Data Governance and Data Integration in Singapore

Statistics Korea (KOSTAT) and Economic and Social Commission for Asia and the Pacific (ESCAP) Joint Conference on Data Governance and Integration for Maximum Development Impact

Paper Presented by Dr. Koh Eng Chuan (Chief Statistician) and Peng Yuxiang (Statistician)

The Evolving Data Landscape

Data governance and integration is an important topic that has seen significant developments in the recent years with changes in both the supply and demand sides of the data landscape.

Supply Side

On the supply side of data, the advancement in technology has led to a data explosion and the enlargement of the data ecosystem. The increase in supply of data comes from many sources such as big data from Internet of Things (IOT) devices and social media. This is relevant to National Statistical Offices (NSOs) as they may tap on these sources to supplement or replace existing indicators. For example, by web-scraping websites of online retails stores, NSOs can reduce or replace price surveys which will help to reduce respondent burden.

Technology has also made it much easier to perform data processing, data fusion and transmission. Machine Learning and Artificial Intelligence algorithms can play a role in data imputation and fusion by identifying patterns or relationships in large sets of integrated data. High-speed internet and communication technologies make it possible to transmit large volumes of data in a short time, allowing for real-time access to datasets. One of the main challenges we face is how to effectively make sense of the vast amount of data that is now so readily available.

Demand Side

The demand for data has also grown significantly.

Businesses want more customised data to make business decisions. For example, they want more demographic breakdown to assess places to set up new business branches, and more location data to gather insights on the potential customer base and competition within the region.

Citizens are requesting better services and more efficient service delivery from government agencies. 'Tell us once' has become a common goal for both citizens and government authorities.

As a result, government agencies must work together to better serve the public, driving the need for seamless data integration. For example, in healthcare services, government agencies can integrate data backend to allow patients to pay for their medical bills from their medical savings (Medisave) and enable hospitals to assess their eligibility for medical subsidies.

Legislative Enablers

With these developments, the Singapore government reviewed the institutional setup and legislation required to better serve the public. In 2017, the **Smart Nation and Digital Government Office** (SNDGO) was formed to lead the digital transformation of the Singapore government to achieve a public service that is 'digital to the core'.

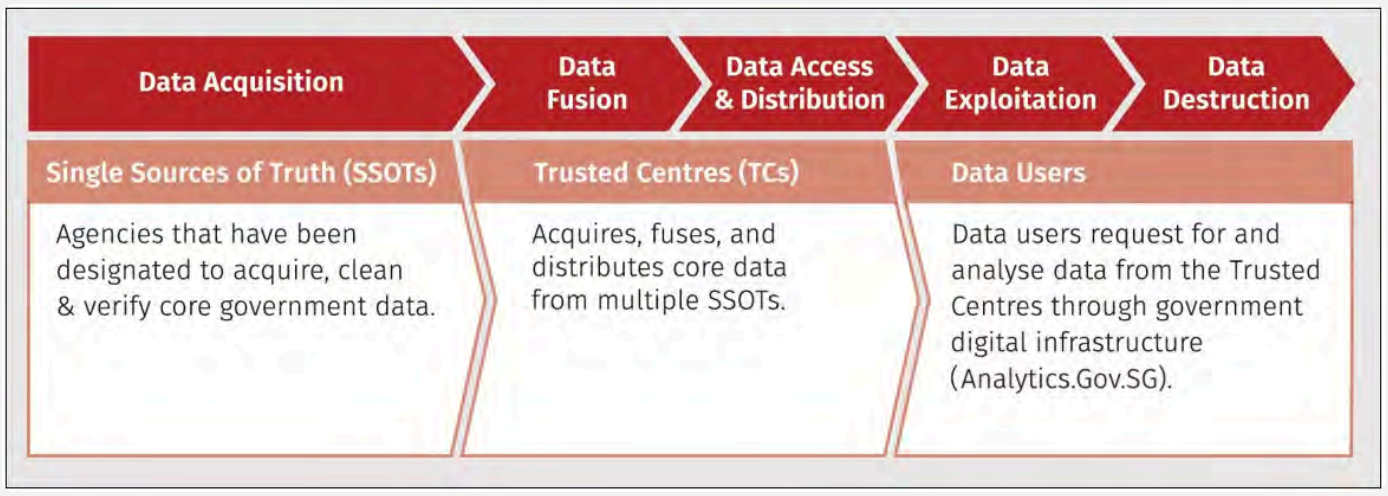
The **Public Sector (Governance) Act 2018** (PSGA) was enacted to provide a consistent governance framework across public agencies. This legislation enables public sector agencies to share data with each other on a needs basis for policy analyses and service delivery. At the same time, there are strong safeguards, such as penalties on public officers who misuse or disclose of data without authorisation.

Government Data Architecture (GDA)

In Singapore, we have unique identifiers for individuals and businesses which make data fusion possible and efficient. To enable a data-driven government, the Government Data Architecture (or GDA, correct as at **2019**) was put in place to facilitate secure data sharing and usage across the public sector. Figure 1 illustrates the efficient sharing of clean and authoritative datasets by establishing Single Sources of Truths and Trusted Centres:

- Single Sources of Truth (SSOTs), or data custodians, are authoritative sources with administrative functions to acquire, clean and maintain high quality Core Data. Core Data refers to administrative data that are frequently used by multiple government agencies for policy analysis, operations, or service delivery. The SSOTS are to provide Core Data to the Trusted Centre regularly.
- Trusted Centres (TCs) perform data fusion functions and serve as data intermediaries to securely distribute Core Data to user agencies.

Figure 1: Government Data Architecture (GDA)



Within the TC structure, there is a need for a central government agency to integrate and disseminate the vast amount of administrative data for policy evaluation, research, and service delivery. As a natural extension of our experience and capabilities, and with the new legislative enabler, the Singapore Department of Statistics (DOS) took on the role as a TC for individual and business data.

Data Sharing with Public Agencies

Under the PSGA, the three designated TCs integrate administrative data from multiple agencies which will be shared with public agencies for research, policy analysis and service delivery. DOS is one of the TCs.

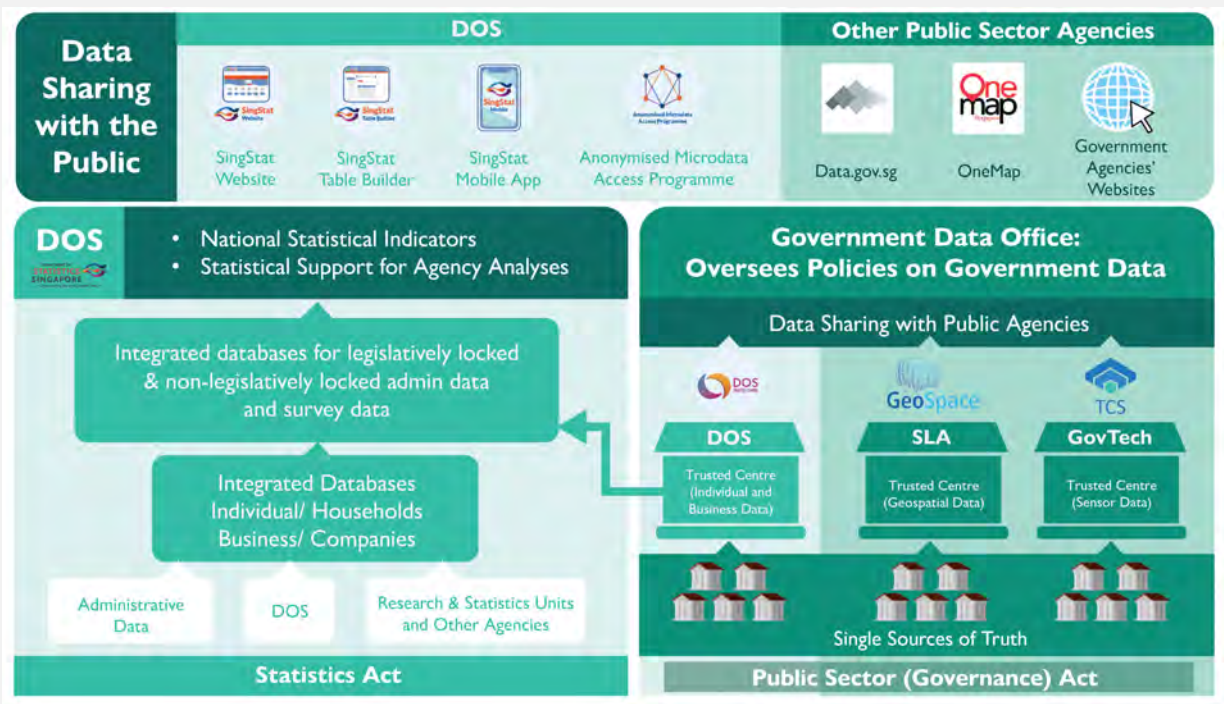
DOS is apt to be a TC for individual and business data for three reasons. Firstly, we are heavy users of data, hence playing the role of a TC allows us to obtain data quickly and directly from other administrative agencies. Secondly, as this role is a natural extension of our experience and capabilities, making it highly efficient and effective for the government to designate us as a TC. Finally, it strengthens our role as a first stop in data provision.

Expanded Data Landscape in Singapore

Administrative data alone is insufficient for the compilation of national statistics; we still need to collect detailed survey data under the Statistics Act to provide the necessary supplementation. The data collected are used exclusively for statistical purposes and are not shared with other public sector agencies. This is in line with the UN Fundamental Principles of Statistics (UNFPOS).

Therefore, DOS plays an important dual role of sharing administrative data under the PSGA and processed statistical data under the Statistics Act. The expanded data landscape in Singapore encompasses the national statistical system and the broader data ecosystem, including the national data bases under the TCs (Figure 2).

Figure 2: Government Data Architecture (GDA)



Key Issues to Navigate Through

DOS's role has evolved from the traditional NSO role of producing statistics to become the first stop for users to obtain data and data services. To fulfil this role, there are various key issues to navigate through. I would highlight just three of these issues.

1 Trust, Legislation and Governance

Firstly, we must recognise that as data become an increasingly important part of the decision-making process, ensuring the trustworthiness and confidentiality of that data becomes even more essential. We must uphold and maintain this trust with stakeholders, and the public to ensure that they will continue to provide us with information. DOS has strengthened our data governance with policies to protect and preserve trust in our management of data. We also openly communicate the legislation, data collection methods, and intended use of data which builds transparency and accountability.

2 Data Innovation and Services

Secondly, we have to understand and respond to users' needs through data innovation and services. To expand our presence in the data landscape, DOS furnishes a wide range of economic and social data, combined with data services, to provide insights for analyses and decision-making for a diverse range of user groups.

Our public facing website, the **SingStat Website**, provides general users with a range of open data, statistical information, and resources. The **SingStat Table Builder** allows users to create customised statistical tables and charts using official statistical data. Our SingStat Mobile App offers mini charts on a wide range of topics to provide users a quick sensing of trends at a glance.

To be effective, these data platforms need to have interesting and useful content presented in a way that is easy to understand. For example, DOS has developed dashboards to provide firms with curated and contextualised data to address questions on their customers and industry. This includes a business benchmarking tool to allow firms to evaluate their business performance against industry performance.

Internally, data innovation means constantly reviewing our processes and technology of the day, to be more productive and cost-effective in producing data. Examples include using open-source programming languages, cloud services as opposed to proprietary software, and automated tools to facilitate self-help.

3 Engagements and Partnerships

Thirdly, we have to constantly engage data users in the public sector, private sector, as well as partners in other countries and international agencies to remain relevant.

These extensive engagements with stakeholders and user groups enable us to gain deeper insights into their evolving needs. Our engagements with data and statistical experts worldwide keep us informed on the latest technological and statistical developments. These are then fed back to refine and improve our data stewardship and governance in the ever–evolving data landscape.

Conclusion

NSOs are operating in an increasingly complex environment with an expanding data ecosystem, accompanied by digital government transformation and the demands of more sophisticated users and stakeholders. The real world has many hurdles that NSOs have to cross, from convincing stakeholders to aligning staff with the shared vision of the future NSO. This includes managing day-to-day work while getting resources to explore new ways of working, developing new products and services, and uncovering new data sources. It is an ongoing challenge for NSOs to remain relevant and on top of their game, and this requires the backing of good data governance, data integration, and data sharing practices.