



# IAEG-SDG WORKSTREAM ON DATA DISAGGREGATION



# Outline

- Goal, Reference and Definitions
- What have we achieved so far
- Scope of Data Disaggregation
- Data availability, sources and consequences
- Results and implications



# Goal

- **Goal: “Leave no one behind”**
- **Call for disaggregation of all indicators by:**
  - income
  - sex
  - age
  - race
  - ethnicity
  - migration status
  - disability
  - geographical location
  - other characteristics relevant in national context
  - **and additional dimensions mentioned in targets and indicators**



# Terms of reference

- **Decision 47/101 (n) on the 47th STATCOM-session requested the IAEG to “develop the necessary statistical standards and tools, including by establishing **a working group to work on data disaggregation as a subgroup** of the Inter-Agency and Expert Group.”**
- **Members:**  
Bahrain, Brazil, Cameroon, Canada, Egypt, France, the Netherlands, Senegal, Co-Chairs: Ghana (from Nov. 2017), Germany and Jamaica (until Summer 2017).



# Definition of disaggregation

- *“Disaggregation is the breakdown of observations within a common branch of a hierarchy to a more detailed level to that at which detailed observations are taken. With standard hierarchical classifications, [...] categories can be split (disaggregated) when finer details are required and made possible by the codes given to the primary observations.”* UN Glossary of Classification Terms
- **Disaggregation dimension:**
  - characteristics by which data is to be disaggregated by (for e.g. sex, age, disability etc.)
- **Disaggregation category:**
  - different characteristics under a certain disaggregation dimension (such as female/ male for the dimension sex)



# What have we achieved so far

- **Stocktaking**
  - Analysis of data availability of disaggregation in the global database
  - Analysis of **disaggregation dimensions stated in the target and indicator** name (“minimum disaggregation” -> to start with) as well as additional dimensions
  - **Overview** of existing and possible standards of disaggregation
  - Identifying key challenges on which to focus



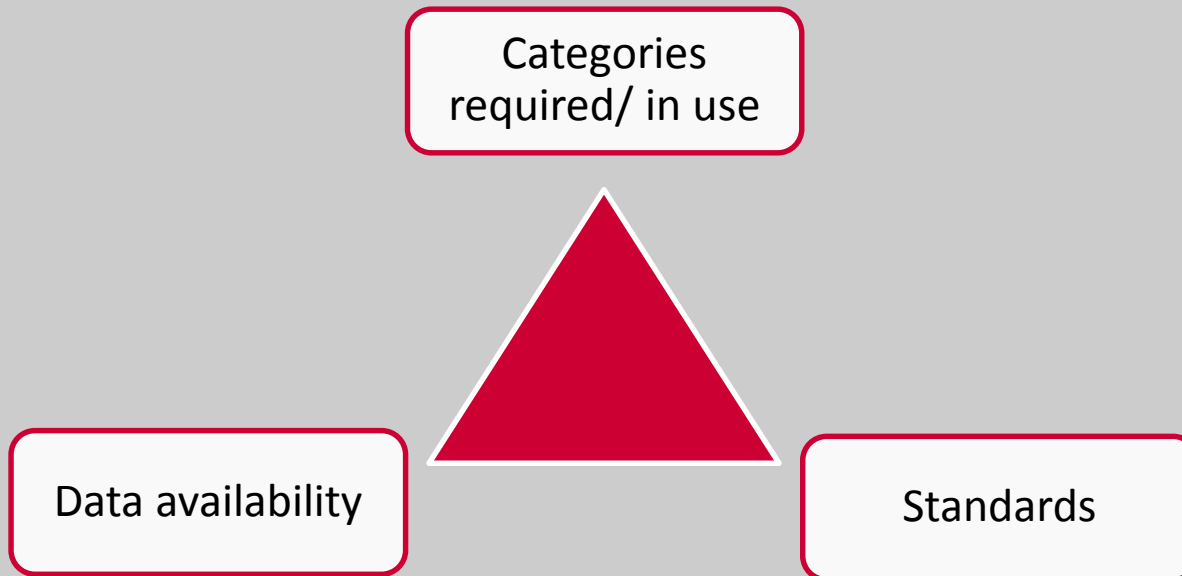
# Data availability - in global database

- **Data available in the Global Database for 30 dimensions**
- **But: data not consistently available for all dimensions**
  - Often individual cases
  - Use of different categories
- **Example: age and age groups**
  - Variety of different categories used in the global database
  - E.g. for age
    - >15, 15-24, >25; <15, 15 -49, >15; 15 years old and older; under 1 year old; under 5 years old; 25-64; 15-64; 16-49; 15-19, 20-24, 25-29, 30-...4-year-steps until 49; 20-24; 5-17; 1-14; 18 – 29; >5; <5; children, retirement age



# Data availability - Areas of conflict

- **Conflict between data availability, commonly used or required categories and harmonized, internationally comparable standards**







# Scope of Data Disaggregation

- Data is to be disaggregated **by 84 dimensions according to the “minimum” disaggregation** (dimensions stated in the targets and the corresponding indicators)
  - The “minimum” disaggregation would require approximately **4000 to 5000 time series**
  - Not included in this number are that beyond the “minimum” disaggregation all indicators should be disaggregated by income, sex, age, race, ethnicity, migration status, disability, geographical location and other characteristics relevant in national context
- This would lead to **approximately 25.000 time series**



# Scope of Data Disaggregation

- **Problem of cross-disaggregated data**  
e.g. 25 year old man, with income x, nationality y, migration background, geographical location z, etc. ...)

→ Obtaining cross-disaggregated data for all indicators and all dimensions and all categories would lead to approximately **700.000** time series or even more



# What does it tell us?

- **Exemplary interpretation (fictive figures!):**

**Indicator 11.2.1 Proportion of population that has convenient access to public transport**

**At the international level, 73.5% of Caucasian women, aged 51 to 55, catholic, with an income of 40 000 to 49 999 US \$, migration background, no disability, living in urban areas have convenient access to public transport.**

**What is the explanatory power of this detailed disaggregation as one number at the international level?**



# Data sources and consequences

- **Household surveys**
  - **Extension of the household surveys**
    - Increase of questions and
    - Increase of sample sizes
    - Large burden on respondents and Statistical Offices
- **Registers**
  - Required political permission to use and combine register data
  - Implementation of unique identifiers required
  - Data security concerns -> “big brother”
- **Certain types of disaggregation **prohibited** in some countries**
  - **Requirement of **political decisions** in the countries**
- **(Econometric) Models**
  - New type of data, results depend on the assumptions made for the model
  - Often no sound and established methods available



# Recommendations and findings

- Disaggregation has to happen (in general) on national level
- “Minimum disaggregation” is starting point
- Decide what disaggregation dimensions and categories are really needed for the time series at global and national level
  - If no global aggregate is required, categories do not necessarily be harmonized across countries
- Treat every indicator separately
  - Do not change categories if an int. established breakdown already exists to harmonise across SDG indicators
- If no breakdown exists, see if a break down for a related indicator exists and use that
- **Keep in mind the economic use of limited resources**



# THANK YOU!

