Informatica

eBook

How Leaders Can Empower Data Teams for the AI Era

Drive positive business results with a modern approach to data and AI governance

informatica.com

Contents

Introduction 3

The Three Pillars of Modern Data Governance 4

- 1. Risk and compliance 5
- 2. Data sharing and democratization 6
- 3. Intelligent data observability 7

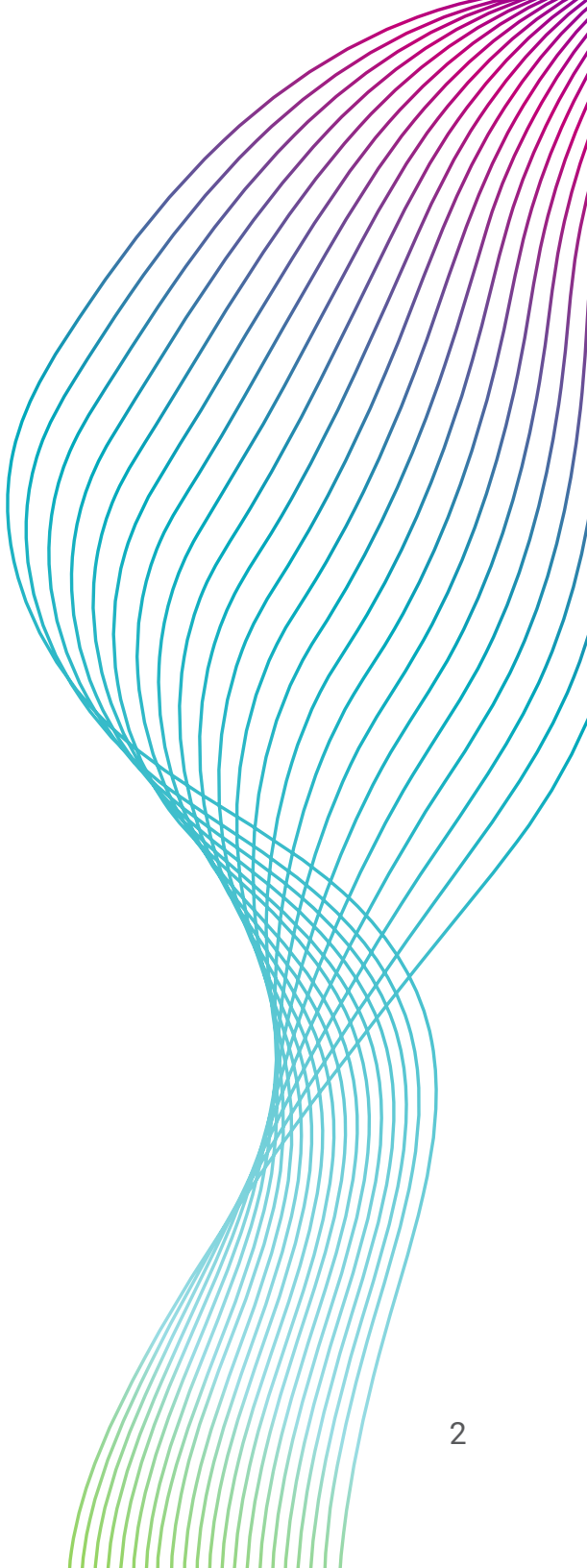
Essential AI-Powered Capabilities that Support Modern Data Governance for Predictive and Generative AI 8

Enabling Modern Data Governance with Predictive Data Intelligence 10

Conclusion 11

Next Steps 12

About Informatica 13



Introduction

In recent years, high-performing enterprises have made it a priority to become more data and AI-driven. The reason for this is simple: companies that have embraced AI technology are enjoying its advantages and outpacing their competitors.[1]

According to the 2022 McKinsey Global Survey on AI, AI is increasingly becoming more prevalent and central for success:

- 50% of organizations have adopted AI in at least one business area
- Nearly 25% of respondents attribute at least 5% of their organizations' EBIT to AI
- 63% of organizations expect to increase their investments in AI over the next three years[2]

Top organizations understand that developing and utilizing AI can help them extract value from data to achieve positive business outcomes. These outcomes include increased productivity, enhanced decision-making and improved customer experience.[3]

Consider BCG's global survey of over 700 business leaders. The study compared the top 25% of these organizations ("data champions" with the highest data maturity scores) to the companies they considered "laggards" (the bottom 25% of organizations). The data champions outperformed the laggards, experiencing revenue growth that was more than twice as high.[4]

But as organizations continue to move from vision to reality, chief data officers (CDOs) face increasing demands from key stakeholders. These stakeholders — who include CFOs, CMOs, sustainability executives and other leaders — want to leverage data and AI to drive business imperatives. And with the development and use of predictive and generative AI, they are seeking answers to questions like these:

- How can we meet compliance and regulatory standards in a changing environment?
- How can we develop and scale ethical and responsible AI practices?
- What is the best way to promote trusted data sharing and democratization to support AI initiatives?
- How can we achieve visibility into the health of our AI data pipelines to scale our digital business?

This eBook will explore how intelligent modern data and AI governance can help data leaders address these challenges. And we'll focus on how enabling AI programs can help accelerate successful business outcomes.

[1]QuantumBlack AI by McKinsey, The State of AI in 2022 — and a half decade in review
[2]Ibid.
[3]PwC AI and Analytics, PwC 2022 AI Business Survey
[4]BCG, Any Company Can Become a Resilient Champion

The Three Pillars of Modern Data Governance

It is widely recognized that **data governance** is critical for the success of any AI endeavor, as it helps to provide the foundation of trusted data and models needed for these initiatives.

However, with the rapid expansion of predictive AI — and, more recently, generative AI — taking a traditional approach to data governance is no longer enough. Navigating this new era requires a modern data and AI governance approach that encompasses three essential pillars:

1. Risk and compliance
2. Data sharing and democratization
3. Intelligent data observability

Let's explore the significance of each pillar in greater detail.

"Having discoverable, trusted and well-managed data is something that's important…it can be an enabler to identify opportunities, whether that means to grow revenue, optimize costs or reduce risk."

Alex Tait
Chief Data and Analytics Officer
BMO Financial Group

The Three Pillars of Modern Data Governance (continued)

1. Risk and Compliance

As the volume and complexity of data have increased over time, so have the challenges related to risk and compliance. Organizations want timely, accurate regulatory reporting even when using data from disparate, siloed systems. There's also the need to address new regulations, reporting laws and mandates concerning data protection and environmental, social and governance (**ESG**).

With the rise of technologies such as predictive and generative AI, there are also emerging AI regulations, such as the Artificial Intelligence Act in the European Union. These have created additional considerations, including the need to:

- **Improve data transparency:** Understanding how AI solutions make decisions can be challenging due to their complexity and lack of transparency. However, organizations must ensure that people understand how their data is used and how AI systems work to increase trust in their decisions.[5] Transparency and explainability are crucial for compliance, too, as regulations such as the European Union's General Data Protection Regulation (GDPR) mandate that organizations be transparent about how they use personal data. They also allow organizations to identify areas where they can improve systems to ensure they work as intended, leading to better, more accurate decisions.
- **Prioritize data privacy:** Organizations need to prioritize data privacy. They often handle large amounts of sensitive data, such as personal information, for purposes like analytics and AI. However, this data is vulnerable to security breaches and unauthorized disclosure of sensitive information. Such incidents can have severe consequences, harming those affected and resulting in hefty penalties. Mishandling sensitive data can also significantly damage the reputation and trustworthiness of the organization.
- **Mitigate bias and drift in data and AI:** Some **AI solutions can be harmfully biased**, leading to unfair or discriminatory decisions. If people cannot trust that AI systems are accurate, they may be less likely to use them (or approve of their use). And this could stifle the adoption of AI and prevent it from reaching its full potential. Organizations can create more reliable and inclusive AI systems by addressing bias and drift.

The modern enterprise must adopt a data governance approach that helps them manage risk exposure. This approach should help them address the need to protect data privacy and reduce the potential for data misuse across complex data landscapes.

In the past, data governance dealt with data access and usage policies, regulatory laws and mandates. Now, however, it's a crucial part of business operations. To manage risk exposure without hindering innovation, a well-balanced approach is necessary. This involves allowing convenient access to reliable data to facilitate business growth.

[5]QuantumBlack AI by McKinsey, Why digital trust truly matters, September 12, 2022

The Three Pillars of Modern Data Governance (continued)

2. Data Sharing and Democratization

To fully leverage predictive and generative AI, organizations will have the greatest success with a strategy that fosters data sharing and democratization. They can do this by making sure that data assets are easily discoverable, trustworthy and accessible to virtually all levels of data consumers. This is critical, as roughly a third of data leaders cite a lack of a complete view and understanding of their data estates as an obstacle to executing their data strategy.[6]

Sharing and democratizing data are crucial for supporting predictive and generative AI. Reliable data helps AI and ML models perform better and enables informed business decisions. Context is vital for data consumers to use AI solutions confidently.

Organizations that encourage collaboration and share information have a better chance of finding new and creative solutions to difficult problems. Sharing data assets, context and insightful analyses is vital to success.

Find	Understand	Trust	Access
Discover data, metadata and AI models across your organization's data landscape in resources, including hybrid cloud systems.	Provide data consumers with context around your organization's data and AI models so they know essential information such as: • What they represent • Where they originate • Who owns them • When and how they have been and can be used	Allow your teams to be confident that they can use your organization's data and AI models to make accurate decisions. Assure data and AI model owners that data consumers are using their data assets appropriately.	Quickly and easily prepare and deliver the right data assets to the right data consumers upon request and approval.

"Organizations that invest in organization-wide data literacy are more than twice as likely to report transformational outcomes across dimensions such as quality of decision-making, innovation, customer experience and more."

Datacamp, "What is Data Literacy? A Comprehensive Guide for Organizations[7]

[6]Deloitte, Fueling the AI transformation: Four key actions powering widespread value from AI, right now. October 2022

[7]Datacamp, What is Data Literacy? A Comprehensive Guide for Organizations

The Three Pillars of Modern Data Governance (continued)

3. Intelligent Data Observability

Keeping track of data processes and pipelines is essential in today's data-driven world. And that is why intelligent data observability is the third pillar in a modern data governance strategy. Data observability allows companies to monitor data usage for data quality issues, improve data flow and help ensure data security. It offers valuable insights into data pipelines for technical and business users and governance teams. And data observability plays an essential role in ensuring data integrity and transparency.

Consider that 55% of leaders report that their organizations have more than 1,000 sources of data. An even larger percentage (91%) predict that the number of sources will increase.[8] These statistics show why many organizations rely on automation and AI to scale their operations, as these technologies can detect **data quality** problems and anomalies at an early stage.

Automated data cleaning and verification using data quality and profiling capabilities can reveal and resolve issues. Data observability features such as threshold alerts and scorecards help track performance and identify pipeline issues for compliance. Reports can be generated for further analysis.

This ensures that high-quality data is accessible and can be shared among various organizations. It's crucial to prioritize data quality — in a recent survey, many tech executives report it as the main obstacle to scaling data solutions, with around 60% citing poor data quality as a significant roadblock.[9]

Business-critical predictive and generative AI models are only as good as the data on which they are trained and fed, so high-quality data is a must. And data observability plays a vital role in meeting this imperative.

Organizations that want to effectively use AI to achieve their objectives need to implement a modern data governance strategy that addresses these three pillars. By adopting this approach, organizations can make intelligent decisions faster and thrive in the data-driven era.

[8]Informatica with Wakefield Research, CDO Insights 2023: How to Empower Data-Led Business Resiliency

[9]McKinsey Digital, How data can help tech companies thrive amid economic uncertainty

Essential AI-Powered Capabilities that Support Modern Data Governance for Predictive and Generative AI

To achieve greater confidence in AI solutions that support business initiatives, organizations can leverage AI and automation for effective risk mitigation, policy compliance, data democratization and sharing and data observability.

AI and automation are pivotal for organizations implementing modern data governance to scale their digital businesses efficiently with predictive and generative AI. Their integration into data governance tools empowers enterprises to work with trusted data and models that drive reliable business outcomes. Here are six areas where automation, AI and machine learning (ML) are essential for supporting modern data governance.

1. **Metadata Scanning and Extraction**
   Advanced data governance solutions can automatically extract metadata from various sources such as SQL scripts, data warehouses, data lakes, ETL tools and business intelligence platforms. These capabilities offer comprehensive intelligence about data relationships in complex and fragmented data landscapes. By efficiently discovering and understanding data throughout the enterprise, organizations can improve data exploration and productivity. This enables businesses to scale and expand insights through AI initiatives.

2. **Semantic Search**
   Modern data governance solutions leverage ML-driven semantic search capabilities. Organizations can easily search across metadata assets, including tables, files, BI reports and business glossary terms. The ability to search using business-friendly language allows data professionals to find relevant data quickly. This significantly improves the discoverability and accessibility of data supporting predictive and generative AI, enabling faster, more agile, data-driven decision-making.

3. **Data Classification and Curation Automation**
   Using AI to categorize and classify data can save organizations time and money. AI can link relevant assets and add business context to data, which aids data consumers in comprehending their data assets more effectively. With the help of automation, productivity can be increased, freeing up time for more valuable tasks.

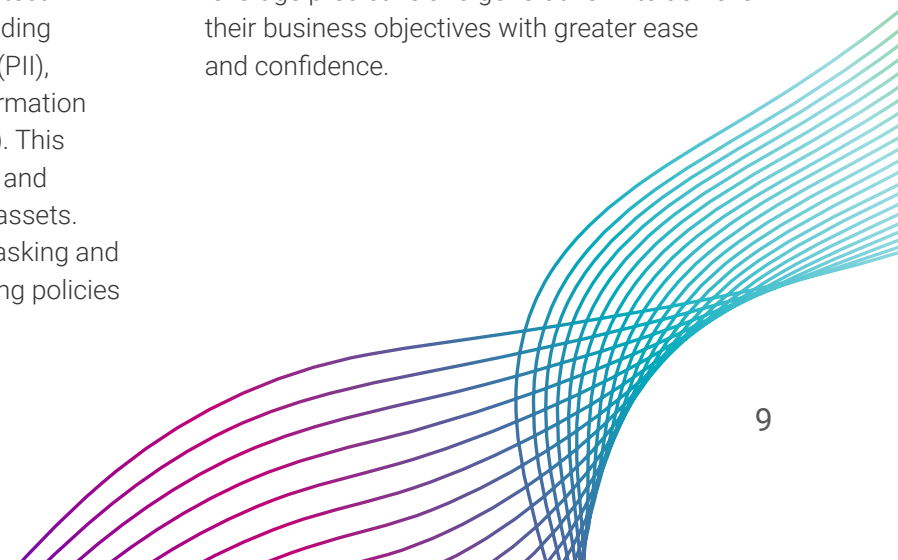
Essential AI-Powered Capabilities that Support Modern Data Governance for Predictive and Generative AI (continued)

4. **Automated Data Quality and Data Observability**
   Modern data governance solutions help maintain data quality by automating certain tasks. These solutions can analyze data, detect irregularities and provide insights to create and implement data quality rules. Some advanced solutions even use natural language processing to generate and apply rules without the need for technical staff. These capabilities allow organizations to quickly identify issues with AI data pipelines and remove obstacles to enable trustworthy data use to scale.

5. **Inferred Relationships**
   Organizations can use machine learning to discover connections in their data without needing to do it manually. This includes automatically identifying primary or unique keys and recommending joins across structured datasets, which enables analysts and data scientists to find relevant and complementary information efficiently. When needed, AI-powered lineage inference can also fill in gaps in data lineage. This makes it easier for organizations to follow regulations and meet objectives related to trustworthy AI, data quality, data privacy and security.

6. **Sensitive Data and Data Protection Policy Management**
   Organizations need data governance solutions that can automatically detect and categorize sensitive data, including personally identifiable information (PII), electronically protected health information (ePHI) and intellectual property (IP). This helps organizations create policies and associate them with relevant data assets. Automated techniques like data masking and data access controls make enforcing policies and protecting sensitive data easier. These capabilities help ensure that data access and use comply with policies and regulations, reducing the risk of data exposure.

These AI- and ML-driven capabilities help organizations manage, share and use their data more effectively. This results in better decision-making, increased operational efficiency and improved policy compliance. By using AI in data governance solutions, businesses can navigate the complexities of the digital landscape and leverage predictive and generative AI to achieve their business objectives with greater ease and confidence.



Enabling Modern Data Governance with Predictive Data Intelligence

Having a successful AI-powered data governance solution involves using predictive data intelligence. This means using metadata intelligence to give recommendations and automate processes. By understanding how businesses use their data, organizations can improve their governance strategies and get better results. Automation and AI are important as more people use data. They can help connect data producers with data consumers more effectively and reduce the workload for data stewards and IT teams. These technologies can streamline the manual work of finding, validating and delivering data and AI models.

To have reliable data, businesses should have a self-service **data marketplace**. This will let business users easily request and access the necessary data and AI models. With this system in place, users will have the power to obtain the data they need quickly and efficiently.

To be efficient, a self-service marketplace must integrate a delivery engine with tools for data quality, as well as for cataloging, governing and mastering data. This integration helps users access relevant and reliable data more easily.

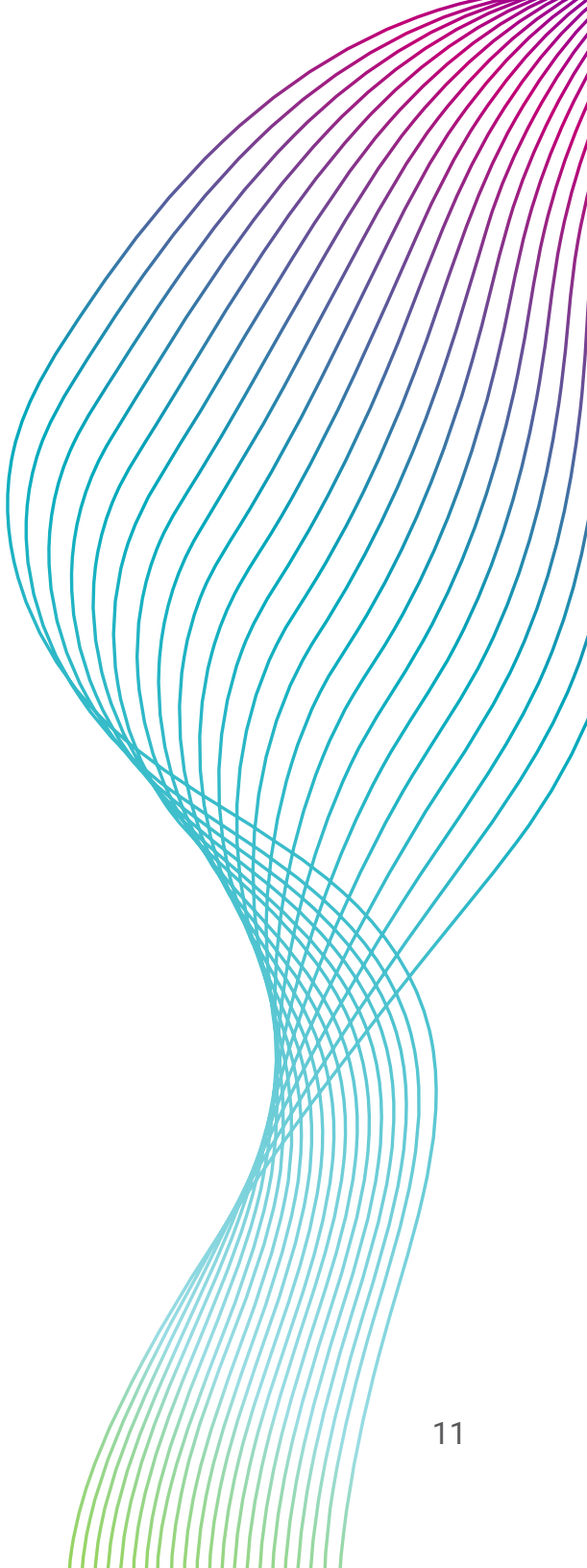
Connecting data intelligence to delivery lets data consumers easily access the information they need for predictive and generative AI projects. This helps the organization create new business opportunities, increase revenue, improve operations, lower risk and enhance customer experiences and loyalty programs.

Conclusion

To succeed in the future with your data, it's important to take a modern approach to data governance. This involves three pillars: managing risk and compliance, sharing data and ensuring data observability. By implementing these methods, you can leverage digital transformation to grow your digital business.

Using automation and AI can simplify the process of managing data in a responsible and ethical manner. It can also enhance data literacy and accessibility for all members of your organization while ensuring high-quality data for AI models.

To truly support modern data governance, it's essential to have AI and ML-driven capabilities. AI can assist organizations in governing and sharing data more efficiently from metadata scanning and extraction, to data quality and data observability. By utilizing predictive data intelligence as part of data governance, you can make informed decisions, improve efficiency, meet compliance requirements and achieve reliable business outcomes.



Next Steps

Ready to bring your data to life? Learn more about how we can help your business grow and scale with modern data governance.

eBook: How to Use Data Intelligence to Drive Better Business Decisions

Whitepaper: Data Observability: The Key to Successful Data and Analytics

Experience our data governance solution in action with this interactive demo.

About Us

Informatica (NYSE: INFA) brings data to life by empowering businesses to realize the transformative power of their most critical assets. When properly unlocked, data becomes a living and trusted resource that is democratized across the organization, turning chaos into clarity. Through the Informatica Intelligent Data Management Cloud™, companies are driving bigger ideas, creating improved processes and reducing costs. Powered by CLAIRE®, our AI engine, it's the only cloud dedicated to managing data of any type, pattern, complexity or workload across any location — all on a single platform, with a simple and flexible consumption-based pricing model. **Informatica. Where data comes to life.**

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
twitter.com/Informatica

CONTACT US

IN19-4614-0823