

Assessment of Data Quality for Census conducted during Covid-19: Singapore's Census 2020

Expert Group Meeting on Impact of Covid-19 Pandemic on
Conducting Population and Housing Censuses and Census Data Quality Concerns

9 - 12 Feb 2021

www.singstat.gov.sg

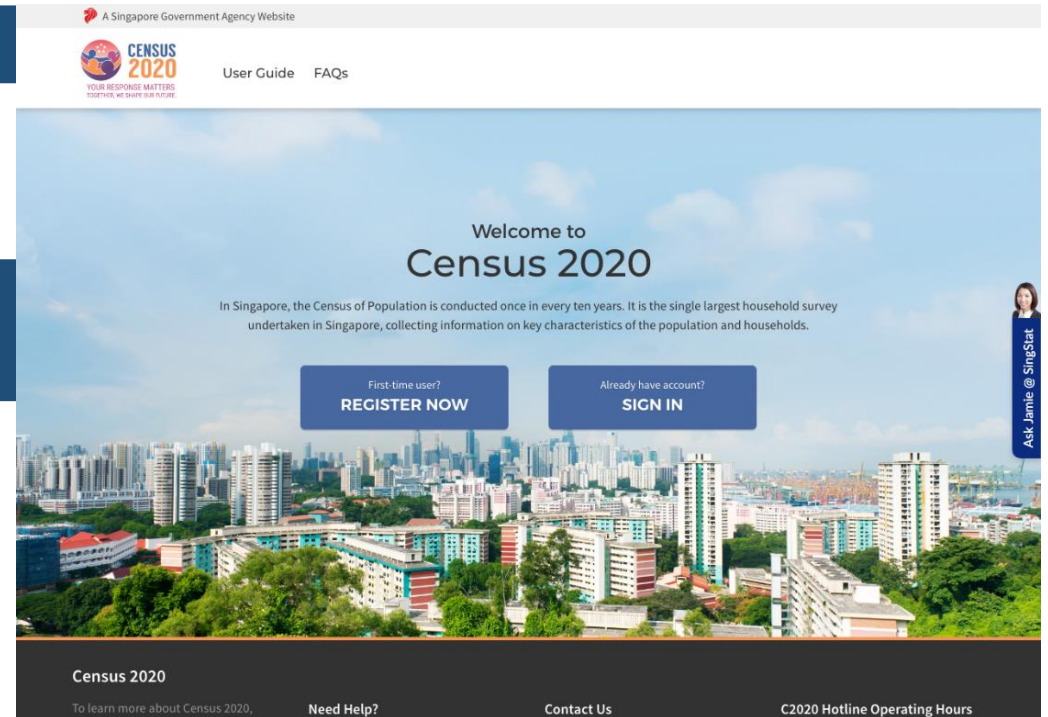
Approach for Census-Taking in Singapore

Background

- Singapore Census is conducted once in 10 years, in years ending '0'

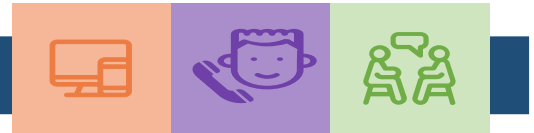
Register-based Approach using Admin Data from Different Sources, Supplemented by Survey

- Register-based Census adopted since 2000
 - Administrative records from multiple sources are merged to provide basic demographic information such as age, sex and ethnic group for the whole population (full coverage)
 - Large-scale sample survey is conducted to capture in-depth information on socio-economic and household characteristics (e.g. language, religion, transport, detailed household living arrangement, disability) that are not available from administrative sources
- Census 2020 survey covered some 150,000 households
- Census Reference Date is as at 30 June 2020



Tri-Modal Data Collection

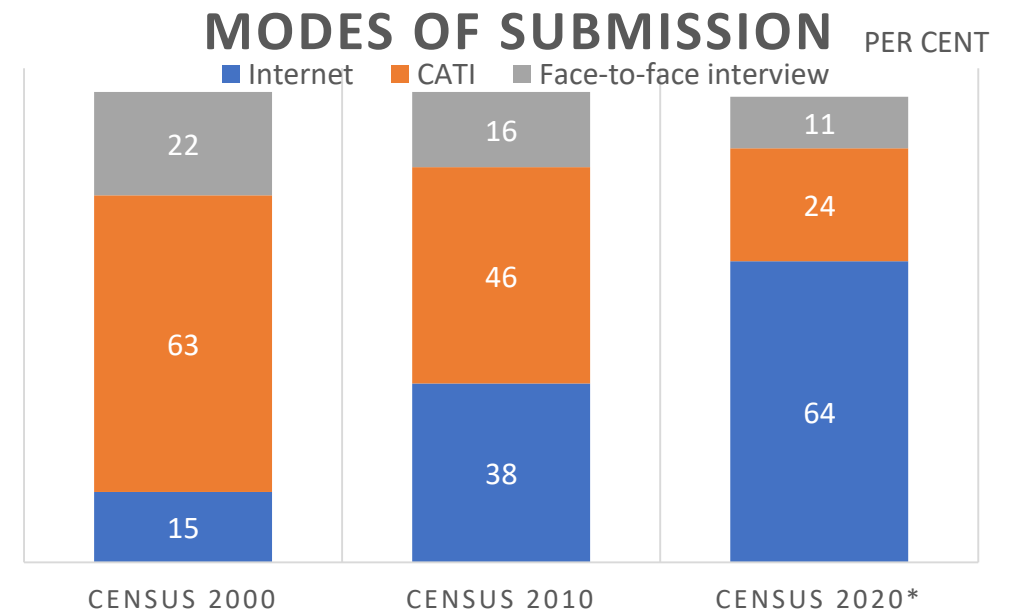
Data Collected via 3 Modes



- Tri-modal data collection strategy adopted since 2000
- Cater to varied profile and needs of population while balancing resource considerations
 - Online Submission via Self-Enumeration
 - Phone Interview through hotline using Computer-Assisted Telephone Interview (CATI)
 - Face-to-face Interview with field interviewers using Tablets

Adoption Rate by Mode of Submission

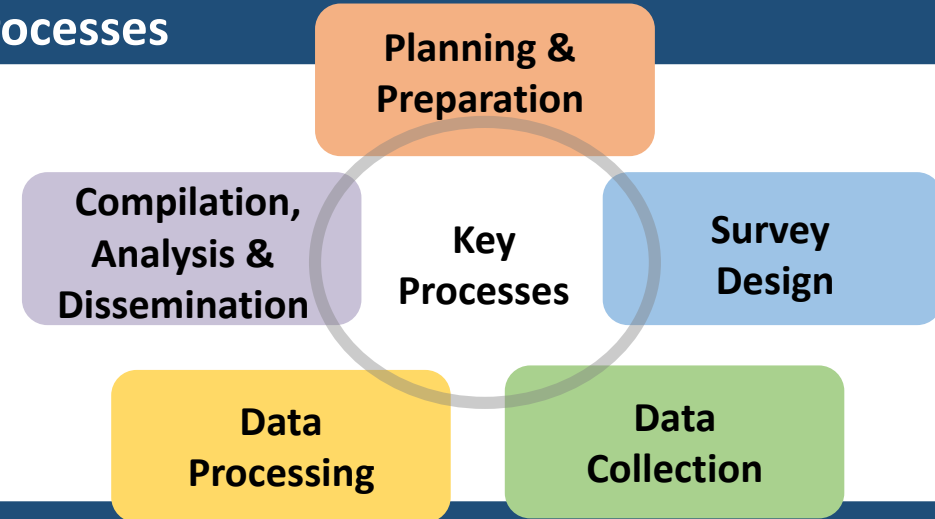
- Response via Online Submission has increased steadily with higher internet penetration rate and computer literacy
- Online submission constituted more than 60% of all responses for 2020



Quality Dimensions

Multi-dimensional Quality Considerations in Key Processes

- Accuracy
- Timeliness
- Interpretability
- Accessibility
- Coherence
- Relevance



Minimising Errors to Ensure Data Quality

- Reducing sampling and non-sampling errors with control measures, including:
 - Careful design of survey questionnaire and data capture instrument
 - Careful planning of operational procedures in data collection, processing and tabulation
 - Standardised concepts and definitions, including conveying to respondents on self-enumeration
 - Training and close supervision of interviewers
 - Verification process to ensure accuracy of survey returns
 - Stringent controls for data editing and coding to ensure good data quality and high data consistency
 - Increased awareness to improve coverage and minimize non-response

COVID-19 in Singapore and Census' Response

Impact on Resources and Mitigation

IMPACT ON C2020 TIMELINE

The C2020 activities, strategies and measures adopted in response to the evolving COVID-19 situation are summarised in the timeline.



COVID-19 Events and Timeline

23 JAN

First confirmed COVID-19 case in Singapore

7 FEB

Disease Outbreak Response System Condition (DOSRCON) Orange declared

26 MAR

Entertainment venues closed, social gatherings limited to 10 pax, mass gathering cancelled

7 APR

Circuit Breaker starts

1 JUN

Exit Circuit Breaker, Phase 1 starts

19 JUN

Post Circuit Breaker, Phase 2 starts

Census 2020 Response



4 FEB

Launch of Census 2020

15 FEB

Fieldwork suspended and only arranged upon request

12 MAR

Third Reminder letters sent to boost responses

29 MAR

Full survey operations carried out in Split Team arrangement

7 APR

Work-From-Home arrangement with scaled down operations. Internet submissions and hotline continue, supported by C2020 staff working from home

21 JUN

Full Call Centre and Data Processing operations resume in Split Team arrangement

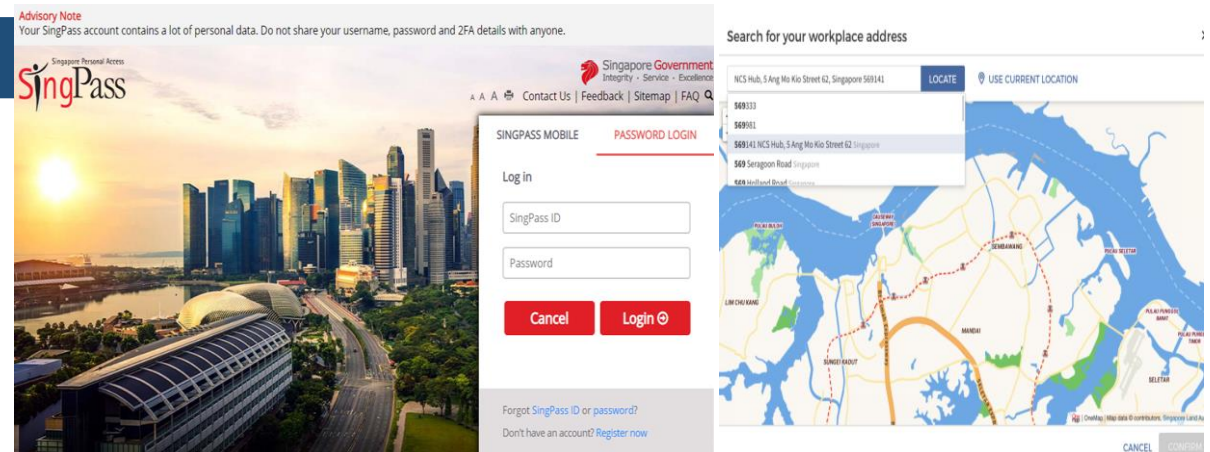
19 JUL

Fieldwork resumes. Field interviewers initiate contact with households which have yet to complete survey. Additional safe management measures, including temperature taking, mask wearing, interviewing at gate/doorstep are put in place.

Online Submission and Data Quality

Benefits for Census Office and Respondents

- Reduce reliance on manpower
- Expand reach to difficult to contact groups
- Available 24/7, flexibility to respond anytime, anywhere, using any device
- Address privacy concerns



Key Features of Online Form

- Leverage nation-wide login credentials used by Government for consistent delivery and legitimacy
- Basic information, e.g. name and unique identification number, preloaded and displayed for verification upon authentication to minimise data entry errors
- Variables not subject to changes, e.g. date of birth, matched backend via unique identifiers
- Tool-tip and Virtual Assistant embedded in form for immediate assistance
- Dropdown selection to facilitate automated coding of Industry and Occupation
- Map application to supplement for Address field
- Automatic branching of questions and simple completeness and validation checks built in prior to submission to prevent omission and minimise accidental misreporting
- Reduce back-end data processing efforts with electronically coded data

Quality Controls at Data Collection

Computer-Assisted Telephone Interviews

- Similar online form accessed by Call interviewers
- Additional language toggle to display questionnaire in different languages to facilitate interviews conducted in different official languages for consistent delivery of questions
- Accessible, lower barriers and particularly useful for those who are illiterate/less internet savvy
- Flexible in catering to different languages and dialects
- Increase response rate by contacting households who did not complete their returns by a stipulated date
- Arising from Work from Home arrangements introduced at an early stage of the data collection, close supervision and checks by supervisors to provide in time training. Checks performed on individual interviewers to identify misconception

Face-to-Face Interviews (resumed in nation-wide Phase 2 reopening)

- Similar online form accessed by field interviewers with the use of mobile devices
- Reach out to respondents who do not respond via Internet and cannot be contacted by phone
- Limited extent with safe-distancing safety measures put in place due to COVID-19, but remained important for selected groups who prefer in-person authentication
- Consistent training delivery with video conferencing, facilitated by supervisors under split team arrangement and safe management measures

Quality Assessment at Data Processing

System and Trend Check

- In-built system checks to flag out errors and inconsistencies across data items
- Customised checks added to address data issues that arose due to the pandemic
 - E.g., Checks on Workplace Address and Transport Mode/Travelling Time to highlight reporting on Work from Home arrangement due to short-term response to COVID control measures
- Detailed cross tabulation of data for comparison with historical years

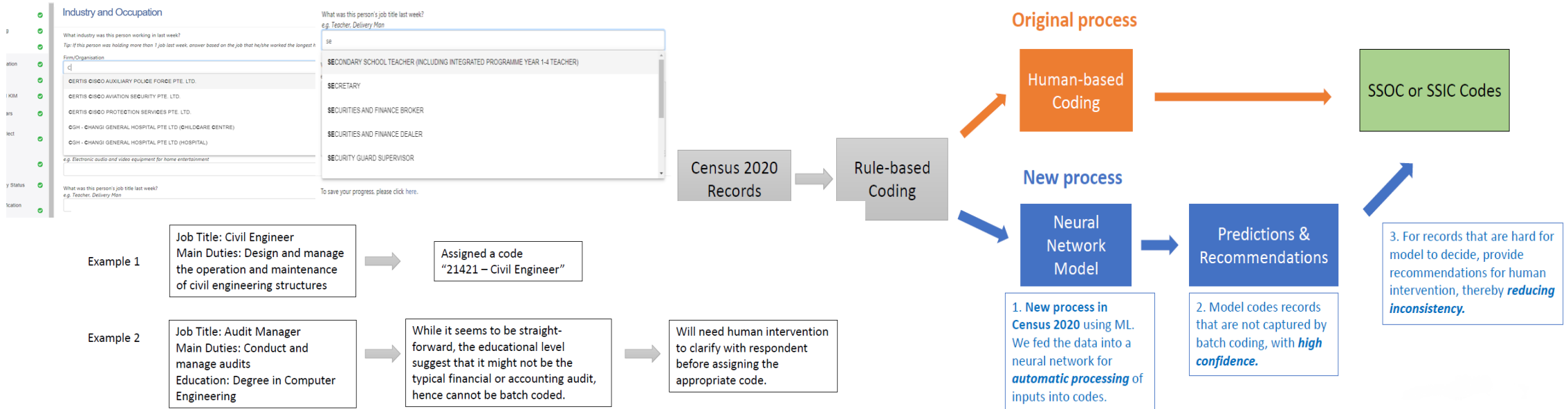
Use of Administrative Data

- To minimise response burden and errors in reporting, information where available are preloaded and not asked again from respondents.
- Administrative records and census data merged for in-depth checks and updating
 - Facilitated by the availability and use of unique identification numbers in government databases, data can be merged for verification
 - E.g. Monthly wage and work pass records to validate employment related data such as Current Activity Status, Industry, Income. Arising from temporary work stoppage/leave, some respondents may report themselves as not working but were still employed and receiving wages. Checks allow such records to be flagged for further verification.

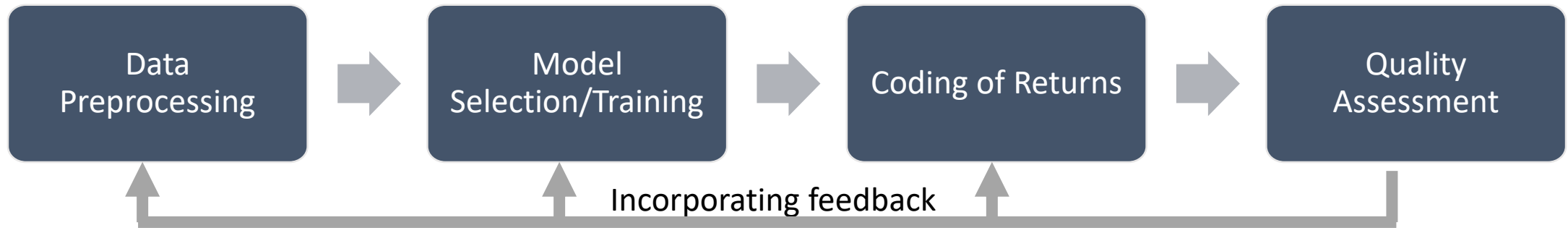
Use of Machine Learning in Data Processing

Application of Neural Network Model for Coding of Occupation

- Machine learning module first developed using past labour force survey data, launched for Census 2020
- Auto-coded for occupations that can be mapped from online selection
- Descriptions in text format go through rule-based batch coding for auto-assign of codes
- Records that cannot be batch coded will go through machine learning module for auto-assignment of codes (high confidence level), code prediction (medium confidence level) or flow through for manual coding (for codes with low confidence level)
- Reduce human-based coding reduced, saving > 4,500 manhours and more consistent code assessment



Use of Machine Learning in Data Processing



On-going Quality Assessment and Improvement

- Periodic checks on prediction results
 - Identify abnormal data behavior or new phenomenon (e.g., COVID-19 affecting returns)
 - Identify unnatural inputs (e.g., unusually small or large incomes)
 - Identify predictions which had good scores but were wrong
- Feedback from users down the data pipeline (Data Collection and Processing teams)
- Modify preprocessing procedures to improve robustness against noise
 - Apply suitable transformations/feature engineering to inputs (e.g., to lower the impact of skewness and extreme values)
- Update model with more timely training data
 - Augment training data with records that were manually coded in earlier periods of coding process

Key Lessons

Implementing Quality Controls and Assessment, Leveraging Admin Register and Technology

- Relatively high online submission take-up rate allowed for majority of the responses to continue to flow in despite the scaled down operations. However, from data collected via the multi-mode approach, quality of responses for online self-enumeration is observed to be poorer due to a lack of understanding of questions (without guide from trained interviewers) or missing details (for descriptive fields). In response,
 - Online help and prompts added to the form design
 - Completeness and validation checks put in upfront
 - More intense consistency checks performed backend, with common errors identified and call-backs to follow up
- With basic data on population estimates compiled from population register, top-line population data continue to be produced on schedule though there is expected to be slight impact on release data for the detailed statistical releases.
- Use of administrative data for consistency checks and application of machine learning in coding help alleviate manpower constraints in data processing, where resources were scaled down.

Thank You

Our Vision

National Statistical Service of Quality, Integrity and Expertise

Our Mission

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

www.singstat.gov.sg

A decorative graphic in the bottom right corner consists of several thick, overlapping, wavy ribbons in shades of blue, purple, orange, and red, creating a dynamic, flowing effect.