

INSIGHT · SAP TRANSFORMATION

Your clean core is only as clean as your data

Why the data clean-up is the most overlooked workstream in an S/4HANA program

01 Introduction

Every SAP Transformation program has a clean core roadmap these days. Fewer modifications. Standard processes. Cloud-ready extensions on BTP, not buried in the core. It is, rightly, the north star for every S/4HANA program right now.

But here is the question I keep asking steering committees, and the one that usually gets a pause before an answer: your core can be as clean as you want on the process and code side, but is it actually clean if the master data running through it is still full of duplicates, dead records, and mismatched reference values?

Because your clean core is only as clean as your data.

Garbage in, in terms of enterprise data, equals garbage out: bad quality reports, slow business processes, and now, increasingly, AI and automation that confidently produce wrong answers at machine speed.

02 Why data quality matters

Three things happen when data quality is left unmanaged, and every SAP PM has lived through at least one of them.

01 It produces waste

Employees spend an enormous amount of time fixing errors, double-checking bad records, and reconciling numbers that should never have needed reconciling.

02 It degrades choices

Leaders make flawed calls based on information that looks authoritative but is not, and rarely find out until the decision has already shipped.

03 It blocks AI success

Every AI or advanced-analytics initiative bolted onto SAP, whether demand sensing, predictive maintenance or agentic workflows, inherits the quality of the master data underneath it. Advanced tools do not fix poor inputs, they scale them.

15–25%

of revenue is the estimated cost of bad data for most companies. Roughly two-thirds of that cost can be identified and permanently eliminated once the root causes are addressed.

THOMAS C. REDMAN, MIT SMR

\$12.9m

is what poor data quality costs the average organisation every year, measured in operational terms rather than as a share of revenue.

GARTNER

That second Redman number matters as much as the first. This is not a permanent tax you are stuck paying. It is a fixable, one-time investment with a payback.

Neither figure is theoretical for an SAP shop. They show up as delayed invoicing, mismatched vendor records, reports that do not consolidate across regions, and go-lives that stall in UAT because nobody flagged the data six months earlier.

03 The SAP-specific reality

Here is what makes this more urgent than a generic "data quality matters" argument: in SAP, some of this clean-up is not optional. The system enforces it.

CUSTOMER-VENDOR INTEGRATION

CVI is the conversion of your legacy Customer and Vendor master records into the unified Business Partner model that S/4HANA requires. This is not a nice-to-have hygiene exercise you can defer to phase two. **SAP's own Software Update Manager will stop and fail the conversion if CVI is not completed.** Every duplicate tax ID, every missing postal code, every inconsistency between a Customer record and its matching Vendor record has to be reconciled before the technical conversion can even proceed.

In other words, SAP already agrees with the thesis of this post. It just enforces it at the CVI checkpoint instead of trusting your governance process to catch it earlier.

Which means the only real choice a PM has is whether that clean-up happens as a planned, funded workstream starting months before cutover, or as an unplanned fire drill discovered during the technical conversion, with the go-live date already fixed and the pressure already on. No prizes for guessing which one is the nightmare scenario.

“

Your clean core is only
as clean as your data

QORELO · SAP TRANSFORMATION



04 The real cost of letting it ride

Duplicate or obsolete master records do not just sit there quietly. They cost money in ways that are easy to miss until someone goes looking.

— Licensing

SAP HANA is licensed on memory, and HANA licences are organised in fixed steps. You cannot buy half a terabyte more. If your bloated, duplicate-riddled dataset pushes you even slightly over your current tier, you are not paying for the overage. You are paying for the entire next block. Cleaning up obsolete records before sizing your HANA environment can be the difference between staying inside your current tier and triggering a full step-up.

— Performance

Every report, every batch job, every reconciliation run gets slower as it churns through records nobody needs anymore.

— Global reporting

Duplication across regions or business units means global numbers do not consolidate cleanly. Someone ends up manually adjusting a "final" number before it goes to the board, every single quarter.

05 Master data clean-up

This is where the actual work lives, and it is more involved than a simple delete.

- Duplicate records need to be marked obsolete and archived, not silently deleted.
- That requires real insight into which transactions are still referencing the flagged, obsolete master data. You cannot retire a vendor record if open POs still point to it.
- Which means the real work is either closing out those transactions, or re-pointing them to the surviving master record, before the obsolete one is retired.

This is the unglamorous part of a transformation program that never makes it into the steering deck, and it is exactly the part that determines whether cutover weekend is smooth or a war room.

06 Reference data clean-up

The same discipline applies to reference data.

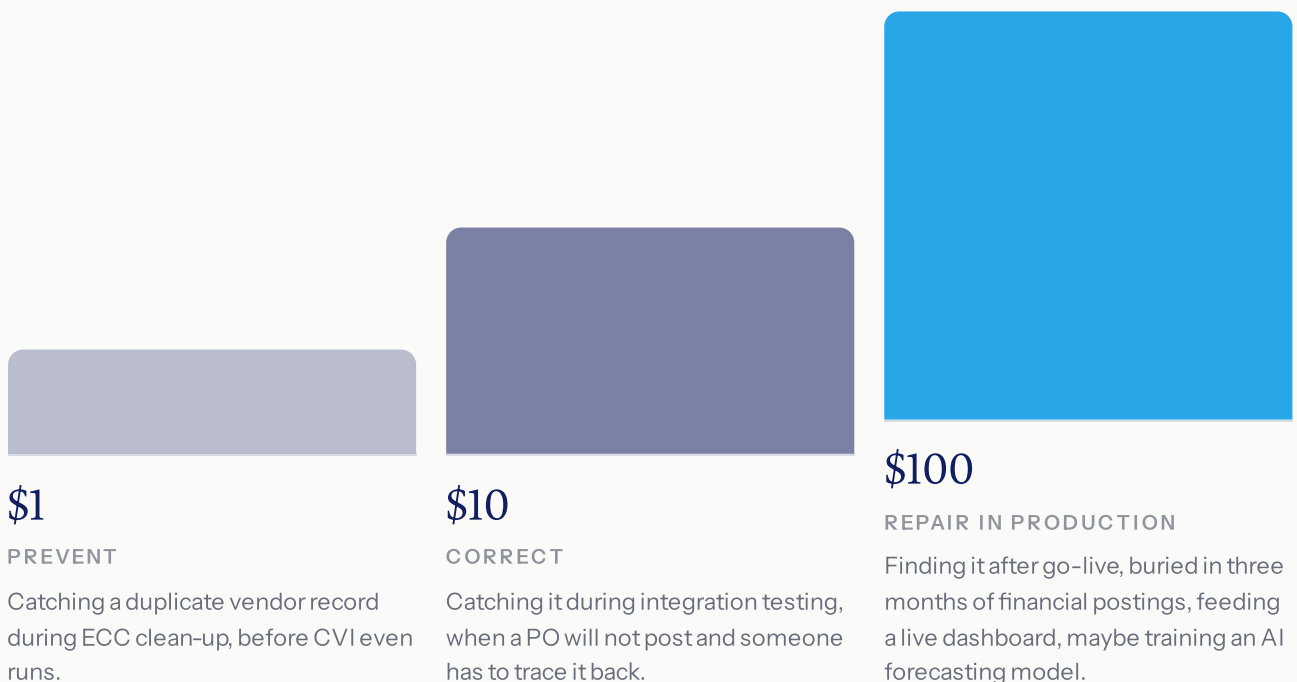
- New Tax IDs delivered in S/4 need to be adopted, with the obsolete legacy values formally retired, not left to coexist and cause reconciliation drift.
- New Region Codes for Country fields need to be retrofitted onto Supplier and Customer Master Data UI forms, with the old codes retired in parallel.

Skip this, and you inherit a system that technically converted but functionally still runs on two versions of the truth.

07 The 1-10-100 rule: why timing is the whole game

If there is one framework worth pinning to the wall of every SAP transformation PMO, it is the 1-10-100 rule, first articulated by Labovitz and Chang in 1992.

RELATIVE COST TO FIX ONE DATA DEFECT



By the \$100 stage it is not really a data problem anymore. It is a credibility problem, a re-audit, a steering committee asking hard questions about a program that was supposed to be done.

08 The takeaway

Every SAP Transformation program invests heavily in process design, custom code reduction, and a clean core roadmap for the technology stack. Far fewer invest with the same seriousness in the data that will run through that clean core on day one.

CVI will force your hand on some of it. HANA's licensing tiers will quietly bill you for the rest. And the 1-10-100 rule tells you exactly what it costs to find out the hard way versus the easy way.

THE BOTTOM LINE

Your clean core is only as clean as your data. Treat the clean-up like the milestone it actually is, not the housekeeping task it gets mistaken for.

If you are scoping an S/4HANA program and the data workstream is still a line item rather than a funded milestone, that is the conversation worth having now, not at CVI.

SOURCES REFERENCED

Thomas C. Redman ("Data Doc"), "Seizing Opportunity in Data Quality," MIT Sloan Management Review. Gartner research on the financial impact of poor data quality. SAP documentation and practitioner guidance on Customer-Vendor Integration (CVI) and SAP HANA memory-based licensing. Labovitz & Chang, "Making Quality Work" (1992), origin of the 1-10-100 rule.