# Data Capture, Transmission and Management

## Introduction

The 2020 round of Population and Housing Censuses (PHC) in Africa marked a pivotal shift from traditional paper-based enumeration to digital methods. This transition brought significant improvements in data quality, timeliness, and operational efficiency. However, it also introduced new challenges, particularly in the areas of infrastructure, training, data security, and system integration. This chapter explores the evolution of data capture, transmission, and management in the digital era, drawing on practical experiences from countries that implemented digital censuses.

### A comparison of data capture, transmission and management in the non-digital versus digital census era

In the non-digital era, census data was collected using paper questionnaires. Enumerators manually recorded responses, which were then physically transported to central processing centers. Data entry was labor-intensive and prone to errors, and the entire process—from collection to dissemination—could take many months. Security concerns were also heightened due to the physical movement and storage of sensitive documents. Fragile documents were at risk of damage during transport, while manual data entry often caused delays and inaccuracies.  
   
In contrast, digital censuses utilize electronic devices such as tablets and smartphones to capture data directly in the field. This approach eliminates the need for manual data entry and allows for real-time validation and transmission. The integration of GIS tools and dashboards enables more effective monitoring and quality control. While digital methods streamline operations and reduce processing time, they require robust ICT infrastructure and careful planning to ensure data integrity and security.

### Considerations for data capture, transmission and management in a digital census

Transitioning to a digital census requires a series of strategic decisions. One of the first considerations is the choice of data collection devices. Some countries opted to purchase tablets, while others borrowed or implemented bring-your-own-device (BYOD) schemes. Each approach has cost and logistical implications, particularly in terms of device standardization and maintenance.  
   
In the 2020 round, according to the survey completed at the 2024 Expert Group Meeting of African Census Managers the majority (83%) of census implementing agencies bought hand-held devices. However, some agencies (9%) were able to borrow devices and the remainder used a mixture of buying and borrowing devices. For countries that responded to the survey, on average just over half the costs were spent on the field enumeration, including field hardware and software, field staff pay and training. Therefore, borrowing devices or implementing use-your-own-device schemes which some countries trialed could be an innovative way to reduce census costs. Note that if the field officers use their own devices, there can be financial benefits, but security and confidentiality must be prioritized. Additionally, any technology adopted must be thoroughly tested during the various testing phases, for necessary adjustments to be made prior to the main census.

Another key decision involves the enumeration method. Countries must determine whether to use Computer-Assisted Personal Interviewing (CAPI), Computer-Assisted Web Interviewing (CAWI), Computer-Assisted Telephone Interviewing (CATI), or a combination of these. The choice depends on factors such as internet coverage, population digital literacy, and the complexity of the questionnaire.  
   
Connectivity is another critical factor. In areas with limited or unreliable internet access, alternative transmission methods such as Bluetooth syncing between enumerators and supervisors must be considered. Data security must also be prioritized, with encryption, access controls, and secure storage solutions forming part of the overall ICT strategy.

## Key Aspects of data capture, transmission and management in a digital census

### Data capture

Data capture in a digital census involves the use of electronic devices such as tablets and smartphones to collect information directly from respondents. This method offers several advantages over traditional paper-based approaches, including real-time data validation, reduced errors, and faster processing times. Enumerators can use Computer-Assisted Personal Interviewing (CAPI) systems to enter data during face-to-face interviews, while Computer-Assisted Web Interviewing (CAWI) allows respondents to complete the census online. The choice of data capture method depends on factors such as internet availability, digital literacy, and the complexity of the questionnaire.

Built-in data validation rules within the digital collection applications (CAPI or CAWI) enable real-time data quality checks, ensuring that data is complete and consistent before it is transmitted. Effective data transmission strategies are crucial, considering factors such as network availability, data security protocols (e.g., encryption), and data volume.

Pre-enumeration planning is crucial for successful data capture. This includes procuring and configuring devices, developing electronic questionnaires, and training enumerators on the use of digital tools. During enumeration, real-time monitoring dashboards can track progress and identify issues, allowing for prompt resolution. Post-enumeration, data is reviewed and validated to ensure completeness and accuracy.

### Data Transmission

Digital enumeration facilitates the efficient and secure transmission of collected data from field devices to central servers. This eliminates the need for manual data entry, reducing the potential for errors and accelerating the data processing timeline.

Data transmission in a digital census involves the electronic transfer of collected data from field devices to central servers. This process can be carried out in real time or in batches, depending on the availability of internet connectivity. In areas with limited or unreliable internet access, enumerators may use alternative transmission methods such as Bluetooth syncing with supervisors, who then upload the data to the server.

Data transmission methods include:

* Mobile networks: Utilizing cellular data connections to transmit data in real-time or at regular intervals.
* Wi-Fi hotspots: Designating specific locations where enumerators can connect to Wi-Fi to upload data.
* Bluetooth transfer: Enabling data transfer between enumerator devices and supervisor devices, particularly in areas with limited connectivity.

Ensuring secure and efficient data transmission requires robust infrastructure and protocols. Encryption should be used to protect data during transmission, and access controls should be implemented to prevent unauthorized access. Supervisors play a key role in verifying data quality before it is uploaded, ensuring that any errors or inconsistencies are addressed promptly.

### Data Management

Data management in a digital census encompasses the storage, integration, and quality control of collected data. Once data is transmitted to central servers, it is stored in secure databases with backup systems to prevent data loss. Integration involves combining data from multiple collection modes (e.g., CAPI, CAWI, CATI) into a unified dataset for analysis.

Quality control mechanisms are essential to ensure the accuracy and reliability of census data. This includes developing validation rules, conducting error detection and correction, and performing regular audits. Trained personnel are required to manage these processes and address any issues that arise.

### Data Security

Data security is a cornerstone of digital census operations. Census data contains sensitive personal information, and any breach could undermine public trust and compromise the integrity of the census. Robust security measures must be embedded throughout the data lifecycle—from collection to transmission, storage, and analysis. This includes the use of encryption, secure file transfer protocols, and device-level protections such as Mobile Device Management (MDM) software. Access to census applications should be restricted through PINs and passwords, and data transmission should be conducted over private networks. Regular security audits and penetration testing are essential to identify and address vulnerabilities. Training all personnel on data privacy and confidentiality obligations is also critical to maintaining data security.

### Data Processing

Data processing in a digital census involves several stages, from pre-enumeration setup to post-enumeration cleaning and analysis. Pre-enumeration setup includes configuring devices, developing electronic questionnaires, and establishing data validation rules. During enumeration, data is validated in real time to identify and correct errors immediately. This reduces the need for extensive post-enumeration cleaning and ensures higher data quality.

Post-enumeration data processing involves cleaning and integrating data from multiple collection modes. This includes resolving inconsistencies, removing duplicate records, and imputing missing values. Advanced data processing techniques, such as machine learning algorithms, can be used to detect and correct errors, further enhancing data quality.

Once data is cleaned and integrated, it is analyzed to generate census results. This includes calculating population estimates, identifying demographic trends, and producing statistical reports. Data processing is a critical step in ensuring the accuracy and reliability of census data, and it requires skilled personnel and robust infrastructure.

## Selected country experiences

**Kenya**’s 2019 census demonstrated the benefits and challenges of digital enumeration. Enumerators used tablets to collect and transmit data, supported by real-time dashboards and supervisory checks. However, issues such as device malfunctions and network instability highlighted the need for robust technical support and contingency planning.

**South Africa** adopted a multi-modal approach, combining CAPI, CAWI, and CATI. The flexibility of this model allowed for broader participation, especially through CAWI, which was promoted as a quick and data-free option for respondents. The success of CAWI in South Africa underscores the importance of public awareness and user-friendly digital platforms.

In **Sierra Leone**, the use of GIS tools and cloud-based systems enabled efficient data collection and transmission. Local mobile providers played a key role in ensuring connectivity, and the integration of mapping tools improved the accuracy of enumeration.

**Botswana** faced challenges related to device procurement and network coverage. The use of multiple tablet models led to inconsistencies in performance, particularly in GPS functionality and battery life. These issues affected data quality and highlighted the importance of adhering to tested specifications during procurement.

**Link to country experiences section**

**Kenya**

The choice to use mobile technology for data collection was based on the existing extensive of mobile technology usage in Kenya, wide network coverage, reduced price of owning mobile devices, and improved ICT literacy levels among other reasons.

Android tablets and phones were chosen due to their affordability and availability of platforms for mobile application development tools. Mobile technologies were used for cartographic mapping, pilot tested and used in the main census enumeration.

Two local universities premised on their proven experience in successful delivery of tablets were used to assemble mobile devices and installed census software and related census files.

The collected data was sent to the central server by enumerators after completion of every household or several households depending on the availability of internet connectivity. Coordinators, ICT, and content supervisors supported by County and Sub County census committee members coordinated and supervised the exercise. In the event of a tablet’s failure to send the data to the server, the enumerators shared their data via Bluetooth with their respective content supervisors, who eventually sent it to the server.

All enumerators were sending their data to the server at the headquarters daily and whenever there was no internet, they would send data to the supervisor via Bluetooth that was configured between supervisor’s and enumerator’s devices. The supervisors would run pre-selected reports and review the data to check for its completeness before uploading the data to the server. A robust server was thus required to handle the magnitude of traffic and hits that were expected during the enumeration process. At the back end, the authorized personnel were pulling the data from the server to evaluate its quality and inform the respective field personnel in case data was inconsistent. In addition, they were also preparing monitoring reports.

The data processing team embarked on the process of assessing the census enumeration coverage immediately after data collection. This process involved the harmonization of EA Geo-codes that had been shared during tablets’ assembly and those used during enumeration. There was need for confirmation that all the EAs created during the mapping exercise were accounted for (with or without data). The next step was extraction of data from tablets whose data did not reach the server by close of enumeration. This was followed by checking and confirming whether all the EAs had data both from pre-enumeration listing of households and enumeration data. A comparison of listed households and actual returns, and accounting for the variances was done. In addition, a comparison of mapped households, listed households and actual enumerated households was also done and any discrepancies addressed.

The next step was to check for data consistency at household and individual levels. Basic edit specifications were developed to check the inconsistencies. The purpose of editing was to ensure that the final dataset did not include training data (data that was collected during training of census personnel) as well as ineligible individuals (empty records). In addition, the editing process took the following into account: households with missing household heads; individual records without age, sex, and relationship for conventional households; individual records without sex and age for non- conventional households; and duplicate records for structure, household numbers, and enumerator codes. Subsequently, tabulation plans were developed and used to generate summary tables for basic and analytical reports. Data in the generated tables was compared with existing data sets to assess the consistency of enumeration coverage and the outputs were validated by key stakeholders and thereafter launched.

Census data editing: Kenya utilized the handbook on Population and Housing Census Editing to come up with the editing specifications for the different modules in the census; Census data editing Team: The editing team was composed of data processors, subject matter specialists (demographic and non-demographic), computer programmers, supervisors and managers; KNBS has the inhouse capacity to program the edit specifications and run them to generate clean data files; We also utilized experienced CSPro data processors (former KNBS staff) who helped in the programming of editing/ imputation codes. Analytical Reports: Subject matter specialists drawn from the 2019 KPHC drafted the reports and then shared with the stakeholders for review; KNBS received TA support through the UNFPA ESARO office to come up with Population Projections report; Looking forward to receiving TA to come up with Geo-Spatial Dissemination System. Data Quality checks: An edit specification document should be prepared by the subject matter specialists and shared before pilot is undertaken; Hard checks in CAPI should be enforced for majority of the questions to reduce content errors.

Data security was achieved through several methods including encryption, secure file transfer and authentication. The system used for data transmission from the mobile devices to the census servers employed industry-leading encryption mechanisms to ensure that data was always encrypted and secure while in transit. In addition, secure sockets layer (SSL) technology and strong encryption keys were implemented at the point of collection and transmission to the server.

To ensure the security of the mobile devices including the application and data, the following was implemented; the mobile devices were running on a tested up-to-date android operating system, there was a logon PIN/Password required to protect against unauthorized access to the device, MDM software was installed to track the devices in case of theft. The MDM had an application locking software to protect installed apps from being uninstalled and introduction of unauthorized apps. Lastly, the devices were protected with a screen protector and flip cover to provide shock protection in case of falls.

To ensure data security, a Private Access Point Name (APN) was configured to ensure that the SIM Cards could only communicate to the Census servers.

**Sierra Leone**

Sierra Leone procured the Enumeration Area Pad software 4.0 license and satellite imagery with recent footprint of all structures and other features. Data was downloaded via google cloud from the mapping assistants in the field to the respective District Offices using onscreen digitization of delineated Enumeration Areas using ARCGIS Mapping software tool.

Each technological instrument intended for use in census data collection activity was available during the preparation period, and fully tested with other related applications.

The local mobile companies provided the internet services at an agreed fee, and this played a vital role working with servers at Statistics Sierra Leone Head Quarters to aid transfer of data directly from the field to main offices.

The mapping PAD Application (a Data acquisition package built by Milsat Technologies) did a direct export of data to CSV format and the Headquarter offices downloaded the data through a cloud server system.

For data security reasons, only accredited functionaries were allowed to use the App, the mapping PAD would periodically run system authentication on every phone the App is installed. The mapping PAD remote submission process utilized a top-level security cloud system to scan and safely deliver uploaded tasks to GIS portal. The GIS portal could only be opened by registered and verified GIS Officers.

**Botswana**

Data transfer to and from the servers was done through mobile data from sim cards loaded with different data packages depending on the role of end users for all the tablets or use of WIFI.

Statistics Botswana used fort iGATE firewall as the first layer of security to protect internal network from external untrusted traffic. They made sure that the antivirus for the organization is installed on all the servers including that of the Census for data protection.

Overall, the 2022 PHC had four end users with authorized access to tablets, laptops, and desktops. In addition, each IT personnel used their Passwords to log in to the servers.

## Challenges and Lessons Learnt

* Several common challenges emerged across countries. Device-related issues, such as hardware failures and inconsistent specifications, disrupted field operations. Inadequate network coverage hindered real-time data transmission, especially in remote areas. Training gaps among field staff limited their ability to troubleshoot technical problems, and in some cases, data quality suffered due to insufficient supervisory checks.
* Security concerns also surfaced. In Kenya, the late introduction of Mobile Device Management (MDM) software meant that some devices were vulnerable to unauthorized access or interference. These experiences underscore the need for early planning and comprehensive training in both technical and operational aspects of digital census implementation.

## Recommendations

* Procurement of devices and servers should be completed well in advance—ideally six months before enumeration—to allow for testing and configuration. Standardizing devices across the field force simplifies training and support.
* Data transmission should follow a layered model, with supervisors reviewing data before it is uploaded to central servers. In areas without reliable mobile networks, alternative connectivity solutions such as satellite dishes should be explored.
* Security must be embedded at every level of the operation. This includes using encryption, implementing access controls, and conducting regular audits. Training should be comprehensive and ongoing, covering both the use of digital tools and the importance of data confidentiality.
* To ensure the security of the entire infrastructure at all levels, the census implementing agency can engage a consultant to work with the ICT team in areas including data center security, transmission security, network security and penetration testing..
* Pilot testing is essential. All systems and tools should be thoroughly tested during pilot exercises to identify and resolve issues before the main census. Public engagement is also critical. Educating the population about digital enumeration methods can increase participation and reduce resistance.

## References and Resources

United Nations Statistics Division. Guidelines on the Use of Electronic Data Collection Technologies in Population and Housing Censuses. <https://unstats.un.org/unsd/demographic/standmeth/handbooks/guideline-edct-census-v1.pdf>

CSPro Users Guide. Synchronization and Data Quality Checks. <https://www.csprousers.org/help/CSPro/synchronization.html>