talend + ncs making IT happen

Compliance considerations for public agencies:

an introduction to Singapore's data governance and protection frameworks

Executive summary

Public agencies in Singapore are governed by the 2018 Public Sector (Governance) Act (PSGA) and the 2011 Government Instruction Manual on IT Management (IM on IT Management). Here is an introduction to some key information practices and concepts that underlie these data governance and protection frameworks, including collection limitation, data quality, security safeguards, and accountability. The paper examines these pillars in light of the technology, service dimensions, and processes that public agencies should take into account as part of their data governance compliance implementation frameworks.

About the authors

Ramkannan Avadainayagam
Associate Director – Big Data Analytics and Data Engineering
NCS

Janet Liao
APAC Product Marketing Lead
Talend

An overview of data governance and protection frameworks in Singapore

Data governance is a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organisation to achieve its data goals. It establishes the processes and responsibilities that provide the quality and security of the data used across an organisation. Data governance defines who can take what action upon what data, in which situations, and using what methods.

A strong data governance programme is a pivotal part of the landscape for data protection and privacy compliance. The traditional data governance disciplines of data ownership, metadata management, data cataloguing, data quality management, and model governance apply to protection and privacy. Data governance programmes that support data protection compliance also support broader adherence to data sovereignty regulations.

Singapore implemented its Personal Data Protection framework in 2014. The act was created to aid in the governance of the collection, use, and disclosure of personal data by private organisations as well as to establish a Do Not Call Registry. The act, while providing governance, also recognises individuals' rights to protect their personal data, and private organisations' needs to collect, use, and disclose personal data for appropriate practices. It specifies some of the stiffest penalties for data protection offenses in the Asia-Pacific region, with fines of up to S$1 million.

Conversely, the public sector in Singapore adheres to different data regulatory frameworks: namely the 2018 Public Sector (Governance) Act (PSGA) and the 2011 Government Instruction Manual (IM) on IT Management. The separate frameworks for the private and public sectors provide for the differences in expectations for the latter — namely integrated services delivery across public agencies for citizens. The PSGA enables criminal penalties to be imposed on public officers who breach data protection regulations.

In addition to regulations, the government had taken steps to implement a set of architectural practices, known as the Government Data Architecture (GDA), based on a single common data governance framework within the public sector. The architecture includes centralised information structures such as single sources of truth (SSOT), trusted centres (TC), and central platforms. SSOTs are databases that contains frequently used high-quality core data for multiple public agencies. TCs are data platforms that fuse and distribute core datasets of both identifiable and non-identifiable data for use by public agencies. The central platforms are secured control points for data users in public agencies to request, download, and analyse datasets.

The latest data governance policies and implementation guidelines come from the Public Sector Data Security Review Committee Report (PSDSRCR; 2019). Recommendations and details are set out following the committee's review of government systems and data management practices, after a series of serious data breaches in 2018 and 2019. The government's target is that by the end of 2021, these recommended measures should be implemented on 80% of government systems, whilst the remaining 20% should be implemented by the end of 2023.

A look at data governance and protection implementation

GDA's implementation of the data governance and protection framework for the public sector touches on information practices and concepts including collection limitation, data quality, security safeguards, and accountability. These principles and their corresponding implementation need to be addressed through a combination of technology, people, and processes in order to derive maximum organisational effectiveness.

The sections below focus mainly on the technology and process dimensions; the people and culture dimension is shaped by the organisational culture and context of each public agency.

Collection limitation

The collection limitation principle limits the collection of personal data. From the PSDSRCR, public agencies are to collect data of value only when it is not part of the GDA SSOT's dataset, and they need to maintain clear retention policies pertaining to purging of data when it is no longer required by the data user. This ensures that there are limits to the collection of personal data, and that any such data is managed by lawful and fair means by the public sector.

From a technology perspective, the data management platform utilised by the public agency must support the different phases of the data lifecycle — namely collect, transform, govern, and share. During the collection phase, agencies that collect data must, firstly, be able to validate that the required data is not already available in the SSOTs, and, secondly, have the ability to ingest across different data sources and formats, as the required information might come from an industry source or an application. In addition, agencies should leverage technologies such as natural language processing for efficient extraction and tagging of personal data from mediums such as free-form text within a document, an application, or user-generated content within a web or mobile application.

Beyond collection limitation, during the transformation and governance phase of the data lifecycle the agency should be able to exercise control over policies pertaining to data access or retention. For example, it should be able to set users' access permissions and the types of activities they can perform on the data. Another key functionality is controlling the volume of data users can download from government information systems and time users have access to the data. Lastly, for the share phase, the platform must be able to support API calls made to the GDA or any relevant external applications or platforms so that data can be exchanged in a standardised fashion across agencies.

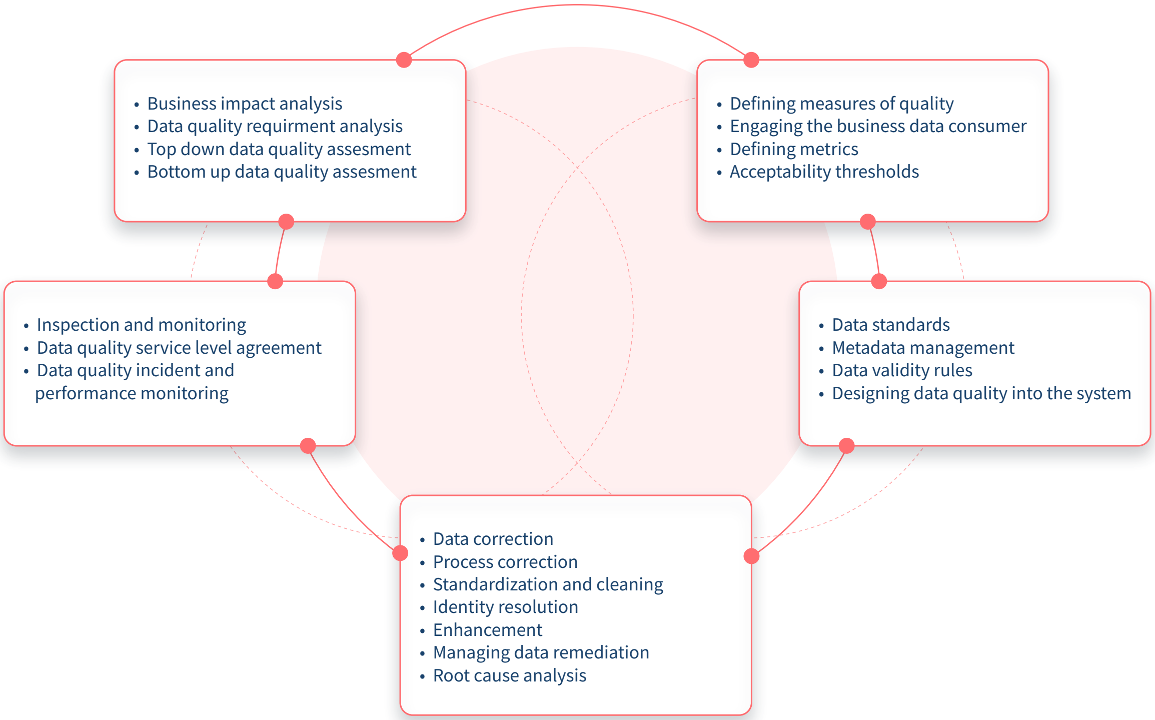
With these criteria in mind, agencies can begin thinking about some key service management dimensions:

a) **Data collection**
Agencies can define business-specific datasets and key points within the data governance operating model. This model has to define how to access the data and stage it for next-level storage and processing.

b) **Data profiling**
Data quality analysts can use standard data profiling techniques to discover the composition and nature of the data contained in the target datasets and critical data attributes defined therein.

c) **Role of critical data elements**
Critical data elements (CDE) are defined as "the data that's critical to success" in a specific business area (line of business, shared service, or group function), or "the data required to get the job done."

Data collection methodology overview



- Business impact analysis
- Data quality requirment analysis
- Top down data quality assesment
- Bottom up data quality assesment

- Defining measures of quality
- Engaging the business data consumer
- Defining metrics
- Acceptability thresholds

- Data standards
- Metadata management
- Data validity rules
- Designing data quality into the system

- Data correction
- Process correction
- Standardization and cleaning
- Identity resolution
- Enhancement
- Managing data remediation
- Root cause analysis

- Inspection and monitoring
- Data quality service level agreement
- Data quality incident and performance monitoring

Data quality

The SSOTs within the GDA require that core data that is accessed often by multiple agencies be cleaned and of high quality. A similar data quality level would be expected of non-core data, such as data that's regularly accessed by users of the public agency within its own organisational domain, especially if the data to the GDA originates from the agencies' databases.

Data quality management is the process of conditioning data to meet the specific needs of business users. Accuracy, completeness, consistency, timeliness, uniqueness, and validity are the chief measures of data quality, and methods to ensure data quality should be built into the data management platform. For example, data quality techniques such as profiling enable data users to explore, identify, and assess whether the data they are using is inaccurate, inconsistent, or incomplete. Techniques such as data deduplication and data matching enable agencies to create a 360-degree view of a citizen's behavioural profile across datasets from the SSOTs or agency-specific databases.

A data quality framework should include the following pillars:

Profiling	A robust profiling tool should allow analysts to dig into the nature of target datasets to understand the current state of data.
Data quality rules	Data quality rules help validate data relationships and let analysts identify and correct the majority of data problems. The key here is to discover all data quality rules and make sure they are well understood.
Metadata and policy information	Metadata is a broad term, and the programme team needs to define what metadata needs to be collected. In the context of a data quality framework, metadata includes information such as governance and roles (e.g., data steward); systems; data dictionaries; extract, transform, load (ETL) and data lineage detail; and data quality profiling and assessment results.
Issues management	Issues management may also be part of the metadata or reporting environments, but should include a workflow and escalation capability. Here, analysts should be able to track data quality issues coming from ongoing assessments as well as other sources, and manage those issues through triage, root-cause analysis, and remediation.
Analytics	As the programme matures and larger sets of data quality results become available, data teams can conduct advanced analytics, taking advantage of AI tools and ML-driven data processing.
Reporting and project management	Reporting and project management tools should automate work and, wherever possible, formalise and simplify project and programme management activities.

Security safeguards

Personal data should be protected by reasonable security safeguards against such risks as loss and unauthorised access, destruction, use, modification, or disclosure. Public agencies are encouraged to minimise their own data management by leveraging the GDA as much as possible to reduce vulnerabilities.

Agencies that intend to manage their own data platforms should leverage “defence-in-depth” capabilities, which apply protection techniques on the data itself to render the data useless to hackers even if it’s extracted. For example, data masking techniques support intra or interagency exchange of production-quality data for development and analysis without disclosing personally identifiable information (PII) to people who are not authorised to see it. Sensitive data such as addresses can be anonymised and made unidentifiable to prevent unauthorised access. In short, masking enables data to be easily shared without violating privacy rules.

Another data security technique is encryption, which protects data by transforming it into unreadable cipher text. Only users with the proper password and cryptographic file can decrypt the cipher text and read the original data. Encryption not only protects data from internal and external leakage, but also reduces the risk of sensitive data being exposed during transmission between systems.

Data governance plays a key role in enterprise data security. A data governance council can determine how and where data may be shared to comply with privacy regulations. Data governance enforces a policy guide across an enterprise for developing a strategy around cyber security and information risk management. Specifically, for emerging digital analytics programmes, organizations may have challenges dealing with external unstructured data sources, including data lakes. An appropriate tool can help manage data security across the entire data analytics stack.

Accountability

Ultimately, for any data governance and protection framework to be successful, data owners and users must be held accountable for their actions. In the public sector, the PSGA prescribes penalties for public officers who violate data security with fines of up to $5,000, and/or up to two years imprisonment and disciplinary actions. These punishments are applicable not just to public officers, but also to third-parties that provide services to public agencies.

Accountability thus requires having a data management platform that provides auditable evidence of negligent acts and intentional data breaches. Functionalities such as data lineage enable public agencies to monitor and trace users' operations on the data over time. Data lineage lets organisations trace data across the landscape of applications and systems they use, and track the sources and types of modifications performed on the data.

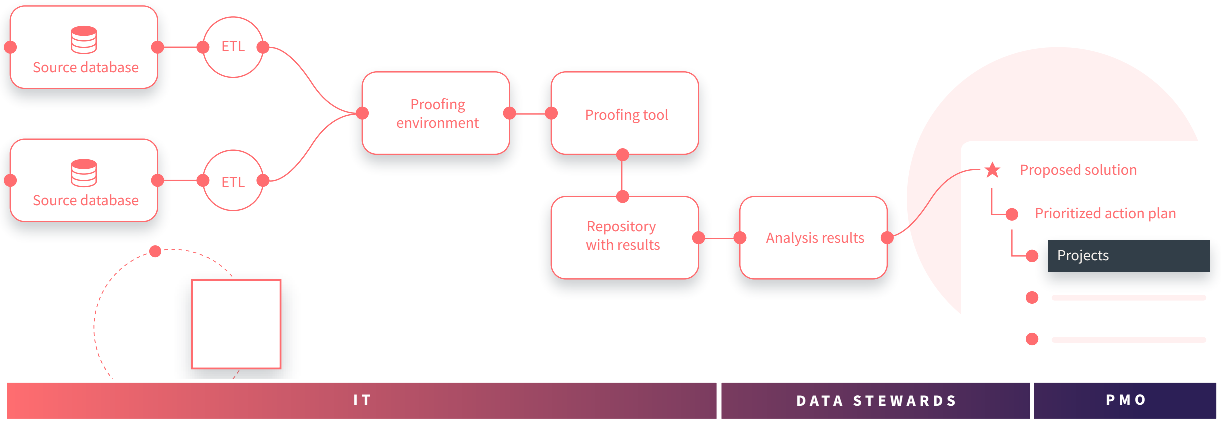
Tracking lineage ensures that data modifications are documented and can serve as evidence in the event that a public officer tampers with personal data.

Data stewardship is crucial to the success of data governance (which in turn is crucial to the success of data management). Data stewards collect, document, and maintain both data and metadata (such as data definitions and business rules).

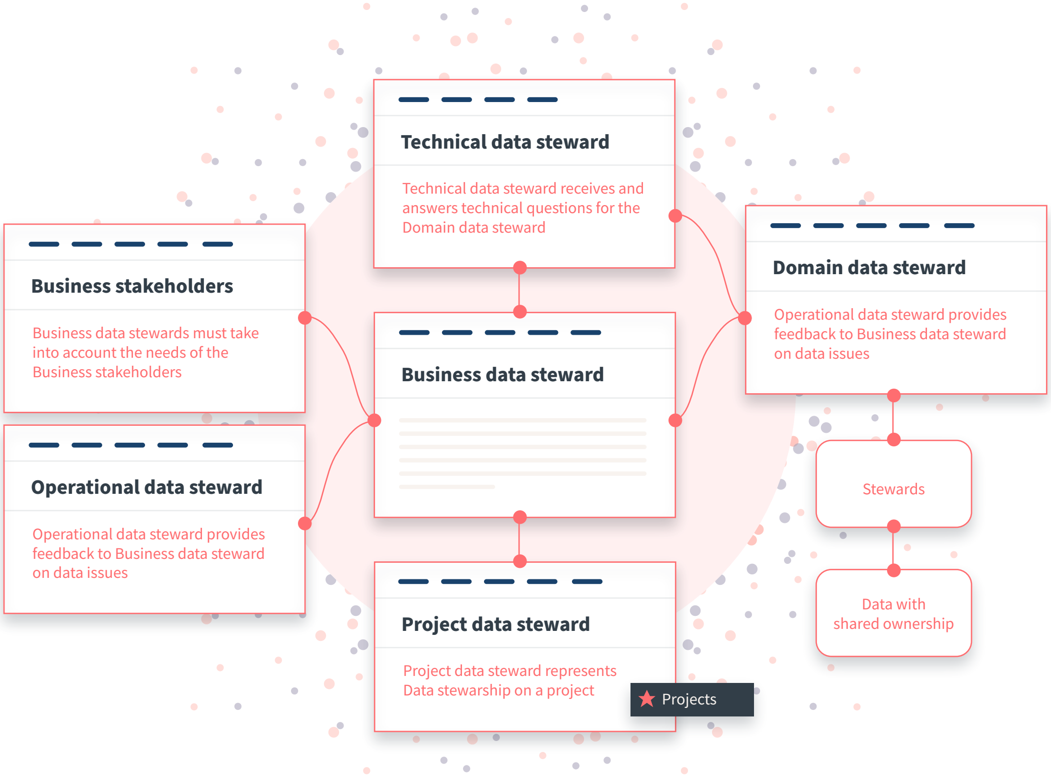
Data stewardship activities may include:

- Define the data
- Create processes and procedures
- Maintain data quality
- Optimise workflows
- Monitor data usage to assist data teams
- Enhance compliance and data security

Data steward role within a data organization



Interaction between data stewards


Technical data steward
Technical data steward receives and answers technical questions for the Domain data steward
Business stakeholders
Business data stewards must take into account the needs of the Business stakeholders
Business data steward
Domain data steward
Operational data steward provides feedback to Business data steward on data issues
Operational data steward
Operational data steward provides feedback to Business data steward on data issues
Stewards
Data with shared ownership
Project data steward
Project data steward represents Data stewardship on a project
★ Projects

Partner with Talend and NCS for compliance systems

To meet the prescribed timeline of 2023 for full-fledged implementation of the PSDSRCR and PSGA, public agencies must select a data management platform with the abovementioned capabilities that supports the GDA framework, and choose a capable service provider to support the system integration and advise on structuring the organisation's processes and changes during the transition.

Talend Data Fabric is a comprehensive data integration and data integrity platform that offers data governance, data quality, and data protection capabilities. It enables public agencies to get complete, clean, and credible data across deployment environments, whether on-premises, in the cloud, or in hybrid configurations. For its part, the NCS Next Digital team has the expertise and a history of supporting public agencies looking to deploy data management platforms with the technology and organizational processes required to comply with Singapore's data governance and protection frameworks.

Several major public agencies are joint customers of Talend and NCS, using Talend Data Fabric for data management, data governance, data ingestion, and ensuring data quality and integrity. To learn more, please email Talend at groupsalesasean@talend.com or NCS at ramkannan.avadainayagam@ncs.com.sg to set up a consultation session.