CUT THE INTEGRATION TAX

Data Integration Is the Hidden Drag on B2B Rollup Returns

Executive Summary

Private equity enters 2026 fueled by over one trillion dollars in dry powder (Lukatsky, Marina and Choi, Jinny and Walters, Kyle, 2025). The insurance brokerage, managed IT, wealth management, field services, and accounting sectors absorbing this capital share a common structural challenge: each acquisition adds new systems, new data formats, and new integration burden, and every month without unified visibility is a month of synergy runway lost.

Both traditional engineering and AI-generated scripting produce maintenance costs that grow quadratically with portfolio size, eventually consuming the integration team's entire capacity. Large IT projects run an average of 45% over budget while delivering 56% less value than predicted (McKinsey & Company, 2012), and nearly 70% of mergers fail to achieve expected revenue synergies (McKinsey & Company, 2018).

A concept-based architecture changes the cost curve by mapping every source to a shared semantic layer rather than building pairwise integrations. In benchmark testing across 60 structural format variants, this approach achieves 99.0% mapping accuracy and compresses typical integration timelines from months to weeks. **Maintenance remains roughly constant regardless of portfolio size.** An illustrative IRR model shows that compressing integration from twelve months to one month improves equity IRR by approximately 70 to 90 basis points – an effect that grows with add-on count and synergy mix.

Intended Audience (Technical Edition)

Target Audience: CDOs, CTOs, and Data Architects in PE-backed B2B rollups
Technical Focus: Scalable integration architecture that prevents maintenance growth from compounding with each add-on acquisition.
Key Framework: Concept-based mapping from $n \times m$ pairwise system links to an $n+m$ semantic layer across ERP, CRM, billing, and support platforms.
Benchmarks: 99.0% mapping accuracy across 60 structural variants with automation for both onboarding and change management.
Maintenance Target: Flat-load upkeep under 1 hour/week for 200+ integrations to preserve operating leverage.

The Integration Tax

A PE-backed insurance brokerage platform closes on its eighth acquisition in two years. Each agency arrived with its own management system: Applied Epic at one site, Vertafore at three others, HawkSoft at two, and two more running legacy systems that predate modern APIs. The CFO needs a consolidated book-of-business report for the board. It does not exist. Commission data is stored in five formats. Customer records use different identifier schemes. Policy structures differ by carrier, by agency, and by AMS vendor. The integration team, now on its third consulting engagement, is still reconciling the fifth acquisition. Meanwhile, the ninth deal is in letter of intent, and the buyer's diligence team will ask for integrated KPIs the platform cannot deliver.

This is the default experience of B2B consolidation today. The gap between the pace of acquisition and the pace of data integration is a measurable drag on rollup returns: an integration tax that compounds with every deal.

The Consolidation Math

US private equity enters 2026 under unusually high pressure to demonstrate realized operational improvements. The inventory of PE-backed companies stands at nearly 12,900, with 30% held seven or more years and another 37% held four to six years (Lukatsky, Marina and Choi, Jinny and Walters, Kyle, 2025). In a recent PitchBook survey, half of respondents identified exiting portfolio companies as their primary focus over the next six months (Lukatsky, Marina and Choi, Jinny and Walters, Kyle, 2025). At the same time, platform buyouts are expected to re-accelerate: LBO activity was 21.6% of US PE deal activity through October 2025, and PitchBook's 2026 outlook calls for platform LBO share to increase to 25% or higher, enabled by over one trillion dollars of dry powder (Lukatsky, Marina and Choi, Jinny and Walters, Kyle, 2025).

The sectors absorbing this capital are fragmented B2B services markets where the rollup thesis is structurally dominant:

- **Insurance brokerage:** 849 announced M&A transactions in 2024, with private-capital-backed buyers comprising the majority (PLANADVISER, 2024).
- **Managed IT services:** 71 PE transactions and 104 strategic rollups in Q1 2024 alone (Founders Investment Banking, 2024).
- **Wealth management:** A record 322 RIA M&A transactions (InvestmentNews, 2025).
- **Field services:** 55 PE-backed HVAC deals in 2024 (a 72% year-over-year increase), plus 466 specialty construction deals across tracked subcategories (PitchBook Data, Inc., 2025).
- **Accounting:** Over $500 million in planned technology-enabled rollup investments by a single platform (Reuters, 2025).

The economic logic of every one of these rollups depends on integration speed. Top-performing PE firms target operational unification within the first 100 days of close, because every month of delayed integration is a month of unrealized synergies. This results in fragmented pricing, invisible cross-sell opportunities, duplicative vendor contracts, and a CFO who cannot produce reconciled KPIs for the board. The operating partner who assembled an eight-agency insurance platform is not managing eight agencies. They are managing the gap between eight agencies and one unified operation.

The Visibility Gap

The "visibility gap" is the period between close and the first date when the sponsor and management team can generate reconciled KPIs across the combined entity: revenue by segment, margin by service line, customer retention, AR aging, utilization, and normalized EBITDA.

The duration of this visibility gap is not driven by deal complexity or management competence. It is driven by system heterogeneity. Each acquired company arrives with its own operational stack:

- Insurance agencies average 5.7 to 11.9 technology platforms depending on size, and while 93% use an agency management system, those systems span multiple vendors with materially different data models (Independent Insurance Agents & Brokers of America, 2024).
- MSPs run layered tool stacks (PSA, RMM, ticketing, security, billing) customized per company and subject to vendor-driven platform changes (Founders Investment Banking, 2024).
- RIAs are increasingly multi-custodial, multiplying the data normalization burden for portfolio reporting, trading, and compliance (WealthMan-

agement.com, 2025).
- Field service companies mix dispatch, work order, inventory, payroll, and local accounting systems, often with no two acquisitions running the same stack.

During the visibility gap, specific synergy levers are disabled:

- **Pricing discipline** requires knowing margin by service line across the portfolio. Without it, unprofitable contracts renew unchallenged.
- **Cross-sell targeting** requires a unified customer view. Without it, the sales team cannot identify which clients of Agency A also need products from Agency B.
- **Vendor rationalization** requires consolidated spend data. Without it, three agencies may pay three different rates for the same carrier or software license.
- **Labor optimization** requires utilization data across sites. Without it, overstaffing at one location coexists with understaffing at another.
- **CFO-grade reporting** requires reconciled financials. Without it, board presentations are assembled by hand from incompatible exports.

The human cost is tangible. Finance and operations teams at portfolio companies spend roughly half their time gathering and reconciling data rather than analyzing it (Ganti, Aditya, 2025). Practitioners describe repeated situations where founder-led platforms with $50 million or more in revenue still operate on basic accounting tools and handwritten AP logs, requiring chart-of-accounts standardization and data normalization before synergies can even be modeled (Cohn-Reznick, 2024).

This is the integration tax: not a one-time project cost, but a recurring drag on the effective hold period available for synergy capture.

Why Current Approaches Fail at Scale

Schema Drift Is a Certainty

In B2B software ecosystems, the systems that acquired companies run are not static. Vendors version their APIs, deprecate older versions, and regularly change schemas across releases:

- **QuickBooks Online:** Intuit formally discontinued support for minor API versions 1 through 74, requiring updates for all applications using older versions (Intuit, 2025).
- **Salesforce:** Maintains a defined API retirement policy, with planned retirements of legacy versions on a regular schedule (Salesforce, 2025).
- **NetSuite:** Release notes are explicitly subject to weekly changes, and Oracle provides schema diff tools between endpoint versions (Oracle, 2025).
- **HubSpot:** Defines "breaking changes" and aims to provide 90 days notice before making them (HubSpot, 2025).

For a growing rollup, schema drift is not an edge case. It is a certainty. Each add-on multiplies the drift surface: every upstream vendor change potentially affects every integration that connects to that vendor. A platform with eight acquisitions running four different AMS vendors is exposed to four independent drift schedules. The exposure grows with each deal.

Traditional Engineering

The default approach to data integration remains labor-intensive. Skilled engineers analyze source schemas, write mapping logic, build transformation pipelines, and test against production data. At fifty source types, the initial project cost runs $1.5 to $2.5 million with a team of fifteen to twenty-five people over eight to ten weeks.

The deeper problem is maintenance. Each source drifts independently. At fifty active integrations, a platform can expect two to five variation events per week, each requiring an engineer to investigate, diagnose, fix, test, and deploy. At two hundred integrations, the maintenance load demands five to ten dedicated engineers, and costs compound quadratically. This is not because each fix is harder, but because the number of things that can drift grows with the square of the number of sources.

The track record of large-scale IT projects reinforces the structural concern. Research on over 5,400 IT projects conducted by McKinsey with the University of Oxford found that large IT projects run an average of 45% over budget while delivering 56% less value than predicted. Seventeen percent become "black swans" with cost overruns exceeding 200% (McKinsey & Company, 2012). Each additional year of planned project duration increases overruns by 15%.

AI-Generated Scripts

Using large language models to generate integration scripts appears to solve the cost problem. Initial per-integration cost drops to $135 to $540. But each integration produces a bespoke script with no shared architecture. At five hundred sources, the “codebase” is five hundred unrelated scripts. Any vendor API update, new compliance requirement, schema modification or other cross cutting change, requires touching all of them.

The maintenance cost at scale actually exceeds traditional engineering, because the generated code is unfamiliar, inconsistent, and difficult to modify reliably. Plausible errors, such as the wrong field mapped or type coercion applied incorrectly, pass along silently and may not surface until downstream analysis or audit fails.

The Core Issue

The cost of any single integration is not the problem. The **shape of the maintenance curve** is the problem. Both traditional and AI-scripted approaches produce maintenance costs that grow quadratically with portfolio size. At small scale, this is manageable. Most PE-backed platforms target a scale where each quarter brings new acquisitions and the portfolio may reach fifty to two hundred active data sources. The integration team becomes entirely consumed by maintaining what already exists. New acquisitions stall. The rollup thesis breaks.

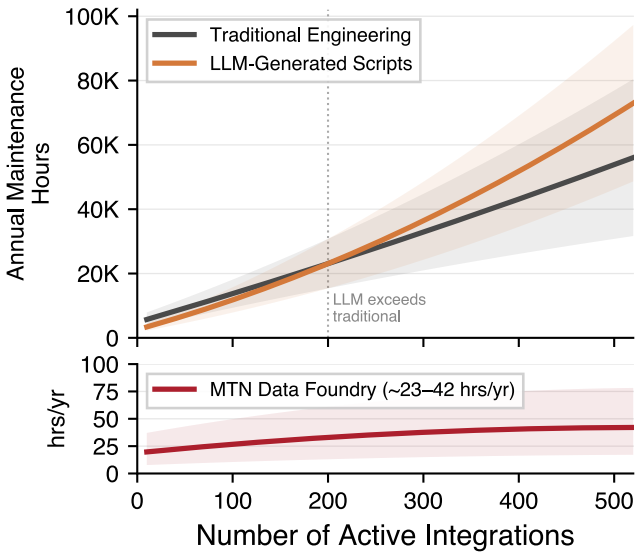

Figure 1. Annual maintenance hours by number of active integrations. Traditional and LLM hours are derived from labor costs at a $65/hr fully-loaded rate; Data Foundry hours are derived from documented variation-event rates (events/week $\times$ review minutes). LLM-generated scripts start cheaper but exceed traditional engineering beyond $\sim$200 integrations because each bespoke script must be touched individually for any cross-cutting change. Shaded regions indicate ranges.

The initial cost of integration is a red herring. What determines whether the rollup thesis holds is the maintenance trajectory. A platform that spends $2 million on integration today but faces millions per year in maintenance at two hundred sources has not solved the problem. It has deferred it.

Nearly 70% of observed mergers fail to achieve expected revenue synergies (McKinsey & Company, 2018), and firms that invest in modern data architecture outperform peers by 15 to 20 percent in portfolio company value creation (Ganti, Aditya, 2025). The gap between these two outcomes is largely an integration gap.

> **The question for any platform operator is straightforward: at your current rate of acquisition, when does the maintenance curve cross the line where integration overhead exceeds the operational**

capacity of your team? For most growing rollups, the answer is sooner than expected.

A Compounding Integration Capability

What a Solution Must Do

1. The solution must adapt to each source system's unique data model without requiring months of manual field-by-field mapping. B2B operational data is too heterogeneous, and too frequently customized per installation, for any approach that depends on predefined schemas or static templates.
2. It must learn and reuse concept knowledge across acquisitions. Mapping "customer," "contract," and "invoice" for one agency management system should reduce the effort required to map the next one, not reset to zero.
3. It must deliver usable integrated data in weeks, not quarters. The economics of a rollup do not accommodate eighteen-month integration timelines. Every month of delay carries a concrete cost in unrealized synergies and an extended visibility gap.
4. The maintenance cost must remain roughly constant as the number of active integrations grows. Any solution whose maintenance burden scales with the square of portfolio size will break.
5. The solution must survive vendor schema drift without reimplementation. In an ecosystem where QuickBooks, Salesforce, NetSuite, and HubSpot change their APIs on predictable schedules, the integration layer must detect and adapt to changes rather than break silently.

Introducing MTN Data Foundry

The MTN Data Foundry is built on a concept-based architecture designed to meet these requirements.

The core principle is straightforward. Rather than producing a combinatorial explosion by building pairwise mappings between every source system and every target system, Foundry maps every source to a shared concept layer. Each concept represents a standardized business object: "Customer.AccountID," "Contract.EffectiveDate," "Invoice.LineAmount," "Employee.Role," and so on. When a new source system arrives, its fields are mapped to this existing concept vocabulary, not to every other system in the portfolio. Adding a new integration does not require touching any existing integration.

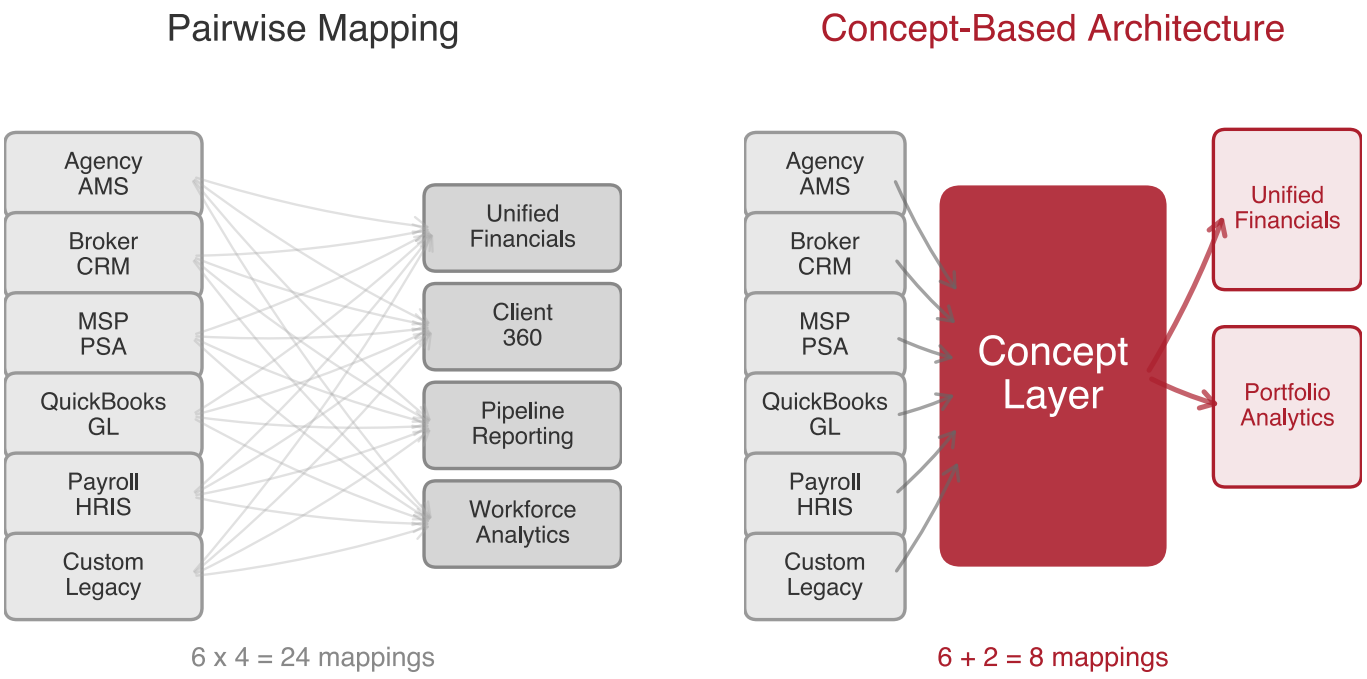

Figure 2. Pairwise mapping (left) produces $n \times m$ connections that grow combinatorially. Concept-based architecture (right) reduces this to $n + m$ mappings through a shared semantic layer.

The compounding advantage emerges over successive acquisitions. Once "customer," "contract," "invoice," and "employee" are mapped and validated for one AMS, the next acquired agency reuses the concept layer and maps only the fields or structures unique to its particular system configuration. The eighth acquisition is faster and cheaper than the first, not harder.

The platform operates through four integrated capabilities:

Schema Discovery. When a new data source is loaded, Data Foundry presents its structure visually: field names, data types, sample values, and structural patterns. This replaces the traditional process of digging through vendor documentation or compiling field inventories in spreadsheets.

AI Concept Mapping. Using schema metadata and sample data, the system automatically proposes mappings between source fields and target concepts. It identifies obvious matches, detects format differences, and flags ambiguous cases for human review. In benchmark testing across 60 structural format variants, the system achieves 99.0% mapping accuracy at an average cost of $3.30 per mapping. An analyst (not a senior engineer) can review and approve a typical integration in five to fifteen minutes.

Deterministic Validation. Every mapping is verified through structural coverage checking before activation. The system refuses to process data it cannot validate – unmapped fields trigger rejection, not silent passthrough. Data is either correct and validated, or explicitly pending human review.

Drift Detection and Governance. When a source system changes with a vendor update, a new field, an unfamiliar code or other drift, the platform detects this deviation automatically and queues it for review. At two hundred active sources generating two to five variation events per week, the total review burden is less than one hour per week. This is what keeps the maintenance curve flat.

Canonical Concepts Across B2B Verticals

The transferability of this approach rests on a structural observation: most B2B operating systems, regardless of vendor or vertical, represent a common set of business objects. For example, the same ten concepts appear in different forms across insurance brokerages, MSPs, wealth management firms, field service companies, and accounting practices:

Table 1The same ten canonical concepts appear across B2B verticals, expressed in domain-specific terminology.

Concept	Insurance	MSP	RIA	Field Services	Accounting
Party	Insured, agent, carrier	Client contact, vendor	Client, beneficiary	Customer, tenant	Client, partner
Account	Agency account	Client organization	Household, account	Service location	Engagement client
Contract	Policy, binder	Service agreement, SLA	Advisory agreement	Service contract, warranty	Engagement letter
Product/Service	Coverage line, endorsement	Managed service tier	Investment product, model	Service type, trade	Service offering
Order/Ticket	Submission, quote	Service ticket, project	Trade order, rebalance	Work order, dispatch	Task, deliverable
Invoice/Payment	Commission statement	Monthly billing, T&M invoice	Fee schedule, custodian stmt	Job invoice, parts billing	Time & billing entry
Employee	Producer, CSR	Engineer, dispatcher	Advisor, planner	Technician, foreman	Staff, preparer
Location/Asset	Agency branch	Client site, endpoint	Office, custodian	Property, equipment	Office, system
Document	ACORD form, certificate	SOW, compliance report	IPS, ADV filing	Permit, inspection	Tax return, workpaper
Event Timeline	Renewal, claim, endorsement	Incident, resolution, SLA breach	Review, rebalance, billing	Dispatch, completion, callback	Filing deadline, review cycle

That is what makes integration compoundable rather than additive. Once the concept layer is learned, each new system maps to it rather than to every other system.

Benchmarked Performance

We evaluated the MTN Data Foundry across 11 structurally diverse data sources and 60 format variants. The evaluation uses structurally equivalent data sources that reproduce the same format diversity, schema complexity, and variation patterns found in B2B rollups.

Benchmarked Source Type	B2B Rollup Equivalent
Mobile devices	Workforce telemetry – GPS trackers, technician apps, field rep geolocation
Industrial monitors	RMM agents (MSPs), fleet telematics, building-management sensors (HVAC)
IoT gateways	Branch network appliances, building-automation controllers, warehouse hubs
Standardized formats	EDI X12, ACORD XML/AL3 (insurance), OFX/BAI2 (banking), NACHA (payroll)
Multi-domain platforms	ERP/CRM/PSA suites – QuickBooks, NetSuite, Salesforce, ConnectWise, Applied Epic

Structural Variant	B2B Data That Looks Like This
Flat (CSV/delimited)	Customer rosters, rate cards, commission schedules, AR/AP aging exports
Nested (hierarchical)	CRM opportunities, policy structures, portfolio account hierarchies, work-order trees
Entity-attribute-value	Custom CRM fields per acquired company, configurable ticket attributes
Columnar batch	Commission runs, financial close packages, payroll batches, custodian files
Pipe-delimited	ACORD AL3, legacy EDI, mainframe GL extracts, bank lockbox files
Wide/sparse-column	Multi-LOB submissions, MSP reports with varying column sets per service type

Every acquisition adds several of these structural variants simultaneously.

Performance Results

Metric	Result
Mapping Accuracy	99.0%
Format Consistency	100.0%
Record Capture Rate	96.2%
Average Mapping Time	442 seconds
Average Cost per Mapping	$3.30
Total Field Mappings Evaluated	480

Mapping accuracy measures the percentage of data fields correctly mapped to the target concept. Format consistency measures whether the same field in the

same source is mapped identically across all structural variants. Record capture rate measures the percentage of expected normalized records successfully produced after transformation.

The Economics

Initial Project Cost

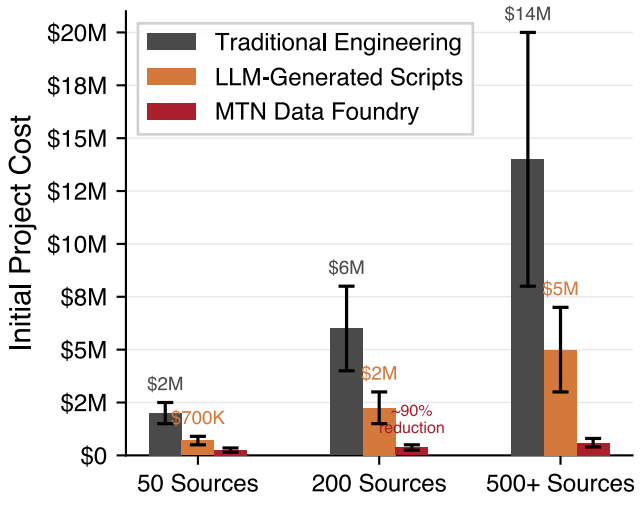

Figure 3. Initial project cost by portfolio scale and integration approach. Data Foundry costs remain sub-linear due to concept reuse across sources.

IRR Sensitivity to Integration Speed

The financial impact of integration speed is directly measurable in PE return terms. The illustrative model below assumes a standard rollup structure: a platform plus ten add-ons over a five-year hold, with add-ons acquired at 6x EBITDA and the combined platform exiting at 12x. Integration-gated synergies include cost savings of 10% of add-on EBITDA (ramping over six months post-integration) and incremental growth of 3% annually once integrated.

Months to Unified Visibility	Equity IRR
1 month	~23.1%
6 months	~22.8%
12 months	~22.4%
18 months	~22.0%
24 months	~21.6%

Moving from twelve months to one month in this scenario improves equity IRR by roughly 70 to 90 basis points. The effect grows as add-on count increases, as acquisition cadence accelerates, and as the synergy mix becomes more revenue-dependent (because revenue synergies compound over time).

The effect also grows when exit multiples are sensitive to integration quality. Practitioner commentary frames this bluntly: “real integration in buy-and-builds” is what avoids “three logos under a holdco,” where incomplete integration can trade at a discount while genuinely integrated platforms earn multiple expansion (CohnReznick, 2024). Even a 0.5x to 1.0x multiple differential on exit EBITDA can dominate the IRR delta.

Conclusion

The Compounding Advantage

The pattern is clear. Private equity capital is flowing into fragmented B2B services at an accelerating pace, with platform LBO share expected to rise above 25% in 2026 and over one trillion dollars in dry powder seeking deployment (Lukatsky, Marina and Choi, Jinny and Walters, Kyle, 2025). Each add-on acquisition brings new systems, new data structures, and new integration burden. The question facing every platform operator is not whether to integrate, but how to integrate at a pace that matches the pace of acquisition.

The core risk is not technical complexity. It is time. When commission data lives in five formats, when customer records use different identifier schemes, when the CFO’s board presentation depends on spreadsheets assembled by hand, the hold period is running but synergy capture is not. That gap is the integration tax, and it compounds with every deal.

Current integration approaches share a structural limitation: their maintenance costs grow quadratically with portfolio size. At small scale, this is manageable. At the scale most PE-backed platforms are targeting, it becomes the binding constraint on growth.

A concept-based architecture offers a fundamentally different trajectory. By mapping every source to a shared concept layer rather than building pairwise integrations, the marginal cost of adding new sources remains roughly constant. Each acquisition makes the next one faster, not harder. The eighth agency joining an insurance platform reuses the concept mappings established by the first seven and maps only the delta. This is not a marginal improvement in integration eco-

nomics. It is a structural change in what is operationally feasible.

Next Steps

We offer an integration complexity assessment: a focused analysis that maps your current data landscape, identifies the specific visibility gaps across your portfolio, and estimates the cost of delayed integration at your current acquisition pace.

Whether you are evaluating a platform acquisition, midway through a multi-site integration, or planning for the next phase of add-on growth, the core question is the same: how many active data sources will you have in eighteen months, and does your current integration approach scale to that number?

We welcome a conversation about how these principles apply to your portfolio.

MTN is a research and technology company with deep roots in computational neuroscience and machine intelligence. Our work has been published in Nature journals, PNAS, JMIR, Chest, PLoS Computational Biology, The Royal Society, and other leading venues. We bring to these conversations the perspective of researchers and advisors with technical and operational backgrounds.

Technical Leadership

Warren Pettine, MD — Co-Founder and CEO. Assistant Professor at the University of Utah where he leads the Medical Machine Intelligence (M^2Int) Lab. Trained in machine learning research at Harvard, Stanford, NYU, and Yale. Leads the research and product teams.

Matthias Christenson, PhD — AI Architect. Investigator with the M^2Int Lab. PhD and post-doctoral research at Columbia University in computational ML, with prior industry experience as a Deep Learning Research Engineer at DeepLife training foundational models on genomic and biometric data. Leads MTN's technical architecture design and data model development.

Samuel Wecker, Lead Systems Engineer. Over twelve years building and scaling production software, including as a founding engineer at a startup that grew to a billion-dollar platform. Specializes in unifying disparate systems and data sources at scale. Leads Data Foundry's core platform development.

Contact: Warren Pettine, MD, CEO of MTNwarren@themtn.aihttps://www.themtn.ai