

Singapore's Administrative and Big Data First Strategy and User-Centric Data Dissemination

**SBJ-SIAP Collaboration Seminar
4 December 2025**

Mr. Edmond Lee, Group Director/Social and Prices and
Director/Business Statistics,
Singapore Department of Statistics

Administrative and Big Data First Strategy

Administrative and Big Data First Strategy

- DOS makes use of a wide range of data sources, prioritizing use of administrative and big data over conducting surveys.
- Reduce respondent burden and data collection costs while supporting the compilation of disaggregated statistics

Administrative Data

- Prioritised for its reliability, comprehensiveness and shareability across government agencies
- Involves collaboration with govt agencies to identify admin sources, build data pipelines, and study data scope and definitions before integrating them into statistical processes.

Big Data

- Other non-survey sources unbound by respondent participation or legislative constraints offer easier access and re-use
- Data collection and processing require new tools and methods

Survey Data

- Surveys conducted to collect data not available from admin or big data sources.
- Higher data collection costs and reliance on cooperation from respondents.

Administrative Data has been Integral to DOS's work – Progress since 1990

Per Cent of Statistics Using Admin Data

1990s – Use of Administrative Data

- Admin data has always been an important data source for National and International Accounts and Input-Output (IO) tables, e.g. merchandise trade, Central Provident Fund (CPF) indicators, Accountant-General's Department (AGD) data, Immigration and Checkpoint Authority (ICA) and Singapore Tourism Board (STB) visitor arrival data, Monetary Authority of Singapore (MAS) banking data
- Admin records used to directly tabulate statistics on births, deaths, marriages and divorces
- Starting from Census 1990, CPF admin data used to augment survey wage data

2000s - Expanding Use of Data Integration

- Register-based approach adopted for Census 2000 and estimates of Singapore population and overseas Singaporeans.
- Estimation of taxes and government transfers received by households and major household wealth components
- Development of register-based cohort indicators and Individual-level and Firm-level Longitudinal Administrative Databases.
- Cross check Household Expenditure Survey (HES) reporting
- Use of AGD's data on Statutory Boards for National Accounts
- Use of admin data for Corporate Sector statistics, inward Foreign Direct Investment (FDI) and Outward Direct Investment

2010 onwards - Improved Access to Administrative Data and Incorporate Big Data

- Increased use of Inland Revenue Authority of Singapore (IRAS) and Accounting and Corporate Regulatory Authority (ACRA) data for economic statistics, e.g, short term and annual indicators, Corporate Sector statistics, IO tables, National and International accounts
- Register-based approach for Annual Industry Survey (AIS) and inward FDI
- Increased use of admin data and tapped on commercial data for Outward Direct Investment estimates. Use of admin and commercial data for Inward Foreign Affiliates Statistics
- Increased use of admin/electronic data for Price indices, e.g. IRAS data, Open Electricity Market data, supermarket prices
- Increased use of admin data for National and International Accounts, e.g. ACRA financial accounts, Customs data, MAS banking, credit card and insurance data, employment and wages data for Compensation of Employees, more frequent transactional data (visitor arrivals, hotel statistics, transport ridership)
- Integrate various admin data to derive firm location
- Register-based Census 2010 and 2020, greater use of admin data to cross check more items in Census and HES.
- Use of web-scraped data for price indices and firm characteristics

1990

1995

2000

2005

2010

2015

2020

2024

Estimated Proportion of Administrative Data used in Statistical Compilation, 1990 to 2024

Subject Matter	1990	1995	2000	2005	2010	2015	2020	2024
Economic Statistics								
Annual Industry Statistics	0%	0%	0%	0%	0%	10%	20%	60%
Wholesale Trade and Business Receipts Indices	0%	0%	0%	0%	70%	70%	70%	70%
Firm-level Longitudinal Administrative Database	N.A.	N.A.	N.A.	N.A.	100%	100%	100%	100%
National Accounts	40%	40%	42%	43%	50%	60%	60%	65%
Balance of Payments & International Investment Position	40%	40%	40%	40%	40%	50%	60%	65%
Direct Investment & Services Trade	0%	0%	5%	5%	5%	30%	30%	65%
Corporate Sector Statistics	0%	0%	10%	10%	10%	100%	100%	100%
Supply-Use and Input-Output Tables	40%	40%	40%	45%	50%	60%	65%	65%
Price Indices								
Consumer Price Index	5%	10%	10%	10%	15%	25%	30%	35%
Producer Price Indices	0%	0%	0%	0%	0%	11%	11%	12%
Population and Household Statistics								
Population Census and General Household Survey	10%	10%	15%	11%	16%	N.A.	20%	N.A.
Population Estimates and Profile	0%	0%	100%	100%	100%	100%	100%	100%
Vital Statistics (births, deaths, marriages and divorces)	100%	100%	100%	100%	100%	100%	100%	100%
Cohort indicators and Individual-level Longitudinal Administrative Database	0%	0%	0%	100%	100%	100%	100%	100%
Household Income, Tax and Benefits	0%	0%	40%	40%	50%	60%	70%	70%
Household Expenditure Survey	0%	0%	0%	0%	5%	10%	20%	25%
Household Wealth	N.A.	N.A.	20%	25%	30%	40%	40%	40%

Key Learning Points

Challenges

Understanding of Source Data

Administrative data consistency and quality

Changes in administrative systems

Data may not be readily available in structured, machine-readable format

Existing statistical systems and officers may not be sufficiently equipped

Actions

- Collaboration with source agencies to identify data sources and study data scope and definitions

- DOS provides regular feedback on data quality and assists data source agencies in data capability building via knowledge sharing
- Use of estimation methods to impute missing data
- Agencies are to notify DOS in advance of changes in admin systems, to allow DOS time to assess the impact and implement measures to minimize disruption to statistical production

- Leveraging AI/ML to extract and process unstructured data

- System modernization to support complex, large scale business processes while upholding data governance standards
- Capability development of DOS officers

DOS will continue to leverage new technology to expand use of administrative and big data

Leveraging new technologies & Develop Officers' Capabilities

- Utilise AI-driven processes to directly extract detailed information from unstructured financial statements and company websites
- Expand use of ML models to estimate services trade and trading partners.
- Further leverage web-scraping and expand to using APIs for items such as prices of clothing, overseas hotels, package tours and mobile phones.
- System modernization while upholding data governance standards— open source solutions, scalable cloud-based systems, ETL tools, co-sourcing approach.
- Continue to develop officers' capabilities

Expanded use of administrative data and new data sources

- Census of Population 2030, ~60% of data items expected to tap on administrative data
- Expand the longitudinal databases to include more data variables.
- New breakdown of Services Trade by Enterprise Characteristics
- Compile price indices using new administrative data sources, such as management fees and sinking funds, taxi fares, childcare and infant care fees, as well as trade data.
- Register-based approach to compile quarterly business receipts

Data Sharing and Communications Strategy with the Key User Groups

Our Key User Groups

- Dissemination Strategy Tailored for Each User Group



Whole-of-Government



Businesses



The Public



Researchers

Meeting Data Needs of Whole-of-Government



- Provide new economic and social **data for forward sensing** in close consultations with user agencies to support policy making
- Undertake economic and social **analytic projects** with government agencies, and support studies using **anonymised microdata**
- Empower agencies to **self-discover data insights** for policy planning, operations and service delivery
- Streamline WOG processes by **productionising and scaling AI/ ML models**
- Facilitate data comparability with the development of **national statistical classifications**
- **DOS Analytics Services Hub (DASH)** – First-stop shopfront for public officers to access DOS's full suite of data products and services for evidence-based policy formulation, evaluation and service delivery

Meeting Data Needs of Whole-of-Government



Visual Statistics and Analytics (VISTA) portal: central platform that enable self-help access to a wide range of data and analytical insights



- **Thematic resource hub** with consolidated information organised by themes (e.g. population, economy) to streamline data discovery and maximise accessibility
- **Visual analytics portal** for creating, sharing, and exploring interactive dashboards that support decision-making
- **No-code analytics tool** to generate customised statistical tabulations and perform regression analyses to support exploratory, deep dive and simulation studies and research



Meeting Data Needs of Whole-of-Government

- In 2019, DOS was designated as the Trusted Centre for Individual and Business Data (DOS TC).
- DOS TC shares **individual and business administrative data** within the government on a need basis under the Public Sector (Governance) Act, to facilitate the use of data for **policy analysis, operations and service delivery**

Improved Data Quality And Timeliness

TC takes in and processes data from source agencies, performing validation checks and working with them to uplift data quality and ensure data timeliness.

Seamless Discovery And Informed Usage

TC maintains a digital metadata catalogue and synthetic datasets on the data sharing platform, to facilitate self-service data discovery and exploratory studies.

Upholding Data Governance Standards

TC shares data on a need-to-know basis, subject to a rigorous request approval process. De-identified data are shared for policy analysis. Identifiable data are shared for operations and service delivery

Wide Range Of Frequently-updated Data

TC makes available a wide range of ready-to-use, frequently demanded datasets across various subject matter refreshed at up to daily frequency.

One-stop Shop For Data Needs

TC works with government users to understand the project's data needs and provide data advice accordingly, acting as the single point of contact for users.

Secure and Efficient Data Sharing

TC distributes data to government users securely and efficiently via various modes of transfer, e.g., on-demand access via API. The requesting agency is responsible for the safe use of data it has received.

Meeting Data Needs of Businesses



BUSINESS INSIGHTS TOOL
FOR ENTERPRISES

BITE-SIZED DATA FOR BIG DECISIONS

Powered by the Singapore Department of Statistics

Provide curated and contextualised content to empower businesses to make data-driven decisions with insights on their customers, industry, and business performance

Developed for Retail Trade, F&B, Wholesale Trade, Training, Tuition and Enrichment Courses, and Personal Care and Repair industries, in consultation with key partner agencies and incorporating feedback from businesses

Scan to Explore



<https://www.singstat.gov.sg/data-for-businesses>

Benchmarking Tool

**Benchmark own business performance
against industry/ peers**

**Benchmark My
Performance**

**Know My
Customers**

**Know My
Industry**

Customer Profiler
**Gain insights into the
profile and trends of
customers and market**

Industry Profiler

**Gain insights into
industry trends, business
outlook, business costs
and labour market
situation**

Meeting Data Needs of the Public

- Promote understanding and data exploration with more infographics and dashboards on the **SingStat Website**
- Facilitate access to around 150,000 data series from 70 public sector agencies via the **SingStat Table Builder**
- Ease access to commonly used data when on the go via the enhanced **SingStat Mobile App** with improved user-interface and functionalities



Meeting Data Needs of the Public

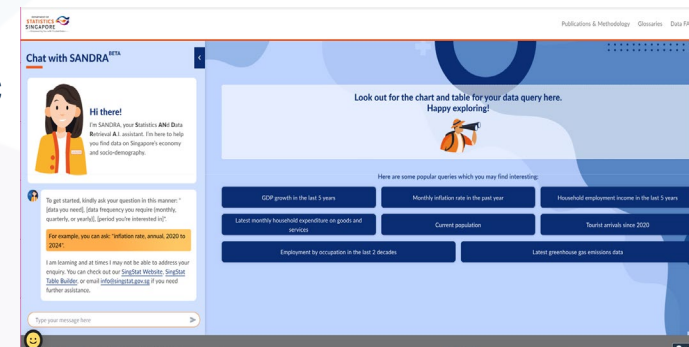
- Assist users with their data enquiries on Singapore's economy and socio-demography with the **Statistics ANd Data Retrieval A.I. assistant (SANDRA)**, an artificial intelligence (AI) chatbot that uses natural language



- Understand context of users' queries through semantic search capabilities and use of Large Language Model

- Retrieves data from the **SingStat Table Builder** and presents the most relevant data in chart and table formats, while suggesting related data to facilitate data exploration

- Expand outreach to a younger population with DOS's Instagram channel, adding to existing **social media channels** (i.e., Telegram, Instagram, YouTube, LinkedIn)
- Expand outreach to the
 - public through **digital displays** in housing estates across Singapore on DOS's data digital services,
 - students via the **Students' Corner** on the SingStat Website



Meeting Data Needs of Researchers



- Enable academic **researchers*** to access wide range of anonymised microdata to conduct advanced analyses on various economic and social topics, through the **Anonymised Microdata Access Programme (AMAP)**



- Aligned with the internationally practised 5-safes framework, approved researchers can access rich detailed anonymised microdata on firms and individuals in a physically secure environment, with only non-disclosive statistical results allowed as output

Annual Industry Surveys/Census of Manufacturing Activities	Census of Population/ General Household Survey	Household Expenditure Survey
• Operating revenue breakdown (e.g. sales of goods, services income, leasing income, royalties) • Operating expenditure breakdown (e.g. purchases of goods, rental costs, freight & transport, maintenance & repair) • Fixed assets breakdown (e.g. building & structure, transport equipment, computers & peripherals, other machinery & equipment) • Total remuneration • Value-added • Operating surplus	• Demographics (e.g. age, sex, marital status, no. of children born for female) • Economic characteristics (e.g. employment status, occupation, industry, monthly income from work, usual hours worked) • Education profile (e.g. highest qualification, field of study, country where highest qualification is attained, level of education attending) • Household characteristics (e.g. house type, tenancy, household income) • Transport patterns (e.g. mode of transport, travel time)	• Demographics (e.g. age, sex, marital status) • Economic characteristics (e.g. employment status, occupation, industry, monthly income from work) • Education profile (e.g. highest qualification, level of education attending) • Household characteristics (e.g. house type, tenancy, household size) • Household income from all sources (e.g. work, rental, investment, government transfers) • Household expenditure details (e.g. food, housing & utilities, health, education, transport, recreation at detailed levels) • Availability of consumer durables/ services (e.g. car, air-conditioner, PC, mobile phone, Internet access)

*Eligible researchers holding full-time appointments in Singapore's local Autonomous Universities

Thank You

Our Vision

National Statistical Service of Quality, Integrity, and Expertise

Our Mission

We Deliver Insightful Statistics and Trusted Statistical Services that Empower Decision Making



www.singstat.gov.sg



info@singstat.gov.sg



[@SingStat](https://t.me/SingStat)



[@singstat_dos](https://www.instagram.com/singstat_dos)



[@SingStatvideo](https://www.youtube.com/SingStatvideo)



[@sg-department-of-statistics](https://www.linkedin.com/company/sg-department-of-statistics)