STATISTICS SINGAPORE NEWSLETTER

ISSUE 2, 2024

Feature Articles

Women-owned Companies in Singapore

International studies found that women-owned companies contributed revenue and jobs to the economy, despite having fewer employees and lower revenue than men-owned enterprises on average. This article provides an overview of the Singapore Department of Statistics' (DOS) study on the prevalence, characteristics, and growth of women-owned companies, highlighting the trends between 2010 and 2022.

Read on page 2

Use of Administrative Data and New Data Sources in Industry Statistics

DOS produces short-term and annual estimates of industry statistics for the services sector, with surveys as the traditional method for collecting data. This article shares DOS's efforts in using administrative and non-traditional data sources to streamline data collection and improve the data compilation process for both short-term and annual industry estimates.

Read on page 5

Integrated Longitudinal Database for Supporting Policy Studies

The Singapore Government adopts a data-driven approach in shaping policies using data analytics to examine issues, identify effective strategies, and assess the impact of policy interventions. This article provides an overview of DOS's integrated Longitudinal Administrative Database and shows how it supports evidence-based research and policy studies in Singapore.

Read on page 9

Use of AI-Enhanced OCR and Machine Learning for Data Processing in the Household Expenditure Survey 2023

In the Household Expenditure Survey 2023, DOS automated traditionally labour-intensive data processing tasks using artifical intelligence (AI)-enhanced optical character recognition (OCR) and machine learning techniques, resulting in higher accuracy and significant time and resource savings.

Read on page 13

More in this Issue:

The Statistics Singapore Newsletter is issued twice a year by the Singapore Department of Statistics. It aims to inform readers on recent statistical findings as well as latest information on statistical methodologies, processes, products, and services.

Submit Feedback on the SSN | Contact Us

Find Us on:

Telegram | Instagram | YouTube | LinkedIn

Women-owned Companies in Singapore

by Jun Hwi Leng and Eric Ng
Business Statistics Division
Singapore Department of Statistics

Introduction

Entrepreneurship plays an instrumental role in the economy by creating jobs, driving innovation, sustaining business dynamics, and contributing to economic growths. Developing entrepreneurial talents in Singapore creates business opportunities and supports inclusive and sustainable growth.

The Singapore Government is committed to achieve this by working with various stakeholders to support entrepreneurs and enterprises in uplifting capabilities and seizing new growth opportunities. For example, to support women entrepreneurs, the Singapore Business Federation and the Action Community for Entrepreneurship, have established the Singapore Women Entrepreneurs Network and Women Entrepreneurship Committee respectively.

International studies [1] found that women-owned companies contributed revenue and jobs to the economy, despite having fewer employees and lower revenue than men-owned enterprises on average.

Against this backdrop, the Singapore Department of Statistics (DOS) conducted a statistical study on the prevalence, characteristics, and growth of companies [2] that are majority-owned by women. This study focused on Singapore resident-owned companies, defined as those that generate revenue or have employees and have over 50% of their ordinary shares held by Singapore residents [3]. These companies are further classified into:

Women-owned Companies

Companies with more than 50% of their ordinary shares held by female Singapore residents.

Men-owned Companies

Companies with more than 50% of their ordinary shares held by male Singapore residents.

Other Companies

Companies where male and female Singapore residents do not separately own more than 50% of their ordinary shares. For example, male residents own 30% and female residents own 21%, with the remaining owned by foreigners.

This study integrated data from DOS's Statistical Business Register (SBR), Firm-Level Longitudinal Database (FLAD), and Individual-Level Longitudinal Database (ILAD).

Key Findings

The number of women-owned companies grew from 16,300 in 2010 to 36,000 in 2022 (Chart 1). In 2022, they accounted for 25.6% of Singapore resident-owned companies, higher than the 20.0% in 2010 (Chart 2). However, men-owned companies continued to be the largest group in 2022.

Chart 1: Number of Singapore Resident-owned Companies by Type of Ownership

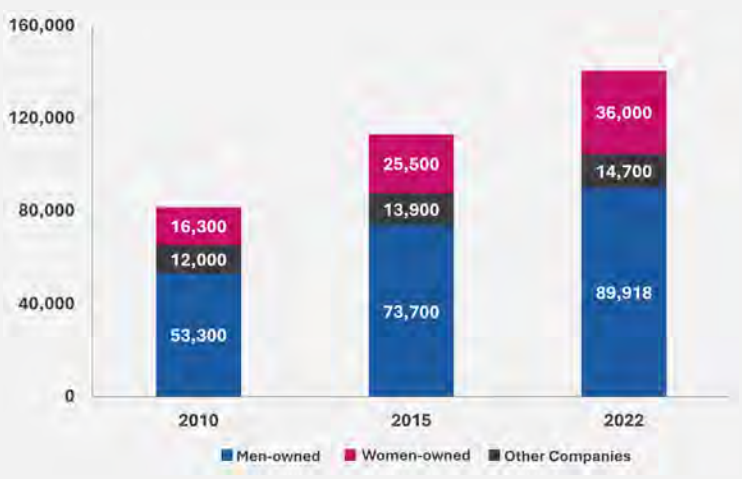
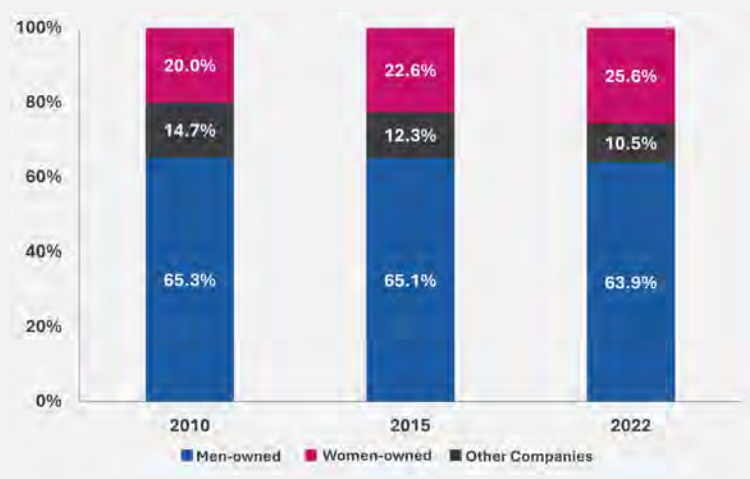


Chart 2: Distribution of Singapore Resident-owned Companies by Type of Ownership



[1] For more information, refer to the References appended at the end of the article.

[2] Companies refer to business entities registered under the Companies Act. A company is business form that is a legal entity separate and distinct from its shareholders and directors. Company shareholder information is as declared by companies to the Accounting and Corporate Regulatory Authority.

[3] Singapore residents comprise Singapore citizens and permanent residents.

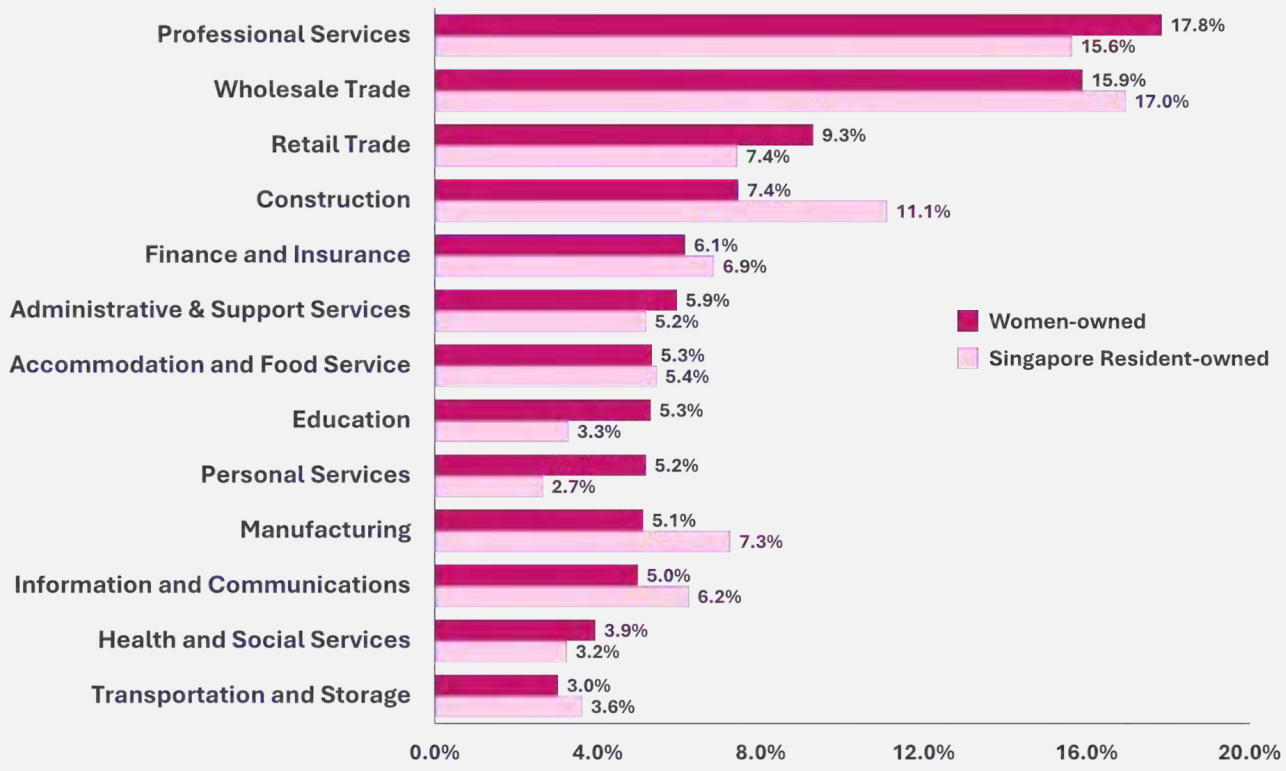
Industry Distribution

In 2022, the top three industries with the highest number of women-owned companies were in the services sector; Professional Services, which accounted for 17.8% of women-owned companies, followed by Wholesale Trade (15.9%), and Retail Trade (9.3%) (Chart 3).

Among all Singapore resident-owned companies, women-owned companies had lower representation in the goods-producing industries. In 2022, only 7.4% and 5.1% of the women-owned companies were in the Construction and Manufacturing industries respectively. The corresponding proportions for all Singapore resident-owned companies were 11.1% and 7.3% respectively.

For the Professional Services, Education, and Personal Services industries, the proportion of women-owned companies was higher compared to all Singapore resident-owned companies with a difference of 2.0 percentage points or more.

Chart 3: Distribution of Singapore Resident-owned and Women-owned Companies by Industries, 2022



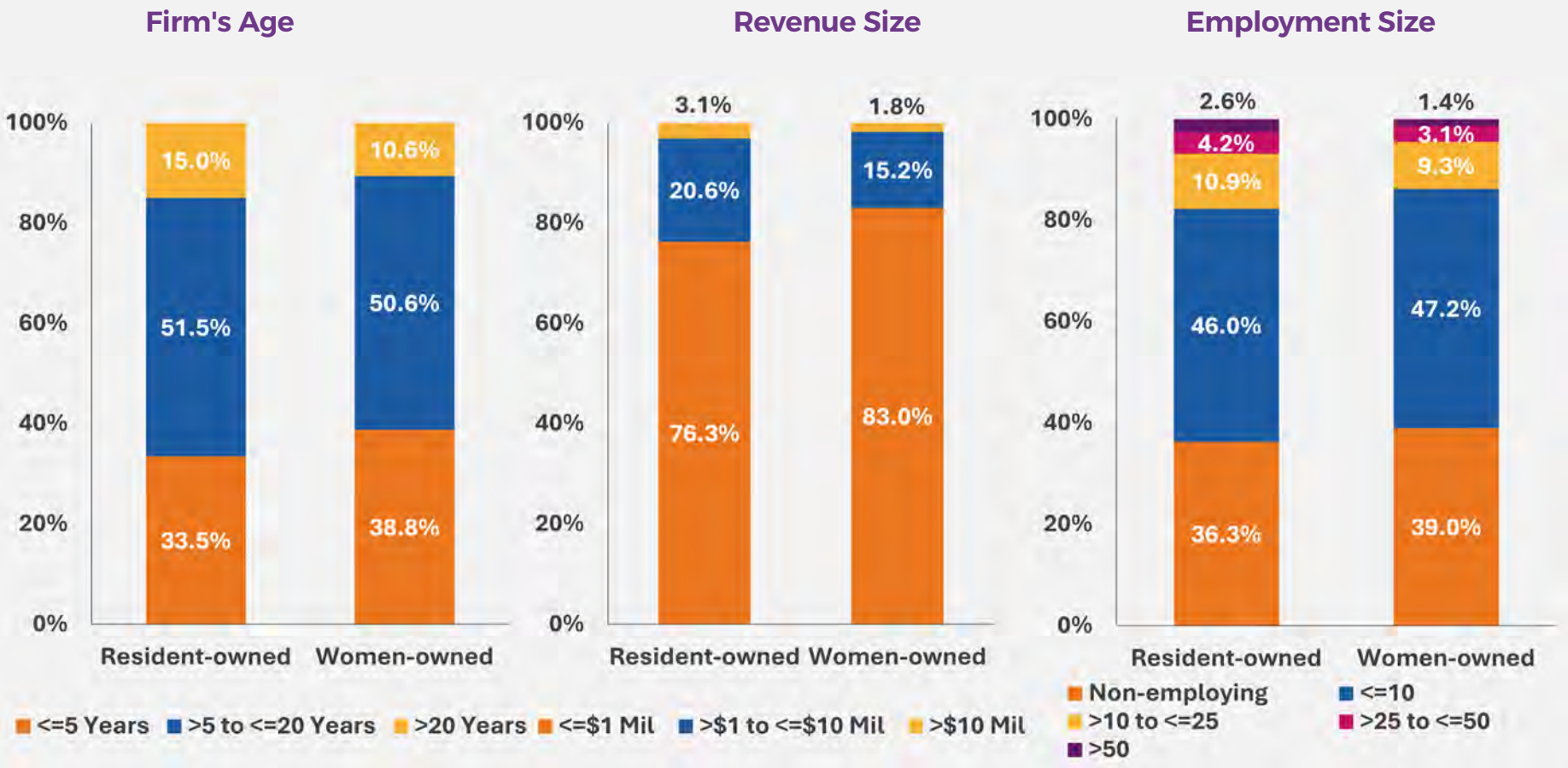
Profile of Women-owned Companies

Women-owned companies are likely to be younger and have smaller revenue and employment sizes compared to all Singapore resident-owned companies.

In 2022, 38.8% of women-owned companies were five years old or younger, while 10.6% were over 20 years old. In contrast, among all Singapore resident-owned companies, 33.5% were five years old or younger, and 15.0% were over 20 years old (Chart 4).

In terms of revenue, 17.0% of women-owned companies earned more than $1 million, compared to 23.7% of all Singapore resident-owned companies in 2022. For employment size, 13.8% of women-owned companies employed more than 10 employees, less than the 17.7% of all Singapore resident-owned companies in 2022.

Chart 4: Profile of Singapore Resident-owned and Women-owned Companies, 2022



Contribution of Women-owned Companies to the Singapore Economy

In 2022, women-owned companies generated $85.3 billion in revenue and employed 202,000 employees. Compared to 2010, this represented an average annual growth of 11.3% and 4.8% respectively.

Women-owned companies accounted for 21.3% of the total revenue of Singapore resident-owned companies in 2022, an increase from 10.3% in 2010 (Chart 5). Similarly, the employment share of women-owned companies rose from 13.2% in 2010 to 18.5% in 2022 (Chart 6).

The revenue and employment share of men-owned companies correspondingly declined over the same period. Nonetheless, men-owned companies remained the largest contributor to revenue and employment among Singapore resident-owned companies in 2022.

Chart 5: Revenue Distribution of Singapore Resident-owned Companies by Type of Ownership

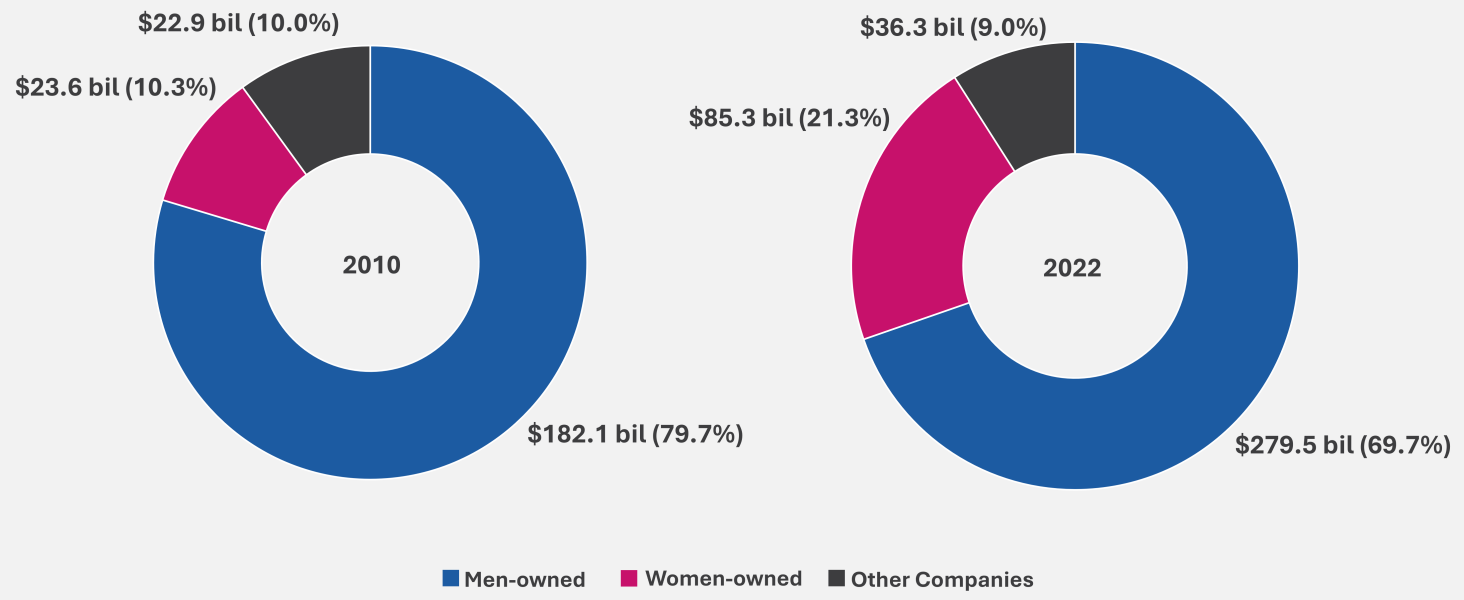
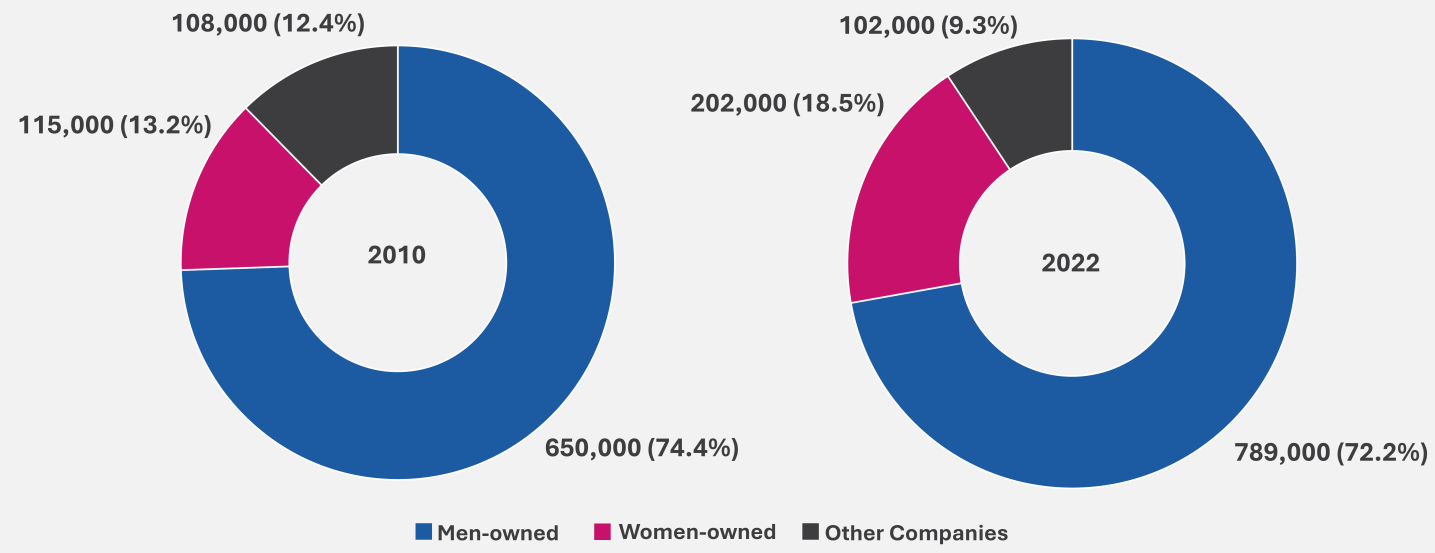


Chart 6: Employment Distribution of Singapore Resident-owned Companies by Type of Ownership



Conclusion

Findings from this study showed that women-owned companies have progressed well from 2010 to 2022. Women-owned companies have grown in number and have contributed revenue and jobs to the Singapore economy. As they continue to grow, women-owned companies are expected to further contribute to the Singapore economy.

References

- HM Treasury (2019), **The Alison Rose Review of Female Entrepreneurship**
- Global Entrepreneurship Monitor (2023), **GEM 2022/23 Women's Entrepreneurship Report: Challenging Bias and Stereotypes**
- MSF (2024), **Progress on Singapore Women's Development 2024**
- MTI (2024), **Singapore Affirms Commitment to Promote Sustainable and Inclusive Growth**
- Nordic Innovation (2020), **Female Entrepreneurship in the Nordics 2020, a Comparative Study**
- OECD (2021), **Entrepreneurship Policies through a Gender Lens**
- US Census Bureau (2023), **Women-Owned Employer Businesses in the United Statistics**
- Women Entrepreneurship Knowledge Hub (2023), **The State of Women's Entrepreneurship in Canada 2023**

Use of Administrative Data and New Data Sources in Industry Statistics

by Wong Yan Ru and Neo Soo Khee
Business Statistics Division
Singapore Department of Statistics

Introduction

The Singapore Department of Statistics (DOS) produces short-term and annual estimates of industry statistics for the services sector [1]. These statistics are used by policymakers, researchers, and the business community, to study the structure and performance of various activities in the services sector.

Surveys have traditionally been the primary method for collecting data to produce industry statistics. DOS is moving away from conventional business surveys by leveraging administrative and non-traditional data sources. Using artificial intelligence (AI) and machine learning (ML), DOS can extract and process new data sources more efficiently. This reduces respondent burden while enhancing the ability to collect and compile more comprehensive data on businesses.

This article shares DOS's efforts in using administrative and non-traditional data sources to streamline data collection and improve the data compilation process for both short-term and annual industry estimates. It also covers DOS's experience in producing new data products based on administrative data and new data sources, as well as the use of AI and ML to utilise new data sources.

Administrative Data

Administrative data are collected by organisations for administrative purposes (e.g., regulatory, accounting, or commercial). For the compilation of industry statistics, the administrative data relating to revenue, profit, and wages acquired from different government agencies in Singapore are listed below.

1 Regulatory Authority of Business Registration & Financial Reporting:
Accounting and Corporate Regulatory Authority (ACRA) of Singapore

Financial Information such as Revenue and Profit

Identification and demographic information of businesses such as:

- Unique Entity Number (UEN)
- Business name and address
- Registration date
- Shareholder information
- Industrial Classification

2 Manpower Authority:
Ministry of Manpower (MOM) & Central Provident Fund Board (CPF) [2]

Employment and Wages

3 National Tax Authority:
Inland Revenue Authority of Singapore (IRAS)

Financial Information such as Revenue and Profit from Goods & Services Tax (GST) [3] and Corporate Income Tax.

Stamp duty on tenancy agreements (e.g., UEN, property address, and tenancy start and end dates)

4 Housing Authority:
Housing and Development Board (HDB)

Rental of commercial properties under HDB management such as:

- UEN
- Property address
- Start and end dates of tenancy

[1] The Services Sector includes enterprises engaged in the Wholesale Trade, Retail Trade, Transportation & Storage, Accommodation & Food Services, Information & Communications, Real Estate, Professional Services, Administrative & Support Services and Community & Personal Services. Enterprises engaged in Finance & Insurance Services and Public Administration Activities are excluded. For more information on industry statistics of the services sector, please refer to the **SingStat Website**.

[2] The Central Provident Fund (CPF) is a mandatory social security savings scheme in Singapore, funded by contributions from employers and employees. Employers of local workers (Citizens and Permanent Residents of Singapore) are required to declare wage information and pay employees' CPF contributions monthly. Enterprises employing foreign workers must apply for relevant work passes from the MOM and inform MOM of changes such as revisions in employee's salary and updates to personal particulars.

[3] GST is a broad-based consumption tax levied on the purchases of goods and services in Singapore. Enterprises with annual taxable turnover exceeding S$1 million must register for GST, while those under S$1 million can voluntarily register for GST. GST data includes timely information on enterprises' revenue, as enterprises are required to file for GST on a quarterly basis, within one month after the end of the accounting period. GST is also known as the Value-Added Tax (VAT) in other countries. For more information, please refer to the **IRAS website**.

As administrative data may differ in coverage and/ or concepts from statistical concepts and definitions, additional work was undertaken to study the differences before they can be incorporated into the production of industry statistics for the services sector.

Short Term Indicators

DOS compiles the monthly Retail Sales Index (RSI), Food & Beverage Services Index (FSI), quarterly Wholesale Trade Index (WTI), and the Business Receipts Index (QBRI). These indices measure the short-term performance of services industries.

These indices are mainly compiled using revenue data collected via monthly or quarterly surveys. Administrative GST data and non-traditional data [4] such as Gross Merchandise Value (GMV) from third-party marketplaces are used to supplement the compilation.

For the quarterly compilation of WTI and QBRI, smaller enterprises are not surveyed as their GST data is sufficient to be used for estimation. Around 80% of the enterprises in the scope of WTI and QBRI are estimated using GST data currently. Larger enterprises will continue to be surveyed in the Quarterly Survey of Services due to differences in coverage and/ or definitions. For example, when reporting for GST, some enterprises may opt to report group level data instead of enterprise level data. Due to the differences in the definition of revenue, government subventions [5] for public and non-profit institutions required for the QBRI compilation are not included in GST revenue. Additional data items such as related party transactions and operating expenditure that are not in GST are collected via surveys.

The monthly compilation of the RSI and FSI includes data on the online proportions of retail and food & beverage (F&B) sales. Data from the monthly surveys are supplemented with the GMV data collected from third-party marketplaces. This ensures a more comprehensive coverage of smaller enterprises, which are more likely to sell their products through third-party marketplaces instead of setting up a website to do so.

Annual Industry Survey

The Annual Industry Survey (AIS) collects information to analyse the structure and performance of enterprises in the Services Sector. The data are used to compile the national accounts, input-output tables, and estimates on indicators such as annual Operating Revenue and Operating Expenditure. These data also support related studies by other economic agencies such as the Ministry of Trade and Industry and Enterprise Singapore.

Transitioning to Register-based Approach

Survey Approach

Previously, the AIS was entirely based on a survey approach. All large enterprises were sampled with certainty, while medium and smaller-sized enterprises were sampled using simple random sampling without replacement. Non-sampled enterprises were represented by sampled enterprises via sampling weights for the compilation of industry statistics.

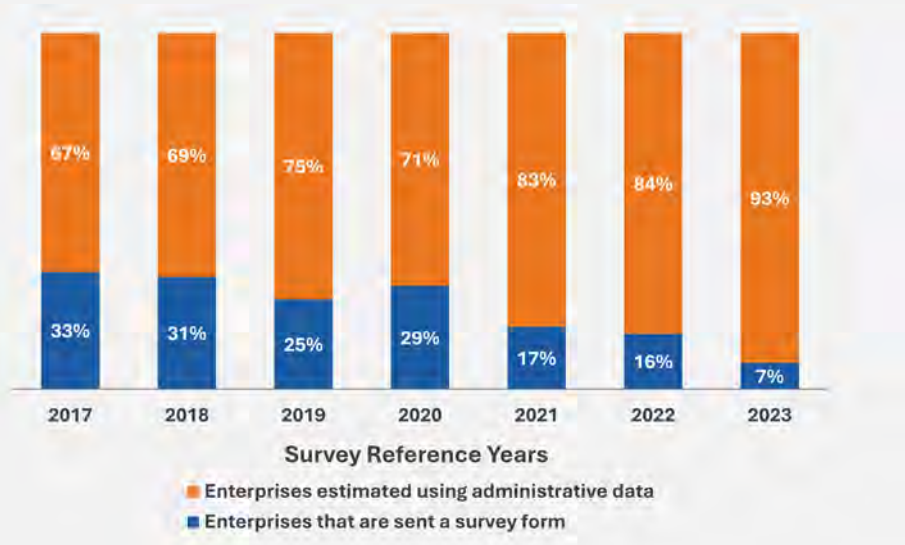
Register-based Approach

DOS is progressively adopting a register-based approach for the AIS where all enterprises are included and estimated using administrative data, compared to relying solely on survey returns. The transition begins with industries where administrative data are assessed to align with statistical concepts and has comprehensive coverage. Larger enterprises are still surveyed to collect detailed breakdowns of their revenue and expenditure items [6] that are not available in administrative data.

By adopting the register-based approach for selected industries, the overall sampling variability is reduced. This in turn improves the quality of the industry estimates developed from the information collected and facilitates data compilation with more granular breakdown. To date, DOS has implemented the register-based approach for around 50% of the services industries in the AIS.

As a result, while the overall number of enterprises covered in the AIS has increased in line with the population size of services enterprises in Singapore over the years, the proportion of enterprises surveyed decreased from 33% in 2017 to 7% in 2023 (Figure 1).

Figure 1: Breakdown of Enterprises by Data Collection Source for AIS, 2017-2023



[4] Non-traditional data refer to information that is not typically captured or analysed using conventional methods or sources.

[5] Government subvention is a revenue item for public and non-profit institutions. It is excluded from the calculation of taxable supplies.

[6] Examples of detailed revenue and expenditure items include freight charges, accounting, auditing and book-keeping fees, legal fees, consultancy fees, etc.

Use of XBRL Data and Machine Learning

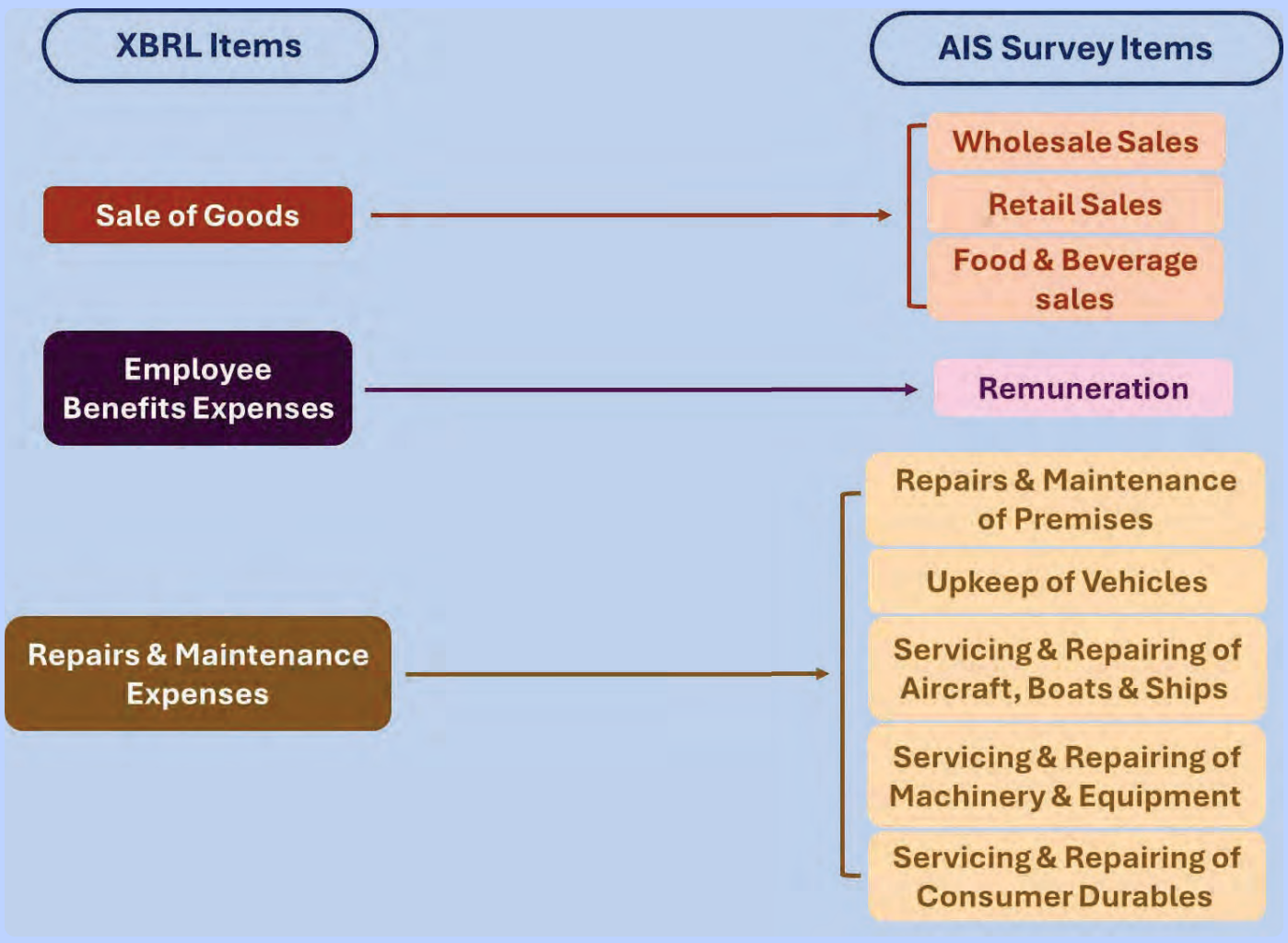
As most administrative data sources only include topline income or expense items at the enterprise level, additional data processing is required when administrative data is used to compile industry statistics. This involves splitting the enterprise-level administrative data to the establishment-level based on the industry proportion, mapping the administrative data items to the relevant income and expense items covered in AIS, and appropriately proportioning the less granular administrative data amongst these survey items.

An example of the data processing steps is illustrated using ACRA's eXtensible Business Reporting Language (XBRL) data [7]. The XBRL format provides a range of detailed income and expense items, such as revenue, profit, sale of goods, rendering of services, employee benefits expenses, repairs & maintenance expenses, and depreciation. Topline items, such as revenue and profits, are mandatory for all enterprises to report when filing financial statements in the XBRL format.

Some of the XBRL income and expense items are non-mandatory and enterprises may opt not to provide these data. In order to fill in the missing information, DOS implemented a random forest ML algorithm to predict their values. For each industry, a random forest algorithm is trained for each non-mandatory data item to predict a value based on the responses from other enterprises that have reported the data and its relationship with other data items.

After running the random forest model on the XBRL data items, these items are mapped as closely as possible to the AIS data items before using them to estimate the values (Figure 2). Some XBRL data items are mapped to more than one AIS item and require further proportioning based on the enterprises' industry classification and historical proportion. For instance, the XBRL income item '*sale of goods'* is mapped to three AIS items, i.e., wholesale sales, retail sales, and F&B sales. If the enterprise is a wholesaler, based on historical industry proportion, '*sale of goods'* could be proportioned into wholesale sales (99%) and retail sales (1%).

Figure 2: Example of Mapping XBRL items to AIS items



[7] Refers to data from financial statements filed in the XBRL format. XBRL is a language for electronic communication of business and financial data worldwide. For more information, please refer to the **ACRA website**.

New Data Products

DOS has produced new data products by leveraging administrative data and new data sources and is exploring the use of AI and ML algorithms to further utilise these sources.

Early Triggering with Employment and GST data

The Early Triggering project detects entries and exits of large enterprises or enterprises with significant changes in revenue and/ or employment in a timely manner. Higher-frequency administrative data on employment from MOM and CPFB, along with data on revenue from enterprises' GST filings with IRAS, are used to identify enterprises with changes in employment and/ or revenue exceeding sector-specific thresholds. Information of these identified enterprises is then shared with the relevant subject domain teams in DOS for incorporation into business surveys and the compilation of economic indicators.

Business Activity by Geographic Area

DOS has developed experimental estimates on retail and F&B activities by geographic area.

The estimates for retail and F&B activities are primarily based on enterprises' operating addresses filed with ACRA and the Singapore Food Agency respectively. The estimates for retail activity are further supplemented with data on rental of commercial properties from IRAS and HDB.

These data are integrated with the enterprises' characteristics (e.g., industry classification) before plotting onto a map within a dashboard [8], which provides insights on the distribution of retail and F&B activities across Singapore.

Weekly Economic Brief

DOS has access to a curation of business news from major media outlets based on keywords covering a range of topics such as business performance, mergers and acquisitions, expansion, and restructuring plans of enterprises. The selected news articles are categorised into pre-defined categories for easy reference.

Indicator on Enterprise's Internet Presence and Other Enterprise Characteristics

DOS text mines web-based data to study how enterprises in Singapore use their corporate websites. Keywords such as 'Shop' or 'Cart' are first extracted from these websites. Supervised ML is applied to categorise the websites into different categories (i.e., enterprises with websites that directly generate revenue, or those with websites but do not generate revenue from them) and derive an indicator on enterprises' internet presence. This indicator is then integrated with enterprises' characteristics from the Statistical Business Register to derive new insights for further analyses. For example, the identification of enterprises with corporate websites are early indications of business activities, and they will be included in the sampling frame for business surveys.

Moving forward, DOS is exploring the use of text mining and ML to obtain other enterprise characteristics like green enterprises, innovative enterprises.

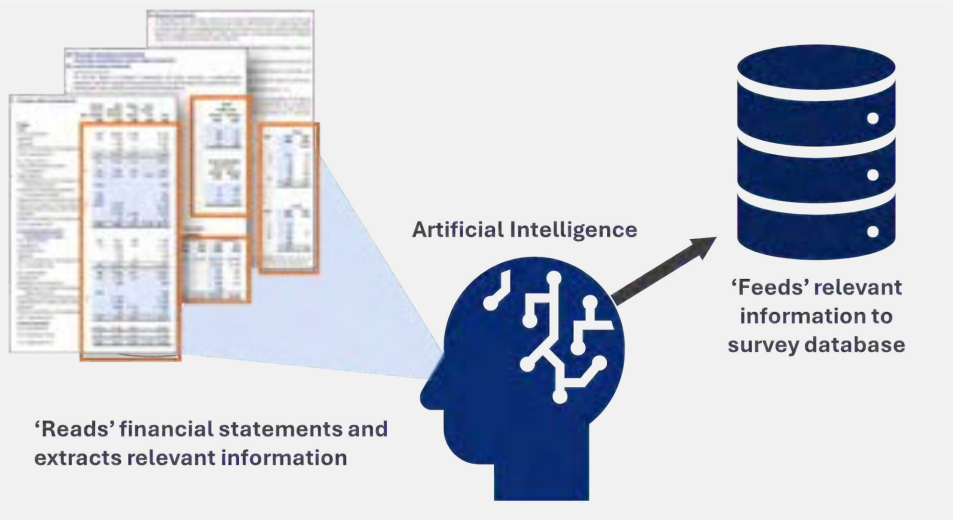
AI on Unstructured Financial Statements

Company financial statements submitted to ACRA or IRAS contain vast information but are presented in an unstructured format and varies significantly across companies. Hence, considerable manual effort is required to extract and interpret relevant business information. For example, one has to manually read, locate, then extract the required information from various sections of the financial statements.

To reduce manual effort, DOS launched a pilot project which uses AI to directly extract detailed information from unstructured financial statements to support the compilation of industry statistics for the services sector (Figure 3). This automates the extraction and interpretation of the unstructured information, which improves operational efficiency and supports the compilation of more comprehensive industry statistics.

DOS is still fine-tuning the algorithm for its eventual use in supporting data compilation. When fully implemented, this will further improve the quality and quantity of administrative data available for data compilation.

Figure 3: Illustration of Using Artificial Intelligence on Financial Statements



Conclusion

While surveys remain an important data source to obtain details not available from administrative sources, DOS has increasingly leveraged and optimised administrative data and non-traditional data sources to improve industry statistics and produce new data products. With the growing availability of administrative data coupled with advancements in data extraction and processing technologies, DOS will continue to explore the use of new data sources to produce relevant statistics for increasingly sophisticated users.

[8] To view the **dashboard** on Business Activity by Geographic Area, make the following selections:
a) For either the Retail or F&B industry, select 'Know My Industry'
b) Within the dashboard, select a Detailed Industry
c) Select 'How is my industry performing?' followed by 'How many firms are engaged in similar activities across Singapore?'

References

- Wallgren, A., & Wallgren, B. (2014). *Register-based Statistics: Statistical Methods for Administrative Data*. John Wiley & Sons.
- EuroStat (2021): **European Statistical System and handbook for quality and metadata reports**
- Singapore Department of Statistics (2022): **Using Big Data to Profile Singapore's Internet Economy**
- Singapore Department of Statistics (2023): **Experimental Use of New Data Sources for Prompt Identification of Changes in Firms' Status**

Integrated Longitudinal Database for Supporting Policy Studies

by Sia Ziyi, Chan Wen Chang and Malcolm Cai
Longitudinal Data Analytics Division
Singapore Department of Statistics

Introduction

The Singapore Government adopts a data-driven approach in shaping policies using data analytics to examine issues, identify effective strategies, and assess the impact of policy interventions. This is done across diverse domains, encompassing the economy, social compact, security, and urban and cultural development. To support this, the Singapore Department of Statistics (DOS) has developed an integrated Longitudinal Administrative Database (LAD).

This article provides a brief overview of longitudinal data and explains why it is a valuable resource for research and policy studies. It then shows the global landscape of such datasets and provides an in-depth look at DOS's LAD. DOS's considerations and measures taken to optimise the dataset's utility while upholding the confidentiality of data are also discussed.

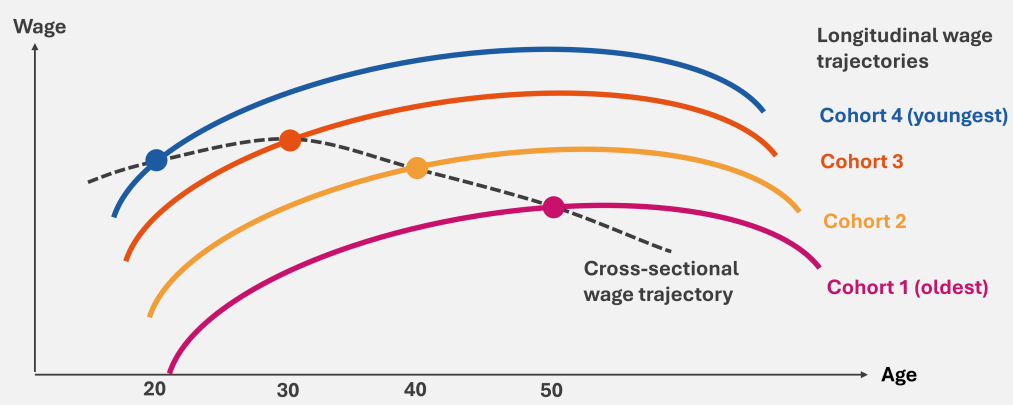
Importance of Longitudinal Data

Longitudinal data refers to data collected on the same research unit (e.g., individual or firm) repeatedly over time. This allows researchers to distil information that may be masked by cross-sectional data.

For example, when studying the lifetime wage trajectory of workers, using cross-sectional data at a particular point in time only captures the wages of different workers at various ages, as depicted by the dotted line in Figure 1, leading to inaccurate conclusions. This is misleading because as a younger worker ages, his wage will not be the same as the wage of a current older worker.

Repeated cross-sectional data can further distort the picture due to compositional changes in the cross-sectional dataset, such as individuals dropping out of older cohorts and the inclusion of new cohorts.

Figure 1: Cross-Sectional vs. Longitudinal Wage Trajectories



Trajectory measured by cross-sectional data can be different from one that is measured by longitudinal data.

In contrast, longitudinal data tracks the same individuals over time, allowing the observation of their actual wage trajectories, as represented by the solid lines in Figure 1, providing a more accurate picture of how wages evolve over a worker's career and better insights into the shifts in wage trajectories across different cohorts.

Using longitudinal data allows Government researchers to better estimate the causal impact of policy levers on outcomes, as the nature of the dataset enables the use of more sophisticated modelling techniques.

- For example, when measuring the economic returns (proxied by wages) to education (a policy lever), other factors such as an individual's innate ability which may also affect the level of wages would need to be accounted for. Individuals with higher ability tend to have both higher education and higher wages, and not accounting for ability would overstate the causal impact of education on wages. With longitudinal data, statistical techniques such as fixed-effects regression modelling can be used to remove confounding effects of time-invariant unobserved characteristics such as ability, providing a better estimate of the effects of education on wages.

International Landscape

International National Statistical Offices (NSOs) construct longitudinal databases for various statistical and research purposes. For example, **Statistics Canada's Longitudinal Administrative Databank** is mainly used by the Canadian Government as a research tool to evaluate programs and support policy recommendations. The Databank can track **low-income persistence** and year-over-year low-income entry and exit rates. The New Zealand Government tapped on its **Integrated Data Infrastructure** to study the performance of their welfare system and social outcomes.

To collect the data required to construct longitudinal databases, NSOs typically tap on either survey data or administrative sources (Table 1). Survey data can be collected through longitudinal surveys, which involve repeatedly surveying the same individuals or firms over time, or by linking survey data from multiple censuses conducted at different points in time. Administrative sources involve linking existing administrative data collected from various sources over time to build a register.

Longitudinal Surveys, such as the Longitudinal and International Study of Adults by Statistics Canada and the National Longitudinal Survey of Youth 1997 by United States Bureau of Labor Statistics, allow for survey questions to be tailored to meet specific data demands. However, conducting large-scale surveys over time can be resource-intensive and impose a burden on respondents, and hence limit their coverage to a portion of the population.

Table 1: Longitudinal Data in National Statistical Offices

NSO/ Country	Longitudinal Database	Period	Scope	Primary [1] Method of Data Collection
Statistics Canada	Longitudinal and International Study of Adults	2011-latest	Sample of households located in ten provinces	Survey (Longitudinal)
United States Bureau of Labor Statistics	National Longitudinal Survey of Youth 1997	1997-latest	Sample of persons born 1980-1984 in 1997	Survey (Longitudinal)
Australian Bureau of Statistics	Australian Census Longitudinal Dataset	2006-latest	5% sample of the population	Survey (Linking Census survey data across years)
Office for National Statistics	Longitudinal Study	1971-latest	1% sample of the population	Survey (Linking Census survey data across years)
Statistics Canada	Social Data Linkage Environment	1926-latest	Generally administrative files to form Derived Record Depository	Administrative
Australian Bureau of Statistics	Person Level Integrated Data Asset	1990-latest	Generally administrative files to form person linkage spine	Administrative
Statistics Canada	Longitudinal Administrative Databank	1982-latest	20% sample of the population	Administrative
Statistics New Zealand	Integrated Data Infrastructure	1840-latest	All New Zealand residents	Administrative
Statistics Sweden	Longitudinal Integration Database for Health Insurance and Labour Market Studies	1990-latest	Persons aged 16 and over	Administrative
United States Census Bureau	Longitudinal Employer-Household Dynamics	1990-latest	95% of the employed population	Administrative

[1] Primary method refers to the main method of data collection. Data can be augmented with other sources (i.e., Administrative source supplemented with survey data, or vice versa).

Agencies such as Statistics Sweden and Statistics New Zealand primarily rely on administrative data to build their longitudinal databases. Administrative sources [2] often provide almost complete coverage of the target population, allowing for studies to focus on specific groups of people (e.g., individuals on social assistance with specific attributes) with less sampling error. The data is also more accurate, as multiple administrative sources allow for cross-checking and validation of the figures. However, the available data is restricted to data collected for administrative purposes, which may limit the range of variables in the longitudinal dataset. Nonetheless, this limitation can be addressed by augmenting the database with additional data (e.g., from surveys).

Maximising the Value of LAD for Policy Analyses

In Singapore, DOS has built the LAD mainly using administrative data augmented with survey data. Administrative processes in Singapore are mostly digital, where data are automatically captured during transactions with the Government, making data collection cheaper, faster, and less burdensome for citizens and firms compared to traditional surveys. Nevertheless, data from surveys are used to augment and update administrative data where necessary.

Context

The LAD was set up to address the policy research needs of the Government and advance the national capability for social science research, particularly in the analysis of social issues such as income mobility and household dynamics. Over time, the LAD has been expanded to support research in a wide range of domains, including analyses of the labour market, education, healthcare, and the corporate landscape.

The LAD is centrally built and managed by DOS, leveraging DOS's expertise in collecting, cleaning, merging, and processing data, as well as its experience in data governance and security. In addition, DOS is familiar with data concepts and definitions of various primary administrative databases within the Government. Considering economies of scale, a centralised LAD is more cost-effective than setting up separate databases in individual agencies.

Over the years, the LAD has supported many research studies within the public sector. In the economic domain, the LAD was used to study the **corporate landscape of Singapore**, aiding Singapore's Central Bank to gauge the financial and productivity performance of firms. Similarly, DOS tapped on the LAD to profile **high growth firms in Singapore**. In the social domain, the Ministry of Finance has leveraged the LAD to estimate **intergenerational mobility** by examining the correlation between the incomes of fathers and their sons.

The LAD has also supported numerous policy evaluation studies. For example, it was used to access the impact of **Workforce Skills Qualifications (WSQ) training on trainees' wages and employment prospects** and to study the effects of Enterprise Singapore's **grants on firms' revenue and exports.** The longitudinal nature of the datasets allows researchers to control for unobserved individual- or firm-specific factors that are time-invariant during the period of analysis.

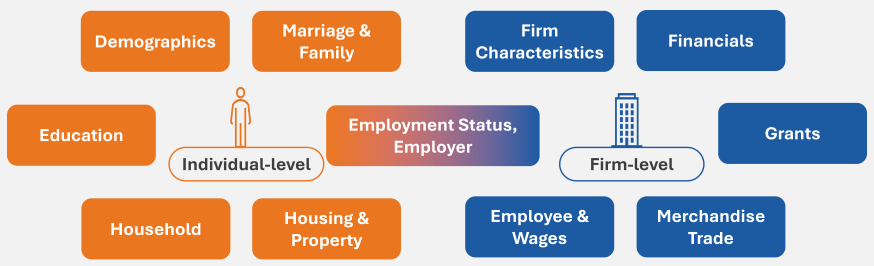
Clean and Rich Data

To ensure that the LAD is logically consistent and coherent over time, DOS developed longitudinal processing capabilities to identify longitudinal anomalies such as inconsistency in records over time. Such anomalies include changes in time-invariant characteristics (e.g., ethnicity), decreases in values that should not decline over time (e.g., highest educational qualification), or cases where deceased individuals erroneously appear in the 'living persons' dataset. Such records would be flagged, investigated, and rectified before storing in the database as clean data.

Besides record-level checks, DOS monitors trends of aggregated statistics (i.e., mean and median) generated through the LAD to detect any trend breaks which may occur from changes in administrative processes, and in turn affecting data collection procedures. DOS ensures that data and definitions remain comparable over time. For example, when the Singapore's Standard Industrial Classification [3] is revised, industry data from earlier years are re-mapped to the latest classification to ensure continuity and comparability over time.

Both individual- and firm-level observations are included in the LAD. At the individual level, it contains variables on employment status, employer, demographic profile, education, households, marriages, and family. Family information includes parental linkages, enabling comparisons between child and parent outcomes to study intergenerational mobility. At the firm-level, the LAD captures key characteristics such as financials, grants, merchandise trade, employees, and wages. Figure 2 illustrates how individual- and firm-level data can be linked to form an employee-employer dataset for hierarchical analysis.

Figure 2: Longitudinal Datasets for Individuals and Firms



The LAD can be fused with external datasets on an ad-hoc basis to create rich, domain-specific datasets for Government researchers to leverage for further deep-dive analyses into specific domains.

[2] Learn more about **Using Administrative and Secondary Sources for Official Statistics [The Advantages of Using Administrative Sources, section 2.5, page 10]**

[3] The **Singapore Standard Industrial Classification (SSIC)** is the national standard for classifying economic activities undertaken by economic units and is used in censuses of population, household and establishment surveys, and in administrative databases.

Consultancy

DOS provides analytical consultancy services to Government agencies to maximise the LAD's value via:

1. **Collaborating with policy agencies**
   to translate their policy questions into empirical analytic problems that can be addressed.
2. **Providing data-related advice,**
   including guidance on the availability and suitability of the data items for their projects, highlighting data caveats that may affect the interpretation of findings, scoping the coverage of the project dataset, and defining appropriate criteria for treatment and control groups for policy evaluation.
3. **Offering statistical and methodological advice**
   to ensure that studies are based on robust methodologies.
4. **Training Government officers**
   in data analytics, econometrics, machine learning, and statistical programming.

Ensuring Confidentiality and Security of the Data

To ensure the integrity and security of the LAD and the confidentiality of the entities within, DOS manages the data using a robust **data governance framework** comprising several components, including legislation, data management policies and process safeguards, and Information Technology (IT) systems.

The LAD is governed by the Statistics Act 1973, one of the key legislation governing statistical activities in the public sector. It mandates data collection and safeguards the confidentiality of information provided. For example, key data items such as income are collected under the Act to study income growth and mobility trends in Singapore. Under the legislation, the LAD can only be used for statistical and research purposes within the Government and cannot be used for administering policies directly.

Data management policies are implemented based on the relevant Government Instructional Manuals, and a high-level of process safeguards are imposed. The LAD is accessible only by a selected team of DOS officers to manage the data and conduct research projects. Identifiable data are not disclosed or shared outside of DOS, even within the Government. Only aggregated data that has passed statistical disclosure controls, or anonymised data, can be accessed by Government officers [4]. DOS's practices are relatively more stringent than those of other NSOs, reflecting a prioritisation of privacy of data records over accessibility.

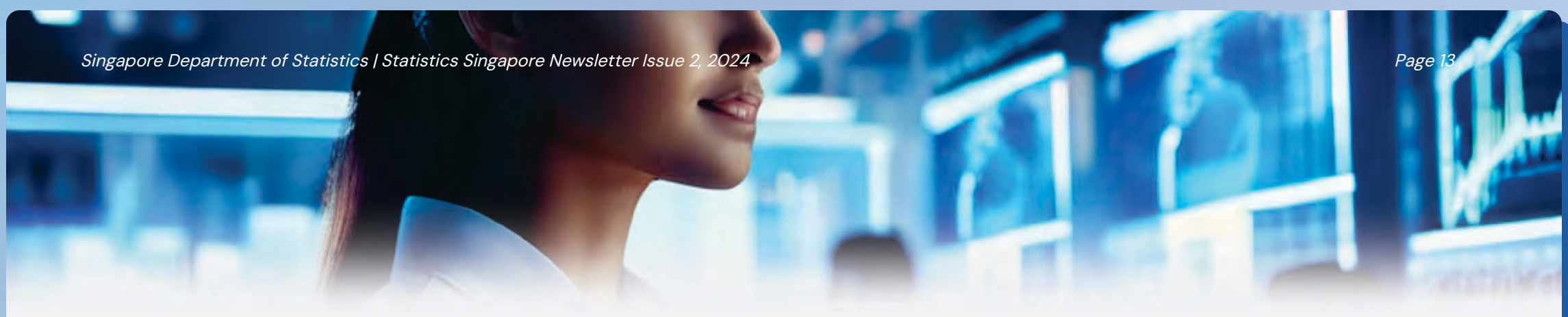
Robust IT security measures are in place to protect the data in storage and in use. A full suite of data loss protection and isolation measures is implemented to prevent exfiltration of data from DOS's systems. Access management systems control identity and access rights ensuring that data can only be accessed by authorised persons; all access and usage of data is consistently managed, logged, and monitored. Data are encrypted at all times, and data lineage records track the usage of sensitive data, the flows between users and systems and the changes made to the data.

Overall, DOS's robust data management processes adhere to, and are more robust than, international standards. This approach balances maximising the use and value of the LAD while maintaining data integrity, security, and confidentiality.

Conclusion

In conclusion, the LAD is a rich, integrated longitudinal dataset, created to advance robust evidence-based policymaking in the Government. DOS ensures strong measures are implemented to safeguard data integrity, security, and confidentiality. The LAD has been used in many research studies across multiple domains, with findings supporting government agencies to formulate or refine their policies to benefit the people of Singapore. DOS remains committed to maintain the LAD to the highest standards, providing reliable and comprehensive data to support policy studies.

[4] Government officers with access to aggregated data may include academics commissioned by Government agencies to provide expertise to the research team.



Use of AI-Enhanced OCR and Machine Learning for Data Processing in the Household Expenditure Survey 2023

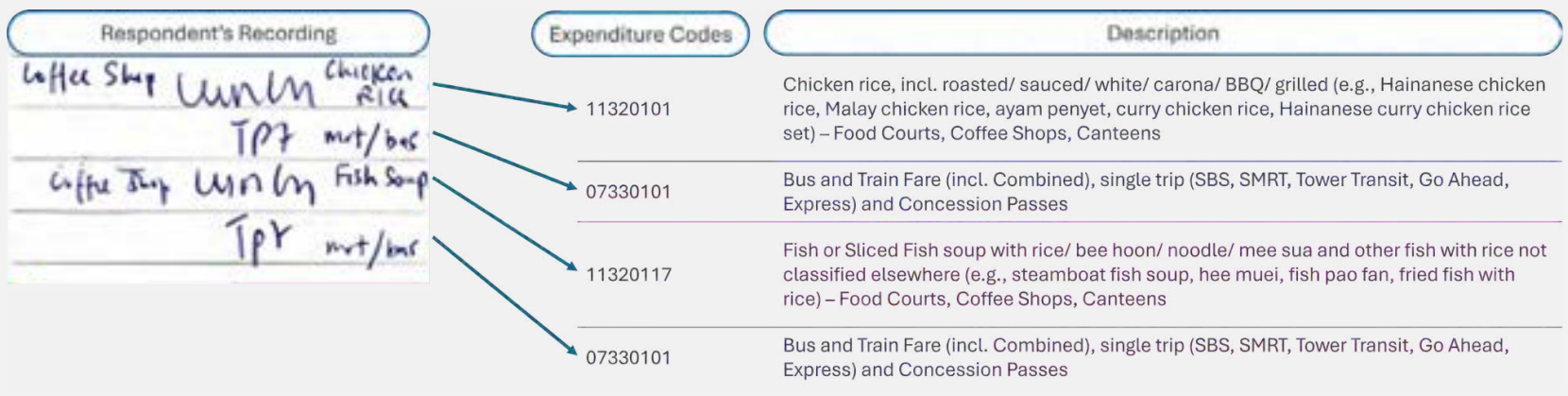
by Boon Kok Ann and Cheng Wan Hsien
Household Surveys and Expenditure Division
Singapore Department of Statistics

Introduction

The Singapore Department of Statistics (DOS) conducts the Household Expenditure Survey (HES) every five years, since 1972/ 73, to collect detailed information on households' expenditure, socio-economic characteristics and ownership of consumer durables. It is carried out over a one-year period to cover different festive and seasonal expenditure of households. Expenditure data collected include day-to-day expenses such as food, groceries, and transport; regular expenditure such as utilities and telecommunication subscription services; and ad-hoc big-ticket expenditure like the purchase of cars and household durables.

Data processing for the HES has traditionally been labour-intensive and time-consuming, requiring extensive manual checks, data entry, and coding. While respondents can submit their returns electronically since the 2017/ 18 survey, many prefer to submit handwritten returns in hardcopy booklets [1] or provide receipts of their regular and day-to-day expenses. In past surveys, data processing clerks manually entered the amount for each expenditure item into the system, then assigned an expenditure code [2] to each expenditure item (Figure 1).

Figure 1: Assigning Expenditure Codes to Expenditure Items



System Redesign and Automation Initiatives

In the HES 2023, the data processing workflow was redesigned, by leveraging AI-enhanced optical character recognition (OCR) and ML modelling techniques, to reduce time spent on manual data entry and coding. For non-electronic returns, hardcopy booklets and receipts were scanned, and the images were processed by the OCR software to extract textual data. This information includes descriptions of expenditure items, dollar amount, payment indicators and date of recording. After verifying and amending any inaccuracies in the extracted information, the data was automatically sent to the data processing system in a machine-readable form.

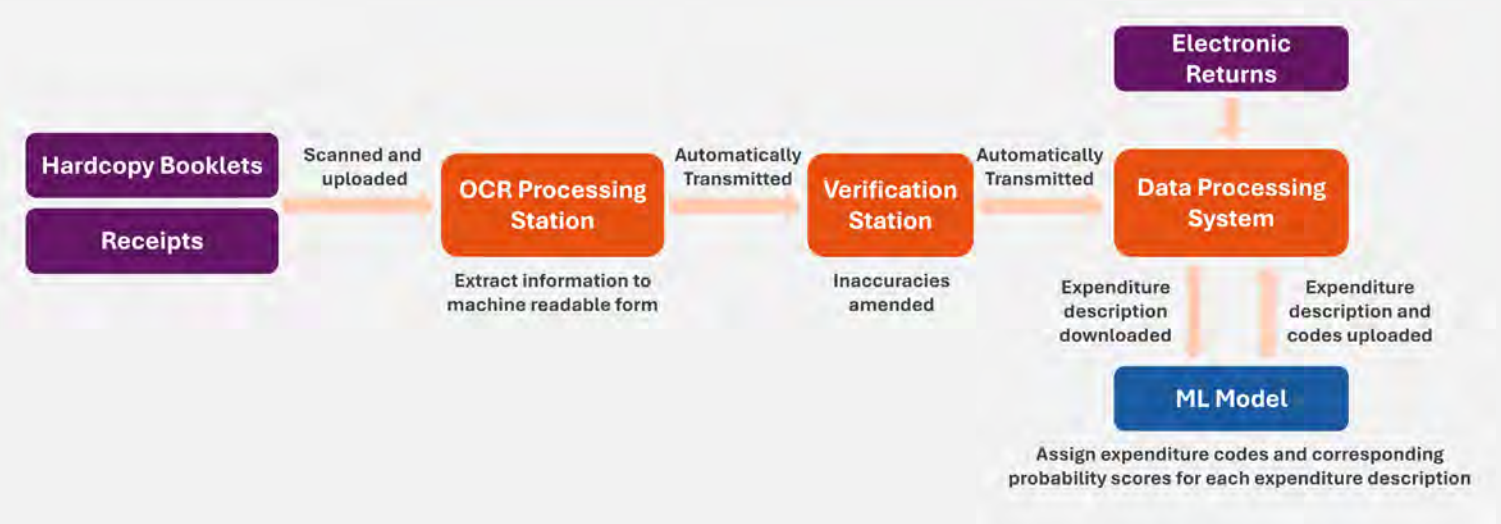
Together with electronic returns, expenditure descriptions were downloaded weekly from the data processing system and passed through an ML model. The model assigns expenditure codes and corresponding probability scores [3] to each expenditure description. If the probability score is above a predefined threshold, the expenditure description and code will be uploaded back into the data processing system (Figure 2). Those with a score below the threshold will not be automatically assigned a code and will require data processing clerks to manually input a code.

[1] In HES 2023, about 85% of respondents submitted some form of handwritten return.

[2] Expenditure codes are based on Singapore Standard Classification of Individual Consumption According to Purpose (S-COICOP).

[3] The probability score refers to the ML model's prediction of the likelihood that a particular expenditure code is the correct code for the description. For e.g., 'bus & train trip' could have a probability score of 90% to be coded as '07330101 - Bus and Train Fare (incl. Combined)' and a score of 8% to be coded '07320101- Bus/Coach fares'.

Figure 2: Redesigned Process for Assigning Expenditure Codes in the HES 2023



Replacing Data Entry with AI-Enhanced OCR

1 ▼ **Recognising Handwritten Returns**

The AI-enhanced OCR software was pre-trained to identify on various handwriting styles, to better interpret respondents' handwritten returns.

2 ▼ **Recognising Fields in Receipts**

Given the different layouts of receipts from various establishments, AI was used to identify fields containing the amount, description, establishment name, payment mode, and other relevant information. The AI model was pretrained on sample receipts and keywords to improve the recognition prior to the data collection for the HES 2023. For example, words such as 'pte ltd' is associated with establishment names, while 'VISA' and 'MASTER' are associated with payment modes.

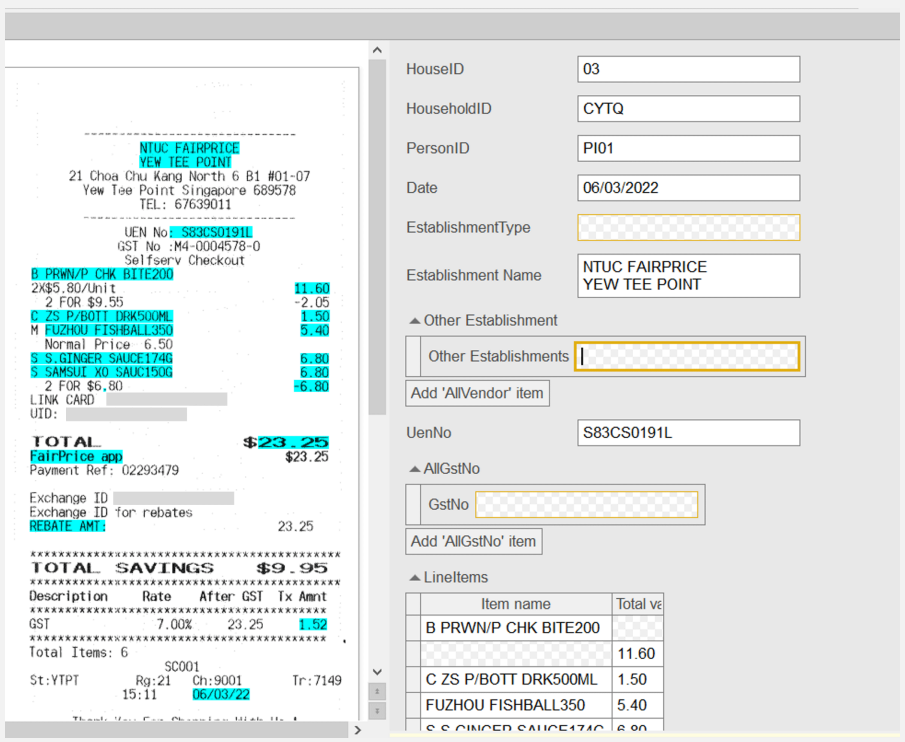
90%

Improved OCR Accuracy for printed text

For faded receipts and poor handwriting, the accuracy rate was lower.

To reduce the risk of inclusion of such inaccurate inputs, data processing clerks used a verification software to compare the scanned image with the extracted information and correct any errors (Figure 3).

Figure 3: Screenshot of a Scanned Receipt in a Verification Software



In the previous HES, the actual expenditure descriptions were not entered into the data processing system due to high resource demands of accurately capturing them. With the use of AI-enhanced OCR, unstructured textual data from booklets and receipts could be captured efficiently, allowing actual expenditure descriptions of expenditure items to be captured in the HES system for HES 2023.

The record of actual expenditure descriptions was very useful for data processing, as the expenditure items may need to be revisited when checking for consistency and accuracy of the data collected from respondents. Manhours were saved as the new process facilitated the review of summarised textual data which contained the expenditure descriptions and codes for the whole household. Whereas in the previous HES, checking involved navigating to stored images of booklets and receipts to view expenditure descriptions, which was time-consuming.

Automating Expenditure Coding with Machine Learning

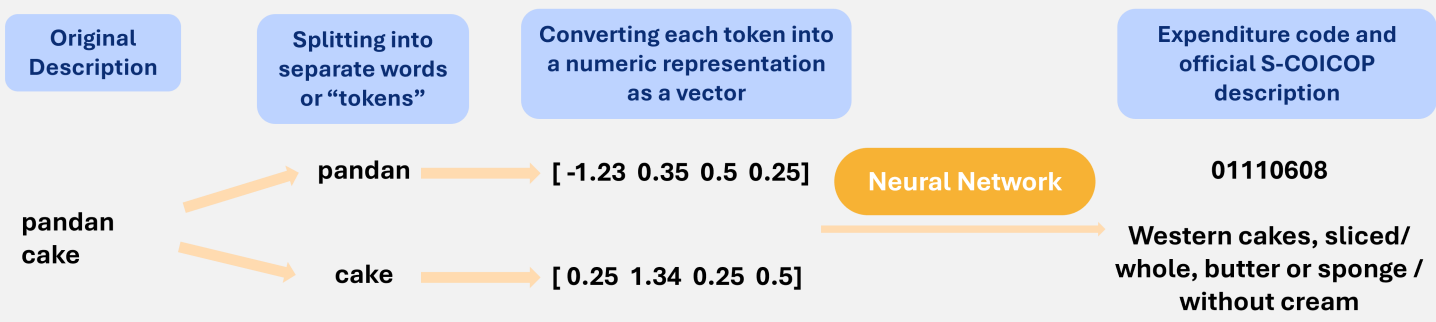
With the expenditure descriptions extracted via OCR, it became possible to apply ML models to automate the assignment of expenditure code to each expenditure item.

To improve the accuracy of the ML models, DOS developed in-house Python scripts to pre-process textual data into a standardised form that can be meaningfully tokenised. Pre-processing involves removing punctuation marks, special characters, unnecessary whitespaces, sizes and weights (e.g., Kg, XXL), and stop words; correcting typographical errors; converting acronyms (e.g., CS = coffee shop, FC = food court); and standardising all characters to lowercase.

After the textual data has been pre-processed, the Recurrent Neural Network (RNN) model is used to assign expenditure codes to each expenditure item. Different methodologies were explored, such as using natural language processing and cosine similarity to perform the coding, trained on data from past HES and the expenditure code dictionary. The RNN model was evaluated to be the best-performing model in terms of accuracy at the most detailed expenditure code level, and was designed for interpreting sequential or temporal information (e.g., text, time series, audio), compared to other neural network models such as Convolutional.

Figure 4 illustrates how the neural network takes in the vector representations of the words as inputs which effectively capture all the information in the sequence of words. The neural network then generates a probability score indicating the likelihood of the expenditure code being the correct code for the description.

Figure 4: Assignment of Expenditure Item to Expenditure Code with Neural Network



Expenditure Code Assignment Process Based on Probability Scores Generated

Automatic Code Assignment

When the probability score of the top-predicted code is very high, the entry will be automatically assigned an expenditure code. The threshold for the score was set to obtain an accuracy rate of 90% for the assigned code.

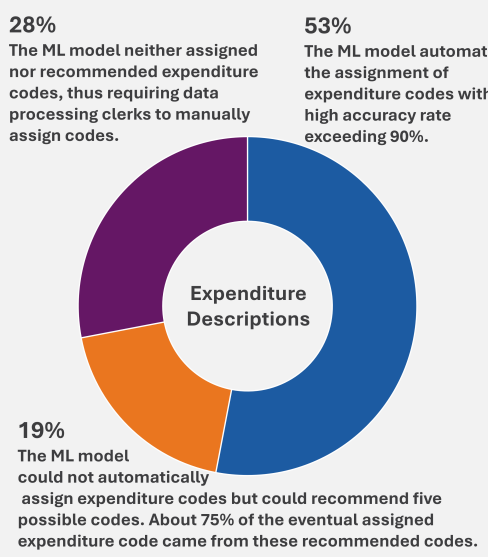
Code Recommendation

When the probability scores of the predicted codes is relatively high, the model will recommend five possible codes with the probability scores. Data processing clerks can either select one of the recommendations or choose a code outside the recommended. The threshold for the score was set such that for the majority of the records, the correct code is within the five recommended codes.

Manual Coding

When the score is low, data processing clerks will manually assign a code.

Effectiveness of ML Model and Conclusion



The RNN model was utilised from the start of the data processing operations and refined with subsequent batches of data. This refinement aimed to capture any nuances and characteristics unique to the HES 2023 data that were not present in the previous HES 2017/ 18. The ML model significantly reduced manual coding.

An additional benefit of for automated coding was the consistency of code assignment. In the previous HES, the accuracy of manual code assignment was largely dependent on the understanding and interpretation of the data processing clerks. In the HES 2023, the ML model ensured that all records with the same description were assigned the same code, making it easier to identify and rectify any incorrect codes.

The use of AI-enhanced OCR and ML techniques in the HES 2023 enabled DOS to automate processes that traditionally required considerable manual effort. As result, DOS achieved significant savings in time and resources. This accomplishment provides further impetus for DOS to explore the use of these tools in other projects.

Singapore's Trade in Services by Industry:

Examining Shifts in Services Trade Patterns in Recent Years

Introduction

International trade in services play a significant role in Singapore's economy, amounting to a total value of about 127% of Singapore's Gross Domestic Product (GDP) at current prices in 2022, up from 84% in 2010. The expansion of trade in services over the years was spurred by globalisation and technological advances which facilitated easier access to services abroad.

The Singapore Department of Statistics (DOS) has been publishing Singapore's trade in services statistics with further breakdown by services categories, export markets, and import sources. In 2024, trade in services by industry breakdown was released, offering insights into the industries that contribute to services trade in Singapore and the type of services traded by the respective industries. This article highlights trends in Singapore's trade in services by examining Singapore's exports and imports of services by industry from 2017 to 2022.

Scope and Coverage

The International Trade in Services survey, conducted annually by DOS, is the main source of data for Singapore's **trade in services** statistics. The survey covers services transactions between firms domiciled in Singapore and overseas trading partners. However, data for certain services categories such as Travel services and Government Goods and Services are compiled via administrative sources instead. Since administrative data sources lacks information on industry breakdown, these services categories are excluded from the estimates on trade in services by industry.

Trade in services by industry statistics are presented based on eight main industries, with an 'Others' category encompassing the remaining industries. These main industries contribute to the majority of services trade in Singapore, whereas industries in the 'Others' category includes either small industries or those driven by domestic consumption.

Singapore Standard Industrial Classification (SSIC) [1] of Industries

Industry	SSIC 2020
Manufacturing	10 to 32
Construction	41 to 43
Wholesale Trade	46
Transport & Storage	49 to 53
Information & Communications	58 to 63
Financial & Insurance	64 to 66
Professional Services	69 to 75
Administrative & Support Services	77 to 82
Others	All remaining SSICs

1] The **SSIC** is the national standard for classifying economic activities undertaken by economic units.

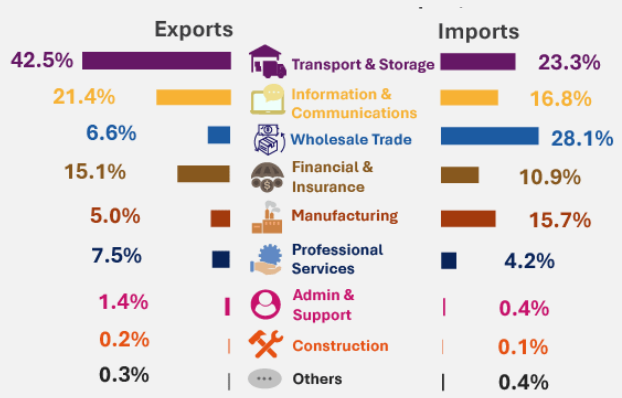
Findings

Services exports by the Transport & Storage Industry was the fastest growing industry in Singapore from 2017 to 2022, closely followed by the Information & Communications Industry.

From 2017 to 2022, Singapore's Transport & Storage industry was the top contributor to Singapore's services trade every year, recording a compound annual growth rate (CAGR) of 19.8%. This industry accounted for 42.5% ($184 billion) of services exports and 23.3% ($85.6 billion) of services imports in 2022 (Chart 1). Within the Transport & Storage industry, majority of its services exports was from the transport services [2], contributing 99% of the industry's total services exports in 2022.

Arising from the impact of the COVID-19 pandemic [3] and the accelerated pace of digitalisation in recent years, the Information & Communications industry had surpassed other industries to become the second largest contributor to Singapore's overall trade in services in 2022, recording a CAGR of 18.9% from 2017 to 2022.

Chart 1: Share to Total Services Exports/ Imports, 2022



With various measures such as remote working and social distancing introduced in response to the COVID-19 pandemic from 2020, the need for new digital tools to support these new arrangements rose sharply. Online conferencing for collaborative work, cloud storage, data security, and remote work solutions, together with digital entertainment such as streaming, proliferated. While these platforms and digitalisation trends were already present in the economy, the pandemic further compounded the demand for new digital products. This surge in demand led to a rapid growth in services exports by the Information & Communications industry from 2020 onwards (Chart 2A). Similarly, the Information & Communications industry experienced growth in services imports, overtaking the Manufacturing industry from 2021 as the third largest services importer (Chart 2B).

Chart 2A: Services Exports of Selected Industries, 2017-2022

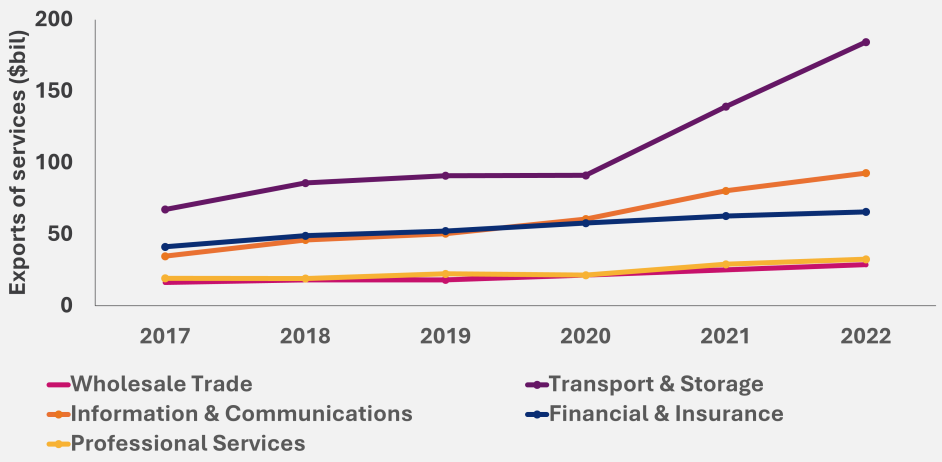
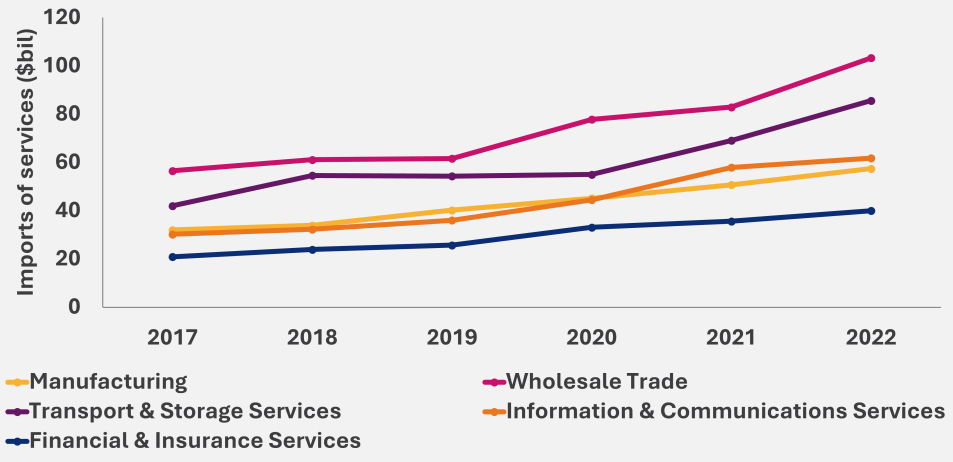


Chart 2B: Services Imports of Selected Industries, 2017-2022



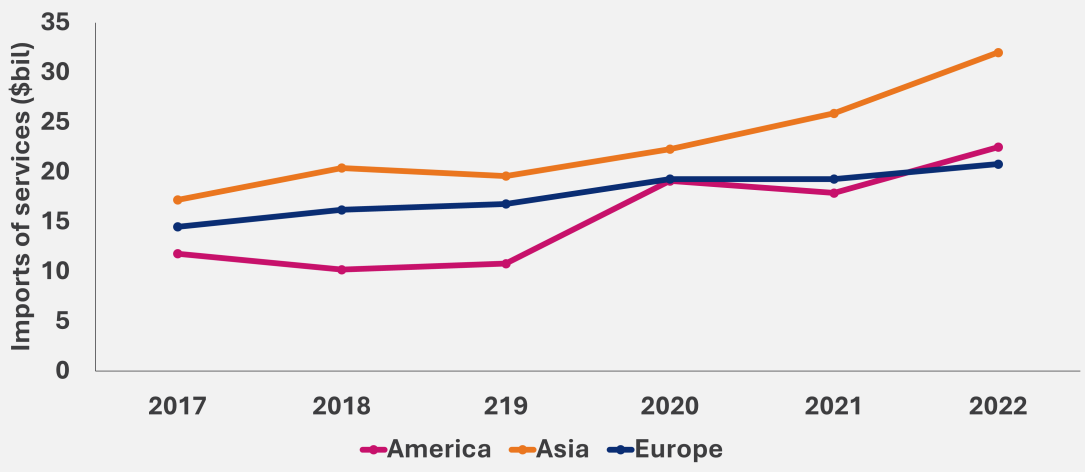
America Overtakes Europe as the Second Largest Source of Services Imported by the Wholesale Trade Industry

The Wholesale Trade industry was consistently the largest importer of services from 2017 to 2022. In 2022, it accounted for 28.1% of Singapore's services imports with transport services contributing 56.6% ($58.5 billion) to the industry's services imports. Trade-related services [4] was the second largest imported service at 12.1% ($12.5 billion).

The Asia region was the largest source of services imports for the Wholesale Trade industry from 2017 to 2022, due to its geographical proximity to Singapore (Chart 3).

Until 2021, Europe was the next largest source of services imports, closely followed by the America region. However in 2022, the Wholesale Trade industry imported more services from Asia and America which may be due to economic uncertainties in Europe stemming from the Russia-Ukraine war.

Chart 3: Services Imports of Wholesale Trade Industry from Selected Regions, 2017-2022



[2] Transport services include shipping of goods between countries and transport of passengers between countries.

[3] The World Health Organisation declared the COVID-19 outbreak a pandemic on 11 March 2020.

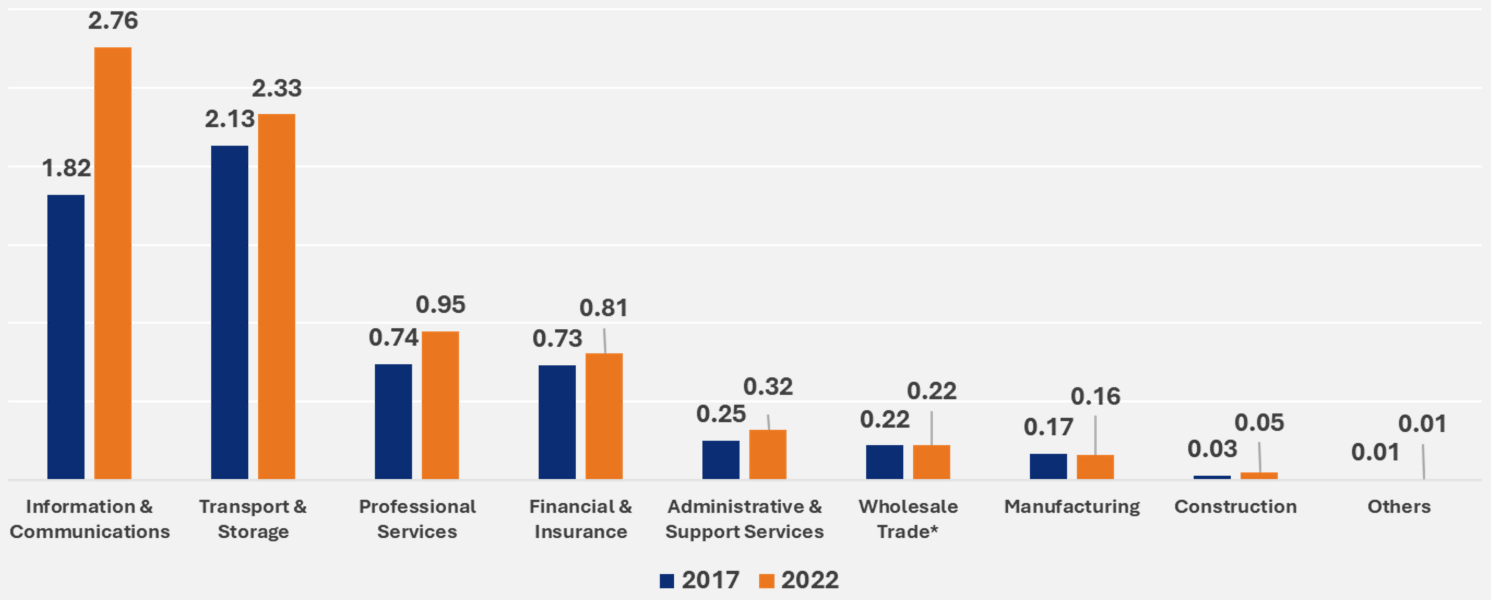
[4] Trade-related services consist of commissions, agency fees and distributor fees.

Export Propensity and Market Diversification: A Tale of Two Industries

By combining trade in services by industry statistics with other statistics such as value added by industry, additional insights can be derived. The services export propensity ratio (services exports by industry divided by **value added by industry**) measures the degree to which industries depend on external demand for services. The Information & Communications and Transport & Storage industries had export propensity ratios of 2.8 and 2.3 respectively in 2022, which suggested that these industries were more dependent on external demand. In contrast, other industries such as Administrative & Support Services had an export propensity ratio of 0.3 (Chart 4A).

A high export propensity ratio may indicate an industry's susceptibility to global economic trends and trade dynamics. For a more holistic picture, export propensity ratios should be analysed in conjunction with other indicators such as market concentration.

Chart 4A: Services Export Propensity by Industry, 2017 and 2022

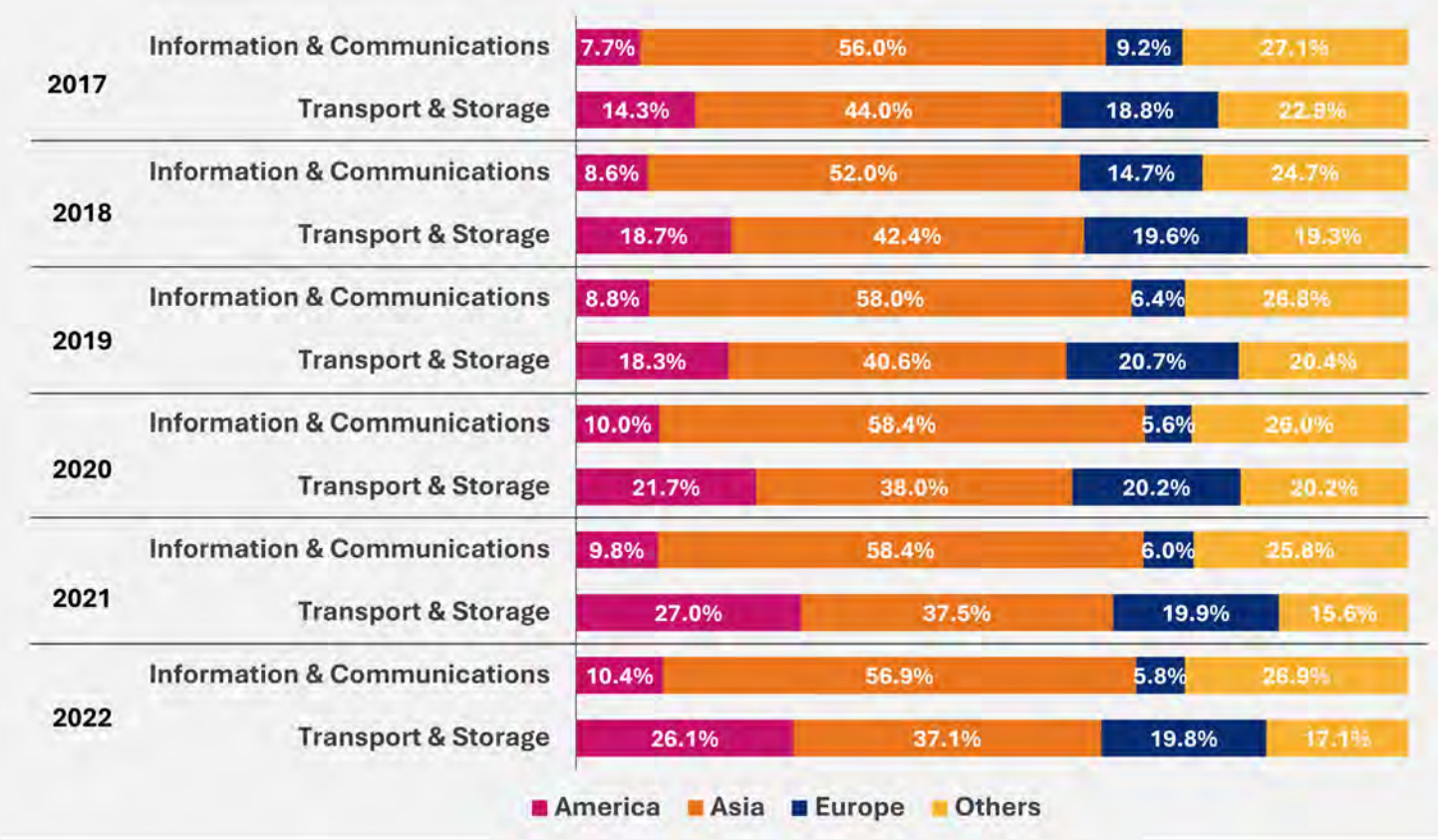


** The services exports propensity for the Wholesale Trade industry is low because the industry is primarily involved in the buying and selling of goods, and not so much in services. As such, the industry would have large exports of goods relative to its services exports.*

The Transport & Storage industry's services exports to the America region increased from 14.3% in 2017 to 26.1% in 2022 while exports to Asia steadily decreased from 44.0% in 2017 to 37.1% in 2022, showing diversification across the export markets (Chart 4B). The Information & Communications industry had high export propensity ratios from 2017 to 2022. On average, 56.6% of the Information & Communications industry's exports were concentrated in the Asia region during the period (Chart 4B).

Analysed together, this suggests that the Transport & Storage industry would likely be more resilient to adverse economic events compared to the Information & Communications industry, due to its more diversified exports.

Chart 4B: Concentration of Services Exports by Region for Selected Industries, 2017-2022



Conclusion

Singapore's trade in services has advanced substantially in recent years, driven by factors such as the services-oriented economy of Singapore, increasing digitalisation of services, and general post-pandemic recovery. Global events have caused disruption to the world economy and led to shifts in industry-specific trends. The newly published statistics on international trade in services by industry provide an additional lens to understand these changes in Singapore's services trade landscape.

Anonymised Microdata Access Programme

As the National Statistical Office (NSO) of Singapore, the Singapore Department of Statistics (DOS) collects data under the Statistics Act [1] for statistical purposes. DOS then compiles the data and publishes aggregated statistics as reports and publications on the SingStat Website. These statistics are also available on the SingStat Table Builder and SingStat Mobile App. To enable greater data-sharing with academic researchers, the Statistics Act was amended in April 2021 to expand anonymised microdata access to researchers who are holding full-time appointments in Singapore Autonomous Universities (AUs). Eligible researchers are now able to participate in DOS's Anonymised Microdata Access Programme (AMAP), which offers researchers access to anonymised microdata compiled under the Statistics Act and conduct complex analyses to address research questions in DOS Innovation Data Lab.

Importance of Anonymised Microdata

Microdata refers to data about persons, households, or establishments at an individual level. It is used for statistical aggregation (e.g., categorising records) and provides the greatest flexibility for statistical analyses and modelling (e.g., transformation of variables, regression analyses, and machine learning). However, issues with data disclosure may arise when researchers are able to identify the person, household, or establishment with the microdata. Therefore, anonymisation of microdata is necessary to reduce the risk of re-identification while minimising the loss of statistical information. Anonymisation techniques such as global recoding and microaggregation are implemented to produce a safe dataset where the identity of a person, household, or establishment cannot be readily discovered or ascertained.

Five Safes Framework

Similar to overseas NSOs such as Statistics Canada, Australian Bureau of Statistics and United Kingdom's Office for National Statistics, DOS's AMAP provides academic researchers with access to anonymised microdata for research purposes. To ensure data confidentiality, DOS adopts the Five Safes framework:

Safe Project

All AMAP projects are reviewed and approved by the multi-agency panel that is chaired by DOS's Chief Statistician (CS).

Safe People

Only approved researchers are allowed to embark on AMAP projects.

Safe Data

Microdata are anonymised to prevent the re-identification of a person, household or establishment.

Safe Settings

Data access is only allowed in DOS Innovation Data Lab.

Safe Output

Only non-disclosive statistical results are allowed to be removed from the Data Lab.

[1] The **Statistics Act** is the primary piece of legislation that governs statistical activities conducted by the Singapore Department of Statistics and gazetted Research and Statistics Units (RSUs) in the public sector. It safeguards the confidentiality of information collected from individuals and companies and spells out the legislative authority and responsibilities of the Chief Statistician and directors of gazetted RSUs.

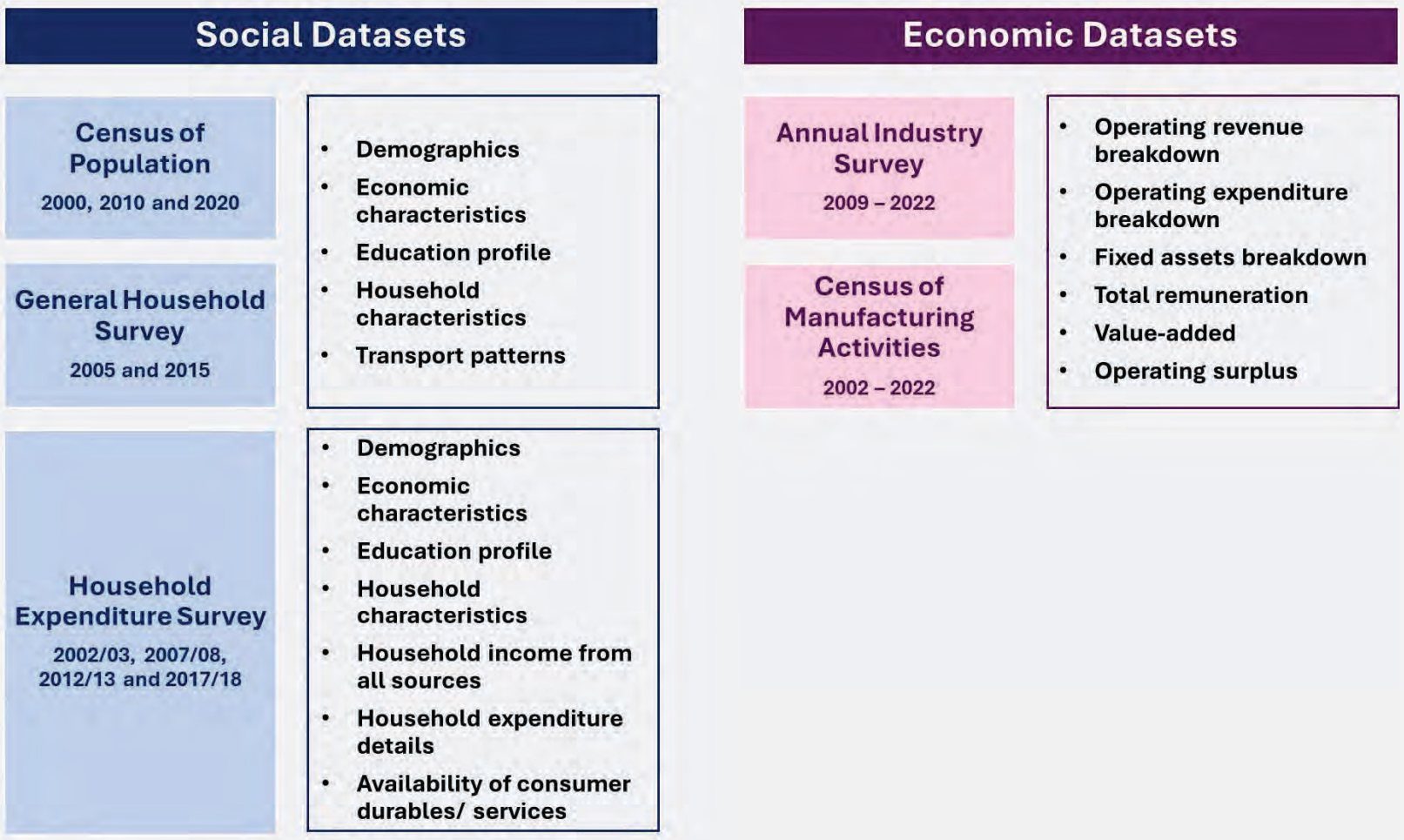
Comprehensive and Rich Microdata Sets on Firms and Individuals Available for Access via AMAP

Anonymised microdata sets available for access via AMAP include:

- Census of Manufacturing Activities (2002 – 2022)
- Annual Industry Survey (2009 – 2022)
- Census of Population (2000, 2010, 2020)
- General Household Survey (2005, 2015)
- Household Expenditure Survey (2002/ 03, 2007/ 08, 2012/ 13, 2017/ 18)

The comprehensive coverage and richness of these datasets on firms and individuals offer a valuable resource for in-depth analysis such as exploratory data analysis (e.g., aggregated statistics tables and distribution charts), econometric models (e.g., linear regressions including those with multiple fixed effects), and machine learning models (e.g., clustering analysis).

Broad information on the anonymised microdata sets and the respective data dictionaries are accessible via the **AMAP webpage** on the SingStat Website.



AMAP Eligibility Criteria

To be eligible for AMAP, researchers must hold full-time appointments in local AUs, have the appropriate security clearance, and sign the legal undertaking form prior to data access.

AMAP Application Cycles

AMAP applications run on a quarterly cycle. Interested applicants may submit a research proposal within the stipulated deadline.

Deadlines for the submission of AMAP applications for the next six months are as follows:

9 November 2024 **25 January 2025** **12 April 2025**

AMAP Application Process

The AMAP application is a two-stage process. The steps for each stage are detailed below.

Stage 1: Application

- Applicants may submit the application form and legal undertaking, signed by the applicants and their Responsible Officer*, to DOS (datalab@singstat.gov.sg)
- DOS assesses the application together with relevant data source and policy agencies
- DOS seeks applicants' agreement to the estimated costs

Stage 2: Approval

- DOS submits recommendations to the Anonymised Microdata Review Panel (AMRP)^ for approval
- If approved, DOS will inform the applicants of the outcome
- Researchers may access anonymised microdata in DOS Innovation Data Lab

* The Responsible Officer must be an officer from the applicants' autonomous university who is the Divisional Director/ Vice Provost or of an equivalent appointment who oversees the research project.

^ The AMRP comprises of members from key data and policy agencies and is responsible for assessing and approving each AMAP application.

AMAP Publicity Efforts

AMAP for AU researchers was launched in 2023.

To promote the awareness of AMAP among AU researchers, DOS has disseminated a series of electronic direct mails. In 2024, DOS presented the AMAP at various platforms such as the 2024 Social Science & Humanities Research Thematic Grant (SSHR-TG) Engagement Session.

In addition, DOS will be conducting a series of Virtual Connect sessions to introduce AMAP to AU researchers. The first Virtual Connect session launched in April and May 2024 was attended by participants from NUS, NTU and SMU. DOS will continue to organise these sessions regularly and keep AU researchers informed through their universities.

Resources

The **AMAP webpage** on the SingStat Website provides useful resources on the AMAP application process, data dictionaries for the anonymised microdata sets available, and frequently asked questions.

Researchers may also reach out to DOS via **datalab@singstat.gov.sg** for any queries relating to AMAP.

Latest Results from the 2021 International Comparison Program

The International Comparison Program (ICP) is a global statistical initiative to collect comparative price and detailed expenditure data to estimate Purchasing Power Parities (PPP) which facilitate international comparisons of macroeconomic aggregates, relative price levels, and standards of living across economies.

Results from the 2021 ICP

PPP-based Gross Domestic Product [1]

PPP-based global output is the sum of gross domestic product (GDP) of all 176 economies that participated in the 2021 ICP cycle. In 2021, it amounted to US$152.4 trillion. The two largest economies were Mainland China and the United States, accounting for 19% and 15% of the world's GDP shares respectively. India, with a world share of 7%, was the third-largest economy, followed by the Russian Federation and then Japan (Table 1). Collectively, these five economies had a PPP-based GDP of US$74.7 trillion and accounted for nearly half of global output in 2021.

Singapore's PPP-based GDP accounted for 0.5% of the global output in 2021. In terms of PPP-based GDP per capita, Singapore was ranked 2nd in the world, after Luxembourg (Table 2). In 2017, Singapore came in 3rd in the world, after Macao SAR, China, and Luxembourg.

Table 1: Relative Sizes of Economies, 2021

Economies	GDP at PPP		
	US$Bil	Share (World=100)	Rank
Mainland China	28,821.6	18.91	1
United States	23,594.0	15.48	2
India	10,963.1	7.19	3
Russian Federation	5,732.4	3.76	4
Japan	5,566.7	3.65	5

Table 2: Economies with the Highest PPP-adjusted GDP per Capita, 2021

Rank	Economies	2021		Economies	2017	
		US$	Index (World= 100)		US$	Index (World= 100)
	WORLD	20,271	100	WORLD	17,006	100
1	Luxembourg	137,948	681	Macao SAR, China	122,838	722
2	Singapore	131,864	651	Luxembourg	114,954	676
3	Ireland	114,451	565	Singapore	95,628	562
4	Qatar	106,491	525	Qatar	92,839	546
5	Bermuda	90,577	447	Bermuda	85,454	502

[1] Compared to the use of market exchange rates to convert GDP based on a reference currency, the use of PPPs adjusts for the differences in the purchasing powers of currencies by eliminating price level differences to better estimate the volume of output across economies in real term.

Relative Price Levels

The Price Level Index (PLI) is often used to assess the relative price levels across economies, reflecting how price levels of a common basket of goods and services compare against one another between economies. The index is computed as a ratio of PPP to market exchange rates. An economy with higher PLI implies that the overall price of a common basket of goods and services is relatively more expensive in that economy compared to a reference economy (where PLI = 100), and vice versa.

An Illustration of PPP and PLI Based on a Single Commodity

The Big Mac is a food item available in most economies and is generally similar in terms of quality and specifications across the globe.

Suppose a Big Mac costs S$5.90 in Singapore and US$5.65 in the United States (US). The PPP for Big Mac (S$ per US$) is computed by taking the ratio of its price in Singapore to its price in the US (i.e., 5.90/5.65 =1.04). The resulting value of 1.04 implies that S$1.04 has the same purchasing power as US$1 for 1 unit of Big Mac.

Suppose the prevailing exchange rate is S$1.37 to US$1. A tourist from the US visiting Singapore exchanges US$5.65 (price of a Big Mac in the US) to S$7.74 at a bank upon arrival (based on the prevailing exchange rate). After buying a Big Mac in Singapore, he will have S$1.84 remaining.

Singapore: S$5.90
USA: US$5.65
PPP = 5.90/5.65 = 1.04S$ per US$

$$PLI = \frac{PPP_{SIN,USA}}{Exchange\ rate_{SIN,USA}} = \frac{1.04}{1.37} = 76$$

Computation of PPP and PLI

	US *	Singapore
Cost of a Big Mac in local currency	US$5.65	S$5.90
PPP	1.00	1.04
Market Exchange Rate	1.00	1.37
PLI for Big Mac	100	76

* Reference Economy

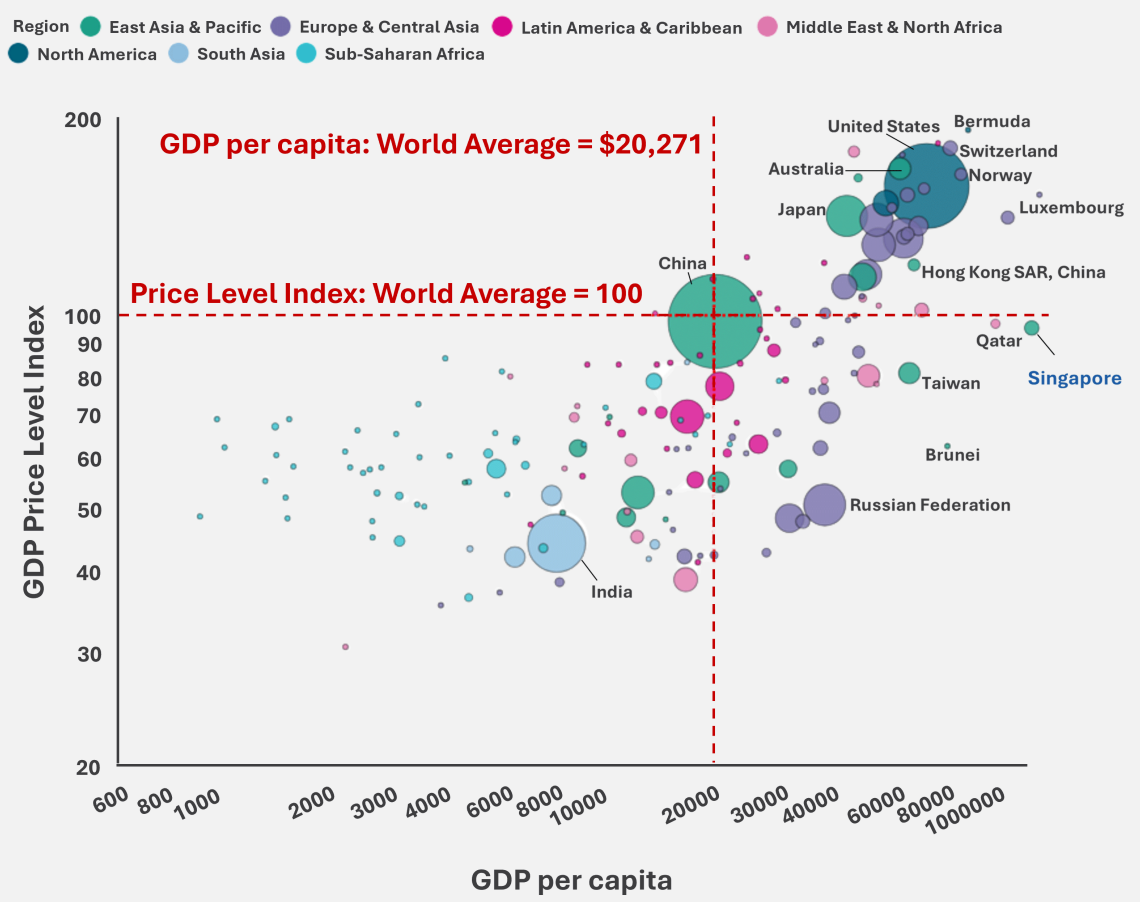
Thus, the PLI for Big Mac in Singapore is 1.04/1.37 * 100 = 76 points. This means that the price of a Big Mac in Singapore is 76%, or three-quarters, of that in the US.

Typically, the PLI exhibits a positive relationship with income (i.e., PPP-based GDP per capita). For example, high-income economies (e.g., Luxembourg, Norway, and Switzerland) tend to have higher price levels.

Singapore's PLI at overall GDP level was 95.6 in 2021. Despite being one of the economies with the highest PPP-based GDP per capita, Singapore's PLI was about 4% lower than the world average of 100.

In the East Asia and Pacific region, Singapore's price level was ranked 8th, behind countries like Australia, Japan, Hong Kong, and Mainland China (Chart 1). This suggests that the general price level in Singapore was relatively lower compared to these countries in the region in 2021.

Chart 1: GDP PLI and PPP-Based GDP per Capita by Economy, 2021



Size of bubble indicates the relative size of PPP-based GDP for each economy. Source: 2021 ICP.

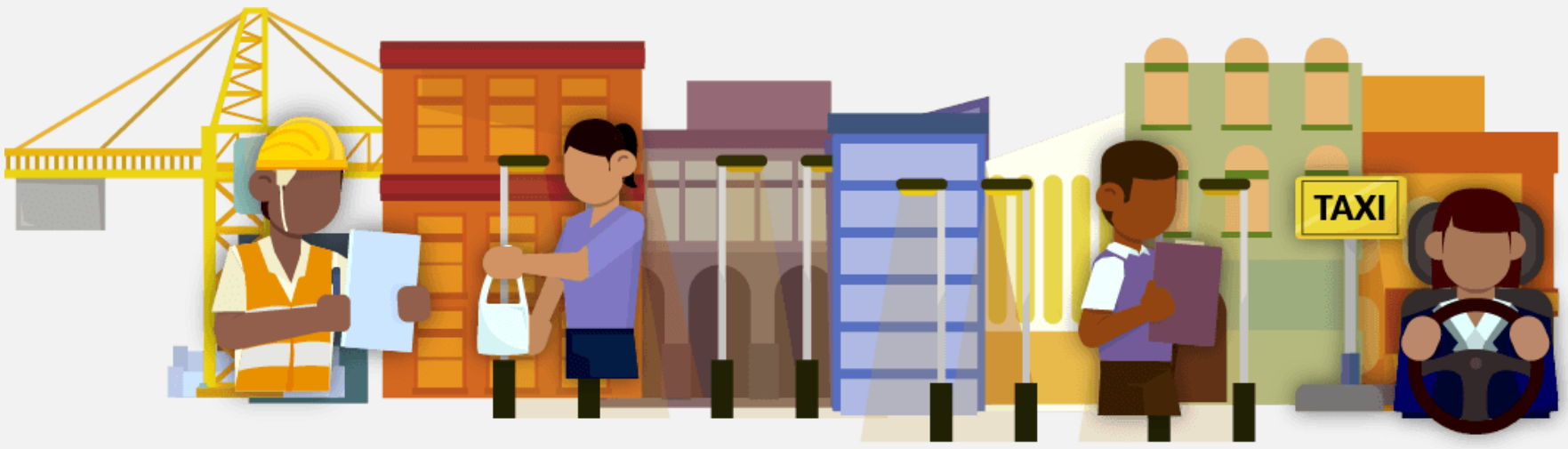
Using the Singapore Standard Occupational Classification (SSOC)

How to Determine an Appropriate Code?

SSOC is the acronym for the Singapore Standard Occupational Classification which is the national statistical standard for classifying occupations.

The SSOC adopts the basic framework and principles of the International Standard Classification of Occupations (ISCO) developed by the International Labour Office (ILO), and is regularly updated to reflect developments in the labour market.

The use of SSOC in the collection, analysis and dissemination of data facilitates data sharing and ensures data coherence (where data are consistent internally with as well as comparable with other data sources and over time). This is one of the six data quality dimensions specified in the **Statistical Best Practices 2020 handbook**.



How to Determine an Appropriate SSOC Code?

SSOC applies a five-digit coding system to reflect the five levels of classification – major group, sub-major group, minor group, unit group and occupation. At the most detailed level, an occupation is defined as a 'set of jobs whose main tasks and duties are characterised by a high degree
of similarity'.

> For instance, a sales assistant working in a fashion store and a salesperson working in a furniture shop are two different jobs but are grouped under the same occupation code SSOC 52202 Shop Sales Assistant as the main tasks and duties of selling goods and explaining the functions and qualities of goods and services to customers
> are similar.

As exemplified, the basic classification principle of the SSOC is to classify the occupations based on the **principal tasks and duties**. Persons who perform the same principal tasks and duties are considered to be doing the same type of work and are classified under the same occupation code regardless of their work experience, employment status, or qualifications.

There are certain occupations which may involve a variety and range of tasks and duties that could fall under the scope of various occupations. There may also be instances where the principal tasks are not possible to be ascertained. In such cases, the following rules shall apply, in the order of precedence given below:

In cases where the tasks and duties performed spans across different broad job levels, the job should be classified according to those tasks and duties related to the highest job level

Rule of Highest Broad Job Level: Broad job level is a function of the complexity and range of the tasks and duties to be performed in an occupation. It takes into account the formal education required to perform the tasks competently, the amount of informal on-the-job training and/ or experience in a related occupation, and the type and complexity of tasks and duties involved. There are four broad job levels with Broad Job Level 4 (comprising occupations with the most complex tasks) being the highest level. See the **SSOC 2024 report** to read more regarding the different broad job levels.

To illustrate, a farm technician reports that he provides technical support to conduct experiments to improve the yield and quality of crops (key task of SSOC 31421 Horticultural Technician) and also cultivates crops in the farm (key task of SSOC 61133 Gardener/ Horticultural Worker). Following this principle, he should be coded to SSOC 31421 instead of SSOC 61133 as the former is associated with a higher broad job level.

In cases where the tasks and duties are associated with different stages of the process of producing and distributing goods and services, the tasks and duties related to the production stage should take priority over those related to the sales, marketing and transportation of the same goods

Rule of Production Taking Precedence: To illustrate, consider a leather craftsman who makes bags and wallets and sells his products at a small pop-up store in a shopping mall. Following this principle, he should be coded as SSOC 75362 Leather Goods Maker/ Assembler rather than SSOC 52202 Shop Sales Assistant as the main task related to the production stage (i.e., making leather products) takes priority.

In cases where the tasks and duties are of the same broad job level and production stage, the appropriate code will be the tasks/ duties which take up the most amount of time

With digitalisation advancements, more occupations increasingly require technology skills to perform the tasks. Nevertheless, the occupation should still be classified based on the main tasks performed rather than the specific competencies involved. For instance, an equity analyst who performs complex financial modelling to facilitate investment recommendations should be coded as SSOC 24131 Financial Analyst instead of SSOC 21222 Data Scientist, despite the use of advance data modelling techniques.

With these classification principles in mind, there are two ways to determine an appropriate SSOC code using the available SSOC resources:

1

Refer to the SSOC Alphabetical Index or use the SSOC Search

to look for the job title by which the occupation is known, followed by selecting the most appropriate code by referring to the detailed definition.

2

Refer to the SSOC classification structure

to broadly classify the job before narrowing down to the most appropriate five-digit code at the occupational level. Confirm the choice of code by referring to the detailed definition.

Important Points to Note when Determining an SSOC Code

The primary objective of the SSOC...

...is to classify occupations of the civilian working population, but with provisions for the collective classification of those in the armed services (under X3000) and foreign diplomatic personnel (under X4000).

SSOC is not applicable to...

...persons outside the labour force such as housewives, full-time students, retired persons, and voluntary/ non-salaried social workers.

Classification Principle:

Persons who perform the same principal tasks and duties are to be classified under the same occupation, regardless of work experience, employment status, or qualifications.

Apprentices and Trainees...

...are classified according to the occupation they are training for, provided they are employed in the capacity of an apprentice or trainee.

SSOC 2024 Publication

To maintain the relevance of the SSOC, the classification is regularly revised and updated to reflect recent developments in the labour market and to align with changes in international standards. The Singapore Department of Statistics led a multi-agency working group for the revision of the SSOC 2020 and consulted various public agencies on the updates to the classification. These efforts culminated in the development and publication of the latest SSOC 2024. The deliberations and consultations ensure that the SSOC continues to meet the classification needs of users and producers of occupational data.

The latest SSOC 2024 publication and related materials are available on the **SingStat Website**. Some of the key changes in the SSOC 2024 include the creation of codes for emerging occupations in various sectors such as the Infocomm Technology and Green Economy. Some occupations across major groups and unit groups have been reclassified due to developments in their main tasks and duties over time. Detailed definitions for certain codes have also been enhanced to provide greater clarity in the descriptions of the main tasks and duties of these occupations.

Conclusion

The SSOC is used in the compilation, presentation and analysis of a wide range of statistics, including demographic, social and labour statistics. Ensuring that appropriate SSOC codes are chosen improves the quality and comparability of statistics which is essential for meaningful analyses and policy implementation.

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'.

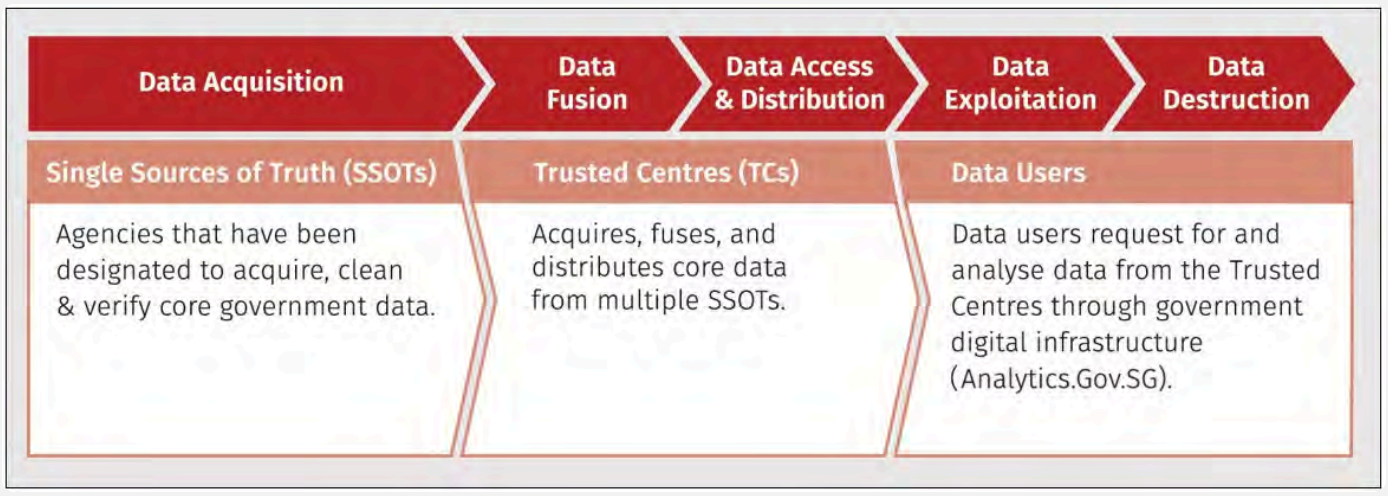
The **Public Sector (Governance) Act 2018** (PSGA) was enacted to provide a consistent governance framework across public agencies. This legislation enables public sector agencies to share data with each other on a needs basis for policy analyses and service delivery. At the same time, there are strong 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.

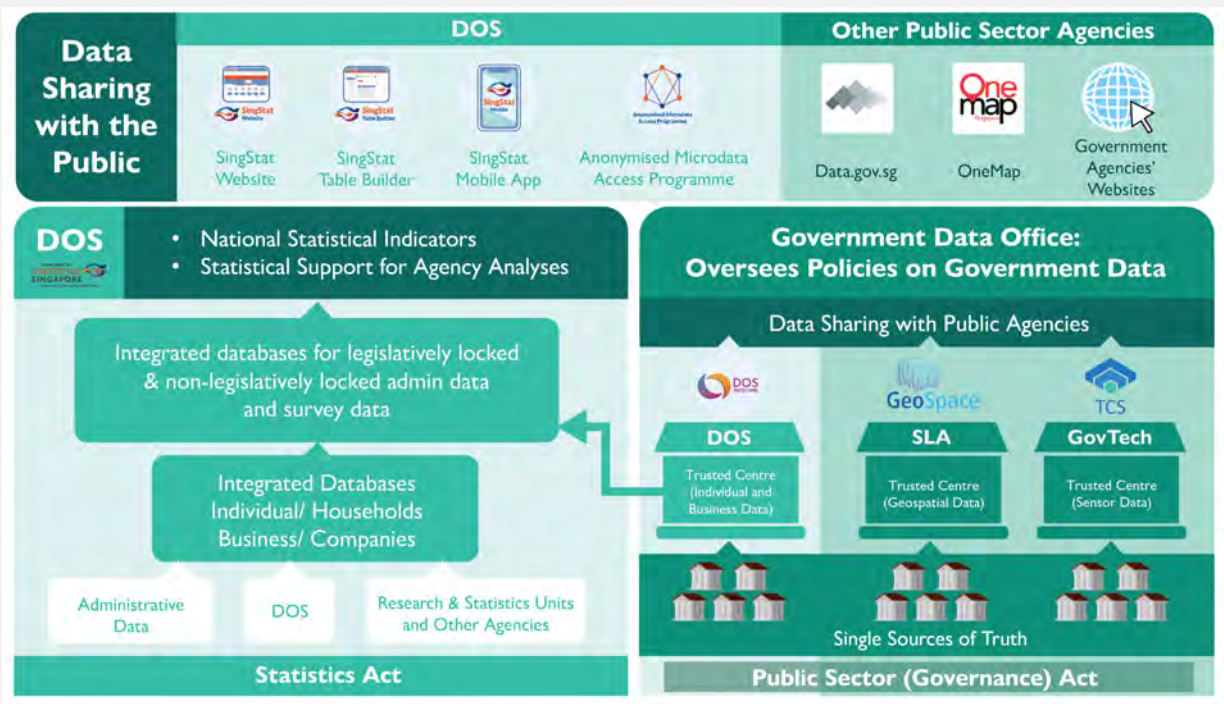
DOS is apt to be a TC for individual and business data for three reasons. Firstly, we are heavy users of data, hence playing the role of a TC allows us to obtain data quickly and directly from other administrative agencies. Secondly, as this role is a natural extension of our experience and capabilities, 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.

Fourth Meeting of the DOS Advisory Panel

The Singapore Department of Statistics (DOS) Advisory Panel was established in 2021 to guide DOS's strategic direction, amidst the changing data and technology landscape, and ensure that we remain relevant and responsive to the diverse needs of our data users. The fourth DOS Advisory Panel (DAP) meeting was held on 4 and 5 September 2024 in Singapore with the theme:

> **Technology Enablers for Digital Transformation of National Statistical Offices and Measuring Well-being Beyond GDP**

The DAP is chaired by the **Chief Statistician, Dr Koh Eng Chuan**, and comprises the following local and international members who are experts in the fields of statistics, data science and technology:

Mr Benny Chan

Managing Director,

Head of Group Channels and Digitalisation,

United Overseas Bank Limited

Mr Lee Bing Yi

Partner, Sustainability & Climate Change, Financial Services Assurance

PwC Singapore

Mr Carsten Ulrik Zangenberg

Director (Communication, Methodology, Analysis, Surveys and Consulting),

Statistics Denmark

Mr Osama Rahman

Director, Data Science Campus,

Office for National Statistics, United Kingdom

Prof Ong Yew Soon

President's Chair Professor of Computer Science, School of Computer Science and Engineering, Nanyang Technological University;

Chief Artificial Intelligence Scientist, A*STAR;

Co-Director, Singtel-NTU Cognitive & Artificial Intelligence Joint Lab

The fourth DAP meeting discussed the opportunities from integrating advanced technology applications and digital tools (e.g., generative AI and cloud computing) into National Statistical Offices' products and processes to meet the evolving demands for data and data services in an increasingly data-driven environment.

The DAP also exchanged views on the need for metrics beyond GDP to measure well-being and sustainability. Members discussed on methodologies and internationally comparable frameworks that can provide a more comprehensive understanding of well-being to better guide policy decisions and resource allocations.

DOS expresses our appreciation to all panel members for the insightful discussions and we look forward to the fifth meeting in 2025.

Instagram

Singapore Department of Statistics (DOS) on Instagram

We had fruitful discussions with our DOS Advisory Panel (DAP) members in early Sep! 🤝 Here's a quick look back! #SingStat #BeyondGDP #DigitalTransformation...

Overseas Visitors

From April to September 2024, the Singapore Department of Statistics (DOS) hosted study visits with New Zealand and Bhutan, and jointly conducted virtual sharing sessions with the ASEAN Secretariat's ASEAN Statistics Division.

Statistics New Zealand

Ms Rachael Milicich,
Deputy Government Statistician, Insights & Statistics

National Statistics Bureau (Royal Government of Bhutan)

Team led by **Mr Sonam Tenzin**, Director General

Sharing Sessions under the ASEAN-Help-ASEAN Framework

National Statistical Offices of ASEAN Member States

Expertise Sharing at International Fora

DOS shared our expertise at the following international fora from April to September 2024:

▼ Asia-Pacific Stats Café Series: Economic and Social Commission for Asia-Pacific (ESCAP) Committee on Statistics: Fit for the Future

> Featured DOS's Chief Statistician as a session chair

▼ Statistics Korea-ESCAP (KOSTAT-ESCAP) Joint Conference on Data Governance and Integration for Maximum Development Impact

> Featured DOS's Chief Statistician as a:
>
> - Speaker on Data Governance and Data Integration in DOS
> - Facilitator for a World Café session: Country perspectives on data integration

▼ Regional Workshop on Agile Statistical Systems Fit for the Future

> Featured DOS's Chief Technology Officer as a:
>
> - Panelist for The Role of National Statistical Offices in Improving and Integrating Administrative Databases
> - Facilitator for a World Café session: Tools for automating metadata-driven processes)

▼ Asia-Pacific Stats Café Series: Economic and Social Commission for Asia-Pacific (ESCAP) Committee on Statistics: Fit for the Future

> Featured DOS's Chief Statistician as a session chair

▼ 38th Meeting of Working Group on International Investment Statistics

> Presented on *Identifying Firms with Direct Investment Abroad Using Machine Learning*

▼ 7th International Conference on Establishment Statistics

> Presented on *Use of Administrative Data and New Data Sources in Industry Statistics*

▼ 5th International Seminar on Big Data for Official Statistics

> Presented on *Development of Singapore's Digital Economy Satellite Accounts*

▼ 22nd Meeting of the Working Group on Data Sharing, Analysis, and Dissemination

> Presented on *Data Stewardship and Governance in DOS*

▼ 32nd Population Census Conference

> Presented on *Singapore's Census of Population*

▼ 11th Meeting of the Working Group on Statistics of International Trade in Services

> Presented on *Measuring Digital Trade and Singapore's Balance of Payments Services Trade by Modes of Supply*

▼ 2nd Webinar of Data Strategy Sprint - Essential Features of a Successful Data Strategy

> Presented on *DOS's Experience on Implementing Data Strategy*

▼ Stats Café Session on Synthetic Data by ESCAP and United Nations Economic Commission for Europe

> Presented on *Use of Synthetic Data in DOS*

▼ 28th Meeting of the ACSS Sub-Committee on Planning and Coordination

> Presented and facilitated discussions for agenda on: *Concept Note for the ASEAN Statistics Virtual Academy*

▼ ABBYY AI Summit in Singapore

> Presented on *Use of AI-Enhanced Optical Character Recognition (OCR) and Machine Learning for Data Processing in the Household Expenditure Survey 2023*

▼ 7th Meeting of the UN Committee of Experts on Business and Trade Statistics

> Presented on *Producing New Insights on Entrepreneurship through Data Integration, Integrating Business and Trade Data to Profile Goods Exporting Firms, Update on the Task Group on Measuring E-commerce Value*

Hear from Our Officers on their International Statistical Involvement

Cui Huimin, Neo Soo Khee, and Peh Li Lin
Business Statistics Division
Members of the United Nations Committee of Experts on Business and Trade Statistics

Expanding Presence and Fostering Collaboration Internationally on Business Statistics

The United Nations Committee of Experts on Business and Trade Statistics (UNCEBTS) was established by the United Nations Statistical Commission (UNSC). Since its establishment, the UNCEBTS has worked on enhancing business and trade statistics to improve overall economic statistics and address emerging needs. Its efforts focus on coordination, development of methodology and concepts, capacity-building, compilation of data, and communication of business and trade statistics.

As of today, there are five Task Teams under the UNCEBTS, each comprising domain experts from various National Statistical Offices and international organisations, to collaborate on methodology development and capability building activities. DOS is a member of the following Task Teams and Working Group:

- Task Team on Statistical Business Register (TT-SBR)
- Task Team on Business Dynamics, Business Demography and Entrepreneurship (TT-BDBDE)
- Working Group on the Handbook of Integration Business and Trade Statistics (WG-HIBTS), under the Task Team of International Trade Statistics (TT-ITS)

In the Task Teams and Working Group, DOS actively shares its experience and knowledge in the international development of statistical methodologies. At the TT-SBR meeting, DOS shared on its experimental use of machine learning and new data sources to update the statistical business register. DOS also contributed to the draft HIBTS. In addition, we actively participate in the strategic planning of areas aligned with DOS's strategic priorities within the Task Team.

At the 7th UNCEBTS meeting in Sep 2024, DOS presented two papers. The first paper shares DOS's experience in profiling women-owned companies by integrating gender and business data. The presentation offered a concrete example to support TT-BDBDE's recommendation to integrate gender data with business statistics to generate new data insights.

The second paper presents DOS's experience in integrating business and trade data to profile goods-exporting firms. The presentation illustrates the methodology and benefits of data integration for new data insights as recommended by the HIBTS.

*Photo source: **UN Committee of Experts on Business and Trade Statistics***

DOS's involvement in the UNCEBTS aligns with our commitment to collaborate, engage, and contribute to capacity building activities in methodology development within the international statistical community. This involvement opens opportunities for DOS to learn about the latest research and development in business statistics and to collaborate with international experts on emerging topics to address new data demands.

DOS's Involvement in the Development of UNCEBTS Handbook on Integrating Business and Trade Statistics

The UNCEBTS has identified the need to develop a HIBTS, which was supported by the UNSC. The HIBTS serves as a reference for countries to produce integrated business and trade statistics that are internationally comparable. It aims to strengthen and build capacities for linking methods and to integrate these two domains.

Alongside other members of the WG-HIBTS, DOS helped draft the HIBTS chapters on the Statistical Business Register (SBR) and the Linking of Business and Trade Statistics. DOS also contributed a write-up on country practices, describing DOS's use of multiple data sources in updating the SBR.

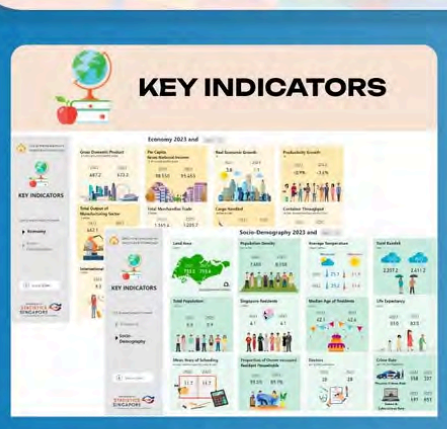
To augment the HIBTS, the Chief Editors of HIBTS are coordinating with the United Nations Statistics Division (UNSD) to create an online repository of national practices on integrating business and trade statistics for knowledge sharing and capability building. Similarly, DOS contributed a write-up sharing DOS's experience in integrating business and trade data through the SBR.

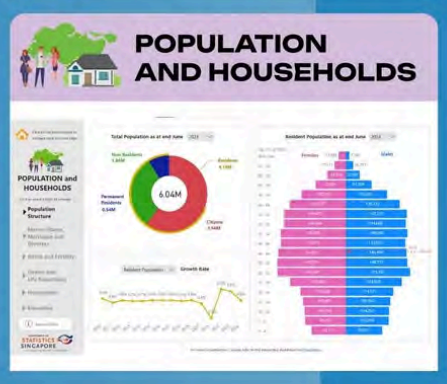
DOS presented a paper on '**Integrating Business and Trade Data to Profile Goods Exporting Firms**' at the 7th UNCEBTS meeting in Geneva, Switzerland from 24 to 26 September 2024. DOS's experience in integrating various business and trade datasets through the SBR to meet emerging data demand was shared. Leveraging the integrated data, the project team examined the enterprise characteristics of goods exporting firms and the impact of exports on firm's performance. The presentation was well received at the meeting as it provided a concrete example on the methodology and benefits of data integration for new data insights as recommended by the Handbook.

The UNCEBTS is conducting a global consultation on the HIBTS draft to seek comments. The HIBTS will be finalised and submitted to the UNCEBTS Bureau and UNSC for their endorsement in early 2025 [1].

[1] Refer to the **Progress report: Handbook on Integrating Business and Trade Statistics** for more details.

SINGAPORE IN FIGURES

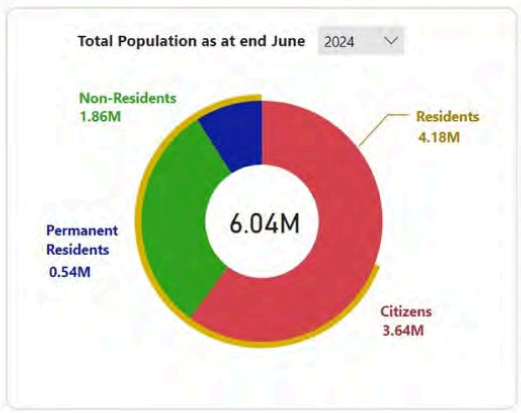


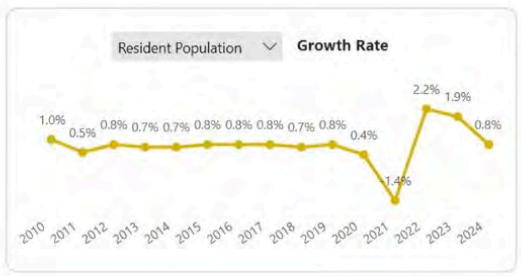


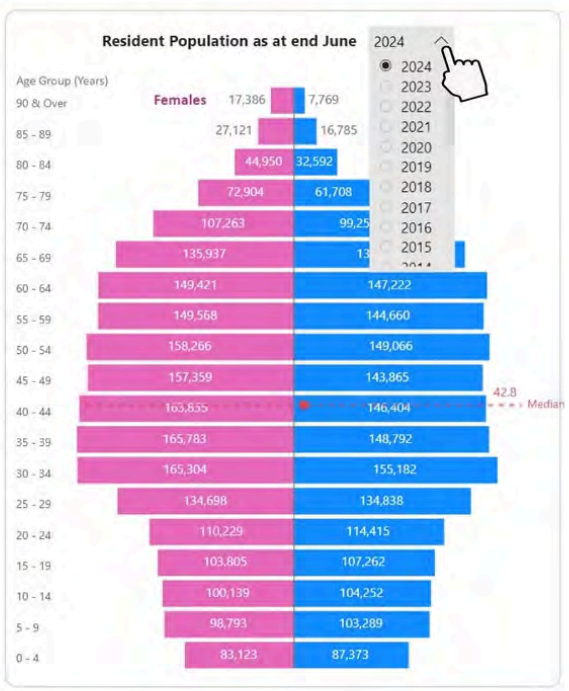
Explore the latest **economic and socio-demographic data trends** through interactive dashboards!

Interactive Features at a Glance

Select and view data by year

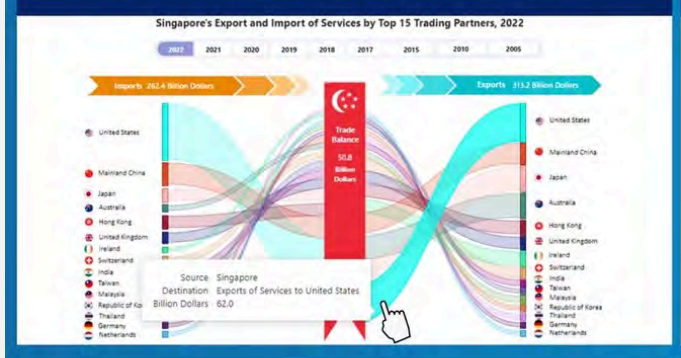






For more visualisations , please refer to the interactive dashboard on Population.

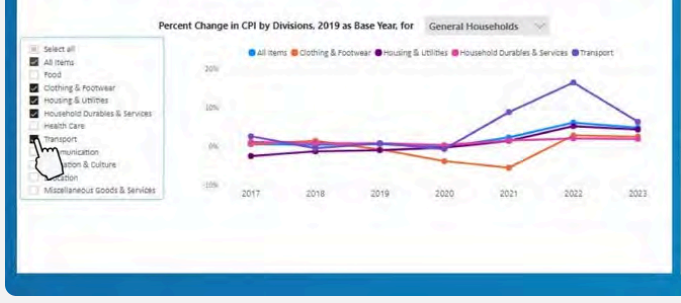
Hover over chart for information



Select multiple years to compare data

View data by year range

Select and compare multiple variables



Toggle between visualisations

Access relevant information

Section	Sources of Data	Download
Population Structure	Singapore Department of Statistics	
Marital Status, Marriages and Divorces	Singapore Department of Statistics	
Births and Fertility	Singapore Department of Statistics, Immigration and Checkpoints Authority	
Deaths and Life Expectancy	Singapore Department of Statistics, Immigration and Checkpoints Authority	
Households	Singapore Department of Statistics	
Education	Singapore Department of Statistics, Ministry of Education	

Explore the Singapore in Figures