



Third-Party Spend Aggregation

WHITE PAPER

Why Your Open Payments Data Is Only as Clean as Your Messiest Vendor

Abstract

More than 80% of pharmaceutical organizations report duplicate or outdated healthcare provider records before undertaking a data cleansing project, and the problem is rarely confined to internal systems. One documented case, a mid-size specialty manufacturer's master data cleanup, uncovered several hundred duplicate HCP records, over 2,000 duplicate or incomplete healthcare organization records, and thousands of missing phone numbers and addresses, gaps that had accumulated because payment and interaction data arrived from speaker bureaus, meeting planners, CROs, and market research vendors, each maintaining its own version of who a given physician actually was. Aggregate spend reporting depends entirely on correctly matching every payment to the right healthcare provider identity, and that matching problem gets harder, not easier, as more third parties touch the data before it reaches the manufacturer's own reporting system. This white paper applies established data reconciliation and identity-management best practices to aggregating clean, reconcilable spend data across a fragmented vendor ecosystem.

The Reporting Obligation That Starts Outside Your Own Systems

A pharmaceutical manufacturer's Open Payments and state transparency reporting obligation feels, on paper, like an internal data problem: gather the payments, match them to recipients, and submit the report. In practice, most of the underlying data never originates inside the manufacturer's own systems at all. Speaker program payments flow through a speaker bureau vendor. Meeting and event spend flows through a meeting planning agency. Market research honoraria flow through a research vendor. Each of these third parties maintains its own record of who the healthcare provider actually is, and none of them were built with the specific goal of matching cleanly against the manufacturer's own master data or a regulator's identifier system.

The result is a spend aggregation problem that starts well upstream of the manufacturer's own compliance team. By the time payment data reaches internal systems for reporting, it may already carry inconsistent name formats, outdated affiliations, duplicate identities under slightly different spellings, and missing identifiers that make clean matching to a specific, correctly identified provider genuinely difficult.

The Scale of the Underlying Data Quality Problem

The scale of this problem is not anecdotal. Over 80% of pharmaceutical organizations report duplicate or outdated healthcare provider records before undertaking a formal data cleansing initiative, a figure consistent enough across surveyed organizations to describe an industry-wide default state rather than an isolated issue.¹ One documented master data cleanup at a specialty pharmaceutical manufacturer found several hundred duplicate HCP records, thousands of missing phone numbers and addresses, and more than 2,000 duplicate or incomplete healthcare organization records, all accumulated from data flowing in through multiple third-party vendor relationships over time.² After a structured master data management remediation, that same organization recovered more than 14,000 HCP and HCO profiles it had effectively been missing, along with tens of thousands of missing addresses and email addresses.³

The consequence for spend aggregation is direct. A payment that cannot be cleanly matched to a single, correctly identified healthcare provider either gets reported against the wrong identity, gets reported as a duplicate alongside a payment already correctly attributed, or gets flagged for manual reconciliation that consumes real compliance staff time during an already compressed reporting window. None of these outcomes is acceptable in a regulatory environment where CMS's own compliance outreach is actively correcting and updating thousands of records annually and increasingly applying risk-based scrutiny to reporting patterns that look inconsistent.

Why This Gets Harder as the Vendor Ecosystem Grows

A manufacturer running a modest speaker program through a single vendor faces a comparatively manageable data quality problem. A manufacturer running multiple therapeutic-area speaker programs, several market research initiatives, a meeting and events function, and a medical affairs engagement program, each potentially through different vendors, faces a genuinely compounding one. Every additional vendor relationship is another party maintaining its own version of provider identity, another data format the manufacturer's systems need to reconcile, and another opportunity for the same physician to enter the aggregate spend pipeline under two or three subtly different identities.



This is not a problem that resolves itself as an organization matures. It typically gets worse, because a growing commercial and medical affairs footprint tends to add vendor relationships faster than it adds the master data governance needed to keep pace with them. An organization that solved its spend aggregation data quality problem three years ago, when it ran two vendor relationships, may be significantly behind that same problem today, running eight.

Applying Best Practices: A Vendor-Agnostic Spend Aggregation Practice

A foundational data governance best practice, one data model that absorbs new inputs rather than requiring a rebuild for each new source, applies directly to the third-party spend aggregation problem. Applying it as a practice built specifically for a multi-vendor ecosystem means four capabilities.

1. **A Single Master HCP and HCO Identity Layer.** Build one authoritative provider identity reference that every vendor's data gets matched against on ingestion, rather than allowing each vendor's version of provider identity to flow independently into the reporting pipeline.
2. **Standardized Data Intake Requirements for Every Vendor.** Define a common data format and required identifier set that every third-party vendor is contractually required to provide, reducing the reconciliation burden created by each vendor submitting data in its own preferred structure.
3. **Automated Duplicate and Mismatch Detection Before Reporting Season.** Run identity matching and duplicate detection continuously against incoming vendor data throughout the year, rather than discovering mismatches during the compressed weeks before a submission deadline.
4. **Vendor-Level Data Quality Scorecards.** Track which vendors consistently deliver clean, matchable data and which consistently require manual reconciliation, giving compliance and procurement teams a concrete basis for vendor management conversations.

None of this requires replacing existing vendor relationships. It requires building the identity reconciliation layer already applied to jurisdictional reporting complexity, extended specifically to absorb the variation multiple third-party data sources introduce.

What a Vendor-Agnostic Practice Buys, and Where It Gets Hard

Advantages

- **Cleaner aggregate spend submissions:** a single authoritative identity layer directly reduces the duplicate and misattributed records that create both compliance risk and manual rework.
- **Less compressed reporting-season workload:** continuous duplicate detection throughout the year moves reconciliation work out of the high-pressure weeks before a submission deadline.
- **Leverage in vendor negotiations:** data quality scorecards give compliance and procurement teams concrete evidence for holding underperforming vendors accountable at contract renewal.
- **Reduced audit exposure:** cleaner, more consistent data directly addresses the kind of inconsistent reporting pattern that risk-based CMS audit selection is specifically designed to flag.



Risks and Barriers

- **Vendor cooperation is not guaranteed:** smaller or less sophisticated vendors may lack the technical capacity to meet a standardized data intake requirement without real support or a phased implementation.
 - **Legacy data cleanup is a genuine project:** years of accumulated duplicate and inconsistent records do not resolve through better intake standards alone. They require a dedicated remediation effort for existing data.
 - **Cross-functional ownership:** provider identity data touches commercial operations, medical affairs, and compliance, and building one authoritative identity layer requires governance agreement across functions that do not always share priorities.
 - **Ongoing maintenance discipline:** a clean master identity layer degrades over time without sustained governance, and organizations sometimes underinvest in maintenance once the initial cleanup project concludes.
-

Case in Point: The Physician Who Was Three Different People

A mid-size pharmaceutical manufacturer running speaker programs through two different vendors and a separate meeting planning agency discovered, during a pre-submission reconciliation review, that one of its most active speaking physicians existed as three separate records across its aggregate spend data: one under a formal legal name from the speaker bureau, one under a commonly used shortened name from the meeting planner, and one with a transposed middle initial from the market research vendor. Each record carried a portion of the physician's total annual payments.

Reported as three separate identities, the physician's payments would have appeared individually unremarkable but collectively understated his true total compensation from the manufacturer, exactly the kind of pattern that invites regulatory scrutiny if discovered externally rather than caught internally. The manufacturer's compliance team caught the discrepancy only through a manual, labor-intensive review shortly before the filing deadline, correcting the records under real time pressure. A standing master identity layer, matching all three vendor records against one authoritative provider reference throughout the year, would have caught the mismatch months earlier with far less last-minute effort.

Recommendations for Compliance and Procurement Leaders

1. **Build a single master HCP and HCO identity reference.** and require every third-party vendor's data to match against it, rather than accepting each vendor's independent version of provider identity.
2. **Standardize data intake requirements in every vendor contract.** specifying the format and identifiers required, and build this requirement into contract renewals for existing vendors.
3. **Run duplicate and mismatch detection continuously.** rather than only during the compressed weeks before an annual submission deadline.
4. **Score vendors on data quality and use it in renewal decisions.** giving underperforming vendors a concrete basis for improvement and a real consequence for not improving.
5. **Fund legacy data cleanup as its own project.** recognizing that years of accumulated duplicate records will not resolve through better ongoing intake alone.



6. **Assign clear cross-functional ownership of the master identity layer.** spanning compliance, commercial operations, and medical affairs, so maintenance discipline survives beyond the initial cleanup effort.
-

Measuring Success: KPIs for Third-Party Spend Aggregation

- **Duplicate record rate:** the percentage of HCP and HCO records identified as duplicates during reconciliation, tracked toward a declining trend.
 - **Vendor data quality score:** a per-vendor measure of how much manual reconciliation their submitted data requires, used in ongoing vendor management.
 - **Pre-submission reconciliation lead time:** how far in advance of the filing deadline duplicate and mismatch issues are identified and resolved, tracked toward earlier detection.
 - **Manual reconciliation hours per submission cycle:** the compliance team labor required to resolve data quality issues before filing, tracked toward reduction as intake standardization matures.
-

Conclusion

Aggregate spend reporting accuracy depends on data quality decisions made largely outside the manufacturer's own walls, in speaker bureaus, meeting planning agencies, and market research vendors that were never built to match cleanly against each other or against a manufacturer's own master data. With more than 80% of organizations reporting duplicate or outdated provider records before any formal cleanup effort, this is not a problem confined to poorly run compliance functions. It is the industry's default state.

Applying disciplined best practices to a genuine third-party spend aggregation practice means building one authoritative identity layer that every vendor's data reconciles against, standardizing intake requirements, and running detection continuously rather than under deadline pressure. Manufacturers that build this discipline are not just reducing manual reconciliation work. They are closing the specific gap, fragmented, unreconciled provider identity across a growing vendor ecosystem, that turns a routine submission into an audit finding.

About OmniTek & Contact Information

For compliance and procurement leaders managing aggregate spend data across a growing vendor ecosystem, OmniTek applies proven best practices to build a practical spend aggregation practice: a master provider identity layer, standardized vendor intake requirements, continuous duplicate detection, and vendor-level data quality accountability. Whether you are building this discipline ahead of your next submission cycle or cleaning up years of accumulated vendor data fragmentation, OmniTek can help you build a reporting process that is only as strong as its cleanest vendor, not its messiest.

Contact us today to explore how OmniTek can support your transformation journey and drive lasting results.



McLean, VA | Serving
clients nationwide



info@omnitekconsulting.com



omnitekconsulting.com

[Click here to book a 15-minute call to see how OmniTek can help your team put these insights into practice.](#)



Endnotes

¹ IntuitionLabs, "HCP Master Data Management: Validation & Pharma Compliance," citing industry survey data and a documented pharmaceutical manufacturer master data cleanup case study, available at <https://intuitionlabs.ai/articles/hcp-master-data-management-pharma-compliance>

