Regional Workshop on the 2020 World Programme on Population and Housing Censuses: International standards and contemporary technologies
Ankara, Turkey, 12-15 March 2019

Session 10

Data Collection with Handheld Electronic Devices

Meryem Demirci
United Nations Statistics Division

Content

- Advantages and challenges
- Planning considerations
- Considerations for selecting handheld devices
- Data collection application
- Data transfer
- Security for data collection with handheld devices
- Use of geospatial technology during enumeration
- Field operation management and monitoring
- Testing the data collection application and systems
- Re-use/disposition of devices

Advantages

- **Real-time quality control-** implementation of built-in edits\
- **Improved questionnaire design-** flexibility in space and language and supporting to respondents/enumerators
- **Added features** -integration with GPS, digital maps, help materials
- **Improved field management** –real time monitoring
- **Reduced time and possibility of reducing costs** –no data capture, less time for editing/imputation

➢ **Disseminating timely and reliable information**

Challenges

- Redesigning census processes
- High cost of equipment
- More time needed for preparation
- Sufficient technical expertise
- Data and system security
- Infrastructure constraints
- Data transfer considerations

Moving from PAPI to CAPI - Planning considerations

- **Critical factors in planning**
  - Census timetable
  - Government support and budget allocation
  - Capacity development for new technology
  - Modernization of IT infrastructure-central and regional
  - Development of systems and applications
  - Data transfer from field to head quarter and security
  - Decision on one mode or mixed modes of data collection

Considerations for selecting handheld devices

- **Overview of devices**
  - Tablet computers
  - Smartphones
  - Laptop (Notebook) computers

Considerations for selecting handheld devices

- **Important features**
  - **Processor performance and operating system -**should not affect the performance of enumerators and data collection application
  - **Storage capacity** –usually no problem for hundreds of interviews, but requires careful review especially if multimedia features are employed
  - **Screen –** affects costs and interview performance
  - **Battery –** more duration is desirable, external batteries can be considered
  - **Connectivity options** –Wi-Fi, Bluetooth, a cellular connectivity
    - ✓ Features of devices that will be used by enumerators, supervisors, IT supervisors are same?

Considerations for selecting handheld devices

- **Acquiring the device**
  - Outright purchase
    - Best way for ensuring it fits with requirements of census operations but can be expensive
  - Sharing/pooling among agencies/countries
    - Jordan shared with Palestine, Malawi will share with Zambia (2020 census)
  - Renting
  - Bring your own device (BYOD)-not advised

Data collection application

- Essential features of CAPI
  - Interface for field users
  - Questionnaire navigation
  - Automated routing (skipping)
  - Precoding
  - Customizing of questions (fill in a data from its memory)
  - Data quality control (real-time editing)
  - Case management (receive assignments, manage completeness and quality of questionnaire)
  - Support and documentation
  - Other features: collection of operational information (paradata)

Data collection application

- **Questionnaire design**
  - Adapting a paper questionnaire using the benefits of electronic data collection
    - The wording and structure of some questions may need to be changed to improve the quality

- ✓ **It is not simply converting a paper questionnaire into an electronic format**

Data collection application

- **Questionnaire design**
  - Two types of design
    - Matrix format or Sequential format
    - Open-ended or closed questions
  - Questionnaire layout and design
    - Easily navigation from one topic to another (one question per page or related questions in one page)
  - Testing testing testing !!!!

Data collection application

- **Real-time editing (or built-in edits)**
  - Types of consistency checks
    - **Range checks -** age, number of children ever-born, etc.
    - **Logic checks** – whether a combination of responses are logically consistent, CEB=CS+CD
    - **Consistency checks**- Whether two or more related responses are consistent with each other (age and completed level of education)
    - **Unknown checks** – where no data is entered

Data collection application

- **Real-time editing –**incorporated into data collection application
  - Consistency checks can be applied with two approaches
    1. **Hard checks**
    2. **Soft checks**

Data collection application

- **Real-time editing –**incorporated into data collection application
  1. **Hard checks – responses must meet with the editing rules** (the interview cannot continue until valid value is entered),

  for example:
  - SEX must have a value- empty space is not allowed
  - AGE must have a value –empty space is not allowed

  - If SEX is MALE, skip the questions on Births
  - If AGE < 15, skip the questions on Labour Force

Data collection application

- **Real-time editing –**incorporated into data collection application
  2. **Soft checks (signals)** – notify users that an item should be assessed for its accuracy
     - Do not prevent users from moving to next question or submitting the questionnaire
     - Warning users for inconsistent and missing responses

➢ **If not solved during enumeration, these types of errors have to be managed during data processing**

Data collection application

- **Real-time editing –**incorporated into data collection application
  2. **Soft checks (signals)**

For example:

- If age =9 and completed level of education is primary, warn users about inconsistency between age and education level
  - Correct age or education level or leave it as it is
- If marital status is missing, warn users
  - Complete the information or keep as missing (all questions related to marital status will be asked)

Data collection application

- **Real-time editing –how to decide hard and soft checks**

  - Key variables- age and sex- should be controlled with hard checks
  - Consider the balance between data quality and burden on respondents/enumerators
  - Do not take any risk to stop the interview

    - Testing testing testing !!!!

Data collection application

- **Development/acquisition of a CAPI system**
  - In-house
  - Outsourcing
  - Configuring existing software
  - Off-the-shelf solution

Data transfer

- **Means of transmission**
  - Networking
    - Cellular 2G/3G/4G
    - WiFi
    - Bluetooth
  - Synchronization options
    - Online -data transfer directly from the field to the server
    - Off-line -data transfer through Wifi at data collection station (local office) or using Bluetooth between devices
  - Servers
    - Web-based- a web application software is requires
    - FTP (File Transfer Protocol) server (requires an account configured on an FTP server)-dedicated for census
    - Cloud-based service

Data transfer

- **Responsibility of field staff for data transfer-two approaches**
  - Data transfer by enumerators
  - Data transfer by supervisors
- **Technical considerations for data transfer**
  - Estimating volume of data flow (for cellular)
  - Challenges of managing mobile devices over cellular networks
    - Network performance constraints (network coverage and reliability
    - Application performance (how application handles connection and response delays)
  - Risk mitigation in data transmission

Security of data collected with handheld devices

- **Major security threats and vulnerabilities**
  - Data stored in the device
    - Physical, disclosure, poor authorization/authentication, use of location services, insecure storage, Bluetooth attack
  - Data in transit
    - Insufficient transport layer protection, network exploits, wifi sniffing, phishing attacks, man-in-the-middle-attacks
  - Aggregated data on the server
    - Insecure data storage, server side controls

Security of data collected with handheld devices

- **Measures to secure data**
  - User and device authentication
  - Encryption of stored data
  - Device use policy-restrictions of particular applications to mobile device
  - Use of MDM (mobile device management) technologies
    - Provide secure access to data transfer networks
    - Can be tracked or remotely locked if devices are stolen

Use of geospatial technology

- Updating and correction of EA maps
- Workforce management –daily optimized routes, estimated travel time/travel costs
- Support enumerator's work in the field through integrating GIS in the data collection process
- Monitoring and operational management

Testing the data collection application and systems

- **Objectives of tests**
  - **Functionality testing** – performance of applications
  - **Usability testing** –user-friendliness
  - **Scenario-based testing** – prepared by subject-matter specialists to test CAPI detects entry problems/warnings/errors messages
  - **Compatibility/integration testing** –other systems such as GIS
  - **Acceptability testing** – for assessing the complexity of systems and corresponding training procedures/materials
  - **Infrastructure stress testing** -stability of systems under extreme use/under the influence of external factors
  - **Security testing** –device security and data security
  - **Protocol testing** – to establish the frequency and sequence of data transfer

Testing the data collection application and systems

- **Testing**
  - Pre-field tests
  - Field tests
  - Pilot Census
    - Field management structure
    - Enumeration process
    - Data collection
    - GIS
    - MIS
    - PES
    - Training of field staff
    - Administrative procedures, including for payment of field staff