DATA GOVERNANCE FOR TAX ADMINISTRATIONS

A PRACTICAL GUIDE

Data Governance for Tax Administrations
A Practical Guide

Cover design:
Illustrates the butterfly effect that data errors can cause in tax administrations, with increasing consequences expanding in various data domains. The cover includes an adapted version of TwoLorenzOrbits by XaosBits CC BY 2.5. It is the trajectory of a Lorenz System.

About the Authors

Andrés Duque

Mr. Duque is an Ecuadorian national. He is a Systems Engineer with a master's degrees in Systems Management and Business Intelligence, as well as Data Management and Enterprise Architecture certifications. For 15 years he has participated and led data management and data governance projects in the public and private sectors.

He owns Datalegio, a data management consulting firm and is president of DAMA Ecuador Chapter. As an independent consultant he has supported the Inter-American Center of Tax Administrations - CIAT.

Wolney Martins

Mr. Martins is a Brazilian national. He holds a degree in engineering with a postgraduate degree in Computer Networks. For 20 years, he worked at SERPRO, the IT company of the Ministry of Finance of Brazil, where he held positions as an analyst, department head, director (operations, technology, and systems development) and president.

Since 2014, he is an independent IT consultant, working in public finance and tax administration systems, crisis management, IT planning and innovation, working for CIAT in several initiatives.

Antonio Seco

Mr. Seco is a Brazilian national from Cabo Frio-RJ. He holds a degree in Electrical Engineering, with a Master of Science degree in ICT Management and a postgraduate in Tax Administration. He was a SERPRO official in the Ministry of Finance of Brazil, where he participated in the implementation of SIAFI and other public finance systems, also having worked as consultant and IT component leader in Public Finance modernization projects in Latin America, the Caribbean and Africa.

He is an Independent Senior Consultant of the Inter-American Center of Tax Administrations (CIAT) and the Inter-American Development Bank (IADB).

Raul Zambrano

Mr. Zambrano is an Ecuadorian national. He is a Systems Engineer with a master's degree in public policy management and serves at the CIAT organization as Director for Technical Assistance and Technology.

He has directed international multidisciplinary teams in modernization projects carried out by CIAT in six countries and has participated in the improvement of processes and in the development and evaluation of information systems as a consultant in over 20 tax and customs administrations at national and subnational levels in Latin-America, the Caribbean and Africa.

Table of Contents

Chapter 1   Setting the Landscape

Information and knowledge are keys for organizations to fulfill their objectives.

The DAMA association[1] emphasizes that organizations with reliable, high-quality data about their users, products, services, and operations can make better decisions than those without. The absence of these properties will result in a waste of opportunities and deficient performance (DAMA-DMBoK2, 2017). This assertion is valid with greater emphasis for tax administrations, where data and its products are fundamental to accomplishing its mission.

1.1   Data, Information, Knowledge

A still current and passionate discussion in information sciences and knowledge management is the differentiation among data, information, knowledge, and (sometimes) wisdom.

Models available often present these concepts as a hierarchy, in which mastery of the lower level provides the opportunity to scale to the next level. This structured ascension is not a point of agreement among scholars, but it can be a starting point to understanding the concepts and establishing more precise communication among different users.

A theoretical model helps in understanding the transformations and relationships among these concepts.

1.2   The DIKW Model

Among the available models, one of the most visible, but not without controversies, is the so-called DIKW (Data, Information, Knowledge, Wisdom), presented in the form of a pyramid (Figure 1-1). One of the high points of the controversies is the inclusion and definition of the last attribute, "wisdom"[2].
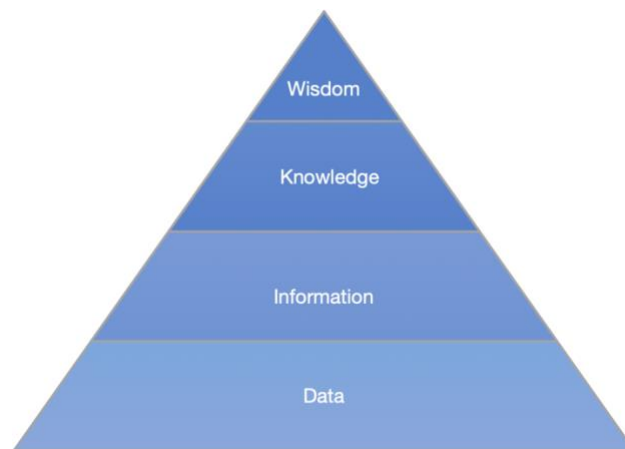
Figure 1-1 The DIKW model. Source: Prepared by the authors

---

[1] DAMA International is a non-profit, vendor-independent, global association of technical and business professionals dedicated to advancing the concepts and practices of information and data management (For more information, see: https://www.dama.org/cpages/mission-vision-purpose-and-goals)

[2] For more information, see http://wiki.km4dev.org/DIKW_model

The implicit assumption of this model is that tax administrations can use data to create information; information can be used to develop knowledge, and knowledge can be used to create wisdom.

The following definitions and associations to different types of information systems can be performed on this model:

Table 1-1 DIKW Model – elements definitions and information systems associations

Element	Definition (Ackoff, 1989)	Association (Rowley, 2007)
Data	Symbols	Transaction Processing Systems
Information	Data processed to be useful; provides answers to who, what, where, and when questions	Management Information Systems
Knowledge	Application of data and information; answers how questions	Decision Support Systems
Wisdom	Evaluated understanding	Expert Systems

Source: Prepared by the authors

1.3 The Growing Importance of Data Governance in Tax Administrations

Tax administrations are related to the automated processing of data from the beginning. After all, they were (along with the census bureau) the first users of the so-called "data processing machines" in government.

Tax returns and the provision of ancillary information in digital format by taxpayers and auxiliary institutions (especially financial institutions) have been part of the life of tax administrations and taxpayers in the recent past.

In those times, the data was structured with a minimal data management schema, consisting fundamentally of a data dictionary[3]. IT[4] personnel had control of the processes of extracting, transforming, and loading the data. The data needed to be cleaned[5], mostly manually.

Data management was the responsibility of the IT area, with occasional advice from the business areas. Thus, organizations merged data management with IT management.

Nowadays, data availability has increased dramatically in quantity and formats, as well as the dependence of tax administrations on its treatment. As established in (Collosa, 2021), this is mainly due to:

- The significant expansion of computer processing and storage capacity associated with the reducing their costs.
- The increasing availability of communications networks and broadband Internet.

---

[3] Defined as a set of information describing contents, format, and structure of a database and the relationship between its elements, used to control access to and manipulation of the database (*Oxford Languages*).

[4] Information Technology

[5] Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

- The development of effective models to capture, store and process massive data and advanced cognitive algorithms.
- The emergence of new data sources and formats e.g., sensors, GPS[6], OCR[7] cameras for truck plates, RFID[8] chips and antennas, social networks, etc. (Arias & Zambrano, 2020), including electronic invoices (Barreix & Zambrano, 2018) and tax information exchange between countries.

A few years ago, the importance of using data in the work of organizations was mentioned with a quote from the famous "total quality guru" W. E. Deming "without data, you're just another person with an opinion" (ETF-Europa, 2018). Currently, KPMG analysts have rephrased this quote: "without trust in your data, you're just another person that consumes data" (KPMG, 2021).

Tax administrations are strongly linked to this reality.

Over the past several years, tax administrations worldwide have started to undergo digital transformation, collecting data from non-traditional sources and formats, and accumulating them in their databases. Tax administrations can rely heavily on data and algorithms for their internal processes and provide more and better services to the taxpayers and other stakeholders, so tax administrations can count on data accuracy, completeness, and availability.

The following numbers illustrate these aspects as presented by the OECD

- From 2014 to 2019, average e-filing rates have increased significantly between 13 and 18%.
- Over 80% of payments (by value and numbers) are made electronically.
- Close to 50% of tax administrations pre-fill PIT (Personal Income Tax) returns with specific deductibles expenses.
- New data sources allow pre-filling to move to VAT (Value-Added Tax) and CIT (Corporate Income Tax) returns.
- A growing number of tax administrations use virtual assistants to respond to taxpayers enquires and support self-service.
- Use artificial intelligence in services supporting taxpayers and tax officials.
- Percentage of tax administrations that allow taxpayers to register online up from 70% (2015) to 97% (2019).
- With the increasing availability of data, compliance work focus can change to prevention.

At the same time, society demands more responsibility from the entities that obtain and consume data from citizens and companies, establishing a series of data protection laws and regulations.

---

[6] Global Positioning System

[7] Optical Character Recognition

[8] Radio Frequency Identification.

In this context, a modern data governance landscape must be set up to ensure data confidentiality, availability, quality, and integrity and reinforce the legal protection instruments (as data protection regulations) and compliance rules.

In other words, data governance must ensure that data are consistent and trustworthy and don't get misused, so as in the transactional operations up to enable the effective use of data analytics helping to optimize operations and drive business decision-making.

This data governance landscape includes all hierarchical levels of a tax administration, intending to define policies, standards, processes, and participating in data governance committees.

1.4   Data Management vs. Data Governance

Data is an essential asset within tax administrations. Data can give tax administrations different benefits through its use and exploitation, as well as through its correct administration.

To generate value, tax administrations require data. It needs to be managed consciously; for this, the organization must put a set of fundamental practices in place to allow it to manage data like any other business asset.

1.4.1   Data Management

According to DAMA (DAMA-DMBoK2, 2017), **Data Management** is defined as the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.
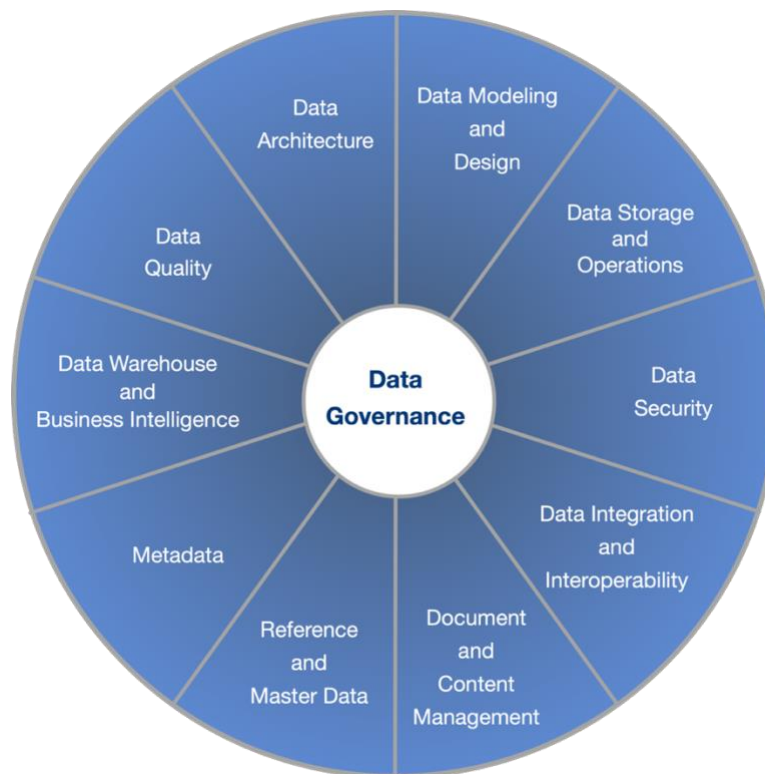
Figure 1-2 The DAMA-DMBoK2 Data Management Framework (The DAMA Wheel). Source: (DAMA-DMBoK2, 2017)

Organizations develop data management practices through different disciplines that cover all activities around the data lifecycles, e.g., Data Governance, Data Architecture, Data Quality, Business Intelligence, etc.

DAMA-DMBoK2 defines 11 disciplines for data management, with **data governance** at the center, as shown in Figure 1-2.

1.4.2 Data Governance in Data Management

As tax administrations face different challenges of information systems implementations, be it to support analytical capabilities, transactional, or business processes, it is recognized that data assets deserve to be managed correctly.

Traditionally, IT departments in organizations have been responsible for promoting data projects. Now, IT departments cannot operationalize these projects in isolation or without the commitment of the whole institution.

To manage data correctly, it is essential to have roles and responsibilities that allow accountability for the problems that data usually present and their inherent definitions. Here is where data governance intervenes as a framework that allows organizations to establish a system of rights and obligations for decision-making throughout the entire data lifecycle.

Data management requires a structure that controls and guarantees the correct administration of data, and that is why the implementation of data governance programs is gaining greater importance.

DAMA-DMBoK2 defines data governance as "the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets" (DAMA-DMBoK2, 2017). On the other hand, Ladley (Ladley, 2020) mentions that the purpose of data governance is to ensure that the data is managed properly, according to policies and best practices.

As we can see from DAMA-DMBoK2 Management Framework (Figure 1-2), data governance is at the **center** of all DAMA-DMBoK2 disciplines because it is crucial to control all kinds of data projects through centered guidance.

Data governance provides the best tools to manage data correctly, e.g., principles, policies, functions, processes, procedures, etc.

1.4.3 What is data governance all about?

Data governance is a key component of data management. Tableau (Tableau Software, 2020) proposes that data governance helps answer questions like:

- Who has ownership of the data?
- Who can access what data?
- What are security measures are in place to protect data and privacy?
- How much of our data is compliant with new regulations?
- Which data sources are approved to use?

Governance models and practices won't be the same across every organization, even among tax administrations, but these models are crucial pieces of the process. As also mentioned in the paper referenced above, the following stand out:

Data quality is a pillar of data management. It doesn't matter how robust your governance program is if you don't have quality data. Having data that is accurate, complete, and reliable is a cornerstone of any data-driven organization.

Data security and compliance is defining and labeling data by their levels of risk and then creating secure access points, keeping a balance between user interaction and safety, considering access levels that can go at the functional, object, or even field level (Martins, Nieto, Seco, & Zambrano, 2020).

Data stewardship helps monitor how teams use data, and stewards lead by example to ensure data access, security, and quality, defining clear interactions and responsibilities of different data stakeholders.

Data transparency matters because every piece of the process and the procedures you put in place should work within a model of clarity.

Analysts and business users should quickly find out where their data comes from and know if there are any special considerations.

1.4.4 Data Lifecycle

The data lifecycle is the sequence of stages a particular data unit goes through, from its initial generation or capture to its eventual archival or deletion at the end of its useful life (Wigmore, 2017).
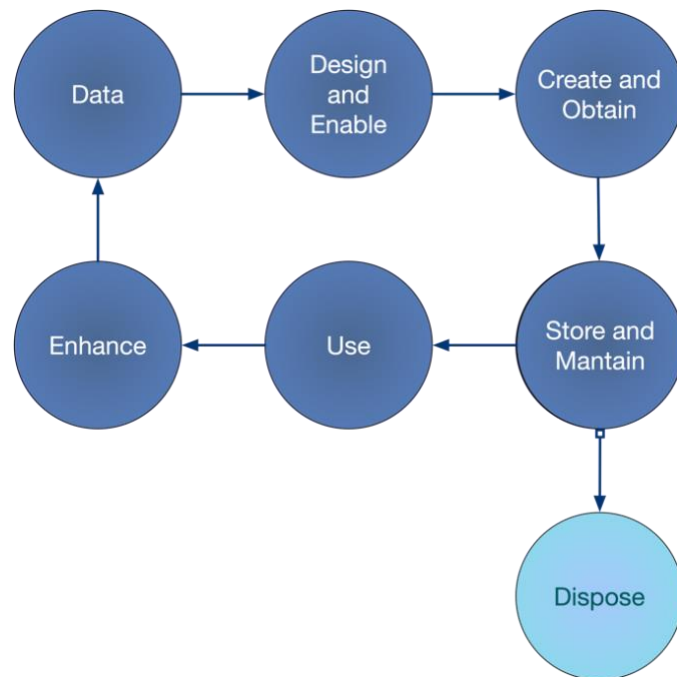
Figure 1-3 The data lifecycle key activities. Source: Prepared by the authors based on (DAMA-DMBoK2, 2017)

The data governance practices must cover all data lifecycle, as it is shown in Figure 1-3.

1.5   Data Attributes

Attributes are specification or characteristic that helps define a data entity. In data management, some attributes refer to the processing characteristics of the data and its lifecycle, use and structure, security requirements, quality parameters, and compliance needs.

The following topics present summaries of several essential data attributes for their management.

Specific chapters of this document will take up these attributes.

1.5.1   Common Business Vocabulary

A typical business vocabulary is a set of commonly defined data names and definitions documented in a business glossary, for example, within a data catalog or independently.

Its purpose is to ensure that data is consistently named and commonly understood, especially when it is shared.

A specialized software may supports creating and maintaining a business glossary with a common business vocabulary of common data names, definitions, and attributes for data entities. This is critical to promoting the proper common understanding and use of tax terms.

Most countries already have some formalization of tax terms, but often in scattered or incomplete documents. These documents can be good sources for everyday business vocabulary.

1.5.2   Master and Reference Data

According to (DAMA-Dictionary, 2009), Master Data is "the data that provides the context for business activity data in the form of common and abstract concepts that relate to the activity. It includes the details (definitions and identifiers) of internal and external objects involved in business transactions, such as customers, products, employees, vendors, and controlled domains (code values)".

Another definition by the consultant company Gartner Group for Master Data is the consistent and uniform set of identifiers and extended attributes that describe the core entities of the enterprise, including customers, prospects, citizens, suppliers, sites, hierarchies, and chart of accounts.

Transaction processing applications and analytical systems need Master Data, so they must be application agnostic.

An example of a Master Data, a subset of the suggested elements for taxpayer identification (Falkenbach, González, Redondo, & Zambrano, 2020), is shown below.

Table 1-2 Master Data (Taxpayer identification)

TIN	Taxpayer name	Taxpayer address	Telephone
07653457	John Bayrd Goode	345 Main Street, 87654	+1 505 5671234
88734509	Art Smith Vandelay	35 Johnson Blvd, 98543	+1 505 555 8765

TIN	Taxpayer name	Taxpayer address	Telephone
00456367	Cosmo K. Benes	42nd Street, 78900	+1 505 555 8245
99976543	Estelle Costanza	112 Aaron Ave, 87320	+1 505 555 1254

Source: Prepared by the authors

Reference Data is any data used to characterize or classify other data or to relate data to information external to an organization. The most basic Reference Data consists of codes and descriptions, but some Reference Data can be more complex and incorporate mappings and hierarchies (DAMA-DMBoK2, 2017).

Reference Data has characteristics that distinguish them from Master Data: they are less volatile; data sets are generally less complex and smaller; they have fewer columns and rows. The management focus differs between Master and Reference Data.

Among the types of Reference Data, we mention Internal Reference Data (created to support internal processes and applications), Industry Reference Data (created and maintained by industry associations or government bodies), and Computational Reference Data (which differs from other types because of the frequency with which it changes).

Reference data could be presented and used in many ways, using a code-value strategy or fixed labels (Zambrano, 2010). A basic Reference Data example is shown below.

Table 1-3 Reference data (list)

Code Value	Description
AR	Argentina
BR	Brazil
PY	Paraguay

Source: Prepared by the authors

1.5.3   Metadata

The Gartner Glossary defines metadata as "information that describes various facets of an information asset to improve its usability throughout its lifecycle" (Gartner Inc., 2012). The DAMA, in (DAMA-DMBoK2, 2017), adds other features: "metadata" includes information about technical and business processes, data rules and constraints, and logical and physical data structures. It describes the data itself (e.g., databases, data elements, data models), the concepts the data represents (e.g., business processes, application systems, software code, technology infrastructure), and the connections (relationships) between data and concepts.

Necessary for structured data, metadata is perhaps most important for unstructured data (see the basics of structured and unstructured data later in this chapter). New practices are emerging for treating unstructured data in data lakes, for example, a minimum set of metadata attributes of ingested objects is collected as part of the ingestion process, such as name, format, source, version, and date received, producing a catalog.

There is also a requirement for a metadata lineage to provide an audit trail to know where the data originated and how it has been transformed in this way to the point of use. It may also trace who or what is maintaining data, including when and where it occurs.

Metadata turns information into an asset, and accurate metadata can help prolong the lifetime of existing data by assisting users in finding new ways to apply it.

Many IT tools are available to deal with metadata, as we will see later in this document.

1.5.4 Operational and Analytical Data

The world of data used to be divided between the applications and processes creating and updating data (operational) and the solutions and processes analyzing data (analytical). The two are structurally different and provide different types of insight.

Operational data is produced by the day-to-day operations of a tax administration, such as changes in the tax registry, tax payments, taxpayers' appeals, etc. Operational data are produced mainly by the OLTP[9] systems, supporting high-volume, low-latency access. These systems create, read, update, or delete one piece of data at a time.

Analytical data is used to support business decisions, instead of recording the data from actual operational business processes. Examples include grouping taxpayers by income or amount of tax due over time. Every organization will have different questions to answer and other decisions, so analytical data is definitely not one-size-fits-all. Analytical data is best stored in a system designed for heavy aggregation, data mining, and *ad hoc* queries, called an OLAP[10] system or a Data Warehouse (Simpson, 2016).

The core of the analytical data is the institution's operational data.

Figure 1-4 Operational and analytical data. Source: Prepared by the authors based on (Simpson, 2016)

---

[9] On-Line Transaction Processing

[10] Online Analytica Processing system

Operational databases contain transactional data, while analytical databases are designed for efficient analysis, as can be seen in Figure 1-4.

1.5.5   Structured and Unstructured Data

According to Talend (Talend Company, 2020) structured data is data that has been predefined and formatted to a set structure before being placed in data storage, which is often referred to as schema-on-write[11]. The best example of structured data is the relational database: the data has been formatted into precisely defined fields, such as tax identification numbers or addresses, to be easily queried with programming languages like SQL.

The same source establishes that unstructured data is stored in its native format and not processed until it is used, known as schema-on-read[12]. It comes in various file formats, including email, social media posts, presentations, chats, IoT sensor data, audio, and imagery.

Structured data is highly specific and is stored in a predefined format, whereas unstructured data is a conglomeration of many varied types of data stored in their native formats.

An intermediary model, semi-structured data, refers to what would typically be considered unstructured data but also has metadata that identifies specific characteristics. The metadata contains enough information to make the data more efficiently cataloged, searched, and analyzed than strictly unstructured data.

An estimated 80 percent of all data is kept outside of relational databases. This unstructured data does not have a data model that enables users to understand its content or how it is organized; it is not tagged or structured into rows and columns (DAMA-DMBoK2, 2017).

Talend proposes the following differentiation between structured and unstructured data:

Table 1-4 Differentiation between structured and unstructured data

	Structured data	**Unstructured data**
Who	Self-service access	Requires data science expertise
What	Only select data types	Many varied types conglomerated
When	Schema-on-write	Schema-on-read
Where	Commonly stored in data warehouses	Commonly stored in data lakes
How	Predefined formats	Native format

Source: Prepared by the authors based on *(Talend Company, 2020)*

1.5.6   Security and Privacy

Data security practices aim to protect information assets in alignment with privacy and confidentiality regulations, contractual agreements, and business requirements.

---

[11] The data needs a schema established to be uploaded.

[12] The data is uploaded in its native format. The schema is created later when the data is read.

All data management experts highlight "data protection" as the primary driver for data governance (Microsoft Corporation, 2020). This is needed primarily to prevent data breaches and remain compliant with data privacy with regulatory legislation such as the European Union General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and other specific regulations for tax administrations.

Data privacy and the growing number of data breaches have made data protection a top priority in the tax administrations' C-level planning. These breaches highlight the risk to sensitive data such as personally identifiable customer data.

Tax administrations are increasingly concerned with the sensitive data they handle since, in addition to general rules, they must follow specific regulations in the tax area.

The consequences of a data privacy violation or a data security breach are numerous and include (Microsoft Corporation, 2020):

- Loss or severe damage to the institution and government images.
- Loss of citizens' trust.
- Significant financial penalties because of audit/compliance failure.
- Legal action.
- The 'domino effect' of the breach, e.g., taxpayers may also fall victim to identity theft because of a breach.

Thus, data security encompasses defining, planning, developing, and executing security policies and procedures to provide proper authentication, authorization, access, and auditing of data and information assets (DAMA-DMBoK2, 2017).

1.5.7 Data Classification

Data can be classified by the type of data, by content, by format, by the level of data protection required, or by how and where it is stored or accessed (DAMA-DMBoK2, 2017), e.g.:

- Type of Data
  - Transactional Data
  - Master Data
  - Reference Data
  - Metadata
- Format
  - Character
  - Float
  - Integer
- Stored
  - Structured Data
  - Semi-structured Data
  - Unstructured Data
- Security/Privacy (Level of Confidentiality)
  - Public
  - Internal use only
  - Confidential

- Sensitive personal data
- Restricted

Throughout the data lifecycle according to its data classification, each may apply different management requirements, for example, according to security/privacy policies and rules are applied according to the label of data confidentiality.

Data classification according to security/privacy is increasingly essential due to existing data protection laws and regulations in countries and increasing information exchange agreements between national tax administrations.

1.5.8 Data Retention

Especially for tax administrations, data preservation is a critical legal requirement, and so data retention policies assume a crucial role in the data lifecycle.

According to DAMA-DMBoK2 (DAMA-DMBoK2, 2017), "the risks of not having defined a proactive litigation response should be assessed and quantified. Sometimes organizations respond only if there is anticipated litigation, and then there is a scramble to find relevant documents and information to review. Most likely, this type of organization either over specifies the amount of data to be kept (i.e., everything) or does not have data deletion policies in place. Not having a retention schedule for data and information can lead to legal liabilities if older unpurged records are required for e-discovery, but not available."

Introducing new ethical behavior also affects a data retention program and practices, such as "the right to be forgotten" (to have information about an individual be erased notably to adjust reputation).

Data retention policies also affect the planning for data storage acquisition, database recovery and business continuity plans, and database performance. Data retention plans will differ by data domain and data type.

If the tax administration can discard the data, it is not enough to delete it. In many cases, legislation requires data to be destroyed.

1.5.9 Data Lineage

Data lineage includes the concept of an origin for the data—its source or provenance—and the movement and change of the data as it passes through systems and is adopted for different uses, i.e., the sequence of steps within the data chain through which data has passed (Sebastian-Coleman, Measuring Data Quality for Ongoing Improvement, 2013).

From data quality and governance perspectives, it is essential to understand data lineage to ensure that existing business rules subsist where expected, calculation rules and other transformations are correct, and system inputs and outputs are compatible. Data traceability is tracking access, values, and changes to the data flow through their lineage (Allen & Cervo, 2015).

Data traceability can be used for data validation and verification, and auditing. In summary, data lineage is the documentation of the data lifecycle, while data traceability is evaluating that the data follows its expected lifecycle.

According to Wikipedia[13], data lineage information includes technical metadata involving data transformations. Enriched data lineage information may consist of data quality test results, reference data values, data models, business vocabulary, data stewards, program management information, and enterprise information systems linked to the data points and transformations.

Data lineage may be part of the data catalog, which allows to have a comprehensive analysis of the data from its sources and what its flows are.

1.5.10 Data Masking

Also known as data obfuscation, de-identifying, or anonymizing, it is a data security technique that copies and scrambles sensitive data, often via encryption, as a means of concealment. Data masking scrambles data to anonymize it.

In general, the more critical data to be masked are the Personally Identifiable Information – PII, which refers to information that can be used to identify, contact, or locate a single person. They can also be used with other sources to identify a single individual.

With the advent of greater regulatory rigor in data protection, this technique is essential for implementing a tax administration's data privacy policies. Data masking also helps to minimize the risk of personal and business-sensitive information being leaked, breached, or used without authorization.

There are some data masking techniques that can be evaluated (Privitar Ltd, 2022), each with their pros and cons: redaction (to delete any value that aren't necessary); hashing (convert an original value into a fixed-length output known as "hash"); encryption (algorithms to replace an original value-plaintext with another value-ciphertext); tokenization (replace an original value with a randomly generated equivalent); generalization (transform an original value into one's that is more general); substitution (replace an original value with another value from a predefined list); perturbation (adds random "noises" to an original value).

There are several technologies and products on the market for automated data masking, static (direct in the database) and dynamic (real-time masking).

It is part of data management to define, in each data domain, the fields that must be subject to masking at the user and application levels. The data exchange with other institutions is another aspect where the masking level must be severely evaluated.

Another area of recent studies is the masking of data in activities related to artificial intelligence (machine learning) to avoid the appearance of bias.

1.5.11 Cloud Systems, Data, and Sovereignty

According to Seco and Muñoz (Seco & Muñoz, 2018), there are concerns in some nations regarding the notions of "cloud sovereignty," which are primarily connected to the physical location of server hardware and cloud storage, local laws, and the rules that will be applied in the event of divergences (primarily related to judicial access to information). As a relatively new concept, cloud sovereignty is

---

[13] For more information, see: https://en.wikipedia.org/wiki/Data lineage

still not clearly defined, but in brief, a sovereign cloud assures that all data, including metadata, stays on sovereign territory and always forbids access to data from outside the country[14].

These concerns stem from rising geopolitical tensions, shifting data privacy laws[15], and the dominance of selected cloud players.

A Capgemini survey (Capgemini, 2022) indicates that 70% of public sector firms are concerned about operational dependency on vendor based outside of their region's jurisdiction; 69% of them believe that a sovereign cloud will be adopted to ensure immunity from extra-territorial laws.

About this topic, the following recommendations were picked-up from Middleton (Middleton, 2022):

- Define sovereignty objectives; understand the laws of the land for digital sovereignty; track key developments in the cloud and data- sovereignty space; continuously assess risk exposure; and set up a compliance organization.
- Assess cloud providers through a sovereignty lens – including **data sovereignty** (for data residency, controls, transparency, storage, back-ups, etc.); **operational sovereignty** (for security, compliance, and operational resilience); and **technical sovereignty** (to assess interoperability, migration features, and clear exit policy/process).
- Align for a flexible cloud architecture: Identify your sensitive workloads and most viable use cases; consider end-to-end encryption, as well as key management solutions. At the same time, evaluate hybrid options, and prepare for a multi-cloud architecture by understanding the potential as well as the challenges it brings

As proposed also by Seco & Muñoz (Seco & Muñoz, 2018), if a tax administration establishes that it intends to use the cloud, a stopgap solution, while legal aspects are discussed, is the classification of data under its management, identifying which information is sensitive to national security or sovereignty, freeing the rest for transfer to the cloud.

1.5.12 Data Domain

According to the DAMA Dictionary of Data Management (DAMA-Dictionary, 2009), a data domain is "a set of allowable values for a data attribute." However, other experts have more aligned definitions with data governance concepts. For example, a description more suitable for data governance would be "a logical grouping of items of interest to the organization, or areas of interest within the organization" (Firican, What is a Data Domain? (Examples Included), 2020) also knows as Subject Area. In data governance terms, data domains are high-level categories of data to assign accountability and responsibility for the data.

The data domain concept's essential phrase is "assign accountability and obligation." Data domains are typically assigned to data owners and other data stewards.

---

[14] For more information, see: https://www.cio.com/article/308751/why-sovereign-cloud-is-a-hot-topic-5-tips-and-the-background.html

[15] As the US "Clarifying Lawful Overseas Use of Data (CLOUD)" Act of March 2018 (For more information, see: https://www.cloudsigma.com/the-cloud-act-what-you-need-to-know/)

A data domain may be formed in tax administrations with a broad vision, such as taxpayer data, data from external financial sources, data from social media, etc., or with a more concentrated vision, such as tax returns, data from external sources through agreements, data from public sector sources, etc.

1.5.13 Data Quality

Quality is one of the most important attributes of data. Data quality may be defined as "a measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability and whether it's up to date. Measuring data quality levels can help organizations identify data errors that need to be resolved and assess whether the data in their IT systems is fit to serve its intended purpose" (Vaughan, 2019).

Many data scientists said to a specialized blog[16] that 90% of its job is just collecting the data, putting it in a consistent form, and dealing with the endless holes or mistakes. Therefore, quality management policies and procedures must be established since data generation or capture; the sooner the data quality issues are detected and solved, the lower costs, and faster they become available.

To avoid focus dispersion, prioritizing solving data quality problems is very important. This can be done, for example, by considering the business impact, frequency, and complexity of the issues.

---

[16] For more information, see https://www.cio.com/article/402076/11-dark-secrets-of-data-management.html

Chapter 2   Data Governance at-a-glance

When deciding to build a data governance system, a tax administration likely has at least the following general objectives:

- Control and supervise the correct management of data throughout the data lifecycle.
- Manage data as an organizational asset.
- Increase the privacy and security of data.
- Improve the quality of the data that is used by information systems.
- Regulate and monitor access to sensitive data.
- Use timely data analytics to improve operations and corporate decision-making.
- Obtain and ensure compliance with data privacy and security standards on an ongoing basis.
- Avoid or reduce data breaches and other cyber security threats.
- Move towards a data-driven culture.
- Define a data responsibilities and accountability agreement for data.

To build such a system, starting with a data governance framework is recommended.

2.1   Data Governance Frameworks

According to NASCIO (NASCIO National Association of Chief Information Officers, 2009):

> *"Frameworks (in general) assist in describing major concepts and their interrelationships. Frameworks assist in organizing the complexity of a subject. Frameworks facilitate communications and discussion.*
>
> *All of these descriptors apply as well to frameworks related to data governance. Additionally, data governance frameworks assist in demonstrating how data governance relates to other aspects of data management, data architecture, and enterprise architecture".*

Talend, a company that works with data health and business objectives, sustains that a framework provides some essential benefits, including[17]:

- A consistent view of — and business glossary for — data while allowing appropriate flexibility for the needs of individual business units.
- A plan that ensures data quality, accuracy, completeness, and consistency.
- An advanced ability to understand the location of all data related to critical entities, making data assets discoverable, usable, and easier to connect with business outcomes — in other words, ensuring.
- A "single version of the truth" that keeps critical business entities aligned across the enterprise.

---

[17] For more information see https://www.talend.com/resources/data-governance-framework/

- Well-defined methodologies and best practices for data assets and data management that can be applied across the organization.
- Easily accessible data that are kept secure, compliant, and confidential according to the demands of legal or regulatory requirements.

A data governance framework describes how all the pieces that compose data governance fit together.

Data Governance and Stewardship

Definition: The exercise of authority, control, and shared-decision making (planning, monitoring, and enforcement) over the management of data assets.

Goals:
1. Enable is organization to manage its data as an asset.
2. Define, approve, communicate, and implement principles, policies, procedures, metrics, tools and responsibilities for data management.
3. Monitor and guide policy compliance, data usage, and management activities.

Inputs:
- Business Strategies and goals
- IT strategies and goals
- Data Management and Data Strategies
- Organization Policies and Standards
- Business Culture Assessment
- Data Maturity Assessment
- IT Practices
- Regulatory Requirements

Activities:
1. **Define Data Governance for the Organization (P)**
   i. Develop Data Governance Strategy
   ii. Perform Readiness Assessment
   iii. Perform Discovery and Business Alignment
   iv. Develop Organizational Touchpoints
2. **Define the Data Governance Strategy (P)**
   i. Define the Data Governance Operating Framework
   ii. Develop Goals, Principles and Policies
   iii. Underwrite Data Management Projects
   iv. Engage Change Management
   v. Engage in Issue Management
   vi. Assess Regulatory Compliance Requirements
3. **Implement Data Governance (O)**
   i. Sponsor Data Standards and Procedures
   ii. Develop a Business Glossary
   iii. Coordinate with Architecture Groups
   iv. Sponsor Data Asset Valuation
4. **Embed Data Governance (C,O)**

Deliverables:
- Data Governance Strategy
- Data Strategy
- Business / Data Governance Strategy Roadmap
- Data Principles, Data Governance, Policies, Processes.
- Operating Framework.
- Roadmap and Implementation Strategy
- Operations Plan
- Business Glossary
- Data Governance Scorecard
- Data Governance Website
- Communications Plan
- Recognized Data Value
- Maturing Data Management Practices

Suppliers:
- Business Executives
- Data Stewards
- Data Owners
- Subject Matters Experts
- Maturity Assessors
- Regulators
- Enterprise Architects

Participants:
- Steering Committees
- CIO
- CDO / Chief Data Stewards
- Executive Data Stewards
- Coordinating Data Stewards
- Business Data Stewards
- Data Governance Bodies
- Compliance Team
- DM Executives
- Change Managers
- Enterprise Data Architects
- Project Management Office
- Government Bodies
- Audit
- Data professionals

Consumers:
- Data Governance Bodies
- Project Managers
- Compliance Team
- DM Communities of Interest
- DM Team
- Business Management
- Architecture Groups
- Partner Organizations

Technical Drivers

Techniques:
- Concise Messaging
- Contact List
- Logo

Tools:
- Websites
- Business Glossary Tools
- Workflow Tools
- Document Management Tools
- Data Governance Scorecards

Metrics:
- Compliance to regulatory and internal data policies
- Value
- Effectiveness
- Sustainability

(P) Planning, (C) Control, (D), Development, (O) Operations

Figure 2-1 DAMA-DMBoK2 data governance and stewardship context diagram. Source: (DAMA-DMBoK2, 2017). Redrawn for clarity.

One of the most known frameworks in data management is the DAMA-DMBoK2 framework, which has data governance as a significant knowledge area/discipline.

Data Management Association International (DAMA) published a data management body of knowledge (DAMA-DMBoK2, 2017) that provides context diagrams that include goals for each objective, with business and technical drivers; activities and roles; and inputs and outputs. An example of the context diagram for data governance and stewardship is presented in Figure 2-1.

> The DAMA data governance knowledge areas has four primary objectives:
>
> - Data governance and Stewardship - guarantee roles and responsibilities that describe and enforce rules of engagement, decision rights, and accountabilities for valuable data and information assets management.
> - Business Cultural Development - the process of influencing a data-driven culture of the tax administration over time.
> - Data in the Cloud – evaluate the impacts of moving data to the cloud.
> - Data Handling Ethics - a code of behavior encompassing the generation, recording, curation, processing, dissemination, sharing, and use of data.

It is complex and almost impossible to follow the DAMA-DMBoK2 framework entirety, but it can serve as a guiding basis for customized models.

Usually, customized data governance frameworks incorporate different aspects of data governance, but the following characteristics are present[18]:

- Accountability and leadership roles in the organization.
- Planning and rules for data handling – quality, integrity, and access.
- Strategic enterprise perspective.
- Cultural change to a data-centric organization.

2.2   Data Governance Policies

A policy is a definite course or method of action selected from among alternatives and considering given conditions to guide and determine present and future decisions[19].

Good data governance policies ensure that your organization's data assets are formally, appropriately, effectively, and proactively managed.

Data governance policies are applied to the entire data lifecycle. They spread from gathering of data to revising and standardizing the information collected (Rouse, 2021) to organizing that information to gain valuable insights into your business and your customers. Good data governance policies

---

[18] See https://cdn.atlantaregional.org/wp-content/uploads/data-governance-best-practices.pdf
[19] See the Merriam-Webster online dictionary - https://www.merriam-webster.com/dictionary/policy

ensure that the right person can access the correct data at the right time and effectively balances that access against security, compliance, and privacy concerns.

The participation of all key stakeholders is essential in terms of policy definition (at least in its conceptualizing). After policies have been detailed, standardized, and disseminated widely in the institution, some stakeholders will need educational events to fulfill or enforce each policy, as part of the change management plan.

As with all data governance aspects, there is no "one size fits all" approach. However, a set can be selected from the standard policies in this type of initiative gradually, depending on the pace of the implementation.

A list of these policies, adapted from Rouse (Rouse, 2021), is presented below:

- Access policy

    Policies for data access permissions are one of the most sensitive points of data security. Depending on the classification of data, access conditions may vary. Access to sensitive information must be recorded. Another critical point is the requirements for revoking access, especially for former employees.

- Usage policy

    Usage policies refer to privacy and compliance. The data processed by tax administrations have, in this area, a legal framework to be followed. Additionally, new laws and regulations passed in several countries (data protection laws) reinforce the special attention given to this area.

- An integrity and integration policy

    Data must be accurate, up-to-date, and easily accessed by the relevant stakeholders.

    Data quality and integration (how information systems will interchange data) policies are also in this category.

- Policies governing the protection, handling, and security

    These policies classify sensitive data and determine how tax administration should handle data and with what safeguards. Some data has laws protecting it and restrictions related to collection and storage.

- Provenance policies

    Critical data needs to be traced back to its origin for compliance reasons. These policies aim to help users reuse data while ensuring that data is safeguarded from improper use, misinterpretation, or non-compliance with data use agreements.

- Storage and retention policy

    Certain sensitive data may be discarded after a specific amount of time, by compliance or agreement, or, on the contrary, kept safe for a defined number of periods.

2.3   Data Governance Processes

To scale the data governance initiative, teams need well-defined and repeatable processes designed for the reality of each task.

Microsoft (Microsoft, 2022) categorizes four types of data governance processes:

Table 2-1 Data governance process categories

Process category	Processes
Data discovery processes to understand the data landscape	A data and data entity discovery, mapping, and cataloging process A data profiling discovery process to determine the quality of data A sensitive data discovery and governance classification process A data maintenance discovery process for CRUD analysis, such as from log files, to understand usage and maintenance of data such as master data across the enterprise
Data governance definition processes	Create and maintain a common business vocabulary: a business glossary defines data entities, including master data, data attribute names, data integrity rules, and valid formats Define reference data to standardize code sets across the enterprise Define data governance classifications schemes to label data to determine how to govern it Define data governance policies and rules to manage data entities and document lifecycles Define success metrics and threshold
Data governance policy and rule enforcement processes	A process to automate the application and enforcement of data governance policies and rules A method to manually apply and enforce policies and rules Event-driven, on-demand, and timer-driven (batch) data governance processes are published as services that tax administration can invoke to govern: • Data ingestion – cataloging, classification, owner assignment, and storing • Data quality • Data access security • Data privacy • Data usage, for example, includes sharing and ensuring licensed data is only used for approved purposes • Data maintenance, such as master data • Data retention • Master data and reference data synchronization
Monitoring processes	Monitor and audit data usage activity, data quality, data access security, data privacy, data maintenance, and data retention Monitor policy rule violation detection and resolution

Source: (Microsoft, 2022)

2.4   Data Governance Roles

To accomplish the data governance goals and principles, a tax administration must engage a diverse audience of professionals throughout the institution, full or part-time, who must fulfill a comprehensive set of roles.

A non-exhaustive list of these roles is presented in Figure 2-2
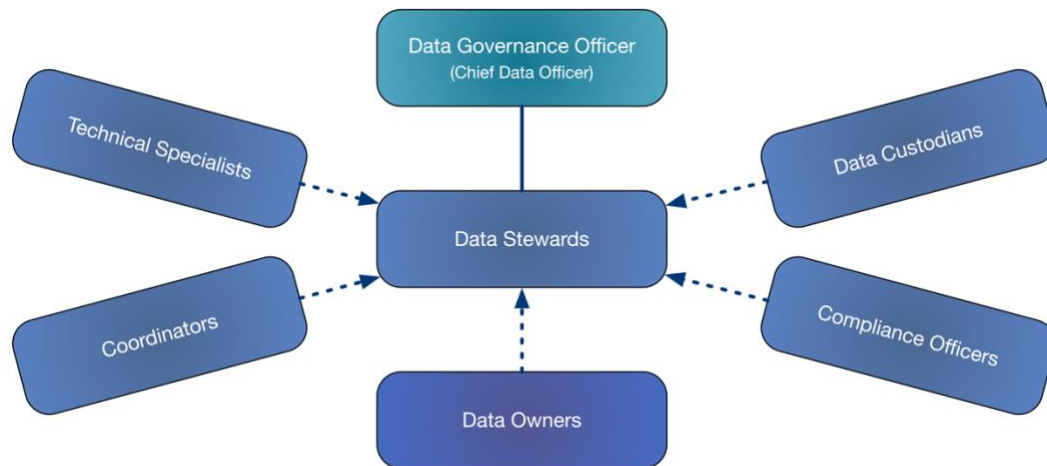
Figure 2-2 Data governance roles. Source: Prepared by the authors

As can be seen, the main line of roles is formed by the **Data Governance Officer** (or a Chief Data Officer), **Data Stewards** (with different task flavors), and the **Data Owners** (which can also be assumed to be a unique type of data steward).

These roles do not necessarily mean boxes on the institution's organizational chart. They are roles that must be performed and could be concentrated on a few corporate boxes or people, depending on the institution's size.

This consideration refers especially to data stewards, whose roles are concentrated or distributed according to the institution's size. The institution's structure - centralized, decentralized, federated, etc. - also influences the transformation of roles into boxes in the organization chart.

These leading roles will be briefly described below[20].

Data Governance Officer or Chief Data Officer

Depending on the institution's maturity or structural organization, the responsibilities for driving a data governance program rests with a C-Level Chief Data Officer or, failing that, a high-level Data Governance Officer. Its responsibilities are:

- Establish an organizational data strategy.
- Align data-centric requirements with available IT and business resources.
- Establish data governance standards, policies, and procedures.

---

[20] The role descriptions were mainly based on DAMA-DMBoK2, Rego (2020), and Herzberg (2021).

- Provide advice (and perhaps services) to the business for data-dependent initiatives, such as business analytics, Big Data, data quality evaluation and improvement, and data technologies adoption.
- Evangelize the importance of sound information management principles to internal and external business stakeholders.
- Oversight of the data usage in Analytics and Business Intelligence.
- Head the Data Governance Steering Committee.

Data Stewards

They are the professionals who work most intensively with data governance. Data stewards are typically subject matter experts who are familiar with the data used by a specific business function or department. They ensure the fitness of data elements, both content and metadata, administer the data and ensure compliance with regulations.

Several types of Data Steward carry out coordinating and operational roles. However, depending mainly on the data governance program's scope and the institution's size, some positions can be concentrated on one person. In general, a Data Steward refers to a data domain. Here are some critical roles:

Business Data Stewards are business professionals, most often recognized subject matter experts, accountable for a subset of data. They work with stakeholders to define and control data.

Technical Data Stewards are IT professionals operating within one of the Knowledge Areas, such as Data Integration Specialists, Database Administrators, Business Intelligence Specialists, Data Quality Analysts, or Metadata Administrators.

Compliance Officer is concerned about data regulatory and statutory issues, such as records retention schedules, location, transport, and destruction. Some data about individuals, for instance, cannot cross international boundaries, and some taxpayer data are protected against exchange or dissemination.

Data Custodian must ensure that access to the data is authorized and controlled; technical processes uphold data integrity; there are methods for resolving data quality issues (in collaboration with other data stewards); technical controls safeguard data; and data added to data sets are consistent with the standard data model. Additionally, versions of master data must be maintained along with a history of changes; change management procedures must be used in the database's upkeep.

Data Owner

Data Owner is a business Data Steward, having the authority to approve decisions about data within their domain. Also known as a data curator or tutor, the Data Owner is a business professional responsible for formally representing a data set or concept before the company and the external public, including regulatory bodies, suppliers, and the community in general. Depending on the institution's characteristics and the sector regulation mechanisms, this role may be liable for any negligence with data under his responsibility. Some commonly assigned responsibilities:

- To sponsor actions to solve data problems.

- To participate as a full member of the Data Governance Steering Committee.
- To authorize access to data under their responsibility, following current data security and privacy policies.
- To authorize the sending of data under his responsibility to companies and external entities.
- To attribute the security classification levels to the data.
- To define priorities related to the acquisition and utilization of new data (supported by the business Data Steward).
- To decide questions on the data usage, together with the business Data Steward.
- To represent the institution in regulatory matters (about the data for which it is responsible).

According to Herzberg (Herzberg, 2021), to fulfill the obligations listed above, a Data Owner needs the authority to make any changes required in terms of workflows, practices, and infrastructure to ensure data quality; and the resources to initiate actions to ensure data quality, such as data cleansing and data audits.

In practice, the institution must assign the Data Owner role to someone relatively senior, typically in upper management. Without adequate authority and access to resources, a Data Owner will be ineffective at fulfilling their role. This shortcoming cascades down the entire data governance chain, defeating the whole initiative.

Stakeholders

Stakeholders are all those with interest in an organization[21]. In a tax administration, there are a lot of stakeholders, beginning with all citizens, especially the citizens who pay taxes (the taxpayers).

Examples of stakeholders in a tax administration data governance initiative are:

- Government and industry bodies/ministries
- Tax intermediaries (accountants, advisors, tax agents, practitioners, bookkeepers)
- Stakeholders in the compliance chain (providers of cash register/POS systems, invoicing solutions, and accounting software)
- Civil society (media, academia, training institutions, unions, civil society organizations)
- Financial organizations
- Tax administration employees

Stakeholders need to be identified and defined as to why it is crucial for the initiative's success. In 8.6 Data Governance Stakeholder Identification Guide, as presented in Chapter 8, there is a guide for identifying new stakeholders.

DAMA-DMBoK2 proposes a stakeholder interest map (Figure 2-3) to help to prioritize based on their influence, their level of interest in the program, or the degree to which the program will impact them.

---

[21] See the Oxford Reference Online (www.oxfordreference.com)
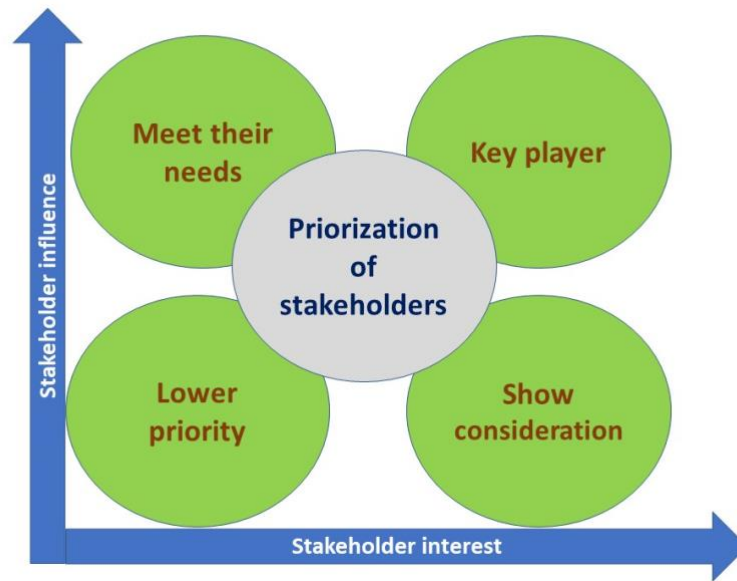
Figure 2-3 Stakeholder interest map. Source: DAMA-DMBoK2

The same reference suggests that the data governance team should investigate why each stakeholder is necessary to the initiative's success. This investigation means understanding their personal and professional goals and linking the output from data management processes to their goals so they can see a direct connection. Without an understanding of this direct connection, they may be willing to help in the short term, but they will not provide long-term support or assistance.

2.5   Data Governance Committees and Councils

In addition to the roles abovementioned, certain committees or councils are recommended to coordinate activities to accomplish the data management objectives. The number and responsibilities of the committees vary according to the size and structure of the institution.

A set of roles, committees and councils, as proposed in DAMA/DMBoK2, is presented in Table 2-2

Table 2-2 Roles, Committees, Councils

Data Governance Body	Description
Data Governance Steering Committee	The primary and highest authority of data governance in an organization, responsible for oversight, support, and funding of data governance activities. It consists of a cross-functional group of senior executives headed by the Chief Governance Officer or Chief Data Officer.
	Typically releases funding for data governance and data governance sponsored programs as recommended by the Chief Data Officer (CDO) or Data Governance Manager (DGM). This committee may, in turn, have oversight from higher-level funding or initiative-based Steering Committees.
Data Governance Council (DGC)	Manage data governance initiatives (e.g., development of policies or metrics), issues and escalations. It consists of executives arranged according to the operating model (centralized, replicated, federation, etc.).

Data Governance Office (DGO)	Ongoing focus on enterprise-level data definitions and data management standards across all DAMA-DMBoK2 knowledge areas consists of coordinating roles labeled as data stewards, custodians, and data owners.
Data Stewardship Teams	Communities of interest focused on specific subject areas or projects, collaborating or consulting with project teams on data definitions and data management standards related to the focus. It consists of business and technical data stewards and data analysts.
Local Data Governance Committees	Large institutions may have divisional or departmental data governance councils working under an enterprise DGC. Smaller organizations should try to avoid such complexities.

Source: (DAMA-DMBoK2, 2017)

2.6   Data Governance Roles and the IT Department

The data governance roles must not be confused with the functions of an IT department.

IT departments hold mainly data management roles. However, some data governance roles must be accomplished by an IT department, particularly (as other business departments) some specialized data stewards. Figure 2-4 depicts a comparative structure of roles between the two areas. IT management roles are focused on managing technology assets, while data management roles are focused on managing the data assets itself throughout its lifecycle.

The figure does not represent an organization chart but a form of functional/operational dependencies.
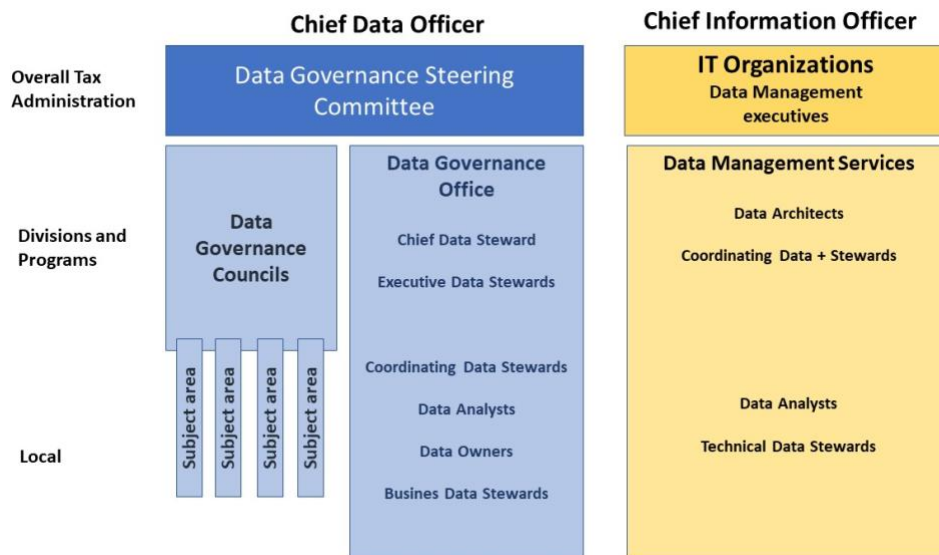
Figure 2-4 Data governance and the IT Department. Source: Adapted from (DAMA-DMBoK2, 2017)

Tax administration might use data governance councils and specific roles in the Data Governance Office, mainly in decentralized or large corporations.

2.6.1 About the organizational titles of a data governance structure

Some new designations for the professionals involved with data governance should be noted, even if several of these functions were already being performed formally or informally within the organization.

Although, as mentioned by Shakespeare[22], "a rose by any other name would smell as sweet," developing and implementing the proper titles can demonstrate to the entire organization that governance and stewardship have brought a new culture to the landscape. Using the correct tags can assist in developing a sustainable appreciation for the beneficial nature of enterprise data governance and management.

2.7 Data Literacy

With the growing reliance on data usage in a tax administration, having a data-literate staff is increasingly important. Data literacy may be defined briefly as the ability to understand, share common knowledge, and have meaningful conversations about data (Panetta, 2021). Figure 2-5 shows this definition graphically.

According to the same source, poor data literacy is ranked as the second-biggest internal roadblock to the success of the Chief Data Officer or equivalent (Gartner Annual Chief Data Officer Survey). By 2023, data literacy will become essential in driving business value, demonstrated by its formal inclusion in over 80% of data and analytics strategies and change management programs.

Figure 2-5 Data Literacy. Source: venngage.com blog

---

[22] "Romeo and Juliet," as quoted by Anne Marie Smith

The significance of data literacy in all life contexts, with special meaning in business, can be summarized in this message (Bersin & Zao-Sanders, 2020): "Data literacy has become important, for almost everyone. Companies need more people with the ability to interpret data, to draw insights, and to ask the right questions in the first place. These are skills that anyone can develop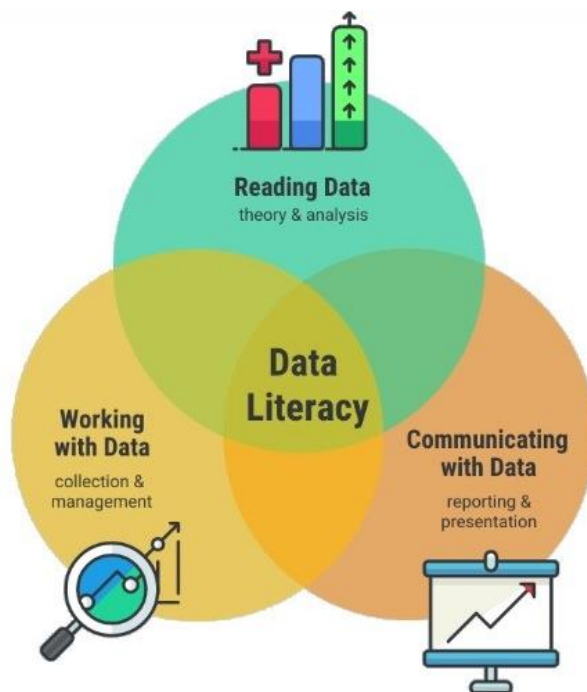, and there are now many ways for individuals to upskill themselves and for companies to support them, lift capabilities, and drive change. Indeed, the data itself is clear on this: Data-driven decision-making markedly improves business performance."

There are different levels of data literacy, but not everyone requires high levels. Some groups in tax administrations, such as tax auditors, may require achieving a high level of data literacy. A proposal for a four-level data literacy scale is presented by (Wills, 2022):

- **Data dexterity**, defined by Gartner as the ability and desire to use existing and emerging technology to drive better business outcomes (lower level)
- **Data democratization**, which makes digital information accessible to more non-technical users of information systems — without requiring IT involvement
- **Greater collaboration**, when different stakeholders (tax auditors, accountants, analysts, etc.) use a common vernacular to discuss data
- **Self-service analytics**, because understanding data is as essential as having quick access to it (highest level)

A data governance program must include a data literacy assessment and improvement plan.

The Data Literacy Project, an initiative supported by various companies and organizations[23], proposes a six-step approach to launching a data literacy initiative.

1. A strong vision and approach to planning

   A strategic plan should include what kind of goals will be achieved, how to fund, and who will lead.

1. Great communications plan

   Two core communications: the first to the whole organization, explaining "why" data literacy is essential; the second to the participants, explaining what and when the project will be happening and what they need to do.

2. An assessment program

   The participants must be assessed to their current comfort level with data literacy.

   An example of a data literacy assessment is proposed by the Data Literacy Project[24].

3. Cultural reinforcement

---

[23] See https://thedataliteracyproject.org/about
[24] See https://thedataliteracyproject.org/assessment

The objective is to evolve the organizational culture so that the language of data becomes second nature. Training everyone on what a culture of data literacy looks like within your organization and highlighting the benefits of working within a data literate environment.

4. To create a learning roadmap (personal training)

    A personal (individual) learning roadmap must be made based on the assessment process.

5. Measurement

    Continually assess the progress of the program, personally, by electronic mail surveys, or short meetings. Some perspectives:

    •   Are more data being used to present arguments and positions?

    •   Is there an increment in the use of applications that deal with large volumes of data?

    •   Are people asking better questions, fueled by data, and making more informed decisions?

This approach must be revised from time to time and adapted to the results obtained.

Wills (Wills, 2022) also proposes some steps to drive a successful data literacy project.

For data literacy, the following topics can be taken as a fundamental guide:

Table 2-3 Data Literacy Syllabus

Topic	Goal
Data analysis and visualization	Understand how to interpret and exploit data to improve decision-making, data democratization, and the search for knowledge.
Data Storytelling	Understand how to use narrative and argumentative techniques supported by data.
Data Governance	Understand the importance of controlling and supervising data through clear roles and responsibilities.
Data Quality	Understand the importance of guaranteeing the trust and credibility of the data throughout the lifecycle.
Data Architecture and Technology Architecture	Understand how to organize data and the technological resources that manage it.
Data Security and Privacy	It is vital to ensure data security and privacy as they have inherent risks.

Source: Prepared by the authors

Chapter 3   Data Governance for Tax Administrations: Strategic Perspectives

The implementation of data governance must be based on a data strategy and alignment with the business strategy of the tax administration. These are essential to obtain positive results.

As mentioned in previous chapters, the increasing dependence of tax administrations on data to fulfill their obligations (data-driven tax administrations) makes the strategy-data relationship mutually influential.

Additionally, after providing several digital services, tax administrations seek to establish a strategic framework for coherent evolution based on digital transformation. According to OECD (OECD, 2022), around 75% of the tax administrations participating in a recent survey[25] have a digital strategy in place at different stages, as shown in Figure 3-1. And a data strategy must be part of this digital strategy.
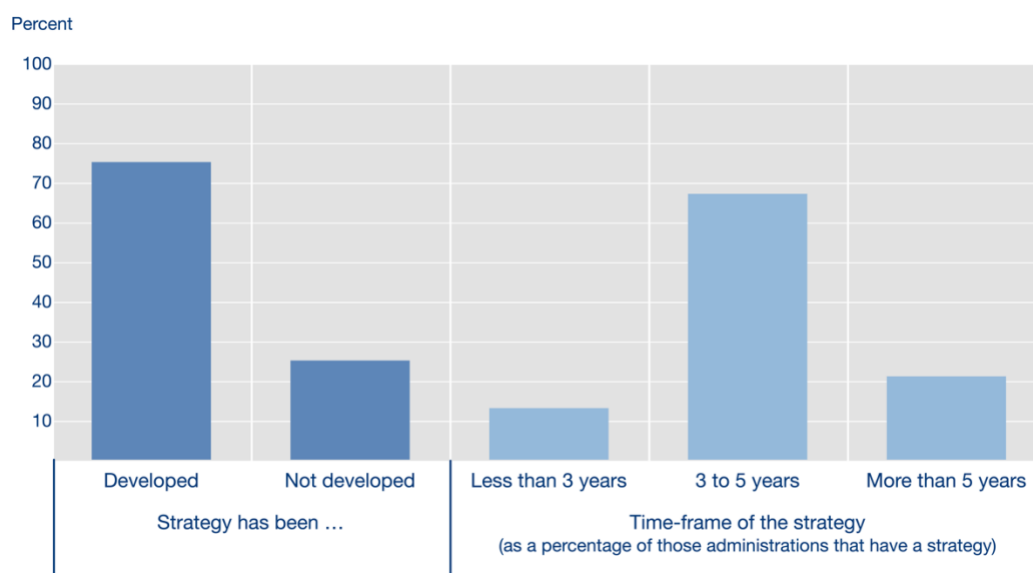
Figure 3-1 Existence of a strategy for digital transformation in tax administrations Source: (OECD, 2022). Redrawn for clarity.

The structure presented in this Chapter in the following paragraphs is based mainly on a proposal presented by (Informatica, 2020).

3.1   Data Strategy

A data strategy is a central, integrated concept that articulates how data will enable and inspire business strategy, as established by the (MIT CISR Data Research Advisory Board, 2018).

Specifically, a data strategy defines how a tax administration achieves institutional goals (business objectives) through strategically using its data assets.

---

[25] The figure is based on data from 52 jurisdictions covered in the referenced report and which completed the global survey on digitalization.

A data strategy supports the overall tax administration strategy by mapping data to business needs, like processes used to run day-to-day operations; analytics used to support decision-making; the technology architecture supporting operations and analytics; and the people and teams accountable for governing and managing data.

It is ultimately about understanding the relationships between data, processes, technology, and people so your organization can maximize its ability to generate the most significant institutional impact from data (Informatica, 2020).

According to the same source cited above, a data strategy is developed to:

- Accelerate all digital transformations.
- Improve business agility.
- Become a taxpayer-centric institution.
- Seize new opportunities.
- Cultural changes, like encourage innovations by "testing hypothesis" using data.
- Focus resources on value creation.
- Earn continued commitment from business partners.

It is essential to emphasize the importance of data to achieve the institutional goals of tax administration. Most people only understand the value of data in their activities. Thus, a table as the one presented below (Table 3-1) can help in the definition and perception of the importance of this strategy.

Likewise, the data strategy requires the support of a data management strategy (DAMA-DMBoK2, 2017); therefore, a data strategy responds to the data needs of the organization and the necessary management activities required to enable the correct provision and administration of data.

Table 3-1 Mapping business outcomes to processes, analytics, and data

Business Objective	Processes/Services	Analytics	Data
Improve tax collection	Tax collection Current tax account Failure to file and failure to pay management	Evaluate and manage discrepancies and errors in the original data Identify trends and compare them among sectors, regional, and national economic growth Identify behavioral changes for individual taxpayers …	Tax returns Tax payments Taxpayer registry Various general and sectoral statistical data from internal and external sources ….
Expand audit results	Risk management Case selection Case results evaluation	Use AI and Big Data analysis to identify and evaluate risks Use AI to select cases	Previous results Financial data gathered from third parties Social networks analysis

Business Objective	Processes/Services	Analytics	Data
		Use AI and social networks to identify possible discrepancies	External sources data Tax administration data
Effectively legal disputes resolution (litigation)	Administrative and legal process Internal judgment process	Legal jurisprudence search and recommendation Similar cases search and recommendation Use AI to identify relevant cases and decisions	Administrative and judicial cases database Trial sentences
Improve relationship with taxpayers	Set of services provided to taxpayers Taxpayer perception surveys	Analysis of taxpayer satisfaction surveys Use AI to identify the mood of taxpayers after interactions with the tax administration Identification of needs for new services	Individual information available for taxpayers Surveys data Information regarding the interactions and contacts between taxpayers and the tax administration
Provide comprehensive and quality information to society	Information generation to make available to citizens Compliance in terms of transparency	Analysis of citizens' satisfaction surveys Identification of needs for new information Statistics on processes and operations	Information available and disclosed in tax administration databases Statistical data generated from tax administration databases and processes

Source: Prepared by the authors

The MIT CISR (MIT CISR Data Research Advisory Board, 2018) proposed the following principles for the creation of a data strategy:

- The journey is as crucial as the destination.

  The action of creating a data strategy is a chance to promote data conversations, educate executives, and identify exciting new data-enabled opportunities for the institution. Creating a data strategy may generate political support, changes in mindset, and new business directions and priorities that are even more valuable than the data strategy artifact itself.

- One size may not fit all.

  Data leaders may need to adapt a data strategy for application across an organization that is large or decentralized. Government institutions also have their own peculiarities.

- Be prepared to change the tires while the car is moving.

A data strategy should support the data activities of an organization to fulfill its business strategy. A tax administration must establish ways to maintain the alignment of data and business strategies to keep the data strategy relevant over time.

In practice, achieving a data and business alignment strategy does not follow a recipe. This fact can be observed in how the alignment of data and business strategies has been achieved in companies that are part of the MIT CISR Data Board, according to a survey on data strategy maturity carried out in 2018:

- Obvious alignment: 33%
- Unclear: 16%
- Embeddedness: 15%
- Governance: 14%
- Business outcome: 12%
- Business unit ownership: 10%

3.2   Metrics to Monitor and Measure the Impact of Data Strategy

Michael Schrage, a researcher at the MIT Sloan School of Management (Schrage, 2019), says, "your KPIs[26] are your strategy; your strategy is your KPIs."

This conclusion makes KPIs central to the success of a data strategy.

One way is to show the relationship between data metrics and strategic KPIs through a hierarchy of metrics.

An example is shown in Table 3-2 below concerning a strategic KPI related to voluntary improvement in tax collection.

Table 3-2 Relationship between Strategic KPIs and Data metrics

Strategic KPI	Voluntary revenue growth		
Process metrics	Reduction of the gap in terms of registered taxpayers	Improved accuracy of pre-filled tax returns	Relevant, timely, and customized information to the taxpayers
Data metrics	Improved quality of the information from financial entities and the taxpayer registry	Expansion of information sources and improved data quality	Accuracy of contact data and other taxpayer information

Source: Prepared by the authors

---

[26] Key Performance Indicators

A comprehensive table associating the strategic KPIs with data metrics, where this is possible, is relevant to communicating this relationship to all levels of the institution and, in addition, to monitoring its development.

Subramanian (Subramanian, 2017) proposes four metrics or critical indicators to initial practical identification of the success of any data governance:

1. Improvement of data quality scores

   Quality, in simple terms, is defined as the Completeness, Accuracy, and Timeliness of the data. There can be a three-dimensional score on each dimension or a consolidated score with appropriate weighting. The key is to ensure that these are measured and monitored for improvement.

2. Adherence to data management standards and processes

   As part of the framework, a tax administration establishes Standards and Policies that need to be followed by all employees under various scenarios. For example, IT should have restricted access to production data. However, in exceptional circumstances, IT can modify data with adequate control procedures and certain approvals. There must be a certification process (either self-certification or other means) by which each department should confirm adherence to Standards and Policies.

3. Reduction in risk events

   A risk event may result from any data quality issue.

   An event could be:

   a) A penalty/fine imposed by a regulator caused by a misreporting.

   b) An inaccurate decision due to insufficient data.

   c) An erroneous refund processed for a taxpayer due to inadequate data quality.

   Once data governance has been implemented, the tax administration should see a reduction in such risk events. If the risk events continue to occur, this shows the malfunctioning of the governance function.

4. Reduction in data rectification costs

   A tax administration incurs costs to rectify bad data or enhance it to meet its needs. The core principle of data governance is to "fix at source," i.e., the erroneous data is not fixed by the consumers of such data but is set at the source where it originates. Sometimes this could be within the organizational boundaries, or it could be from an external provider. An organization should track the rectification costs to ensure that it is kept to the minimum. Costs could also be involved when due to a lack of available quality, efforts must be made to validate that data is good even when it is.

As institutional maturity in data governance progresses, other more sophisticated metrics could be established and evaluated.

3.3 Mapping Technical Capabilities to Processes and Analytics

The data architecture and technology infrastructure are fundamental to scaling the data and analytics activities. The tax administration must have the right technical capabilities to develop these areas.

The following technical capabilities may be needed and must be evaluated depending on the outcomes expected from the data strategy, as summarized and adapted from (Gallant & Fleet, 2020):

1. Data discovery and cataloging

    These capabilities involve documenting and categorizing data assets, finding new data sources, understanding its contents, and disseminating among target areas.

2. Data governance

    These capabilities have to do with defining and documenting organizational structures, policies, rules, glossaries, processes, and people required to govern data.

3. Data quality and enrichment

    These capabilities involve cleaning, standardizing, and enhancing data to ensure its suitability for use in analytical and operational activities.

4. Data integration, interoperability, and APIs

    These capabilities involve moving, combining, and syndicating data across sources, applications, processes, and, if necessary, external use.

5. Master data management

    These capabilities have to do with ensuring the quality of the core entities like taxpayers' identification, tax returns, tax payments, and the chart of accounts used in analytical and operational activities.

6. Data privacy and protection

    These capabilities involve implementing policies to enforce controls and demonstrate compliance with regulations.

7. Business intelligence and reporting

    These capabilities involve reporting what happened, analyzing why, modeling what to do, and planning execution.

8. Data science and AI

    These capabilities involve creating models of what is likely to happen and using them to improve risk management capabilities and automate decision-making and business process workflow.

9. Data warehouse and lakes

These capabilities involve consolidating and storing data for use in reporting and analytics.

Technical capabilities are added to the program as needed. In smaller institutions, an expert or group may initially accumulate several capabilities.

3.4   Mapping Organizational and Program Capabilities to the Data Strategy

An executive survey on Big Data and AI (NewVantage, 2020) with more than 70 leading private firms shows that the principal challenge of an organization to become data-driven is about people, business processes, and culture (90,9%), not technology (7,5%).

Roles, structures, and processes need to be aligned with the strategy: if not, responsibilities can be overlooked, staffing can be inappropriate, and people and even functions can battle among themselves.

The following key points are proposed by (Gallant & Fleet, 2020):

1. Start with the roles

    The roles must be outlined around defined outcomes, not around people. Sometimes a role is illustrated with a view of a particular person. Only after a role is described, based on a set of competencies that someone must have to deliver a set of defined metrics, a specialist must be sought to occupy it.

2. Recruit the right talent

    The right people to be assigned to the roles must come from internal recruitment, internal staff training, or external recruitment.

3. Go beyond organizational charts and hierarchy

    Structure dictates the relationship of roles in an organization and how people behave, and teams collaborate. It must be considered what work should be designed around a centralized, structured functional organization and what work can be distributed in a more team-oriented matrix design to balance between centralized economies of scale and decentralized flexibility and agility.

4. Design processes to facilitate collaboration

    All stakeholders must be able to weigh in on how their priorities fit into the company's larger plan. When there is a defined process for discussion and resolution, it's easier to manage the operational trade-offs by setting priorities for the long term and coordinating activities across functions.

5. Develop a communication plan

    Communication is essential in any project; in implementing a data strategy, even more so. The Chief Data Officer or equivalent must translate the data strategy vision into messages addressed to different teams and stakeholders and make them reach the recipients properly, periodically, and in time.

According to the Data Governance Institute, at the industry's first Data Governance Conference, in December of 2006 (Orlando, Florida), leaders of successful data governance programs declared that, in their experience, data governance is between 80% and 95% communications!

(Gallant & Fleet, 2020) also proposes the following questions to be answered in a communication plan:

- What are the stakeholders' attitudes and behaviors that need to change to be successful?

- What barriers prevent them from fully supporting and participating in the required work?

- What communications channels work best—face to face, email, corporate portal?

- What activities, events, or materials—to be used in your selected channels—will most effectively carry your message to the intended audiences?

- What is the time frame for first sharing the message and, how often will you reinforce your message?

3.5 Change Management

Organizational change refers to how an institution or business alters a significant component of its organization, such as its culture, the underlying technologies or infrastructure it uses to operate, or its internal processes. Organizational change management is leveraging change to bring about a successful resolution (Harvard Business School, 2020). This management model focuses on the transformation process's significant impacts to ensure its benefits are continuously superior. Organizational change management focuses on helping individuals impacted by these changes adapt and succeed.

Best practices in information systems management, for example, ITIL[27], already explore change management at an operational level.

It is precautionary to assume that changes, mainly organizational as a data governance initiative, will suffer resistance, and it is necessary to be prepared for this. By the way, it is not too much to recognize that resistance to change in complex organizations is essential and healthy.

There are many forms of resistance to change (Juneja, 2020):

- Rational versus irrational resistance
- Justified versus unjustified resistance
- Overt versus disguised resistance

The specialized literature helps to understand the forms of resistance better. This knowledge helps select ways to take care of resistances, which, in general, can be classified as follows:

---

[27] Information Technology Infrastructure Library

- Broad education and intense communication.
- Facilitate participation and involvement.
- Support and safety for employees to face fear, resentment, or conflicts of interest.
- Agreement and negotiation.
- Cooptation and manipulation.
- Coercion.

No "recipe for cake" can be used by any organization in any situation. It is essential to know the history and culture of the organization.

There are, however, many systematic techniques and methods for managing change. One of the most adopted models was proposed by John Kotter (Kotter, 2014), based on research of 100 organizations undergoing a change process. It's a general-purpose approach, organized in 8 steps, and applicable to any change, as shown in Figure 3-2.
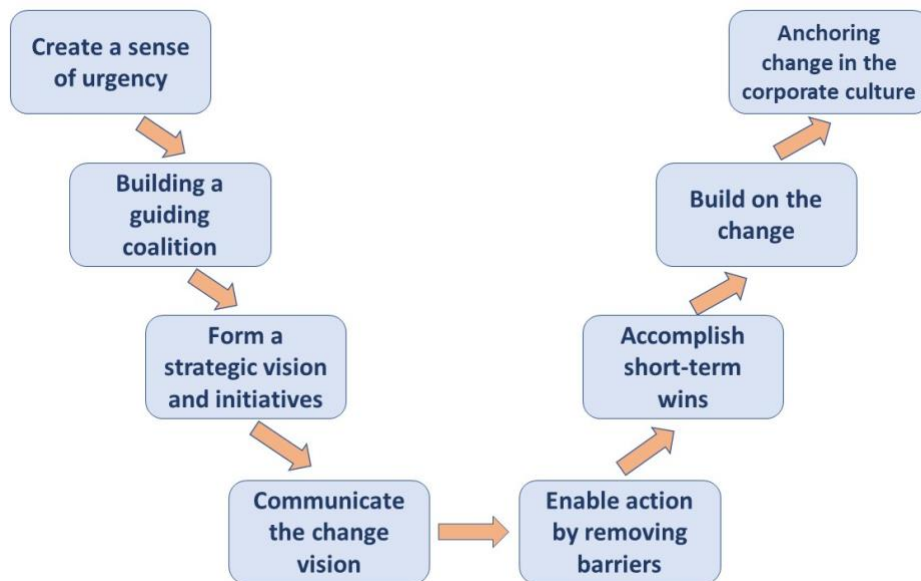
Figure 3-2 Improvement of data quality scores. Source: Adapted from (Kotter, 2014)

Step 1 – Create a sense of urgency

Organizational managers must carefully evaluate complex changes. Once the change is approved and its importance to the organization is understood, it must be carried out with high priority and ensure that people are on board. A sense of urgency can serve at least two goals. First is the rationale for the need for change. Second, it is essential to group people around the idea and takes care of opposing positions. A sense of urgency should not be used to accelerate change. Change execution must respect the time and maturation of the actions.

Step 2 – Building a guiding coalition

No one implements organizational change alone. It takes a team with change agents who do not always occupy positions in the organization's hierarchy. Change agents need motivation and, above all, leadership and sponsorship from the organization's top management. It is vital to act on

eventual weaknesses in the team and ensure the diversity and breadth of the group (various organizational units and different hierarchical levels).

Step 3 – Form a strategic vision and initiatives

Initially, the description of the change may be general and vague, including the diagnosis and coping options. Once the change is approved, you must ensure that your description is accurate and clear. The vision of change and its benefits is the basis for seducing people. Leaders and change agents must be aligned and ready to communicate the essence of change within minutes. These people must "preach" the vision.

Step 4 – Communicate the change vision

Communicate the vision consistently. Use every opportunity to communicate the idea. Address fears and anxieties related to change honestly and transparently. The change must be integrated and coherent with the day-to-day of the organization. Sponsors of change should exercise clear communication of vision and, preferably, practice "leading by example."

Step 5 – Enable action by removing barriers

The change leaders must address any obstacles and resistances. Removing barriers helps engage people and avoids the unproductive burning of energy. If required, actions to adjust norms, job descriptions, or organizational structure are an excellent opportunity to remove barriers and encourage people.

Step 6 – Accomplish short-term wins

Success is motivating. The people involved cannot be encouraged by long-term results alone. Short-term and intermediate victories need to be shared and rewarded. Working on short-term goals, in addition to long-term goals, helps anticipate the perception of possible failures and deviations. Low-cost actions that can produce good results in the short term and that do not harm change should be selected.

Step 7 – Build on the change

The effects of change must be sustainable. Effective and lasting change is profound – victory cannot be declared too soon. Each positive result reinforces what is right and indicates improvement opportunities. The kaizen principle of continuous improvement has to be permanently present. Ideas and renewed disposition can be stimulated by rotating the team of change agents.

Step 8 – Anchoring change in the corporate culture

Data governance has no end. The shift to implementing data governance only ends when its principles are solidly embedded in the organization's culture. Top management support must go beyond executing change. The values and ideals of change must be practiced daily, including in the selection, hiring, and training of people. The use of positive reinforcement tools – awards, recognition, and references to intermediate successes – is strongly recommended.

Kotter's model is step-by-step and describes the entire process of implementing the change. It emphasizes preparing and building acceptability for change and leaves the details of each specific change to experts. Skipping a single step can result in severe problems as the method is structured

in stages. As it is a sequence that generates social effects, change cannot be treated as a "project" and has a maturation that requires time, persistence, and determination.

Despite its wide acceptance, some disadvantages of this model are pointed out in the literature (Juneja, 2020), such as the high cost in time and its top-down approach.

A successful change management initiative for data governance involves planning the initiative, metrics, identification of cultural challenges or constraints, stakeholders' identification, and communication.

3.6   Final Comments

A data strategy mapped to the institutional goals, processes, and outcomes, adequately communicated to all stakeholders, supports a cultural change where everybody thinks of data as an asset.

Defining a data strategy is one of the main tasks in the evolution towards a data-driven tax administration. The leadership of the Chief Data Officer or its equivalent will create the basis for the institution's gradual improvement of data management.

Chapter 4   Data Governance for Tax Administration: Modeling Proposal

This chapter proposes a data governance model adjustable to tax administrations. The model can be a starting instrument to be evaluated and adapted to each reality.

This chapter also proposes the data governance capabilities required by tax administrations; the fundamental generic principles for applying a data governance initiative; an operating model (with Data Governance Organizational Structure); and data stewardship for the data governance in the organization.

4.1   Data governance principles and policies

Data governance principles help stakeholders to work together to achieve common goals. The following principles were developed using 8.2 Data Management Principles and Policies Definition Guide, as presented in Chapter 8, and two additional sources:

- The principles of use and management of information in tax administrations, disseminated through the goals and challenges of revenue authorities proposed by the (OECD, 2001)
- The goals and principles for data governance, suggested by the Data Governance Institute[28].

From these general principles, the following specific regulations are initially proposed for data governance in a tax administration:

Principle 1: Data as a tax administration asset

> **Statement**: Data is a resource and asset of the tax administration.
>
> **Justification**: The tax administration requires the use of data to guarantee compliance control, and design and provide tailored services.
>
> **Implication**: Guarantee the treatment and quality of data as a valuable resource in the tax administration throughout its entire lifecycle.

Principle 2: Privacy and data protection

> **Statement**: Promote taxpayer data privacy compliance by following laws and regulations.
>
> **Justification**: The data of taxpayers and the tax administration must be treated/used per what is dictated by tax, transparency, and data protection laws and regulations.
>
> **Implication**: The processes, technologies, and the tax administration, in general, must guarantee compliance with what is dictated by data protection and tax laws. Under no other circumstances, will data be used for different purposes.

---

[28] See http://datagovernance.com/goals-and-principles-for-data-governance/

Principle 3: Transparency in management

Statement: Data management must show transparency throughout the entire tax administration.

Justification: Data management activities need to be transparent to the different stakeholders.

Implication: Provide clear and precise evidence of the management activities on the data, the controls used, the treatment carried out, data definitions, models, and processes, among others.

Principle 4: Control and auditability in management

Statement: Data management (and governance) is susceptible to audit and control.

Justification: The decisions, processes, and controls related to data management must be auditable and evidence documents that support their compliance.

Implication: The processes and operating model must be formalized, controlled, and evidence of compliance.

Principle 5: Responsibility and data stewardship

Statement: To govern the data, the tax administration must define the limits of responsibilities of the actors in the management and governance of the data.

Justification: For data governance, it is essential to maintain the responsibilities and administration model clearly and precisely.

Implication: Adjustment in management processes; organizational structures suitable for managing data correctly; integration of management practices in the tax administration.

The tax administration can seek other principles to meet the specific needs.

A data policy is a documented set of guidelines for ensuring that an organization's data and information assets are managed consistently and used adequately, aligned with the established data principles. Each tax administration should focus on the data policies best suited to its context. Expanding on the concepts discussed in Chapter 2, two usual data policies are exemplified below:

Data Security Policy:

This policy is aligned with principle 2, with the following guidelines:

- Manage the security of data assets following information security guidelines.
- Tax administration will protect the data generated, transmitted, processed, and stored throughout the institution.
- Implement the necessary access controls to safeguard data assets.

Data Quality Policy:

This policy is aligned with principle 1, with the following guidelines:

- Define the appropriate dimensions of data quality for the institution that allows it to be managed and measured correctly.
- Any data quality initiative should be focused on determining the root cause of data quality problems.
- Perform periodic data quality measurements on the critical data managed by the tax administration and align with what is defined in the data strategy.
- Data quality rules must be aligned with the business rules and respond to the defined data quality dimensions.

  Note: According to the needs of the tax administration, the above policies can be adopted or adapted to their reality.

Each policy has a general scope of application to the entire tax administration; in the case of more detailed policies, these can be defined for a specific scope, for example, one or several organization processes.

Each tax administration can set up the documentation requirements for putting the data governance model into practice in a variety of ways. For instance, it can define a single document for all data and data management policies or create individual documents for each procedure.

4.2   Data Governance Capabilities

According to TOGAF® (The Open Group Architecture Forum, 2018), a capability is an organization, person, or system's ability. Capability is a management term and refers to the skills that the organization requires to guarantee a specific practice, in this case, "data governance" practice. Capabilities show "WHAT" needs to be done, while the processes say "HOW" to do it. Under this premise, capabilities allow an abstract definition of what an organization is capable. There are different advantages of a capabilities-based approach, as can be seen below (Seet, 2018):

1. It's a top-down, whole-of-organization approach. It breaks through departmental silos by shifting from a functional to a capability view.

    2. It focuses directly on what an organization needs to do to execute its strategy.

    3. It provides a map of the organization's overall capabilities to ensure nothing is missed.

    4. It directly links initiatives and projects back to capability changes and, in turn, to the organization's objectives. No more random initiatives that seemed like a good idea at the time but, in hindsight, don't align with the strategy.

    5. It cuts the wheat from the chaff by helping you determine the highest priority capabilities needed to develop and related initiatives. In doing so, it clarifies and optimizes business investment.

    6. It stops you from jumping to conclusions about solutions too early. Delaying the definition of solutions and doing it in the context of capabilities opens alternatives rather than simply incrementing existing deployed equipment, processes, and people.

    7. It provides a systematic way of identifying change initiatives. Many business planning approaches define mission, goals, and objectives and spawn initiatives and projects.

> Looking at what capabilities are required to meet your objectives clarifies your endeavors.

Capability-based modeling is a technique used by enterprise architects to align strategy and promote agile thinking to bring out all that the business does and will do. A business capability is a structured construct that contains processes, technology, people, and data needed to execute a business task; for example, strategic planning needs resources such as people (to do something) and processes or projects (to deliver value), supported by technological tools and the information and data required to perform the planning.

According to Ladley (Ladley, 2020), the data governance operating model comprises two components: the **capabilities model**, which states WHAT is happening, and the **workflow model**, which dictates HOW information flows, how parties interact within the program, and how decisions are made.

The proposal of a data governance model for the tax administrations focuses on the capabilities rather than the workflows because workflows must be integrated into each tax administration's organizational processes and structures. From this perspective, the following map of fundamental capabilities of data governance for the tax administration is proposed.
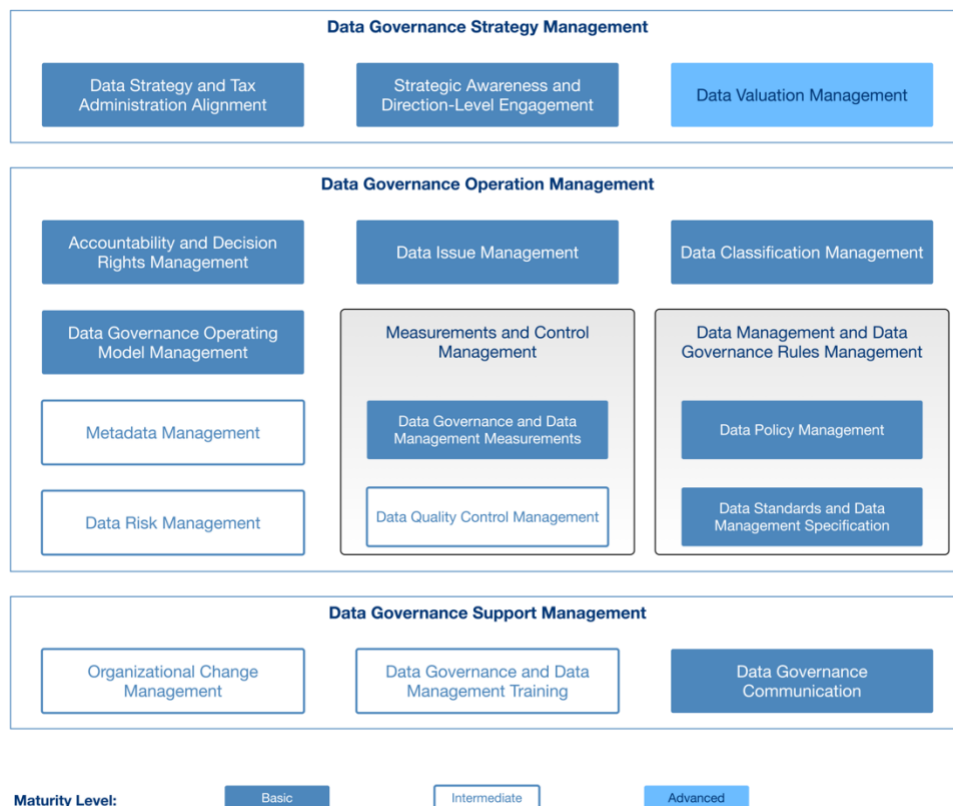

Figure 4-1 Tax administration's data governance Capability Map. Source: Prepared by the authors

The Capability Map comprises three capabilities of level 1: Strategic Management, Operation Management (value), and Support Management. Each one has a different purpose that this work will address later. Likewise, each level 1 has second-level capabilities; depending on the complexity

and relevance, some level 2 capabilities can have a third level. The map proposed in Figure 4-1 has three level 1 capability, 14 level 2, and four level 3.

The primary purpose of grouping capabilities is to maintain an order of the proposed abilities that the tax administration must have to ensure the practice of data governance.

The map of data governance capabilities (Figure 4-1) proposes a capability maturity guide that shows what capabilities to cover as the governance practice matures in the organization. For example, when the tax administration is initially adopting data governance, it is essential to focus on basic capabilities (Basic level boxes). In contrast, if a tax administration has already covered specific basic capabilities, it can concentrate on intermediate capabilities (Intermediate level boxes) and, finally, advanced capabilities (Advanced level box). However, despite being an evolutionary suggestion of capabilities, the tax administration can develop them according to its needs.

4.2.1   Data Governance Strategy Management

This capability covers the ability to align, plan and provide strategic data governance vision within the tax administration vision.

Table 4-1 Tax administration's data governance Strategy Management Capabilities

Sub Capability	Description
Data Strategy and tax administration alignment	Ability to define, monitor, and support the data governance strategy and data strategy and align them with the tax administration strategy.
Strategic awareness and C-level engagement	Ability to define, implement and socialize strategies that allow the commitment, support, and promotion of the awareness of the C level of the tax administration that investments in data governance programs serve to help their business strategy succeed in the long run.
Data Valuation	Ability to provide methods and calculate the value of the tax administration's collected, stored, analyzed, and marketed data.

Source: Prepared by the authors

4.2.2   Data Governance Operation Management

This capability covers managing the operation model and core data governance function.

Table 4-2 Tax administration's data governance Operation Management Capabilities

Sub Capability	Description
Accountability and decision rights management	Ability to create, manage and assign decision rights over data governance functions and processes activities. Also, the roles and responsibilities that the data governance model needs to govern data.
Data governance Operating Model Management	Ability to create and maintain data governance within the organization structure and processes to support the data governance capabilities and data governed controls and supervision
Metadata Management	Ability to collect, control, provide and use information about data to be managed.
Data Issue Management	Ability to centralize, categorize, prioritize, and resolve tax-related issues and requirements.

Sub Capability	Description
Data management and data governance Rules Management	Ability to create, maintain, publish, and socialize principles, policies, standards, and specifications related to data management and data governance. 3rd level: **Data policy management**: Ability to create, maintain, publish and socialize business principles and policies related to data management. **Data standards and data management specification**: Ability to design, develop, maintain, socialize, and promote documented agreements on the representation, format, definition, structuring, labeling, transmission, manipulation, use, and management of data.
Data Risk Management	Ability to mitigate, treat or eliminate risks when data is created, stored, transformed, used, and destroyed (e.g., poor data quality, Data security breaches, etc.).
Data Classification Management	Ability to identify, define, organize, and catalog data according to criteria specified by the tax administration, e.g., Data Domains.
Measurements and Control Management	Ability to identify, define, establish, and control data management and data governance measurements. 3rd level: **Data governance and data management measurements**: Ability to identify, define, establish, and monitor management indicators related to data governance and data management. It can be coordinated and integrated into centralized management indicators of the tax administration. **Data Quality Control Management**: Ability to establish and ensure practices and control measures of data quality requirements of the tax administration.

Source: Prepared by the authors

4.2.3   Data governance support management

The ability to support the operation and strategic data governance functions

Table 4-3 Tax administration's data governance Support Management Capability

Sub Capability	Description
Organizational Change Management	Ability to implement strategies for effecting change, controlling change, and helping people to adapt to change when tax administration adopts data governance capabilities. The organization must coordinate this capability with similar functions already in place (e.g., change management).
Data governance and data management training	To enhance awareness and guarantee that data is managed correctly, it is possible to train individuals in data management techniques. Data literacy programs can also help tax administrations become more data-driven.

| Data governance communication | Ability to create, socialize and maintain communication of data management activities and practices towards the tax administration's stakeholders |

Source: Prepared by the authors

Tax administration can operationalize the capabilities according to its reality through people, processes, technologies, and information needed to execute each capability. For example, if we take one of the capabilities, Data Issue Management, it can be operationalized with the following process flow, presented in Figure 4-2:
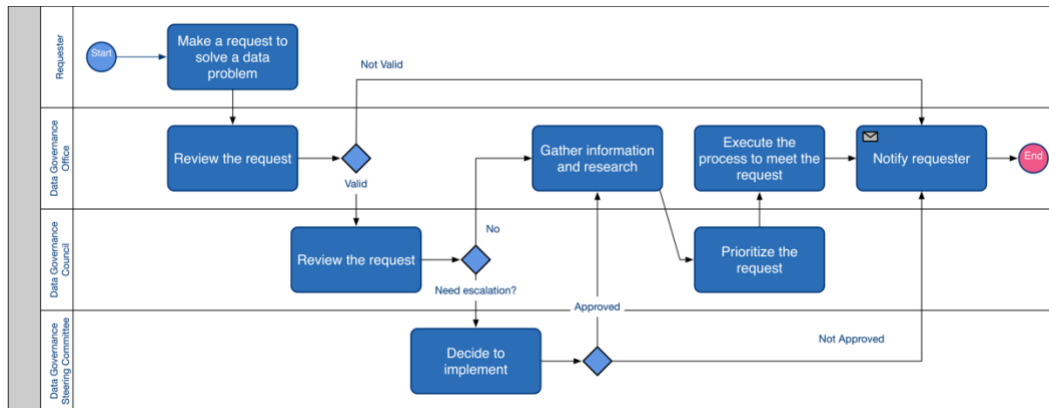

Figure 4-2 Tax administration's Data Issue Management Process. Source: Prepared by the authors

4.3   Data Governance Organization

It is essential to assess how data governance fits into a tax administration organizational architecture to articulate and assign responsibilities and operations.

Kidd (Kidd, 2010) proposes the following primary considerations for a tax administration design, no matter the size of the organization:

- Fundamental principles for tax administration organization: (1) should be function-based, (2) should integrate the delivery of all tax types, and (3) should address the specific needs of taxpayer groups by segmenting the taxpayer population, for example, into large, mid-size companies, small companies and self-employed, employees, etc. Tax administration can adapt all these principles for small and micro-economies.
- Integration of the administration of all taxes is possible, no matter what the size of the tax administration. The challenges created by small sizes make integration even more critical to ensure efficiencies.
- Function-based and segmentation remain solid principles to be followed and can be readily adapted to the specific needs of small and micro administrations.

Nonetheless, there are currently variations around two models in force for the organizational structures of tax administrations (OECD, 2008).

The "functional" model, where the staff is organized principally by functional groupings (e.g., registration, accounting, tax returns processing, audit, arrears collection, objections, and appeals, etc.) and generally work across different taxes.

The "taxpayer segment" model, where the service and enforcement functions principally around segments of taxpayers (e.g., large businesses, small/medium businesses, individuals, etc.).

And, of course, a "mixed" model, made up of the two models.

Considering that whatever the model, a tax administration should internally share a good part of the available data, there is a need for solid data policies in the institution and coordination between Data Owners and Data Stewards in multiple departments, possibly with the support of intermediary Data Governance Councils.

To illustrate how the data office, committee, councils, and stewards may spread across the organization, we present a simplified tax administration organizational model and discuss how committees and roles can apply across the organization, depending on the data stewardship approaches adopted (functional or data domain), as described below. Remember that this is a theoretical exercise because, in addition to not having a unique way for this, it depends, among other things, on the size, maturity, organizational peculiarities, and culture.

For the construction of the data governance organizational structure and stewardship model, it is essential to define the approach to be adopted: (a) stewardship based on functions of the tax administration ("Functional Model"); or based on data domains ("Data Domain/Subject Area Model") (Plotkin, 2020); and (b) if the data governance organization model is centrally organized or distributed (DAMA-DMBoK2, 2017).

The data governance organization model largely depends on maturity and choices of the tax administration. For this document, both stewardship and structural organization approaches are proposals to be evaluated.

Figure 4-3 Data Governance Decentralized Organizational Structure (with Functional Stewardship Approach) in tax administration. Source: Prepared by the authors

In the first approach, represented in Figure 4-3, the members that participate in the Data Governance Council are Data Stewards that describe the functions of the tax administration

We can observe the focus on large taxpayers, with other taxpayers controlled mainly by central and regional services. A possible fit of the data governance structure into the organizational structure is shown above.

The main features of this proposal are:

- In more extensive or mature tax administrations, there may be a need to create a level of coordination between the Data Governance Councils and the Data Governance Office. Otherwise, this coordination is made directly in the Data Governance Office.
- Currently, the collection processes use the operational support of financial institutions, and the control is made centrally. Arrangements for exchanging data with financial entities are carried out through contracts, monitored at an operational level (compliance with contractual clauses). Instead of an internal council, the Data Owners can dialogue directly with the Data Governance Office on data governance issues.
- In tax administrations, multiple stakeholders are concerned with the same data set. It is important to designate one individual who will assume the Data Owner role, and then they may consult and collaborate with other stakeholders as closely as necessary.
- The centralized IT area in central services is the primary location of the Technical Data Stewards, although similar profiles may be inserted in other regions in larger organizations.
- More and more, compliance has become very important for tax administrations. The central compliance area often needs agents (Compliance Data Stewards) in the institution's business areas.
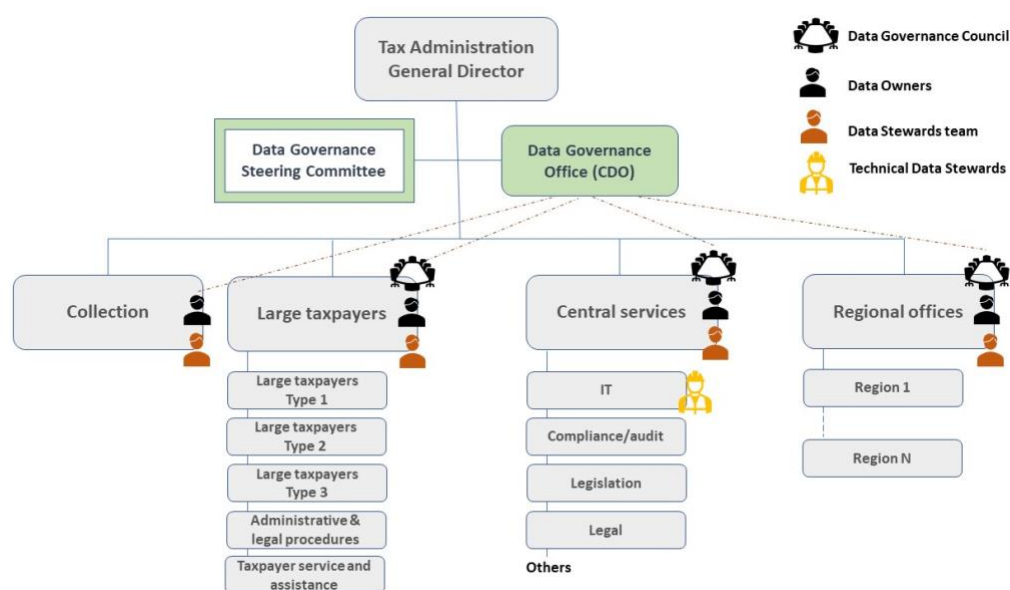

Figure 4-4 Data Governance Centralized Organizational Structure (with Functional Stewardship Approach) in tax administration. Source: Prepared by the authors

In the same "Functional Model" approach, a centralized data governance organizational structure is proposed below for tax administrations interested in centrally controlling data management. This approach is also used in the case of data governance models that initially adopt the data governance practice.

Figure 4-4 is a proposal for an operating model of centralized data governance in the tax administration; it shows three government bodies: Data Governance Steering Committee, Data Governance Council, and Data Governance Office, where they all comply with a centralized approach.

The Data Governance Council is vital in centralizing tactical efforts and bringing together the different data owners (optional) and data stewards to manage data governance needs and requirements. This approach may require the participation of a coordinating Data Steward (optional) to participate in the Council on behalf of functional data stewards.

The Council informs and escalates needs to the Data Governance Steering Committee if required. Likewise, the Data Governance Office coordinates the operational executions and participates in the Data Governance Steering Committee and Data Governance Council sessions through the Chief Data Officer (or Data Governance Manager).
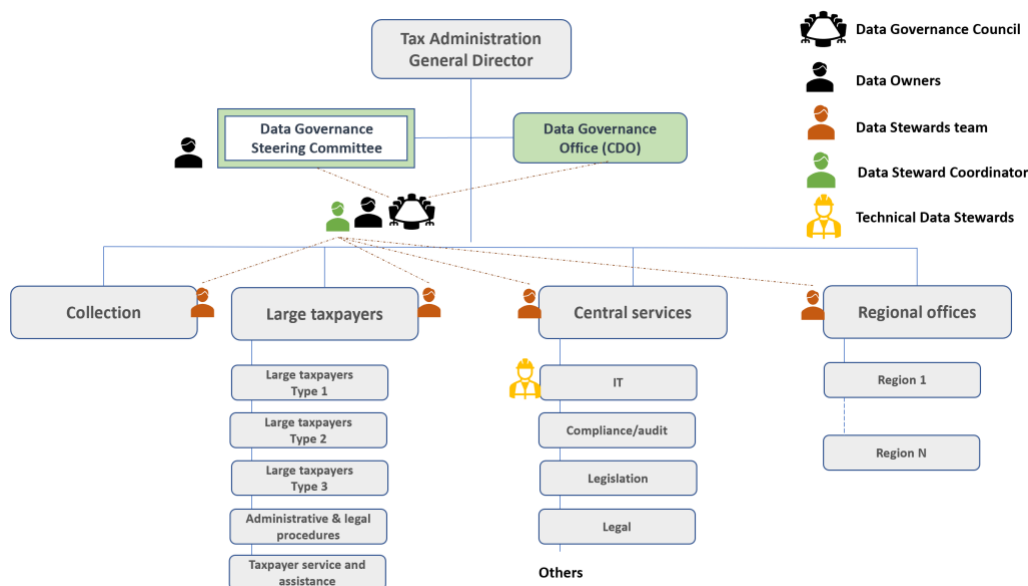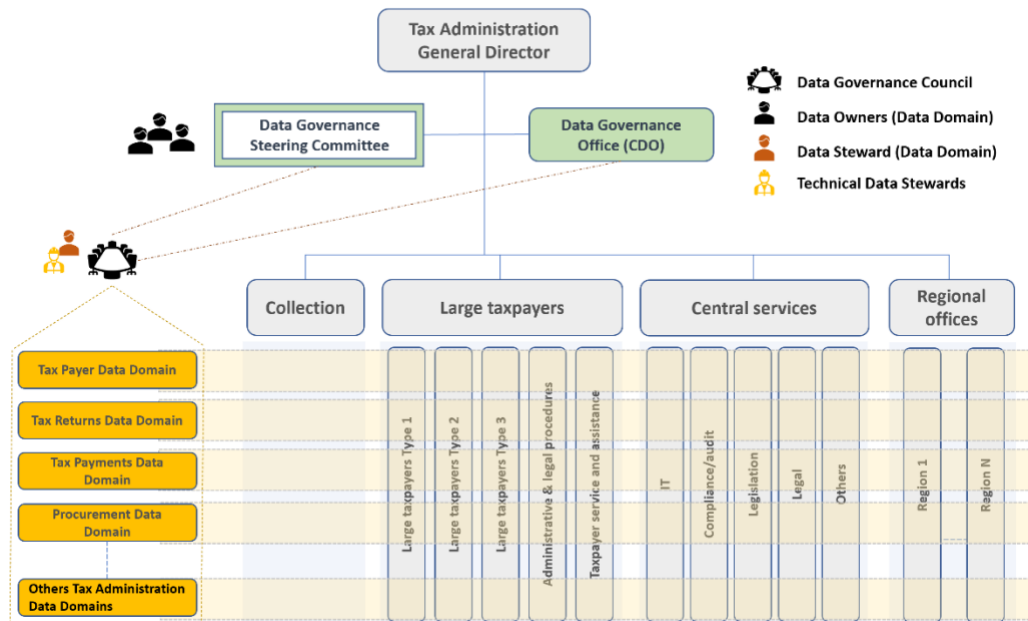

Figure 4-5 Data Governance Centralized Organizational Structure (with Data Domain Stewardship Approach) in tax administration, Source: Prepared by the authors

Figure 4-5 shows a centralized data governance organizational structure with an operation similar to that shown in Figure 4-4 with the difference in the stewardship approach in which data management prevails according to the data domains of the tax administration, that is, representatives (Data Stewards) of the data domains participate in the data governance council. For example, the Data Steward of the Taxpayer Data Domain and the other Data Stewards. The Technical Data Stewards, the IT support for the data domain of interest, can participate in the council if the organization or the data stewards require it.

4.3.1   Basic Data Governance Implementation and Evolution in Small Economies

Empirical works show that three variables are usually associated with the size of an economy: the size of the population, the size of the land area, and the country's GDP in question (Briguglio, 1994).

These characteristics affect the operational reality of tax administrations; for instance, in smaller economies, the workforce is usually small; the domestic labor market does not have the educated and experienced professionals needed; the budget is small; technical capacity is relatively low (a consequence of a small labor market); few training opportunities; major political influences[29]; and difficulties in obtaining and retaining legal counsel (Kidd, 2010)[30].

Other aspects of tax administrations in small economies that could influence data governance strategies include the presence of foreign-owned and managed businesses, who frequently use automated enterprise resource planning systems (ERPs) or at least robust information systems to support their operations and tax-related matters that may not be local, even though the number of large taxpayers may be small but have greater significance in terms of total revenue collection. Tax administrations of small economies may also have to deal with the administration of permits and local fees, which might reach a large part of the population and small businesses, requiring the participation of a large part of the available employees that would be focused on routinary activities only.

Although the principles of data governance remain the same, some aspects of its implementation are impacted by the context mentioned above and must be observed and considered:

- The governance structure must be more compact. One person assumes multiple roles.
- One data council may be enough to discuss data governance strategies and policies.
- A lesser number of data domains might be prioritized.
- A comprehensive data literacy program is essential to the data governance initiative.
- Data quality, security, and compliance are always important, but opportunities for improvement in those aspects might be significant.
- Automatization is critical.
- In-house development of IT solutions might be complex, and the solution may be to establish a low-cost maintenance open-source or software as a product environment.

Implementing a data governance system in smaller tax administrations might rely on technical and financial support from international agencies in all aspects, including identifying available solutions, constructing capabilities, and providing resources for the implementation and the future maintenance of the initiative.

According to what was mentioned above, a minimum viable data governance initiative (DG-MVI) for a tax administration is a basic version of a data governance model. It could be released as soon as possible for early adopters that would rapidly evolve but would also respond to tax administrations as a way to go in countries with small economies. The design of the DG-MVI focuses on having a simple model with a data domain approach that can be gradually expanded as the institution matures in data governance.

The DG-MVI, when adopted, must be the result of a tax administration data strategy, as presented in Chapter 3.

---

[29] Given the small population base, civil society is relatively small, and many government officials and Private Sector business people are well known to each other.
[30] Some of these effects can be alleviated through international technical cooperation initiatives.

Like any data governance model, this DG-MVI must be based on principles, policies, roles, committees, processes, and indicators conveniently inserted into the organizational structure of the tax administration. At a minimum, it will consist of a centralized approach, as shown in Figure 4-5:

- A basic Data Governance Steering Committee.
- A central Data Governance Council.
- A basic Data Governance Office.
- A basic assessment method with indicators.
- Identification of the data domains to be initially governed ("Data Domain Model Approach").
- Data Owners for the designated domains.
- Data Stewards (at least one technical Data Steward for each data domain).
- Focus on data quality and compliance policies and processes.
- Evaluate which data security policies and mechanisms are in place and define an evolutionary strategic plan.
- Identification of which priority data governance main processes should be adjusted or developed.

4.4   Organizational Structure Roles and Responsibilities

Depending on the needs of the tax administration and the starting point of the data governance model, it can evolutionarily cover its maturity aligned along the proposed Data Governance Capability Map.

As we can see from the different approaches proposed in the structure and stewardship, the following fundamental examples of the structures for data governance are presented, based on a centralized organizational structure and through a data domain stewardship approach.

Table 4-4 Tax administration's data governance organizational structure roles and responsibilities proposal

Organizational Structure	Description	Responsibilities
Data Governance Steering Committee	Body with the highest authority for data governance in the tax administration Composed by: Executive Managers (C-level executives responsible for data-intensive processes, plus the CIO) minimum one and up to three. Chief Data Officer or Data Governance Manager (C-level executive responsible for the data governance in	Define and approve the budget for data management and data governance projects and activities and monitor its execution. Prioritize strategic decisions related to data management and data governance to address activities and resolutions relevant to the tax administration. Inform the tax administration's steering committee or senior managers about data governance activities and data management and data governance projects. Define the strategy, principles, policies, procedures, objectives, and goals of data governance in the tax administration.

Organizational Structure	Description	Responsibilities
	case a dedicated official has not been designated). Data Owners (in a ratio to the number of data domains).	To be the highest body for resolving data-related problems in the tax administration. Review and/or approve the data strategy aligned with the tax administration strategy. Collaborate and coordinate with other high-level bodies within the tax administration.
Data Governance Council	Body responsible for data management and data governance activities, data problems or incidents, and their escalation. Composed by: Data Governance Manager. All Data Stewards responsible for the data. All Technical Data Stewards, All Data Architects. The council has a tactical and frontline approach to data management.	Collaborate with different stakeholders on definitions and handling of data issues. Solve first-level problems or conflict resolutions related to data and its management. Identify potential improvements to manage data throughout the entire tax administration data lifecycle. Collaborate with other interested parties in the direction of definitions and data problems. Collaborate with the Data Governance Office to guarantee the satisfactory execution of data management and data governance policies in the tax administration. Ensure that data governance efforts are aligned with the defined data strategy and the objectives of the tax administration.
Data Governance Office	Body responsible for leading the tax administration's definitions, control, and data management standards, which promotes documentation, communication, and compliance with data policies. In small tax administrations, an independent organizational unit might not be possible, and the function should be shared. It should not be within the IT structure.	Document, support, publish and lead the activities and resolutions of the Data Governance Steering Committee and Data Governance Council. Promote good data management and data governance practices throughout the tax administration. Manage and document risks and issues related to data. Document, publish and maintain policies, procedures, and standards related to data governance. Enforce policies and procedures related to data management and data governance, and scale when necessary.

Source: Prepared by the authors

Depending on the tax administration's needs and availability of human resources, the Data Governance Organizational Structure can have Data Working Groups or Data Forum, that support operational data governance activities to overall bodies.

About the meeting sessions, both the Data Governance Steering Committee and Data Governance Council, which have a collegiate body approach, can hold sessions with specific recurrence or on demand. It is recommended that the sessions be held periodically and scheduled according to the workload of the participants and the session policies that are counted in the tax administration. For example, the Steering Committee meets monthly or quarterly, while the Council meets every 15 days or monthly.

It is recommended that Data Governance Steering Committee and Data Governance Council have designated the roles of president and secretary so that they can manage both collegiate bodies. The president mainly chairs the sessions and signs the resolutions of the collegiate body. At the same time, the secretary primarily handles documentation, including a repository of resolutions and session minutes and organizational aspects such as scheduling non-programmed sessions and their agendas. The responsibilities of such roles may vary according to the rules of the tax administration. For example, in the case proposed in Figure 4-5, the CDO of the Data Governance Office can hold the role of president in the Data Governance Council or appear as secretary of the Data Governance Steering Committee.

4.5   Light Data Governance Model

When the first steps in data governance are taken by a tax administration, it might not be suitable or convenient to change the organizational structure. In some cases that might require an amendment to a law with the list of functions and responsibilities of the unit, plus the identification of job titles, job description, profiles, and budgetary provisions. Although, when possible, a formal implementation will facilitate the assimilation of data government within the culture of the tax administration it might be a lengthy process that could impose significant delays.

A light data governance model approach could be beneficial where responsibilities, functions and powers could be distributed within current units or established collegiate bodies. An example of such a distribution is listed below:

- Control and compliance functions to the internal control unit.
- Technology-related definitions to the corporate IT unit.
- Data quality and metadata definition to the data governance council.
- Data ecosystem and data modeling review to the enterprise or IT architecture units
- Development of data strategy to the highest strategy committee within the organization
- Monitoring of data governance programs to a group of delegates to the data governance council
- Promote the culture of data management, good practices, and integration with other management models to a group of delegates to the data governance council.

Data management and data government responsibilities should be assigned taking two decisive factors into account: the hierarchy level and scope, to gain sufficient decision-making power, and the existent knowledge (security, compliance, issue management, etc).

As it is implied before, a data governance council is still required, even with a minimal composition of delegates, but that council could be initially established as a project team. In that case some typical responsibilities of the council could be delegated to ad-hoc groups or even user forums.

The implementation of this light data governance model could benefit of both project management and change management practices. In the process, the tax administration could seize the opportunity to train managers and reduce the capabilities gap and improve data literacy in the whole organization.

This light data governance model should not be perpetuated, and should be treated as an interim model, being gradually oriented towards a more structured data governance management model, as proposed in this document. The formalization of a Data Governance Office is suggested as a first step towards the gradual evolution from this light data governance model.

4.6   Data Stewardship

Data Stewardship mainly includes Data Stewards from the functional area's side and Technical Data Stewards from the IT side. Data stewardship handles data responsibly, consistently, and reliable (Bhansali, 2014).

The Roles of Data Stewards are recommended to respond according to the tax administration's proposed Data Domain Stewardship Model approach. A proposal of responsibilities on which the tax administration could be based are:

Table 4-5 Tax administration's data stewardship roles and responsibilities proposal

Role	Role Description	Responsibilities
Data Governance Officer	This role leads the Data Governance Office. It is responsible for promoting the tax administration's data governance model and good data management practices.	Design the data strategy and submit it to the consideration of the Data Governance Steering Committee for its approval and adoption in the tax administration.
		Define and monitor data governance programs in the tax administration.
		Appoint the members of the Data Governance Steering Committee and the Data Governance Council.
		Lead and coordinate the decisions and resolutions taken by both collegiate bodies.
		Guarantee the timely involvement of different interested parties and support areas in the sessions of the collegiate data governance bodies (e.g., Information Security, IT, Audit, etc.).
		Promote the timely and consensual identification of the information/data needs of the tax administration.
		Promote the continuous improvement of the data governance model in the tax administration (e.g., policies, capacities, processes, etc.)
		Promote the integration of the data governance model with existing management models in the tax administration (e.g., Project Management, Risk Management, Security Management, etc.)

Role	Role Description	Responsibilities
		Promote the development and communication of data governance products throughout the tax administration (e.g., policies, processes, etc.)
		Evangelize good data management practices throughout the institution and in front of other institutions.
Data Owner	Is the leader of a data domain, and its accountable for data governance-related issues within it. This type of role must be assigned at a managerial level and should be knowledgeable about the processes and operations of the organization.	Approve definitions of attributes/data elements within their data domain (e.g., acquisition or utilization of new/existing data). Approve definitions of data quality dimensions and acceptable thresholds. Be accountable for data definitions and quality within their data domain. Approve and lead the necessary data changes within the data domain. Approve business specifications and data definitions related to their data domain. Approve definitions and data remediation and correction actions. Authorize the access and/or sending of the data in accordance with the data security and privacy policies within the institution as well as with other institutions. Responsible for the data that is shared with other institutions. Participate as a full member of the Data Governance Council and on demand according to the needs of the Data Governance Steering Committee.
Data Steward or Functional Data Steward	This type of role is recommended to be assigned at the coordination level and to be knowledgeable about the processes and operation of the functional area, function, or domain.	Execute or coordinate the execution of action plans for the remediation of data quality problems. Coordinate the efforts to identify and assess the root cause of data quality issues. Support the Data Owner in the definitions related to the data in their domain, for example, definitions of authorized sources, dimensions, or data quality rules, etc. Support definitions for data classifications within their domain, e.g., security or data protection classifications. Help define meanings and concepts related to data within their domain.
Data Custodian or	This role is generally located in the IT areas. They are the most knowledgeable people	Support functional data stewards with information and experience in information systems/applications, ETL[31], database, data warehouses, BI, etc.

[31] Extract, transform, and load

Role	Role Description	Responsibilities
Technical Data Stewards	in charge of specific data sets in the information systems and data stores; they are responsible for the technology and know-how of how the data is created, manipulated, and stored in the systems.	Support or execute data quality remediations on systems and data sources. Note: Depending on the IT human resource availability and the knowledge of the information systems, a Technical Data Steward may be assigned to one or more Data Domains.

Source: Prepared by the authors

The proposed roles and responsibilities are essential for the practice of data governance in the tax administration. Depending on the resources and maturity of the practice, the focus, scope, and attributions of each can be changed, or more roles can be included in the data governance operating model. Chapter 2 shows other examples of roles that can be adapted or adopted as needed, particularly in more extensive tax administrations.

Depending on the reality/need, institutions can extend the scope of action of the Data Governance Officer (e.g., strategy, government, control, privacy, quality), and add the exploitation of data assets to create value for the institution. In this case, in which the management of the entire data lifecycle is covered, it is under the responsibility of the position known as Chief Data Officer (CDO).

4.7   Data Quality Dimensions

Data quality dimensions allow the tax administrations to have reference concepts to improve data quality. Low-quality data generates direct risks to the organization ending in a general loss of trust in the information systems and automated processes. The quality dimensions allow for measuring the data quality against a standardized scale.

Dimensions allow data administrators to monitor their quality through minimum tolerance thresholds. The selection of the data quality dimensions must respond to the characteristics that best represent the current situation of the tax administration.

Data quality dimensions are characteristics that differentiate a data item. Since a data element can be characterized in various ways, there may not necessarily be a fixed set of data quality dimensions. This dimension may vary and largely depend on the requirements of the tax administration in terms of the contexts in which the data is used and how it contributes to the tax administration's needs.

However, tax administration could build a fundamental data quality dimension set for its context upon those defined in the DAMA-DMBoK2. For any data domain, a relevant set of dimensions can be defined as a subset from the following list:

Table 4-6 Common Data Quality Dimensions

Dimension of Quality	Description
Accuracy	Accuracy refers to how well data represents 'real-life' entities. For example, if a taxpayer address as recorded in a database is the effective taxpayer address.

Dimension of Quality	Description
Completeness	Completeness refers to whether all required data is present. Completeness can be measured in the data set, record, or column. For example, if all tax obligations of taxpayers are recognized.
Consistency	Consistency can refer to ensuring that data is logically in sync with other data. For example, a closed business or dead person should not file tax returns.
Integrity	Integrity refers to the relationships among data in terms of connections. It can be understood as referential integrity. For example, when a taxpayer identifies a legal representative, that individual should be a registered taxpayer.
Reasonability	Reasonability asks whether a data pattern meets expectations.
Timeliness	Timeliness refers to several characteristics of data. Timeliness measures need to be understood in terms of expected volatility – how frequently data is likely to change and for what reasons. For example, data related to electronic invoices must arrive at the tax administration in real-time. A tax return must arrive on the due date of the corresponding period, usually yearly or monthly.
Uniqueness	State that no entity exists more than once within the data set. Asserting the uniqueness of the entities within the data set implies that a critical value relates to each unique entity, and only that specific entity, within the data set. Measure uniqueness by testing against a vital structure.
Validity	Refers to whether data values are consistent with a defined domain of values. A domain of values may be a limited set of valid values, a range of values, or values that can be determined via rules. The data type, format, and precision of expected values must be accounted for in defining the domain. The data may also be valid only for a specific time. For example, all dates must be represented in the same way.

Source: DAMA-DMBoK2 with examples from the authors

Of all the characteristics of data, "quality" is paramount.

It's impossible to keep a one hundred percent data accuracy. Then the goal must be to improve quality to an acceptable level, according to tax administration data quality threshold acceptance.

Throughout this document, the topic "data quality" will be presented from several perspectives.

The quality dimensions presented by DAMA-DMBoK2 (above) may be illustrated and expanded from (Qureshi, 2022), with practical issues and examples encountered in data quality management mapped to DAMA-DMBoK2 Common Data Quality Dimensions:

Table 4-7 Matrix common data quality dimensions (DAMA-DMBoK2, 2017) and data quality management examples (Qureshi, 2022)

Dimension of Quality	Data Issue	Examples
Accuracy	Unclear data definitions	Unclear data definitions lead to different interpretations through the institution. Align descriptions contained in the data glossary minimize this problem.

Dimension of Quality	Data Issue	Examples
Completeness	Incomplete data	Key columns are missing information, failing ETL jobs, or causing downstream analytics impacts. The load programs should notify these flaws and can be remedied or alleviated by understanding its pattern and how default values are used. Other data sources, if available, may be used to adjust data.
Consistency	Cross-system inconsistencies	Multiple legacy systems are the primary sources of this flaw. Occur primarily, as examples, in the taxpayer's full name, address, and date of birth. All different information must be matched in a single record, for example, by direct analysis or using a fuzzy-based algorithm.
Consistency	Orphaned data	Consistency relates to data inconsistency problems when data exists in one system and not the other. For example, a taxpayer exists in the taxpayer registry but has no current account. A data quality rule that checks for consistency when data is ingested in tables will help. The data stewards must check this inconsistency.
Integrity	Inconsistent keys	This problem concerns the growth of data warehouses when new elements are added, and specific keys may lose their uniqueness. Lack of integrity can lead to problems in the data model, with loss of referential integrity. To avoid this problem, data must be profiled to ensure the key on which the surrogate key is dependent is always unique.
Reasonability	Drastic data quantity changes	Loading data have a pattern for each source, depending on the day, hour, etc. If these patterns change drastically, they may denote problems and must be evaluated by the data stewards.
Timeliness	Old and stale data	Keeping some data beyond a certain period can be useless. Thus, data must have its useful life defined and then be erased or stored in secondary memory. Also, personal data protection laws require that specific data be removed after a specified time or on demand.
Timeliness	Data received too late	Delays must be identified and reported to the respective providers, or the internal processes that generate them must be reviewed.
Uniqueness	Duplicate data	Relatively easy to detect but difficult to fix, it may cause problems with all further processing. A uniqueness check must detect and purge one record, notifying this event to the data stewards.
Uniqueness	Redundant data	Across the organization, multiple sources may provide the same data. The control must be established to avoid this redundant data dispersed through different databases.

Dimension of Quality	Data Issue	Examples
Validity	Default values	It is a problem mainly if documentation is missing. To fix this problem, data must be profiled to understand the pattern of why the default values were used.
Validity	Data format inconsistencies	Predominant in string columns (e.g., names in different cases, electronic mail addresses) and data coming from uncontrolled (external) sources. Data coming through tax administration's provided programs suffer less of this problem.
Validity	Irrelevant data	Capturing irrelevant data occupies storage space and management time unnecessarily. Each data should have an end goal.

Source: Prepared by the authors

This list is not exhaustive but can guide an initial data quality strategy. Each tax administration must assess its biggest data quality problems and set a priority to tackle them.

If the tax administrations do not adopt all of the quality dimensions proposed above, the Data quality dimensions definition guide in Chapter 8 can guide the selection of one or several quality dimensions.

Chapter 5   Data Governance for Tax Administration: Maturity Assessment

5.1   Maturity Models

Maturity models are instruments to systematize the diagnosis of an organization's capacity and stage in carrying out a particular activity. There are maturity models applicable to various IT activities – software development, information security, service management, data governance, etc. Maturity models are powerful tools to support strategic actions and the planning of new moves.

The DAMA-DMBoK2 (DAMA-DMBoK2, 2017) specifies that "maturity models are defined in terms of a progression through levels that describe process characteristics. When an organization understands process characteristics, it can evaluate its level of maturity and put in place a plan to improve its capabilities."

Data governance is not a project that ends after implementation cycles. The implementation is a project, while data governance is a permanent and lasting activity perfectible over time. Hence the importance of using a maturity model with a long-term view. OvalEdge, a company specialized in data governance, addresses the pursuit of maturity in the following text: "A data governance maturity model is a tool and methodology used to measure your organization's data governance initiatives and communicate them simply to your entire organization. In a mature organization, all the processes to manage, access, and innovate using data assets are in place. Less progressive organizations can use the maturity model to achieve this objective" (Varshney, 2021).

Data governance maturity models usually have two parts. One specifies maturity levels – usually 4, 5, or 6 (Baltassis, Coulin, Gourévitch, Khendek, & Quarta, 2020) – and the main characteristics of each. It is common for levels to be treated according to themes, domains, critical dimensions (people, processes, technology, etc.), or objectives (data integration in the portfolio, data quality, metadata, etc.). The other part is the assessment instrument, usually presented as a scorecard. Few maturity models offer both parts.

Figure 5-1 presents a diagram with the insertion of the maturity model and the maturity assessment. The data asset technical and operational management operations are part of the maturity model.

With maturity models, the Data Governance description can use language that does not require IT expertise. The use of assessment tools allows for a simplified explanation of the evolution and progress of the work, in addition to facilitating benchmarking with other organizations.
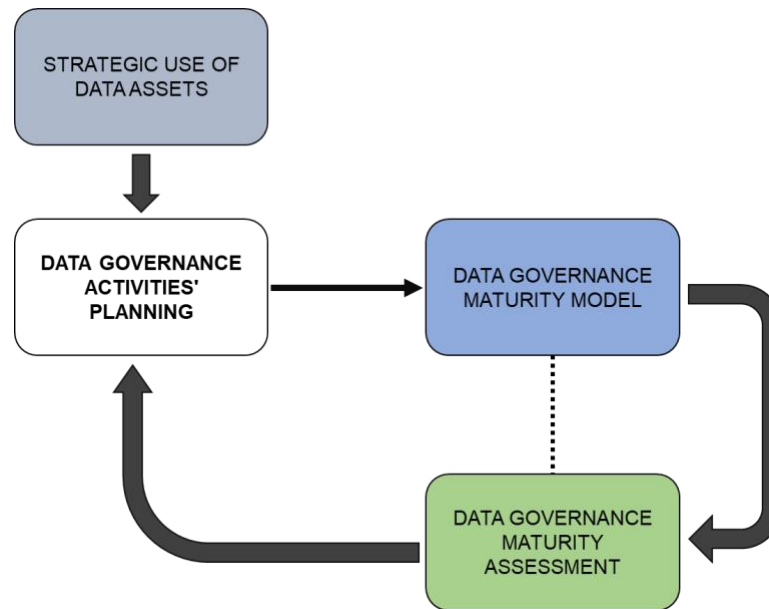
Figure 5-1 Maturity model and the maturity assessment. Source: Author's elaboration

5.1.1   Lack of Precision in the Description of Maturity Models

Strangely, discussions around data governance and data management do not use rigorously precise definitions and words, as expected in a context that usually values clarity and precision. Unfortunately, things are like that. When reading technical publications and conversations with suppliers, it is necessary to be attentive. The diversity of understandings is present from fundamental concepts such as "data governance" and "data management" to elements such as "data governance framework" and "data governance maturity model" (McSweeney, 2013).

There is no clear, accurate, and indisputable description of the data governance maturity models (Addagada, 2018). Essential elements that make the epistemology of matter viable are lacking. Discussions come to question whether a proposal is a model of maturity.

Considering the pragmatic objectives of the present work, the text will not deepen the theoretical discussions. With this, the fundamental intention is to help with the selection of tools to support data governance implementation. An emblematic occurrence of the meetings refers to the DAMA-DMBoK2, which will be recommended later as the data governance maturity model. There are, however, many experts who understand that the DAMA-DMBoK2 is not a maturity model.

5.1.2   Maturity Model: It is not about "how to do" data governance

A data governance maturity model does not outline the proper way to carry out tasks. It simply distinguishes what needs to be done in accordance with each organization's development stage.

Data governance maturity models can be vendor independent or tied to a particular software or consulting services vendor (Basker, 2016). In general, there are costs in implementing data governance derived from contracting software, consulting services, or both.

Some maturity models are data management-oriented, such as DAMA-DMBoK2 and DCAM. Other models include Gartner (Firican, 2018), IBM (Firican, 2018), OvalEdge, and Stanford, which use data

and information governance. On initial examination, this focus only generically serves to position maturity models.

Each maturity model contains a central trait that can be recognized. The knowledge areas are the centerpiece of DAMA. It is business capability in DCAM. Processes are the basis of the Stanford model. Competencies are the cornerstone of the IBM model. The different extant data governance maturity models share a lot of similarities, but their associations are complicated and only partially complete.

There are dozens of maturity models in data governance (ANNEX 5.1 presents a sample of maturity models). Even so, using a specific model makes it possible to compare an organization's maturity stage qualitatively and broadly with other organizations that also measure maturity. Although the assessment is not universal, it is possible to benchmark with organizations in the same industry, organizations in the same region, and organizations of the same size, among other references. The comparison helps plan actions and allocate resources to improve the maturity level.

5.1.3   ISORA and TADAT

In the specific case of tax administrations, there are valuable broad models – such as ISORA (International Survey on Revenue Administration) (ISORA, 2021) and TADAT (Tax Administration Diagnostic Assessment Tool) (TADAT, 2019) – for the general assessment of these organizations. These models aim to assess the general situation of the Tax Authority, including IT services. Each covers typical items for using IT services, systems, and data, but they are not specialized. In any case, the joint analysis of data governance maturity and the general situation of the tax authority is helpful for the search for a balanced evolution. It is not advisable to spend resources on a "super powerful engine" (the data governance) for a precarious vehicle (tax administration with a very low ISORA or TADAT rating).

The Innovation Digitalization and Technology Index, developed by CIAT (Díaz de Sarralde & Morán, 2022), is of particular interest, which uses the data of over 150 countries under four thematic areas: technological innovation, compliance improvement, operational digitalization, and budgeting.

5.1.4   Using an Existing Maturity Model

While there is great variety and diversity in data governance maturity models, many organizations choose to develop their models (Palmer, 2021). Virtually all maturity models use the CMM[32] as a reference. The CMMI[33], its current name, was initially designed to assess maturity in software development. Subsequently, CMMI expanded the scope, and, among other activities, the model has also been used in the maturity of data governance. (Steenbeck, 2021)

A hypothetic "tailor-made model" will probably use parts of existing models. However, this approach is not recommended. There will hardly be a model that exactly meets the needs of an organization (where it is, objectives and goals, deadlines, challenges, etc.). But it is still possible to find good approximations among the available models. The existing models are of general use,

---

[32] Capability Maturity Model
[33] Capability Maturity Model Integration

encourage systematization, incorporate good practices, and minimize the typical risks of self-assessments. It makes no sense to "reinvent the wheel."

5.2   Data Governance Maturity Models

5.2.1   Why use DAMA-DMBoK2?

The DAMA-DMBoK2 maturity model (Sebastian-Coleman, 2020) is a good choice based on a few factors:

- The DAMA-DMBoK2 model is broader than just maturity and covers almost all disciplines and activities related to good data management.
- The DAMA-DMBoK2 model is the most used, with about 50% of users (half of the organizations that work with data governance use the DAMA-DMBoK2 model (DataCrossroads, 2021)).
- The DAMA-DMBoK2 model harmoniously guides all data-related actions. It covers operational data management, quality, security, privacy, evaluation, etc.
- The scope of the DAMA-DMBoK2 model enhances the consistency of activities related to the use of data, from the technical-operational level to the strategic level.
- The DAMA-DMBoK2 model offers the possibility of professional certification, which facilitates the training of personnel, assessment, and hiring of specialized services.
- The DAMA-DMBoK2 model was developed to serve different types of organizations with no specific focus on the line of business.
- Uniform data governance model can facilitate benchmarking with other tax administrations such as ISORA and TADAT.
- Tax administrations tend to operate as "data-driven" organizations and greatly benefit from using DAMA-DMBoK2.

An alternative to maturity models is to use "scorecards" techniques. There is literature on the development and use of scorecards. However, this approach is not "systematic," in addition to having other shortcomings with benchmarking, staff training, etc.

5.2.2   The Importance of Measuring

As has been stated on numerous occasions, management requires measuring the existing situation and the results achieved. Due to its nature and complexity, organizations could not measure data governance with traditional instruments, such as operational efficiency gains, return on investments, and cost reduction. In cases like this, maturity models are widely used.

The DAMA-DMBoK2 states that "the primary goal of a data management capability assessment is to evaluate the current state of critical data management activities in order to plan for improvement" (DAMA-DMBoK2, 2017).

Measuring data governance maturity helps to track and assess the evolution of an organization's maturity level. This quality is present in all data governance maturity models. The assessment also provides elements that indicate actions to increase the maturity level so that the organization reaches a more advanced stage in data governance. The use of a maturity model is indispensable in implementing data governance. In addition to the predicates indicated above, the maturity model

offers data governance systematization and the opportunity to use best practices developed by other businesses. Data governance implementation is a multi-year undertaking that often takes three years or longer. The maturity model supports systematic measurement of progress and action options for the next steps. Quantitative knowledge of the distance between the intended and current situation is a decisive benefit.

The maturity model and assessment feature influence the variables that are measured as well as the measurement process. Each maturity item level often fall into bands. In order to lessen subjectivity and facilitate measurement, all models aim to set precise standards for classifying levels. The act of measuring is simple. But the subjective component is always there. It is not unusual to encounter radically demanding and critical viewpoints at one extreme and tolerant and beneficent circumstances at the other. Balance, realism, and systematization are important, just as they are in other situations.

Due to the differences between maturity models, it is almost impossible to compare the results of assessments that use different maturity models. There are variations in the number of maturity levels and their meaning, focuses, domains, and items measured in the assessment. Thus, even using existing maturity models, it is practically impossible to achieve one of the primary objectives of maturity models: the availability of instruments to perform universal benchmarks.

According to DAMA-DMBoK2, chapter 15 (DAMA International 2017), "based on assessment results, an organization can enhance its Data Management program, so it supports the organization's operational and strategic direction. Typically, Data Management programs develop in organizational silos. They rarely begin with an enterprise view of the data. A Data Management Maturity Assessment (DMMA) can equip the organization to develop a cohesive vision that supports overall organizational strategy."

5.2.3   How to Measure

Maturity measurement is based on filling in a points table. Even so, the risk arising from subjectivity persists – with extremes of benevolence or undue demand. Also, a planned action or an intention is often incorrectly classified as "accomplished." Planned activities or desires, however valuable, should not be included in the maturity measurement. Maturity assessment is about what exists.

With few exceptions, maturity model assessments are not fully automatable; they cannot be executed exclusively by machines. The evaluator's judgment must always be present. This fact recommends systematically carrying out the maturity assessment using the same model to compare the results obtained in each measurement. More than one person must assess because they can harmonize the evaluation of each item through discussion and consensus-seeking or by calculating average values.

Although there are dozens of maturity models and assessment resources, access to technical documentation is not simple. Much of the material is not freely available, hampering the preliminary study and evaluation. Given the pragmatic nature of the present work, the Stanford model is recommended for assessment - only for the evaluation (OMES - Oklahoma Office of Management & Enterprise Services, 2016). Stanford has two parts - maturity model and assessment - but its maturity model does not offer the breadth of benefits and market share of DAMA-DMBoK2.

It is recommended that the assessment feature be carefully selected and maintained throughout the implementation of Data Governance. The use of the same assessment instrument facilitates the analysis of the evolution of the organization's maturity. In addition, the assessment provides subsidies for planning subsequent actions.

Regarding the "existing Data Management Maturity Assessing frameworks," the DAMA-DMBoK2 informs that "the data management maturity assessment framework is segmented into discrete data management topics. Framework focus and content vary depending on whether they have a general or industry-specific focus. However, most address subjects that can be mapped to DAMA-DMBoK2 Knowledge Areas. Many vendors have developed their own models. Organizations should evaluate several models before choosing a vendor or before developing their own framework" (DAMA-DMBoK2, 2017). The DAMA-DMBoK2 expressly cites the following models: CMMI Data Management Maturity Model (DMM), EDM Council DCAM (EDM Council, 2014), IBM Data Governance Council Maturity Model, Stanford Data Governance Maturity Model, and Gartner's Enterprise Information Management Maturity Model.

It is essential to carry initial maturity assessment before the implementation of Data Governance begins. The result of this evaluation is a reference for a comparison of future reviews, which must be carried out at least every year.

The assessment schemas proposed by Marchildon et al. (Marchildon, 2018) and by OvalEdge (Varshney, 2021) are exciting, especially concerning completeness and ease of use. The schemas, even if not used, are good training resources.

5.2.4   DAMA-DMBoK2 in a Nutshell

Many data governance experts do not consider DAMA-DMBoK2 a framework or maturity model. However, the scope, consistency, and breadth of use make it the broadest technical reference in data management (on the order of 50% - variations depending on measurement criteria). For the implementation of data governance, DAMA-DMBoK2 is an indispensable guide. From a technical point of view, it addresses the main issues.

The involvement and commitment of senior management establish the strategic importance of data governance. The use of DAMA-DMBoK2 as a technical guide complements the strategic approach and, in addition, offers guidance on world best practices and the certainty that no vital topic will be overlooked.

DAMA-DMBoK2 has five maturity levels (Figure 5.2) and level 0 (lack of capability).

Among other factors, DAMA-DMBoK2 offers the possibility of professional certification, associated with training and services available in the market.
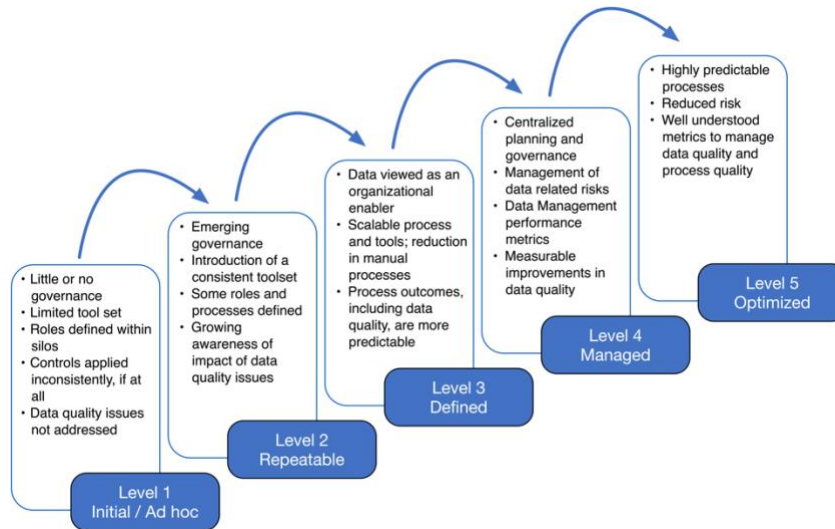
Figure 5-2 Adapted from DAMA-DMBoK2 maturity levels. Source: DAMA-DMBoK2

DAMA-DMBoK2 is just a recommendation, but each organization can select the method that best suits its needs and convenience. There are commercial solutions – frameworks, tools, consultations, training, etc. – and free software solutions. DAMA-DMBoK2 presents what is essential in each area used to structure the practices. It does not establish "how" to do it.

Some organizations cannot do without implementing data governance and must choose a technical reference and a data governance model. DAMA-DMBoK2 can certainly suit most organizations (Kempe, 2011) (Kempe, 2011).

The maturity model is not always directly linked to a framework or guide. Furthermore, many frameworks do not have a specific maturity model. Many maturity models can be used together with different technical materials and guides.

There are Data Governance models from Gartner, IBM, Stanford, DataFlux, Oracle, OvalEdge, Data Crossroads, Talend, etc. The list is not exhaustive and does not indicate preferences or recommendations. Usually, specialized consulting services offer a framework as an instrument for carrying out the work, along with the data governance implementation roadmap. In the case of software vendors, the framework is part of packages of solutions aimed at data governance.

DAMA-DMBoK2 has almost 1,000 pages, essentially technical. The material has chapters dealing with operational topics, information use, administration, and a specific chapter on maturity.

Figure 1-2 shows the DAMA Wheel, with 11 knowledge areas that structure it. Data Governance is at the center, interconnecting the other disciplines "since governance is required for consistency within and balance between the functions." According to DAMA-DMBoK2 (DAMA-DMBoK2, 2017), this way of presenting the 11 areas of knowledge is described as follows:

"Because data moves horizontally within organizations, Knowledge Area activities intersect with each other and with other organizational functions.

1. Data Governance [*at the center of the wheel*] provides direction and oversight for data management by establishing a system of decision rights over data that accounts for the needs of the enterprise.

2. Data Architecture defines the blueprint for managing data assets by aligning with organizational strategy to establish strategic data requirements and designs to meet these requirements.

3. Data Modeling and Design is the process of discovering, analyzing, representing, and communicating data requirements in a precise form called the data model.

4. Data Storage and Operations includes the design, implementation, and support of stored data to maximize its value. Operations provide support throughout the data lifecycle from planning to disposal of data.

5. Data Security ensures that data privacy and confidentiality are maintained that data is not breached, and that data is accessed appropriately.

6. Data Integration and Interoperability includes processes related to the movement and consolidation of data within and between data stores, applications, and organizations.

7. Document and Content Management includes planning, implementation, and control activities used to manage the lifecycle of data and information found in a range of unstructured media, especially documents needed to support legal and regulatory compliance requirements.

8. Reference and Master Data includes ongoing reconciliation and maintenance of core critical shared data to enable consistent use across systems of the most accurate, timely, and relevant version of the truth about essential business entities.

9. Data Warehousing and Business Intelligence includes the planning, implementation, and control processes to manage decision support data and to enable knowledge workers to get value from data via analysis and reporting.

10. Metadata includes planning, implementation, and control activities to enable access to high quality, integrated Metadata, including definitions, models, data flows, and other information critical to understanding data and the systems through which it is created, maintained, and accessed.

11. Data Quality includes the planning and implementation of quality management techniques to measure, assess, and improve the fitness of data for use within an organization."

DAMA-DMBoK2 covers all data management topics, with a focus on technical aspects. To use that material in implementing data governance, each tax administration can prioritize the data governance module and some modules that cover the other areas. The choices can be based on the assessment, or the selection made by the organization, considering the existing situation, legal requirements, challenges, and goals. Tax administrations must implement all disciplines in the long term. Still, the organization can adjust and prioritize the sequence of domains to implement and develop according to the challenges, difficulties, and availability of resources.

The option for DAMA-DMBoK2 provides a combination of strategic and technical factors. In the strategy, DAMA-DMBoK2 is adaptable to the form of implementation chosen by the organization, such as the establishment - or not - of a Data Governance Committee, data owners, etc. In the technical part, DAMA-DMBoK2 promotes the solid implementation of data management and related disciplines, with a consistent structuring of data governance support. Data governance can be long-lasting and sustainable through the implementation of DAMA-DMBoK2 in tax administrations.

5.2.5 Short Description of the Stanford Data Governance Maturity Model

Like many other maturity models, the model is based on the Capability Maturity Model (CMM), created by the Software Engineering Institute at Carnegie Mellon University in 1986. By the way, the CMM was initially developed for managing software development processes. The CMM is currently known as the Capability Maturity Model Integration (CMMI) after modifications and additions.

"The Stanford Data Governance Maturity Model was developed for use by the University; it was not intended to be an industry standard. Even still, it serves as a solid example of a model that provides guidance and a standard of measurement. The model focuses on data governance, not data management, but it nevertheless provides a basis for evaluating data management overall" (DAMA-DMBoK2, 2017). The Stanford Data Governance Maturity Model serves the purposes of maturity guidance and assessment (OMES - Oklahoma Office of Management & Enterprise Services, 2016). In this opportunity, it is indicated as a resource to assess data management maturity, while the DAMA-DMBoK2 is recommended as a guide.

Stanford Model differentiates between "foundational (Awareness, Formalization, Metadata) and project (Data Stewardship, Data Quality, Master Data) components." Within each component, the Stanford Model "articulates drivers for people, policies, and capabilities" and "provides qualitative and quantitative measurements for each level." (DAMA-DMBoK2, 2017)

Table 5-1 presents the two components – foundational and projects – and the three elements in each. The foundational aspects are linked to the core competencies of data governance. The project group encompasses factors related to using data governance concepts in ongoing projects. The three dimensions – People, Policies, and Capabilities – are used in assessing each of the six maturity elements.

Table 5-1 Data Governance Maturity Model. Guiding questions for each Component-Dimension Source: (OMES - Oklahoma Source: Office of Management & Enterprise Services, 2016). Redrawn for clarity

Data Governance Maturity Model
Guiding Questions for each Component-Dimension

Foundational	People	Policies	Capabilities
Awareness	What awareness do people have about their role within the data governance program?	What awareness is there of data governance policies, standards and best practices?	What awareness is there about data governance enabling capabilities that have been acquired or developed?
Formalization	How developed is the data governance organization and which roles support data governance activities?	To what degree are data governance policies formally defined, implemented, monitored and enforced?	How developed is the toolset that supports data governance activities and how consistently is that toolset used?
Metadata	What level of cross-functional participation is involved in the development and maintenance of metadata?	To what degree are metadata creation and maintenance policies formally defined, implemented, monitored and enforced?	What capabilities are in place to actively manage metadata at various levels of maturity?

Project	People	Policies	Capabilities
Stewardship	To what degree have stewardship roles been defined and filled?	To what degree are stewardship policies defined, implementad and enforced?	What capabilities are implemented to support stewardship?
Data Quality	To what degree have data quality competences developed?	To what degree are data quality policies defined, implemented and enforced?	What capabilities are implemented to support the production and maintenance of high quality date?
Master Data	To what degree has a formal master data management organization been developed and assigned consistent responsibilities across data domains?	To what degree are master data policies defined, implemented and enforced?	What capabilities are in place to actively manage metadata at various levels of maturity?

The operationalization of the evaluation, involving qualitative and quantitative aspects, is presented in Chapter 8 (Data Governance Guides). The implementation of data governance can skip no maturity level, as the evolution of maturity follows an established order.

5.2.6   Data Governance Matters

It can be desperate to see, on the one hand, how the tax administration depends on the use of data and, on the other hand, how the data could be deficient in terms of systematization, documentation, modeling, quality, security, credibility, internal knowledge, literacy, etc. For many years, even having unquestionable knowledge about data management, the imperious demands of everyday life led to the data being "filtered" by the systems – usually, users only see what the systems show. Changes in technologies, corrective, adaptive, and evolutionary maintenance, new legal requirements, and security adjustments, among other causes associated with short deadlines and lack of resources, often lead to precarious solutions.

From the point of view of IT professionals, data management techniques and methods are known and used – data dictionaries, data modeling, standardization, etc. Likewise, technical solutions for data integration, processing, and presentation and services independent of the intervention of IT professionals – such as Data Warehousing, Business Intelligence, and Data Visualization – are available and tempting.

Even in cases where technical resources and knowledge are employed, it is not always possible to guarantee that there are high-quality data. Many organizations do not have even minimal satisfactory conditions with their data.
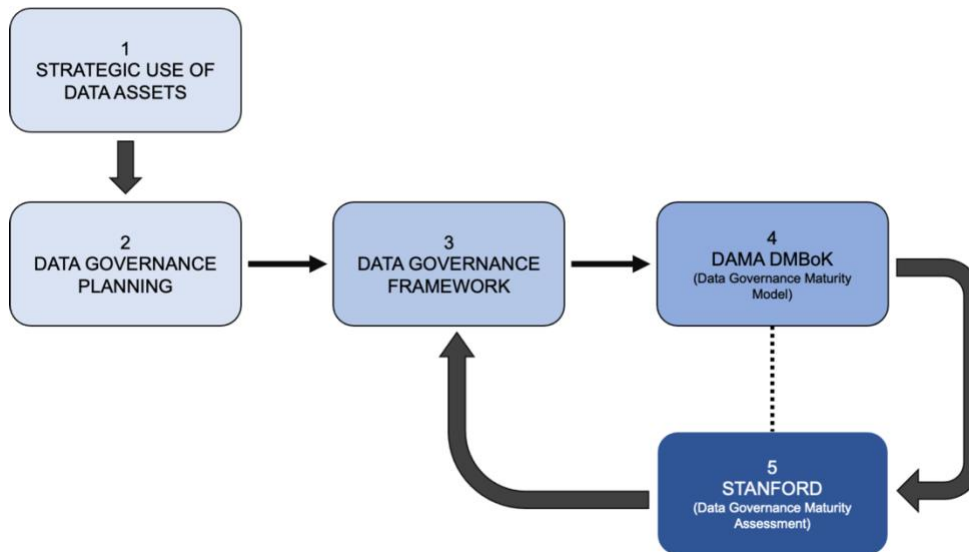

Figure 5-3 Data Governance System. Source: Prepared by the authors

The obstacle arising from difficulties with data – quality, documentation, literacy, etc. – may seem insurmountable. In situations like this, an approach is recommended based on the systematization of activities, effective data management, consistency of processes, and standardized assessments. Even more important is recognizing the impossibility of buying a general-purpose technical solution and acknowledging that solutions consume financial resources and time and require internal changes.

The five blocks in Figure 5-1 seek to summarize the structuring of the data governance system.

Block 1 deal with the strategic use of data assets. Assets can span existing data and data not yet available. Data can be structured and unstructured; internal and external; transactional and analytical; etc.

Data governance planning, Block 2, encompasses the activities that must be performed to meet the strategic objectives. The range of activities can be vast and go beyond just IT-related topics.

The framework (Block 3) presents the approach (wide vision and blueprint) used to achieve the objectives and goals. The list shown below is just an example:

- Consolidation of data management.
- Implementation of data governance functions (Chief Data Officer, Data Governance Committee, Data Stewardship, etc.).
- Use of ETL resources (Extract-Transform-Load) and data integration (DW, BI, Visualization, etc.).
- Treatment of data as a "product."
- Use of approaches offered by vendors and consultants.

Block 4 represents the maturity model, which takes care of the systematized activities. DAMA-DMBoK2 offers this approach where the consolidation and systematization of typical data management activities stand out. Consistent implementation of data governance depends on sound data management. Any deficiency or weakness in data management will undermine Data Governance.

Block 5 performs the feedback function in the Data Governance system. The "maturity assessment" serves as the systematic collection of the results achieved and offers indicators for prioritizing the following activities. The use of the Stanford assessment feature is indicated according to the reasons already exposed.

Data governance solutions offered by product or service providers generally cover blocks 3, 4, and 5. It is usual for these solutions to recommend DAMA-DMBoK2 for data management.

5.2.7   Data Governance and COBIT

The tax administration does not always have the resources and conditions to face the challenges of data governance through the implementation of data governance, the selection of a framework and maturity model, definition of roles and responsibilities, hiring specialized services, etc.

The tax administration often needs to improve activities in IT service management (Martins & Seco, 2020), total quality, security, personnel training, documentation, processes, etc. The lack of resources and deadlines are also present. It takes clarity of direction, determination, and persistence to act. The situation is more difficult about data, as the improvement measures are specific to each organization and each situation.

Considering all the factors mentioned, highlighting the urgent need for the tax administration to produce good results and generate sustainable value with data, implementing data governance can be indispensable. Even so, it is critical to consider the overall situation of the organization. While it is recommended to prioritize resource allocation and take care of data governance, in cases where it is not possible to adopt this approach, it is reasonable to consider using a "palliative" approach. The "palliative" approach produces results that are immediately useful and, at the same time, preparatory to implementing data governance.

Figure 5-4 presents a diagram that summarizes the sequence of decisions on the approach. If the tax administration is going to implement data governance, this text is pragmatic and can help the execution of the work.
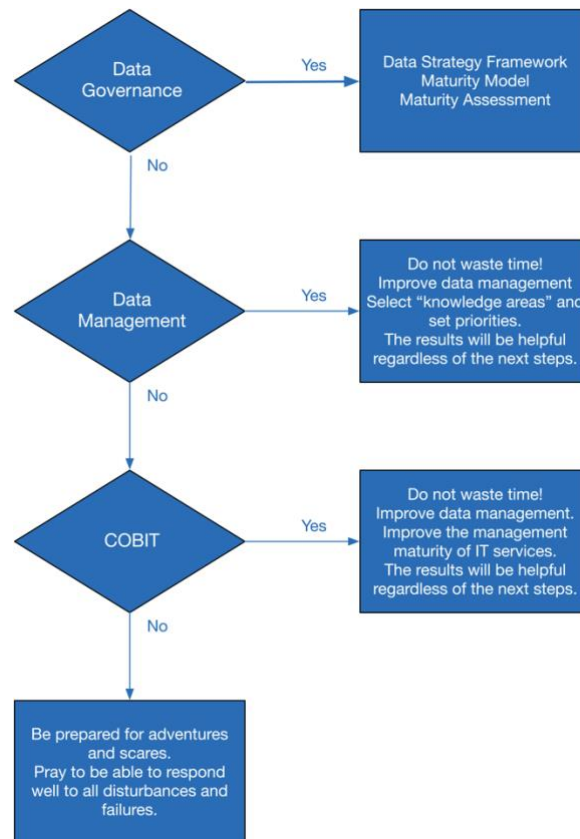
Figure 5-4 Data governance. A sequence of implementation. Source: Prepared by the authors

If the tax administration decides it will not start implementing data governance, some actions can be taken to improve data management. In this case, the DAMA-DMBoK2 can be a good solution, and activities can be selected and prioritized in line with the most significant challenges or needs related to data.

Figure 5-4 still presents a more straightforward approach based on the data management practices of ITSM-type solutions. This alternative only makes sense if the tax administration is implementing ITSM. While data governance frameworks offer specific functionality for taking care of data, ITSM frameworks are broad for handling data assets without significant specializations (Addagada, 2016). ITSM frameworks provide other general functionality (security, software development, project management, strategic alignment, etc.) along with powerful specialized functionality (incident management, change management, problem management, performance management, and others).

This workaround – using the capabilities of the ITSM framework to take care of the data – is only palliative and can take advantage of cyclical situations. The result does not fully meet data governance requirements, but they are undoubtedly valuable for addressing fundamental deficiencies in data usage. If the workaround is adopted, the benefits can be realized when there is a later decision to implement data governance.

The tax administration can extend this reasoning to other disciplines, such as security, continuity, software development, etc. Measures based on specific and specialized solutions can address these issues. ITSM solutions provide general functionality for each area without deep specialization.

It is essential to understand that these approaches are palliative, although the results can be later used in implementing data governance.

There are many options for ITSM frameworks, such as ITIL, COBIT, MOF, FitSM, etc. ITIL and COBIT solutions are the most used (Martins & Seco, 2020). The ITIL solution is usually implemented from a more IT-focused "bottom-up" view. The COBIT solution, on the other hand, is initially driven by needs at the tax administration at the strategic level. COBIT and ITIL provide general-purpose measures for data management.

Thus, following the logic adopted in these considerations, it is recommended to use the COBIT solution palliatively if possible. This solution is most powerful when it comes to meeting the needs of strategic levels, which also guides actions in implementing data governance.

COBIT uses five principles, which can link directly to data governance objectives:

- Meet the needs of interested parties.
- Covering the enterprise from end to end.
- Applying a single integrated framework.
- Allowing a holistic approach.
- Separating governance from management.

The organization can adequately exploit the principles to meet data governance requirements:

- Clear ownership of information,
- Timely and correct information,
- Clear management schemas and efficiency, and
- Compliance and security.

The various frameworks – ITSM, data governance, security, etc. – may not be directly interconnected but offer good possibilities for coexistence. In this way, it facilitates eventual transitions.

The tax administration can also use the ITIL framework alternatively to anticipate the strategic start of data governance implementation. However, in a decision between COBIT and ITIL, COBIT has the advantage of being better suited to the top-down approach and corporate interests. ITIL focuses on IT infrastructure and operations and is tailored to the "bottom-up" approach.

ANNEX 5.1

DATA GOVERNANCE MATURITY MODELS (examples)

IDENTIFICATION	VINCULATION	MATURITY LEVELS	DIMENSIONS	ASSESSMENT	REFERENCE
DAMA-DMBoK2 (DAMA-DMBoK2, 2017)	DAMA International	No capability Initial/Ad Hoc Repeatable Defined Managed Optimized	Activity Tools Standards People and resources	No	https://www.dama.org/cpages/home
DataFlux (Smith, 2011) (DataFlux Corp (SAS), 2007)	SAS	Undisciplined Reactive Proactive Governed	People Policies Technology Risk and reward	Yes	https://www.sas.com/content/dam/SAS/en_us/doc/servicebrief/sas-data-governance-maturity-assessment-106383.pdf https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-data-governance-framework-107325.pdf
DataOrchard (Data Orchard, 2022)	DataOrchard	Unaware Emerging Learning Developing Mastering	Uses Analysis Data Tools Leadership Culture Skills	Yes	https://www.dataorchard.org.uk
Dattamza	Dattamza	Initial Conceptual Defined Actively measured Optimized	People Process Technology capabilities Data monetization Managed risk	Yes	https://www.dattamza.org/
DCAM	EDM Council	Non initiated Conceptual Developmental Defined Achieved Enhanced	-	Yes	https://edmcouncil.org/
Data Maturity Model (DMM) – retired	Capability Maturity Model Institute (CMMI)	Initial Managed Defined Quantitatively managed Optimizing	-	Yes	https://cmmiinstitute.com/
Gartner (Firican, 2018)	Gartner	Unaware Aware Reactive Proactive Managed Effective	Data integration across the IT portfolio Unified content Integrated master data domains Seamless information flows Metadata management and	Yes	https://www.gartner.com/en

IDENTIFICATION	VINCULATION	MATURITY LEVELS	DIMENSIONS	ASSESSMENT	REFERENCE
			semantic reconciliation		
IBM (Firican, 2018)	IBM	Initial Managed Defined Quantitatively managed Optimizing	-	Yes	https://www.ibm.com/
Kalido (Chen, 2022), (Firican, 2019)	Kalido	Application centric Enterprise repository-centric Policy centric Fully governed	Organization Process Technology	Yes	https://docplayer.net/2788287-Kalido-data-governance-maturity-model.html
Open Universiteit Nederland Data Governance Maturity Model (Firican, 2019)	Open Universiteit Nederland	No process Beginning process Established process Managed process Optimizing process	Corporate governance Risk management & compliance People Processes Technology Data assets Business alignment Data governance organization Data management	Yes	https://www.ou.nl/ https://datagovernancematurity.wordpress.com/data-governance-maturity-self-assessment/
Oracle (Oracle, 2015)	Oracle	None Initial Managed Standardized Advanced Optimized	People Process Technology	No	www.oracle.com/assets/oea-best-practices-data-gov-1357848.pdf
OvalEdge (Varshney, 2021)	OvalEdge	Unaware Aware Defined Implemented Optimized	Data quality Data access management Data literacy	Yes	https://www.ovaledge.com/solutions/data-governance
Stanford (Firican, 2018)	Stanford University	Awareness Formalization Metadata Stewardship Data quality Master data	People Policies Capabilities	Yes	http://web.stanford.edu/dept/pres-provost/cgi-bin/dg/wordpress/dgc/
TDWI (Firican, 2020)	TDWI	Prenatal Infant Child Teenager Adult Sage	Organizational 1. Maintain a cross-functional team and process 2 Align with data-intense business initiatives	No	https://tdwi.org/Home.aspx

IDENTIFICATION	VINCULATION	MATURITY LEVELS	DIMENSIONS	ASSESSMENT	REFERENCE
			Technical 3. Govern data usage via technical implementations and 4. Automate data governance process via technical implementations		

Chapter 6   Data Governance Tools

Data governance and management tools are not mandatory components of a framework, but they can support and facilitate almost all aspects of the initiative. They can be used, for example, to support areas such as program and workflow management, collaboration, development of governance policies, process documentation, the creation of data catalogs, and other functions. Tools can also be used in conjunction with data quality, metadata management, and master data management (MDM) tools.

These tools are usually paid, but there are some open-source alternatives. Some software suits with large scopes might contain a subset of data governance tools.

Concerning data governance tools, to make a good choice, a tax administration must consider, among other things, available budget, priority areas, integration with other tools, implementation strategy, long-term maintenance costs (as is done for other software), and requirements of human resources.

Next, an overview of the types of tools available will be described, noting that it is a dynamic area and new products are constantly appearing in the market.

6.1   Glossary of Terms

A business glossary is an essential tool that tax administration must have to identify and maintain the terms, concepts, and definitions that are common throughout all the functions and that are used within the organization. Through this tool, users can consult the terms and definitions. In the glossary of terms, we can find most of the knowledge of what certain concepts mean. It is considered the semantic basis for the organization of data.

This tool has certain features, for example:

- For its construction, it is not necessary to invest in specialized software; glossary can accomplish it through centralized spreadsheets or office tools that allow the terms and definitions to be documented in a structured manner and shared in a corporate repository.
- Reflects the semantics or lexicon of the tax administration.
- Serves as an artifact to promote knowledge in people.
- Promotes communication under a single language in the tax administration.

A glossary of terms is functional/business metadata. It differs from technical metadata such as the data dictionary, through which specifications related to technical data sources such as database tables are described.

The tax administration has two options for creating the glossary: (a) if there is no glossary of terms elsewhere in the organization, a top-down method can be used to produce it from scratch; or (b) when there are initiatives within the organization, a bottom-up method can be used to integrate them and come to a consensus.

In general, a tax administration can carry out the following steps for its construction:

1. Define a work team under a data governance framework.

2. Identify and determine where the terms are used and who uses them.

3. Identify and consolidate the most critical terms of the organization. In case of not having terms, reference can be made to glossaries of terms related to taxes, such as the one proposed by the OECD[34]. Another good approach is to build them covering the scope of tax administration systems architecture as the one offered by CIAT (Gascón Catalán & Redondo, 2020).

4. Coordinate with users and related parties that use the terms and reach a consensus.

5. Put issues for approval of collegiate bodies of data governance.

As part of constructing the glossary of terms, additional information (metadata) can be included to help enrich the understanding of the terms, for example, providing associated synonyms, data owners responsible for managing the term, etc.

Another source of interest for financial and tax terms definition and compatibilization is the glossary published by the International Monetary Fund in an English, French, and Portuguese[35] edition as well as in an English, French, and Spanish[36] edition.

It is essential, like most data governance tools, to automate it while increasing more data collected in this way to be able to keep the information (metadata) sustainable over time. Otherwise, its maintenance can become a big problem.

Table 6-1 Tax glossary terms example

Tax Term	Definition	Synonymous	Data Owner	…
Taxpayer Identification Number	Identification number, which must be used when filing a tax return and assessing taxes and for all other correspondence between the taxpayer and the tax authorities.	Taxpayer Id	…	…
Taxable Base	Amount on which the tax rate is applied, e.g., added value, corporate income, personal income, and real estate.	Imposition, levy, impost, taxation		
…				

Source: Prepared by the authors

---

[34] Glossary of Tax Terms, OECD 2022, https://www.oecd.org/ctp/glossaryoftaxterms.htm

[35] For more information, see IMF GLOSSARY, https://www.elibrary.imf.org/view/books/074/03303-9781589061064-en/03303-9781589061064-en-book.xml

[36] For more information, see IMF GLOSSARY, https://www.elibrary.imf.org/view/books/074/03314-9781589066465-pt/03314-9781589066465-pt-book.xml

6.2   Data Catalog

Provide a single reference point to find the tax administration's different data sources and repositories. In this way, it is easy for officers and officials to search all sources where data assets are found in the tax administration and all relevant information.

Data catalogs are metadata that may or may not be combined with other functional data management capabilities such as semantic searches, data lineages, glossaries of terms or data dictionaries, which brings value through the integration in a single tool. These capabilities are highly dependent on software manufacturers.

The data catalog acts as a hub for all pertinent information about the data that the tax administration has access to, enriching each data type with metadata, such as details about data origins, database engines, instances where they are used, schemas, tables, fields, and data types, as well as quality levels by source and associated quality rules, data domains, associated business terms, and ETL processes.

Although initiatives of manual surveys of data inventories (data catalogs) might be completed as a static picture, it might become unmanageable in large organizations since greater scopes of the ecosystem and data architecture must be covered and that would require significant collaboration between different types of officials or interested parties within the tax administration.

Because of this, efforts to establish projects of this nature should preferably be assisted by automated technologies. For metadata collecting, reverse engineering, semantic inference and interpretation, and tagging, it is possible to utilize advanced tools based on AI, particularly machine learning. This maximizes the value of automation and reduces the need for manual involvement. However, monitoring and control (data governance) will always be required to ensure that the automatic survey and relationship is accurate and compatible with the organization's reality (technology and functional areas).

The data catalog helps gathering the knowledge of the data that is often not documented or when that knowledge resides only in the experience of tax administration officials. For this reason, the data catalog allows the use of information and data throughout the organization.

A practical example of a data catalog tool can be identified through the World Bank data catalog, where you can find (search) World Bank[37] development data, including data from other catalogs. Figure 6-1 shows the initial search functionality of the data inventories held by the World Bank, while Figure 6-2 shows the detail of one of the selected data sources, in this case, the World Development Indicators database

---

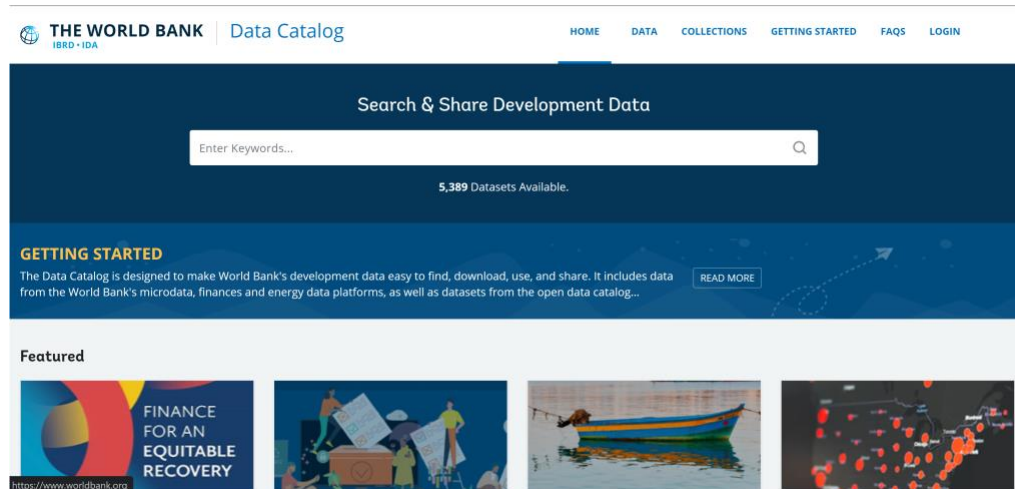[37] Data Catalog (worldbank.org), https://datacatalog.worldbank.org/home

86

Figure 6-1 World Bank Data Catalog. Source: The World Bank Data Catalog
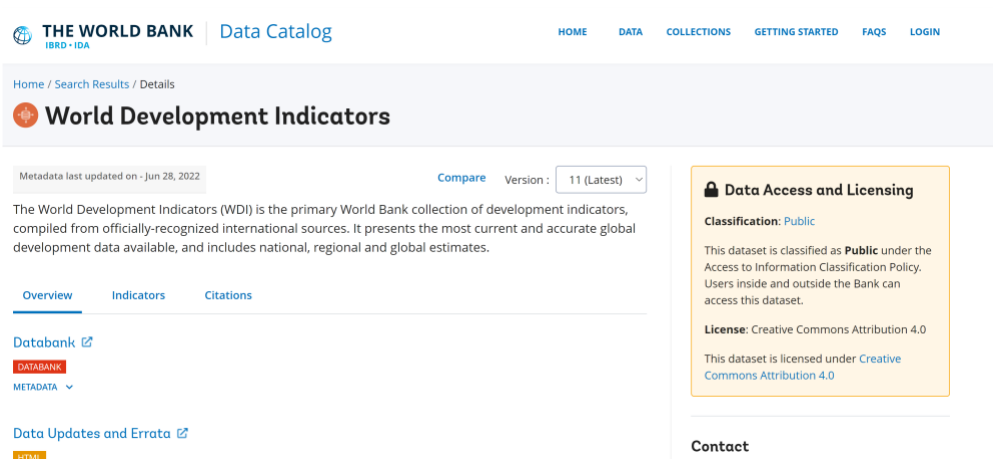

Figure 6-2 World Development Indicators. Source: The World Bank Data Catalog - World Development Indicators[38]

6.3   Data Lineage

Data lineage tools describe a data asset's history from the time it was created to the end of the data flow, making it possible to determine where the data comes from, the systems or processes it passes through, and how it has affected the asset.

A data lineage is a powerful tool for the exercise of data governance because it makes transparent the information sources that participate within the entire tax administration data ecosystem, and typifies how data flows between data sources and destinations, and the treatments they undergo. To a large extent, depending on the type of software used and the software manufacturer, there may be more or less features that help automate data flows as part of the critical metadata for the data governance exercise.

By following the lineage of data through various systems, tax administrations may assist their data governance operations and efforts more effectively, which heavily rely on data traceability. Different

---

[38] World Development Indicators | Data Catalog (worldbank.org),
https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators

data control points can be identified for the organization to apply actions like root cause analysis on data quality problems to rule out causal data errors.

Tax administrations may attempt a survey-based approach to construct the lineage of data manually; however, documenting data lineage requires significant resources, both financial and human.

As part of the characteristics to be considered when evaluating data lineage tools, the following could be considered:

- Native access to several data sources, that is, connectors for metadata scanning of the different information systems that process data (e.g., database engines, ETL systems, data viewers and reports, etc.).
- Visual representation of data traceability between information systems and database engines.
- Deep details in the analysis of the data through different data assets (e.g., database engines, database schemas, tables, and fields).
- Descriptions that enrich the impact analysis and integration with other data governance tools and schemes (e.g., Business Glossary, Data Catalog, etc.).

As one of the key value outcomes from the practice of data governance, data traceability should be made obvious by data lineage tools, as shown in the accompanying picture, where you can track the data regarding taxes, taxpayers, and tax returns from the beginning to the finish.
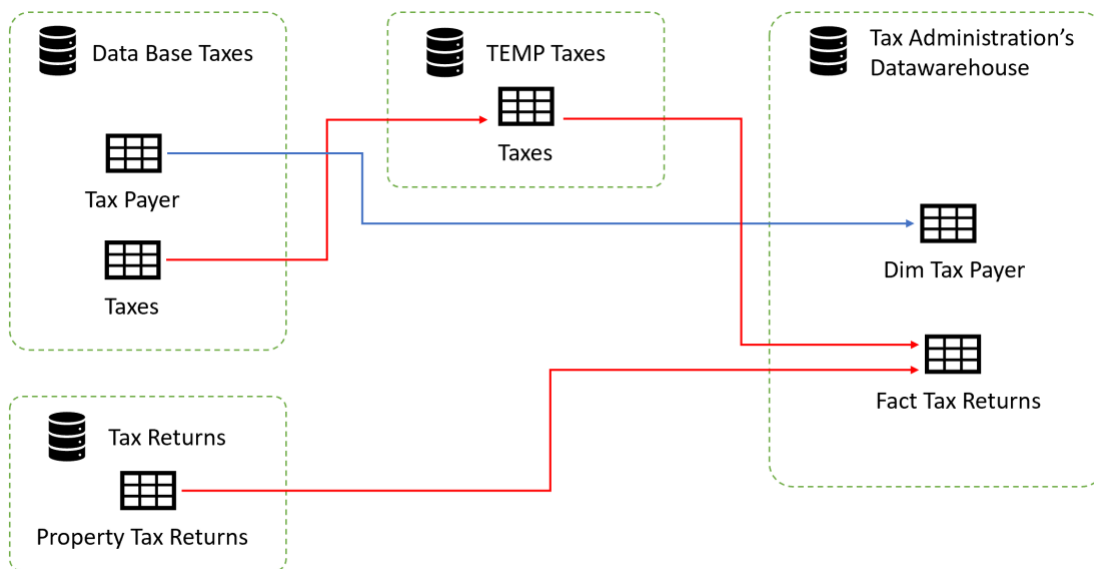

Figure 6-3 Tax Returns Data Lineage. Source: Author's elaboration.

6.4   Document Management and Collaboration Portals

As part of the execution of data governance programs, communication and collaboration between the different stakeholders and the tax administration community is highlighted. Document

management and collaboration portals allow different governance initiatives to be visible and allow active collaboration. (DAMA-DMBoK2, 2017) mentions the importance of documentation and collaborative tools with workflow management capabilities that help to exercise data governance. Some recommended capabilities:

- Document and maintain principles, policies, processes, procedures.
- Control of workflows for involved parties to interact in line with established processes and procedures, as well as data management and governance policies, such as approval of definitions in connection to data assets.
- Active communication throughout the tax administration of data governance activities and initiatives.
- Control the execution of policies and procedures related to data management and governance.
- Contact point for the publication and continuous management of the communication of data governance activities in the tax administration.

6.5 Other beneficial tools

As was already noted, implementing data governance is a difficult task for which there is no universally effective method.

The installation of data governance is a project that is manageable. Data governance must be implemented over an extended period to be effective and permeate corporate culture.

A detailed evaluation of the size and stage of tax administration in terms of data use maturity, data quality, data literacy, challenges, and strategy, is necessary for data governance. To achieve the intended objectives, it is also crucial to evaluate the resources and timelines available.

There is, therefore, a pressing task to estimate the costs involved and the expected benefits. The two factors need to be carefully balanced by tax administrations.

Along with those already mentioned in this work, several additional technologies may be useful and appropriate in this situation for developing and operationalizing data governance.

Most of the time, possessing a tool does not ensure that the desired outcomes will be obtained. In general, the tools are a component of an effort to find a technical answer to the problems. However, data governance is more than just a technological issue. The benefits it delivers in support of corporate results and its integration into organizational culture are the best and highest expectations in relation to data governance.

Consequently, there is not only a technological issue (this approach is typical in IT problems) but also a social challenge, which calls for social solutions. People must be involved in social solutions to adopt new attitudes and habits. It makes sense that the corporate community would adopt even novel linguistic components to successfully implement social solutions.

The tax administration can evaluate the following disciplines and instruments to aid in the implementation of data governance considering the objectives of the current work:

- Project management[39].
- Change management[40].
- BSC (Balanced Scorecard)[41].

Tools to help with data governance integration into company culture:

- Assessment of data literacy[42].
- Assessment of the data quality[43].

Utilizing tools to facilitate actions to incorporate data governance into organizational processes is the goal of the first group (implementation support).

The second group (integration support) tries to make data governance sustainable.

It is appropriate to argue that data governance does not solely depend on the application of tools and technical solutions, regardless of their level of sophistication and expense. The long-term viability of data governance depends on social solutions, which address attitudes and behaviors. The cultural impact of data governance may be leveraged by investments in data quality and data literacy. On the one hand, using high-quality data increases the security of carrying out procedures and making decisions that rely on the data. Personnel, on the other hand, must know and be aware of how data is used, which is a factor that promotes its use. The cultural internalization of data governance can be facilitated by data quality and literacy.

6.6 References: Market Research

Some consulting companies of recognized reputation in the market are dedicated to evaluate products in different areas of information technologies, also including data governance.

These assessments have their own parameters and forms of presentation, which will be summarized below through two well-known assessments: Gartner Group and Forrester Consulting.

It should be highlighted that, considering the unique requirements and circumstances of each tax administration, these evaluations complement but do not take the place of the requirement for own assessments.

6.6.1 Gartner Magic Quadrant

The Magic Quadrant

Gartner's Magic Quadrant allows companies that are interested in having technology-based services or products to have an overview and study of the technological tools in the area of interest. Gartner mentions that its Magic Quadrant offers visual snapshots, in-depth analysis, and practical advice

---

[39] Project management information can be found at https://www.pmi.org/.

[40] Information on change management tools can be found in **Invalid source specified.**, **Invalid source specified.**, 2021), and **Invalid source specified.**.

[41] Information on BSC can be found at www.hbs.edu/ris/Publication%20Files/10-074_0bf3c151-f82b-4592-b885-cdde7f5d97a6.pdf and various websites.

[42] Information on data literacy can be found in **Invalid source specified.** and **Invalid source specified.**.

[43] Information on data quality assessment can be found in **Invalid source specified.** and **Invalid source specified.**.

that provide information on the direction and maturity of market participants. For this analysis, Gartner presents its results through a two-dimensional matrix (completeness of vision and ability to execute), in which four types of technological competitors are allocated: challengers, niche players, leaders and visionaries.
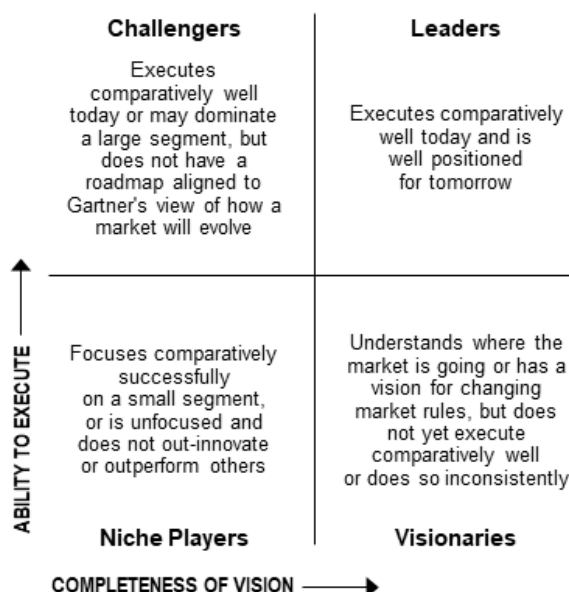

Figure 6-4 Magic Quadrant Two-Dimensional Matrix. Source> Gartner Magic Quadrant Official Page

For more information on the Magic Quadrant components and how they are evaluated, see the Magic Quadrant Official Page[44].

Example: The Magic Quadrant for Data Quality Solutions (2021)

Gartner assumes that (a) by the end of 2022 60% of the organizations will leverage machine learning enabled data quality technology for suggestions to reduce manual tasks for data quality improvement[45]; and (b) through 2024, 50% of the organizations will adopt modern data quality solutions to better support their digital business initiatives[46].

With these strategical assumptions and defining a set of capabilities required by stand-alone data quality software products, the following Magic Quadrant emerges:

[44] See https://www.gartner.com/en/information-technology/glossary/magic-quadrant
[45] See https://intelligent-ds.com/blog/opportunties-with-augmented-data-quality-strategy
[46] See https://blogs.gartner.com/andrew_white/2021/01/12/our-top-data-and-analytics-predicts-for-2021/
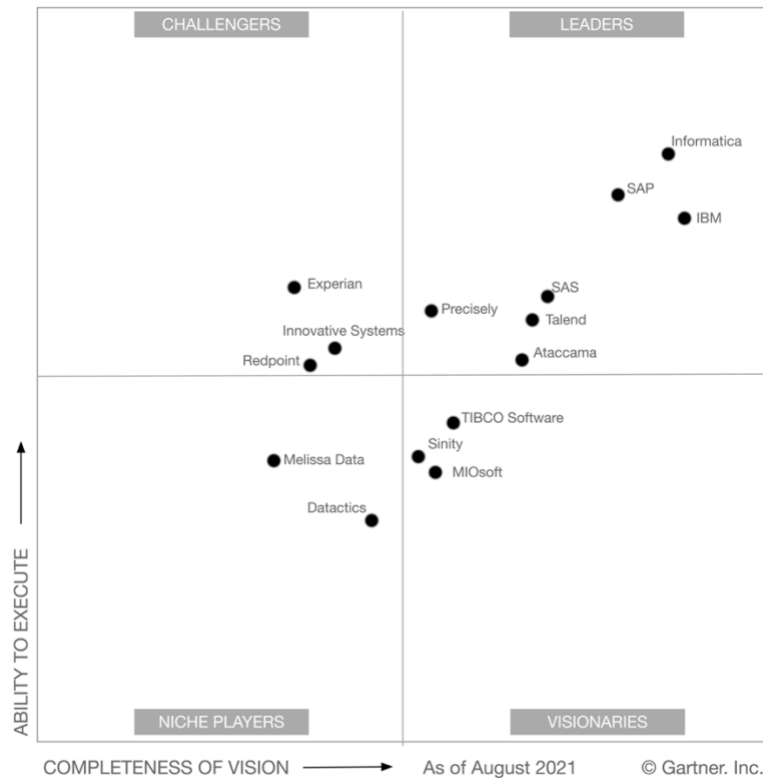
Figure 6-5 Gartner Magic Quadrant for Data Quality (2021). Source: Gartner Inc. Redrawn for clarity.

For more information on capabilities definitions and evaluated products, a reprint of the Data Quality Gartner Report 2021 may be downloaded from Talend[47].

6.6.2   The Forrester Wave™

The Wave

Forrester Wave is a study conducted by Forrester for those interested in acquiring technology products (software and hardware) and services based on analysis and opinions. The study it carries out uses its own methodology[48] in which the different suppliers are evaluated in accordance with its supplier policies – The Forrester Wave and The Forrester New Wave™[49]. The intention of the Forrester Wave study is to provide the buyer with information to support their purchase decisions.

According to Forrester, the evaluation criteria have the following entries:

- Questionnaire: Forrester send questions to suppliers according to the points that the analysts address.
- Strategy and product demo session: demonstration of functionality considering questions and demo scenarios.
- Client references: development of interviews and surveys to clients provided by the provider.

---

[47] See https://www.talend.com/lp/gartner-magic-quadrant-data-quality/
[48] https://www.forrester.com/policies/forrester-wave-methodology/
[49] https://www.forrester.com/policies/wave-vendor-nonparticipation-policy/

Example – The Forrester Wave for Data Governance solutions (2021)

The two axes used to illustrate the Forrester Wave reports' evaluations of companies and their products (current offering and strength in strategy). Providers are classified into one of four "waves" based on their location: Challenger, Contender, Strong Performer, or Leader, each of which is represented by a distinct industry. Strong Performers and Leaders receive stronger marks, while Challengers and Contenders receive lesser scores, so long as the supplier is further to the right and above. A larger dot indicates that the supplier is more represented in the market in terms of more clients and more earnings. The dots along the axis are also varying sizes to reflect market presence.
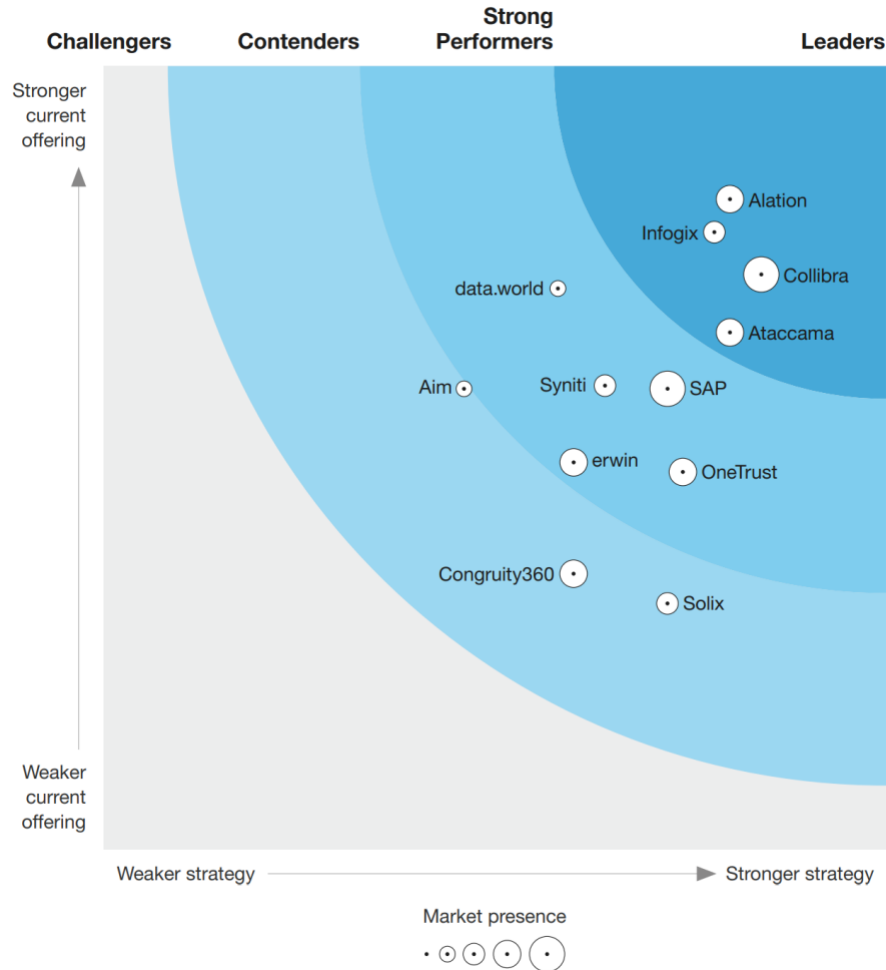

Figure 6-6 Forrester Wave for Data Governance Solutions, Q3 2021. Source: Forrester Wave™

For more information on criteria definitions and evaluated products, a reprint of the Data Governance Forrester Wave Report 2021 may be downloaded from Alation[50].

---

[50] See https://www.alation.com/forrester-wave-data-governance-q3/

Chapter 7   Roadmap for the implementation of data governance in a tax administration

7.1   First Activities

Data governance is not a project (Askham, 2022). Implementing data governance is typically a project. Ideally, at the end of the project, data governance should be an operation, continuous improvement, and ingrained in the organizational culture (Ancick, 2022).

The size and intricacy suggest an evolutionary strategy with clearly defined goals and a definite beginning. Following the completion of a Proof of Concept (PoC), the initial data governance implementation activities should analyze the execution plan and make any adjustments and improvements (PoC). This step should also help to strengthen the changes in communication and achieve quick wins to inspire everyone's engagement and dedication.
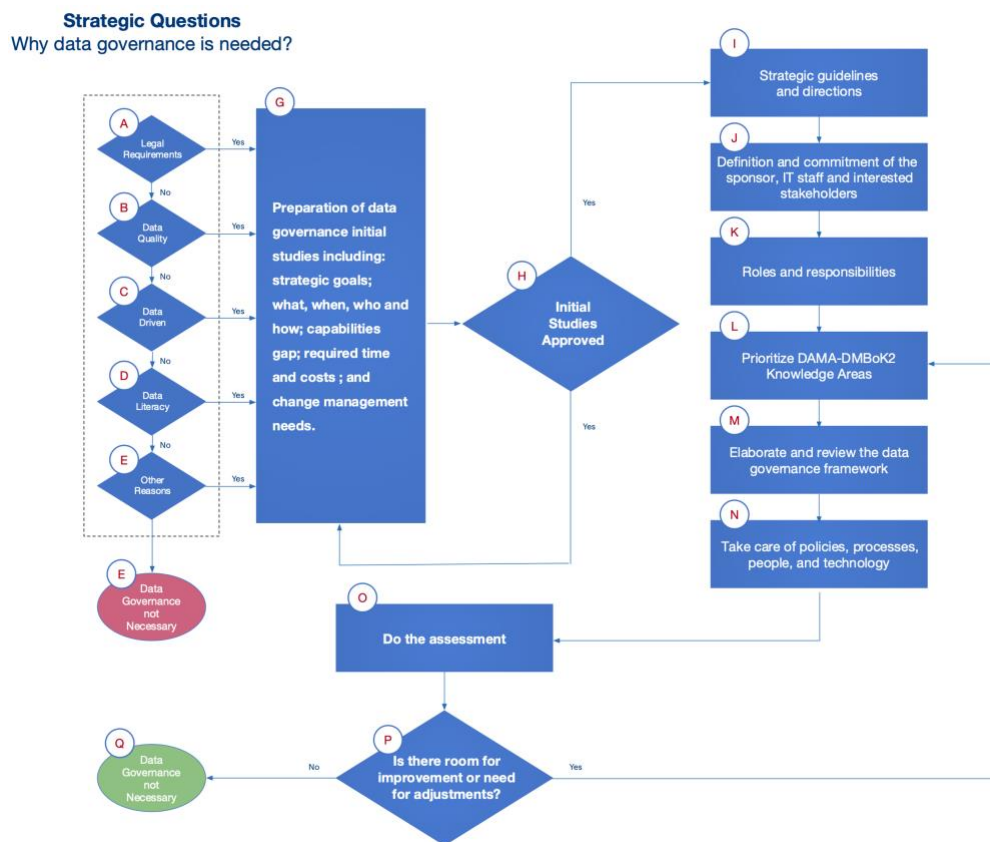

Figure 7-1 Data governance Roadmap. Source: Prepared by the authors.

There are many similarities between the tax administrations, especially regarding activities and organizational structure. However, the detailed examination of the current situation shows that each is quite peculiar. The number of factors related to challenges, strategic direction, processes, people, and technologies, makes each different from the others. The suggested roadmap for implementing data governance, presented in Figure 7-1, is for general use and can be adapted to each tax administration's challenges, capabilities, and availability.

7.2   How to Implement Data Governance

Implementation options, such as using in-house staff or external services, free or commercial software, expert advice, and timeframes, must be considered in the roadmap adequacy (Peters, 2021).

It is imperative to understand that data governance is not a challenge for the IT unit exclusively – the intense participation and commitment of the entire organization is essential (Alation, 2020) (Informatica, 2021). The IT unit must perform many cautions and data handling activities. There are, however, other activities related to the use of data that are carried out by the business units. Data governance must harmonize the two sets – IT infrastructure activities and the effective use of data across business units. There are tax administrations that have made significant progress in IT activities. Such a situation is an advantage for the implementation of data governance.

7.3   Why deploy data governance?

The initial blocks of the roadmap seek to answer the following question: why does the tax administration need to implement data governance? (Walery, 2021) This question can have more than one answer. Among the most common situations, those presented in Blocks A to D Figure 7-1 stand out:

- Legal requirements (Block A): Many countries have established legislation that determines cautions and responsibilities related to the lifecycle of personal data. For instance, the European Community approved this type of legislation in 2016. Since then, countries in other parts of the world have followed it. Compliance with legal requirements often needs action on the data and can be enough to justify the implementation of data governance. Another legal issue stems from compliance laws, those that are related to the tax systems and others. The tax administration is affected by these two matters.
- Data quality (Block B): the loss of credibility in the data and the deficiencies in the data lifecycle can require significant resources and efforts. Instead of just implementing isolated efforts to improve the quality – and, consequently, the trust – in the data, the tax administration must evaluate the action, viability, convenience, and opportunity of dealing with the issue with the implementation of data governance.
- Data-driven (Block C): the tax administration is an organization that works essentially with data. In general, taxes and processes establish the structure and operations of the tax administrations, following best practices of recent decades. However, with technical and organizational innovations, tax administrations can currently improve their functioning and increase their efficiency and effectiveness by becoming data-driven organizations through a digital transformation. Once again, the resources and efforts for data adequacy are significant and, in general, justify the implementation of data governance.
- Literacy (Block D): The new technological solutions can expand the capacity to use IT in tax administrations. As a result, some business units can achieve autonomy in the use of data, often without depending on the IT unit (Wills, 2022) (Panetta, 2021). For this, business units must have the knowledge and secure access to data and master new

tools. In most cases, however, action to disseminate data knowledge and secure access may require significant resources and efforts, which justify the adoption of data governance.

Regarding the four situations described for Blocks A to D, data governance can meet the needs exposed and, in addition, offer more benefits. Most tax administrations live with the Blocks A to D issues. However, given each tax administration's peculiarities, size, and maturity, there are other issues - covered by Block E- which may justify the implementation of data governance.

7.4   Why not implement data governance or implement it only partially?

Unless there is a strategy redirection, the tax administration won't have to handle data governance if all data issues are addressed and fixed. On the other hand, if a tax administration is overwhelmed with existing challenges and deficiencies and has minimal resources and capacity, it must most likely postpone the implementation of data governance. Considering that the tax administration is dependent on data, it seems inevitable that the implementation of data governance will be necessary at some subsequent point.

Until the right time comes, the tax administration can take advantage of valuable opportunities for data management and take some critical steps. One of these possibilities is the adoption and implementation of some knowledge areas that make up the DAMA-DMBoK2, in an isolated way, instead of a complete implementation of data governance. The choice of knowledge areas must consider the challenges, deficiencies, and short-term goals of the tax administration.

Another opportunity is to take advantage of implementing techniques, processes, and tools for the systematic management of IT services. It is essential to clarify that this measure does not achieve the results of data governance but helps the organization prepare for the journey toward data governance. These techniques, processes, and tools are called ITSM (IT Service Management) solutions.

There are several ITSM solutions. Each has different disciplines or practices. Part of these disciplines or practices deals with data. Most of the time, ITSM solutions follow the operational and technical vision of the IT unit, but they are still helpful in improving data management. As presented in Chapter 5, if the implementation of data governance still must not be done, using data management features of an ITSM solution – such as COBIT or ITIL – can be very beneficial for tax administration.

Finally, there are solutions based on the composition of ETL (Extract-Transform-Load) solutions and Business Intelligence (BI) or data visualization resources. This approach does not fall into data governance. It is a palliative way of dealing with data quality deficiencies and lack of "literacy." It can be a step-in data preparation, appropriate in a context of scarcity of capacity and resources, but it is not data governance.

7.5   Initial Studies

Suppose the implementation of data governance ("why?") is decided. In that case, the tax administration must prepare the initial studies (Block G), covering strategic goals, how to do it, when to do it, resources and capabilities required, cost estimates and deadlines, and change management

(which is very complex), and expectation of benefits. These preliminary studies should also estimate the roles and responsibilities needed to implement data governance (Informatica, 2021).

It should be clear that data governance is not an IT topic. Data governance encompasses the entire tax administration and, in a unique way, the use of data to support strategic goals. The IT responsibility includes fundamental activities for data to be available with quality and security (Seiner, 2020). In this scenario, the tax administration should carry out the evaluation and approval of the implementation of data governance (Block H) at the highest management level of the organization (Benthien, 2022) (data.world, 2021).

After approval of the data governance implementation plan, the tax administration can deepen the preliminary studies, and the first actions must be carried out (Wray, 2016).

7.6   Pay Attention to Change Management and Communication

Block I deals with strategic guidelines and direction, adjusting them to the disclosure of the decision in the organization. Intensive communication is an essential part of managing change. Employees must be motivated and, therefore know in detail what will be done, why and who will do it, how they can participate, and the expected gains.

It is necessary to create a group of players responsible for change management concerning the implementation of data governance. This group involves sponsors, stakeholders, and IT staff (Block J). Each one should know the meaning and importance of data governance for tax administration, according to their field of activity and specialization. Each one must be convinced and deeply committed to the decision to implement data governance.

7.7   Roles and Responsibilities

The tax administration must formalize roles and responsibilities according to its organizational culture, size, situation, and challenges (Block K). Not always will there be the exclusive dedication of staff. The allocation of roles and responsibilities does not have to follow the administrative structure strictly. Informal leaders often produce extraordinary results in the implementation of data governance.

7.8   Address a maximum of four knowledge areas at a time

The DAMA-DMBoK2 recommends a framework involving eleven knowledge areas, including "data governance." It is recommended that the implementation should start with at most four of those areas. In exceptional situations, when the tax administration is advanced in one of those areas of knowledge, a more significant number of areas can be undertaken (Block L). The assessment of the current situation of the tax administration, its challenges, and strategic direction helps with the prioritization of knowledge areas. Ideally, the implementation can start with a specific delimited topic or tax administration area. This way, it will be possible to make corrective adjustments and seek quick wins to motivate everyone.

After each maturity assessment is carried out, the prioritization of knowledge areas can be reviewed and adjusted to the needs and circumstances of the tax administration. The limitation to work in up to four knowledge areas simultaneously remains valid for the total implementation effort.

7.9   Framework

Block M indicates the preparation or revision of the framework. Many frameworks are presented and accessible via the Internet, as seen in the technical references (SAS, 2018). On the one hand, the frameworks are specific to implementing data governance (Askham, 2022). On the other hand, the organization's current state determines the framework's proper configuration. The resources identified by Talend (Talend, 2020) may benefit from balancing the planned activities to achieve results.

The activities selected from the framework will establish actions in terms of policies, processes, people, and technology (Block N). Policies must be formalized and widely communicated (Thomas). The methods stem from the DAMA-DMBoK2 Knowledge Areas. The systematization of processes helps incorporate data governance into the organizational culture (Microsoft, 2020).

7.10   Maturity assessment

There are options for the maturity assessment (Block O). Among them, the Stanford Maturity Model assessment's indication follows this work's intentions of simplicity and pragmatism. The essential recommendations are to maintain the use of the same assessment resource throughout the implementation of data governance and to carry out the assessment systematically and periodically. A tax administration can find more information on the respective tab in Chapter 8.

Using the same assessment feature makes assessing the evolution of data governance maturity easier. The tax administration must conduct the maturity assessment at least once a year. The organization can shorten this period by following the general deadlines recommended for implementing data governance. The maturity assessment must follow the standards that, among other things, establish that "intentions" and "plans" should not be counted as "actual accomplishments."

7.11   Progressive Implementation of Data Governance

Block P is the decision point in the data governance implementation loop. If there are still opportunities for improvement or need for adjustments, the tax administration must execute the cycle one more time (Blocks L to P). There is no obligation for the tax administration to reach the highest level of data governance. In the case of DAMA-DMBoK2 maturity levels, level 5 is the highest. For example, the tax administration can set a goal of reaching level 3 within a specific time frame. In this case, the reference for the assessment of Block P would be level 3, and when level 3 is reached, the tax administration will finish the implementation of data governance.

7.12   Final Comments

The purpose of the roadmap is to organize the activities for implementing data governance so that no work is forgotten. The graphic form of the roadmap facilitates communication with all those involved, including personnel at the top management level of the tax administration, sponsors, and the data owners and stewards. The tax administration can prepare the roadmap following other mechanisms queried in the technical references.

A tax administration should place special emphasis on two actions: data literacy and data quality, considering the requirement to integrate data governance into the business culture. Employees who are more data literate will be more empowered, proactive, and secure while using technologies and data. They will feel more assured if you provide high-quality data. Data governance will be put into place in a more effective, long-lasting manner that will directly help the tax administration.

Chapter 8   Data Governance Guides

Data Governance Guides include practical guidance about topics needed to know how to implement a data governance framework within tax administrations. These guides contain practical steps to implement the main activities that support this objective. Many aspects dealt with here are applied in the previous chapters, which will be indicated when necessary.

8.1   Data Strategy Definition Guide

A data strategy is a mid to long term plan designed to identify the resources needed to improve all data related processes within the tax administration, including how to acquire, store, manage, share, use and dispose data and information assets and how to generate value from them. It's important to think about the whole data lifecycle and the tax administration needs when designing the data strategy.

Data Strategy Goals

- Help tax administration to improve how organization handle (acquire, store, manage, share, use and dispose) data.
- Align data actions to strategic tax administration needs.
- Improve general data quality

Data is an important asset within tax administrations. Tax administrations rely completely on data when fulfilling their mission; therefore, it is essential to define and execute a data strategy according to the strategic needs. The data strategy is based on the definition of the strategic course of action (Data Strategy Route), which is the reason for this guide, and is complemented by its execution (Data Strategy Execution).

Figure 8-1 Data Strategy Route. Prepared by the authors.

8.1.2   Data Strategy Route

First, data strategy is defined to ensure that the data generates value for the tax administration, has the appropriate data management capabilities to support it, and the use case that reflects the interest of the stakeholders.

Align data and the tax administration's strategy

Any activity to manage data must contribute to value generation within the tax administration; for instance, a data quality strategy might align with the broad implementation of electronic tax

domiciles for all taxpayers as a way to improve tax compliance. The efforts of the data management and analysis initiatives in the tax administrations must be directed through a data management and data analysis strategy that points towards the strategic objectives of the organization. In a similar vein, cutting-edge data analysis techniques may enable well-informed choices in risk management tactics that address possible tax evasion. Understanding the data requirements of the tax administration is the first step in obtaining the optimum data strategy.

Assess data management capabilities

In organizations, such as tax administrations, it is important to know the direction they must take for the development or execution of actions related to data. In this sense, it is complemented with the identification of information needs of the organization with an evaluation of the degree or level of maturity that the tax administration has in relation to data management.

For the maturity, an assessment framework that is aligned with the organizational needs should be selected and applied for different capacities. With the maturity analysis completed and after determining the gaps, the practices that the organization must undertake for adoption will be determined, for its implementation and future development until a desired maturity level is reached.

Prepare the organization and engage stakeholders

Once the needs and level of maturity of the tax administration have been identified, interested parties should be involved to reach a commitment at both tactical and strategic levels. As a suggestion, involved parties must jointly develop a use case that gather the needs of the tax administration and that would be used as a vehicle to deploy the data strategy. The business case will provide real benefits to the organization, since it should be aligned to solve an actual problem that the tax administration is facing. Among the aspects to be considered in a use case are:

- Executive summary
- Project definition
- Tax administration requirements (Organization's needs and maturity results)
- Strategic alignment
- Expected benefits
- Risk

The use case should become part of the data strategy document.

Define the data roadmap

After the data strategy has been formulated, the tax administration must establish long-term and short-term objectives and goals applicable to the entire organization or specific institutional units. The tax administration needs to plan how to achieve the stated purposes. These plans must be aligned with the institutional strategic plan and executed in coordination with annual operational plans, particularly those that involve IT. These blueprints will shape the data strategy roadmap. The plans must be specific and detailed and include aspects such as: who is responsible for the objective; what process and technology will be used; costs and expected return on investment; timeline; indicators for tracking progress; expected results; change management planned activities and responsible people; communication artifacts and frequency.

8.1.3   Data Strategy Execution

Once the data strategy is defined, it is executed and followed up.

Implement and control

For the execution of the data strategy, the following aspects are taken into account:

- Deploy planned data initiatives and control their execution.
- Measure the execution of the initiatives, the value produced from the data strategy and metrics related to data management (e.g., data quality).
- Development a change management and communication plan.

8.2   Data Management Principles and Policies Definition Guide

Existing laws, policies, and regulations must be considered to establish principles and policies, especially those that directly pertain to data and its handling (tax secrecy, data protection, transparency, statistical treatment, administrative responsibility and accountability, internal control).

8.2.1   Principles

The principles of data management and data governance are general rules that must remain constant over time. They help guide and maintain the actions the tax administration deliberately takes when dealing with data.

The principles can be established focusing on different areas or levels within the tax administration. They can be seen as hierarchy with some principles covering the full organization (handling cloud-based storage, for instance) or principles focused on specific departments (limiting data management permissions and functions to certain organizational units).

Review of strategic principles of tax administrations

When documenting data-related principles, the implementation team must consider any mandates or elements that might influence them, including external regulations such as national IT policies that reach the whole Public Sector, or internal aspects, including the tax administration mission, vision, and values as well as any strategic initiatives that might rely heavily upon data.

Definition of the data management and governance principles

The definition of principles should have the following characteristics:

- Must be contextualized.
- Cannot be ambiguous.
- Relate and complement each other.
- Be clear, specific, and understandable.
- Be stable over time.

The following standard attributes can be considered a minimum:

Table 8-1 Characteristics of principles.

Characteristics	Description
Name	Word or short set of words that represents the rule, must be clear, concrete, and easy to remember.
Statement	Succinctly and unambiguously describe the meaning of the principle.
Purpose/Justification	Adopting the idea has several advantages for the organization, and it is related to and aligned with fundamental tax administration principles.
Implication	When possible, the resources (human, financial, technological, etc.) and activities required to put the concept into practice; as well as the overall impact and consequences that the tax administration will experience as a result of adopting the principle.

Source: Prepared by the authors.

Review and formalization of principles

Once the set of principles have been stated, it must be approved by the competent entity or body within the tax administration, ideally by the Data Governance Steering Committee if already in place. As a good practice, a communication program should be put in place to disseminate the set of principles with all interested parties or even with all the tax administration.

8.2.2   Policies

Policies define scopes of action and their limits in terms of the right to access or control data. Policies can be defined with a broad institutional approach or targeting specific segments (e.g., organizational unit, process, information system, etc.).

Identify the need for policies

As a preliminary step before defining policies, it is necessary to gather all the tax administration stakeholders that are related to the management of data or directly or indirectly affected by problems derived from data problems to set a schedule. The process should consider the tax administration's current maturity has in terms of data governance.

Develop the policy

Start with recommendations aimed at resolving issues or those that amend current policies while developing or changing a policy. Make sure a policy does not conflict with any other policies at the same level before drafting it. The tax administration's corresponding data governance body (Data Governance Steering Committee or Council), or if one has not yet been established, the mechanisms established for other general policies within the tax administration, must approve the policy when it has been produced.

Implement policy

The implementation of a policy is carried out by actors that did not approve it, usually by IT units or data stewards within functional units.

A policy is more complex than a principle and its success depends on the coordination between the parties for its execution and the degree of compliance with all existent principles and policies.

Policy evaluation and maintenance

Policies should be assessed after they are put into place to see how well they are accomplishing their intended goals. This analysis can clearly show where the policy needs to be changed to mitigate risks and increase advantages while handling the data.

8.3   Data quality dimensions definition guide

This guide aims to provide a guideline to follow when selecting one or several dimensions of quality, trying to focus on the usefulness and value it generates, but not on the quantity. The following steps to define data quality dimensions could be followed.

Understand the need for the tax administration

Defining the quality dimensions to be adopted in the tax administration should consider the data problems that the tax administration faces, and the existing quality requirements. To achieve this, it is required to:

- Identify needs of the organization in terms of strategy, products, and services.
- Identify the data and information environment including data specifications (e.g., context, structure, meaning, and rules), technology, processes, and data lifecycle.

The metadata associated with data specifications must be considered as extensively as feasible. Otherwise, there is a chance that a poorly specified dimension will either be minimally useful or utterly useless.

Determine key stakeholders.

Identify existing data quality problems in the tax administration to be considered within the scope of a potential data quality project (e.g., duplication errors, gaps, unrealistic data, etc.).

It is recommended to categorize that data quality problems, thus grouping data problems that can be addressed together and measured by the same dimension.

Identify new or existing dimensions

List current data quality dimensions and identify sources of other commonly used dimensions (for example those describe in the DAMA-DMBoK2 and included in the Chapter 4).

Define data quality dimensions

Once the needs of the tax administration and the potential dimensions have been identified, the dimensions that will ultimately serve for data quality projects must be determined considering their relevance (what is measured actually matters and align with the strategy) and its feasibility (it can be measured at reasonable effort and cost).

Formalize and socialize the Data Quality Dimensions

Once the data quality dimensions have been defined, the tax administration must institutionalize them as a standard. The dimensions should then be channeled through the data governance structures so that they are agreed upon, approved, applied, and monitored throughout the organization.

8.4 Data Management Maturity Assessment Guide (data governance focus)

As presented in Chapter 5, there are several options for performing the data governance maturity assessment. In line with the previously recommended Stanford assessment, this guide illustrates how to use that assessment tool.

Overall, data governance maturity assessment tools are good and straightforward to use. However, there are three critical points of attention:

1. If there is interest in using another assessment tool, it is recommended that the choice considers the most significant factors for implementing data governance (long-term strategy, deficiencies, challenges, and opportunities for data use), alignment with eventual suppliers, and the framework.

2. Once the assessment tool is selected, keep it stable during the implementation of data governance. Continued use of the same tool facilitates assessment of deployment progress.

3. The tax administration must complete the data governance maturity assessment before starting the implementation of data governance. This assessment will be the baseline throughout the implementation.

8.4.1 Stanford Assessment

The Stanford assessment tool follows the structure of the Stanford maturity model, as shown in Table 8-2. There are six components divided into two groups of three components. The first group is called Foundational and comprises Awareness, Formalization, and Metadata. According to Firican (Firican, 2018), "the foundational aspects focus on measuring core data governance competencies and development of critical program resources." The second group is called Project and includes Stewardship, Data Quality, and Master Data. The components of the Project group, also according to Firican (Firican, 2018), "measure how effectively data governance concepts are applied in the course of projects." The assessment must examine the six components from the perspective of three dimensions: People, Policies, and Capabilities.

Table 8-2 Stanford - Wide view

FOUNDATIONAL	PEOPLE	POLICIES	CAPABILITIES
Awareness			
Formalization			
Metadata			

PROJECT	PEOPLE	POLICIES	CAPABILITIES
Stewardship			
Data Quality			
Master Data			

Source: Prepared by the authors

Permanent questions

Table 8-3 and Table 8-4 present questions that guide the application of the assessment, respectively, for the Foundational and Project components. The questions should conduct the assessment at all maturity levels for each element.

Table 8-3 Guiding questions for each component-dimension - Foundational

FOUNDATIONAL	PEOPLE	POLICIES	CAPABILITIES
Awareness	What awareness do people have about their role within the data governance Program?	What awareness is there of data governance policies, standards and best practices?	What awareness is there of data governance enabling capabilities that have been purchased or developed?
Formalization	How developed is the data governance organization and which roles are filled to support data governance activities?	To what degree are data Governance policies formally defined, implemented and enforced?	How developed is the toolset that supports data governance activities and how consistently is that toolset utilized?
Metadata	What level of crossfunctional participation is there in the development and maintenance of metadata?	To what degree are metadata creation and maintenance policies formally defined, implemented and enforced?	What capabilities are in place to actively manage metadata at various levels of maturity?

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-4 Guiding questions for each component-dimension - Project

PROJECT	PEOPLE	POLICIES	CAPABILITIES
Stewardship	What awareness do people have about their role within the data governance program?	What awareness is there of data governance policies, standards and best practices?	What awareness is there of data governance enabling capabilities that have been purchased or developed?
Data Quality	How developed is the data governance organization and which roles are filled to support data governance activities?	To what degree are data overnance policies formally defined, implemented and enforced?	How developed is the toolset that supports data governance activities and how consistently is that toolset utilized?
Master Data	To what degree has a formal master data management organization been developed and assigned consistente responsibilities across data domains?	To what degree are metadata creation and maintenance policies formally defined, implemented and enforced?	What capabilities are in place to actively manage metadata at various levels of maturity?

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2016)

Foundational components

Table 8-5 specifies the minimum requirements for each of the three factors in the Foundational set (OMES - Oklahoma Office of Management & Enterprise Services, 2016). The requirements are both qualitative and quantitative. Each tax administration can add quantitative requirements

appropriately to its situation, challenges, and goals. The quantitative requirements serve, above all, as evidence of the quantification of each component.
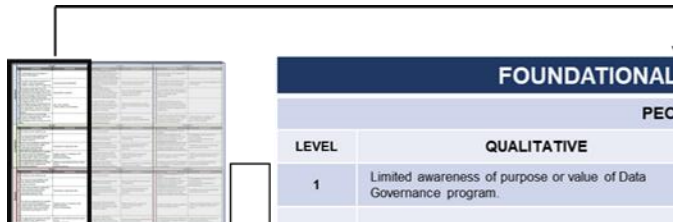
Table 8-5 Foundational components

Data Governance Foundational Components

Awareness

Level	People Qualitative	People Quantitative	Policies Qualitative	Policies Quantitative	Capabilities Qualitative	Capabilities Quantitative
1	Limited awareness of purpose or value of DG program.		Most existing data policies are undocumented and there may be inconsistent understanding of data policies within a department.		Little awareness of DG capabilities and technologies.	
2	Executives are aware of existence of program. Little knowledge of program outside upper management.	Training Sessions*attendees	Existing policies are documented but not consistently maintained, available or consistent between departments.	Policies documented by functional area, business subject area.	A small subset of the organization understands the general classes of DG capabilities and technologies.	Training sessions on DG capabilities and technologies.
3	Executives understand how DG benefits/impacts their portion of the organization, knowledge workers are aware of program. Executives actively promote DG within their groups.	Newsletters*recipients	Common data policies are documented and available through a common portal. Most stakeholders are aware of existence of data policies that may impact them.	Hits on Policy Management Content. Unique visitors on Policy Management Content.	A small subset of the organization is aware of the specific DG capabilities that are available at the organization.	
4	Executives understand long-term DG strategy and their part in it. Knowledge workers understand how DG impacts/benefits their portion of the organization. Executives actively promote DG beyond the immediate group.	Hits on DG website. Unique visitors on DG website.	All data policies are available through a common portal and stakeholders are actively notified whenever policies are added, updated or modified.	Number of stakeholders on RACI matrices by functional area, subject area.	A targeted audience has been identified and a significant portion of that audience is aware of the DG capabilities that are available at the organization.	
5	Both executives and knowledge workers understand their role in the long-term evolution of DG. Knowledge workers actively promote DG		A history of all data policies are maintained through a common portal and all stakeholders are made part of the policy development process.	Non-executive leadership participants in policy development.	A significant portion of the targeted audience understands how to utilize relevant DG capabilities that are available at the organization.	Training sessions on usage of DG technologies and capabilities (person*tech trained).

Formalization

Level	People Qualitative	People Quantitative	Policies Qualitative	Policies Quantitative	Capabilities Qualitative	Capabilities Quantitative
1	No defined roles related to DG.		No formal DG policies.		Classes of DG capabilities are not defined.	
2	DG roles and responsibilities have been defined and vetted with program sponsors.		High-level DG meta-policies are defined and distributed.	Meta-policies defined, documented and approved.	Classes of DG capabilities are defined and homegrown technical solutions are used within some organizational functions.	DG capabilities with solutions by functional area. Reuse of technical solutions by functional area.
3	Some roles are filled to support DG needs and participants clearly understand responsibilities associated with their roles.	Participants in approved roles.	Data policies around the governance of specific data are defined and distributed as best practices.	Best practices/standards/policies identified, documented and approved.	Homegrown technical solutions are adopted as best practices for some classes of capabilities and made available throughout the institution.	Capabilities approved as organization recommended solutions.
4	DG roles are organized into reusable schemas which are designed to support specific data and functional characteristics. There is broad (but inconsistent) participation in DG.	Program areas in compliance with defined schemas. Percent of roles filled.	Data policies become official organization data policies and compliance with approved data policies is audited.	Official data policies approved. Audits are done to ensure compliance.	All defined classes of DG capabilities have an available solution.	Usage of standard solutions by project. Uses of non-standard solutions by project.
5	5 DG organizational schemas are filled as defined, meet regularly and document activities.	Staff from each defined schema meets to plan. Minutes produced.	Compliance with official organization data policies is actively enforced by a governing body.	Number of exceptions to official data policies (lower is better).	All defined classes of DG capabilities are mandatory for assigned systems or critical data.	Use of non-standard solutions by project (lower is better).

Metadata

Level	People Qualitative	People Quantitative	Policies Qualitative	Policies Quantitative	Capabilities Qualitative	Capabilities Quantitative
1	No defined roles related to DG.		No formal DG policies		Classes of DG capabilities are not defined.	
2	DG roles and responsibilities have been defined and vetted with program sponsors.		High-level DG meta-policies are defined and distributed.	Meta-policies defined, documented and approved.	Classes of DG capabilities are defined and homegrown technical solutions are used within some organizational functions.	DG capabilities with solutions by functional area. Reuse of technical solutions by functional area.
3	Some roles are filled to support DG needs and participants clearly understand responsibilities associated with their roles.	Participants in approved roles.	Data policies around the governance of specific data are defined and distributed as best practices.	Best practices/standards/policies identified, documented and approved.	Homegrown technical solutions are adopted as best practices for some classes of capabilities and made available throughout the institution.	Capabilities approved as organization recommended solutions.
4	DG roles are organized into reusable schemas which are designed to support specific data and functional characteristics. There is broad (but inconsistent) participation in DG.	Program areas in compliance with defined schemas. Percent of roles filled.	Data policies become official organization data policies and compliance with approved data policies is audited.	Official data policies approved. Audits are done to ensure compliance.	All defined classes of DG capabilities have an available solution.	Usage of standard solutions by project. Uses of non-standard solutions by project.
5	DG organizational schemas are filled as defined, meet regularly and document activities.	Staff from each defined schema meets to plan. Minutes produced.	Compliance with official organization data policies is actively enforced by a governing body.	Number of exceptions to official data policies (lower is better).	All defined classes of DG capabilities are mandatory for assigned systems or critical data.	Usage of non-standard solutions by project (lower is better). No use of solution by project.

Source: OMES (OMES - Oklahoma Office of Management & Enterprise Services, 2016)

Table 8-5 is presented below divided in three parts to facilitate reading:

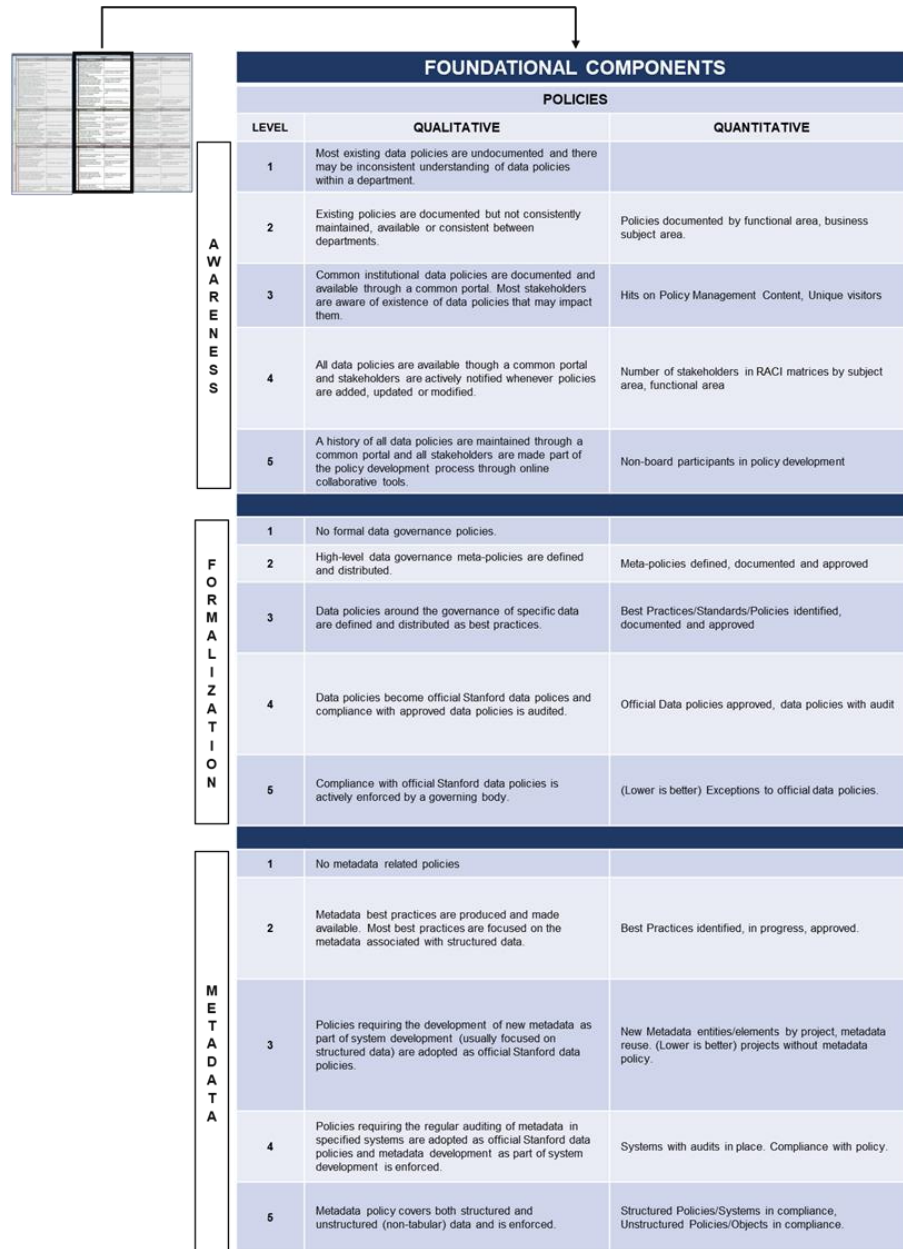*Table 8-6 Stanford data governance Foundational components - People*

FOUNDATIONAL COMPONENTS		
PEOPLE		
LEVEL	**QUALITATIVE**	**QUANTITATIVE**
AWARENESS		
1	Limited awareness of purpose or value of Data Governance program.	
2	Executives are aware of existence of program. Little knowledge of program outside upper management.	Training Sessions*attendees
3	Executives understand how program benefits/impacts their portion of the organization, knowledge workers are aware of program. Executives actively promote program within their groups.	Newsletters*recipients
4	Executives understand long-term program strategy and their part in it. Knowledge workers understand how the program impacts/benefits their portion of the organization. Executives actively promote program beyond the immediate group.	Hits on DG website, Unique visitors on DG website
5	Both executives and knowledge workers understand their role in the long-term evolution of the program. Knowledge workers actively promote program.	
FORMALIZATION		
1	There are no defined roles related to data governance.	
2	Data governance roles and responsibilities have been defined and vetted with program sponsors.	
3	Some roles are filled to support data governance needs and participants clearly understand responsibilities associated with their roles.	Participants in approved roles
4	Data governance roles are organized into reusable schemas which are designed to support specific data and functional characteristics. There is broad (but inconsistent) participation in Data Governance Organization.	Boards in compliance with defined schemas, % roles filled
5	Data governance organizational schemas are filled as defined, meet regularly and document activities.	Board meetings to plan, Minutes produced
METADATA		
1	Limited understanding of types and value of metadata.	
2	AS roles responsible for production of technical metadata on structured data are defined during system design.	(Lower is better) Projects without documented technical metadata
3	The responsibility for developing institutional business definitions and storing them in a central repository is assigned to and continually performed by subject matter experts.	Unique individuals creating/updating metadata. Qualitative rating of metadata.
4	Metadata collection/validation responsibilities assigned to named individuals for all projects.	Projects with metadata responsibility assignment
5	A dedicated metadata management group is created to strategically advance metadata capabilities and more effectively leverage existing metadata.	ROI of Metadata Competency Center

Source: Authors elaboration, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2016)

Table 8-7 Stanford data governance Foundational components - Policies

FOUNDATIONAL COMPONENTS		
POLICIES		
LEVEL	QUALITATIVE	QUANTITATIVE
AWARENESS		
1	Most existing data policies are undocumented and there may be inconsistent understanding of data policies within a department.	
2	Existing policies are documented but not consistently maintained, available or consistent between departments.	Policies documented by functional area, business subject area.
3	Common institutional data policies are documented and available through a common portal. Most stakeholders are aware of existence of data policies that may impact them.	Hits on Policy Management Content, Unique visitors
4	All data policies are available though a common portal and stakeholders are actively notified whenever policies are added, updated or modified.	Number of stakeholders in RACI matrices by subject area, functional area
5	A history of all data policies are maintained through a common portal and all stakeholders are made part of the policy development process through online collaborative tools.	Non-board participants in policy development
FORMALIZATION		
1	No formal data governance policies.	
2	High-level data governance meta-policies are defined and distributed.	Meta-policies defined, documented and approved
3	Data policies around the governance of specific data are defined and distributed as best practices.	Best Practices/Standards/Policies identified, documented and approved
4	Data policies become official Stanford data polices and compliance with approved data policies is audited.	Official Data policies approved, data policies with audit
5	Compliance with official Stanford data policies is actively enforced by a governing body.	(Lower is better) Exceptions to official data policies.
METADATA		
1	No metadata related policies	
2	Metadata best practices are produced and made available. Most best practices are focused on the metadata associated with structured data.	Best Practices identified, in progress, approved.
3	Policies requiring the development of new metadata as part of system development (usually focused on structured data) are adopted as official Stanford data policies.	New Metadata entities/elements by project, metadata reuse. (Lower is better) projects without metadata policy.
4	Policies requiring the regular auditing of metadata in specified systems are adopted as official Stanford data policies and metadata development as part of system development is enforced.	Systems with audits in place. Compliance with policy.
5	Metadata policy covers both structured and unstructured (non-tabular) data and is enforced.	Structured Policies/Systems in compliance, Unstructured Policies/Objects in compliance.

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-8 Stanford data governance Foundational components - Capabilities

FOUNDATIONAL COMPONENTS

CAPABILITIES

	LEVEL	QUALITATIVE	QUANTITATIVE
AWARENESS	1	There is little awareness of data governance capabilities and technologies.	
	2	A small subset of the organization understands the general classes of data governance capabilities and technologies	Training Sessions on DG capabilities and technologies
	3	A small subset of the organization is aware of the specific data governance capabilities that are available at Stanford.	
	4	A targeted audience has been identified and a significant portion of that audience is aware of the data governance capabilities that are available at Stanford.	
	5	A significant portion of the targeted audience understands how to utilize relevant data governance capabilities that are available at Stanford.	Training Sessions on usage of DG technologies and capabilities (person*tech trained)
FORMALIZATION	1	Classes of data governance capabilities are not defined.	
	2	Classes of data governance capabilities are defined and home-grown technical solutions are used within some institutional functions.	Data governance capabilities with solutions by functional org. Reuse of technical solutions by functional org.
	3	Home-grown technical solutions are adopted as best practices for some classes of capabilities and made available throughout the institution.	Capabilities approved as Stanford recommended solutions.
	4	All defined classes data governance capabilities have an available solution.	Usage of standard solutions by project. Uses of non-standard solutions by project
	5	All defined classes of data governance capabilities are mandatory for assigned systems or critical data.	(Lower is better): Uses of non-standard solutions by project, (Lower is better) no use of solution by project.
METADATA	1	Metadata is inconsistently collected and rarely consolidated outside of project artifacts.	
	2	Metadata templates are adopted to provide some consistency in content and format of captured metadata. Metadata is consolidated and available from a single portal. Capabilities focus on capture of metadata of structured content.	Metadata entities in portal. Edits by users to metadata
	3	The collection of metadata on structured content is automated and scheduled extracts are performed for selected systems.	Systems with automatic collection of metadata.
	4	A centralized metadata store becomes the primary location for all institutional metadata. Metadata is automatically collected from most RDBMS and vendor packaged systems.	(Lower is better) Systems not loading to metadata repository.
	5	A metadata solution provides a single point of access to federated metadata resources including both structured and unstructured data.	Unstructured objects linked to metadata repository. (Lower is better) systems out of compliance with load SLA.

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-9 Foundational components - example

Data Governance Foundational Components								
People			**Policies**			**Capabilities**		
	Qualitative	Quantitaive		Qualitative	Quantitaive		Qualitative	Quantitaive
Awareness								
1	Limited awareness of purpose or value of DG program.		1	Most existing data policies are undocumented and there may be inconsistent understanding of data policies within a department.		1	Little awareness of DG capabilities and technologies.	
2	Executives are aware of existence of program. Little knowledge of program outside upper management.	Training Sessions*attendees	2	Existing policies are documented but not consistently maintained, available or consistent between departments.	Policies documented by functional area, business subject area.	2	A small subset of the organization understands the general classes of DG capabilities and technologies.	Training sessions on DG capabilities and technologies.
3	Executives understand how DG benefits/ impacts their portion of the organization, knowledge workers are aware of program. Executives actively promote DG within their groups.	Newsletter recipients	3	Common data policies are documented and available through a common portal. Most stakeholders are aware of existence of data policies that may impact them.	Hits on Policy Management Content. Unique visitors on Policy Management Content.	3	A small subset of the organization is aware of the specific DG capabilities that are available at the organization.	
4	Executives understand long-term DG strategy and their part in it. Knowledge workers understand how DG impacts/benefits their portion of the organization. Executives actively promote DG beyond the immediate group.	Hits on DG website. Unique visitors on DG website.	4	All data policies are available through a common portal and stakeholders are actively notified whenever policies are added, updated or modified.	Number of stakeholders on RACI matrices by functional area, subject area.	4	A targeted audience has been identified and a significant portion of that audience is aware of the DG capabilities that are available at the organization.	
5	Both executives and knowledge workers understand their role in the long-term evolution of DG. Knowledge workers actively promote DG		5	A history of all data policies are maintained through a common portal and all stakeholders are made part of the policy development process.	Non-executive leadership participants in policy development.	5	A significant portion of the targeted audience understands how to utilize relevant DG capabilities that are available at the organization.	Training sessions on usage of DG technologies and capabilities (person*tech trained).
Formalization								
1	No defined roles related to DG.		1	No formal DG policies.		1	Classes of DG capabilities are not defined.	
2	DG roles and responsibilities have been defined and vetted with program sponsors.		2	High-level DG meta- policies are defined and distributed.	Meta-policies defined, documented and approved.	2	Classes of DG capabilities are defined and homegrown technical solutions are used within some organizational functions.	DG capabilities with solutions by functional area. Reuse of technical solutions by functional area.
3	Some roles are filled to support DG needs and participants clearly understand responsibilities associated with their roles.	Participants in approved roles.	3	Data policies around the governance of specific data are defined and distributed as best practices.	Best practices/standards/policies identified, documented and approved.	3	Homegrown technical solutions are adopted as best practices for some classes of capabilities and made available throughout the institution.	Capabilities approved as organization recommended solutions.
4	DG roles are organized into reusable schemas which are designed to support specific data and functional characteristics. There is broad (but inconsistent) participation in DG.	Program areas in compliance with defined schemas. Percent of roles filled.	4	Data policies become official organization data policies and compliance with approved data policies is audited.	Official data policies approved. Audits are done to ensure compliance.	4	All defined classes of DG capabilities have an available solution.	Usage of standard solutions by project. Uses of non-standard solutions by project.
5	5 DG organizational schemas are filled as defined, meet regularly and document activities.	Staff from each defined schema meets to plan. Minutes produced.	5	Compliance with official organization data policies is actively enforced by a governing body.	Number of exceptions to official data policies (lower is better).	5	All defined classes of DG capabilities are mandatory for assigned systems or critical data.	Use of non-standard solutions by project (lower is better).
Metadata								
1	No defined roles related to DG.		1	No formal DG policies		1	Classes of DG capabilities are not defined.	
2	DG roles and responsibilities have been defined and vetted with program sponsors.		2	High-level DG meta-policies are defined and distributed.	Meta-policies defined, documented and approved.	2	Classes of DG capabilities are defined and homegrown technical solutions are used within some organizational functions.	DG capabilities with solutions by functional area. Reuse of technical solutions by functional area.
3	Some roles are filled to support DG needs and participants clearly understand responsibilities associated with their roles.	Participants in approved roles.	3	Data policies around the governance of specific data are defined and distributed as best practices.	Best practices/standards/policies identified, documented and approved.	3	Homegrown technical solutions are adopted as best practices for some classes of capabilities and made available throughout institution.	Capabilities approved as organization recommended solutions.
4	DG roles are organized into reusable schemas which are designed to support specific data and functional characteristics. There is broad (but inconsistent) participation in DG.	Program areas in compliance with defined schemas. Percent of roles filled.	4	Data policies become official organization data policies and compliance with approved data policies is audited.	Official data policies approved. Audits are done to ensure compliance.	4	All defined classes of DG capabilities have an available solution.	Usage of standard solutions by project. Uses of non-standard solutions by project.
5	DG organizational schemas are filled as defined, meet regularly and document activities.	Staff from each defined schema meets to plan. Minutes produced.	5	Compliance with official organization data policies is actively enforced by a governing body.	Number of exceptions to official data policies (lower is better).	5	All defined classes of DG capabilities are mandatory for assigned systems or critical data.	Usage of non-standard solutions by project (lower is better). No use of solution by project.

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-9 is a sample of how to use Table 8-5: the Formalization component, in the Policies dimension, for example, received grade 2, which corresponds to the fulfillment of "high-level DG meta-policies are defined and distributed" (qualitative requirement) and "meta-policies defined, documented and approved" (quantitative requirement).

Project components

Table 8-10 specifies the minimum requirements for each of the three factors in the Project set. Likewise, requirements are both qualitative and quantitative.
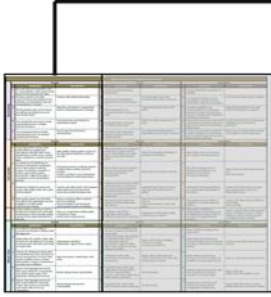
Table 8-10 Project components

Data Governance Project Components								
	People			**Policies**			**Capabilities**	
	Qualitative	Quantitaive		Qualitative	Quantitaive		Qualitative	Quantitaive
Stewardship								
1	Few well-defined stewardship roles or responsibilities. Data requirements driven by the development team.		1	Limited stewardship policies documented.		1	Limited stewardship capabilities are available.	
2	Business analysts drive data requirements during design process. Definition of stewardship roles and responsibilities is limited.	Projects with explicit data design.	2	Policies around stewardship defined within a functional area.	Functional areas with policy. Functional data entities with policy.	2	A centralized location exists for consolidation of and/or access to stewardship related documentation.	Count of policies (by status) in registry.
3	All stewardship roles and structures are defined and filled but are still functionally siloed.	Stewards, participants in stewardship boards, stewardship board meetings.	3	Stewardship policies are consistent between functions and areas.	Organizational data entities with policy.	3	Workflow capabilities are implemented for the vetting and approval of institutional definition, business metadata and stewardship related documentation.	Organizational definitions through process (completed, in progress).
4	The stewardship structures include representatives from multiple business functions.	Functional areas represented on stewardship boards.	4	Stewardship teams self-audit compliance with policies.	Audits and audit compliance are in place.	4	Stewardship dashboards report data quality levels and data exceptions to support the auditing of stewardship effectiveness.	Dashboards by function program area. Qualitative score included on dashboard.
5	The stewardship board includes representatives from all relevant institutional functions.	Boards with AS and business representation.	5	Compliance with stewardship policies are enforced for key institutional data.	Key organizational data without stewardship policies (lower is better).	5	A common stewardship dashboard enables managed issue remediation as part of data quality reporting and data exception reporting.	Data issues are reported and resolved. Time it takes to resolve data issues.
	People			**Policies**			**Capabilities**	
	Qualitative	Quantitaive		Qualitative	Quantitaive		Qualitative	Quantitaive
Data Quality								
1	Individuals perform ad hoc data quality efforts as needed and manually fix identified data issues. Identification of data issues is based off its usability for a specific business task.	Data quality implies quality in terms of formally defined definitions of fit-for- use data.	1	Data quality efforts are infrequent and driven by specific business needs. These efforts are usually large one-time data cleansing efforts.	Data cleansing efforts identified, in progress or completed.	1	Data quality is done on an ad hoc basis usually using SQL and Excel.	
2	A small group of individuals are trained in and perform profiling to assess data quality of existing systems to establish a baseline or justify a data quality project. Downstream usage of the data is considered in issue identification process.	Individuals trained in profiling, systems profiled, tables profiled, elements profiled. Profiles resulting in recommendations, recommendations spawning projects.	2	Best practices have been defined for some data quality related activities and followed inconsistently.	Data quality best practices defined.	2	Basic data profiling tools are adopted and available for use anywhere in the system development lifecycle.	Data profiles by system and functional area. Rows are profiled.
3	People are assigned to assess and ensure data quality within the scope of each project.	Projects with data quality roles assigned. Data quality fixes at project level. Issues documented and approved.	3	Profiling and development of data quality standards are adopted as part of the standard application development lifecycle and become scheduled activities on project plans.	Application development projects without profiling effort (lower is better).	3	Data quality reporting capabilities are implemented and available to any system.	Systems with data quality reporting, approved elements reported on. Raw quality metrics.
4	Data quality experts are identified throughout the organization and are engaged in all data quality improvement projects.	Systems analyzed, tables analyzed, elements analyzed. Recommendations proposed and spawning data quality remediation.	4	Data quality best practices are adopted as official organization data policies.	Approved organizational data quality policies. Data quality policies in place with audits.	4	Data quality issue remediation is integrated into quality reporting platform.	Systems with data quality remediation functionality. Issues resolved.
5	A data quality competency center is funded and charged with continually assessing and improving data quality outside of the system development lifecycle.	Return on Investment of data quality competency center. System team endorsements.	5	Compliance with official organization data quality is tracked and reported on centrally.	Exceptions to official data quality policies (lower is better).	5	Data quality remediation is implemented on both data at rest (in databases) and data in flight (in ETL and as messages between systems).	Systems without data quality reporting, and/or remediation (lower is better). Interfaces without reporting and/or remediation (lower is better).
	People			**Policies**			**Capabilities**	
	Qualitative	Quantitaive		Qualitative	Quantitaive		Qualitative	Quantitaive
Master Data								
1	Inconsistent understanding of concepts and benefits of Master Data Management.		1	No formal policies defining what data are considered institutional data.		1	There is limited management of master data.	
2	Stakeholders for specific master data domains are identified and consulted to develop basic definition and model of master data.	Stakeholders identified. Stakeholders' agreements in place.	2	Institutional master data domains are defined and the systems storing master data are documented. Usage of master data in these systems is actively being documented.	Master data entities identified. Functions consulted. Perspectives identified.	2	Master data are identified and manually managed and provisioned via extracts, file transfers or manual uploads.	Systems using master data by transport method.
3	Owners of institutional master data are identified and drive resolution of various perspectives of master data. Owners establish and run master data boards to support maintenance and data issue mediation.	Approved owners, stakeholders with input.	3	Institutional master data perspectives are resolved and documented.	Master data models approved. Distinct perspectives of master data entities (lower is better).	3	Master data are provisioned through services but management capabilities are still largely manual.	Systems using master data via services.
4	Master Data Management boards take responsibility for reviewing the use of their master data in the application development process.	Boards taking review responsibility.	4	Compliance with master data usage policies and standards is enforced. Synchronization frequency with master data hub at system owner's discretion.	Results of audit.	4	Multiple single domain master data hubs handle provisioning and management of master data.	Master data hubs. Master data hub capability score.
5	Master Data Management boards take responsibility for enforcing master data policies around their own master data across the organization.	Boards taking enforcement responsibility.	5	Compliance with master data synchronization policy is enforced.	Results of audit.	5	Multidomain master data hub handles all provisioning and management of master data.	Master data hubs (lower is better). Master data hub score (lower is better).

Source: OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-10 is presented below in three parts to facilitate reading:

Table 8-11 Stanford data governance Project components - People

		PROJECT COMPONENTS	
		PEOPLE	
	LEVEL	QUALITATIVE	QUANTITATIVE
STEWARDSHIP	1	Almost no well defined data governance or stewardship roles or responsibilities. Data requirements are driven by the application development team.	
	2	Business analysts drive data requirements during design process. Definition of stewardship roles and responsibilities is limited.	Projects with explicit Data Design.
	3	All stewardship roles and structures are defined and filled but are still functionally siloed.	Stewards, Participants in Stewardship boards, Stewardship board meetings.
	4	The stewardship structures include representatives from multiple business functions.	Functional Areas Represented on Stewardship Board
	5	The stewardship board includes representatives from all relevant institutional functions including AS.	Boards with AS and Business Representatio
DATA QUALITY	1	Individuals perform ad hoc data quality efforts as needed and manually fix data identified data issues. Identification of data issues is based off of its usability for a specific business task.	*data quality implies quality in terms of formally defined definition of fit-for-use data.
	2	A small group of individuals are trained in and perform profiling to assess data quality of existing systems to establish a baseline or justify a data quality project. Down stream usage of the data is considered in issue identification process.	Individuals trained in profiling, systems profiled, tables profiled, elements profiled. Profiles resulting in recommendations, recommendations spawning projects.
	3	People are assigned to assess and ensure data quality within the scope of each project.	Projects with DQ roles assigned. "DQ fixes" at project level. Issues documented and approved.
	4	Data quality experts are identified throughout the institution and are engaged in all data quality improvement projects.	Systems analyzed, tables analyzed, elements analyzed. Recommendations proposed, Recommendations spawning DQ remediation.
	5	A data quality competency center is funded and charged with continually assessing and improving data quality outside of the system development lifecycle.	ROI of DQCC. System Team endorsements.
MASTER DATA	1	Inconsistent understanding of concepts and benefits of Master Data Management.	
	2	Stakeholders for specific master data domains are identified and consulted to develop basic definition and model of master data	Stakeholders identified, stakeholders' agreement.
	3	Owners of institutional master data are identified and drive resolution of various perspectives of master data. Owners establish and run master data boards to support maintenance and data issue mediation.	Approved owners, stakeholders with input.
	4	Master Data Management boards take responsibility for reviewing the use of their master data in the application development process.	Boards taking review responsibility.
	5	Master Data Management boards take responsibility for enforcing master data policies around their own master data across the institution.	Boards taking enforcement responsibility.

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-12 Stanford data governance Project components - Policies

	PROJECT COMPONENTS	
	POLICIES	
LEVEL	QUALITATIVE	QUANTITATIVE
STEWARDSHIP		
1	Limited stewardship policies documented.	
2	Policies around stewardship defined within a functional area or subject area.	Functional areas with policy, Functional Data Entities with policy
3	Stewardship policies are consistent between functions and subject areas.	Institutional Data Entities with policy
4	Stewardship teams self-audit compliance with policies.	Audits, Audit Compliance
5	Compliance with stewardship polices are enforced for key institutional data.	(Lower is better) Key Institutional Data without stewardship policie
DATA QUALITY		
1	Data quality efforts are infrequent and driven by specific business needs. These efforts are usually large one-time data cleansing efforts.	Data Cleansing Efforts identified, in progress, completed.
2	Best practices have been defined for some data quality related activities and followed inconsistently.	Data Quality Best Practices defined.
3	Profiling and development of data quality standards are adopted as part of the standard application development lifecycle and become scheduled activities on project plans.	(Lower is better) Application development projects without profiling effort.
4	Data quality best practices are adopted as official Stanford data polices	Approved Stanford DQ Policies, data quality policies with audit
5	Compliance with official Stanford data quality policies is tracked and reported on centrally.	(Lower is better) Exceptions to official data quality policies.
MASTER DATA		
1	No formal policies defining what data is considered institutional master data.	
2	Institutional master data domains are defined and the systems storing master data and is documented. Usage of master data in these systems is actively being documented.	Master Data Entities Identified. Functions consulted. Perspectives Identified.
3	Institutional master data perspectives resolved and documented.	Master Data Models approved. (Lower is better) distinct perspectives of master data entities.
4	Compliance with master data usage policies and standards is enforced. Synchronization frequency with master data hub at system owner's discretion.	Results of audit.
5	Compliance with master data synchronization policy is enforced.	Results of audit.

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-13 Stanford data governance Project components - Capabilities

PROJECT COMPONENTS			
CAPABILITIES			
	LEVEL	QUALITATIVE	QUANTITATIVE
STEWARDSHIP	1	Limited stewardship capabilities are available.	
	2	A centralized location exists for consolidation of and/or access to stewardship related documentation.	Count of policies (by status) in registry
	3	Workflow capabilities are implemented for the vetting and approval of institutional definition, business metadata and approval of other stewardship related documentation.	Institutional Definitions through process (completed, in progress)
	4	Stewardship dashboards report data quality levels and data exceptions to support the auditing of stewardship effectiveness.	Dashboards by function, subject area. Qualitative score on dashboard.
	5	A common stewardship dashboard enables managed issue remediation as part of data quality reporting and data exception reporting.	Issues reported, Issues resolved, Time to resolution.
DATA QUALITY	1	Data quality is done on ad hoc basis usually using SQL and Excel.	
	2	Basic data profiling tools are adopted and available for use anywhere in the system development lifecycle.	Data Profiles by system, functional area. Rows profiled.
	3	Data quality reporting capabilities are implemented and available to any Stanford system.	Systems with data quality reporting, approved elements reported on. Raw Quality Metrics.
	4	Data quality issue remediation is integrated into quality reporting platform.	Systems with data quality remediation functionality. Issues resolved.
	5	Data quality remediation is implemented on both data at rest (in databases) and data in flight (in ETL and as messages between systems).	(Lower is better) Systems without DQ reporting, remediation. Interfaces without reporting, remediation.
MASTER DATA	1	There is limited management of master data.	
	2	Master data are identified and manually managed and provisioned via extracts, file transfers or manual uploads.	Systems using master data by transport method
	3	Master data are provisioned through services but management capabilities are still largely manual.	Systems using master data via services.
	4	Multiple single domain master data hubs handle provisioning and management of master data.	Master Data Hubs. Master data hub capability score.
	5	Multidomain master data hub handles all provisioning and management of master data.	(Lower is better) Master data hubs, master data hub score.

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Table 8-14 is a sample of the application of Table 8-10. The Data Quality component, in the People dimension, for example, received grade 3, which corresponds to the fulfillment of "People are assigned to assess and ensure data quality within the scope of each project" (qualitative requirement) and "Projects with data quality roles assigned - Data quality fixes at project level - Issues documented and approved" (quantitative requirements).

Table 8-14 Project Components - Example

Data Governance Project Components					
People		**Policies**		**Capabilities**	
Qualitative	Quantitaive	Qualitative	Quantitaive	Qualitative	Quantitaive
1 Few well-defined stewardship roles or responsibilities. Data requirements driven by the development team.		Limited stewardship policies documented.		Limited stewardship capabilities are available.	
2 Business analysts drive data requirements during design process. Definition of stewardship roles and responsibilities is limited.	Projects with explicit data design.	Policies around stewardship defined within a functional area.	Functional areas with policy. Functional data entities with policy.	A centralized location exists for consolidation and/or access to stewardship related documentation.	Count of policies (by status) in registry.
3 All stewardship roles and structures are defined and filled but are still functionally siloed.	Stewards, participants in stewardship boards, stewardship board meetings. **2**	Stewardship policies are consistent between functions and areas.	Organizational data entities with policy. **3**	Workflow capabilities are implemented for the vetting and approval of institution definitions, business metadata and stewardship related documentation.	Organizational definitions through process (completed, in progress). **2**
4 The stewardship structures include representatives from multiple business functions.	Functional areas represented on stewardship boards.	Stewardship teams self-audit compliance with policies.	Audits and audit compliance are in place.	Stewardship dashboards report data quality levels and data exceptions to support the auditing of stewardship effectiveness.	Dashboards by function program area. Qualitative score included on dashboard.
5 The stewardship board includes representatives from all relevant institutional functions.	Boards with AS and business representation.	Compliance with stewardship policies are enforced for key institutional data.	Key organizational data without stewardship policies (lower is better).	A common stewardship dashboard enables managed issue remediation as part of data quality reporting and data exception reporting.	Data issues are reported and resolved. Time it takes to resolve data issues.

People		**Policies**		**Capabilities**	
Qualitative	Quantitaive	Qualitative	Quantitaive	Qualitative	Quantitaive
1 Individuals perform ad hoc data quality efforts as needed and manually fix identified data issues. Identification of data issues is based off its usability for a specific business task.	Data quality implies quality in terms of formally defined definitions of fit-for- use data.	Data quality efforts are infrequent and driven by specific business needs. These efforts are usually large one-time data cleansing efforts.	Data cleansing efforts identified, in progress or completed.	Data quality is done on an ad hoc basis usually using SQL and Excel.	
2 A small group of individuals are trained in and perform profiling to assess data quality of existing systems to establish a baseline or justify a data quality project. Downstream usage of the data is considered in issue identification process.	Individuals trained in profiling, systems profiled, tables profiled, elements profiled. Profiles resulting in recommendations, recommendations spawning projects.	Best practices have been defined for some data quality related activities and followed inconsistently.	Data quality best practices defined.	Basic data profiling tools are adopted and available for use anywhere in the system development lifecycle.	Data profiles by system and functional area. Rows are profiled.
3 People are assigned to assess and ensure data quality within the scope of each project.	Projects with data quality roles assigned. Data quality fixes at project level. Issues documented and approved. **3**	Profiling and development of data quality standards are adopted as part of the standard application development lifecycle and become scheduled activities on project plans.	Application development projects without profiling effort (lower is better). **3**	Data quality reporting capabilities are implemented and available to any system.	Systems with data quality reporting, approved elements reported on. Raw quality metrics. **1**
4 Data quality experts are identified throughout the organization and are engaged in all data quality improvement projects.	Systems analyzed, tables analyzed, elements analyzed. Recommendations proposed and spawning data quality remediation.	Data quality best practices are adopted as official organization data policies.	Approved organizational data quality policies. Data quality policies in place with audits.	Data quality issue remediation is integrated into quality reporting platform.	Systems with data quality remediation functionality. Issues resolved.
5 A data quality competency center is funded and charged with continually assessing and improving data quality outside of the system development lifecycle.	Return on investment of data quality competency center. System team endorsements.	Compliance with official organization data quality is tracked and reported on centrally.	Exceptions to official data quality policies (lower is better).	Data quality remediation is implemented on both data at rest (in database) and data in flight (in ETL and as messages between systems).	Systems without data quality reporting, and/or remediation (lower is better). Interfaces without reporting and/or remediation (lower is better).

People		**Policies**		**Capabilities**	
Qualitative	Quantitaive	Qualitative	Quantitaive	Qualitative	Quantitaive
1 Inconsistent understanding of concepts and benefits of Master Data Management.		No formal policies defining what data are considered institutional master data.		There is limited management of master data.	
2 Stakeholders for specific master data domains are identified and consulted to develop basic definition and model of master data.	Stakeholders identified. Stakeholders' agreements in place.	Institutional master data domains are defined and the systems storing master data are documented. Usage of master data in these systems is actively being documented.	Master data entities identified. Functions consulted. Perspectives identified.	Master data are identified and manually managed and provisioned via extracts, file transfers or manual uploads.	Systems using master data by transport method.
3 Owners of institutional master data are identified and drive resolution of various perspectives of master data. Owners establish and run master data boards to support maintenance and data issue mediation.	Approved owners, stakeholders with input. **2**	Institutional master data perspectives are resolved and documented.	Master data models approved. Distinct perspectives of master data entities (lower is better). **2**	Master data are provisioned through services but management capabilities are still largely manual.	Systems using master data via services. **1**
4 Master Data Management boards take responsibility for reviewing the use of their master data in the application development process.	Boards taking review responsibility.	Compliance with master data usage policies and standards is enforced. Synchronization frequency with master data hub at system owner's discretion.	Results of audit.	Multiple single domain master data hubs handle pro-visioning and management of master data.	Master data hubs. Master data hub capability score.
5 Master Data Management boards take responsibility for enforcing master data policies around their own master data across the organization.	Boards taking enforcement responsibility.	Compliance with master data synchronization policy is enforced.	Results of audit.	Multidomain master data hub handles all provisioning and management of master data.	Master data hubs (lower is better). Master data hub score (lower is better).

Source: Prepared by the authors, based on the OMES publication (OMES - Oklahoma Office of Management & Enterprise Services, 2020)

Joining the numbers

Table 8-15 Assessment grades transcription - example

Foundational	People	Policies	Capabilities	Average
Awareness	1	3	1	**1.7**
Formalization	2	2	1	**1.7**
Metadata	2	2	2	**2.0**
Average	1.7	2.3	1.3	
Project	**People**	**Policies**	**Capabilities**	**Average**
Stewardship	2	3	2	**2.3**
Data Quality	3	3	1	**2.3**
Master Data	2	2	1	**1.7**
Average	2.3	2.7	1.3	

Source: Prepared by the authors

Table 8-15 contains the transcription of the values recorded in Table 8-9 and Table 8-14 (examples) for the components and dimensions. It also shows the average values calculated for each component and dimension.

The values in Table 8-15 can be used for other calculations, such as the average of the Foundational components (in the example, the standard is 1.8) and the Project components (in the model, the average is 2.1). In the same way, the average value of maturity, in the example summarized in the Table 8-15, is 1.9 or almost REPEATABLE in the DAMA-DMBoK2 maturity levels. Note that calculation can be weighted to adapt to the tax administration's reality. However, the mentioned average values only serve as a general indicator because the specific values of each component are more useful for the evaluation and planning of actions.

The diagram in Figure 8-2 graphically records the average value of each component, as shown in Table 8-15.
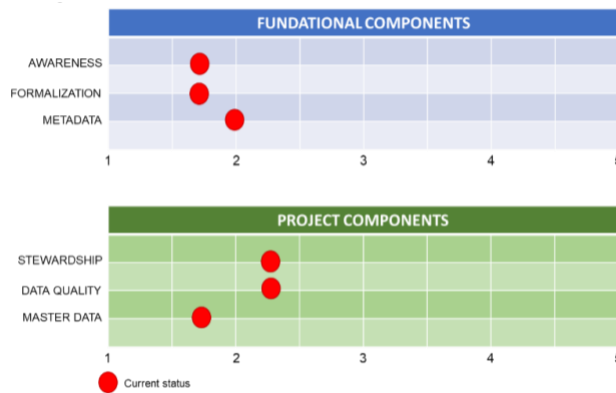
Figure 8-2 Average values for components. Source: Prepared by the authors.

The diagram in Figure 8-3Figure 8-1 shows the current situation (red) and, still considering the example, a hypothetical goal to be achieved by the following assessment cycle (grey).
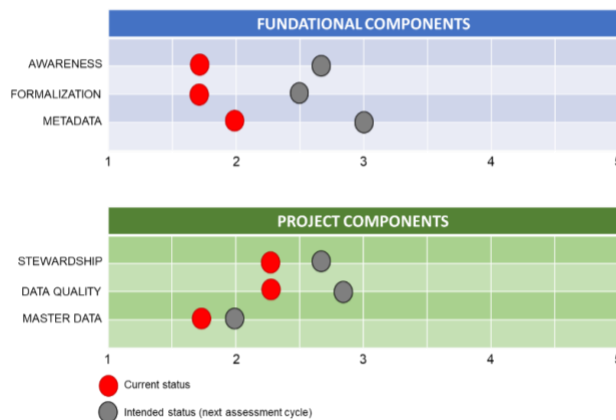
Figure 8-3 Current and Goal values. Source: Prepared by the authors.

Incidentally, the goal for the next assessment cycle should be sensible, challenging, and doable. These characteristics help in planning the actions that will lead to achieving goals.

8.4.2 Other assessment tools - examples

Numerous data governance maturity models and assessment tools are available, as was already indicated. Some resources are available online and are explored below to demonstrate the potential use of alternative assessment tools.

The first is in the article published by Marchildon et al. (Marchildon, 2018), which presents an assessment tool developed using the Design Science Research technique, which "aims to create and evaluate artifacts and tools to solve problems identified in organizations." The result is exciting and is in the attachment of the article. The tool covers 11 dimensions and 72 questions. At the paper's conclusion, the authors say, "at the beginning of this paper we set out to design an artifact that would help organizations assess their own level of data maturity governance. Specifically, our objective regarding this artifact was threefold. First, our artifact needed to help organizations know, before the realization of their data governance initiatives, which data governance processes, policies, practices and/or structure should be developed and prioritized. Second, our artifact needed to help organizations evaluate, after the implementation of their data governance initiatives, if those initiatives allowed them to evolve in terms of data governance maturity. Third, our artifact needed to be aligned with existing data governance maturity frameworks" (Marchildon, 2018).

The second assessment tool belongs to OvalEdge (Varshney, 2021), a leading company in data governance, which works with "Progressive Data Governance." OvalEdge states that a data governance model "is a blueprint that defines the roles, responsibilities, policies, and procedures of the data governance initiative, so everyone in the organization knows the plan and is in agreement ... traditional governance follows the DAMA ... (that) has been in the data governance game for over three decades, and they have done some incredible things during that time" (Varshney, 2020). The tool is an Excel spreadsheet with seven tabs – one with instructions, five with 69 questions, and one with results (radar chart) (Varshney, 2021). Each question has a drop-down menu, which facilitates the assessment.

8.5   Data Governance Roles Designation Guide

The purpose of this guide is to define a set of steps that help organize the mapping and designation of roles and responsibilities as a guide for its application within the data governance framework.

Delineate Data Governance Roles

To define the roles that will participate in the data governance framework, it is ideal to rely on the structuring of the data governance operating model. The roles are oriented to organic structures such as collegiate bodies at different levels (e.g., Data Governance Steering Committee or Council), as well as roles related to people (e.g., Data Governance Officer and Data Owner).

Determine activities/functions

As part of the definition of roles, the set of activities and functions related to data governance processes, e.g., manage meetings of data governance bodies, define data quality teams, formulate the data strategy, etc. Depending on the organization and the processes of the governance model,

tax administration must describe functions and responsibilities that those roles must maintain as part of the execution of the management model.

Identify activities/functions gaps

The tax administration must check the expected set of actions with the actual ones after compiling the list, and any discrepancy between what is being done and what should be done must be noted. This activity aids in directing the tasks or functions that the management model is missing.

8.5.1   Assignment of Roles

The assignment of existing roles to officers and officials of the tax administration should be focused on:

- The profile that each person has (e.g., technological areas, functional areas).
- Organizational levels (e.g., strategic, tactical, operational).
- Information to govern (e.g., taxpayer information, tax compliance information, etc.).
- Knowledge of the area, processes, or data domain.

When there are actors with significant or unique knowledge, an effort should be made to assure their involvement with administrator roles.

8.5.2   Use of RACI Matrix

As a good practice and depending on the need for the data governance model, roles and responsibilities (general or detailed) can be mapped through a RACI matrix. The RACI matrix is a tool that enables a person to see and recognize in a structured manner which (role) reacts to what (responsibility) is in terms of activities, tasks, achievements, and deliverables.

The RACI matrix maintains the following acronym:

- R - Responsible

  People who are responsible for executing and delivering the assigned or requested task.

- A - Accountable

  Actor with authority, who is in charge and accountable for the execution or request. They are not necessarily the ones who execute. Ideally, this category should be exercised by a particular role, avoiding power conflicts and maintaining a role with sufficient authority to exercise the necessary control.

- C- Consultant

  Actors and experts in a topic or the activity, who can provide valuable information to the execution of the task. They are an important support figure for the execution of the task.

- I - Informed

  People who will be informed about the different stages of the execution of the task.

Through the creation of the RACI matrix, the set of activities to be developed in the data governance model is clearly defined, and the role (RACI code) of each of the roles is assigned to each of the tasks.

If there is any potential for responsibility overlap, it may be found using the RACI matrix and managed by the data governance bodies.

Table 8-16 Data governance RACI matrix

	Data Governance Steering Committee	Data Governance Council	Data Governance Office	Data Owner	…
Manage meetings of Data Governance bodies	R	R	A	I	…
Leadership Data Quality teams	I	I	I	A	…
Plan Data Strategy	C	R	A	I	…
Data Issue Resolution/Definitions	I	R	I	A	…
Data remediation	I	I	R	A	…
Promote best data management practices	I	R	A	I	
…		…	…	…	…

Source: Prepared by the authors

8.6   Data Governance Stakeholder Identification Guide

Prior to the start of any initiative or data governance project within the tax administration, the list of stakeholders must be determined including the relationship of each actor with its initiative and their commitment.

8.6.1   Identify Stakeholders

Determine the stakeholders in data management and governance; that is, people who are or may be directly or indirectly affected; who have some degree of influence or interest. The tax administration can do this activity by brainstorming potential candidates.

Stakeholders can come from any level of the tax administration (strategic, tactical, or operational), regardless of the data management or governance initiative.

8.6.2   Analyze and Map Stakeholders

When preparing the list of stakeholders, the relevance of each actor for a data project must considered. Depending on the number of people, interested parties can be grouped to common interests or levels of influence. Consider the following:

- What kind of people are they?
- How much can they contribute or influence the project?
- What is the level of commitment?
- What concerns do they have?
- What results do they expect?
- What is their level of participation?
- What is their availability?

Once the stakeholders have been identified and analyzed, a map or a chart could be used to document their effective participation in the process. The chart may consider the following:

- Interested – Name of the individual or group
- Concerns – Declaration of the concerns of the interested party
- Outcome – Expected result for the interested party.
- Influence – What is the level of influence towards the project
- Intervention – What is the level of participation or interest in the project

Table 8-17 Stakeholders' map.

Interested	Concerns	Outcome(s)	Influence	Intervention

Source: Prepared by the authors

8.7 Practical Implementation Guide

This guide identified a sequence of practical steps that could be used by a tax administration when adopting data governance from scratch.

Concepts and details regarding the proposed steps are described previously in this document, particularly in chapters 2 to 6.

Contextualize and prepare the organization

- Responsible units and individuals:

  - Project team

  - Data architects

  - Corporate planning bodies

- Activities

  - Identify all interested parties (units and individuals) and their current responsibilities.

- Prepare an assessment of existing capacities, capabilities, and major data related issues.

- Prepare an assessment of the maturity level of data governance. (See section 8.4)

- Identify the strategy and requirement of the organization and align the objectives of the project to those needs.

- Develop an implementation strategy, including a project plan, establishing the roadmap and identifiable milestones to be achieved.

- Develop a Communication Plan and start its first steps to promote the initiative and its benefits.

Define a data governance scheme

- Responsible units and individuals

  - Project team

  - Data Government Council

- Activities

  - Identify individuals in the units with the largest stakes (major current data problems and biggest potential gains).

  - Evaluate the knowledge and competences of the candidates and train them with basic data governance related knowledge

  - Constitute the Data Governance Council

  - If the light data governance model approach is not selected continue by creating of the steering committee, the Data Governance Office. Identify and train data stewards within each relevant unit.

  - Create awareness of data quality importance in all the organization with focus in areas that deal with data intensity processes.

  - Prepare a data literacy program for all officers and officials in the organization and start its execution.

Develop data governance capacities

- Responsible units and individuals

  - Project team (light data governance model)

  - Steering Committee, Data Governance Council and Data Governance Office

- Activities

  - Identify de gap of capabilities of each candidate data steward

  - Train the candidates in the lacking competences and practices

- Prepare the whole organization in data literacy aspects

Implement the data governance model

- Responsible units and individuals

  - Project team (light data governance model)

  - Steering Committee, Data Governance Council and Data Governance Office

  - Data owners, data custodians and other data stewards.

- Activities

  - Identify the data that should be governed

  - Define and prioritize data domains and data owners

  - Institutionalize the involvement of data governance bodies in data management related decisions.

  - Assure the alignment of the data governance strategy with the tax administration strategy.

  - Consolidate the participation of data management roles in different processes within the tax administration, including the design and execution of new projects.

  - Verify compliance of data policies during software development or acquisition, software maintenance, deployment, and production.

  - Coordinate with legal and risk management units the adjustment to regulatory guidelines.

  - Execute compliance checks for data quality, data modeling standards, data architecture, data privacy regulations, data disposal, etc.

  - Manage the ongoing change.

  - Periodically apply the maturity assessment of data governance.

  - Define and monitor metrics for data governance control. Table 8-18 include a set of metrics that respond to certain capabilities presented in chapter 4.

Table 8-18 Data Governance Control Metrics

Data Governance Capability	Metric
Data Quality Control Management	Percentage of complete data (for a given data domain / data element)
	Number of certified quality data sources
Metadata Management	Percentage of data sources inventoried
	Number of data domains raised
	Number of agreed business terms

Data Governance Capability	Metric
Data Risk Management	Number of data privacy controls implemented Number of privacy risks identified vs. mitigated
Data Governance Training	Percentage of courses taken in data management and analytics Number of managers trained in data strategy Number of people trained in data storytelling
Data Policy Management	Percentage of data quality policies complied with in data projects Percentage of data architecture standards complied
Issue Management	Number of data issues identified Number of resolved vs. unresolved data incidents

Source: Prepared by the authors

Tables

Figures

Bibliography

Ackoff, R. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis, 16*, pp. 3-9.

Addagada, T. (2016, October 11). *Applying COBIT principles to call your Data Governance, a success.* Retrieved July 28, 2022, from www.dattamza.org: https://www.dattamza.org/dattamza-blog/applying-cobit-principles-to-call-your-data-governance-a-success

Addagada, T. (2018, October 22). *Data Governance and the Maturity Assessment Model.* Retrieved July 27, 2022, from Dataversity: https://www.dataversity.net/data-governance-maturity-assessment-model/

Alation. (2020). *Active Data Governance Methodology.* Retrieved July 26, 2022, from www.alation.com: https://www.alation.com/resource-center/whitepapers/data-governance-methodology

Allen, M., & Cervo, D. (2015). *Multi-Domain Master Data Management,.* Amsterdam, NL: Elsevier Inc.

Ancick, T. (2022, June 6). *Data Governance Roadmap.* Retrieved July 26, 2022, from www.ovaledge.com: https://www.ovaledge.com/blog/data-governance-roadmap

Arias, I., & Zambrano, R. (2020, October). Technology and Tax Administration Control in Latin America. *IBFD - Bulletin for International Taxation.*

Askham, N. (2022, March 4). *Do you know what is in a data governance framework?* Retrieved July 26, 2022, from www.nicolaaskham.com: https://www.nicolaaskham.com/blog/2022/3/4/do-you-know-what-is-in-a-data-governance-framework?rq=framework

Askham, N. (2022, February 18). *Five Common Data Governance Misconceptions.* Retrieved July 26, 2022, from www.nicolaaskham.com: https://www.nicolaaskham.com/blog/2022/2/17/nxts01aohpqc3kyvby72azv9kep0j7

Baltassis, E., Coulin, A.-D., Gourévitch, A., Khendek, Y., & Quarta, L. (2020, December 30). *Rough road to data maturity.* Retrieved July 27, 2022, from www.bcg.com: https://www.bcg.com/pt-br/publications/2019/rough-road-to-data-maturity

Barreix, A., & Zambrano, R. (2018). Electronic Invoicing in Latin America: Process and Challenges . In *Electronic Invoice in Latin-America. English summary of the Spanish Document.* Washington D.C.; Panama City: IADB; CIAT.

Basker, S. S. (2016, May 16). *Review of data maturity models.* Retrieved July 28, 2022, from www.dataorchard.org.uk: https://www.dataorchard.org.uk/resources/review-of-data-maturity-models

Benthien, C. (2022). *Six Steps for Developing a Strategy Roadmap.* Retrieved July 26, 2022, from www.jibility.com: https://www.jibility.com/six-steps-develop-strategy-roadmap/

Bersin, J., & Zao-Sanders, M. (2020, February 12). *Boost Your Team's Data Literacy.* Retrieved from Harvard Business Review: https://hbr.org/2020/02/boost-your-teams-data-literacy

Bhansali, N. (2014). *Data Governance. Creating Value from Information Assets.* CRC Press. Taylor / Francis Group.

Briguglio, L. (1994). *Some characteristics of small economies.* University of Malta. Islands and Small States Institute.

Capgemini. (2022). *The Journey to Cloud Sovereignty.* Retrieved from Capgemini: https://prod.ucwe.capgemini.com/wp-content/uploads/2022/07/CRI_Cloud-sovereignity_web10mb.pdf

Chen, W. (2022, Juky 28). *Kalido.* Retrieved from docplayer.net: https://docplayer.net/2788287-Kalido-data-governance-maturity-model.html

Collosa, A. (2021). *CIAT Blog.* Retrieved from Centro Ineramericano de Administraciones Tributarias: https://www.ciat.org/use-of-big-data-in-tax-administrations/?lang=eng

DAMA-Dictionary. (2009). *DAMA Dictionary of Data Management, 1st Edition.* NJ, USA: Technics Publications LLC.

DAMA-DMBoK2. (2017). *DAMA-DMBoK Data Management Body of Knowledge 2nd Edition.* Basking Ridge, NJ, USA: Technics Publications.

Data Orchard. (2022, January). *Data maturity framework for the not-for-profit sector.* Retrieved July 28, 2022, from www.dataorchard.org.uk: https://www.dataorchard.org.uk/resources/data-maturity-framework

data.world. (2021). *Agile Data Governance Playbook.* Retrieved July 26, 2022, from data.world: https://data.world/resources/reports-and-tools/confirmation-agile-data-governance-playbook/

DataCrossroads. (2021, March 31). *DAMA DMBoK2 vs DCAM 2.2 usage statistics.* Retrieved July 27, 2022, from datacrossroads.nl: https://datacrossroads.nl/2021/03/31/dama-dmbok2-vs-dcam-2-2-usage-statistics/

DataFlux Corp (SAS). (2007). *DataFlux.* Retrieved July 27, 2022, from dataflux.com: https://www.fstech.co.uk/fst/whitepapers/The_Data_Governance_Maturity_Model.pdf

Díaz de Sarralde, S., & Morán, D. (2022). *Innovation, Digitalization and Technology Index (INDITEC) A tool for benchmarking Tax Administrations at the international level (Based on data from ISORA 2020 Survey).* Panamá City: CIAT.

EDM Council. (2014, July 30). *Data Management Capability Assessment Model (DCAM).* Retrieved July 27, 2022, from edmcouncil.org: https://dgpo.org/wp-content/uploads/2016/06/EDMC_DCAM_-_WORKING_DRAFT_VERSION_0.7.pdf

ETF-Europa. (2018). *without-data-youre-just-another-person-opinion*. Retrieved August 31, 2022, from www.etf.europa.eu: https://www.etf.europa.eu/en/news-and-events/news/without-data-youre-just-another-person-opinion-w-edwards-deming

Falkenbach, R., González, I., Redondo, J., & Zambrano, R. (2020). The Registry. In CIAT, *ICT as a Strategic Tool to Leapfrog the Efficency of Tax Administration.* Panama City: CIAT.

Firican, G. (2018, October 3). *Gartner data governance maturity model.* Retrieved July 28, 2022, from www.lightsondata.com: https://www.lightsondata.com/data-governance-maturity-models-gartner/

Firican, G. (2018, August 8). *IBM data governance maturity model.* Retrieved July 28, 2022, from www.lightsondata.com: https://www.lightsondata.com/data-governance-maturity-models-ibm/

Firican, G. (2018, August 29). *Stanford data governance maturity model.* Retrieved July 28, 2022, from www.lightsondata.com: https://www.lightsondata.com/data-governance-maturity-models-stanford/

Firican, G. (2019, June 12). *Kalido data governance maturity model.* Retrieved July 28, 2022, from lightsondata.com: lightsondata.com/data-governance-maturity-models-kalido/

Firican, G. (2019, February 6). *Open Universiteit Nederland data governance maturity model.* Retrieved July 28, 2022, from www.lightsondata.com: https://www.lightsondata.com/data-governance-maturity-models-open-universiteit-nederland/

Firican, G. (2020, July 15). *TDWI data governance maturity model.* Retrieved July 28, 2022, from www.lightsondata.com: https://www.lightsondata.com/tdwi-data-governance-maturity-model/

Firican, G. (2020). *What is a Data Domain? (Examples Included).* Retrieved July 28, 2022, from LightsOnData Blog: https://www.lightsondata.com/what-is-a-data-domain-examples-included/

Gallant, J., & Fleet, K. (2020). *The data strategy playbook.* Informatica LLC.

Gartner Inc. (2012, February 27). *Gartner Glossary. Information Technology*. Retrieved from Gartner.com: https://www.gartner.com/en/information-technology/glossary/metadata

Gascón Catalán, J., & Redondo, J. F. (2020). Conceptual Framework: The Tax Administration Functions. In *ICT as a Strategic Tool to Leapfrog the Efficiency of Tax Administrations.* Panama: CIAT.

Harvard Business School. (2020, January 21). *ORGANIZATIONAL CHANGE MANAGEMENT: WHAT IT IS & WHY IT'S IMPORTANT.* Retrieved from Business Insights: https://online.hbs.edu/blog/post/organizational-change-management

Herzberg, B. (2021, November 11). *The Datamasters: Data Owners vs. Data Stewards vs. Data Custodians.* Retrieved from Satori: https://blog.satoricyber.com/the-datamasters-data-owners-vs.-data-stewards-vs.-data-custodians

Informatica. (2021). *Data Governance Workbook.* Retrieved July 26, 2022, from www.informatica.com: https://pdfcoffee.com/data-governance-workbook-pdf-free.html

ISORA. (2021, November 3). *ISORA 2018: Understanding Revenue Administration*. Retrieved July 28, 2022, from data.rafit.org: https://data.rafit.org/?sk=ba91013d-3261-42f8-a931-a829a78cb1ec&sId=1445908451587

Juneja, P. (2020). *Contingency Model of Change Management: Dunphy and Stace's Model of Change.* Retrieved from Management Study Guide: https://www.managementstudyguide.com/contingency-model-of-change-management.htm

Juneja, P. (2020). *Reasons for Resistance to Change.* Retrieved from Management Study Guide: https://www.managementstudyguide.com/reasons-for-resistance-to-change.htm

Kempe, S. (2011, December 27). *Assessing Data Management Maturity Using the DAMA Data Management Book of Knowledge (DMBOK) Framework - Part 2.* Retrieved July 28, 2022, from www.dataversity.net: https://www.dataversity.net/assessing-data-management-maturity-using-the-dama-data-management-book-of-knowledge-dmbok-framework-part-2/

Kempe, S. (2011, November 29). *Assessing Data Management Maturity Using the DAMA DMBOK Framework – Part 1.* Retrieved July 28, 2022, from www.dataversity.net: https://www.dataversity.net/assessing-data-management-maturity-using-the-dama-dmbok-framework-%E2%80%93-part-1/

Kidd, M. (2010). *Tax Administration in Small Economies.* Washington D.C.: IMF Fiscal Affairs Department.

Kotter, J. (2014). Leading changes: why transformation efforts fail. *Harvard Business Review*.

KPMG. (2021). *Tax Data Management.* Retrieved August 31, 2022, from meijburg.nl: https://meijburg.nl/sites/default/files/2021-02/KPMG%20Meijburg%20%26%20Co%20-%20Article%20Tax%20Data%20Management.pdf

Ladley, J. (2020). *Data Governance. How to Design, Deploy, and Sustain an Effective Data Governance Program. 2nd Edition.* Elsevier.

Marchildon, P. B. (2018). *Data Governance Maturity Assessment Tool: A Design Science Approach.* Retrieved July 27, 2022, from www.cairn.info: https://www.cairn.info/revue-projectique-2018-2-page-155.htm

Martins, W., & Seco, A. (2020). *ICT as a Strategic Tool to Leapfrog the Efficiency of Tax Administrations.* Panama City: CIAT.

Martins, W., Nieto, Ó., Seco, A., & Zambrano, R. (2020). Enablign Digital Technology and Services. In CIAT, *ICT as a Strategic Tool to Leapforg the Efficiency of Tax Administrations.* Panama City: CIAT.

McSweeney, A. (2013, October 23). *Review of Data Management Maturity Models.* Retrieved July 28, 2022, from www.researchgate.net: www.researchgate.net/profile/Alan-Mcsweeney-2/publication/333485812_Review_of_Data_Management_Maturity_Models/links/5cefe568299bf1fb184a99b5/Review-of-Data-Management-Maturity-Models.pdf

Microsoft. (2020). *A Guide to Data Governance.* Retrieved July 26, 2022, from www.microsoft.com: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4GEtu

Microsoft. (2022, April 22). *Data Governance Processes*. Retrieved August 30, 2022, from Microsoft: https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/scenarios/cloud-scale-analytics/govern-components

Microsoft Corporation. (2020). *A Guide to Data Governance: Building a Roadmap for Trusted Data.* Retrieved from https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4GEtu

Middleton, C. (2022). *Cloud sovereignty a growing strategic concern, says Capgemini.* Retrieved from Diginomica: https://diginomica.com/cloud-sovereignty-growing-strategic-concern-says-capgemini

MIT CISR Data Research Advisory Board. (2018, November 1). *How to create a successful data strategy?* Retrieved from MIT CISR website: https://cisr.mit.edu/publication/how-create-successful-data-strategy

NASCIO National Association of Chief Information Officers. (2009, May). Data Governance Part III: Frameworks – Structure for Organizing Complexity. *NASCIO Governance Series*, p. 1.

NewVantage. (2020). *Big Data and AI Executive Survey 2020.* USA: NewVantage LLC.

OECD. (2001). Principles of Good Tax Administration – Practice Note. *Tax guidance series*.

OECD. (2008). *Tax Administrations in OECD and Selected non-OECD Countries: Comparative Information .* Paris: OECD Publishing.

OECD. (2022). *Tax Administration 2022.* Paris: OECD Publishing.

OMES - Oklahoma Office of Management & Enterprise Services. (2016). *Oklahoma Office of Management & Enterprise Services.* Retrieved July 27, 2022, from https://oklahoma.gov/omes.html: https://hnu.edu/wp-content/uploads/2020/03/Data-Governance-Maturity-Model.pdf

Oracle. (2015, September). *Data governance with Oracle.* Retrieved August 20, 2022, from www.oracle.com: https://www.oracle.com/a/ocom/docs/middleware/data-integration/oracle-data-governance-wp.pdf

Palmer, K. (2021, January 20). *Building a data maturity model + the 4 stages of data maturity.* Retrieved July 28, 2022, from safegraph.com: safegraph.com/blog/the-four-stages-of-data-maturity#:~:text=Data%20maturity%20is%20a%20measurement,is%20often%20measured%20in%20stages.

Panetta, K. (2021, August 26). *A Data and Analytics Leader's Guide to Data Literacy.* Retrieved August 10, 2022, from Gartner Inc.: https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy

Peters, R. (2021, September 8). *How to build your data strategy roadmap.* Retrieved July 26, 2022, from www.hellersearch.com: https://www.hellersearch.com/blog/how-to-build-your-data-strategy-roadmap#.YuCANT3MJhE

Plotkin, D. (2020). *Data Stewardship An Actionable Guide to Effective Data Management and Data Governance.* Academic Press.

Privitar Ltd. (2022). *The Top 7 Techniques for De-identifying Data (E-book).* Retrieved from Dataversity: https://content.dataversity.net/PrivitarQ22021WP_DownloadWP.html

Rouse, J. (2021, Jan 6). *Data Governance Policies Every Business Needs.* Retrieved from The Tech Break Blog: https://www.birdrockusa.com/blog/9-data-governance-policies-for-business

Rowley, J. (2007). The Wisdom Hierarchy: Representations of the DIKW Hierarchy. *Journal of Information Science. No. 33*, pp. 163-180.

SAS. (2018). *Data Governance Framework.* Retrieved July 26, 2022, from www.sas.com: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-data-governance-framework-107325.pdf

Schrage, M. (2019, September 16). *Smart Strategies Require Smarter KPIs.* Retrieved from MIT Sloan Management Review: https://sloanreview.mit.edu/article/smart-strategies-require-smarter-kpis/

Sebastian-Coleman, L. (2013). *Measuring Data Quality for Ongoing Improvement.* Amsterdam: Elsevier Inc.

Sebastian-Coleman, L. (2020, September 16). *Data Management Capability Maturity Models.* Retrieved July 28, 2022, from damanewengland.org: damanewengland.org/images/downloads/CDMP_Study_Group/data_management_capability_maturity_models.pdf

Seco, A., & Muñoz, A. (2018). *Panorama del Uso de las Tecnologías y Soluciones Digitales Innovadoras en la Política y la Gestión Fiscal.* Discussion Paper IDB-DP-602, Inter-American Development Bank, Fiscal Division, Washington DC.

Seet, C. (2018, April 29). *What Is Capability-Based Planning?* Retrieved from Jibility: https://www.jibility.com/what-is-capability-based-planning/

Seiner, R. (2020, July 16). *Building a data governance roadmap.* Retrieved July 26, 2022, from www.dataversity.net: https://www.dataversity.net/rwdg-slides-building-a-data-governance-roadmap/#

Simpson, K. (2016). *The Difference Between Operational and Analytical Data Systems.* Retrieved from The Arkatechture Blog: https://www.arkatechture.com/blog/the-difference-between-operational-and-analytical-data-systems

Smith, M. (2011, May 10). *SAS.* Retrieved August 30, 2022, from www.sas.com: https://www.sas.com/content/dam/SAS/en_ca/User%20Group%20Presentations/Calgary-User-Group/Smith-DataFluxDataManagementApproach-May2011.pdf

Steenbeck, I. (2021, May 26). *Data Management Maturity 2020: Data Governance.* Retrieved July 27, 2022, from datacrossroads.nl: https://datacrossroads.nl/2021/05/26/data-management-maturity-2020-data-governance/

Subramanian, S. (2017, April 21). *Simple Metrics for a Successful Data Governance.* Retrieved from Towards Data Science: https://towardsdatascience.com/simple-metrics-for-a-successful-data-governance-ea55b1887d46

Tableau Software. (2020). *Data Management vs. Data Governance: The Difference Explained.* Retrieved from https://www.tableau.com/learn/articles/data-management-vs-data-governance

TADAT. (2019, April). *The TADAT Framework - a summary*. Retrieved from www.tadat.org: https://www.tadat.org/home

Talend. (2020). *Data Governance Framework - Guide and Examples.* Retrieved July 26, 2022, from www.talend.com: https://www.talend.com/resources/data-governance-framework/

Talend Company. (2020). *Structured vs. Unstructured Data: A Complete Guide.* Retrieved from Talend.com: https://www.talend.com/resources/structured-vs-unstructured-data/

The Open Group Architecture Forum. (2018). *The TOGAF® Standard, Version 9.2,.* USA: The Open Group Publications.

Thomas, G. (n.d.). *DGI Data Governance Framework.* Retrieved July 26, 2022, from datagovernance.com: https://datagovernance.com/the-dgi-data-governance-framework/#:~:text=The%20DGI%20Data%20Governance%20Framework,taking%20action%20on%20enterprise%20data.

Varshney, S. (2020, November 24). *What is Data Governnce? The Ultimate Guide.* Retrieved August 25, 2022, from ovaledge.com: https://www.ovaledge.com/blog/what-is-data-governance

Varshney, S. (2021, July 20). *Data Governance maturity models and how to measure it?* Retrieved July 28, 2022, from www.ovaledge.com: https://www.ovaledge.com/blog/data-governance-maturity-model

Varshney, S. (2021). *Maturity Model Questionnaire.* Retrieved July 28, 2022, from www.ovaledge.com: https://www.ovaledge.com/maturity-model-questionnaire-download-form

Vaughan, J. (2019). *Data Quality.* Retrieved from Techtarget WhatIs: https://www.techtarget.com/searchdatamanagement/definition/data-quality

Walery, J. (2021, March 31). *Why a Data Governance Roadmap is Important.* Retrieved July 26, 2022, from blog.idatainc.com: https://blog.idatainc.com/why-dg-roadmap-important

Wigmore, I. (2017). *Data Life Cycle.* Retrieved from Techtarget WhatIs: https://www.techtarget.com/whatis/definition/data-life-cycle

Wills, J. (2022, July 14). *How to Drive Data Literacy with 4 Simple Steps.* Retrieved August 10, 2022, from www.alation.com: https://www.alation.com/blog/how-to-drive-data-literacy/

Wray, S. (2016, August 6). *The 6-step roadmap for good data governance.* Retrieved July 26, 2022, from inform.tmforum.org: https://inform.tmforum.org/features-and-analysis/2016/08/6-step-roadmap-good-data-governance/

Zambrano, R. (2010, October 12). *CIATBlog.* Retrieved from Centro Interamericano de Administraciones Tributarias: https://www.ciat.org/mitos-leyendas-tradiciones-y-otras-cosas-asi-ii/