ANNEX 2

PART 3 – DATA GOVERNANCE FRAMEWORK

Contents

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 byproduct 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.

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.

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. This part of the guide is dedicated to discussion on the guiding principles, standards, policies, structures and procedures of data collection, processing, storage, classification, use, security, archival and destruction.

CHAPTER 5 – 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

5.1 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.

5.2 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.

5.3 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' behavior. Tax evasion, fraud, under-declaration can all be detected using unstructured data, as much as structured data.

5.4 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. Instill 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.

5.5 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 analyzing 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 analyze 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' behavior. 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.

5.6 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 analyze 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 analyzing 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;
- Preprocess and clean data;
- Classify and segment data;
- Visualize data; and
- Draw conclusion.

5.7 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 document 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.

5.8 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:

i. 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.

ii. 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 minimum. Tax authorities rations 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.

CHAPTER 6 – USE OF DATA

6.1 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 center 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).[1]

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

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

infrastructure, including network, servers, Operating Systems (OSs), or any other application-related issues. The consumer just uses the software as a service on demand.

- 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.[2]

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.

6.2 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.

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

6.3 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 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.

6.4 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.

6.4.1 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.

6.4.2 Physical Security

When data is stored in a datacenter 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 Datacenter 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 datacenter 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 datacenter 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 datacenter 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 center 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 center and computer rooms.
ix. The deactivation of USB ports on computer hardware to avoid the theft of data with USB drive.

6.4.3 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.

6.4.3.1 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.

6.4.3.2 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.

6.4.3.3 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.

6.4.3.4 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 datacenter 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.

6.4.3.5 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.

6.5 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.