Concept Note

Online Training workshop

on Data science for Official Statistics and the SDGs

6–8 November 2023

Introduction:

Big data has emerged as a valuable resource for producing official statistics and monitoring progress towards the Sustainable Development Goals (SDGs). As the world becomes increasingly interconnected, vast amounts of data are generated from various sources, providing new opportunities for policymakers and statisticians. In addition, Decision 53/124 of the United Nations Statistical Commission supported the proposed direction of work of the UN Committee of Experts on Big Data and Data Science for Official Statistics (UNCEBD) to mainstream big data and data science in the daily work of national statistical offices (NSOs), to create a network of data science leaders of NSOs, and to strengthen the collaboration between UNCEBD and the geospatial community.

The exponential growth in technology is disrupting NSO work like other areas. There is pressing need to modernize NSO work using data science in the Arab region. Therefore, the below program compromises the capacity building to provide participants with a solid foundation in the fundamental concepts of big data for official statistics and its relevance to the SDGs.

Objectives:

The workshop will provide an opportunity to enhance the capacity of NSOs in the Arab region and to provide participants with an introduction in the fundamental concepts of data science and big data for official statistics and its relevance to the implementation and monitoring of the Sustainable Development Goals (SDGs).

As such, the workshop will aim specifically to:

- ✓ Introduction in Data science and Big Data for official statistics and the SDGs
- ✓ Big data and UN initiatives for SDGs (Definition, UN initiatives, Forums, data storage and management)
- ✓ BIG Data and data science in Official Statistics: Issues, Methodologies, ethics, and tools (Data Ecosystem: concepts and sources, Bias type and solutions, existing methods, and tools)
- ✓ Introduction to Big Data and its relevance to Official Statistics
- ✓ Big data analytics: Descriptive analytics (Descriptive statistics, correlations, regression)
- ✓ Big data analytics: Predictive analytics (Classification, clustering)
- ✓ Big data analytics: Predictive analytics (clustering, association rules)
- ✓ Imputation methods to fill the gap in SDG.
- ✓ Practical exercises and use cases in python

Expected Outcomes:

Enhanced skills of participants, specifically:

- ✓ concepts of data science and its applications in official statistics
- ✓ familiarity with data science and big data and UN initiatives for the SDGs, issues, methodologies, ethics
- ✓ hands on tools for big data analytics
- ✓ potential of non-traditional data sources for indicators
- ✓ theoretical and practical cases.

Target Audience:

The program is designed for statisticians, IT, data analysts, and professionals working in national statistical offices.

Program Delivery:

The program will be delivered through a combination of lectures, interactive discussions, case studies, and hands-on exercises. Practical training based on experiences, and insights on utilizing data science and big data for official statistics and the SDGs will allow participants to apply the concepts, tools and techniques learned during the course.

Duration:

This program will be conducted online over a total period of 3 days during the period 6-8 November 2023 (3 hours per day).