City of Baltimore Data Governance Handbook



**General Information**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Information | Version | Date | Status | Author |
| Initial Draft | Version 1.0 | July 4, 2016 | Draft Status | Nithya / Evette |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Project Contacts**

|  |  |  |
| --- | --- | --- |
| Contact Name | Title | Email |
| Evette Munro | Deputy, CIO | evette.munro@baltimorecity.gov |
|  |  |  |
|  |  |  |

## Table of Contents

[Table of Contents 3](#_gjdgxs)

[**1.1 Purpose** 4](#_30j0zll)

[**1.2 Overview** 4](#_1fob9te)

[**1.3 Focus of Data Governance** 4](#_3znysh7)

[**1.3 Objectives** 5](#_2et92p0)

[**2.1 Process Workflow** 6](#_tyjcwt)

[***2.1.1 Data Discovery*** 7](#_3dy6vkm)

[***2.1.2 Data Profiling*** 7](#_1t3h5sf)

[***2.1.2 Data Inventories*** 7](#_4d34og8)

[***2.1.3 Process Inventories*** 9](#_2s8eyo1)

[***2.1.4 CRUD Analysis*** 9](#_17dp8vu)

[***2.1.5 Capabilities Assessment*** 9](#_3rdcrjn)

[***2.3.6 Measure & Monitor – Proactive Monitoring*** 11](#_26in1rg)

[***2.3.6.1 Example for Proactive Monitoring - Health Monitor Dashboard for the data warehouse*** 11](#_lnxbz9)

[***2.3.7 Operational Dashboards*** 12](#_35nkun2)

[***2.3.8 Data Lineage Analysis*** 13](#_1ksv4uv)

1 Introduction

**1.1 Purpose**

The Data Governance Handbook is designed as a reference guide to help Mayor’s Office of Information Technology (MOIT) employees become familiar with Data Governance as implemented by MOIT and to inform them of associated responsibilities and processes.

**1.2 Overview**

Data Governance can be defined as an approach by which an organization can leverage data as an enterprise asset. Data Governance establishes the decision rights, data management policies, data procedures and data accountability that in turn foster a formal process in valuation, creation, storage, use, archival and deletion of data and information.

**1.3 Focus of Data Governance**

The focus on data policies, data standards, and overall data strategies are usually the first steps when an organization initiates a data governance function. The main activities of a data governance program with a focus on policy, standards, and strategy include:

* Identify stakeholders, establish decision rights, and clarify accountabilities
* Establish, review, approve, monitor policy
* Establish, review, approve, and monitor standards
* Establish Enterprise data strategies and governance.

**1.3 Objectives**

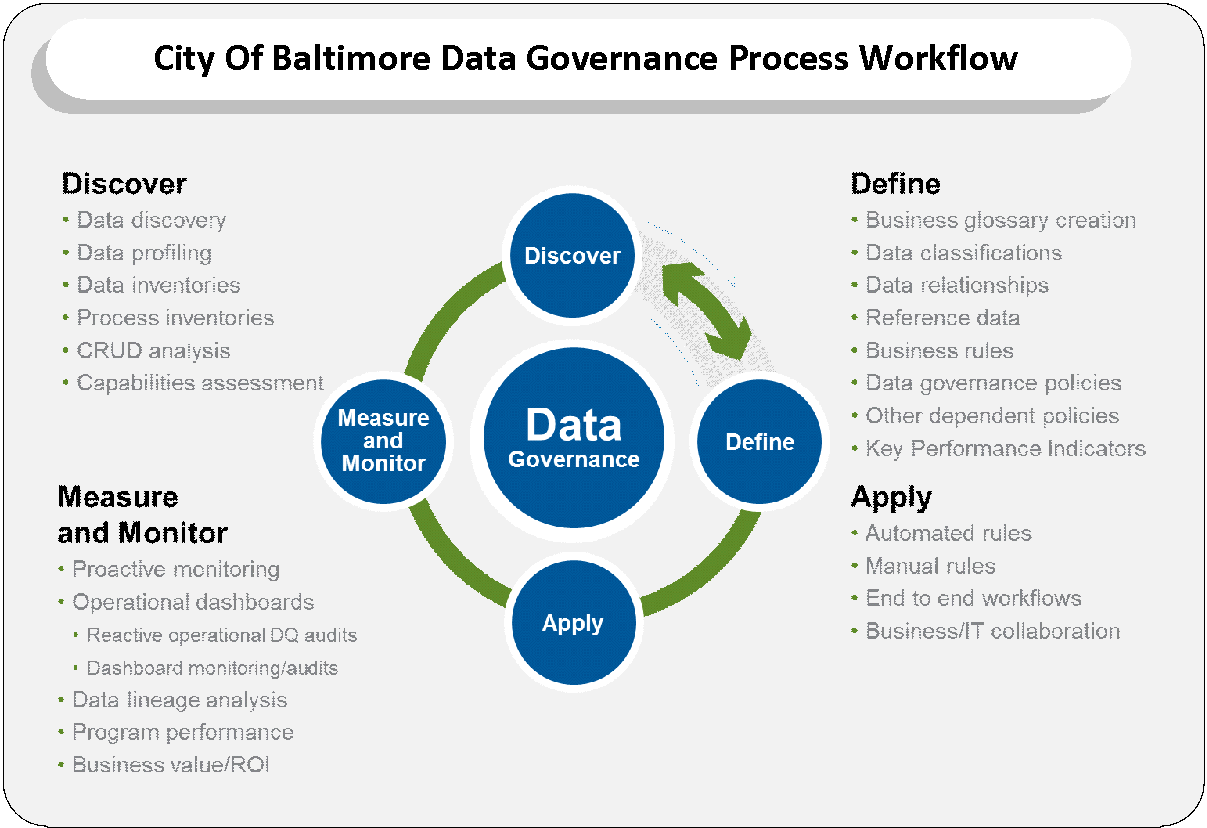
The main objectives of Data Governance are to -

* Establish tools, policies and processes to improve data quality and reduce data redundancy
* Protect sensitive data and develop agreements on data sharing
* Develop data usage agreements between agencies for inter-agency data
* Ensure data and IT compliance with federal , state, city and local regulations
* Encourage correct use of data
* Provide platform for robust data analytics and predictive analytics, thereby enhancing anticipatory intelligence
* Unlock the potential of Government data by strictly adhering to the Open Data Policy
* Facilitate better decision-making
* Reduce operational friction
* Protect the needs of data stakeholders
* Train management and staff to adopt common approaches to data issues
* Build standard, repeatable and reusable processes, thereby reducing costs
* Reduce costs and increase effectiveness through coordination of efforts
* Ensure transparency of processes
* Implement and establish the ‘Privacy for Design’ Data Architecture
* Ensure that the public is informed about what kinds of personal information City of Baltimore maintains; to limit the uses and disclosures of the information to those compatible with the law permitting its collection

2 Data Governance Process

**2.1 Process Workflow**

The below diagram illustrates the process workflow to achieve effective data governance in the City of Baltimore



***2.1.1 Data Discovery***

Data discovery is a business intelligence process for creating and using interactive reports for exploring data from multiple data sources. Data Discovery provides easy exploration across a big variety of data to provide the users with a broad new visibility into the agency’s performance.

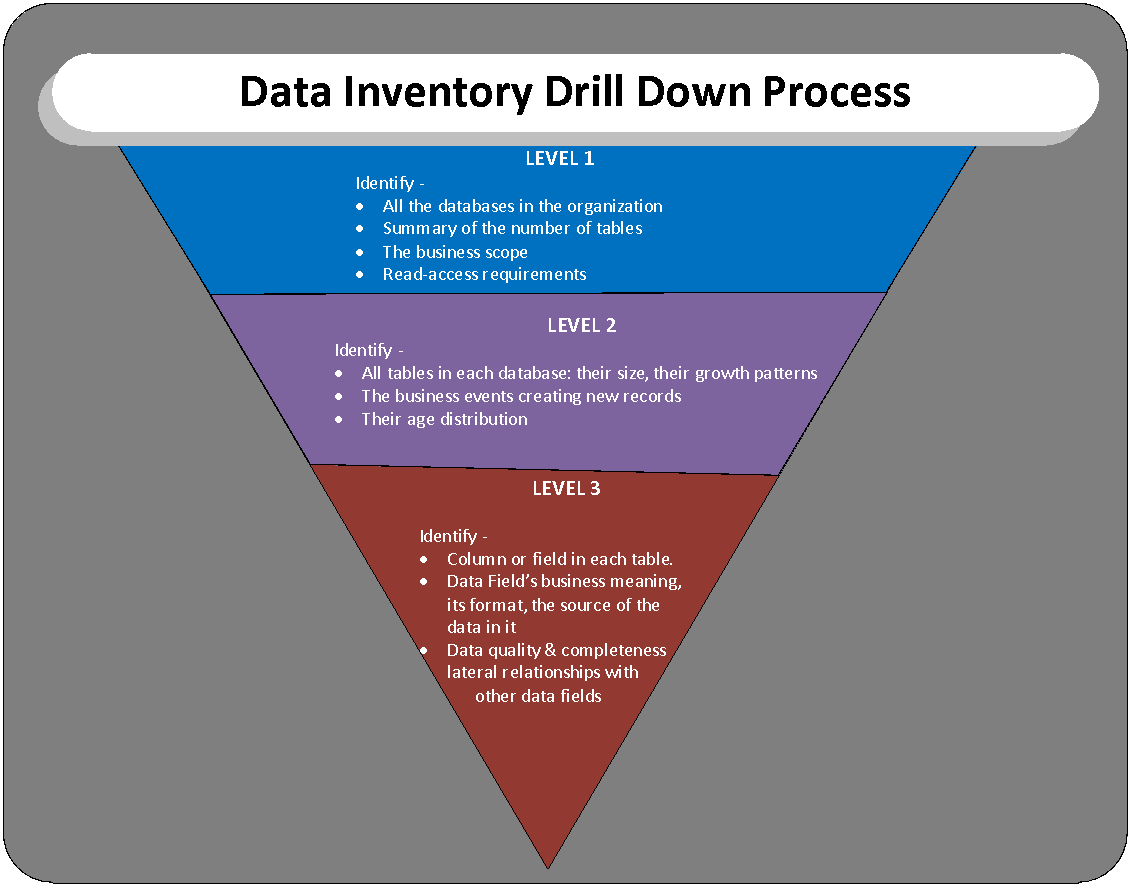
***2.1.2 Data Profiling***

Data Profiling is a process of identifying data issues in source data before they turn out to become big data problems. Data Profiling can be done with a variety of tools and solutions like Data Profilers and Data Explorers that can scan every single record from a particular source to find anomalies and hidden relationships. Data Profiling is helpful to give a complete and accurate picture of the agency’s data.

Implementing an effective Data Profiling process instills business confidence of providing an accurate, actionable data.

***2.1.2 Data Inventories***

The data inventory activity helps to survey the actual data behavior by inspection and is done by the data analyst in the team. The ‘Data Inventory drill down process’ has three levels as per the diagram below –



***2.1.3 Process Inventories***

Process Inventory Assessment is the process of identifying and documenting an inventory of the business processes that run in an organization. Process inventories are usually maintained at multiple levels, with major processes – high level processes, with more granular process steps defined beneath.

***2.1.4 CRUD Analysis***

Current state “Create, Read, Update, Delete (CRUD)” analysis maps the business processes, supporting applications and systems to the data to provide easy to understand visibility to the life cycle of critical data.

Understanding the current state processes, applications and stakeholders that have the ability to create, read, update or delete/purge/archive critical data is necessary to identify behavioral, system and policy changes required to improve the trustworthiness and security of that data.

***2.1.5 Capabilities Assessment***

Capabilities assessment is used to assess the current state of organizational competencies that data governance depends upon. This encompasses multiple dimensions including technology/architecture, functional, people skills and process/policies.

This assessment will allow data governance drivers to understand what organizational strengths can be leveraged, as well as provide insight to organizational weaknesses where investments can be made to mitigate risk and improve ROI of the Data Governance efforts.

The most relevant processes that comprise the Define stage include:

***2.1.6 Business Glossary Creation***

Collaborative process to capture and share full business context around critical data. In addition to the expected definitions of core data entities and attributes, context can also include rules, policies, reference data, free form annotation, links, and data owners. Ensures everyone is on the same page –data architects, modelers, developers, stewards and data consumers: business process owners as well as operational and strategic decision-makers.

***2.1.7 Data Classification***

The act of capturing relevant supporting business and IT context about data in the form of metadata. For unstructured content, classification plays a critical role in tagging and categorizing content with appropriate context to deliver relevant search results. Effective data classification delivers context and fast tracks information access to business users, allowing them to respond quickly to regulatory and compliance requirements, reduce costs and inefficiencies and gain improved insights into the business and customers. Trusted data classification benefits IT by reducing integration complexity, providing transparency often missing from black box/custom coding and ultimately improves collaboration, agility and time to value.

***2.1.8 Data Relationships Definition***

Process of defining data relationships, mappings and hierarchies at both the metadata (data modeling) and data (business hierarchies) levels. A data model without relationships is just a data inventory. Defining the expected relationships between master data, transactional data and reference data – and the applications and processes that depend on them – ultimately defines an organization’s business model.

***2.1.9 Reference Data Management***

Process for defining and standardizing reference data to be used within and across applications to ensure consistent data capture and usage.

***2.1.10 Business Rules Definition***

The process of creating and documenting logical business requirements to build the rules and policies for data validation, cleansing, enrichment, matching, merging, masking, archiving, standardization, etc. These rules define both the automated machine-supported and manual human-centric processes. When operationally implemented in the Apply process stage, business rules are the key to ensuring data is trusted, secure, and ultimately fit for business usage.

***2.1.11 Data Governance Policies Definition***

Process of defining data governance-driven policies such as data accountability and ownership, organizational roles and responsibilities, data capture and validation standards, data access and usage, as well as data masking, archiving, sub setting and retention.

***2.1.12 Key Performance Indicator (KPI) Definition***

Processes that define service level agreements (SLAs), operational baselines metrics for DQ and policy compliance, return on investment (ROI) and total cost of ownership (TCO) measures, and other measures used to define the effectiveness and value delivered from the data governance efforts.

2.2 Apply Process

The Apply process aims to operationalize and ensure compliance with all the data governance policies, business rules, stewardship processes, workflows, and cross-functional roles and responsibilities captured through the Discover and Define process stages. The processes that comprise the Apply stage include:

***2.2.1 Automated Rules***

The data quality, data privacy and other business rules and policies captured during the Define process stage are implemented into automated systems, processes and rules engines. These rules – once implemented – are the key to ensuring data is trusted, secure, and ultimately fit for business usage.

***2.2.2 Manual Rules***

Process that implements data quality, data privacy and other business rules and policies captured during the Define process stage into human-centric workflows and processes to manage exceptions and make decisions requiring high levels of confidence that automated rules are unable or untrusted to make. These rules – once implemented – are the best way to minimize the risk that a data exception will negatively impact the efficiency of your business processes, the productivity of your workforce, the quality of your decisions or the satisfaction of your stakeholders.

***2.2.3 Employment of End to End Workflows***

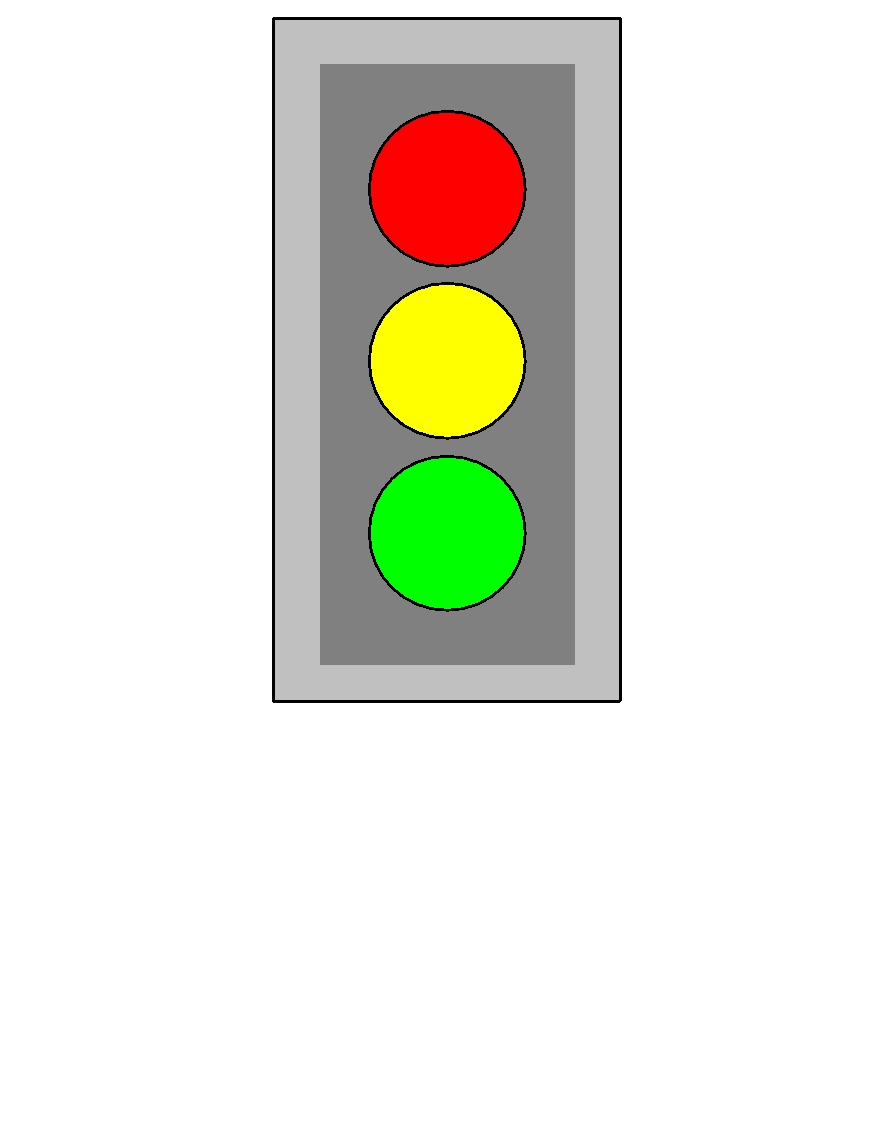
Apply and maintain workflow designs to support all four stages of data governance and stewardship processes. These workflows enable the necessary handoffs and collaboration required across business and IT stakeholders to enable truly holistic data governance.

***2.3.6 Measure & Monitor – Proactive Monitoring***

Proactive monitoring used in information technology permits teams to understand how information services are performing, along with identifying potential areas of risk 7 days a week, 24 hours a day. Monitoring can be applied to applications, networks, security, computers and data centers. Operationally, proactive monitoring provides a view into the health and status of IT services. They can act proactively while using the tools to benchmark performance. It also proactively monitors data quality or compliance exceptions as they are identified in real time as transactions and interactions are captured, in order to more quickly identify and mitigate critical issues that can cause costly process breakdowns. The City of Baltimore’s data warehouse architecture will comprise of a Health Monitor Dashboard for proactive monitoring purposes.

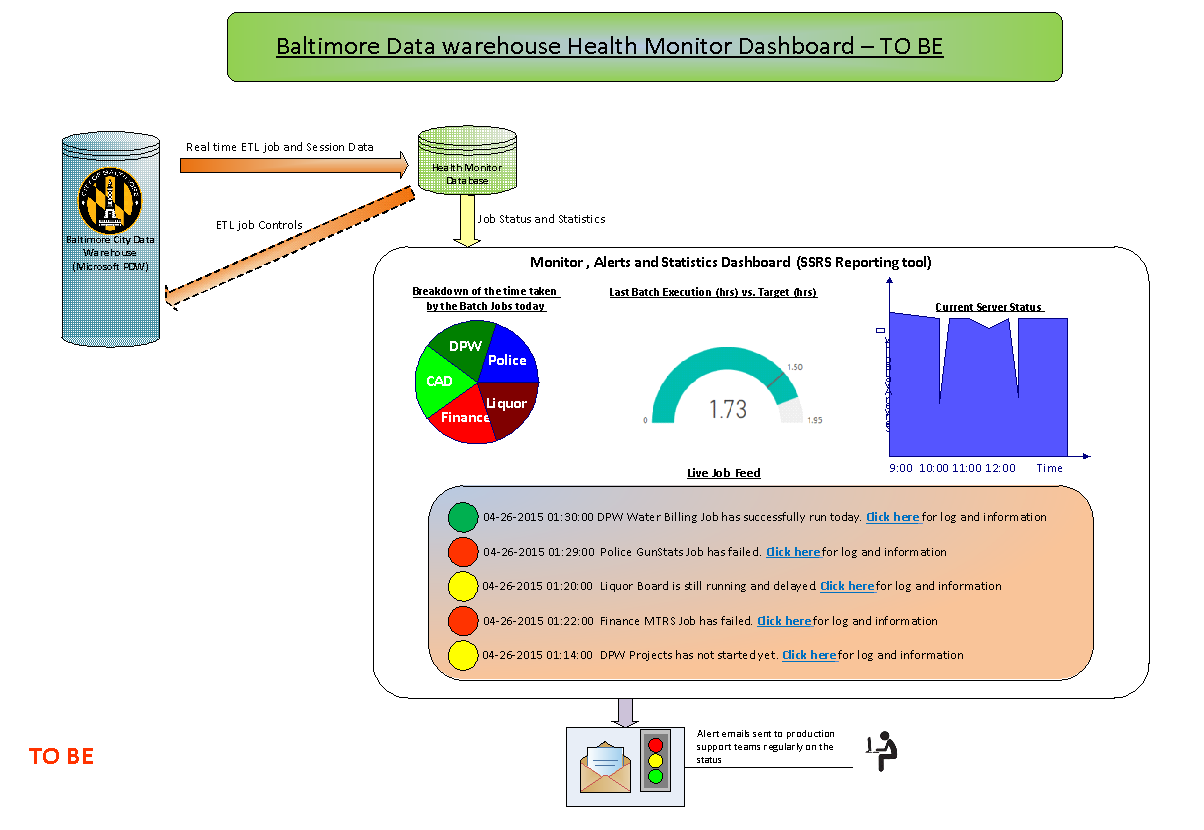
***2.3.6.1 Example for Proactive Monitoring - Health Monitor Dashboard for the data warehouse***

Health Monitor Dashboard is an essential component of the data warehouse. The dashboard has an underlying suite of tables, usually in star schema and they contain the details of every ETL (Extract, Transform and Load) job run at various stages of the process. The Health Monitor Tables are loaded for every run and they have dashboards built on top of them so that the development and production support team can monitor the status of the jobs everyday easily without having to go to the ETL tool monitor every time. Emails are set up from the Health Dashboard tables accordingly.



Dashboard Status:

* Red - ETL job failed. DEV & Production support team notified via email about the failure.
* Yellow- ETL job delayed / still running. An expected completion time is set for every ETL job. When it crosses the cut-off time, the status is ‘Delayed’. On the dashboard, it shows yellow in front of the job name
* Green- ETL job completed successfully within the expected time frame range.



***2.3.7 Operational Dashboards***

***2.3.7.1 Reactive operational DQ audits***

This process provides data stewards with visibility to reactively mitigate any data quality-related issues routed to them through predefined stewardship workflows implemented in the ‘Apply’ process stage.

Business and IT stewards alike are responsible for ensuring compliance with data policies, rules and standards, and when necessary are required to mitigate or reconcile a data quality, privacy or security issue. Proactive and reactive monitoring capabilities provide the visibility stewards need to observe and mitigate any issues.

***2.3.7.2 Dashboard monitoring/audits***

Data monitoring acts as an early warning system for catching data quality, security or privacy compliance problems before they wreak havoc on the dependent applications, reports, and processes. Combined with facilities to report on the state of data quality or data security problems, data monitoring ensures the right level of checks and balances are in place to quickly react to changes as needed.

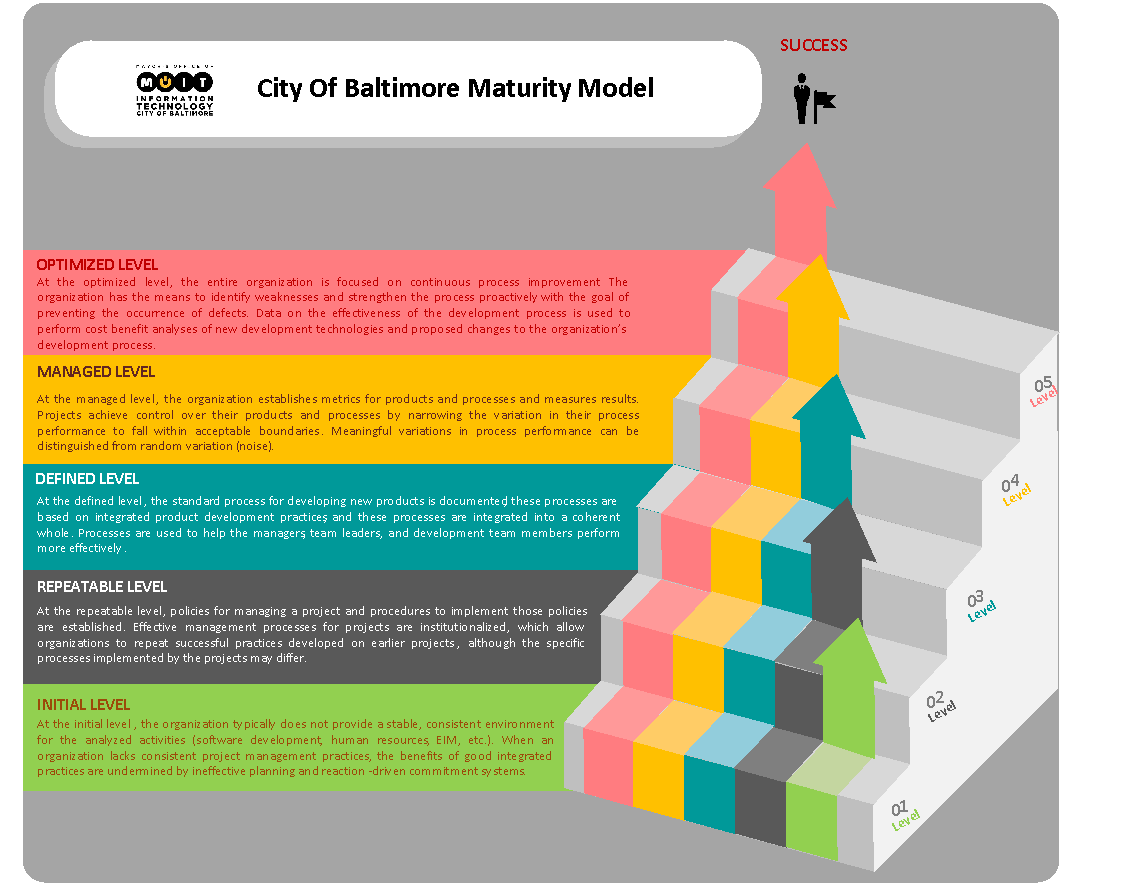
Themes for these operational metrics include data accuracy, completeness, integrity, uniqueness, consistency, standardization, and audits ensuring compliance with privacy and security policies.

***2.3.8 Data Lineage Analysis***

It is the process of identifying root cause, impact and performing data lineage analysis of data throughout its lifecycle. Having the ability to reconcile and provide transparency and visibility to the supporting metadata of the most critical data is a foundational element of the organization’s data management reference architecture. Data lineage visualization and auditing capabilities also allow data architects and stewards to effectively assess impact analysis of potential changes to data definitions, rules or schemas – as well as root cause analysis capabilities when responding to a data quality or security failure. This capability also provides transparency necessary to support auditability requirements of many regulatory edicts.

3 City of Baltimore Information Maturity Model (IMM)

The information maturity model (IMM) of an organization is a driver for data governance. The key aspect of any discussion around maturity is that the discussions in turn influence the definition of what the intended level of information maturity needs to be. There are definite stages of IMM along the way that can be described and measured. The IMM diagram is illustrated below. Best viewed at 250 % resolution -



3.1 Initial Level

At the Initial Level, the organization typically does not provide a stable, consistent environment for the analyzed activities (software development, human resources, EIM, etc.,). When an organization lacks consistent project management practices, the benefits of good integrated practices are undermined by ineffective planning and reaction-driven commitment systems.

3.2 Repeatable Level

At the repeatable level, policies for managing a project and procedures to implement those policies are established. Effective management processes for projects are institutionalized which allow organizations to repeat successful practices developed on earlier projects, although the specific processes implemented by the projects may differ.

3.3 Defined Level

At the defined level, the standard process for developing new products is documented, these processes are based on integrated product development practices and these processes are integrated into a coherent whole. Processes are used to help managers, team leaders, and development team members perform more effectively.

3.4 Managed Level

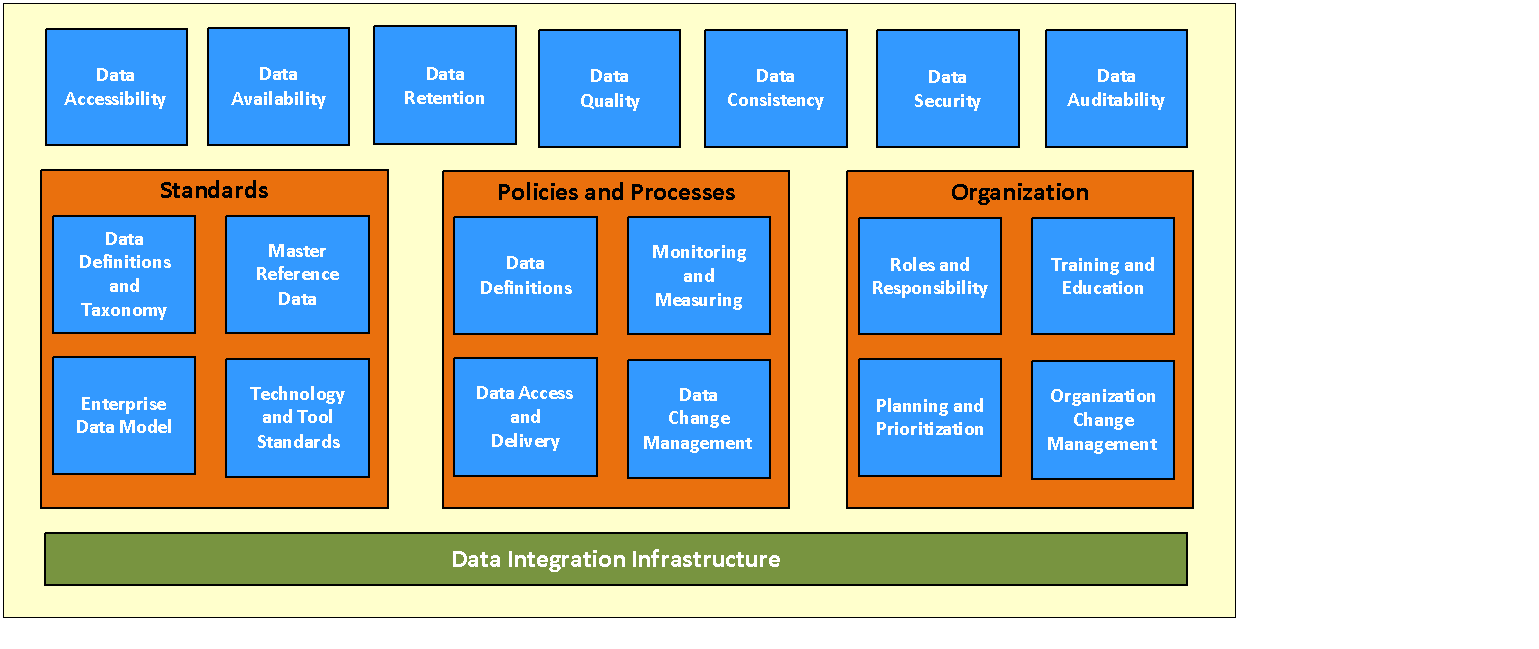
At the managed level, the organization establishes metrics for products and processes and measures results. Projects achieve control over their products and processes by narrowing the variation in their process performance to fall within acceptable boundaries. Meaningful variations in process performance can be distinguished from random variation (noise).

3.5 Optimized Level

At the optimized level, the entire organization is focused on continuous process improvement. The organization has the means to identify weaknesses and strengthen the process proactively, with the goal of preventing the occurrence of defects. Data on the effectiveness of the development process is used to perform cost benefit analysis of new development technologies and proposed changes to the organization’s development process.

4 Data Governance Model

The Data Governance Model is given below -



**4.1 Data Accessibility**

Data Accessibility is the capability to precisely and quickly analyze all the data within an organization so that key business decisions can be analytically-driven

**4.2 Data Availability**

The enterprise data warehouse is being accessed in real time. Hence, the data warehouse system is designed to be highly available. Data is transformed, loaded, validated and backed up in a very short time frame to provide around-the-clock and real time access.

**4.3 Data Quality Management**

Data Quality Management can be defined as the planning, implementation and control activities that apply quality management techniques to measure, assess, improve and ensure the fitness of data for use. Leveraging “systems of record” when delivering data while also supporting data quality practices like enrichment and standardization ensure the highest data quality.

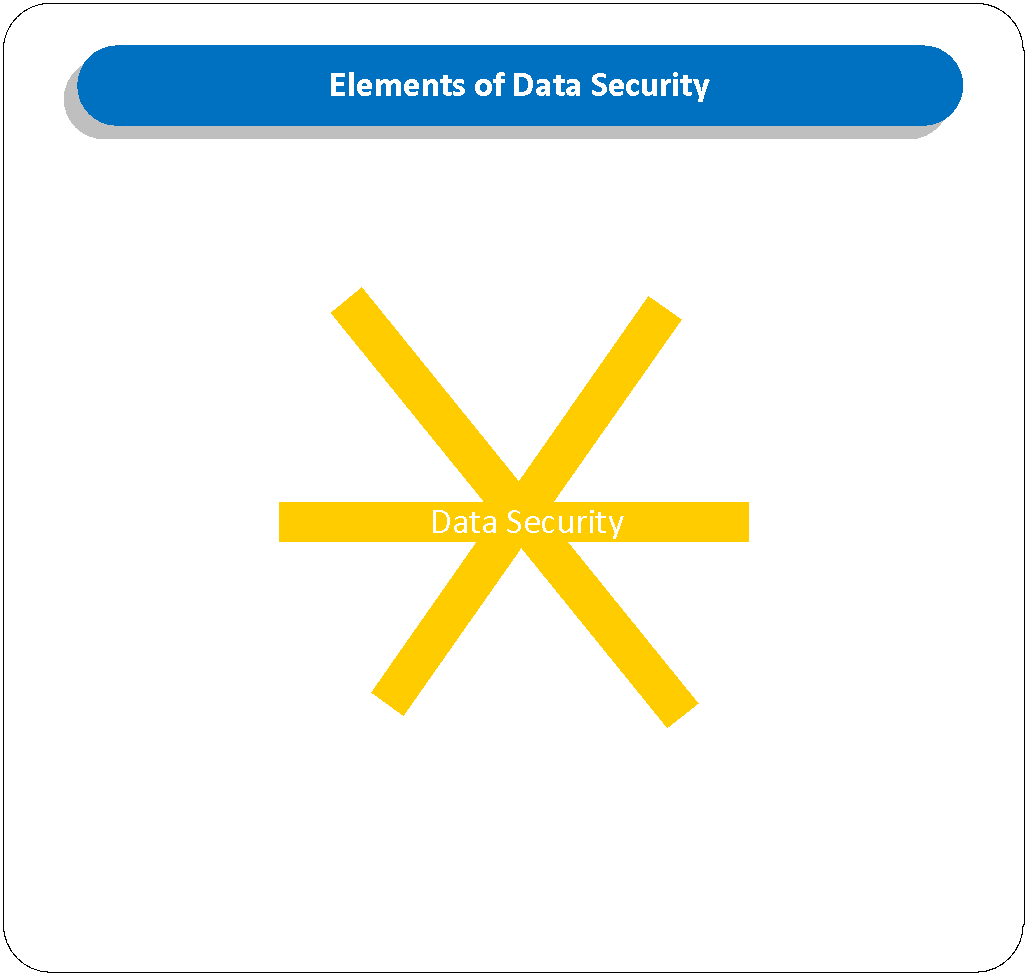
**4.4 Data Consistency**

Provide consistent and complete data applicable to all aspects of the business. By bringing data from disparate data sources into a centralized repository, the data warehouse provides a homogenized view of the organization's data. Users from across the organization making use of the data warehouse all view a single and consistent version of the truth. The data redundancy is prevented by utilizing highly normalized tables in the data warehouse schema, thereby ensuring that there is data no repetition of data across multiple tables and databases. The data warehouse utilizes the third normal form (3NF) to achieve normalization.

**4.5 Data Security**

Planning, implementation and control activities to ensure privacy and confidentiality and to prevent unauthorized and inappropriate data access, creation or change. Data Security is one of the core aspects of the data warehouse.

The data warehouse has its own set of challenges for security – the City’s Enterprise Data warehouse is a very large system that serves several user communities with varying security needs. The City data warehouse should be designed with a flexible, scalable, robust, stringent and powerful security infrastructure which enables security capabilities to seamlessly operate in the environment.



There are six elements of Data Security as mentioned in the diagram. They are explained as follows:

1. Network & Systems
   1. The front-end city applications must prevent unauthorized users from accessing or modifying data.
   2. The applications and the underlying databases must not be susceptible to data-theft by hackers
   3. Data must be available to the right users at the right time
2. Data Protection
   1. Data Protection is the process of safeguarding important information from corruption and / or loss.
   2. A data protection strategy includes operational backup of data, disaster recovery and business continuity.
3. Vulnerabilities
   1. Identifying vulnerabilities and protecting the network from highly sophisticated attacks is part of Data Security Hygiene
   2. Keeping software patched and updating regularly with anti-virus and anti-malware on a regular basis protects the network from attacks.
4. Access Control
   1. Data Security can be achieved by using role-based and attribute-based security administration.
   2. The roles should be assigned based on job capacity and function.
   3. The datasets will be available only as attribute-level datasets where certain fields can be accessed only by certain users
   4. Asking for use case for requested access will be one of the ways to limit and restrict access
   5. The city agencies may only view the data they have subscribed to. They should never be permitted to see the data of other agencies unless there are clear agreements between the parties
   6. The security infrastructure should ensure that employees of each division or agency should be able to view the data that is relevant to their own division, while employees in executive leadership positions / corporate positions will be allowed to view the overall picture
5. Monitoring
   1. Constant Proactive Monitoring of violations in data confidentiality, data contamination and limit unwanted data exposure. The system should keep a record of activities performed by its users
6. Policy
   1. Since the city’s transactional databases and the enterprise data warehouse (EDW) store personal identifiable information (PII) and Personal Health Information (PHI), privacy laws govern the use of such personal information. These privacy laws must be strictly adhered and implemented in the enterprise data warehouse.

**4.6 Data Auditability**

Data-related decisions, processes, and controls subject to Data Governance will be auditable; they will be accompanied by documentation to support compliance-based and operational auditing requirements.  When auditing is required, full lineage will be readily available anytime. The Data Audit Board will regularly audit and check for any vulnerabilities, unwanted exposure of personal identifiable information (PII) and Personal Health Information (PHI) in the city’s datasets.

**4.7 Data Standards**

While data policies guide what to do and what not to do with regard to data, data standards are more detailed rules on how to do it. Sample data standards include naming standards, data modeling standards, and other data architecture standards.

**4.8 Data Definitions and Taxonomy**

The general approach to defining data includes the following:

* Defining the tables, rows, and columns.
* Inserting index keys.
* Creating table relationships.
* Assigning data types.

Taxonomy is about "semantic architecture" - it is about naming things and making decisions about how to map different concepts and terms to a consistent structure. Taxonomy includes mechanisms for understanding context and making meaning precise.

**4.9 Master Data Management / Reference Data**

MDM comprises a set of processes, governance, policies, standards and tools that consistently defines and manages the critical data of an organization to provide a single point of reference. Alternatively, it can be defined as the process of consolidation resulting in one master record per entity or an individual.

**4.10 Technology and Tool Standards**

Upstream on-premises transactional/operational applications, systems, and processes that create, update or import data.

* Downstream on-premises analytical applications, systems, and processes that consolidate, reconcile, deliver, and consume data.
* Growth of off-premises sources and targets of data, including cloud-based applications and platforms, social data, mobile devices, third-party data feeds, sensor data, and Hadoop analytic environments.
* Supporting data management infrastructure investments that enable and ensure compliance with the organization’s unique requirements for delivery of “the right data at the right time with the right latency of the right quality and security in the right context.”

**4.11 Enterprise Data Model**

An Enterprise Data Model is an integrated view of the data produced and consumed across an entire organization. An Enterprise Data Model (EDM) for a traditional data warehouse represents a single integrated definition of data, unbiased of any system or application. It is independent of “how” the data is physically sourced, stored, processed or accessed. The model unites, formalizes and represents the things important to an organization, as well as the rules governing them.

For data modeling on Hadoop, the enterprise data model’s star schema can be flattened out into a few larger files and joining them with fewer key relationships in order to increase Hadoop’s performance.

**4.12 Data Policies & Processes**

Data policies are a collection of statements that describe the rules controlling the integrity, security, quality, and use of data during its lifecycle and state change.

For example, the Personal Information Protection Act (PIPA), Md. Code Ann. Comm. Law 14-3504 was enacted to make sure that Maryland consumers' personal identifying information is reasonably protected, and if it is compromised, they are notified so that they can take steps to protect themselves. PIPA contains provisions for notification of consumers in the event of a data security breach and for reasonable security measures to protect consumers' personal identifying information.

**4.13 Monitoring and Measuring**

The Measure and Monitor processes have the following functions -

* Capture and measure the effectiveness and value generated from data governance and stewardship efforts,
* Monitors compliance and exceptions to defined policies and rules, and
* Enables transparency and auditability into data assets and their life cycle.

**4.14 Data Access and Delivery**

Increasing access to data, across all members of the enterprise, including external stakeholders, members of the community, and especially users is a critical function of the Data Governance Committee. While the data security committee tends to protect data and restrict access to data, the Data Governance Committee should create a productive tension in the opposite direction. The data governance and data security committees are combined, thus forcing the members to balance the tension internally and streamlining what can otherwise be lengthy decision making and reconciliation between the two committees.

**4.15 Data Change Management**

Data Governance will support proactive and reactive Change Management activities for reference data values and the structure/use of master data and metadata. Policies and procedures need to be defined to govern change management.

**4.16 Training and Education**

Training and education is necessary and important to ensure that the organization has the knowledge needed to maintain high quality data for all the operations.

**4.17 Planning and Prioritization**

The key steps in planning and developing a data warehouse can be summarized as follows:

* Project initiation
* Requirements analysis
* Design (architecture, databases and applications)
* Construction (selecting and installing tools, developing data feeds and building reports)
* Deployment (release & training)
* Maintenance

Prioritization of BI efforts is the responsibility will take into consideration the following factors:

* Importance of business problems (both the strategic and operational importance of the organization's business needs)
* Data availability
* Organizational readiness

**4.18 Data Retention**

Data retention is a very important aspect of compliance. The data retention controls how the history of data is maintained, stored and made accessible when there is a need. The Data retention is performed for the following mandatory requirements –

1. To comply with data storage and retention laws/regulations
2. Data Backup and recovery refers to the copying and archiving of data so that it can be used to restore the original data after a data loss event
3. Retaining versions of data as it goes through its life cycle – creation, modification and deletion by the end-users. Also, it helps identify and store each and every version of a data item so that any version of data in the past can be retrieved at any time. The versions are implemented and stored using an incremental logic
4. In addition to data retention, the data records are filtered through a process of block-level de-duplication. This process identifies duplicate data, removes redundancies and reduce the overall storage and bandwidth usage (up to 90% in some scenarios)
5. Ability to reproduce (old) data as evidence- Example in cases of a litigation or court case
6. Data Retention policy can be applied at user level, group level or organization level and
7. Data Retention policy allows the administrator to control the following –
   1. The number of versions that can be created
   2. The specific duration of how long they can be stored
   3. The retention period for the versions for the datasets that have been deleted needs to be set by the administrators
8. Data history maintenance – The data in the data warehouse maintains history via Change Data Capture mechanism (CDC) and Slowly Changing Dimensions (SCD)

5 Role of Data Governance in Data Management Functions

The below diagram outlines the role of Data Governance –

Data Quality Management – specification, analysis, measurement and improvement

Data Architecture Management – Enterprise Data Modeling, Value Chain Analysis, Other Related Data Architectures

Data Development – Analysis, Data Modeling, Database Design and Implementation

Database Operations Management – Acquisition, Recovery, Tuning, Retention and Purging

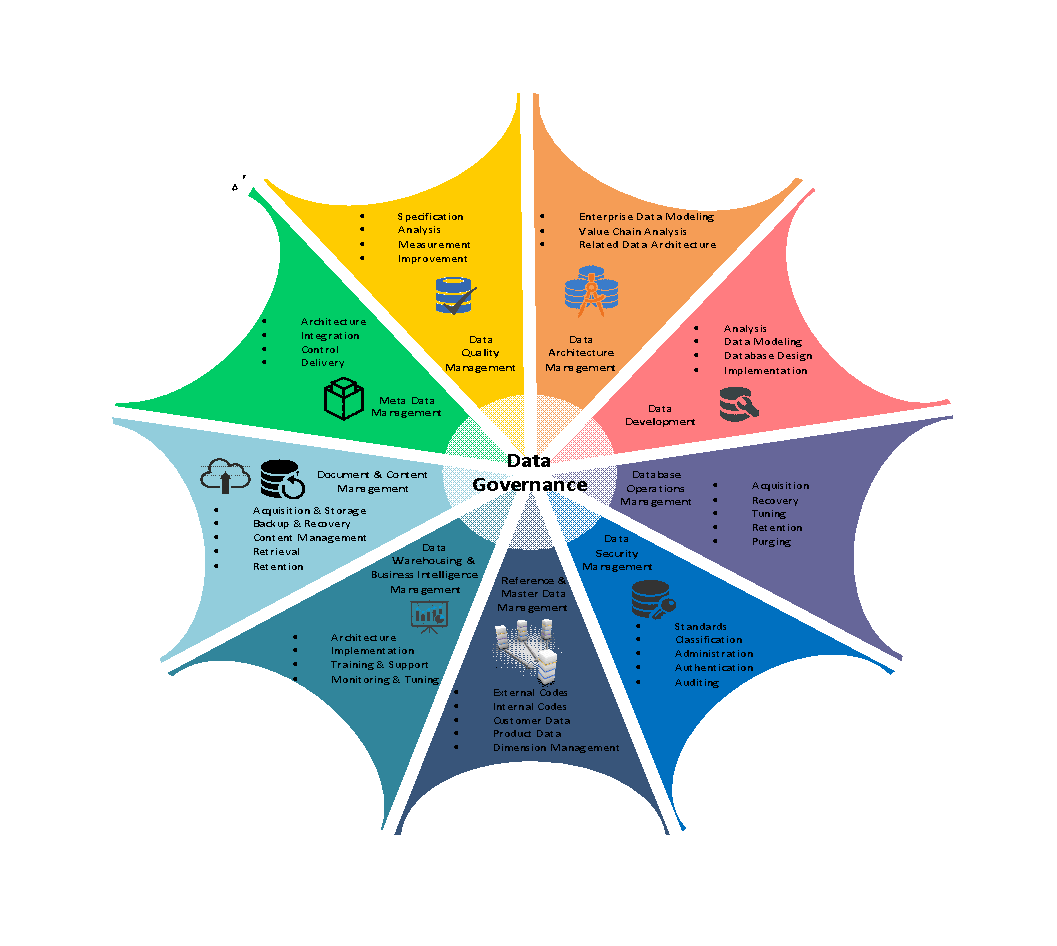
Data Security Management – Standards, Classification, Administration, Authentication and Auditing

Reference & Master Data Management – External Codes, Internal Codes, Individual Data and dimension management

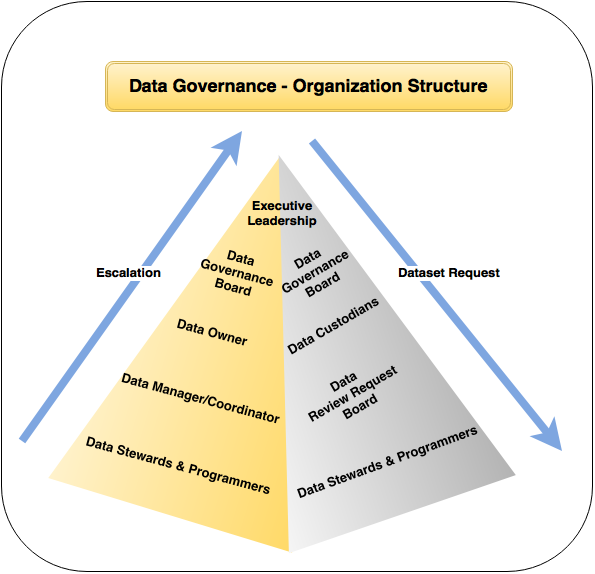
Data warehouse & Business Intelligence Management – Data warehouse Architecture, Implementation, Training & Support, Monitoring & Tuning

Document & Content Management – Acquisition & Storage, Backup & Recovery, Content Management, Retrieval & Retention

Metadata Management - Database Architecture, Integration, Control & Delivery



6 Organization - Roles and Responsibilities



**6.1 Data Stewards**

This group is comprised one or more individuals from each functional area who has knowledge of the business and data entry processes within their area.   Their primary responsibilities include:

* Participate in the data standards process
* Ensure that data are created and maintained according to business standards
* Work with technical teams to ensure data integrity and accuracy.

**6.2 Data Managers & Coordinators**

The Data managers and coordinators are comprised of the leadership from all functional areas of the organization that create and maintain data.  The data manager manages information resources internal to the organization and manages relationships with data consumers and data suppliers, both internal and external. Primary responsibilities include

* Participate in the data governance steering process
* Oversee the integrity of the data

**6.3 Data Owners**

The data owner role includes those members who have direct responsibility for entering and using data as part of their daily tasks.  Data owners participate in the data governance process as follows:

* Attend training relating to data standards and data entry.
* Be aware of standards
* Ensure manual data entry follows business standards
* Help identify and correct data which may not be entered correctly.
* Some data owners may participate in the data standards committee as data stewards for their functional area.

**6.4 Data Custodians**

Information Technology fulfills the role of data custodians. They are responsible for the security and availability of data.  They oversee the systems used to collect, manage and provide access to institutional data.  The primary responsibilities of a data custodian are:

* Maintain physical and system security including physical security of servers and data user security as determined appropriate by the Data Managers or Data Stewards.
* Ensure adequate system backups and disaster recovery plans.
* Ensure adequate system response times and system availability.
* Participate in setting priorities by providing details on technical, systems and staffing requirements related to Data Governance initiatives.

**6.5 Data Review Request BOARD**

The Data Review Request Board (DRRB) is a group of team members who are responsible for reviewing each dataset request, and by consensus, decide whether to approve or disapprove the release of the data being requested.

**6.6 Data Governance Board**

The Data Governance Board (DGB) usually coordinates and facilitates requests, conducts communications and serve as liaisons to technology and architecture teams. A member of the DGB will typically keep records of data stewards and other program participants.

**6.7 Executive Leadership**

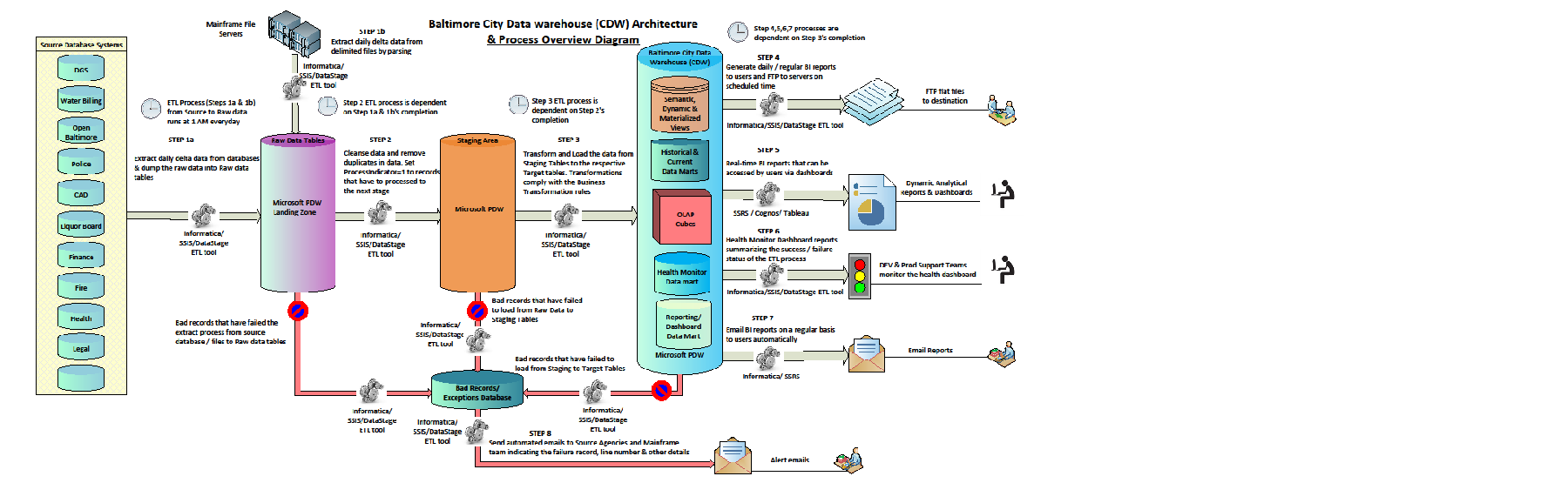
The Chief Information Officer (CIO) is the chief holder of accountability for enterprise information and is responsible for decisions regarding the acquisition, storage, and use of data. He or she is the ultimate arbiter with respect to dispute resolution between areas of ownership and is the ultimate manager of the definition and enforcement of policies.

**6.8 Identifying Data Governance Participants**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| City of Baltimore Team | Data Owners | Data Governance Board | Data Request Review Board | Data Steward | Data Steward Backup |
| MOIT | TBD | TBD | TBD | TBD | TBD |
| DPW | TBD | TBD | TBD | TBD | TBD |
| Liquor | TBD | TBD | TBD | TBD | TBD |
| CitiStat | TBD | TBD | TBD | TBD | TBD |
| Police | TBD | TBD | TBD | TBD | TBD |
| Other Agencies, etc., | TBD | TBD | TBD | TBD | TBD |

7 APPENDIX

**7.1 City of Baltimore data warehouse Architecture – to be**

****

**7.3 A Concise Data Governance Glossary**

**Access Management**

A discipline that focuses on ensuring that only approved roles are able to create, read, update, or delete data – and only using appropriate and controlled methods. Data Governance programs often focus on supporting Access Management by aligning the requirements and constraints posed by Governance, Risk Management, Compliance, Security, and Privacy efforts.

**Assurance**

Activities designed to reach a measure of confidence. Assurance is different from audit, which is more concerned with compliance to formal standards or requirements.

**Audit**

Audit is an independent examination of an effort to determine its compliance with a set of requirements. An audit may be carried out by internal or external groups.

**Audit Trail**

The audit trail is record that can be interpreted by auditors to establish that an activity has taken place. Often, a chronological record of system activities is used to enable the reconstruction and examination of the sequence of events and/or changes in an event. An audit trail of system resource usage may include user login, file access, and triggers that indicate whether any actual or attempted security violations occurred.

**CRUD**

Create, Read, Update & Delete. It is used to describe access rights for data.

**Change Control**

A formal process used to ensure that a process, product, service, or technology component is modified only in accordance with agreed-upon rules. The organization has a formal Change Control Board (CCB) that reviews and approves proposed modifications to technology infrastructures, systems, and applications. Data Governance programs strive to extend the scope of change controls to include additions, modifications, or deletions to data models and values for reference/master data.

**Compliance**

Compliance is a discipline, set of practices, and/or organizational group that deals with adhering to laws, regulations, standards, and contractual arrangements. Also, it is the adherence to requirements. Data Governance programs support many types of compliance requirements: Regulatory compliance, contractual compliance, adherence to internal standards, policies, and architectures, and conformance to rules for data management, project management, and other disciplines.

**Control**

Control is a means of managing a risk or ensuring that an objective is achieved. Controls can be preventative, detective, or corrective and can be fully automated, procedural, or technology-assisted human-initiated activities. They can include actions, devices, procedures, techniques, or other measures.

**Customer Data Integration**

It is an approach to managing multiple records containing information about an organization’s customers. In this approach, instead of combining all information into a single repository, a combination of technologies, processes and services are used to align information in multiple repositories.

**DGO**

Data Governance Office

**Data Architecture**

It is a discipline, process, and program focusing on integrating sets of information.

**Data Dictionary**

It is a database about data and database structures. A catalog of all data elements, containing their names, structures, and information about their usage, for the benefit of programmers and others interested in the data elements and their usage.

**Data Element**

The smallest piece of information considered meaningful and usable. It is also the single logical data fact- the basic building block of a Logical Data Model.

**Data Governance**

It can be defined as “The exercise of decision-making and authority for data-related matters, the organizational bodies, rules, decision rights, and accountabilities of people and information systems as they perform information-related processes”. Data Governance determines how an organization makes decisions — how we “decide how to decide.”

**Data Governance Framework**

It is the logical structure for organizing how we think about and communicate Data Governance concepts.

**Data Governance Methodology**

A logical structure providing step-by-step instructions for performing Data Governance processes.

**Data Governance Office (DGO)**

DGO is a centralized organizational entity responsible for facilitating and coordinating Data Governance and/or Stewardship efforts for an organization. It supports a decision-making group, such as a Data Stewardship Council.

**Data Mapping**

The process of assigning a source data element to a target data element

**Data Modeling**

The discipline, process, and organizational group that conduct analysis of data objects used in a business or other contexts, entities the relationships among these data objects, and create models that depict those relationships.

**Data Privacy**

The assurance that a person’s or organization’s personal and private information is not inappropriately disclosed. Ensuring Data Privacy requires Access Management and other data protection efforts.

**Data Stakeholders**

Those who use, affect, or are affected by data. Data Stakeholders may be upstream producers, gatherers, or acquirers of information; downstream consumers of information, those who manage, transform, or store data, or those who set policies, standards, architectures, or other requirements or constraints.

**Data Steward**

A person with data-related responsibilities as set by a Data Governance or Data Stewardship program. Often, Data Stewards fall into multiple types. Data Quality Stewards, Data Definition Stewards, Data Usage Stewards, etc.

**Decision Rights**

The system of determining who makes a decision, and when, and how, and under what circumstances. Formalizing Decision Rights is a key function of Data Governance.

**Enterprise Architecture**

Enterprise Architecture (EA) is a comprehensive framework used to manage and align an organization’s business processes, information technology (IT) software and hardware, local and wide area networks, people, operations and projects with the organization’s overall strategy. Enterprise Architecture is often subdivided into four architectural domains: Application Architecture, Business Architecture, Data Architecture, and Systems Architecture. Other types of architectures (security, compliance, controls, etc.) may be considered as part of EA, or they may be aligned with EA.

**GRC**

An acronym for Governance, Risk, and Compliance used by management to acknowledge the interdependencies of these three disciplines in setting policy

**GRC-SQ**

An acronym for Governance, Risk Management, Compliance, Security, and Data Quality, used often by Data Governance and Data Quality programs to acknowledge the interdependencies of these five disciplines in managing data.

**IT Governance**

Information Technology governance can be defined as “the leadership, organizational structures, and processes that ensure that the enterprise’s IT sustains and extends the enterprise’s strategies and objectives.”

**IT Infrastructure Library (ITIL)**

A series of publications providing Best Practice guidance for IT Service Management

**IT Portfolio Management**

A key function of IT Governance, IT portfolio management is the formal process for managing IT assets such as software, hardware, middleware, an IT project, internal staff, an application or external consulting.

**IT Service Management (ITSM)**

The implementation and management of quality IT Services that meet the needs of the Business. IT Service Management is performed by IT Service Providers through an appropriate mix of people, Process and Information Technology.

**Information Architecture**

In its broadest definition, a discipline, process, and/or program focusing on the design and organization of data, unstructured information, and documents. In the context of Enterprise Architecture, it is a synonym for Data Architecture, which is one of the four Enterprise Architectures (with Application Architecture, Business Architecture, and System Architecture). In the context of designing documents and web pages, it is the structuring of large sets of information, as opposed to the development of the content of any content unit within the larger set.

**Issue Framing**

A process for scoping and defining a problem prior to solving it. How a decision is framed limits the possible choices that are seriously considered.

**Issue Resolution**

A structured process for reaching a solution to a problem while considering the needs of all stakeholders.

**Master Data**

Master Data are the “nouns” upon which business transactions take action. Master Data describes core entities of an enterprise that are used by multiple business process and IT systems.

**Master Data Management (MDM)**

A structured approach to defining and managing an organization’s Master Data.

**Metadata**

Data about data. The definition and scope of metadata depends upon context. In the context of Information Management, metadata is generally thought of as providing information (what database stores it? what data type is it? how long is the field? etc.) about a data element. Within the context of Data Governance, the term also includes “business” metadata such as the names and roles of Data Stewards. Metadata repositories are employed to store and report on metadata.

**Post-Compliance Paradigm Shift**

Change in expectations that say that it’s no longer acceptable to simply “do” work. Instead, for work that exists in an environment with compliance requirements, the work is not complete until you 1. Do it, 2. Control it, 3. Document it, and 4. Prove compliance.

**Risk Management**

In a broad sense, to assess, minimize, and prevent negative consequences posed by a potential threat. The term “Risk Management” has significantly different meanings that can affect Data Governance programs. At an organization level, “risk” refers to many types of risk (operational, financial, compliance, etc.); managing risk is a key responsibility of organization Boards and Executive Teams. At a project level, “Risk Management” is an effort that should be undertaken as part of Project Management, focusing on risks to the successful completion of the project. From a Compliance/Auditing/ Controls perspective, “Risk Assessments” and “Risk Management” are high-effort activities included in the COSO, and COBIT frameworks and required by Sarbanes-Oxley and other compliance efforts. Data Governance programs may be asked to support any of these Risk Management efforts, and may need input from these efforts to resolve data-related issues.

**Rogue Data Usage**

Accessing or using information in a manner that is not authorized or proper.

**Sensitive Data**

Data that is private, personal, or proprietary and must be protected from unauthorized access.

**Tone from the Top**

Explicit or implicit messages sent by an organization’s leadership. To be successful, compliance and governance programs generally require a strong tone from the top about expectations for participation.

**Workflow**

The movement of data, documents, or tasks through a work process; generally used in the context of technologies that automate workflows. Data Governance programs often strive to address workflows by embedding governance controls (e.g., approvals, decision steps) or by providing loop-outs to governance processes (e.g., issue resolution, change control)