Data Governance Toolkit for NGOs

pwc

Content

Data is a valuable asset to all businesses and organisations, including Non-governmental organisations (NGOs). NGOs generate and capture data in each and every aspects of their operations, from internal administration, service delivery, to engagement with donors, volunteers and other stakeholders. Data gathered from these diversified sources, if harnessed properly, could be used to support the attainment of NGOs' strategic objectives and enhance their overall operations.

With the increasing focus of Hong Kong being a Smart City and the implementation of various data-driven public initiatives, it is important for NGOs as the key partners in social services delivery to understand and appreciate the value of data and for that matter, benefits of good data governance; and start planning to become a data-driven organisation.

To support NGOs to jumpstart this process, PwC Hong Kong (PwC) has developed this data governance toolkit primarily for NGOs which may wish to:

- understand the general concept and comprehend the value and benefits of data governance;

- learn how good data collaborative practices could improve the effectiveness and efficiency of services and/or overall operations; and

- have an initial understanding of the organisation's data governance maturity level and based on which plan for future improvements.

What is data governance?

Data governance is a collection of practices and processes which enable structured governance and management of data assets to support attainment of strategic objectives of an organisation. Data governance, in principle, consists of the following key tasks and activities:

- definition of roles, responsibilities and an operating model to ensure a proper governance structure for data management;

- establishment of data policies, procedures, standards and metrics;

- compliance to data policies, procedures and standards by establishing a proper workflow;

- management of data related risks to ensure data quality;

- monitoring and resolution of data related issues, such as,

  – duplication of data (e.g. manual registration of programmes with some in paper forms while others are handled via digital channels);

  – data inaccuracy (e.g. information provided by data subjects is not always cross-checked against proof of identity unless with specific service/programme requirements); and

- cultivation of a culture of data governance and, therefore, creation of a data-driven organisation.

2   Definition of Data

Data, which is usually categorised as structured, semi-structured or unstructured data, consists of information, facts and figures gathered for the purposes of reference and/or analysis. NGOs which would like to understand the value of their data must start by identifying and categorising the data they already have and data they would like to collect. The process of categorising existing data provides an overview of data across different operating units and functions.

Data Sources

Captured Data
- Data collected through data entry into a system by users.
- Records captured via software applications, e.g. online donation/social service registration.

Processed Data
- Generated by combining, aggregating and processing other data, e.g. aggregate service needs analysis, data subject characteristics profiling.

Data Categories and Examples

Master Data
- People, places, and other data items that are critical to an organisation's operations.
- Data of service users/funders/volunteers/staffs (e.g. name, contact number, e-mail address, correspondence address, HKID number)

Administrative Data
- Transactions executed within the organisation and those with external parties/entities.
- Accounting and finance (e.g. programme-related payments, donations received)
- Human resources (e.g. staff recruitment)
- Programme execution (e.g. co-payment by service users)
- Facilities usage (e.g. Reading corner of the centre)

Activity Data
- Activities of or interactions with an individual or entitie (e.g. activities of a service user or other relevant stakeholders).
- Service user initiated (e.g. programme enquiries or registration, attendance registration of after-school care programme)
- Fundings, donations and sponsorship agreements (e.g. Key Performance Indicators (KPIs), programme delivery terms and conditions)

While NGOs may collect large amount of data on a daily basis, intrinsic value of data comes from the ability of the organisation to use those data assets to achieve well defined strategic objectives and driving incremental benefits along the way. Furthermore, it is important to note that for data to have value, it must be identifiable, definable, accurate and timely with good integrity and quality.

3   Data Journey

All organisations which aspire to become a data-driven entity would go through a data journey, each with its own pace and setting. Data journey is a development path in their quest to develop capabilities taken into consideration of their data vision, operating environment and resources constraints. Data journey typically covers four key stages: starting with the formulation of a data strategy; establishment of data governance; implementation of data architecture; to adoption of analytics and/or AI.

The following explains the objective and tasks of each stage of the data journey.

1. Data Strategy

Definition of an organisation's data vision to become data-centric or data-driven which requires conducting the following tasks:

- definition of data missions and initiatives;
- assessment of data maturity;
- identification of potential opportunities to become a data-driven organisation; and
- development of short-term, mid-term and long-term roadmap.

2. Data Governance

Establishment of a data governance structure with the objective of aligning management of data with the strategic objectives of the organisation and enhancing support to critical operation(s). Requisite tasks include defining the:

- target operating model;
- data governance model;
- data quality framework; and
- data policies and standards.

3. Data Architecture

Implementation of data architecture with cloud technologies to enable streaming and processing of large volumes of data into a single source as an integrated inventory. Requisite tasks for this stage include:

- selection of platform/technology/tool;
- design of enterprise architecture;
- preparation of data flow and data exchange readiness; and
- assessment of cloud readiness.

4. Analytics and/or AI

Adoption of data analytics and/or AI tools to strengthen capabilities, enhance operation efficiencies and improve service delivery, for example:

- self-service analytics;
- process and performance analytics; and
- descriptive, diagnostic, predictive and prescriptive analytics.

4 Value of Data

Data is a valuable asset to businesses and organisations, including NGOs. What is the value of data? Why is data important?

Value of data, in general, can be recognised through five perspectives:
- facilitate effective decision making;
- support evidence-based process improvement;
- support reliable measurement of efficiency of service provision;
- offer insights on stakeholders' feedback; and
- enable transparency in reporting, hence, building trust with stakeholders.

The following diagram sets out some of the key benefits for an NGO in more detail.

Facilitate decision making

- Data can help organisations to gain a better understanding of their service users and service needs of the society; monitor effectiveness and efficiency of internal processes; and provide fact-based information to facilitate timely decision making.
- Trusted data asset is critical in facilitating organisation to leverage the value of data.Understanding areas that need prioritising and targeted resources investment is essential in helping organisations to grow and progress.

Measure efficiency of service provision

- Performance related data enables better understanding of performance of different operating units/functions and overall service delivery (e.g. KPIs on homecare system and fundraising campaign).
- Provides data-driven analysis on organisations' operational capabilities with a view to leverage those strengths across different operating units/functions; and identify inefficiencies and bottlenecks.

Enable transparency in reporting

- Well-maintained and quality data assets enable an accurate and trusted measurement of value of services provided to users through more advanced outcomes/impact measurement.
- Data-driven analysis offers transparency of performance (output/outcome/impact) of programme supported by funders which helps to instill trust in the capability of the organisation.

Support process improvement

- Proper utilisation of data assets facilitates optimisation of operational processes (e.g. time and resources spent on each key tasks of service delivery).
- A well designed database minimises unnecessary duplication of data entry, hence, enable staff to focus on important tasks/activities (e.g. a centralised master dataset enables centre staff to do single sign-on for service users registered with multiple programmes).

Offer insights on stakeholders feedback

- Well developed data collection process enables organisations to analyse users' feedback and extract insights for service improvement and planning.
- Analysis on donor's preference and response rate could improve performance of fundraising campaigns.

5   Components of Data Governance

Data governance can be dissected into different components with each having its own well-defined requisite tasks and activities. The ability and extent of the completion of each required tasks/activities will determine the maturity level of data governance of an organisation. The diagram below provides an overview of the eight components of data governance.

Governance Framework

| Low | /////// | Maturity | *
|---|---|---|

- Data governance strategy and principles
- Process flow, issue escalation procedures, classification of decision making rights and documentation

Policy and Standards

| Low | /////// | Maturity | *
|---|---|---|

- Control policies, standards for processes and procedures, <u>metadata</u> standards, format and definition
- Data issue resolution process (e.g. what are the policies and procedures that organisations should follow when data issue arises)

Data Quality

| Low | Medium | /////// | Maturity | *
|---|---|---|---|

- Establish data quality framework
- Setup guidelines, operation rules and thresholds (KPI) for data quality monitoring and reporting

Data Architecture

| Low | Medium | High | Maturity | *
|---|---|---|---|

- Data infrastructure architecture
- Data integration
- Data security

Master Data Management

| Low | Medium | High | Maturity | *
|---|---|---|---|

Policies and procedures to manage:
- Master data
- Administrative data
- Activity data

Page 10 provides a simple assessment for NGOs to evaluate their data governance maturity level along the Data Governance Maturity Steps. Based on the result of the initial assessment, NGOs can make reference to the recommended practices on how to enhance their maturity through implementation of requisite tasks and activities.

Culture and Change

| Low | Medium | /////// | Maturity | *
|---|---|---|---|

- Identify and assign "Champions" to drive change, training, communication and awareness
- Adopt self-service analytics tool
- Upskill staff

Tools and Templates

| Low | Medium | /////// | Maturity | *
|---|---|---|---|

Adopt tools or templates to manage:
- <u>Data inventory</u>
- <u>Critical data elements</u>
- <u>Data lineage</u>
- <u>Data standards</u>
- <u>Data profiling</u> and <u>data mapping</u>
- Data quality

Operating Models

| Low | Medium | High | Maturity | *
|---|---|---|---|

- Target operating model
- Data organisation structure
- Roles and responsibilities (e.g. data ownership)
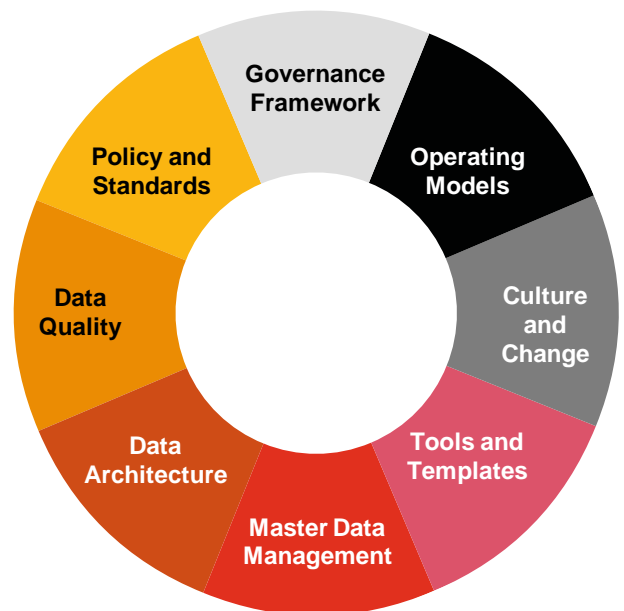
* Denotes the data governance maturity level assigned to each principle.
Please refer to the Data Governance Maturity Steps for details.

Data Governance Toolkit for NGOs   6

6   Data Quality Dimensions

In order to maximise the value of data, it is essential that data captured is of good quality. NGOs can review quality of data through eight Data Quality Dimensions. These dimensions are important yardsticks to gauge and assure the quality of their data. Insights to enhance operations and decision making can only be harnessed from good quality data.

The diagram below describes the eight Data Quality Dimensions and what each of those prescribes in connection to the setup and operation of an NGO.

Timeliness
- Data must be made available in a timely manner with respect to the intended applications; data must be updated to the extent possible in consideration of its relevancy; date of last update should always be recorded for the purpose of reference and clarity of audit trails.

Accuracy
- Accuracy is defined as the degree of data trueness and precision which could correctly describe the 'reality'.
- Data should be accurately captured at the data creation stage (e.g. incorrect data entry should be revised at the source system instead of the downstream data warehouse to ensure consistency of data accuracy across the data lineage).

Completeness
- Data must be complete (e.g. service records should capture all services provided to all users, 20 homecare visits delivered should generate 20 service records and not less or more).

Consistency
- Consistency refers to the extent of which data complies with its specified format and values of variables (e.g. if the gender of the data subject is specified as either male or female, the data recorded should not have other values like Y or N).

Uniqueness
- Extent of uniqueness in which data should be recorded and stored (e.g. same data subject should not have duplicate identifiers; service record for the same service user should have the same unique reference ID across different operating units).

Integrity
- Linkages between data (within or across applications/databases) are properly established and integrated (e.g. service records must align with a validated service user ID).

Compliance
- Personal data that can identify an individual must be protected by policies/processes which comply with relevant regulations such as the Personal Data (Privacy) Ordinance (the "PDPO") and the Company Ordinance.
- An audit trail needs to be put in place to track authorised access.

Security
- NGOs should take all practicable steps to protect data (especially personal data and identifiable data) against unauthorised or accidental access, processing, erasure, loss or use. Access rights should only be granted on a need-to-have basis.

Building Blocks of Data Governance Capabilities

Well developed data governance of an organisation requires solid support from four important building blocks of capabilities, namely, people, processes, technologies and metrics. Design and execution of each building block are based on the overarching data governance regulatory compliance requirements, ethical considerations and the organisation's data vision and objectives.

The following diagram sets out what each building block encompasses.

The following diagram (house structure) contents:

Governance and Oversight

Senior Management

People	Processes	Technologies	Metrics
Staff responsible for enforcing the management of data and administering data related activities	Policies and standards that govern creation, management and usage of data	Standardised templates and tools used to monitor creation, management and usage of data	Measures for monitoring data quality and overall governance in achieving a data-driven organisation
Operating Model	Governance Processes	Systems	Key Performance Indicators
Roles and Responsibilities	Data Issue Management	Data Tools	Segregation of Duties Matrix
Data Ownership	Data Quality Monitoring	Data Templates	Data Quality Metrics

Data Governance Regulatory Requirements and Ethical Considerations

Data Governance Vision and Objectives

The following are some examples of metrics:

- Key Performance Indicators – total issue identified, total training conducted
- Segregation of Duties Matrix – authority to create, authority to change
- Data Quality Metrics – completeness, accuracy, timeliness

Data Governance Maturity Steps

Data Governance Maturity Steps is an overview of the development pathway of an organisation in its' quest to implement data governance. Data gov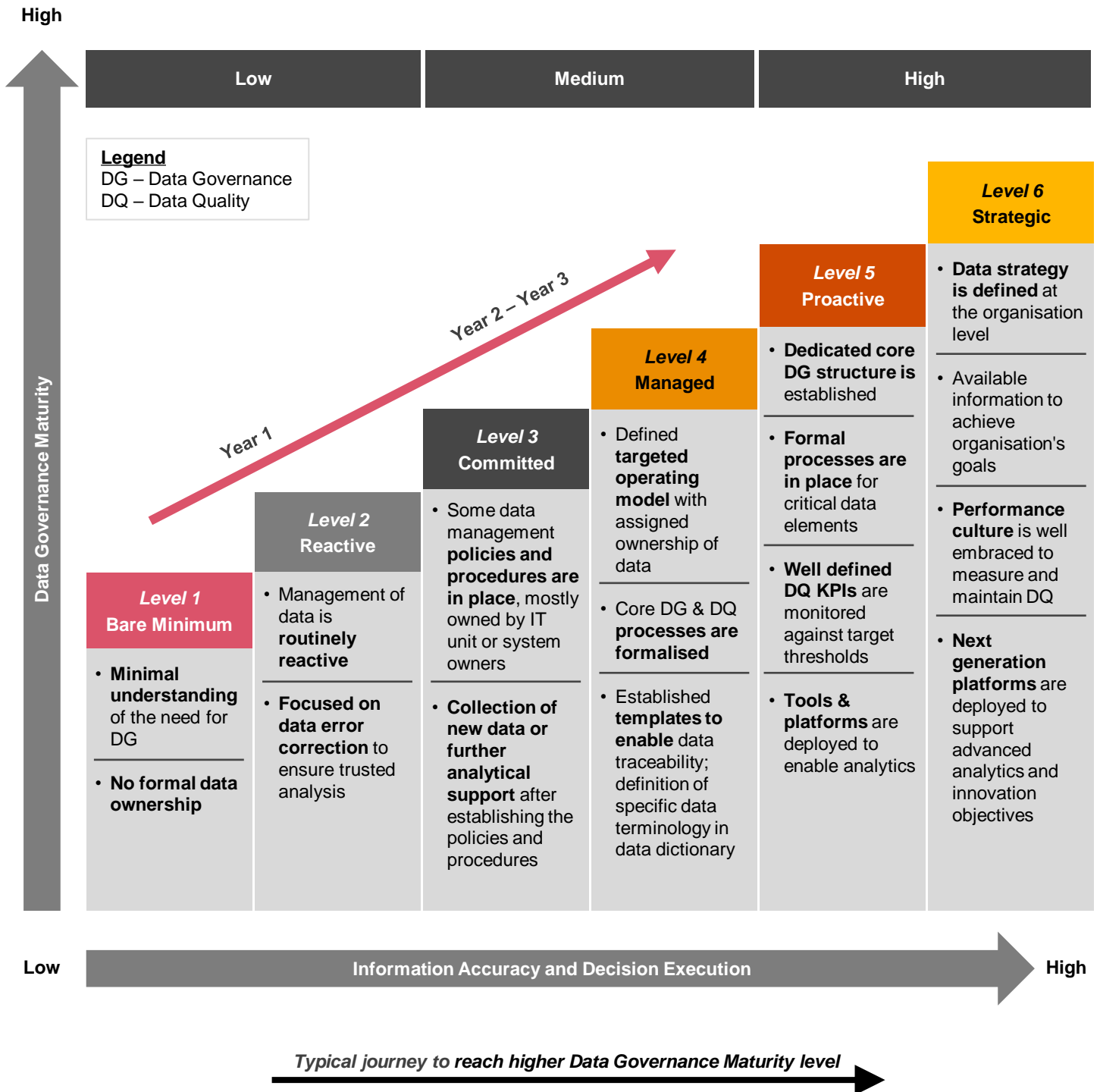ernance maturity levels of different operating units may vary within an organisation. In order to govern and manage data effectively and efficiently, it is recommended that NGOs should achieve Level 3 or Level 4 within a defined timescale taking into consideration of its resources and capabilities availability.

High

	Low	Medium	High

Data Governance Maturity

Legend
DG – Data Governance
DQ – Data Quality

Year 2 – Year 3

Year 1

Level 6
Strategic
- **Data strategy is defined** at the organisation level
- Available information to achieve organisation's goals
- **Performance culture** is well embraced to measure and maintain DQ
- **Next generation platforms** are deployed to support advanced analytics and innovation objectives

Level 5
Proactive
- **Dedicated core DG structure is** established
- **Formal processes are in place** for critical data elements
- **Well defined DQ KPIs** are monitored against target thresholds
- **Tools & platforms** are deployed to enable analytics

Level 4
Managed
- Defined **targeted operating model** with assigned ownership of data
- Core DG & DQ **processes are formalised**
- Established **templates to enable** data traceability; definition of specific data terminology in data dictionary

Level 3
Committed
- Some data management **policies and procedures are in place**, mostly owned by IT unit or system owners
- **Collection of new data or further analytical support** after establishing the policies and procedures

Level 2
Reactive
- Management of data is **routinely reactive**
- **Focused on data error correction** to ensure trusted analysis

Level 1
Bare Minimum
- **Minimal understanding** of the need for DG
- **No formal data ownership**

Low ←→ **Information Accuracy and Decision Execution** → **High**

*Typical journey to **reach higher Data Governance Maturity level***

9   Data Governance Maturity Assessment

NGOs are encouraged to perform this Data Governance Maturity Assessment

Data governance maturity assessment is a quick evaluation tool with 10 simple questions that enable NGOs to have a general understanding of their maturity along the data journey. It also indicates where they stand in managing and using data to support and enhance their operational performance.

Scoring from this simple assessment provides a preliminary baseline for an organisation to define future improvement plan referencing good practices of data governance set out in the following sections.

NGOs are reminded to ensure proper execution of the foundational practices recommended for lower level maturity even when they have attained a higher level maturity score based on the assessment.

	Mark your score	
	Yes =1	No = 0
1. Does your organisation know what "Data" is and treat it as an asset?		
2. Does your organisation have a vision/view on how to use data?		
3. Does your organisation assign resources with defined responsibilities to manage data quality?		
4. Does your organisation have specific policies or standards in relation to data management?		
5. Does your organisation have resources and the capability to review compliance of regulatory data requirements?		
6. Does your organisation use any centralised systems/platforms to store the data?		
7. Does your organisation have a methodology framework or organisation structure to manage the data?		
8. Does your organisation ensure proper access to data is being managed and monitored?		
9. Does your organisation use any data analytics tools/platforms to generate reports?		
10. Does your organisation have any data analytics focusing on specific service delivery?		

Legend:

Low Maturity:	Total score 0-5	**Low** ///////// **Maturity**
Medium Maturity:	Total score 6-8	**Low** **Medium** ///// **Maturity**
High Maturity:	Total score 9-10	**Low** **Medium** **High** **Maturity**

Recommended Data Governance Practices

Essentially, there are eight good data governance practices. While each recommended practice requires execution of specific tasks and activities, advancement of an organisation along the Data Governance Maturity Steps could only be possible if implementation of these practices are aligned in terms of completeness and sophistication.

01 Develop Data Principles

Low ///// **Maturity**

Development and implementation of data principles which, at least, enable minimum compliance and regulatory requirements of personal data and other sensitive information (e.g. financial, health). Data principles should apply to all data policies and procedures within the organisation.

02 Define Roles and Responsibilities

Low ///// **Maturity**

Establish data governance roles and responsibilities which includes deploying those staff already executing data related activities to take on data governance as part of their official day-to-day responsibilities.

03 Establish Data Quality Monitoring Procedures

Low ///// **Maturity**

Establish continuous data quality monitoring controls. Create a data issue log to record all quality issues and conduct continuous assessments of data quality standard.

04 Establish Core Data Governance Workflow

Low ///// **Maturity**

Implement core data governance controls and workflow as a safeguard to ensure foundational processes for data activities are in place.

05 Establish a Data Inventory

Low   Medium ///// **Maturity**

Establish a structured and well-defined data inventory mechanism to record data collection touchpoints, definition of each data field, variables and descriptors of information, usage (e.g. operational, analytical, marketing), Personally Identifiable Information (PII) or non-PII and corresponding data subject consent.

06 Identify Critical Data Elements

Low   Medium ///// **Maturity**

Identify and prioritise critical data elements for effective operation. Data elements which could have material impact on operations or performance with respect to service delivery, need to be governed with more controls (e.g. address, preferred time and mode of hot meal delivery for seniors).

Eight Recommended Data Governance Practices for NGOs
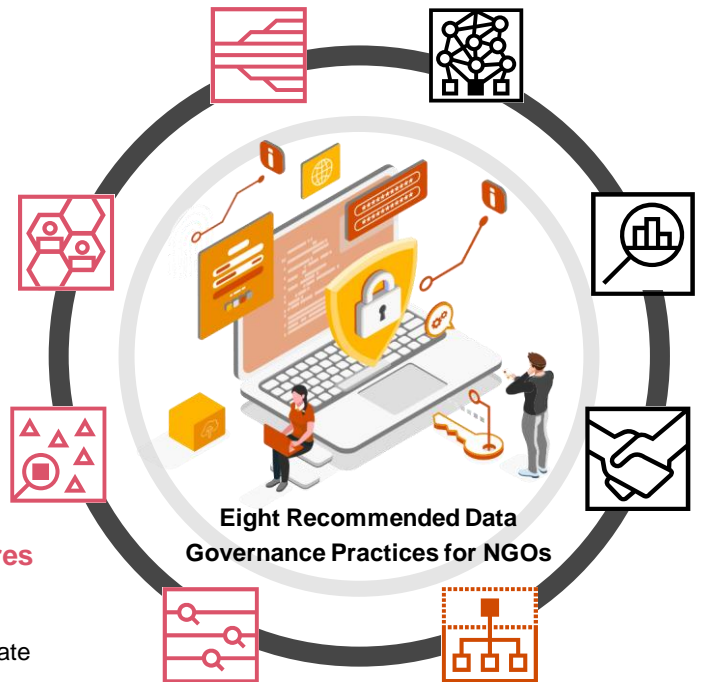
07 Promote a Data Governance Culture

Low   Medium ///// **Maturity**

Foster a data governance culture across divisions/departments through workshop and training on operation differentiation and strengthening of trust through rigorous approach to data governance and protection of personal information of data subjects; and leveraging insights from data analytics for improvement of overall service delivery.

08 Define Data Operating Model

Low   Medium   High   Maturity

Adopt a centralised model to align the strategic value of analytics; define requisite policies and standards; enhance data transparency and sharing of best practices.

01   Develop Data Principles

Recommendations

Development and implementation of data principles which, at least, enable minimum compliance and regulatory requirements of personal data and other sensitive information (e.g. financial, health). Data principles should apply to all data policies and procedures within the organisation.

Develop Data Principles, for example,
1. Data is a core asset of an NGO and needs to be managed accordingly.
2. Data shall be governed by individuals with designated roles and responsibilities.
3. All operating units and functions shall develop and upskill the staff's data and analytics capabilities.
4. Data quality dimensions shall be clearly defined.

Benefits

A precise and concise set of policies and procedures enables clarity of guidance and directions on data management, and effective review with an organisation ensures sustainable data management capability.

The following set of guidelines derived from each data principle will enable data and information assets are managed consistently and being used properly.

Data Principle 1 - Data is a core asset of an NGO and needs to be managed accordingly.

- Proper roles with well defined authorities and responsibilities to manage data need to be established.
- An inventory of data shall be established and maintained. This enables capturing of all the data from different source systems, processes and touchpoints.
- All staff need to be educated on the value and risks of data and are required to be properly trained and supported in the use and sharing of data.
- The top down commitment from Senior Executives of all divisions/departments is required to enable smooth execution of the data principles.

Data Principle 2 - Data shall be governed by individuals with designated roles and responsibilities.

- Data governance shall be organised in all operating units to integrate data silos from different operating units.
- Data areas may include 'data subjects', 'services' and 'staffs'.
- Data Stewards shall be identified to ensure appropriate governance and quality of data in each data area.

Data Principle 3 - All operating units and functions shall develop and upskill the staff's data and analytics capabilities.

- Lead Data Analyst shall support operating units and functions to ensure all staff are sufficiently trained and aware of relevant data policies, procedures, processes, techniques, and tools.

Data Principle 4 - Data quality dimensions shall be clearly defined such as timeliness, accuracy, completeness, consistency, security, compliance, integrity and uniqueness to the extent that all staff of the organisation has clarity on the expectation.

02 Define Roles and Responsibilities

Recommendations

Establish data governance roles and responsibilities which includes deploying those staff already executing data related activities to take on data governance as part of their official day-to-day responsibilities.

Benefits

Enhance data quality

- Ensure data format is standardised and consistent.
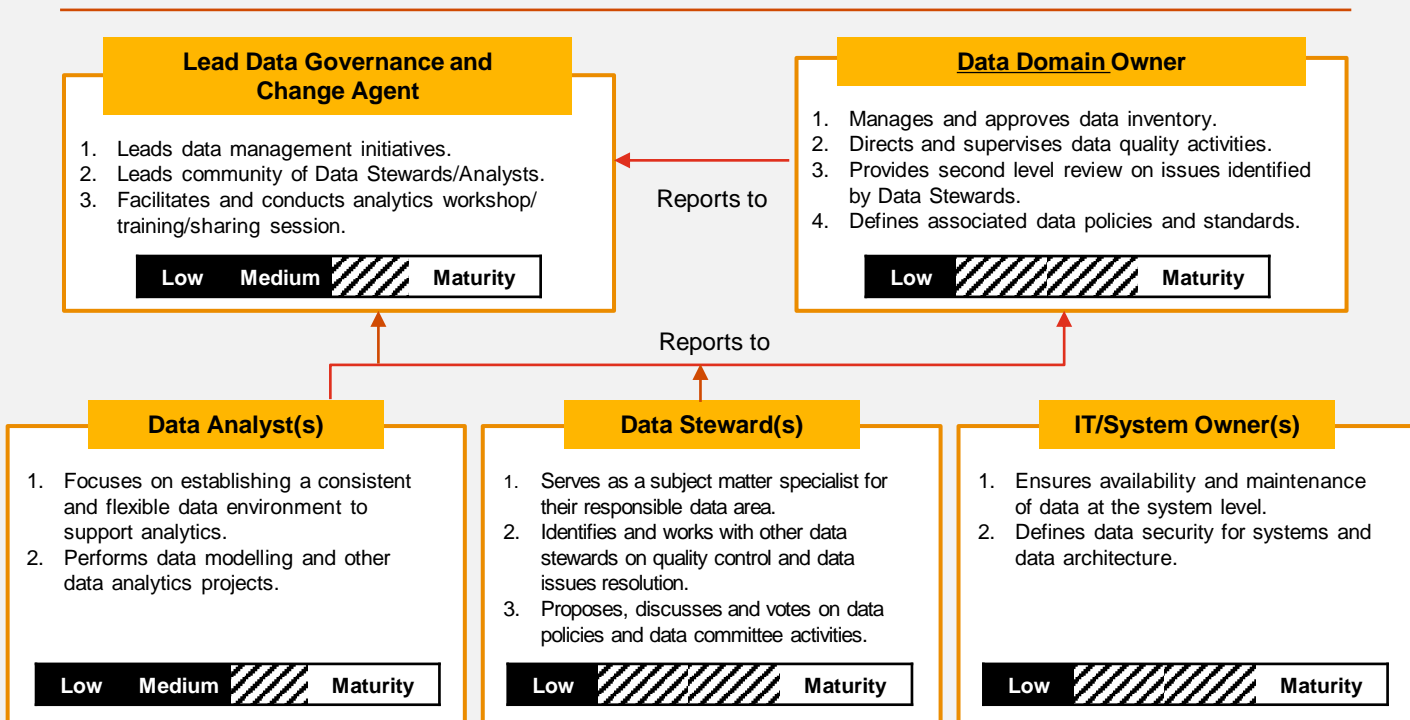- Improve overall data quality through clarity of ownership and location/medium of storage of data.

Improve data discoverability

- Discoverability (i.e. accessibility and visibility) of well-governed data facilitates evidence based analysis which will bring about useful insights and objective perspectives for rational decision making and helps to embed trust in the deliberation process.

Facilitate data-driven innovation

- Brainstorming sessions and collaborative workshops, sharing sessions amongst teams with different knowledge and skillsets can spark ideas for new data use cases (e.g. donor behaviour analytics, operational analytics) across different data collection touchpoints/platforms.

There are *5 main data ownership areas* for NGOs to focus on. For smaller NGOs, it is recommended that one individual could undertake multiple roles (e.g. the same individual can be assigned as data steward and data analyst). The diagram below sets out the responsibilities and tasks of each role.

Lead Data Governance and Change Agent

1. Leads data management initiatives.
2. Leads community of Data Stewards/Analysts.
3. Facilitates and conducts analytics workshop/ training/sharing session.

Low   Medium   ///// Maturity

Data Domain Owner

1. Manages and approves data inventory.
2. Directs and supervises data quality activities.
3. Provides second level review on issues identified by Data Stewards.
4. Defines associated data policies and standards.

Low ///// Maturity

Reports to

Reports to

Data Analyst(s)

1. Focuses on establishing a consistent and flexible data environment to support analytics.
2. Performs data modelling and other data analytics projects.

Low   Medium   ///// Maturity

Data Steward(s)

1. Serves as a subject matter specialist for their responsible data area.
2. Identifies and works with other data stewards on quality control and data issues resolution.
3. Proposes, discusses and votes on data policies and data committee activities.

Low   ///// Maturity

IT/System Owner(s)

1. Ensures availability and maintenance of data at the system level.
2. Defines data security for systems and data architecture.

Low   ///// Maturity

03 Establish Data Quality Monitoring Procedures

Recommendations

Establish continuous data quality monitoring controls. Create a data issue log to record all quality issues and conduct continuous assessments of data quality standard.

Data Quality Monitoring Control

- Data Quality Monitory Control consists of two key components: Data Quality Processes and Data Quality Issue Resolution Process.
- Data quality processes refer to a set of processes necessary to continuously define, assess, improve and sustain the quality of data in an NGO throughout the data lifecycle.
- Data quality issue resolution process encompasses of data issue identification, tracking, and the resolution of issues through root cause analysis.
- Data issue log shall be created and maintained which includes details such as date of issue raised, persons identified the issue (staff name and division/department), description of the issue, impact, status, target resolution date and Data Domain Owner.

Benefits

The monitoring and issue resolution processes allow greater transparency and clear delineation of responsibilities. With both processes in place, an NGO can track, report and analyse relevant information for any data quality related decision making.

The table below illustrates the typical roles and responsibilities of a data user, data steward and data domain owner throughout the data lifecycle to ensure continuous data quality monitoring and issue resolution.

Data Lifecycle		Data Created	Data Stored → Transformed → Used	Data Retired
Roles and Responsibilities	Data Domain Owner	• Defines associated data policies and standards. • Sets data quality metrics/targets. • Manages and approves the data inventory.	• Ensures data policies and data standards are applied. • Reviews data quality reports and issue logs. • Resolves escalated data issues. • Performs root cause analysis to improve data quality.	• Ensures disposal of data is in line with established policies and standards.
	Data Steward(s)	• Implements data quality metrics. • Introduces addition to Critical Data Element library (if appropriate).	• Maintains compliance against data standards, policy, and maintenance of the definition. • Monitors data quality and prepare data quality reports. • Manages data issues and issue escalation. • Maintains a data issue log.	• Manages disposal of data.
	Data User(s)	• Creates data.	• Reports all identified data issues.	

Example of root cause analysis that can be performed to resolve data quality issues:

Root Cause	Example	Potential Impact	Remediation Options
Data sourcing deficiencies	Lack of standardisation of common data from different sources.	Inaccurate data aggregation and reporting issues.	One-time cleansing.
Application or human errors	Lack of controls over human input on key data fields.	Dependence on failbacks and downstream reconciliation, resulting in operational inefficiencies.	On-going monitoring and data entry controls at the source system.

04 Establish Core Data Governance Process Workflow

Recommendations

Implement core data governance controls and workflow as a safeguard to ensure foundational processes for data activities are in place.The following core processes are recommended to be established:
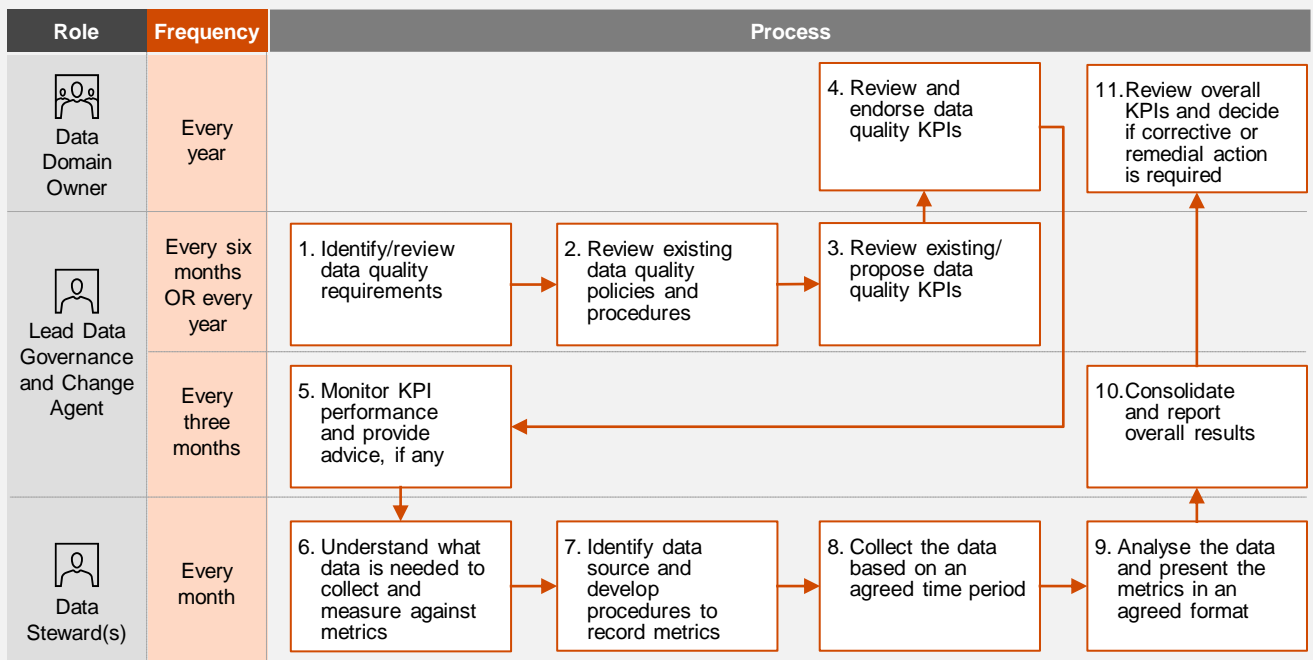
- KPI definition - by implementing data quality KPIs, the quality of data can be reviewed and measured against a set standard for further improvement. Refer to process flow below for illustration.
- Data access - manage formal requests for access to controlled data when stakeholders require additional or new data to enhance reporting or decision making capabilities.
- Regulatory compliance - address any new regulatory requirements which are pertinent to data management and governance, such as Personally Identifiable Information.
- Data retention - review and approve data retention or archive strategy in relation to critical data, or any data potentially subject to regulatory scrutiny.

Benefits

Standardising and streamlining data related processes improve process efficiencies (though minimising redundancies and errors during data activities).

Data quality KPIs can be essential key success factors for an organisation's implementation of data governance. These metrics can help to evaluate organisations' performance and effectiveness in managing data quality. Data quality metrics should also be aligned with the organisation's overall strategic goals and objectives, while continuously being measured, monitored and reviewed to identify any potential gaps and mitigate respective risks.

The following diagram is an example of how to introduce and implement data quality metrics in an organisation.

Role	Frequency	Process			
Data Domain Owner	Every year			4. Review and endorse data quality KPIs	11. Review overall KPIs and decide if corrective or remedial action is required
Lead Data Governance and Change Agent	Every six months OR every year	1. Identify/review data quality requirements	2. Review existing data quality policies and procedures	3. Review existing/ propose data quality KPIs	
	Every three months	5. Monitor KPI performance and provide advice, if any			10. Consolidate and report overall results
Data Steward(s)	Every month	6. Understand what data is needed to collect and measure against metrics	7. Identify data source and develop procedures to record metrics	8. Collect the data based on an agreed time period	9. Analyse the data and present the metrics in an agreed format

05 Establish a Data Inventory

Recommendations

Establish a structured and well-defined data inventory mechanism to record data collection touchpoints, definition of each data field, variables and descriptors of information, usage (e.g. operational, analytical, marketing), Personally Identifiable Information (PII) or non-PII and corresponding data subject consent.

Benefits

Improve data visibility

- Enables organisations to search data assets swiftly and supports discovering 'easy-to-understand' new datasets across all divisions/departments.
- Better understanding of data value as data inventory can be used to explain information about the data.
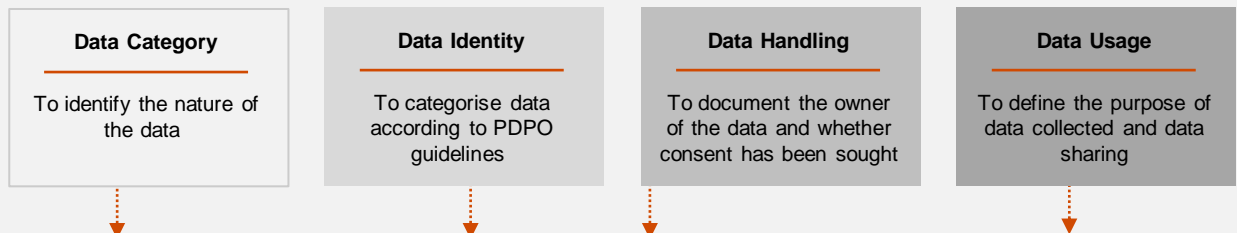
Facilitate advancement to be a data-driven organisation through data sharing and collaboration

- Supports data users to tag, document or annotate datasets to assist internal and external data users to understand how data is being used.

Greater accessibility

- Enable data users to process and extract insights from data assets while minimising their reliance on IT and/or data team.

Data inventory is used to identify types of data available and definition (meaning) of each data field. Data inventory, when established correctly, may serve as a single source of truth in an organisation to facilitate internal and external data handling. It also improves sharing and collaboration across divisions/departments. The following diagram shows an example of a good data inventory.

Data Category	Data Identity	Data Handling	Data Usage
To identify the nature of the data	To categorise data according to PDPO guidelines	To document the owner of the data and whether consent has been sought	To define the purpose of data collected and data sharing

Data Domain	Data Category	Data Name	Data Definition	Personally Identifiable Information	Data Handling Division	Consent Request	Data Usage				External Disclosure	
							Finance	Service Planning / Program Mgt	Program Comms	Insight Analysis	3rd Party	Reason
Service User	Service User Personal Info	Full Name	Full name that identifies the individual	Yes	Service Planning or Program Mgt.	Yes		✓	✓			
Service User	Service User Personal Info	Gender	Sex of the individual either male or female	Yes	Service Planning or Program Mgt.	Yes		✓		✓		
Service User	Service User Personal Info	HKID Number	Identity card number issued in HK used to identity the individual	Yes	Service Planning or Program Mgt.	Yes		✓				
Service User	Service User Personal Info	Phone Number	A phone number which can contact the individual	Yes	Service Planning or Program Mgt.	Yes		✓	✓			
Service User	Service User Personal Info	Email Address	An email address which can contact the individual	Yes	Service Planning or Program Mgt.	Yes		✓	✓			
Service User	Service User Personal Info	Home Address	A home address where can contact the individual	Yes	Service Planning or Program Mgt.	Yes		✓	✓	✓		
Service User	Service User Program Registration	Program Registered	Name and code of program registratered by the service user	No	Service Planning or Program Mgt.	No	✓	✓		✓	✓	For program audit and reporting
Employee	Employee Info	Staff ID	Staff number used to identify an employee	Yes	Human Resources	Yes	✓					
Employee	Employee Info	Job Title	Position held by the staff within NGO	No	Human Resources	Yes	✓				✓	For operation purposes

Recommended Data Governance Practices

06 Identify Critical Data Elements

Recommendations

Identify and prioritise critical data elements for effective operation. Data elements which could have material impact on operations or performance with respect to service delivery to users need to be governed with more controls (e.g. address, preferred time and mode of hot meal delivery for seniors).
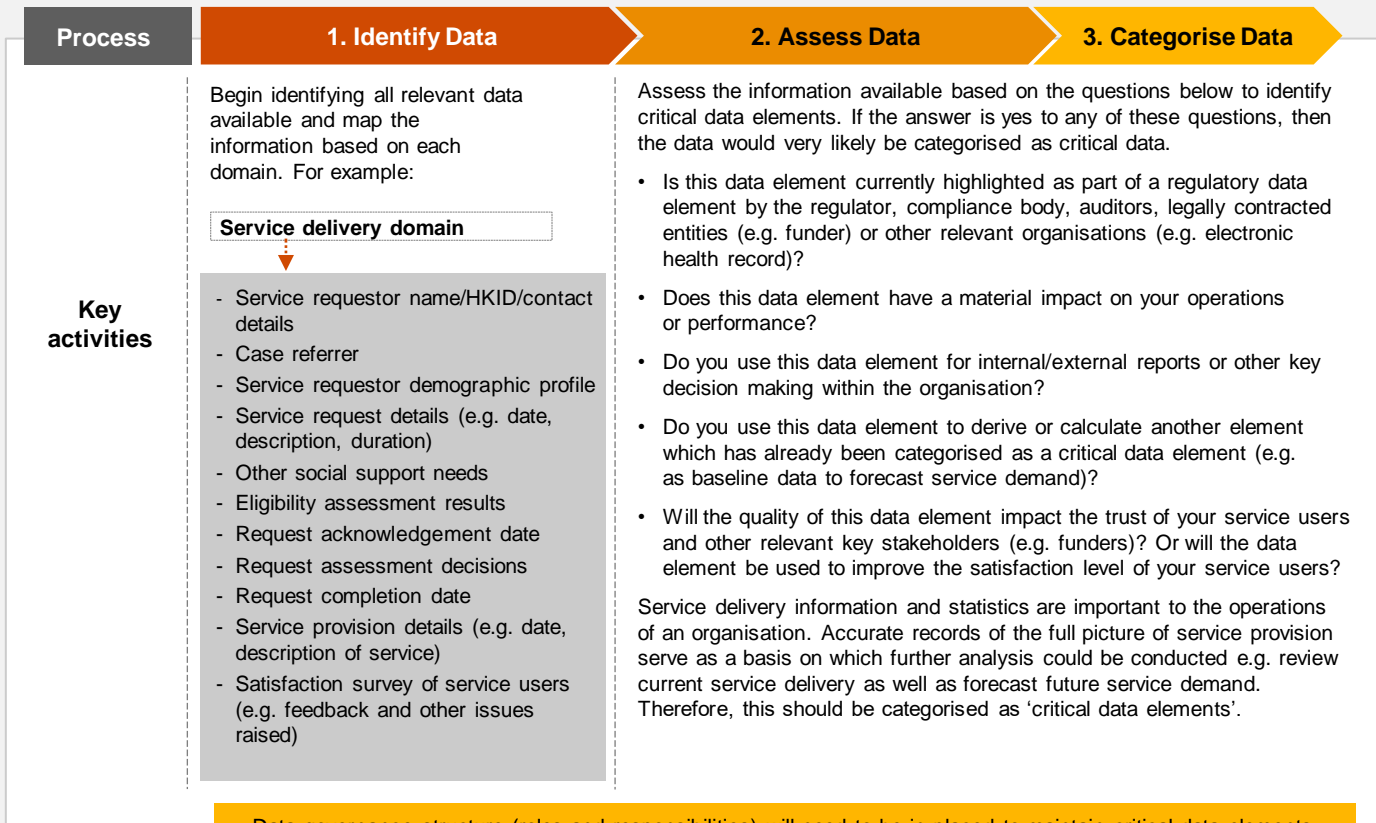
Managing changes to critical data elements

- It is highly recommended that periodic review is performed to ensure the information of critical data elements is still valid, complete, and accurate.

Benefits

Facilitate data analytics on key aspects of operations

- Facilitate organisations to focus on critical data elements which are important for operations and performance; and, extract insights from relevant data analytics (e.g. identification and quantification of service needs, measurement of service delivery efficiency).

The following process illustrates a typical example on how to identify critical data elements.

Process	1. Identify Data	2. Assess Data	3. Categorise Data
Key activities	Begin identifying all relevant data available and map the information based on each domain. For example: **Service delivery domain** - Service requestor name/HKID/contact details - Case referrer - Service requestor demographic profile - Service request details (e.g. date, description, duration) - Other social support needs - Eligibility assessment results - Request acknowledgement date - Request assessment decisions - Request completion date - Service provision details (e.g. date, description of service) - Satisfaction survey of service users (e.g. feedback and other issues raised)	Assess the information available based on the questions below to identify critical data elements. If the answer is yes to any of these questions, then the data would very likely be categorised as critical data. - Is this data element currently highlighted as part of a regulatory data element by the regulator, compliance body, auditors, legally contracted entities (e.g. funder) or other relevant organisations (e.g. electronic health record)? - Does this data element have a material impact on your operations or performance? - Do you use this data element for internal/external reports or other key decision making within the organisation? - Do you use this data element to derive or calculate another element which has already been categorised as a critical data element (e.g. as baseline data to forecast service demand)? - Will the quality of this data element impact the trust of your service users and other relevant key stakeholders (e.g. funders)? Or will the data element be used to improve the satisfaction level of your service users? Service delivery information and statistics are important to the operations of an organisation. Accurate records of the full picture of service provision serve as a basis on which further analysis could be conducted e.g. review current service delivery as well as forecast future service demand. Therefore, this should be categorised as 'critical data elements'.	

Data governance structure (roles and responsibilities) will need to be in-placed to maintain critical data elements.

07 Promote a Data Governance Culture

Recommendations

Foster a data governance culture across divisions/departments through workshop and training on operation differentiation and strengthening of trust through rigorous approach to data governance and protection of data of service users and other stakeholders; and leveraging insights from data analytics for improvement of overall service delivery.

Develop a curriculum of multi-level training that caters to staff at different data maturity levels. This can be designed as follows:

- Level 1 – Data Awareness for All Staff
- Level 2 – Data User
- Level 3 – Data Analyst
- Level 4 – Advanced Analyst

Benefits

Upskill data capabilities
- All staff can be trained with foundational skills including strengthening their data knowledge and equipping them with skills to leverage data to enhance the performance and efficiency of their work.

Transform to a data-driven culture
- Organisations would be able to make better decisions based on insights extracted from data.

The following diagram illustrates the target audience of different levels of data training, overview of training curriculum and recommended form of knowledge dissemination.

Levels of training	Level 1 Data Awareness	Level 2 Data User	Level 3 Data Analyst	Level 4 Advanced Analyst
Target audience	All staff	Any staff who wants to explore value of data for their day-to-day work	Data consumer who wants to extract new insights from data but is currently restricted by available tools (e.g. Excel spreadsheet)	Data analyst who wants to leverage advanced data analytics tools for their use case
Training content/ Tools and technologies	Overview of data governance concept to develop awareness and educate on the importance of data governance	Illustration of how data analytics tools could enhance current analysis/reporting (generated by traditional means, e.g. Excel)	Introduction to self-service tools for data analytics and data visualisation, including practical hands-on use case exercises	Advanced analytics that may include using out-of-the-box tools to make predictions and identify trends/patterns which are crucial for operations
Method(s) of training	Self-paced e-Learning	Self-paced e-Learning + Classroom Training	Self-paced e-Learning + Classroom Training	Classroom training

10 Recommended Data Governance Practices
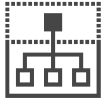
08 Define Data Operating Model

Recommendations

Adopt a centralised model to align the strategic value of analytics; define requisite policies and standards; enhance data transparency and sharing of best practices.

Benefits

A centralised model is more efficient with following benefits:

- Develop key strategic analytics solutions in one place.
- Centrally define and manage guidelines, standards, policies and procedures for data security, privacy and compliance.
- Provide one-stop support to self-service analytics users across divisions/departments.
- Share of best practices or lessons learnt.
- Larger organisations could consider deploying a centralised platform to promote data visibility and integration initiatives. Alternatively, low or medium maturity organisations could consider having a shared database/cloud platform that allows collaboration and data sharing based on level of access rights.

Strengthen overall analytics capabilities

- Leverage Citizen Data Analyst from various departments to facilitate development and rollout of organisation-wide analytics capabilities.

There are different strategic and operational aspects to be considered in selecting the appropriate data operating model. The following page presents different types of models that could be considered by organisations depending on their maturity level and preferred data operating environment.

What are the factors in considering the appropriate type of data operating model?

There are mainly two important factors in assessing the best fit data operating model. First of all, whether the decision making process of each operating unit in the organisation is independent or not. For the prior case, a decentralised model could be considered. Secondly, the extent of governance and control which the organisation would prefer to establish. For those organisations which prefer stringent governance and control, a centralised or hybrid model would likely be more appropriate.

Recommended Data Governance Practices
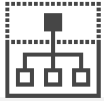
08 Define Data Operating Model

The following sets out the key points of different types of Data Operating Model:

Centralised Model (Recommended)

| Low	////////	Maturity

The centralised model is one of the most common models adopted by medium and high maturity organisations. This model requires that all governance functions shall be defined and driven centrally with minimum resources support.

Maturity level requirement of a centralised model is typically low, hence, appropriate for organisations with limited staff and funding to adopt in its data journey.

Key analytics functions would need greater governance and control to:
• Align strategic priority;
• Promote sharing of best practices;
• Formulate policies and define standards;
• Standardise data lifecycle; and
• Establish standard and consistent datasets.

Hybird Model (Targeted future)

Low	Medium	High	Maturity
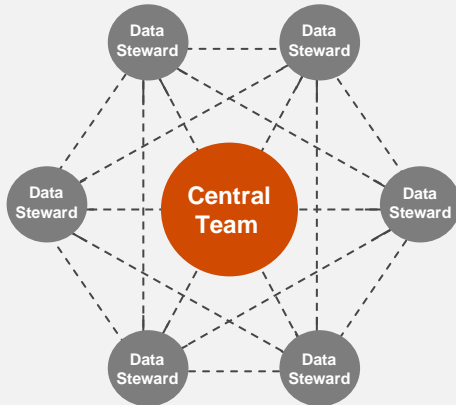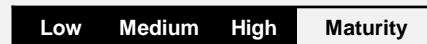
Centralised IT/data department to develop and issue guidelines, but each operating unit and function will manage data activities individually.

Data Steering Committee

Domain functional team e.g. Fund raising
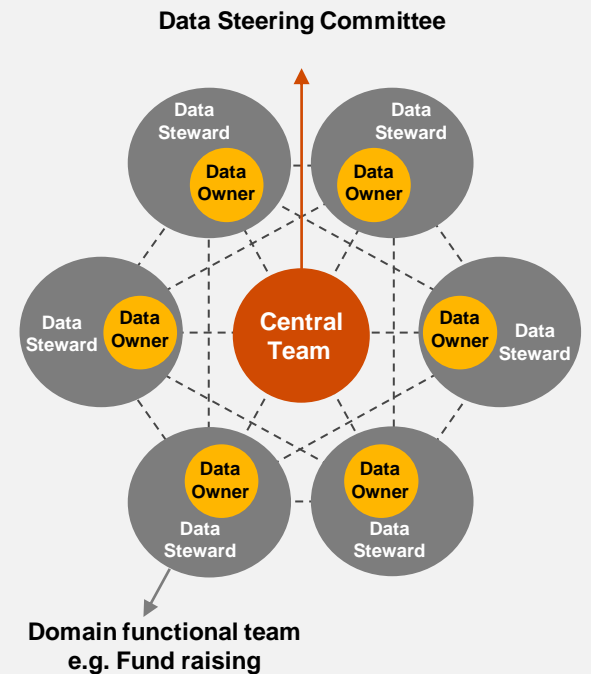
Decentralised Model

Low	Medium	//////	Maturity

Localisation of data functions and capabilities due to specific focus that relies on data available only to respective divisions/departments.

Decentralised model could be appropriate for those organisations which require a quick response to changes in their data and/or reports within a very short period of time from individual operating units/functions.

However, it would be difficult to standardise practices across all operating units/functions.

Glossary
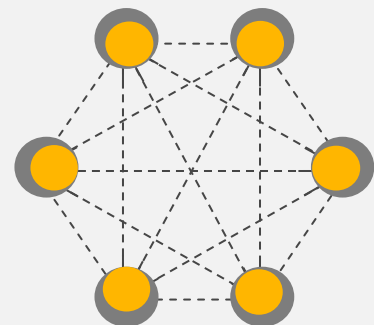
Terms	Definition
Citizen Data Analyst	Citizen data analyst is a person who creates advanced analytics models, but whose primary job function is not analytics related. - Based on "Gartner's" definition
Critical Data Elements	Critical data elements are key elements of the datasets that are used to drive key decisions and are considered vital to the organisation.
Data Architecture	Data architecture is a set of rules, policies, standards and models that govern and define the type of data collected and how it is stored, managed and integrated within an organisation and its system of databases.
Data Domain	Data domain is high level data categories of operational areas/key processes for the purposes of assigning responsibilities for the data.
Data Inventory	Data inventory is a list of datasets with metadata that describes their contents, sources, licensing and other useful information.
Data Lineage	Data lineage refers to the mapping of the data journey including each point of transition from data creation, storage, transformation to disposal.
Data Lifecycle	Data lifecycle refers to the different stages that a typical data go through from creation, storage, transformation, usage to retirement.
Data Mapping	Data mapping refers to the process of matching fields from one dataset to another into a centralised database. Data mapping is usually performed to facilitate data integration, data migration or any other data management activities.
Data Profiling	Data profiling refers to the process of examining or reviewing data available, understanding structure, content and interrelationships, and create summaries or statistics about the data.
Data Quality	Data quality is an intricate way of measuring data properties from different perspectives. It is a comprehensive examination of application efficiency, reliability and fitness of data.
Data Standard	Data standard is a rule that is used to measure the quality of data. - Based on the definition of "Data Management Body of Knowledge"
Data Subject	Data subject is the individual who is the subject of the personal data. - Based on "The Personal Data (Privacy) Ordinance"
Descriptive Analytics	Descriptive analytics is the analytics that describe, summarise and analyse historical data (What happened?) .

Glossary

Terms	Definition
Diagnostic Analytics	Diagnostic analytics is the analytics that identify causes of trends and outcomes (Why it happened?) .
Master Data Management	Master data management (MDM) is a technology enabled discipline in which operating units and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and responsibility of the official shared master data assets.
Metadata	Metadata is structured reference data that helps to sort and identify attributes of the information it describes. Metadata summarizes basic information about data, which can make it easier to find, use and reuse particular instances of data. For example, author, data created, date modified and file size are examples of very basic document file metadata. Having the ability to search for a particular element (or elements) of that metadata makes it much easier for someone to locate specific document. - Source from "whatis.techtarget.com"
Predictive Analytics	Predictive analytics is the analytics that are able to predict future outcomes based on facts from the past historical data (What could happen?) .
Prescriptive Analytics	Prescriptive analytics is the analytics that recommend the best course of actions or decisions based on descriptive analytics' results (What should be done?) .
Self-service Analytics	Self-service analytics is the analytics in which staff are enabled and encouraged to perform analysis and generate reports on their own using existing business intelligence tools with minimal IT support.
Semi-structured Data	Semi-structured data is a combination of structured and unstructured data.
Structured Data	Structured data is the data in a predefined row and column format.
Unstructured Data	Unstructured data is the data stored in its natural format that is data which are not in a proper row and column format.
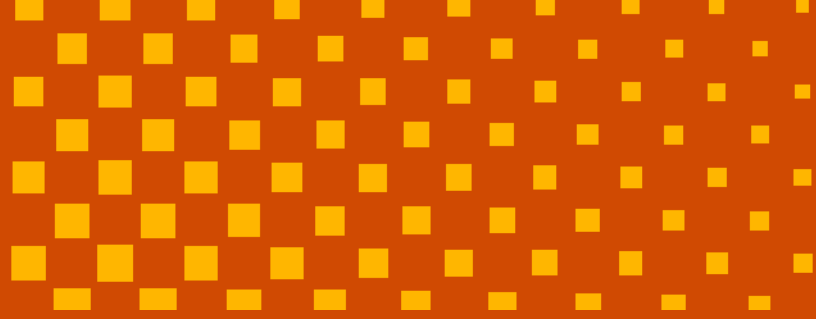
If you have any further questions about this toolkit, please contact:

Jennifer Ho
Risk Assurance Leader
PwC Mainland China and Hong Kong
Jennifer.cw.ho@hk.pwc.com

Chris Mo
Partner
PwC Hong Kong
chris.yw.mo@hk.pwc.com