

Tejasvi Addagada

# Simple but Efficient Data Management Strategy

*First published by Tejasvi Addagada in 2017*

*Copyright © Tejasvi Addagada, 2017*

*All rights reserved. No part of this publication may be reproduced, stored, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise without written permission from the publisher. It is illegal to copy this book, post it to a website, or distribute it by any other means without permission.*

*This book was professionally typeset on Reedsy*

*Find out more at [reedsy.com](https://reedsy.com)*

# Contents

Pre-face

## I. Part One

Focus of Data Management, Stewardship and Governance

Discovering the current state environment

Define a target state service model

Creating a Metadata Service

Running a Data Quality Service

Decide to drive by Policy or Guidelines

Assessing Value from Data Management & Governance

Measuring Benefits

# Pre-face



# Part One

# Focus of Data Management, Stewardship and Governance

The needs of firms including financial services are seeing an increased need for regulatory preparedness as regulations around data are quickly evolving. The business environmental changes including strategic changes and advancements in technology and business models are some primary drivers for these evolving needs. Data Governance is now a new normal in most enterprises as demanded by regulations similar to BCBS 239, GDPR, EU No 1024/2013, EMIR, GDPR and MiFID2.

In most of the published books by leaders, on Data Management Industry, the context in which Data Management plays an important role in enabling business decisions is stated. But, to understand the Data Management space better, going back to the origins is required.

## Back to Origins

**Data Management** is often said to be the science of actively managing Data definitions, Quality, privacy, Architecture, life cycle along with the value and Risk related to data. It also entails the process of obtaining people, process and technology capabilities to manage data consistently across Enterprise.

**Data Governance** is mostly understood to be an oversight on data management activities to ensure that policy and ownership is enforced. The emphasis is on formalizing the data management function along with the roles and responsibilities. In addition, Governance also ensures that data management as a function is in place in order to carry out required activities and to make sure that data management is sustainable in an organization.

**Data Stewardship** by contrast assists by enforcing formalized roles, responsibility and accountability on people like data owners. This helps push active Data Management through the Grass roots up to leadership while also ensuring that the required data management activities are carried out.

In the 1800s, horses were trained to pull carriages from London to Farringdon, which is to say that these trained horsemen were in the business of transferring people from place to place. This is an active process that needs to be managed so that passengers reach their destination on time, a service that also includes the proper management and care of the horses. The outcome of this process is clear: the passenger reaches the desired destination, but the *benefit* of the service is that the passenger reaches the destination on time and remains safe and secure upon arrival.

There also needs to be an oversight over all the horses, drivers, and carriages owned by the horsemen or perhaps borrowed from a third party. Next, there is the need to manage the risks, which includes safety and security issues as well as missed timelines, all of which affects the profitability of the enterprise. Think of all the details involved that the company running the travel business is responsible for: established guideposts all along the way to assess, monitor, and guide the drivers in scenarios of bad weather, faulty carriages, and alternate routes to take when inclement weather interferes with the standard route.

Likewise with the analogy just described, governance is akin to evaluating, monitoring and directing all the aspects of Data Management: planning, acquiring capabilities, building the service, and then implementing and supporting the service or enterprise.

A common question that I often ask is "*Does your organization have a model that evaluates the benefits of Data Governance?*". Sometimes, firms assume that "*building value from data*" is similar to "*getting the right data fit for purpose*". Though, the latter can be quoted as a necessity to get value from data. Data Management professionals in the industry could state the value from data governance, with ease. But, there is quite a change needed in the perspective of viewing value from data governance activities.

An understanding of Benefits

A benefit is commonly described as – “*An outcome of change which is seen positive by a stakeholder*”. Comprehending the value from an outcome is in-fact making it objective and measurable. Often, coming up with outcomes like “*Making data fit for purpose*” or “*increased awareness*” is not the ask to mature data governance divisions. But, how the regulatory losses are controlled with good quality data used in reporting.

Why don't we think of a cook, you have employed as providing you with a service. The value from this service is just not to satisfy your hunger but also to provide you with rich nutrition, on time, while also keeping you away from the risk of food poisoning. Value is realized only when it is monitored and measured. In the above analogy, one can convert nutrition to daily required limit in mg; “*poison free*” is nothing but mitigating the risk which can be measured by number of food poisoning incidents; A count of missed instances of having to serve plates on time, to get your enzymes right is another measure.

## Focus of Data Governance today

The lack of focus is vividly seen in data governance divisions and is constantly affecting how the organization pursues data activities. There are immediate and cumulative benefits from Data Management and Governance dimensions either metadata management or data quality. Some data governance divisions kick-start their initiatives with the right Business cases that rightly overcome organizational challenges, these business cases should clearly articulate tangible benefits of using these data services. What is it to the finance division, if the Data Accuracy improves from 60% to 98%? Does that mean that the financials that are made public need not be corrected after they are published, thus savings on compliance costs. This is feasible in firms that have known issues with Data. But some organizations have challenges in trusting their data, though their data is of sufficient quality. In such cases, benefits would be monitored and measured on a continuous basis in accordance to an assessment plan which captures the qualitative benefits as well. Most organizations are not orchestrating data management activities as projects or programs but rather as continuous push or pull based services like Data delivery services or Data Definition services. It is

strongly recommended that the function have an assessment plan before starting a data governance service be it data quality or metadata management. This plan should bring out the approach to monitor and measure the value of orchestration over specific timelines.

Well, think of every data management activity like data quality assessment as an enabler. An enabler, in simple sense, is a new or an improved capability made available to the organization to fulfill a part or a full need. These enablers can be further classified into Business, Process and Technology enablers. For example, “Policy making” is a business enabler, “Metadata service management” is a process enabler while “Data profiling” is a technology enabler.

A common challenge that CDOs today face is having these Business divisions own the metrics that monitor the controls as well as the value from data management activities. While there are common enterprise benefits like reduced operational costs and risk, there are benefits that weigh in directly with the value chains of the divisions like client service effectiveness.

## Discovering the current state environment

It was a pleasant news that I was requested to travel to where the most of the horses live, to perform a current state analysis of the enterprise. The investment management firm I was working with, facilitated the travel booking. The COO always had a strong opinionation on increasing the trust in data used for regular operations. The executive owner of Data management, Robin was based out of Rancho. She always said "*we do not have a data quality problem, but we have problem with the trust in data*". I was to put an operating model at the end after the analysis that will help the firm gain trust from grassroots.

It was september, with the winter nights relatively dry when I took a flight from hyderabad to Sacramento. I had to take a hop as Seattle with a layover of 8 hours, giving me a good nap. I toured the airport in Seattle, got a good nap in the lounge. With a long flight of 38 hours, I was particularly not energetic, but thankfully excited, Emirates helped with a seat with ample leg space.

As I landed in Sacramento, I did take a quick cab to Marriott Inn where I was put up, close to two weeks. To my disappointment, my travel card was not punching in the Point of Sale machine in the reception. Technology has advanced while it has brought us challenges along as well. I did not get myself a phone as it was late in the night, though I could have scooted to wal-mart hailing an Uber. But, I decided to check in, got to set up my wifi, skyped the bank. I understand from the bank that the block is a extra security capability enabled to stop cross border fraud. Well, I always worked on Financial Crime Compliance and could not complain.

Robin could not come to pick me up in the airport as her daughter was getting engaged the same day I travelled to Rancho. I helped myself with a good night's sleep to ease myself of the lag. I was startled to see her coming over the next day

early morning to pick me up to the office. We had an early lunch to get rid of the Monday blues, and we started discussing the approach to strategy analysis, over some Sushi. I was intrigued by the Sous chef's ownership of having to sculpt the fish to prepare my sushi plate. Later, I came to know from Robin, the standards the restaurant had in sourcing the catch, cooking the rice, cooling, adding vinegar, and maintaining it at the perfect temperature and assembling the plate together delicately. I was rather amused by the communication between the head chef, the sous chef and the rest of the restaurant kitchen personnel. This reminded me of the outcome that I was expecting from the Data Quality strategy. We further were discussing over fried Calamari along with Wasabi cubes, the approach to the strategy analysis. The wasabi always has a burst of flavours in my mouth ranging from sweet to spice that tingles. This is when I got myself thinking why not use a range of techniques from interviews, focus groups and Documentary analysis along with questionnaires to understand the current state of Enterprise.

The Strategy analysis for Data Management and Governance emphasizes on defining the transition and future data governance states that the organization must harness. This target control environment is needed to address the business needs related to data. The analysis covers the strategic thinking in Data Analysis and Business analysis as well as discovery of solutions to enable the organization to harness incremental value. If the outcome of a cultural change is difficult to predict in an organization, the strategy needs to focus more on mitigating this risk, testing assumptions, and changing course until the capabilities that will succeed in reaching the objectives can be identified. It was autumn and though the climate was predictable, it has not rained or snowed over the mountains for two years now. People were experiencing drought like situations with forest fires, lack of water pools and dried up grass in the golf courses. What about preparedness for such scenarios? I heard that even veteran firefighters have challenges predicting fire behavior. A range of techniques from satellite mapping, weather information, remote sensing data collection, cleansing, evaluation, modelling, and communication have changed the nuances fire behaviour analysis. I understand the importance of quality data and its operations as it is used to save lives and reduce threat to properties.

It is required to understand the reasons why an enterprise needs to change some aspects of how it operates today. Is the data quality being assessed for regulatory related data while customer data which is a priority for business enablement, is not assessed. Today, the Bank's customers, shareholders and regulators require a documented approach to data management to understand how the data operations are undertaken.

## Organizational Structure and Culture

An Organizational model defines the formal relationships and functional hierarchies among people working in the enterprise. While communication channels and existing relationships are not limited to the organizational model, they are highly influenced by it.

Organizational culture is the beliefs and values shared across the people in the firm. These beliefs drive the actions performed by an organization. The beliefs of people also decides whether the wanted behavior for data related operations, in an organization, should be driven by Guidelines or policy. But, there is always a need to perform a cultural assessment to:

1. Identify if cultural changes apart from the data driven cultural change is required to better achieve the outcomes.
2. Identify whether stakeholders understand
  - Rationale for the future state of the enterprise
  - Value delivered by it
  - Identify the attitudes of the stakeholders to the current needs
  - Capabilities and Processes

The existing Capabilities like Data profiling along with processes and activities like Metadata management describe how an organization operates today. A capability-centric view is required that must leverage existing capabilities to produce a new outcome with less costs. A capability-based view is useful in this situation because capabilities are generally organized in a functional hierarchy with relationships to other capabilities, making it easier to identify any gaps.

## Technology and Infrastructure

Information systems used by the data operations personnel in executing processes, making decisions, and in interactions with data suppliers and customers need to be assessed.

## Policies

Policies define the boundaries and guidance to decision making in an organization. They in-fact harness predictable behaviors in the personnel through guidance, using best practices, values and ethical ways of taking a decision that produces a desired outcome. They generally address routine operations rather than strategic changes. But this is also fast changing with policy defining decisions to be taken in strategic changes. An analysis is required of the current policies that have an overlap with the current data governance program. This is to address any policy protection aspects or duplication of outcomes between divisions.

## Internal assets

There is a need to identify the organizational assets used in the current state of operations. The resources can be tangible or intangible, such as financial resources, patents, reputation, and brand names which need to be a part of the analysis. Further, the relationship among the assets as put forth by people to develop funding models is to be analyzed.

## External Influencers

There are external influences on the enterprise that do not participate in a change. But, they might present constraints, dependencies, assumptions or drivers on the current state. To state a few, the influencers are Industry Structure, Competitors, Customers, Suppliers, Political, Regulatory, Environment, Technology and Macroeconomic Factors

## Techniques

There are a plethora of techniques that one can use to come up with a current state strategy analysis. Most of these techniques in detail are published in the BABOK from IIBA.

## Business Capability Analysis

This technique is used to identify gaps and prioritizes them in relation to value and risk associated with each capability.

## Business Model Canvas

This technique provides an understanding of the value proposition along with the critical success factors in delivering that value. It also helps analyze the resulting cost and revenue streams. It is also helpful in understanding the context for any change and identifying the challenges and opportunities that may have significant impact.

## Business Cases:

This is more of an artifact or deliverable while it is also a technique used to capture information regarding the business needs and opportunity.

## Concept Modelling

This technique is used to capture key terms and concepts in the business domain and define the relationships between them.

## Document Analysis

This technique focuses on simple analysis of existing documentation about the current state, including documents created during the implementation of a solution, training manuals, issue reports, competitor information, supplier agreements, published industry benchmarks, technology trends, and performance metrics.

## Financial Analysis:

used to understand the profitability of the current state and the financial capability to deliver change.

## Focus Groups

This technique focuses on having to understand the the current state of the control environment.

## Functional Decomposition

This technique breaks down complex systems, processes or relationships in the current state that will assist in further understanding of the organization.

## Interviews

A common technique that facilitates dialogue with stakeholders to understand the current state and any needs evolving from the current state.

## Lessons Learned

This technique enables the assessment of failures and opportunities for improvement in past initiatives, which may drive a business need for process improvement.

## Metrics and Key Performance Indicators (KPIs)

This technique assesses the performance of current state of an enterprise. The metrics are used to assess the current state capabilities of the organization.

## Organizational Modelling

This technique describes the roles, responsibilities, and reporting structures that exist within the current state organization.

## Process Analysis

The technique identifies the current state of the processes and activities. If required the technique can be clubbed with other techniques like interviews and focus groups to discover the existing process in case, existing documentation is not available. While performing process analysis, the opportunities to improve the current state are un-covered.

## Process Modelling

This technique describes how work occurs within the current solution.

## Risk Analysis and Management

This technique assists in putting the risks associated with the current capabilities to deliver their success factors. Further, any compensatory controls needs to made available.

## Root Cause Analysis

This technique provides an understanding of the underlying causes of any problems in the current state to further clarify a need.

## Scope Modelling

The technique helps define the boundaries on the current state description of the capabilities.

## Survey or Questionnaire

The technique assists in gaining an understanding of the current state from a large, varied, or disparate group of stakeholders.

## SWOT Analysis

This technique evaluates the strengths, weaknesses, opportunities, and threats to the current state enterprise.

## Vendor Assessment

This technique determines whether all vendors are adequately meeting commitments as put forth in their agreements like Data Quality thresholds, or if any changes are needed.

## Workshops

This technique engage stakeholders to collaboratively describe the current state and their needs.

Finally, with use of these techniques, I was able to capture the information that has helped me craft a current state analysis of the control environment. Next, I started using the heat map to classify the capabilities (organizational or divisional) based on their performance gaps and risks. This helps in planning in the future state and in the transition state.

\*\*\*

On the Friday before I had to travel back to Hyderabad, Robin surprised me by taking me on a wine tour as we ended our day in the office early. The weather was

perfect; there were trees shedding bright-orange leaves all around the buildings. It was a sunny and a particularly dry day as we started on the wine tour at around 12 noon. Our first stop was 10 miles into the country side that overlooks usually snow ladden mountains. As we arrived at the vineyard, there was a huge gathering of people around the tasting area. We decided to get ourselves off there and drive further to an older vineyard that hosted a winery that dates back 200 years. Back in my hotel, as an avid news buff, I was watching the pope's visit to States. And the vineyard that we were in, has its wine specially made for the travelling Pope. There was a young lady in her late 20's who was dressed in delicate hues of blue, as she was to join someone on a date. Or, perhaps, it is to appeal to the visitors, to get to enjoy the tasting better, is what came to my mind.

I quickly reached to my room after the wine tour. Later, in the evening, I had a great dinner on the pier with Robin's family. The pier was wonderfully decorated, with private cruises and Yachts adorning the bay. During sunset, the seating space clad in white cloth waving in the light wind, played with the vibrant colors of the sky. The restaurant adjoining the pier served us a soulful dinner but I particularly liked the minestrone soup and the Risotto that accompanied it. Robin was hosting her daughter's marriage in the same restaurant by the river. That day I ended the documentation of current state analysis of the control environment.

Based on the current state analysis, below are some challenges that firm was going through –

- The customers and shareholders do not have a documented approach to data management that would help them understand how the data operations should be performed in line with the best practices.
- Regulatory preparedness from regulations like MIFID II, require the focus to be on having quality data provided.
- A lack of emphasis on the ownership and accountabilities for their data. Tackling shared ownership between Business Units.
- Based on the culture, the guidelines are a best way to implement data management, but the guidelines do not clearly state responsibilities and accountabilities of data stakeholders, in a consistent way.

- The internal customers' reduced trust in data is leading to quality analysis in silos over the same redundant data across Business Units, and that is increasing the operating costs.
- There are no published processes for Data Quality monitoring, data definitions, and architectural decisions integrated into the projects and regular operations.

## Define a target state service model

After performing the current state analysis, I moved on with my task of assessing the necessary conditions and capability improvements to meet the business needs and outcomes. To add more to the context, I am writing my experience and best practices I have followed for a global firm with operations in more than 60 countries.

\*\*\*

From the plush landscapes of the office in Bangalore, I was enjoying a mid-day mocha with the practice manager while he briefed me on the consulting engagement with a large bank, headquartered in UK. The next day, I set up calls with the engagement management based out of London and Chennai. It was particularly difficult to perform a current state analysis, when not co-located, as it requires a great deal of endorsement from various partners at every stage of progress. These stakeholders included C-Level and Business Unit heads along with Application teams influenced or even impacted by the onset of new Data Governance Program.

Although there was not much I could do from Bangalore, I was particularly enjoying my leisure and the time I got for myself. I'd tied the knot to my fiancee, three months prior to getting this program engagement. Early morning in a September, on landing in Heathrow, I took a cab to East Ham where I was to stay in a house share. I courteously declined to stay in a service apartment in London, as I wanted to experience the culture and make some friends outside of work. East Ham always had an old lingering essence of India, no different from a typical tropical city, with all its markets, temples, churches and electronics stores. There was no dearth of good Indian curry or breakfast including the high stores like

Saravana Bhavan, Taste of India while there were other restaurants from other Asian countries as well. Then, there was West Ham which boasted some old taverns for a quick drink along with a football club that helped me cool over game and few drinks. I still remember having to walk early on weekends, to McDonald to have my regular hash browns, pancakes and omelets while I could see the shopkeepers opening their shacks and placing the eye-catching electronics and toys in the displays while people stomped markets

I Ubered myself to offices in Canary Wharf and Angel for the initial weeks. I got the meetings and focus groups set up from various SMEs representing Business Units, Business Analysts and Architects from Data Management practice. The bank grew in-organically over the past few years through acquisitions and mergers which resulted in data redundancy from duplicated Business processes. Mortgage products for instance have distinct processes, leveraging same data, duplicated across the landscape while it had to be maintained separately. The costs were plummeting in maintaining redundant not-fit-for-purpose data and nourishing its quality. At the same time, this was getting the attention of regulators and the customers due to interrupted service.

The purpose of coming up with a current state analysis is to determine the necessary conditions to meet the business needs of the bank. It is required to ensure that the future state control environment is well defined, and is achievable with the resources available, and that key stakeholders have a shared consensus of vision on the outcome. Once the future state is defined, the existing capabilities like the processes and organization structure will undergo a gap analysis. This is to reveal the capability gaps that will provide the desired outcomes.

I was having a conversation with the program manager, and perhaps startled him by saying that "*At the outset, a change may be needed to many components in the organization including Business processes, Functions, Lines of business, Organization structures, Staff competencies, Knowledge and Skills training, Services, Governance Organization locations, Data and information, Application systems, Technology and Infrastructure.*"

## Business Objectives

I gathered the initial objectives from my interviews with the nominated stakeholders from the executive sponsor of data management and business unit stakeholders.

- Design and implement processes which manage and Govern data for the benefit of all stakeholders leveraging it.
- Benchmark maturity of the data management capabilities with external benchmarks and the needs of regulators
- Have clear, context based definitions of critical data while having that data be fit for purpose
- Establish ownership and Stewardship of data in Business, in the grassroots of the organization
- Establish effective collaboration between Business and Technology stakeholders.
- Ensure that benefits of data management are realized by the business units.

## Scope of Solution

The needs along with the new capabilities or capability improvements have been scoped from interviews, focus groups and questionnaires. These needs, capabilities were consolidated into a solution scope.

## Constraints

Constraints are a best way to showcase aspects of current state and planned future state that may not be changed by the solution. A few constraints may include time restrictions, technology and infrastructure policies, privacy and security considerations.

There can be protection mechanisms from other policies that trigger a check on overlap with new data policies being published. An example of one such constraint can be protection against services to be set-up for data delivery in profiling environment. The reason being, there are already existing data delivery services existing for a different purpose. While, data management office can be Business or IT owned, it is still bound to have the constraint. Some other constraints include

- limits on the number of resources available
- restrictions based on the skills availability.

A skills analysis will be conducted to understand the skills needed for implementation. If there are constraints on re-skilling or up-skilling stakeholders, it must be well documented and closed.

An additional constraint can be certain stakeholders not having to be impacted by the implementation of the solution.

There can be constraints on ownership of activities and metrics for data quality by business units. Such constraints should also be captured, as it helps in incremental and prioritized development of the solution capabilities, including the operating model.

## Organizational Structure and Culture

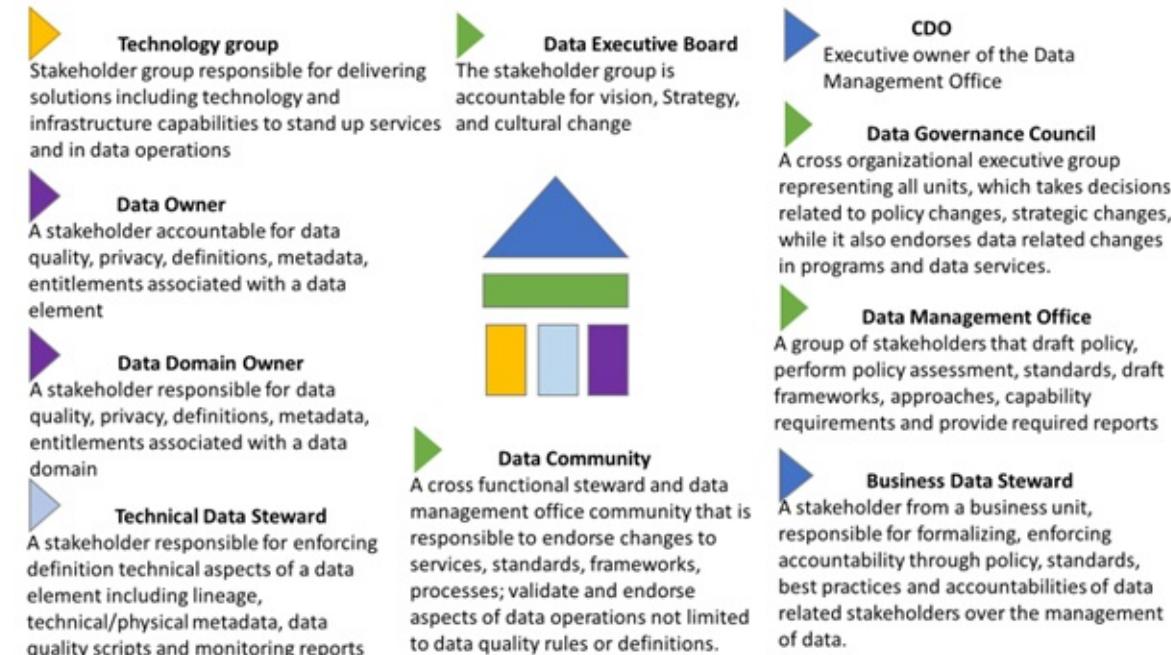
The formal and informal working relationships that exist within the enterprise can require a change, to facilitate the desired future state. A data steward can be reporting to the Data Management Office today, but to enable business ownership of data through stewardship, the connect should be strong with the business unit. This might require a change to the reporting structure of Data Stewards. Changes to reporting and organizational structure thus can encourage teams to work more closely together and facilitate alignment to goals and objectives.

Putting an organizational model required me to have extended interviews over iterations with Data Management stakeholders and business unit representatives. I had to get onto early morning flights from London to Edinburgh. I preferred travel from London City airport as it was just a 15 minute drive on early Monday mornings. Gogarburn was fun to travel to, with its vast plush landscapes in campus. This was the time when winter was setting in, with occasional snowing. I got a flu, got a jab as well but I reckon the chef from my hotel, offering me some herb soups that did Magic. I had my room, which boasts a balcony that opens to the river changed. Now it faces the art gallery.

Later, offered with a service apartment in the city center, I had much to explore with the queen's walk, the eateries around, Christmas markets, the portrait and landscape galleries. The sight of kids playing enjoying the first snows, in the

garden just opposite to where I was living, was sight to behold. Thursday's, I flew to London offices for the next engagements with the distributed team.

Note: Data stewardship roles should be standardized for alignment across business units or across geographies (in case of difference in operations across globe)



*Exhibit 1: A sample Organizational structure with role descriptions*

There are some essential roles and responsibilities quoted in the table-1, that clearly differentiate responsibilities from accountabilities. Most of the firms today, fail in Governance and Stewardship due to the lack of differentiation between responsibility and accountability. Some organizations document the functional hierarchies to showcase the accountability. For example, in a scenario like data domain management, the data owners can be responsible for managing definitions of the data-sets or concepts while the data domain owners are accountable for the same.

<b>Roles</b>	<b>Responsibility</b>
<i>Chief Data Officer</i>	<ul style="list-style-type: none"> <li>➢ Is the executive owner of the Data Management Office</li> <li>➢ Accountable for the policies, standards, frameworks drafted by DMO</li> <li>➢ Responsible for the development and sustainability of the funding model for Data Governance and management services</li> <li>➢ Accountable to the Project Management Office of Data Management Office</li> <li>➢ Accountable reporting to Data Governance council and Data Executive Board</li> </ul>
<i>Business Data Steward</i>	<ul style="list-style-type: none"> <li>➢ Responsible for enforcing accountabilities of data owners</li> <li>➢ Responsible to ensure policy, standards are adhered to within the division/country Demography</li> <li>➢ Accountable for managing Data Quality, Metadata, Entitlements, Architecture, Privacy for non-enterprise data within the business unit</li> <li>➢ Responsible for reporting on the data within the Business Unit (domains) to the data governance council</li> <li>➢ Responsible to enforce policy, standards in the division/country related to Enterprise data/Critical Data elements</li> </ul>
<i>Technical Data Steward</i>	<ul style="list-style-type: none"> <li>➢ Responsible for data lineage, physical or technical metadata capture across organization</li> <li>➢ Responsible to create data quality scripts from requirements/rules.</li> <li>➢ Responsible to execute data quality rules against systems of records/golden sources, systems of reference/golden copies</li> <li>➢ Certifies golden sources and golden copies</li> <li>➢ Accountable for the management of data analysts</li> <li>➢ Responsible to provide blueprints or architectural changes</li> </ul>
<i>Data Domain Owner</i>	<ul style="list-style-type: none"> <li>➢ Responsible for the definition of the domain and management of the definitions for data sets or concepts within the domain</li> <li>➢ Responsible for suggesting data owners for the data within the domain</li> <li>➢ Responsible for the overall reporting for non-enterprise data across all dimensions of data management</li> </ul>
<i>Data Owner</i>	<ul style="list-style-type: none"> <li>➢ Accountable to conform to policy associated with dimensions of data management, associated with the data elements that are owned</li> <li>➢ Responsible to partner with the business data steward to implement processes that will enable the active management of Metadata, data rules and entitlements.</li> <li>➢ Responsible to review privacy classifications applied to data and as changes are implemented</li> <li>➢ Responsible to escalate to data stewards and council in case breach of policy is noted.</li> <li>➢ Responsible to produce data rules not limited to policy enforcement rules, data quality rules, transformation rules, notification rules and thresholds rules</li> <li>➢ Escalates issues based on the notifications received for data quality</li> <li>➢ Accountable in drafting data agreements between data producers and data consumers</li> <li>➢ Accountable to maintain metadata, entitlements, classification, thresholds, data rules, lineage and taxonomies</li> </ul>

**Table 1: Roles and Responsibilities in a Data Management and Governance organization model**

To have an organizational structure that supports the future operational processes, there is a need to have certain activities acted upon by a collective group of stakeholders. This primarily can include various functional levels and have representation from diverse divisions as well. One such stakeholder group can be a Data stewarding council which can include data stewards from divisions, geographies along with the representation from the Data Office. These stakeholder groups are often referred to as named working groups like master

data management group. They are leveraged in setting up services like data quality or can be referred to as councils. There are some working groups mentioned in Table-2, that are required to build the initial set of services, and can be decommissioned based on the need.

<b>Working Group</b>	<b>Composition</b>	<b>Responsibility</b>
<i>Data Governance</i>	CDO Data Management Office Data Stewards	Create Vision, mission, plan  Standardize dimensions to be pursued  Draft Policy, review and get endorsement  Publish approach and mechanism to policy self-assessment  Establish Project Management Office  Implement organizational structure  Classify data logically into domains and concepts  Create an approach to classify data elements into CDE's based on data management characteristics along with risk and value characteristics  Create working groups for Metadata, Data Quality, Architecture, Stewardship, Privacy, maturity assessment, policy self assessment, Tooling
<i>Metadata Management</i>	CDO Data Management Office Data Stewards Architects Content Analysts	Define standards and best practices for Business and Physical Metadata management  Publish a Metamodel that includes attributes of Business and physical metadata  Initiate a vendor analysis for tooling  Define a SAAB Business Information Model (CDM).  Crowdsource concepts and baseline them  Standardize horizontal and vertical relationships across data modeling levels  Define standards for capturing lineage  Standardize templates and user guide with mandatory and optional fields as put in metamodel  Define standards for taxonomies and ontologies along with templates  Liaise with risk management to identify privacy classification allowable values.
<i>Data Quality</i>	CDO Data Management Office Data Stewards Data Quality Analysts	Define data quality strategy, target state objectives and approach  Standardize data quality dimensions  Identify accountable and responsible stakeholders

*Table 2: Working groups and Composition*

## Overcoming challenges in Data Ownership

Having the business stakeholders own their data is a crucial aspect of Governing data. Having to differentiate the responsibilities of the data steward from the accountability of data owners is a stepping stone in ownership and accountability.

Data Ownership as put forth by many leaders in the industry is all about having to identify, enable and empower the stakeholder preferably the Business stakeholders who own the operational processes. There are many challenges in having to imbibe ownership into an organization. Many regulators including PRA have surveyed the firms and published common challenges in data ownership. Data ownership is often not a full-time job for most data owners while it is for data stewards. I have given my advice below on gaining advantage in assignment of data ownership roles & responsibilities.

- A common challenge while identifying data owner is "Should this be a process owner or an application owner or a people (users') manager". Data owners are often Business analysts, Process owners, Application owners, Project managers, SMEs or knowledge workers. They can support processes, people and applications that leverage data in scope and have familiarity and knowledge of the data in their scope. In firms that are just starting up capabilities for data management, the ownership lies with IT while for mature organizations, it lies with the COO or CDO organization. The stakeholders who have substantial knowledge of the business processes along with the data that is created, used by the process will be apt for the role of a data owner.
- In order to complement the technical know how – the data flows between systems, processes and people; the role of the SME can be standardized.
- Data owner has the responsibility in ensuring controls related to data are managed, and he/she can be complimented by knowledge from an SME who has shared knowledge of that data.
- SME roles should be standardized and consulted by stewards and data owners whenever required. The responsibilities of governing data should still lie with the data owner.
- There are potential benefits to having an ownership model that defines contributing and viewing data owners for data elements along the data life cycle. A contributing data owner is the owner of the data domain like customer domain in whose purview the customer KYC data is created. There is Risk and Compliance domain, that has shared ownership of the data based on the context in which KYC data is being used in risk modeling. In this scenario, the Risk and Compliance function can be a viewer domain. It is beneficial to have

one contributing data owner and multiple data viewers, unless the same data element is being enriched for context across various processes.

- Knowledge of data management and governance processes, techniques, tools are required by data owners to orchestrate data governance activities. These are common challenges that need to be addressed by up-skilling the data owners which will enable them to take responsibility of managing the data.
- If data owners have business familiarity of the data but do not necessarily understand the flow of data along its value chain or life-cycle, it will be a challenge in leveraging metadata management and data quality services.
- A culture of business ownership along with every data governance service of data should be promoted that builds awareness in enterprise. Promoting the data quality, metadata, architecture, risk and privacy services should have significance on par with Service usage and Service improvement.
- Data owners should be aware of the benefits of leveraging data governance services. They can then assist other stakeholders in their sphere of influence to look for value beyond their regular line of sight.
- Cascade Goals of enterprise in association with Data Governance to the data owners so that the overall assessment plan takes their performance into account.

## Capabilities and Processes

### Operating Model

A well-defined operating model plays a vital role in achieving expected business benefits, to embark on data governance in an enterprise that spans divisions, geographies and diverse stakeholders. The essential aspects of a winning business operating model should encompass –

- Identification of responsible and accountable stakeholders;
- defining handshakes and hand-offs;
- Discovering and standardizing processes, procedures based on policies and guidelines;
- Motivation, goals and performance assessment plan to measure progress and report;
- Change control with well-planned stakeholder communication strategy;

- Implementation road map with a work breakdown structure;

A benefits based perspective is required, if an operating model needs to be centralized or a decentralized. If benefit arising from centralizing data management operations across the enterprise is more than 20% compared to the decentralized model, the focus should be taken to centralize.

Often with a centralized model, below are some benefits

- skills gap is bridged
- processes yield consistent outcomes irrespective of the business unit
- reporting is much easier
- decision making and executive buy in at C-level is synonymous continuous feedback elicitation
- improvements and change management can be planned for effectiveness.

At the same time, operation rigidity, reduced motivation in mid-level managers and bureaucracy often outweigh the benefits. That's where the socio-cultural aspects have a major say in formulating the operating model. Attaining a fine equilibrium between control and management of required capabilities is required. This allows stakeholders to embrace the data quality services with reskilling. This equilibrium also enables self-service.

Approach	Operating model	Emphasizes on
Top down	Centralized	Policy, Guidelines, costs
Bottom up	Federated or distributed	Self-service & management
Balanced or Hybrid approach	Centralized control and distributed management	Needs, impact & maturity (improvement)

*Exhibit 2: Emphasis on approach to operating model*

Each of the variants (eg: centralization or decentralization) of Data quality management operations has its merits and demerits as showcased in table 3.

Features	Centralized	Federated
<b>Benefits management</b>	Easy to manage & communicate to sponsors and data executive board	Can be managed well by business unit's data management committee if the unit is operating independently.
<b>Ownership/ Accountability</b>	Rules, profiles and results elicited and owned by a central cross-functional team. Only for "data at rest"	Rules, profiles, results elicited, owned by application owners & business units. For data "at rest" and "in motion"
<b>Change management</b>	Leadership Intensive program Top down approach	Data stewards/data owners and application owners are change agents at grassroots
<b>Turnaround time for orchestration</b>	High	Low
<b>Skills</b>	Highly skilled team required	Regular skills with re-skilling will suffice
<b>Personnel and process Efficiency</b>	Central team's efficiency will be high, Process efficiency is high	Teams' efficiency depends on localized skills
<b>Process maintenance</b>	Easy to maintain	Often ignored
<b>Stakeholder handshakes and handoffs</b>	High in number and documentation	Less in number and based on LoB standards
<b>Establishing cross dimension levers (metadata management)</b>	Nimble to address in framework	Not simple to address
<b>Continuous feedback elicitation</b>	Planned and time intensive to attain from knowledge workers	Easy to elicit in silos, but often reactive
<b>Rules elicitation &amp; management</b>	Dependent on SMEs, Highly Systematic	Localized knowledge exists, Less systematic
<b>Framework View and maintenance</b>	Centralized view and continuous improvement possible	Siloed view, improvements often neglected
<b>Service monitoring, reporting</b>	Easy at enterprise level	Cumbersome at enterprise level
<b>Prioritizing DQ issues</b>	Effort intensive	Less intensive
<b>Tool access &amp; management</b>	Lies with central team, Vendor relationship easy to maintain	Dependent on the governance and quality tool owner
<b>Data sourcing</b>	Easy if using a centralized datamart for profiling	Easy if using the application data store but not central mart
<b>Knowledge management</b>	Driven by global communication strategy	Driven locally and depends on effective local management
<b>Attitude towards data quality</b>	Positive as orchestrated by central team	Less positive across distributed teams, can face resistance in pockets
<b>Group policy implications</b>	Easy to analyze impact	Difficult to analyze impact
<b>Infrastructure</b>	Central control over infrastructure	Decentralized control if using distributed data systems
<b>Service improvements</b>	Easy to elicit, change and push for subscription	Less systematic and application agnostic

*Table 3: Centralized vs Federated Model features*

On completing the current state analysis of the firm and its culture, the outcomes from this section will be used to customize an operating model. It is always suggested to have an operating model that provides centralized control and distributed management approach as showcased in Exhibit 2. Centralized control stresses on having data management in grass roots. This can be achieved by balancing the responsibility and accountability across data ownership and stewardship in the data management office and operational level of divisions. Some high level processes along with activities are provided in the table 4.

High Level process	Process	Activities
Define, operationalize data governance	Data governance structure is created	Data Governance Office is created An executive Owner (CDO) is assigned the privileges The Data Governance plan is created
	Content Governance is defined	Project Management office is established The organization structure is implemented, roles assigned Authorized data domains are identified, defined and published Criteria to define critical data element is established Alignment with Metadata Management is established Data Classifications including Data management characteristics (aka Master, reference), Data security and privacy classification are established Definition and use of taxonomies is put forth
	Policy and standards are written and approved	Policy is drafted, documented, reviewed and published Policy and standards have been reviewed and approved by senior executive governing bodies, executive board and governing councils
	Program governance is established	Approach and mechanism to policy self-assessment internally is established Project Management Office commissions Program/Project Governance
	Program governance controls are in place	Project Governance structure established, aligned with Project Management Office of organization Funding Model is operational Escalation Procedures are developed and documented Project engagement, review and approval processes are established Policy and standards are enforceable and auditable Metrics are in place to track program adherence, progress and outcomes Formal training programs have been designed and implemented
	Tool Stack governance is aligned	Platform governance is established Data distribution governance is established Data storage governance is established BI, ETL and data tool governance is established
Cross-organizational enterprise data governance is aligned		Data governance is aligned with information security policy Data governance is aligned with privacy and cross-border policy

Table 4: High level Data Governance Process

## Data Management process

Some high level processes along with activities and associated capabilities are provided in the table 5

Data Management and Governance Process		Privacy Management activities	Capabilities
Define, communicate and endorse Data Management Strategy	Data Management Strategy is defined, communicated and endorsed	<ul style="list-style-type: none"> <li>Data Management Strategy is developed, updated to include the Data Privacy management aspects</li> <li>Data Management Strategy and Data Privacy Management are aligned with Organizational objectives</li> <li>Include a mechanism for approval</li> <li>Evaluate strategy : as being enforceable by Audit</li> </ul>	<ul style="list-style-type: none"> <li>High level business requirements are elicited, prioritized, communicated and packaged</li> <li>Requirements are signed off by the relevant stakeholders using the mechanism in the strategy</li> </ul>
	Define the importance of identifying, prioritizing and assuring the appropriate use of authorized data domains	<ul style="list-style-type: none"> <li>Set the need to identify logical classification of Data Domains</li> <li>Articulate the importance of establishing policy to enforce appropriate use of authorized data domains</li> </ul>	
	Define the importance of establishing risk management, integrating it with Data Management	<ul style="list-style-type: none"> <li>Define the need behind having alignment with Risk Management</li> <li>Align Data Risk Management objectives with Data Management Strategy</li> <li>Describe Target structure and organizational structure of Data Risk Management</li> <li>Define Roles and responsibilities</li> </ul>	
	Aligns with architectural, IT and operational capabilities	<ul style="list-style-type: none"> <li>Data architecture concepts have been incorporated</li> <li>Technology concepts have been incorporated</li> <li>Operational concepts have been incorporated</li> </ul>	
	Create formally established governance program	<ul style="list-style-type: none"> <li>Define the purpose and objectives for establishing data governance</li> <li>Describe the data governance target state organizational structure</li> <li>Describe the governance roles and responsibilities</li> </ul>	
	Define how the data management program will be measured and evaluated	<ul style="list-style-type: none"> <li>Define the importance of developing outcome metrics to determine the effectiveness of the data management program</li> <li>Define the importance of developing tracking and adherence metrics to determine how the data management program will be measured</li> </ul>	
	The data management strategy calls for the creation of a communication and training program	<ul style="list-style-type: none"> <li>Describe the importance of establishing a communication strategy</li> <li>Describe the need for an education and training program to ensure stakeholder understanding, buy-in and compliance to the data management program</li> </ul>	
	The data management business case is aligned to strategic value and risk drivers and tangible business outcomes	<ul style="list-style-type: none"> <li>The Data Management Business Case is mapped to and aligned with drivers, requirements and strategy</li> <li>High level business outcomes are defined and sequenced</li> <li>Business Case is socialized and validated by program stakeholders</li> </ul>	
	The data management funding model has been established, approved and adopted by the organization	<ul style="list-style-type: none"> <li>The funding model is mapped to prioritized business requirements, implementation timelines and operational capabilities</li> <li>Total expense for the Data Management Program is captured, maintained and analyzed</li> <li>A standard methodology for calculating ROI is established</li> <li>ROI is measured, monitored and used for making Data Management Program decisions</li> </ul>	
	Define and operationalize Data Management program	<ul style="list-style-type: none"> <li>The data management program is defined and documented</li> <li>The data management program is communicated and provided rights of enforcement for compliance</li> </ul>	
Define and operationalize Data Management program	The roadmap for the data management program are developed, socialized and approved	<ul style="list-style-type: none"> <li>Program roadmaps are defined, developed and aligned with the Data Management Strategy</li> <li>Program roadmaps are established and endorsed by Program Stakeholders</li> <li>Project plans are developed detailing deliverables, timelines and milestones</li> </ul>	
	Stakeholder engagement established and confirmed	<ul style="list-style-type: none"> <li>Identified stakeholders including sponsors, data analysts, architects and others commit and are held accountable to the Data Management Program deliverables</li> <li>Funds are allocated and aligned to program roadmaps and workstreams</li> </ul>	
	Communication program is designed and operational	<ul style="list-style-type: none"> <li>Internal communication plan have been created, channels established, plan published and approved</li> <li>Communication plan with external regulators bodies are created and approved</li> <li>Active engagement with external industry and standards bodies are in place</li> </ul>	
	Data Management Routines are Established, Operational and Measured	<ul style="list-style-type: none"> <li>Routines for support of the data management program have been established</li> <li>Issue identification, prioritization, escalation and conflict resolution are defined and operational</li> <li>Metrics (i.e.: KPIs, KRIs) are defined and used to track Program progress</li> </ul>	
	Metadata Management	<ul style="list-style-type: none"> <li>Logical domains of data have been identified, documented, endorsed and communicated</li> <li>Underlying physical repositories of data have been identified, documented and inventoried</li> <li>Physical domains of data into Master, Reference and Transaction are classified</li> </ul>	
	Define the data (semantically) and structurally	<ul style="list-style-type: none"> <li>Data definitions and associated Business, Operational and Technical metadata is documented</li> <li>Vertical lineage between various levels of data - Semantic/Conceptual/Logical/Physical is established</li> <li>Relationships across horizontal levels including synonyms, alternate names, specific, generalizes are established</li> <li>Lineage is documented, verified and attributed to the data</li> <li>Taxonomies and ontologies are created, documented, maintained and governed</li> </ul>	
	Govern the data (establish sustainable data architecture governance)	<ul style="list-style-type: none"> <li>Data architecture governance procedures are established to ensure authorized as well as controlled use of data</li> <li>Data architecture governance procedures are in place and aligned with Business and technology</li> </ul>	
	Define Technology Architecture strategy and service	<ul style="list-style-type: none"> <li>Technology architecture strategy is defined, communicated and endorsed by relevant stakeholders</li> <li>An actionable roadmap is developed and implemented</li> <li>Platform governance structure and processes are defined and operational</li> </ul>	
Define and operationalize Data Quality Service	Technology tool stack is identified and governed	<ul style="list-style-type: none"> <li>Technology tool selection strategy is defined and endorsed by relevant stakeholders</li> <li>Technology tool roadmap is developed and implemented</li> <li>Tool selection governance structure and process are operational</li> </ul>	
	Data storage lifecycle management strategy defined and governed	<ul style="list-style-type: none"> <li>Data storage management strategy is defined and endorsed by relevant stakeholders</li> <li>Data storage management roadmap is developed and implemented</li> <li>Storage governance structure and processes are operational</li> </ul>	
	Operational risk planning is in place	<ul style="list-style-type: none"> <li>Data infrastructure continuous planning is defined and operational</li> <li>Operational Risk Governance Structure and processes are in place and operational</li> </ul>	
	Data quality program is established	<ul style="list-style-type: none"> <li>The data quality strategy, target state objectives and approach is defined, communicated and endorsed</li> <li>Accountable parties have been identified and roles and responsibilities have been assigned</li> <li>Data Quality operating model and processes are defined, and operational</li> </ul>	
	Quality of existing stores of data are identified and assessed	<ul style="list-style-type: none"> <li>Data is profiled, analyzed and described for Data Quality against the dimensions, in enterprise repositories and golden sources</li> <li>Data Quality rules are extended, enriched and published</li> <li>Data Quality monitoring is operational</li> <li>Data quality remediation has been planned, prioritized and actioned</li> </ul>	
	Quality of new data is monitored, analyzed and reported	<ul style="list-style-type: none"> <li>Data Quality controls are in place across the entire lifecycle of the data</li> <li>New data (internal and external) is profiled and profiles stored in a central repository</li> <li>Remediation is invoked for data in error</li> <li>Data Quality processes are audited by risk function</li> </ul>	

*Table 5: Data Management Process*

The future state analysis thus provides the needs and capabilities required in the target state. In further chapters, a step by step approach to set up Business Data Glossary and Data Quality services will be presented.

## Creating a Metadata Service

Again, I was working for a global bank, for its chief data office, based out of Bangalore. The bank headquartered in London, has operations in more than 60 countries. At the same time, it has operations servicing retail, commercial, private banking customers. The breath of the operations heaves in confusion in having to understand the data and its related context in which it is produced or applied. The bank has commissioned my services to get the data definitions captured for the critical customer and compliance data. But, the services that existed could not help much in enabling the divisions in capturing the metadata.

As I started on this engagement, I got a deep dive into the existing Operating model for Metadata. I see several challenges as below

1. Non-Integration between data quality, Metadata and Architecture dimensions
2. A federated operating model with stakeholders having difficulty accepting their responsibility to capture definitions and associated Metadata.
3. A meta-model that was not customized to the divisions based on its needs

Metadata in actual terms is defined as “*a set of data that describes and gives information about other data.*” Niso defines it as “*structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource*”. Is that how the Business stakeholders understand it?

**Metadata management** is a series of activities ensuring that metadata is properly acquired when data is planned for, obtained, stored, shared, maintained, applied and deleted. On collection and management of metadata, the challenges in understanding the meaning and the context of data are reduced considerably. If there is a Business Analyst or a reporting analyst, he/she need not stumble to understand what a customer name as a data element really means in the

organization. The organisation requires technology capabilities like a one stop shop repository that captures Business, Technical Metadata in defined templates or organized meta-models. Also, required are process and people capabilities including workflows and authorization roles. All these capabilities will be used to define services which will operationalize management of metadata.

Standardizing Metadata as a service will assist the organization in pushing or pulling the service capabilities based on their needs. The current state requires emphasis to be on promoting existence of Metadata capabilities. This is a stepping stone when the cultural change of “Managing Data as an Asset” is being trickled into grass roots of the firm. The awareness handshakes should be customized based on the stakeholder groups. One wouldn’t want to pitch the word Metadata to business users, operational and front end staff. As this often gives a technical fervor. I prefer to use “Business Terms or associated characteristics” to refer to data elements and related Metadata.

Data has a life cycle, POSMAD, (Plan, Obtain, Store/Share, Maintain, Apply, Decay) which helps us understand the lineage –

- Where it comes from and originates – sources
- Which processes it is applied in
- Who uses the business term
- Which system uses the data element for logic

Aspect of Metadata management	Commentary
<b>Business Metadata</b>	relates to active management of data definitions, relationships like synonyms, examples, data domain, Business rules and data quality rules. The ontologies and taxonomies assist in developing relationships across data assets in the organisation, vertically in a hierarchical fashion.
<b>Operational Metadata</b>	relates to active management of risk classification, controls around data lifecycle, entitlements, Governance and Data ownership.
<b>Technical Metadata</b>	relates to active management of information about physical characteristics of data stores, transformations, derivations, timeliness and lineage.

*Table 6: Types of Metadata*

Metadata does not generate information on its own and requires every responsible stakeholder, system and process to consistently produce Metadata. This can include definitions, owners, length, lineage, security classification, product association and other characteristics that relate to the context of the data. This further enables holistic Governance of other data dimensions like quality. A service comprising of repeatable metadata activities helps in better managing and governing the Metadata while ensures sustainability in the organization. The service also would ensure that metadata management activities are formalized and operationalized as defined by the policy.

## Metadata Service Catalog

A service catalog that provides the list of the activities that are encapsulated into specific service operations. These boundaries must be used to align and arrange activities into discrete (non-overlapping) functional partitions with clear interfaces that are well suited for service enablement. The model defines the role of the Service Domain in two facades – a type of business function performed ("functional pattern") and a type of object that is acted on. A best practice is to fulfill the asset life cycle that starts with "Metadata capture" and ends with Metadata management as shown in Exhibit 3 below.

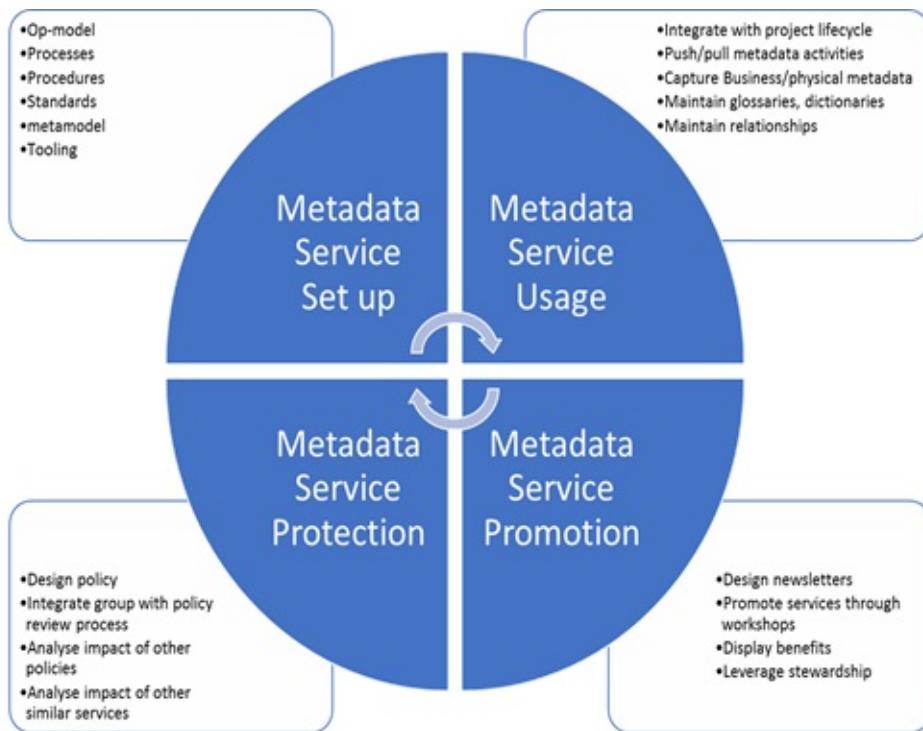


Exhibit 3: Metadata Service Catalog

# Metadata Management Integrated with high level Software Development Life cycle steps

Metadata activities can be mapped to waterfall, Agile and Business transformation life-cycle stages for better capture, integration, maintenance and governance. A sample mapping of Metadata stages to Software Development Lifecycle stages is shown in Table 7.

Service Phase	SDLC Phase	Metadata Phase
Service Promotion	Project Planning/Business Analysis planning	Metadata scoping
Service Usage	Requirements analysis	Business Metadata Capture
Service Usage	Design, Architecture and Coding	Technical Metadata Capture Lineage capture
Service Usage	Testing	Validate Metadata
Service Usage	Implementation	Publish/Integrate Metadata
Service Usage	Post Implementation	Manage Metadata

*Table 7: Metadata life cycle integrated with SDLC stages*

At the outset, Metadata is enabled by namespaces that are logical placeholders that hold required information in defined templates.

- **Metadata repositories** — Used to store, publish and share the metadata along with the relationships. People can leverage metadata stored in repositories during the requirement, data analysis, design and other phases of a change life cycle.
- **Data dictionary** — Holds business terms along with their associated definitions. Data Dictionaries will as well contain business, operational and technical metadata.
- **Data Glossary** — Holds terms of reference used in definitions along with acronyms and alternate names
- **Data lineage** — Attributes that have details on the data producer, systems to which data is distributed, systems in which data is stored and decayed. It also showcases the processes as data traverses across systems.
- **Rules management** — Holds data rules including data quality rules, policy enforcement rules, transformation rules, derivation rules, notification rules, threshold rules pertaining to a data element.
- **Taxonomies** — Support entity relationship (ER) models, ontology and modeling and relationships.

## Meta-model & Structure

It is quite a necessity to understand the relations between the data assets to better manage them. The aspects detailed below in Table 8, are adopted from Collibra and can be customized to the needs operating model.

Community	A community is a grouping of users. It often corresponds to functional divisions in a company and should be aligned with the company's governance organization. A community can contain various domains.
Domain	A domain is a logical grouping of assets and belongs to exactly one community. It has a domain type that specifies which asset types can be created in the domain.
Asset	An asset is the fundamental building block or resource for which you want to capture information. An asset belongs to exactly one domain. An asset has a unique name within its domain.
Asset Type	An asset is the instance of exactly one asset type. Asset types allow you to formally specify what type an asset is, as a kind of template. Examples of asset types are Business Term, Code Value, Database, Policy, etc.  An asset type is assigned to one or more domain types.
Domain Type	A domain is the instance of exactly one domain type. A domain type allows you to formally specify what type a domain is, as a kind of template. Examples of domain types are Glossary, Code List, Data Asset Domain, etc.  By assigning asset types to a domain type, you can specify which types of assets can be created within a domain of that domain type. For example, in a Glossary domain type, you can create assets of types Business Term, KPI and Acronym.
Attribute	An attribute is a specific piece of information that you capture for an asset. An asset can contain many attributes. Attributes contain the actual information you want to capture about an asset such as the text of its definition.
Attribute Type	An attribute is an instance of exactly one attribute type. Attribute types allow you to specify what type of information is contained in an attribute. Examples of attribute types are Definition, Description, Example.
Relation	A relation is a link between exactly two assets. It captures how two assets are related to each other. One asset can have many relations.
Relation Type	A relation type specifies the type of a relation between two asset types. A relation is always the instance of exactly one relation type.  A relation type is bidirectional; going from the source to the target of the relation and the other way around. The source and target of a relation are both asset types.  An example of a relation type is <i>Data Asset is parent of/is child of Data Asset</i> . The source and the target of this relation type are both asset type Data Asset. One direction, from source to target, reads <i>Asset Type is parent of Asset Type</i> . The other direction, from target to source, reads <i>Asset Type is child of Asset Type</i> . This relation type thus allows Asset Types to have children of the same type.
Complex Relation	A complex relation is a link between two or more assets. A complex relation may also have attributes. It is an extended, more powerful variant of a regular relation.

Table 8: A description of assets & types in Metadata structure

Classifying data into semantic, logical and physical models helps in deriving ownership at a high level as shown in Table-6.

**Semantics** refer to the adoption of precise, shared and consistent Business meaning of data across enterprise. This is what Governing data as a meaning emphasizes. The challenge for organizations has been the inability to harmonize disparate data across an enterprise or a business function. This is because, data is referred to with different meaning in a same division or firm. Does a customer mean "The person who is opting for financial services from the bank" or does it also include the brokers, Beneficiaries of an account as well. Does it really depend on the context in which a beneficiary is considered a customer such an account closure.

The inconsistency in understanding the data has been primarily due to the in-organic growth of data. This has been stemming from the non-existence of planning for data management during mergers, Acquisitions and organizational growth. Semantics is the discipline of assigning unambiguous meaning to data across its lifecycle. Semantics are usually associated with Business Architecture where a *Business object* is used to represent significant informational and conceptual elements in business context. The other way of referring to a *Business Object* is that it is a passive entity that represents business, products or services in its business context.

Techniques like Semantic models, Business Metadata management, Logical classification, as shown in Table 9, and Data Entity views can be leveraged to capture precise, consistent definitions and also aid simplification of landscape. Further simplification can be achieved by labeling logical classifications like Customer or Finance domain else privacy classifications like Restricted. Also, physical classifications like master, reference or transaction can be used based on the need.

A **Conceptual level of data** describes data at its highest form, identifying critical data objects required to satisfy business objectives while also defining their relationships. *Business Object* is a passive entity that represents business, products or services in its business context. It is modeled in Business Architecture. At an application architecture level, The passive counterpart of *Business Object* is the *Data Object*.

Techniques like Logical models, Business & Operational Metadata are used to capture precise, consistent definitions of data objects.

**Logical level of data** is a fully attributed conceptual model that has been abstracted from physical implementation or instantiation of data. The logical model represents business requirements in terms of what is needed to satisfy the objectives of the business function. Logical data model can also be associated with the *Data objects* in the application Architecture. Techniques like Logical models, Business & Operational Metadata can be used to capture precise, consistent definitions of data objects and it's attributes.

**Physical level** of Data is the instantiation of meaning, relationships and attributes of data at a physical level as data exists in storage devices. Physical models are usually attributed to the Technology Architecture. A physical piece of data is called a *Data Artifact* as is produced, consumed or stored by the software application process. It models typical messages, extracts, scripts, database tables and documents. Example – A Lead campaign extract.

Levels of data classification	Primary Drivers	Standard definitions	Lineage(L), relation(R)	Usage	Techniques	Examples
Semantic	Business ownership and accountability of data Managing Data as a meaning	Business term	Business term to business name (L)  Business term to business term (R)	Common business language used in enterprise.  Usually this is the Business vocabulary, Regulatory or legal vocabulary used by the Organization or the Industry.	Semantic model, ontology model, taxonomy	Product Name – Legal  Product Name – Contractual  Product Name – Sales  Product Name – Regulatory filings
Logical	Business ownership and accountability of data Best practices & Guidelines to rationalize definitions and usage of data elements across lifecycle.	Standard Business name	Business name to system name (L)  Business name to business name(R)	Data in the BI, Business metadata environment, Data Mapping	Conceptual model, logical model, synonym  Data Dictionary, Metadata repository, Data Directory, Data Book	Savings Account name – Short  Checking Account name – Long  Savings Account – Temenos  Savings Account – T24
Physical	Best practices & Guidelines to rationalize definitions and usage of data elements across lifecycle.	System name	System name to system name(R)	Names used in the Physical data structures	Physical model Database structures	svngs-Nme, Acc Type-Name  Should be rationalizing the system names to the standard Business names.

*Table 9: Semantic, Logical and Physical Models*

A meta-model defines standardized attributes and characteristics that need to be captured to manage and govern data. The Meta-model is customized and aligned to Metadata operating model, based on the needs and objectives of data management. The tool capabilities will also undergo a change to accommodate the meta-model in it's operations.

Quoting an example, at a conceptual level, a data element is known as a business term. The business metadata associated with a business term includes data definition, synonym, alternate name, data domain, data concept, sample values to name a few attributes.

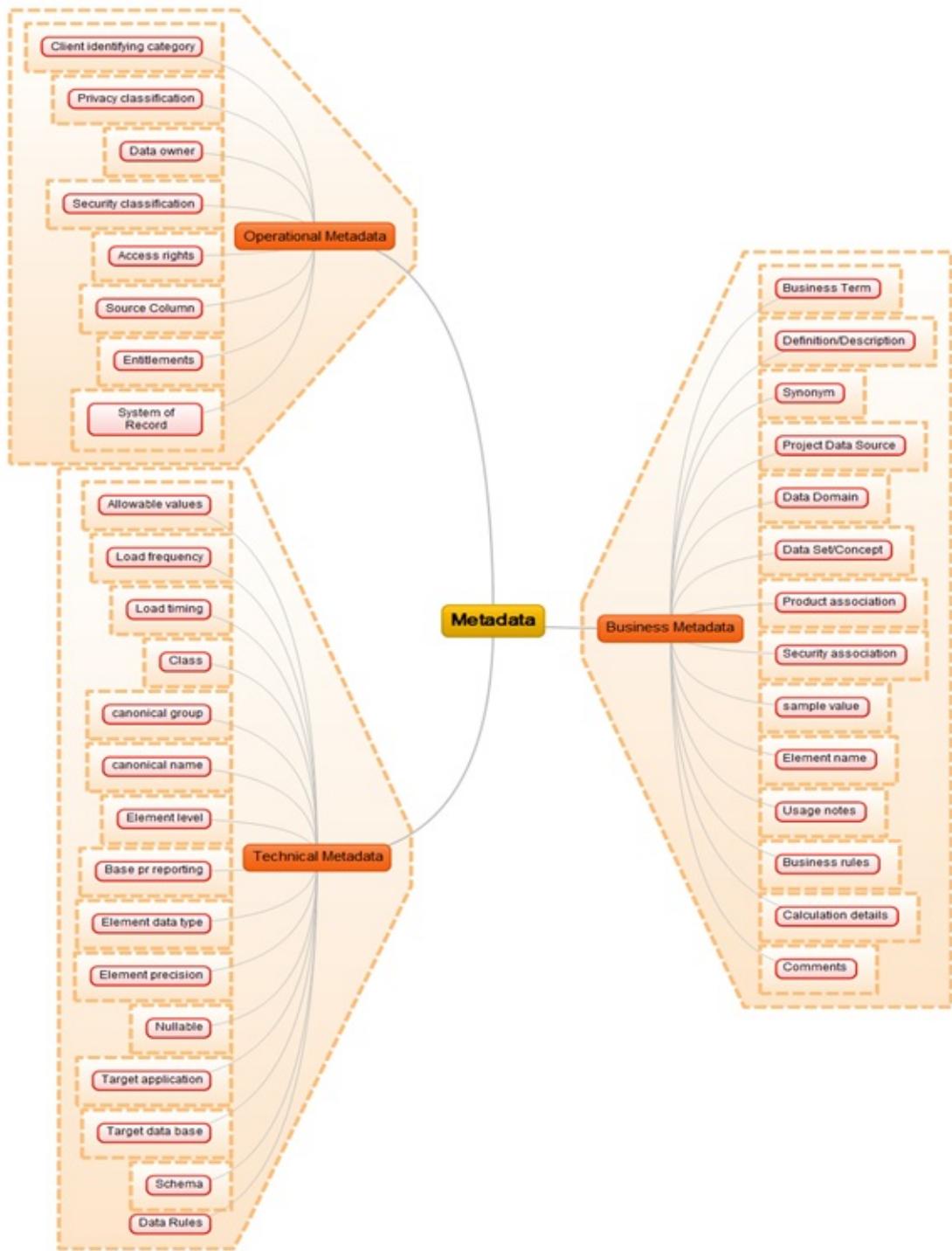
For a logical data element, informational attributes include description of the data element, synonym, conceptual representation of logical name, sample values to name a few.

Based, on the asset type say a Business term at conceptual level or Data Element at logical level, the attributes can be mandatory or non-mandatory. The meta-model further defines the structural requirements, including type of data characteristics required, length, relationship to other asset types.

## Meta-model

### Conceptual Data Models

A conceptual model must be hosted delineated from the logical model. FIBO standards are usually embraced for conceptual modeling by Banks or Financial institutions. **Conceptual level** as described earlier depicts data at its highest form, identifying the critical Business objects needed to satisfy a business purpose. A sample template or Meta-model that captures the essential information about data is shown in Exhibit 4 below.



*Exhibit 4: Attributes of Meta-model*

Generally, the first activity is to identify logical and physical data domains and data sets. For example, the logical classification for any industry would be Customer and Finance domains that hold data related to customer and finance operations. BIAN landscape also provides a simplified service based views that can be used to classify data landscape into domains, based on it's operations. Physical domains are defined based on physical characteristics including master, reference and transaction data. This in-fact satisfies the need to manage data

based on the physical characteristics and pace with which data changes. This will create clear boundaries based on understanding of Business ownership. Thus, the high level classification of business areas into data domains is the first step of simplification.

There are two approaches to simplify a data domain further into logical entities called concepts or data sets. This can be done based on

- logically breaking down the data domain further into data-sets
- Grouping data into an associated data-set or concept
- Generalization and specialization of an existing concept

**Approach 1:** For each concept, identify a suitable label, find a suitable

- definition in a dictionary, and use this definition to represent the concept

*Note:* Add properties and restrictions as required

**Approach 2:** To support industry-specific concepts for banking (BIAN standards, FIBO), arrive at a high level abstract concept of under which other concepts are specializations.

1. Note: Think of a suitable label which could reasonably label that abstraction.

Techniques:

1. Class modeling
2. Business Landscape modeling
3. Service modeling
4. Ontologies and taxonomies modeling

Standards:

1. FIBO
2. BIAN

Metadata Naming Standards

The best practices and guidelines to standardizing data element or business terms can be customized from the Industry standards including ISO 11179 standards.

## Data Element

A data element naming can be of the following formats for brevity and consistency

- **Object [Qualifier] Property Representation Term**
- **[Concept] <Qualifier> Class**

Further, considerations while naming a data element or a business term are below

- o "<>" for Qualifier are optional
- o The Concept can be captured along with the data domain in data dictionary as well
- o For example: **[Import Loan Facility] <Last Paid> Installment Amount**

A sample set of standards to define a data element should

1. be stated in singular
2. State what the data element is and should not specify what it is not
3. be stated as a descriptive phrase or sentence(s)
4. contain only commonly understood abbreviations
5. be expressed without embedding definitions of other data elements or concepts

## Technology and Infrastructure

Based on the requirements set forth for technology and Infrastructure, a vendor analysis can be initiated to cater to the basic and the additional capabilities required to collect and actively manage Metadata. If current technology and infrastructure are insufficient to meet the business needs, requirements will identify the changes necessary for the desired future state. The existing technology may impose technical constraints on the design of the solution. These may include development languages, hardware and software platforms, and application software, data stores that must be used.

As Governance needs usually cater to the people or process capabilities, the same are represented in the operating model and associated processes. The same need to be documented and agreed upon. Then they will be taken to closure

through solution definition and procurement or build. For example, the meta-models, processes, environments, organization structure and workflow in a metadata management solution must be customized based on the defined operating model.

## Running a Data Quality Service

When an organization widely explores the benefits of standardizing Data Quality management, they look to find efficiency and scalability in their data quality operations. At the same time, the industry standards including DAMA, EDM and Cobit provide best practices and guidelines to start. Getting a framework that has the dimensions of quality including Completeness, Accuracy and others is not much of a challenge but maturing across Data Quality operations, while making it sustainable is a challenge today.

Identifying the critical and key data elements is at the forefront of the data quality assessment and monitoring activities. Table 10, shows a sample framework using the criteria of risk, value, privacy and data management characteristics to force score the data elements to arrive at an objective score. The combination of the score and thresholds will be used to arrive at the key or critical data element classifications. Based on the needs the Level-1 key data elements (scored above high threshold) will be managed for quality by a central data quality team. While the Level-2 can be managed by the business unit's data quality management team, while level-3 can be prioritized and managed by Business Unit.

Weightage	10	7	10	7	7	10	10	7
Criteria	<i>Risk</i>		<i>Value</i>				<i>Privacy</i>	<i>Characteristics</i>
Classification	Regulatory Risk (Low = 4 Medium = 7 high = 10)	Operational Risk (Low = 4 Medium = 7 high = 10)	Critical in > 1 Business Unit ( 1 BU = 4 2 BU = 7 > 2 = 10)	Used in Financial calculations (1 calc = 4 10 calc = 7 >10 calc =10)	Used in Reporting (1 report = 4 10 report = 7 >10 report =10)	Number of Incidents (1 = 4 5 = 7 >10 = 10)	Privacy classification (Internal = 4 Restricted = 7 Highly Restricted = 10)	Transaction = 4 Master = 7 Reference = 10
Data Element - 1								
Data Element - n								
Score								

*Table 10: Risk & Value Realization framework*

**Business rules**, an integral component of consistency Quality checks are often ignored in operational Data Quality processes, implemented by Business data stewards, data owners and Data Quality analysts. A business rule can be a very apt consistency rule while data is distributed between end points. It is not only the business rules that have specific use in Data Quality but other classification of data rules as well. Data is the life plasma of the business functions; business rules define how the organizational policies and decisions are enforced by leveraging data through specific rules. Data Rules rules thus channelise the active functioning of an organization. The various data rules that should be captured with clear delineation to manage data quality better are shown in Table 11.

## Business Rules Classification

Decision rules by which the business derives the conclusions from conditions.

- e.g. For a value, inference and conclusion from business

Data Rules that define the precise characteristics that data needs to adhere to

- e.g. valid values or ranges for particular fields, relationships between fields or records, etc.

Target Rules that define the thresholds for Data Quality Indicators

- e.g. red-amber-green status (for health check based on thresholds, will be used for representation in scorecards, dashboards)

Notification Rules that define alerts that should be triggered under particular circumstances

- e.g. notifying a data steward if a record fails a validation check, alerting a data owner if data quality falls below a defined threshold, etc

Transformation/Derivation Rules that define operations that should be applied to data.

- e.g. logic establishing the conditions under which one statement can be derived or validly deduced from one or more other statements

Policy enforcement rules

- Defines constraints on process/functions using data that is based on policy

---

*Table 11: Data Rules Classification*

The policy enforcement and the business rules will be realized as data consistency and validity rules. Quoting an example of country code which can be a two letter ISO code rather than a three letter code as enforced by policy or reference data guideline. It is required to check where there is wanted redundancy of data say a country code that is three letter but means the same as represented by two letter ISO code across the landscape. The next step is to validate the scope of the remediation required.

In Table 12 is sample illustration for broker segment on grouping business rules (decision rules) from multiple logic statements into a two dimensional tabular view. It is much easier to manage and scale rules for Data quality as they are managed as metadata. I was crowd sourcing and eliciting ideas on better management of rules in the IIBA group that aids better management of Data.

Element Name	Conditions					Conclusions			
Broker Segment	Sales		Frequency of consistent sales in years	1st sale date		Derivation/ Transformation	Business Rule (Conclusion on product sales)		Data Rules
Segment-A	Is Greater than	\$10 M	10	Is greater than	11 Years	Platinum	May sell/Will be	All products at \$10M level	Valid Values for Broker segment are Segment-A, Segment-B, Platinum, Gold, Silver, New Broker
	Is Greater than	\$10 M	5	Is greater than	6 Years	Gold	May sell/ Will be	5 products at \$10M level	
	Is Greater than	\$10 M	1	Is greater than	0.6 Years	Silver	May sell/ will be	only one product at the \$10M LEVEL	
Segment - B		\$0M		Is Less than	0.6 Years	New Broker	May sell/ will be	In the same segment for 6 months. promoted based on sales, in 7th month	

Table 12: Decision rules illustration

Every change in the state of an entity like a "Lead" to "Customer" is associated with business rules like "*A lead on completing a product purchase transaction is a customer*". If there is a business rule that has multiple business terms to be a part of the rule; the business terms appearing in the rule can be related with a relationship attribute like "*Relates to*" after importing them to Glossary.

Most of the organizations are fairly maturing their data management and Governance practices to meet their goals. For example, data quality service would have been started and grown in-organically based on the "*then needs*" like reducing data related risks and meeting regulatory needs. But, the recent organizational drivers both internal and external requiring "*managing data as a meaning*", "*simplifying data landscape*" and "*Data Risk management*" is pushing the need for alignment of Data Quality management, Metadata management, Data architecture and other dimensions to the risk management and corporate Governance structure. A simple target state environment highlighting "*Data Quality*" is showcased in the Exhibit 5 below.

## Data Governance

- Group policy is operational and supported by standards, guidelines and enforcement of accountability
- Formalization of responsibilities of the data owners, stewards and the reporting structure to CDO
- Project Governance integrated Data Quality in regular changes through toll gates
- Toll Gates to provide formal check points during SDLC, Agile and Transformation phases
- Project Governance representative to provide consultation where required
- Formalization, Oversight through assessment, policy assessment and reporting integrated to the CDO structure

## Data Quality

- Centralized Data Profiling services to provide consistent and efficient characteristics of data
- Data Quality rules are stored at enterprise level and for projects/adhoc use are enhanced to context
- Promoted re-use and applicability of common rules as group rules
- Consolidation of issues to be remediated across divisions
- Detective and Predictive control environment to monitor data quality on criticality of data elements
- Consulting services to assist the divisions to assess and monitor data quality
- Integration with other data management and governance dimensions for better management

## Metadata Management

- Data rules stored in the data dictionaries at a group level and at a division level for localized needs
- Data Definitions and associated lineage are documented for data elements in the repository

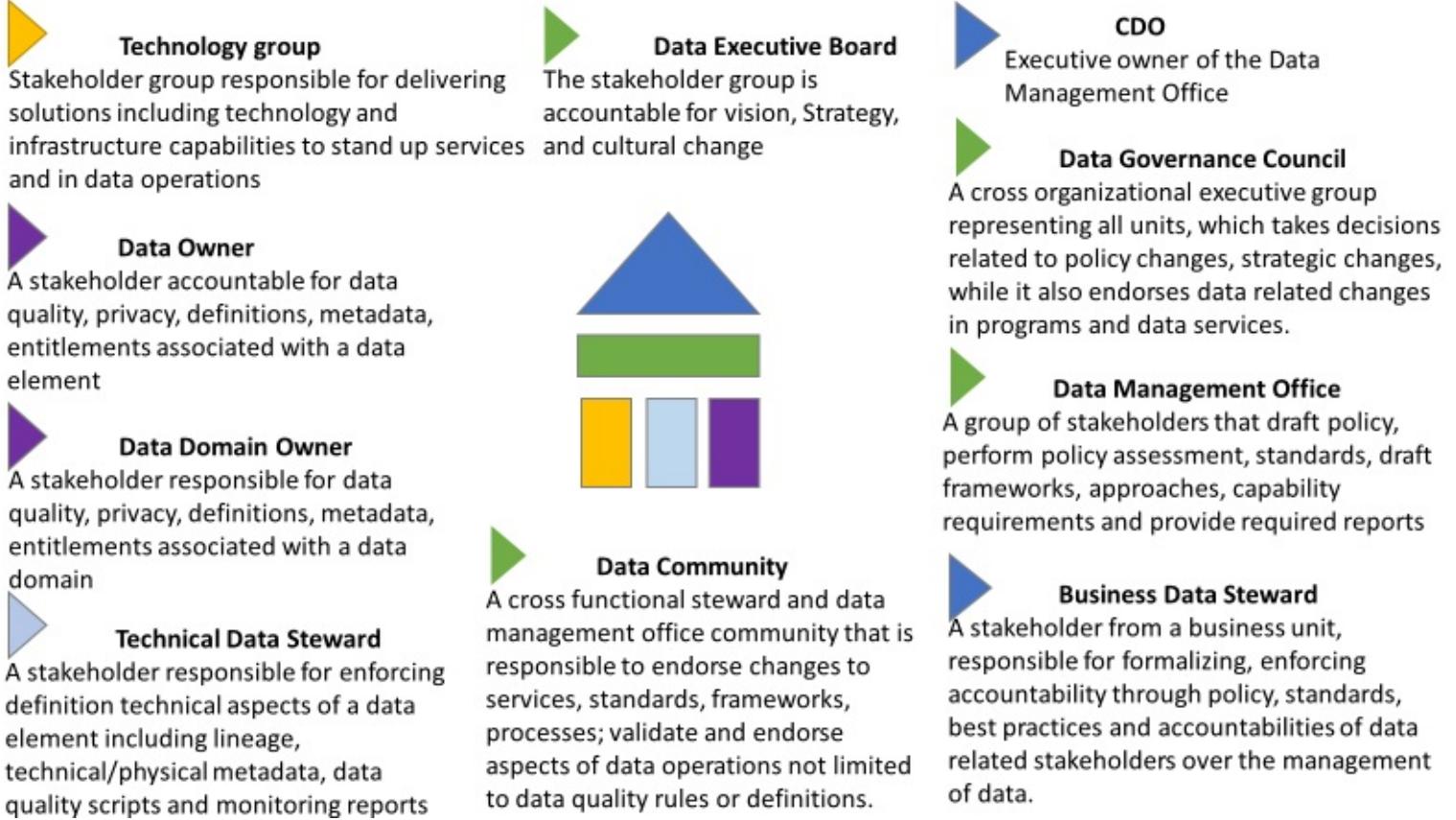
## Data Management

- Data delivery services to provide data acquisition, sourcing capabilities including cross border
- Environment abstraction based on special regulatory or legal needs
- Reference data management provides consistent data quality across group and divisional needs

### *Exhibit 5: Target State environment for Data Management & Governance*

## Organizational Structure and Culture

The formal and informal working relationships that exist within the enterprise may need to change to facilitate desired future state. A sample stakeholder and stakeholder groups in future state is described in Exhibit 6. Changes to reporting lines of Business Data stewards can encourage Business teams to work more closely while facilitating alignment to goals. Elements of the organizational structure and culture may need to change to support the future state.



*Exhibit 6: Organizational structure for Data Quality Services*

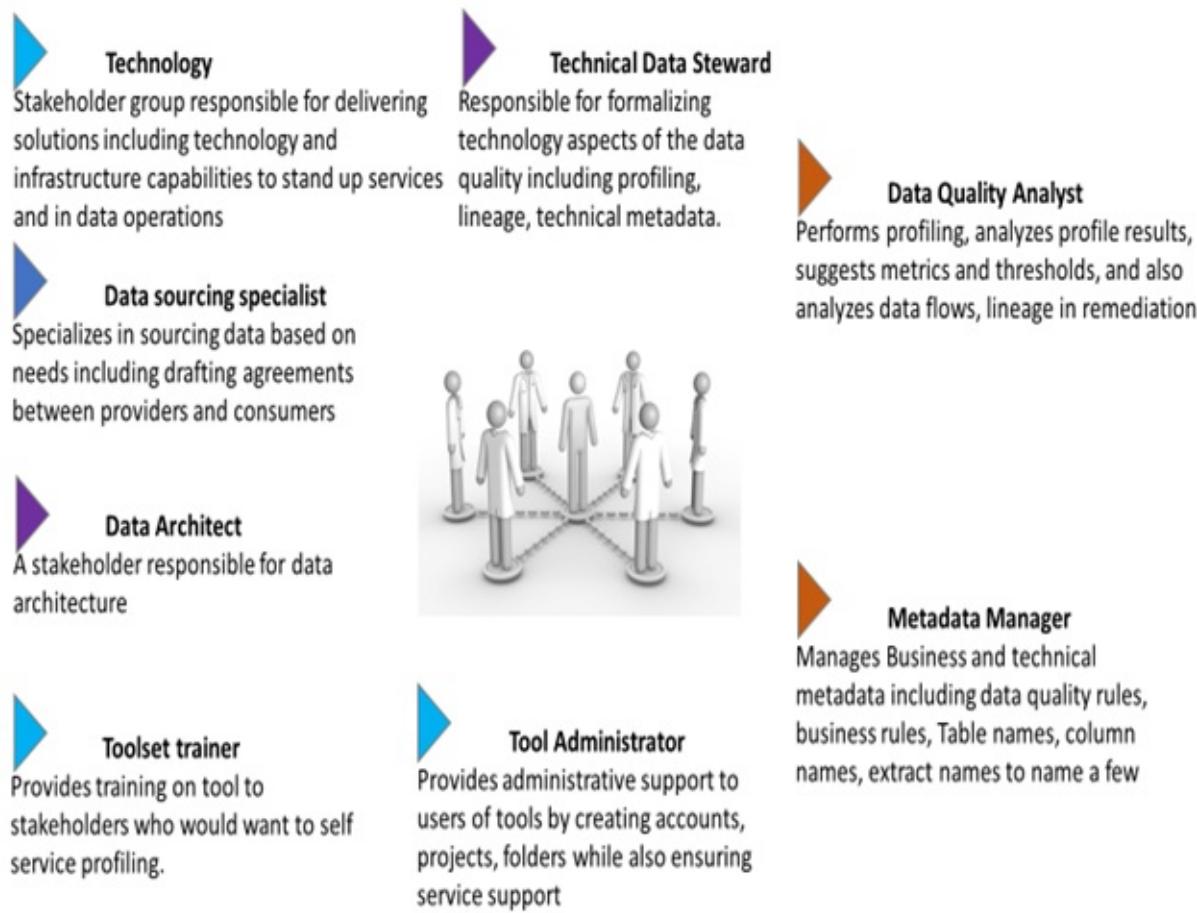
*Note: Data stewardship roles should be standardized for aligned either across business units or across geographies (in case of difference in operations across globe)*

Some roles and responsibilities are presented in the table 13 below –

Roles	Responsibility
Chief Data Officer	<ul style="list-style-type: none"> <li>Is the executive owner of the Data Management Office</li> <li>Accountable for the policies, standards, frameworks drafted by DMO</li> <li>Responsible for the development and sustainability of the funding model for Data Governance and management services</li> <li>Accountable to the Project Management Office of Data Management Office</li> <li>Accountable reporting to Data Governance council and Data Executive Board</li> </ul>
Business Data Steward	<ul style="list-style-type: none"> <li>Responsible for enforcing accountabilities of data owners</li> <li>Responsible to ensure policy, standards are adhered to within the division/country/Demography</li> <li>Accountable for managing Data Quality, Metadata, Entitlements, Architecture, Privacy for non-enterprise data within the business unit</li> <li>Responsible for reporting on the data within the Business Unit (domains) to the data governance council</li> <li>Responsible to enforce policy, standards in the division/country related to Enterprise data/Critical Data elements</li> </ul>
Technical Data Steward	<ul style="list-style-type: none"> <li>Responsible for data lineage, physical or technical metadata capture across organization</li> <li>Responsible to create data quality scripts from requirements/rules.</li> <li>Responsible to execute data quality rules against systems of records/golden sources, systems of reference/golden copies</li> <li>Certifies golden sources and golden copies</li> <li>Accountable for the management of data analysts</li> <li>Responsible to provide blueprints or architectural changes</li> </ul>
Data Domain Owner	<ul style="list-style-type: none"> <li>Responsible for the definition of the domain and management of the definitions for data sets or concepts within the domain</li> <li>Responsible for suggesting data owners for the data within the domain</li> <li>Responsible for the overall reporting for non-enterprise data across all dimensions of data management</li> </ul>
Data Owner	<ul style="list-style-type: none"> <li>Accountable to conform to policy associated with dimensions of data management, associated with the data elements that are owned</li> <li>Responsible to partners with the business data steward to implement processes that will enable the active management of Metadata, data rules and entitlements.</li> <li>Responsible to review privacy classifications applied to data and as changes are implemented</li> <li>Responsible to escalate to data stewards and council in case breach of policy is noted.</li> <li>Responsible to produce data rules not limited to policy enforcement rules, data quality rules, transformation rules, notification rules and thresholds rules</li> <li>Escalates issues based on the notifications received for data quality</li> <li>Accountable in drafting data agreements between data producers and data consumers</li> <li>Accountable to maintain metadata, entitlements, classification, thresholds, data rules, lineage and taxonomies for the data elements.</li> <li>responsible to publish agreements from the data providers regarding data quality, consolidation, integration and rationalization in advance of distribution of data</li> <li>Responsible to review client identifying categories and privacy classifications applied to data on a regular basis</li> </ul>
Content analyst	<ul style="list-style-type: none"> <li>Responsible to create data definitions and associated business metadata</li> <li>Responsible to maintain data definitions and associated business metadata</li> <li>Responsible to maintain relationships horizontal and vertical across data landscape.</li> <li>Responsible to maintain taxonomies and ontologies</li> <li>Responsible to maintain criteria to define KDEs/CDEs</li> </ul>
Data Architect	<ul style="list-style-type: none"> <li>Responsible to create blueprints for architectural changes</li> <li>Accountable to certify the golden copies/golden sources</li> <li>Responsible to define Technology architecture strategy</li> <li>Responsible to define Platform governance structure and processes</li> <li>Responsible to plan for Data infrastructure contingency</li> <li>Responsible to implement Data storage management roadmap and standards</li> <li>Responsible to operationalize tool selection governance structure and process</li> </ul>
Data Management Office	<ul style="list-style-type: none"> <li>Responsible to draft policy and perform policy self-assessments</li> <li>Responsible to draft standards, frameworks and present them to the council for endorsements</li> <li>Accountable for the review, quality and applicability of standards of the Metadata in the enterprise data dictionaries and glossaries</li> <li>Responsible to ensure that SAAB CDM related data dictionaries and glossaries for critical data elements or enterprise data have golden copies, golden sources designated, published for metadata, data rules and privacy entitlements and classifications.</li> <li>Publish profiles and reports on data quality monitoring for critical data elements or enterprise data</li> <li>Provides auditable deliverables, assessment plans, audit schedules and audits</li> <li>Responsible to report on the policy self-assessments, escalations, quality reporting and other reporting requirements to the Governance council and executive board</li> </ul>

*Table 13: Roles & Responsibilities*

To have a data quality service working, specific roles as stated in the exhibit 7 below are required.



*Exhibit 7: Data Quality Roles & Responsibilities*

It is required to sustain Data Governance and integrate it into daily data operations in such a way that these services are no longer perceived as an overhead. This is one of the major challenges.

Further challenges most organizations are facing today are detailed.

1. *Challenges in cultural adoption* of risk management, in a distributed way in the organization, in the perspective of data management and Governance. Most of the organizations take a value driven approach to data governance and management but fail to understand that a risk based approach is also required, in interest of Enterprise goals. Currently, risks related to “*data being considered as enterprise asset*” are not identified actively in the Business Units.
2. *Awareness of data management & governance services* is a major challenge in

Enterprises. With awareness comes adoption and enablement of taking Governance & Risk management forward in an enterprise.

3. *Project Governance* structure along with well-defined deliverable at every stage, in a project life cycle, are already defined. This has *limited adoption* and needs to be enforced through Data Governance, risk management and Project-Management-Office standardization.
4. *Data Privacy & Security* needs to be dealt differently than Information security & privacy.
5. There is a need for *continuous enhancement* of data quality services, operating model, and thresholds to complement the current needs of stakeholders, organizations, challenges.

Most of the challenges can be overcome by re-discovering and standardizing the current data management and governance services or processes in operations. There is a need for a target operating model that consists of discrete functional modules that collaborate through service calls.

Each Governance function can be considered as a *Governance area* as shown in exhibit 8, to orchestrate coherent activities within a *service domain* called *service operation*. Service operations describe a high level dependency in a Service Domain or between two service domains.

1. ***Data Governance Area*** – is the highest-level classification of the Governance domain. An area groups a set of business, process and technology capabilities to achieve an end goal for data management.
2. ***Service Domain*** – is the finest level of partitioning, each defining unique, discrete business, and process, technology capabilities. The Service Domains are the ‘elemental building blocks’ of a Data Governance service landscape.
3. ***Service Operation*** – is at a high level what business, process or technology functionality it should contain and what functionality it needs to access through delegated service operation calls to other service Areas or domains. Usually, this would be through a combination of functional pattern and Asset in focus.

A data quality service can be well defined with a set of service domains including Service setup, service promotion, service usage, service protection, service monitoring and improvement



*Exhibit 8: Data Quality Service catalog*

Quoting an example from Data Quality service, the service domain "*Data Quality service set up*" in Table 14, details the Service operations including the functional pattern and Asset. Quoting an example from Data Quality service, the service domain "*Data Quality service set up*" in the below model details the Service operations including functional pattern and Asset.

Service Phase	Service ID	Functional Pattern	Asset	Description
Data Quality Service set up	SS1.1	Plan	Strategy	DQ strategy and approach has been designed and developed
Data Quality Service set up	SS1.2	Communicate	Strategy	Communicate strategy to relevant stakeholders in councils, Committees, LOB leadership, CDO & Board.
Data Quality Service set up	SS1.3	Plan	operating model	Communicate operating model to relevant stakeholders in councils, LOB leadership, CDO & Board.
Data Quality Service set up	SS1.4	Communicate	operating model	Operating model has been communicated to relevant stakeholders
Data Quality Service set up	SS1.5	Administer	Feedback solicitation & incorporation for strategy	Feedback from stakeholders has been incorporated into the final version of the DQ strategy
Data Quality Service set up	SS1.6	Endorse	Strategy and operating model	Stakeholders and Senior Management endorse and support the DQ program and strategy.
Data Quality Service set up	SS1.7	Align	Stakeholders with Org structure	Accountable parties have been aligned to the organizational structure
Data Quality Service set up	SS1.8	Align	Assessment plan with performance plans	Individuals are held accountable for the performance of their data quality function via annual reviews and compensation considerations
Data Quality Service set up	SS1.9	Allocate	Roles & Responsibilities	Accountable parties have been identified. Data quality responsibilities have been assigned
Data Quality Service set up	SS1.10	Align	Roles & responsibilities with Governance structure	Accountable parties have been aligned to the organizational data management governance structure
Data Quality Service set up	SS1.11	Administer	Support solicitation on roles	Stakeholders and Senior Management endorse and support the defined roles and responsibilities
Data Quality Service set up	SS1.12	Design	Validation Routines	Understand the needs for validation routines based on landscape, the industry standards as well. Baseline the validation routines.
Data Quality Service set up	SS1.13	Deploy	Operating Model	Orchestrate the Data Quality operating model through the defined plan or roadmap.

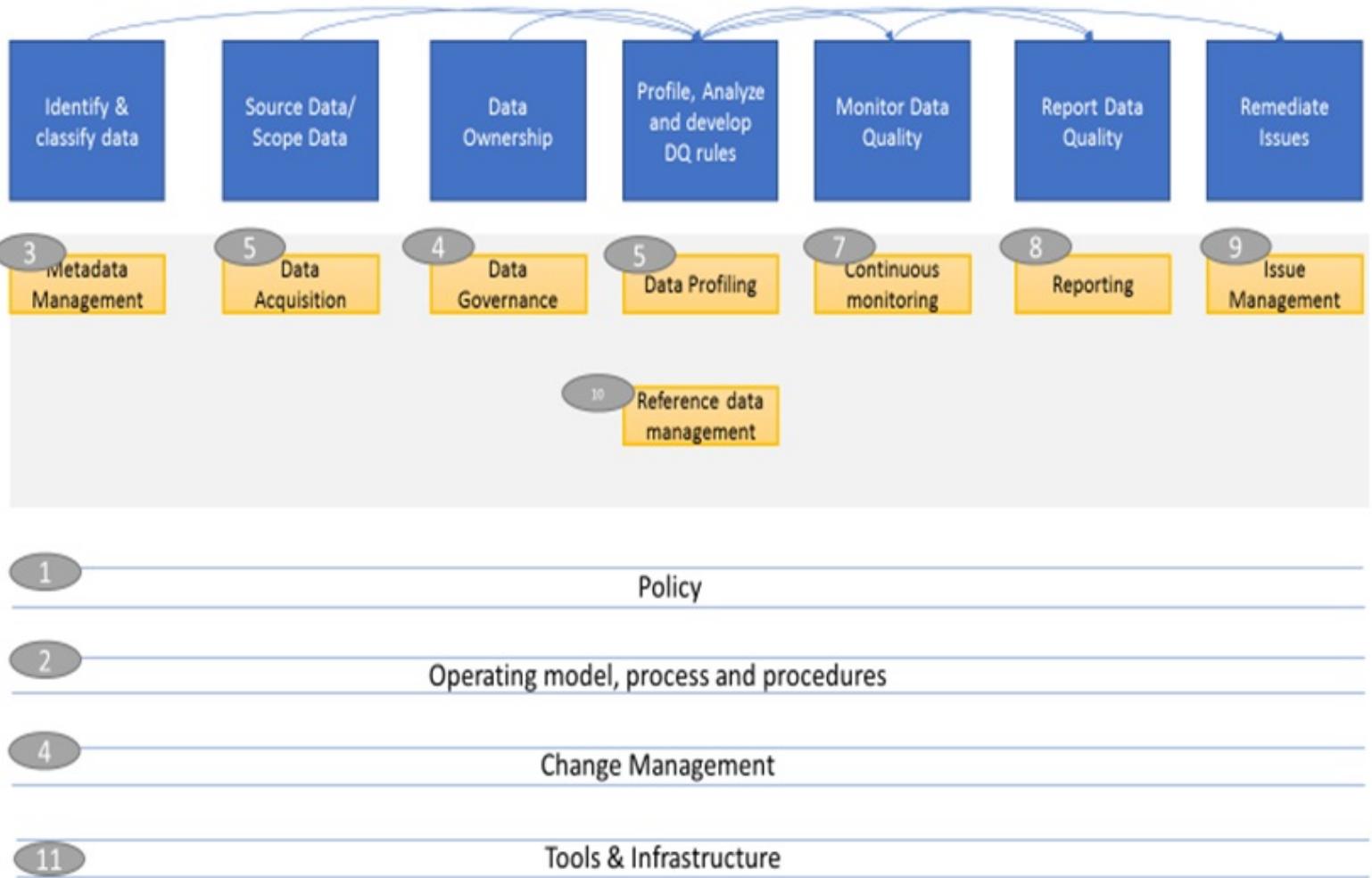
*Table 14: Data Quality Service set up service operations*

COBIT provides an industry accepted framework, once implemented, the executives can ensure that data governance is aligned effectively with business enterprise goals and better directs the use of data for business advantage. COBIT provides best practices, controls activities and tools for assessment, monitoring and governing data management (IT) activities. Cobit processes can be mapped to service domains to clearly differentiate management from Governance as shown in Table 15.

Service domain	COBIT Domain / Processes	Sample COBIT Practice Name
Data Quality Service Set up	Evaluate, Direct and Monitor (P) Align Plan and Organize (P) Build, Acquire and Implement (P)	Evaluate risk management. Understand enterprise direction.
Data Quality Service Promotion	Build, Acquire and Implement (P)	Nurture and facilitate a knowledge-sharing culture.
Data Quality Service Usage	Align Plan and Organize (P) Deliver Service and Support (S) Monitor Evaluate and Assess (S) Evaluate Direct and Monitor (S)	Prioritise resource allocation.
Data Quality Service protection	Monitor Evaluate and Assess (P) Evaluate Direct Monitor (S)	Optimize response to external requirements.
Data Quality Service Monitoring & improvement	Evaluate Direct and Monitor (P) Monitor Evaluate Assess (S)	Direct the governance system.

*Table 15: Cobit Practices mapped to Data Quality service catalog*

The existing data quality capabilities will be standardized into discrete functional modules. The focus on the current capabilities in an organization can be in the order stated in the Exhibit 9 below.



*Exhibit 9: High Level Capabilities grouped for implementation in order*

## Processes (LO)

The high level processes to set up and operationalize data quality are quoted in the Table 16 below. Efficient processes that are lean and provide consistent outcomes are required for data quality management.

Data quality program is established	The data quality strategy, target state objectives and approach is defined, communicated and endorsed
	Accountable parties have been identified and roles and responsibilities have been assigned.
	Data Quality operating model and processes are defined, and operational
Quality of existing stores of data are identified and assessed.	Data is profiled, analyzed and described for Data Quality against the dimensions, in enterprise repositories and golden sources
	Data Quality rules are extended, enriched and published
	Data Quality monitoring is operational
	Data quality remediation has been planned, prioritized and actioned.
Quality of new data is monitored, analyzed and reported	Data Quality controls are in place across the entire lifecycle of data
	New data (internal and external) is profiled and profiles stored in a central repository
	Remediation is invoked for data in error
	Data Quality processes are audited by risk function

*Table 16: Data Quality processes along with activity descriptions*

New or improved processes can be required to implemented by leveraging existing or new target state organizational structure. The questions, Who, Where, How, What will be answered in the level-1 and Level-2 process as in the Table 17 below.

Process (L1)	Description
Initial Engagement	Determines the scope, needs and skills available in the Business Unit
Define Data Requirements	Provides data elements to be sourced for profiling, along with the format for presenting the data in. For example, tables, or flat files with a row count, agreed delimiters and column headings.
Metadata Input	The Business definition, table, column details, extract details and other technical metadata required is provided
Access to Data	Understand the constraints and availability of data for profiling.
Source Data	If access to data is restricted, the data is sourced with agreements from providers
User Access	If the Business Analyst, Data Analyst or the profiler has access to profiling tool or not is assessed. The required access is provided. This is a one time activity
Profile Data	Data in the staging environment is profiled to bring the characterises of data in a report
Identify Potential Issues	The characteristics of data are analysed to identify issues.
Rule Development	The data rules including consistency, Accuracy, Integrity, Completeness and Timeliness are developed.
Execute Rules	Run data quality rules against the data
Review Results	Review results from running the rules against the entire set of data in scope
Publish Rules	The rules are published as baselined
Manage Rules	Data Rules are managed as Metadata
Schedule Rule Runs	Schedule rules to be runs on data refresh or definitive frequency
Notify Issues	Provide timely scorecards to Data Steward, Owners and other stakeholders
Remediate Issues	Prioritize and remediate issues based on criticality and impact

*Table 17: Level-1 Data Quality processes and descriptions*

The Business Unit, Project or Data management office has a need to understand the characteristics of data. The technical data stewards can be engaged by Business Data Stewards to determine the type of engagement. If there is a need for assistance in sourcing data, use of a data quality solution to profile data, developing basic and advanced business rules; the skills are assessed and the services and stakeholders are engaged as per the process. The basic profile of the data is created and shared with the requestor.

## Data Quality Exploration and Analysis

A firm needs to understand where the quality of its data is lacking. It is also important to understand where the data is being applied, as it helps the firm

understand and estimate the impact of poor data quality. It is important to know where the data is created or produced to have it corrected.

A cross integration between Metadata and Data Quality, brings out information like Lineage, System of Origin, System of Record etc. This information can be leveraged to Explore further into the data value chain, and to provide an analysis. But to start leveraging this information from data dictionaries, it should be captured initially and Chapter 4 will provide details of a metamodel that is self sufficient to aid in Data Quality Exploration and Analysis.

There are multiple data quality dimensions that aid the data quality analysts in identifying the type of validation to be performed. Below are some validation routines. As showcased in the table-N, A dimension like *Completeness* ensures "*Availability of required data in a column, field, object or records*". There can be further sub dimensions like "*Count of Records*", "*Frequency Distribution*", "*Fill Rate*" that can be associated as sub dimensions of Completeness.

"*Count Of Records*" ensures that if data is being distributed from a lead generation system to a customer database on conversion to a customer, there is no loss of data. This is ensured by having to perform a check sum or count of records between transmitting and receiving systems.

"*Fill Rate*" ensures that if a KYC due or enhanced diligence is being performed on customer information, the fill rate of the customer name, address, beneficiary name should be 100% complete.

"*Frequency Distribution*" is more of a technique rather than a validation routine, which can be used to look at a distribution of values across a time period. If there is particular period, where the data value is not present, the impact can be analyzed. Or, if a table is analyzed for frequency distribution, the number of blank or null values in a column or set of columns in a table can be presented.

Further, we will explore on the right validation routine, which needs to be applied in a particular phase of the POSMAD.

Dimensions	Description	Sub Dimension
Accuracy	Refers to the degree that the value represents a real life entities they model	
Completeness	Availability of required data attributes and records	Count of Records Frequency distribution Fill rate
Consistency	Ensuring values in one dataset are consistent with another	Record consistency Structural consistency Semantic Consistency Pattern Analysis
Currency	Degree to which information is current to be applied	
Precision	Level of details on the characteristics of data elements	
Reasonableness	Level of consistency in operational context	
Referential Integrity	Condition that exists when all intended references from data in one column of a table to data in another column of same or different table	
Timeliness	Expectation of availability of the data by specified interval	Manual Float System Float
Uniqueness	No entity exists more than once	Duplicacy Redundancy
Validity	Ensures data values conform to numerous characteristics like range of values	Range of Values Decimals Content check

*Table 18: Data Quality Dimensions*

It is required to clearly identify the type of the data quality checks that need to be performed in a particular scenario and life cycle stage of Data. It makes sense to perform a consistency check if data is being distributed but not at rest. The validation routines are shown in Table 18 above.

Data Quality assessment scope can include identifying any issue that prevents the data user from being properly applied or impacts a business process. The profiling results from the data quality assessment or continuous rule runs provide a platform to explore and analyze the data quality using the dimensions stated above. The issues related to the Accuracy, Completeness, Consistency, currency, Precision, Reasonableness, Referential Integrity, Timeliness, uniqueness and validity will be validated, prioritized and taken forward through the remediation life cycle with the ownership from data owner and guidance of Data Steward.

Issues related to Integration between systems will be fixed by ensuring consistency in the data values across systems through Business rules, transformation, derivation and consistency rules.

## **Data Quality Remediation**

A Data Quality Remediation solution can be termed as a service domain. The capabilities should range from having to identify and document data quality issues to start with. Next, comes the process of assigning them to the data owner or SMEs, which require workflow capabilities. Further, the capabilities of having to collaborate, add analysis, perform Root Cause Analysis (RCA), document Business impact as well as reporting, closure should be available. There are many tools in the market that have been capabilities and can be customized based on the operating model and organization structure. Throwing in some scorecard and dashboard capabilities will definitely help.

Where there is an implementation of Master and Reference data management solutions, the third party data and internal data will be leveraged to automatically correct data to it's accuracy and identify the Golden Record.

It is necessary to put in policies and procedures for manual data cleansing and remediation, procedures and guidelines along with integration with other data management services like Metadata as well as Governance will be established.

### **Analysis of Reference Data quality:**

Reference data values are a slow changing dimension and do not require the same kind of data quality monitoring once they are standardized. The Metadata Model will be leveraged to capture the current and future Reference data standards. This would include the data element name including the capture of allowable values, standard name along with other Metadata. The existing solution for reference data can be leveraged or a schema can be built in the data management owned database to store reference data and a front end that provides self service capabilities to show the reference data standards and values. The data ownership and Stewardship processes will cater to the governance of reference data while specific reference data management process will cater to the management of the values and labels.

Reference data will be classified into

- Global reference data standard across Group
- Local Reference Data Standard across a Business Unit

Promotion from local to global reference data standards will follow the proposal, vetting and endorsement process. This will be a Governance process and can be initiated by the Data Owner or Data Steward or a cross functional stewardship group. With a goal of rationalization of Reference data across the organization, a standard should be enforced. Based on the criticality of the data element, they can be prioritized for rationalization.

## Differentiating Data Quality assessment from Monitoring

Today, the Chief Data Offices have put in Data Quality processes that have the minimum necessary activities like profiling based on need, remediation based on impact. But, the data quality process is much more larger than the activities just quoted above. The processes of having to assess data quality comes as a first step. To have to assess data quality, either the data should have been in error or should have been identified as a critical. It is required to understand the characteristics of the data like type, length, requirements for nullability, min and max values, that can come out of profiling of data. On pulling characteristics of data like an age, where the range can be between 10 to 120, the policy can be checked to see what is the age of having to invest into a certain product with the bank. If the same data element age needs to be used for a context of a specific product, the data owner and steward can come up with a specialized or contextualized data quality requirement that states age should be between 30 to 60 for a retirement finance product. This is the outcome of having to come up with Data Quality assessment which ends by having data rules documented. Further, the rules can be promoted to the group, in case they are common across. While, any localized rules can be associated with the domain alone.

Now, starts the need for data quality monitoring, where the data tends to change frequently. If this changing data is prone to errors, the same can be monitored continuously through rules run. However, all the data rules need not be run continuously, and it is only required to run only the data quality rules like completeness or integrity which pull out the data quality issues as they appear. One rule of thumb, is to have the data sampled and profiled based on the

frequency of refresh, to sense the occurrence of data quality issues associated with specific validations.

## Decide to drive by Policy or Guidelines

If current policies are not sufficient to meet the business needs, there is a necessity to identify the changes necessary for the desired future state. The policy can be enforced through the business rules in processes or workflow. While guidelines can guide the behaviour of the personnel as they are embedded in procedures. Organizations create data policy to ensure enforcement of compliance with mandatory regulatory, internal compliance, best practices, legal and ethical requirements along with need for managing risk. These requirements are embedded into the policy and privacy statements to provide guidance to personnel on their accountabilities and responsibilities. This assists the personnel in carrying out any activity that include data related operations. This helps manage the risks in data processes thus aligning with risk appetite and tolerance levels.

A sample Data Sourcing policy has been presented below which clearly articulates the summary of the policy, what is an outcome of this policy, the responsibilities and accountabilities of various roles in sourcing data.

### **Policy Summary:**

- Data must be made available based on the enterprise needs to processes, people and applications
  - There must be only one primary source of each data element
  - Data must be sourced in agreement with the data transfer protocols, country specific policies.
  - Data that is derived or transformed must be documented from the sources, both internal and external

### **Guidelines**

- There must be an authoritative source System of record or golden source identified for all Critical data
  - A data element should have a golden copy or a system of reference, which is a read only source

- Data should be sourced only from a certified system of record or a system of reference
  - The data consumer should get into an agreement with the data provider mentioning any transfer agreements as per data movement policy; quality requirements as per data quality policy.

### **System of Record or Golden Source**

- Is the data storage system that is considered an authoritative source of a data element
  - It can be external to the enterprise
  - Other copies should be considered secondary but must replicate the data within SLA as per agreement, to ensure timeliness and currency.
  - Data that is enhanced or transformed outside the golden source will be the responsibility of the stakeholders or application, performing the transformation.

### **Calculated data**

- Golden copies or systems of reference should not be the System of record for any data created by complex calculation
  - A simple calculation can include addition, subtraction, multiplication, division of 2 or 3 variables.
  - complex calculations include more than three variables. ex: calculation of Risk exposure for Basel-II

### **Data Management Office**

- Accountable to ensure that Metadata repository is populated with systems of record or golden source as well as golden copy or system of reference
- Responsible to ensure the Key Performance Indicators for the workflow to certify the golden sources and golden copies
- Responsible to standardize framework, workflow to manage the attributes related to golden sources and golden copies
- Accountable to publish reports to the council and board on the data availability in golden sources and copies.

### **Data Owner**

- Accountable to accept proposals, certify a data store as golden source or golden copy for critical data elements
- Responsible to provide data dictionaries with the information of golden

sources or copies

- Responsible to determine a change to the designation of golden source or golden copy by initiating the workflow
- Responsible to trigger the generation of the profile for a data element from a data store

### **Data Steward**

- Responsible to determine a change to the designation of golden source or golden copy by initiating the workflow
- Responsible to propose data store as golden source or golden copy for critical data elements
- Responsible to ensure that data owners are certifying data stores
- Accountable to liaise with business data stewards from other units to understand the wanted and the unwanted redundancy in the data, that is shared with the data owner
- Accountable for continuous elicitation of feedback on golden source, copy standardization process
- Accountable to ensure the certification is complete on time within SLA.

### **Project, program group**

- Responsible to notify the governance representative if the solution capabilities include acquiring, creating, storing, sharing, integration and distributing data from data source not certified as golden sources or copies.
- For data that does not have a designated golden source or copy, responsible to coordinate with the data owner to identify the proposed golden source or copy.

Thus a policy or guideline should clearly state as mandated –

- Purpose of the policy
- Outcome
- Who will be in scope of the policy
- Role of a stakeholder
- Responsibility
- Accountability
- Processes or links to procedures that provide guidance to personnel
- SLA and escalation mechanism

- Balance of responsibility and Accountability
- Requirement of creation of cross functional forums for decisioning
- Related policies or procedures
- Audit frequency

# Assessing Value from Data Management & Governance

Every data governance service and activity will be considered as governance enabler. An enabler, in simple sense, is a new or an improved capability made available by data management and governance division. These enablers can be further classified into Business, Process and Technology enablers.

- “*Policy making*” is a business enabler
- “*Metadata service management*” is a process enabler
- “*Data profiling*” is a technology enabler.

Every Governance enabler should have a metric associated for measurement like person hours spent on Metadata management or number of business terms included in the glossary. The data stewards then, within each division, should get to document and publish divisional business and data value chains.

The stewards along with divisions discover the success factors and metrics used to measure the commercial success of the division like time to service, customer service effectiveness, cross sell ratio and many more. Then, a trace is established between governance enablers and the divisional value chains. Get the divisions to agree and own the metrics. This is what creates dialogue and awareness in the divisions where Governance needs to seep in. The value realization framework should clearly outlay the traceability between ***Governance enablers -> Technology and process impact -> Business impact and value.***

For example, Metadata capture and data mapping activities as enablers

1. Lead to reduction in person hours spent on performing data analysis
2. Further, reduces the requirements turnaround time
3. Finally, puts you in a competitive position in the market with product or Service time to market.

The lack of focus is common in data governance divisions across industries and is constantly affecting how people think of data governance activities. There are immediate and cumulative benefits from having to manage metadata or data quality. But, you need a Benefits Management and Assessment framework to realize the benefits from these services.

Some data governance divisions kickstart their initiatives with insightful business cases that rightly overcome organizational challenges. These business cases should clearly articulate tangible benefits of using the data management services. In other organizations, the data management services are implemented Big Bang across all divisions. In this case, benefits would be monitored and measured on a continuous basis in accordance with an assessment plan. Most organizations are not orchestrating governance activities as initiatives but rather as continuous push or pull based services. It is strongly recommended that one have a performance assessment plan before starting a data governance service be it data quality or metadata management. This plan should bring out the approach to monitor and measure the value of orchestration over specific timelines.

Quoting a second example; what does a 10% increase in the accuracy of the leads dataset mean to your marketing division and organization?

1. It directly impacts the ROI.
2. The division efficiently utilizes the budget allocated to the campaign and also embraces high conversion rate which leads to revenue increase.

And a third example as well; what does de-duplication of the customer primary accounts mean to the organization?

1. Less time is spent by the service executive on call with customer to get to the right customer account by not having to struggle with duplicate accounts.
2. This directly impacts your customer satisfaction score, cross sell ratio while also reducing operating costs with maintaining duplicate accounts and time spent in getting the right account.

# Measuring Benefits

Every Governance enabler should have a metric associated for measurement like person hours spent on Metadata management or number of business terms included in the glossary.

The data stewards then, within each division, should get to understand the divisional business and data value chains.

The stewards along with divisions discover the success factors and metrics used to measure the commercial success of the division like time to service, customer service effectiveness, cross sell ratio and many more.

Then, establish a trace between governance enablers and the divisional value chains.

Get the divisions to agree on these metrics and have them own the metrics.

This is what creates dialogue and awareness in the divisions where you would want to have governance to seep in. The framework as I said earlier should clearly outlay the traceability between governance enablers – Technology and process impact – Business impact and value.

For example, metadata capture and data mapping are governance enablers – which lead to reduction in person hours spent on data analysis – which further reduces the requirements turnaround time – finally, puts you in a competitive position in the market with product or service time to market.

Quoting a second example; what does a 10% increase in the accuracy of the leads dataset mean to your marketing division and organization? It directly impacts the ROI. The division efficiently utilizes the budget allocated to the campaign and also embraces high conversion rate which leads to revenue increase.

And a third example as well; what does de-duplication of the customer primary accounts mean to the organization? Less time is spent by the service executive on call with customer to get to the right customer account by not having to struggle

with duplicate accounts. This directly impacts your customer satisfaction score, cross sell ratio while also reducing operating costs with maintaining duplicate accounts and time spent in getting the right account