● Whitepaper

The Pharma Data Journey: From Compliance to Innovation in the Age of AI

eraneos

Dr. Jörg Werner
Associate Director Data Management, CureVac Corporate Services GmbH

Dr. Werner has worked for over 30 years in the pharmaceutical industry, specializing in IT, digitalization, and data strategy. He holds a Ph.D. in experimental physics from the University of Wuppertal and conducted research at CERN, where he honed his skills in data science and management. His focus is on maximizing data value through effective data management and governance, fostering enterprise-wide, integrated data environments to support AI readiness and informed decision-making.

Dr. Peter Rehbein
Senior Manager Life Sciences, Eraneos Group

Advisor in digital transformation for the Life Sciences industry with a focus on pharmaceutical and biotech innovation. With a doctorate in Biochemistry and deep industry experience, he supports clients in advancing laboratory automation, data management, and AI integration. His work emphasizes fostering data strategies that are aligned with people, processes, and long-term business value.

Contents

Executive Summary

Pharmaceutical companies are managing more data than ever, but most struggle to extract its full value. While early efforts focused on infrastructure and compliance, the industry is now facing a broader challenge: aligning data strategy with governance beyond traditional safety, compliance and regulation activities, with technology, and most importantly, human capability.

This whitepaper outlines a six-level maturity model describing how pharma organizations evolve from siloed, compliance-oriented data management to integrated, business-led models that drive innovation, agility, and improved patient outcomes. It reveals that technology and governance, while essential, are not sufficient on their own. True return on data investment depends on data literacy, cultural readiness, and effective cross-functional collaboration.

The most advanced organizations treat data as a strategic capability, embedding it in decision-making, enabling self-service access, and fostering a culture of accountability and continuous learning. By aligning people, processes, and platforms, these companies are transforming data into a true competitive advantage.

Success in this new paradigm will not be defined by the adoption of AI or adherence to regulatory requirements alone, but by the ability to empower people with the skills, confidence, and mindset to use data effectively. In the pharmaceutical data journey, culture is the final and most decisive frontier.

Introduction

Pharmaceutical companies have long understood the value of data, leveraging it throughout the pharmaceutical value chain, from Research and Development through clinical operations to manufacturing and supply chain management.

Traditionally, the industry has adopted a technical and compliance-focused data management approach, emphasizing data warehousing, infrastructure, and the establishment of controlled environments to support auditability and traceability.

As data volumes grow and complexity increases, organizations began to recognize that managing data effectively required more than just storing it. This realization led to the rise of strategic data governance-formal structures, ownership models, and processes designed to improve data quality, consistency, and trust.

Data Governance introduced more clarity and discipline; however, it frequently faced challenges in achieving broad adoption, partly due to the perceptions of being overly theoretical and complex, as well as limitations in demonstrating substantial business impact on its own.

In recent years, the focus has shifted once again - this time toward the human side of data. Companies are recognizing that even with strong governance and modern tools, value is limited if employees lack the skills and mindset to work confidently with data. Data literacy and an overarching data culture have emerged as critical enablers of transformation, empowering teams across functions to use data as a strategic asset rather than a technical resource.

The key takeaway is clear: sustainable return on data investment in pharma will not be achieved through technology or governance alone. True impact comes from aligning technical capabilities paired with governance frameworks and human readiness - building a culture where data is understood, trusted, and used to drive better decisions at every level of the organization.

Data as the Lifeblood of Pharma

The pharmaceutical industry occupies a unique space at the intersection of science, healthcare, and regulation. It is one of the most data-intensive sectors in the world -reliant on rigorous research, evidence-based decision-making, and stringent compliance standards. From molecular biology to market access, virtually every stage of the drug development lifecycle depends on data.

Over the past two decades, the volume and variety of pharmaceutical data have grown exponentially [1,2,3].

Clinical trials are increasingly digitized, capturing millions of data points per patient through wearables, electronic case report forms (eCRFs), and remote monitoring tools. Genomic sequencing and molecular profiling have generated massive biological datasets, opening new frontiers in precision medicine. Meanwhile, the broader healthcare ecosystem contributes vast streams of real-world data-from electronic health records (EHRs) to patient-reported outcomes and insurance claims.

Advances in the Internet of Things (IoT) have added even more complexity and opportunity. Smart inhalers, connected injectables, and mobile health apps now offer real-time data on medication adherence, physiological response, and environmental exposure. Each of these data sources, when managed effectively, holds the potential to accelerate development timelines, reduce risk in clinical trials, and deliver more targeted and effective therapies to patients.

The promise is immense. With the right data infrastructure and analytics capabilities, pharmaceutical companies can identify drug candidates faster, optimize trial design, detect adverse events earlier, and improve patient outcomes. Predictive safety models, adaptive trial protocols, and personalized treatment pathways all hinge on the ability to harness high-quality, well-integrated data at scale.

Yet despite this potential, a persistent gap remains between the availability of data and its effective use. Many organizations continue to struggle with siloed systems, fragmented data ownership, disconnected data assets and inconsistent data quality. Business users often lack awareness of and visibility into available data assets, while IT and data science teams face challenges in sourcing, cleaning, and contextualizing the data they need. Regulatory complexity further complicates efforts to standardize and share data across borders and functions.

As a result, much of pharma's data remains underutilized-isolated in siloed systems, difficult to access to key decision-makers, and lacking sufficient context for insights or machine actionable readiness.
In many cases, sophisticated AI models or dashboards are deployed atop weak or incomplete foundations, leading to disappointing results and lost opportunities.

To realize the full value of data, pharmaceutical companies must address not only the technical and regulatory challenges but also the organizational and cultural barriers that stand in the way of effective data use. This requires a shift from treating their data as a byproduct of operations to viewing it as a core strategic company asset - managed, governed, and activated with the same discipline as any other critical business asset.
To support this transformation, it's essential to understand where an organization currently stands in its data journey and what steps are needed to advance. Data maturity of pharmaceutical organizations can be modelled into 6 distinct levels that organizations usually progress on their data journey.

Level 1 Early Data Strategy: Foundations Without Focus

“These initiatives are typically driven by IT departments and compliance teams.”

In the early 2000s, pharmaceutical companies faced mounting regulatory pressures and growing complex operations, leading to increased investments in centralized data infrastructures. Accordingly, the early data strategies were shaped largely by regulatory (establishing rules and laws by governing bodies) and compliance (adhering to these rules and laws) demands and technical feasibility rather than a cohesive, business-aligned vision of utilizing data as a strategic asset. Even today, it is observed that many smaller CROs or CDMOs and even mid-sized pharmaceutical companies are still struggling to move beyond this low level of data maturity, focusing on the bare necessities, such as maintaining data accuracy for audit trails and meeting basic reporting standards for regulatory submissions. In addition, compliance efforts might be limited to merely storing data in formats acceptable for submission, without leveraging it for broader decision-making or insights.

At the core of data maturity level 1 is the enterprise data warehouse. These large, monolithic systems are built to collect, store, and report on data from various operational sources. Their primary goal is to enable traceability and audit readiness-especially in the context of global regulatory bodies such as the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), and local health authorities.

These initiatives are typically driven by IT departments and compliance teams. Focus areas include building validated systems to support GxP activities, meeting 21 CFR Part 11 electronic records requirements, and ensuring that data submitted in regulatory dossiers can be easily retrieved, verified, and defended during audits or inspections.

While these efforts are essential to fulfil external expectations, they are often reactive and narrowly scoped. The resulting architectures are rigid, expensive to maintain, and not designed to support cross-functional collaboration or advanced analytics. Business users often encounter significant barriers to intuitive access to these systems, as data is frequently stored in formats prioritizing compliance over ease of exploration, sharing or reuse.

One of the most significant limitations of these early strategies is **organizational fragmentation**. Data is organized and managed according to departmental or project boundaries, resulting in significant duplication and inconsistency. Research and Development (R&D), for example, often maintains its own internal databases to manage preclinical results, clinical protocols, and trial data. Meanwhile, commercial teams may develop separate platforms to track sales performance, physician engagement, and market access metrics.

These silos make end-to-end data integration demanding to establish. Key questions - such as whether early trial results were predictive of commercial success, or how patient adherence affected long-term outcomes - must often remain unanswered, not due to a lack of data, but because connecting and contextualizing them is difficult, if not impossible.

Case in Point: R&D vs. Commercial Silos

Consider a global pharmaceutical company attempting to assess the market viability of a late-phase oncology drug. The R&D team has collected extensive data on clinical endpoints, adverse events, and biomarker responses. However, that data resided in isolated clinical trial management systems (CTMS) and statistical analysis software, accessible only to specialized teams.

On the Commercial side, teams are building forecasts based on market research, sales analogues, and payer feedback - stored in CRM systems, Excel sheets, and vendor reports. The two groups operate independently, with minimal data sharing or alignment on key definitions (e.g., what constituted a „target patient“ or a „responsive subgroup“).

The result will not just be inefficient, but strategically misaligned: Important opportunities to refine go-to-market strategies using real-world clinical insights are lost. Decision-making will be slower and more fragmented, and confidence in data across groups will remain low.

Metadata Confusion and Ownership Gaps

Another issue that usually emerges in this early phase is the lack of consistent metadata standards representing an important pillar in building FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) data. Different teams start using different naming conventions, taxonomies, and reference data. Without governed sources of truth, questions as simple as „Which version of the dataset is correct, and which is the leading source?“ become routine and difficult to resolve.

Moreover, ownership of a data set is often unclear. Is the IT department responsible for maintaining data quality? Or the business function that produced it? Without clear stewardship roles, accountability suffers, and efforts to improve data quality are ad hoc at best and won’t scale.

Limited Business Alignment

Perhaps the most critical and most common shortcoming of early state data strategies is their detachment from broader business goals. These systems are designed for storage, not insights, and are often seen as solely compliance infrastructure rather than strategic enablers. As a result, investments in data will yield little competitive advantage beyond meeting regulatory thresholds. Compliance alone and fragmented systems will not unlock the potential of data.

Most pharmaceutical companies take the next step on the maturity scale once these realizations reach the business. A concept originally already conceived in the 2010s is then suddenly receiving attention (and funding), one that is suited to move data from the basement of IT systems into the boardroom, requiring new roles, frameworks, and mindsets: Data Governance.

Level 2
The Rise of Data Governance

"This shift marked the rise of formal data governance in the pharmaceutical sector."

As data volumes continued to grow and the regulatory landscape became more complex, pharmaceutical companies reached a critical inflection point. The limitations of early data warehousing strategies - especially poor data quality, fragmented ownership, and inconsistent standards - began to hinder progress. In parallel, new regulations and industry standards such as IDMP, PQ/CMC, CDISC, Allotrope, ISA-S88 forced organizations to take a more structured, enterprise-wide approach to data management. This shift marked the rise of formal **data governance** in the pharmaceutical sector.

Unlike earlier efforts that focused primarily on technical infrastructure, the governance movement emphasized **accountability**, **transparency**, and **standardization**. It was no longer enough to collect and store data; companies now had to demonstrate control over its provenance, context, and use-particularly in regulated domains like clinical research, manufacturing, and pharmacovigilance.

Regulatory Pressure as a Catalyst

Several major regulatory frameworks served as catalysts for the adoption of data governance. Among the most impactful were:

- **GxP requirements**, which emphasized traceability and data integrity across Good Clinical, Manufacturing, and Laboratory Practices.
- **General Data Protection Regulation (GDPR)** in the EU, which introduced strict rules around personal data privacy, consent, and transparency.
- **ISO Identification of Medicinal Products (IDMP) Standards** developed within the International Organization for Standardization, to facilitate the unique identification of medicinal products in the context of pharmacovigilance and the safety of medications as well as the reliable exchange of medicinal product information in a robust and consistent manner between global regulators, manufacturers, suppliers and distributors.

These regulations require companies to know where their data is located, who owns it, how it is being used, and whether it adheres to recognized data standards - all fundamental questions that early data strategies failed to answer. Data governance emerged as a structured approach to address them, ensuring not only accountability and transparency but also consistency and interoperability across systems.

Establishing the Governance Framework

In response, many pharma organizations are launching enterprise data governance programs. These initiatives typically include:

- **Data Governance Councils**, composed of cross-functional leaders tasked with defining policies and prioritizing initiatives.
- **Data Owners**, usually business leaders accountable for key data domains containing such as clinical trial data, product master data, or sales data.
- **Data Stewards**, operational roles focused on maintaining data quality, resolving data issues with the business teams, and enforcing standards and policies.
- **Enterprise Data Glossaries, Taxonomies and Ontologies**, ensuring shared understanding of definitions and context across regions and functions.

Companies also began investing in data governance platforms - software tools that provided lineage tracking, data catalogues, and workflow automation. Examples are Collibra, Informatica Axon, Talend Data Fabric, Alation, or MS Purview. These systems help identify data ownership, redundancies, validate data relationships, and support data sharing with appropriate data access and use.

A key area of focus of data governance is Master Data Management (MDM) - the creation of authoritative, trusted sources for core data entities such as compounds, patients, investigators, healthcare providers, and products. This allows for different systems to „speak the same language“ and reduces the risk of conflicting information.

Governance also brings attention to metadata management and lineage tracking, enabling organizations to understand how data moves through systems, how it is transformed, and whether it can be trusted for decision-making or regulatory submission.

“Without governance, data remains fragmented and unreliable.”

Real Benefits, But Uneven Adoption

The benefits of data governance become evident quickly in high-stakes areas like regulatory submissions and clinical trial reporting. Organizations typically see improvements in:

- **Data quality and consistency**, especially in structured domains.
- **Regulatory compliance**, with greater audit readiness and fewer data-related findings.
- **Operational efficiency**, as standardized data enabled better system integration and reuse.
- **Risk reduction**, through improved visibility into data ownership and use.

Yet, data governance efforts still encounter significant challenges - chief among them perception and adoption.
Beyond regulatory and compliance teams, which are inherently aligned with policies and controls, data governance is frequently viewed as overly bureaucratic, burdensome, and time-consuming. Business teams often view it as a barrier to agility, while data stewards struggle to gain cooperation from colleagues focused on day-to-day operations. The tension between centralized control and local flexibility is a common and well-known friction point.

Moreover, data governance programs are frequently launched as standalone and isolated initiatives or even worse regarded as a project with a defined beginning and end, disconnected from broader business strategies or cultural change efforts. As a result, they risk becoming shelfware-well-documented frameworks that are poorly adopted or inconsistently applied. This is often the reason why pharmaceutical companies make several attempts over the years to implement data governance, but repeatedly fail.

Governance as a Prerequisite for AI Readiness

Companies at maturity level 2 generally recognize that governance is a crucial foundation, but unlocking the full potential of data and AI requires more. This “more” is reflected in an emerging emphasis on data enablement, AI readiness, and ultimately, improving data quality.

> “Without adoption, even the best governance frameworks fall short.”

Level 3 AI, Advanced Analytics, and the Quality Imperative

“These problems are particularly acute when attempting to combine datasets across departments or sources.”

In recent years, artificial intelligence (AI) and machine learning (ML) have become central to the pharmaceutical industry’s vision of a more efficient, predictive, and patient-centred future. [4,5]

From accelerating drug discovery to optimizing clinical trials and unlocking insights from real-world evidence (RWE), advanced analytics has offered an enticing promise: faster decisions, better outcomes, and lower costs.

Pharma’s investment in AI has surged accordingly. Machine learning algorithms are now used to identify drug targets from genomic data [6], predict adverse events during clinical trials, flag at-risk patients in real time, and improve supply chain forecasting. Generative AI is also beginning to impact medical writing, regulatory submissions, and patient engagement.

However, as organizations scale their ambitions, they encounter a harsh reality: AI is only as good as the data that feeds it. Clean, contextualized, well-labelled data - at scale - is not a nice-to-have, but a prerequisite for any reliable machine learning model. And for many pharmaceutical companies, this requirement painfully reveals serious cracks in the data foundation.

AI Highlights Governance Gaps

While level 2 in the data maturity model introduces important governance structures, many of them are still built for compliance and control, not for analytical agility. As data scientists begin training models on historical clinical trial data, EHR records, or real-world datasets, they often discover that:

- **Labels are inconsistent or missing**, making it difficult to train supervised learning models.
- **Metadata is incomplete** and not standardized, preventing accurate interpretation of variables across systems.
- **Lineage is unclear**, leading to questions about data transformations and trustworthiness.
- **Data quality checks are absent**, allowing errors to propagate into model training datasets.

These problems are particularly acute when attempting to combine datasets across departments or sources. For instance, linking R&D trial data with post-market safety data or commercial claims data introduces severe inconsistencies in patient identifiers, adverse event definitions, and dosage formats.

As a result, many early AI initiatives are currently stalling - not because the algorithms fail, but because the underlying data lack integrity and context. Data governance may lay important groundwork, but it may not always adequately support the dynamic needs of data science teams.

When Models Go Wrong: Real-World Consequences

The risks of poor-quality data in AI contexts are not hypothetical. Several organizations already encountered high-profile setbacks that underscore the consequences of weak data foundations [7]:

In one case, a machine learning model was developed to predict patient drop-out in oncology trials. The model showed high performance in internal tests but failed when deployed. A post-mortem revealed that training data was labelled inconsistently across study sites - some classified early withdrawals as adverse events, while others did not. The algorithm had effectively learned to distinguish sites, not patients, leading to spurious predictions.

In another example, a predictive model for treatment response in cardiology was trained on claims data and EHRs. But due to missing values and biased historical prescribing patterns, the model amplified pre-existing disparities, suggesting that patients in certain demographics were less likely to benefit from standard-of-care therapies. The flawed recommendation wasn't just inaccurate - it posed an ethical and clinical risk.

These cases are highlighting a painful insight: sophisticated analytics amplify data problems. Unlike static reports, AI models "learn" from data. If the data is noisy, biased, or mislabelled, those flaws are encoded into predictions and recommendations, often in ways that are difficult to detect.

Quality measures Governance effectiveness, but Quality Isn't Everything

In response to these challenges, many companies are doubling down on data quality efforts. They define KPIs for data quality, improve stewardship, sharpen accountability, introduce data profiling tools, and expand metadata management. These efforts help reduce noise and improve confidence in analytical results.

But a second realization soon emerges: Even high-quality data won't deliver business value if it's not contextualized, understood, and accessible. A pristine dataset stored in a silo is no more useful than a messy one. Quality is necessary - but not sufficient - for effective AI usage. Data must also follow the FAIR principles: It should be findable, accessible, interoperable, and reusable supported by a semantic approach [8] to truly support advanced analytics and AI-driven decision-making.

At data maturity level 2, data scientists often still face challenges interpreting legacy datasets, domain experts still lack visibility into how data is being used, and business teams still aren't equipped to assess model outputs. Without strong collaboration between governance, analytics, and business functions, models often remain in "pilot" status, never fully operationalized.
Ultimately, this level of data maturity clarifies a foundational principle: There is no trustworthy AI without clean, well-governed, and context-rich data. Data quality must be prioritized upstream, not retrofitted at the model-building stage. And governance must evolve to support iterative, cross-functional workflows, not just static definitions and controls.

Organizations that invest in integrated data ecosystems - combining stewardship, lineage, quality controls, and agile governance – will see better results: AI models become more accurate, reproducible, and scalable. Regulatory teams gain confidence in model traceability. Business leaders trust insights generated by analytics teams.

The Data Literacy Imperative

But still, another barrier remains. Even with good data and solid models, many organizations find that business adoption is limited. Users from the business functions struggle to interpret analytics outputs, mistrust recommendations, or don't know how to properly integrate insights into decision-making. The implications are clear: Technology and governance are vital, but people - and their ability to work with data - are the ultimate enablers of success.

Organizations reaching the next level of data maturity therefore bring the focus squarely onto human capability: Data literacy, organizational behaviour, and cultural change.

Level 4 The Human Factor: Data Literacy and Cultural Gaps

“The ability to read, interpret, and act on operational, commercial, or patient-level data remains unevenly distributed across functions.”

As pharmaceutical companies mature their data strategy, implement data governance and maximize data quality for analytics and AI readiness, another critical obstacle comes into focus: People. Even with well-structured data, modern platforms, and robust controls, organizations struggle to convert data into insights - and insights into impact. The root cause? A widespread lack of data literacy and a misaligned or inexistent data culture - critical aspects, that are unfortunately usually completely neglected when companies build and scale their data strategies.

Despite remarkable investment in systems, platforms and compliance processes, many disciplines along the pharmaceutical value chain still lack a comprehensive skill set, mindset, and proficiency in data management that would be expected in a highly regulated, science-driven industry.

Pharmaceutical companies are among the industries with the highest proportion of highly educated employees - including scientists, engineers, specialists, and other highly qualified professionals. But formal scientific training doesn‘t always translate into data fluency in a digital business context. The ability to read, interpret, and act on operational, commercial, or patient-level data remains unevenly distributed across functions. In many cases, this gap has quietly limited the value of even the most advanced data initiatives. For example, several Life Sciences companies made substantial investments in self-service analytics platforms to empower employees and unlock the value of their data assets. However, many of these initiatives failed to deliver a meaningful return on investment, as widespread adoption lagged due to limited data fluency and lack of awareness across the workforce.

In one case, a global specialty pharma company launched a company-wide analytics dashboard to support commercial decision-making, but fewer than 10% of sales and marketing staff used it regularly - many reported difficulties interpreting the data and drawing actionable insights. In another instance, a CDMO deployed an advanced manufacturing analytics suite to optimize batch performance, yet frontline teams continued relying on manual tracking, having received little training on how to read statistical indicators or trust algorithm-driven suggestions.

Introducing Data Literacy

At its core, **data literacy** is the ability to read, work with, analyse, and argue with data. It's not about writing code or building models, but about asking the right questions, interpreting visualizations, recognizing bias, and understanding the implications of data-driven decisions.

In a pharma setting, data literacy this means for example:

- Gathering accurate information regarding the context, quality, condition, and characteristics of the data, including research methods and the data generated and reported.
- Understanding clinical endpoints and patient-level variability in datasets.
- Recognizing how real-world data differs from clinical trial data.
- Being able to validate whether a dashboard tells the full story.
- Identifying when a model output may be misleading or statistically unsound.
- Using post-marketing surveillance data to refine, confirm, or define the safety of product candidates and evaluate their potential commercial effect

Importantly, literacy is not a one-size-fits-all capability. What a regulatory reviewer needs differs from what a sales manager, medical affairs lead, or manufacturing head needs. This underscores the importance of role-based literacy programs, rather than generic data training.

The Cultural Challenge

Beyond skills, data-driven transformation requires a **cultural shift**. As noted in section “Level 1 - Early Data Strategy: Foundations Without Focus”, the pharmaceutical industry has historically operated in functional silos, with tightly guarded data assets and top-down decision-making. This legacy creates several cultural and behavioural barriers:

- Over-reliance on IT or data science teams to “pull the data” or “run the numbers.”
- Lack of ownership of data context, associated quality and interpretation in business units.
- Resistance to change, especially when data challenges longstanding assumptions or workflows.
- Fear of misuse, which sometimes leads to an overly cautious, control-oriented approach to data access.

Many governance models unintentionally reinforce this culture by emphasizing restriction and compliance over enablement and exploration. To address this, leading organizations have begun to expand the role of Chief Data Officers (CDOs) beyond policy and control. Today’s CDOs are increasingly focused on empowerment - bridging the gap between data teams and business users, and cultivating a workforce that can confidently engage with data.

Democratization vs. Control

A key tension lies between data democratization and data control [9]. On one side, democratization means giving more users access to data, tools, and training so they can self-serve insights and make faster, more informed decisions. On the other side, pharma companies must maintain compliance, consistency, and data integrity-especially when working with sensitive clinical or patient-level data.

Striking the right balance requires clear data access policies, robust metadata management, and targeted education programs that reduce misuse while encouraging exploration. Done right, democratization enhances agility without compromising control. Done poorly, it results in errors, duplicated efforts, or confusion about “which number is right.”

Literacy as a Risk Mitigation Strategy

Ironically, one of the most compelling reasons to invest in data literacy is risk reduction. When employees don't understand the data they're using - or avoid using it altogether - the organization is exposed to avoidable risks:

- Misreporting results in regulatory submissions.
- Misinterpreting trial data, leading to flawed portfolio decisions.
- Misusing predictive models, damaging trust with stakeholders.

In contrast, a data literate workforce can challenge assumptions, spot anomalies, and collaborate more effectively with data teams. Data Literacy also enables better governance - not by adding rules, but by fostering understanding.

Literacy as a Cost Reduction Strategy

One of the most overlooked benefits of data literacy is its potential to reduce operational costs [10]. When teams understand the origins, structure, and limitations of the data they handle, they are less likely to duplicate efforts or pursue flawed analyses. Without this literacy, organizations waste time and resources:

- Collecting data that already exists in usable form.
- Failing the dissemination of data-driven insights.
- Reanalyzing datasets without understanding prior methodologies.
- Making decisions based on misunderstood or low-quality data.

Batch failures and product recalls due to poor data quality. By fostering literacy, companies promote smarter data reuse—drawing more value from existing assets while avoiding redundancy. Informed employees know what to trust, what to question, and when to repurpose, leading to leaner, more efficient operations.

Building Literacy and Culture: What Works

Pharmaceutical leaders are increasingly recognizing that building a data-driven culture is a long-term investment. Some of the most effective tactics include:

- Internal Data Academies: Structured, role-specific training programs that teach data fundamentals, business-relevant analytics, and ethical use.
- Champions and Ambassadors: Engaged employees who act as translators between data teams and the business, helping peers understand and adopt data tools.
- Embedded Data Coaches: Assigning analysts or stewards to departments to support day-to-day data work and mentoring.
- Gamification and Recognition: Celebrating data-driven wins, sharing success stories, and rewarding critical thinking with data.

Perhaps most importantly, leaders must model the behaviour they want to see ("Data Leadership") - asking for data, questioning assumptions, and being transparent about decision-making. Without visible leadership support, cultural change efforts often stall.

In conclusion, to maximize returns on their data investments, pharmaceutical companies need to prioritize capability and culture alongside data governance, systems and standards. Achieving an integrated, adapting operating model in the next phase of the evolution requires strategic alignment of governance, technology, and people - because in the end, data doesn't drive change. People do.

Level 5 A New Data Operating Model: Strategy, Governance, Literacy, Culture

As the pharmaceutical industry moves deeper into the digital era, it's clear that success in data-driven transformation requires more than isolated initiatives or tool deployments. Organizations that consistently extract value from their data - whether through operational excellence, innovation, or AI-driven insights - do so because they have built and deployed an **integrated data operating model**. This model harmonizes strategy, governance, technology, literacy, and culture into a coherent system.

The final stage of data maturity in pharma is thus not driven by any single discipline. Instead, it will come from how these disciplines interact - how they are embedded across teams, aligned to business value, and sustained by behaviour and leadership. The shift is away from central command-and-control functions toward **networked, cross-functional capabilities** with clearly defined roles, accountability, and feedback loops.

The Five Pillars of the Modern Data Operating Model

1. Strategy - Business-Aligned, Outcome-Driven

At the core of any sustainable data model is a **clear and business-aligned strategy**. In the past, data strategies were often written in technical language, disconnected from operational priorities. Modern strategies articulate how data will support growth, improve speed to market, reduce risk, or enhance patient engagement.

This means:

- Prioritizing data investments based on business outcomes.
- Co-developing use cases with business leaders and identifying appropriate data assets contributing to it.
- Measuring impact not just in terms of data quality, but in **decision quality** and **time to value**, and data (re)usage. Business is interested in decisions, not in data.

In successful pharma organizations, data strategy is treated as a core business strategy, owned and sponsored by the executive team - not just the CIO or CDO.

2. Governance - Enabling Through Ownership and Policy

Data governance remains foundational - but its tone and scope have evolved. Modern data governance is about enablement, not enforcement. It provides clarity around:

- **Ownership** (who is accountable for data domains),
- **Stewardship** (who maintains and improves the data life cycle incl. modelling, access, sharing, documentation),
- **Policy** (how data is classified, shared, protected, and (re)used).

Modern governance models often adopt **computational** (the governance function is invisible for data providers and consumers) **federated** (decision-making model led by the data domain owners) **approaches**, where accountabilities are distributed across business units with central coordination. Rather than bottlenecking access, governance enables safe **self-service**, balancing agility with control.

Leading organizations embed governance into daily workflows, not just quarterly reviews. Data quality monitoring, lineage tracking, and metadata management are increasingly automated and integrated into analytics pipelines, ensuring relevance and adoption.

3. Technology - Scalable, Secure, and Interoperable

Technology is a crucial enabler, but it must be aligned to strategy, not drive it. The modern data stack in pharma is evolving rapidly, moving from monolithic platforms to **modular, interoperable ecosystems** that support real-time analytics, AI, and data sharing.

Key characteristics include:

- **Cloud-native architectures** that scale flexibly and support hybrid data environments.
- **Interoperability** across clinical, manufacturing, commercial, and external data sources (e.g., EHRs, RWE).
- **Security and privacy by design**, especially for handling sensitive patient data in compliance with GDPR, HIPAA, and other frameworks.

Tech teams increasingly work in partnership with business and data governance to ensure that platforms are not only functional, but **usable** by those who need insights most.

“Changing culture requires deliberate effort.”

4. Literacy - Empowering the Workforce

As highlighted in the previous chapter, data literacy is no longer optional - it is essential. The new data operating model treats literacy as a strategic capability that must be actively built and sustained.

This involves:

- Offering tailored training to different roles existing along the pharma value chain (e.g., from researcher, clinical leads to commercial analysts).
- Creating internal data academies, workshops, and e-learning paths.
- Appointing data coaches, mentors and analytics translators within teams to level-up peers.
- Promoting experiential learning-giving employees the tools and support to explore real data, ask questions, and apply insights.

Data literacy is also linked to risk mitigation and cost reduction - reducing the likelihood of misinterpretation or inappropriate model use-and fostering innovation, as employees gain the confidence to test new hypotheses and workflows.

5. Culture - Seeing Data as a Shared Asset

Culture is the invisible glue that connects the other pillars. A high-performing data culture is one where:

- Data is seen as a shared asset owned by the company, not a departmental resource.
- Employees feel safe to ask questions, challenge assumptions, and make data-informed decisions.
- Leadership models data-driven behaviour, celebrating evidence over intuition.
- Accountability is shared, and data quality is inherently linked to specific data domains, as its assessment may vary depending on the perspective from which it is evaluated.

Changing culture requires deliberate effort. Pharma organizations are making progress by appointing data culture ambassadors, launching change management programs, and linking data use to performance incentives and recognition.

Best Practice: Cross-Functional Data Teams and Value Streams

Traditional organizational structures, where IT manages systems and business teams request reports may no longer suffice in a world of fast-moving insights and agile experimentation.

Modern pharma companies are embracing cross-functional data teams organized around value streams: therapy areas, trial design, patient engagement, or market launch. These teams bring together data engineers, stewards, analysts, clinicians, and commercial leads to co-own outcomes.

Value stream approaches allow for:

- Tighter alignment to business priorities.
- Faster iteration and feedback.
- Shared accountability for quality and impact.

Rather than building data capabilities in silos, these organizations embed them into end-to-end workflows, where data creation, transformation, and consumption happen within a coordinated system.

Level 6 Continuous Improvement and Feedback Loops

The most mature data organizations will treat their operating model as a living system - one that evolves through feedback and learning. They apply continuous improvement principles to data:

- Monitoring **data usage and sharing patterns** and iterating tools and governance accordingly.
- Measuring **business impact** of data initiatives, not just compliance KPIs.
- Creating **closed-loop feedback systems**, where users can report data issues, request enhancements, or suggest improvements.

This mindset fosters agility, resilience, and long-term relevance.

Real-World Examples: Federated Ownership and Data Ambassadors

In one global pharmaceutical company, data domains are managed through a **federated ownership model**, where each domain (e.g., clinical trial data, product data, HCP data) has a clearly defined business owner and dedicated stewards. These domain teams meet monthly with the central data office to align standards and share best practices.

Another firm has deployed a network of **data culture ambassadors** - employees embedded in local markets and departments who advocate for data use, mentor peers, and surface adoption challenges. These ambassadors have proven instrumental in driving cultural change and bridging the gap between strategy and execution.

The future of data in pharma will not be defined by any one capability, but by how capabilities come together. The new operating model is **holistic, human-centered**, and **business-led**. It treats data as a dynamic asset, to be governed, shared, understood, and acted upon by everyone.

Pharma companies that embrace this model will be best positioned to deliver on the promise of data-improving outcomes for patients, accelerating innovation, and driving sustainable growth.

“These ambassadors have proven instrumental in driving cultural change and bridging the gap between strategy and execution.”

Practical Recommendations and Roadmap for Pharma Companies

The evolution of data management in the pharmaceutical industry - from compliance-focused warehousing to AI-driven insight and cultural transformation - demands a new approach to execution. For companies seeking to realize sustained return on their data investments, the path forward requires intentional design, clear prioritization, and the ability to adapt over time.

This chapter provides a practical set of recommendations and a roadmap to guide pharma organizations as they align data strategy, governance, technology, literacy, and culture. While no two companies will follow the exact same path, the core principles outlined here are broadly applicable across the industry.

"This diagnostic phase should inform prioritization and sequencing of investments."

1. Assess Maturity Holistically

Before launching new initiatives or technologies, organizations should begin with a **comprehensive assessment** of their current data and AI maturity. Too often, efforts are scoped narrowly-focused only on infrastructure or governance, without understanding broader organizational readiness. The assessment should cover:

- **Governance**: Are ownership, policies, stewardship, and lineage well-defined and adopted?
- **Technology**: Are data platforms scalable, interoperable, and accessible across functions?
- **Data Literacy**: Do employees understand and trust data? Can they analyze and act on it?
- **Culture**: Is data used for decision-making across the enterprise? Are behaviours aligned?

Tools such as maturity models, stakeholder interviews, and data usage audits can reveal critical gaps (such as underused tools, unclear stewardship, or uneven literacy) that undermine the broader data strategy. This diagnostic phase should inform prioritization and sequencing of investments.

2. Build Literacy Frameworks Tailored to Roles

Improving data literacy requires more than ad hoc training. Pharma companies should develop structured **data literacy frameworks** aligned to job roles and business domains. These frameworks define the specific skills and knowledge required for different audiences, such as:

- **Clinical Operations Leads**: Understanding trial data structures, endpoints, and statistical outputs.
- **Medical Affairs Teams**: Interpreting RWE, model outputs, and dashboards used in HCP engagement.
- **Regulatory Affairs**: Assessing data provenance (who, where), traceability (how, where), and submission-ready formats.
- **Commercial Analysts**: Analysing multi-channel data, forecasts, and AI-powered segmentation.

This role-based approach ensures that literacy efforts are relevant and actionable-helping employees gain confidence and reduce errors. Learning pathways should include a mix of self-paced content, live sessions, and hands-on use cases, ideally embedded within day-to-day workflows.

Organizations should also create roles for **data coaches, mentors or translators** who support colleagues in applying these skills, particularly during onboarding, major system changes, or analytics initiatives.

3. Treat Change Management as a Core Capability

Deploying a new analytics platform or governance policy is not enough. Pharma organizations must invest in change management to ensure adoption, engagement, and long-term behaviour change.

Key components include:

- **Stakeholder mapping and engagement**: Identify champions and resistors early.
- **Communication plans**: Tailor messaging by audience, using clear examples of value and impact.
- **Training and enablement**: Align timing and content to when people need it-not just during rollouts.
- **Reinforcement mechanisms**: Celebrate data-driven wins, create peer-to-peer learning, and link data goals to performance reviews.
- **Measurement of maturity and progress and its communication of key components**: Change management should be treated as a core capability - resourced and funded alongside technology and process initiatives. Without it, even the best-designed solutions can fail to scale.

4. Redefine Success Metrics Beyond Compliance

Historically, pharma organizations have measured data initiatives by compliance indicators: audit readiness, system validation, or policy adoption. While important, these metrics are insufficient to track progress toward a modern, value-driven data organization.

Success metrics should expand to include:

- **Business impact**: Time saved in trial design, improved forecast accuracy, reduced cost of poor data.
- **User adoption and confidence**: Are tools being used? Do users trust the outputs?
- **Data quality and reuse**: Are data assets being shared across silos? Is duplication decreasing?
- **Speed to insight**: How quickly can teams generate and act on evidence? Are data assets ready to be shared and can be easily “shopped”?

Companies should regularly collect feedback from data consumers, not just data producers. Pulse surveys, user analytics, and stakeholder interviews can help quantify qualitative shifts in data culture and behaviour.

5. Start Small with High-Impact, Cross-Functional Pilots

To build momentum and demonstrate value, companies should avoid trying to "boil the ocean." Instead, they should launch pilot programs in focused, high-impact areas-ideally where data is already critical and cross-functional collaboration is essential.

Examples include:

- **Clinical Operations**: Improving protocol feasibility using predictive analytics.
- **Supply Chain**: Using real-time data to optimize inventory and reduce waste.
- **Medical Affairs**: Enhancing field communication with unified, trusted data sources.
- **Regulatory Affairs**: Using standardized (e.g. IDMP, SPOR), semantically interoperable frameworks, improving data integration, automation, and compliance

These pilots should include business leads, data stewards, analysts, and IT partners. Co-creating use cases, testing solutions, and refining governance in these environments creates repeatable patterns that can be scaled across the enterprise.

Crucially, any pilots should be evaluated not just for technical feasibility but for **business value, adoption, and sustainability**. Lessons learned should feed back into broader strategy and roadmap development.

Conclusion: Building for Sustainability

Pharmaceutical companies stand at a pivotal moment: Equipped with more data than ever before, yet often held back by fragmentation, misalignment, and underutilization. The roadmap forward requires more than tools or policies - it requires orchestration across disciplines, sustained investment in people, and a willingness to challenge long-standing assumptions about ownership, access, and control.

Those who succeed will treat data not as an asset to manage, but as a capability to grow - one that is woven into the fabric of how they discover, develop, and deliver healthcare innovation.

The Next Frontier is Human

The pharmaceutical industry has made remarkable strides in evolving its data capabilities - from basic data warehousing to complex governance frameworks and cutting-edge AI applications. Yet, as this whitepaper has explored, the greatest challenge and opportunity now lies beyond infrastructure and technology: It lies in the people, working together on building a sustainable, future-proof operating model.

Pharma companies must recognize that data strategies focused primarily on compliance, technology or governance are necessary but not sufficient. Without empowering the workforce - the scientists, clinicians, commercial teams, and analysts who create, interpret, and act on data - ROI will remain elusive. This human dimension is the next frontier in data management.

From Infrastructure to Empowerment

Historically, investments in pharma data have prioritized systems, repositories, and compliance frameworks. These foundational elements were essential to meet compliance and regulatory demands and establish data provenance and traceability. However, focusing narrowly on infrastructure has often left users disconnected or ill-equipped to fully leverage the data assets at their disposal.

True value emerges when employees at all levels can confidently access, interpret, and use data asset to make better decisions. Data literacy - the ability to read, analyse, and communicate with data - is no longer a nice-to-have skill reserved for data science teams but a core competency of all data citizens across the enterprise.

Pharma companies must therefore embed continuous education, role-based training, and cultural reinforcement into their data strategies. This empowers individuals to be active participants in data governance and innovation, reducing risk and accelerating insights.

Culture: The Catalyst for Sustainable ROI

Data technology and governance set the stage, but culture drives performance. A culture that views data as a shared asset, where curiosity, transparency, and evidence-based decision-making are rewarded, creates fertile ground for transformation.

Such a culture nurtures:

- Trust in data and analytics outputs.
- Collaboration across traditionally siloed functions.
- Ownership of data quality and (re)usage.
- Agility to adapt as new tools and insights emerge.

Pharma leaders must champion this cultural shift by modelling data-driven behaviours, recognizing data champions, and creating safe spaces for experimentation and learning.

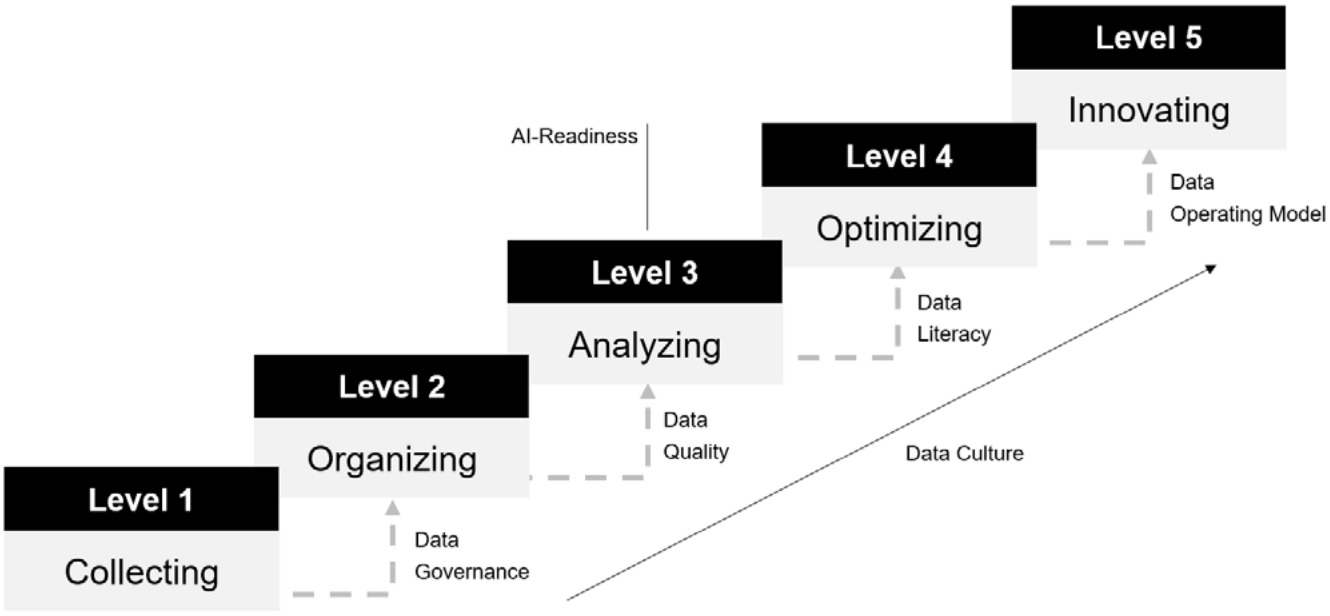
Looking Ahead

As the pharmaceutical industry continues to navigate complex healthcare challenges, from personalized medicine to real-world evidence and beyond, the capacity to harness data effectively will be a critical differentiator.

The companies that succeed will not only have advanced technology and rigorous governance but will have cultivated a data-literate workforce and a culture that embraces data as a strategic asset.

In this exciting next phase, the human factor will be the true competitive advantage. Empowering people to unlock the potential of data is where pharma's future lies - not just in becoming data-driven, but in becoming data culture-enabled.

Appendix



Life Sciences Data Maturity Journey

Due to a multitude of factors, organizations typically progress through a defined maturity journey on their path to becoming data-driven.

References

[1]
https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion

[2]
https://healthtechmagazine.net/article/2023/05/structured-vs-unstructured-data-in-healthcare-perfcon

[3]
https://doi.org/10.1016/j.drudis.2024.104009

[4]
https://www.weforum.org/stories/2025/01/2025-can-be-a-pivotal-year-of-progress-for-pharma/

[5]
https://web-assets.bcg.com/pdf-src/prod-live/biopharma-path-to-value-with-generative-ai.pdf

[6]
https://doi.org/10.1017/pcm.2024.4

[7]
https://www.wsj.com/articles/johnson-johnson-pivots-its-ai-strategy-a9d0631f

[8]
https://doi.org/10.1186/s13326-025-00327-4

[9]
Davenport, T. H., & Bean, R. (2017). What's your data strategy? Harvard Business Review, defense versus offense data activities

[10]
https://www.bioprocessintl.com/information-technology/the-paradox-of-data-overabundance-in-biomanufacturing-data-literacy-is-key-to-unlocking-value

About Eraneos

Eraneos is an international strategy, transformation and technology consulting group, dedicated to empowering organizations to thrive in an ever-changing digital age.

By bringing together top-tier experts from business and tech, we help clients to continuously raise the bar in successful transformations, from strategy to execution. Whether we're designing future-ready organizations, unlocking the potential of data and AI, or securing businesses with cutting-edge cyber-security, we deliver results pragmatically. Our team of around 1,200 dedicated professionals is based in offices across Switzerland, Germany, the Netherlands, Spain, Denmark, Sweden, Austria, the UK, Singapore, and the USA, and seam-lessly blends global perspectives with strong local roots. In 2024, Eraneos realized a turnover of EUR 263 million.

Get in touch

Eraneos Switzerland AG
Andreasstrasse 11
8050 Zurich
Switzerland

+ 41 58 411 95 00
info.ch@eraneos.com