

# Fonteum Composite Quality Methodology

v1.2 · Snapshot 2026-05-26 · DOI 10.5072/fonteum/methodology-v1.2 (reserved)

---

This document specifies the formulas, denominators, per-source freshness targets, versioning policy, limitations, and references behind the composite quality score published at <https://fonteum.com/quality/methodology>. Every number is reproducible against the machine-readable twin at <https://fonteum.com/quality.json>.

## What the composite means

The composite quality score is a single 0-100% number summarizing how faithfully Fonteum's federal data layer reflects its primary sources. It is a weighted mean of up to four families: per-source completeness, inter-source consistency, ingestion timeliness, and the OIG LEIE byte-level match rate. Only families with a reproducible, published value are included; a family that cannot be computed against the current schema is dropped and the remaining weights are renormalized (see §3). It measures Fonteum's fidelity to the source files, not the source files' fidelity to ground truth (see Limitations).

This document is pinned to methodology version v1.2, anchored to the 2026-05-26 snapshot. Every number on the /quality scorecard corresponds to a formula below and can be replayed against the machine-readable twin at /quality.json.

## The four sub-metrics

### Completeness

For each row in the latest snapshot we count the public-displayable required fields that are non-null and non-empty, divide by the size of the required-field set, and report the median across rows. The required-field set is the public-displayability contract (the columns Fonteum renders), intentionally narrower than the upstream schema. Snapshots over 100k rows are sampled to 10k deterministically (seed = SHA256 of source\_id || snapshot\_date) so repeated reads return the same sample.

### Consistency (not currently published)

Where two independent federal feeds describe the same NPI, do they agree? The intent is a cross-source agreement rate joined on NPI. The two checks originally specified — specialty agreement (NPPES taxonomy vs a PECOS enrollment-specialty field) and active-status agreement (NPPES active flag vs a PECOS enrollment-status field) — are not currently published, because the PECOS provider table holds no enrollment-specialty and no enrollment-status column to join against. Rather than publish a figure we cannot reproduce from the schema, the scorecard reports no consistency rate and the composite renormalizes over the families that are published. A consistency rate will appear here only once it is computed against columns that exist; the page reads any such published rows from source\_consistency\_metrics.

### Match rate

The strongest single accuracy proof Fonteum can publish: SHA256 equality between Fonteum's archived copy of the OIG LEIE CSV and the SHA256 the OIG itself publishes alongside the file. The file is polled weekly on Mondays at 09:00 UTC; the metric is  $\text{matched\_weeks} / \text{total\_weeks}$  over the trailing 52 polls. A mismatch means our copy diverged byte-for-byte from what the OIG served — the only thing this metric claims to detect.

### Timeliness

Wall-clock hours between an upstream publication (source\_release\_date) and Fonteum ingesting it (ingested\_at), computed over the trailing 90 days. The per-source timeliness sub-score is  $\text{clamp}_{01}(1 - \text{median\_lag\_hours} / 168)$ : a median lag of zero hours scores 1, and any median lag of a week or longer scores 0. Snapshots without a known release date are excluded from the percentile calculation.

## Composite formula

The composite is a weighted mean of the four families, clamped to [0, 1]. The weights are pinned at this methodology version; the timeliness ceiling is one week (168 hours).

```
composite = ( sum of weight_i * score_i ) / ( sum of weight_i )
            over the families actually published this snapshot

weights:
  0.35 completeness  median(median_field_completeness across sources)
  0.3  consistency   mean(rate across published cross-source checks)
  0.2  timeliness    clamp01(1 - median(median_lag_hours) / 168)
  0.15 match_rate    matched_weeks / total_weeks

// A family with no published, reproducible value is dropped and the
// remaining weights are renormalized to sum to 1 - a missing family is
// never read as a zero. Consistency is currently UNPUBLISHED (see below),
// so the live composite renormalizes over completeness, timeliness and
// match_rate. Result clamped to [0, 1], rounded to 4 decimals for display.
```

The composite is intentionally a weighted arithmetic mean rather than a product or harmonic mean: each family is a distinct, separately-published guarantee, and a buyer can recompute the headline number from the four family scores published on the same page.

## Per-source weighting and freshness targets

The composite weights apply per metric family, not per source — within completeness and timeliness, each source contributes through the median across sources, so no single source is weighted above another. The table below documents each headline source's upstream refresh cadence (its freshness target) and which sub-metrics it currently feeds. Consistency is omitted from every row because no consistency check is published at this snapshot (see §2).

DATASET	FRESHNESS TARGET	FEEDS SUB-METRICS
NPPES (NPI registry)	Monthly	Completeness, Timeliness
OIG LEIE exclusions	Weekly	Completeness, Timeliness, Match rate
CMS PECOS PPEF	Quarterly	Completeness, Timeliness
CMS Open Payments	Annual	Completeness, Timeliness
CMS Care Compare	Quarterly	Completeness, Timeliness

## Versioning policy

The methodology is append-only: it is never silently amended. Every change to a formula, weight, or required-field set ships with a new version string, and each version is git-tagged so an old published number stays reproducible against the code revision that produced it. The current version is v1.2.

v1.2 keeps the v1 composite formula and weights unchanged; it adds the per-source sub-score decomposition surfaced on the scorecard, this per-source freshness table, and the downloadable PDF. A future version that changes any weight will publish the prior weights in this changelog.

Correction (2026-06-11): earlier revisions of this scorecard published two cross-source consistency rates (a specialty-agreement and an active-status-agreement figure) joined against PECOS enrollment-specialty and enrollment-status fields. Those columns do not exist in the provider schema, so the rates could not be reproduced and have been withdrawn. The consistency family is now shown as unpublished and the composite renormalizes over the remaining families until a consistency check is computed against columns that exist. The weight reserved for consistency is unchanged.

This methodology version is citable as DOI 10.5072/fonteum/methodology-v1.2 (reserved — DataCite test prefix; not yet minted, so it does not resolve and is not presented as a live credential). The 14-tuple provenance `_doi` field stays null until DOI minting is active.

## Limitations

This scorecard measures Fonteum's accuracy against the source files, not the source files' accuracy against ground truth. The OIG's own 2018 review found PECOS provider data inaccurate in 58% of records and NPPES in 48%; Fonteum's normalization, cross-source reconciliation, and per-field provenance contract address that gap separately (documented at /methodology).

Specifically, this page does not assert:

- That every provider in NPPES is real or currently practicing.
- That the upstream agency's required-field set matches the public-displayability set used here.
- That a cross-source disagreement means one side is wrong — taxonomy and specialty mappings legitimately drift.
- That a 100% OIG LEIE byte match would mean the exclusion data is free of false negatives at the upstream layer.

What it does assert: every computation published here runs as described, against the snapshots described, on the cadence described — and any consumer can replay it against /quality.json. A family that cannot be reproduced is shown as unpublished, never as a placeholder number.

## References

- NPPES Data Dissemination (NPI files) — [https://download.cms.gov/nppes/NPI\\_Files.html](https://download.cms.gov/nppes/NPI_Files.html)
- CMS Provider data and PECOS enrollment — <https://data.cms.gov/provider-data>
- CMS Open Payments — <https://openpaymentsdata.cms.gov>
- OIG LEIE downloadable exclusions — [https://oig.hhs.gov/exclusions/exclusions\\_list.asp](https://oig.hhs.gov/exclusions/exclusions_list.asp)
- OIG, Improvements Needed to Ensure Provider Enumeration and Medicare Enrollment Data Are Accurate (OEI-09-18-00410, 2018) — <https://oig.hhs.gov/oei/reports/oei-09-18-00410.asp>