BADAN PUSAT STATISTIK

Sensus Penduduk 2020
#MencatatIndonesia

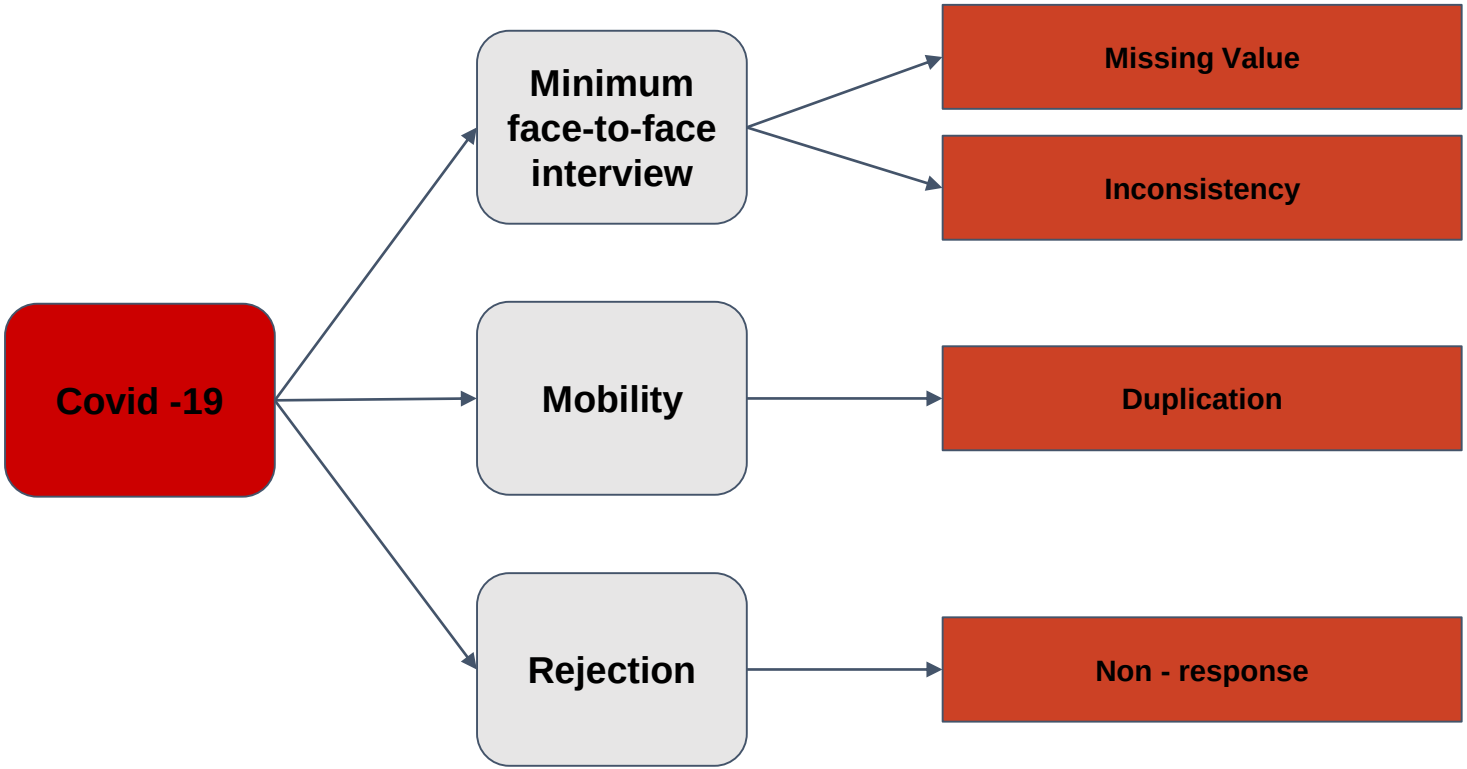
Source:
http://www.divorcelawyer.id/blog/item/64-child-custody-in-indonesia

2020 Population Censuses Data Processing

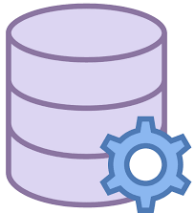
Statistics of Indonesia

United Nations Expert Group Meeting, 9-12 February 2021

Covid-19 Pandemic



Data Sources and Profiles



Administrative Data

292.979.718 record / 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

51.360.072 record

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 record

Attributes :

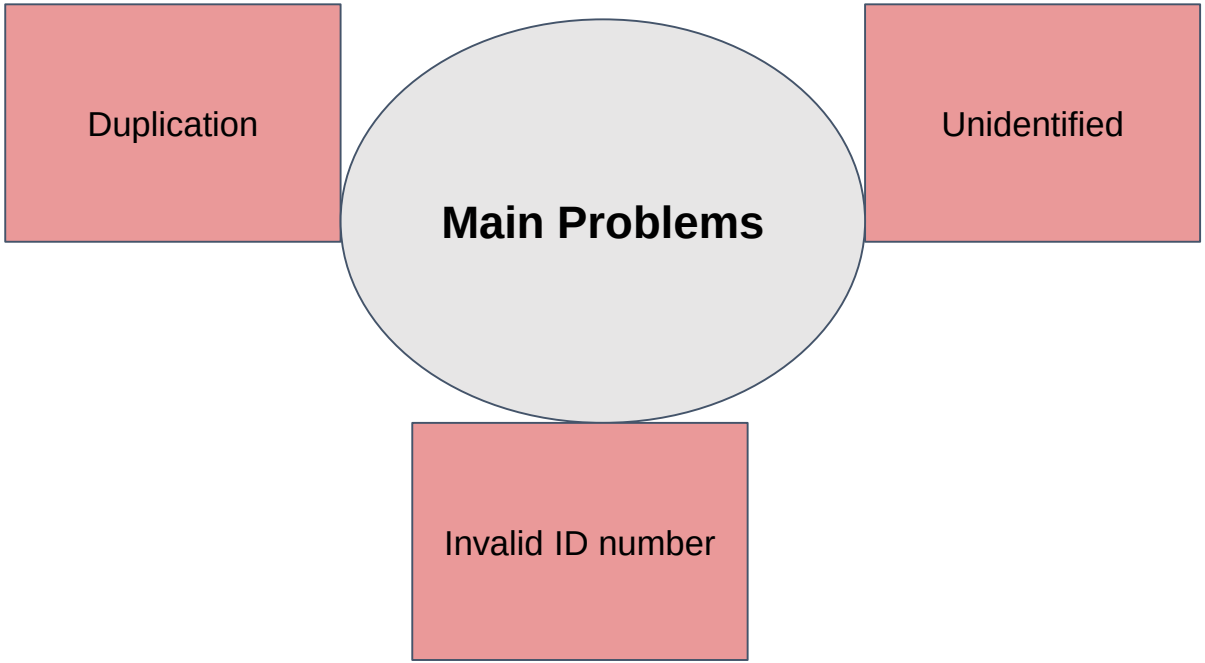
- ID number
- Family card number
- Full name
- Sex
- Family relationship
- Address

Data Quality Assessment

Data Quality Dimension

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

Data Quality Assessment



What Do We Do?

Data Source Profiling

Data structure, data type, data summary, interrelationship

01

02

Data Quality Assessment

Validity, consistency, completeness, uniqueness, integrity, etc

Define Algorithm and Cleansing Rule

Data matching, data pairing

03

04

Data Cleansing

Remove duplicate, finding individual, etc

Evaluation

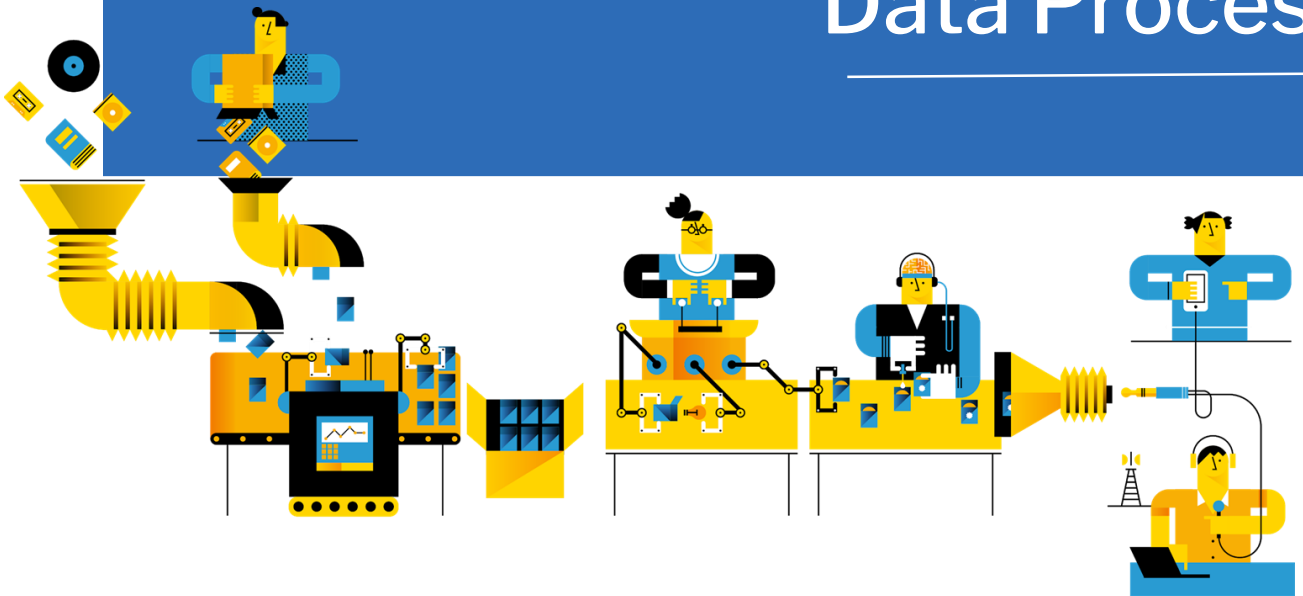
Data checking

05

BADAN PUSAT STATISTIK

Sensus Penduduk 2020
#MencatatIndonesia

Data Processing Details



Defining Processing Strategy - Algorithm and Rules

01
Levelling validity of ID number based on administrative data

02
Running matching algorithm for record with valid ID number

03
Running pairing algorithm for specified record

04
Integration of matching and pairing result

05
Remove duplication because of integration

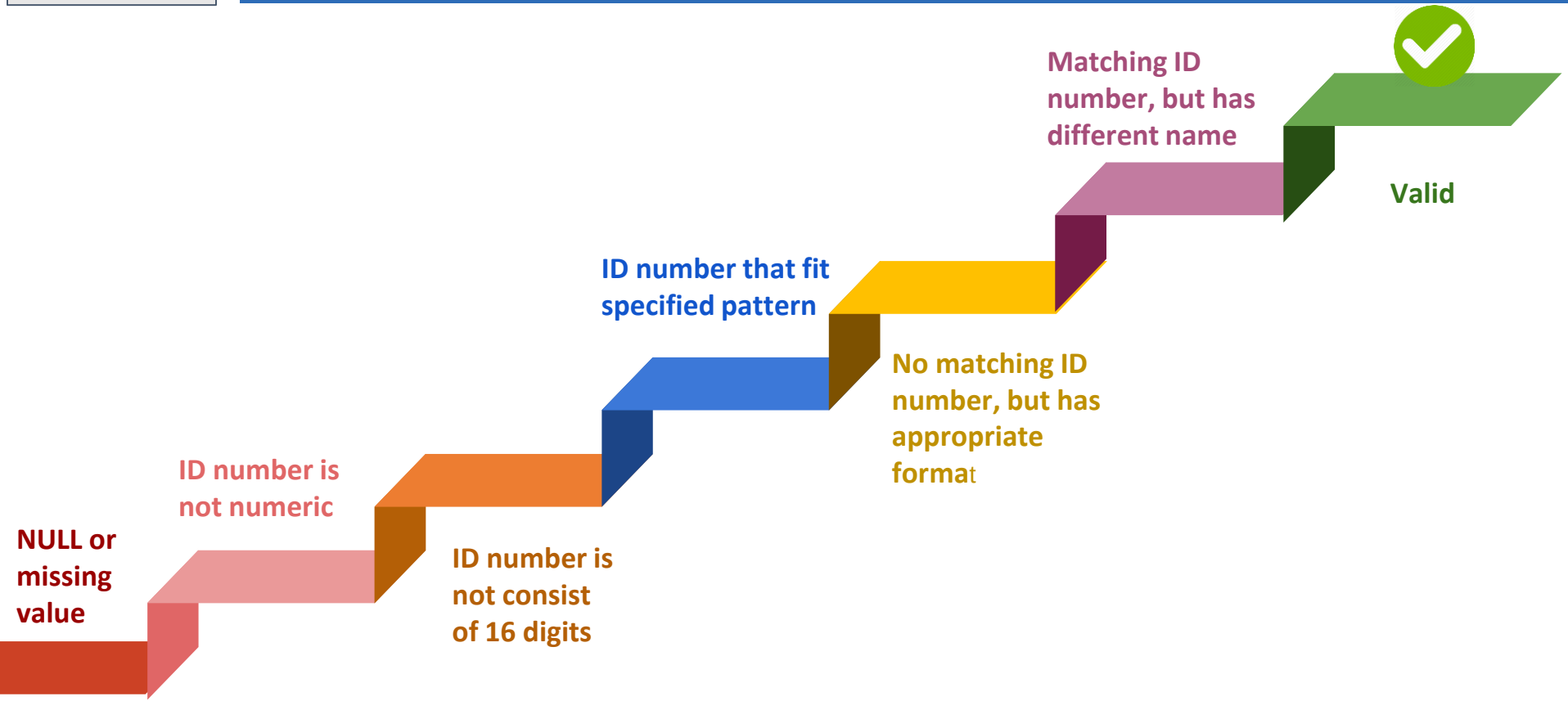
06
Consolidate with Department of Population and Civil Registration

07
Finalizing data

08
Continues Improvement

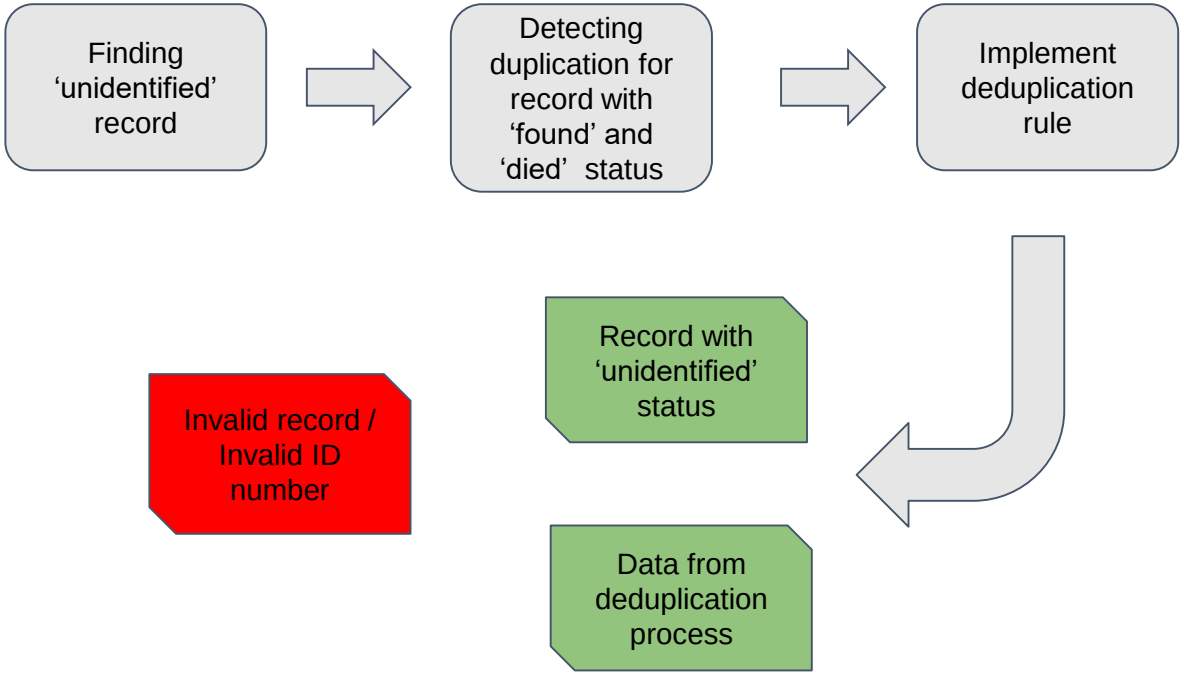
01

Levelling validity of ID number based on administrative data

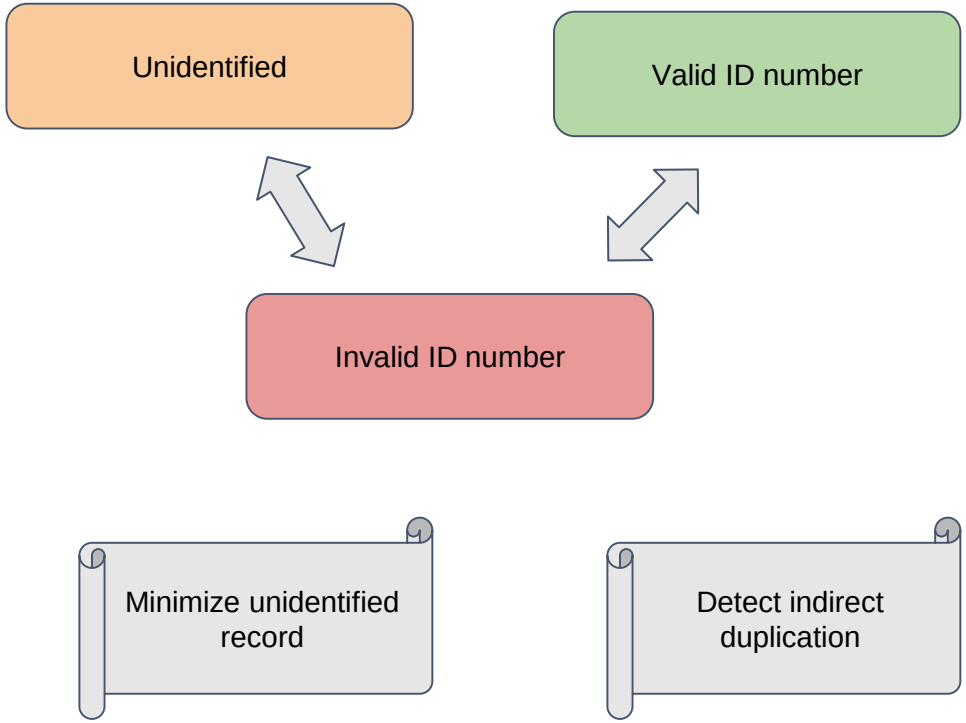


02

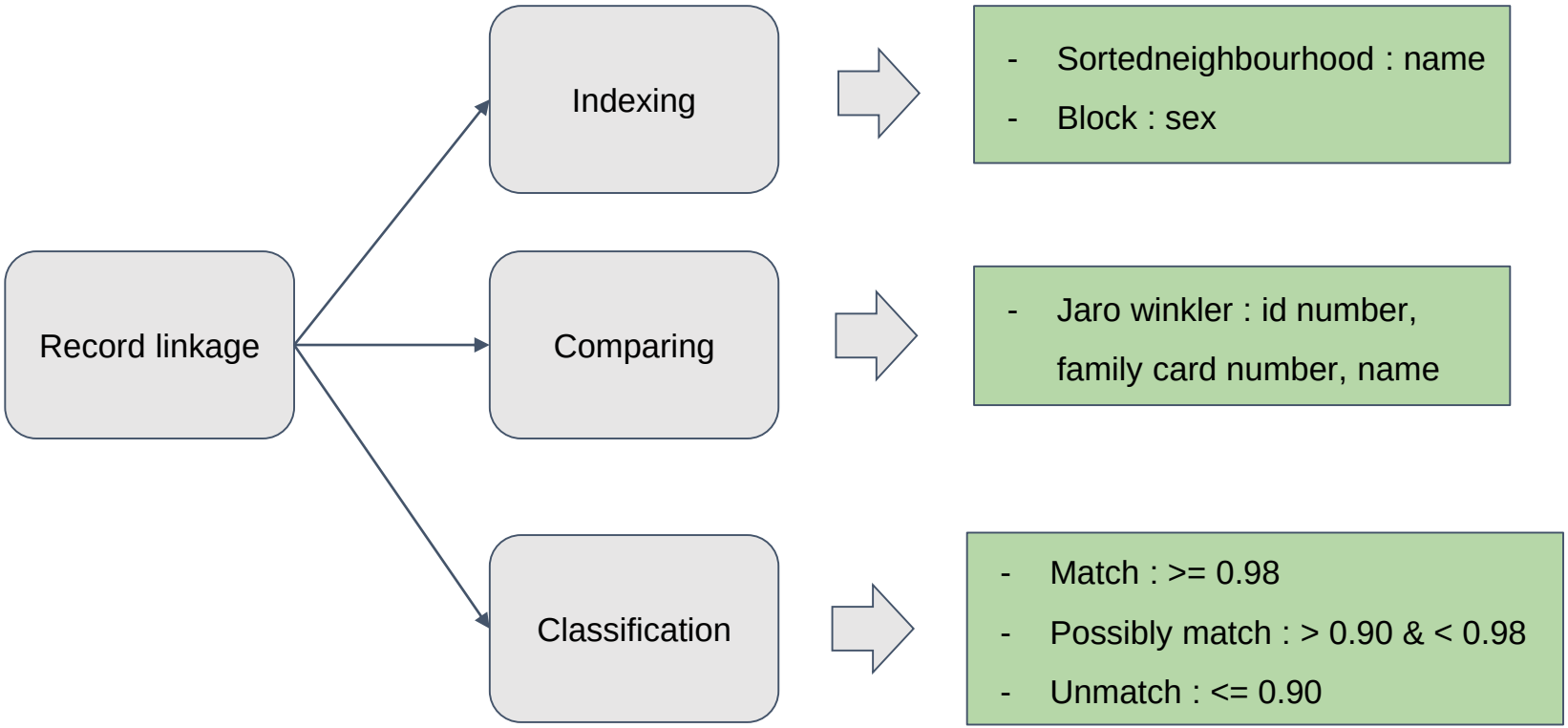
Matching Algorithm



03 Pairing Algorithm



03 Pairing Algorithm



03 Pairing Algorithm

kk_x	nik_x	nama_x
1272030806170001	1208130009850010	HOTMAULI PARLINDUNGAN S.M MANILLANG

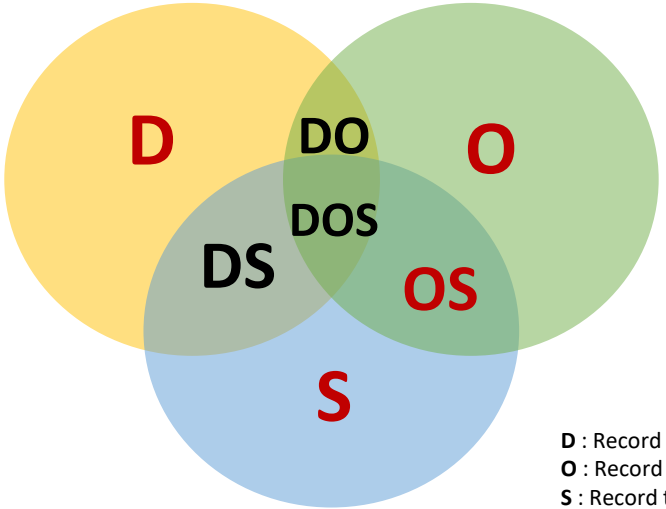
kk_y	nik_y	nama_y
1272030806170001	1208013009850010	HOTMAULI PARLINDUNGAN S.M. MANULLANG

score_kk	score_nik	score_nama	mean_score
1	0,9875	0,983174603	0,990224868

- Match : >= 0.98
- Possibly match :

  > 0.90 & < 0.98
- Unmatch : <= 0.90

Final Result



ONE DATA PLATFORM

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

DOS : Record that exist in all data source

Our Team

Ade Koswara
(Team Leader)

Kemas Muhammad Irsan Riza
(Integrator)

Amin Fathullah
(ETL and Dashboard)

Wa Ode Zuhayeni Madjida
(Data structure and Matching Alg.)

I Komang Yudi Hardiyanta
(Data structure and Matching Alg.)

Khalilur Rahman
(Pairing Algorithm)

Gesang Wibawono
(Pairing Algorithm)

Miswar
(Infrastructure)

Thank You

Sensus Penduduk 2020

#MencatatIndonesia