Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

NPPES Data for Healthcare Market Mapping and Resolution

> Build a more reviewable acquisition pipeline with Docket. Our deal-origination software and managed research help private equity teams screen companies against a mandate and understand the evidence behind each finding.

Published by Docket · 2026-10-13 · nppes data for healthcare market mapping · nppes · npi registry · provider entity resolution · healthcare acquisition research · practice location data · type 1 npi · type 2 npi · healthcare market mapping

Original article: https://docket.capital/articles/nppes-healthcare-target-maps-provider-resolution



In this article

1. Executive Summary
2. Introduction and Background
3. Key Changes
4. Provider Identity and Location Model
5. Implementation Considerations and Process Changes
6. Corroborating Organizations and Ownership
7. Data Analysis and Evidence
8. Hypothetical Worked Examples
9. Implications and Future Directions
10. Frequently Asked Questions (FAQs)
11. Conclusion

Executive Summary

NPPES data for healthcare market mapping works best as a provider-identity and reported-location layer. The **National Plan and Provider Enumeration System (NPPES)** assigns a **National Provider Identifier (NPI)**, a unique **10-digit** provider identifier. (www.cms.gov) Its **Type 1** and **Type 2** categories distinguish individuals from non-person providers. (www.federalregister.gov) Organizational subparts can also have identifiers without being separate legal businesses. (www.cms.gov) The useful output is a dated crosswalk connecting provider observations, locations, and independently verified legal entities, with ownership recorded separately.

The operational baseline has changed. The **Centers for Medicare & Medicaid Services (CMS)** lists the **October 12, 2026 Version 2** monthly release and states that Version 1 support ended effective **March 3, 2026**. (download.cms.gov) The Registry updates **daily**, alongside monthly bulk replacements and weekly increments. (www.cms.gov) Supporting files add organization names and non-primary practice locations. (download.cms.gov) Address interpretation needs care: the federal notice effective **April 3, 2024** permits specified home-based individuals to report a mailbox instead of their home location. (www.federalregister.gov) A reported location therefore needs separate classification before appearing on a physical-site map.

External corroboration should preserve relationship meanings. The **National Uniform Claim Committee (NUCC)** describes taxonomy codes as **self-selected**, rather than evidence of services rendered or licensure. (www.nucc.org) University of Washington workforce guidance recommends crossing NPI supply records with other provider sources. (familymedicine.uw.edu) Medicare reassignment permits another entity to bill, while **Agency for Healthcare Research and Quality (AHRQ)** system linkages can represent common ownership **or joint management**. (data.cms.gov) (www.ahrq.gov) Those relationships should remain separately labeled unless further evidence supports ownership, employment, or legal-entity consolidation.

This report supplies a field dictionary, update workflow, and resolution worksheet. Record **deterministic**, **reviewed**, and **unresolved** links separately. The **Office for National Statistics (ONS)** cautions that match rates do not establish linkage quality. (www.ons.gov.uk) Census definitions also distinguish physical establishments from enterprises under common control. (www.census.gov) Historical reconstruction requires a

missing-file inventory because the **National Bureau of Economic Research (NBER)** warns that some months and years are incomplete or missing. (www.nber.org) Accepted-link counts are operational measures, not market share or acquisition availability. The decision rule is to accept the relationship the source actually establishes, preserve conflicts and collection dates, and obtain separate evidence before rolling records into a target organization.

Introduction and Background

A private equity healthcare-services analyst may begin with an apparent list of practices and discover that it actually mixes clinicians, incorporated providers, organizational subparts, mailing addresses, and old identifiers. NPPES contains both individual and organization records; the **Research Data Assistance Center (ResDAC)** makes that mixed population explicit. (resdac.org) (www.federalregister.gov) The first analytical task is therefore to define what a row represents before counting it or attaching it to a company.

An **enumerated provider** is the person or organization assigned an NPI. A **reported practice location** is an address observation attached to that identifier. A **candidate legal entity** is a business that may correspond to an organization record. A **target organization** is the analyst's researched acquisition unit. These are proposed working definitions for this report, and the crosswalk should preserve their differences rather than force them into one table of companies.

That distinction has a statistical analogue. The Census Bureau defines an establishment as a single physical location, while its enterprise concept aggregates domestic establishments under common ownership or control. (www.census.gov) An NPI is a provider identifier rather than either Census unit. The report consequently addresses provider resolution and reported-location coverage, with independently evidenced legal-entity links, rather than estimating the number of businesses available for acquisition.

Docket is a direct provider of sourced acquisition-target research. Its published workflow assigns identity checks to Triage, answer collection to Scout, and evidence review to Audit. Its broader library covers market mapping and acquisition-data sources. This analysis addresses the narrower import decision: when an NPPES observation can be linked to an organization, what kind of relationship is supported, and which questions must remain open.

Source collection occurred at publication; source-specific vintages are retained below. A retrieval date describes when evidence was collected; it does not make an older study, filing, or address current. The operating recommendations are a proposed research protocol, not the results of a tested NPPES matching system.

Key Changes

Version 2 requires a deliberate schema check

CMS's download index says Version 1 monthly and weekly files are no longer supported and that Version 2 extends the character limits for **First Name** and **Legal Business Name**. (download.cms.gov) A pipeline that truncates names to a legacy width risks creating avoidable collisions or losing distinguishing text. The

appropriate response is to validate the current archive's header and bundled documentation before importing, then retain the complete source values.

The download index is authoritative for the release offered, but this report does not publish unverified numerical field widths. Legacy documentation can help explain familiar field names without establishing the current maximum length. Treat header changes as a schema event with an explicit review, rather than letting a successful parser imply that every column still means the same thing.

Recommended release controls are:

- **Release manifest:** Record the publisher label, archive filename, retrieval time, byte count, and checksum.
- **Schema comparison:** Compare header names, order, data types, and accepted null values with the installed schema.
- **String preservation:** Import NPIs, Zone Improvement Plan (ZIP) codes, license numbers, and registry identifiers as text.
- **Change review:** Hold unexpected columns or truncation for review before publishing a target map.
- **Rollback:** Retain the previous accepted release and the transformation version used to create it.

The string recommendation also has a documented practical basis: AHRQ warns that spreadsheet software may drop leading zeros when opening linkage comma-separated value (CSV) files. (www.ahrq.gov) The archive should be treated as a source dataset, with a smaller analyst extract generated after controlled ingestion.

Practice-address interpretation has changed

The federal NPPES notice allows a Type 1 provider whose only physical location is a home address to enter a post office box or a private-delivery-service personal mailbox as its provider location address. (www.federalregister.gov) That exception makes an address label insufficient evidence that a street address represents a patient-facing office.

The practical rule is to attach an **address-purpose classification** to every observation: reported practice, reported mailing, confirmed service site, mailbox, or unresolved. A mailbox can remain useful for identity reconciliation while being excluded from a physical-site map. Do not convert that exclusion into an assumption that the provider has no service footprint.

More locations and names create relationships, not extra targets

The supporting files include additional organization names and non-primary practice locations. (download.cms.gov) Their presence creates more observations attached to existing NPIs. It does not itself create more independent acquisition organizations. Preserve the original name type and whether a location came from the main record or a reference file.

The endpoint reference file adds another information category associated with both provider types. (download.cms.gov) Keep it separate from site observations. An electronic endpoint should not become a physical branch merely because it shares an NPI. The overall schema change is to model repeatable attributes as child records, each carrying its own provenance.

Provider Identity and Location Model

Type 1 and Type 2 answer different questions

The federal description defines Type 1 as an individual human being furnishing health care and Type 2 as an entity other than an individual human being furnishing health care. (www.federalregister.gov) An individual does not become an organization merely because a business name appears nearby. CMS's enumeration guidance includes sole proprietors among Type 1 providers and describes an individual as eligible for only one NPI. (www.cms.gov)

Conversely, an organization identifier need not identify the complete economic business. CMS's guidance describes organizational subparts that can furnish care without being separate legal entities. (www.cms.gov) Keep a **provider-to-legal-entity relationship** distinct from a **legal-entity-to-parent relationship**. Multiple organization NPIs can remain separately traceable even after evidence supports grouping them into the same researched target.

This approach also prevents address changes from becoming artificial new providers. CMS states that an assigned NPI remains the same when the provider changes its name or address. (www.cms.gov) Store new observations against the existing identifier, with their source dates, instead of creating an unrelated entity simply because normalized text differs.

Reported fields and current relationships

NPPES data should be read as reported observations, with the limits of provider maintenance made explicit. CMS describes the information as reported by the provider or someone acting on its behalf. (www.cms.gov) University of Washington guidance says providers are responsible for updating their records. (familymedicine.uw.edu) This makes a populated field useful evidence of what was reported, while leaving current practice activity as a separate question. Taxonomy has its own self-selection limitation. (www.nucc.org)

Search behavior can broaden the candidate set without establishing equivalence. The Registry **application programming interface (API)** searches organization names across legal business names, doing-business-as names, former legal names, and other names. (npiregistry.cms.hhs.gov) An alias hit should retain the matched name type. NBER also cautions that identifier crosswalks depend on the other identifier having been supplied. (www.nber.org) Missing values should create a research gap rather than a negative finding about affiliation.

An address is a candidate relationship

Normalize addresses into comparable components while retaining the complete raw text. USPS guidance requires secondary unit designators where applicable and prefers them at the end of the delivery-address line. (pe.usps.com) A suite is therefore a meaningful component, not noise to remove during deduplication.

Use separate keys for a **building** and a **unit**. Matching the building can generate a research lead; matching a unit can strengthen a candidate relationship. Neither establishes common ownership. The operating policy should distinguish address equivalence from entity equivalence, even when the same reviewers assess both.

The API distinguishes the primary practice location from the mailing address in its Addresses array. (npiregistry.cms.hhs.gov) Preserve that source role before creating address comparison keys. A mailing address should not enter a service-site count merely because it geocodes successfully. USPS also advises against replacing a known unit designator with a pound sign. (pe.usps.com) Retaining the actual unit supports comparisons under the published secondary-unit convention. (pe.usps.com)

An additional safeguard is to preserve missing-unit information explicitly. A blank suite and a populated suite do not constitute contradictory values in every case, but they also do not justify a precise unit match. Flag the comparison as incomplete and seek a company location page or other dated corroboration.

Taxonomy supplies classification evidence

NUCC maintains the taxonomy code set. (www.nucc.org) It describes a **10-character** alphanumeric identifier and a hierarchy of provider grouping, classification, and specialization. (www.nucc.org) Retain all reported codes, their descriptions, and the primary indicator rather than filtering exclusively on a single preferred code.

Taxonomy is useful for generating and segmenting candidates, with boundaries. NUCC says codes are self-selected, do not define services rendered, and are outside the scope of establishing licensure. (www.nucc.org) An analyst can therefore use a code to open a specialty research queue, but should verify actual service offerings separately before applying a mandate criterion.

For implementation, distinguish the provider's reported classification from the analyst's assessed business activity. A multi-specialty group may require several reported categories; an individual specialist's code should not silently become the only classification of a related organization. Preserve both the evidence and the transformation that produced the assessed category.

Implementation Considerations and Process Changes

Download and update workflow

ResDAC warns that the complete downloadable file is too large to open in Microsoft Excel. (resdac.org) Use a database or another controlled bulk-processing environment for ingestion, then export a scoped worksheet for analysts. NBER's reshaped files illustrate separating repeated values from core variables and linking them by NPI. (www.nber.org)

The recommended update sequence is:

- **Install the baseline:** Download the current monthly replacement and inventory its constituent files.
- **Validate the release:** Confirm headers, row parsing, duplicate identifiers, and the manifest before acceptance.
- **Load repeated attributes:** Keep other names, non-primary locations, and endpoints in separate child tables.
- **Apply weekly changes:** Process increments in publisher-period order, retaining the source week for each observation.
- **Reconcile the next month:** Replace the operational baseline with the next accepted monthly replacement.

- **Verify selected records:** Use Registry lookups for specific unresolved or recently changed identifiers.
- **Track interruptions:** Record missing weeks, failed imports, and the last fully accepted baseline.
- **Preserve research links:** Re-evaluate affected links while retaining earlier reviewer decisions and their evidence.

CMS describes weekly files as supplements containing new NPIs, updated records, and newly deactivated NPIs. (www.cms.gov) Do not treat an indefinitely accumulated sequence of weekly extracts as a substitute for monthly reconciliation. For supporting reference files, inspect the release documentation and scope before deciding whether a child table is replaced or amended.

The API is useful for targeted checks, with **200 results per request** and a documented maximum of **1,200 records across six requests** for a search using its limit and skip parameters. (npiregistry.cms.hhs.gov) Its documentation directs applications to **Version 2.1**. (npiregistry.cms.hhs.gov) A bounded name-search response is consequently not evidence that a geographic market has been exhaustively enumerated.

Field dictionary and entity-resolution worksheet

Table 1 translates source attributes into a proposed worksheet. Source field naming and the separation of secondary locations can be checked against the Registry's CSV-to-JavaScript Object Notation (JSON) mapping; the worksheet's review fields are analyst-created. (npiregistry.cms.hhs.gov)

Worksheet column or field group	Source or working definition	Resolution use and control
NPI type/status	NPI, Entity Type Code, and current deactivation or reactivation evidence	Keep individual and organization populations separate; preserve status as of the release.
Enumeration/deactivation date	Provider Enumeration Date, NPI Deactivation Date, NPI Reactivation Date, and Last Update Date	Preserve separate lifecycle dates; an update date does not validate every field.
Organization/other name	Provider Organization Name (Legal Business Name), other organization names, and their name types	Keep legal-name and alias observations distinct; normalize for comparison only. (npiregistry.cms.hhs.gov)
Taxonomy	Reported taxonomy codes, primary switches, and code descriptions	Preserve multiple classifications; use them for candidate generation. (npiregistry.cms.hhs.gov)
Primary/non-primary location	Main practice-location fields and practice-location reference rows	Label the source of each observation; do not turn an extra location into an extra provider.
Raw address	Original address lines, unit, city, state, postal code, country, and address purpose	Retain spelling and missing components for review.
Normalized address	Parsed building and unit components plus normalization version	Support candidate matching while preserving distinctions between suites.
Candidate legal entity	Proposed legal name, registry jurisdiction, registry ID, and operating website	Leave unresolved until identity evidence supports the link.
Match rule	Exact identifier join, reviewed identity match, or unresolved candidate	Describe the specific relationship accepted, including its direction.

Worksheet column or field group	Source or working definition	Resolution use and control
Conflict	Competing names, locations, entity types, dates, or relationship evidence	Preserve both observations and the reason for any disposition.
Corroborating URL	Exact external source and relevant supporting passage	Store source scope: license, legal identity, billing, management, or ownership.
Collection date	Retrieval timestamp, source vintage, release ID, and reviewer decision date	Keep retrieval and effective dates separate.

The table's purpose is reproducibility. Every accepted organization link should be explainable without reconstructing the analyst's browser history. Store the link as a record with evidence rather than writing the candidate legal entity back over the original NPPES name.

The **World Wide Web Consortium (W3C) Provenance (PROV)** model offers a relevant provenance pattern: activities create or change entities, and agents can include people or software. (www.w3.org) Applied here, the input release, normalization process, and reviewed crosswalk are separate objects. A rerun can then explain which links changed because of new source data and which changed because of a revised matching rule.

Deduplicate observations before consolidating organizations

An exact repeated NPI within the same accepted source population should prompt an import-grain check before a fuzzy match. Repeated taxonomy or location rows may be valid child observations, not duplicate providers. NBER explicitly separates repeated fields from nonrepeated core fields. (www.nber.org)

Recommended match states are:

- **Deterministic:** The same authoritative identifier connects records within its defined namespace.
- **Reviewed:** A reviewer accepts a candidate link after checking entity-specific, dated evidence.
- **Unresolved:** A plausible relationship remains unaccepted because evidence is incomplete or conflicting.
- **Rejected:** Available evidence identifies different entities or a relationship outside the rule's scope.
- **Rule scope:** Separate provider identity, affiliation, legal identity, and ownership tests.
- **Review queue:** Assign unresolved conflicts an owner and an explicit next evidence request.

The Office for National Statistics describes deterministic methods as exact matching. (www.ons.gov.uk) Exactness is a property of the rule, not a guarantee that the relationship means ownership. An exact NPI join resolves provider observations; it does not turn the provider into its employer or a parent company.

The rule register should specify the namespace, required fields, excluded cases, and supported edge type. Exact matching is an operational technique. (www.ons.gov.uk) Link volume remains distinct from quality. (www.ons.gov.uk) Where a rule is evaluated, measure precision against reviewed true relationships. (www.ons.gov.uk) Preserve revised outputs alongside prior decisions so later reviewers can reconstruct the change, consistent with the PROV revision model. (www.w3.org)

Name and address similarities should generate candidates for reviewed matching. Retain name suffixes and meaningful specialty terms in the raw value, even when comparison keys remove punctuation or common legal suffixes. A former name is an alias observation rather than permission to merge every current business

using similar words.

Corroborating Organizations and Ownership

Compare evidence by the question it answers

Table 2 compares the roles of the principal evidence sources and the research service. Docket appears because it directly provides sourced target research; its documented offering is research with supporting evidence rather than an official provider register.

Source or research layer	Useful evidence	Boundary for target consolidation
NPPES	Enumerated individual or organization, reported names, taxonomy, and locations. (resdac.org)	A provider observation needs a separate legal-entity and ownership decision.
State professional-license sources	Jurisdiction-specific license identity and professional information. (www.mbc.ca.gov) (www.op.nysed.gov)	Verify the person or applicable license holder; avoid turning professional licensure into company ownership.
State business registry	Legal identity, document number, filing status, and named roles. (dos.myflorida.com)	Role definitions and the relevant jurisdiction determine what is established.
CMS Medicare enrollment sources	Enrollment and billing relationships at the dataset's vintage. (data.cms.gov)	Reassignment supports a billing relationship, with separate ownership verification.
AHRQ Compendium	Dated system and organization linkage resources. (www.ahrq.gov)	Its system relationship can include joint management, and its files have calendar-year scope.
Securities and Exchange Commission (SEC) filings	Reporting-company disclosures and subsidiary lists where applicable. (www.sec.gov)	Confirm the exact entity and filing vintage; absence of a matching disclosure remains unresolved.
Docket	Sourced target research through Triage, Scout, and Audit, with identity checks, answer collection, and evidence review.	Research findings remain dependent on their retained sources, dates, and unresolved questions.

These sources should be combined as distinct relationship types. Agreement on a name strengthens an identity candidate, while disagreement on a corporate role or effective date needs explicit resolution. Counting agreeing pages does not replace checking whether they refer to the same legal entity and the same relationship.

License and legal-entity identifiers need namespaces

California's physician-license profile explains that the license number becomes unique to an individual when its license type is included. (www.mbc.ca.gov) Build keys from jurisdiction, issuing authority, license type, and license number. Do not compare an unqualified number across states as though it were an NPI.

New York's Office of the Professions describes its verification database as updated daily and a primary source for license verification. (www.op.nysed.gov) California's search also supports practice-location queries, but includes self-reported profile information. (www.mbc.ca.gov) Record which field came from which mechanism

rather than treating every profile item as equally validated.

For organization identity, Florida's search guide reports document numbers and filing statuses and warns that multiple entities may appear under one **Federal Employer Identification (FEI)** number. (dos.myflorida.com) A **jurisdiction-qualified registry ID** is therefore preferable to using a shared name or an unqualified tax-number match as the automatic consolidation key.

Registry roles need equivalent care. Florida distinguishes limited liability company (LLC) managers from members and says a manager may or may not be a member. (dos.fl.gov) (dos.myflorida.com) It separately defines a registered agent by the role of accepting service of process and defines the registered office as that agent's designated address. (dos.fl.gov) Do not map a manager, registered agent, or shared registered office directly to an ownership relationship.

Medicare relationships support a different crosswalk

The **Provider Enrollment, Chain, and Ownership System (PECOS)** supplies Medicare enrollment data. CMS's public enrollment guidance distinguishes an enrollment application identifier from the provider identifier, and the main enrollment file lists one NPI per enrollment. (data.cms.gov) Model provider, enrollment, and additional-NPI links separately instead of flattening every enrollment into a new company.

CMS describes reassignment as permitting another entity to bill Medicare on an individual practitioner's behalf. (data.cms.gov) Record that as **billing reassignment**. It may help identify organizations for further research, but the accepted edge should retain that billing meaning until external evidence supports another relationship.

The Doctors and Clinicians downloadable file likewise has a clinician, enrollment, group, and address row grain. (data.cms.gov) Its dictionary distinguishes a PECOS individual identifier and flags addresses that may be incomplete. (data.cms.gov) Joining these rows to an NPI can enrich a provider map while also multiplying observations. Count unique providers and candidate organizations after joining, with a documented denominator.

Ownership remains an independently sourced claim

AHRQ's Compendium covers a single calendar year, rather than a continuously current graph. (www.ahrq.gov) Its health-system definition accepts common ownership or joint management. (www.ahrq.gov) This is valuable system-linkage evidence, but a management relationship should remain labeled as management when assessing an acquisition unit.

The SEC's explanation of Form 10-K notes that the Business section describes subsidiaries and that exhibits include subsidiary lists. (www.sec.gov) Those are possible corroborating sources where a relevant reporting-company filing exists. For private organizations without such evidence, record ownership as **not established** rather than inferring independence from a missing parent match.

A consolidation decision should also state why its grouping unit is appropriate. Census's enterprise concept requires common ownership or control, rather than a common address. (www.census.gov) AHRQ's linkage definition has a different boundary because joint management can qualify. (www.ahrq.gov) Florida's manager

definition and shared-FEI warning likewise show why apparently strong matches still need role interpretation. (dos.fl.gov) (dos.myflorida.com) The analyst should preserve these competing grouping concepts rather than importing whichever produces the smallest target count.

Before a multi-entity roll-up, require an evidence packet showing the exact entities, the proposed relationship, source vintage, supporting passage, and conflicts. Obtain counsel's interpretation when transaction structuring or the legal effect of an ownership arrangement matters. This report's matching protocol is an origination research tool, not a legal conclusion about acquisition structure.

Data Analysis and Evidence

Separate measured evidence from proposed operating metrics

The strongest available numerical evidence is bounded by its date and population. A **2015** physician-address study randomly sampled **3,000 physicians**, with **500 in each of six specialties**, and selected **1,200**, with **200 per specialty**, for telephone verification. (pmc.ncbi.nlm.nih.gov) The authors reported confirming **65%** of physicians' address information by phone; their abstract reported NPPES correct-address rates of **72% to 94%** across specialties in telephone verification. (pmc.ncbi.nlm.nih.gov)

Those measures describe a historical sampling-frame study. They are not current NPPES population accuracy, an organization-match benchmark, or an ownership-verification success rate. The telephone-confirmation measure and specialty-specific correctness measure also have different contexts; they should not be combined into one claimed dataset error rate. No independently measured performance for the proposed acquisition-target workflow was established in this research.

A separate scale illustrates the importance of source scope. AHRQ's **2020** Compendium contains **629 U.S. health systems**, under a definition requiring a hospital and a physician group providing comprehensive care. (www.ahrq.gov) That population is not the universe of healthcare-services acquisition targets. Use it for its defined linkage purpose, and keep independent practices and other organizations outside its system definition visible.

A worked reporting denominator

Table 3 is a **Hypothetical Example** of an entity-resolution worksheet summary. The values are invented solely to demonstrate arithmetic, not observations from NPPES, Docket, or an executed research test. Link categories are mutually exclusive at the relationship-decision level. The distinction between link volume and link quality follows the ONS policy. (www.ons.gov.uk)

Proposed NPI-to-legal-entity link state	Hypothetical decisions	Share of all candidate-link decisions	What can be reported
Deterministic	180	60%	Identifier-supported links under the stated rule, with namespace and relationship retained.
Reviewed	75	25%	Accepted links with a dated evidence packet and reviewer disposition.

Proposed NPI-to-legal-entity link state	Hypothetical decisions	Share of all candidate-link decisions	What can be reported
Unresolved	45	15%	Candidate links awaiting adequate evidence or conflict resolution.
Total	300	100%	Candidate-link decisions, not providers, companies, or market share.

The arithmetic is explicit: accepted decisions are **180 + 75 = 255**; accepted share is **255 / 300 = 85%**; unresolved share is **45 / 300 = 15%**. These are calculations from the hypothetical worksheet. A provider with competing candidates could contribute more than one decision, so the denominator must not be renamed as unique NPIs.

Precision is the share of made links that are true matches; recall is the share of all true matches found. (www.ons.gov.uk) Reporting an acceptance rate does not establish either quantity. A real evaluation would require adjudicated reference relationships, the sampling method, reviewer disagreements, and results by rule and source vintage.

Recommended report lines are:

- **Provider observations:** Distinct NPIs by type and source-status date.
- **Location observations:** Reported primary and non-primary addresses, with duplicates and address purpose stated.
- **Accepted entity links:** Deterministic and reviewed decisions reported separately.
- **Open relationships:** Unresolved candidates and conflicts, with research owners.
- **Evaluation results:** Precision and recall only where reference evidence supports their calculation.
- **Historical completeness:** Expected, available, incomplete, and missing source periods.

NBER cautions that some archived months and years contain incomplete or missing data. (www.nber.org) A blank historical month should therefore produce a coverage flag, not an inference that a provider disappeared.

Hypothetical Worked Examples

The following are **Hypothetical Examples**. All organizations and provider situations are fictional, and no example describes an actual target, client engagement, or measured result.

Group practice with individual providers (Hypothetical Example)

Harbor Grove Clinical Group appears as an organization record, while individual clinicians report addresses at the same building. The analyst retains separate Type 1 and Type 2 records, using the federal entity distinction. (www.federalregister.gov) The shared address creates candidate links to the practice, not automatic employment or ownership assignments.

A clinician's dated practice biography and applicable enrollment relationship may support a reviewed affiliation. The worksheet records the supported relationship and its vintage. Where only co-location exists, the candidate remains unresolved. If a billing reassignment appears, it is stored with CMS's billing meaning rather than

promoted to ownership. (data.cms.gov)

The practical deliverable is an organization record with separately evidenced provider relationships. The research team can assess the group without asserting that every clinician sharing its building belongs to it.

Multi-location organization (Hypothetical Example)

Juniper Community Therapy appears with a main address and additional practice-location rows. Those rows are linked to the same provider identifier; the source's non-primary-location file supplies the additional observations. (download.cms.gov) The analyst retains a single provider identity while preserving each reported location and its source.

Address normalization reveals different units in a shared building. The analyst keeps the unit values using USPS's secondary-unit guidance. (pe.usps.com) Company location evidence then determines which observations can be labeled current service sites. An older address remains in history instead of being silently counted as an active branch.

The resulting map can show reported footprint and verified footprint separately. It cannot infer patient volume, staffing, or site economics from the number of location observations.

Management brand and legal-entity collision (Hypothetical Example)

Cedar Span Practice Services operates a shared brand website, while differently named clinical entities appear in organization records. Name similarity generates candidates. The analyst retrieves registry identifiers and checks whether the published relationship is management, ownership, or simply brand participation.

The reason to preserve these categories is illustrated by AHRQ's system definition, which includes either common ownership or joint management. (www.ahrq.gov) The worksheet therefore keeps a management edge when that is all the evidence supports. Separately verified legal entities remain separately identifiable even when the analyst groups them visually under a brand.

The output is an explicit relationship diagram expressed in records, with open ownership questions retained. A shared website does not erase distinct entities or create an unsupported consolidated acquisition target.

Implications and Future Directions

Preserve lifecycle and history without projecting old details forward

CMS says that for deactivated NPIs it discloses only the identifier and associated deactivation date. (www.cms.gov) A current deactivation record should consequently be treated as a status observation, not a full current company profile. If historical information is retained for legitimate reconstruction, label it as historical and avoid presenting it as current contact or practice evidence.

CMS also describes reactivated providers returning to the Registry and downloadable data. (www.cms.gov) Re-evaluate the current observations and affected links after reactivation. Do not infer a new legal entity from the status transition, and do not assume every earlier relationship automatically becomes current again.

The history workflow should include:

- **Period inventory:** List required snapshots, including gaps and incomplete archives.
- **Observed transition:** Record the releases bracketing a change without inventing a precise event date.
- **Lifecycle evidence:** Preserve source-stated enumeration, deactivation, and reactivation dates where supplied.
- **Historical identity:** Keep previously observed names and addresses dated and separate from current views.
- **Decision revision:** Retain the prior match decision and the evidence supporting its replacement.

W3C PROV treats revisions as new entities related to earlier ones and supports timing information for significant events. (www.w3.org) That model is useful for keeping a match reversible when evidence changes.

Improve evidence coverage before claiming market coverage

NBER crosswalks depend on providers supplying relevant other identifiers. (www.nber.org) An absent crosswalk is therefore not affirmative evidence that a provider lacks a license, enrollment, or related organization. Similarly, an operational snapshot represents its source vintage; ResDAC describes updated files as the most current NPPES information available in that distribution. (resdac.org)

A useful next step is a controlled evaluation of the proposed link rules within a defined specialty and geography. Record ambiguous cases, false links, missed links, and evidence gaps by rule. Evaluate both person-to-practice and organization-to-legal-entity decisions separately, using the defined precision and recall concepts. (www.ons.gov.uk) Improvements should be measured against an adjudicated reference set rather than rewarded solely for merging more records.

Frequently Asked Questions (FAQs)

Can NPPES alone identify a practice owner?

It should not be treated as ownership proof. NPPES provider categories identify individuals and non-person providers; legal-entity and ownership links require their own evidence. (www.federalregister.gov) (resdac.org) An organization record can be a useful starting point without being a complete acquisition unit.

Should Type 1 providers be converted into company records?

Preserve them as individual providers, including sole proprietors, and create separate evidenced relationships to candidate organizations. CMS's enumeration guidance includes sole proprietors in Type 1. (www.cms.gov) A business-like name does not change that classification.

Does an additional practice location mean another organization?

No such conclusion follows from the file. The practice-location reference contains non-primary locations associated with existing NPIs. (download.cms.gov) Review the location separately and obtain entity evidence before treating it as another target.

How should missing months or deactivated records be handled?

Flag archive gaps and distinguish missing observations from source-stated deactivation. NBER warns about incomplete historical periods, while CMS limits current disclosure for deactivated identifiers. (www.nber.org) (www.cms.gov) Keep historical evidence dated and open relationships visible.

Conclusion

NPPES can make healthcare target research more organized when the unit of observation stays explicit. Start with the enumerated provider, preserve its reported names and locations, and attach independently supported legal-entity relationships. Ownership, employment, current service activity, and acquisition availability should remain separate research questions.

The recommended operating model is a versioned bulk baseline, controlled updates, targeted lookups, and a reviewed relationship ledger. Normalize text for comparison while preserving raw values. Treat co-location, name similarity, and shared branding as candidate-generation signals. Retain exact identifiers and the relationship they actually establish, with source vintages and conflicting evidence.

The worksheet provides the handoff standard: NPI type and status, lifecycle dates, organization and other names, taxonomy, location purpose, raw and normalized addresses, candidate legal entity, match rule, conflict, corroborating URL, and collection date. An unresolved link is a useful result when the available evidence does not justify consolidation.

Report deterministic, reviewed, and unresolved decisions with their denominators. Keep provider counts, location observations, legal entities, and target groups distinct. The historical study and system figures discussed here illustrate bounded evidence rather than a current performance promise. A future evaluation should test the proposed rules against reviewed reference relationships before claiming precision, recall, or market coverage.

The resulting asset is a traceable healthcare identity map that an origination team can revisit as records change. Its value rests on keeping accepted relationships explainable and open questions visible, so a target organization is assembled from evidence rather than implied by a collection of provider rows.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.cms.gov/priorities/key-initiatives/burden-reduction/administrative-simplification/unique-identifiers/npis
2. https://www.federalregister.gov/documents/2024/03/04/2024-04517/national-plan-and-provider-enumeration-system-nppes-data-changes
3. https://www.cms.gov/medicare/provider-enrollment-and-certification/medicareprovidersupenroll/downloads/EnrollmentSheet_WWWWH.pdf
4. https://download.cms.gov/nppes/NPI_Files.html
5. https://www.cms.gov/medicare/regulations-guidance/administrative-simplification/data-dissemination
6. https://www.nucc.org/taxonomy
7. https://familymedicine.uw.edu/chws/wp-content/uploads/sites/5/2026/01/WA_WorkforceData_UWCHWS_2025.pdf

8. https://data.cms.gov/sites/default/files/2026-07/PPEF_Data_Guidance.pdf
9. https://www.ahrq.gov/chsp/data-resources/compendium-2020.html
10. https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/datalinkageandmatchingpolicy
11. https://www.census.gov/programs-surveys/susb/about/glossary.html
12. https://www.nber.org/research/data/national-plan-and-provider-enumeration-system-nppesnpi
13. https://resdac.org/articles/overview-nppesnpi-downloadable-file
14. https://npiregistry.cms.hhs.gov/api-page
15. https://pe.usps.com/text/pub28/28c2_003.htm
16. https://npiregistry.cms.hhs.gov/help-api/json-conversion
17. https://www.w3.org/TR/prov-primer
18. https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/developingstandardtoolsfordatalinkagefebruary2021
19. https://www.mbc.ca.gov/License-Verification
20. https://www.op.nysed.gov/verification-search
21. https://dos.myflorida.com/sunbiz/search/guides/corporation-records
22. https://www.ahrq.gov/chsp/data-resources/compendium.html
23. https://www.sec.gov/answers/reada10k.htm
24. https://www.mbc.ca.gov/License-Verification/License-Verification-Sample.aspx
25. https://dos.fl.gov/sunbiz/about-us/faqs
26. https://data.cms.gov/provider-data/dataset/mj5m-pzi6
27. https://data.cms.gov/provider-data/sites/default/files/data_dictionaries/physician/DOC_Data_Dictionary.pdf
28. https://pmc.ncbi.nlm.nih.gov/articles/PMC4512968

THE PUBLISHER

About Docket

Docket provides deal-origination research software and managed research for private equity firms. We help investment teams investigate acquisition targets using structured screening criteria, retained sources and reviewable company evidence. Teams can work through a self-serve platform or use managed research, depending on how they want research delivered.

Research against a defined mandate

A useful target list needs more than company names. Docket focuses on the questions that determine whether a company fits an investment mandate, including the evidence needed to support or qualify each answer. Our research approach makes the connection between screening criteria, source material and conclusions visible to the team reviewing the work.

Triage, Scout and Audit

Docket's three named research agents perform complementary tasks. Triage resolves company identity and screens fit. Scout collects sourced answers against the mandate. Audit checks retained evidence, addresses contradictions and leaves unsupported answers visibly unresolved. This structure helps reviewers distinguish established findings from missing information and questions requiring further investigation.

Evidence that supports investment-team judgment

Our research library covers market mapping, screening criteria, private-company data, succession and ownership, source evaluation and evidence standards. These resources explain the methods and limitations behind origination research. Findings support a team's judgment; they do not establish that a company is for sale or guarantee a transaction or investment outcome.

Work with Docket

Visit Docket to explore the platform, managed research and the current contact path. Read about screening criteria, evidence standards, AI research agents and market mapping.

Public examples are illustrative unless explicitly identified otherwise. Research preparation and authorized outreach are separate activities; confidential target lists and customer outcomes should never be inferred from an educational example.

Website: https://docket.capital

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.