

The State of Data Platform Consolidation: Findings from the 2025 Meridian Enterprise Data Survey

Enterprises spent the last decade assembling data stacks out of best-of-breed components, and they are now spending this one paying the integration bill. The 2025 Meridian Enterprise Data Survey collected evidence on what that bill actually contains, and its central finding is blunt: the tooling footprint of large organizations has drifted far beyond what their own architects consider defensible, and the drift is funded by budget lines nobody reads.

The scale of the drift is easiest to state as a single pair of numbers. The median respondent operates eleven distinct data tools in production, while the number their architecture teams consider justified is five. The gap between those two numbers is where budgets, reliability, and morale quietly erode, and every section of this paper returns to it from a different angle.

The survey behind these findings was fielded between February and April of 2025, targeting organizations with more than one thousand employees across Europe and North America. Respondents held titles of head of data or above in sixty-eight percent of cases, and the sample was screened so that no single industry exceeded eighteen percent of responses; financial services contributed seventeen percent, manufacturing fifteen, healthcare twelve.

Median company size was 4,200 employees, and annual platform spend ranged from 300 thousand to 41 million euros with a median of 2.7 million. Differences reported between segments are significant at the ninety-five percent level under a two-proportion test, and segments under thirty responses are marked directional in the appendix. In total, the instrument gathered structured responses from 412 organizations, every one of them screened by a qualifying interview.

Cost of ownership is the theme respondents raise first when asked open-endedly what is wrong with their platform, and the anatomy of that cost is not what invoices suggest. Licence fees are the visible layer, but the survey asked respondents to decompose their fully loaded spend across licensing, integration engineering, incident response, and enablement, and the visible layer turns out to be the minority.

Direct licensing accounts for only forty-four percent of the reported total, with integration engineering at twenty-nine, incident and reliability work at fourteen, and training at thirteen. The integration share grows with tool count: above fourteen tools it reaches thirty-eight percent, against nineteen below eight tools.

One logistics group in the sample turned this decomposition into policy. Its finance function computed that every additional tool costs a recurring 190 thousand euros per year in glue work before any licence fee, and now applies that figure as an internal tax during procurement, a practice the recommendations section returns to.

Duplication of function is the second face of sprawl, and it is mostly unplanned. Sixty-one percent of organizations run at least two enterprise-grade schedulers, and forty-three percent run three or more overlapping catalog products. Asked to name the origin, respondents pointed overwhelmingly at organizational seams rather than technical disagreement.

Acquisitions dominate that story: mergers brought incompatible stacks in fifty-two percent of cases, autonomous business units purchased overlapping tooling in thirty-nine, and only nine percent describe their duplication as a deliberate multi-vendor strategy. Retirement stalls for organizational reasons too, with migration risk ranked first and the absence of a

project owner second, ahead of contractual lock-in.

Migration behavior contradicts the industry's public enthusiasm for replatforming. Seventy-three percent of organizations call consolidation a priority for the next eighteen months, yet only twenty-two percent completed a major migration in the past twenty-four, and the completed ones carried a characteristic signature of pain.

The median completed migration overran its planned duration by sixty percent. The successful minority shared three practices: they froze feature development on the legacy system at the start, they migrated by data domain rather than by technology layer, and they published a shutdown date for the legacy system that survived executive turnover.

Parallel operation without a shutdown date turns out to be the trap. Where both stacks ran open-ended, the temporary state became permanent: forty-one percent of organizations report at least one such zombie system still billing licences today, and the finance reviews quoted later in this paper suggest how those systems end.

Governance maturity clusters into three groups the survey names ad hoc, documented, and enforced. Ad hoc organizations, thirty-one percent of the sample, decide access case by case and track no lineage. Documented organizations, the majority at fifty-four percent, hold written policy whose application depends on individual diligence, a posture one respondent summarized as compliance by hope.

Enforcement is rare and different in kind. Only fifteen percent are enforced, meaning the platform itself executes policy through access controls, automated lineage capture, and quality gates that block promotion of failing datasets; in this group, audit preparation is measured in days rather than weeks.

The outcome gap is the striking part. The enforced group reports forty percent fewer data incidents reaching consumers, and the survey found no correlation between governance outcomes and governance tooling spend: the correlation is with whether checks execute inside the pipeline or beside it.

Security posture around data platforms shows one specific, repairable weakness, and it is not encryption. Encryption at rest is effectively universal at ninety-six percent adoption, which the survey reads as evidence that the industry solves what its compliance checklists name and neglects what they omit. The neglected item is machine identity.

Human passwords rotate on schedule in sixty-six percent of organizations, but only twenty-eight percent rotate the service-account keys that connect pipelines to warehouses, and nineteen percent admit to a shared credential used by more than five systems. Breaches reported in the previous three years traced to compromised service credentials in more than half of cases, ahead of misconfigured storage at twenty-three percent.

Data quality economics received a dedicated section for the first time this year, and the numbers explain why quality initiatives struggle for sponsorship. Respondents estimate that analysts spend twenty-seven percent of their week validating or reconciling figures they do not trust, yet only twelve percent of organizations can attribute a monetary cost to a specific data incident after the fact.

Where attribution exists, behavior changes fast. The eight percent of organizations running incident cost accounting report quality budgets approved at twice the sample rate, and one retailer credits a single attributed incident, a mispriced promotion costing 640 thousand euros in eleven hours, with unlocking its entire quality programme.

The skills picture inverts a common assumption: organizations do not primarily lack specialists in individual tools, they lack people who understand the seams. Pipeline debugging across system boundaries was rated the scarcest capability at fifty-eight percent, ahead of any single-product expertise, and training budgets point the opposite way, with seventy percent of spend going to vendor certifications.

Onboarding time makes the seams cost concrete. New data engineers need a median of nine months to full productivity in stacks above fourteen tools, against four months below eight; the two organizations running an internal seams rotation, embedding newcomers with adjacent teams for two weeks, both beat the median for their size class.

Team topology showed more variation than any technical dimension. Centralized data teams still serve forty-six percent of the sample, embedded analysts report into business units in thirty-one percent, and a platform-team model with federated ownership covers the remaining twenty-three, with satisfaction scores favoring the platform model by a wide margin.

The ratio behind the platform model is consistent enough that the survey publishes it as a planning number: one platform engineer sustains the tooling of roughly nine data producers and consumers at the median, and organizations below a one-to-fifteen ratio report the highest burnout and the longest change lead times in the sample.

Vendor strategy divides the sample into three camps of roughly equal size: single-vendor adopters, deliberate multi-vendor architects, and accidental multi-vendor inheritors, the last being the largest at forty-four percent. Lock-in fear remains the stated reason to resist consolidation, and the survey's contribution is to price that fear.

Accidental multi-vendor organizations pay an integration premium that exceeds the worst-case switching cost they themselves estimate within roughly three years. The paper does not call lock-in fear irrational; it calls it mispriced, because a certain continuous cost hidden in payroll is discounted against an uncertain one-time cost that would appear on an invoice.

Artificial intelligence workloads entered the survey this year and are already reshaping platform requirements. Fifty-seven percent of organizations report that at least one production AI use case reads from the analytical platform daily, and this group ranks retrieval quality and data freshness above dashboard latency, reversing every previous edition of the survey.

The AI wave restarts sprawl unless resisted deliberately. Embedding pipelines, vector stores, and evaluation harnesses arrived as three or more new tools in seventy percent of AI-active organizations; the minority that held the line did so by extracting roadmap commitments from incumbent vendors and accepting a capability lag they estimate at two quarters.

Data contracts moved from conference talk to measurable practice. Twenty-four percent of organizations now enforce schema and semantics agreements between producing and consuming teams at deploy time, and they report a distinctive benefit: breaking changes caught before production rose to eighty-three percent, against thirty-one percent for teams relying on downstream tests alone.

Adoption of contracts correlates with the enforced-governance cluster rather than with company size, supporting the paper's broader thesis that the execution point of a control matters more than its existence on paper. Contract tooling remains young, and half of the enforcing organizations built theirs internally.

Cost allocation practices split the sample cleanly. Fifty-one percent of organizations still

treat the platform as a shared overhead, thirty-five percent run showback reports that inform but do not bill, and fourteen percent charge consuming teams directly; the charging group reports the lowest idle-resource waste, at seven percent of compute spend against nineteen for the overhead group.

Showback appears to be the pragmatic middle: it captures most of the waste reduction, eleven percent idle at the median, without the organizational friction of internal billing, and it is the only allocation model whose adopters grew year over year in the survey.

A regional insurance group anonymized as Case Delta illustrates the consolidation playbook succeeding. Delta entered 2023 with sixteen production tools after two acquisitions and a 5.1 million euro budget, of which integration consumed an estimated thirty-five percent, and it committed to the domain-sequenced approach with published shutdown dates.

Eighteen months later Delta had retired seven systems, migrating claims, then policy, then customer, freezing legacy feature work at the start of each domain. The programme cost 1.9 million euros, reduced annual platform spend by 1.4 million, and cut median pipeline incident recovery from six hours to under two, an outcome its architecture lead attributes to sequencing discipline rather than to the destination platform.

A North American retailer anonymized as Case Kilo illustrates the failure mode with equal clarity. Kilo attempted the same scope as a single two-year programme, ran legacy and target in parallel without shutdown dates, and let feature development continue on both stacks to placate stakeholders, so the migration backlog grew faster than domains moved.

At month thirty the finance function froze the programme in a review whose finding the survey quotes with permission: we are paying twice to stand still. Kilo restarted in 2025 on the domain-sequenced model with mandatory shutdown dates, adopting the Delta playbook it had initially rejected as too rigid.

Looking forward, respondents expect tool-count pressure to intensify before it eases. The median organization forecasts adding two AI-related components in twelve months while retiring one, a net increase, and only the enforced-governance group forecasts a net decrease; budgets meanwhile stay flat, with sixty-two percent expecting spend within five percent of current levels.

The squeeze lands on the lines that already hurt. Integration and reliability absorb forty-three percent of spend today, and flat budgets with a growing component count imply that share rises, which the survey reads as the setup for a consolidation cycle in 2026 and 2027 executed under pressure rather than growth, historically the harder variant.

The recommendations close the paper in the order respondents rated their expected impact. Pricing integration explicitly, the Delta finance practice generalized, came first: respondents rated this highest at 8.1 out of 10, because it changes procurement behavior without requiring a migration. Migrating by domain with published shutdown dates followed at 7.8.

Moving governance checks inside the pipeline rated 7.4, as the prerequisite for raising change velocity safely during consolidation. Rotating service credentials with the discipline applied to human ones rated 7.2, described by one respondent as the cheapest material risk reduction available. Extracting vendor roadmap commitments before buying AI point solutions closed the list at 6.9, and Meridian expects the AI section of the 2026 edition to double in length.

Regional differences within the sample are larger than the industry's uniform marketing

suggests, and they concentrate in three areas: procurement cadence, sovereignty requirements, and the appetite for managed services. European respondents run procurement cycles a median of two months longer than North American ones, driven by works-council consultation and data protection review, and they compensate with contract terms a full year longer on average, trading agility for stability in a pattern the survey has observed in every edition since 2021.

Sovereignty is no longer a niche concern of regulated industries. Forty-eight percent of European organizations now carry a contractual or regulatory requirement pinning at least one dataset to a national territory, up from twenty-nine percent two years ago, and a third of those report having disqualified a vendor during selection specifically over the location of support personnel rather than of the data itself. North American respondents report territorial pinning at only seventeen percent, but the trajectory points the same direction.

The managed-service split follows headcount rather than geography once controlled for company size. Organizations below the sample's median size buy managed offerings for sixty-two percent of their platform components, while the largest quartile self-hosts fifty-five percent of components, citing unit economics at scale and negotiating leverage. The middle of the distribution is migrating toward managed: of respondents changing hosting mode for any component this year, four in five moved toward the managed variant.

Reliability engineering practices around data pipelines lag their software counterparts by roughly a decade, and the survey quantifies the lag precisely. Only twenty-one percent of organizations define availability objectives for their critical datasets, against eighty-six percent that define them for customer-facing applications, and fewer than one in ten attach an error budget that data teams actually spend. Where objectives exist, they were set by the platform team alone in most cases, without consumer input, which respondents themselves flag as a defect.

Incident handling shows the same immaturity. The median organization detects a broken pipeline through a consumer complaint rather than an alert in thirty-eight percent of incidents, and post-incident reviews happen for fewer than half of data incidents against ninety percent of application outages. The organizations that close this gap share one practice: they route data incidents through the same on-call and review machinery as application incidents, rather than maintaining a parallel, softer process for data.

Observability tooling adoption is broad but shallow. Seventy-nine percent of organizations collect pipeline execution metrics, yet only thirty-one percent trace a record's journey across systems well enough to answer where a specific figure came from without manual archaeology. Respondents rated lineage completeness as the largest gap between what their tools promised and what they deliver, scoring the promise-delivery gap at 3.4 points on a five-point scale, the worst of any capability surveyed.

The semantic layer, defined in the survey as a governed mapping from business terms to physical data, crossed from aspiration to practice this year. Twenty-eight percent of organizations now serve at least one critical dashboard exclusively through a semantic layer, and those organizations report metric disputes in executive meetings at half the sample rate. The dissenting view, held by nineteen percent, is that the layer adds a bottleneck team; the survey notes that every organization holding that view runs the layer as a centralized function rather than a federated one.

Real-time processing turns out to be narrower than its mindshare. Only fourteen percent of organizations operate true streaming pipelines with sub-minute freshness in production, and

half of those apply them to exactly one use case, typically fraud screening or operational alerting. The much larger population, sixty-one percent, refreshes hourly or faster in micro-batches and reports that this satisfies their stated business requirements; respondents repeatedly distinguished between the freshness the business asked for and the freshness engineering assumed it wanted.

The cost profile of streaming explains the caution. Organizations running streaming report its per-pipeline operating cost at roughly three times the batch equivalent, driven by always-on capacity and the scarcity of engineers comfortable with stateful stream semantics. The survey's freshness recommendation is procedural rather than technical: write the required freshness into the data contract of each dataset, and let the number, not the architecture fashion, choose the processing mode.

Open-source components hold a stable share of the surveyed stacks at forty-one percent of components by count, but the governance of that share is changing. Sixty-four percent of organizations now require a supported distribution or a commercial support contract for any open-source component on the critical path, up nine points year over year, and license scanning at build time has become the norm at seventy-one percent adoption. The motivation respondents cite first is not legal exposure but the operational risk of unmaintained dependencies.

Contribution flows the other way remain rare. Seven percent of organizations report employees contributing patches upstream on work time, concentrated almost entirely in the technology and financial sectors, and three organizations in the sample maintain a component they originally adopted, having hired its maintainer, a pattern the survey expects to grow as foundational projects age.

Mergers and acquisitions deserve their own playbook, because they are the single largest generator of the duplication documented earlier. The organizations that integrate data platforms fastest after a merger run a standing integration team with a pre-approved target architecture, decide within ninety days which stack survives for each domain, and measure integration progress by retired systems rather than by connected ones. The median organization without such a team still runs both inherited stacks three years after closing.

The survey's own limitations deserve statement. The sample skews toward organizations willing to discuss their platforms, which plausibly biases maturity upward; spend figures are self-reported and were not audited; and the significance testing applies to segment comparisons, not to year-over-year deltas, which the panel design cannot support at equal rigor. Meridian publishes the full instrument, the anonymized aggregates, and the segment tables so that readers can judge the evidence directly.

Methodologically, two changes from the 2024 edition matter for comparability. The tool-count question now explicitly excludes libraries and includes only deployed services with their own operational surface, which removes roughly two items from the average respondent's previous count, and the cost decomposition question moved from percentages to absolute figures converted centrally, which respondents found harder but which removed the anchoring artifact the 2024 edition documented.

Reading the findings together, the paper's throughline is that the industry's data platform problem is not a technology selection problem. Every failure pattern documented here, from zombie systems to unrotated credentials to disputed metrics, is an execution and ownership pattern, and every success pattern is a sequencing and enforcement pattern. The organizations that do well buy roughly the same technologies as the ones that struggle; they

retire more, enforce more, and price the invisible costs.

The practical corollary for a reader planning 2026 is to treat the consolidation cycle as an organizational programme with a technical component, not the reverse. Fund it from the integration line it will shrink, sequence it by domain, publish shutdown dates, move the controls into the pipeline while change velocity is elevated, and rotate the machine credentials before an incident makes the case. None of these steps requires a new purchase, which may explain why they are rated highest by the people who have tried them.