Data Governance
for the AI Enterprise

Table of Contents

*Disclaimer: This document addresses the current and future functionality of Data Cloud. Any unreleased services or features referenced here are not currently available and may not be delivered on time or at all. Customers should make their purchase decisions based upon features that are currently available.

Data Governance in the AI Era: The Strategic Imperative for CIOs and CDOs

Today, CIOs and CDOs are at the forefront of two transformative forces reshaping the business landscape. One is the explosive growth of data, with the world expected to generate and store nearly 300+ zettabytes of structured and unstructured information by 2027. This exponential volume and diversity of data require meticulous management to unlock its true value, which is critical to drive growth and enhance strategic decision-making for organizations.

> **291ZB of data will be created, consumed, and stored in the world by 2027.**
>
> **Worldwide IDC Global DataSphere Forecast,** 2023-2027

On the other side is the rapid advancement of AI, which over 64% of CEOs aim to make a central pillar of their business strategy. As AI capabilities continue to evolve, it relies on vast sources of structured and unstructured data to power large language models (LLMs), applications, and agents that can unlock business efficiency and growth.

The convergence of these trends underscores the critical importance of robust data governance. As data volumes grow and AI becomes integral to business strategies, having strong data governance is not merely a regulatory requirement but a strategic necessity. Effective data governance ensures that your data is secure, accurate, and accessible for your organization to use to unlock the benefits of AI and achieve a sustainable competitive advantage.

However, many organizations struggle with data governance because it's not a 'set it and forget it' task. The ever-evolving nature of data–constantly being generated, updated, and used–demands ongoing oversight to uphold its quality, usability, security, and privacy through continuous reviews. And like a living organism, data governance must also evolve to keep pace with the difficult and complex regulatory landscape so that you can protect the organization's most valuable asset: customer trust.

In this white paper, we'll delve into the principles of effective data governance and the key pillars that are necessary to build a successful strategy. You'll discover the organizational structures required to implement a governance strategy that balances innovation with control. We'll also explore how solutions such as Data Cloud and the Salesforce Platform can support your governance strategy by providing the tools to unify your data sources and securely manage them across the enterprise. Together, these principles and technologies will empower your organization to navigate the complexities of data governance in the AI era with confidence.

What is Data Governance?

Data governance refers to the rules and policies that dictate how information is securely collected, managed, stored, and used within an organization. Gartner defines it as "the specification of decision rights and an accountability framework to ensure appropriate behavior in the valuation, creation, consumption, and control of data and analytics." At its core, data governance can be distilled into four fundamental questions:

What data do you have?

Where is your data located?

Who needs to manage your data?

Is your data sensitive?

As Kendra Chandra, Head of Data Governance and Quality at Salesforce, puts it, "These are foundational questions, and once answered, they guide how you implement governance within the tools and solutions your organization uses." These questions are not only fundamental to your organization's data governance strategy but also crucial in shaping its design and implementation across the following five core pillars:

Data Access
Who needs access to your data? Who is authorized to view, edit, or use it?

Data Accuracy
What safeguards ensure your data is clean, reliable, and up-to-date?

Data Privacy
How do you handle data related to customer privacy, particularly personally identifiable information (PII)?

Data Security
What policies and procedures protect your data from external and internal threats?

Data Retention
What legal and business requirements dictate data retention? How can data be safely disposed of when it's no longer needed?

Why Data Governance Matters

Trust is the cornerstone of every successful business, and in today's AI-driven landscape, it's more complex than ever. Trust extends beyond merely securing data–it encompasses its accuracy, reliability, and ethical use, especially as AI becomes more integrated into business operations. Without a strong foundation of trust, customer confidence and loyalty can quickly erode.

92% of analytics and IT leaders say the need for trusted data is higher than ever.
Salesforce: The State of Data and Analytics Report

As AI technology advances, global lawmakers are under increasing pressure to establish regulatory guardrails around the security, management, and accuracy of data used in AI applications. Initiatives like the EU's AI Act are designed to ensure the safe and ethical development and use of AI. These regulations are only becoming more stringent, and companies must be diligent in adhering to them. As a result, promoting strong data governance and security has become the top priority for Chief Data Officers according to Salesforce research.

It is under this rapidly changing landscape that having a robust data governance strategy is essential. It ensures that your structured and unstructured data sources are managed and protected in compliance with both current and future regulations. By implementing rigorous data governance policies, businesses can safeguard the integrity of their data, which can minimize risks like inaccurate AI outputs or regulatory penalties. More importantly, data governance solidifies trust with customers and stakeholders, reassuring them that their data is handled responsibly.

A strong data governance strategy, which focuses on the key pillars of access, accuracy, privacy, security, and retention enables organizations to confidently scale their AI initiatives. It not only protects against risk but also positions organizations to unlock the full potential of AI by ensuring that the data feeding these systems is trustworthy. In turn, this trust translates into sustainable business growth and long-term success.

85% of analytics and IT leaders use data governance to ensure and certify baseline data quality.
Salesforce: The State of Data and Analytics Report

Who Advances Data Governance

Implementing a strong data governance strategy isn't the responsibility of a single person or function – it's a unified commitment where key stakeholders from every department collaborate to protect and manage your data. Below are the essential roles for a robust data governance strategy:

1. Executive Sponsor

Typically the CIO, CDO, and/or IT/Security Lead, this leader drives the data governance program, ensuring it aligns with organizational goals. They secure necessary resources– budget, personnel, and technology–and advocate for the program's importance across the C-suite. Their role is crucial in maintaining momentum and ensuring data governance remains a strategic priority.

2. Data Stewards

Data Stewards ensure that data quality standards are met within their business units. They enforce governance policies, address issues proactively, and transform data into a strategic asset, making stewardship a shared responsibility. Their hands-on approach ensures data is usable, trusted, and consistently improved.

3. Data Custodians

Data Custodians manage specific data sets or systems, ensuring proper access and compliance. They align data usage with organizational standards and, overlap with Stewards in safeguarding data quality and accessibility. Their deep understanding of both data and business needs helps bridge gaps between technical and strategic goals.

4. Data Architects

These technical experts build and maintain the infrastructure for data governance. By embedding data quality into the architecture, they ensure secure, seamless data collection, storage, and integration across the organization. Their designs lay the foundation for scalable, reliable data operations that support long-term growth.

5. Data Scientists

Data Scientists and Analysts rely on high-quality data for actionable insights. They collaborate with Stewards and Custodians to ensure data meets the standards needed for accurate analysis and decision-making. Their work turns raw data into valuable business intelligence, guiding critical decisions and innovations.

Salesforce Platform and Data Cloud: The Trusted Foundation For AI Innovation

AI-driven enterprises seeking to build AI innovations on a foundation of strong data governance require a secure, scalable platform that not only simplifies the enforcement of data governance policies but also provides the tools to seamlessly integrate AI capabilities across business workflows. Salesforce addresses this dual challenge with Agentforce – a groundbreaking solution that empowers organizations to deploy intelligent, autonomous AI agents across their operations.

Building AI on a Foundation of Trust and Data

Agentforce's transformative potential lies in its foundation: the Salesforce Platform and Data Cloud.

The **Salesforce Platform** acts as the backbone for creating, deploying, and managing intelligent AI agents and applications across the enterprise. With native security, privacy features, and low-code development tools such as the new Agent Builder, the platform gives organizations the unique flexibility and control to build and scale AI agents and applications while ensuring that the data that powers these solutions is managed in adherence to both current and future industry regulations.

Data Cloud is the heartbeat of the Salesforce Platform. As a hyperscale data engine, it breaks down silos by harmonizing diverse data formats–both structured and unstructured–into comprehensive customer profiles that are stored as metadata within a unified data foundation. Fully integrated within the Salesforce Platform, this metadata flows seamlessly into any Salesforce applications, based on set data governance policies. This gives every team a 360-degree view of the customer that is needed to power AI agents with trusted customer context and drive automation and analytics across every touchpoint – without compromising data security.

The Salesforce Platform and Data Cloud provides organizations with the secure foundation to fully unlock AI's transformative potential in adherence to the key principles of data governance–access, accuracy, privacy, security, and retention. In the following sections, we'll use this data governance framework to dive deeper into some of the key technology capabilities that can be utilized to build a trusted AI-driven enterprise.
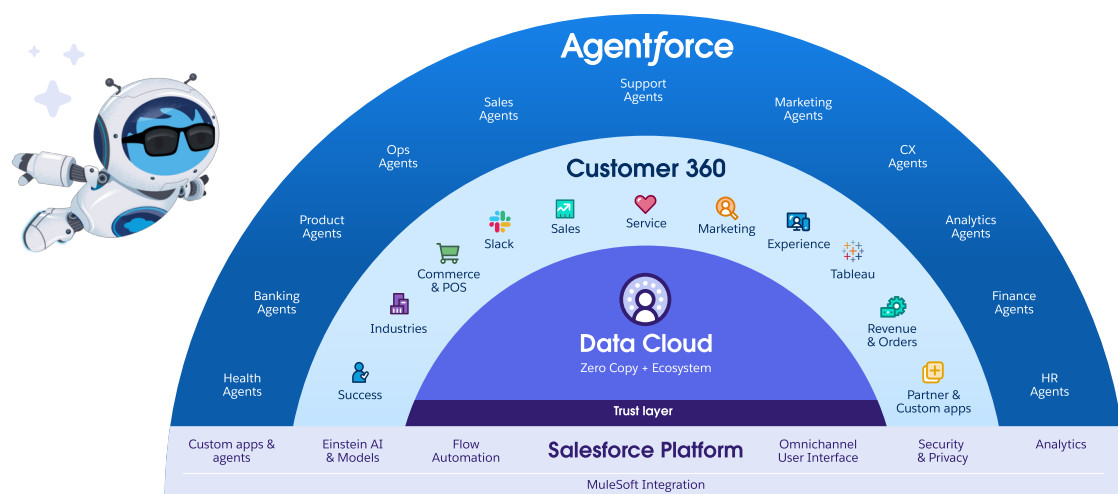
The Five Pillars of Effective Data Governance

1 **Tighten Data Access**

Why this matters: Controlling who can access your structured and unstructured data is fundamental to effective data governance. Tightening data access involves defining and enforcing permissions to ensure that only authorized individuals and autonomous AI agents can view, edit, or use sensitive information. This not only safeguards your data but also enhances accountability and compliance.

Here's how to get started with Salesforce:

Access Governance:
- **Centralize Control Over User Permissions:** Data Cloud centralizes control over user access, ensuring only authorized users can access sensitive data. Integrated with the Salesforce Platform, permissions are managed through Profiles and Permission Sets, with options for both out-of-the-box updates and custom configurations. Security Center further enhances these capabilities by allowing you to easily audit and manage permissions across all Data Cloud instances from a single pane of glass.
- **Deploy Granular Access Policies:** Admins can create detailed access policies at the object, field, and row levels. This fine-grained control ensures that only authorized users, including AI, can interact with specific data, enhancing security and compliance throughout the organization.

Data Spaces:
- **Organize and Secure Data by Category:** Data Spaces are logical partitions within Data Cloud that allow for the efficient organization of data based on departments, brands, or regions. By categorizing structured and unstructured data into distinct spaces, you can ingest information from multiple sources simultaneously while maintaining focused access and compliance. Each Data Space enables identity resolution, insight generation, and segment creation, ensuring users only access data within their designated space, thus protecting sensitive information and enforcing compliance.

Private Connect for Data Cloud:
- **Ensure Secure and Compliant Data Transfers:** Boost data security by enabling federated, bi-directional access to data between Data Cloud and customers' private and public cloud networks.

② Improve Data Accuracy

Why this matters: Accurate data is crucial for making reliable decisions and maintaining operational efficiency. Improving data accuracy means implementing processes and tools to cleanse, validate, and standardize your structured and unstructured data sources, ensuring it is correct, complete, and consistently up-to-date across all systems so that they can be used to power trusted AI.

Here's how to get started with Salesforce:

Ingest Data From Any Source:
- **Stream Data into Data Cloud:** Ingest data from various external sources into Data Cloud through custom data streams, ensuring accurate integration into Data Lake Objects (DLOs).
- **Federate Data with Zero Copy:** Data Cloud allows you to virtually access and integrate structured and unstructured data from external data lakes and data warehouses such as AWS, Databricks, Snowflake, and Google BigQuery without moving or duplicating it, via the Zero Copy Partner Network (ZCPN). This method ensures real-time access to a single source of truth, which minimizes the need to track and audit multiple copies of data that are spread across your systems.

Unify Data Across Systems:
- **Organize with Customer 360 Data Model:** Structure imported data into Data Cloud's Data Model Objects (DMOs) across business areas for consistent integration and governance.
- **Unify Data Streams with AI Models:** Use AI-powered models to consolidate data streams, insights, and external inputs into unified DMOs, ensuring consistency and accuracy across departments.
- **Track Data Lineage:** Maintain data integrity by tracking how data moves and transforms across systems. Use lineage tracking to tag source information and ensure accuracy through detailed comparisons.

Transform Data for Accurate Results:
- **Apply Transformation Methods in Data Cloud:** Implement row-based formulas, batch transformations, and streaming transformations to standardize and cleanse data at ingestion. These methods automate tasks like record matching, insight generation, and real-time updates, ensuring continuous accuracy.

Monitor Data Health in Real-Time:
- **Automate Alerts and Tracking:** Once data is ingested via a Data Stream, you can quickly view data freshness on your Data Cloud homepage, including the last refresh and any data sources with zero density. This gives you confidence in the accuracy and relevance of your data. Additionally, you can monitor the status of your streams in real time, instantly identifying any ingestion failures.To access real-time alerts, Data Cloud's integrations with the following Platform capabilities gives you key insights on the health of your data:
    - **Salesforce Flow:** Automate notifications for data errors and failures to stay informed about critical issues.
    - **Security Center:** Manage permissions and monitor data insights through a centralized dashboard.
    - **Event Monitoring:** Track data movements comprehensively to detect and respond to unexpected changes in real-time.

Manage Metadata for Consistency:
- **AI Tagging and Classification:** Automate the labeling and organization of structured and unstructured data using AI-recommended tags that are aligned with your data governance policies. This ensures data consistency and allows both AI agents and users to easily search and analyze data at scale.
- **Streamline Metadata with Data Kits:** Simplify metadata management using reusable schemas and organized kits, allowing updates without disrupting runtime operations.
- **Data Cloud Catalog (Metadata Graph):** Leverage an AI-powered unified view of how data is used and related across systems, improving accuracy and governance through better metadata management.
- **Access Metadata API:** Data Cloud utilizes Salesforce Metadata API which allows developers to manage and transfer org configurations, such as custom objects and layouts, between orgs during development. It handles metadata but doesn't interact with business data.

3 **Implement Data Privacy**

Why this matters: Data privacy is essential for maintaining customer trust and complying with government and industry regulations. It involves protecting personal information, managing consent, and adhering to data protection laws to secure customer identities and sensitive data.

Here's how to get started with Salesforce:

Define and Enforce Consent and Preference Policies:
- Data Cloud enables you to gather and unify consent and preference signals from various sources into a single view. The built-in Privacy Data Model tracks and manages privacy preferences, with customizable options tailored to your organization's data governance standards. Key privacy subjects include consent for communication channels, contact points, and data usage purposes. Examples of available Privacy DMOs include Authorization Form Consent, Communication Subscription, Contact Point Consent, and Engagement Channel Type Consent.

Manage Consent Across Different Levels:
- Consent is managed in the data model on four levels: global consent (1), engagement channel consent (2), contact point consent (3), data use purpose (4):
  - **Global Consent:** General consent for communications.
  - **Engagement Channel Consent:** Consent for specific channels (e.g., email, mobile).
  - **Contact Point Consent:** Permission for specific contact points (e.g., work email only).
  - **Data Use Purpose Consent:** Consent tied to specific purposes (e.g., product updates).

Streamline Consent Policies with APIs and Preference Manager:
- The Consent Write API allows you to update consent across multiple records with a single API call, syncing data and populating the Consent Data Model. It also writes consent directly to the Data Cloud Individual record.
- Preference Manager, part of Salesforce's Privacy Center allows you to manage customer preferences and consent across multiple systems, including Data Cloud. It enables organizations to capture consent for various purposes–such as marketing, data sharing, and AI usage–while ensuring compliance by restricting data processing when necessary. This tool provides flexibility in managing consent across clouds and will continue to evolve with expanded capabilities in the future.

4 Increase Data Security

Why this matters: Ensuring data security is crucial for protecting against breaches and unauthorized access to your structured and unstructured data. Strengthening data security involves implementing comprehensive safeguards such as data masking, encryption, and threat detection to defend your data from internal and external threats.

Here's how to get started with Salesforce:

Protect Sensitive Data in Sandboxes:
- **Secure Your Application Testing Environments:** Data Mask for Data Cloud Sandboxes allows you to safely test Data Streams, federated data from zero-copy integrations, Data Models, and Calculated Insights without exposing sensitive data. Data Mask anonymizes structured and unstructured data to prevent unauthorized access to PII from unauthorized individuals and autonomous AI agents.
- **Safe Testing and Training:** With Data Cloud Sandboxes, you can safely validate and train users on your Data Cloud solutions. Use familiar tools like Command Line Interface, Change Sets, and DevOps Center to promote changes from isolated Sandboxes to production, maintaining a secure development lifecycle.

Enforce Robust Security Measures:
- **Encrypt Data at Rest:** Data Cloud provides robust data protection with encryption at rest. Data that is masked using Classic Encryption remains masked in Data Cloud unless viewed by users with specific permissions. For more advanced encryption, Shield Platform Encryption for Data Cloud, allow you to encrypt data stored in the database and Data Cloud.
- **Encrypt Data with Customer Managed Keys (CMKs):** Platform Encryption for Data Cloud uses Customer-Managed Keys (CMKs) to encrypt sensitive data for enhance security & compliance. Utilizing CMKs gives you full control over your encryption keys in compliance with internal and external regulations while preserving the power of the platform to build amazing experiences on it.
- **Hashing and Salting:** Batch and streaming transformation processes within Data Cloud can include hashing (SHA256) and salting (MD5 + CONCAT) to further protect data during processing.

Monitor Unusual Data Cloud Activity
- Use Shield:Event Monitoring to track key Data Cloud UI interactions including page views, report exports, and API calls to understand how your team uses the platform. Event Monitoring Plus offers deeper insights with CRM Analytics to help understand Data Cloud page loads. These capabilities are useful to detect unusual activity and mitigate threats with Data Cloud that can cause data loss.

⑤ Ensure Data Retention

Why this matters: Data retention policies define how long data is kept and how it is securely disposed of when no longer needed. Clear retention schedules ensure compliance with legal and business needs. Retaining data is also critical for protecting against external disasters, ensuring business continuity, and minimizing the risk of permanent data loss.

Here's how to get started with Salesforce:

Standardize Data Subject Rights:
- Data Cloud provides built-in support for critical data subject rights, including Right-to-be-Forgotten, Restriction-of-Processing, and Data Export and Portability. This supports compliance with data protection regulations and facilitates effective data management.

Create/Automate Retention Policies with Privacy Center:
- For a more comprehensive approach, Privacy Center enables you to create and manage data retention policies with ease. It automates the execution of policies and offers robust tracking and auditability features, which makes it easier for you to maintain compliance across the Salesforce ecosystem.

Enforce Zero-Retention with the Einstein Trust Layer
- While Data Cloud surfaces structured and unstructured data to power predictive and generative AI, the Einstein Trust Layer enforces a Zero-Retention policy to ensure third-party models never store or use your data.

Safeguard Against Data Loss with Hyperforce:
- Hyperforce ensures data continuity for Data Cloud by replicating and storing data across multiple regional data centers. This distributed architecture safeguards against data loss from disasters or system failures, providing redundancy and ensuring data remains secure and accessible even during unexpected events.

Conclusion

In today's AI-driven world, where data is the backbone of organizational success, robust data governance is essential. A proactive and comprehensive approach transforms data from a potential risk into a strategic asset–strengthening compliance, safeguarding customer trust, and driving AI innovation. Discover how Salesforce provides the trusted foundation to responsibly utilize your data and the tools that empower you to securely build AI innovations that drive growth and deliver exceptional customer experiences.

Learn more about Agentforce

Discover Data Cloud

Explore Salesforce Platform Security and Privacy

Contact an expert to learn more