



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

27-30 October 2025

Bangkok

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, UN Women Knowledge and Partnership Centre, UN Women Regional Office of Asia and the Pacific, and the Statistical Institute for Asia and the Pacific (SIAP), are organizing a 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, analyse, 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 4-day 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, and constraints 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

This introductory training is designed specifically for members of the Sub-Group on Gender Statistics Training (SGGST). The SGGST was established in 2018 to support national statistics offices in the Asia-Pacific region to develop statistical capacities to effectively produce and use gender statistics for gender policy decisions, advocacy, and tracking progress of the Sustainable Development Goals. The SGGST comprises representatives from national statistics offices, national statistical training institutes, academia, the media, and civil society. Since its formation, the SGGST has been instrumental in providing guidance and leadership to several activities, most notably, the development of Gender Statistics Training Curriculum. The curriculum was launched in 2020 as a key resource to boost efforts at building and strengthening institutional capacity in the production and use of gender statistics in the Asia-Pacific region. This training aims to address members' identified needs related to emerging data sources and methodologies for gender analysis.

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?

It is important to note that since this course has a practical goal and will use handouts in R, participants

will benefit from familiarity with R programming. [A tutorial¹](#) has been put together to facilitate this process for participants to complete before the course.

To facilitate exchange of experiences, each participant will need to submit a short presentation of no more than 5 slides before **16 October 2025** to sujung.an@unwomen.org. The presentation should not be more than 10 minutes long. It should briefly introduce the current or future plans for the use of Big Data; present main challenges encountered and outline specific plans to incorporate new data sources into statistical production processes.

¹ This tutorial is designed to be readily accessible for participants with prior experience in analyzing statistical data using a modern programming language, such as Stata, Python, or R.