

End-to-End data lineage

Building trust and powering reliable
reporting
May 2025

Click to enter



The better the question. The better the answer. The better the world works.



Shape the future
with confidence

01

Background

02

Effective data management

03

Decoding data lineage

04

Overcoming challenges
in implementation

05

Essential features in a
data lineage tool

06

The road ahead



01

Background

The 2008 Global financial crisis reshaped regulatory oversight, compelling banks to meet stricter capital, leverage, and liquidity requirements. However, beyond the financial resilience, the crisis also highlighted the gap that existed in the information technology and data architecture of the banks, which were inadequate to support broad management of financial risks. In response, the Basel Committee on Banking Supervision (BCBS) introduced the BCBS 239 framework in 2013¹, establishing principles for risk data aggregation and reporting to enhance the risk management and decision-making processes in banks.

Over the past decade, regulators have consistently emphasized the need for robust regulatory reporting through their publications and reviews being performed by skilled persons to assess the accuracy of banks' returns and opine on their governance and controls.

In their recent publication, 2025 priorities, the Prudential Regulation Authority (UK PRA) points out that poor data continues to be a root cause in a number of risks requiring remediation within firms². It expects firms to continue to improve their ability to aggregate data to support holistic risk management,

robust Board decision-making, and accurate regulatory calculations. The PRA also confirms it will continue to use the full set of its supervisory tools, including Section 166 reviews, to determine progress is being made.

In the light of these expectations, one of the key priorities for banks remains data management- focusing on understanding and identifying risks associated with how data is collected, used, stored, and shared, as well as how it is protected from misuse. Data governance, data lineage, data quality and technological advancements are set to be central pillars of banks' evolving data frameworks.

This document highlights the importance of data lineage, the associated challenges, emerging trends, and the way forward for global banks and their global capability centers (GCCs) in India.

¹ [BCBS 239](#)

² [International banks Supervision: 2025 priorities BOE PRA](#)



02

Effective data management

1 2

Data management is the foundation of how an organization creates, stores, uses, and shares data effectively. It involves defining systems, processes, and standards to guide data through its entire lifecycle. A key part of this journey is data governance, which fosters data accuracy, security and proper handling.

Successful data management requires collaboration between IT teams, business teams and subject matter experts. A Chief Data Officer (CDO) can guide these efforts by managing data projects in an organization and lead the cultural change that enables a strategic approach to data. Together, they build a data environment that supports transparency, integrity, and trust.

To implement a successful data governance program, below are a few points to consider³:

- Documenting data dictionaries to create a repository of the organization's metadata
- Documenting data lineage to track the origin, movement, and transformation of data through its lifecycle
- Determining existing controls and establishing new controls to enhance data quality and integrity

Financial institutions must execute data governance on all data sets or individual data elements that are used in key business processes, support critical reports, and have potential financial, regulatory, or reputational risks involved if the data was inaccurate or compromised.

2.1

Data dictionaries and metadata

Data dictionaries and metadata are foundational elements in modern data governance as they provide critical context that helps organizations manage and trust their data.

A data dictionary defines the structure and content of data sets, often for a single database, application, end-user computing (EUC) tool or data warehouse. The dictionary is used to manage the metadata - names, descriptions, structure, characteristics, storage requirements, default values, relationships, uniqueness, and other attributes of every data element in a model. Simply put, a data dictionary is a centralized repository of metadata, and the metadata is the data about data³.

Types of metadata:

01

Business metadata

Business metadata focuses on the meaning, purpose, and usage of data from a business standpoint.

It offers non-technical definitions of concepts, entities and attributes and supports data governance by ensuring clarity and consistency across organization.

Examples include definitions, business rules, transformation logic, data quality standards, data lineage, ownership, and known data issues.

02

Technical metadata

Technical metadata provides detailed information about the structure, storage, and movement of data across systems and is essential for IT teams.

Examples include physical database structures, column names and properties, access permissions, object properties, schema definitions, documented relationships between the data models and the physical assets.

³ DMBok - Data Management Body of Knowledge



2.2

Data lineage

Data lineage refers to the process of tracking the flow of data across an organization's systems and EUCs, from its point of origin to its final use. It provides a clear visual representation of how data moves, transforms, and interacts within various processes and systems. For data lineage to be effective, it should be integrated with the metadata, data quality initiatives and data governance frameworks.

Successful data lineage requires both these perspectives³:

- **Business focus:** Tracing back data elements critical to business from target systems or reports (for example, regulatory or financial reports) to original sources. When combined with data quality metrics, the insights help stakeholders identify where data accuracy or integrity is impacted.
- **Technical focus:** Beginning at the source and following data forward through each source of processing and usage, it enables technology teams to assess the impact of potential changes and any anomalies.

By aligning both these perspectives, data lineage empowers organizations to troubleshoot issues, optimize data flows and strengthen compliance.



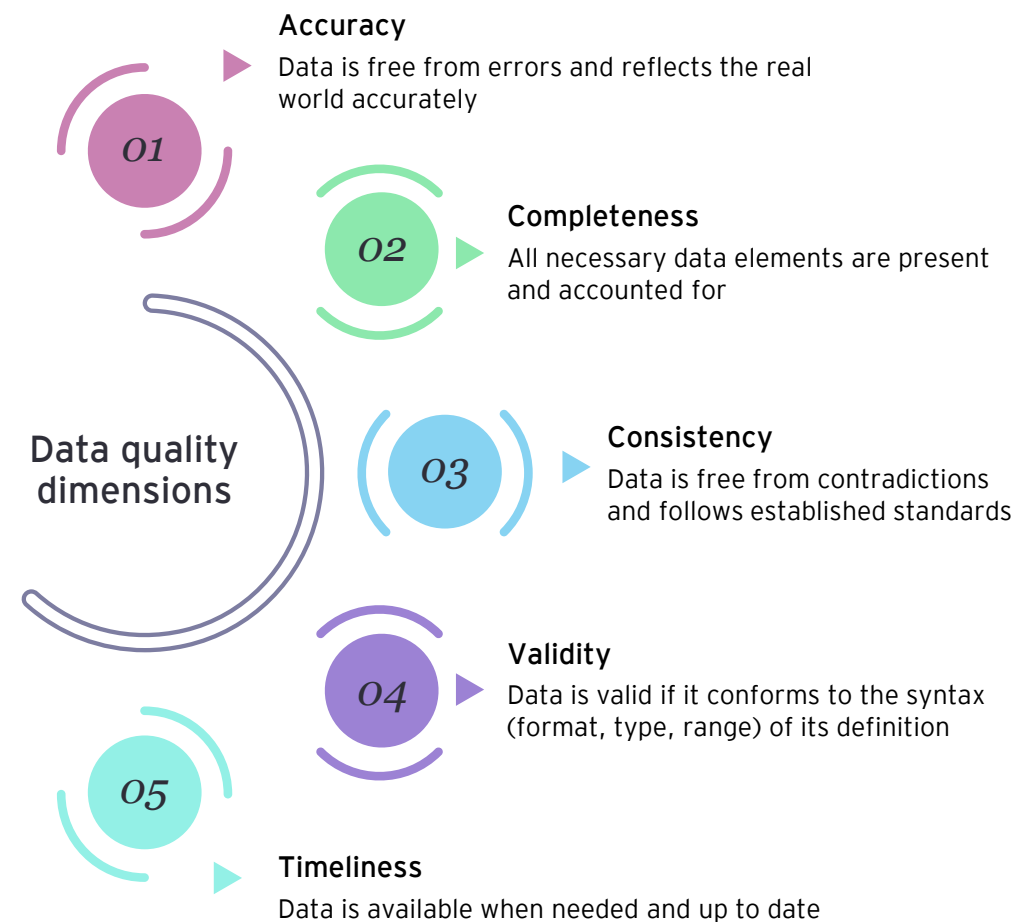
³ DMBOK - Data Management Body of Knowledge

2.3

Data quality

Data quality is another foundational element of effective data governance, ensuring that information is accurate, complete, consistent, and reliable for business use, i.e., it is fit for purpose based on data consumers requirements. Building data quality involves applying standards, validation rules and cleansing techniques to minimize errors and inconsistencies.

A key aspect of assessing and improving data quality lies in evaluating it across defined data quality dimensions that provide measurable attributes³.





03

Decoding data lineage

1 2 3

3.1

Key steps in implementing data lineage

Data lineage requires documenting the origin of data, as well as their movement and transformation through systems where they are accessed and used. There are several steps involved in this process. Below is a detailed guide covering all the key steps in data lineage implementation:

01

Identify data sources

This involves creating a comprehensive inventory of all data sources within the organization, which may include databases, applications, EUC tools, manual inputs, and external feeds. This can be achieved by collaborating with business and technical teams to gather insights into how they generate and utilize data.

02

Identify critical data elements

Not all data holds the same level of importance for the organization. The organization needs to identify the data elements that are deemed critical by applying criteria of materiality, role in regulatory or financial reporting, extent of usage across different teams or impact on business operations. By doing so, the organization can concentrate its time and resources on the most significant data elements.

03

Establish relationships between data elements

Identify how data elements interact and traverse through various processes and systems within the organization. For example, deep diving into data transformations at various stages in a data warehouse that is based on the concept of the ETL- Extract, Transform, and Load.

04

Gather and document the metadata

This involves documenting various metadata attributes like data ownership, definitions, underlying sources and applications, data type, logical names that are meaningful to business users, business context, etc., for each critical data element identified.

05

Create visualization in lineage tools

The process of connecting the pieces of data lineage is called stitching. This results in a holistic visualization of the data, its metadata, and its journey from source to its final destination. Organizations use various visualization tools, for example, Solidatus, Informatica, Collibra, and Alation, for creating graphic data flow diagrams.

06

Establish data quality controls

This involves collaborating with data stewards and business teams to establish data quality standards and robust controls (for example, validation rules, reconciliation checks, etc.) around entire data flow, ensuring its integrity and accuracy.

07

Regularly review and update

Regularly monitor and update your data lineage to incorporate any changes in the data flow, data sources, transformations, and changes in regulatory requirements, etc.

³ DMBok - Data Management Body of Knowledge



3.2

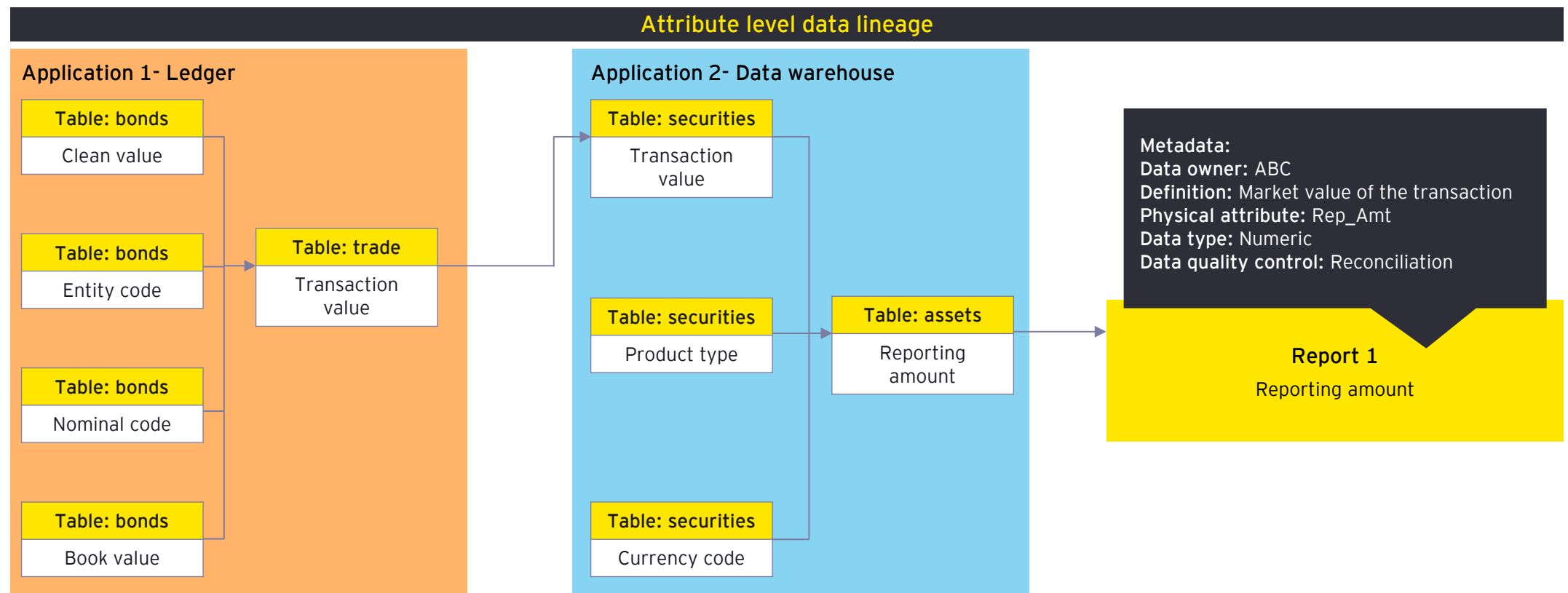
Seeing data in motion

The goal of data lineage is to turn data into information and information into insight. This is possible through data lineage tools that help visualize the data flows graphically. Data flows can be documented at different levels of detail:

- **Subject area:** Capturing data flows for a specific regulatory report or submission
- **Business entity:** Capturing data flows for a specific entity across functions
- **Attribute level:** Capturing data flows for a specific data element across systems and reports
- **System or application level:** Capturing data flows across different systems or applications

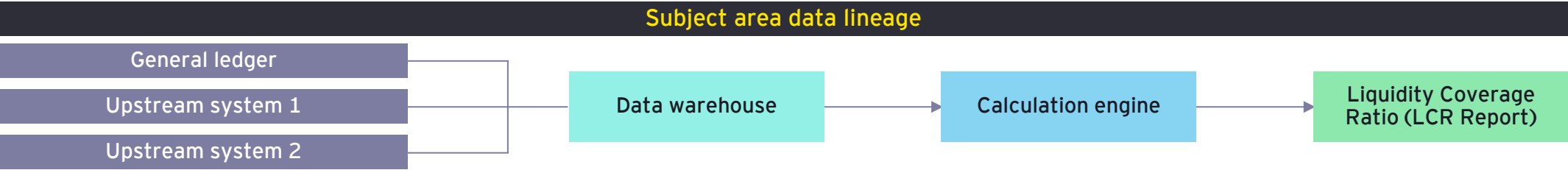
Below are a few examples of such visualizations:

This illustration provides a detailed representation of the data lineage of an individual data element level - Transaction value - highlighting how specific data attributes flow and transform across systems to support reporting requirements.



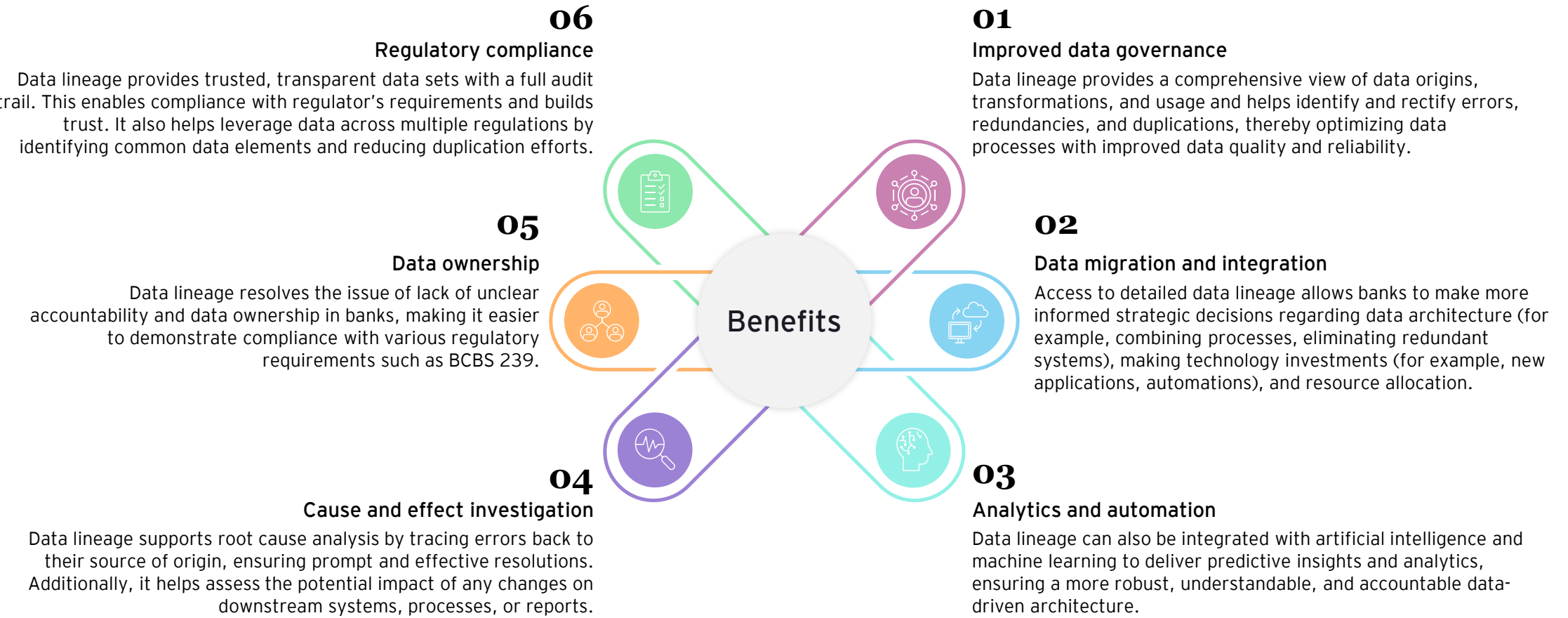


The illustration below provides a high-level data flow at a subject area level - for liquidity coverage ratio report - depicting the journey of all the data sources involved in producing the final reporting outputs.



3.3

Benefits of data lineage implementation





04

Overcoming challenges in implementation

1 2

Achieving end-to-end data lineage is challenging but essential for compliance, risk management, and operational efficiency. Financial institutions that invest in automation, governance, and proactive monitoring can transform data lineage from a regulatory burden into a strategic asset.

Below are some of the challenges that banks observe in implementing data lineage. By overcoming these challenges banks can improve reporting accuracy, and drive business success in an increasingly data-driven world.

Areas	Challenges	How can banks address them?
Defining the scope of the Data Program	Unclear boundaries - Determining which entities, business units, and external reports should be included can be complex.	A centralized data governance function regulated by the CDO and data stewards may agree on the scope of entity inclusion and external reports across business units.
	Regulatory compliance - The requirements can vary across regions, making it difficult to standardize processes.	Data stewards across functions may consider collaborating and leveraging information to avoid any redundancies and duplication of efforts in documenting data dictionaries and data lineage.
	Competing priorities - Different upstream and downstream departments may have conflicting views on data lineage priorities.	Foster a collaborative planning approach to prioritize data lineage efforts across compliance, IT, and business units.
	Lack of ownership - Unclear data ownership and accountability can hinder decision-making.	Create clear data stewardship roles to encourage accountability and streamline decision-making.
Legacy systems and data silos	Lack of interoperability , hinders integration between old and new technologies.	Leverage middleware or Application Programming Interface (APIs) to facilitate seamless integration between legacy and modern technologies.
	Unstructured data like manual reports, emails, screenshots used in data processing present a significant challenge for data lineage.	Leverage advanced data mapping tools and artificial intelligence to establish robust data management on unstructured artefacts enabling better tracking and understanding of these data sources.

Areas	Challenges	How can banks address them?
Labor-Intensive without automation	Manually documenting and tracking data flows, transformations and dependencies is time consuming and prone to errors.	Using data lineage tools that provide an open API, read SQLs and utilize artificial intelligence to automatically track, store and visualize data lineage can reduce the manual intervention.
	Lack of standardization leads to inconsistent documentation .	Implement organization-wide data documentation standards to build consistency in lineage records.
Ever-evolving processes and data flows	Frequent data pipeline changes disrupt lineage continuity.	Develop an operating model to periodically review changes in the data pipeline management frameworks and adapt the changes into the lineage.
	New data sources and integrations increase documentation complexity.	Establish scalable documentation practices at implementation stage for new systems to enable ease of documenting the metadata and lineage.
High costs	High implementation costs for tools and infrastructure.	Adopt cloud-based lineage solutions to reduce initial costs and scale with the organization's needs.
	Ongoing costs incurred for manually maintaining data lineage documentation.	Use automated data lineage tools to reduce manual documentation updates and lower maintenance costs.
Complexity of documentation	Mapping complex data transformations across multiple platforms requires both functional domain and data expertise.	Collaborate with IT teams and data SMEs to perform mappings, in addition to this firms could also invest in AI-driven tools (for example, Alation, Informatica etc.) to automate and simplify lineage mapping.
	End-User Computing (EUC) tools , particularly Excel workbooks, are often used for ad-hoc analysis and adjustments, leading to data silos and inconsistency.	Establish a governance framework for EUCs to include monitoring the EUC inventory for changes, integration of EUCs with centralized data platforms, designing effective controls to avoid any errors. Additionally, use tools that automate the tracking of data flows in EUCs.



05

Essential features in a data lineage tool

1 2

A data lineage tool is vital for banks aiming to improve data governance and compliance. It offers a clear view of data flow, enabling users to track the origins and transformations of data across systems. Here we have covered a few key features that we believe would be beneficial in a data lineage tool:

Automated end-to-end real-time lineage

The tool automatically identifies and maps the sources of data, transformations, and data flows by connecting to various data sources without manual input. This process enables real-time updates of data maps and other lineage documents, ensuring that the most current information is always readily available. As a result, data becomes available in a timely manner and is less susceptible to errors, as it eliminates the need for manual intervention.

Mapping of various metadata attributes

A key feature in a data lineage tool is the ability to integrate various metadata attributes such as risk, privacy sensitivity, ownership, data quality controls, etc. This integration offers vital context to the data flow. By associating these attributes with data as it progresses through its lifecycle, banks can bolster data governance, improve compliance, and maintain high data quality.

Dynamic visualization of data maps for ease of understanding

This involves utilizing interactive visual representations, such as process maps and diagrams, to illustrate the movement and transformation of data across various systems, tools, and processes. Such visualization simplifies the understanding and analysis of complex data flows, creating a more intuitive and user-friendly experience.

Multiple views and filters

An effective data lineage tool offers multiple views and filtering capabilities to improve user experience and streamline data analysis. This could include keyword search functionality, enabling users to refine their searches based on criteria like data type, source, destination, and more. It should also provide a data element-level view, illustrating how data is sourced and transformed, as well as an application-level view, showing how different applications interact with the data.

AI-powered analytics

The platform should empower business users to access and analyze data independently. Using AI to preform validation checks, provide insights, highlight inconsistencies, automate data quality checks, and suggest improvements in data governance.

Data privacy and security

The data privacy and security features of a data lineage tool are essential for protecting sensitive information and ensuring compliance with regulations. These features may include:

Access controls

Role-based access controls to restrict who can view or manipulate data lineage information

Data masking

Utilizing data masking techniques to obscure sensitive information in visualizations

Audit trails

Maintaining audit logs that track changes and access to data lineage information

Encryption

Encryption of data to protect information from unauthorized access during storage and transmission

Periodic reviews and sign-offs

An efficient data lineage tool should enable users to schedule regular reviews of data lineage such that data flows and transformations are accurately documented and remain current. Users should receive automatic alerts on approaching reviews, and reviewers should have the capability to assess and provide sign-off on the data lineage documentation directly within the tool. It should also be capable of generating reports on the status of reviews and signoffs, assisting management in understanding compliance levels and identifying areas that require attention.

These features can significantly enhance the effectiveness of a data lineage tool, making it more valuable for banks looking to manage their data effectively.



06

The road ahead

As banks navigate the complexities of an evolving data landscape, data lineage is emerging not only as a regulatory necessity, but also a pivotal strategic asset. The foundational elements – spanning from lineage capture to visualization – have laid the groundwork. However, unlocking the true potential of data lineage demands a visionary approach that extends beyond compliance to drive innovation and value creation.

To achieve this, banks must align their data lineage strategies with emerging industry trends, advanced enterprise use cases and technological advancements. The next stage in this journey calls for scalability, automation, and seamless integration with broader data governance frameworks and business transformation initiatives. In the section below we outline our perspective on how banks can shape and accelerate their data management roadmap.





Adopting scalable, automated lineage frameworks

Manual data lineage efforts are resource-intensive and prone to errors. To scale lineage management efficiently, banks can leverage AI-powered, metadata-driven automation tools that reduce costs, enhance accuracy, and allow continuous updates in dynamic ecosystems. As businesses adopt emerging technologies like data mesh, cloud-native platforms, and advanced analytics, lineage solutions must evolve to support decentralized architectures, real-time data flows, and interoperability across hybrid environments.

1

2

Integrating data practices enterprise -wide

To sustain momentum and establish robust data management, it is essential for banks to create a well-defined operating model to cultivate data practices across all levels. This requires moving beyond IT and compliance silos to foster a "lineage-first" mindset among business users, data stewards, and decision-makers. A structured operating model should outline a clear framework defining roles, responsibilities, ownership, processes in scope, review frequency and overall data policies ensuring a seamless integration of data lineage and data management practices across the enterprise, driving the banks long-term governance goals, and promoting better data ethics.

3

Cultivating data stewardship practices

Data stewards are the backbone of data governance. Banks must establish a robust structure that empowers data stewards to fulfil their critical role in data governance. This includes formalizing their responsibilities, such as managing the business glossary, ensuring alignment of business terms, addressing data quality issues, and maintaining regulatory compliance. Providing them with the right tools, cross functional collaboration frameworks, and decision-making authority is essential. Additionally, banks should invest in training, standardized data quality policies and audit readiness support to enable data stewards to drive consistency, compliance, and value across data assets effectively.

4

Empowering employees through data literacy

Banks need to prioritize employee trainings to embed data management practices across the organization effectively. Every employee, from the Chief Data Officer (CDO) to a front-line employee, plays a vital role in identifying data risks and ensuring compliance with governance standards. Regular training sessions and clear communication about data management policies and procedures not only fosters accountability but is also essential to keep employees informed and aware of their responsibilities, ensuring an ongoing compliance and overall data quality.



5

Building transparency through data traceability

While banks emphasize data lineage to understand the flow and transformation of data, equal focus is needed on data traceability. Data traceability is the process of evaluating whether the data is following its life cycle as expected. It is particularly crucial during system upgrades or interface replacements to confirm seamless integration without disruption. It also supports regulatory audits by demonstrating visibility and regulatory alignment across the entire data lifecycle.

6

Simplifying data ecosystems

Banks should prioritize reducing their reliance on EUCs, especially Excel spreadsheets by embracing automation. EUCs often rely heavily on manual processes, increasing risk of errors while complicating data lineage tracking. Moreover, banks should focus on leveraging AI-driven solutions to streamline any unstructured data. This will help banks streamline their operations, improve data quality, and alleviate the burden on data lineage tracking.

7

Ensuring audit preparedness

Banks should prioritize audit readiness by preparing for skilled personnel reviews and regulatory exams. This includes maintaining comprehensive documentation, adhering to established data quality and lineage standards and undergoing regular reviews with internal audit. Internal audit plays a critical role in assessing the effectiveness of data governance framework, detecting compliance gaps, and identifying any inefficiencies thereby helping banks to align with the regulatory requirements.

8

Setting up governance forums and feedback loops

Governance forums and working group meetings play a crucial role in ensuring the effectiveness of data governance models and architectures. It is important to assess how well the data governance model aligns with existing architecture and data standards. To enable comprehensive evaluation, governance forums should involve a diverse group of subject matter experts, data stewards, CDO, and IT experts, each bringing different backgrounds, skills, expectations, and opinions. This collaborative approach will help refine the data governance model and enable implementation of best practices.

Contact us



Adarsh Ranka

India Region FAAS Leader and Partner
with Indian member firm of EY Global
adarsh.ranka@srb.in



Shiva Iyer

Partner
Financial Services and Financial Accounting
Advisory Services (FAAS), EY India
shiva.iyer@in.ey.com



Smridhi Arora

Senior Manager
Financial Accounting Advisory Services
(FAAS), EY India
smridhi.arora@in.ey.com



Prashant Agarwal

Manager
Financial Accounting Advisory Services
(FAAS), EY India
prashant2.agarwal@in.ey.com



We acknowledge contributions from

Mohit Megotia

Senior Manager
FAAS, EY India

Krantikumar P Dhuri

Senior Manager
FAAS, EY India

Rohit Mulchandani

Manager
FAAS, EY India

Kunchi Sravani

Manager
FAAS, EY India



Our Offices

Ahmedabad

22nd Floor, B Wing, Privilon
Ambli BRT Road, Behind Iskcon Temple
Off SG Highway
Ahmedabad - 380 059
Tel: + 91 79 6608 3800

8th Floor, Building No. 14A
Block 14, Zone 1
Brigade International Financial Centre
GIFT City SEZ
Gandhinagar - 382355, Gujarat
Tel +91 79 6608 3800

Bengaluru

12th & 13th Floor
"UB City", Canberra Block
No.24 Vittal Mallya Road
Bengaluru - 560 001
Tel: + 91 80 6727 5000

Ground & 1st Floor
11, 'A' wing
Divyasree Chambers
Langford Town
Bengaluru - 560 025
Tel: + 91 80 6727 5000

3rd & 4th Floor
MARKSQUARE
#61, St. Mark's Road
Shantala Nagar
Bengaluru - 560 001
Tel: + 91 80 6727 5000

1st & 8th Floor, Tower A
Prestige Shantiniketan
Mahadevapura Post
Whitefield, Bengaluru - 560 048
Tel: + 91 80 6727 5000

Bhubaneswar

8th Floor, O-Hub, Tower A
Chandaka SEZ, Bhubaneswar
Odisha - 751024
Tel: + 91 674 274 4490

Chandigarh

Elante offices, Unit No. B-613 & 614
6th Floor, Plot No- 178-178A
Industrial & Business Park, Phase-I
Chandigarh - 160 002
Tel: + 91 172 6717800

Chennai

6th & 7th Floor, A Block,
Tidel Park, No.4, Rajiv Gandhi Salai
Taramani, Chennai - 600 113
Tel: + 91 44 6654 8100

Delhi NCR

Aikyam
Ground Floor
67, Institutional Area
Sector 44, Gurugram - 122 003
Haryana
Tel: + 91 124 443 4000

3rd & 6th Floor, Worldmark-1
IGI Airport Hospitality District
Aerocity, New Delhi - 110 037
Tel: + 91 11 4731 8000

4th & 5th Floor, Plot No 2B
Tower 2, Sector 126
Gautam Budh Nagar, U.P.
Noida - 201 304
Tel: + 91 120 671 7000

Hyderabad

THE SKYVIEW 10
18th Floor, "SOUTH LOBBY"
Survey No 83/1, Raidurgam
Hyderabad - 500 032
Tel: + 91 40 6736 2000

Jaipur

9th floor, Jewel of India
Horizon Tower, JLN Marg
Opp Jaipur Stock Exchange
Jaipur, Rajasthan - 302018

Kochi

9th Floor, ABAD Nucleus
NH-49, Maradu PO
Kochi - 682 304
Tel: + 91 484 433 4000

Kolkata

22 Camac Street
3rd Floor, Block 'C'
Kolkata - 700 016
Tel: +91 33 6615 3400

6th floor, Sector V,
Building Omega, Bengal Intelligent Park, Salt
Lake Electronics Complex, Bidhan Nagar
Kolkata - 700 091
Tel: +91 33 6615 3400

Mumbai

14th Floor, The Ruby
29 Senapati Bapat Marg
Dadar (W), Mumbai - 400 028
Tel: + 91 22 6192 0000

5th Floor, Block B-2
Nirlon Knowledge Park
Off. Western Express Highway
Goregaon (E)
Mumbai - 400 063
Tel: + 91 22 6192 0000

3rd Floor, Unit No.301
Building No.1, Mindspace-Gigaplex
IT Park, MIDC, Plot No. IT-5
Airoli Knowledge Park
Airoli West, Navi Mumbai - 400 708
Tel: + 91 22 6192 0003

Altimus, 18th Floor
Pandurang Budhkar Marg
Worli, Mumbai - 400 018
Tel: +91 22 6192 0503

Pune

C-401, 4th Floor
Panchshil Tech Park, Yerwada
(Near Don Bosco School)
Pune - 411 006
Tel: + 91 20 4912 6000

10th Floor, Smartworks
M-Agile, Pan Card Club Road
Baner, Pune - 411 045
Tel: + 91 20 4912 6800



Ernst & Young Associates LLP

EY | Building a better working world

EY is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fueled by sector insights, a globally connected, multi-disciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

All in to shape the future with confidence.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via ey.com/privacy. EYG member firms do not practice law where prohibited by local laws. For more information about our organization, please visit ey.com.

Ernst & Young Associates LLP is one of the Indian client serving member firms of EYGM Limited. For more information about our organization, please visit www.ey.com/en_in.

Ernst & Young Associates LLP is a Limited Liability Partnership, registered under the Limited Liability Partnership Act, 2008 in India, having its registered office at 6th Floor, Worldmark-1, Asset Area 11, Hospitality District, Indira Gandhi International Airport, New Delhi-110037, India.

© 2025 Ernst & Young Associates LLP. Published in India.
All Rights Reserved.

ED None
EYIN2505-013

This publication contains information in summary form and is therefore intended for general guidance only. It is not intended to be a substitute for detailed research or the exercise of professional judgment. Neither EYGM Limited nor any other member of the global Ernst & Young organization can accept any responsibility for loss occasioned to any person acting or refraining from action as a result of any material in this publication. On any specific matter, reference should be made to the appropriate advisor.

SA

ey.com/en_in

