UNITED NATIONS
ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC

STATISTICAL INSTITUTE FOR ASIA AND THE PACIFIC (SIAP)

Big Data: Innovative Methods and Applications for Achieving SDGs
21-25 October 2024
Chiba, Japan

CONCEPT NOTE

I. Background and rationale

The achievement of the Sustainable Development Goals (SDGs) requires the availability of high-quality, timely and reliable data to produce the relevant SDG indicators and other statistics, disaggregated as relevant. To meet this need, official statistics must modernize and incorporate new data sources, including Big Data. At its 53rd session in 2022, the Statistical Commission underscored the importance of mainstreaming *the use of big data and data science into the work programmes of national statistical offices and* the necessity *to include training in big data and data science into the training curricula of national statistical offices*. This course is a first response to this injunction by sharing experience, and providing the pedagogical activities required for understanding the process leading to the production and dissemination of official statistics and SDGs with new data sources.

The course *Big Data: Innovative Methods and Applications for Achieving* organized by the United Nations Statistical Institute for Asia and Pacific (UNSIAP), with the contribution of the UN Women Centre of Excellence for Gender Equality (Seoul, Korea) will be conducted from 21 to 25 October 2024, in Chiba, Japan.

II. Objectives

The course aims to provide a platform for demonstrating methods, best practices, and tools for addressing challenges in incorporating Big Data into the production and dissemination of official statistics and SDGs. It will cover essential methods like machine learning, classifiers, and AI, focusing on practical applications through case studies, handouts, and hands-on sessions. Additionally, the course will highlight how Big Data can fill gaps in Sustainable Development Goals data and be used to analyse international and national trends.

As a result of this course, participants are expected to be able to:

- **Identify** the concepts and technologies needed for analysing Big Data.
- **Recognize** the different types of Big Data and their use for Official Statistics and SDG Indicators.
- **Analyze** through case studies, how Big Data can address data gaps and improve the accuracy of SDG indicators.
- **Demonstrate** the ability to use Big Data for identifying and interpreting international and national trends in various statistical domains.
- **Recognize** the limitations and challenges when using Big Data

III. Target participants

The course is designed for mid-level statisticians and technical staff from national statistical offices whose main responsibilities include the production and dissemination of official statistics with new data sources to include private sector data, administrate data and data collected from non-traditional sources such as social media data, geospatial data and online pricing data. Approximately 14 participants are expected to attend. The participants are expected to have a basic proficiency in the analysis and modelling of statistical data and in the use of a modern statistical software or language (R, Python).

IV. Course design and content

The course is comprised of 7 modules, consisting of presentations on the main topic and related issues/problems; illustrative good practices from resource persons presenting practical approaches and solutions to the issues/problems; sharing of relevant country experiences and concern. Plenary and group discussions will be organized throughout the course. The course is not based on any specific software nor programming language. However, some experience and knowledge in programming would help participants understand some examples and case studies.

To facilitate exchange of experiences, **each participant will need to submit a country presentation** of no more than 5 slides on or **before 4 October 2024** to escap-staff@un.org and christophe.bontemps@un.org**.** The presentation should be 10 minutes long. It should briefly introduce the current use of Big Data, present main challenges encountered, and outline specific plans to incorporate new data sources into statistical production processes. Please include relevant detail regarding, (a) the current or planned use of Big Data (including administrative data sources, private sector and crowd-sourced data); (b) the scope, sources, frequency, and quality checks protocols used for these data; (c) SDG indicators that will be compiled using Big Data sources or a combination of sources to include Big Data.

The course will cover 7 modules of different lengths[1]

Module	Coverage
Module 1: Overview of Big Data	• Understanding Big Data • Typologies of Big Data • Big data for Official Statistics and SDGs
Module 2 : Methods for Big Data	• Reproducible Analytical Pipelines • Data Visualization • Machine Learning
Module 3 : Countries examples	• Current and future uses of Big Data *(Participants' presentations of NSOs projects)*
Module 4 : Use cases of Big Data	• Web scrapping methods and applications • Web scrapping for CPI • AIS data – Applications • Scanner data – Applications
Module 5 : Use cases of Big Data-Geospatial	• Geospatial methods & Applications • Applications to Land Cover estimation • Applications to Poverty estimation
Module 6: Gender Equality and Big Data	• Big data sources for gender-related SDG indicators • Using Artificial Intelligence to Quantify Misogyny and Offensiveness • Mapping indicators of women's welfare • Using Big Data to Measure Gender and Environment Nexus
Module 7: Challenges in Big Data	• Big Data for Official Statistics: Integration & limitations • Countries examples and project

[1] The final detailed program of the training will be sent prior to the course.