

BEYOND THE GOLDEN RECORD

The Twenty-Year Problem: Why Master Data Management Keeps Failing

How a discipline that got its technology right kept getting its politics wrong

Giovanni Leonardi

October 2019



The Twenty-Year Problem: Why Master Data Management Keeps Failing

Transformation · No. 236

Originally written October 2019 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

"A single version of the truth is a political settlement before it is a data model, and no tool has ever been able to negotiate one on our behalf."

1. When the Board Asks a Simple Question

Somewhere in most large organisations, a version of this scene has already taken place. A non-executive director asks what ought to be the easiest question in the building: how many customers do we have? The room goes quiet, because everyone senior enough to be sitting in it knows the honest answer is "it depends who you ask."

Sales reports 2.3 million active accounts. Finance can defend 1.7 million billable legal entities. Marketing holds 4.1 million contactable individuals. Each figure is correct — rigorously, auditably correct — against its own definition of the word customer. And somewhere along the corridor sits a master data management programme, now in its third year and on its second sponsor, whose entire purpose was to end exactly this confusion. Its contribution to the discussion is a fifth number that none of the three functions recognises and none of them will use.

I have watched this scene resolve in the same way more than once, and the tell is always the same: the argument is never really about the data. It is about who has to change. The moment a single figure is agreed, one function's reports break, one function's targets have to be recut, and one function has to explain to its own leadership why the number it has reported for years was apparently wrong. That is the true subject of the meeting, and no data model has ever settled it.

We have now been trying to master this problem for roughly two decades. The customer-data-integration projects of the late 1990s became the master data management programmes of the mid-2000s, which became the data-governance functions and the golden-record platforms of the years since. Twenty years of investment, and the board still cannot get one number. The comfortable explanation is that the technology was not yet ready. That explanation is wrong, and its wrongness is the whole point.

Master data quality is not a technical problem wearing a political mask. It is a political problem wearing a technical one — and we keep sending technical people to negotiate it.

2. Why the Definition Refuses to Settle

Master data is the place where an organisation's internal disagreements are written down. The reason "customer" cannot be pinned to one meaning is not that three teams are being sloppy; it is that they genuinely mean three different things, and each is measured and paid on the meaning it holds. Sales is compensated on accounts opened. Finance is audited on legal entities that can be billed. Marketing is judged on individuals it is permitted to contact. The divergence between their numbers is not dirt to be cleaned — it is the faithful residue of three legitimate operating models sitting inside one company.

This is why the single-definition ambition runs aground so reliably. When a programme proposes one canonical meaning of customer, it is not proposing a schema. It is proposing that one function's worldview should win and that the others should absorb the cost of conforming to it — new mandatory fields, re-keyed records, broken management reports, retrained staff — in exchange for a benefit that accrues mostly to the centre and to people they will never meet. Asked to pay a certain local cost for a diffuse corporate good, rational managers decline, quietly and indefinitely. The golden record gets built, and then it gets starved.

The mechanics of the starvation are always recognisable. A programme spends its first year on match and merge — survivorship rules, probabilistic matching, deduplication — and proudly collapses 400,000 duplicate customer records into 260,000 mastered ones. It is genuine work and it photographs well on a steering-committee slide. Six months later the duplicate count is climbing again, because the operational system that produced the duplicates is still in production, still rewarded for opening accounts quickly, still measured on speed rather than uniqueness, and still feeding the same mess downstream every night.

"The programme cleaned the lake while leaving the tap running."

That is the shape the twenty years actually take: heroic one-time remediation, no change to the incentives that generated the defects, and a steady regression to the mean. We treat the symptom as the disease, and we are surprised, every time, when the fever returns.

3. "But the Tools Finally Work Now"

The strongest objection to all of this deserves to be stated at full strength, because it is largely true. The tooling of today really is better than anything we had ten years ago. Cloud-hosted platforms remove the eighteen-month infrastructure prelude.

Graph-based models hold the tangled relationships between people, households, accounts and organisations that the old relational hubs flattened and lost. Machine-assisted matching resolves entities that hand-written rules never caught. Application programming interfaces let systems ask for a mastered record in real time instead of waiting for an overnight batch. A capable engineer in 2019 can stand up a matching capability in weeks that would have taken a programme in 2006 the better part of a year.

All of that is real, and all of it is worth having. But it answers a question almost nobody was truly stuck on. Matching accuracy was never the binding constraint. Even a decade ago, most programmes could get the records to roughly the right level of correctness; the residual errors were rarely what killed them. What killed them — what still kills them — is that no algorithm can make a business unit want to surrender its own version of the truth, and no platform can make a sponsor accept the standing operating cost of maintaining someone else's. Point the best matching engine ever built at an unchanged operating model and you get a more accurate golden record that still nobody uses: a sharper, faster, more elegant answer to a question the organisation has never agreed to ask.

If anything, the past two years have made the stakes plainer rather than the problem easier. Since the data-protection regime tightened, a single subject access request or erasure request forces precisely the question these programmes were meant to answer — which of our forty systems actually holds this person, and can we prove they are the same person across all of them? The new obligations did more to expose the true cost of unmastered data than a decade of business cases ever managed. But exposure is not resolution. Faced with it, most organisations have reached, once again, for another remediation project — a bigger clean, a better hub — rather than for the thing that was actually missing.

4. Master the Decision, Not the Architecture

What was missing is not a tool. It is a change in how the work is chartered and owned, and it comes down to two moves.

The first is to stop running master data as a project and start running it as an operating capability with named ownership. A project has an end date and a golden record as its deliverable; a capability has a persistent owner, a budget that does not expire, and a standing mandate to arbitrate. The arbitration is the part that matters. The centre should own the definitions and the process for resolving disputes about them, but each domain — customer, product, supplier — needs a single named owner who is accountable for that data the way a profit-and-loss owner is accountable for a number, and who feels a consequence when it is wrong. Governance that cannot make anyone, anywhere, feel a consequence is not governance. It is theatre with a steering committee.

The second move is narrower and more heretical: master the data a decision depends on, not the data an architecture admires. The totalising ambition — master everything, one hub, one truth for the whole enterprise — is exactly what makes these programmes simultaneously enormous and unfinishable. Invert it. Begin from a decision that a named person is accountable for — a regulatory return that must reconcile, a credit exposure

that must not be double-counted, a cross-sell campaign that must not be aimed at a customer who left last month – and master only the attributes that that decision genuinely relies on, to only the accuracy it genuinely needs. Then let the rest stay messy.

The golden-record project	The mastered-decision capability
Deliverable is a complete central record	Deliverable is a trustworthy answer to a specific decision
Funded once, to an end date	Funded continuously, as a running cost
Owned by the centre and by IT	Owned by a named domain accountable for consequences
Success measured in records mastered	Success measured in decisions it can be trusted for
Scope is the whole enterprise	Scope is the attributes a live decision depends on

A narrow mastered set that a real decision leans on will be maintained, because someone notices the moment it breaks and that someone has a name. A comprehensive golden record that no live decision depends on will rot quietly, because nobody is watching the parts that are decaying. The instinct to master everything feels like rigour. In practice it is how we guarantee that nothing stays mastered.

5. The Twenty-Year Lesson

Return to the boardroom and the simple question. The organisations that eventually learn to give one answer are not, in my experience, the ones that bought the most capable hub or ran the largest cleanse. They are the ones that decided who owns the word customer, gave that person real authority over its meaning, and made them carry the consequence of getting it wrong. The technology to support them has been adequate for years. What was scarce, and remains scarce, is the willingness to treat master data as a question of governance and accountability rather than of tooling.

A single version of the truth is a political settlement before it is a data model, and no tool has ever been able to negotiate one on our behalf. Until we accept that – until we stop asking engineers to resolve, in software, a disagreement that is really about ownership, incentives and cost – we will spend the next twenty years exactly as we spent the last: buying steadily better answers to a question we have not yet agreed to own.