AI in data management

| Hype or help?

CDQ
SHARING DATA EXCELLENCE

AI changes the economics of data management. Done right, it cuts manual effort, lowers cost, and improves control. Done wrong, it scales errors, rework, and risk.

KEY TAKEAWAYS

1 Cut cost, not control

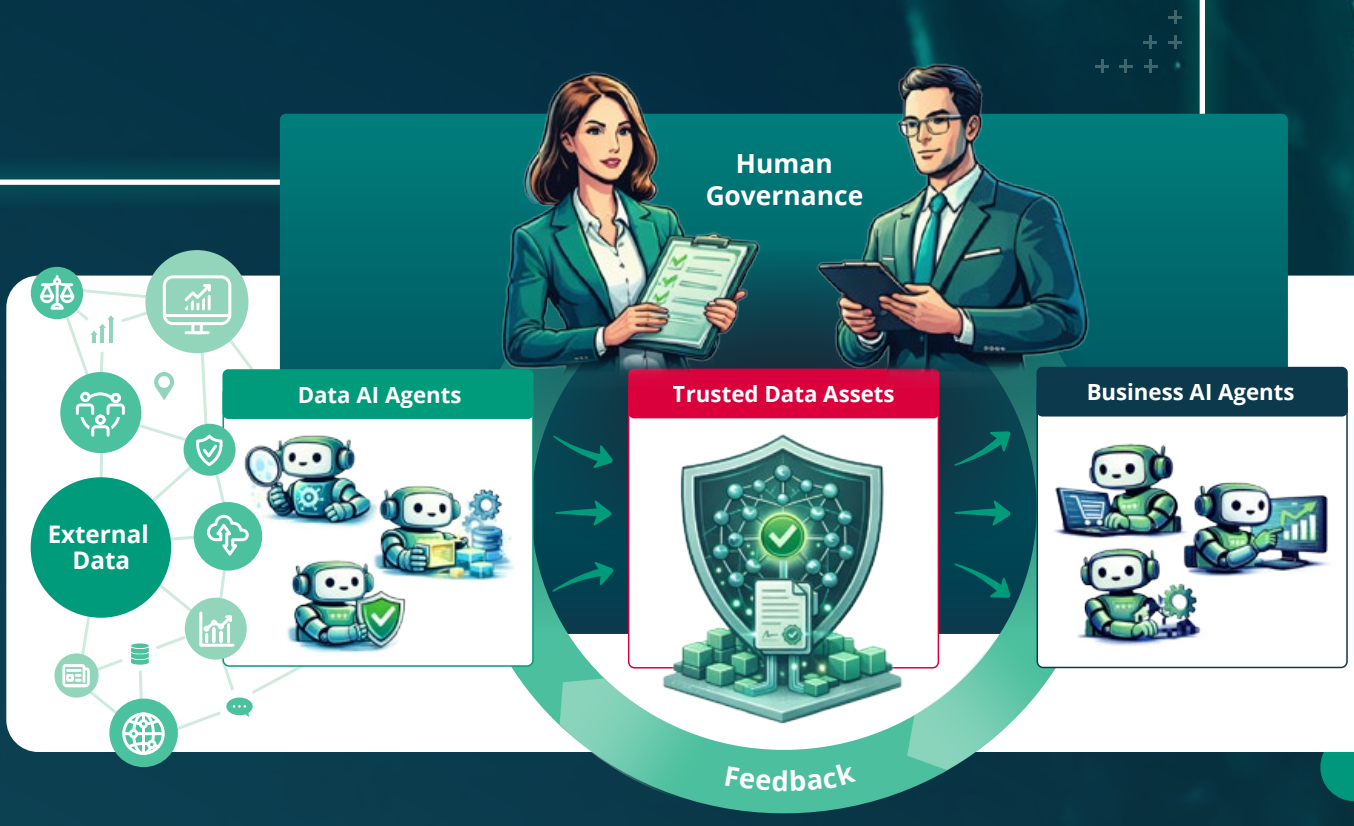
AI shifts repetitive data maintenance work from manual teams and standardized routine tasks to governed agents. This lowers operating cost, increases throughput, and frees experts for exceptions and policy decisions.

2 Trusted data, real value

AI creates business value only when it runs on trusted sources, governed logic, and accountable ownership. Weak data foundations turn AI into a multiplier of defects and downstream cost.

3 Governance decides the outcome

AI delivers reliable results in clearly governed tasks with explicit boundaries. Uncontrolled AI autonomy turns local data issues into systemic operational and compliance risk.



Executive summary

AI creates value in data management only when it works on trusted source data and governed rules. Web-trained LLMs are not a trusted source, they are reasoning layers that must operate on authoritative inputs.

KEY TAKEAWAYS

1 Trusted inputs determine outcomes
AI delivers value only when it works on authoritative data. Weak inputs produce weak results, while trusted inputs create the basis for reliable automation, lower effort, and better business decisions.

2 Hybrid data management sets the model:
The strongest operating model combines deterministic controls, AI agents, and human accountability. Rules secure correctness, agents handle scale, and people own policy, exceptions, and release decisions.

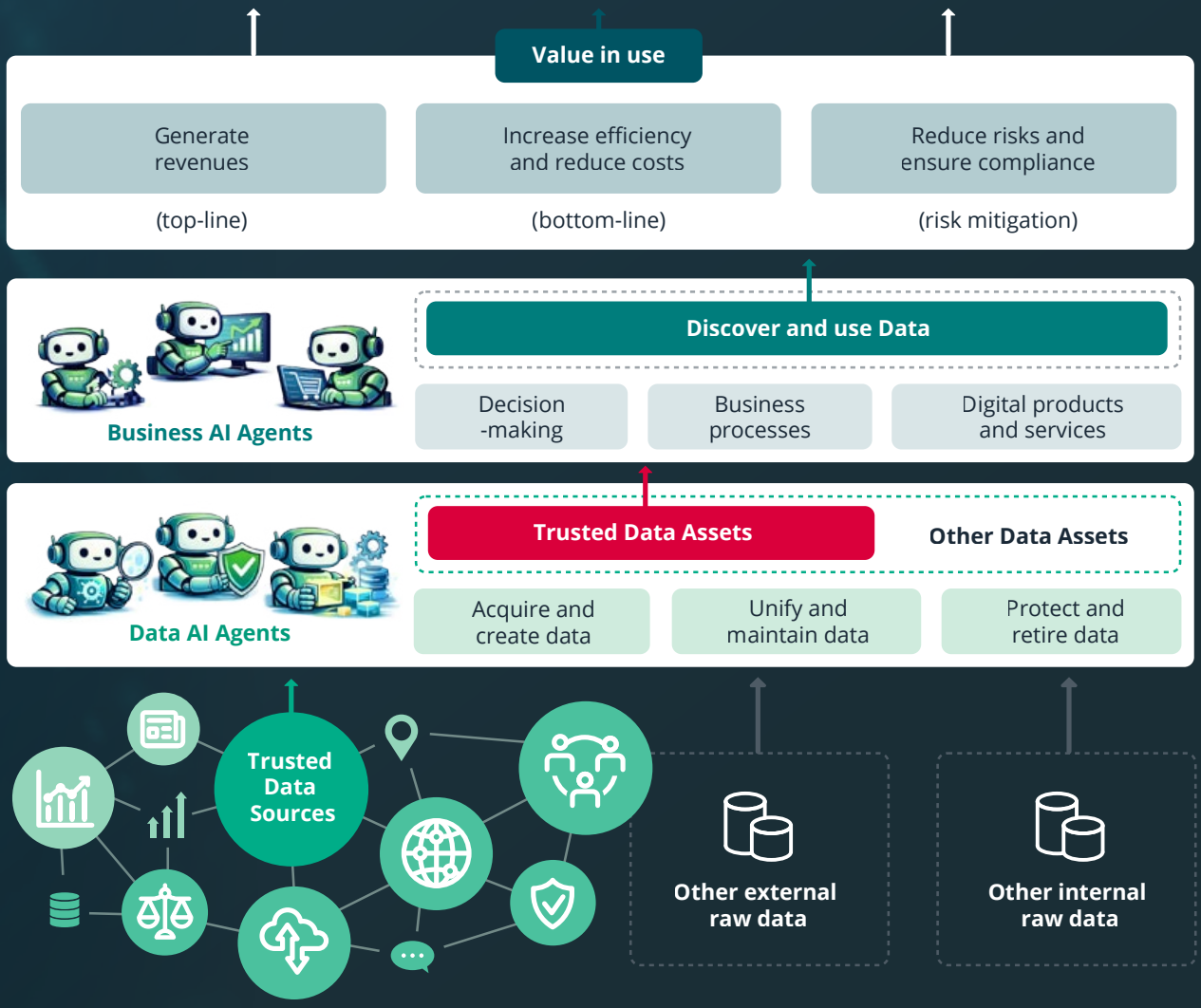
3 Governed scale pays off
Organizations that apply AI in governed data workflows reduce manual work, increase processing speed, and improve consistency across data operations.

AI changes the economics of data management. It shifts large volumes of repetitive maintenance work from manual teams to agents, reducing effort, lowering cost, and increasing speed and scale. At the same time, it raises the price of weak data foundations, because poor inputs, unclear rules, and weak controls scale errors, rework, and risk just as fast as efficiency gains.

This paper makes one core point: AI creates value in data management only when it runs on trusted data, governed rules, and clear accountability. Web-trained LLMs are not a trusted source. They add value as reasoning layers inside a controlled operating model, where authoritative data, deterministic checks, and policy controls determine what is true, what is allowed, and what moves into operations.

A new operating model is taking shape in data management. Agents handle routine execution, people define policy and resolve exceptions, trusted external data strengthens decisions, and shared intelligence improves speed and coverage across company boundaries. This model reduces manual work, improves consistency, sharpens compliance, and strengthens control at scale. It turns AI from hype into measurable business impact, and it gives managers a practical path to modern, closed-loop data management.

The topics in focus



Why AI changes data management now

AI changes the economics of data management. It moves high-volume maintenance work from manual teams to agents, reduces cost, and increases speed and scale. But when data foundations are weak, it scales errors just as fast.

KEY TAKEAWAYS

1 Business pressure keeps rising
Regulatory change, geopolitical volatility, and faster business cycles increase the pressure on companies to keep business partner data accurate, current, and usable at all times.

2 The new operating model is agentic
Companies scale data management by running well-governed agents, not by adding more people across regions and service centers. This increases consistency and lowers marginal cost.

3 Agents take over routine work
Generative AI and agents handle simple, well-defined, repetitive data tasks autonomously. With trusted sources and governed rules, this reduces manual effort and increases scale.

Many managers still treat data management as a housekeeping function. In reality, it now sits on the critical path of compliance, resilience, and AI performance. Regulatory checks happen more often, ownership structures change faster, and business processes depend on data that stays accurate and current at all times. At the same time, management expects lower cost, faster execution, and better control.

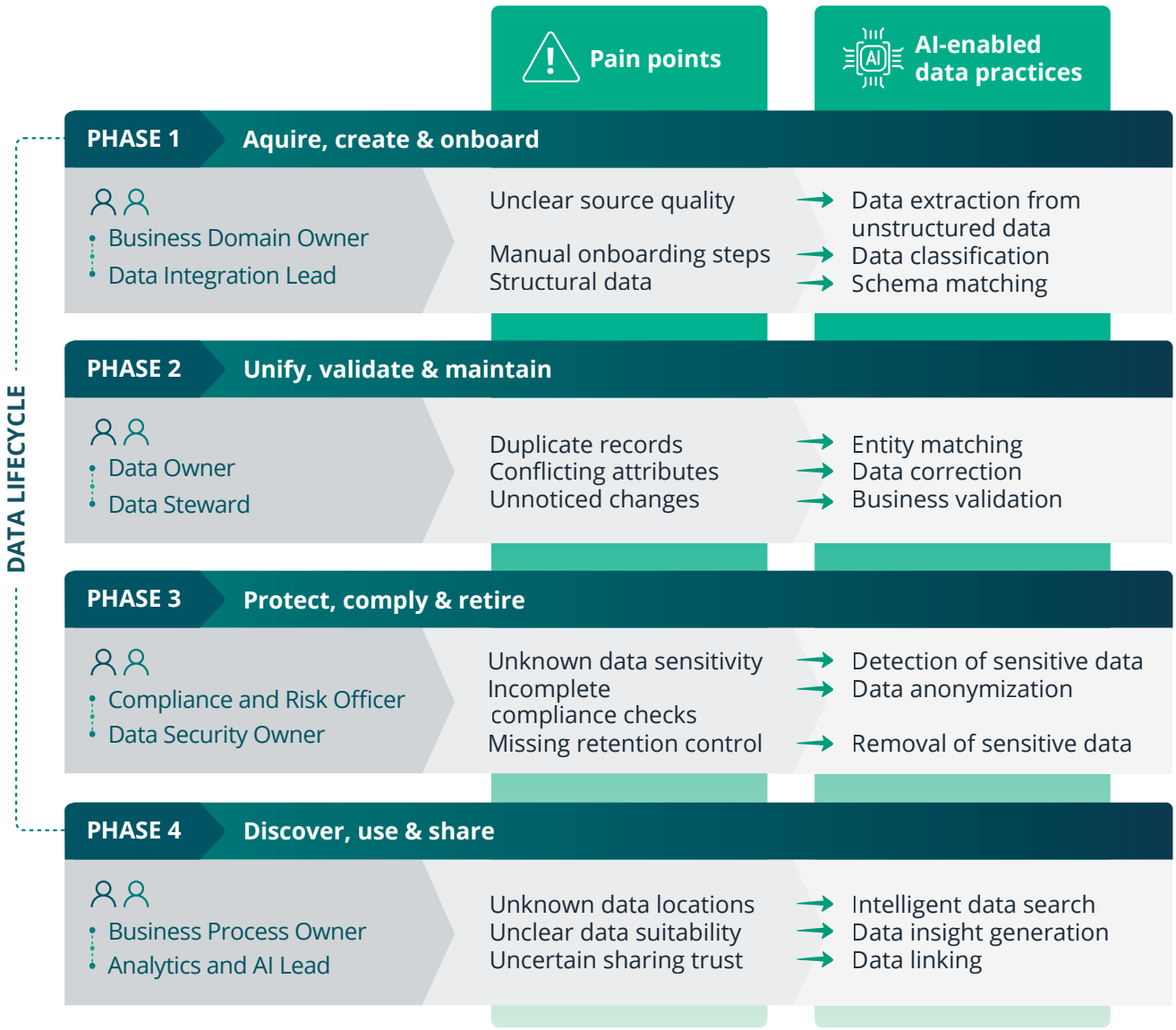
This pressure changes the operating model. Large enterprises still handle many repetitive data tasks through manual teams, regional service centers, and fragmented workflows. That model adds cost and struggles to keep pace with the volume and speed of change. AI shifts this work to agents. Matching, monitoring, consistency checks, and routine maintenance move into governed automation, while people focus on policy, exceptions, and release decisions.

The pattern is already visible in practice. A global enterprise started with a high-friction problem, manual duplicate handling, and introduced AI-supported matching and resolution. Rework dropped, and data experts moved from repetitive corrections to high-risk exceptions and policy decisions. The company then extended the same approach to anomaly detection and consistency checks across additional domains, increasing coverage without proportional headcount growth. In the next step, it established governed, reusable data assets across source systems, which improved decision speed in functions such as finance and operations.

This example shows why isolated fixes do not scale. The limiting factors sit across the whole lifecycle: weak source trust, inconsistent records, fragmented controls, and limited transparency for reuse. Point solutions produce local gains, but they do not change the system. Sustainable AI value appears when organizations manage data as one governed lifecycle with shared standards, clear ownership, trusted external data, and continuous improvement.

Research from the Competence Center Corporate Data Quality (CC CDQ), based on more than 150 use cases, reinforces this point.

External pressure creates urgency, improved AI creates leverage, and governed lifecycle automation turns both into measurable business impact.



The dangerous overestimations of AI

The biggest mistake in AI-driven data management is to expect AI to fix data problems on its own. LLMs are not intelligent by themselves, and they are not an authoritative source. AI delivers strong results when organizations give it trusted inputs, clear objectives, and governed boundaries.

KEY TAKEAWAYS

1 AI is not a trusted source

Web-trained LLMs do not establish truth. They cannot confirm whether a tax number is currently valid for a specific company, whether a legal status has changed, or which source takes precedence.

2 AI does not replace judgment

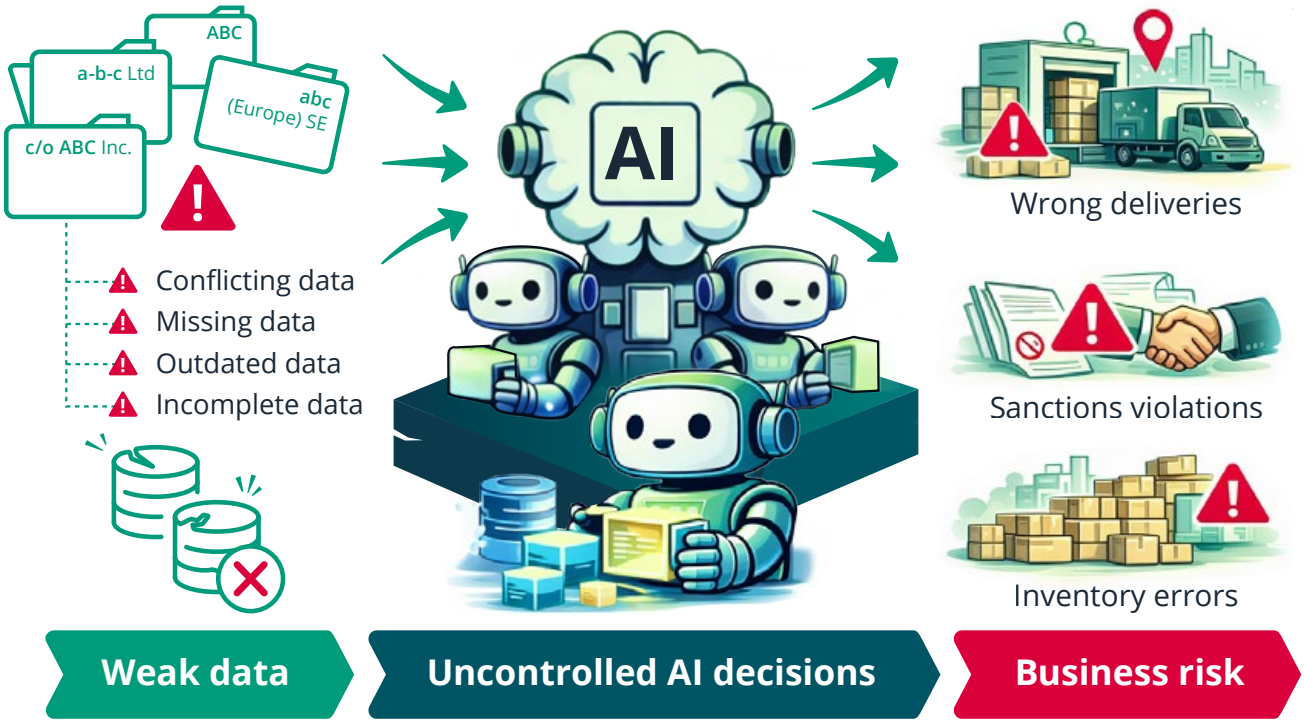
LLMs and agents are strong in pattern recognition, similarity detection, and synthesis. They do not define authoritative truth, ownership, or policy. Organizations must set those rules.

3 AI performs best with a clear plan

Agents deliver the strongest results when objectives, source hierarchies, thresholds, and review rules are explicit. Without that structure, they scale inconsistency instead of reducing it.

AI gets overestimated in data management when organizations apply it beyond its valid scope. LLMs and agents are strong in pattern recognition, similarity detection, and synthesis. They do not define authoritative truth, ownership, or policy. Organizations must provide that structure through trusted sources, clear rules, and governance.

This shows up clearly in day-to-day operations. In procurement, supplier onboarding usually includes formal checks before activation, such as mandatory fields and tax ID validation. The bigger problems appear later, when supplier data changes in external registers but internal records stay behind. Legal name changes, ownership updates, or address changes push contracts, purchase orders, and compliance records out of sync. AI detects and proposes those changes early, but without clear source ranking and release rules, teams still end up with conflicting supplier states and repeated validation work.



A similar pattern appears in sales. Customer creation often starts with basic controls, but the real risk builds over time as billing, shipping, and legal entity data drift across CRM, ERP, and invoicing processes. AI improves completeness and helps reconcile signals, but weak thresholds and unclear stewardship lead to overly aggressive changes in commercially sensitive fields. That creates disputes, credit holds, and invoicing delays instead of better data.

The same issue appears across functions. Procurement, sales, finance, and compliance often work with the same entity through different process views, refresh cycles, and ownership rules. Agents reduce manual effort and connect those signals, but only when governance defines who owns each attribute, which source takes precedence, and where human review is mandatory.

The lesson is simple. AI accelerates data work, but it does not replace judgment or governance. Without trusted sources, clear objectives, and explicit boundaries, it magnifies weaknesses instead of resolving them.

From hype to impact, what actually works in practice

AI creates measurable impact in data management when companies apply it in a governed operating model. Deterministic controls secure correctness, agents handle ambiguity and scale, and governance turns output into reliable business action.

KEY TAKEAWAYS

1 Value starts with the right use cases
AI delivers the strongest results in matching, enrichment, anomaly detection, and evidence synthesis, where data is heterogeneous, patterns matter, and manual effort is high.

2 Hybrid models win in practice
Deterministic checks provide consistency and auditability, while AI handles tasks that require interpretation, comparison, and judgment at scale.

3 Governance turns output into impact
Confidence thresholds, provenance, and approval rules ensure that only justified changes move into operations, reduce manual effort, and create measurable business value.

Reliable AI impact in data management comes from a hybrid operating model. Deterministic controls secure correctness, auditability, and repeatability. AI handles ambiguity, pattern recognition, matching, enrichment, and evidence synthesis at scale. Governance controls what moves into operations and under which conditions.

High-performing organizations apply this model across the data lifecycle. They do not run AI as a side tool or isolated experiment. They embed it into governed workflows with clear decision rights, evidence requirements, thresholds, and escalation logic. That operating discipline turns AI output into business action, reduces manual effort, and improves consistency at scale.

Research from the Competence Center Corporate Data Quality, based on more than 150 AI use cases, confirms this pattern. The most mature and highest-value applications are enrichment, validation, cleaning, and matching. The lesson is clear: deterministic controls secure the baseline, AI scales judgment-intensive work, and governance creates reliable release into operations.

The three cases on the next page show how that model works in practice. They also show that impact does not come from model choice alone. It comes from disciplined operating design, clear boundaries, and integrated execution in real business environments. Read more about these cases in **this blog post**.

CASE A

GenAI for Business Partner Master Data

CDQ Good Practice Award 2025 winner

Kuehne+Nagel presented how generative AI supports the management of millions of business partner records that previously required large-scale, manual, error-prone review. Large language models support duplicate detection, cleanup, matching, and grouping across heterogeneous records.

The solution reaches about 93.5 percent automated match decisions and reduces the effective duplicate rate to 1.7 percent. That delivers multi-million-euro efficiency gains through lower manual effort, faster processing, and more consistent duplicate handling.

On top of that stabilized foundation, up to 90 percent of new business partner records now get created automatically within seconds, under controlled governance. The case shows how governed AI moves from pilot to measurable operational impact.

CASE B

Next Generation Intelligent Tool Suite for Master Data

CDQ Good Practice Award 2025 finalist

Bosch presented an AI-enabled tool suite for master data cleansing and structural analysis across multiple domains, including bills of materials. AI supports pattern recognition, anomaly detection, and consistency checks in areas where manual processes struggle to keep pace.

By embedding generative AI into governed master data processes, Bosch increases processing speed and coverage while maintaining accountability and data quality. The value comes from combining AI scale with clear process control.

The result is a scalable and reusable foundation for intelligent master data management across the enterprise. The case shows that strong outcomes come from disciplined operating design, not from model choice alone.

CASE C

Driving Networking Excellence through Scalable Data Products

CDQ Good Practice Award 2025 finalist

thyssenkrupp Steel Europe presented how a data-as-a-product approach, combined with AI-supported analytics, transforms financial steering and net working capital management. The initiative integrates more than 70 terabytes of finance and operations data from over 30 source systems.

More than 150 source-aligned and 40 target-aligned data products deliver diagnostic, predictive, and prescriptive insights in near real time across more than five departments. This improves visibility across functions and strengthens decision support in daily operations.

The result is stronger transparency, higher data quality, faster decision cycles, and measurable liquidity improvements that support strategic priorities. The case shows that governed data foundations and AI-supported analytics together create business impact beyond core data maintenance.

Data sharing and shared intelligence

No company detects every relevant data change early enough on its own. Shared intelligence closes that gap. It turns validated updates into a collective advantage, reduces duplicate work, and improves speed and confidence across companies.

KEY TAKEAWAYS

1 More eyes, faster signals
Relevant changes across external sources happen too often and too broadly for any one company to catch early and consistently on its own.

2 Shared signals scale better
When companies reuse validated updates under clear rules, they stop checking the same changes again and again. This cuts effort and speeds up action.

3 Trust makes sharing work
Consent, anonymization, source hierarchy, and shared rules turn data sharing into reliable intelligence. Without them, it creates noise.

Even highly automated data management hits structural limits. No company and no agent detects every relevant change early enough across legal registers, sanctions lists, insolvency notices, ownership disclosures, and other external sources. Many organizations monitor the same signals in parallel, invest similar effort, and still react too late to changes that affect onboarding, compliance, and risk exposure.

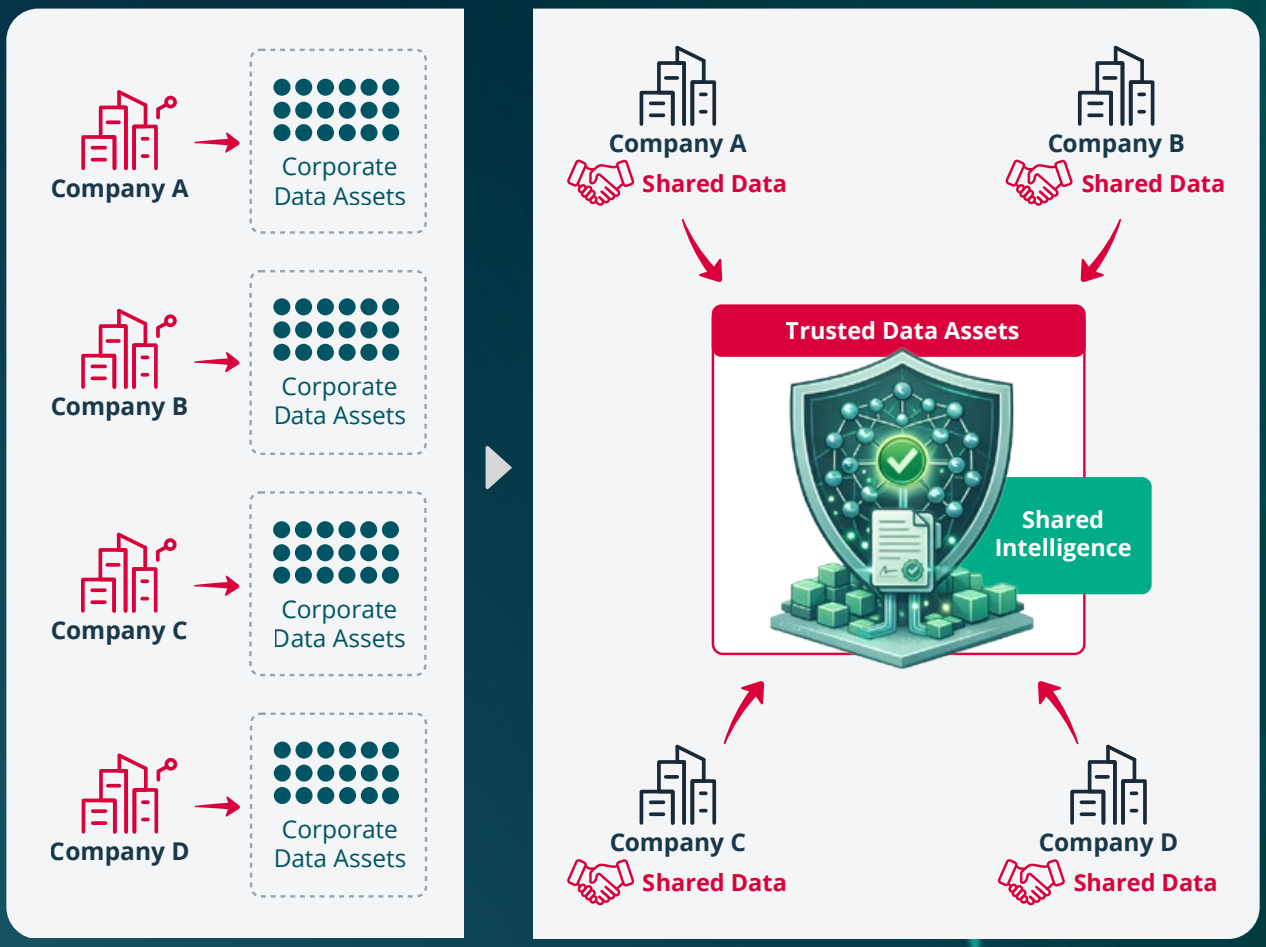
Data sharing changes that equation. When companies contribute validated updates into governed sharing pools, detection shifts from isolated effort to collective intelligence. Instead of hundreds of organizations verifying the same change on their own, one validated observation becomes reusable for many.

What scales is not raw data access, but trusted interpretation under shared rules.

In an AI-driven operating model, this effect becomes even stronger. Agents monitor external signals continuously, but their performance depends on signal quality, confirmation, and context. Shared intelligence pools strengthen all three. Agents contribute validated findings, consume trusted updates from others, and detect relevant changes earlier. That shortens reaction times, reduces duplicate validation work, and improves automation outcomes without increasing local effort.

This model does not replace authoritative sources. It builds on them. Official registers and regulatory lists remain the baseline for correctness. Shared intelligence adds timeliness, cross-company validation, and operational context. Because participants contribute under clear governance and with strong incentives for correctness, these pools develop into trusted layers of intelligence rather than noisy data exchanges.

Governance makes the whole model work. Consent protects control, anonymization protects sensitive data, and source hierarchy ensures that authoritative inputs override weaker signals. Shared semantics and interpretation rules keep participants aligned. Without these mechanisms, data sharing creates noise. With them, it becomes a force multiplier for speed, trust, and scale.



Two futures, compounding advantage or failure

AI is not an advantage in data management per se. It amplifies the operating model underneath it. With trusted data, governed agents, and clear accountability, it compounds quality, speed, and efficiency. Without them, it compounds errors, rework, and loss of trust.

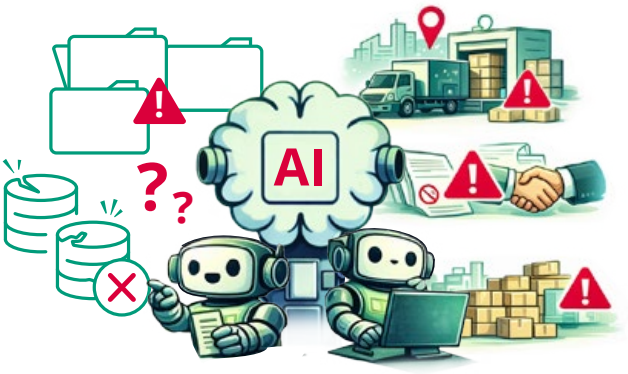
KEY TAKEAWAYS

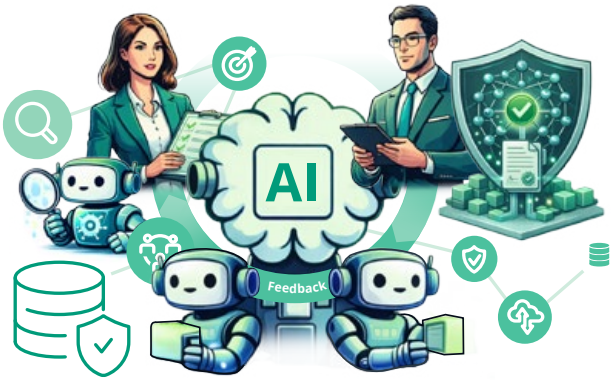
1. **The payoff starts with people**
   AI delivers value only after teams set the rules, establish review points, and assign a clear role to each agent.
2. **Trusted data powers the system**
   Well-organized agents deliver the strongest results when they run on trusted internal and external data, under clear governance.
3. **Discipline creates scale**
   Organizations that invest early in governance, trusted data, and well-scoped agent designs build compounding advantage over time.

AI does not move data management in one fixed direction. It acts as a multiplier. The same technology that reduces manual effort, speeds up execution, and improves consistency also scales weak governance, poor data, and unclear accountability. That is why AI creates a real fork in data management.

One path leads to a quality flywheel. Trusted data, governed agents, and clear operating rules improve speed, cost, control, and resilience over time. The other leads to an error factory. Automation amplifies inconsistency, creates hidden conflicts, and pushes more rework and risk into business processes.

The two scenarios on the next page make this contrast explicit. They show that AI success does not depend on the model alone. It depends on the operating model around it.





Error factory

Automation amplifies weakness and erodes trust

Organizations start to rely on AI too quickly. People lean back, expect automation to take over, and assume the system will sort out inconsistencies on its own. Management reduces attention, manual controls loosen, and unresolved data issues move into automated workflows instead of being fixed at the foundation. At first, this feels like progress. Work moves faster, fewer manual checks happen, and AI appears to absorb the complexity.

The underlying mess does not disappear. It spreads. Automation runs on inconsistent, incomplete, and weakly governed data. Ownership stays unclear, authoritative sources are missing or ignored, and agents operate on conflicting records without dependable thresholds, source precedence, or release rules. To keep processes moving, AI fills gaps with inferred values, implicit conflict resolution, and updates that look plausible but lack sufficient evidence. Silent overwrites become normal. Conflicting states spread across systems. What looked like efficiency turns into hidden operational debt.

The burden then returns to people in a more expensive form. Teams spend their time on exceptions, revalidation, escalations, and downstream failures. Business functions protect themselves with spreadsheets, extra approvals, and local workarounds because they no longer trust the data.

Costs rise again through rework, delays, and process friction. Instead of creating scale, automation weakens the operating model.

Quality flywheel

Automation compounds quality and capacity

Organizations set up AI the right way from the start. They do not expect automation to fix weak foundations on its own. They invest early in trusted data, clear ownership, explicit rules, and well-scoped agent roles. People stay actively involved where it matters most: they define policies, set thresholds, decide which sources to trust, and determine where review remains necessary. That early discipline creates the conditions for AI to operate with speed and precision rather than guesswork.

Agents then run within a governed operating model. They work on high-quality internal and external data, monitor relevant changes continuously, and handle routine tasks such as matching, enrichment, validation, and low-risk updates at scale. They follow clear source hierarchies, apply defined release logic, and escalate unclear or conflicting cases instead of pushing them silently into operations. Every action is grounded in evidence, which keeps automation explainable, auditable, and under control.

This setup changes the economics of data management. Manual effort shifts away from repetitive maintenance and toward policy, exceptions, and continuous improvement. Teams stay leaner, but more focused and more effective.

Trusted data improves agent performance, better agent performance improves data quality, and higher data quality supports faster decisions and more reliable automation. The result is a compounding advantage in cost, speed, quality, compliance, and resilience.

How CDQ operationalizes agent-based data management

CDQ turns agent-based data management into a controlled operating model. Trusted external data, governed agents, and continuous feedback loops let companies scale AI in data management without giving up control, auditability, or accountability.

KEY TAKEAWAYS

1. **Control by design**
   CDQ does not treat agents as open-ended automation. Evidence, rules, and release controls are built in from the start.
2. **Trusted data is the differentiator**
   CDQ combines trusted external data, internal data, and domain rules so agents work on authoritative inputs, not guesswork.
3. **Governed agent autonomy at scale**
   Agents act autonomously within clear boundaries, escalate unclear or conflicting cases, and prepare changes for controlled release into operational systems.

CDQ embeds agents into a controlled operating model, not into open-ended automation. Policies, trusted sources, and human oversight shape how the system runs from the start. The goal is not maximum autonomy. The goal is reliable, scalable outcomes in real data environments.

At the center of this model sits the Data Mirror. Each customer uses it as a staging environment between external signals, AI agents, and operational systems. Agents do not write directly into production data. They collect signals, evaluate evidence, enrich attributes, and prepare structured update proposals inside this controlled space. That separation keeps uncertainty and conflicts out of operational systems until the right checks are complete.

CDQ agents monitor authoritative registers, regulatory lists, commercial sources, and customer system changes on a continuous basis. Instead of waiting for periodic cleanup cycles, the system evaluates signals as they appear. That shortens the time that incorrect or outdated data stays in circulation and shifts data management from reactive correction to proactive readiness.

Agents also operate with explicit domain knowledge. Identifier structures, legal forms, jurisdictional rules, and country-specific logic shape how they interpret signals and prioritize actions. Trusted sources take precedence over weaker inputs, and clear policies govern how conflicts get resolved. Every proposal includes evidence, timestamps, confidence, and context, which keeps decisions explainable and auditable.

People remain in control throughout the process. Agents handle repetitive checks and routine proposals at scale. People define policy, review sensitive cases, and retain accountability for release decisions. High-confidence changes move forward automatically where policy allows. Unclear or conflicting cases move into short review loops, which preserve control without recreating manual bottlenecks.

Throughout this model, CDQ applies the right tool for each task. Deterministic services handle verifiable checks such as identifiers or registry validations. AI handles interpretation, similarity assessment, and synthesis where reasoning adds value. Together, these components create governed agent autonomy at scale.

What managers should do now

AI impact in data management is a leadership execution challenge. Organizations create measurable value when they start with a focused scope, put governance in place early, and scale only what delivers control, quality, and business results.

Pick the battlefield

Start where business value is high, manual effort is visible, and governance matters from day one.

- ✓ Start with a data domain that directly affects critical business processes, such as **supplier** or **customer** data used in **onboarding**, compliance, invoicing, or payment reliability.
- ✓ Set hard guardrails before automation starts, such as **no direct AI write-back** into core systems and **mandatory review** for sensitive attributes.
- ✓ Assign **clear ownership** from the start across business, governance, and technology.

CDQ supports in identifying the highest-value use cases and translating them into a clear, actionable roadmap built on our AI for Data Management framework.

STEP 1

Narrow the scope

Focus on a small number of high-friction processes and prove value fast.

- ✓ Prioritize two or three use cases with high volume, high repetition, and clear pain points, such as **duplicate detection**, **onboarding checks**, or **lifecycle monitoring**.
- ✓ **Limit** the rollout to **one** region, **one** business unit, or **one** process area to keep complexity under control.
- ✓ Establish a clear **baseline** before launch, including **manual touches**, cycle time, error rates, and rework.

CDQ helps in co-creating agentic AI solution approaches and validating their feasibility and value through a proof of concept with measurable outcomes.

STEP 2

STEP 3

Set the guardrails

Translate governance into rules, thresholds, and evidence requirements that agents follow in practice.

- ✓ Define which sources are trusted for which attributes, for example **legal name from the commercial** register and **tax status from the tax authority**.
- ✓ Set clear thresholds for automatic release, **human review**, and rejection.
- ✓ Require every proposed change to carry **full evidence**, including source, timestamp, confidence, and **lineage**.

CDQ offers support in designing and implementing an AI-policy framework that operates within strict deterministic policies ensuring trusted data.

STEP 4

Control the pilots

Run in a controlled staging setup, measure results closely, and tighten the model as you learn.

- ✓ Use a **staging layer** such as a **Data Mirror** before anything reaches operational systems.
- ✓ Combine **deterministic checks** with AI reasoning, so **hard validations** stay hard and AI focuses on interpretation and ambiguity.
- ✓ Track **weekly progress** with a small set of business and control KPIs, such as cycle time, exception rate, **false positives**, and audit completeness.

CDQ supports enterprises in designing KPI-driven multi-step agentic AI workflows with research expertise and agentic software capabilities.

STEP 5

Scale the system

Expand only after proof, then reuse the same control model across domains and teams.

- ✓ Extend into adjacent domains only after the first scope shows stable gains, for example from **supplier onboarding** into **customer lifecycle maintenance**.
- ✓ Reuse the same **governance pattern**, including **trusted sources**, thresholds, approval logic, and escalation paths.
- ✓ Add **shared intelligence** where possible to detect changes earlier, reduce duplicate validation work, and strengthen automation over time.

CDQ enables customers to scale their AI-driven journey safely and sustainably with enterprise-grade software for AI-ready data.

Who we are and why our perspective matters

CDQ brings a distinctive perspective to AI in data management because it combines long-standing AI practice, trusted external data, governed operations, and applied research. This gives companies a proven way to turn AI into modern, closed-loop data management with real business impact.

CDQ as a provider for innovative data services and AI agents

CDQ provides enterprise-grade cloud software for managing and continuously improving business partner data. The focus is on trusted, explainable, and auditable data as the foundation for efficient operations, compliance, and automation. For more than two decades, CDQ has helped organizations build a governed core for legal entities, addresses, identifiers, and other compliance-relevant attributes.

This capability now extends into agent-based data management. CDQ combines trusted external data, internal data, deterministic services, and domain-specific agents in one controlled operating model. Agents monitor relevant sources continuously, validate evidence, detect meaningful changes, and prepare structured update proposals under clear scopes, thresholds, and approval rules. The Data Sharing community strengthens this model further by enabling anonymized reuse of validated updates under clear governance, reducing duplicate effort while improving speed and confidence.

CDQ therefore delivers more than software features. We deliver a modern data management system built for continuous operation, governed automation, and measurable business impact, optionally integrated with customers' agent-based workflows through agent-to-agent interfaces.

Competence Center CDQ and the evolution of data management with AI

The Competence Center CDQ is a long-standing research initiative dedicated to the advancement of data management. For more than twenty years, it has worked with leading companies across industries and with the University of Lausanne to study how organizations manage, govern, and trust data. Its work follows real business problems closely and develops in step with changing enterprise requirements.

This research has shaped important concepts in data management over time, from governance and accountability to quality models, data sharing ecosystems, and now AI-driven and agent-based data management. That long-term perspective matters because it places AI in a broader context. It separates lasting shifts in operating models from short-term technology cycles and gives managers a more grounded basis for decisions.

This research does not stay theoretical. It shapes the concepts, methods, and design principles behind CDQ's software and operating model.

That gives customers more than advanced functionality, it gives them software built on a strong and continuously evolving research foundation.

Build AI on trusted data, not assumptions.

Assess your data foundations, define governance guardrails, and operationalize AI under control.

Take the next step now

Agree or disagree? Get in touch.

Dr. Kai Hüner
Chief Technology Officer
kai.huener@cdq.com

Dr. Richard Lehmann
Head of Community
richard.lehmann@cdq.com

Dr. Tobias Pentek
Head of Strategic Business Development
tobias.pentek@cdq.com

Dr. Simon Schlosser
Head of Product
simon.schlosser@cdq.com

CDQ
SHARING DATA EXCELLENCE

CDQ AG
Lukasstrasse 4
CH-9008 St. Gallen

Follow us on social media:

LinkedIn YouTube

Visit CDQ website:

www.cdq.com