



Regional Training Workshop on Transition to Register-based Approaches for Population and Housing Censuses

Ankara, Türkiye
12-15 June 2023

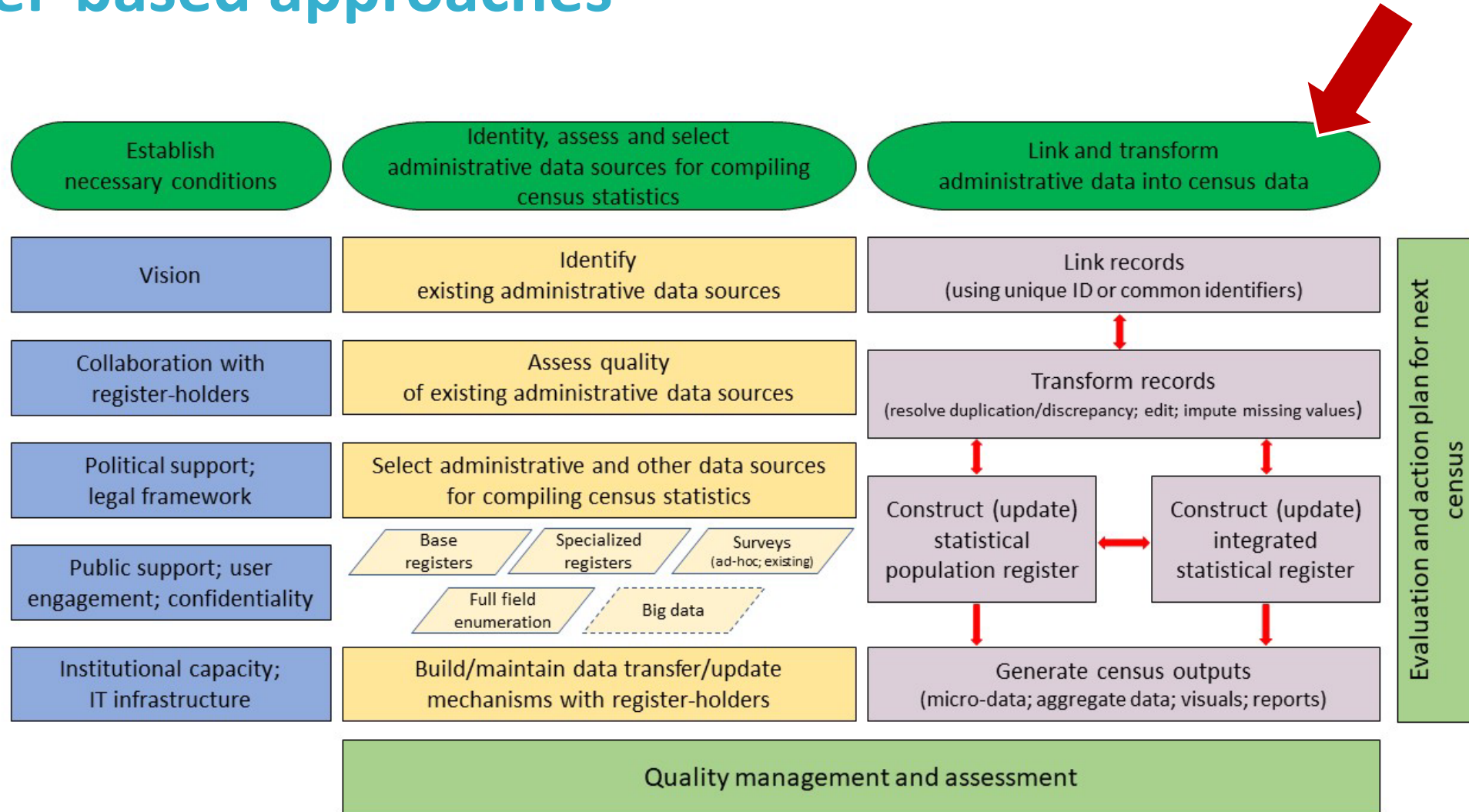
Data integration and transforming administrative data to census data

Regional Training Workshop on Transition to Register-based Approaches for Population and Housing Censuses

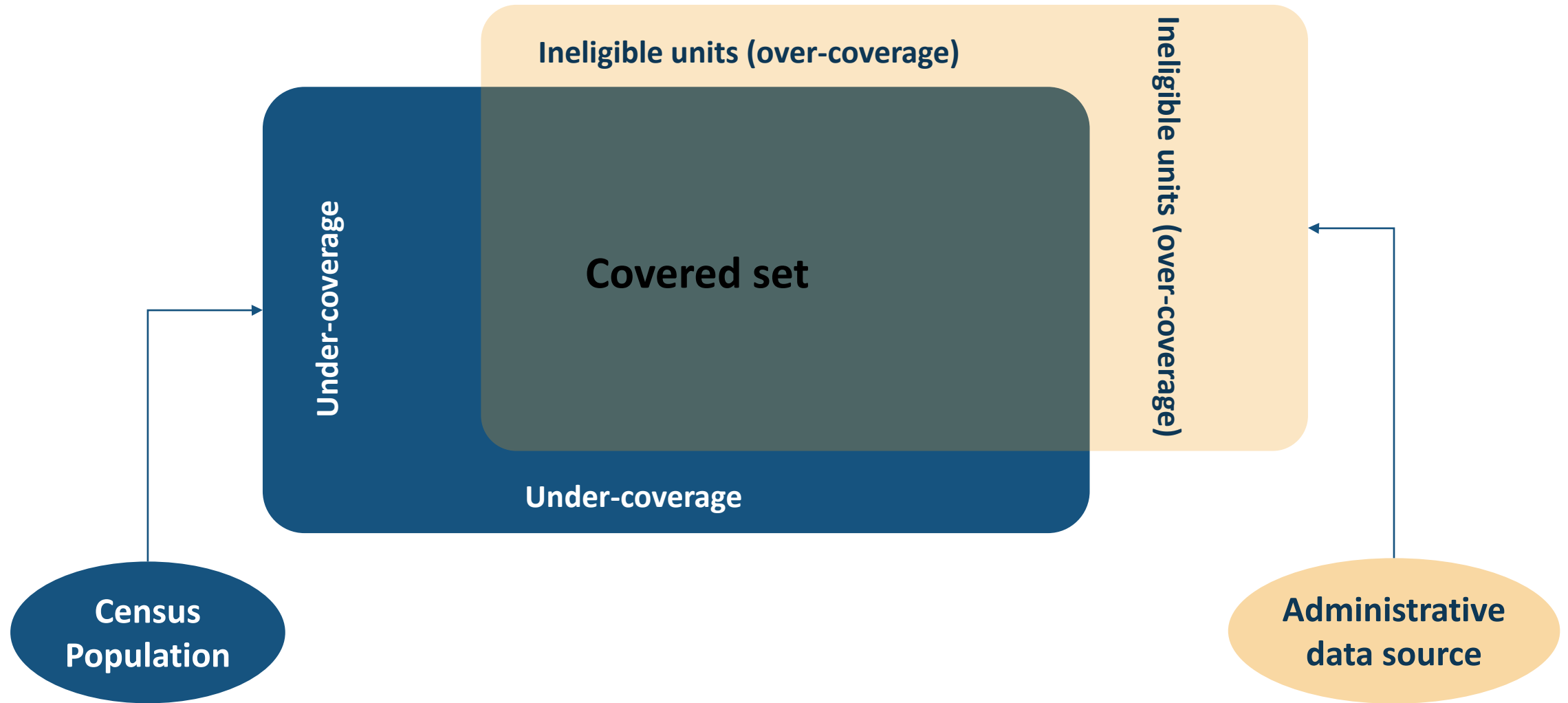
12-15 June 2023, Ankara, Türkiye

Afsaneh Yazdani, ESCAP Statistics Division

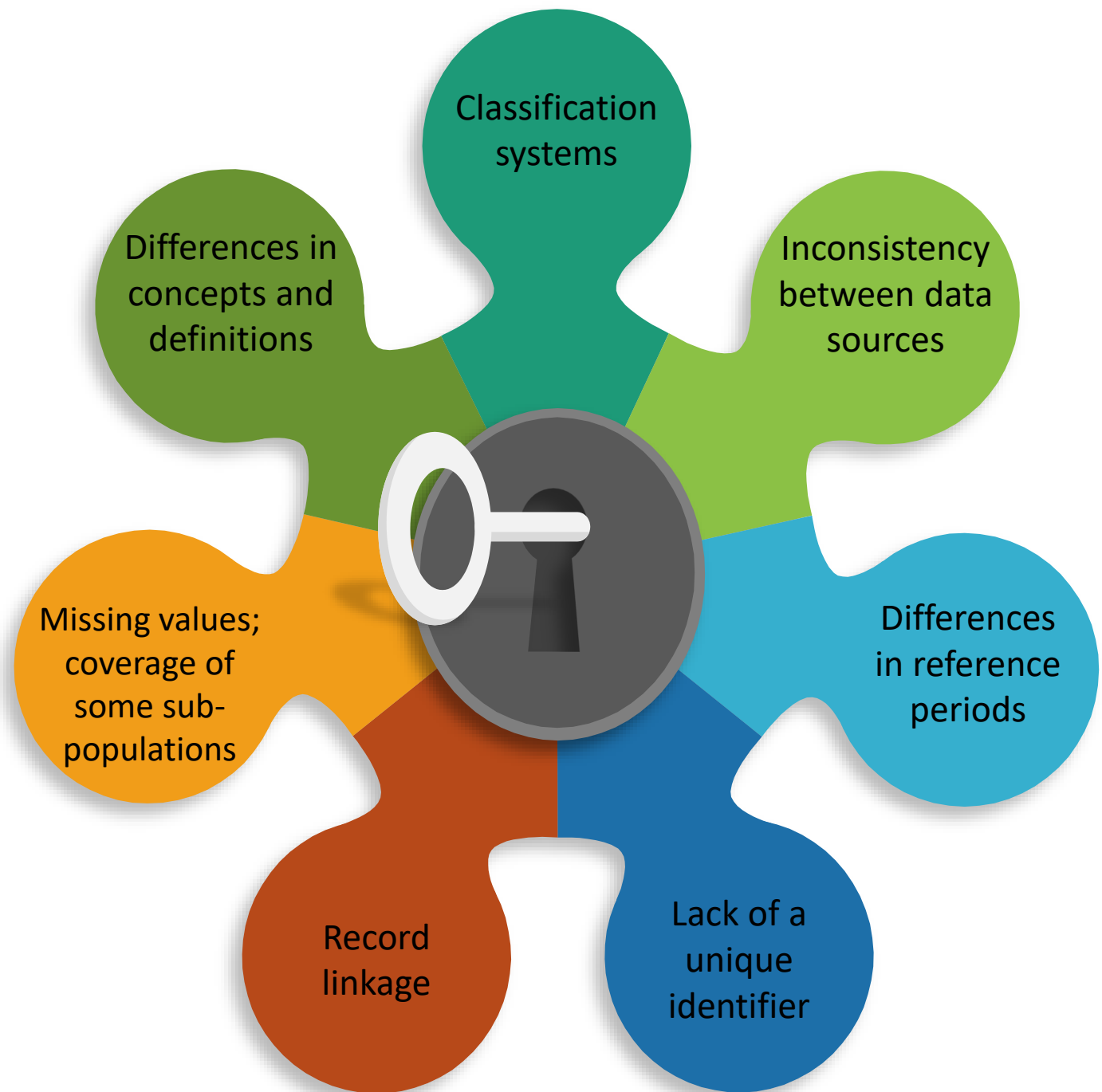
Generic model for the transition from a traditional census to register-based approaches



Coverage of census population



Some of the key technical challenges and issues of using administrative data sources



Constructing integrated statistical registers for the purpose of census

➤ In addition to identifying the data sources to be used, some of the key processes involved in the construction of a statistical register are:

- Data linkage
- Dealing with duplication
- Conflict resolution
- Updating and the “Signs of Life” methodology
- Editing and imputation



Data linkage

- Record linkage refers to the identification and combination of records corresponding to the same entities – for example, persons, enterprises, dwellings and households – throughout two or more data sources.
 - **Deterministic or exact matching** is when a formal decision rule, typically in the form of unique IDs, such as personal identification numbers (PIN), exists in data sources to be used for matching.
 - **Probabilistic matching** is when strict decision rules are not applicable. Instead, complex probabilistic decision rules are established based on a set of key variables that are common in data sources, such as name, sex, date of birth, and address, to assign similarity scores.
 - **Combined approach** can be applied, with exact/deterministic linking used first for as many records as possible, followed by probabilistic linking for the remaining records.

1. Data linkage – identification numbers

➤ “Unique identification numbers” greatly facilitate the linkage of several administrative data sources.

- The unique identifier should be common across all relevant registers.
- ID numbers are often created for administrative purposes to be used in population registers, civil registers, national identification systems or other administrative registers. Though, sometimes, they have been introduced for statistical purposes.
- The ways ID numbers are implemented vary among countries. Sometimes the number relates to the attributes of the individual, and sometimes is a unique information-free number. Sometimes include check digits. It may be issued for citizens upon reaching legal age, or when they are born.
- To increase personal data protection, ID numbers are encrypted to prevent information from being read by unauthorised parties.

1. Data linkage – linking persons to dwellings/households

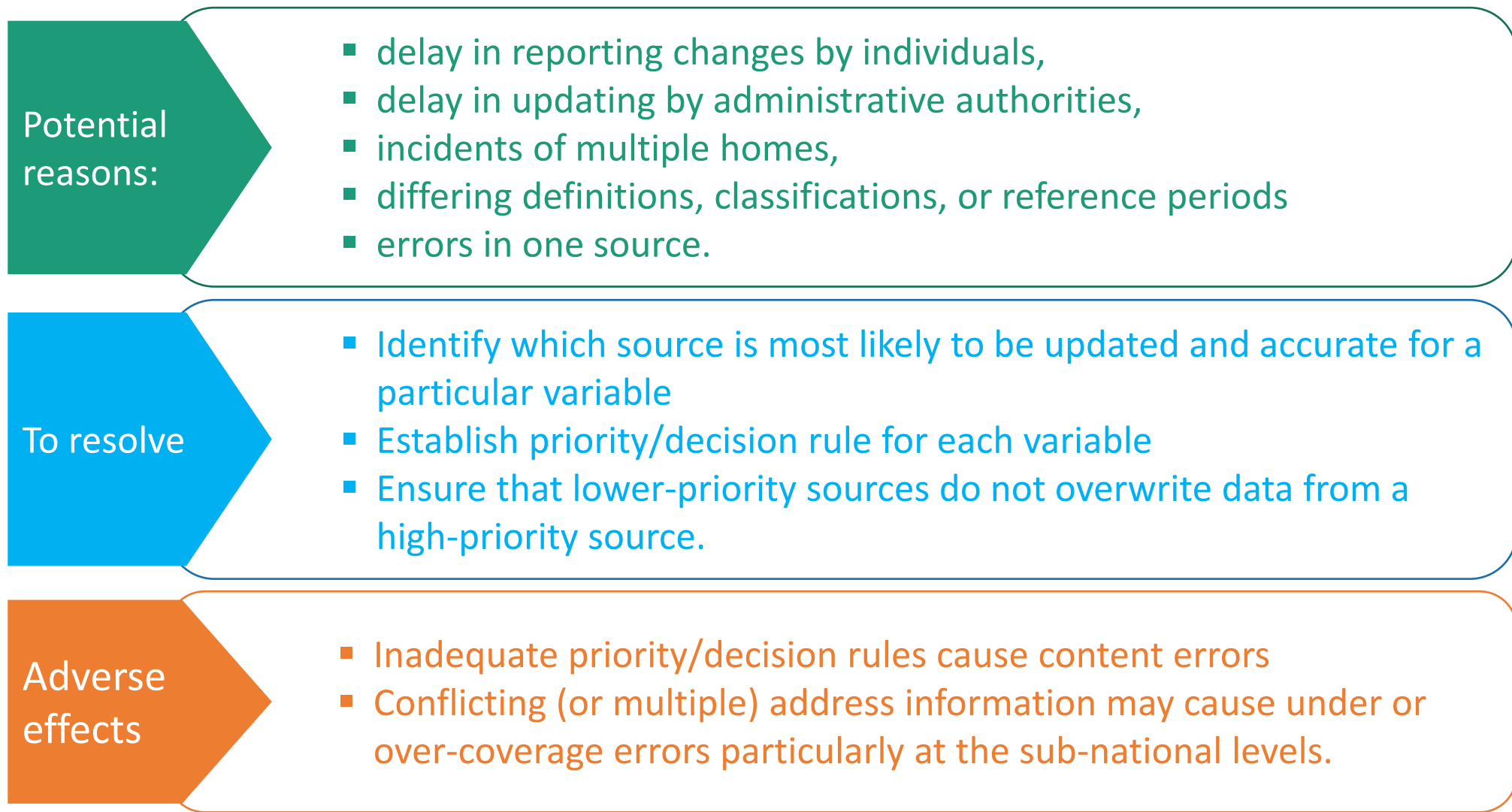
- The basic counting units of a census include persons, households, families and dwellings; this should be considered in a register-based census:
 - The minimum necessary identifiers are for persons (PIN) and dwellings (address code, and/or spatial coordinates). The dwelling ID is assigned to each person residing there.
 - In a traditional census, households are built using the “housekeeping concept”. This is challenging in a register-based census; thus, many countries instead use the “household dwelling concept” which considers all persons living in the same housing unit, as a household.
 - In some countries a distinct household register exists, which eases the process of building households.

2. Dealing with duplications

- To avoid significant coverage issues, an adequate process for detecting and removing duplications should be in place. Duplication of persons in the statistical population register can occur, if:
- Good record linkage methods are not in place.
 - Identifiers are not of good quality across data sources, causing a false match/mismatch of records.
 - Newly created units (true births or immigrants) and deleted units (true deaths or emigrants) are not well reflected in administrative data sources.

Maintaining a log of changes to the registers is helpful.

3. Conflict resolution



4. Updating and the “Signs of Life” methodology

- The basic minimum input for keeping SPR up-to-date is the information obtained from the civil registration of births, deaths, marriages, and of any changes of address resulting from either internal or international migration.



SPR should cover the census population, i.e., only include persons who are alive and meet a set of predefined residency criteria.



Signs of Life (SOL) is a commonly used tool to help minimize the almost inevitable shortcomings of coverage in SPR.

4. Updating and the “Signs of Life” methodology

- Signs of Life (SOL) is a set of "activity rules" that can be used to check across various administrative data sources available to NSO in order to determine whether or not a person is alive and resident at a particular period in time
- The list of SOL markers can never be absolute; however, the more markers that can be used the more accurate will be the judgement.
- SOL markers may be determined by using data from CRVS, tax registers, social security, unemployment database, education database, etc.

If a person has been at least once active (has a record) in a register during a specific year, then the value of SOL for him/her is 1; otherwise, 0.

4. Updating and the “Signs of Life” methodology

- To use these markers, two key pre-conditions need to be satisfied, at least at the national level:
 - a) a set of administrative registers should be held in which,
 - all persons in all registers are identified by their unique ID codes; and
 - all living quarters (dwellings, family houses, etc.) in all registers are identified by their unique address IDs;
 - b) all registers should cover the whole population and be regularly updated at least annually

5. Editing and imputation

- The SPR should ideally be clean and consistent, with no conflicts between individual data items, no missing or improbable values.
 - Administrative data sources should be carefully investigated to detect and resolve “systematic errors” (either coverage or content errors).
 - Availability of metadata, particularly on editing procedures within the administrative authority, is very helpful, particularly in understanding the existence of any systematic limitation.

Resolving approaches

- Editing: to correct values that are clearly erroneous or implausible
- Imputation: to insert plausible values where data items are missing
- Simultaneous editing and consistency checks of several data sources
- Close contact with admin authorities to make improvements in the way that data are collected and recorded.

Research and testing

- Never integrate new administrative data sources into census data production without “feasibility research” by NSO. Feasibility research involves:
 - developing a detailed understanding of the administrative authority’s data collection processes, the population covered, and the variables included within the administrative source as well as how accessible these data are.
 - Test data should be obtained and examined in detail to identify quality issues and define cleaning and harmonisation, along with validation checks.
 - Combine with other available registers to verify data quality and select the most reliable variables and values in accordance with developed methodological rules.
 - Produce estimates using test data and evaluate the estimates by comparing them with previous census results or other sources.
 - Using feasibility research for developing methods for the derivation of census characteristics

Research and testing

- NSOs should address the following main challenges when deriving census characteristics.
 - ascertain the international census standard (definition, classification, etc.) applicable to the target census characteristic;
 - compare and contrast census definitions and classifications with the definitions and classifications used in the administrative source;
 - test the accuracy of the administrative data recorded against alternative sources and work collaboratively with data suppliers to eliminate/mitigate any shortcomings;
 - determine which, and how many, sources are required to derive and assure the quality of each target census characteristic;
 - establish optimal rules for deriving each census characteristic and develop the necessary data processing software, optimised for the quality of outputs sought; and
 - where characteristics are not covered by any administrative sources, take steps to ensure creation of the necessary register or register part (e.g. suggest amendments in register procedures, the legal environment, etc.).

Takeaways

- Constructing the Statistical Population Register (SPR) is a complicated process that involves several steps that must be followed with extreme caution.



UNECE

Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses

UNECE

Guidelines on the use of registers and administrative data for population and housing censuses

Statistical Commission
Fifty-third session
1–4 March 2022
Agenda item 3(e) of the provisional agenda
Items for discussion and decision: population and housing censuses

Background document
Available in English only

Handbook on Registers-Based Population and Housing Censuses

Prepared by the United Nations Statistics Division
1st draft – subject to final substantive and copy editing

United Nations Economic Commission for Europe

Register-based statistics in the Nordic countries

Review of best practices with focus on population and social statistics

Asia-Pacific Guidelines to Data Integration for Official Statistics

Data Integration Community of Practice (DI-CoP)

WILEY SERIES IN SURVEY METHODOLOGY

Register-based Statistics

Statistical Methods for Administrative Data

Anders Wallgren
Britt Wallgren

Second Edition

A decorative graphic featuring a blue square border. Inside the border, a yellow smiley face with a wide, open-mouthed grin is positioned above the words "Thank you". The text is rendered in a dark blue, rounded, cursive-style font. The entire graphic has a white drop shadow, giving it a 3D effect.

Thank you

