



Unlocking Generative AI's Potential: The Imperative of **Data Trust**

Wipro and Informatica's Digital
Data Trust solution

Author: Deepa Jagmohan Behl





Table of contents

01	Introduction.....	03
02	The importance of Data Trust in AI	03
03	Lifesciences industry: A case in point	05
04	Wipro & Informatica's Digital Data Trust: A one-stop solution	07
05	Key components of digital data trust solution	08
06	Implementation journey	00
07	Delivering business and IT value with Digital Data Trust	11
08	Conclusion	11

Introduction

Organizations are adopting a data-driven culture making analytics and Generative AI (GenAI) initiatives crucial in shaping business strategies, improving customer experiences, and driving innovation. Most companies are exploring the potential of these solutions, with some even developing custom Generative AI solutions to drive business differentiation.

A significant barrier to scaling and adopting AI initiatives is inaccessibility to high-quality, trusted, governed data. Further, the advent of GenAI initiatives increase the focus on data privacy and data access controls. The consequences can range from financial losses and reputational damage to ethical concerns and regulatory penalties. Data and AI leaders are proactively reevaluating their data platforms to address these challenges.

Wipro & Informatica's Digital Data Trust provides a joint solution addressing the various dimensions of data trust within the organization. In this whitepaper, we address the importance of Data Trust in AI, the key components of the Digital Data Trust framework.

The Importance of Data Trust in AI

The power of GenAI hinges on the quality and trustworthiness of the data it learns from. This data falls into two categories: core data for model training and contextual data for generating prompts and guiding the output. Core data is key in training LLMs to recognize patterns and structures, providing the necessary knowledge for generating logical outcomes. On the other hand, contextual data, which includes specific information about the current situation or environment, supplements this process by offering prompts that guide the generation of outcomes.

Below are some of the key considerations for core data leveraged for Generative AI:

- **Contextualization** - Data must directly relate to the specific problem your GenAI model aims to solve. For example, a Generative AI model predicting insurance risk needs data such as client age, disease history, and pre-existing conditions, but not their occupation details.
- **Comprehensiveness** – Data must include all possible groups and scenarios to ensure adequate representation. For example, if the AI model will be used in multiple countries, your training data should include information from all those countries, not just a select few.
- **Bias Mitigation and Consistency** – Bias in data leads to bias in model outcomes. Your training data should reflect the full diversity of your target population, including groups often under-represented in datasets. Apart from tools and strategies for bias detection in models, iterative review of datasets is also needed to ensure the inclusion of less-represented classes or attributes. Further, data from various sources must be consistent for AI models to make reliable predictions. For example, if one data source lists blood pressure in mmHg and another in kPa, this needs to be aligned.
- **Regulatory Compliance** – Personally identifiable information (PII) – sensitive data, organizational confidential data, and other such data must be handled according to regulations like General Data Protection Regulation (GDPR). This might involve excluding, masking or anonymizing such data before it's used for training.



On the other hand, contextual data for crafting prompts for the GenAI model should have the following considerations to ensure optimal performance and generate relevant outputs:

- **Clarity** – Unambiguous prompts will provide more accurate and consistent outcomes from GenAI models. The more specific you are, the more focused and relevant the AI's response will be.
- **Specificity** – Helps AI understand the context and provide a more accurate response. Don't assume the AI has background knowledge that you want it to consider.
- **Bias Consciousness** – The language used in prompts can unintentionally introduce bias. For example, asking "What are the typical characteristics of a successful clinical trial participant?" might lead the AI to generate responses based on over-represented demographics such as young, healthy males. Instead by using "What factors contribute to diverse and inclusive participant recruitment in clinical trials?" can lead to a more balanced response.

However, most AI initiatives are marred with data challenges such as:



Data Integration

Combining data coming in from multiple different sources and in various formats can lead to data inconsistencies and complicate the model training process.



Data Quality

Missing values, inconsistencies, inaccuracies, and outdated information can compromise the reliability and trust in data, thereby hindering the training and performance of generative AI models.



Data Annotation

GenAI often requires large amounts of annotated data for training. Manually adding context and labels to the existing datasets is time-consuming and expensive, creating a significant bottleneck for AI initiatives.



Data Privacy

Leveraging PII-sensitive data for training models can lead to compliance issues with GDPR or other industry-related regulations for data protection resulting in possible regulatory fines and reputational risks for the organization.

The integrity and trustworthiness of data are foundational to the success of Generative AI initiatives. Overcoming these obstacles is crucial for organizations to fully leverage the power of AI, necessitating a strategic approach to data management and governance.



An Industry Example: Lifesciences

In the life sciences industry, accuracy and reliability are extremely important, a robust data framework is essential for navigating the complex regulatory landscape and driving innovation. In the following sections, we will delve deeper into the specific challenges and opportunities for data trust within this sector.

The life sciences industry is at the cusp of a Generative AI revolution. A McKinsey survey revealed that 79% of life sciences organizations plan to build their own GenAI solutions tailored to their unique requirements. This enthusiasm is driven by the potential of GenAI to transform various functions such as:

- **Accelerating research and development (R&D):** GenAI can analyze vast datasets and synergize human expertise with AI to unlock biological insights, improve understanding of disease progression and identify potential new targets for therapies significantly speeding up the drug discovery process.

- **Optimizing clinical trials:** By leveraging real-world data (RWD), GenAI can help recruit diverse and appropriate participants for clinical trials and assist in drafting regulatory submissions.
- **Enhancing manufacturing and supply chain:** GenAI can help minimize factory-to-market lead time in a global distribution model, predict and prevent supply chain disruptions, and enable track-and-trace capabilities for pharmaceuticals.
- **Personalizing commercial strategies:** GenAI can analyze patient and healthcare professional (HCP) data to personalize marketing, tailor digital omnichannel strategies, and identify high value HCPs.

Realizing these transformative use cases depends on access to high quality, integrated data. From clinical trials to 3rd party insights, life sciences organizations are using a variety of datasets to power their data analytics and GenAI applications. The table below provides a representative overview:

Internal data	External data
• Customer data (CRM) • ERP data • Sales databases • Digital platforms (web analytics, social media) • Regulatory submission data • Compliance data • Adverse event data • Real-World Evidence (RWE) data • Manufacturing & distribution data • Patient data • Post-market surveillance (PMS) data • Electronic health records (EHR) • Patient registries • Clinical trial management systems • Case Report Forms (CRFs) • Patient self-tracking devices • Administrative data • Devices data (Lab data, ECG data, MRI results, etc.) • Provider notes • Patient Reported Outcomes (PROs)	• HMO networks data • Birth & death stats • Disease registries • Funded Research & Study material (Framingham Heart, NHLBI funded studies) • National Institutes of Health Datasets • Market research firms • Health Economics and Outcomes Research (HEOR) Data • Healthcare databases

RWD data, like data from patients, clinicians, hospitals, and payers, is extremely valuable for life sciences organizations. By combining RWD with CRM insights or HCP engagement history, companies can:

- **Identify and target high-value HCPs:** By analyzing patient demographics, payer mix for their patients and referral patterns across networks, companies can segment and identify influential HCP segmentation and tailor marketing efforts accordingly.

- **Enhance pharmacovigilance:** RWD data such as patient-reported outcomes (PROs), electronic health records (EHRs), and clinical and hospital data can be leveraged to monitor the real-world effectiveness and safety of drugs, enabling early detection of potential adverse events.

These examples highlight how integrating internal and external data sources enables life sciences organizations to make informed decisions and improve patient outcomes.

However, this is easier said than done, and most organizations are struggling with data readiness as they scale their AI ambitions:

1 Data silos and discoverability - Data is often scattered across disparate systems and lacks clear and common definitions and labels, making it difficult and time-consuming for the end users to access the required datasets.

2 Data harmonization issues - Data collection through different touchpoints and in their different formats require harmonization to be leveraged in the context of a given business use case.

3 Data redundancy - Duplication of data across multiple systems (like CTMS, EHRs PROs etc.) leads to storage inefficiencies, increases the risks of errors, and hampers data management.

4 Poor data quality - Inaccurate, incomplete or inconsistent data entry practices can cascade downstream, undermining the usefulness of the AI models and leading to unreliable results.

5 Data access and privacy controls - Life sciences organizations need to deal with the challenge of balancing data accessibility for research and innovation with the need to comply with stringent patient data and privacy protection regulations. Excessively restrictive access can lead to missed business opportunities, while inadequate controls can lead to regulatory fines and reputational risks.

Wipro & Informatica's Digital Data Trust: A One-Stop Solution

Together, Wipro and Informatica provide a one-stop solution to address the multidimensional challenges of building and maintaining data trust within your organization. Our joint solution empowers you to unlock the full potential of your data for AI and analytics initiatives while ensuring compliance and mitigating risks. Below are some of the key tenets of the Digital Data Trust joint solution:

Cloud native at scale — Scale as needed for virtually all enterprise workloads with elastic and serverless processing. Run, interoperate and support virtually all combinations of multi-cloud and on-premises hybrid infrastructures.

Persona-based access and control — Empower both technical and business personas with role-based access to search and access relevant data sets, view data lineage, curate the data, assess and maintain data quality, and govern access tailored for specific needs.

Holistic data governance — Get a 360-degree view of overall data relationships for each asset, encompassing technical, business, semantic, and usage-based perspectives. This can be further enriched with business glossary terms, domains, policies, and processes for complete understanding and control. The solution also provides multiple levels of data lineage views, from technical column-level to more summarized business-level lineage overviews. It also integrates the data quality scores for comprehensive monitoring and insights.

Curation and collaboration — Foster a data-driven culture with a centralized data marketplace where users can publish, discover, rate, and review data assets/ collections, promoting data curation and collaboration.

AI for data management — Automate and accelerate data curation, glossary mapping, thus improving the data steward productivity.

How it works

Our Digital Data Trust solution establishes a robust data foundation by:

Centralized metadata repository

All ingested datasets are cataloged and enriched with business context, including glossary terms, domain, data ownership etc.

Proactive data quality management

Data quality controls are established to ensure data completeness, data consistency as well as reconciliation with the data sources. Data would be observed for any unexpected data or schema drift to capture anomalies or special bias scenarios well in advance.

End-to-end data lineage

Track data origin, transformations and usage throughout the data lifecycle, providing transparency and accountability.

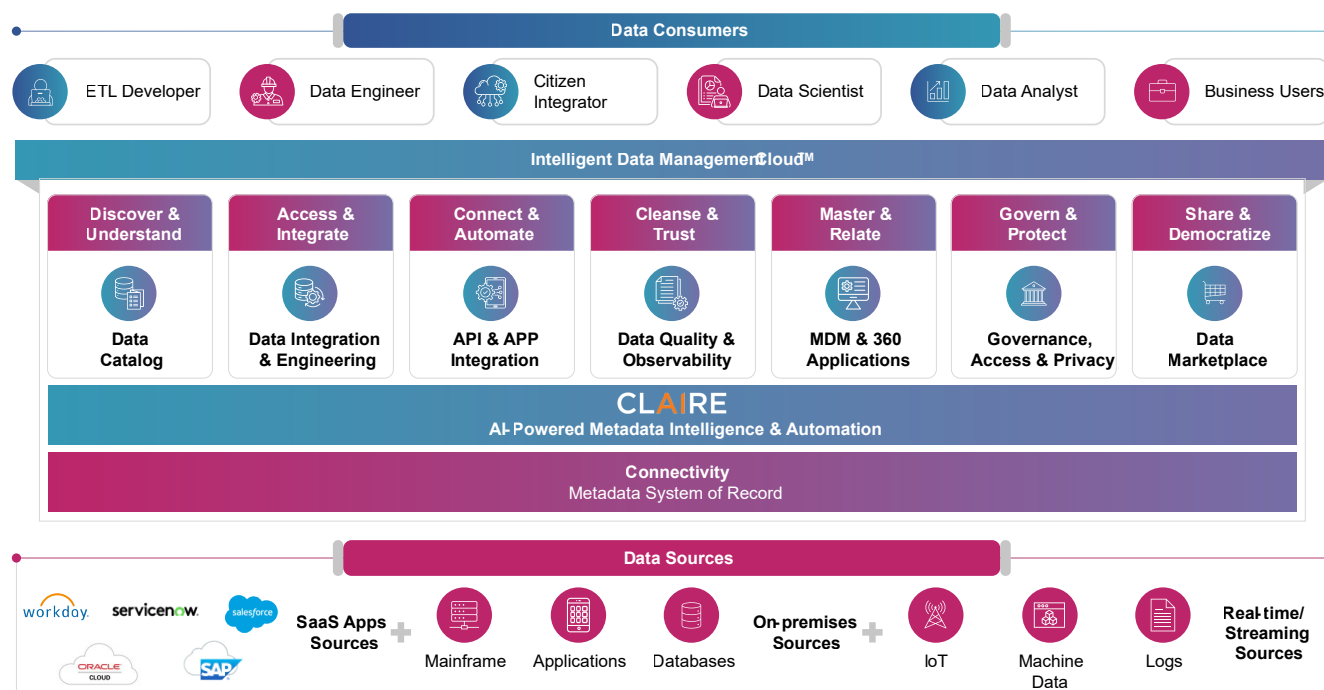
Data security and privacy

Sensitive data will be identified, classified and protected with granular access controls to ensure compliance and make the data accessible to the right users at the right time.

Key Components of Digital Data Trust solution

Digital Data Trust is built on a modular architecture, representing the power of the Informatica Intelligent Data Management Cloud™ (IDMC) and Wipro's deep technology and

domain expertise. The figure below illustrates the key components within the Digital Data Trust solution.



Informatica Cloud Data Governance & Catalog

Effortlessly manage and trust your data with AI-powered cataloging, automatic metadata classification, and enrichment. Explore asset relationships and analyze impacts with comprehensive, automated data lineage.

Informatica Cloud Data Quality and

Observability Ensure timely, trusted data by identifying and resolving issues with data observability. Continuously analyze data to detect problems and improve accuracy at scale using AI-powered rules. Understand data health through multiple lenses including data, pipelines, and business.

Informatica Cloud Data Assess Management

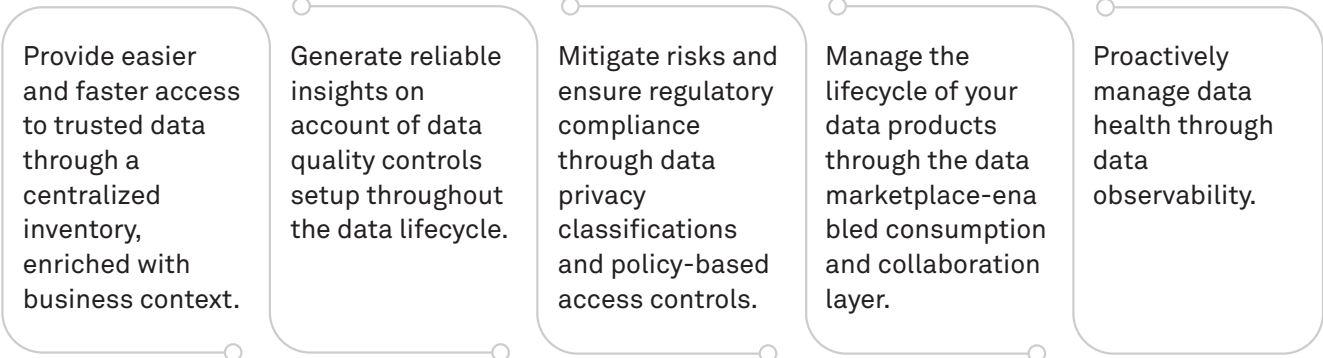
It enables policy-based data protections for efficient and responsible data use. By leveraging metadata conditions, it streamlines data access and enforces privacy policies across your organization, ensuring responsible use, regulatory compliance, and customer trust

Informatica Cloud Data Marketplace

Enable fast, safe data sharing with a data shopping experience for confident access. Maximize data use and value by sharing trusted data products. Empower users with self-service data access and boost operational efficiency by automating data sharing in alignment with governance policies.

Digital Data Trust combines the strengths of these various components with Wipro's data governance frameworks and accelerators to

address your unique needs towards data readiness for AI. We empower you to:



Maturity Model

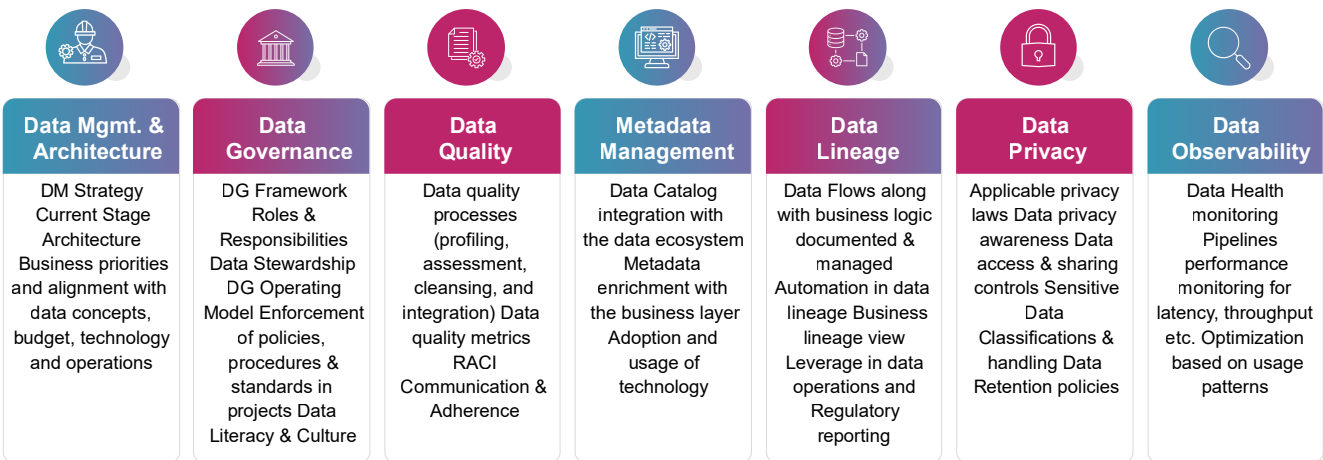
We understand that addressing data challenges can feel overwhelming. That's why Wipro takes a collaborative and phased approach to implement Digital Data Trust, ensuring your organization builds a solid data foundation for AI and analytics success.

The first step in this journey is **assess and prioritize**. We begin by understanding your

organization's unique data landscape and challenges. Our Data Trust Rating Model, depicted in the figure below, provides a comprehensive assessment model that helps set the baseline of your data trust journey, uncovering hidden issues and helping to build a compelling business case for data trust initiatives with various stakeholders.

Digital data trust rating model

Assess & Rate AS-IS data landscape across various dimensions, rating the people, process, and technology capabilities and sub-capabilities.



Provides a tangible view of the current state that would help understand the key gaps and potential areas of focus for the organization.

The next series of steps cover strategizing, piloting and scaling depicted in the figure below. Based on the assessment findings, we provide a clear digital data trust strategy and roadmap aligned with your business objectives and AI/analytics aspirations. We select the most

appropriate use case for pilot that addresses a critical business need and showcases the value of Digital Data Trust. We gather feedback from the pilot, refine the solution and further expand to onboard additional use cases, data sources, and business units.



Real-world success: Empowering a telecom giant

For a leading European telecom organization, Wipro & Informatica leveraged Digital Data Trust to enable business data products and establish a robust data governance framework. The solution delivered tangible results:

- **Strong data governance framework:** Established a strong data governance framework across critical domains, including Networks, Products, HR, and nine others.
- **Centralized data inventory:** Onboarded over 100 systems spanning databases, cloud platforms, and ERP systems into a unified technical catalog.

- **Enhanced data quality:** Implemented data quality controls as the data traversed through the medallion architecture of the data platform, ensuring data accuracy and consistency.
- **Data democratization:** Empowered users across the organization to discover, access, and utilize curated datasets through a centralized data marketplace.

Wipro's collaborative approach and deep expertise helps you navigate the complexities of data trust and establishes the foundation for AI and analytics initiatives.

Delivering Business and IT Value with Digital Data Trust

Digital Data Trust empowers organizations to build trustworthy AI and Generative AI by addressing the critical pillars of people, process,

data, and technology. The solution delivers key benefits which include:

Enable data-driven decision making and promote data literacy to help foster a data-driven culture.



50% improvement demonstrated in Data Trust Ratings for implemented use cases.



30% automation in business layer enrichment tasks leveraging Wipro's Business Glossary Generator and Informatica's embedded intelligent domain classification and tagging.



Wipro's Generative AI (WeGA) framework & accelerators are integrated to extend data governance principles to Generative AI, ensuring responsible and ethical AI development and deployment.



As organizations increasingly leverage their proprietary and external data for AI and Generative AI, concerns around data privacy and security with frequently surface. Digital Data

Trust directly addresses these concerns by enabling secure data sharing and collaboration, and mitigating risks associated with LLM models.

Conclusion

The transformative potential of Generative AI depends on a solid foundation of high-quality, trusted and governed data. As organizations work towards defining GenAI objectives and use cases, data readiness cannot be treated as an afterthought. Procrastinating data readiness will only lead an organization to scale down on AI ambitions. Wipro and Informatica's Digital Data Trust solution provides a comprehensive solution to accelerate your data governance journey.

Don't let data challenges hold back your organization's Generative AI ambitions. Reach out to learn how Digital Data Trust can be tailored to your specific needs and requirements.

Why Informatica?

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 pioneered 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 any multi-cloud, hybrid system, democratizing data to modernize their business strategies.

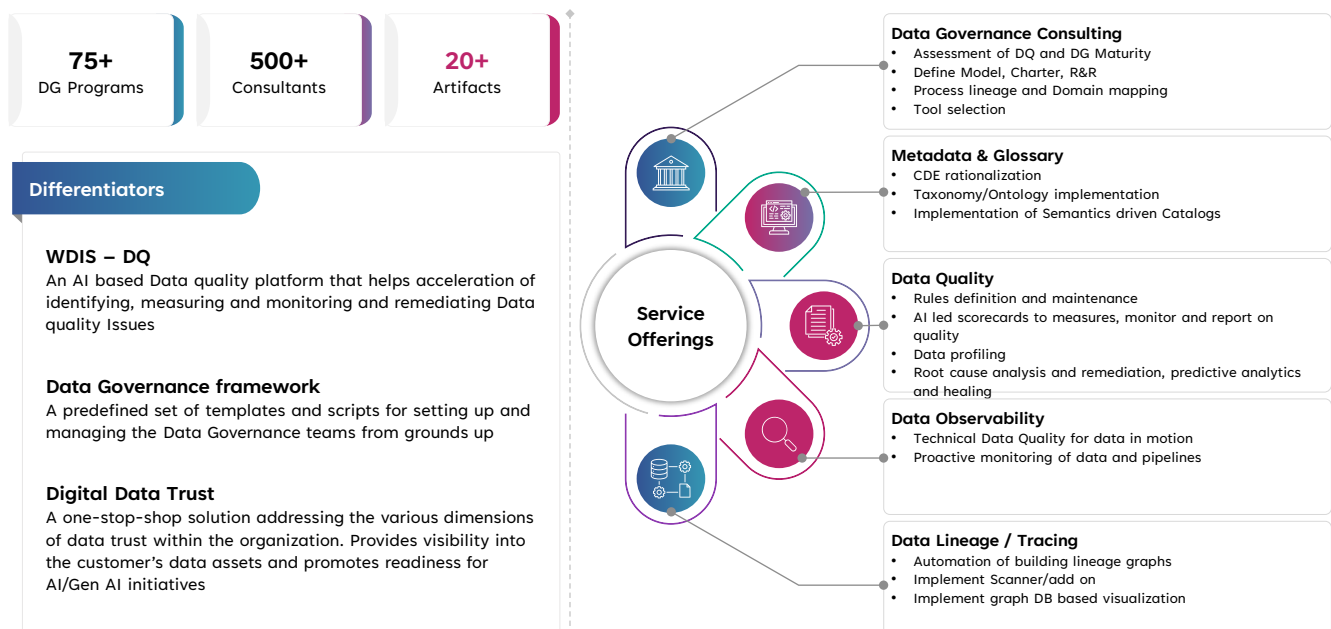
Informatica's IDMC provides a fully unified, cloud-native data management platform that integrates capabilities such as data governance, data quality, and data integration under a single

architecture. By centralizing these critical components, IDMC eliminates the need for fragmented solutions, offering a seamless experience for managing, integrating, and governing data across hybrid, multi-cloud, and on-premises environments. This unified approach ensures faster time-to-value and reduced operational complexity for AI and analytics initiatives.

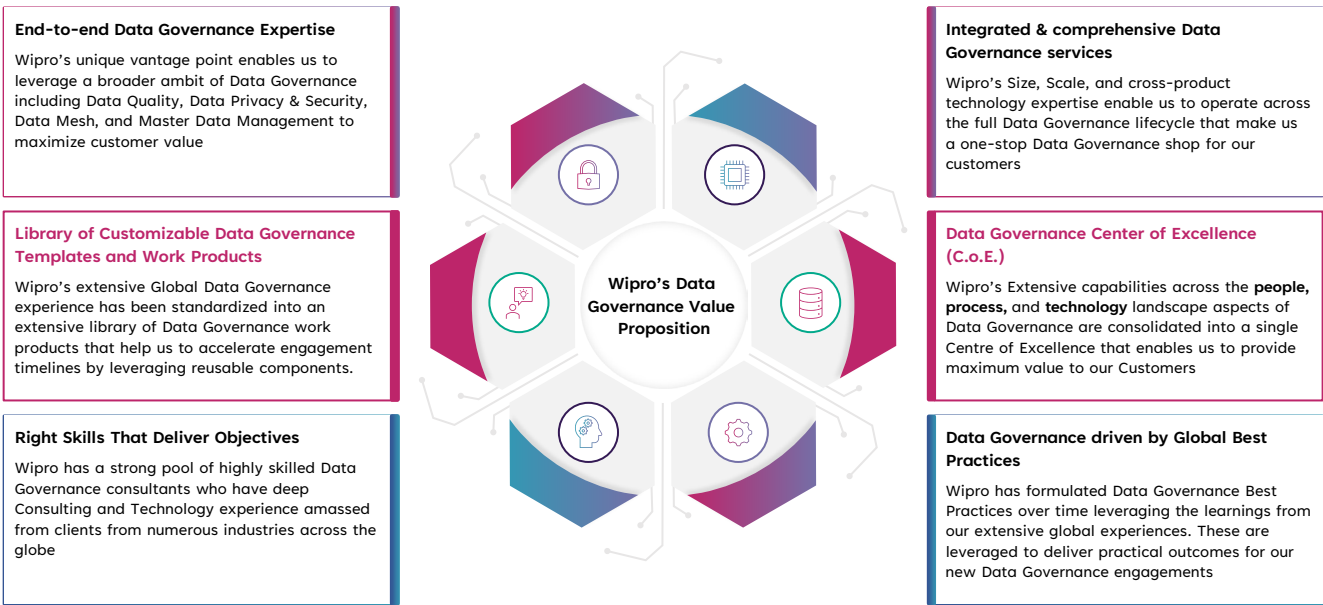
Customers in approximately 100 countries and more than 80 of the Fortune 100 rely on Informatica to drive data-led digital transformation. www.informatica.com. Connect with us on LinkedIn, X, and Facebook. Informatica. Where data and AI come to life.™

Why Wipro?

Explore Wipro's distinctive approach to data quality and governance, highlighting our service offerings and what sets us apart.



Data governance – A driver of trust & Value to deliver Governed, Secure, and Insightful data for the organization





Ambitions Realized.

Wipro Limited
Doddakannelli
Sarjapur Road
Bengaluru – 560 035
India

Tel: +91 (80) 2844 0011
Fax: +91 (80) 2844 0256
wipro.com

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading technology services and consulting company focused on building innovative solutions that address clients' most complex digital transformation needs.

Leveraging our holistic portfolio of capabilities in consulting, design, engineering, and operations, we help

clients realize their boldest ambitions and build future-ready, sustainable businesses. With over 230,000 employees and business partners across 65 countries, we deliver on the promise of helping our clients, colleagues, and communities thrive in an ever-changing world.

For additional information, visit us at **www.wipro.com**