The Use of Technology in Statistical Production

Ang Boon Heng[1], Eugenia Goh[2], Jeremy Heng[3]

[1] Ministry of Manpower, Singapore; ang_boon_heng@mom.gov.sg
[2] Ministry of Manpower, Singapore; eugenia_goh@mom.gov.sg
[3] Ministry of Manpower, Singapore; jeremy_heng@mom.gov.sg

Abstract
In an advanced information age, National Statistical Agencies need to continuously review and improve their processes in order to meet the growing demand for statistics in policy-making and research. It is important that the production and dissemination of official statistics are objective, timely and reliable so that the government and community-at-large can make informed decisions. As the National Statistical Agency responsible for the collection and dissemination of official labour market statistics, the Manpower Research and Statistics Department (MRSD) of Singapore Ministry of Manpower has explored and harnessed technology to enhance operational efficiency and disseminate quality information. This enables the government to formulate policies and programmes to secure a better future for its people. It also helps job-seekers, employers and employees to make informed decisions on career and human resource matters.

Keywords: technology, statistics, efficiency, quality.

1. Introduction

As the main statistical agency for official manpower statistics in Singapore, the Manpower Research and Statistics Department (MRSD) of Singapore's Ministry of Manpower produces a wide range of labour statistics that seeks to inform the government and community-at-large.

In recent times, technology has played an increasingly important role in the statistical production process. In this paper, we examine several initiatives of the past, present and future in which MRSD has made use of technology in its statistical product cycle – data collection, processing, analysis and dissemination.

2. Use of Technology in Statistical Production

2.1. Data Collection

MRSD conducts around 30 surveys annually at different frequencies, for both households and establishments. Apart from in-house surveys, MRSD also draws upon administrative data from external sources to supplement or validate data from survey sources.

In 2008, the Integrated Manpower Survey System (IMSS) was developed to provide an integrated platform for data collection. MRSD was then one of the first statistical agencies in the world to have a comprehensive online survey system for the conduct of major surveys including the Labour Force Survey.
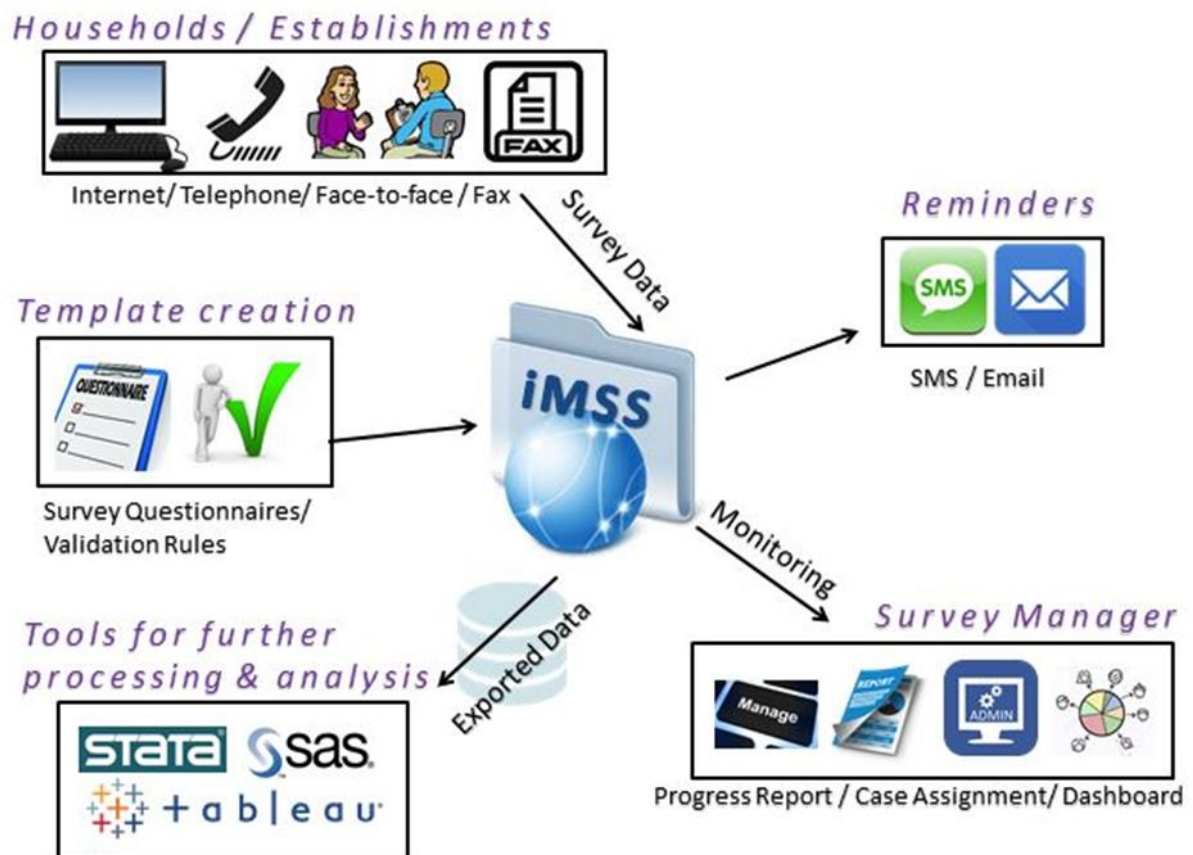

Fig. 1: IMSS Framework

Through IMSS, survey questionnaires and validation rules can be built into the system. This provides a convenient platform for survey respondents to log in and complete the questionnaires online. For responses obtained via fax, phone and face-to-face interviews, survey interviewers can input the information collected directly into the system. Interviewers would then be required to validate the responses of each completed case using the built-in validation rules to ensure data accuracy.

Survey managers are able to utilize the case assignment feature to allocate and manage workload among the various interviewers. In addition, a monitoring dashboard is also available for survey managers to track the overall progress of the survey and its individual cases.

With fast-changing respondent behaviour and increasing usage of smartphones and tablets, there are ongoing plans to revamp the system to incorporate more advanced mobile capabilities and survey analytics. For instance, mobile apps would be developed to allow respondents and interviewers to access the surveys on-the-go. Built-in survey analytics would be able to predict respondent behaviour which can provide insights to response patterns. This allows survey managers to fine-tune processes to enhance operational efficiency.

2.2. Data Processing

The survey data exported from IMSS has to be processed into usable information. A more comprehensive set of validation rules is run using software such as SAS to flag out any data inconsistencies. These are then transmitted via spreadsheets to survey interviewers for further review. This manual intervention would occasionally result in omission or duplication in checks, thus creating inefficiency and delays in timelines.

In 2014, the Unified Data Processing System was developed to provide a physical structure that brings together data validation and editing, data transformation and analytical functions. It seeks to reduce the need for manual intervention through automation and ensures better data governance by formulating and adhering to a set of rules and controls at each stage of data management.
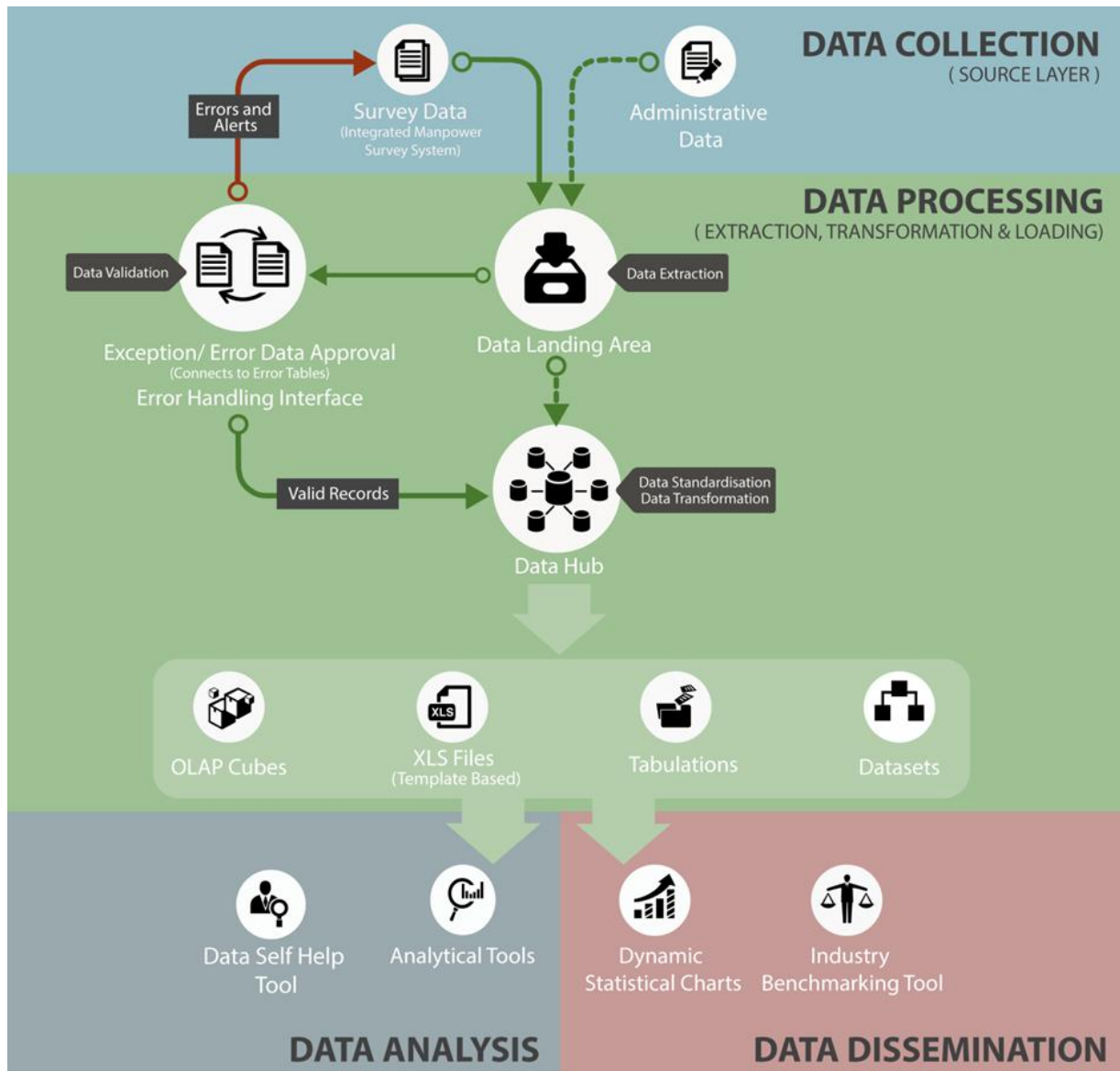
Fig. 2: Unified Processing System Framework

The data validation process begins by identifying survey responses requiring interviewer attention. Any data inconsistencies are flagged as error or alert cases and are automatically routed to the appropriate interviewer for corrective action. The process of error and alert identification, validation and editing leaves an audit trail of record changes as well as approval of such changes. These record changes can be easily retrieved if there is a need to trace back past decisions.

The data validation rules are written and coded in-house and integrated into the system. These pre-determined rules ensure that validation checks are consistently applied to the raw data iteratively until the data is robust and fit for analysis.

The maintenance of data comparability and coherence is efficiently achieved through a consistent system-driven Extraction, Transformation and Loading (ETL) process consisting of pre-determined formats and master lists. Data from multiple sources are pulled together to create datasets and Online Analytical Processing (OLAP) cubes ready for the next stage of data analysis or dissemination to data users.

Another type of data standardisation performed is the application of formats and master lists to be used across most data sources. An example of a type of master list is the national standard classification of industry, occupation and education. These data dimensions of industry, occupation and education are common variables used for further slicing and dicing of data. The ETL process pulls in these master lists and incorporates them into datasets to ensure consistent data generation.

As national standard classifications are regularly reviewed to better reflect the structure of the economy, master lists mapping past versions of standard classifications are also maintained and incorporated into the time series datasets to allow for easy generation of comparable data and OLAP cubes.

With the implementation of the new system, the collection and processing workflows can be automated and streamlined from the data validation stage to the production of OLAP cubes and other data products.

As less manual intervention is needed at each stage of data processing, more resources can be channelled to enhancing the design and efficiency of the system. For example, more master lists and variable formats can be built and added at the transformation stage to ease the creation of OLAP cubes and other data products. The unified system of data processing allows for the development of more sophisticated data visualisation software, as an extended layer of data analytics. The unified framework also provides a single interface to perform the administrative tasks required to create and maintain an integrated environment across multiple platforms.

2.3. Data Analysis

Due to greater complexity in data requirements over time, there is a need to move to more advanced analytics when analysing the various information. Apart from the demands of the general public about understanding the labour market issues and how it affects them, policy-makers are also increasingly using data as the main driver for policy decisions.

Since 2015, MRSD has adopted the use of analytical software such as Tableau to enhance the depth of data analysis and reporting. Graphs and charts can be better customised to identify visual patterns and spot trends, which will aid in data-driven policy-making. The various types of graphics can also be combined into a single dashboard and then managed dynamically. In this way, presentations are made more interactive which helps in forming the narrative, thus facilitating a deeper engagement with the audience.

2.4. Data Dissemination

Every year, MRSD releases around 14 publications about the state of the labour market on its statistical website, accompanied by detailed tables and charts. While comprehensive, there are

certain segments of the public who find these publications difficult to digest and sieve out useful information.

In 2010, MRSD became the first statistical agency in Singapore to develop and publish its own brand of infographics and videographics. Infographics summarise the key takeaways of each report through visuals and bullet points while videographics consist of short two-minute clips that offer insights into interesting labour market topics. Through these data visualisation tools, users would be able to better understand the wide range of labour market information.

MRSD's statistical website also contains useful features such as interactive charts and benchmarking tools, which are powered by Adobe Flash and offers a large degree of customisation for users. Interactive charts provide users the flexibility to select the type of data breakdowns that are of interest to them. Benchmarking tools allow users to compare their organisation against its industry counterparts in wages, employment conditions and staff turnover.

Through public feedback and engagement sessions, MRSD would be able to better cater to the needs and preferences of various users. As such, MRSD would continuously review the content on its website[1] and improve upon it.

---

[1] http://stats.mom.gov.sg

3. Conclusion

Good data quality and governance are fundamentals in producing statistics that facilitate informed decision-making for a broad range of economic and social outcomes. MRSD seeks to harness the potential of technology to help achieve these goals. It is important that the production and dissemination of official statistics is objective, timely and reliable so that it inspires confidence and trust in the community. MRSD seeks to harness the potential of technology to help achieve these goals. As new frontiers in technology emerge, MRSD will review and adopt new initiatives to aid in its statistical production process.

While technology is an essential component in the statistical production process, it is important to be mindful of the impact of the over-reliance of technology. There should be contingency plans to be able to produce the required statistics if technology breaks down. As we rely more and more on technology, there is also a danger of losing the aspect of human interaction and personal touch. Therefore, it is essential for us to utilize technology to aid in statistical production, not replace us.