

# Virtual UNECE Workshop on Statistical Data Editing

31 August – 4 September 2020

## **Future work**

*Discussant:* Daniel Kilchmann (SFSO),  
InKyung Choi (UNECE)

# Procedure

1. Participants were asked their input through a web-survey
2. Short presentation of ideas with respect to E&I (following slides)
3. Participants should comment on ideas during the WS:
  - Proposal for selecting E&I domains for international collaboration
  - Topics of the next workshop (provisional commitment for potential contribution)
4. Discussion of ideas and fixing future work including provisional contribution commitment: Friday, 4 September

# Participants' inputs: domain to following up

## Potential collaboration

1. ML and AI for E&I, pre-processing guidelines + metadata for new data sources, guidelines where edge computing is used.  
→ HLG-MOS project on ML?
2. R for statistical data editing and imputation → uRos?
3. Development of e-learning courses e.g. for GSDEMs, validation, influence observation + more complex methods.
4. Sharing tools via github + comparison with standard tools e.g. CANCEIS, BANFF, etc.
5. Design of E&I, discuss explicitly situations where "no intervention" or a "simple methodology" would be fit for purpose.

# Participants' inputs: domain to following up

## General E&I domains

1. Validation and imputation for household surveys.
2. Editing of geographic statistical data.
3. Imputation Data, outlier detection.
4. E&I for balance statistics.
5. Treatment of unit nonresponse in SDE; bias selection and SDE.

# Participants' inputs: domain to following up

Not directly related to E&I

1. Time series, Classification, Visualization and analysis (Tableau, SQL, SPSS), Interpretation of data (Excel, Tableau, R, SQL, and SPSS), etc...

# Participants' inputs: topics SDE2022 (?)

1. ML, AI for E&I.
2. Imputation (for 'Covid-19 mass nonresponse') + variance.
3. Modernisation of data editing and statistical production.
4. Use of Deep Learning on Big Data + synchronize it with E&I.
5. Use of admin data for E&I.
6. Selection bias and SDE.
7. Infrastructure needed for open source SDE packages.
8. Anthropometric E&I.
9. Missing prices and outlying price data.

# Participants' inputs: topics SDE2022 (?)

Not directly related to E&I

1. Synthetic data sets, Forecasting (time series models), data analysis, ML methods for communication, analysis (time series, AI, PCA, ...), etc...