Data Governance Awareness

General

Produced by the Data Governance Team

Course Objectives

By the end of this training, you will have a solid understanding of the following:

1. **Data Governance**
   a) What is Data Governance?
   b) Data Governance Risk Types and Risk Mitigation
2. **Data Governance Framework**
   a) Roles & Responsibilities – 1st, 2nd and 3rd Lines of Defense
   b) Framework & Organizational Structure
3. **Data Governance Standards**
   a) Standards Overview
   b) Data Dictionary/Data Lineage
4. **Data Quality Issue Management & Remediation**
   a) Data Quality Control Dimensions Overview
   b) What is a Data Quality Issue?
   c) Issue Management & Remediation Process

You are required to **acknowledge** your understanding of the material presented at the end of this training.

Section I: Data Governance

What Is Data Governance?

- Data governance is a **methodology that ensures important data assets are formally managed throughout Apple Bank.**
- **Data governance is not just about technology.** It's about people taking responsibility for the information assets of their organization **by looking at the processes they use to interact with information** as well as how and why it's being used.
- The overall goal of the Bank's Data Governance Program is to **make quality and reliable data readily available to the right people,** to make business decisions, optimize operations, offer services, and improve profitability.

What Is Data Governance? (Continued)

Without effective Data Governance, organizations may suffer from inconsistencies in data availability, collection, usability, integrity, and security which prevent them from realizing the full-value of their data.

Below are some examples of ineffective Data Governance and how each adversely impacts the Bank:

- Null/Blank values in the Payee field for select monetary instrument transactions impacting the quality of transaction monitoring alerts
- Outgoing wire transactions containing incomplete or incorrect country codes impacting transaction monitoring for large wires to high risk or tax haven countries
- A critical data feed that is not received on a timely basis for input to a key business process
- Data entry of incorrect NAICS codes (North American Industry Classification System) on select Commercial & Industrial loans impacting CECL (current expected credit loss) reserve model calculations
- Missing Customer Date of Birth during new Customer Account setup impacting Financial Crimes Compliance screening capabilities

Data Governance – Risk Types

It is important for the Bank to monitor and control the **following types of Data Risks** to ensure the on-going availability, integrity and security of enterprise data.

Risk Type	Risk Type Definition
Governance	– The risk of the **incomplete or ineffective oversight** of the data lifecycle, associated data assets and controls
Data Quality	– The risk of the **erosion of data accuracy, completeness, consistency and integrity** (for example, a data file is corrupted during data transmission)
Use / Fit for Purpose	– The risk that both the business definition and the quality of the data being consumed is **insufficient to meet the requirements of their intended business use or purpose**
Availability	– The risk that a file/service is **not accessible to the end-user** at the time that person/department is scheduled to use it
Change Management	– The risk of the **ineffective management and communication of changes** made to the availability, delivery and consumption of Critical Data Elements (CDEs)** between Data Providers and Data Consumers

** Refer to the **Glossary** that defines the Data Governance Key Terms used throughout this training course.

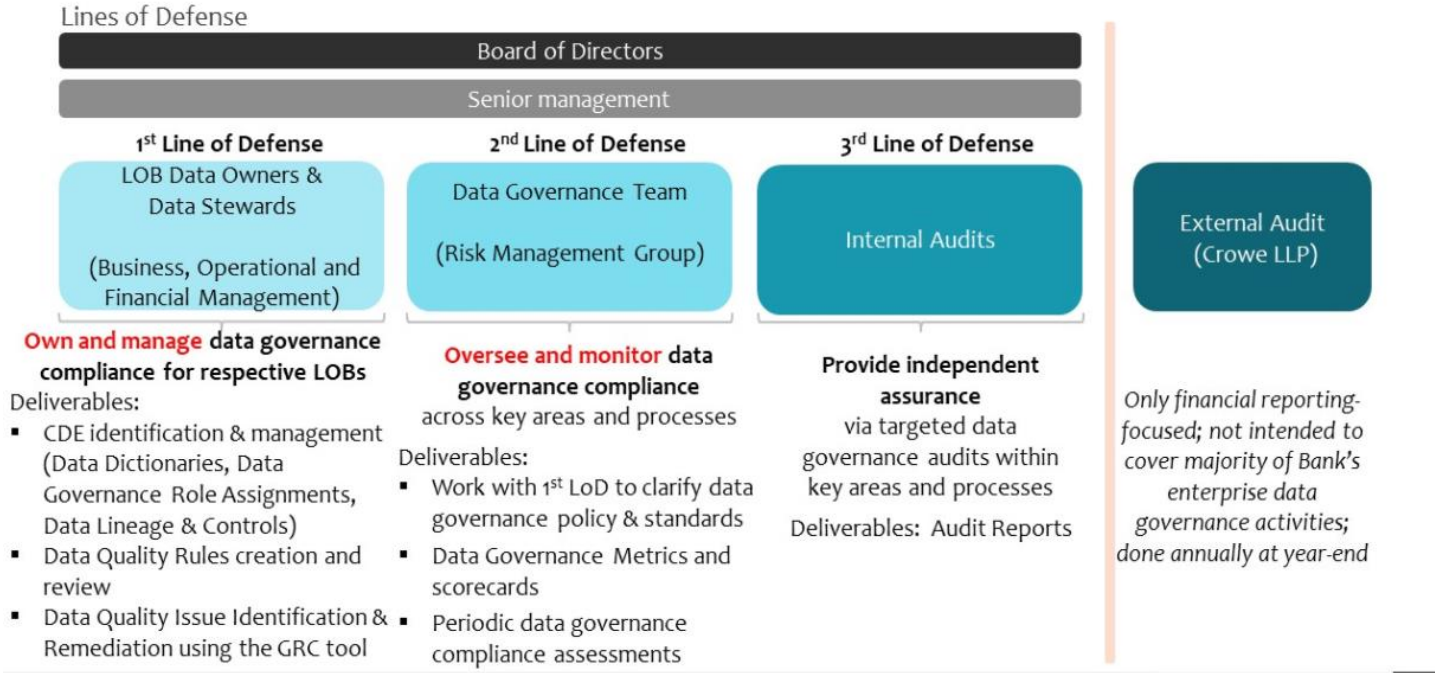
Data Governance – Risk Type Mitigation

The Data Governance Policy & Standards provide several tools to both monitor and control these Data Risks

Risk Type	Risk Type Mitigation – Data Governance Standards Deliverable
Governance	– Data Governance Framework – Data Governance Standards Assessment Scorecard (periodic assessment rating of critical Bank functions based on a 4-point rating scale of Below, Meets, Above, or Exceeds Standards compliance levels)
Data Quality	– Data Quality Rules development and implementation – **Data Reconciliations to "Golden Copy" Data Sources** – Data Quality Scorecard delivery for **Critical Data Elements** (CDEs) – Data Quality Issue Tracking and Remediation (GRC tool) and Data Quality Issue Management Forum
Use / Fit for Purpose	– Data Dictionaries for CDEs – Data Lineage & Controls documentation – Data Quality Testing and Data Quality Issue Management Forum – Service Level Agreements
Availability	– Service Level Agreements
Change Management	– Service Level Agreements; Formal Change Impact Notification communications between Data Providers and **Direct** Data Consumers; Data Governance Committee Meetings

Section II: Data Governance Framework

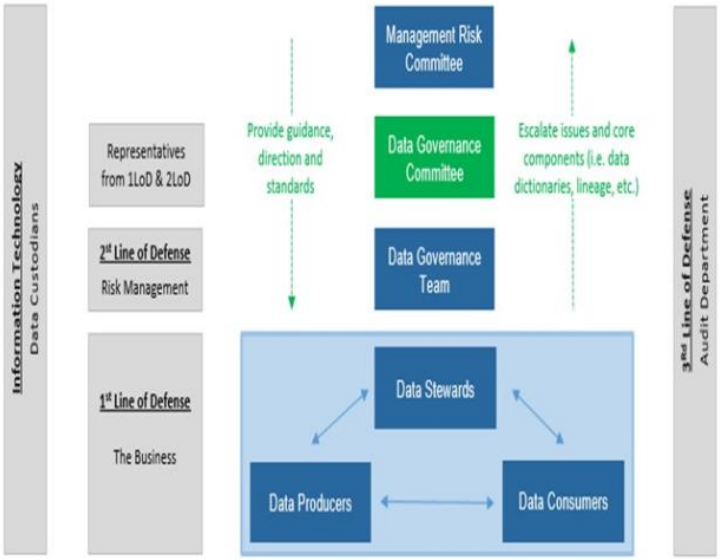
Data Governance Roles & Responsibilities



Data Governance Framework & Organizational Structure

The Bank has established the following Data Governance framework and organizational structure to monitor and enforce the Bank's compliance with the Firm's Data Governance Policy & Standards

- **Data Owners & Stewards in the 1st LoD** will act as the **hands-on resource within the business** to:
  - **create / manage Critical Data Elements**
  - **report data quality issues "in the spirit of … "If you see something say something"**
  - conduct other data governance responsibilities around control and use of data
- **Data Stewards on the Data Governance team** within Risk Management (2nd LoD) will:
  - **collaborate** with LOB Data Owners and LOB Data Stewards to **provide clarity** to Data Governance policy and standards
  - **drive overall data governance compliance**
- The **Data Governance Committee** is accountable for setting the plans, objectives, priorities, and performance measures for the data governance effort. The Committee includes key representatives from both the 1st & 2nd LoD



Section III: Data Governance Standards

Data Governance Standards Overview

The Data Governance Standards represent a set of operational guidelines used across the Business and Support areas of the Bank to achieve the goals and objectives of the Data Governance Policy and Program

Data Standard	Data Standard Guidelines	Key Deliverables
Data Dictionary	– The Bank will compile data dictionaries of **Critical Data Elements (CDEs)** for prioritized processes and functions	– **Identification of CDEs** for prioritized processes, reports, models and analytics – **Data Dictionary of key business terms and associated business definitions** for CDEs, inclusive of Data Governance Role Assignments (CDE Data Owner, Business Data Steward) – **Full front-to-back data asset level Data Lineage documentation** across the data lifecycle showing each CDEs data origin, what happens to it, and where it moves over time to the point of final consumption
Data Quality Controls & Testing	– Expectation of the Bank having adequate **Data Controls in place** for prioritized processes and functions – **Validation of these Data Controls at key points in the data lifecycle** to ensure ongoing data quality and data integrity – Introduction of the **six dimensions of data quality**** used at **key data control points** to test, measure and monitor ongoing data quality	– **Data Lineage and Controls documentation** – **Data Quality Rules** development and implementation leveraging one or more data quality dimensions** – **Data Reconciliations** to "Golden Copy" Data Sources – **Data Quality Scorecard** delivery for CDEs highlighting DQ Rule test results and spotlighting data quality issues **** The six dimensions of data quality will be discussed in more detail under Section IV of the Data Governance Training Program**

Data Governance Standards Overview (Continued)

The Data Governance Standards represent a set of operational guidelines used across the Business and Support areas of the Bank to achieve the goals and objectives of the Data Governance Policy and Program

Data Standard	Data Standard Guidelines	Key Deliverables
Data Quality Issue Remediation	– Data Quality Issues will be escalated as per the Bank's data quality issue remediation program – Data Quality issues involve data discrepancies across one or more of the **six dimensions of data quality**** – **These are the same six dimensions** used at **key data control points** to test, measure and monitor ongoing data quality of **Critical Data Elements**	– To identify data quality issues (potential and actual) in the spirit of "if you see something say something" – To **log issues into the Risk Management's GRC (MetricStream) data quality incident management tool** – To triage, **manage and remediate** data quality issues through the **Data Quality Incident Management Forum** **** The six dimensions of data quality and the Data Quality Issue Remediation Process will be discussed in more detail under Section IV of the Data Governance Training Program**
Change Impact Notification	– Establish a methodology between Data Providers and Data Consumers to **manage and communicate changes** made to the **availability, delivery and consumption of Critical Data Elements**	– Formal **Change Impact Notification communications** between Data Providers and **Direct** Data Consumers – Discussion and communication of change impacts leveraging the **Data Governance Committee Meetings** – Formal **Service Level Agreements** between Data Providers and Data Consumers for **critical data processes and feeds**

Data Governance Standards – Data Dictionary

The first step in creating Data Dictionary and Data Lineage Governance documentation is to identify **Critical Data Element "Crown Jewels" associated with key business processes, reports, models and analytics**

As was discussed in **Section I,** Critical Data Element (CDE) selection should consider the:

- Risk priority ranking of the Bank's products
- Dollar materiality (using both point in time and multi-period trend analyses)
- Level of manual effort required to calculate, aggregate, and report on underlying data

After all CDEs have been identified, a **Data Dictionary** of **key Business Terms** and associated **Business Definitions** can be prepared. Below is an example of a Data Dictionary for CDEs associated with the Bank's Call Report regulatory filing prepared and submitted quarterly to the FDIC.

Critical Data Element / Business Term | **Business Term Definition**

FFIEC 041 Call Report Term Name	FFIEC 041 Call Report Schedule	FFIEC 041 Call Report Term Definition	Critical Data Element (Y/N)	CDE selection criteria based on meetings with Shyler and Jasmin of Regulatory Reporting
Non-Interest Income: Trading revenue (RIADA220)	Schedule RI - Income Statement	Represents the net gain or loss from trading cash instruments and derivative contracts (including commodity contracts) that has been recognized during the calendar year-to-date.	Y	CDE based on critical product and dollar materiality

Data Governance Standards – Data Lineage

The first step in creating Data Dictionary and Data Lineage Governance documentation is to identify **Critical Data Element "Crown Jewels" associated with key business processes, reports, models and analytics**

Next, **Data Lineage** documentation can be prepared. **The purpose of Data Lineage is to trace a Critical Data Element's data origin from the point of final consumption back to its initial origin.**

Full front-to-back system level Data Lineage documentation for the Bank's Call Report is shown below. Note that, in addition to showing the data lineage flow back from the **final reporting system** (Fed Reporter) through to the **initial system of origin** (Banktel), that **CDE data governance roles of Data Owner and Data Steward have been identified****

Data Lineage Documentation Example

Critical Data Element / Business Term | **Data Governance Role Assignments** | **Final Data Provisioning Point Fed Reporter** ← **System of Record MISER General Ledger** ← **System of Origin TPG Software**

FFIEC 041 Call Report Term Name	Business Data Owner	Business Data Steward	Authorized Data Source (Target)	System of Record Data Source (1 Hop Back from ADS)	System of Origin Data Source (2 Hops Back from ADS)
Non-Interest Income: Trading revenue (RIADA220)	Douglas Vanhorne	Zinaida Gonta	Fed Reporter	Miser (General Ledger)	TPG Software

** Refer to the **Glossary** for additional information on these roles.

Section IV: Data Quality Issue Management & Remediation Program

Data Quality Control Dimensions Overview

Apple Bank

Ensuring good data quality is an essential part of the Data Governance Program. The Data Governance Policy and Standards focus on the following **six Data Quality Control Dimensions** to monitor and control the **data quality of Critical Data Elements** consumed in key business processes, reports, models and analytics:

The degree to which the **data is consistent with an authoritative source of the data** (e.g., accounting details reconcile to the Bank's General Ledger)

The assurance that the data has been provided or is **available within the documented timeframe required by the data consumer** (e.g., email confirmation from data provider)

The degree to which the data **conforms to defined business rules for acceptable content** (e.g., Tax Identification Number field only contains numbers (no text))

The degree to which the data remains consistent when in motion, and within or across data stores. The assurance that the data components were transmitted and received in total and **were not corrupted in transmission** (e.g., compare record counts and hash totals pre and post data transmission)

The degree to which **data may not be duplicated within a set of data** (e.g., a customer account number must always be a distinct value)

The degree to which the **data was transmitted and received in total** (e.g., same record counts and control totals pre and post data transmission)

What is a Data Quality Issue?

- A data quality issue represents any matter that causes the high quality of the data to be in dispute (i.e., the **data is no longer fit for its intended use**)
- A data quality issue represents a **repeated event** impacting the accuracy, completeness and timeliness of the data shared between Data Providers and Data Consumers
- A data quality issue would **NOT** include:
  - **One-off events** that are not repeated issues (these are considered data quality observations)
  - Issues identified during testing processes in **non-production environments**

Data Governance Issue Management & Remediation Process

Data Quality issue management and remediation consists of the following eight process steps:

Process Step	Who is Responsible	Data Governance Issue Management – Phase Deliverable
Issue Identification	– **All Bank Personnel**	– Identification of potential data quality issue (in the spirit of … "if you see something say something") – Description of the problem, effect and risk – Document controls and workarounds to mitigate risk
Logging of Issue in GRC tool	– 1st or 2nd line of defense Data Steward, Business SME or IT SME	– The data quality issue is formally entered in the GRC tool along with all relevant details
Triage Data Quality Issue	– 2nd line of defense Data Steward, Data Governance Team – Data Quality Issue Management (DQIM) Forum	– Perform an initial review to determine if the issue is a one-time occurrence or a repeated event; **a repeated event would qualify for additional data quality issue management and remediation activities** (i.e., DQ issue root cause identification, remediation strategy, execution and close out in GRC tool)
Data Quality Issue Prioritization	– 2nd line of defense Data Steward – DQIM Forum	– The data quality Issue is prioritized as **very high, high, medium or low** based on the following: • Issue Severity • Value and Complexity • Business Impact

Data Governance Issue Management & Remediation Process (Continued)

Process Step	Who is Responsible	Data Governance Issue Management – Phase Deliverable
Root Cause Analysis	– LOB Data Owner, 1st & 2nd line of defense Data Stewards and IT Subject Matter Expert	– Collectively research and identify the root cause factor for the reported data quality issue – Update the GRC tool to document root cause analysis work and update of controls and workarounds to mitigate risk
Remediation Plan & Owner Assignment	– LOB Data Owner, 1st & 2nd line of defense Data Stewards and IT SME – DQIM Forum	– Remediator and impacted stakeholders address the following: • Record of remediation path (Remediate or No Action) • Justification for No Action (Closure or Deferral) • Milestone based action planning (Remediate)
Remediation Plan Execution	– LOB Data Owner, LOB Data Steward and IT SME	– Execute plan to remediate root cause of data quality issue and update progress notes in the GRC tool – Conduct post-remediation testing to ensure the data quality issue has been fixed
Test and Close Data Quality Issue	– LOB Data Owner, LOB Data Steward and IT SME – DQIM Forum	– Complete post-remediation testing – Secure formal approvals to close the data quality incident from the remediator and affected stakeholders – Update the GRC tool to fully document close out activities and formally close the data quality incident in GRC