

# Building modern and resilient population data systems to enhance data quality, improve cost efficiency and policy relevance

03 October 2024

Angela Ngyende

Statistics South Africa

## Background (1)

- ▶ Data echo systems have evolved in the most recent times and will continue as more data sources emerge-including big data, increased use of admin data such vital events (births, deaths, marriages and divorces) e.t.c.
- ▶ Emerging data needs and increased demand for regular, quality data not only on demographic aspects such as population size, structure and composition but also on households' socioeconomic status and living circumstances.
- ▶ More complex data needs are eminent and require adaptive data systems that comprise of various sources to inform population development agenda and environment nexus.
- ▶ Globally, there has been a shift in the last decade from reliance on primary sources – Censuses, Household based Surveys and some admin data sources in some regions and countries
- ▶ Although most countries largely depend on full- field enumeration censuses to generate data on key indicators, emerging data needs and systems are about to disrupt the old order!

## Evolution of data echo systems (1)

- ▶ The evolution of data echo systems pose both opportunities and challenges to different nations, depending on level of development, technological advancement, investment in capacity building and technical know-how on data integration and modelling.
- ▶ Challenge -most admin data sources not yet meeting quality requirements- Level of completeness for most sources in many countries not yet at desired level, other countries not having quality assessment frameworks to determine quality of admin data sources- thus cannot form part of official statistics.
- ▶ It is against such existing data echo systems that for example in Africa, 39 out of 54 countries in this region conducted full-field based censuses despite fiscal challenges. South Africa is one of them that concluded Census 2022 and is in the process of disseminating results on small areas.

## Evolverment of data echo systems (2)

- ▶ Dissemination of data at the lowest appropriate geographic level remains one of the essential features of a population and housing census. Census 2022 is the only data source across spheres of government, business and academia at lower levels of geography to inform planning and resource allocation, development, training, policy-making and programme monitoring.
- ▶ Conducting a successful Census 2022 in South Africa despite the challenges posed by COVID-19 Pandemic and fiscal pressures is one of the latest country's milestones .

# Evolution of data echo systems (3)

## Census Landscape in South Africa

### ▶ Traditional Model: Full field based Censuses

- Full enumeration
- Post Enumeration Survey conducted to evaluate quality of census

### ▶ Fully digital Census in 2020 round of censuses

- Census geo frame
- Census data collection instruments-Long form-HH based pop and Short Form- Homeless and Institution based pop

### ▶ Transition to multi-mode data collection approaches

- CAPI
- CAWI
- CATI

## Evolution of data echo systems (4)

### CAWI Environment

- Use of online questionnaire
- Contingency planning to promote self enumeration
- Direct benefits: Respondents complete the questionnaire privately, responses are entered directly into a system-expectation is reduced errors and improved data quality as well as improvements in coverage of hard to reach- affluent areas.
- Reduces the need for face-to-face interviewers, lowering enumerator costs and logistical challenges- Zero user cost through reverse billing

### CATI

- Minimal uptake

## Opportunities (1)

- ▶ Exponential increase in mobile connectivity, along with broader advancements in technology—such as smart devices, GPS, and web-based GIS—offers new possibilities for collecting census data timely and with greater accuracy.
- ▶ International guidelines on field based Census
- ▶ Use of dashboards for effective real-time monitoring of the data collections process
- ▶ Well defined Census methodologies and processes
- ▶ Well developed quality assessment framework that is being used to assess admin data for completeness and other quality dimensions
- ▶ Changes in legislative framework –will enable Stats SA to access alternative data sources that can be used to augment primary sources

## Challenges

- ▶ Upward trend in population and household undercount – attributed to a number of factors:
  - ❑ Complexities in living arrangements,
  - ❑ Reduced public interest in participating in Census exercise,
  - ❑ Timing of Census 2022 – COVID-19 pandemic environment which affected both planning and execution phases of the census
- ▶ Decreased public trust in Census results
- ▶ Increased costs in census undertaking Vs fiscal challenges and competing demands for public resources
- ▶ Apart from Censuses, hardly any data sources that avail indicators at lower levels of geography

## Lessons learnt

- ▶ Try and maintain sequential application of data collection modes during data collection
  - ❑ Only go parallel once you have control and you want to plug gaps
- ▶ Apply secondary editing concurrently with data collection
  - ❑ Field comparison tables with complementary census/admin information
  - ❑ Do secondary data validations –soft/hard checks
- ▶ Design and develop Census output database before data collection commences and TEST/TEST/TEST
- ▶ Increased costs Vs fiscal challenges—over spending due to extended census enumeration period- begin looking at alternative, cheaper methods of producing estimates

## Lessons learnt (2)

DIGITAL WORLD

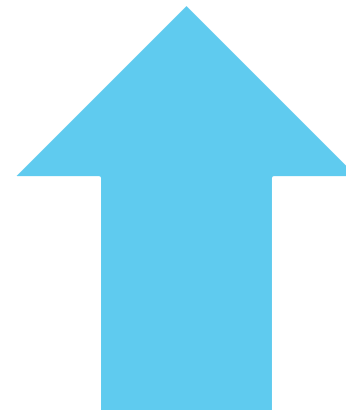


90%

Pre-planning: geo  
frame, data  
collection  
instruments design &  
operations

- Frame updates
- Logical Skips, data Validations
- Data structure editing

10% downstream  
( post data  
collection  
processes)- data  
integration,  
editing, coding,  
analysis



- Automated coding
- Limited editing
- Timely dissemination

## Recommendations

- ▶ When mixed mode collections are to be applied- should be properly planned, implemented, monitored and controlled.
- ▶ Digital censuses need thorough testing! All systems should be thoroughly tested before deployment
- ▶ Countries should determine the data deficiencies and their quantitative impact on indicators to inform dissemination strategies.

## Conclusion (1)

- ▶ With the ever increasing data needs and complexities, countries to need to develop and maintain resilient population data systems which require the following contingencies:
  - ❑ Innovation towards data collection, data integration, processing and data analysis ,
  - ❑ Continually invest in capacity building in data modeling techniques
  - ❑ Creation of partnerships with other statistical organizations to share experiences in latest data modeling techniques used to produce population estimates using admin data and alternative data sources
  - ❑ Ensuring that data quality remains priority irrespective of data sources used by establishing and enforcing standards and quality frameworks

## Conclusion (2)

- ❑ Investing in research prior to venturing into use of alternative data sources to conduct hybrid censuses
- ❑ Adopting cost-effective approaches to avoid/reduce cost of conducting full –field based Censuses
- ❑ NSO investment in development of small area estimates

**Thank you**

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the frame, creating a modern, layered effect against the white background.