

BEYOND THE GOLDEN RECORD

The Twenty-Year Problem: Why Master Data Management Is Never Solved, Only Rediscovered

On ownership, incentives, and why the enterprise keeps rebuilding a golden record it will not maintain

Giovanni Leonardi

December 2019



Contents

1. Executive Summary	3
2. The report that is always amber	3
3. A twenty-year-old problem in successive suits	4
4. Why it is structural, not technical	5
5. The programme delusion	6
6. The regulation that changed the stakes	7
7. The steelman: it is solved, you simply will not pay	8
8. What twenty years should have taught us	8

The Twenty-Year Problem: Why Master Data Management Is Never Solved, Only Rediscovered

Transformation · No. 291

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

“Master data is unowned, contradiction is rational, investment is invisible, and attention is transient.”

1. Executive Summary

Master data management is the problem the enterprise keeps rediscovering as though for the first time. Every few years a new programme is chartered to build the single customer view, the product master, the golden record — and every few years, once the funding thins and the sponsor moves on, the master data quietly begins to rot again. By the close of this decade the technology to solve it is abundant and unglamorous: matching engines, survivorship rules, cloud hubs, probabilistic algorithms that resolve identities better than any human clerk. And yet the problem persists, almost exactly as it did twenty years ago, when we called it data warehousing and believed the warehouse would settle it once and for all.

This essay argues that master data has remained unsolved for two decades not because it is technically hard — though it is — but because it is structurally unowned. Master data sits in the seams between functions, and organisations are built to fund and reward what happens inside functions, not between them. The recurring MDM programme is the symptom of a deeper mismatch: we treat as a project something that is only ever a capability, and we ask a temporary initiative to hold back a permanent tide. The regulation now bearing down on data — the obligations that arrived last year and the ones arriving in the new year — has raised the stakes sharply, turning an efficiency nuisance into a matter of legal exposure. That pressure may finally change the economics. But it will only do so if we stop asking how do we fix the data and start asking who is accountable for it, forever. The twenty-year lesson is that master data is not a thing to be built. It is a thing to be stewarded, and stewardship is precisely what our operating models are least equipped to sustain.

2. The report that is always amber

There is a particular slide that anyone who has spent time near a large data programme will recognise. It is the data quality dashboard, and its headline indicator has been amber

for as long as anyone can remember. Not red — red would provoke a decision. Amber. Amber says we are aware, we are managing it, it is under control, and amber is where master data goes to be ignored in plain sight.

I have watched that amber survive three consecutive transformation programmes at a single organisation. The first was going to consolidate the customer systems. The second, some years later, was going to build a data lake into which everything would flow and from which truth would emerge. The third rebranded the whole endeavour under the banner of digital and promised a 360-degree view of the customer as a foundational capability. Each was well funded. Each delivered something. And at the end of each, the customer master was marginally better than before and structurally no more owned than it had ever been. The amber remained.

What is striking, looking back across two decades, is not that these programmes failed — most delivered real improvements — but that the problem did not change shape. The vocabulary changed. In the late nineties it was the single version of the truth and the enterprise data warehouse. By the mid-2000s it was customer data integration and the golden record. In this decade it became the data lake, then the customer 360, and now, with regulation in the room, it is data governance and lineage. Different words, the same intractable thing underneath: an organisation that cannot agree what a customer is, held together by systems that each hold their own partial, confident, contradictory answer.

3. A twenty-year-old problem in successive suits

It helps to be honest about the age of this. The discipline we now call master data management is not new thinking dressed in new tools; it is old thinking that has never been allowed to finish.

The warehouse builders of the 1990s understood the core difficulty perfectly well. They knew that if two source systems disagreed about a customer's address, someone had to decide which was right, and that the decision was a business judgement, not a technical one. They built conformed dimensions and reconciliation routines and they argued, famously and at length, about whether truth should be modelled top-down or assembled bottom-up. What they could not do was make the argument stop — because the disagreement between systems was not a bug to be fixed once, it was a standing condition of an organisation whose parts each captured the customer for their own purposes.

The 2000s gave the problem its own name and its own market. Master data management became a category, with platforms, quadrants, and a consulting practice attached. The framing was seductive: buy the hub, define the golden record, let survivorship rules arbitrate between sources, and the master would maintain itself. A great deal of genuine capability was built in this period. But the framing carried a quiet lie — that the hard part was the arbitration logic. The hard part was never the logic. The hard part was getting the marketing function and the finance function and the operations function to accept the same customer, when each had spent years optimising its own version for its own numbers.

By this decade the pattern had acquired yet another suit. The data lake promised to sidestep the whole reconciliation problem by simply keeping everything, in raw form, and resolving meaning later, on read. It was an appealing evasion. In practice, deferring the reconciliation did not remove it; it relocated it downstream, to every analyst who now had to work out, alone and repeatedly, which of nine customer feeds to trust. The swamp that so many of these lakes became was not a failure of engineering. It was the old master data problem, unowned as ever, now pooled in one place where its unownedness was simply more visible.

Every generation of tooling has promised to make the arbitration automatic, and every generation has discovered that the arbitration was never the point. The point was the accountability for the answer – and that has never been automatable.

4. Why it is structural, not technical

If the technology were the binding constraint, twenty years of Moore's Law and a mature software market would have loosened it. They have not. That tells us the constraint lies elsewhere. Four structural forces, in my experience, sustain the problem far more powerfully than any technical limitation.

The first is that master data lives between functions, and organisations fund within them. A customer record is created by sales, enriched by service, billed against by finance, and analysed by marketing. It belongs, therefore, to none of them. Every function depends on it and every function treats improving it as somebody else's cost. A budget line has an owner; the space between budget lines does not. Master data is the archetypal thing that falls into that space.

The second is that the incentives reward local optimisation over shared truth. A sales team is measured on pipeline, and a duplicated customer that inflates the pipeline is not, from where they stand, a problem worth fixing. A billing team needs the address that gets the invoice paid, not the address that is globally correct. Each function's version of the customer is rationally maintained for that function's objectives. Asking them to subordinate their working copy to a shared master is asking them to accept a small local cost for a diffuse global benefit they will never be measured on. People do not do this reliably, and they should not be expected to.

The third is that we account for data as a cost, not as an asset. This sounds abstract; it is the most consequential force of all. An asset gets a custodian, a maintenance budget, and a line on which its condition can be tracked and someone held to account for its decline. Data gets none of these. It appears in no ledger, depreciates on no schedule, and its deterioration shows up only indirectly – in a marketing campaign that misfires, a regulatory return that has to be reworked, a risk exposure that turns out to have been understated because the same counterparty was counted as three. Because the decline is invisible in the accounts, the investment to arrest it must be justified afresh every single time, as a special case, against competing demands that have the great advantage of being visible.

The fourth is the half-life of sponsorship. Master data improves slowly and its benefits are cumulative and unspectacular. Executive attention is fast and rewards the spectacular. A programme chartered under one sponsor with a two-year horizon will, more often than not, outlive that sponsor's tenure or that horizon's patience. When the sponsor moves and the horizon closes, the capability is not maintained — it is concluded, marked as delivered, and the standing work of stewardship it was supposed to establish evaporates with the programme structure that briefly held it.

- Between functions, so unowned by any of them
- Locally optimised, so globally contradictory by design
- Accounted as cost, so re-justified from zero every cycle
- Sponsored in bursts, so never continuously maintained

Put these four together and the persistence of the problem stops being a puzzle. Master data is unowned, contradiction is rational, investment is invisible, and attention is transient. A problem with those four properties will not be solved by a better matching algorithm. It will be managed, forever, or it will rot.

5. The programme delusion

Here is the mistake that recurs so reliably it deserves a name of its own. Organisations keep treating master data as a programme when it is only ever a capability. The difference is not semantic; it decides the outcome before the work begins.

A programme has a start, an end, a budget, and a definition of done. It is the right instrument for building something that will then exist — a new billing platform, a migrated data centre. A capability has none of those things. It is a standing function that must be resourced in perpetuity, because the thing it maintains decays continuously the moment maintenance stops. Master data is unambiguously the second kind of thing, and we keep buying it with the instrument designed for the first.

Master data as a programme	Master data as a capability
Has a definition of "done"	Is never done; decay is continuous
Funded as a one-off investment	Funded as a standing operating cost
Success is the golden record built	Success is the golden record maintained
Ends when the sponsor's horizon closes	Persists across sponsors by design
Owned by a temporary team	Owned by a permanent, accountable role
Measures delivery of an artefact	Measures the ongoing quality of decisions

The programme framing is not chosen out of ignorance. It is chosen because it is fundable. A capability that costs money every year forever, and whose benefit is the absence of problems that never quite become visible, is nearly impossible to defend in a budget round against initiatives with a return you can point to. A programme, by contrast, has a beginning and an end and a business case with a number at the bottom. So the organisation reaches, again and again, for the fundable instrument — and each time gets a

golden record that is briefly built and then, unmaintained, resumes its decline. The delusion is not that anyone believes master data has an end. It is that the funding model can only pay for things that do.

6. The regulation that changed the stakes

For most of these twenty years, poor master data was an efficiency problem. It cost money in duplicated effort, wasted mailings, and analyses that had to be redone. It was expensive, but it was survivable, and its survivability was precisely why the amber never turned red.

That calculus has shifted. The data protection regime that came into force across Europe last year did something the efficiency argument never could: it attached legal consequence to not knowing who your customers are. When a person exercises their right to see everything an organisation holds about them, or asks to be erased, the organisation must be able to find every instance of that person across every system – and answer completely, within a deadline, on pain of penalty. That obligation is, whether anyone framed it this way in the business case, a master data requirement. You cannot honour a subject access request against a customer you have stored eleven different ways and cannot reconcile. The comparable regime arriving on the western side of the Atlantic in the new year will press in the same direction for a great many more organisations.

“Regulation did what two decades of efficiency arguments could not: it made not knowing who your customer is a liability rather than a nuisance.”

This matters because it changes the kind of argument that funds the work. For twenty years the case for master data was an efficiency case, and efficiency cases lose to growth cases in every budget round. A compliance case is different in nature. It is a case about avoiding a downside that is now concrete, dated, and enforceable, and downside-avoidance can command funding that upside-pursuit cannot. For the first time, the structural force that starved master data – its invisibility in the accounts – has a partial counterweight, because the cost of neglect has been made visible by law.

We should be clear-eyed about the risk in this, though. A capability funded by fear is funded only as far as the fear reaches. Compliance-driven master data tends to be narrow: it fixes what the regulation can see and leaves the rest amber. An organisation that builds its customer master purely to survive a subject access request will have a master good enough for the regulator and no better, and will have missed the far larger prize – that the same reconciled view that satisfies the regulator is the one that lets you understand a customer, price risk against them, and serve them as a single relationship rather than a scatter of accounts. Regulation can open the door to funding the capability properly. It cannot be trusted to walk the organisation through it.

7. The steelman: it is solved, you simply will not pay

The most serious objection to everything above is worth stating in its strongest form, because it is at least half right. The objection runs: master data is a solved problem. The methods are known, the tools are mature, the reference architectures are published. The reason it appears unsolved is not that anyone lacks the knowledge; it is that organisations will not fund the boring, permanent stewardship the solution requires. Stop intellectualising a problem of ordinary corporate discipline.

This deserves respect, because at the level of technique it is correct. We do know how to build a customer hub. We know how to write survivorship rules, how to tune a probabilistic matcher, how to stand up a stewardship workflow. None of that is a research problem. If will and money were unlimited, the customer master would be built and kept, and this essay would be unnecessary.

But the objection mistakes a description of the failure for an explanation of it. To say "organisations will not fund permanent stewardship" is not a rebuttal of the structural argument — it is the structural argument, stated as a complaint rather than a diagnosis. The interesting question is not whether organisations under-fund stewardship; everyone agrees they do. The interesting question is why the under-funding is so universal and so durable across firms of every size, sector, and level of sophistication. A failure that consistent is not a failure of individual discipline. It is a property of the system. When the same behaviour appears everywhere regardless of who is in charge, the explanation almost never lies in the character of the people in charge; it lies in the incentives they are all responding to. The steelman is right that the answer is money and will. It is wrong to treat money and will as things a determined leader simply supplies. They are outputs of a structure, and until the structure changes, exhortation will not supply them.

8. What twenty years should have taught us

If the problem is structural, the response has to be structural too. Tooling helps — modern matching genuinely is better than what we had a decade ago — but tooling has never been the binding constraint, and buying more of it in programme form will reproduce the same result with higher resolution. What changes the outcome is a small number of moves that attack the structure rather than the symptoms.

1. Give master data a permanent owner with a budget, not a programme with an end. The single most decisive act is to make a standing role accountable for the condition of the customer master the way a treasurer is accountable for cash — not to deliver it once, but to be answerable for its state indefinitely. The recent rise of a senior data role in many organisations is the raw material for this, provided the role is given a maintenance budget and not merely a mandate to write policy.
2. Fund it as an operating cost, and measure its decline. Stop financing master data as episodic capital investment and carry it as a standing line, the way any other maintained asset is carried. Put a small number of honest quality measures on it — duplication rate, reconciliation coverage, time to resolve a subject request — and track their trend, so that neglect becomes visible in a number before it becomes visible in a crisis.

3. Push ownership of each domain to the function that lives it. The newest and most promising reaction to the failure of the central hub is to stop pretending a single team in the middle can own everyone's data. The function that creates and depends on a domain — the people closest to what the data actually means — should own its quality as a product they publish for others to consume, with the central capability providing the standards, the tooling, and the arbitration of last resort rather than doing all the work itself. Ownership follows meaning; meaning lives in the domains.
4. Let compliance open the door, but do not let it define the room. Use the new regulatory pressure to win the funding that efficiency arguments never could — and then deliberately build past the compliance minimum to the reconciled view that also serves the customer and prices the risk. The regulation is the lever; it must not become the ceiling.

None of this is a matching algorithm, and that is the point. The reason master data has been unsolved for twenty years is that we kept answering a structural question with a technical solution, and a permanent problem with a temporary team. The organisations that will finally break the pattern are not the ones with the best matching engine. They are the ones that stop chartering the fourth programme and instead do the unglamorous thing the first three never did — appoint someone to own the customer master forever, fund them the way an asset deserves, and hold them to account for a number that measures not what was built but what is being kept.

The amber on that dashboard was never really a data quality signal. It was an ownership signal, and it has been telling us the same thing for twenty years. We have simply preferred to read it as a problem for the next programme.