ANNEX 4: PART 3

Table of Contents

PART 3 – DATA GOVERNANCE FRAMEWORK

Introduction

In the era of digitization and data-driven decision-making, Data and Information have become a valuable asset of modern tax administration. This was not the case in time past where data was seen more as a by-product of operations. Data provides valuable insights into all facets of revenue administration. The quality of decision making in Tax Administration (RA) is impacted greatly by the quality of data and information available to decision makers.

In an era where data has become a valuable asset, tax authorities must establish a robust data governance framework to effectively manage, protect, and leverage taxpayer data. Such a framework ensures data integrity, privacy, and security while promoting better decision-making and compliance. By investing in data governance, revenue authorities can optimize their operations, improve service delivery, and strengthen their role in revenue collection and economic development.

The digitalization of tax administration in many countries through the use of technology is accelerating the value of data on an ever-increasing scale. Evenly, technology is contributing to the increasing risk of unauthorized data access and disclosures. Data management and governance, an essential for sustainable revenue administration, enables growth in revenue and facilitates seamless operation. The increasing value-addition of data and its associated risk as well as the challenge of managing and leveraging vast amounts of data for effective tax administration makes an appealing case for a holistic data governance framework in revenue administration.

A sound data governance framework is important for the digital transformation of tax administration, and is crucial to ensuring the quality, accessibility, security, and ethical use of data. The prime objective of data governance in a digitalized tax administration is to create value to tax administration by ensuring that data attributes (ownership, availability, quality and relevant) and risk mitigation controls are in place.

Implementing a robust data governance framework offers several benefits to revenue authorities:

- Improved Data Quality: By establishing data governance practices like data standards, validation procedures, and data cleansing processes, revenue authorities can enhance data quality, reduce errors and inconsistencies in taxpayer data. This ensures the reliability and integrity of data, leading to more accurate reporting, analysis, and decision-making processes.

- Enhanced Compliance: A well-defined data governance framework ensures compliance with legal, regulatory, and data protection requirements. This enables revenue authorities to collect and maintain accurate taxpayer data, facilitating effective compliance management and reducing tax evasion. Tax authorities can demonstrate accountability and transparency in their data management practices, fostering public trust and confidence.
- Efficient Operations: Effective data governance streamlines data-related processes, such as data integration, sharing, and reporting. This leads to increased operational efficiency, faster response times, and improved service delivery to taxpayers.
- Informed Decision-Making: With reliable and consistent data, revenue authorities can make informed decisions, identify tax evasion patterns, detect non-compliance, and allocate resources effectively. Data-driven insights support evidence-based policymaking and enhance overall revenue collection efforts.
- Stakeholder Collaboration: A data governance framework promotes collaboration among internal and external stakeholders. It enables sharing and integration of data across different departments, facilitating better coordination, knowledge sharing, and collaboration between revenue authorities and other government agencies.

A comprehensive data governance framework for tax authorities should incorporate the following key elements:

- Data Strategy and Governance Framework: Establish a clear data strategy aligned with organizational objectives, defining roles, responsibilities, and decision-making processes related to data governance. This includes the creation of data governance policies, guidelines, and frameworks to guide data management practices.
- Data Architecture and Infrastructure: Develop a robust data architecture that outlines the technical infrastructure, data models, and data flow diagrams to facilitate data integration, storage, and retrieval. It should consider scalability, data security measures, and compliance with relevant data protection regulations.
- Data Quality Management: Implement mechanisms to ensure data accuracy, completeness, and consistency. This involves defining data quality metrics, conducting regular data audits, and establishing data validation and cleansing procedures.
- Data Security and Privacy: Implement stringent security measures to protect sensitive taxpayer data from unauthorized access, breaches, or misuse. Establish data privacy policies, consent frameworks, and protocols for data anonymization and pseudonymization to comply with privacy regulations.
- Data Access and Sharing: Develop protocols and procedures for data access and sharing within the revenue authority and with external stakeholders. This includes defining access

controls, user permissions, and data sharing agreements to maintain data confidentiality and prevent unauthorized use.

- Data Ethics and Governance Committee: Establish a dedicated committee responsible for overseeing data governance practices, ensuring ethical use of data, and addressing any potential ethical dilemmas or concerns. The committee should consist of multidisciplinary experts and stakeholders to provide guidance and recommendations.
- Capacity Building and Training: Invest in training programs to enhance data literacy among staff members, enabling them to understand and navigate the complexities of data governance. This includes training on data protection, privacy, data analytics, and emerging technologies.

This part of the guide discusses data governance strategy, the guiding principles, standards, policies, structures and procedures of data collection, processing, storage, classification, use, security, archival and destruction. It comprises of three chapters, that is,

- Chapter 5: Data Governance Strategy and Framework
- Chapter 6: Data Collection and
- Chapter 7: Use of data.

CHAPTER 5 – DATA GOVERNANCE STRATEGY AND FRAMEWORK

Digitized tax administrations are powered by data, which underpins the construction of taxpayer profiles and the administration of tax functions. The application of technology to big data flows allows for the "translation" and organization of data into usable information associated with individual taxpayers. But access and use of taxpayer data generates risk . To manage this data effectively and ensure its quality, security, and compliance, Tax administrations implement data governance strategies frameworks. Data governance strategies set out the objectives and goals of the frameworks. Data governance frameworks provide a structured approach to managing data and data flows, ensuring that data is accurate, accessible, used responsibly, and protected.

5.1 Background and foundation

5.1.1 Purpose: Why develop a Data Governance Framework?

In today's data-driven world, Tax administrations are inundated with vast amounts of data. This data, if managed effectively, can provide invaluable insights, drive strategic decisions, and offer a competitive edge. However, without proper oversight, data can become a liability, leading to inefficiencies, security breaches, and compliance issues. This is where a data governance framework becomes essential.

5.1.2 Data Governance Strategy

One of the primary goals of a data governance framework is to ensure the quality and consistency of data across the organization. High-quality data is accurate, complete, and reliable, which is crucial for making informed decisions. A governance framework establishes standards and procedures for data entry, storage, and maintenance, ensuring that all data adheres to these quality benchmarks. This consistency allows different departments to collaborate effectively, as they can trust the data they are using is accurate and up to date.

With the increasing prevalence of cyber threats and stringent data protection regulations, safeguarding sensitive information has never been more critical. A data governance framework outlines the policies and controls necessary to protect data from unauthorized access, breaches, and other security risks. It ensures that data is classified according to its sensitivity and that appropriate measures are in place to protect it. Additionally, it helps Tax administrations comply with legal and regulatory requirements, such as GDPR or CCPA, by defining how data should be handled and protected.

Data governance frameworks provide a structured approach to managing data, which in turn facilitates better decision-making. By ensuring that data is accurate, consistent, and readily

available, Tax administrations can leverage it to gain insights and make strategic decisions. This framework also promotes a culture of data-driven decision-making, where employees at all levels understand the importance of data and are equipped to use it effectively. This leads to more informed, timely, and effective decisions that can drive business growth and innovation.

Effective data governance can significantly improve operational efficiency. By standardizing data management practices, Tax administrations can reduce redundancies and streamline processes. This not only saves time and resources but also minimizes the risk of errors and inconsistencies. Furthermore, a well-implemented data governance framework can automate many data management tasks, freeing up employees to focus on more strategic activities. This increased efficiency can lead to cost savings and improved productivity.

A data governance framework clearly defines roles and responsibilities related to data management. This promotes accountability and ownership, ensuring that data is managed responsibly throughout its lifecycle. By assigning specific roles, such as data stewards or data custodians, Tax administrations can ensure that there are dedicated individuals responsible for maintaining data quality, security, and compliance. This accountability fosters a culture of responsibility and ensures that data governance is a shared priority across the organization.

As Tax administrations grow and evolve, their data needs and challenges also change. A robust data governance framework provides the scalability and flexibility needed to adapt to these changes. It allows Tax administrations to scale their data management practices to accommodate increasing volumes of data and evolving regulatory requirements. Additionally, it provides the flexibility to incorporate new data sources and technologies, ensuring that the organization can continue to leverage data effectively as it grows.

5.1.3 Key Components of a Data Governance Framework

A robust data governance framework typically consists of the following areas:

a. Data Governance Policies and Standards: This section contains the rules and guidelines that dictate how data should be managed and used within the organization. Policies cover areas such as data quality, data privacy, and data security, ensuring that all data-related activities comply with legal and regulatory requirements.
b. Data Ownership**:** This component delineates the roles and responsibilities for those in charge of the data program. These “data stewards” are responsible for overseeing the implementation of data governance policies and ensuring that data is managed

according to established standards. They play a crucial role in maintaining data quality and resolving data-related issues.

c. Data Quality Management: This component focuses on ensuring that data is accurate, complete, and reliable, by delineating metrics and KPIs against which data performance is judged. It involves regular data quality assessments, data cleansing, and the establishment of data quality metrics.

d. Data Architecture and Metadata Management: This design section delineates the structure and organization of data's journey within the tax administration, defining data access points, flows, and storage. Along with the data architecture, a metadata management plan is needed and sets the processes and protocols for maintaining information about the data, such as its source, format, and usage. Together, these components help ensure that data is organized and easily accessible.

e. Data Security and Privacy: As a key component as a part of a data governance framework, data security and privacy provisions define access rights and responsibilities. Protecting data from unauthorized access and ensuring that it is used in compliance with privacy regulations is a critical aspect of data governance. This involves implementing security measures such as encryption, access controls, and regular security audits.

f. Data Governance Committee: A data governance committee, typically composed of representatives from various departments, oversees the data governance program. The committee is responsible for setting data governance priorities, resolving conflicts, and ensuring that data governance initiatives align with the organization's strategic goals.

5.1.4 Possible benefits derived from implementation of a Data Governance Framework

Implementing a data governance framework offers several benefits:

a. Improved Data Quality: By establishing clear standards and processes for managing data, Tax administrations can ensure that their data is accurate, complete, and reliable.

b. Enhanced Data Security and Privacy: A data governance framework helps protect sensitive data from unauthorized access and ensures compliance with privacy regulations, reducing the risk of data breaches and legal penalties.

c. Better Decision-Making: High-quality, well-managed data provides a solid foundation for making informed business decisions, leading to improved operational efficiency and competitive advantage.

d. Regulatory Compliance: A data governance framework helps Tax administrations comply with legal and regulatory requirements related to data management, reducing the risk of fines and reputational damage.
e. Increased Data Accessibility: By organizing data and maintaining comprehensive metadata, a data governance framework makes it easier for employees to find and use the data they need, enhancing productivity and collaboration organization's data assets. It encompasses the policies, procedures, and standards that ensure data is accurate, consistent, secure, and used responsibly.

A well-implemented data governance framework helps Tax administrations make better decisions, improve operational efficiency, and maintain trust with customers and stakeholders.

5.2 Moving to Practice: Setting Up a Data Governance Program

In the era of big data, establishing a robust data governance program is crucial for Tax administrations to manage their data assets effectively. A well-structured data governance program ensures data quality, security, and compliance, while also enabling better decision-making and operational efficiency. Here's a comprehensive guide on how to set up a data governance program.

5.2.1 Defining Objectives and Scope

The first step in setting up a data governance program is to define its objectives and scope. This involves identifying the specific goals the organization aims to achieve through data governance, such as improving data quality, ensuring regulatory compliance, or enhancing data security. It is also essential to determine the scope of the program, including the data domains, business units, and processes that will be governed. Clear objectives and scope provide a foundation for the program and guide subsequent steps.

5.2.2 Establishing a Governance Structure

A successful data governance program requires a well-defined governance structure. This includes forming a data governance council or steering committee composed of key stakeholders from various departments. The council is responsible for overseeing the program, making strategic decisions, and ensuring alignment with organizational goals. Additionally, roles such as data stewards, data owners, and data custodians should be assigned to manage

specific aspects of data governance. These roles ensure accountability and ownership of data management tasks.

5.2.3 Developing Policies and Standards

Once the governance structure is in place, the next step is to develop data governance policies and standards. Policies provide guidelines on how data should be managed, accessed, and protected, while standards define the specific requirements for data quality, metadata, and data security. These policies and standards should be aligned with industry best practices and regulatory requirements. It is also important to establish procedures for policy enforcement and compliance monitoring.

5.2.4 Implementing Data Management Processes

Effective data governance requires the implementation of data management processes that ensure data quality, consistency, and security. This includes processes for data collection, storage, integration, and usage. Data quality management processes, such as data profiling, cleansing, and validation, help maintain high data quality. Data security processes, including access controls, encryption, and auditing, protect data from unauthorized access and breaches. Additionally, metadata management processes ensure that data is well-documented and easily accessible.

5.2.5 Leveraging Technology and Tools

Technology plays a critical role in supporting data governance initiatives. Tax administrations should invest in data governance tools and technologies that automate and streamline data management processes. These tools can include data cataloging, data lineage, data quality, and data security solutions. By leveraging technology, Tax administrations can enhance the efficiency and effectiveness of their data governance program. It is also important to ensure that these tools are integrated with existing IT infrastructure and systems.

5.2.6 Promoting a Data-Driven Culture

A successful data governance program requires a cultural shift towards data-driven decision-making. This involves promoting awareness and understanding of data governance principles across the organization. Training and education programs should be conducted to equip employees with the necessary skills and knowledge to manage data effectively. Additionally, fostering a culture of collaboration and communication between business and IT teams is essential for the success of the program.

5.2.7 Monitoring and Continuous Improvement

Data governance is an ongoing process that requires continuous monitoring and improvement. Tax administrations should establish metrics and KPIs to measure the effectiveness of their data governance program. Regular audits and assessments should be conducted to identify areas for improvement and ensure compliance with policies and standards. Feedback from stakeholders should be gathered and used to refine and enhance the program. Continuous improvement ensures that the data governance program remains relevant and effective in the face of evolving data challenges and requirements.

Setting up a data governance program is a complex but essential task for tax administrations seeking to harness the power of their data. By defining clear objectives, establishing a governance structure, developing policies and standards, implementing data management processes, leveraging technology, promoting a data-driven culture, and continuously monitoring and improving the program, Tax administrations can ensure effective data governance. This not only enhances data quality, security, and compliance but also drives better decision-making and operational efficiency.

5.3 Challenges in Setting Up a Data Governance Framework

Establishing a data governance framework is a critical step for digitized or digitizing tax administrations aiming to manage their data assets effectively. However, this process is not without its challenges. Here are some of the key obstacles tax administrations may face when setting up a data governance framework:

5.3.1 Lack of Executive Support

One of the most significant challenges is securing buy-in from top management. Without executive support, it can be difficult to allocate the necessary resources and prioritize data governance initiatives. Executives need to understand the value of data governance and how it aligns with the organization’s strategic goals.

5.3.2 Cultural Resistance

Implementing a data governance framework often requires a cultural shift towards data-driven decision-making. Employees may resist changes to established processes and practices, especially if they do not see the immediate benefits. Overcoming this resistance requires effective communication, training, and demonstrating the value of data governance.

5.3.3 Data Silos

Data silos occur when data is isolated within different departments or systems, making it difficult to achieve a unified view of the organization's data. Breaking down these silos and ensuring data integration across the organization is a significant challenge. It requires collaboration between departments and the implementation of data integration tools and processes.

5.3.4 Complex Data Landscape

Tax administrations often deal with a complex data landscape that includes various data sources, formats, and systems. Managing this complexity and ensuring data consistency and quality across the organization can be daunting. It requires robust data management processes and technologies to handle diverse data types and sources.

5.3.5 Data Quality Issues

Ensuring high data quality is a fundamental aspect of data governance, but it can be challenging to achieve. Data quality issues such as inaccuracies, inconsistencies, and missing data can undermine the effectiveness of data governance efforts. Addressing these issues requires ongoing data quality management processes, including data profiling, cleansing, and validation.

5.3.6 Regulatory Compliance

Compliance with data protection regulations such as GDPR, CCPA, and others is a critical aspect of data governance. However, keeping up with evolving regulatory requirements and ensuring compliance can be challenging. Tax administrations need to stay informed about regulatory changes and implement processes and controls to meet compliance requirements.

5.3.7 Resource Constraints

Implementing a data governance framework requires significant resources, including time, budget, and skilled personnel. Tax administrations may struggle to allocate these resources, especially if they have competing priorities. It is essential to demonstrate the long-term benefits of data governance to justify the investment.

5.3.8 Technology Integration

Integrating data governance tools and technologies with existing IT infrastructure can be complex. Tax administrations need to ensure that these tools are compatible with their current systems and can support their data governance objectives. This may require additional investments in technology and expertise.

5.3.9 Defining Clear Roles and Responsibilities

Establishing clear roles and responsibilities for data governance is crucial for accountability and ownership. However, defining these roles and ensuring that employees understand their responsibilities can be challenging. It requires careful planning and communication to ensure that everyone involved knows their role in the data governance framework.

5.3.10 Continuous Improvement

Data governance is not a one-time project but an ongoing process that requires continuous monitoring and improvement. Tax administrations may struggle to maintain momentum and ensure that data governance practices evolve with changing business needs and data challenges. Establishing a culture of continuous improvement and regular assessments is essential for long-term success.

While setting up a data governance framework presents several challenges, addressing these obstacles is crucial for effective data management. By securing executive support, fostering a data-driven culture, breaking down data silos, managing data quality, ensuring regulatory compliance, allocating resources, integrating technology, defining roles, and committing to continuous improvement, Tax administrations can overcome these challenges and establish a robust data governance framework

5.4 Challenges faced by Developing Countries in Setting Up a Data Governance Framework

Developing countries face additional challenges in establishing robust data governance frameworks. These challenges stem from various socio-economic, technological, and infrastructural factors that can hinder the implementation and effectiveness of data governance initiatives. Here are some of the special challenges faced by developing countries:

5.4.1 Limited Financial Resources

One of the most significant challenges for developing countries is the lack of financial resources. Implementing a data governance framework requires substantial investment in technology, infrastructure, and skilled personnel. Developing countries often have limited budgets, and data governance may not be prioritized over other pressing needs such as healthcare, education, and infrastructure development. This financial constraint can impede the establishment and maintenance of an effective data governance framework.

5.4.2 Inadequate Technological Infrastructure

Developing countries often struggle with inadequate technological infrastructure, which is essential for data governance. Reliable internet connectivity, data storage facilities, and advanced data management tools are often lacking. This technological gap makes it difficult to implement and sustain data governance practices. Additionally, the lack of infrastructure can lead to data silos and inconsistencies, further complicating data management efforts.

5.4.3 Shortage of Skilled Personnel

A successful data governance framework requires skilled personnel, including data scientists, data stewards, and IT professionals. Developing countries often face a shortage of such skilled individuals due to limited educational and training opportunities. This skills gap can hinder the implementation of data governance initiatives and affect the overall quality and reliability of data management practices.

5.4.4 Regulatory and Legal Challenges

Developing countries may lack comprehensive data protection laws and regulations, which are crucial for data governance. The absence of clear legal frameworks can lead to inconsistencies in data management practices and make it challenging to ensure data security and privacy. Additionally, even when regulations exist, enforcement can be weak due to limited resources and institutional capacity.

5.4.5 Cultural and Organizational Resistance

Cultural and organizational resistance to change can be a significant barrier to implementing data governance in developing countries. Tax administrations may be accustomed to traditional ways of managing data and may resist adopting new practices and technologies. This resistance can stem from a lack of awareness about the benefits of data governance or fear of the unknown. Overcoming this resistance requires effective communication, education, and change management strategies.

5.4.6 Data Quality Issues

Ensuring high data quality is a fundamental aspect of data governance, but developing countries often face significant data quality issues. Inaccurate, incomplete, and outdated data can undermine the effectiveness of data governance efforts. Addressing these issues requires robust data quality management processes, which can be challenging to implement without the necessary resources and expertise.

5.4.7 Political and Economic Instability

Political and economic instability can pose significant challenges to data governance in developing countries. Frequent changes in government, policy shifts, and economic crises can disrupt data governance initiatives and lead to a lack of continuity and consistency. Additionally, political instability can affect the allocation of resources and priorities, making it difficult to sustain long-term data governance efforts.

5.4.8 Limited Awareness and Understanding

There is often limited awareness and understanding of data governance principles and practices in developing countries. This lack of awareness can result in a lack of support from key stakeholders, including government officials, business leaders, and the general public. Raising awareness about the importance of data governance and its benefits is crucial for gaining support and driving successful implementation.

Setting up a data governance framework in developing countries presents unique challenges that require tailored solutions. Addressing these challenges involves securing financial resources, improving technological infrastructure, building skilled personnel, establishing clear regulatory frameworks, overcoming cultural resistance, ensuring data quality, and navigating political and economic instability. By recognizing and addressing these special challenges, developing countries can establish effective data governance frameworks that enhance data management, drive better decision-making, and support sustainable development.

5.5 Experience: Regional and multilateral data governance programs

- **Africa**: Many African countries have made significant strides in data governance. For instance, between 2012 and 2021, the number of African countries with at least one form of data protection law tripled from 12 to 28. These frameworks focus on both safeguards (e.g., data protection, privacy) and enablers (e.g., data portability, localization). However, implementation remains a challenge due to regulatory inertia and capacity issues.[1]

- **Global Data Governance Mapping Project:** This project, conducted by the Digital Trade and Data Governance Hub, provides insights into how various countries, including developing ones, govern personal, public, and proprietary data at national and international levels. The project highlights the diverse approaches and the importance of robust legal and regulatory frameworks to sustain a data-driven economy.[2]

- **Open Government Data Projects:** Several developing countries have active open government data projects. These initiatives aim to enhance transparency, accountability, and public service delivery. A study identified 12 case studies from diverse geographic regions, showcasing the impact and challenges of these projects.[3]

5.5.1 Country-specific programs

- Kenya: Kenya has made significant progress with its data governance framework, particularly through the implementation of the Data Protection Act, 2019. This act establishes the Office of the Data Protection Commissioner, which oversees data protection and privacy, ensuring that personal data is processed in accordance with the law.[4]

- Brazil: Brazil's General Data Protection Law (LGPD) is another example of a robust data governance framework in a developing country. The LGPD regulates the processing of personal data and establishes guidelines for data protection, ensuring that individuals' privacy rights are respected[5].

- South Africa: South Africa's Protection of Personal Information Act (POPIA) is another example of a robust data governance framework. POPIA regulates the

[1] https://www.usitc.gov/publications/332/journals/jice_africa_data_protection_laws.pdf
[2] https://globaldatagovernancemapping.org/
[3] https://oecd-opsi.org/guide/open-government/open-government-implementation/
[4] Kenya Data Protection Law (2019) DataProtectionAct24of2019.pdf (kenyalaw.org)
[5] https://iapp.org/news/a/brazil-s-new-regulation-on-international-data-transfers

processing of personal information and establishes guidelines for data protection, ensuring that individuals' privacy rights are respected.[6]

[6] https://cms.law/en/int/expert-guides/cms-expert-guide-to-data-protection-and-cyber-security-laws/south-africa

CHAPTER 6 – DATA COLLECTION

This chapter discusses data source, data attributes, data mining and cleansing, unstructured data management and techniques to avoid duplication. Data collection, which can be used interchangeably with data creation, is the first step of the data life cycle.

Data collection mechanism in tax authorities can be categorized into the following:

i. Data acquisition - obtaining data that has been produced by a third-party organization.
ii. Data entry - manual input of data by humans or devices.
iii. Data capture – generation of data by devices such as sensors

6.1. Sources of Data

Tax Authorities collect data from multiple sources. Primarily, taxpayers and third parties are the two major external sources of data.

Taxpayers' data are obtained at various stages of engagement. From the onset, data is obtained at registration. Subsequently, data is sourced from taxpayers through filing, surveys, and inquiries. Data from taxpayers are collected though either electronic or manual means.
In order to administer taxes efficiently, tax authorities cannot rely on data supplied by taxpayers only. They must utilize data from third parties to enhance taxpayers' compliance. With third party data, tax authorities can detect potential taxpayers that have not registered and identify false or under-declaration of taxpayers' incomes. Third parties include (but not limited to):

- Treasury Department;
- Banks;
- National Registry;
- Social Security Office;
- Immigration Department;
- Labor Department;
- Transportation Department.

6.2. Data Collection Tools

Tax Authority can collect data in a variety of ways, including:

- Forms: Web forms, paper forms and applications forms are some of the most common ways Revenue Administrations generate data.
- Surveys: Surveys can be an effective way to gather vast amounts of information from a large number of respondents.

- Interviews: Interviews and focus groups session conducted with taxpayers offer opportunities to gather qualitative and subjective data that may be difficult to capture through other means.
- Direct Observation: Observing how taxpayers interact with a website, application, or product can be an effective way to gather data that may not be offered through the methods above.
- Device: Use of sensors on products or web metrics to gather information about taxpayers such as the number of visits on tax authority's website, location of goods etc.
- Web service: use of software protocol or web application programming interface (API) to obtain data from a third party.

6.3. Types of Data available to Tax Authorities

The set of data available to a tax authority has expanded to include unstructured data. Before, tax authorities harnessed structured data alone for analysis and decision making. What is the difference between structured data and unstructured data? Structured data are data that conform to a pre-defined format and is therefore straightforward to analyze. Unstructured data are information that either does not have a predefined data model or is not organized in a pre-defined manner.

Structured data include data such as taxpayer registration information, filing and payment information. These are stored easily in relational database. Storage of unstructured data, such as social media post, chats, emails, and sensor data, in relational database is very difficult and impractical. Most times, unstructured data are stored in non-relational (NoSQL) databases.
Unstructured data is a rich source for data analytics to detect events and predict taxpayers' behaviour. Tax evasion, fraud, under-declaration can all be detected using unstructured data, as much as structured data.

6.4. Data Quality

For data to create value for tax authorities, it must have integrity. If data are collected only to satisfy volume requirements, less value will accrue to tax authorities from data collected. Tax authorities should ensure quality controls are in place to enhance the quality of data collected, processed, and stored. Quality data are accurate, complete, and relevant (timely). tax authorities should strive to use digital tools for data collection. Collection of data with non-digital tools (paper or manual) minimizes the quality of data. The following should be considered to ensure quality data is collected and maintained

i. Instil data validation control at time of data capture.

This includes the use of field validation of data type (Number, letters, email format, phone number format, use of check digit)

ii. Provide facility to update information

Some taxpayer information is dynamic. Over time, information provided may become outdated due to changes in taxpayer situation, status, location etc. outdated information needs to be updated and tax authorities should ensure taxpayers or tax officers can update information seamlessly.

iii. Compares data across multiple sources

With the use of technology, tax authorities can validate information supplied by taxpayers with similar information at other sources. Inconsistent information across multiple sources should be investigated and the necessary correction should be made. Continuous data cleansing shall ensure that data is relevant, accurate and complete.

6.5. Data Mining and Analytics

Many Tax Authorities collect data from a variety of sources to develop a more complete picture of taxpayers' profiles. Taxpayers are increasingly required to submit client invoices, statements of accounts, customs declarations, vendor invoices and bank records, mostly in real or near-real time to the tax authority. In addition, tax authorities collect data from sources other than directly from taxpayers.

As tax authorities gather troves of data mostly through the use of digital platforms from many sources, the task of analysing data becomes cumbersome for tax officers. Data analytics techniques and tools are necessities for tax authorities to derive insight from data. With more data available to tax authorities, analytics is no longer an optional tool. Data Analytics has become the driver of compliance risk management program. Risk engines are built using technology to analyse taxpayers from a 360-degree angle. Tax authorities should use data analytics to mine data to help increase tax collections, target compliance initiatives, and improve overall efficiency. Analytics reveals patterns, trends, and associations in tax and taxpayer data.

Moreover, tax authorities should leverage Big Data and analytics to detect fraud or predict taxpayers' behaviour. Big data refers to the increasing volume of data available, the variety of formats and the speed at which it can be processed. Data mined from social media, taxpayers' registers, filing, and other third-party sources can help to:

i. validate taxpayers' invoices;

ii. verify sales and purchase declarations;

iii. verify payroll and withholding declarations;
iv. compare data across jurisdictions and taxpayers;
v. examine taxpayer lifestyle;
vi. predict tax residency;
vii. manage tax debt; and
viii. combat evasion.

6.6. Unstructured Data in Tax Administration

The participation of Tax authorities in the social networking space has only increased the data available for analysis. Unstructured data is becoming a treasure trove of insight and data-driven value addition. Posts, tweets, logs, chats, location data, email, video, audio, and other types of unstructured data are rich data sources with leverage for tax administration. But how can one analyse unstructured data to draw conclusions?

The analysis of unstructured data is possible through the use of technologies and concepts such as Big Data, analytics, NoSQL database, artificial intelligence (AI), machine learning, the Internet of Things (IoT), mobility and cloud computing. On their own, unstructured data are meaningless. They have to be transformed with some semblance of structure to be meaningful. In analysing unstructured data and extracting insights, tax authorities should consider the following:

- Determine analysis goal: For example to detect taxpayer sentiment, to identify evasion etc;
- Identify data source;
- Select analytics technique and technology;
- Pre-process and clean data;
- Classify and segment data;
- Visualize data; and
- Draw conclusion.

6.7. Avoiding Duplication of Taxpayer Registration Data

In all tax authorities, registration of taxpayers is the prerequisite for all other taxpayer responsibilities. Taxpayer register forms the basis of any taxpayer compliance program. Therefore, a credible taxpayer register is sine qua non to effective tax administration. Credible tax register adheres to tax legislations that forbid multiple Tax Identification Numbers for a unique person (legal or natural). Often times, tax authorities are faced with duplication in registration data of taxpayers – same person registering more than once and obtaining multiple

TIN. Duplicate registration may be attributed to many factors – ranging from manual registration process to a lack of sophisticated registration systems.

To address issues of duplication in taxpayer registration data, tax authorities should strongly consider the following:

i. **The use of automated system with robust duplicate check algorithm to record registration data.**

It is almost impossible to maintain a manual taxpayer registration process and ensure unique taxpayers. An automated system with robust duplicate check algorithm ensures taxpayers are unique by cross-checking registrant information with existing taxpayers' data for similarity. This check should be on taxpayer identity document (primarily), location, email, name, telephone number, address and/or relationships.

ii. **Requirement of a single set of government issued identity documents for taxpayer verification.**

The use of a single set of documents to verify an individual or organization improves the credibility of taxpayer registration. Government issued identity documents may include National Identification card, passport, social security card, driver license, company registration certificate, business registration certificate etc. Allowing the use of multiple sets for taxpayer registration eviscerates credibility of taxpayer register especially when national systems are not linked. For instance, an individual may be registered twice if they present two sets of identification documents at different times for registration. However, with singular national document requirement, even an unsophisticated system will detect duplication of a national document reference number when it is entered more than once.

iii. **Interface Taxpayer registration system with other national systems.**

Taxpayer register is credible if the data provided by taxpayers during registration has integrity. One way to ensure data integrity is to link registration systems with other national systems to validate the information captured from taxpayer and/or retrieve information for national systems. The interfacing of tax register with national systems may simplify the registration process by requiring taxpayers to provide data that is not available in other national systems.

iv. **Maintain a single registration database**

For duplication check to be effective, taxpayer registration data must be stored in a single database. The use of multiple databases may pose resource constraints on the system and may even be ineffective in ensuring unique registration. Only in extreme cases where a single

database may be impractical should a tax authority allow multiple registration databases. In such a case, duplicate check algorithm should apply to all data regardless of where it is stored – a resource intensive process.

v. Generate unique Tax Identification Number (TIN) with validation control.

Tax authorities should have control over the issuance and allocation of Taxpayer Identification Number. A unique TIN controlled by the tax authority and used for all taxes is essential for effective tax administration. TIN should be generated, preferably by an automated system, with validation controls such as check digit. Tax Authorities should ensure that a single TIN is not assigned to multiple taxpayers and that no taxpayer is assigned multiple TINs.

vi. Maintaining Taxpayer Register

The credibility requirement of the taxpayer register does not end with the entry of credible data at registration. Over time, taxpayers' information changes because of changes in taxpayer address, legal status, contact information, or business status. Changes in taxpayer information should be reflected in the taxpayer register. The tax authority should provide mechanisms for taxpayers to make changes and tax officers to validate taxpayers' changes as well as capture changes identified internally.

6.8. Obstacles to obtaining Data

Since data is the panacea of insightful decision making and value creation, should tax administration collect data from every source possible? Ideally, the answer will be in the affirmative. However, there are obstacles to obtaining data directly or indirectly. Some obstacles to obtaining data are grounded in legal precepts while others may border on capacity constraints. Below are two legal constraints to obtaining data:

6.8.1. Privacy, Confidentiality, and other data protection laws

Most tax authorities are bound to adhere to privacy laws which prohibit the sharing of personal information with third parties unless with express consent of the data subject. Similarly, confidentiality laws put a requirement on organizations to protect data against unintentional, unlawful, and unauthorized access, disclosure or theft. Violation of these laws has far-reaching effects on the survivability of organizations. As such, organizations exercise extreme care to avoid violation of these laws and regulations. Refusal to share data or limiting of the amount of data shared with tax authorities may be part of the safety measures enacted by an organization.

6.8.2. Data localization and data sovereignty laws

One of the important sources of data for tax authorities is other tax authorities under Information Exchange arrangement. A tax authority's ability to freely obtain information from other tax authorities is greatly impacted by data location laws in other tax authorities' jurisdictions. Data localization and data sovereignty laws require data about natural and legal persons to be collected, processed and stored locally. Some data localization laws are flexible and allow transfer to other national jurisdictions, but others are inflexible and prohibit transfer of data across national borders. Where data localization and data sovereignty laws are inflexible, access to relevant information for other tax administrations to investigate cross-border tax evasion cases may not be possible.

Besides these legal obstacles, operating incapacity may also pose a challenge to data collection. Operationally, a tax authority's data collection capacity is impaired by the following:

i. **Lack of or limited processing and storage capacity**

   The capacities of information systems and data infrastructure are factors in determining the volume of data to be collected. Where information system and data infrastructure capacities are limited, the quantum of data collected will be minimal. Tax authorities with manual processes are less likely to obtain large volumes of data.

ii. **Lack of or limited data security**

   Most times, third parties assess tax authorities' data security arrangements and risk management processes before finalizing exchange of information agreements. These assessments are necessary because third parties are still accountable for protection of data shared with tax authorities Where tax authorities' data security arrangements and risk management are inadequate, third parties will be reluctant to share data with them. Third parties' assessment looks at areas of data encryption, access control, auditability, roles and responsibility for data ownership and custody.

Data collection is essential to the successful digitalization of revenue authorities, enabling more efficient tax administration, better compliance, and enhanced policymaking. As it is the first step towards the utilization of a data-driven approach, it is critical that authorities ensure that they put in place mechanisms that drive proper data collection to ensure they have clean data that will feed into the digitalization process.

CHAPTER 7 – USE OF DATA

7.1. Data Storage

After data is collected by tax authorities, it is then stored for processing and use. Data storage is the second stage of the data lifecycle. In todays' digital world, data are mostly digital. Data are stored in either a database or a file system on a server hosted in an on-premises data centre or in the cloud.

Decision on whether to store data in the cloud or on premise depends on multiple factors including legal and regulatory compliance, cost, security, and scalability. Cloud storage is advantageous with respect to cost, scalability, availability whereas on-premises storage is advantageous in respect of security, control, and data regulatory compliance.
Even with a decision on a cloud storage, tax authorities need to decide further on the cloud deployment model and cloud service model. There are four cloud deployment models:

- Private cloud: The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- Community cloud: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party.
- Public cloud: The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- Hybrid cloud: The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).[7]

Similarly, there are three cloud service models:

- Software-as-a-Service (SaaS): The cloud service provider (CSP) provides software for the user, which is running and deployed on cloud infrastructure. In this case, the user (consumer) is not responsible for managing or maintaining the cloud infrastructure, including network, servers, Operating Systems (OSs), or any other application-related issues. The consumer just uses the software as a service on demand.

7 J.R. Wrinkler, Securing the Cloud: Cloud Computer Security Techniques and Tactics. (Massachusetts: Syngress, 2011), 29

- Platform-as-a-Service (PaaS): The CSP provides a platform to the consumer to deploy consumer-created applications written in any programming language supported by the CSP. The consumer is not responsible for managing or maintaining the underlying infrastructure, such as the network, servers, OSs, or storage. However, the consumer controls the deployed applications and the hosting environment configurations.
- Infrastructure-as-a-Service (IaaS): The CSP provides the consumer with the processing, storage, networks, and other essential computing resources to enable the consumer to run his or her software, which can be OSs and applications. This model involves managing the physical cloud infrastructure by the provider.[8]

Data storage is also contingent on the type of data tax authorities collect. Structured data are suited for relational databases while unstructured data typically makes use of non-relational or NoSQL databases.

7.2. Data usage

This is the third stage of the data lifecycle. During this stage, data is available to tax authorities, taxpayers, and other stakeholders. Access Control Management enables tax authorities to define users and usage. Usage of data may include computation of taxpayers' tax liability, generation of reports, data analysis and visualization. Data-driven tax authorities leverage data and analytics to solve perennial problems of fraud, low filing ratio, undetected taxpayer, under-declaration, and low payment ratio. Some analytics available to tax authorities include clustering, predictive modelling, network analysis, visualization and many more. Data analytics results in decision-making and reporting to various stakeholders.

Additionally, sharing of data to third parties is also considered usage of data. Other tax authorities, who request and obtain data, use data for similar purposes as the primary tax authority Other third parties such as banks, Treasury Department, Purchasing and Procurement Department, may use data to validate clients and foster regulatory compliance.

7.3. Data archival and destruction

Statutes of limitation on collection of assessed tax and assessment of tax obligation and data retention laws and regulations may render some data no longer useful for everyday operations. Such data are removed from active production and archived and may be destroyed where

[8] John R. Vacca, Computer and Information Security Handbook, 3rd Edition. (Massachusetts: Morgan Kaufmann, 2017), 109

retention period has elapsed. Data Archival and Data Destruction are the fourth and fifth stages of data lifecycle respectively.

Before data is destroyed, it is critical to confirm adherence to all policies on retention. Even more, data may need to be kept beyond retention where it may be useful for potential litigation and investigation. Tax authorities should avoid destroying data and maintain a rich data warehouse for analytics purposes.

Destruction of data should be handled with extreme care otherwise the data may be exposed to unauthorized access. Where data is stored on a media which is no longer needed, the data should be destroyed by physical destruction of the media. However if the media is needed for an economic reason, then data should be destroyed by demagnetizing or degaussing of the media. Demagnetization avoids recovery of deleted data from a media.

7.4. Data Safeguards

The protection of information assets is mandatory to ensure compliance with legal, regulatory and contract agreement and derive the intrinsic value of data. In order to protect data, tax authorities should consider a combination of a legal and institutional framework, physical security, and logical security to safeguard data.

7.4.1. Legal and Institutional Framework

i. Taxpayer Confidentiality and Privacy Laws: Tax laws in many jurisdictions designate taxpayers' information as confidential and prohibit unintentional, unauthorized, and unlawful disclosure and theft.

ii. Exchange of Information (EOI) and Automatic Exchange of Information (AEOI) standards: EOI and AEOI standards ensure that data recipient tax authorities or Global Forum members (on transparency and exchange of information for tax purposes) are pre-assessed and post-assessed (periodically) to ascertain the adequacy of confidentiality laws and information security management framework to safeguard data.

iii. Security Policy: security policy is a comprehensive document that outlines the rules and procedures for accessing, using, and protecting information technology and data assets.

7.4.2. Physical Security

When data is stored in a datacentre on-premises, tax authorities should consider some or all of the following physical access and safety controls:

i. The use of Bolting Door Lock at the Datacentre entry door: Bolting Door Lock is a traditional control that requires the use of metal key to open and lock a door. The keys should be under strict control and duplication of keys should be prohibited.

ii. The use of Cipher Lock at the datacentre entry door: Cipher lock is a numeric keypad that requires the preset combination of access code to allow access to authorized persons. The access code should be changed at periodic intervals and when a member of staff is transfer or terminated.

iii. The use of Electronic Door Lock at the datacentre entry door: Electronic Door requires the use of a magnetic or embedded chip-based plastic access card to gain access to an enclosed space. Access card issuance and maintenance process should be carefully controlled. When an employee's service is terminated or a card is lost, such card should be deactivated.

iv. The use of Biometric Door Lock at the datacentre entry door: Biometric door grants access through any of the biometric features of the authorized person such as voice, retina, iris, fingerprint, or hand geometry.

v. The use of Deadman Door or mantrap: Deadman door or mantrap reduces the risk of tailgating wherein an unauthorized person follows an authorized person to gain unauthorized access to an enclosed space. Two doors are set up with a space between the doors where a single person fits. The first door must be closed and locked for the second door to open.

vi. The use of Closed-circuit TV (CCTV) camera at data centre entrance to monitor entry and exit. Videos and images must be kept for a period of time to facilitate any future investigation.

vii. The use of fire suppression to detect and suppress the spread of fire or heat. FM-200 is the most commonly used fire suppression gas.

viii. The placement of water detector under raised floor in data centre and computer rooms.

ix. The deactivation of USB ports on computer hardware to avoid the theft of data with USB drive.

7.4.3. Logical Security

Unlike Physical Security that requires the use of physical object to control access and secure resource, Logical Security controls access to and secures data and information technology resources on the basis of computerized logic.

7.4.3.1. Identification and Access control

Logical Access Controls can be classified as either Mandatory or Discretionary. Tax authorities should implement more Mandatory Access Controls (MAC) and less Discretionary Access Control (DAC). Under MAC, control rules are governed by an approved policy and users or data owners cannot modify the access role. DAC allows activation and modification of access control based on data owner discretion.

Tax authorities should ensure that logical access to data and information resource satisfies the following principles:

- **Identification**
  Identification is the ability to uniquely identify a user (an individual or system). An individual or system is enrolled with a username or biometric feature for subsequent recognition.
- **Authentication**
  Authentication is the process of verifying the claim of identity. It is the combination of identification and verification.

  There are 3 common factors of authentication:
  - Knowledge (something you know) – such as Password, PIN or Passphrase
  - Possession (something you have) – such as Access Card, Token, or phone
  - To Be (something you are) – such as Fingerprint, retina.

  Authentication of users can be single-factor (use of one of the above factors) or multi-factor authentication (use of two or all of the above factors). Single sign-on is a trending authentication technique that tax authorities can leverage as part of identity and access management program. Single sign-on permits the use of the same authentication information across multiple applications and systems. While the impact of a compromise of authentication information is severe, single sign-on reduces the risk of users writing down passwords where multiple passwords are used for multiple systems, and it is difficult

for each user to keep all passwords in memory. It also improves system administrator's ability to manage user accounts.

- **Authorization**
  Authorization is the level of access an identified and authenticated user can have or perform. It is best practice to institute Role-based access control (RBAC). RBAC allows access to data and system based on user role, job description and responsibility. It facilitates least privilege (exact access needed and no more) and need-to-know (authorized based on user needs) principles.

- **Accountability**
  Accountability is the capability to identify actions performed by each unique user who was granted privileges. Accountability is assured by the use of audit logs where every activity on data and system is monitored and recorded.

7.4.3.2. Data Encryption

One means by which tax authorities can secure data is to implement Public Key Infrastructure (PKI). PKI covers encryption of data while in transit or at rest. Encryption is the process of converting data into an unreadable form so it cannot be accessed or read by any unauthorized person. This unreadable data can again be converted into readable form by a process of decryption. Data is encrypted by the sender and decrypted by the recipient. There are many types of encryption algorithms. Some are AES, 3-DES, SNOW, RSA, Blowfish, Twofish etc. Encryption algorithms can be categorized into two types: Symmetric and Asymmetric Encryption. Symmetric Encryption involves the use of a single key to encrypt and decrypt data. Asymmetric Encryption, on the other hand, involves two keys – public key and private key. Either public or private key can be used to encrypt data, but decryption is only possible with the corresponding key.

Public Key Infrastructure facilitate adherence to the following digital identity principles:

a. Confidentiality – access to view or use data is granted to only the authorized person or system.
b. Authentication– is the verification of the identity of the source of the data.
c. Non-repudiation – is the indisputability of the source of data or action of a user.
d. Integrity – ensure that data is original, correct, and complete and is not modified by an unauthorized person or system.

The most efficient use of Public Key Infrastructure (PKI) is to combine the best features of asymmetric and symmetric methods. To achieve enhanced security of data, tax authorities should consider the use of both symmetric and asymmetric encryption in its PKI implementation.

7.4.3.3. Data Backup and Recovery

Tax authorities should institute a comprehensive Business Continuity and Disaster Recovery Planning. Business Continuity Planning (BCP) involves the conduct of risk assessment and business impact analysis to identify mission-critical business processes and system and identify risk to information assets.

As part of measure to ensure resilience under a full or partial BCP, tax authorities should adopt and implement a data backup and restoration strategy. Data backup and restoration strategy should be documented in a data backup policy. Data backup is the process of copying data to a separate device or remote location (generally) so that it may be used in the event of original data loss. There are many factors that may cause data loss –internal or external. Some factors include computer viruses, hardware failure, fire, natural calamities, and hacking attacks.

Generally, there are 3 types of backup strategy:

- Full Back up: the entire database is copied up every time, regardless of previous backups. The Backup consumes a lot of time and space, but it is the fastest in recovery.

- Differential Backup: only the new data created since last full backup is copied. It requires less time and storage capacity when compared with a full backup but requires more time and storage capacity than an incremental backup. On the other hand, it is faster for restoration when compared with incremental backup but slower in restoration when compare with Full Backup.

- Incremental Backup: only the new data created since the last full back up or incremental backup is copied. It requires less time and storage capacity when compared with a full backup and differential backup, but it is the slowest of the three in data restoration.

The frequency of data backup should be an important component of tax authorities' data backup plan. The frequency of data backup is determined by an organization Recovery Point Objective (RPO). RPO is the measure of an organization tolerance for data loss. For example, an RPO of 0 hours indicates that data loss is unacceptable, and backup procedure is carried out in real-time. Similarly, an RPO of 6 hours indicates an organization accepts a maximum data loss of 6 hours and will carry out backup procedures every 6 hours.

It should be noted that there is a trade-off between cost and business resilience. The more resilient tax authorities become (low RPO), the higher the cost of operation and maintenance.

7.4.3.4. Network security

In today's digital world, access to data and resources of tax authorities is mostly accomplished through network connection. Whether data is stored in a datacentre on-premises or in the cloud, the network to access data must be secured.

a. Use of a Firewall

The use of firewall to secure the network is common and necessary practice in many organizations and tax authorities must ensure that network is secured with a firewall. A firewall is a device or software that monitors and controls incoming and outgoing network traffic as per defined rules. It is designed to allow authorized users and disallow unauthorized users. Application-level firewall is the most secure and highly recommended for tax authorities to protect network and data.

In the definition of firewall policy, tax authorities should ascertain the trustworthiness of its sources of network traffic. It is recommended to implement a Default Deny Access Control Policy where network traffic is from untrusted sources. Default Deny Policy restricts all network traffic and allows only pre-approved traffic. Unlike Default Deny, Allow All policy allows all traffic except for predefined restricted traffic.

b. Use of a Virtual Private Network

Virtual Private Network, VPN for short, allows remote access to data and resources through a secure channel using the internet. It extends a private network over the internet in a secure manner. The VPN server is configured at head offices or branches and a client software is installed on end-user computers. End-users can connect to data or resources at head office or branches from a remote location by logging into the VPN Client application on their laptop or desktop.

The VPN encrypts and encapsulates data in a tunnel when in transit over the internet to safeguard the data from intruders. It is a cost-effective option as it relies on public infrastructure (public internet) to transmit data. The use of a dedicated lease line is another option for remote communication, but it is very expensive.

7.4.3.5. Education and Awareness Program

While automated controls are highly recommended to safeguard data, they alone cannot prevent or mitigate risk to data assets. Security awareness programs and training should be

conducted for tax authorities' staff and other stakeholders to play a key role in mitigating information security risk.

Tax authorities' stakeholders should be educated on various aspects of security events to minimize any impact of security breaches. Security awareness programs should include topical areas of the security policies including password standard, email usage, internet usage, social engineering, and other relevant factors.

Tax authorities can offer security awareness to staff and other stakeholders by any of the following ways:

- Workshop and training programs;
- Security tips via email;
- Documented security policies and procedures;
- Non-Disclosure Agreements with employees and third-party vendors;
- Awareness through newsletters, posters, screensavers, and suchlike;
- Documented security roles and responsibilities; or
- Simulated drills and security scenarios.

Security Awareness programs are most effective in curbing social engineering attacks. Social engineering attacks are less sophisticated but rely on human intelligence, that is, the ability of the user to identify fake information designed to grant unauthorized users access to information or network. Baiting, scanware, pretexting, phishing and spear phishing are common social engineering attacks.

7.5. Data policy

Tax authorities collect a lot of data from natural and legal persons. These data are used for operational purposes and are routinely exchanged with other tax authorities and third parties. TAs have a responsibility to ensure confidentiality of data obtained by national laws, bilateral and multilateral information exchange agreements. Data Policy is a must-have for tax authorities to meet with their confidentiality responsibility and at the same time derive value from data.

Data Policy is a document that outlines guidance for the management of tax authorities' information assets in accordance with laws and regulations. It covers the data life cycle and offer guidelines on attributes of data quality, roles and responsibilities, access, usage, security and privacy.

Data Policy should be developed by a high-level committee comprising of senior executives and process owners of tax authorities The committee should oversee the implementation of the policy. At a high-level, data policy should cover the following:

a. Background.
b. Policy Purpose.
c. Policy Scope.
d. Policy Principles.
e. Roles and Responsibilities.
f. Review Process.
g. Resources.
h. Contacts.
i. Terms and Conditions.

The effectiveness of a Data Policy depends on how well the policy is communicated to all stakeholders. It should be communicated to all relevant stakeholders. Communication may be achieved through meetings, training, seminars, focus-group discussions and email circulation.

CASE STUDIES ANNEX

A. Kenya's Data Protection Act, 2019[9]

Kenya's Data Protection Act, 2019, serves as a comprehensive data governance framework designed to protect personal data and ensure privacy. The act establishes the Office of the Data Protection Commissioner, which oversees the implementation and enforcement of data protection laws.

- **Policies and Standards**: The act outlines clear policies and standards for data processing, ensuring that personal data is handled lawfully, fairly, and transparently. It mandates Tax administrations to obtain consent from individuals before collecting their data and to use the data only for specified purposes.

- **Data Stewardship:** The act assigns specific roles and responsibilities to data controllers and processors. Data controllers determine the purposes and means of processing personal data, while data processors handle the data on behalf of the controllers. Both parties are required to implement appropriate technical and organizational measures to protect data**.**

- **Data Quality Management**: The act emphasizes the importance of data accuracy and integrity. Tax administrations are required to ensure that personal data is accurate, complete, and up to date. Individuals have the right to request corrections to their data if it is inaccurate or incomplete.

- **Data Security and Privacy**: The act mandates Tax administrations to implement robust security measures to protect personal data from unauthorized access, loss, or destruction. This includes encryption, access controls, and regular security assessments. Additionally, the act provides individuals with rights to access their data, request its deletion, and object to its processing.

- **Data Lifecycle Management**: The act requires Tax administrations to manage the entire lifecycle of personal data, from collection to deletion. Data should be retained only for as long as necessary to fulfill the purposes for which it was collected. Once the data is no longer needed, it should be securely deleted or anonymized.

- **Compliance and Audit**: The Office of the Data Protection Commissioner conducts regular audits and compliance checks to ensure that Tax administrations adhere to the data

[9] Kenya Data Protection Law (2019) DataProtectionAct24of2019.pdf (kenyalaw.org)

protection laws. Non-compliance can result in significant penalties, including fines and imprisonment.

Kenya's Data Protection Act, 2019, serves as a model of a comprehensive data governance framework in a developing country. By establishing clear policies, assigning roles and responsibilities, and implementing robust security measures, the act ensures the protection of personal data and enhances data-driven decision-making. As data continues to play a pivotal role in the digital economy, the importance of effective data governance frameworks cannot be overstated.

In-depth analysis of Kenya's Law[10]

What does the Act say?	**What are the potential implications?**
'Consent' to the processing of personal data by the data subject must be an express, unequivocal, free, specific and informed indication of the data subject's wishes by a statement or by a clear affirmative action.	It is clear that data controllers and data processors will no longer be able to rely on implied consent to process personal data. However, it is not yet clear whether or not a company may be able to rely upon pre-ticked boxes or any other default method of consent, or whether or not a positive opt-in will be required instead. We would recommend that data controllers and data processors review their existing consent practices.
The definition of '*sensitive personal data*' has been widened to include property details, marital status and family details, including names of the person's children, parents, spouse or spouses.	The collection, processing and transfer of sensitive personal data is a category of data that is afforded a higher level of protection (as is the case throughout the world). By extending the definition to include property details, this could possibly include a person's physical address provided for the purposes of receiving a delivery, for example. Companies may need to review the

[10] Bowmans SNAPSHOT: Analysis of the Data Protection Act 2019 | Bowmans (bowmanslaw.com)

	information collected from data subjects and the manner in which such information is used; for example, address or honorific titles (which could indicate marital status).
The Act applies to the processing of personal data by a data controller or data processor by automated or non-automated means. Where personal data is processed by non-automated, means, the Act applies where the recorded data forms a whole or part of a filing system by a data controller or data processor who: (a) is established or ordinarily resident in Kenya and processes personal data while in Kenya; or (b) not established or ordinarily resident in Kenya, but processes personal data of data subjects located in Kenya.	The previous versions of the Act applied certain geographical provisions and qualifications to processing by both automated and non-automated means. The geographical restrictions now only expressly and specifically apply to processing by non-automated means. As a result and in the absence of these clear geographic references, there is arguably some ambiguity as to whether or not the Act applies to foreign data processors or data controllers. However, we would err on the side of caution and recommend that all data controllers and data processors carrying out any processing activities involving the personal data of Kenyan data subjects, ensure that they comply with the provisions of the Act.
All data controllers and data processors (processing by both automated and non-automated means) must hold a valid registration with the Data Commissioner. The Act specifies the information to be provided by the data controller and data processor in the application for registration.	There are clear indications throughout the Act that data controllers and data processors must have adequate and sufficient safeguards, security measures and mechanisms in place (although this obligation may be somewhat tempered by the amount of personal data collected, the cost of processing, and the extent of processing activities, which may be of comfort to small and medium sized enterprises).

	Included within the application requirements is a new proviso requiring the applicant (i.e. the data controller or data processor) to indicate what measures are in place to indemnify the data subject from unlawful use. The indemnification obligation is a further sign that data controllers and data processors will be held accountable for any encroachment of a data subject's rights and interests to his or her personal data.
The principles of data protection are well ensconced in the field of data protection legislation throughout the world. The principles, as set out in the Act, are similar to those applying to international standards (the GDPR in particular). Importantly, the Act also contains a new data protection principle prohibiting the transfer of data outside of Kenya unless there is proof of adequate data protection safeguards or, consent from the data subject has been obtained.	All data controllers and processors must adhere strictly to the principles of data protection. These principles must form the backbone of an organization's standard operating procedures when it comes to the collection, processing, storage and use of personal data. The cross-border transfer of personal data is a sensitive topic, with opposing views on the issue voiced by public bodies, business organizations and individuals. The concerns touch on security and technical considerations, compatibility with the current digital global marketplace within the context of multinational organizations, and the right of a data subject to determine where his or her data should be stored. This data protection principle that we have highlighted here will need to be interpreted alongside the cross-border transfer provisions in the Act; additionally, there are certain ambiguities in the drafting of this provision that may cause some confusion when implementing the Act.

	Finally, the Act gives certain rights to the Data Commissioner to suspend or prohibit any cross-border transfers. Further, the Cabinet Secretary may prescribe, on grounds of strategic interests of the State or for protection of revenue, that certain types of processing be effected through a server or data center located in Kenya.
A data controller or data processor *may* designate or appoint a Data Protection Officer (DPO) where the processing is carried out in the context of certain activities, for example, where the core activities of the data controller or data processor require the regular and systematic monitoring of data subjects.	The primary role of the DPO will be to ensure that the relevant organisation processes personal data in compliance with the provisions of the Act. The Act specifies the minimum qualification criteria that the DPO must hold. It will be possible for Group entities to appoint a single DPO but the DPO must be accessible by each entity. International data controllers and data processors who do not have a presence in Kenya may consider appointing a local DPO to deal with any compliance issues that may arise. It should be highlighted that the requirement to appoint a DPO is not an absolute obligation.
Any persons processing the personal data of a child will be required to take steps to incorporate appropriate mechanisms for age verification and consent.	The choice of mechanisms to be incorporated can be guided by the available technology, the proportion of such personal data that is likely to be processed and the volume of personal data to be processed. That said, a data audit may enable companies to determine whether or not specific actions should be implemented prior to the processing of any personal data relating to a child.

The Act sets out a prescribed response to be provided where an unauthorized person has accessed or acquired any data and there is a 'real risk of harm' to the data subject whose personal data has been accessed. The Data Commissioner must be notified within 72 hours of becoming aware of the delay and, unless provided for otherwise under the Act, the data controller must then communicate the occurrence of such breach to the affected data subject in writing within a reasonably practicable period.	There are detailed requirements as to what the notice to the Data Commissioner and data subject must entail, including where applicable, the identity of the person who may have accessed or acquired the personal data.
There are new provisions which seek to protect the processing of health data and impose restrictions on who may collect such data.	Companies collecting health data, such as pharmaceutical companies, must ensure that the data is collected in accordance with the Act. It specifically restricts processing to be carried out by or under the responsibility of a healthcare provider or by a person subject to the obligation of professional secrecy.
Data subjects are entitled to file complaints with the Data Commissioner. Where the Data Commissioner finds that an individual or organisation is found to have failed or is failing to comply with the provisions of the Act, the Data Commissioner	The enforcement notice will set out the steps to be taken to remedy any failure identified and the period within which the breach must be remedied. The Data Commissioner has a further right to issue a penalty notice to accompany the enforcement notice. The amount of the penalty

will have the right to serve an enforcement notice on the party in breach as well as a penalty notice.	will be determined by various factors relating to the severity of the failure, the duration of the failure, the degree of cooperation with the Data Commissioner, and the manner in which the Data Commissioner was notified of the breach. For example, it would be preferable for the data controller or data processor to have notified the Data Commissioner of the breach before the data subject did. In addition, the Data Commissioner may impose an administrative fine for an infringement of the Act. The maximum amount of the penalty that may be imposed in a penalty notice is KES 5 million or, in the case of an undertaking, up to 1% of annual turnover in the preceding financial year, whichever is lower.
The Act makes provision for certain regulations that may be further prescribed by the Cabinet Secretary under the Act. This includes further requirements to be imposed on a data controller or data processor when processing personal data, regulations on the processing of data through a data server or data center in Kenya and the issuing of codes of practice and guidelines; for example, a code of practice containing practical guidance in relation to the processing of personal data for	It is clear that the Data Protection Act will be a moving piece of legislation and whilst the Act will provide the framework for compliance, we would expect that additional regulations and the codes of practice that may be issued in the future will affect the way in which the Act will be implemented according to the sector, the undertaking in question and the data processing activities undertaken.

purposes of Journalism, Literature and Art.	

B. European Union Data Governance Frameworks[11]

- **Introduction**

The European Union (EU) has established a comprehensive set of data governance frameworks to ensure the secure, efficient, and ethical use of data. These frameworks are designed to foster innovation, protect individual privacy, and promote economic growth. Key among these are the General Data Protection Regulation (GDPR), the Data Governance Act (DGA), and the Open Data Directive.

- **General Data Protection Regulation (GDPR)**

The GDPR, implemented in 2018, is one of the most stringent data protection regulations globally. It aims to protect the personal data of EU citizens by setting high standards for data privacy and security. The regulation applies to all Tax administrations processing personal data of individuals within the EU, regardless of the organization's location. Key principles of the GDPR include data minimization, purpose limitation, and the rights of individuals to access and control their data. Non-compliance can result in hefty fines, making GDPR a critical consideration for businesses operating in the EU.

- **Data Governance Act (DGA)**

The DGA, adopted in 2022, aims to enhance data sharing across the EU by creating a framework for the reuse of certain categories of data held by public sector bodies. It also regulates data intermediation services and promotes data altruism, where individuals and Tax administrations voluntarily share their data for the common good. The DGA supports the creation of Common European Data Spaces in strategic sectors such as health, environment, and mobility, facilitating cross-border data flows and fostering innovation.

- **Open Data Directive**

The Open Data Directive focuses on the reuse of public sector information, making data held by public bodies more accessible and usable. This directive encourages the development of new products and services by providing businesses and individuals with access to valuable datasets. The directive also promotes transparency and accountability by ensuring that public sector data is available to the public.

- **European Data Innovation Board (EDIB)**

The EDIB, established under the DGA, plays a crucial role in facilitating the development of data-sharing practices and standards across the EU. The board brings together representatives

[11] https://commission.europa.eu/system/files/2020-07/summary-data-governance-data-policies_en.pdf

from member states, the European Commission, and other stakeholders to promote best practices and ensure the interoperability of data systems.

The EU's data governance frameworks are designed to create a robust and trustworthy data economy. By balancing the need for innovation with the protection of individual privacy, these frameworks aim to foster a data-driven economy that benefits both businesses and citizens.