A BASE LIFE SCIENCE WHITE PAPER

Preparing Life Sciences R&D Platforms for Scalable AI

A Framework for Building AI-Ready Digital Foundations

March 2026

www.baselifescience.com

Executive Summary

Across Life Sciences, Generative AI is transitioning from experimental pilot to a core strategic capability[1]. Yet most organizations struggle to move beyond experimentation.[2] The primary constraint is not model performance but the readiness of the underlying digital foundation. For example, years of rapid Veeva deployments have often produced configuration and data complexity that, while functional for human users, are not optimized for machine readability, cross-domains orchestration, or AI-driven workflows.

At the same time, the pharmaceutical industry is entering a period of intensifying pressure to improve R&D productivity. McKinsey Global Institute estimates that generative AI could create $60-110 billion in annual value across the pharmaceutical sector, driven largely by improvements in R&D productivity, clinical development, and regulatory processes.[3] The organizations that capture this value will not necessarily be those with the most advanced models, but those that have built the most robust and well-governed digital foundations.

Achieving this requires a shift in perspective. To enable AI at scale, organizations must align across the core dimensions of platform maturity. Weaknesses in any one of these areas can undermine the reliability, scalability, and regulatory acceptability of AI-driven workflows.

This paper explores why platform complexity has become a critical barrier to AI adoption in Life Sciences. BASE life science R&D Value Realization Assessment provides an AI-accelerated diagnostic and a prioritized roadmap to help organizations prepare their Veeva R&D platforms for the next wave of AI-driven innovation.

The AI Imperative in Life Sciences R&D

The pharmaceutical industry is facing a convergence of structural pressures that are making AI adoption a strategic imperative. Development timelines exceed a decade on average[4]; R&D costs per asset surpass $2 billion, and fewer than 13 percent of Phase 1 assets reach market. At the same time, the industry faces what analysts term a "patent cliff" $236 billion in revenue at risk through 2030 from expiring exclusivities.[5]

Against this backdrop, AI is emerging as a primary pathway to fundamentally alter R&D economics. Research suggests that large biopharma companies could gain $5-7 billion over five years by scaling AI, with R&D offering the largest value opportunity at 30-45 percent of total potential.[6] McKinsey's analysis of next-generation pharma technology stacks emphasizes that modern, modular architectures built on FAIR data principles are the foundational enabler of this transformation.[7]

A Rapidly Maturing Landscape

The AI in pharma landscape has matured significantly over recent years. In 2024, industry analysts estimate that more than 150 drug candidates discovered using AI technologies have now entered clinical development globally, demonstrating the increasing maturity of AI-enabled discovery platforms. At the same time, partnerships between pharmaceutical companies and AI technology firms have expanded rapidly, reflecting growing industry confidence in AI as a driver of future R&D productivity.

The KPMG 2025 CEO Outlook survey found that 71 percent of life sciences executives now prioritize AI investments.[8] Vendor platforms are increasingly embedding automation, orchestration capabilities, and programmatic data access into their enterprise systems. The most future-ready pharmaceutical companies combine strong R&D capabilities with digital platforms and advanced data infrastructure to accelerate innovation and adapt to emerging technologies.[9]

These developments highlight an important shift: modern R&D platforms are evolving from document repositories into programmable data environments across medicine's lifecycle.[10] However, the ability to leverage these capabilities depends heavily on the structural integrity of the configurations, metadata and semantic models, and governance processes implemented within each organization's environment.

The AI Paradox: Why Most Initiatives Stall

Despite this growing momentum, many organizations struggle to move beyond isolated pilots. They approach AI as a technology layer that can simply be added on top of existing systems. The implicit assumption is that current platforms, having served the organization well for regulatory submissions, trial management, and quality processes, are adequate foundations for AI-driven workflows. This assumption is the root cause of the AI paradox in life sciences.

AI performance depends on the quality of the underlying digital foundation. In regulated R&D platforms, technical debt* rarely arises from a single source. It typically emerges from the interaction of fragmented metadata models, inconsistent governance practices, evolving configuration layers, and workflows designed for manual execution rather than scalable digital processes. Unlike humans, who can apply critical judgment to interpret incomplete or inconsistent information, AI systems depend heavily on structured, high-quality inputs to generate reliable outcomes. When these inconsistencies exist, the inconsistencies propagate through models and analytics pipelines, leading to unreliable outputs or failed automation initiatives.

McKinsey identifies several root causes for slower AI adoption in pharmaceutical organizations compared with other industries. These include organizations clinging to legacy processes that undermine standardization, neglecting change management, implementing technology without clear business benefits, conducting transformations in departmental silos, and relying on inflexible systems plagued by low-quality, siloed data[6] Each of these failure modes maps directly to the technical debt accumulated in Veeva environments.

Understanding this paradox requires looking beyond individual AI initiatives and examining the structural foundations of the platforms on which they depend.

---

* Technical debt refers to the accumulation of structural issues across platform configuration, data models, operational processes, and governance practices that have evolved over time.

Understanding Platform Complexity Beyond IT

Platform complexity in Life Sciences is systemic, spanning four tightly coupled dimensions: technology, data, process, and people. These dimensions must be assessed together rather than in isolation. Addressing one dimension in isolation, such as standardizing object models without aligning the processes that generate and manage data, typically delivers limited value and can create new issues elsewhere.

Weak platform design can create poor data practices, inconsistent data can drive process workarounds, and unclear ownership allows these issues to accumulate over time. For that reason, root causes need to be assessed rather than isolated symptoms, especially as AI-driven systems introduce new forms of debt that compound when managed in isolation.

While these structural issues are often viewed as technical concerns, their consequences extend far beyond IT architecture. They directly influence the economic performance of R&D organizations.

Table 1: The four dimensions of technical debt in Life Sciences Veeva environments

Dimension	Example in Veeva	AI Impact
Technology	Over-customization of objects, lifecycles with 15+ states, heavily overridden security/DAC models, excessive custom Java SDK code	Makes platform structures harder to interpret, maintain, and integrate, limiting the ability of AI solutions to reliably access and use data across domains.
Data	Poor metadata quality, inconsistent taxonomies, non-adherence to FAIR principles, unstructured content without classification, gaps in data accessibility.	Leads to unreliable outputs, hidden patterns, and AI models that reproduce existing data quality issues or bias.
Process	Manual workarounds, high variability across areas, undocumented tribal knowledge, SOP-to-configuration misalignment	Prevents AI from automating end-to-end workflows; forces AI into a reporting role rather than an active participant
People & Governance	Weak change control, unclear data ownership, no federated governance model, reactive compliance posture, and too much time spent handling changes.	Complicates AI validation and audit readiness; creates regulatory risk when AI-driven decisions lack traceable data provenance

The AI Performance Ceiling

There is a common misconception that a "smarter" AI model can overcome "messy" data. In regulated R&D environments, there is a structural performance ceiling for AI. When data models are fragmented, metadata is inconsistent, processes vary across teams, or governance is weak, AI performance inevitably reaches a plateau. Even highly sophisticated models struggle to compensate for structural inconsistencies in the data, workflows, and controls that underpin regulated R&D platforms.

This ceiling manifests in several ways. Fragmented data structures and inconsistent governance limit the reliability of AI outputs. For example, AI-assisted dossier assembly relies on consistent document metadata and lifecycle states; when classifications vary across Vault environments, automated pipelines cannot reliably assemble submission packages. Similarly, predictive safety analytics may miss correlations when safety, clinical, and quality data remain siloed. In each case, the limiting factor is not model sophistication but the structure and consistency of the platform environment.

Deloitte's analysis of data quality in pharmaceutical R&D confirms this dynamic: high-quality FAIR data is critical for reliable AI, enabling higher-performing models, faster discovery cycles, improved reproducibility, and greater regulatory confidence. Conversely, poor data quality affects the entire AI lifecycle from training and retrieval to inference and validation, undermining every downstream application.[11]

Raising this performance ceiling requires strengthening the R&D platform environment: standardizing data models, harmonizing domain processes, improving metadata governance, and simplifying platform configurations. This enables advanced use cases: automated regulatory dossier generation, predictive safety signaling, AI-assisted quality event investigation, and cross-domain analytics that connect clinical, safety, regulatory, and quality data into a unified intelligence layer. For that reason, organizations need a practical way to assess readiness before scaling AI initiatives.

The Economic Case: The Cost of Inaction

For business leaders, system complexity creates hidden liabilities across the R&D platform. Every poorly optimized component of the digital R&D landscape acts as a compounding tax on innovation. As platforms evolve and organizations invest in analytics and AI, these structural inefficiencies increase the cost and complexity of transformation initiatives.

Inflated Implementation Costs
Organizations with fragmented platforms and poor data quality spend significantly more on AI pilot programs as a large share of effort is consumed by manual data cleaning, metadata harmonization, and configuration remediation. Industry research consistently shows that the primary barrier to scaling AI is not model development but data readiness. The 2025 CDO Insights survey found that 43 percent of data leaders cite data quality and data readiness as the top obstacle to successful AI initiatives.[11] In heavily customized Veeva environments, the burden can be greater still, as each non-standard object, lifecycle variation, and overridden security model must be individually mapped and reconciled before an AI model can be trained.

Opportunity Cost of Delayed Insights
In a sector where speed-to-market defines a competitive advantage, the cost of delayed AI adoption is measured in months of lost productivity and deferred value capture. A Deloitte report indicates that lab-of-the-future investments have led to 30 percent greater cost efficiency for early adopters. The implication is clear: organizations that wait for "perfect" conditions before beginning their AI journey risk being overtaken by competitors who have invested in foundational readiness.[12]

The Maintenance Trap
Complex and highly customized environments consume scarce expert capacity that should otherwise be directed toward innovation. Veeva specialists, data architects, and compliance leads are instead pulled into troubleshooting, release impact analysis, and remediation of legacy customizations. With enterprise R&D platforms evolving rapidly and introducing new automation, analytics, and data access capabilities with each release cycle, the cost of maintaining non-standard environments adds quickly. Every release widens the gap between what the platform can deliver, and what heavily customized environments can realistically adopt.

Beyond the Technical Audit: A Holistic Framework

Traditional assessments, including vendor-provided configuration reviews, typically focus on technical parameters such as object counts, lifecycle complexity, API usage, and system performance. While valuable, they do not answer the question that matters most for AI readiness[3]: Can this platform serve as a reliable foundation for machine-driven decision-making?

Answering that question requires more than a technical audit. It requires a multidimentional assessment that bridges the gap between "it works for human users" and "it's AI-ready." The BASE Value Realization Assessment therefore assesses four dimensions, each through an explicit AI readiness lens:

Table 2: The BASE life science Holistic Value Realization Assessment

Dimension	The AI Readiness Lens
Technology	Moving from opaque, highly customized configurations to standardized and interpretable platform structures that are easier to maintain, integrate, and govern. Evaluating the balance between standard and custom components as an indicator of long-term platform readiness.
Data	Ensuring high-quality metadata, structural consistency, and traceable relationships across the R&D landscape. Assessing whether data adheres to FAIR principles so it can be reliably found, accessed, connected, and reused by machines.
Process	Standardizing workflows and clarifying decision points so that AI can support execution in a controlled and compliant way. Assessing alignment between SOPs, business processes, and platform configuration to avoid reinforcing non-compliant or inefficient practices.
People	Building the roles, capabilities, and governance needed to manage AI responsibly. Assessing data stewardship, change readiness, and cross-functional oversight across business, compliance, and technology teams.

FAIR Data Principles as the Foundation for AI

Among the structural foundations required for AI readiness, data architecture is often the most critical. The FAIR principles provide a practical framework for designing data environments that machines can reliably interpret and reuse.

The FAIR data principles - Findable, Accessible, Interoperable, and Reusable - have emerged as the foundational framework for data management in pharmaceutical industry.[13] Originally formulated in 2016[14], these principles have gained renewed urgency in the context of AI adoption, because they describe the data characteristics required for machine learning systems to reliably reuse scientific datasets and support advanced analytics.[15]

In the Veeva context, FAIR compliance means ensuring that every document, record, and workflow generates metadata that is machine-readable, consistently structured, and linked across domains. In practice, this means creating the conditions for a clinical trial record, a safety case and a quality event to be connected through standardized data models and governed relationships rather than manual cross-referencing.

Table 3: FAIR Data Principles Applied to Veeva AI Readiness

Principle	What It Means in Veeva	AI Readiness Implication
Findable	Every asset has unique, persistent identifiers and rich metadata	AI models can reliably locate, classify and index relevant data without extensive manual preparation
Accessible	Data can be retrieved through governed interfaces, integration services, and appropriate access controls	AI pipelines can access relevant platform data programmatically and at scale, within compliance boundaries
Interoperable	Common vocabulary, shared taxonomies, and consistent structures across Clinical, Safety, Quality, and RIM	Cross-domain AI use cases become feasible without extensive reconciliation or remapping
Reusable	Clear provenance, context, and usage constraints for all data assets	Models can be trained, validated, and audited against traceable and trustworthy data sources

The BASE life science Value Realization Methodology

To move from diagnosis to action, organizations need more than a conceptual framework. They need a structured methodology that translates platform observations into quantified maturity findings, clear priorities, and an executable remediation roadmap. The Assessment combines Veeva domain expertise, structured scoring, and AI-enabled analysis to evaluate how ready an R&D environment is for scalable AI adoption.

The methodology is built around four dimensions: technology, data, process, and people. It is supported by AI accelerators that deepen the analysis and reduce manual effort across configuration review, metadata profiling, process assessment, and stakeholder insight synthesis.

These tools do not replace expert judgment. They accelerate pattern detection, expose inconsistencies, and surface improvement opportunities more quickly and at greater scale.

The Four Dimension Maturity Model

The assessment findings are then consolidated into a structured maturity model scored on five levels, from ad-hoc/reactive operations through to fully optimized, AI-driven decision-making.

The outcome is a prioritized remediation roadmap that helps organizations address structural constraints and prepare the platform for scalable AI adoption.

BASE life science Proprietary AI Accelerators

These proprietary AI tools compress delivery timelines and deepen analytical rigor:

ConfigHealth

Analyses Veeva Vault configurations across objects, lifecycles, workflows, and document hierarchies to identify differences from the standard baseline, custom complexity, and improvement opportunities. This provides an objective, data-driven view of how the platform has been configured, delivering actionable insights to support AI adoption.

DataScan

Evaluates the quality, completeness, and structure of Vault documents, including metadata usage, document relationships, template usage, field population rates, and lifecycle patterns, to identify hidden gaps and risks across content and configuration.

AuditLens

Analyzes user behavior and process efficiency across Vault, surfacing workflow rejection and cancellation patterns, task compliance rates, access anomalies, and usage trends to highlight operational bottlenecks and governance risks.

The Path Forward

The objective is not standardization for its own sake. It is to create an R&D environment capable of supporting AI-driven operations at scale, with platform configuration, data architecture, process design, and organizational capability aligned to capture value from AI investments

By strengthening platform configuration, data foundations, process alignment, and governance now, organizations position themselves to:

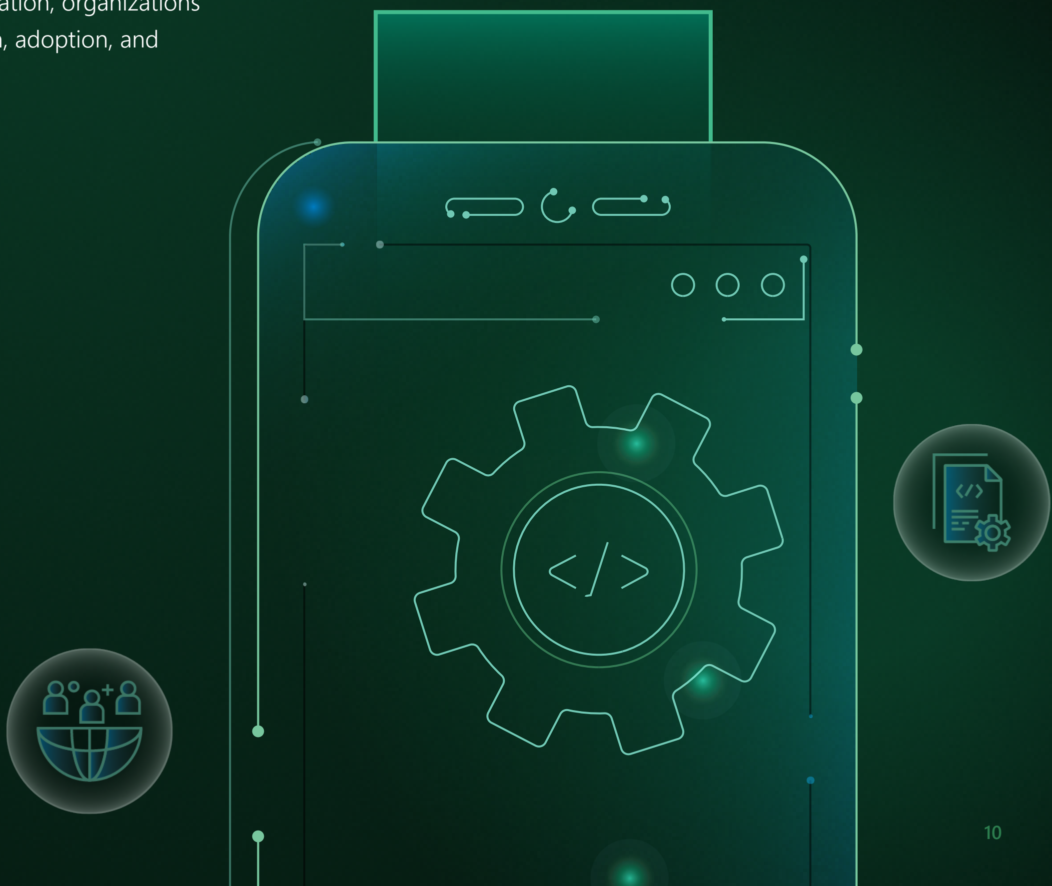
Reduce Maintenance Costs: Standardized environments are significantly easier to support, upgrade, and govern. When configuration models, metadata structures, and governance practices are aligned, platform releases become opportunities to adopt new capabilities rather than exercises focused solely on risk mitigation.

Accelerate AI ROI: High-quality data, aligned metadata, and standardized processes reduce deployment friction and shorten the path from use-case design to production. Instead of spending months on cleanup and reconciliation, organizations can focus resources on validation, adoption, and value realization.

Strengthen Continuous Compliance: Strong governance, traceable data provenance, and controlled AI processes make it easier to maintain alignment with GxP expectations[16] and emerging guidance on AI model documentation, validation, and monitoring.[17]

Unlock Cross-Domain Intelligence: When Clinical, Safety, Quality, and Regulatory data follow consistent structures and shared metadata principles, organizations can support AI use cases that span the full R&D lifecycle, from signal detection and quality investigations to submission preparation and post-market surveillance.

Leverage Innovation: The industry is rapidly introducing new capabilities for automation, data orchestration, and advanced analytics. Organizations with clean, well-governed environments can adopt these capabilities faster; those with technical debt must first remediate foundational issues before they can benefit.



Conclusion: From Digital Maturity to AI Value

AI readiness is not primarily a technology problem; it is a maturity problem. The leaders of the next era will not be those with the largest AI budgets or the most advanced models. They are the ones that have systematically prepared their platforms, data, and governance to support reliable, compliant, and scalable machine-driven decision making.

As you plan for 2026 and beyond, consider: Is your R&D platform optimized to power AI-driven workflows, or is it primarily serving as a digital archive? Is your metadata machine-readable, or does it require human interpretation at every step? Are your configurations, data and processes aligned across domains, or do they create silos that AI cannot bridge?

The ability to answer these questions with data rather than intuition is the first milestone of digital maturity. The BASE Veeva R&D Assessment provides a clear view of your current platform maturity and the priority actions needed to support compliant AI at scale.

Organizations that address foundation complexity proactively will transform their R&D platforms into strategic intelligence engines. Those that delay will find that AI initiatives remain trapped in isolated pilots rather than delivering enterprise-scale impact.

References

1. CB Insights, "Pharma AI Readiness: How the 50 Largest Companies by Market Cap Stack Up," 2025.

2. MIT, "The GenAI Divide: State of AI in Business 2025. MLQ.ai Research Report", 2025.

3. McKinsey Global Institute, "The Economic Potential of Generative AI: The Next Productivity Frontier", 2023.

4. Research & Development Policy Framework, PhRMA, https://phrma.org/policy-issues/Research-and-Development-Policy-Framework

5. McKinsey & Company, "Accelerating Clinical Trials to Improve Biopharma R&D Productivity," 2024.

6. Deloitte, "Realizing Transformative Value from AI & Generative AI in Life Sciences," 2024.

7. McKinsey & Company, "Boosting biopharma R&D performance with a next-generation technology stack", 2025.

8. KPMG, "KPMG Global CEO Outlook", 2025.

9. IMD Business School, "Future Readiness Indicator: Pharmaceutical Industry 2025", IMD Center for Future Readiness, 2025.

10. European Medicines Agency (EMA), "Review of AI/ML Applications in Medicines Lifecycle (2024): Horizon Scanning Short Report", EMA/571739/2024, 2025.

11. Informatica, "CDO Insights 2025: Data Leadership in the Age of AI", 2025.

12. Deloitte, "Future-Proofing Pharma R&D Labs", 2025.

13. McKinsey & Company, "Boosting Biopharma R&D Performance with a Next-Generation Technology Stack," 2025.

14. Wilkinson, M.D. et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," Scientific Data, 2016.

15. Gadiya, Y. et al., "FAIR Data Management: What Does It Mean for Drug Discovery?", Frontiers in Drug Discovery, 2023.

16. ISPE, GAMP® Guide: Artificial Intelligence, 2025.

17. U.S. Food and Drug Administration, "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products," Draft Guidance for Industry, 2025.

Author Perspectives

This white paper reflects the combined perspectives of experts in R&D transformation, platform architecture, and data governance in life sciences.

Manuella Dautan

Manuella has over 15 years of experience supporting life sciences organizations in R&D. Her work focuses on how digital platforms, governance models, and operating structures must evolve to enable scalable analytics and AI adoption.

Thomas Videbech

Thomas brings nearly 15 years of experience advising pharmaceutical companies on enterprise R&D transformation and digital platform strategy. He focuses on helping organizations modernize operating models and technology foundations to support advanced analytics, automation, and AI-driven decision making.

Flemming Jensen

Flemming brings close to a decade of experience in data architecture and information management within regulated life sciences environments. His expertise focuses on designing scalable data models, strengthening metadata governance, and establishing robust data foundations that support reliable analytics and AI-ready R&D ecosystems.

Pierre-Guillaume Amiel

Pierre-Guillaume is a specialist in Veeva Vault platform architecture and configuration within the life sciences industry. His work focuses on optimizing platform configurations, reducing structural complexity, and aligning environments with scalable, compliant, and AI-ready operating models.

Abi Wilkinson

Abi has nearly a decade of experience advising pharmaceutical organizations on operational and digital transformation across R&D. Her work focuses on aligning business processes, governance models, and organizational capabilities with evolving digital platform strategies.

About BASE life science

BASE Life Science, an Infosys company, is a leading Veeva implementation partner with deep expertise across the entire Veeva R&D and Commercial suite. With over 200 Veeva engagements, 100+ certified Veeva specialists, and a team that includes former Veeva product and professional services staff, BASE combines product-level configuration depth with industrialized delivery models.

Trusted by 18 of the top 20 pharmaceutical companies globally, BASE life science Centre of Excellence provides standardized playbooks, templates, and AI accelerators that ensure consistent quality and faster time-to-insight across every engagement.

Request an Assessment

Contact our experts to obtain a data-backed evaluation of your Vault landscape and define a clear path to scalable, compliant AI adoption.

Get in contact now