CENSUS 2022

HOW THE COUNT WAS DONE

stats sa

Department:
Statistics South Africa
REPUBLIC OF SOUTH AFRICA

NDP 2030

Census 2022
How the count was done

How the count was done / Statistics South Africa

A complete set of Stats SA publications is available at Stats SA Library and the following libraries:

> National Library of South Africa, Pretoria Division
> National Library of South Africa, Cape Town Division
> Library of Parliament, Cape Town
> Bloemfontein Public Library
> Natal Society Library, Pietermaritzburg
> Johannesburg Public Library
> Eastern Cape Library Services, King William's Town
> Central Regional Library, Polokwane
> Central Reference Library, Mbombela
> Central Reference Collection, Kimberley
> Central Reference Library, Mmabatho

This report is available on the Stats SA website: www.statssa.gov.za

For technical enquiries please contact:

Mosidi Nhlapo
Tel: 012 310 8223
Email: mosidin@statssa.gov.za

Angela Ngyende
Tel: 012 310 4699
Email: angelan@statssa.gov.za

Abbreviations and Acronyms

ACRONYM / ABBREVIATION	DESCRIPTION
CAPI	Computer Assisted Personal Interview
CAWI	Computer Assisted Web Interview
CATI	Computer Assisted Telephonic Interview
EA	Enumeration Area
GIF	Geospatial Information Frame
IEC	Independent Electoral Commission
MDB	Municipal Demarcation Board
PCMA	Publicity, Community Mobilisation and Advocacy
SDI	Special Dwelling Institutions
Stats Act	Statistics Act No.6 of 1999
Stats SA	Statistics South Africa

Table of Contents

List of tables

List of figures

1.   INTRODUCTION

1.1   Background/history of Census in South Africa

Censuses provide reliable data at various levels of planning, essential in helping countries monitor development programs. Census data are fundamental for informed planning, policy-formulation and decision-making in various sectors as nations address socio-economic and service delivery challenges. Since 1994, South Africa has had four Population and Housing Censuses (in 1996,2001, 2011 and 2022). The information collected during this process plays a crucial role in planning at the national, provincial and municipal levels, by providing policymakers with the information required to formulate rational policies and implement service delivery. This includes building and maintaining critical infrastructure such as hospitals and schools. The census data is also critical in determining budgetary allocations for various spheres of government.

Census 2022 was the first digital census to be conducted by Statistics South Africa (Stats SA). The fundamental benefits of a technology-driven census cannot be overemphasized; notably, record time quality data, with census results released within 12 months after data collection. This achievement is attributed to the deployment of real-time data collection instruments/tools, geography frame, data collection applications and systems, data processing and automated coding.

In preparation for Census 2022, a multi-mode data collection approach was adopted and tested in various tests including Census Pilot, before implementation during the main census. Three methods of data collection were used in this census, namely:

1. Face to face interviews- Computer-assisted Personal Interview (CAPI);
2. Telephonic interviews -Computer-assisted Telephone Interview (CATI); and
3. Computer-assisted Web Interview (CAWI).

Use of multi-mode data collection approach became an advantage in conducting a census in the COVID-19 pandemic environment, which affected Census key phases of geography frame finalisation and data collection. The repeated COVID-19 pandemic waves in South Africa led to the implementation of strict regulations in population movement and interactions between and across households. The restrictions brought interruptions in the census project activities forcing Stats SA to postpone the Pilot Census to July 2021. Prior to the full implementation of the Pilot Census, COVID-19 pandemic infection rates increased drastically and government enforced stricter regulations between May 2021 and August 2021. Owing to the importance of the Pilot Census exercise, Stats SA management recommended implementation of remote data collection methods of Computer-assisted Telephone Interview (CATI) and Computer-assisted Web Interview (CAWI). However, it is important to note that successful implementation of these two collection methods depended on access to households' contact details. Although the two collection modes were deployed, not many households within the sampled Pilot Census areas participated. This was mainly due to poor household registrations, a process that preceded data collection. In order to ensure that all Census 2022 processes, methods, tools and systems/applications were fully tested, face-to-face data collection method had to take place despite the unconducive COVID-19 environment. The face-to-face data collection method (CAPI) thus did not take place in the stipulated enumeration period.

1.2   Legal basis for conducting Census

Stats SA's mandate to conduct censuses is derived from the Statistics Act (Act 6 of 1999). This legislation tasks the statistical organisation with collecting, producing and disseminating official statistics, including periodic national population censuses. The Act also provides for the establishment of the Statistics Council representing a variety of interests in the statistical community.

1.3   Census 2022 goal and objectives

The goal of the Census 2022 project was to count everyone within the borders of South Africa without omission and duplication. Census 2022 key objectives were linked to three questions;

- **How many are we?** Determining population size per locality/area, a critical indicator used for resource allocation, measurement of the extent of service delivery, decision making and budgeting, among others.
- **Who are we?** Census 2022 data provides the current picture in terms of population dynamics of the South African population including demographics and some socio-economic characteristics. The information on population characteristics such as age and sex composition, educational attainment and employment status is pertinent to planning and resource allocation
- **Where do we live?** Census 2022 data provides insights on living conditions of South Africans in terms of the number of households and average household sizes and the type of dwelling structures (housing), access to water, availability of essential services and facilities, and access to Internet etc. This information is critical in understanding and addressing development challenges at all levels of geography and communities.

  Central to answering the three questions is how the information collected in Census 2022 compares to previous censuses. Comparing Census 2022 results with previous censuses highlights changes such as population growth, and measurement of progress in addressing various development challenges and identifying emerging data gaps and issues. This information is fundamental for data users (stakeholders) such as national departments, development partners, Civil Society Organisations (CSOs), the private sector, researchers and the general public.

In summary, a population and housing census provides detailed statistics on population composition, characteristics and spatial distribution to the lowest geographical area.

2.   CENSUS 2022 PLANNING, ORGANISATION AND ADMINISTRATION

2.1   Overview

A successfully executed census is dependent on well planned and implemented census strategy, goal and objectives and linked programs for each census phase. It is paramount to identify and define all census activities linked to pre-enumeration, enumeration and post-enumeration processes, methods and procedures. A digital census particularly requires well planned, coordinated, tested systems and applications linked to each census phase.

In order to plan for, and implement a successful Census 2022, Stats SA management put in place a census structure constituting 12 workstreams and various committees. The role of each workstream and committee is outlined in sections 2.2 to 2.4.

2.2   Census management and operational structures

In preparation for Census 2022, Stats SA enacted a census structure to plan for and coordinate all activities during project implementation. This was to ensure that the census objectives and methodologies are executed accordingly, and monitor progress towards a complete and successful population count.

Census 2022's management structure comprised of twelve managerial workstreams, each responsible for the planning and implementation of census activities. Each workstream was mandated with specific objectives and outputs which were implemented through the various tests and census pilot in preparation for main census. The twelve workstreams included:
- Project Management Office (PMO);
- Secretariat;
- Census Inputs and Outputs;
- Data Operations;
- Governance;
- Corporate Services;
- Census Geography (Frame Update);
- Information Technology (IT);
- Census End-to-End Systems Development (CEESD);
- Publicity, Community Mobilisation and Advocacy;
- Field Logistics and Specification Development; and
- Provincial Coordination and Quality Assurance.

2.2.1   Project Management Office

Project Management Office (PMO) workstream was responsible for the development and application of best project management practices to ensure a successful planned and executed Census 2022 project. Workstream specific objectives included:

o To ensure that the census project was planned and managed in a structured manner and that the principles of good project management were applied throughout the project life cycle.

o To facilitate project planning, monitoring and reporting support during Census 2022 planning and implementation.

o To serve as the gatekeeper of the project management standard, processes and systems, and provide support for the Census 2022 production of high quality deliverables. This is done by ensuring the implementation of project management tools, common procedures and progress monitoring and measurements.

o To ensure overall project management, coordination and monitoring of workstream activities and all strategic, policy and governing issues pertaining to the project.

2.2.2   Data Operations

The Data Operations workstream was responsible for planning and implementation of Census data collection, data processing and integration. In preparation for data collection, the workstream was also responsible for training of field staff. Key deliverables included operation plan, data collection methodology and procedures, trained census staff, business requirements for data collection applications and systems, training plan and manuals, monitoring of data collection and progress reports, edited and integrated census data.

Data Ops had to give form to the requirements of the end-to-end census. Some of the outputs of the workstream were: operational plan, data operations methodology, training plan, fieldworkers manual and reports.

2.2.3   Field Logistics and Specifications Development

The Census exercise involves determination of logistical requirements, sourcing and distribution of census materials within specified timeframes. Field Logistics and Specifications Development (FLSD) workstream was responsible for planning, execution and control of the procurement, movement, and stationing of census materials. All procured materials were distributed from a central warehouse at head office (HO) to the different district offices based on the distribution plans.

The deliverables of the workstream included facilitation and consolidation of the Census 2022 project's logistical requirements, specifications for the field logistics management tool/application, logistics officer's manual/guide, facilitation of cost-effective procurement of Census 2022 materials through Supply Chain Management (SCM), and implementing the field logistics management tool (FLOS) for the purpose of distribution, tracking and monitoring of the Census 2022 materials.

Further, the workstream coordinated forward and reverse logistics between HO and the provincial/district offices providing efficient asset/inventory management and the training of the district logistics officers (DLO) on the forward and reverse logistics and FLOS. The workstream also provided logistical support to all Census fieldwork related processes. It worked collaboratively with the national, provincial and district offices; receiving, storing, packaging and distributing Census 2022 materials to these offices. It additionally sought to enforce efficient quality control measures for all processes in the logistics value chain.

The successful implementation of the FLSD workstream programme depended on other workstreams meeting their obligation of providing their logistical requirements. The FLSD team liaised with other workstreams to compile their needs into a comprehensive Logistics Master Plan, which was used to advise the Supply Chain Management (SCM) that independently procured the items/materials. The materials were delivered to FLSD working in partnership with SCM to verify that the items delivered were correct in numbers and quality as specified. The team was also responsible for storage and maintained record for issuing out and distribution through courier or by other specified means to the provinces and districts.

2.2.4 Census Geography Frame Update

Frame Update workstream (Frame Update) was responsible for Geospatial Information Frame (GIF). The Census Geography Frame Update workstream was responsible for providing the Census 2022 geographical frame that included Enumeration Areas (EAs) with place names on digitally captured structures and maps. The workstream played a number of roles including identification and assessing of the appropriateness of external data sources towards constructing a GIF, creation of geospatial platform, preparation of data for mobile navigation, development of fieldwork procedures and training manuals and conducting frame update training. For the purpose of recruitment, field operations and resource planning, Frame Update was responsible for creating Fieldwork, Supervisor, Field Operations Officer (FOO) and District Census Coordinator (DCC) Units.To support the multi-mode data collection, the workstream also provided resources to customise the online registration and unpacking of structures.

The process of updating the frame in preparation for Census 2022 involved a number of projects including capturing of Census 2011 listing books and field verification of Community Survey (CS) 2016 sampled areas. The frame update had been planned to be finalised in May 2021, however there were some delays due to COVID-19.

2.2.5 Facilities, Transport and Security Management

The Facilities, Transport and Security Management (FTSM) workstream was responsible for screening of contract staff applicants in order to identify and exclude those with criminal records from the census project, securing vehicles, airtime/data, safe storage of tablets and registering Census 2022 with National Joint Operations Committee (NATJOC) and Provincial Joint Operations Committees (PROVJOCS). Fleet sourcing was one of activities faced with some challenges due to closure of most vehicle dealerships during the COVID-19 hard lockdown. This resulted in shortages and delayed delivery of vehicles, impacting on downstream census activities.

2.2.6   Information Communication and Technology

Information Communication and Technology (ICT) workstream was responsible for the development, implementation and maintenance of an efficient and effective integrated ICT infrastructure and architecture to enable the organisation to conduct a successful Census 2022. This included the provision of technical support throughout the Census 2022 exercise and implementing internal/external service provider's requirements.

The workstream was also responsible for the configuration of tablets for the Census project and procurement of servers. The ICT team also provided ICT infrastructure for virtual training and support during the national, provincial and district training, as well as during data collection.

Census 2022 was digital and central to this was applications and systems that required ICT environment that was agile and adaptive to cater for an increased network load. To achieve this, the workstream increased Access Point Name (APN), internet link and also engaged service providers to boost network coverage in some affected areas throughout the data collection period.

2.2.7   Census End to End Systems development

The Census End-to-End Systems (CEESD) workstream was responsible for developing, implementing and maintaining efficient and effective integrated application architecture to enable the organisation to conduct a successful Census 2022.

The workstream developed in-house applications and provided support during implementation of the census project. In a digitally driven census, both ICT and CEESD workstreams played a pivotal role and were thus critical at all census phases; pre-enumeration (for the preparation of census geography frame, development and testing of data collection instruments/tools, development of training materials and actual training of field staff), enumeration and post- enumeration (data editing, coding and integration). The applications developed included among others data collection tools of CAPI, CATI and CAWI. Other workstream objectives included:
 o Development, implementation and maintenance of quality approved systems to enable Stats SA to conduct a successful digital Census 2022;
 o Testing the efficacy of the deployed end-to-end census systems, integration of census systems and automation of all identified census processes;
 o Delivering real time reporting to enable project stakeholders to make informed decisions; and
 o Providing continuous technical support during all levels of training and during data collection.

2.2.8   Census Inputs and Outputs

The workstream was responsible for Census 2022 Inputs (data collection instruments/questionnaires and tabulation plan for main statistical release) and census outputs (data editing specifications and basic print products). The key role played by the workstream included:

- Conducting engagements with stakeholders on census data items and census products;
- Development of business requirements for all census questionnaires;
- Development of the Census 2022 household questionnaire, Special Dwelling Institutions (SDIs) questionnaires, Homeless questionnaire and Transient questionnaire;
- Development of questionnaire validation rules and error messages for online questionnaire;
- Development of editing specifications and imputations;
- Quality assurance of all electronic questionnaires including translations;
- Data assessment and evaluation in collaboration with subject matter specialists; and
- Data analysis and product development.

2.2.9   Corporate services

This workstream was responsible for all aspects of human resource management, supply chain management, contracting and financial management, transport, security, accommodation, and legal services. The human resources sub-workstream recruited, appointed, paid and terminated contract staff. This workstream also managed permanent staff deployment to provinces and districts during Census field operations.

2.2.10   Publicity, Community Mobilisation and Advocacy

The Census 2022, being the first ever digital census to be conducted in the country, required a more elaborate communication strategy entailing the extensive use of technology to reach out to various audiences and adopting new media, such as social media, online and mobile communication. The Publicity, Community Mobilisation and Advocacy (PCMA) workstream was responsible for the development and implementation of a communication strategy that encompasses among others, educating communities about Census 2022 so as to see the value of participating or providing information. Other objectives included:

- Coordination of internal and external communication activities to ensure awareness of the Census 2022 project by key stakeholders;
- Creation of effective media relations and use relevant advertising that reach and call targeted audiences to action;
- Conduct publicity and advocacy campaigns supporting recruitment drive, stakeholder partnerships, educating the public about their participation and issue of data privacy; and
- Promoting greater buy-in at community level.

2.2.11 Secretariat

The Secretariat workstream was responsible for Census 2022 documentation on methodologies, instruments/data collection tools, processes and procedures.

2.2.12 Project Governance

The Governance workstream was responsible for ensuring that the Census 2022 was managed with care and integrity and the culture of the good governance was practised by all teams. Its objectives were, among others, to coordinate and facilitate the overall development and monitoring of Census 2022 risks and their mitigations; ensure coordination of Census 2022 Internal Audit planning, execution and reporting; and coordination of efforts on compliance activities and provide advice on governance related matters. The workstream also advocated for compliance and adherence to Census 2022 project activities to set standards and requirements.

A sub-committee to assess project risks and associated mitigation plans was created within the Project Governance workstream. A governance and risk management model was developed to guide the processes for the Census 2022 project. Outputs included consolidated project risk registers, mitigation plans and progress reports to be used as monitoring and compliance tools in relation to the identified project risks. Governance workstream further initiated collaborative efforts with Financial Internal Control (FIC) to assist with an oversight role on compliance by performing quality assurance functions in line with the normal operational activities of these units.

2.2.13 Provincial Integration and Quality Assurance

Provincial Integration and Quality Assurance (PIQA) workstream was created to ensure the efficiency and effectiveness of Census 2022 operations by coordinating and integrating Census 2022 project operational plans and activities across provinces, and ensure quality outputs, particularly at the field level. Among others, the workstream sought to facilitate the finalisation and approval of the Census 2022 project structures at provincial and district levels in support of integration of operational activities across provinces, districts and the head office. It ensured that resources needed for Census 2022 were distributed in accordance with the workload for each province and district offices. PIQA also managed monitoring, quality assurance and oversight of Census 2022 activities in the provinces and facilitated a coherent and consistent approach for timeous communication and implementation of project decisions across provinces.

2.2.14 Census dissemination workstream

This workstream is responsible for development census dissemination plan and coordination of census launch activities in liaison with relevant internal and external key stakeholders. The dissemination plan is designed to create and maintain relationships with target stakeholders for the purpose of achieving organisational and Census 2022 strategic objectives. It encompasses a process of identification, consultation, product conception, promotion and distribution of those products and services in a way that satisfies both individual and organisational stakeholder objectives.

2.3   Census 2022 project planning and implementation committees

The success of any census does not only depend on how the various activities are implemented and monitored, it also relies greatly on governance bodies/committees which enhance the efforts of the teams that are mandated to execute the activities. It is in this regard that the following oversight and advisory bodies/committees were set up and required to assist with monitoring the project processes and implementation.

2.3.1   Technical committee

The purpose of this committee was to coordinate discussions and approval of Census project documents pertaining to planning, processes and methods presented by project workstreams.

2.3.2   National Advisory committee

The Census 2022 National Advisory Committee (NAC) was inaugurated in November 2020 to serve as an oversight body to advise and assist Stats SA to deliver a historic technology-driven census with improved coverage and response rates.

2.3.3   Project Steering committee

Stats SA's Executive Council (EXCO) played the role of Census project steering committee. The steering committee assisted in the monitoring and implementation of various census value chain activities and phases. This committee was the primary decision-making body with a strategic mandate of ensuring the alignment of census project with expectations from internal and external stakeholders.

2.3.4   Rapid Response Committee

The purpose of this committee is to provide a forum for process owners to discuss urgent interventions to Census processes, procedures and methodologies.

2.4 High level Census 2022 process flow

The Census 2022 high-level flow is presented in Figure 1 below.

Figure 1: High level Census 2022 process flow
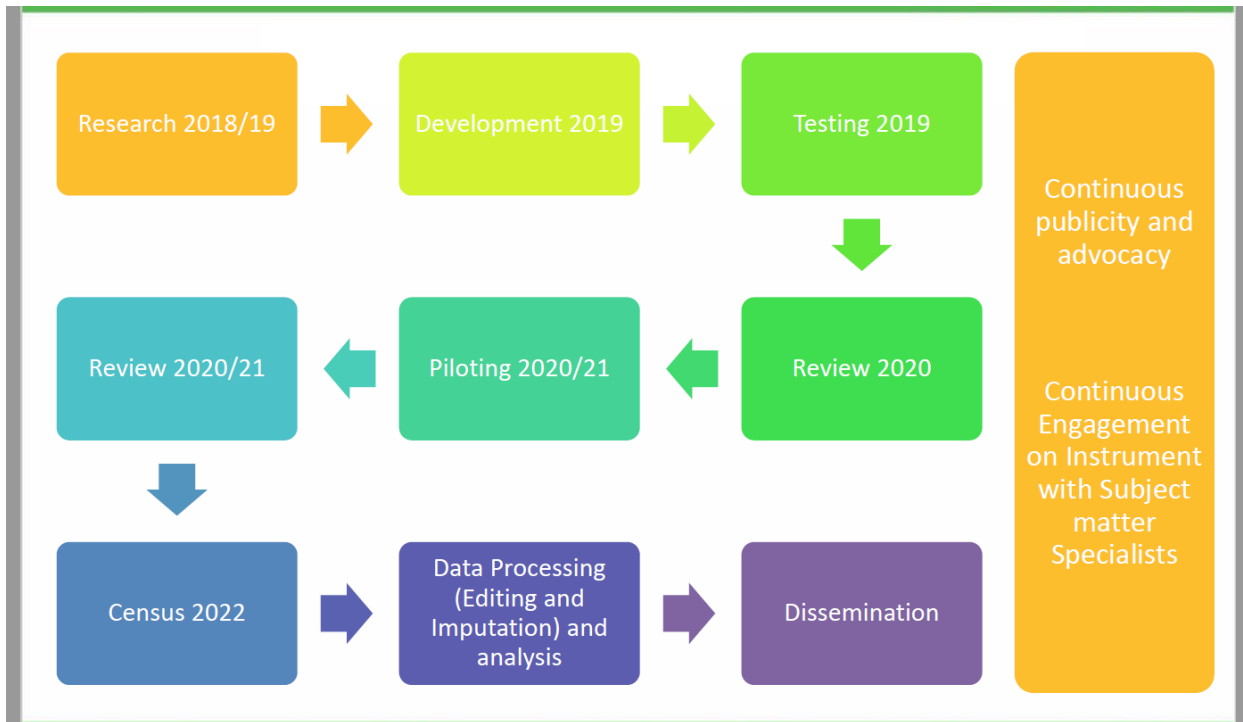
Planning for Census 2022 commenced with research on the use of multi-mode data collection approaches, followed by development of Census content, methods and systems, all which were subjected to testing their practicality, relevancy and user friendliness. All key Census phases were planned and fully tested, with revisions made to ensure successful implementation during the main Census.

3. CENSUS 2022 HIGH LEVEL PLANNING PHASES

Figure 2: High level planning phases over the five-year period: 2018–2022
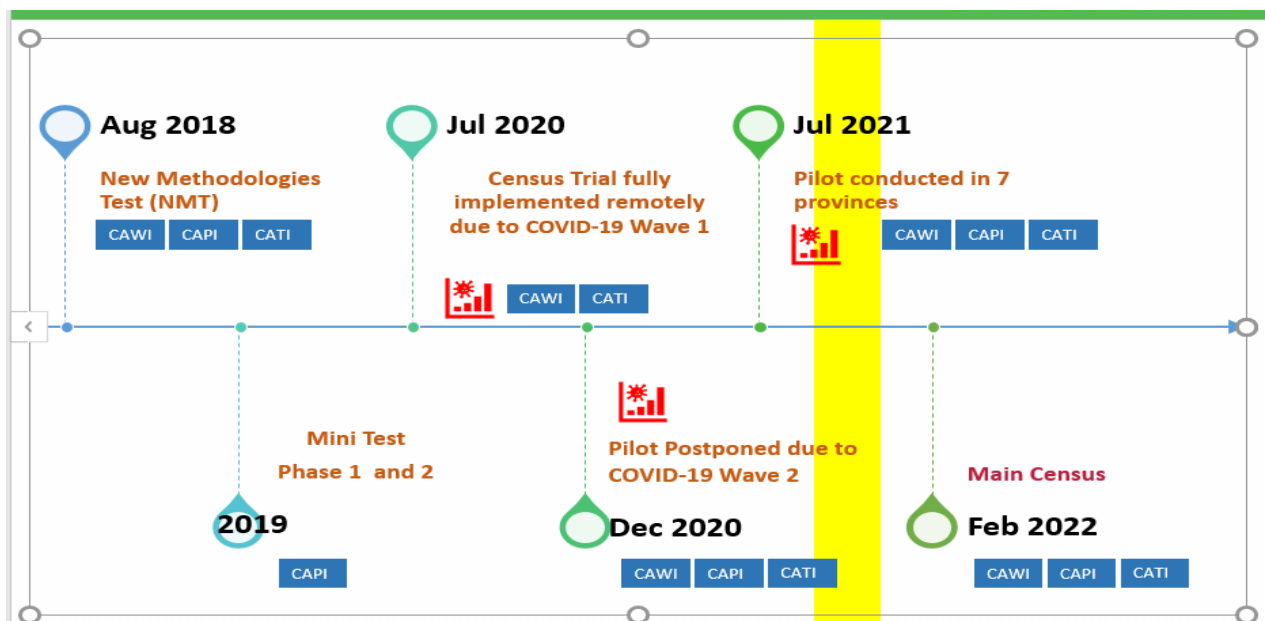
Census 2022 was part of the overall organisational strategy to shift from paper to digital survey/census operations as a measure to improve cost effectiveness, timeliness of releases, optimise contact with households and improve coverage, relevance and quality. This new approach of a technology driven Census was thus characterised by a series of tests to ascertain the feasibility and viability of proposed new approaches, methods, applications, processes and systems over a period of five years as shown in Figure 1.

The first test was conducted in August 2018, which aimed at establishing respondent's preferred mode of data collection. Two mini tests were conducted in 2019 to achieve a number of objectives including testing the newly developed in-house data collection applications, implementation of multi-mode data collection methods, data collection timeframes and digital census geography frame (GIF), its adequacy, completeness and navigational facilities.

The census pilot was scheduled to take place in October 2020 and the main Census in 2021. However, due to COVID-19 disruptions, the pilot was moved to 2021 and main census to 2022. With the advent of the COVID-19 pandemic, another test termed Census Trial was conducted in 2021 to test remote data collection methods and virtual training methodology in the COVID-19 environment in preparation for Census Pilot.

Census Pilot was conducted in two phases. The first phase covered only remote data collection methods of CATI and CAWI, due to level 4 lockdown which was announced on 28 June 2021 and lasted for 28 days. All nine provinces were included in the first phase of the enumeration. The second phase of the data collection covered the CAPI (face to face) data collection following the downgrade of COVID-19 level 4 lockdown to level 3. Seven provinces participated in the second phase of the data collection with the exclusion of Western Cape and KZN due to the state of the pandemic in the former and violence in the latter province. Furthermore, due to the COVID-19 environment and its associated challenges, the population in institutions and homeless population were not enumerated in the Census pilot.

4. CENSUS 2022 PRE-ENUMERATION OPERATIONS

4.1 Publicity and advocacy

Census 2022 publicity campaigns were driven by Publicity, Community Mobilisation and Advocacy (PCMA) workstream. The team developed a communication strategy that encompassed multi-pronged outreach strategies to reach all sectors of society.

Noted during planning was that consumption of print media had declined immensely due to increased access to digital devices which audiences use to access content. There was thus a focus on increased use of digital media. Most platforms used to inform, educate, and mobilise society were radio and television, driven by the audience segmentation as outlined in the government segmentation model developed by GCIS. However, print media was used to reach out to some communities that still preferred to consume media through print as well as those who had challenges with access to digital media. The communication and outreach approach adopted for Census 2022 was planned and implemented focusing on three main phases: Pre- enumeration phase, during enumeration and post enumeration.

The core messages crafted for Census 2022 were in line with the mandate of the organisation and sought to ensure that every sector of society participated in Census. During census data collection, a fully functional call centre was established to enhance swift response to the public's census related queries.

Central to Census 2022 publicity campaigns was the branding theme, ***"#Get Counted,"*** which was adopted to depict a digital census. Throughout the census phases, Census 2022 brand was very visible on Stats SA's website. The Stats SA website was one of the effective outreach tools visited by different users. The website was adopted as a key census communication platform. Inauguration of countdown activities including "100 days countdown to Census day," was key in engaging the public on the days remaining before the big census exercise. The landing page also had pop-up message directing visitors to the Stats SA recruitment site to know about recruitment timeframes and most importantly application closing dates. The webpage content also included information about the census and its objectives to encourage maximum participation in Census.

4.2 Census Geography frame

One critical aspect to a successful census is a complete, accurate and current geographic frame. The organisation adopted a digitally enabled approach, which required an accurate and up-to-date digital Geospatial Information Frame (GIF) to enable a seamless digital collection and dissemination of data. GIF refers to a spatially referenced hierarchy of geographical features (points and polygons) comprising of dwelling frame, Enumeration Areas (EAs) and place names. In this geography frame, Geo-point and unit records uniquely represent dwellings, while polygons represent geographic hierarchies such as Enumeration Areas (EA), place names, local municipality boundaries, etc.

During intercensal period, there was a continuous update of frame to accommodate changes in settlement status. The updating of frame has been ongoing using data from a number of sources including municipal data, Census 2011 listing records and Spot Building Count using the web-mapping tool. Central to having georeferenced dwelling frame in preparation for Census 2022 was the verification and unpacking of all structures prior to commencement of enumeration. The unpacking process took place in the office and field environment. Both processes involved recording of all dwelling units and other structures attached to a geo-point within a yard, stand, segment, block or cadastre.

As per geography hierarchy depicted in Figure 3, Stats SA is responsible for the development and management of main place, sub-place, enumeration area, small area and dwelling unit layers. The Enumeration Area (EA) was one of the critical geography layers for resource planning and allocation. The process of EA layer demarcation formed a critical input into all upstream and downstream Census activities and was done according to approved specifications. The main aim was to demarcate EAs in such a manner that an enumerator can walk or drive from one geo-point to the next, covering the entire EA with ease.

Census 2022 geography frame preparation did not alter the 2011 EA boundaries; but rather sub-divided 2011's EAs that had grown over the 10-year period. The sub-divided 2011 EAs were termed Sub-EAs and utilised to plan and implement data collection.

Figure 3: The Geospatial frame layers
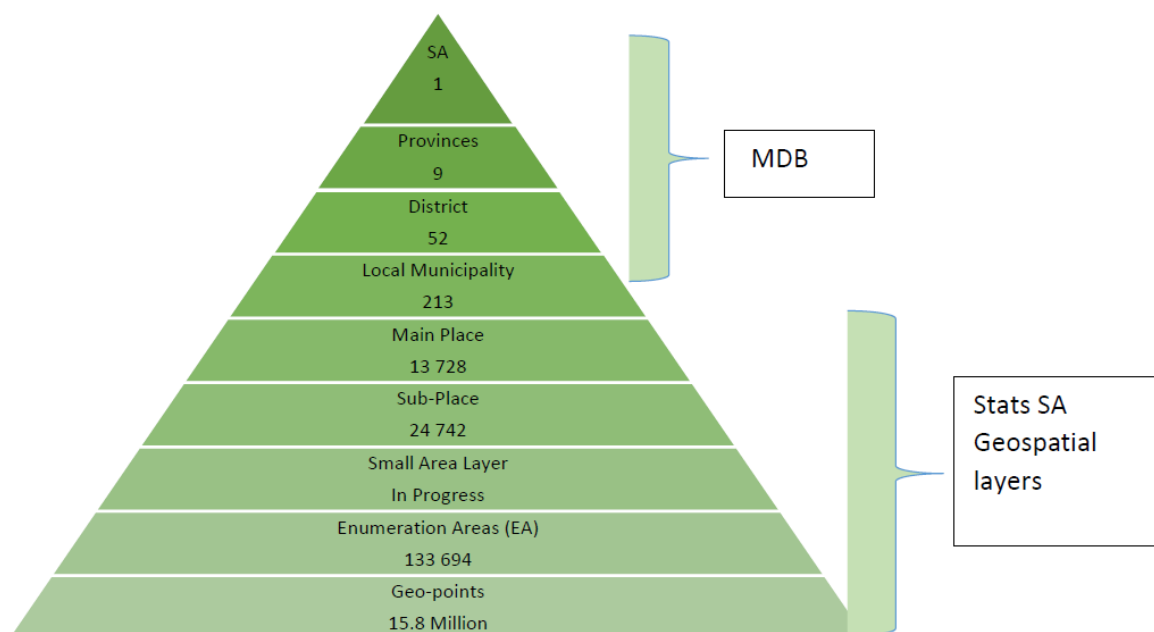
4.3   Online registration and self-unpacking

To prepare the geography frame for multi-mode data collection, online registration was introduced, where the public registered their address and detailed information on their residential structures. The registration process was aimed at those who were interested in completing the online census questionnaire (CAWI households) and those who wanted to be enumerated telephonically (CATI households). Identifying the location of households interested in completing the online Census questionnaire in advance to facilitate the process of creating a unique link with security code was a critical step. The process of online registration also entailed self-unpacking by respondents including providing information on their contact details (Cell phone or email address). The process of registration was also open for Special Dwelling Institutions (SDIs) that opted for online submission of the required information.

4.4   Systems/Applications Development

A number of systems/applications were developed in-house as part of planning for a successful digital census.

- The recruitment application was integrated with google maps and coordinates. The application assisted in ensuring that Census contract staff applicants were identified within areas they resided. Use of the application enabled Stats SA to identify gaps in terms of areas with no applicants or few applicants in advance and mitigate the challenge through targeted recruitment.
- The online fieldworker verification system was created to enable the identification of field staff for security reasons, by displaying images of field staff, supervisors, areas of work and contact details on the website, this online function allowed respondents to check and verify the field staff before giving them access to their homes and providing confidential information on the Census questionnaire.
- The Corporate Services Systems which Integrated various functions including HR Database, Recruitment, Appointment and Payment system, Data Collection Systems for the three data collection modes of CAPI, CATI, and CAWI, Field Logistics Operations Systems (FLOS), Payment Trigger System (PTS) and PERSAL System (Bulk Appointment and Bulk Payment Processes).
- The self-registration and self-unpacking system for respondent to register their residential structures and number of households residing at that particular address.

  Data Collection applications for the three data collection modes of CAPI, CATI, and CAWI were developed and successfully used during Census 2022 enumeration. The online Census questionnaire was developed with in-built data security measures to enable secure online-household questionnaire access and completion using cell phone, tablet or computer.

- Special Dwelling Institutions (SDI) that registered online to submit information on persons and services had access to a system where they could complete required census information and send to Stats SA with ease. The system enabled the institution to indicate the institution address, contact person and number of persons that spent Census night at the institution.
- Dashboard reporting system that enabled progress reporting on the number of households and SDIs enumerated per day throughout the enumeration period.

4.5   Recruitment of Census staff

Recruitment remains one of the activities that impacts critically on project performance, with specific reference to census. It is of paramount importance that it is done efficiently and effectively within the prescribed frameworks, policies and procedures. This section focuses on the recruitment, appointment, remuneration and termination of Census 2022 Project contract staff.

There were a number of key considerations in relation to the recruitment of contract staff for the Census, among these was the requirement for FWs and FWSs to be recruited strictly from within the EAs in which they live and will work for the duration of the project. However, targeted recruitment was conducted in areas where there were few or no suitable applicants, particularly high-walled areas, farms and small holdings, etc. Competency exercises were used as selection criteria upon completion of training. Bulk appointment and payment was done through PERSAL.

All potential candidates who met the minimum selection requirements were selected for training and all trained candidates were assessed through the Learner Management System. Training scores were used to appoint successful candidates. In instances where there were no suitable candidates identified in an EA during training, successful candidates from adjacent EAs were considered. During training, reserves were also identified to cater for the replacements where people declined the offer, resigned or absconded. Critical to the recruitment process was setting of selection criteria for enumerators including, Matric certificate or equivalent, proficiency in relevant computer applications such as Microsoft Office, proficiency in English and at least one or two South African official languages.

Recruitment of Census contract staff commenced with the compilation of HR database that started in 2019 with the aim of having a pool of potential candidates in preparation for Census 2022. Although the database is used for all survey projects, it was updated on a continuous basis. Figure 4 below shows the recruitment process for Census contract staff.

Figure 4: The process flow for recruitment of field staff

PROCESS FLOW FOR RECRUITMENT OF FIELD STAFF
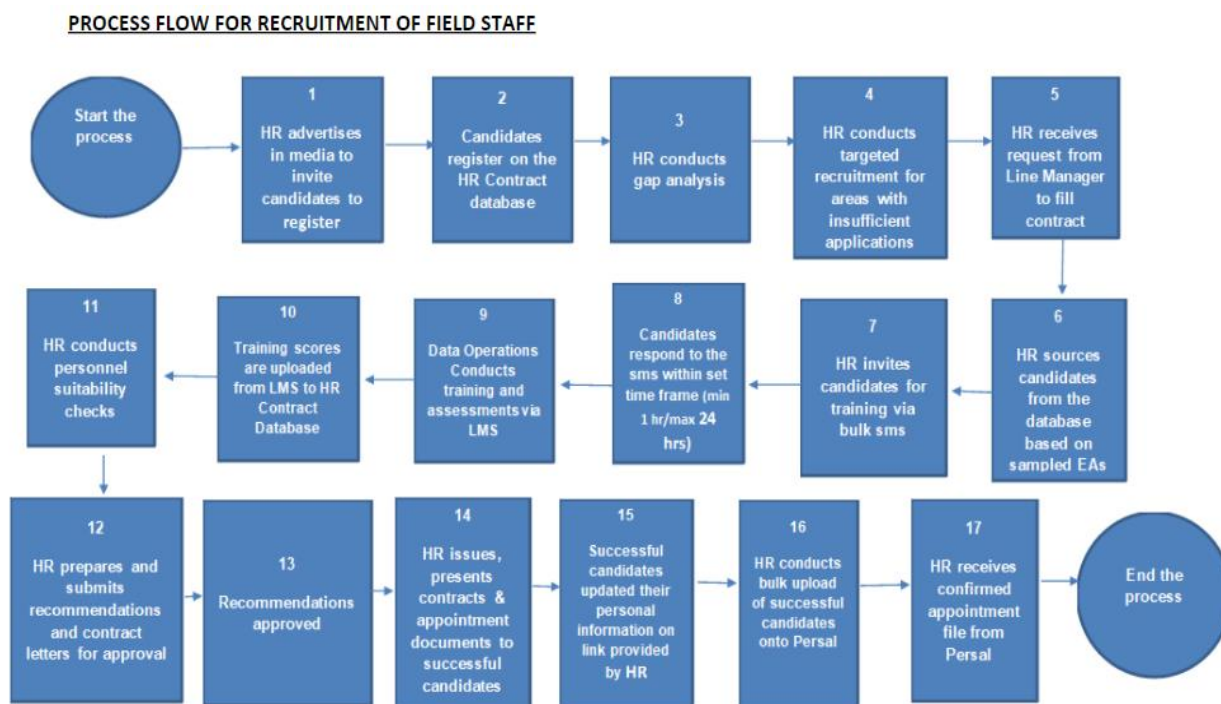
The recruitment strategy for Census 2022 also highlighted the importance of recruiting youth and women (50 per cent youth and 50 per cent female as contract staff). The recruitment plan also highlighted the need to over-recruit by 10 per cent for training purposes – in cases where trained candidates declined the offer, some of the appointed resigning during the course of enumeration or others absconding. Despite all these plans, certain provinces could not meet these targets. This resulted in extension of the enumeration period beyond the initial planned completion dates. In some instances, in order to address the shortages of field staff in various provinces, those who had completed their work were taken to other provinces to assist in finalising data collection. In instances of resignations, Human Resource Management and Development was notified and replacements from reserves followed. Terminations due to misconduct and poor performance were dealt with through the disciplinary process as prescribed by the Labour Relations prescripts (Resolution 1 of 2003). During census operations, a number of permanent staff with experience in surveys and census operations were also deployed/seconded to the Census project at province and district levels depending on need.

4.5.1   Finance and payment of contract staff

The processes of paying census contract field staff were largely automated with some exceptions. Candidates' information was verified including their banking details. On terminations, contract end dates per category were captured and verified on the payment system.

4.6   Census questionnaire development and testing

4.6.1   User needs determination

The key process of developing a census data collection tool is extensive consultations with data users. During the process of determining which data items will be included in the Census 2022 questionnaire, current and potential data users were consulted for their views on the type of information the Census should provide.

The main factors that determine the topics to be included in the Census 2022 questionnaire included the national priorities of the country, historical comparisons, international comparability, suitability of topics for collecting reliable information and the resources available for the Census. Furthermore, the length and complexity of the questionnaire were also taken into account. In the *Principles and Recommendations for Population and Housing Censuses, Revision 3*, it is emphasised that a census should not collect information that is no longer required only because it was collected in the past, but rather focus on key demographic, social and socio-economic variables that are relevant to users. It is against this background that Stats SA conducted user consultations to determine the data items to be collected via the Census 2022 questionnaire.

4.6.1.1 Overview of user consultation process

In order to ensure that a wide range of stakeholders have the opportunity to participate in the user consultations, various stakeholders were consulted through varied platforms in an attempt to reach existing and potential data users. The objectives of user consultations were (i) to solicit inputs on the proposed Census 2022 data items, (ii) to understand user needs in terms of data required (iii) to strengthen partnerships with users and other stakeholders and (iv) to assist users in understanding what data the Census can and cannot provide and (v) to provide advice on alternative data sources.

Internal stakeholders representing various chief directorates at Stats SA with expertise in the various areas that the questionnaire covers were consulted.

National departments are key users of census information. Their strategies, planning and monitoring of basic services such as housing, provision of electricity and education as well as employment and population dynamics are often based on data collected from the census. The discussions with departments provided an opportunity to make input and recommendations on the data items to be included.

The provincial user consultation workshops took place at various venues in the provinces. Participants included government officials from provincial government, representatives from municipalities, academics, researchers, analysts and representatives from the private and NGO sector. Stats SA provincial staff and district managers also attended the workshop.

In an attempt to broaden the reach to census data users, an invitation for users to provide their comments on the recommended core questions for the 2020 round of Censuses as well as their specific data needs was made available on the Stats SA website. The online submission tool allowed for users to comment on each data item as well as make general suggestions.

The inputs from the user consultations were discussed by census management and amendments were made prior to the Census 2022. The decision to revise questions, response categories and the addition of new questions was carefully considered. Factors such as comparability, resources available, understanding of questions and respondent burden were taken into account. Engagement with subject-matter specialists at Stats SA, researchers, academics and national departments on specific census themes continued until the Census Pilot was conducted. Behind-the glass tests, door step tests, mini tests and the pilot test all provided useful input on the suitability of the questionnaire for Census 2022.

4.6.2 Questionnaire types and development

South Africa conducts a *de facto* census, meaning that the population is counted based on where they were found on census reference night (midnight of 2–3 February 2022). Based on the location of persons on the reference night, four sets of questionnaires were developed for Census 2022, each set administered to a targeted group. The four groups identified were:
- The population that live as households
- The population living in special dwelling institutions (SDIs)
- The transient population - persons who were on the move during the Census night and did not return to their normal place of residence the following day (e.g. persons leaving the country on Census night)
- The homeless population – persons with no form of shelter on the reference night and no known residential address, including those who were accommodated in tents erected during the COVID-19 pandemic.

The paper questionnaires were designed using Microsoft Word and used as a reference for the development of the electronic questionnaires using the in-house data collection application developed by CEESD. The data items for each questionnaire are shown in Tables 1–4. The approved questionnaires were developed electronically depending on the three modes of data collection used, face-to-face interviewing via Computer-Assisted Personal Interview (CAPI), online self-enumeration via Computer-Assisted Web Interview (CAWI) and telephonic interviewing via Computer-Assisted Telephonic Interview (CATI). In addition, the SDI schedule was developed using Microsoft Excel.

Given that the population living in households forms the basis for planning, the household questionnaire was the longest (see Table 2). It was designed to collect comprehensive information on individuals in the households and household goods and services, agriculture, food security and mortality. The SDI data items are shown in Table 3. The SDI questionnaire was developed to target the population living in collective living quarters (people that spent Census night, midnight of 2–3 February 2022, at boarding schools, prisons, hotels, hospitals etc.). In addition, two very short questionnaires were developed for enumerating the transient and

homeless population. Basic information was collected on these two groups (see Tables 4 and 5). The transient questionnaires were used at points of entry and exit such as airports, harbours and border posts. The homeless questionnaires were used at homeless hot spots on Census night.

Table 1: Data items for household questionnaire for Census 2022

Questionnaire section	Data item
Cover page	Particulars of the dwelling
	Particulars of the household
	Contact details of respondent
	Final result code
Demographics	Sex
	Date of birth
	Age
	Relationship to head of household
	Marital status
	Population group
	Language
	Religious affiliation/belief
Migration	Citizenship
	Province of birth
	Country of birth
	Year moved to South Africa
	Usual residence
	Province of usual residence
	Municipality of usual residence
	City/town of usual residence
	Movement since 2011
	Year and month moved
	Province of previous residence
	Municipality of previous residence
	City/town of previous residence
	Main reasons for moving
General health and functioning	Difficulty in seeing
	Difficulty in hearing
	Difficulty in communicating in his/her usual language
	Difficulty in walking or climbing stairs
	Difficulty in remembering or concentrating
	Difficulty with self-care such as washing, dressing or feeding him/herself
	Assistive devices
Parental Survival	Mother alive
	Father alive
Education	Attendance at an Early Childhood Development Center (ECD)
	Attendance at an educational institution
	Highest level of education
	Field of education

Questionnaire section	Data item
Income	Income category
Employment	Employment status
	Farming/animal product
	Other work
	Looking for work
	Available to work
	Occupation
	Main tasks and duties
	Main activity of place of work
	Main goods or services
	Employer/employee type
Fertility	Ever given birth
	Age at first birth
	Total children ever born
	Total children surviving
	Total children no longer alive
	Date of birth of last born child
	Sex of last child born
	Last child born alive
Housing, household goods and services	Type of living quarter
	Type of main dwelling
	Tenure status
	RDP/government subsidised dwelling
	Access to piped water
	Source of water
	Reliability of water supply
	Water interruptions longer than two conservation days
	Toilet facilities
	Energy/fuel for cooking
	Energy/fuel for lighting
	Refuse disposal
	Household goods and services
	Internet services
Agricultural activities	Type of agricultural activities
	Current quantity of livestock/poultry
	Main purpose of agricultural activities
	Place of agricultural activity
Food security	Adult hunger
	Child hunger
Mortality	Death occurred
	Number of deaths
	Sex of the deceased
	Age of the deceased
	Year and month of death

Table 2: Data items for special dwelling institutions (SDIs) questionnaire for Census 2022

Questionnaire section	Data item
Cover page	Particulars of the SDI
	Name of the Institution administrator
	Capacity of the Institution
	Total number of persons in institution
Demographics	Sex
	Date of birth
	Age
	Population group
Basic services	Access to piped water
	Source of water
	Toilet facilities
	Energy/fuel for lighting
	Refuse disposal

Table 3: Data items for homeless questionnaire for Census 2022

Questionnaire section	Data item
Cover page	Physical identification of homeless location
	Type of homeless location
Demographics	Sex
	Date of birth
	Age
	Marital status
	Population group
Migration	Citizenship
	Province of birth
	Year moved to South Africa
Education	Highest level of education
Reason for homelessness	Main reason for homelessness

Table 4: Data items for transient questionnaire for Census 2022

Questionnaire section	Data item
Cover page	Physical identification of transient location
	Type of transient location
Demographics	Sex
	Date of birth
	Age
	Marital status
	Population group
Migration	Citizenship
	Province of birth
	Year moved to South Africa
	Province of usual residence
	Municipality of usual residence
Education	Highest level of education

After drafting of paper questionnaire, requirements for electronic questionnaire versions of all four questionnaire types were defined and provided to the CEESD. The electronic questionnaires were subjected to various tests before finalization.

4.6.3   Questionnaire Translations

South Africa has 12 official languages i.e. English, Afrikaans, Tshivenda, Xitsonga, Sesotho, Setswana, isiZulu, isiXhosa, isiNdebele, siSwati, Sepedi and Sign Language (Sign language was declared one of the official languages in 2023). One of the initiatives for Census 2022 was to cater for language diversity within the country. The questionnaire translation exercise was only done for the household questionnaire while the transient, SDI and homeless questionnaires were excluded. Different groups consisting of employees from both provincial and head office were constituted to fulfil the translation task. The translation focused on CAPI and CAWI mode of collection. To ensure quality of the translated questionnaire in the various languages, a number of processes were undertaken both in the field and off field including testing during the Census trial test conducted in 2020, during Pilot Census conducted in 2021 and through quality improvement exercise.

On CAPI translations, the plan was to have all versions of 11 of the 12 official languages on the system for the fieldworker to select the preferred one. Due to time constraints, the CEESD delayed putting all the versions on the system. To circumvent this a PDF format of the 10 official languages was provided as documents on the tablet used by fieldworkers during enumeration. The challenge with PDF document was that it was difficult to open two screens while conducting enumeration.

All four questionnaire types were presented to census management for approval and sign off after the pilot Census was conducted.

4.7   Training on Census processes, content and methods

Training strategy was initially to implement a three-tier cascade face to face training approach. However, due to COVID-19, some plans had to be changed including change of training from face to face to virtual training. During Census pilot national training, about 12 positive cases of COVID-19 were identified at a training venue in Pretoria where about 180 Stats SA employees had gathered to attend the training. This incidence led to the indefinite suspension of pilot census training.

Subsequently, the organization recommended that large-scale physical attendance training sessions were to be avoided. This was a major shift to be implemented amidst preparations for a census project of such great magnitude.

Census training was conducted at three levels (national, provincial and district level) to equip field staff with relevant and adequate skills to effectively conduct face-to-face and telephonic interviews. This was done to ensure that accurate and reliable data are collected, and respondents' queries were handled correctly during enumeration period. Each level of training was allocated ten days.

Training involved use of subject matter specialists selected from relevant work streams and Chief Directorates. These conducted training at national level to provincial trainers who in turn conducted training at provincial level. Those trained at provincial level conducted training at district level. The overall training strategy entailed the following:

- Virtual instructor-led, web-based training in recorded live sessions, made accessible to trainees with materials and files, through MS Teams;
- Thorough training on the content and COVID-19 safety regulations adhered to during training and fieldwork;
- Loading all training materials and manuals on all gadgets/tablets accessed during training and enumeration for reference purposes;
- 10% over-training to cater for resignations during enumeration;
- All trainers and trainees tablets prepared and configured before the start of training and;
- Training evaluation exercises and assessments administered online via Learn Management System (LMS).

4.7.1   Training challenges

During Census 2022 training, a number of challenges were experienced, ranging from poor network and connectivity and load shedding, to late replacement of trainees due to the vetting process. Network issues also caused delays in training. With the introduction of virtual training, some of the trainees had difficulty in grasping the concepts, technology and content. It was also difficult to manage and track attendance and level of understanding. This is in contrast to the face-to-face training, which could have been more interactive. In addition, there were no field visits as part of practical training and this may have limited trainees in understanding of project concepts and content.

5.   ENUMERATION

5.1   Enumeration approaches

The enumeration strategy for Census 2022 outlined a multi-mode data collection approach. Census 2022 planning ensured that enumeration procedures, approaches and methods were appropriate for a complete population and housing count without any omissions or overlaps. Census 2022 was affected by unprecedented challenges including the COVID-19 pandemic and climate change issues such as flooding in some parts of the country, and these affected the Census 2022 data collection which was conducted between February and May 2022, with provinces completing enumeration in a staged manner over this period.

The EAs were designed to allow fieldworkers to complete enumeration within the enumeration period. Fieldworkers were to visit all geo-points in their respective EAs. Fieldworkers were to confirm and/or update required information accordingly. Among other core responsibilities, FWs were expected to do the following before commencement of the interview:

- Follow the systematic way of moving from geo-point to geo-point without omission;
- Identify the boundaries of the EA;
- Identify a starting point on the map;
- Plan the route to follow to ensure that no structures were missed;
- Navigate to the starting point;
- Capture location for each household and enumerate, keeping to the left until the whole EA was completed;
- Update new structures without geo-points and DU feature classification before enumeration; and
- Enumerate households using household questionnaire.

According to the enumeration plan, a FOO was to have five supervisors, and each supervisor, five fieldworkers, making it a total of 25 fieldworkers under an FOO. Due to COVID-19, it was a challenge to secure sufficient field staff as per stipulated ratios in the strategy. This led to changes in workload allocation for the purposes of complete coverage.

5.1.1   Multi-mode data collection procedures

Population in private dwellings, SDI, the homeless population and transients were enumerated using CAPI as the main mode of data collection. In the case of private dwellings and SDI, where respondents opted to self-enumerate, CAWI methodology was applied. That is, households had to register in advance and indicate that they wanted to complete the Census questionnaire online. Thereafter, they were given links to access and complete the online questionnaire using either a smart cell phone, tablet or computer at no cost to the respondent.

All households in DUs within an allocated EA that did not register for self-enumeration or telephonic interviews were to be visited and counted through face to face interviews.

Face to face interviews (CAPI)

The FW conducted enumeration in their assigned EA by visiting all geo-points on the maps uploaded on their tablets as their assignment. To get to each point, the FWs tapped on the geo-point on the gadget, with the help of the navigation tool. The FWs then had to verify the geo-point and update information where necessary (DU and household unpacking). They then activated CAPI to generate the number of questionnaires that match the number of households and enumerate accordingly.

Telephonic interviews (CATI)

All households that opted for telephonic interviews during registration, received a confirmation of registration and a message to indicate that they will be called by a Stats SA representative on specific dates. A database was compiled and the CATI workload was distributed among the CATI enumerators.

Self-enumeration (CAWI)

Respondents who registered for self-enumeration were able to access the CAWI system (online questionnaire) on their personal computers and smart devices. A questionnaire link was sent to the respondents on Census day through their cell phone numbers or email addresses that they provided during registration. Reminder messages were sent to respondents who had not completed the questionnaire by the fifth day of data collection. Respondents who completed the online questionnaire were sent a reference number that they were expected to show to FWs as confirmation of enumeration. Those who had not completed the questionnaire by the stipulated date were automatically treated as CAPI respondents. FWs visited all CAWI respondents to capture reference numbers.

Gated communities that did not initially register for self- enumeration were provided with a CAWI questionnaire during enumeration.

5.2   Methodology for enumerating the population in institutions

The procedures for enumerating this population sub category included SDIs submission of administrative records in line with the reference period (Census night) or FW enumerating population in SDIs in their respective EAs. In instances of the latter approach, procedures similar in enumerating persons in DUs were applied. SDIs were given opportunity to provide administrative records physically to FWs or submit to Stats SA online. Those that selected online questionnaire completion during their registration were to receive links for access of online SDI questionnaire completion and submission.

Some institutions chose to submit excel spreadsheets of persons that spent census night in the institution. They were received and transcribed into SDI application system.

5.3   Methodology for enumerating the homeless population

A number of enumeration procedures aimed at complete count of the homeless were planned for and implemented in Census 2022. Prior to enumeration, provincial offices identified homeless hotspots in liaison with Non-Government Organisations (NGOs) that work with the homeless. The planning for the homeless required all district offices to provide the list of EAs with identified hotspot locations. This information assisted in work allocation for counting homeless persons on Census night. During enumeration, which was conducted on Census night, the sweeping method was applied.

The outlined procedures to conduct enumeration of the homeless included the following steps:

- Navigate to the homeless hotspot;
- On arrival at the hotspot, the FWs introduce themselves to the homeless persons and requested to administer the Homeless questionnaire using CAPI mode. If there are still other hotspots to visit, the team moved to the next hotspot until all were covered;
- FW Login into the gadget;
- The FWS generate multiple homeless questionnaires based on the number of homeless individuals at the hotspot. The team members (FWs) then login to access questionnaires and enumerate homeless persons; then
- All FWs, FWSs and team members thereafter synchronised questionnaire after each interview.

During Census night, however, there were system glitches that were experienced and this made it impossible to cover all homeless hotspots and enumerate this population sub category. In consultation with methodologists, a re-count of homeless persons was arranged during the course of the Census enumeration period. Each province arranged and communicated dates when a recount was to take place.

5.4   Methodology for enumerating the transient population

Transient Population refers to persons who were on the move during Census night and were not going to their usual place of residence the following day, e.g. truck drivers, persons leaving the country after midnight of the reference night, etc. Key points were considered in relation to counting this population category. Most importantly, consultations with key stakeholders prior to enumeration were made to ensure access to ports of entry.

The outlined procedures to conduct enumeration of transients included the following steps:

- Transient population spots were identified before the reference night and staff members to enumerate this group were identified. Transient spots included international airports, border posts and overnight truck or bus stops.
- Prior arrangements were made in order to be given permission to access and conduct interviews at both international airports and border posts. On Census night, teams responsible for enumerating this group contacted the relevant authorities.
- Navigated to the transient spot e.g. truck overnight stop, bus stop locations

- On arrival to the spot, the FW checked/verified if there was indeed a transient spot. If there was a point, the FW logged in to the tablet and selected the point and thereafter launched the questionnaire for transients.
  - o In instances where there was no geo-point, FWs generated the point and allocated correct feature category of "Transient spot". Thereafter, FW synchronized the record in order to receive a "Transient spot number" generated by system and launch the transient questionnaire.
- FWs were expected to generate multiple transient questionnaires based on the number of individuals in transit at the spot and to conduct the interview.
- All FWs were expected to synchronise completed questionnaires after each interview.

5.5   Quality assurance procedures and processes

The first three days of fieldwork were characterised by fieldwork monitoring to identify potential problems with FWs and address them immediately. This was to ensure that FWs had navigated and visited the correct geo-point, launched the right questionnaire type, updated/unpacked correctly, captured the correct number of DUs and households, administered the questionnaire correctly, and adhered to COVID-19 protocols. Quality Assurance (QA) during the enumeration phase was thus two-fold; through in-built quality assurance mechanisms within electronic data collection tools and during data collection. The in-built measures included implementation of validations in the electronic questionnaires to eliminate possible skip errors during questionnaire completion. Field QA on the other hand involved the checks and verifications that were conducted by FWSs and FOOs during data collection.

After every interview, the respondent received a system-generated reference number that played a dual role i.e. as a quality control measure and also to allow linkage with the post enumeration survey (PES) which seeks to estimate the levels of under or over counting. The FW/CATI enumerator informed the respondent to keep the reference number safe. In the absence of a valid cell phone number, the FW passed on the reference number to the respondent to keep. For CAWI respondents, the generated reference number was to be shown to the FW to confirm that the household was counted through CAWI or CATI.

As part of addressing quality issues during enumeration, a query log on technical issues was compiled and escalated to the relevant work streams to provide responses timely.

5.6   Enumeration progress tracking

Fieldwork monitoring was conducted through visits to selected locations and field teams, and tracking the work through the dashboard. The heads of workstreams visited a few places. Provinces were using the dashboard to zoom into the areas that were lagging behind. The dashboard provided diverse information; where enumerators have been, and if people were not working, management saw places coloured by red points which meant the areas were not touched yet; yellow or orange were safer and green was the acceptable colour in terms of EAs that were on track to finish on time. There was also a dashboard for management. DCCs were to monitor FOOs, PCCs and the districts. The DCC unit had FOO units and supervisors to monitor directly. Some of the FOOs were not supplied vehicles, though they had 10-15 fieldworkers.

6.   DATA PROCESSING

The data processing phase involved retrieving data from the cloud, conducting data structuring, integration, editing and coding. During data collection, data was saved and stored in the cloud environment. Upon completion of data collection, data was moved to on-premises infrastructure. SQL and CSPro software were used to structure the data according to the metadata specification.

6.1   Census data editing

Structure editing involved applying minimum processability rules defined for different questionnaire types (CAPI/CATI vs CAWI) for the four sub-populations (households, homeless, institutionalised and transients). Questionnaire records not meeting minimum processability such as those with no information at all on persons were deleted whilst those with information (e.g. questionnaires with name and surname and additional information on other demographic information were kept for further processing. Overall minimum editing was applied to the data:

- o Logical editing applied to data to remove inconsistencies from raw data;
- o Minimum hot deck editing method applied to resolve few undefined cases in critical variables such as age and service delivery questions; and
- o Variables with missing information were coded to unspecified;

Amongst other steps, the data operations team checked if all questionnaires had geographical hierarchy. Content input and output workstream provided editing specification to inform programs for checking and addressing data inconsistencies and imputing missing information where necessary). CSPro was used for content editing. The use of electronic questionnaires with in-built validations assisted greatly in minimising errors in the census data.

Comparison of raw and edited variables showed minimal imputation rates for all variables (less than 5%, implying limited changes to census raw data).

6.2   Census data coding

Some open ended census questions required coding of responses during data processing. In the Census 2022 questionnaires, few variables in the employment module were coded, namely occupation and industry. The coding of both occupation and industry was based on the International Standard Classifications. Both auto coding and manual coding were applied.

7. DATA ASSESSMENT AND EVALUATION

Evaluation of population census processes is a requirement of a good census according to international standards on census undertaking. Part of the process of assessing whether the Census 2022 exercise was executed successfully was to conduct an evaluation of both Census processes and final data. Paramount to note is that Census 2022 was the first digitally-enabled census conducted by Stats SA. Secondly, the Census implementation phases were conducted during the COVID-19 pandemic environment. The combination of these factors (use of technology, multi-mode data collection approach and COVID-19 disruptor) dictated doing things differently in the implementation of Census value chain processes. To ascertain the impact of factors on overall performance of the census project, evaluation of both processes and data became imperative.

Consistent with previous censuses, evaluation of census data was conducted by two groups; Stats SA team constituting of subject matter specialists and independent consultants.

In addition, a Post-Enumeration Survey (PES) to evaluate census results by determining the extent of under or over coverage of the population was conducted and informed adjustment of the census results.

8. POST-ENUMERATION SURVEY (PES)

8.1 Background

A post-enumeration survey (PES) is an independent survey used to measure the accuracy (content error) and reach (coverage error) of the census. The survey aims to identify and measure how many households and persons were erroneously included, missed or counted more than once in the census. PES data collection is undertaken shortly after the completion of census enumeration. Data collection for PES 2022 was conducted from June to August 2022 in 840 sub-enumeration areas (sub-EAs).

A population census is a massive exercise, hence it is inevitable that some people might be missed, and some might be counted more than once. A PES serves to identify the following types of errors:

- Coverage error: this includes both omissions (e.g. a household that was not enumerated) and erroneous inclusions (e.g. a household that moved into the EA after the census but was still enumerated, or a household that was enumerated more than once).
- Content error: this refers to the errors on the reported characteristics of the people or households enumerated during the census.

People may be missed or counted more than once during census data collection, so the census count of the population might be lower or higher than the estimated true population. The difference is called the net undercount or overcount, respectively. Rates of net undercount or overcount can vary significantly for different population groups depending on factors such as sex, age and geographic location. Statistics South Africa (Stats SA) obtains estimates of the net undercount or overcount, including the type and extent of content errors (reported characteristics of persons and households enumerated in the census) using information collected through the PES.

8.2 Objectives of the PES

The objectives of a PES are:

- to evaluate the accuracy of census data by providing quantitative information on coverage and content error at specified levels of estimation; and
- to provide a statistical basis for adjustment of census data if a decision is taken to make an adjustment as per United Nations (UN) guidelines (UN, 2017).

8.3   Scope of PES 2022

The scope of the PES 2022 was to estimate the total number of persons and households in private dwelling units (DUs) on Census 2022 reference night (midnight of 02–03 February 2022). The units of observation were persons and households who spent the census reference night and/or the PES reference night in these DUs. The scope of the PES excludes persons and households living in the following establishments:

- Homes for the aged unless they are structured into separate households;
- Student residences;
- Tourist hotels/motels/inns;
- Institutions, e.g. prisons and hospitals; and
- The homeless.

8.4   Preparations for the PES

Planning involved the development of documents outlining the goal and objectives of the PES, timelines for the project, identification of resources (financial and non-financial) required for implementing the project, and the development of methodology documents. Timelines for the PES were synchronised with those of Census 2022 in adherence to international best practice for maintaining a closed population between census and PES data collection. This is to ensure that the impact of natural population changes, such as births, deaths and migration, as well as lapses in respondent recall do not complicate the PES exercise. Activities of the PES included the following:

- Sample design and sample selection;
- Development of data collection methodologies, procedures and systems for data collection (publicity, unpacking of dwelling units, household registration, and enumeration), including quality control measures applied during data collection;
- Development of matching and reconciliation procedures and systems inclusive of guidelines and rules for matching and determining the match status of households and persons, as well as computer-based systems for matching household and person records in the census and PES;
- Development of data collection instruments which involved selection of data items that allowed measurement of coverage and content error, including layout design and user interface of the questionnaire for Computer-assisted Personal Interview (CAPI) and Computer-assisted Web Interview (CAWI) enumeration;
- Undertaking data collection activities of publicity, unpacking of structures and registration of the head or acting head of household's contact information, and enumeration of households in selected sub-EAs;
- Matching is a desktop-based exercise which involved the comparison of census and PES household and person records;
- Reconciliation visits involved the revisit of unresolved cases from matching in order to confirm or get more information;
- Adjustment of census household and person data entailed the creation of homogenous classes, and calculating the undercount/overcount within the adjustment classes that were used to adjust the census data; and
- Analysis and reporting involved the compilation of tables and report on PES results.

8.5   Methodology

Two independent data sources, i.e. census and the PES datasets, are used to arrive at the estimate of the *true population*. The first attempt at measuring the *true population* yields the *census-enumerated population*, based on a census enumeration. The second attempt yields *the PES estimate of the total population*, based on PES sampling and the estimation techniques.

Instead of assuming that the estimates from one or the other is better (Bureau of the Census, 1985), both census and PES estimates are used to derive a third, composite estimate of the true population called the 'Dual-system Estimate (DSE) of the true population'. The dual system estimation provides an estimate of the persons and households included in one source (PES) and excluded from the other (census), and vice versa. Both estimates contribute to the dual system estimate of the true population, which is more comprehensive than either the census or the PES true population estimates.

In the end, the dual system estimate of the *true population* is compared with the census-enumerated population and the difference is the *final net undercount or final net overcount* of the census. The *preliminary true population* is an estimate based on the PES sample while the *final true population* is based on the application of PES adjustment factors on the in-scope census data. The PES adjustment factors are calculated and used to correct for coverage errors.

8.5.1   Sampling

The PES sampling frame was a list of Census 2022 sub-EAs after the exclusion of sub-EAs that were not in-scope for the PES. The primary sampling units (PSUs) were the census sub-EAs. The principle for selecting the PES sample is that the boundaries for the sampled sub-EAs should be well-defined, and these boundaries should correspond with those of census sub-EAs to allow for item-by-item comparison of census and PES records. To ensure comparability with previous post-enumeration surveys, a single-stage stratified sample design was used. The stratification and sampling process followed would allow for the provision of estimates at national, provincial and geography type (i.e. urban, traditional and farm) levels. The sample of 840 sub-EAs was selected. The sample design methodology for the PES 2022 followed a similar methodology and assumptions to what was used for the PES 2001 and PES 2011. In general, the methodology considered the achieved precision levels at provincial level of the key indicator, the preliminary net undercount rate, from the last PES.

8.5.2   Questionnaire development

The approach to questionnaire design focused on capturing the main elements for measuring coverage and content errors. Only a few elements of the Census 2022 questionnaire that were not likely to change within a short period (that is, between the census and PES reference nights) were retained. The data items for the PES questionnaire included Name and surname, Date of birth, Age, Sex, Population group, Relationship to head/acting head of household, Country of birth, and Presence of person in dwelling unit on census night and/or PES night. The questionnaire content enabled the classification of each listed person as 'non-mover', 'in-mover', 'out-mover', or 'out-of-scope' with regard to their household presence on census night.

8.5.3   Fieldwork methodology

The PES 2022 data collection methodology was adapted from the Census 2022 data collection methodology. The data collection methodology had three main phases: publicity; unpacking of structures and registration of the head or acting head of household's contact information; and enumeration of households in selected sub-EAs. In order to fulfil the objectives of the PES to accurately measure coverage error and content error, the adopted methodology had the following key requirements:

- Publicity focused on informing and educating respondents and relevant stakeholders about the purpose of the PES to ensure successful coverage of all dwelling units in selected sub-EAs.
- All geo-points were independently visited, profiled and unpacked in order to identify dwelling units.
- All households within each dwelling unit were identified and the head of the household or the acting head of the household's contact information was registered.
- Households and persons who were eligible to be counted in the PES (based on their presence on the PES and/or the census nights) were enumerated.
- A multi-modal data collection approach with CAPI was used as the primary mode, while Computer-assisted Telephone Interview (CATI) and CAWI were used as supplementary modes.

8.5.4   Matching and reconciliation visit methodology

The matching exercise involved the comparison of household and person information between the census and PES data. A two-way case-by-case matching was conducted using the two sources: PES and census data. Reconciliation visits were conducted to confirm or get more information that would assist in finalising matching of unresolved cases. Guidelines for matching, including rules for determining the match status of households and individuals, were developed. Two (2) computer-based matching systems were developed for matching purposes: an automated matching system and a Computer-assisted manual matching system.

8.6   PES data collection

PES data collection commenced immediately after the completion of Census 2022 data collection. The PES is a much smaller scale operation (and hence easier to control) than the census. These features enable the PES to deliver a more accurate estimate of the percentage of people and dwellings missed or counted more than once by the census. By design, PES field operations are supposed to be independent from census field operations. To achieve this, the following measures were put in place:

- Independent publicity, unpacking of structures and registration of the head or acting head of household's contact information, and enumeration of households in EAs in the PES sample;
- Using separate/independent office staff in the PES and census;
- Ensuring that PES field staff were not employed as census field staff, and vice versa; and
- Maintaining the confidentiality of the PES sample so that census field and office staff were not aware which areas were included in the PES sample.

The PES implemented an integrated approach towards fieldwork. Temporary personnel (Fieldworkers, Fieldwork Supervisors and Quality Assurance Monitors) were recruited from the sub-EAs or districts in which they would be working. They underwent rigorous training of data collection procedures to ensure that they would deliver work of high quality at the end of field operations. Experienced permanent staff in Stats SA's Provincial and District Offices were seconded to the project to fulfil supervisory roles, and Head Office staff were deployed to the provinces as subject matter specialists for the duration of PES field operations to ensure high quality data and minimise costs.

8.7   Matching and reconciliation visits

The matching process involved the comparison of household and person records in census and PES data. The main phases of the matching process were:

- Automated initial matching was conducted using a computer-based automated matching system which searched through the census records in order to find the corresponding cases from PES enumeration records, and vice-versa (a two-way match), using a defined algorithm. The automated matching system also assigned an initial match status for all household and person records.

- Computer-assisted manual matching commenced immediately after the completion of automated initial matching. Computer-assisted manual matching was used to verify and/or revise the initial match status for all households and persons within the household who had been assigned 'possible match' and 'non-match' statuses during the initial automated matching. Matching staff assigned the revised initial match status for each household and person by applying the set matching rules.

- Reconciliation visits (RVs) are follow-up visits to households in the PES sampled sub-EAs. The purpose of reconciliation visits was to collect relevant information in order to determine the final match status for all 'possible match' and 'non-match' cases that could not be resolved during CAM matching. Furthermore, persons with an initial match status of 'matched' and with a moving status of 'in-mover', 'born after census' or 'out-of-scope' were also referred for reconciliation visits since their moving status was not consistent with the initial match status.

- Final matching involved the use of the results obtained from the reconciliation visits and initial matching phases to assign a final match status for person and household records.

8.8   Adjustment of the census data

The estimates of coverage by geographic or demographic variables such as province, geographic area type, sex, age group or population group tend to be more skewed when analysed individually. This is due to persons and households not being evenly missed over these subgroups during data collection. This necessitates the creation of homogeneous adjustment classes within which the adjustment of coverage errors can be implemented. CHAID analysis was used for deriving the adjustment classes and factors.

The adjusted Census 2022 count (excluding population in institutions, transient and homeless individuals) for the total population was obtained by summing up the adjustment factors across the adjustment classes. Only the in-scope census population was adjusted; the balance of the population (i.e. population in institutions, transient and homeless individuals) was not adjusted.

8.9   Estimation and tabulation

Coverage measures were calculated only for cases belonging to the PES universe. The initial estimates – weighted estimates of totals from the sample – include the following:

- a) Estimated number of non-movers;
- b) Estimated number of out-movers;
- c) Estimated number of in-movers;
- d) Estimated number of matched non-movers;
- e) Estimated number of matched out-movers;
- f) Estimated number of erroneous inclusions in the census; and
- g) Estimated number of correctly enumerated persons missed in the PES.