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

25 – 29 June 2018
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 2014 session, the United Nations Statistical Commission recognized the importance of Big Data with the establishment of a Global Working Group on Big Data for Official Statistics. At its 2018 session, the Statistical Commission further underscored the importance of Big Data with the adoption of the Bogota Declaration, which calls for the innovation of current statistical production processes and the creation of new ways of compiling SDG indicators.

Innovations are needed in the daily production of official statistics, which requires partnerships with the private sector, new skills and infrastructure, and clear links between available Big Data sources and the SDG indicators. This training aims at sharing experience and building the skills required for 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 23-27 April 2018 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 good practices and 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 administrative data (Big Data collected by other government departments), publicly available data and private sector data sources.

III. Learning outcomes

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

- Understand the concepts and technologies use for analysing Big Data.

- Discuss the tasks required to introduce administrate data into the statistical process.

- Understand methodologies for linking private sector data to existing survey and administrative data.

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

- Understand the architecture and infrastructure required to process Big Data.

IV. Course design and content

The course is comprised of six 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; hands-on exercises; sharing of relevant country experiences and concerns; and plenary discussions.

To facilitate exchange of experiences, **each participant will need to submit a country presentation of no more than 10 slides on or before 8 June 2017 to staff@unsiap.or.jp**. The presentation should be 10-15 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 use of administrative data sources, private sector and crowd sourced data; (b) the scope, sources, frequency, and quality checks protocols used for these data; (d) SDG indicators that will be compiled using Big Data sources or a combination of source to include Big Data.

Course Outline[1]

Module	Coverage
Module 1: Overview of Big Data	- Understanding Big Data - Uses of big data
Module 2 : Big Data Analysis using R	- Hands on: Introduction to data analysis - Hands on: Introduction to data visualization

---

[1] The final list of topics and detailed program of the training will defined by the trainers.

Module 3: Administrative Data	• Data acquisition, preparation and storage • Case study: Statistics production using administrative data
Module 4: Statistics Production using digital economy	• Collecting and analysing online retail price data • Linking administrative data to private data sources • Case study: Analysis of digital economy data
Module 5: Country Presentations & Field Trip	• Country Presentations • Field trip
Module 6: Statistics Production Using Private Data	• Accessing personal data • Linking private data sources. • Case study: Statistics production using personal data • Hands on: Analysing simulated telecommunication data
Module 7: Crowd sourced data and social media data	• Analysis using crowd sourced data • Hands on: Processing, analysis and visualization • Analysis using social networks data • Hands on : Social Media data analysis
Module 8: Next Steps	• Discussion: Advancing Big Data within your organization

IV.   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 and online pricing data. Approximately 20 participants are expected to attend. The participants are expected to be have a basic proficiency in the use of R. Students without R experience are encouraged to complete SIAP's online training course, "R for Official Statistics".