Sensus Penduduk 2020
#MencatatIndonesia

Indonesian Population Census 2020 Editing and Validation

BPS – Statistics Indonesia

Sesric Webinar Series On Statistical Experience Sharing :
"2020 Round of Population & Housing Censuses: Census Editing & Validation"

Jakarta, 7 September 2022

POPULATION CENSUS 2020

SHORT FORM PC 2020

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 with PAPI

LONG FORM PC 2020

Provides demographic parameters (fertility, mortality, and migration) as well as other individual characteristics for **population projection** and **SDGs indicators.**

▶ **May - June 2022 (field enumeration)**

DATA COLLECTION MODE

Main Mode : CAPI & PAPI
Mitigation Mode : CATI & CAWI

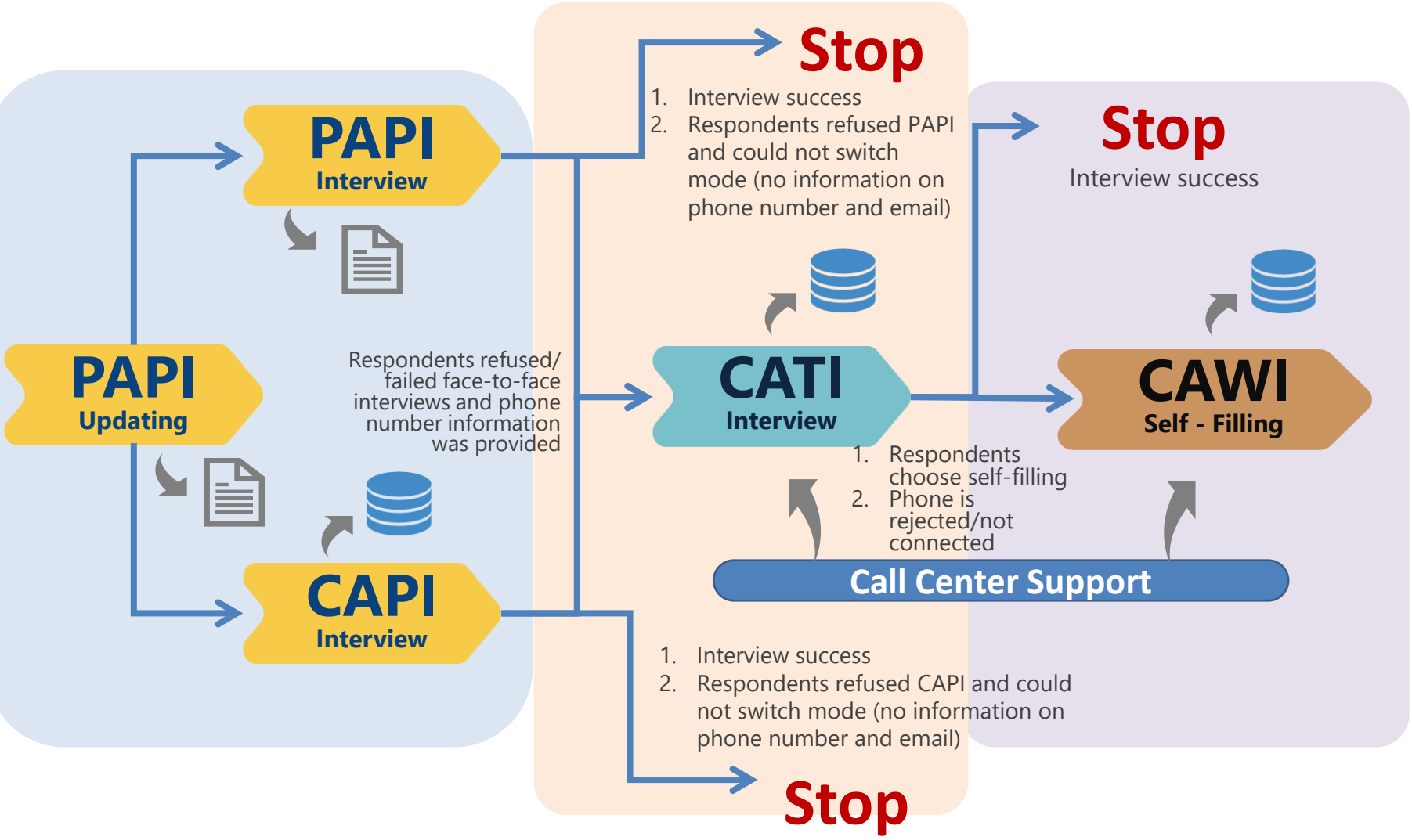
SAMPLES AND LEVEL OF ESTIMATION

Sample of 5% household in Indonesia (4,3 million HH)
Level estimation: regency

VARIABLES COLLECTED

83 questions

IMPLEMENTATION OF MULTIMODE DATA COLLECTION



CATI & CAWI is used to mitigate respondents who refuse face-to-face interviews or who cannot be found

Dashboard With Dynamic Weighting
(as A Data Quality Monitoring and Early Warning System)

1. Monitor all stages from preparation to data processing in Long Form.
2. Generating tables and indicators quickly thereby shortening processing time. Preliminary tables and indicators will be displayed in real time on the dashboard.
3. The dashboard adopts **dynamic weight**. Weight calculation is conducted dynamically based on the submitted data.
4. Data will be loaded and displayed in dashboard regularly regarding field data collection progress.

Perdesaan, dan Jenis Kelamin

						Daerah Perkotaan+Pedesaan			
						+Perempuan	Laki-laki	Perempuan	Laki-laki+Perempuan
						1.327.115	2.313.003	2.213.930	4.526.93
						48		48	4
						31		55	5
						85		85	8
						1.574.277	2.829.284	2.723.220	5.552.50
						1.774.686	3.254.851	3.078.430	6.333.28
						1.849.364	3.476.195	3.323.863	6.800.05
						1.930.644	3.602.133	3.438.310	7.040.44
						1.575.873	2.805.432	2.716.071	5.521.50
						1.707.899	3.076.014	2.944.770	6.020.78
						1.705.397	3.081.513	2.944.995	6.026.50
						1.716.074	3.101.646	2.965.330	6.066.97
						1.790.664	3.241.589	3.107.248	6.348.83
						1.758.227	3.183.790	3.009.360	6.193.15
11	2.165.828	2.066.351	4.232.179	874.806	814.768	1.689.574	3.040.634	2.881.119	5.921.75
12	2.181.391	2.067.692	4.249.082	881.711	821.469	1.703.179	3.063.101	2.889.160	5.952.26
13	2.194.294	2.087.857	4.282.151	884.472	823.785	1.708.257	3.078.766	2.911.642	5.990.40
14	2.193.881	2.093.074	4.286.954	879.067	820.327	1.699.394	3.072.948	2.913.401	5.986.34
15	2.245.062	2.150.354	4.395.416	925.008	861.233	1.786.241	3.170.070	3.011.587	6.181.65

TABULATION MENU AND INDICATOR

Field data collection monitoring progress

Tabulation design:
95 preliminary tables with RSE, age pyramid

Methodologi*:*
Sampling calculation template
RSE calculation procedure

Demographic indicators
Mortality*: IMR, CMR, U5MR, e0*
Fertility: *TFR dan ASFR;* ***Sex Ratio; Dependency Ratio; % Elderly Population***

Labor/employment Indicators
employment-to-population ratio (EPR)

Migration Indicator
Recent Migrant

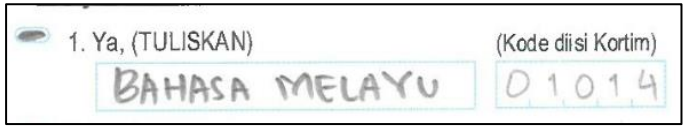
Editing & Coding PAPI Document Before Data Processing

a. Correction of handwriting in document for further processing (scan or entry).

b. Consistency between variables

Example : age less than 10 years old but graduated from junior high school

c. Code checking, such as region code, language code, ethnicity code, occupation code, etc



d. Fairness of the contents

Example : number of children ever born according to the age of woman

e. Consistency between documents (updating document, main document and verification document)

EXAMPLE OF INCONSISTENT/ANOMALY DATA

a. Age

- Out of range : more than 95 yo (maximum age in LF is 95)
- Age with decimal
- Missing age

We check with date of birth or identity (ID) number

b. Sex, relationship with the head of household, and marital status

- If the spouse has the same sex with the head of household
- If the marital status of husband or wife is “not married”

c. Age, Sex, and Marital Status

- Wife but sex is male
- If age of married is less than 10 yo
- Mother but the sex is male or never married

d. Number of Children Ever Born, Survive, and Death

- If the number of children ever born is not equal to the sum of children surviving and dead children.

e. Age and the highest completed level of education

- Age is 10 but completed level of education is primary school, etc

Etc……

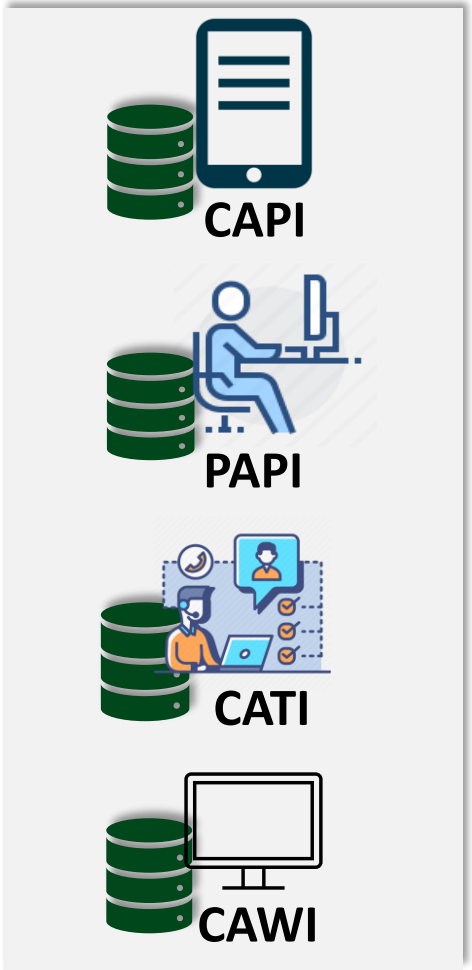
DATA PROCESSING

Sensus Penduduk 2020

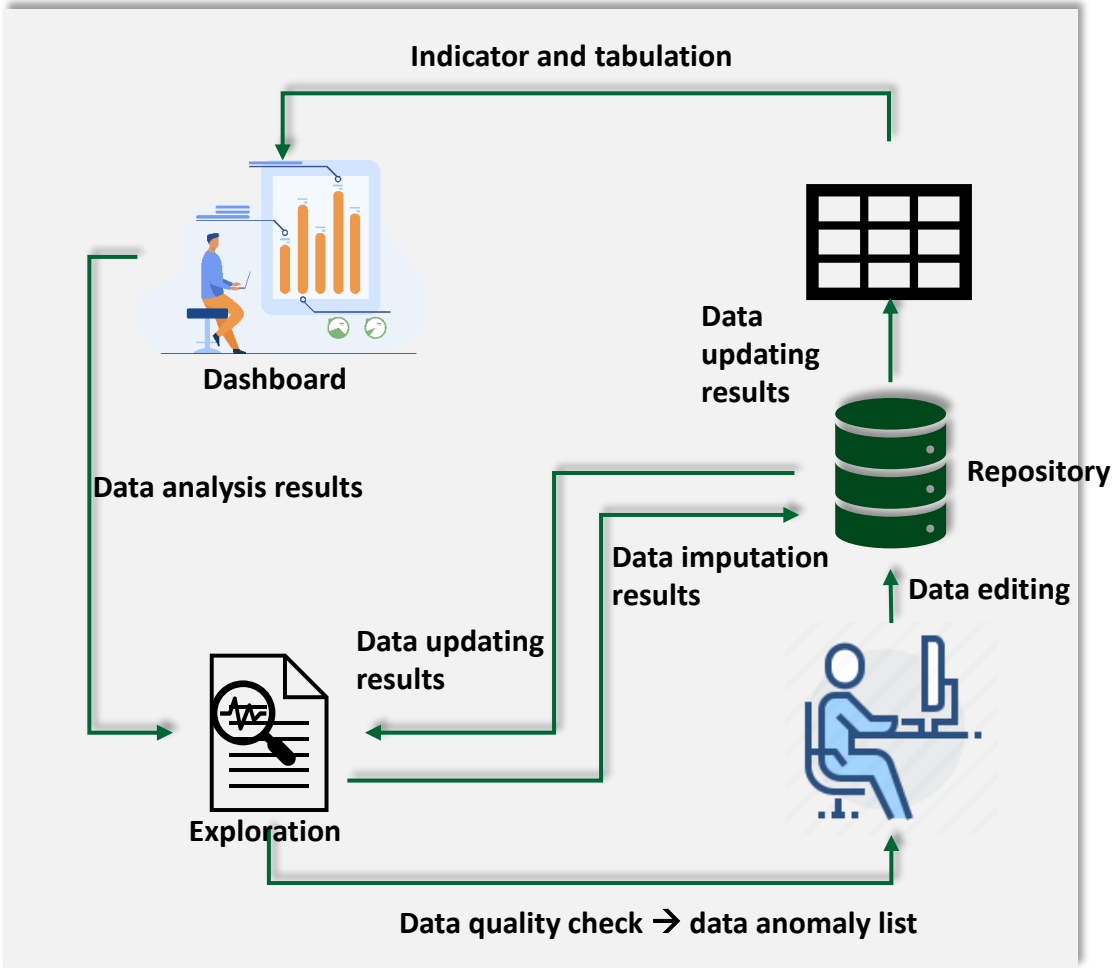
#MencatatIndonesia

DATA PROCESSING FLOW

DATA SOURCES



SINGLE SOURCE OF TRUTH

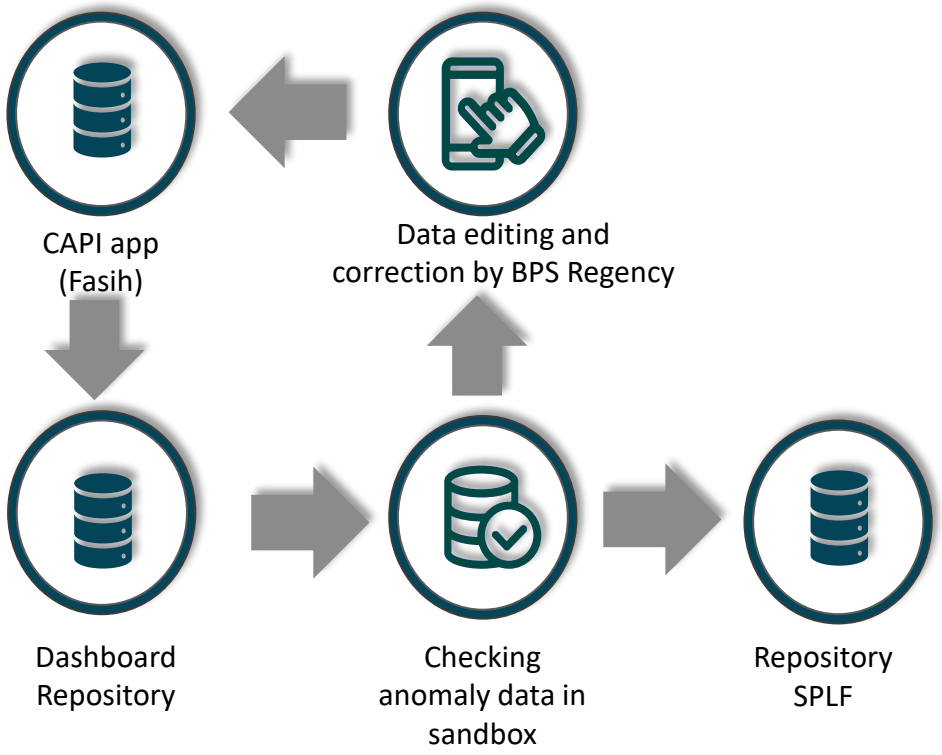


DISSEMINATION

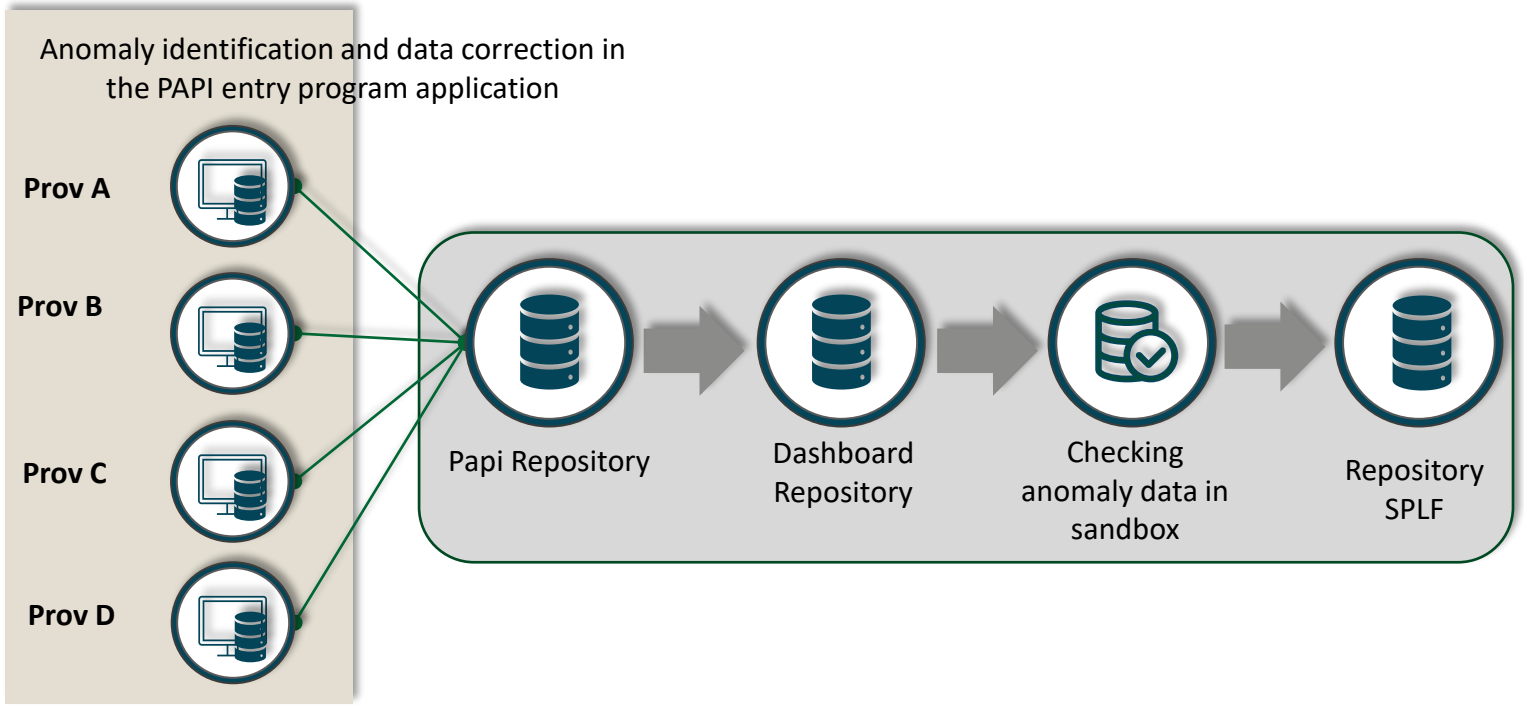
- Access data from data warehouse
- Published indicator and tabulation must follow the rule of anonymization and disclosure control
- Data is related with metadata

DATA EDITING

CAPI



PAPI



DATA IMPUTATION

Validation Rule
- Range value of variable
- Consistency between variables

Imputation Rule
- Validation Rule
- Value that used for imputation

Single Source of Truth

Data Sandboxing

- Access data from single source of truth for exploration
- Create a script based on rule validation
- Run the script to get anomaly recapitulation for each variable and each record
- List of anomalies and record identity will be used in data editing
- Revalidation and data imputation

Example of Recapitulation Table

Record	Anomaly Status			
	Gender not in (1,2)	Marital status not in (1,2,3,4)	Age not between 0 and 95	....
1	✓		✓	
2	✓	✓	✓	
3		✓	✓	

EXAMPLES OF DATA IMPUTATION

Case I: Ensure value of variable in raw data comply with rule validation

Example:

1. Range value of sex variable is 1 (male), 2 (female)
2. Range value of age is 0 – 95. The value should be consistent with month and year of birth

BLOK III. SUSUNAN ANGGOTA RUMAH TANGGA

301. No. Urut	302. Nama Anggota Rumah Tangga (TULISKAN NAMA SEMUA ANGGOTA RUMAH TANGGA MULAI DARI KEPALA RUMAH TANGGA)	303. Hubungan dengan Kepala Rumah Tangga (KODE)	304. Jenis Kelamin 1. Laki-Laki 2. Perempuan	305. Tanggal, Bulan, dan Tahun Lahir	306. Umur (TULIS DALAM TAHUN)	307. Status Perkawinan 1. Belum Kawin 2. Kawin 3. Cerai Hidup 4. Cerai Mati	308. Nomor Urut Ibu Kandung (TULISKAN '00' JIKA IBU KANDUNG BUKAN ART)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1		0 1					
2							
3							
4							
5							
6							
7							
8							
9							
10							

Rule:

- ✓ IF gender value is null OR r304 NOT IN (1,2) THEN false ELSE true
- ✓ IF age NOT BETWEEN 0 AND 95 AND month of birth BETWEEN 1 AND 6 THEN 2022-YEAR of Birth ELSE age

Case II: Ensure consistency between variable and record

Example: check the consistency for total of household member in recapitulation part (r112) and number of person in individual question (number of rows in third part)

112. Jumlah Anggota Rumah Tangga (ART) :

BLOK III. SUSUNAN ANGGOTA RUMAH TANGGA

301. No. Urut	302. Nama Anggota Rumah Tangga (TULISKAN NAMA SEMUA ANGGOTA RUMAH TANGGA MULAI DARI KEPALA RUMAH TANGGA)	303. Hubungan dengan Kepala Rumah Tangga (KODE)	304. Jenis Kelamin 1. Laki-Laki 2. Perempuan	305. Tanggal, Bulan, dan Tahun Lahir	306. Umur (TULIS DALAM TAHUN)	307. Status Perkawinan 1. Belum Kawin 2. Kawin 3. Cerai Hidup 4. Cerai Mati	308. Nomor Urut Ibu Kandung (TULISKAN '00' JIKA IBU KANDUNG BUKAN ART)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1		0 1					
2							
3							
4							
5							
6							
7							
8							
9							
10							

Rule:

IF r112 <> number of rows in third part THEN false ELSE true

TECHNOLOGY USE IN CENSUS EDITING AND VALIDATION

STANDARDIZATION

Data from multi-mode data collection will be integrated in general data structure through standardization process, such as mapping of variable, data type, column length, etc.

- ✓ RDBMS
- ✓ Extract, Transform, Load (ETL) Tool
- ✓ Metadata Management System

SINGLE SOURCE OF TRUTH

Standardized data will be stored in centralized repository that also has role as processing environment in data integration

- ✓ RDBMS

DATA PROCESSING

It includes data validation, editing, agregation, and tabulation. Engine processing will be used to perform data processing.

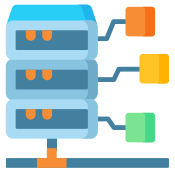
- ✓ RDBMS
- ✓ ETL Tool
- ✓ Engine Processing
- ✓ Data editing tool

VISUALISATION

Display the indicators and tabulation that will be analyzed for data quality checking

- ✓ RDBMS
- ✓ Analysis and Visualization Tool

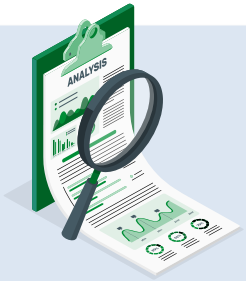
CHALLENGES



Understanding the differences of data structure and data quality in each of data sources

Each of data collection mode has difference standard to ensure the data quality that usually constraint with performance requirement. Therefore, revalidation is needed to get list anomalies as input for data editing

OUTPUT LF PC2020



In 2023
BRS, Booklet, Website

*contains **important indicators as follows**:*

FERTILITY
CBR; TFR; ASFR

MORTALITY
CDR; ASDR; IMR;
CMR; U5MR; MMR

MIGRATION
Lifetime migration, total migration, risen migration, international migration, commuter, migration stock

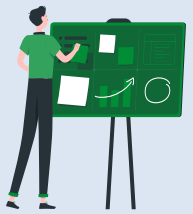
Employment
residents work by business field, type of work, employment status

Disability
Population by disability level

Education
highest level of education, language

Housing
the widest main type/material of roof, wall, floor, land

Micro Data



Thematic Publications (fertility, mortality, migration)

BADAN PUSAT STATISTIK

THANK YOU!

Sensus Penduduk 2020

#MencatatIndonesia

Pictures from: https://www.pexels.com/photo/asian-children-bridge-children-clouds-2132062/