



Guidelines on Use of Electronic Data Collection in Censuses: Use of Handheld Devices and Internet

Meryem Demirci

United Nations Statistics Division

“Experiences, good practices and challenges for the National Statistical Systems of Latin American countries in the light of the 2020 Population and Housing Census round”

UN ECLAC, Santiago, Chile, November 27-29, 2017



Content of the Guidelines

- Decision making process
 - Considerations for Planning and Preparations
 - Electronic Data Collection Application
 - Use of Handheld Devices for data collection
 - Use of Internet for data collection
-



Decision making process

- ❑ Steps for making a decision
 1. Developing strategic objectives
 2. Identifying key factors affecting decision
 3. Information required for decision making
 4. Analysis of information (SWOT and Cost-benefit analysis)
 5. Components of decision making
 6. Documentation and evaluation
-



Decision making process

○ Drivers for adoption of electronic data collection

- Improving census coverage and data quality
- Disseminating timely census results
- Reducing response burden
- Potential cost saving, cost-effectiveness

Indirect drivers

- Modernization of business process of statistical operations
- Meeting public expectations

○ Possible challenges

- Systems infrastructure and building a institutional capacity
- High initial cost
- Coordination of multi-modes
- Confidentiality and security of data

➤ **Develop strategic objectives based on national priorities**



Data collection methods

❑ Face-to-face interview data collection

- Paper questionnaire with face-to-face interview (PAPI)
- Computer-assisted personal interviewing (CAPI)
- Computer-assisted telephone interviewing (CATI)

❑ Self-interview data collection

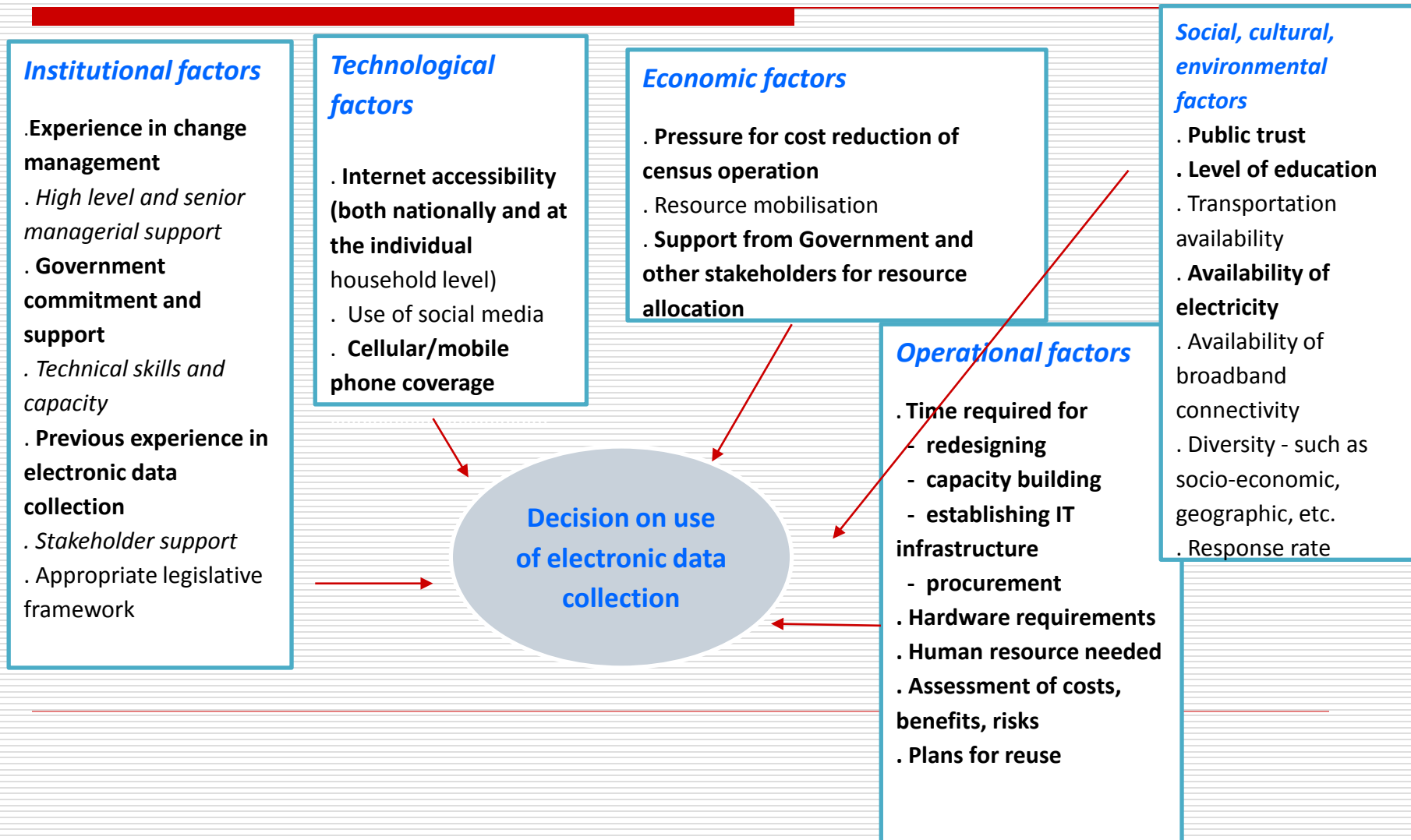
- Paper questionnaire with self-enumeration (PASI)
- Computer-assisted self-interviewing (CASI)

❑ Multimode data collection methods

- Paper-based data collection (PAPI/PASI)
 - Electronic data collection (CAPI/ CASI/ CATI)
 - Combination of electronic and paper-based data collection (PAPI and at least one of CAPI/CASI/CATI or PASI and at least one of CAPI/CASI/CATI)
-



Key factors affecting decision-general framework





Information required for decision making

a. Quantitative information

- Population size and density
- Literacy rate
- Proportion of population having basic computer skills
- Number and proportion of population who can access to the Internet
- Number and percentage of households with Internet connection
- Proportion of households with electricity
- Cellular/mobile phone coverage
- Pattern of non-response rates for previous censuses and surveys
 - By geographical area and rural/urban

b. Qualitative information

- How ready is the statistical agency to adopt new technology for enumeration?
- How will Government and main stakeholders (particularly data users) react to the use of new technology?
- Does existing legislation allows for a change in the method of enumeration, and, if not, is a requisite change practicable?
- What is the public's reaction to adoption of Internet as the medium for self-response?
- What are main expected challenges? What would be strategies to achieve these challenges?
- What are the risks in procuring hardware and outsourced services?



c. Estimate of Total Cost Ownership (TCO)

- Fixed costs and variable costs
- Three components of TCO:
 - Hardware/ software
 - Operational costs
 - Personal costs
- Total costs of the multi-mode can be assessed by adding the cost estimate of each possible method
- Comparison with the costs of previous census

An example of a template

Items	Method of Data Collection 2010		Method of Data Collection 2020				
	PAPI	PASI	PAPI	PASI	CAPI	CAWI	CATI
Number of EAs							
Number of questions							
Number of households							
Fixed costs:							
Hardware							
Software							
Technical support/services							
Other Electronic equipment							
Maintenance Fees							
Variable costs:							
Recruitment of temporary staff							
Training/workshops, etc.							



Information required for decision making

United Nations Statistics Division

-
- d. **Benefits**- improving data quality, coverage, timeliness of delivery of outputs and future uses of same technology for statistical operations
 - e. **Risks and challenges**- such as building a capacity, data transfer, resource mobilization, hiring a qualified field staff, data security
 - f. **Timetable**-sufficient time such as for building a capacity, procurement, system development, end-to-end testing (usually extensions to the timetable required for decision making and planning)
 - g. **Reuse of technology**- other censuses, household surveys, business registers, the use of hardware by other organizations
 - h. **Best practices**-practices of countries which have similar conditions
-



Decision making

- ❑ *Analysis of information*
 - **SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis:** a better understanding of the balance between the benefits, risks and challenges in the introduction of new technology
 - **Cost-benefit analysis:** comparison of total costs with its benefits
 - If a multi-mode data collection methodology is an option, the cost estimates, benefits and risks should be determined for each type of data collection method under consideration
 - ❑ *The components of decision making:*
 - Data collection mode or modes
 - Type of devices and equipment
 - Development of data collection application (commercial, in-house, out-sourced, customized)
 - Understanding and balancing an investment in refreshing or updating existing systems
-
- What new skills are needed- based on existing skills



Electronic Data Collection Application

- Essential features of data collection application
 - **Data quality control**-validation between responses
 - **Automated skip patterns**- improve data quality,
 - Residency status
 - Occupancy status

Benefitting most with asking additional questions for clarification
 - **Pre-coding**- selection of appropriate codes from drop-down menu
 - What variables can be automatically coded or coded during data processing (occupation/industry?)
 - **Validation messages**: Non-response, invalid response, partial response, consistency verification, etc.
 - **Customizing of questions**- recalling a piece of information and inserting appropriate place
 - **Collection of operational information** –dates, time and duration of the interview and GPS coordinator, etc.
 - **Multiple language capabilities**
-



Electronic Data Collection Application

- ❑ Allows
 - **Case management-** collecting and updating status of individual enumeration-persons/housing units and controlling the quality of records
 - ❑ Design consideration
 - **Matrix format:** each question appears once and response options are repeated under the name of each person in the household. This structure is more appropriate for sections of the questionnaire that are not dependent on the responses to previous questions
 - **Sequential format:** all questions are asked about one person at a time. This structure is more appropriate in sections of the questions contain more automated skip
-



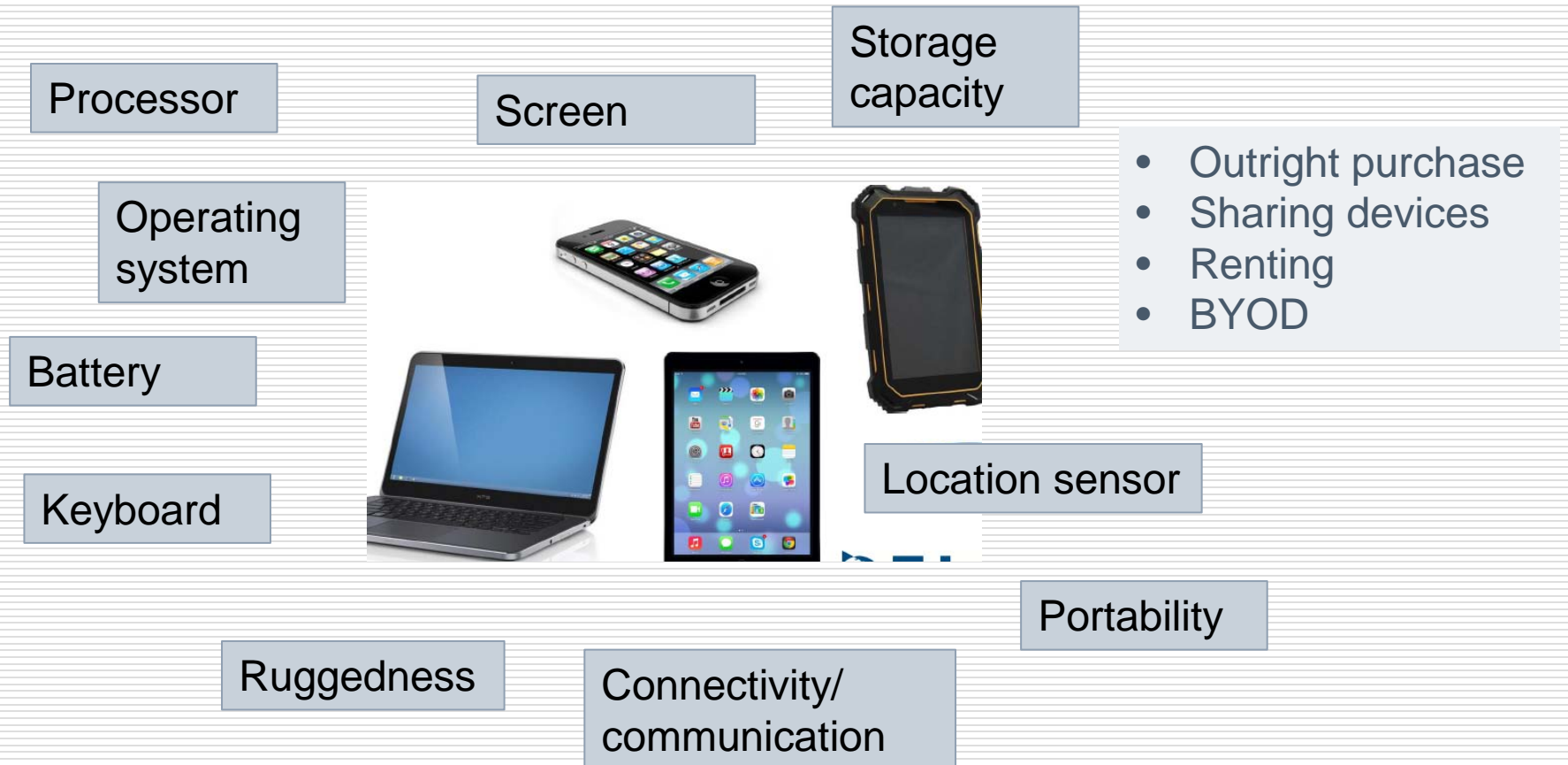
Electronic Data Collection Application

- ❑ Testing programme should include a series of tests for data collection application
 - **Functional testing**-all aspects of the questionnaire design (texts, response options, missing values, instructions, error messages)
 - **Usability testing** – testing user-friendliness of application
 - **Scenario-based testing**- using the input of subject matter specialists for testing typical and common profiles, entry problems, warning/error messages
 - **Extreme-scenario testing**- similar tests for extreme cases such as large households including several generations, multiple occupations, multiple types of disabilities

- ✓ Pre-field and field tests to finalize data collection application
 - ✓ Pilot census for testing whole processes
-



Data collection with electronic devices





Use of Handheld Devices –Data transfer

Networking

- WiFi
- Cellular
- Bluetooth

Data transfer options:

- Online- Direct transfer from devices to main server
 - Partially online-Due to limitation of cellular coverage/availability of Wifi, data transfer occurs when the Internet is available
 - Off-line – Due to rare or never access to the Internet, data transfer occurs between devices
- **How the method of data transfer affects data quality, possibility of duplication, removing ?**

Servers:

- Web server
- Cloud based service

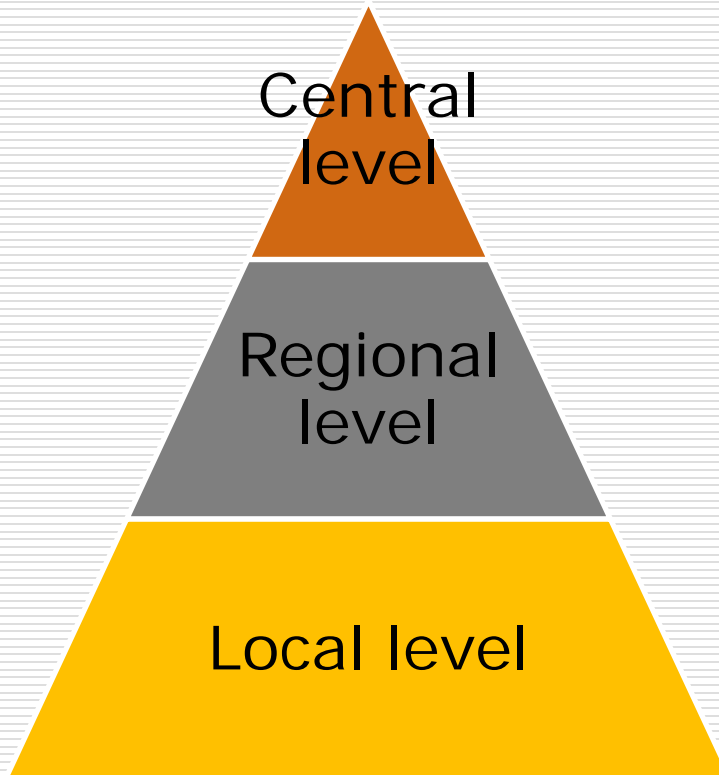


Use of Handheld Devices –Security measures

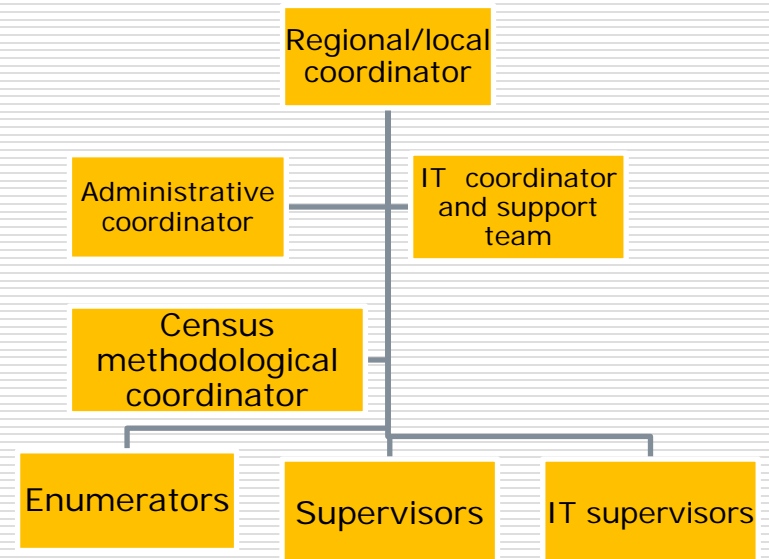
- ❑ Measures for securing data on devices
 - User and device authentication - A device password and/or other authentication; Under the direction of administrator, remotely lock the device
 - Data storage – strongly encrypt stored data on both built-in storage and removable media storage
 - General device use policy – Restrictions of particular application to mobile device
-



Use of Handheld Devices –Field organization



Field structure





Use of Handheld Devices –Field organization

□ Technical support to field staff

- **IT support team-** to resolve technical problems, operational problems related to devices, applications and data transfer encountered by enumerators and supervisors
 - **Call center** – provide logistical support, IT related support and methodological support to field staff and information to publics
 - **Support in cases of emergency** – consider GPS solution to provide immediate support to field staff in case of emergency
-



Data collection with Internet

- ❑ Requirements for data collection via Internet
 - High literacy rate
 - High rate of access to computers and computer literacy
 - High rate of access to Internet
 - Increasing use of mobile devices
 - Public trust and acceptance of the Internet for official business
 - Making reasonable assumptions about take-up rates
 - Availability and quality of address/building/dwelling register or list
 - Legal authority, especially in case of moving from traditional field enumeration to online data collection
-



Data collection with Internet

- ❑ Conduct and communication strategy
 - How the households are invited to respond by Internet
 - How many time they are contacted
 - Whether the option of self-responding by completing a paper questionnaire is offered at initial stage

- ❑ Offering the paper questionnaire for self-enumeration can be grouped into three categories:
 - **Paper questionnaire sent to at initial contact**, where an Internet option is offered including a unique Internet access code
 - **Paper questionnaires available on request**-to maximize the overall response rate
 - **Paper questionnaire sent during the reminder phase**-sending only invitation letter providing Internet access code initially and then providing a paper questionnaire with a reminder



Data collection with Internet

Identification and authentication of respondents

- Unique identifiers to access the online questionnaire is necessary for data confidentiality and identification of responding households
- Different approaches for sending unique identifiers

Please enter your personal internet access code to begin your online questionnaire, or to finish one you have already started.

BD8C 47F2 SNVN ZVRD LFNN|

Sign in

You will find your personal internet access code on your paper questionnaire

Household Questionnaire

Complete online

XXXX XXXX XXXX XXXX XXXX

Return to: FREEPOST 2011 Census, Processing Centre, UK



Data collection with Internet

- ❑ Development of data collection portal
 - Capacity of the infrastructure (such as servers, storage devices and communication vehicles) should be decided based on potential users
 - The most important factors are:
 - The average number of pages a respondent will see
 - The average time spent per page
 - The number of simultaneous users at the busiest time
 - The total number of responses expected
 - ❑ Support for respondents
 - **Online support** – instructions and conceptual help materials for completing the online questionnaire; frequently asked questions; a help button; questionnaire guide
 - **Census Helpline (Call center)**- in addition to questions regarding the census, the use of the Internet application, the login process or other technical difficulties, new Internet access code, etc.
-



Management and Monitoring for Electronic Data Collection

- Real time monitoring is possible
- Up-to-date information can be collected for monitoring and managing the progress of the enumeration
- Sharing performance indicators , reports and alerts among field managers/staff
 - Key indicators
 - Status of enumeration –Completed, Refused, Follow up, No contact, Non-residential
 - Enumerated population
 - Average number of population/housing unit per day
 - Number of data transmissions per day by geographical area
 - Number and percentage of devices that have to be renewed
 - Counts of submitted completed questionnaire/incomplete questionnaires
 - Comparisons between actual return and planned returns
 - Housing units enumerated by mode of data collection for timely monitoring non-response

CAPI

CASI



Multi-mode data collection- paper+electronic

- ❑ Three factors concerned with data quality and associated with mode-effect include:
 - Coverage
 - Item non-response
 - Measurement differences
- ❑ Minimizing mode effect on data quality
 - Optimize design
 - Sequential design- main data collection with additional method for non-response follow up
 - Concurrent design - the different methods are equally important and respondents are given a choice
 - Use a generalized design to make them insensitive to the effects of data collection method with the risk of not using a mode to its fullest potential
 - Conduct empirical studies-understanding causes of mode effects



THANK YOU...
