Expert Group Meeting on Time-Use Statistics

New York
20-22 June 2018

Session 3: Mode of data collection

Use of electronic data collection technologies in population and housing censuses

Overview

- Modes of census data collection
- Data collection with handheld electronic devices
- Data collection with Internet
- Data collection with multi-mode approach

Modes of census data collection

- Modes
  - Interviewer-administered data collection
    - PAPI *(paper q w face-to-face interview)*
    - CAPI *(computer-assisted personal interview)*
    - CATI *(computer-assisted telephone interview)*
  - Self-administered data collection
    - PASI *(paper q w self-enumeration)*
    - CASI/CAWI *(computer-assisted self interviewing)*
- Multi-mode: use of more than one mode of data collection

Data collection with handheld electronic devices

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

Advantages

- Validation checks
- Automated routing
- Automated coding
- Customization of questions
  - Reduced data entry errors and reduced time and costs
- Easier handling (vs paper)
- Improved field operation management (eg. ability to collect GPS, date/time stamp)

Challenges

- High cost of equipment
- More time needed for preparation
- Infrastructure constraints (electricity, connectivity)
- Sufficient technical expertise
- Confidentiality

Planning considerations

- **Critical factors in planning**
  - Census timetable
  - Budget
  - Infrastructure considerations
  - Systems and software design (incl. questionnaire)
  - Data transfer
  - Operation management and monitoring
  - Data security
  - Technical skills and capacity development

Considerations for selecting handheld devices

- **Important features**
  - Processor performance
  - Operating system
  - Storage capacity
  - Keyboard
  - Screen
  - Battery
  - Connectivity options
  - Portability
  - Ruggedness
  - Sensors and peripherals

- **Evaluating requirements for selection**
  - Criteria – Security, Manageability, Productivity, Performance

Data collection application

Essential features of CAPI

- Interface for field users
- Questionnaire navigation
- Automatic routing (skipping)
- Precoding
- Customising of questions
- Data quality control (validation)
- Case management
- Data management
- Support and documentation
- Development environment/tools for design
- Other features: collection of operational information

Evaluation requirements

- Evaluation areas - performance in the field, tools for managing cases and data, tools for design, support and documentation

Data transfer

- **Means of transmission**
  - Networking
    - Cellular 2G/3G/4G
    - WiFi
    - Bluetooth
  - Synchronization options
    - Online
    - Off-line

Security of data collected with handheld devices

- **Major security threats and vulnerabilities**
  - Data stored in the device
    - Physical, disclosure, application/web-based, poor authorization/authentication, use of location services, insecure storage, Bluetooth attack
  - Data in transit
    - Insufficient transport layer protection, network exploits, wifi sniffing, phishing attacks
  - Aggregated data on the server
    - Insecure data storage, server side controls
  - Other security-related challenges
    - Enumerators fabricating data, low budgets, not fully developed infrastructure, authentication not possible where no cellular coverage

Security of data collected with handheld devices

- **Measures to secure data**
  - Authentication
  - Encryption of stored data and data in transit
  - Use of MDM (mobile device management) technologies
  - Device use policy
  - Development of security policies

Testing the data collection application and systems

- **Testing**
  - Pre-field tests
  - Field tests
  - Pilot Census (procedures& processes, applications, systems, training)
- **Types of tests**
  - Functionality testing
  - Usability testing
  - Scenario-based testing
  - Compatibility/integration testing
  - Acceptability testing
  - Infrastructure stress testing
  - Security testing
  - Protocol testing

Field operation management and monitoring

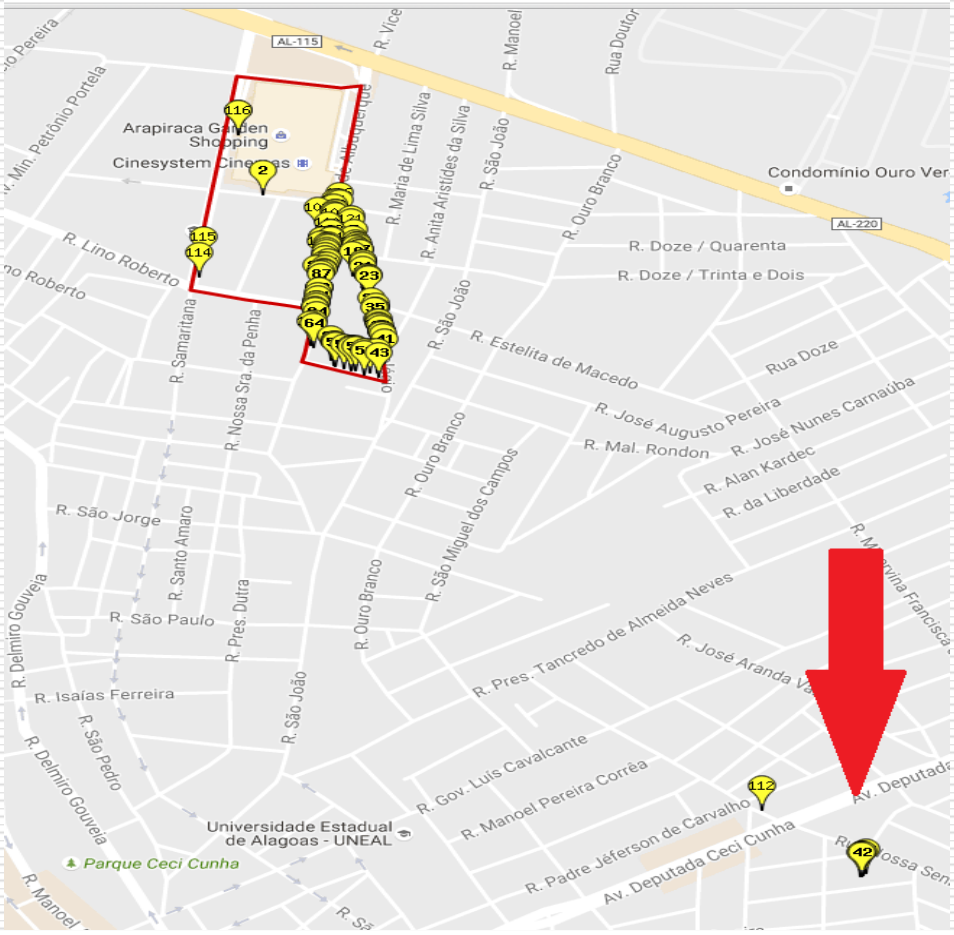
- **Organization of field enumeration**
  - Technical support to field staff: IT support team, Call Centre, support in case of emergency
  - Technology coordinator and team
    - Installation of system on device
    - Checking all accessories
    - Assisting in case of hardware/software problems
    - Ensuring that devices synchronize
    - Additional training of supervisors and enumerators, as needed
  - Recruitment and training (content and technical)

Field operation management and monitoring

- **Management and monitoring system**
  - **Field management tools**
    - HQ and regional manger tools
    - Supervisor's module
    - Enumerator's module
  - **Management information system for operation control**
    - Performance indicators (Daily returns; Cumulative returns; Comparison with predicted returns; etc)
    - Supervision alerts
    - Geo-tracking tool

During enumeration: Monitoring and operation management

- Example of an attempt to complete a questionnaire at a location beyond the boundary of an enumerator's EA



Field operation management and monitoring

- **Device use policy**
  - Protection and proper handling of the devices
  - Ensuring return of the devices in good condition
  - Securing data confidentiality
  - Preventing unauthorized use of the devices
  - Confining device use to specific work areas
  - Preventing enumerators from entering fabricated data
  - Measures for non-compliance

Re-use/disposition of devices

- **Benefits of re-use**
  - Optimizing scarce financial resources
  - Reducing environmental footprint
  - Reduce cost of subsequent data collection operations
- **Challenges of re-use**
  - Degradation
  - Battery life
  - Warranty life
  - Removal of data
  - Wiping software
  - Restoring disabled features
  - Compatibility with IT infrastructure for re-use

Data collection using the Internet

- Advantages, Challenges and Requirements
- Contact and communication strategy
- Identification and authentication of respondents
- Design of questionnaire and data collection portal
- Support for respondents
- Managing and monitoring response
- Security
- Testing

Using Internet self-response as a collection mode

- **Advantages**
  - Improved coverage and data quality
  - Potential cost saving
  - **Greater convenience and lower respondent burden**
  - Improved timeliness
  - **Protecting privacy of respondents**
  - **Other benefits (environmentally friendly, generating a large amount of paradata)**

Using Internet self-response as a collection mode

- **Challenges**
  - **Providing reliable linkage between household and location**
  - **Coordination of multi-mode collection**
  - Developing systems infrastructure
  - Protecting data security –portal may be crashed/respondent data may be corrupted or lost
  - Extended collection period (hard to capture picture as at census day)
  - **Mode effect** and non-response bias
  - High initial cost

Using Internet self-response as a collection mode

- **Requirements**
  - High literacy rate
  - High level of access to computers and the Internet, and computer literacy
  - Making reasonable assumptions about take-up rate
  - Availability of address/building/dwelling list
  - Legal authority
  - Questionnaire design for mobile devices
  - Public trust and acceptance of Internet for official business

Contact and communication strategy

- Multi-phase contact approach
- Communication strategy (invitation letter/reminder letters, etc.)
- Offering a paper questionnaire
  - at initial contact or;
  - during the reminder phase or;
  - on request

Identification and authentication of respondents

- Link households to address of dwellings
- Reduce risk of impersonation
- Reduce risk of duplicate responses
- Better security, both actual and perceived

Development of data collection application and portal

- **Questionnaire design and application features**
  - **Log-in screen (user experience starts here)**
  - Screen visibility
  - Intuitive and easy to complete
  - Format: matrix or sequential
  - Automated skip patterns
  - Response options and menus
  - Validation messages
  - Progress through/save/submit questionnaire
  - **Mobile friendly, responsive design (multiple browsers)**

Support for respondents

- Embedded help text in the online questionnaire
- Frequently asked questions
- Online help
- Census helpline (Call Centre)

Management and monitoring of Internet response

- **Metadata -- *During enumeration***
  - Daily returns
  - Cumulative returns
  - Comparison with predicted returns
  - **Concurrent users**
  - Saved, abandoned or incomplete returns
  - **Reports on IT infrastructure stability and capacity**

Management and monitoring of Internet response

- **Coordination with non-response follow-up**
  - Critical in multi-mode collection
  - Central and integrated data collection operation control system,
  - Non-response follow-up procedures need to have flexibility

Security

- To maintain security of personal information:
  - Secure log-in
  - Internet application should ensure zero footprint on respondent computer
  - Timing-out after period of inactivity
  - Encryption
  - Powerful firewalls, intrusion detection
  - Strong access control procedures
  - Contingency plans for temporary service interruptions
  - Communication strategies to assure respondents

Testing

- Questionnaire -- cognitive/qualitative testing
- **Planning -- experimental testing to estimate take-up rate, various metrics for planning**
- IT systems -- infrastructure and system testing

Multi-Mode Data Collection Approach

- Benefits of multi-mode data collection
- Factors impacting decision for mixing modes
- Challenges of using multi-mode collection
  - Operational management
  - Data integration
  - Mode-effect

Benefits of multi-mode data collection

- **Benefits of using multi-mode approach include:**
  - Improve coverage
    - Especially to reach people difficult to enumerate, such as people living alone, living in buildings difficult to access
    - Provide an alternative to people reluctant to participate (due to concerns about privacy)
  - Reduce respondent burden – offering alternative means to respond
  - Reduce data collection cost (esp. those associate with field work)

- **Mixing modes could potentially minimize the cost and quality issues associated with a single mode**

Factors impacting decision to use of multi-mode

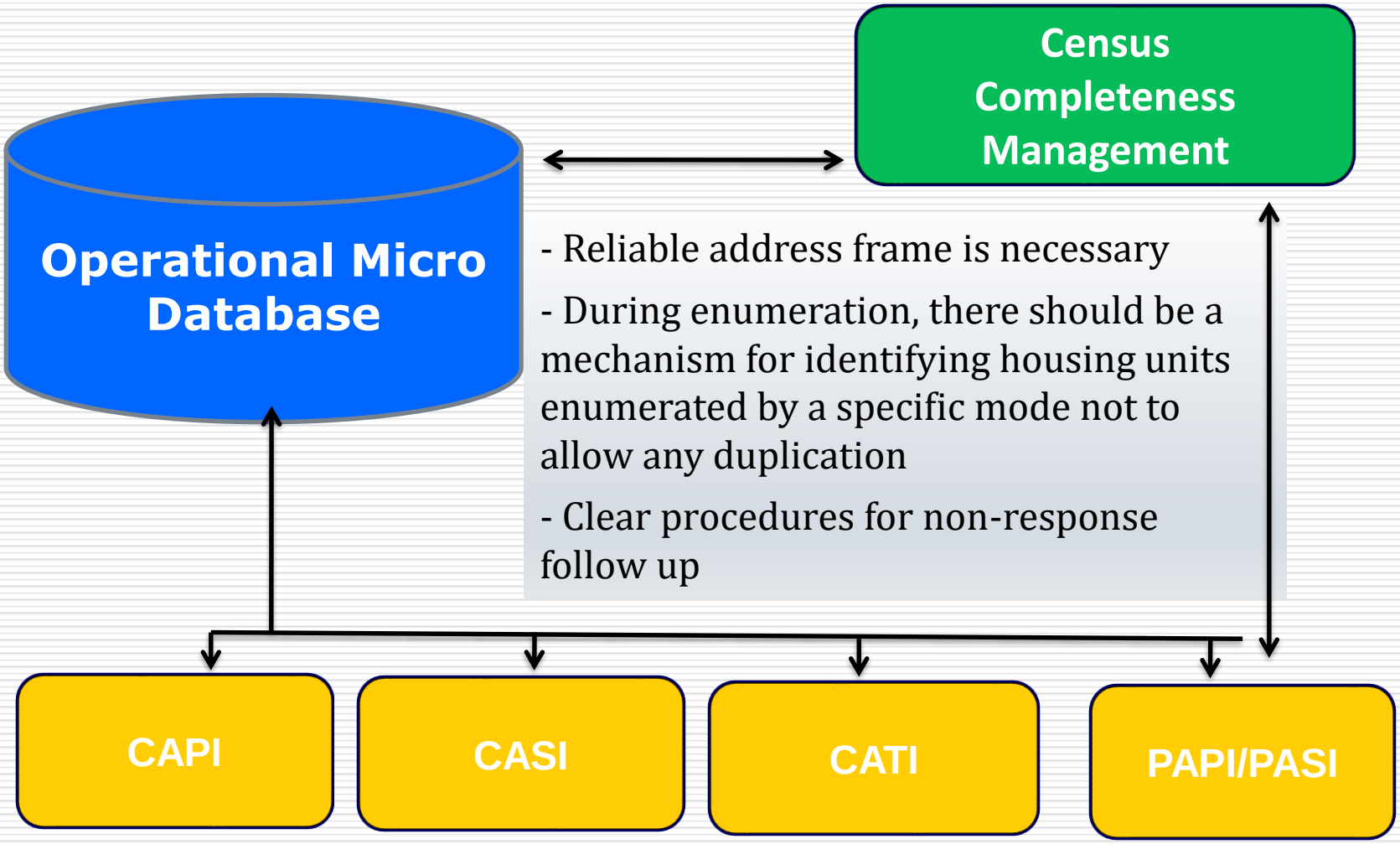
CAPI →
CASI →
CATI →
PAPI →
PASI →

Administrative and resource factors:
- Available budget
- Time period available
- Public acceptability and respondent burden
- Existing infrastructure
- What are additional burdens on operation?
- Is this approach cost-effective ?

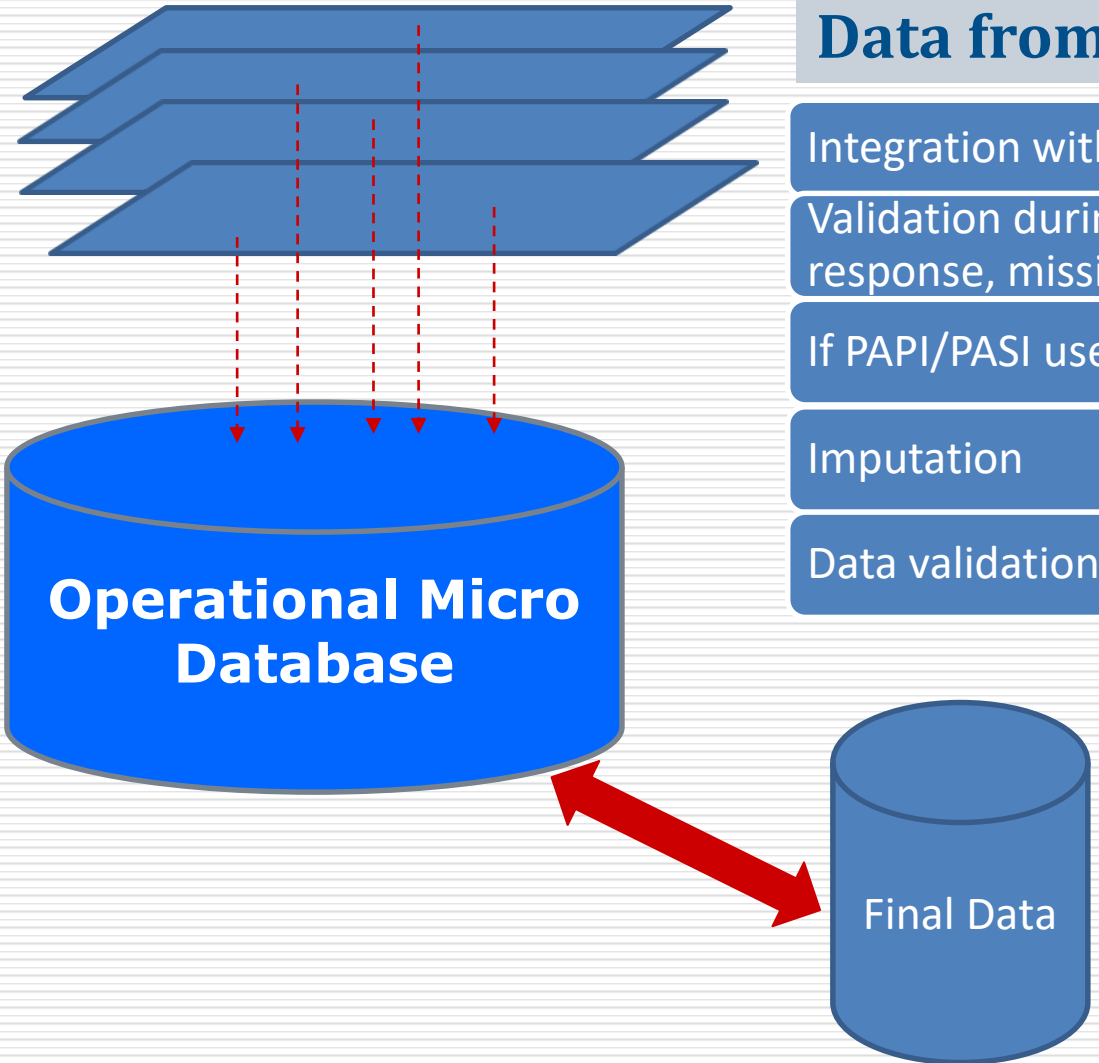
Design factors:
- Sequential or concurrent approach

Ability to access different population groups, and ability to improve coverage:
- What geographical areas/population groups are targeted with additional mode?
- What is the expected proportion of population that will be enumerated with specific method?
- To what extent response rate will increase?

Challenges - Operational management



Challenges - Data integration



Data from different modes

- Integration with Census Frame
- Validation during the field enumeration (non-response, missing, inconsistency)
- If PAPI/PASI used, data capture, coding, editing
- Imputation
- Data validation/analysis

Challenges – Mode effect

- Multi-mode data collection has implications for the quality of the collected data, particularly for data comparability
- **'Mode-effect'** is the bias caused by the mode of the data collection used -- due to the delivery of different results as a consequence of using different means of collection
  - Mode effect creates artificial differences in the population (ie. differences observed are due to how the data are collected rather than real differences in the population)
  - Mode effect varies depending on the type of design chosen
- Four factors contribute to/associated with mode-effect: ***Coverage, Response rate, Item non-response and Measurement differences***

Challenges – Mode effect

- **Minimizing mode effect on data quality**
  - **Optimize design**
    - *Sequential design* - one mode is used as main data collection method supported by additional data collection method for non-response follow up
      - Primary mode data collection should be used to its maximum potential to enumerate the majority of population
    - *Concurrent design* - the different methods are equally important and respondents are given a choice
      - There is a risk of not using a mode to its fullest potential (for example, the potential benefit from use of validation checks in electronic questionnaires)
    - In general, sequential design has less pronounced mode-effect as compared to concurrent design

Challenges – Mode effect

- **Minimizing mode effect on data quality**
  - **Conduct empirical studies** - understanding causes of mode effects on coverage, response and measurement can provide information that can help to minimize mode-effects
    - This kind of experimental study can be conducted with pre-tests and pilot censuses to understand mode effects on the data quality especially on item-non-response and measurement error
    - Findings from such experimental studies can be used to develop editing and imputation strategies for decreasing mode effect on the data quality.
    - Where experiments are not possible, matching studies (ie. comparing information from respondents collected from different modes) is another option to assess mode effect