasions) and vice versa.
8. **CONUS-only coverage.** Alaska, Hawaii, U.S. territories, and international footprints are out of scope. Chains with material exposure outside CONUS are scoped to CONUS only, with the scope disclosed in the PDF.
9. **Quarterly refresh cadence.** Store footprints and demographics are refreshed quarterly. Custom-tier customers can request out-of-cycle refresh; the underlying data sources (ACS, OSM) update on their own schedules and a custom refresh does not produce data more recent than the upstream source.
10. **Methodology versioning.** This methodology document is version 1.0. Material methodology changes trigger a version increment and a 30-day notice to active Custom-tier customers. Version 1.0 outputs remain reproducible from the inputs documented in §2.

---

10. Reproducibility Specification

Every delivered PDF includes a provenance footer allowing an external auditor or sell-side analyst to reproduce the result from the underlying data.

10.1 Data Vintage Stamps

Every delivery cites the exact vintages used:

- Census ACS vintage and download date.
- Census TIGER vintage (2020 for 49 states + DC + PR; 2023 for Connecticut).

- OpenStreetMap planet-extract date and CONUS subset specification.
- LODES vintage (currently LODES8 2022).
- Chain store coordinates — scrape date, reconciliation source (10-K, 10-Q, or 8-K), and the disclosed unit count cited.

Vintage fields are populated automatically at generation time and printed in the methodology footer of the PDF.

10.2 Software Versions

- **OSRM:** version 5.27, `car.lua` profile (unmodified default).
- **H3:** version 4.1.
- **Routing graph:** Contraction Hierarchies.

10.3 Output Specification

Every analysis produces a structured output package:

- **PDF deliverable** with branded cover, methodology footer, and the customer's logo (Custom tier).
- **Underlying tables in CSV on request** — every chart in the PDF backed by a CSV row-and-column dump.
- **H3 cell IDs** for any whitespace, cannibalization, or trade-area output, allowing the customer to re-plot the geography in their own GIS environment if desired.

10.4 Methodology Attestation

The methodology described in this document is attested to be the procedure used to generate the corresponding PDF on its generation date, by Nick Gill, Founder, Nan Point Group. Attestation is included as the final page of every Custom-tier delivery.

10.5 Version-Pinned Reproducibility

Custom-tier customers receive a version-pinned URL for every deliverable. The version-pinned URL holds the methodology version, data vintage, and software stack active at the time of generation, immutable for the duration of the contract. If a customer cites a figure from the PDF in an SEC filing, an analyst-day deck, or a published research note, the version-pinned URL serves as the citation anchor.

10.6 Auditor's Appendix — The Self-Serve CSV Audit Trail

Every figure the platform reports is backed by a live, licensed export endpoint that returns the underlying rows — not "on request," but as a URL the auditor runs themselves with their license token:

Figure under audit	Export	Rows returned
Store count, store locations	`/api/store-list/export?chain={C}`	One row per store: id, address, lat/lng, format flags, `snapshot_date`
Trade-area population	`/api/trade-area/export?chain={C}&drive_min={T}`	One row per H3 cell: cell id, lat/lng, ACS population, median household income
Cannibalization buckets	`/api/overlap/export?a={A}&b={B}&drive_min={T}`	One row per cell in A∪B with its bucket label (a_only / b_only / both) and population
Whitespace candidates	`/api/whitespace/export?chain={C}&drive_min={T}`	Full ranked candidate list with composite score components
Site-selection scoring	`/api/site-selection/export?h3={cell}&chain={C}`	Every fingerprint row, score component, demographic, and workforce metric for the cell

Each CSV is stamped with five footer lines: the methodology line, the data refresh date, the exact source URL (the reproduction recipe), the generation timestamp, and the license holder. Summing the population column of the trade-area export reproduces the headline trade-area figure exactly — the export and the headline are computed from the same H3 res-8 rows, not parallel pipelines. An auditor who disputes any figure in a delivered PDF can re-derive it from these endpoints without contacting us.

10.7 License to Publish

Narrative Geo deliverables are licensed for attributed external use through the License-to-Publish Rider attached to every paid agreement. The rider grants, without further approval per use:

1. **SEC filings** — citation of platform figures in 10-K, 10-Q, 8-K, and proxy materials, with attribution "Source: Narrative Geo (methodology v{X}, data refresh {date})".
2. **Investor-relations channels** — IR websites, investor-day decks and livestreams, earnings-call prepared remarks and Q&A.
3. **Sell-side distribution** — figures and charts reproduced in third-party research notes covering the licensee.

4. **Board and internal strategy materials** — unrestricted internal use.

Conditions: attribution with the version-pinned citation anchor (§10.5); no resale or redistribution of the underlying cell-level data; figures must be quoted with their stated drive-time tier and threshold. The rider survives contract termination for any figure published during the contract term — a cited figure never becomes retroactively unlicensed.

Appendix A — Chain Coverage

The platform covers 106 U.S. retail and restaurant chains. The full list, with industry grouping and most-recent company-disclosed unit count, is maintained at `narrativegeo.nanpointgroup.com/coverage`. Custom chain ingestion (private companies, regional concepts, international parents with U.S. footprints) is available on Custom-tier contracts with a 5-business-day SLA.

Appendix B — Glossary

- **H3:** Uber's hierarchical hexagonal geographic indexing system.
- **H3 resolution 8:** A grid cell ~0.74 km² (~12.9 million cells in CONUS).
- **H3 resolution 7:** A grid cell ~5.16 km² (~1.84 million cells in CONUS).
- **OSRM:** Open Source Routing Machine, an open-source routing engine.
- **Contraction Hierarchies (CH):** A pre-computation step that accelerates shortest-path queries on a road network by hierarchically contracting the graph.
- **ACS 5-year:** U.S. Census American Community Survey, 5-year estimates — the standard small-area demographic dataset.
- **LODES:** Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics.
- **Trade area:** The set of H3 cells whose centroids are reachable within a specified drive-time threshold of at least one store in the chain's footprint.
- **Saturation curve:** The function relating cumulative population reach to cumulative unit count.
- **Whitespace cell:** An H3 cell not currently covered by the chain's trade area at the default drive-time tier.
- **Cannibalization (in the platform):** The geographic overlap between two chains' trade areas, measured as $|A \cap B| / |A|$ at a specified drive-time tier. Not a behavioral measure of cross-shopping.

© 2026 Nan Point Group · Narrative Geo · Methodology version 1.1, published 2026-06-09 (v1.0: 2026-05-15). Changes in v1.1: worked threshold-sensitivity table (§7.3), five-tier cannibalization reporting as default with worked example (§8.3), POI coordinate-vintage disclosure (§2.4), self-serve auditor's appendix (§10.6), license-to-publish terms (§10.7). Comments and questions: nick@nanpointgroup.com.