Informatica

Workbook

Generative AI Adoption Guide

6 key data management strategies to conquer GenAI fragmentation and drive success

Where data & AI come to LIFE™

Contents

Foreword

A step-by-step workbook that empowers organizations to create relevant, responsible and robust Generative AI (GenAI) solutions that seamlessly integrate both structured and unstructured data for better business outcomes.

Introduction

Striking the Perfect Balance

Generative artificial intelligence (GenAI) primarily runs on unstructured data, such as text documents, images, audio and video. These diverse formats enable large language models (LLMs) to generate rich and dynamic outputs. However, this highly fluid nature also makes GenAI difficult to govern, validate and operationalize in an enterprise context.

Because unstructured content lacks explicit schemas, lineage or fine-grain access controls, GenAI systems built on them are prone to:

- Hallucination and contextual error
- Privacy and intellectual property risks
- Opaque decision logic
- Challenges in tracing lineage or enforcing policy

In contrast, the structured data world has long been supported by technologies and practices to clean, integrate and govern data assets effectively. Tools such as master record matching, tracking data origins and rules based on metadata allow control and accountability possible in AI workflows atop tables and semi-structured files. Because structured data systems are well-developed, it's smart **for organizations to reintegrate their GenAI with structured data pipelines to reduce risks mentioned earlier.**

Structured data is the reliable "scaffolding" for GenAI, stabilizing its fluid yet unpredictable nature. Just as scaffolding provides the framework that allows a building to rise safely and in alignment with blueprints, structured data enables GenAI systems to generate functional, compliant and context-aware outputs, while remaining anchored to enterprise knowledge and rules.

Harnessing Unstructured and Structured Data for GenAI Success

This convergence of structured and unstructured intelligence represents a new frontier for enterprise architecture. It requires platforms that can unify data pipelines, metadata, integration flows and observability layers — without compromising speed or innovation. However, most data platforms still lack an integrated way to manage this convergence, resulting in a phenomenon that we refer to as "The GenAI Fragmentation."

In enterprise GenAI, a multi-vector fragmentation arises from the existence of isolated data silos, internal competition among departments to expedite GenAI project deployment, and the divergent responsibilities between IT teams handling GenAI development based on unstructured data and those managing the traditional structured data platforms. This is compounded by the rapidly evolving landscape of GenAI models and the vendor ecosystem. Hence, cohesive GenAI strategy and integrated technology across the enterprise has been elusive. Addressing these issues requires strategic governance, enhanced collaboration and agile adaptation.

While GenAI architectures may involve dozens of tools, services and models, what's needed is a **centralized, vendor-neutral data management platform** that serves as the connective tissue — an organizing backbone across all those moving parts. IDMC delivers that. It provides a common language for how GenAI components are developed, reused and deployed while enforcing the governance and policy controls that keep those components responsible and secure.

This workbook explores how to connect structured governance with unstructured GenAI experiences through a combination of data integration, quality, MDM, metadata orchestration, policy enforcement and lifecycle control. The following steps are essential for turning a fragmented data and AI landscape of experiments into relevant, responsible and robust GenAI solutions, while overcoming the crucibles of the GenAI Fragmentation!

Informatica's Intelligent Data Management Cloud (IDMC) has long been recognized as the industry's most complete platform for enterprise-scale structured data management. While GenAI has shifted the spotlight toward unstructured data, semantic context and real-time interactivity, the foundation for scalable, trustworthy AI remains rooted in well-governed, high-quality data. This is where IDMC provides unmatched value: it is designed to be **the enterprise hub for data and the central orchestrator of GenAI** model services.

While many GenAI developers and integrators value experimentation and mass customization, few address the underlying chaos of fragmented data. IDMC brings **structure to that chaos,** ensuring that every model, agent or automation is grounded in accurate, trusted data. Its extensible design enables seamless interoperability with vector databases, GenAI APIs, semantic search tools and external AI orchestration layers.

IDMC's Eight Capabilities for Responsible, Relevant and Reliable AI

At its core, IDMC brings together **eight essential capabilities** that enterprises need to anchor GenAI innovation in a trusted data fabric:

1. **Data Discovery and Metadata Management** – Unified cataloging, lineage tracking and active metadata propagation across systems and clouds
2. **Data Governance** – Stewardship workflow to manage policy-driven controls for privacy, usage rights, compliance and ethical oversight
3. **Data Quality** – Rules-based profiling, cleansing, enrichment and continuous observability to ensure reliable, accurate inputs
4. **Data Integration** – High-performance, low-code ingestion and transformation pipelines across hybrid environments
5. **Application Integration** – Seamless orchestration of GenAI and enterprise application services, APIs, events and business processes across platforms
6. **Master Data Management (MDM)** – Golden record creation and stewardship for critical business domains
7. **Data Access Management** – Policy and metadata-driven access control, secure data sharing and anonymization
8. **Data Marketplace** – A user-friendly storefront for data and GenAI model-provisioning, with transparency-promoting features such as model card, usage metering and reuse via governed data exchange

Step 1

GenAI Strategy: Defining Purpose with Precision

Too often, enterprises treat GenAI as a scattered collection of experiments. Without a clear strategy, these initiatives compete for resources, drift from business priorities and fail to scale. A GenAI wish list is not a well-crafted strategy. We need an intentionally curated portfolio of GenAI initiatives with clearly measured outcomes and accountability.

At the foundation of that strategy should be a **categorization framework**. Every GenAI initiative must be classified into one of three strategic buckets:

- **Foundational:** Investments that build reusable capabilities, such as enterprise search, data enrichment and agent orchestration platform
- **Transformational:** Initiatives that change core process, impact product roadmap or alter customer journey
- **Compliance-Driven:** Initiatives tied to risk, regulatory compliance and contractual obligations

Each initiative must be traceable to the strategic business unit goals, the KPI it improves and the executive steering committee and/or sponsor responsible for achieving those outcomes. Strategic roadmaps must be developed with short-, medium- and long-term horizons, considering dependencies and synergies across initiatives.

SELF-ASSESSMENT

What's Your GenAI Strategy?

For each question, rate your organization's readiness:
(0 = Not in place, 1 = In planning, 2 = Partially implemented, 3 = Fully implemented)

Area	Question	Score (0-3)
Categorization	Have you classified your GenAI use cases as being foundational, transformational or compliance-driven?	
Alignment	Is each use case mapped to strategic business KPIs and owned by a specific business unit?	
Prioritization	Have you assessed each initiative's size in terms of effort versus impact to guide resource allocation?	
Planning Horizon	Do you have a short-, medium- and long-term roadmap that balances quick wins with long-term value?	
Integration	Have you reviewed overlaps or dependencies between GenAI initiatives to create cohesion?	

In Practice

Challenge

A corporation with extensive customer-facing content faces challenges due to multiple departments maintaining their own unstructured data repositories, using platforms such as Sharepoint, Onedrive and Dropbox. Despite frequent collaboration, the separate platforms lead each department to pursue independent GenAI initiatives resulting in fragmented systems and data silos.

Solution

To address these challenges, the corporation implemented IDMC's rich data and application integration capabilities to integrate unstructured data into a centralized, curated enterprise knowledgebase. This approach involved several strategic layers:

- **Bronze Layer:** Using retrieval-augmented generation (RAG), the corporation ingests, classifies and enriches documents from all departments, removing outdated or redundant content.
- **Silver Layer:** The output from the bronze layer, such as classified information chunks, is organized into logically clustered embeddings and stored in domain-specific vector databases.
- **Gold Layer:** Data access management is applied to enable granular GenAI agent security, ensuring only authorized users can retrieve responses from appropriate domain vector DBs for specific use cases.
- **Data Catalog:** Metadata and lineage among these layers enhance traceability, allowing GenAI responses to be traced back to their originating source documents.

Outcome

By utilizing IDMC, the corporation eliminated conflicting and outdated content and achieved unified knowledge management with context-aware access control across departments. This integration has enabled the departments to synergize their GenAI strategic roadmap across multiple functions and business units, fostering collaboration and enhancing operational efficiency.

Step 2

Evolving Data Governance for Today's GenAI Reality

GenAI introduces new layers of complexity that stretch traditional data governance to its limits. At its core, GenAI governance draws from three pillars: data governance, model risk management and AI ethics. While this workbook does not cover model risk or ethical oversight in depth, it focuses on the foundational layer — data governance for GenAI — where most enterprises must begin.

Unlike structured data governance, which was designed for known data elements in well-defined systems, GenAI consumes and produces unstructured, semi-structured and operational metadata across analytical, developmental and runtime environments. This means governance must expand beyond the data platform into analytical workflows, pipelines and even the prompts and agents orchestrating responses.

Before extending governance, leaders must assess the maturity of what they already have. In highly regulated industries like banking, organizations have followed risk data aggregation frameworks such as BCBS 239 for over a decade, yet self-assessments still reveal dissatisfaction with data governance maturity. That's not failure — it's the norm. But GenAI raises the stakes.

Governance must now prioritize new forms of data risk, such as:

- Inferred context from unstructured data
- Governance of prompt libraries and intermediate artifacts
- AI validity and explainability
- Exposure to adversarial prompts and synthetic data attacks

The key is reprioritization, not reinvention. Enterprises must shift from generic data policies to purpose-built governance that is contextual, adaptive and enforceable across the GenAI stack's structured and unstructured layers.

SELF-ASSESSMENT

Are you governing the right things for GenAI?

Area	Question	Score (0-3)
Baseline Maturity	Do you have a formal data governance program in place for structured data?	
Governance Gaps	Have you assessed which parts of your existing governance framework are insufficient for GenAI (e.g., unstructured data, prompt artifacts)?	
Reprioritization	Have you reprioritized governance efforts to account for GenAI-specific risks and analytical pipelines?	
Extension	Is your data governance extended into analytical development and deployment environments (not just the data lake or warehouse)?	
AI Risk Awareness	Do you plan to evolve governance to include AI-specific risks like explainability, bias and adversarial threats?	

Step 3

Architecture Integration: Designing for Composability and Control

The good news? GenAI architecture isn't as foreign as it looks. From a governance and design perspective, the principles that once shaped classic data warehouse and business intelligence (BI) solutions still apply: reusability, integration, observation and user-centricity. These are not new ideas — they're just being applied to new patterns, like prompt flows, agent coordination and unstructured context processing.

What has changed is the diversity of threats and the expansion of data shapes. GenAI introduces new architectural layers: vector stores, prompt templates, agent memory and orchestration runtimes. However, enterprises that have built mature data ecosystems already have a head start if they can intentionally apply those same architectural disciplines.

A core question emerges: **What shapes your architecture — your tooling or your organization?** Most companies find that their digital architecture mirrors their org chart. Expect fragmentation if IT is siloed into separate data, application and infrastructure teams. If the organization is structured only by a business unit, expect duplicated effort and tool sprawl.

The right model? Usually, a federated structure blends business-aligned ownership with shared technical standards and platform layers. Some industries may prefer a more centralized approach, such as finance and defense, while others, like media and retail, may favor democratization. Regardless of the model, the goal remains to align organizational design with enterprise architecture, rather than solely focusing on solution architecture.

Only with this alignment can GenAI evolve from a patchwork of pilots to a composable, governed and scalable intelligence layer across the enterprise.

SELF-ASSESSMENT

Does your architecture support scalable and governed GenAI?

Area	Question	Score (0-3)
Component Reusability	Do you have a unified platform that supports reusable software components via cloud application integration, iPaaS and API lifecycle management?	
Platform Neutrality	Does your data integration layer operate independently of hyperscalers, ISVs and analytics tools — serving as a neutral backbone across your ecosystem?	
Metadata and Stewardship	Is your architecture designed to connect and manage both technical and business metadata, including lineage and stewardship workflows?	
Data Protection and Policy Enforcement	Do you have integrated capabilities to enforce data quality, security access, privacy controls and consent/legal use of data across your GenAI stack?	
Intelligent Automation and Democratization	Does your platform use GenAI to intelligently automate data and metadata management, empowering technical and non-technical users?	

In Practice

Challenge:

A global manufacturer faces challenges in integrating generative AI (GenAI) into its supply chain due to incomplete data and poor context. Multiple versions of product master data result in hallucinations and inconsistencies, while the absence of data lineage hampers the ability to trace data provenance and workflow processes. Additionally, the lack of connection between large language model (LLM) outputs and transactional application insights undermines the system's ability to reason effectively over relationships and context, leading to invalid and unreliable model results.

Solution:

To address these issues, the manufacturer used IDMC to connect LLMs with master data from Product and Customer 360 MDM and curated business definitions, ensuring context-rich responses. Key is IDMC CDI RAG Ingestion, which links insights from unstructured documents to Reference 360 taxonomies for semantic coherence across vector stores.

Outcome:

With IDMC, the manufacturer improves its GenAI assistant by providing structured product master data, delivering accurate product comparison responses based on up-to-date information. Vendor contract clauses are classified and linked to business terms in the data catalog. Using Reference 360's taxonomy, the RAG maps legal entities to contractual terms in a vector database, enhancing GenAI's reasoning. This ensures reliable answers for product catalog and terms queries, boosting supply chain efficiency.

Step 4

Trusted GenAI Data Pipelines: The Regimen of Exercising Data Management

Before GenAI can be scaled, trusted or even meaningfully evaluated, an organization must return to the basics — the fundamental practices of data management. The steps are not radical departures from long-standing principles; they are updated applications of the same disciplined regimen, now adapted for the age of intelligent agents and unstructured data.

It begins with ingesting and classifying data from diverse sources, such as documents, images, customer interactions and system logs. That data must then be standardized, integrated and enriched with business meaning through knowledge modeling, taxonomy definition and semantic scaffolding. From there, metadata becomes the connective tissue, linking technical and business contexts, surfacing ownership and lineage and powering intelligent search and retrieval.

GenAI initiatives must define clear use cases to turn data into action and build modular, reusable agents that connect prompts to purpose. These agents require fine-grained policy metadata to ensure secure, legal and ethical operation, especially as they move from pilot to production. Finally, orchestration, testing and delivery mechanisms must ensure that agents work together reliably and transparently with human oversight when needed.

There is no perfect data management platform that meets every GenAI requirement. However, the need for a **central, intelligent and extensible platform** is non-negotiable. That platform must integrate with best-of-breed third-party services, support multi-layered architectures and be backed by a robust ecosystem of system integrators and accelerators.

SELF-ASSESSMENT

Are your data pipelines ready for GenAI?

Area	Question	Score (0-3)
Data Ingestion and Classification	Do you systematically ingest and classify data from diverse sources, including documents and images?	
Standardization and Integration	Is your data standardized, integrated and enriched with business meaning through taxonomy and semantic tagging?	
Metadata Connectivity	Does your metadata link technical and business contexts effectively, providing clear ownership and lineage?	
Use Case Definition	Have you defined clear GenAI use cases and developed modular, reusable agents to connect prompts to purpose?	
Policy Metadata and Security	Does fine-grained policy metadata govern your agents to ensure secure, legal and ethical operations?	
Orchestration and Testing	Do you have robust orchestration, testing and delivery mechanisms to ensure reliable agent operations?	
Central Platform Integration	Does your platform integrate with third-party services and support multi-layered architectures for GenAI?	

Step 5

AI Ops Automation: Preparing for the Transition from Code to Enterprise Control

While full-scale GenAI operations (AI Ops) warrant their own dedicated playbook, there is one foundational reality every enterprise must face now: most GenAI development is happening outside the boundaries of your centralized data platform — and that's not necessarily wrong.

Developers prefer tools they can control — Python, Langchain or the PaaS / SaaS-specific AI agent builders (e.g., Azure AI Foundry, Amazon Bedrock, Salesforce Agentforce). These are flexible, fast and tailored to the needs of data scientists and domain experts. However, what starts as a promising experiment can quickly become a deployment liability if left unmanaged and not integrated across use cases. Once GenAI pilots begin transitioning to user testing and production use, they must gradually shift from experimental codebases to **low-code, metadata-aware, policy-enforced components** integrated into a centralized platform like IDMC.

This shift isn't just about standardization — it's about **ensuring observability, trust and long-term operability**.

Here's a readiness check to assess your ability to scale responsibly:

SELF-ASSESSMENT

Are you set up to transition GenAI development into production?

Area	Question	Score (0-3)
Development Visibility	Do you know who is building GenAI pipelines and what tools they use (e.g., python code, PaaS/SaaS, internal apps)	
Toolchain Vetting	Have those tools been reviewed for security, privacy and integration alignment with your data management platform?	
Transition Standards	Do you have a documented process or guidebook for transitioning GenAI experiments into enterprise-grade, governed components?	
Refactoring Support	Are you enabling teams with low-code or metadata-driven accelerators to replace brittle custom code with reusable, platform-native patterns?	

In Practice

Challenge

A global manufacturing company sought to leverage Python and LangChain for generative AI (GenAI) development within its supply chain automation. However, the enterprise faced challenges in balancing risk and benefit. The seemingly cost-effective hand-coding option raised concerns among AI governance teams about increasing technical debt due to undocumented code, API key leaks, and the lack of enterprise-grade scalability, particularly for multi-agent development. Furthermore, the development team's lack of AI fluency and data literacy posed systemic risks, as GenAI data pipelines were often created without clear lineage, quality management, privacy enhancements, governance, and stewardship needed to meet responsible AI objectives.

Solution

To address these challenges, the company implemented IDMC's tools and environments:

Robust Testing and Deployment: CAI offers a low-code platform for building GenAI agents and RAG with enterprise-grade orchestration, audit logging, and API management, ensuring secure, governed deployment and testing.

GenAI Ops Ecosystem: Informatica's network of engineers, architects, and partners provides accelerators that help migrate Python pipelines to IDMC as projects advance from pilot to deployment.

Collaborative Governance: CLAIRE GPT and Copilot enable real-time data checks, observability, and integrated metadata and lineage, supporting collaborative GenAI data stewardship.

Outcome

The global manufacturing company secured its GenAI by wrapping its LLM in a hardened API with full lifecycle management in CAI. After prototyping in Python, development shifted to CAI for robust APIs and observability via GenAI Ops automation. The data pipeline was finally deployed in CAI and CDI, allowing data engineers and AI governance managers to use CLAIRE for streamlined GenAI change management, governance, and compliance through conversational natural language, boosting efficiency and reliability.

Step 6

GenAI-Driven Data Governance and Management

Governance that learns, adapts and scales GenAI doesn't just consume data governance — it can now contribute to it. Across the modern data landscape, we're seeing the emergence of GenAI-powered tools and platform-native capabilities that assist with everything from cataloging and metadata generation to data quality assessment and policy mapping. These agents support various roles: data analysts, BI developers, application developers, data engineers, data scientists and compliance officers.

However, organizations must approach this landscape wisely. Many GenAI tools are now offering trial access or lightweight implementations. This is a strategic opportunity to test-drive emerging capabilities without premature lock-in. The goal is to evaluate each tool in isolation and assess how they work together under a consistent data governance framework. In a multi-agent environment, it's critical that your GenAI components "speak the same language," or you risk exacerbating the big fragmentation all over again — this time inside your governance layer.

Equally important is **human enablement**. Despite recognizing the benefits of GenAI, many technical professionals — especially in data management — remain hesitant to adopt it. They're uncertain about job security, confused by fast-moving tool changes or frustrated by unreliable outputs. This silent resistance can become a roadblock. Now is the time to invest in your people through targeted **training, collaborative hackathons and community-driven experimentation**. Create a culture of empowerment and curiosity before fear and fatigue take root.

When implemented with care and cohesion, GenAI governance turns what was once a bottleneck into a catalyst for trust, agility and scale.

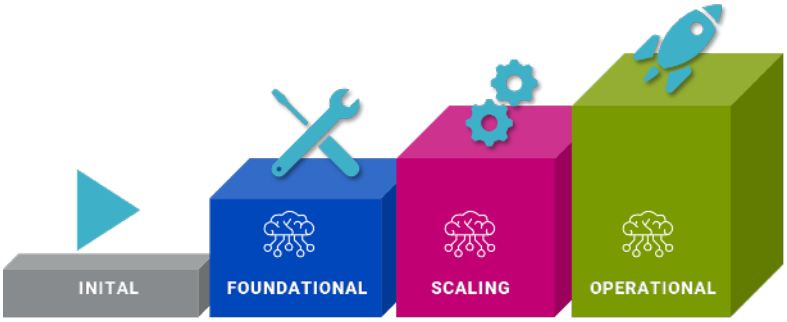
SELF-ASSESSMENT

Are you using GenAI to strengthen — not fragment — data governance?

Area	Question	Score (0-3)
Tool Exploration	Before committing to any, have you tested GenAI-powered data governance tools across different roles (e.g., cataloging, quality, compliance)?	
Interoperability	Have you validated that your GenAI governance tools can work together — or be orchestrated under a shared metadata framework?	
Framework Consistency	Are all GenAI agents operating from a standard policy, taxonomy and metadata foundation to avoid fragmented or conflicting actions?	
Platform Integration	Do these GenAI capabilities integrate with your central data management platform to ensure continuity, oversight and traceability?	
Human Oversight	Have you defined where GenAI agents should act autonomously versus where human-in-the-loop review is required for governance decisions?	
Talent Enablement	Are you equipping your data management professionals with GenAI literacy through training, hands-on workshops and community-based hackathons to reduce fear and accelerate adoption?	

Assessing Your Results

Total up all your results for each self-assessment to see where you fall on the GenAI maturity scale.



GenAI Maturity Level	Range	Description
Operational	80-96	Your data is clean, structured and accessible in a production environment that is continuously updated and integrated into real-time systems to power AI applications for immediate decision-making.
Scaling	64-79	You have a data foundation but have yet to make it operational for AI.
Foundational	32-63	Either you have started or you're just beginning your AI journey and have yet to scale
Initiating	0-31	You are in the early stages of your AI journey.

Fuel GenAI with AI-Ready Data

In the world of GenAI, there are no perfect tools — yet. The landscape is still evolving, and and no single platform today offers a complete, end-to-end solution for every GenAI need. That's not a flaw — it's a fact. But enterprises can and must now secure a strong, stable foundation that promotes trust, enables composability, and and scales with governance. This is where Informatica's **Intelligent Data Management Cloud (IDMC)** stands apart.

The IDMC Difference

Holistic Data Management

Provides you with a full-spectrum data operating platform, combining everything from metadata intelligence and lineage tracking to data integration, quality, governance, MDM and marketplace management.

Neutral and Extensible by Design

Unlike platforms locked to a single hyperscaler or tool stack, IDMC is inherently neutral, integrating across AWS, Azure, GCP, Snowflake, Databricks, Salesforce, OpenAI, Hugging Face and more.

Metadata Intelligence + GenAI Automation

Automates areas once dominated by manual curation, using GenAI internally to enrich metadata, streamline governance, auto-map lineage and power intelligent recommendations.

Policy and Protection, Built In

Grants IP protection, consent enforcement, data sovereignty and privacy policies with fine-grained controls that work across cloud and on-premises environments.

Global Ecosystem of GSIs and SIs

IDMC is supported by a global network of trained system integrators, ensuring enterprises can scale responsibly with partners who know the platform and its accelerators inside and out.

Accelerator-Ready

From ingestion blueprints to trusted agent orchestration patterns, IDMC enables rapid implementation through reusable assets that align with your existing tools and pipelines, from ingestion blueprints to trusted agent orchestration patterns.

IDMC is not a GenAI engine. It doesn't try to replace your LLMs, prompt engineering tools or or orchestration frameworks. What it does is far more critical: It enables all of them to work together **securely, responsibly and at scale.**

Begin Your Trusted GenAI Journey

Technology leaders have observed that generative AI is reshaping the very nature of SaaS. It's no longer just about the product — it's about how it empowers people, especially the less technical users who now have direct access to intelligent tools.

This shift increases the value of human-centric services that **complement the product** — advisors, curators and implementation partners who know how to activate GenAI within the real complexity of an enterprise. The pendulum may eventually swing back to a more product-led strategy. Still, it is prudent to engage with a **professional services team that lives and breathes IDMC** — one that's embedded in a **global community of best-practice curators** across system integrator (SI) and global SI (GSI) networks.

That's why we invite you to partner with us.

Guiding Your Steps

Informatica Professional Services offers our GenAI Strategy, Governance, and Architecture Assessment — a focused initiative designed to help you:

- Understand where you're at in your GenAI journey
- Prioritize what matters most to your business
- Align strategy, data governance and technical architecture in a clear, actionable roadmap

This workbook addresses the challenges of the GenAI fragmentation and how to overcome them. Let us help you in implementing your plan.

Contact Informatica Professional Services and we'll get you started on your GenAI journey.

About Us

Informatica (NYSE: INFA), a leader in enterprise AI-powered cloud data management, brings data and AI to life by empowering businesses to realize the transformative power of their most critical assets. We have created a new category of software, the Informatica Intelligent Data Management Cloud™ (IDMC), powered by AI and an end-to-end data management platform that connects, manages and unifies data across virtually any multi-cloud, hybrid system, democratizing data and enabling enterprises to modernize their business strategies. Customers in approximately 100 countries and more than 80 of the Fortune 100 rely on Informatica to drive data-led digital transformation. **Informatica. Where data and AI come to life.™**

Where data & AI come to

Worldwide Headquarters
2100 Seaport Blvd.
Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/company/informatica
x.com/Informatica

CONTACT US

IN18-5173-0525