UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(The Hague, Netherlands, 24-26 April 2017)

Information Notice No. 3

The Work Session on Statistical Data Editing will be held at the office of Statistics Netherlands in The Hague, from 24 to 26 April, starting at 09.30 on Monday 24 April, and ending by 13.00 on Wednesday 26 April.

I. ENTRANCE TO STATISTICS NETHERLANDS BUILDING

1. The main entrance to the building of Statistics Netherlands is located at Henri Faasdreef 312. On the first day of the Work Session, all participants should register at the reception desk at the main entrance. Registration will be open from 8.30 onwards. Please bring a copy of the registration form received from the UNECE website. For directions to the office of Statistics Netherlands by public transport, see Information Notice No. 2.

II. MEALS AT THE VENUE OF THE WORK SESSION

2. Lunch and coffee breaks will be offered by Statistics Netherlands near the meeting room. Delegates who have special dietary requirements (vegetarian, gluten-free etc.) should inform organisers about it in the survey below.

III. SOCIAL EVENT

3. All participants are invited to the social dinner on 25 April. The dinner will be held at a restaurant that is located in the city centre of The Hague:

   Restaurant “Luden”
   Plein 6-7
   2511 CR The Hague
   http://www.ludendenhaag.nl

   The dinner will start at 19.00 and is offered by Statistics Netherlands to the participants at no cost. To participate in the dinner you must register through the survey below.
   A Map of the city centre which indicates the location of the restaurant could be found in Annex 1.

IV. SURVEY

4. In the survey below please confirm your participation in the dinner, your dietary requirements and select a group for the mini sprint that is of most interest to you: https://www.surveymonkey.com/r/895H2Z5

   Please respond to the survey by 24 March 2017 at the latest.
V. MINI SPRINT FOR TOPIC 3

5. As mentioned in Information Notice 1, on Wednesday 26 April participants of the Work Session will participate in the interactive mini sprint for Topic (iii): ‘How to foster the implementation in NSOs of good practices from other organisations, areas for international collaboration, and priorities for joint work’ prepared by the Organising Committee of the Work Session.

6. The mini sprint was encouraged by the High Level Group of Modernisation of Official Statistics (HLG-MOS) who asked this Work Session for a tangible contribution to the implementation of good practices, and to identify and elaborate priorities for joint work on international collaboration.

7. Participants will be assigned to small groups in advance, based on their interest in a particular theme.

8. Each group will develop tangible proposals in an area of statistical data editing for their selected theme. Given the time limitations, groups will not be expected to produce a deployable product with professional polish, but rather the well-conceived outline of a solution.

9. The Organising Committee in collaboration with the UNECE will produce a report summarising the findings of the discussions.

10. You can find below a list of themes for discussion and groups:

11. **Theme 1:** What are the priorities for future work on data editing by the international statistical community? Which domains, methods or data sources will be your highest priorities in the coming years? How could international collaboration help?

12. **Group A:**

   - Which method presented at the Work Session or already well-established could/should be implemented as a CSPA service, how to do it, which are the benefits compared to sharing tools (SAS Macros, R packages, etc.) and whether an international collaboration might be beneficial. Further information on CSPA can be found in the Annex 2.

**E&I CSPA services available**
- Linear Rule Checking  
  ([https://webgate.ec.europa.eu/fpfis/wikis/display/CSPACatalog/Linear+Rule+Checking](https://webgate.ec.europa.eu/fpfis/wikis/display/CSPACatalog/Linear+Rule+Checking))
- Data Set Re-code  
- Error Correction  
- Linear Error Localization  
  ([https://webgate.ec.europa.eu/fpfis/wikis/display/CSPACatalog/Linear+Error+Localization](https://webgate.ec.europa.eu/fpfis/wikis/display/CSPACatalog/Linear+Error+Localization))

   - Would it be beneficial for your institution to have the following methods or tools implemented in CSPA?
     i. Selective editing.
     ii. Further Fellegi-Holt based imputation tools like BANFF.
     iii. Nearest Neighbour Imputation method like CANCEIS.
     iv. Outlier detection of unit evolutions (exceptional evolution of the values of a unit compared to the past).
     v. Distributional outlier detection (extreme values, univariate and multivariate).
     vi. Process monitoring and quality indicators.
     vii. E&I methods based on machine learning.
     viii. Others?
• Could your institution contribute to an international collaboration for implementing the above discussed methods and tools in CSPA (please get the commitment of your management beforehand)?

• List the above findings with the estimated amount of personal resources which are planned to work on them and prioritize them.

• Several institutions promote the exchange of SAS Macros and/or R packages. The discussion should result in a common understanding and outline how these exchanges could interact with CSPA services.

  - How to share SAS Macros or R packages (cran, sourceforge, github, etc.)?
  - How to ensure the support of SAS Macros?
  - Sustainability vs. quick version changes of R?
  - Is a catalogue of E&I Macros and/or R packages needed and if yes, how to maintain it?

• Please get the commitment of your management beforehand if your institution is planning to promote the exchange of SAS Macros or R packages with the estimated amount of personal resources which are planned to work on that.

• List the pros and cons with respect to CSPA, SAS Macros and R packages and report on experiences how a SAS Macro or R function was integrated as a CSPA service.

13. **Group B:**

• Detecting missing domains and suggestions for future topics, including how this could be done and how they are related to the contributions of the work session.
  
  i. What are the priorities of your institution with regard to developments and enhancements in the domain of E&I?
  
  ii. Are the above mentioned developments already part of a cooperation with other countries or institutions?
  
  iii. Are all domains of data preparation listed in GSDEMs perfectly covered in your institution and/or your survey? Is there a need for further development?
  
  iv. Is your institution, with the support of your management, interested in cooperation with other countries (please get the commitment of your management beforehand)?

• List the above findings with the estimated amount of personal resources which are planned to work on them and prioritize them.

• Show the relationship to the contributions of the work session.

14. **Theme 2:** There are various standards and tools for data editing created by this group over the years. Examples include the Generic Statistical Data Editing Models, and the Data Editing Glossary (for both, please see http://www1.unece.org/stat/platform/display/kbase).

15. **Group C:**

Revision and updating of standards, discussion of gaps and potential further development of standards:

• What standards or tools are missing?
• What should be expanded or improved?
• How may the adoption to these standards be enhanced?
Could your institution contribute to an international collaboration with respect to the above discussed issues (please get the commitment of your management beforehand)?

VI. CONTACTS

16. For further information, please contact the following organizers:

**Statistics Netherlands:**

Mr. Jeroen Pannekoek  
Senior Methodologist  
Statistics Netherlands, P.O. Box 24500  
NL-2490 HA THE HAGUE  
Netherlands  
Phone: +31-70 337-4919  
Email: J.Pannekoek@cbs.nl

Mr. Sander Scholtus  
Methodologist  
Statistics Netherlands, P.O. Box 24500  
NL-2490 HA THE HAGUE  
Netherlands  
Phone: +31-70 337-4926  
Email: s.scholtus@cbs.nl

**UNECE:**

Ms. Tetyana Kolomiyets  
Statistical Information and Methodology Unit  
Statistical Division  
United Nations Economic Commission for Europe  
Palais des Nations  
1211-GENEVA 10  
Switzerland  
Tel.: +41 22 917-4150  
email: tetyana.kolomiyets@unece.org
Annex 1 – Map of the city centre which indicates the location of the restaurant.
Annex 2 - CSPA

- **CSPA Global Artefacts Catalogue** ([http://www1.unece.org/stat/platform/display/CSPA/CSPA+Global+Artefacts+Catalogue](http://www1.unece.org/stat/platform/display/CSPA/CSPA+Global+Artefacts+Catalogue)): A primary aim of CSPA is to support efficient sharing and reuse of process patterns, information and services at an organization and international level. One key requisite in achieving this goal is an ability to reliably and efficiently discover what is available for reuse to support a particular business need. This includes an ability to efficiently assess whether a potentially reusable artefact is, in fact, "fit for purpose" in practice when it comes to supporting that particular business need.

  The Global Artefact Catalogue fulfils this requisite - enabling the CSPA community to access and share CSPA services. It is a composite conceptual catalogue representing 5 layers of information:

  - Knowledge base ([http://www1.unece.org/stat/platform/display/VSH/Virtual+Standards+Helpdesk](http://www1.unece.org/stat/platform/display/VSH/Virtual+Standards+Helpdesk))
  - Investment Catalogue ([http://www1.unece.org/stat/platform/display/CSPA/Investment+Catalogue](http://www1.unece.org/stat/platform/display/CSPA/Investment+Catalogue))
  - Capability Catalogue ([http://www1.unece.org/stat/platform/display/BCC](http://www1.unece.org/stat/platform/display/BCC))
  - Technical/Supporting Services layer (example: [https://github.com/edwindj/cspa_rest/tree/ea9d59b5afbb4764cca6b34f5d3120f3cf7c16e6](https://github.com/edwindj/cspa_rest/tree/ea9d59b5afbb4764cca6b34f5d3120f3cf7c16e6))