# Use of Administrative Data in Statistical Compilation at the Singapore Department of Statistics

by **Neo Soo Khee**
Business Statistics Division
Singapore Department of Statistics

## Introduction

Administrative data refer to information collected by government agencies as part of their regulatory operations or record-keeping processes. Compared to surveys, the use of administrative data for statistical compilation helps to reduce respondent burden and lowers data collection costs.

In recent decades, the digitalisation of systems and advances in technology have provided the Singapore Department of Statistics (DOS) with improved access to administrative data, enabling its use for more sophisticated applications, rather than solely for tabulations. Government agencies are producing more administrative data when transacting with businesses and individuals.

### Administrative-Data-First Strategy

As administrative data become more readily available, DOS has adopted an Administrative-Data-First strategy, where the wealth of information in administrative data are leveraged prior to the use of survey methods.
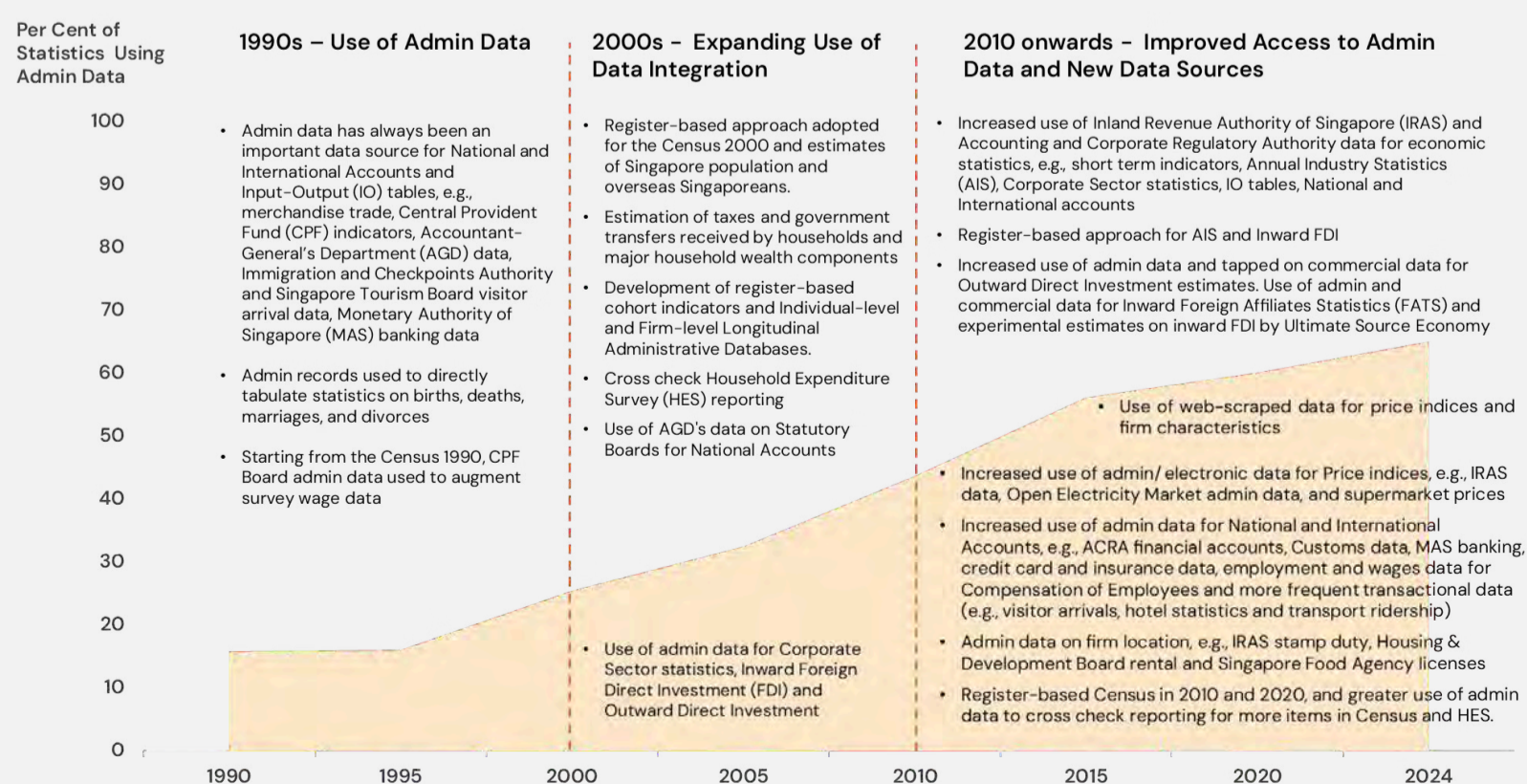
This strategy requires DOS to review and re-design statistical processes to incorporate administrative data. It also involves close collaboration with government agencies to identify their various administrative data sources, build data pipelines, and study their scope and definitions before integrating them into statistical processes.

This paper shares DOS's progress in using administrative data to streamline data collection and improve the data compilation process over the years. It also covers DOS's future plans for utilising administrative data and exploring new data sources.

## Progress on the Use of Administrative Data

Over the years, DOS's use of administrative data has grown. It has become an important data source for DOS and is now integral to its statistical processes. The progress and key milestones of DOS's administrative-data-first strategy are illustrated in Figure 1. On average, the proportion of DOS's key indicators that incorporated administrative data jumped from about 15% in the 1990s to about 65% in 2024.

### Figure 1: Progress of DOS's Administrative-Data-First Strategy, 1990-2024

**1990s – Use of Admin Data**

- Admin data has always been an important data source for National and International Accounts and Input-Output (IO) tables, e.g., merchandise trade, Central Provident Fund (CPF) indicators, Accountant-General's Department (AGD) data, Immigration and Checkpoints Authority and Singapore Tourism Board visitor arrival data, Monetary Authority of Singapore (MAS) banking data

- Admin records used to directly tabulate statistics on births, deaths, marriages, and divorces

- Starting from the Census 1990, CPF Board admin data used to augment survey wage data

**2000s – Expanding Use of Data Integration**

- Register-based approach adopted for the Census 2000 and estimates of Singapore population and overseas Singaporeans.

- Estimation of taxes and government transfers received by households and major household wealth components

- Development of register-based cohort indicators and Individual-level and Firm-level Longitudinal Administrative Databases.

- Cross check Household Expenditure Survey (HES) reporting

- Use of AGD's data on Statutory Boards for National Accounts

- Use of admin data for Corporate Sector statistics, Inward Foreign Direct Investment (FDI) and Outward Direct Investment

**2010 onwards – Improved Access to Admin Data and New Data Sources**

- Increased use of Inland Revenue Authority of Singapore (IRAS) and Accounting and Corporate Regulatory Authority data for economic statistics, e.g., short term indicators, Annual Industry Statistics (AIS), Corporate Sector statistics, IO tables, National and International accounts

- Register-based approach for AIS and Inward FDI

- Increased use of admin data and tapped on commercial data for Outward Direct Investment estimates. Use of admin and commercial data for Inward Foreign Affiliates Statistics (FATS) and experimental estimates on inward FDI by Ultimate Source Economy

- Use of web-scraped data for price indices and firm characteristics

- Increased use of admin/ electronic data for Price indices, e.g., IRAS data, Open Electricity Market admin data, and supermarket prices

- Increased use of admin data for National and International Accounts, e.g., ACRA financial accounts, Customs data, MAS banking, credit card and insurance data, employment and wages data for Compensation of Employees and more frequent transactional data (e.g., visitor arrivals, hotel statistics and transport ridership)

- Admin data on firm location, e.g., IRAS stamp duty, Housing & Development Board rental and Singapore Food Agency licenses

- Register-based Census in 2010 and 2020, and greater use of admin data to cross check reporting for more items in Census and HES.

Per Cent of Statistics Using Admin Data: 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100

Years: 1990, 1995, 2000, 2005, 2010, 2015, 2020, 2024

The increase in the proportion of administrative data used is also evident across various subject matter domains (Table 1). Today, some domains are purely compiled using administrative data. These include the basic population estimates and profile, statistics on births, deaths, marriages and divorces, corporate sector statistics, cohort indicators, as well as the firm-level and individual-level longitudinal administrative databases.

### Table 1: Estimated Proportion of Administrative Data Used in Statistical Compilation, 1990 to 2024

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

## Use of Administrative Data in the 1990s

In the 1990s, the proportion of key indicators that incorporated administrative data was around 15%. For certain domains, administrative data was already an important data source. For example, civil registration records were directly used to produce aggregated data on births and deaths, as well as marriages and divorces.

Likewise, administrative data has always been an important data source for the compilation of national accounts and international accounts. Examples include merchandise trade from Singapore Customs, visitor arrivals from the Singapore Tourism Board and Immigration and Checkpoints Authority, wage data from the Central Provident Fund Board (CPFB), and financial data from the Monetary Authority of Singapore.

In the Census of Population 1990, CPFB wage records were obtained by DOS and integrated with data collected through the Census to cross-check and correct under-reporting of wages. This approach was later applied to other surveys and subsequent Censuses.

## Expanding the Use of Data Integration in the 2000s

The 2000s saw an acceleration in DOS's use of administrative data, driven by the growing application of data integration.

A key milestone was the Census of Population 2000, which was the first register-based census in Singapore's history. Under this approach, administrative data from various sources were integrated to produce population counts and a basic demographic profile, rather than surveying all individuals in Singapore [1]. Subsequent Censuses of Population also adopted the register-based approach. Since 2000, administrative data has been used to produce Singapore's annual population counts and population profile.

Another key milestone was the production of new cohort indicators (e.g., marriage cohort dissolution rate) and the creation of individual-level and firm-level longitudinal databases by linking administrative records over time. This enabled DOS to **support studies conducted by government agencies that require a longitudinal perspective**.

DOS continued to supplement survey data with administrative data such as financial accounts from the Accounting and Corporate Regulatory Authority (ACRA) for statistics on the Corporate Sector, inward Foreign Direct Investment (FDI), and Outward Direct Investment. Likewise, administrative data (e.g., road tax data from the Land Transport Authority, migrant domestic worker levy data from the Ministry of Manpower) were used to supplement the Household Expenditure Survey. Administrative data from various sources were also integrated with survey data to estimate the taxes paid and government benefits received by households.

[1] The Census of Population 2000 surveyed some 20% of households for additional and more detailed information that were not available from administrative sources.

### Improved Access to Administrative Data and New Data Sources from 2010

From 2010, DOS's use of administrative data continued to grow as more data became available. In recent years, the rapid development of new Artificial Intelligence (AI) and Machine Learning (ML) technologies has unlocked new data sources. Access to firms' financial accounts from ACRA was also enhanced with the introduction of machine-readable format (i.e., XBRL [2]). With these developments, DOS made a concerted effort to review and redesign the compilation process of various statistics to incorporate more administrative data and new data sources.

Data from ACRA and the Inland Revenue Authority of Singapore are now widely used in the compilation of economic statistics, including the Wholesale Trade Index (WTI), Business Receipts Index (BRI), Input-Output tables, Annual Industry Statistics (AIS), Corporate Sector statistics, National Accounts, and International Accounts statistics. DOS has also progressively adopted the **register-based approach for AIS** and inward FDI estimates [3].

In addition to administrative data, DOS has started utilising new data sources such as commercial data and web-scraped data. From 2022, commercial data [4] have been incorporated in the compilation of Outward Direct Investment estimates and Inward Foreign Affiliates Statistics. Since 2015, DOS has been utilising web-scraping techniques [5] to automate data collection. ML methods are then applied on the web-scraped data for statistical compilation. For example, **web-scraping of online prices** is used for the compilation of price indices, while **web-scraped data from firms' corporate websites** help to profile their Internet presence.

# Future Plans

## Leverage New Technologies

DOS intends to further enhance its utilisation of administrative data and new data sources with web-scraping, Application Programming Interfaces (APIs), AI, and ML. These plans include:

Utilising AI-driven processes to directly extract detailed information from unstructured financial statements and companies' websites to support the compilation of industry statistics and acquire timely insights into Singapore firms' overseas investments. The application of ML models will also be expanded to estimate services trade and trading partners.

Further leveraging web-scraping and expanding the use of APIs to gather data on items such as the prices of clothing, overseas hotels accommodation, package tours and mobile phones for the compilation of price indices, where available.

## Expand Use of Administrative Data and New Data Sources

DOS has plans to expand its use of administrative data and new data sources to improve current processes or to release new indicators. These plans include:

Expanding the individual-level and firm-level longitudinal databases to include more data variables.

Developing new breakdown of Services Trade by enterprise characteristics, to be released in 2025.

Expanding the use of Goods and Services Tax data to cover more firms in the compilation of the quarterly WTI and BRI.

Compiling price indices using new administrative data sources, such as management fees and sinking funds, taxi fares, childcare and infant care fees, as well as trade data.

Expanding the use of administrative data in the Census of Population 2030, with about 60% of the data items expected to be obtained from administrative sources.

# Conclusion

Over the years, DOS's use of administrative data has significantly grown and evolved, leading to enhanced data compilation processes and the release of new data products. Looking ahead, DOS will maintain its Administrative-Data-First strategy and expand the use of administrative data and new data sources to produce statistics and data insights.

[2] 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**.

[3] With the register-based approach, all enterprises are included and estimated using administrative data. Larger enterprises are still surveyed to collect detailed breakdowns of their revenue, expenditure, and net lending items that are not available in the administrative data.

[4] Examples of commercial data include Mergers & Acquisitions news, as well as company financials from overseas business registers.

[5] DOS adopts web-scraping principles to assure that web-scraping is carried out consistently, ethically and transparently. DOS's web-scraping principles can be found on the **SingStat Website**.