UNITED NATIONS
ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC

STATISTICAL INSTITUTE FOR ASIA AND THE PACIFIC (SIAP)

Regional Training Course on Big Data for Sustainable Development

19-23 June 2023
Daejeon, Republic of Korea

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. In order 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 with new data sources.

> The *Regional Training Course on Big Data for Sustainable Development* will be held from 19 to 23 June 2023 in Daejeon, Republic of Korea, and will be conducted by the Statistical Institute for Asia and the Pacific (SIAP), in collaboration with Statistics Korea (KOSTAT).

II.   Objectives

The course aims to provide a platform for demonstrating methods, good practices and tools and for exchanging experiences in dealing with challenges and issues in the introduction of Big Data into the production and dissemination of official statistics. The course will focus on the use of publicly available data, private sector data sources, administrative data and other type of non-traditional data.

III. Learning outcomes

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 type of Big Data and their use for Official Statistics and SDG Indicators.

- Discuss the processes required to introduce Big Data into the statistical process.

- Retrieve and perform simple analysis of publicly available Big Data.

- Recognize the limitations and challenges when using Big Data

IV. Course design and content

The course is comprised of 6 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 2 June 2023** to escap-siap@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.

V.   Course Outline[1]

Module	Coverage
Module 1: Overview of Big Data	• Understanding Big Data • Uses of big data
Module 2 : Methods of Big Data	• Reproducible Analytical Pipelines, Machine Learning • Data Visualization
Module 3 : Countries examples	• Current and future uses of Big Data at participants' NSOs
Module 4 : Web scrapping methods	• Methods, process and examples • Web scrapping with APIs • Web scrapping for CPI
Module 5 : Use cases of Big Data	• AIS data – Applications • Scanner data methods-Applications • Administrative data- Applications • Mobile phone data – Applications • Geospatial methods - Applications
Module 6: Challenges in Big Data	• Big Data for Official Statistics: Integration & limitations
Module 7: Field Trip	• Field trip

VI.   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 20 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).

---

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