Information Notice No. 1

I.   PURPOSE AND TARGET AUDIENCE OF THE MEETING

1.   Machine learning (ML) draws increasing attention among statistical organisations. ML can help carry out tasks that used to be perceived as something only humans can do such as reading and understanding texts. This allows statistical organisations to perform manual tasks in a more efficient manner. ML is also an indispensable tool for processing and analysing the vast amount of information in big data that can be used to produce statistics on a more disaggregated level as well as new types of statistical products.

2.   While the specific business environments may vary depending on the country, statistical organisations face similar types of challenges and can therefore benefit from sharing knowledge and experiences, and collaborating on developing common solutions within the broad official statistical community. To facilitate the investigation of the use of ML for official statistics, and to provide a platform where users and practitioners in statistical organisations can share experiences and lessons learned, several international initiatives have been launched in the past few years, notably the [HLG-MOS Machine Learning Project (2019-2020) and the ONS-UNECE Machine Learning Group (2021-2022)](.).

3.   The Machine Learning for Official Statistics Workshop 2023 aims to bring together experts in national and international statistical organisations who are using ML to improve the quality of official statistics, share new developments in the field and discuss issues and challenges.

4.   The target audience are those who develop and use ML (e.g., data scientists, statistician, methodologists, subject matter experts, product/project managers) as well as those who manage and design foundational infrastructure needed for the integration of ML in the organisations (e.g., IT specialists, business architects, data officers, quality managers, metadata specialists).

5.   The meeting will be hosted by UNECE in Geneva, Switzerland from 5 to 7 June 2023.

II.   AGENDA OF THE MEETING

6.   The meeting will have the following main sessions:

Session 1: Machine Learning Applications

ML holds great potential to contribute to the work of statistical organisations in various ways. It can automate the tasks that were primarily done by humans, assist humans to do the work more efficiently, and allow the organisations to make use of new data sources (e.g., websites, satellites, social media) which can ultimately increase the relevance and timeliness of statistics produced. Additionally, ML can also be applied beyond traditional production areas, for example, to provide data services to other government agencies or to improve customer service efficiency. This session aims to explore the application of ML with various data sources (e.g., survey, register, big data) and their added value for statistical organisations. Potential topics include, but are not limited to:

- Text classification;
- Edit, imputation and validation;

- Record linkage;
- Imagery classification.

Please note that the main interest of this session is on the value addition of ML application in the statistical organisation rather than a deep dive into the technical and methodological aspects.

Session 2: Quality Aspects of Machine Learning in Official Statistics

Quality is the most important principle of official statistics. Without high-quality standards, trust in official statistics would be jeopardized and it would be difficult to maintain the privilege that official statistics enjoy in many countries. However, the concept of quality is by no means one-dimensional, as work in recent years has shown.

On the one hand, machine learning offers new possibilities for analysis and insight against the background of the novel data sources, improving the quality of statistics (e.g., relevance, timeliness) and improving the efficiency of the production processes. On the other hand, it is not yet generally clear whether and how the typically prediction-oriented and non-model-based approach of machine learning methods can be reconciled with the quality requirements and framework conditions of official statistics.

This session will cover quality aspects regarding the use of machine learning in official statistics. Possible topics include, but are not limited to:
- Quality aspects of machine learning algorithms (e.g., random forests, neural networks) and pipelines;
- Quality aspects in the production (including architecture and processes) when there is a machine learning component;
- Quality aspects of training data;
- Metrics and indicators for measuring machine learning quality;
- Quality frameworks and guidelines;
- Quality aspects of implementations of machine learning in software;
- Best practices in statistical offices.

Session 3: Toward System-wide Transformation of Statistical Production

The core of machine learning is the models, they can empower system-wide transformations of statistics production, but they may fall short of expectations. One critical reason is that the machine learning study focuses largely on modelling aspects (i.e., which algorithms are used, how the models are trained). However, when a machine learning model is to be taken into production, it requires much more than just the model and the predictions it produces.

This session will discuss the experiences of adopting machine learning in the statistical organisation such as strategic management, technical/operational arrangement, organisational capability building. We also welcome lessons learned from 'fantastic failures'": circumstances and factors that turned out particularly challenging in implementing ML in official statistics and knowledge gained from these experiences (i.e., what should have been done differently?).

Potential topics covered under this session include, but are not limited to:
- Strategies for building organisational capabilities for ML;
- Strategic management and empowering of ML-driven innovations;
- Key factors for integrating ML in statistical organisations (e.g., communication, addressing resistance from stakeholders, external collaboration);
- MLOps for statistical organisations - can we adopt best practices from the big IT companies?;
- Emerging techniques important for the use of ML (e.g., privacy-preserving techniques, cloud);
- Responsible AI, ML ethics and explainability;
- Centralised platforms for MLOps (ML as service).

7.   Based on the abstracts received (see Section IV for more details about the submission of abstracts), the programme of the meeting will be determined and communicated closer to the date of the meeting.

8.   The meeting may include interactive elements such as small group discussions (e.g., ML failures, typical reasons behind, and potential solutions) and live polling.

III.   PARTICIPATION AND ACCREDITATION

9.   This meeting is planned as in-person. However, the presentations given during the meeting will be recorded and made available online (subject to the approval of the presenters).

10.   Representatives of Member States of the United Nations and of international organisations, are welcome to participate in the workshop. All participants must be accredited by the competent authorities of their country or international organisation.

11.   All meeting participants should register by **20 May 2023** by completing the Indico on-line registration form at: https://indico.un.org/e/ML2023. Please note that if you do not already have a user account on the Indico system, you will first need to create an account, and then activate it by clicking on the account activation link that is sent to you by email. Once you have a user account, and have logged in, you can submit your registration for this meeting (please note that the registration requires passport information and photo). You will receive a notification confirming this, and then a further notification email when your registration is approved.

12.   Your timely registration will ensure that you receive emails with further information that we may send out. Please note that the registration may close earlier if the number of registered participants surpass the room capacity.

13.   Participants attending the meeting are requested to have a valid passport and, if required, a visa. Applications for visas should be made as soon as possible to the Swiss Embassy in the country in which the participant resides, with a reference to the UNECE Machine Learning for Official Statistics Workshop 2023. A letter to facilitate obtaining a visa can be requested from the UNECE Secretariat by 20 April 2023.

IV.   CALL FOR CONTRIBUTIONS

14.   National and international statistical organisations can make a contribution in any format (e.g., giving presentation, conducting training or sprint, running small discussion) as long as they are in line with the sessions as described in Section II.

15.   Those who wish to give a presentation at the meeting are invited to submit the abstract via the following online form: https://forms.office.com/e/fqnPi1TTkh

16.   The Organising Committee will notify in due time whether the submission is accepted or not, and might request changes. Information about the selection of contributions for the meeting, guidelines on formatting, and means of submission will be sent to authors by email.

17.   Following deadlines and requirements apply:

   (i) A short abstract of the proposed contribution (200-400 words) should be submitted online, as described above in paragraph 15, **by 15 February 2023** at the latest;

   (ii) Any written paper (3-10 pages depending on the topic) must be provided **by 30 April 2023** at the latest. A link will be sent to the authors where documents can be uploaded. Authors are encouraged to share links to any code repositories containing code, if applicable (e.g., GitHub);

(iii) Any presentation slides should be provided **by 30 May 2023** at the latest. Any equipment required for practical demonstrations must be provided by the participants. A link will be sent to the authors where presentations can be uploaded.

V. TRAVEL AND ACCOMMODATION

18. Participants are requested to make their own travel arrangements and hotel reservations. The UNECE secretariat does not provide any financial support for travel or accommodation. More details about logistics (e.g., transport from the airport) will be provided in Information Notice No.2.

19. Although we anticipate an improved COVID situation by the time of this meeting, participants must ensure they have insurance to cover all associated eventualities, including cancellation of flight and hotels if the public health situation seriously deteriorates. Participants attend at their own risk, and should stay up-to-date with any requirements that may be needed for those who travel to Switzerland, including for example any documentation that might be required to demonstrate vaccination against COVID.

VI. DOCUMENTATION, METHODS OF WORK AND OFFICIAL LANGUAGE

20. To minimise paper use, participants are encouraged to download meeting documents and, where feasible, to use electronic devices to read them. Documents will not be distributed in the conference room.

21. Please note that the official language of the meeting will be English, and therefore all contributions should be submitted in English only. No translation or interpretation will be provided during the meeting.

22. Prior to the meeting, abstracts, papers and presentations will be made available online via the following wiki page: https://statswiki.unece.org/x/xQDpFg.

VII. FURTHER INFORMATION

23. A second information notice with practical information will be shared with registered participants, and loaded onto the meeting website in due time before the meeting. For further information, please contact:

Secretariat of the United Nations Economic Commission for Europe
Ms. InKyung Choi (choii@un.org)

DEADLINES	
15 February 2023	**Submission of abstracts**
30 April 2023	**Submission of papers**
20 May 2023	**Registration for the meeting**
30 May 2023	**Submission of presentation materials**