UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Data Editing
(Geneva, Switzerland, 15-17 April 2020)

INFORMATION NOTICE No.1[*]

The workshop will be held at the Palais des Nations, Geneva, Switzerland from 15 to 17 in April 2020.

I.   Purpose of the workshop

1.   The workshop aims to progress work on statistical editing in the wider context of the High-Level Group on Modernisation of Official Statistics (HLG-MOS) work programme. In particular, the workshop will:

- Identify new methods that can improve the quality and efficiency of Editing and Imputation (E&I);
- Investigate the statistical quality risks arising from using new methods and data sources and the ways to address them;
- Develop approaches to standardizing and implementing statistical editing functionalities;
- Facilitate the sharing of experiences, ideas and tools for modernizing statistical data E&I process.

2.   The target audience of the workshop includes methodologists, statisticians and researchers in official statistics working on editing and imputation of statistical data derived from surveys, censuses, administrative and external sources in various subject-matter areas.

II.   Agenda of the meeting

3.   The programme of the workshop will consist of the following substantive topics.

(i) Methods: for machine learning and time series data, and new/emerging methods

This topic covers new and emerging methods for improving or optimizing the methods of data editing and imputation. It aims to bring together contributions that promote innovative ideas and applications related to various aspects of editing and imputation.

Of particular interest are contributions on:

- The use of machine learning techniques for data editing and imputation (for instance, algorithms for content extraction, classification, clustering, random forests, neural networks, association rules);
- Methods for editing and imputation of time series data;
- Methods for editing and imputation of aggregated data;
- Methods for the detection and treatment of errors in data sources other than sample surveys, such as administrative data or big data, including integrated data sets;

---

[*] Information about the venue and local arrangements in Geneva will be provided in Information Notice No. 2.

- Progress for methods in a more classical setting.

Contributions could report on developments in theory and techniques, empirical comparisons and evaluations of methods, exploration of potential bias of given methods, and ways of combining different methods in an editing process. They could also highlight the expected impact that new methods might have on a statistical agency, including how they contribute to standardizing concepts, terminology, methods, data structures and improving the quality of its data products.

(ii) Data: 2021 Census, administrative data, geospatial data, big data and other alternative data

This topic aims to highlight specific features of various data other than that from sample surveys, and how to address their challenges to editing and imputation.

Examples include how to cope with mode or source effects in 2021 Census, which can lead to bias in imputation if not dealt with correctly; how to improve record linkage procedures and impute for multi-entity microdata of person, household and dwellings; how best to triangulate and micro-integrate multiple administrative sources, possibly with additional survey or big data, and apply editing and imputation efficiently to these datasets; how to identify or establish target statistical units when making use of non-survey big data that are organised around different objects; how to apply and assess machine learning algorithms for extracting measures of interest from text, image or even speech, and develop specialised editing methods for these sources; how to match geospatial data employing different location-coordinate systems and effectively use the underlying spatial structure of the data during imputation.

Contributions could focus on theoretical frameworks for integrating various unconventional data sources into official statistics, and the development of effective micro- and macro-editing strategies. They could report on best practice and practical examples of how the challenges and opportunities afforded by new and disparate sources have been addressed in terms of changes to processes and metadata management, including pre-processing and documentation. They could also evaluate and compare different methods and approaches for dealing with the specific features, which are characteristic of non-survey data. Lastly but not least, they may highlight preparations for the upcoming 2021 Census round, pertaining to mixed mode data collection, integration of new sources with collected census data, imputation of one-number census file, and so on.

(iii) Software: open source software and software demonstrations

Usage of suitable software tools is inevitable when designing and implementing modern data editing strategies. Development of such tools can be a complex, long-term and very costly job. For smaller organisations, such development often exceeds their capability and budget constraints. The principle of sharing seems to be a natural solution in such cases. Although the statistical community has shared knowledge and established concepts, methods and best practices, there is still a lack of successful collaboration in the field of exchange of operational ready-to-use software tools. Stronger international collaboration in this respect is considered a key opportunity for improving the efficiency of developing high quality statistical software.

Contributors are encouraged to give practical demonstrations of the usage of the mentioned tools. If a live demonstration of the software is foreseen, this should be indicated in the abstract.

Papers focusing on the following issues are especially welcome under this topic:
- Open source tools for data editing;
- Shareable data editing tools developed for national purposes;
- Development or (re-)use of portable (plug-and-play) software tools for editing and imputation functions;
- Software implementations of the Common Statistical Production Architecture (CSPA);
- Implementation experiences of generic software modules.

(iv) Quality: assessing data quality and indicators

For a long time, quality in statistical offices was equivalent to accuracy. A lot of the quality work had to do with estimating error rates and controlling error levels. The concept of quality used now has been extended to several dimensions, such as relevance, timeliness, accessibility, interpretability or coherence. Moreover, the world is increasingly flooded with statistical data, produced by different producers in the most diverse way. In this situation, official data must offer a quality label that users can recognize. The aim of Editing and Imputation (E&I) processes is to improve data quality. Hence it is important to evaluate whether an E&I process is achieving this objective, and whether the resulting data are of sufficient quality. This allows statistical agencies to make informed decisions about allocating resources to improve existing E&I or other business processes.

The main focus of this topic is on indicators and techniques that allow the evaluation of the quality of E&I and related processes, and of the data resulting from these processes. Papers in this topic may address, for instance, the following issues: how to estimate the uncertainty due to E&I, empirical comparisons of different methodologies using the same input data, quality indicators for different groups of users, or how to use the information from E&I to improve the survey process as a whole. Any other issue related with assessing or monitoring the statistical data quality with respect to E&I is also welcomed.

(v) Processes: editing in a process-oriented set up, standardisation and meta-data driven processes

The use of standards for concepts, methods and processes for data editing is a prerequisite to develop and apply reusable modern methods and tools and to support sharing knowledge and systems. Under this topic, contributions can report on the standardisation of editing methods and practices. This can apply, for instance, to the standardisation of editing (and imputation) procedures across the production processes of different statistics within NSI's or to the adoption of international standards for sharing methods and tools between NSI's.

Metadata provide not only additional information on collected data but can serve as driving engine in a thoroughly planned and well-functioning data processing system. Therefore, metadata elements are needed to structure the process steps and the process flows. Contributions may discuss the importance of metadata driven data editing procedures and processes as well as the challenges of building and maintaining such procedures.

III.   Call for contribution

4.   Statistical offices, international organizations, universities and research institutes that wish to contribute in form of a presentation or a poster on the topics mentioned above should send an abstract to the UNECE secretariat (choii@un.org) **by 30 November 2019**. Please send a short abstract with participant(s) name(s) and contact information with indication: i) whether the contribution is for a paper or for a poster; ii) which topic the paper or poster is about; and iii) whether the contribution comprises a software demonstration.

IV.   Participation and accreditation

5.   Representatives of all Member States of the United Nations and of interested intergovernmental organizations are welcome to participate in the workshop. Researchers at universities or research institutes working on statistical data editing applicable in the context of official statistics are also welcome to participate.

6.   All participants attending the workshop are requested to have a valid passport and, if required, a visa. Applications for visas should be made as soon as possible to the embassy of Switzerland in the

country in which the participant resides, with a reference to the UNECE Workshop on Statistical Data Editing. A letter to facilitate obtaining a visa can be requested during registration (see paragraph 7).

7.   Participants should register **by 14 February 2020** by completing on-line registration form available at: https://reg.unog.ch/event/31130/ If you need a letter to facilitate obtaining a visa, make sure to click "visa support requested" during registration.

8.   To foster collaboration among countries, a small group discussion or a sprint can be organized during the workshop. Participants who work on topics that they would like to discuss with other participants in different organizations can submit their ideas by completing the on-line form at https://www.surveymonkey.com/r/RCB6CD8 **by 14 February 2020**. Topics that are recommended for small group activity are more specific and practical (e.g. distance function for kNN-based methods for donor imputation) that can lead to concrete discussion results. The organization of the small group discussion or sprint is subject to the minimum number of interested participants.

V.   Documentation, methods of work and official languages

9.   The working language of the workshop is English. All contribution should, therefore, be submitted in English only. The following deadlines and requirements apply:

- A short abstract of the paper or of the poster should be submitted to the UNECE secretariat (choii@un.org) **by 30 November 2019**;
- A paper (in MS Word or PDF formats, max 10 pages) should be sent to the UNECE secretariat **by 14 February 2020**. Templates for the paper can be found on the UNECE event page: https://www.unece.org/index.php?id=52553;
- An electronic copy of the poster should be sent to the UNECE secretariat **by 27 March 2020.** The size can be A2, multiple A3 or multiple A4. **Please note that authors should print their own posters and bring them to the workshop**.

10.   About 15 minutes will be allocated for presentations for selected papers as time permitting. The following deadlines and requirements apply:

- Presentations in PowerPoint or PDF should be sent to the UNECE secretariat **by 27 March 2020**. These will be loaded on the conference room computer prior to the meeting;
- Participants intending to demonstrate software should contact the UNECE secretariat **by 27 March 2020** to ensure the necessary technical facility is available.

11.   All papers, posters and presentations will be made available on the UNECE event page: https://www.unece.org/index.php?id=52553

12.   Participants are encouraged to download the documents from the website and bring their own copies to the workshop. Documents posted on the website will **not** be distributed in the conference room.

VI.   Travel and accommodation

13.   Participants and/or their offices are requested to make their own travel arrangements and hotel reservations. The UNECE secretariat cannot provide any financial support for travel or accommodation.

VII.   Further information

14.   For further information, please contact:

Ms. InKyung Choi	Mr. Taeke Gjaltema
UNECE Statistical Division	UNECE Statistical Division
Palais des Nations	Palais des Nations
1211 Geneva 10 Switzerland	1211 Geneva 10 Switzerland
Tel: +41 22 917 3968	Tel: +41 22 917 1272
E-mail: choii@un.org	E-mail: taeke.gjaltema@un.org