Data integration for disaggregated statistics on international migration

Training Workshop on the Use of Administrative Data for Migration Statistics in Maldives

10-12 December 2024, Malé, Maldives

Elena De Jesús, UNSD

United Nations | Statistics Division

Why data integration?

International student

How data integration can improve migration statistics

- Improves coverage and richness of migration data (including hard-to count populations)

- Provides more timely, accurate, and granular data than relying on a single data source

- Fills in missing characteristics

- Reduces operational costs compared to new data collection

- Facilitates understanding of complex migration patterns

- Improve data quality ("triangulation")
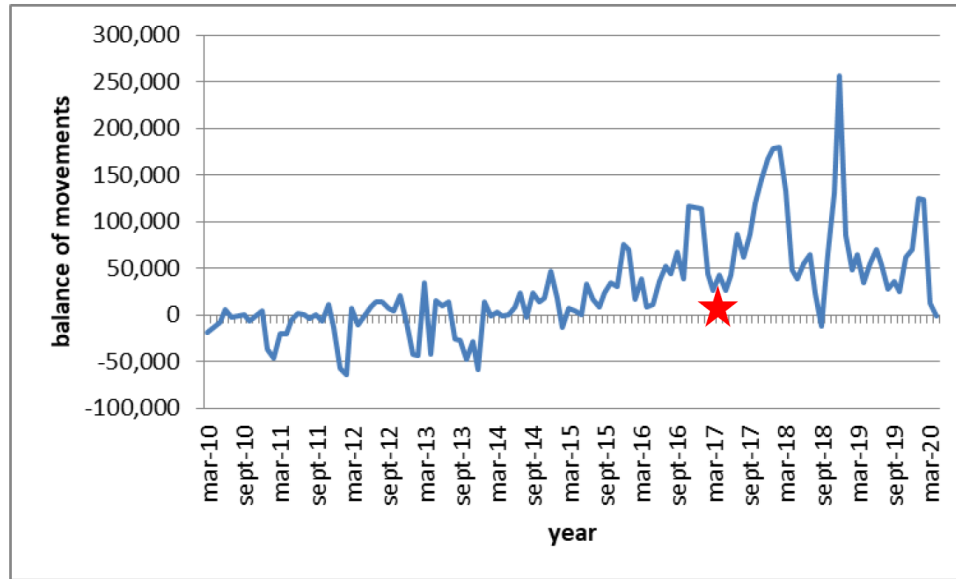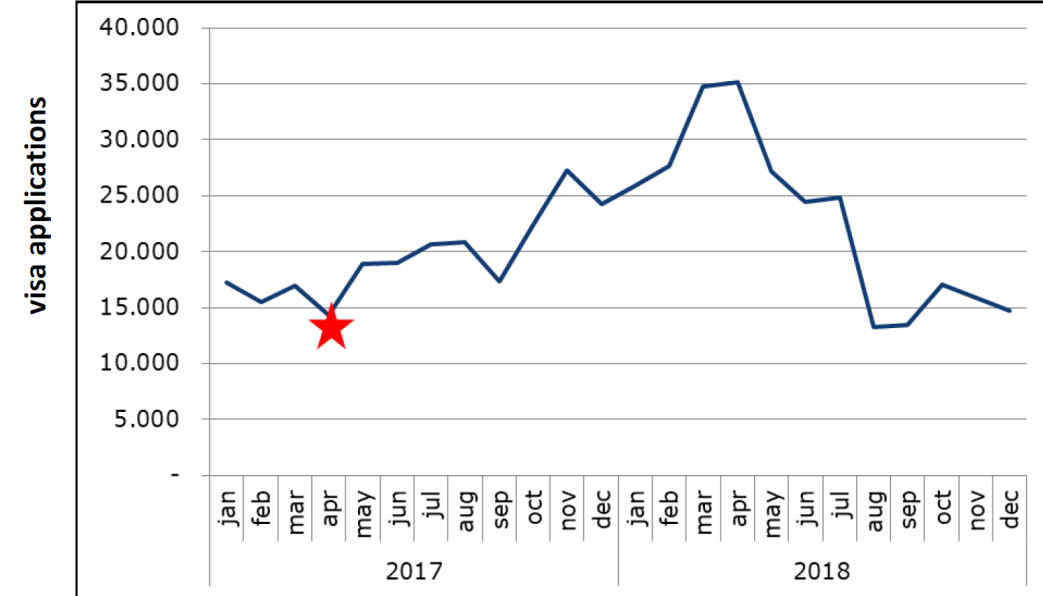
- Facilitates longitudinal analysis

Why data integration? Chile example

Graph 2. Balance of entries and exits, Chile (2010-2020)

Source: PDI

Graph 3. Applicants for residence permits inside the country, Sermig 2017-2018

Source: Sermig

1. **Significant increase in immigration after 2017 census**
2. **Different official figures for the stock of immigrants in Chile in 2018**
3. **Need for information to generate public policies**

Data integration definitions

DATA INTEGRATION

Combining data from multiple sources to create enriched datasets

MICRO-DATA INTEGRATION
creates new combined datasets, which can produce statistics

- individual-level linkage of microdata

- feasible when record-level data is available and linkable with key identifying variables

- e.g. population registers

MACRO-DATA INTEGRATION
creates new statistics from outputs from multiple datasets

- aggregated data combination

- e.g. demographic accounting

United Nations | Statistics Division

Macro-data integration methods

Typically used when micro-data integration is not possible.

Compilation
Pull from various existing subcomponents

Statistical adjustment
Adjust numbers based on strongly correlated input data
For example, trends in one data set inform another (flight, visa, border, etc.)

Statistical modeling
- Information from one dataset used to supplement/enhance data missing or of low quality in a second dataset
- Shared characteristics in both data sets can be used to estimate characteristics missing from one of the datasets

Other methods
- Hybrid (macro- and micro-)
- Big Data

United Nations | Statistics Division

Macro-data integration methods (continued)

Challenges for macro-data integration

- Internal consistency (differences between integrated data sources)
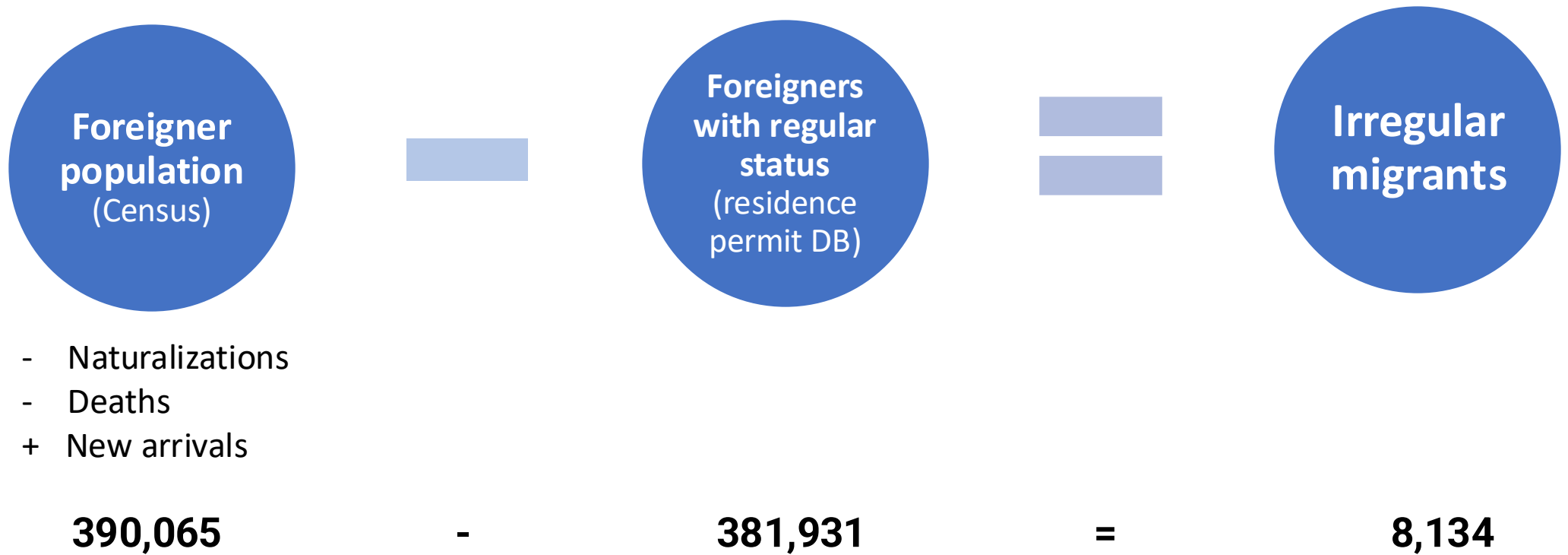- Estimate validation (finding the "true" value)

Mitigation
- Make adjustments prior to integration
  - Temporal cutoff
  - Harmonization of definitions

Macro-data integration methods (example)

Mexico Case Study: Estimate of irregular migrant stocks

Foreigner population (Census)

Foreigners with regular status (residence permit DB)

Irregular migrants

- Naturalizations
- Deaths
+ New arrivals

390,065   -   381,931   =   8,134

United Nations | Statistics Division

Micro-Data Integration Methods

Linking individual data at the record level

Creating / enabling the legal framework
- Govern the transfer of personally identifiable information/data-sharing agreements
- Ensure data confidentiality
- Dictate use and retention of linked data

Creating / enabling the technical framework
- Financial / technical resources
- "Spine" selection / creation
- Selected variables / events
- Data updates, data validation, etc.

Micro-Data Integration Methods (continued)

DETERMINISTIC

"exact", based on common identifier or set of variables

PROBABILISTIC

create unique identifier based on likelihood the same person

Hybrid linkage approaches common

- Post-linkage evaluation / cleaning of matches (co-mingled data, i.e. multiple people combined into one person, multiple records for same person)

- "Signs of life" principle used to ascertain the presence of an individual in a country through activities such as tax filling, work, or education

United Nations | Statistics Division

Micro-Data Integration Challenges

Challenges for micro-data integration

- Data access

- Data privacy

- Sufficient personal identifiable information for matching

- Unequal coverage

- Data quality

United Nations | Statistics Division
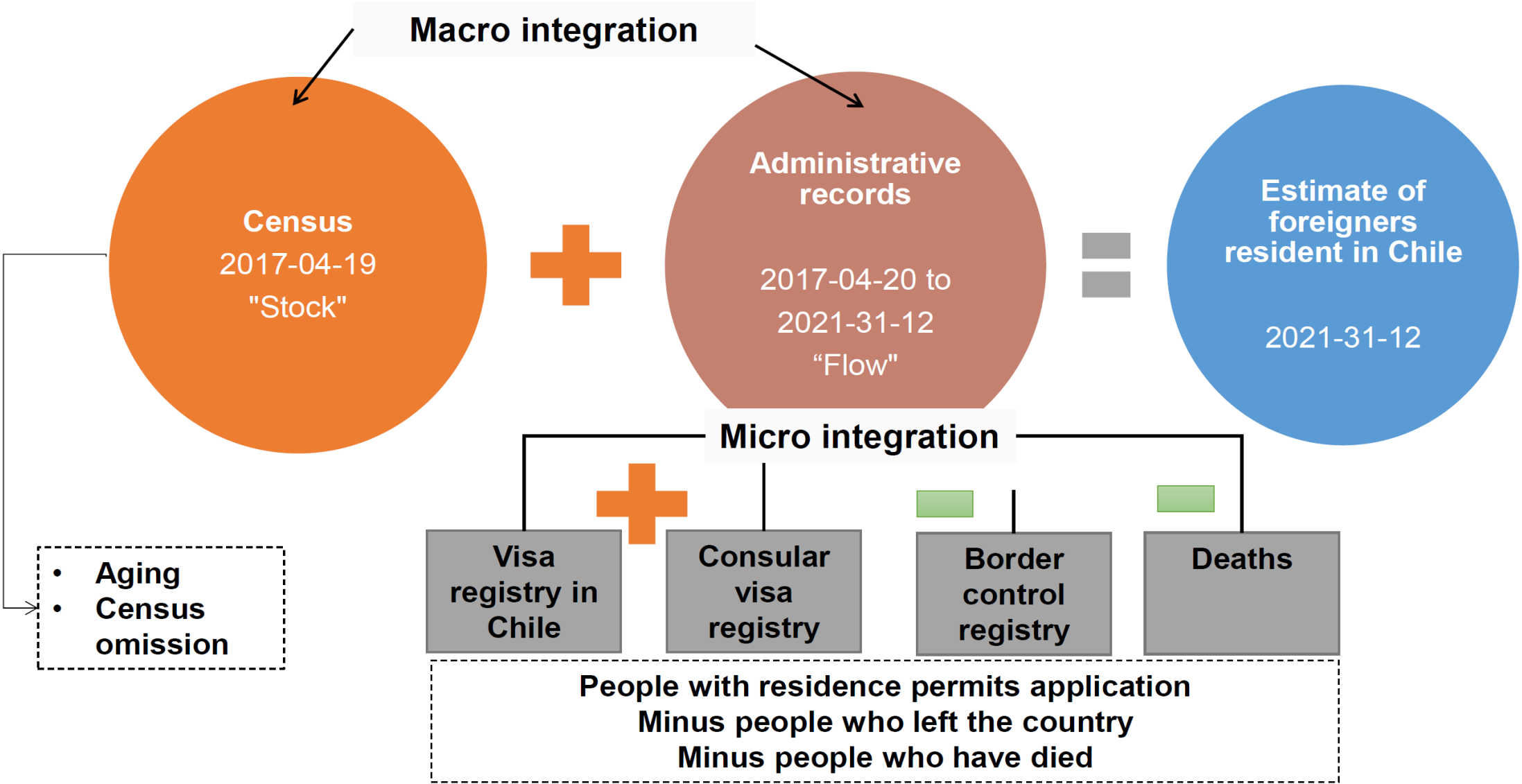
Micro-Data Integration Methods (example)

United States: Micro-data integration and the Integrated Database on International Migration (IDIM)
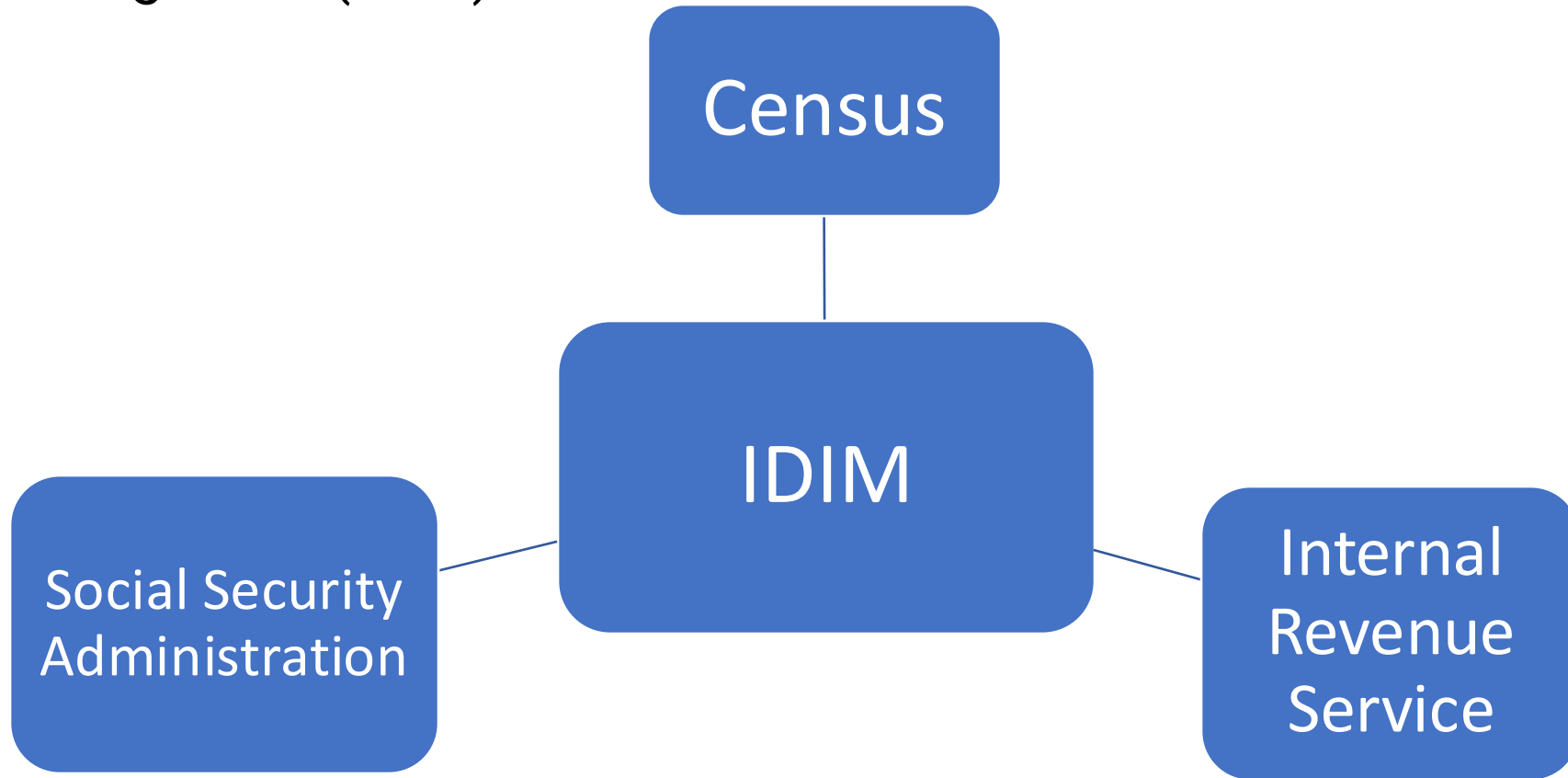
Methodology for estimating foreigners residing in Chile (2018-2021)

Technical report on data integration

UN **Expert Group on Migration Statistics**
https://unstats.un.org/unsd/demographic-social/sconcerns/migration/index.cshtml

21 countries

17 organizations

6 independent experts

5 Task Forces

Technical report that supports countries to produce sufficiently disaggregated data for the measurement of international migration by means of integration of micro- and macro-data techniques

The report covers:
    (a) variety of methodologies to integrate and analyze migration data on the macro and micro level
    (b) legal, policy and technical prerequisites
    (c) major challenges and opportunities in mainstreaming the use of integrated data in official statistics

https://unstats.un.org/unsd/demographic-social/sconcerns/migration/task-forces/Task%20Force%203%20Report-Web.pdf

United Nations | Statistics Division

Country case studies

Country case studies: Annex to the technical report

Each example of data integration is organized around six main themes:
        (1) Data needs
        (2) Collaboration framework for data integration
        (3) Data sources
        (4) Integration methodology
        (5) Data evaluation strategy
        (6) Dissemination/communication of results

Country examples: Canada, Chile, Georgia, Mexico, Moldova, Switzerland, United Kingdom, and the United States

https://unstats.un.org/unsd/demographic-social/sconcerns/migration/task-forces/TF%203%20-%20Appedix_case_studies.pdf

United Nations | Statistics Division

Thank you!

Questions?

Group work: Harmonizing variables

Is micro-data integration possible? Which datasets and variables?

Which datasets can be integrated using macro data integration methods? Using which methods?

Which variables need to be harmonized before data integration. What is needed for this to happen?

United Nations | Statistics Division