

BEING RESILIENT

HOW
INSURERS CAN
EFFECTIVELY
MANAGE
THEIR DATA IN
THE WAKE OF
DATA PRIVACY
REGULATIONS

Infosys®
Navigate your next



BEING RESILIENT. THAT'S LIVE ENTERPRISE.



INSURANCE

PROVIDE EXPERT GUIDANCE



Apply Now

Learn More

Insurance is a complex industry with multiple business lines that covers different types of risks and challenges. Today, customers are seeking new types of cover and new ways of being serviced. Insurers can only satisfy these demands if they can efficiently, intelligently, and effectively leverage their data assets.

Insurers have a wealth of data about customers, risks, losses, policies, and claims. However, to harness the value of

their data, they have to accurately describe the business relevant data, trace it to its origin and be able to trust it while assuring customers that data is secure. The recent introduction of regulations such as CCPA, NYDFS, GDPR, and PIPEDA has added pressure on insurers to find a holistic solution that addresses these challenges.

If an organization has robust data governance policies and access management process, effective data

catalog and metadata management solution, and a clear understanding of data lineage, then it is far simpler and faster for it to comply with privacy regulations. This is also accelerate its digital transformation journey and extract value faster and more reliably than its peers. The foundation of these attributes is effective data management, and this article describes how to achieve it.

A Time of Rapid Change in Insurance

Rapid digitization, social media and disruptive innovations such as the Internet of Things and artificial intelligence are causing the insurance industry to go through fundamental changes in product design, channels to market, and customer management. The top five trends impacting insurance companies are:

1. Hyper-connected customers and devices. Customers now expect to have omnichannel experiences and multiple touch points with a company, in which the data and experiences are consistent.
2. InsurTech and the convergence of analytics and AI/ML. Many InsurTechs are using analytics, artificial intelligence, and machine learning to solve specific issues in the insurance business, such as developing a new product area or improving customer experience.
3. Modernization of insurance technology. Many insurers have long used home-grown IT systems, but now they are buying off-the-shelf systems, moving to the Cloud, and leveraging third-party data.
4. An explosion of data-driven products. These include usage-based insurance, virtual claims, and embedded home insurance, which serve customers who increasingly expect tailored products.
5. Data privacy regulatory reforms, such as CCPA, NYDFS, GDPR, PIPEDA. Insurers are faced with the costs and challenges of implementing these measures, as well as the threat of higher penalties for data breaches.

To address these challenges, insurers are embarking on strategic programs such as:

1. Customer experience transformation. Providing a differentiated customer journey and experience across channels with a high degree of personalization in real-time.
2. Product modernization. Designing tomorrow's insurance products for



connected consumers, such as hyper-personalized policies, cyber insurance, autonomous car insurance, and AI device insurance.

3. Realizing modern platforms in the Cloud leveraging InsurTech, off-the-shelf policy administration products, and API solutions.
4. Data landscape transformation. Driving digital transformation and automation through modernized data platforms, leveraging the Cloud, big data, APIs, and AI/ML.
5. Monetizing data with new service and business models that drive behavior, with usage-based or demand-adjusted pricing.

The success of these programs depends heavily on Insurer's ability to leverage its data assets which are reliable, transparent and available on demand.

Privacy regulations require Insurers to comprehensively answer questions about customer data, such as:

- What personal information does the organization store?
- Where is the personal information stored?
- How is the personal information processed?
- How long is the personal information needed?

To address the privacy regulations and strategic program needs, Insurers must embark on a journey to transform their data management capabilities holistically. However, there are numerous architectural challenges, which need to be accounted for as Insurers build their solution, including: Multiple legacy data assets, with inconsistent data and data quality issues.

1. Inconsistent data definitions, requiring human intervention to resolve.
2. Incomplete data lineage, jeopardizing data privacy compliance with regulations in multiple geographies.
3. Siloed data quality management, and an absence of streamlined processes to assure data quality, rendering it untrustworthy.
4. Multiple reporting and analytics tools and platforms.
5. Product-based administration systems that have their own data models.
6. Migration to the Cloud, requiring a strong understanding of where data lies, how it flows and is governed, which is not possible without data catalogs, data lineage, governance, and privacy.

Next-Generation Data Management - Principles

Leading insurers are adopting a holistic data management discipline to build a

strong foundation to accelerate market differentiation while addressing the privacy needs.

The next-gen data management solution should adopt privacy by design approach

while embracing self-service driven consumption, delivered through a hybrid scalable architecture. Overall guiding principles should include those in Table 1.

Solution Principle	Description
Privacy by design	Address Rights of consumers (Right to Know, Access, Delete) and Consent (Right to Opt out/Opt in) requirements Data Inventory and lifecycle management Data Protection and Breach handling Change Management & Organization Readiness
Consumption-Driven	Governance not only for data that feeds regulatory reports, but also analytics, discovery, reporting, Cloud, and APIs.
Self-Service Enabled	A self-service model, especially for data analysts and data scientists to easily discover, curate, accelerate, and share data for specific use-cases without being dependent on IT. Automated data quality rules using a self-service business rule engine.
Agile	Ownership of insights and business decisions derived from them. Agile data governance that focuses on the value produced, and proactively measures long-term benefits.
AI/ML Powered	Smart data catalog using AI/ML algorithms for automated cataloging of business & technical metadata, understanding of dark data, data rationalization, data quality, and more. Leverage emerging technologies such as chatbots, mobile apps, and blockchain for self-service governance.
Hybrid Cloud-Based	Governance policies, procedures, and accountabilities extended to on-prem, multi-cloud, and hybrid-cloud architectures. Preference for Cloud-native tools, especially for data security and access controls.

Table 1: Next-Gen Data Management Solution Principles

Next-Generation Data Management - Components

The high-level components of a next-gen data management solution include:

Data Governance

Data governance ensures that the most important data is easy to access, understand,

and use. While enterprises need to build data governance policies, communities, access, and stewardship to support consumption patterns such as reporting, analytics, data science, partner collaboration and data marketplaces, but data governance also need to account for privacy needs such as Consumer Right to Know, Access, Delete or Consent such as Opt out/Opt etc. The data governance policies should

include data protection and breach handling procedures as well.

Leading product tools such as Informatica Axon, Collibra Data Governance, Talend Data Governance and IBM Infosphere Data Governance are enhancing their products to incorporate capabilities to self-manage governance including privacy and security needs.

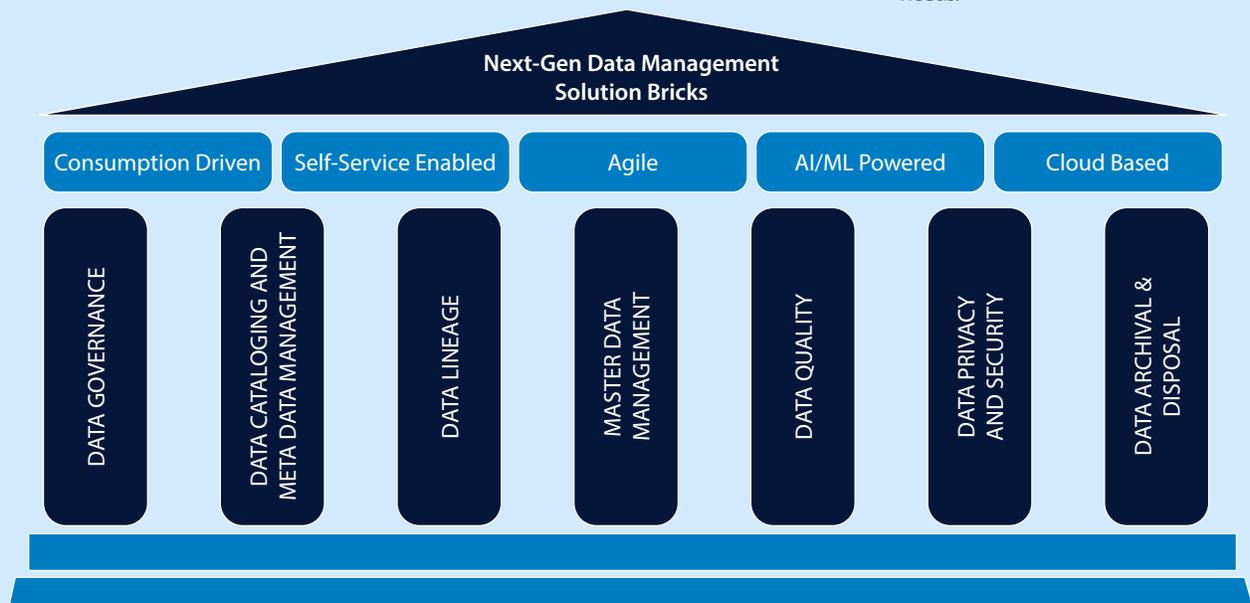


Figure 1: Next-Generation Data Management Solution Bricks

Data Catalog and Meta Data Management

As data sources and consumption points kept increasing, data catalog became a necessity, and privacy regulations made it unavoidable. Personal data discovery, classification for personal, situational or sensitive data are core pillars to build a privacy compliant ecosystem. The privacy needs extended the data catalog scope to go beyond traditional data warehouse and analytics apps to all source systems to provide the context for privacy and governance.

Data catalogs that are tightly integrated with a governance platform will provide full access to relevant data which in turn will help insurance companies to comply with regulations and policies. Integrated product suites from Informatica such as Axon and EDC, or from Collibra such as Collibra Data Governance and Data Catalog, or from Talend such as Talend Data Governance and Data Catalog, or from IBM such as Infosphere Data Governance and Data Catalog, are thus becoming popular implementation choices.

Irrespective of tool suites, next-gen catalog and metadata management must strive to be self-service oriented, agile, and Cloud-based. Smart cataloging using AI/ML algorithms for data discovery, data matching and data tagging will reduce overall implementation efforts by 80-90%. AI/ML based routines can interpret data patterns and enable automatic tagging of business metadata. Crowd-sourced annotations and user comments will help gather tribal knowledge and establish trust in the data.

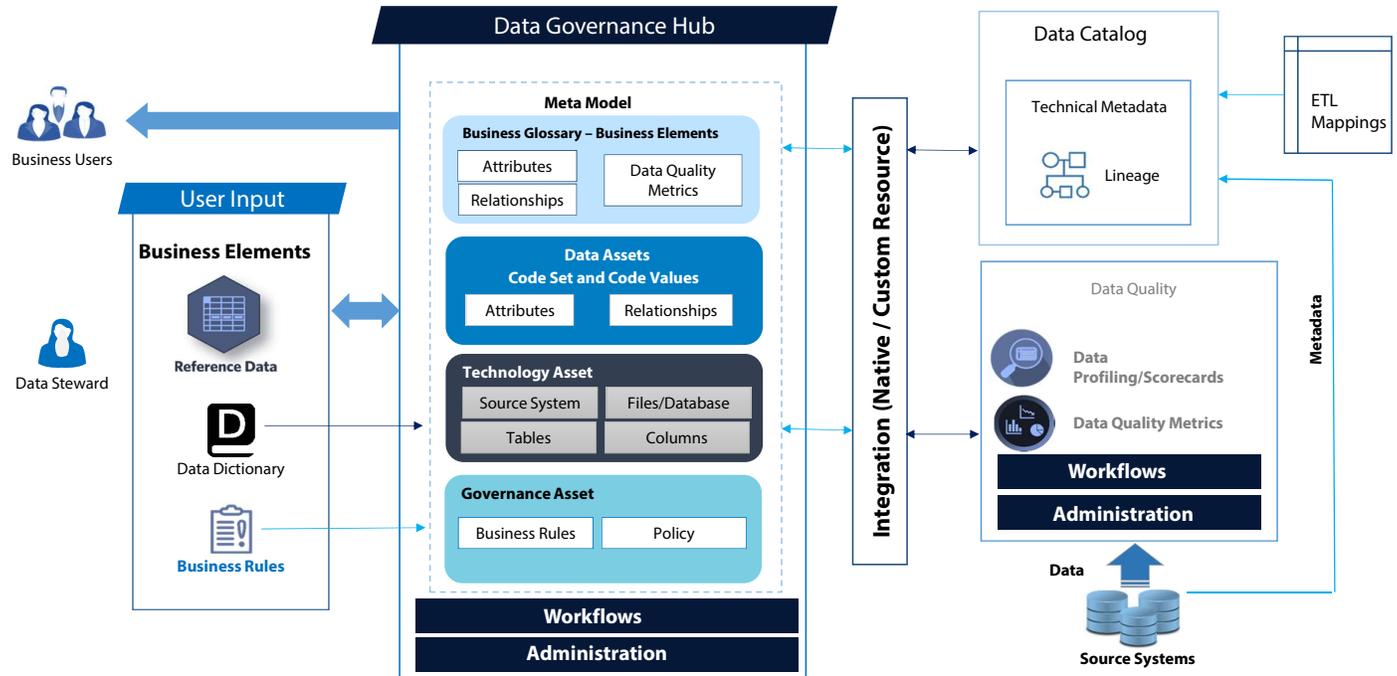


Figure 2: High-Level Architecture of an Integrated Data Management Solution

Infosys helped a large P&C insurer in the US Midwest to implement data governance, data cataloging, and metadata solutions using Informatica Axon and EDC in agile methodology. The integrated data governance and catalog solution took 60—70% less effort to implement than would have been the case for separate applications.



Data Lineage

Regulatory compliance requirements such as Customer's right to know, delete, or Opt-Out etc. have necessitated the enterprises to have a complete trace of data across systems so that the customer request can be serviced in totality. Data lineage is critical in doing that.

Large product vendors are including data lineage capabilities within their data catalog products. However there are niche products such as Manta that offer dedicated capabilities for lineage across data assets whether in on-prem platforms, data warehouses, or data lakes in the Cloud. Adoption of these solutions are becoming mainstream due to the expanded scope of privacy needs

Best practice involves building a graphical multisource, multifamily lineage with Graph DB. Automated lineage extraction from the

stored procedures, ETL codes, database tables, or programming languages is critical to building end-to-end lineage. Lineage metadata in graph format and support for ontology-based search with NLP and AI/ML provides an edge in creating a comprehensive lineage map. GRAKN and Neo4j are some of the tools commonly used to enable these capabilities.

Master Data Management

Data Privacy is concerned about protecting individual's data rights. MDM solutions are supposed to be the guardian of individual's data. While majority of the Insurers have built MDM solutions, but many lack coverage from all LOBs, causing lack of visibility of individual's data thereby seriously jeopardizing compliance ask.

Though many are implementing various facets of MDM, they tend to focus

on incremental efficiency and ignore the potential benefits of strategically modernizing the MDM solution to accommodate tomorrow's consumption patterns along with addressing the privacy needs.

Insurers should build a next-gen MDM solution which will be privacy context aware, consumption-driven, agile, and self-service enabled. Although we are seeing product vendors investing into integrated solution such as Informatica coming up with an integrated DPM (Data Privacy Manager) and MDM Edition, Collibra enhancing data privacy capabilities within its product suites, however since the technologies are evolving fast the MDM solution need not be constrained to traditional MDM product-based solutions but can leverage emerging technologies and AI/ML-based solutions much more effectively. Below are some of the options:

Cloud MDM	Big Data & MDM	Graph MDM	Intelligent MDM	Multi-Domain MDM	Quick MDM
<ul style="list-style-type: none"> • A new level of flexibility and scalability on Cloud • Keeping up with insurance consumerization and ubiquity of smart connected devices 	<ul style="list-style-type: none"> • Mining data on big data platforms to provision 360° customer view and other data domains to create a consistent omnichannel experience 	<ul style="list-style-type: none"> • MDM on graph DB • Flexible data modeling & relationships • Ease of implementation and faster time to value • Tracking of metadata 	<ul style="list-style-type: none"> • Rules-based source data • Less data stewardship & governance activity • Automatic capture of source data variation & business rules 	<ul style="list-style-type: none"> • Lower total cost because it can be implemented in a single instance • Better cross-functional collaboration 	<ul style="list-style-type: none"> • Faster creation of "golden record" in an agile approach • Virtualization, and real-time consolidation of master data into an enterprise master view

Table 2: Next-Gen MDM Solution Options

Infosys enabled an American insurer to redesign its MDM solution on multi domain big data pattern.. The result was a single source of truth not just for the

customer master data, but also for the claims, product, agent, and other master data sets. Besides providing a 360-degree customer platform with in built privacy

attributes, the data domains are helping the Insurer accelerate their transaction systems development by leveraging the new MDM-enabled data lake.



Data Quality

Data privacy laws requires companies to rectify inaccurate personal information and complete missing personal data without undue delay - a perfect catalyst to include data quality as part of the overall data governance solution.

Insurers always have had vast amounts of

structured data, but a flood of unstructured and semi-structured data from internal and external systems has increased the challenge of keeping the data clean and reducing risk. Some of these challenges are i) Data is dynamic, so DQ rules have to be too, ii) rules written in SQL are not scalable, and are labor-intensive to monitor and

maintain, and iii) traditional DQ products are inefficient and ineffective. The modern data quality solution must address these challenges by combining best-of-breed DQ tools with an orchestrator to unify different DQ processes and tools. An AI-powered approach for data quality will look like below:

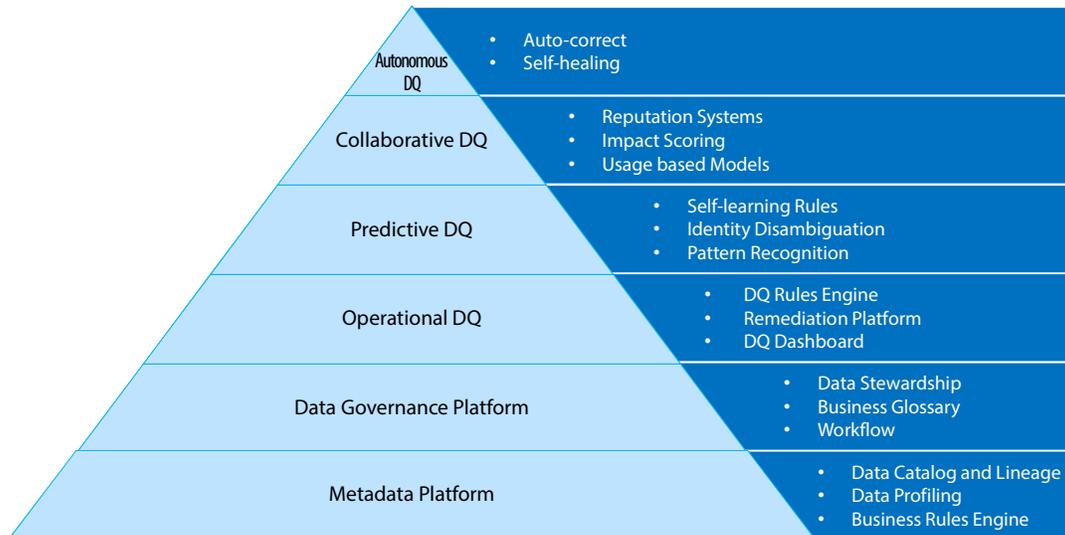


Figure 3: An AI-Powered Approach for Data Quality

Using machine learning, data quality matching can be fully automated. Automating the building of data quality business rules significantly reduces development time. AI/ML-based descriptive and prescriptive analytics can identify anomalies in the data and treat outliers and missing values.





Data Privacy and Data Security

Privacy laws require insurers to know where and how data is used and allow customers to decide on the use and retention of their data in the enterprise.

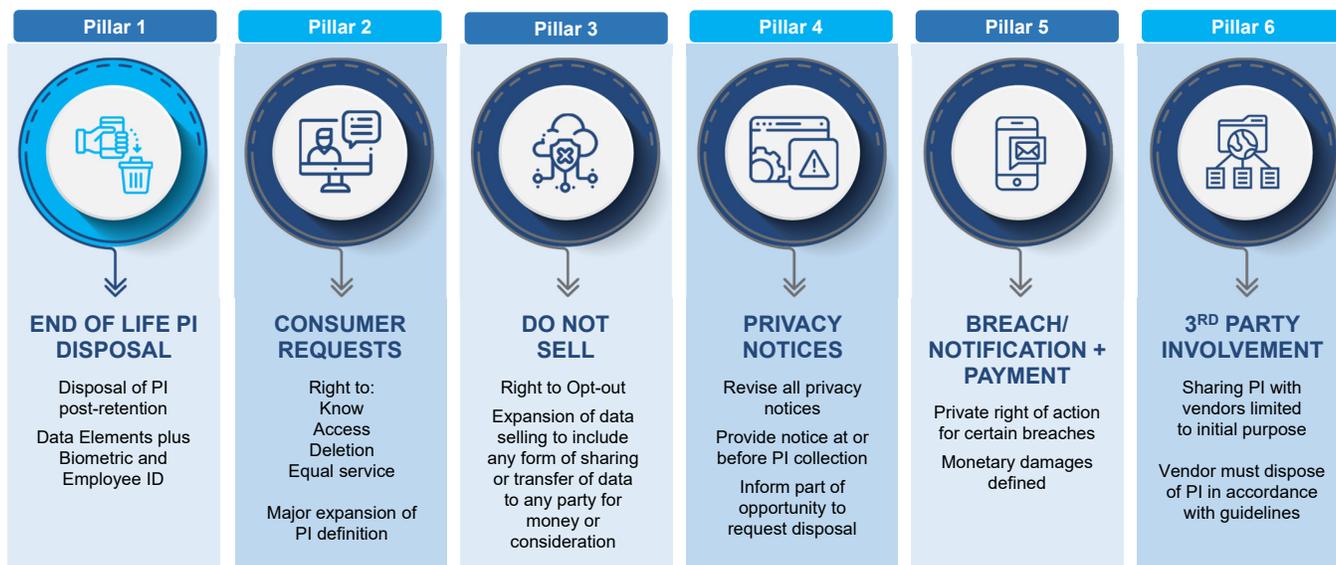


Figure 4: Data Privacy Pillars

To cover all the pillars, isolated data privacy solutions are ineffective at best. It is difficult to treat data uniquely and to protect sensitive data without a comprehensive data management solution, hence data privacy needs to be integrated into the overall data management strategy of the organization.

A large mutual life insurer has implemented a hub-and-spoke-based solution to centrally address the customer data privacy requirements of CCPANYDFS, and GDPR. The hub was built using a data lake, a custom user interface and rules framework, and data encryption and access control were achieved with Dataguise and Sentry.

Data Archival & Disposal

Traditionally, data that is no longer actively used was moved to dedicated storage to increase systems performance. Long-term retention also enabled seamless access to

archived data for business usage. However, the new data privacy regulations are now mandating insurers to consider a solution where data archiving, disposal, and retrieval are comprehensive and operate in real-time.

Insurers are now modernizing solution for data archiving, disposal and retrieval along with data privacy and security to not just cater to regulatory needs, but also to integrate with the overall data management solution.

Information lifecycle management product vendors such as Informatica and IBM have strong product suites to support data archival, retrieval and restoration, purging, compression, and encryption. Any solution should support both structured and unstructured data since the prevalence of unstructured data is expected to continue increasing. It should also leverage AI

and ML to automatically apply retention schedules and legal holds and dispose of unnecessary data.

Infosys Data Governance Workbench Solution

Infosys has developed the Infosys Data Governance Workbench solution to address the entire range of data management requirements including -

- Single point of access to address all data governance needs.
- Privacy Solution
- Flexible, loosely-coupled architecture.
- Available in open source as well as industry-leading data governance vendor formats.
- Supported by additional services, including consulting, customized development, and hosted SaaS.

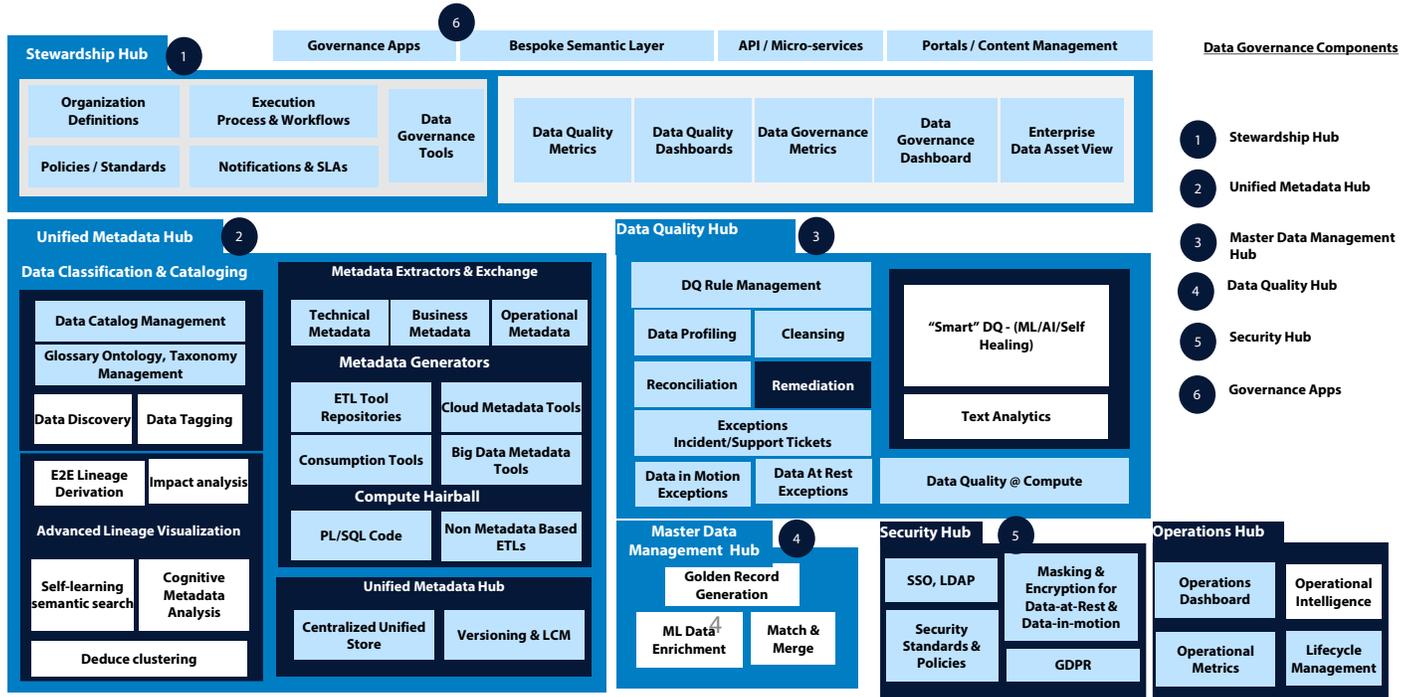


Figure 5: The Infosys Data Governance Workbench Overall Architecture



The key components of the Infosys workbench are the stewardship hub, unified metadata hub, MDM hub, data quality hub, security hub, and governance apps. The functionality of each component is as follows:

Infosys Data Governance Workbench Component	Features
Stewardship Hub	<ul style="list-style-type: none"> • Repository of governance metrics • Creation of derived metrics • Management and reporting of governance metrics • Seamless access to enterprise metadata
Unified Metadata Hub	<ul style="list-style-type: none"> • Unified metadata management • Lineage of data across the entire data landscape • Consignment lifecycle journey – source to report • Cloud-based UMDH offering agility in deployment
Master Data Management Hub	<ul style="list-style-type: none"> • ML-based match-merge engine • Golden record generation
Data Quality Hub	<ul style="list-style-type: none"> • The Infosys Smart Data Quality solution enables self-service capabilities such as data profiling and cataloging across data asset layers (lake, non-lake, and Cloud), builds data quality dashboards for better insights, and establishes DQaS (Data Quality as a Service) for both on-prem and on-Cloud data assets. • ML-based data quality checks and cleansing • De-duplication of data from multiple sources and golden record generation.
Privacy and Security Hub	<ul style="list-style-type: none"> • Sensitive data discovery, static and dynamic masking, Big Data masking, data subsetting • Security standards and policies • Infosys iEDPS Solution - Privacy compliance for GDPR, CCPA etc.
Governance Apps	<ul style="list-style-type: none"> • Plug and play model, registering only apps which are needed by the enterprise, such as advanced lineage and impact analysis

Table 3: Infosys Data Governance Workbench Component Features



Implementing a Holistic Data Management Solution

Infosys believes in establishing a sustainable practice and culture for data management capabilities and recommends a 4-stage approach for implementing a comprehensive data management solution for the enterprise which can support data privacy needs as well.

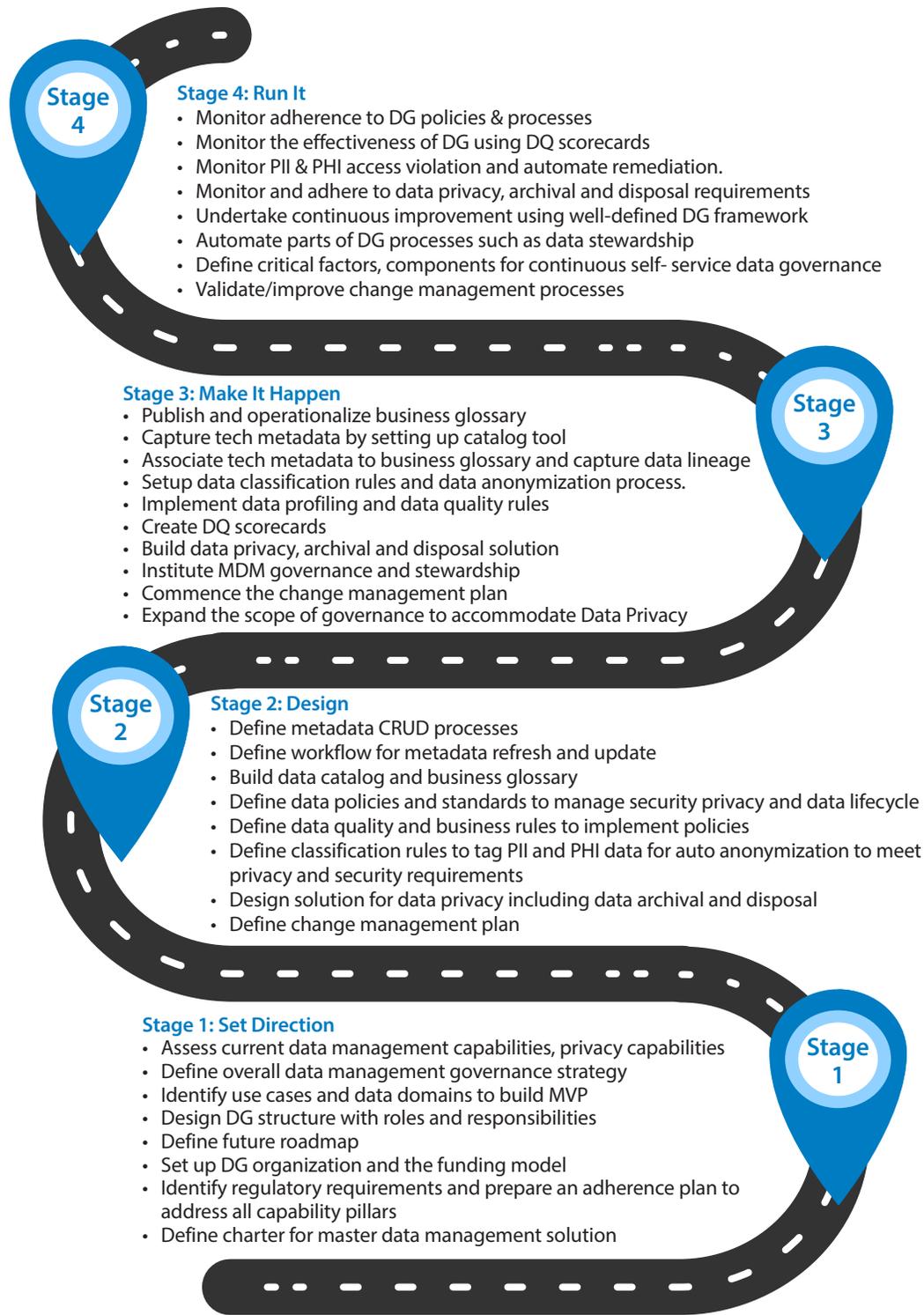


Figure 6: 4-Stage Approach for Implementing a Comprehensive Data Management Solution



Integrated Data Management - The Need of the Hour

A systematic approach for building a privacy compliant, consumption-driven, self-service enabled, agile solution for data management is the need of the hour. A comprehensive data management implementation will ensure that information stays secure and does not end up in the wrong hands, thus avoiding the huge costs of data duplication, data loss, or data privacy litigation. It also ensures that

important information is backed up and retrievable if something happens to the original, or to meet regulatory needs such as CCPA or GDPR.

Thus a sustainable effective, and efficient data management will be the foundation for success for insurers in the connected data world. With proper data management, insurers will be more organized and productive, users will have an easier time finding and understanding the information and insurers will be able to

react quickly to market changes as well as respond to federal regulators.

Emerging tech and AI/ML-based solutions will future-proof against a world driven by data and AI, and help accelerate the implementation of the solution itself. Hence, Insurers must prioritize data management as a key foundation program and choose a partner with deep experience in implementing end-to-end data management solution.

About the Authors



Prabir Majumdar

Director, Data and Analytics, Insurance

Prabir Majumdar, is a Director and leads Infosys Data and Analytics Practice for Insurance. He has more than 20 years of diverse experience of solving problems across the data value chain from Data Strategy and Data Management to Big Data and Analytics Advisory to Cloud Modernization of Data and Analytics.



Aaman Lamba

Sr Industry Principal, Data and Analytics Practice

Aaman Lamba, is a Sr Industry Principal and leads Infosys Data Governance and Master Data Management Consulting for global clients across industry verticals. He has more than 23 years of experience in solving problems in the areas of Data Governance, Data Privacy, Data Quality, Data Catalog and Master Data Management.

For more information, contact askus@infosys.com



© 2020 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.