



# Regional Training on Big Data and Data Science for Gender Statistics in Asia and the Pacific

21-25 April 2025

Daejeon, Republic of Korea

## I. Background

Closing gender data gaps is essential for achieving the Sustainable Development Goals (SDGs). Gender statistics are critical for understanding the distinct realities of women and men, girls and boys, and for informing policies aimed at addressing inequalities, but data and information gaps are often impediments to achieving this. While efforts to enhance gender data production have advanced, significant gaps still remain.

The intersection of Big Data, data science, and gender statistics holds immense potential to address persistent gender data gaps by leveraging advancements in technology. With the rise of digital data sources such as social media, satellite imagery, and mobile devices, coupled with increased computing power and sophisticated analytical tools, Big Data enables more granular, timely, and cost-effective insights than ever before. These advancements offer transformative possibilities for measuring and addressing gender inequalities in ways that traditional methods cannot.

Despite these opportunities, limited Big Data knowledge and data science capacity among practitioners remains a significant challenge. Many lack the technical expertise, resources, or institutional frameworks needed to effectively harness Big Data for gender statistics. Without targeted efforts to build this capacity, the full potential of Big Data and data science to advance gender equality will remain unrealized.

Recognizing these challenges and opportunities, the Statistical Institute for Asia and the Pacific (SIAP), Statistics Korea (KOSTAT) and UN Women are organizing a Regional Training on Big Data and Data Science for Gender Statistics in Asia and the Pacific. The training will equip a core group of individuals with the necessary Big Data skills and knowledge, to enhance their capacity to produce, analyze, and utilize gender statistics. This investment in human capital is expected to contribute to the generation of high-quality, reliable gender data, which could lead to more informed decision-making, effective policies, and accelerated progress towards gender equality.

## II. Objectives

The overall objective of the week-long training is to help participants better understand the use and benefits of Big Data in the production of gender statistics. To that end, the training program will present

recent developments in the compilation of gender statistics as well as new methods, case studies and processes that can facilitate the integration of Big Data in this process. The training will also emphasize the limitations, constraints and privacy issues inherent to the use of Big, non-traditional or administrative data.

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 potential use for Official Statistics and gender-relevant SDG Indicators.
- **Analyse** through case studies, how Big Data can address filling gender data gaps
- **Demonstrate** the ability to use Big Data for identifying and interpreting international and national trends in gender-related development outcomes
- **Recognize** the limitations and challenges when using Big Data for filling gender data gaps

### 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 utilizing new data sources, including private sector data, administrative data and data collected from non-traditional sources such as social media, geospatial information systems and online footprint. Approximately 25 participants will be invited 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).

- Experience handling large-scale datasets, Big Data, or geospatial data in their professional work
- Prior experience using R, Python, or other statistical software
- Likelihood of working on Big Data applications to address gender data gaps

### IV. Course design and content

The course is comprised of several modules, consisting of presentations, exercises, and handouts on methods and topics relative to Big Data for compiling or augmenting gender data and statistics 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 concerns. Plenary and group discussions will be organized throughout the course. Some prior experience and knowledge in programming would also help participants understand some examples and case studies. The topics covered in this course include:

- Big data collation for gender statistics
- Gender statistics concepts and definitions
- Reproducible analytical pipelines and data science good practices
- Machine learning methods and applications
- Geospatial data and methods to augment gender-relevant statistics
- Methods and practices to collate and analyse online data
- Biases, ethics and good practices for Big Data and gender statistics
- Challenges in Big and non-traditional data integration
- Assessment tool for evaluating organizational maturity in Big Data and data science

## V. How do I prepare for this training?

Prior to the training, participants will be required to [complete a preliminary tutorial](#) on the use of R. This self-paced online tool will provide instructions on how to install R and RStudio as well as recall a few elements of the R programming language.

Additionally, to facilitate exchange of experiences, **each participant will need to submit a country presentation** of no more than 5 slides on or **before Wednesday 16 April 2025** to [escap-siap@un.org](mailto:escap-siap@un.org) and [christophe.bontemps@un.org](mailto:christophe.bontemps@un.org). Participants will be required to deliver their country presentations on day one of the training. Each participant will be allocated a maximum of 10 minutes for their oral presentation. It should briefly introduce the current use of Big Data, including for the production of gender statistics, present main challenges encountered and outline specific plans to incorporate new data sources into statistical production processes. More information will be provided prior to the course.