Introducing Administrative Data to the Census in Indonesia

BPS – Statistics Indonesia

Jakarta, 4 March 2022
PC2020 is the 7th Population Census

Legal Basis

- Law No. 16 of 1997 on Statistics
- Law No. 23 of 2006 on Population Administration
- Law No. 52 of 2009 on Population Development and Family Development
- Law No. 24 of 2013 on The Ammendment of Law No. 23 of 2006 on Population Administration
- Government Regulation No. 51 of 1999 on Statistical Governance

United Nations Recommendation 2017
Population and Housing Census 2020

Presidential Decree No. 39 of 2019
One Data Indonesia

Presidential Decree No. 62 of 2019
National Strategies For The Acceleration of Population Administration For The Development of Vital Statistics

“Documenting Indonesia toward ONE POPULATION DATA for Advanced Indonesia”
Provides demographic parameters (fertility, mortality, and migration) as well as other individual characteristics for population projection and SDGs indicators.
- Planned to be conducted in September 2021
- Postponed to 2022

Provides data on number, composition, distribution, and characteristics of Indonesian toward ONE POPULATION DATA IN INDONESIA
- Completed in 2020

COMBINED METHOD
The use of relevant administrative data (data managed by the Ministry of Home Affairs (MOHA)) completed by sample survey

DATA COLLECTION MODE
Online population census, Drop Off Pick Up (DOPU), and interview
FIELD ACTIVITIES

September 2020

1. VERIFICATION OF SPR
   - Confirm the existence of residents
   - Mapping residents address: by domicile and by ID Card
   - Adding new residents (who are not in the pre-list yet)
   - Knowing the dead residents

2. GROUND CHECK
   - Updates the population characteristics

3. POPULATION DATA COLLECTION

TOWARDS ONE POPULATION DATA OF INDONESIA
DATA PROCESSING PHASES:

1. Preparation of population database from MOHA for online census purposes
2. Preparation of population database for field enumeration
3. Release population database
PREPARATION OF POPULATION DATABASE FROM MOHA

- Coordination with MOHA
- Population data replication from MOHA
- Cleaning and region encoding to BPS standard
- Population database for Online Census
Why do we need region matching?

The region encoding of BPS Statistics and MOHA are different.

What do needs to be matched?

✓ Master of the region
✓ SLS
✓ Individual

What is the focus problems of data matching?

✓ The name of the SLS of MOHA and BPS are different.
✓ Inconsistent code and area names (from province to the village).
✓ No standard in writing answers for the same purpose (education, type of work, and relationship with the head of the family). So, BPS need to make standardization. For example, PC2020 can't use ISIC category of work but MOHA's.
Matching area code (province, district, sub-district, village)

Matching SLS
- SLS from BPS is the 2019 mapping results
- SLS from MOHA is taken from the address in the registered family card (KK)

Enter the information from registered family card (KK) from MOHA into the BPS’ SLS

Ground check for the unmatched registered family card (KK) by BPS Regional

MOHA submits individual data to BPS that used as a database in the Online Census (November 2019)

SLS (Local Neighbourhood Unit below the village)
51,36 Million Residents or about 13,63 Million Families of Indonesia have been participated in the PC2020 Online Census.
VARIABLE IMPUTATION

Online PC Variable Imputation based on filled variable and MOHA Variable

GRADING

- To determine the household visited or verified in later field enumeration
- Confirm the difference between census and MOHA variables value

POPULATION DATABASE FOR FIELD ENUMERATION
**GENERATING STATISTICAL PEOPLE REGISTER (SPR)**

**Preparation of data from the result of the Online Census**

- Compiled from the result of the online census and those who are not online census but their data is registered in MOHA
- Several Online Census status will state in SPR, such as:
  - Not Online Census yet
  - Already, but needs to be re-entered
  - Already, move out
  - Already, live in this SLS
  - Already, declared dead
  - Already, need confirmation (date of birth, sex, or ID number)

**SPR Printing**

- SPR made on SLS level
- We also made village level of SPR as the list of names of residents who live in the Village area but whose SLS area is not known
- Digital SPR was built to facilitate census enumerator in searching resident’s name/address
Data Integration
Population data from field enumeration, Task Force, and overseas population census

Deduplication
By using deterministic and probabilistic approach

Consolidation
Consolidation of processing results with MOHA

Integration
of processed data with MOHA based on consolidation

Release Population Database

1 2 3 4 5
DATA SOURCES AND PROFILES

ADMINISTRATIVE DATA
292,979,718 records/distinct ID Number

Attributes:
- ID number
- Family card number
- Full name
- Sex
- Date of birth
- Place of birth
- Family relationship
- Marital status
- Profession

ONLINE POPULATION CENSUS
48,266,123 records

Attributes:
- ID number
- Family card number
- Full name
- Sex
- Date of birth
- Family relationship
- Marital status
- Profession
- etc (activity, housing)

FIELD ENUMERATION
363,478,036 records

Attributes:
- ID number
- Family card number
- Full name
- Sex
- Family relationship
- Address
COVID-19 PANDEMIC

COVID-19

Minimum face-to-face interview

Mobility

Rejection

Missing Value

Inconsistency

Duplication

Nonresponse
**WHAT DO WE DO?**

- **Data Source Profiling**
  - Data structure, data type, data summary, interrelationship

- **Data Quality Assessment**
  - Validity, consistency, completeness, uniqueness, integrity, etc

- **Define Algorithm & Cleansing Rule**
  - Data matching, data pairing

- **Data Cleaning**
  - Remove duplicate, finding individual, etc

- **Evaluation**
  - Data checking
DATA QUALITY ASSESSMENT

DATA QUALITY DIMENSIONS

- Completeness
  - Completeness of record total
  - NULL value

- Validity
  - Range value of each variable
  - ID number validity (compared to administrative data)

- Consistency
  - Different name with same ID Number
  - Variable consistency between data source

- Integrity
  - Is every ID number in field result based on administrative area?

- Uniqueness
  - No individual duplication between regions

- Timeliness
  - The latest information has a higher confidence level
MAIN PROBLEMS

DATA ASSESSMENT

Data Duplication

Invalid ID Number

Unidentified
DOS: Record that exist in all data source

D: Record that only exist in Administrative Data
O: Record that only exist in Online Population Census
S: Record that only exist in Field Enumeration

DO: Record that only exist in Administrative Data and Online Population Census
DS: Record that only exist in Administrative Data and Field Enumeration
OS: Record that only exist in Online Population Census and Field Enumeration
Lesson Learned [1]

Planning Process
✓ Careful and detailed planning, not easily changed
✓ Instrument identification with fast and accurate decision

Mitigation
✓ Preparing mitigation and backup plans in every stages
✓ Explore the latest data collection methods, including methods of delivering material to prospective census/survey officers, which are adapted to the development of scientific and technological advances accompanied by strengthening risk mitigation for these modes
Implementation

- Detailing the implementation of activities in accordance with the output to be achieved
- **Strengthen the cooperation and collaboration with ministries and agencies in building statistics within the framework of strengthening the National Statistical System (SSN)**
- Optimizing the use of administrative data for census and survey activities with the main objective of building One Data Indonesia
- Strengthening community participation to participate in BPS census and survey activities, through statistical education from an early age
THANK YOU!

Sensus Penduduk 2020

#MencatatIndonesia