

UN Committee of Experts on Big Data and Data Science for official statistics



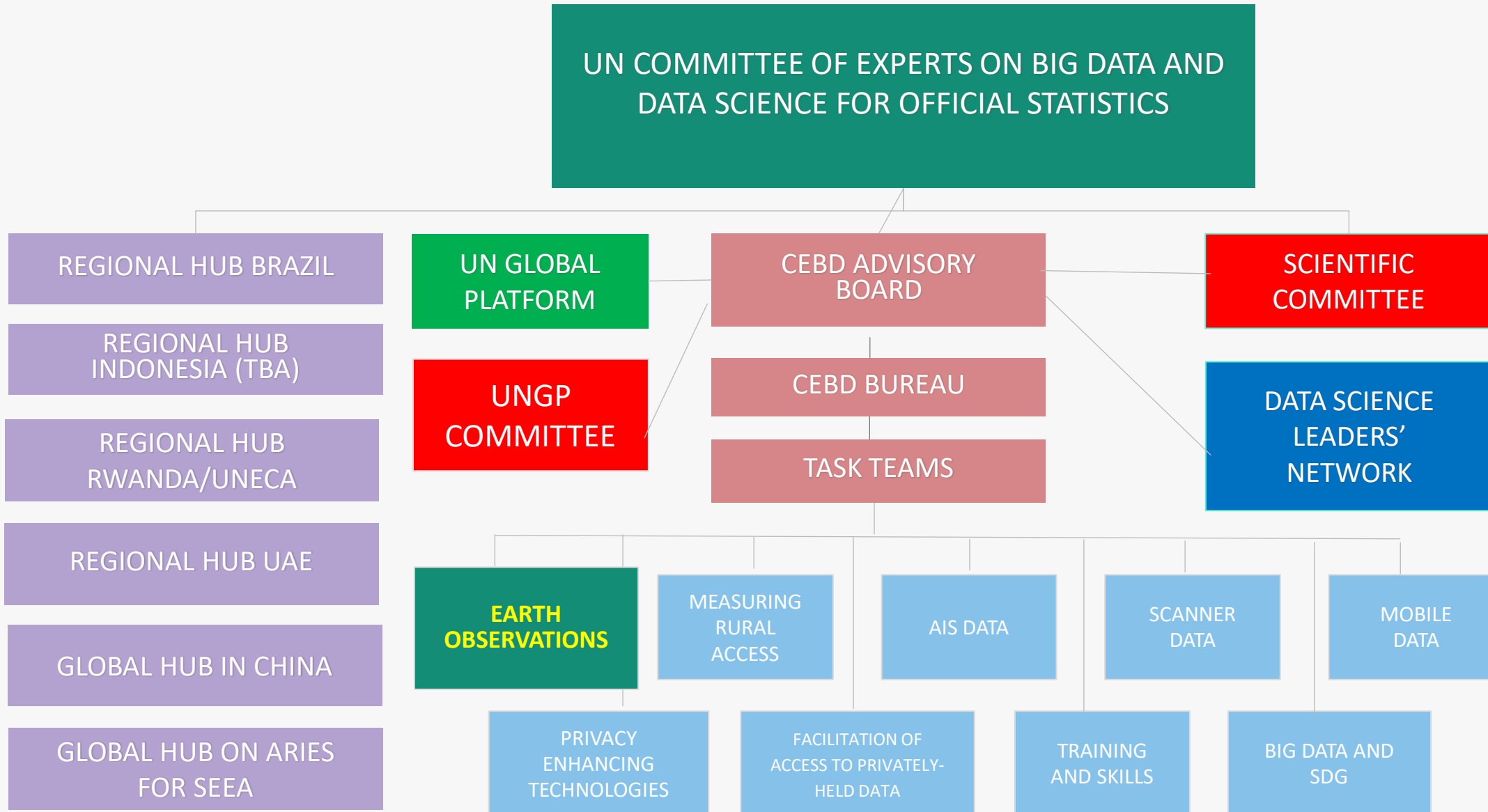
Ronald Jansen
Assistant Director
United Nations Statistics Division
Email: jansen1@un.org ; BigData@un.org
<https://unstats.un.org/bigdata/>

UNCEBD

- Created in March 2014 by the UN Statistical Commission
- Inter-governmental body with 31 countries and 16 international organizations
- Collaboration of more than 400 experts from all stakeholder communities



Organization of Committee



Task Team on Use of Mobile Phone Data for Official Statistics

About Mobile Phone Data



Introduction

The statistical community has the obligation of exploring the use of new data sources, such as Big Data, to meet the expectation of the society for enhanced products and improved and more efficient ways of working. Use of Big Data could also support the monitoring of the Sustainable Development Goals (SDGs) by improving timeliness, frequency, detail and relevance of indicators without compromising their

Methodological Guides on the use of Mobile Phone Data (2022)

- [Displacement and Disaster Statistics](#)
- [Dynamic Population Mapping](#)
- [Measuring the Information Society](#)
- [Migration Statistics](#)
- [Tourism Statistics](#)

Publications

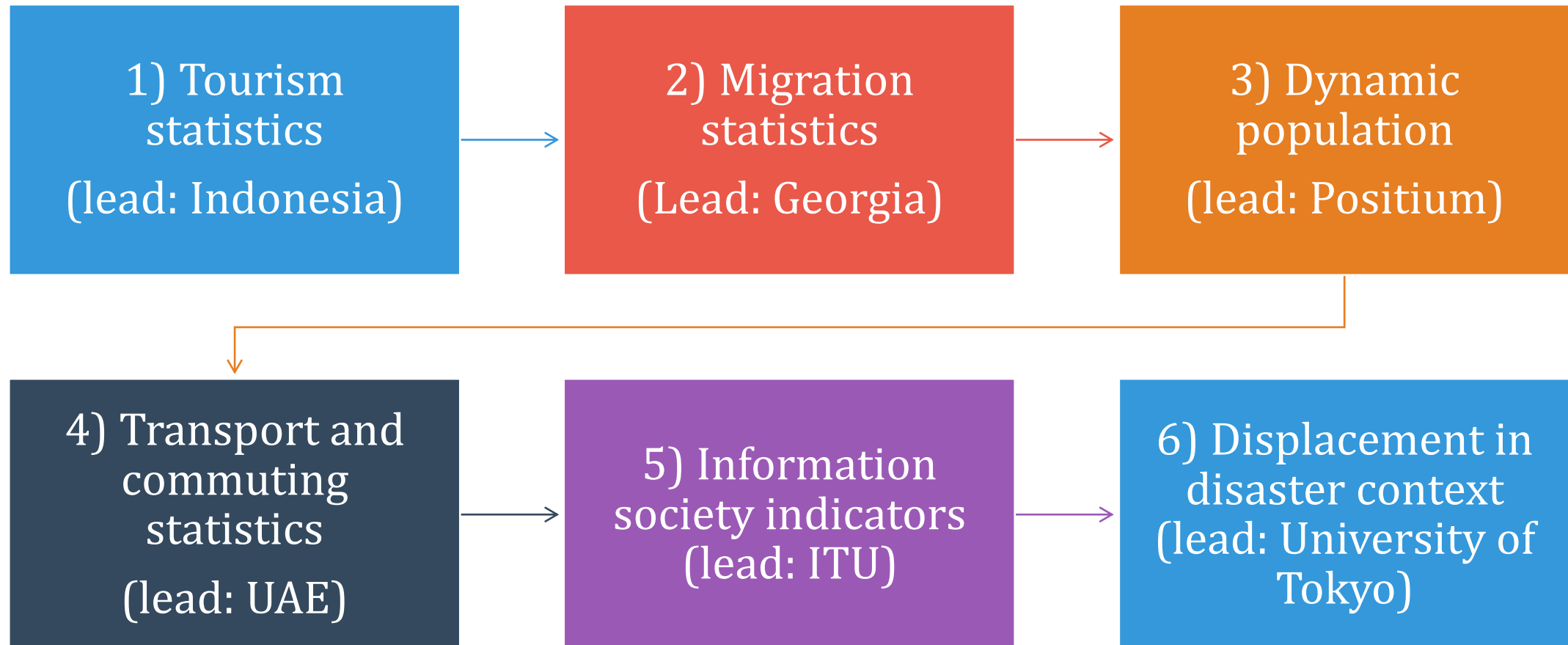
Handbook on the use of Mobile Phone data for Official Statistics (2019)

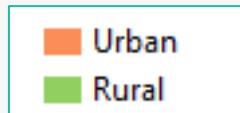
- [English](#)
- [Russian](#)

Events

[MPD Session at the 7th International Conference on Big Data](#)

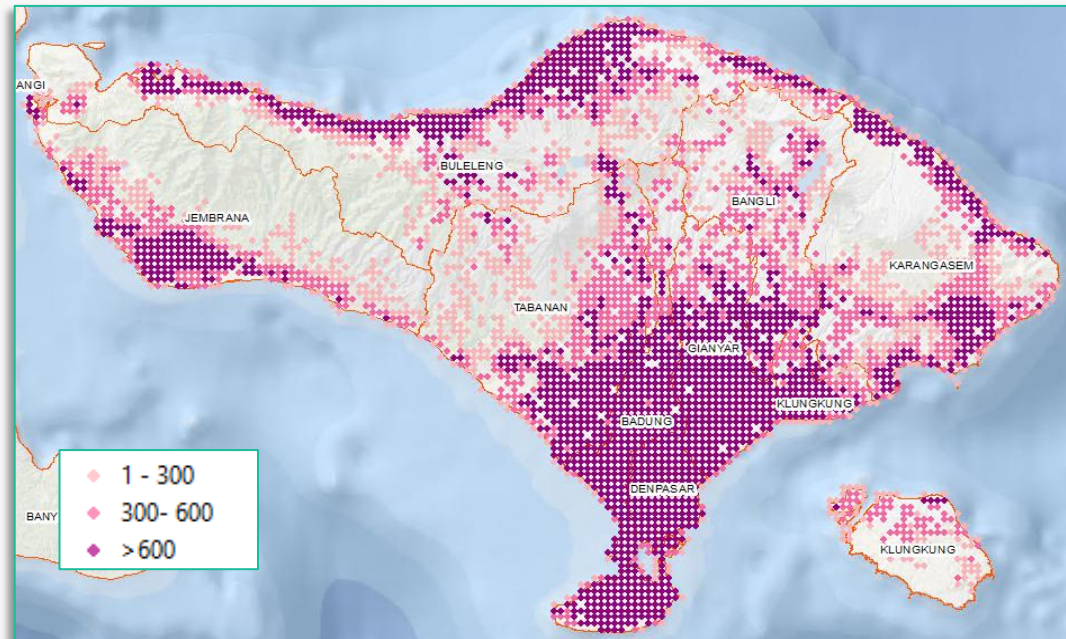
📍 Yogyakarta, Indonesia 📅 8 Nov 2022





Land Area Covered by Mobile Cellular Network (4G) in Bali Province

Population Covered by Mobile Cellular Network (4G) in Bali Province



Task Team on Use of Earth Observations for Official Statistics



Mission and Strategies

The demand for more diversified, sophisticated and rapid statistical services could be met by leveraging the emerging sources of Big Data, such as those relating to remote sensing imagery, transactional and social media data and mobile device data.

Statistical agencies around the world have a strong interest in investigating the viability of using satellite imagery data to improve official statistics on a wide range of topics spanning agriculture, the

Task Team Report

[Satellite Imagery and Geospatial Data Task Team report](#)

Use Cases

- [Digital Earth Africa](#)
- [Use of EO data in Agriculture Statistics of Statistics Canada](#)

More information

References

- [Report of the Global Working Group on Big Data for Official Statistics](#)
- [Results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data](#)
- [Big data and modernization of statistical systems](#)

Publications

2022



UNITED NATIONS

United Nations Statistical Commission

United Nations Committee of Experts on Big Data and Data Science for
Official Statistics (UN-CEBD)

Earth Observation Joint Task Team on Agricultural Production Statistics

Research Sub Task Team

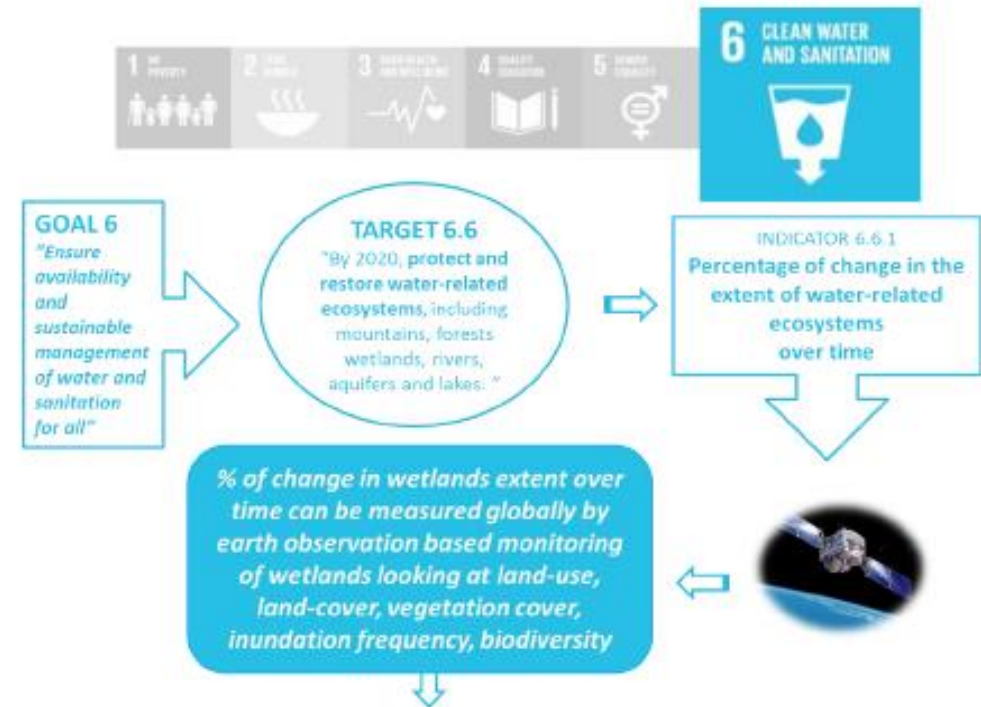
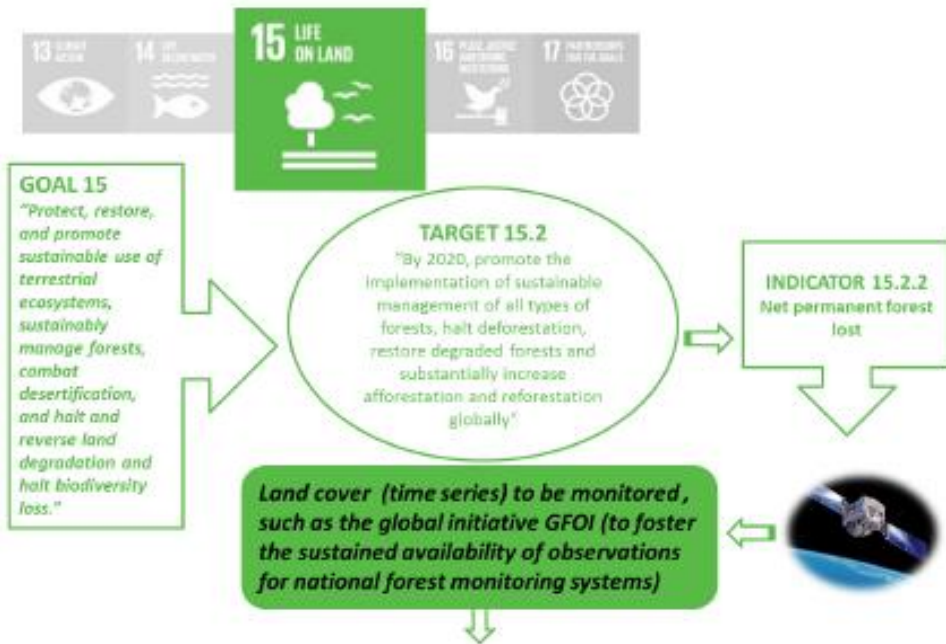
Trusted methods: Lessons Learned and Recommendations from Select Earth Observation
Applications on Agriculture

Earth Observations for Official
Statistics


Satellite Imagery and Geospatial Data
Task Team report

5th December 2017

Examples of EO for SDGs




Task Team on Use of Scanner Data and Webscraping for Official Statistics

 UN Statistics Wiki

Spaces ▾ People Calendars Analytics Cockpit

Create ...

Search

 UN-CEBD - Scanner Data Wiki

Pages

Blog

Calendars

PAGE TREE



> Handbook on utilising new data sources

> Members section

> Classification archive pages

• Allocation of wiki pages

> Demo

Pages   894 views

UN-CEBD Task Team on Scanner Data

Created by UNSD Clarence Lio, last modified by Tanya Flower on May 27, 2022

Welcome to the handbook on utilising new data sources in the production of consumer price statistics!

This handbook has been developed by the [UN Task Team on Scanner Data](#) to provide a useful source of information for anyone looking for guidance and support on using new data sources and web scraped data in the production of consumer price statistics. There is no single 'correct' approach in which to use these new data sources, and the purpose of this handbook is to provide an overview of relevant information for colleagues to refer to in their own work and decide what is best for the particular environment/project which they are working within.

We have drawn on a lot of material from existing handbooks and manuals but we have also included some topics where there is less guidance available currently (for example, data acquisition and classification). Given the pace at which this topic evolves, this handbook has been designed as a living document so that when any new publications or research are made available it can be updated to reflect the most recent analysis and findings.


When this handbook becomes available to a public audience, colleagues will be invited to use the comment functionality on each page to provide feedback, these will be reviewed by a working group who will be responsible for any updates of the content based on this feedback. The group will also be responsible for ensuring that any new publications or research are reflected in the handbook so that it remains relevant to users.

The full content of the handbook is displayed below:


Contents:

- Glossary
 - Example of chain drift
- Initial considerations
 - Introduction to the new data sources
 - Selection of categories
 - Selection of retailers for alternative data sources
 - Quality assurance
 - IT system requirements
- Data acquisition
 - Scanner data
 - Web scraping


Noticeboard

 CPI Expert Group 2023 - Geneva, Switzerland


Jun 23, 2023 • updated by Serge Goussev • [view change](#)

 Seventeenth Meeting of the Ottawa Group - Rome, 07-10 June 2022: UN Task Team

May 11, 2023 • updated by Serge Goussev • [view change](#)

 CPI Expert Group 2023 - Geneva, Switzerland

Mar 16, 2021 • created by UNSD Clarence Lio

 Seventeenth Meeting of the Ottawa Group - Rome, 07-10 June 2022: UN Task Team

Feb 22, 2021 • updated by UNSD Clarence Lio • [view change](#)

(to be updated)

Implementing more online data in the future

Web Scraped

- Online vs instore
 - Clothing
 - Electronics

Application Programming Interfaces (APIs)

- Travel



Sharing Economy

- Netflix & Spotify
- Uber & Lyft
- Airbnb



Future Trends

- Scope definition could extend beyond country
- Growth online marketplace
- New product offers: Internet Of Things



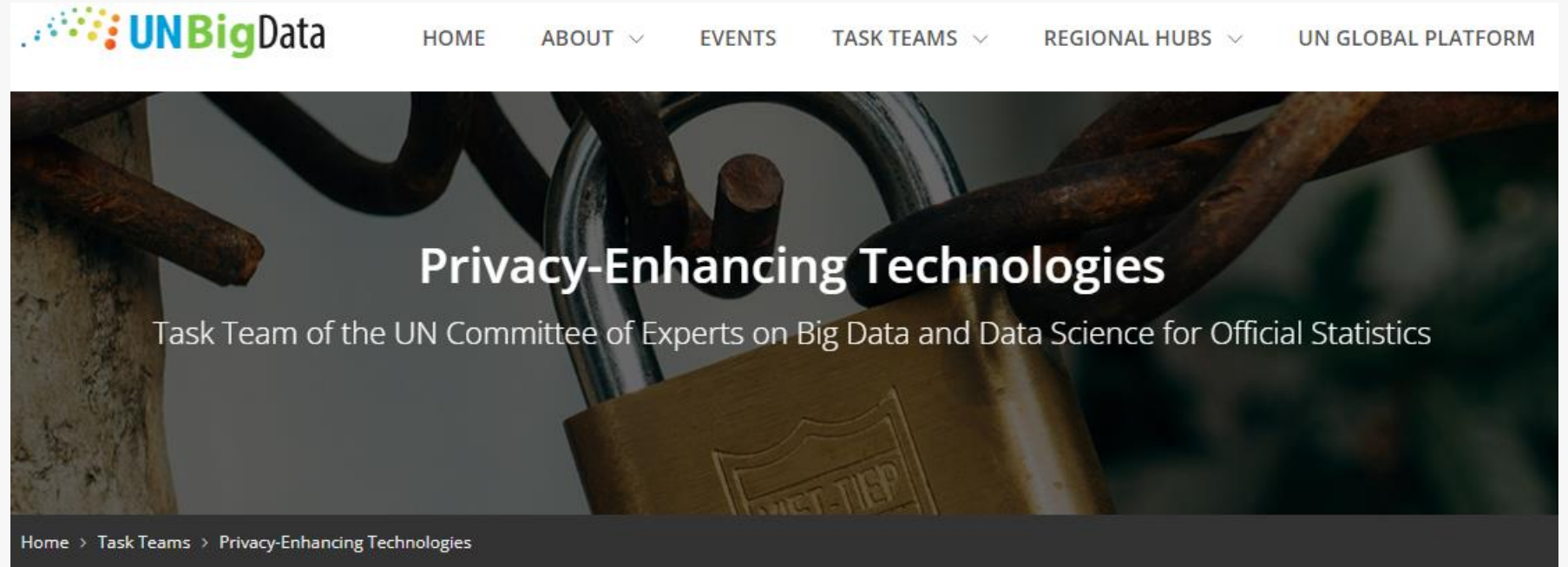
Statistics
Canada

Statistique
Canada

www.statcan.gc.ca

Canada

Task Team on Privacy Enhancing Technologies for Official Statistics



Introduction

The Privacy-Enhancing Technologies Task Team (PETTT) is advising the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) on Big Data on developing the data policy framework for governance and information management of the global platform, specifically around supporting privacy enhancing technique.

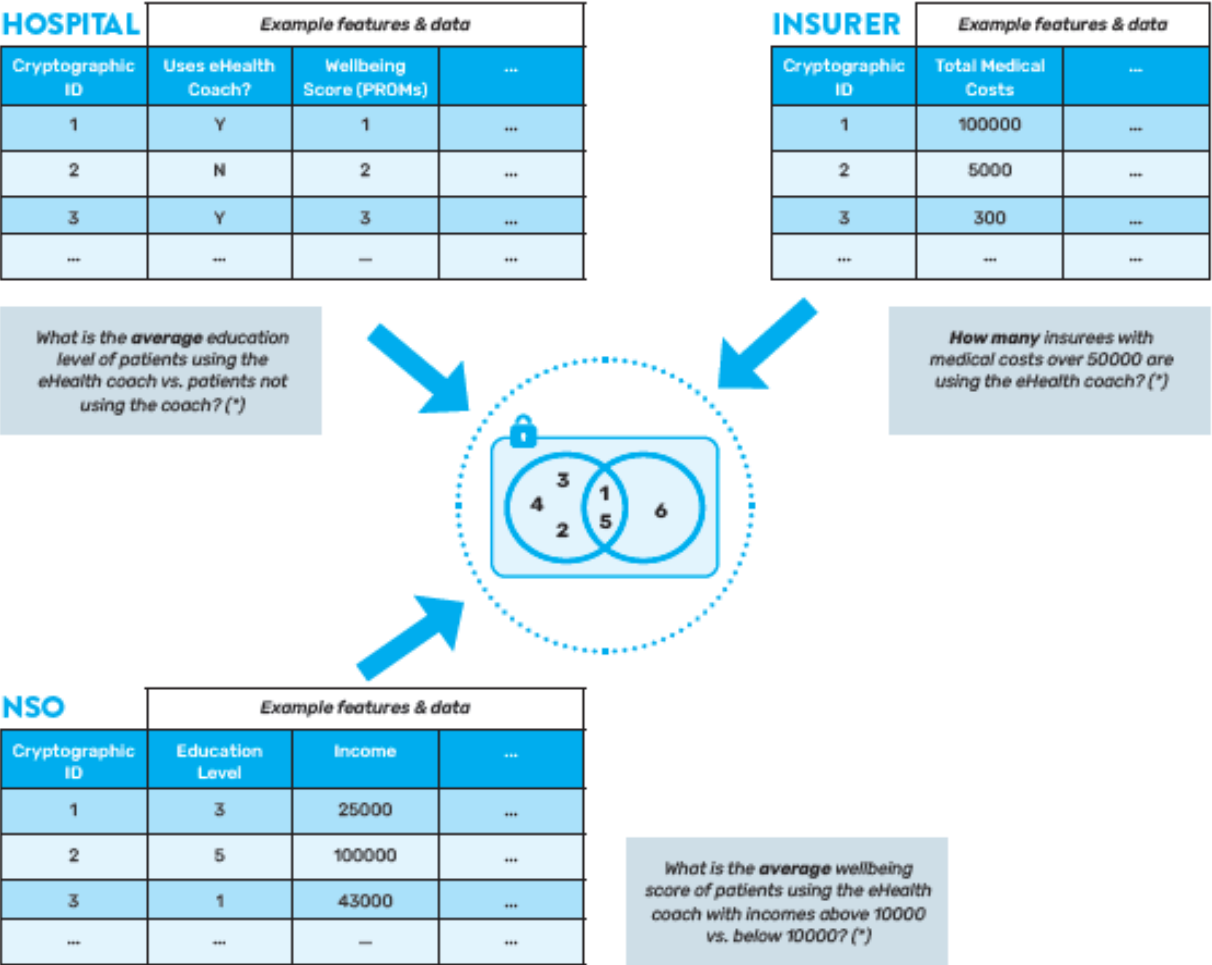
The task team has been active since April 2018 and has released the UN Privacy Preserving Techniques Handbook. This document describes motivations for privacy-preserving approaches for the statistical analysis of sensitive data; presents examples of use cases where such methods may apply; and describes relevant technical capabilities to assure privacy preservation while still allowing analysis of sensitive data. It currently covers techniques that support the protection and sharing of sensitive information: Secure

Publications



CASE STUDY DESCRIPTION

HIGH LEVEL FUNCTIONAL PERSPECTIVE



**For illustration purposes only, actual allowed queries are subject to implemented smart contract business rules*

Figure 3.13: An example of Private Set Intersection with Analytics (PSI-A)

Training in Big Data and Data Science for official statistics



Big Data Training Curriculum



E-Learning Courses



Big Data Maturity Matrix



**Training of data scientist in
academic centers**



**Big Data Competency
Framework**



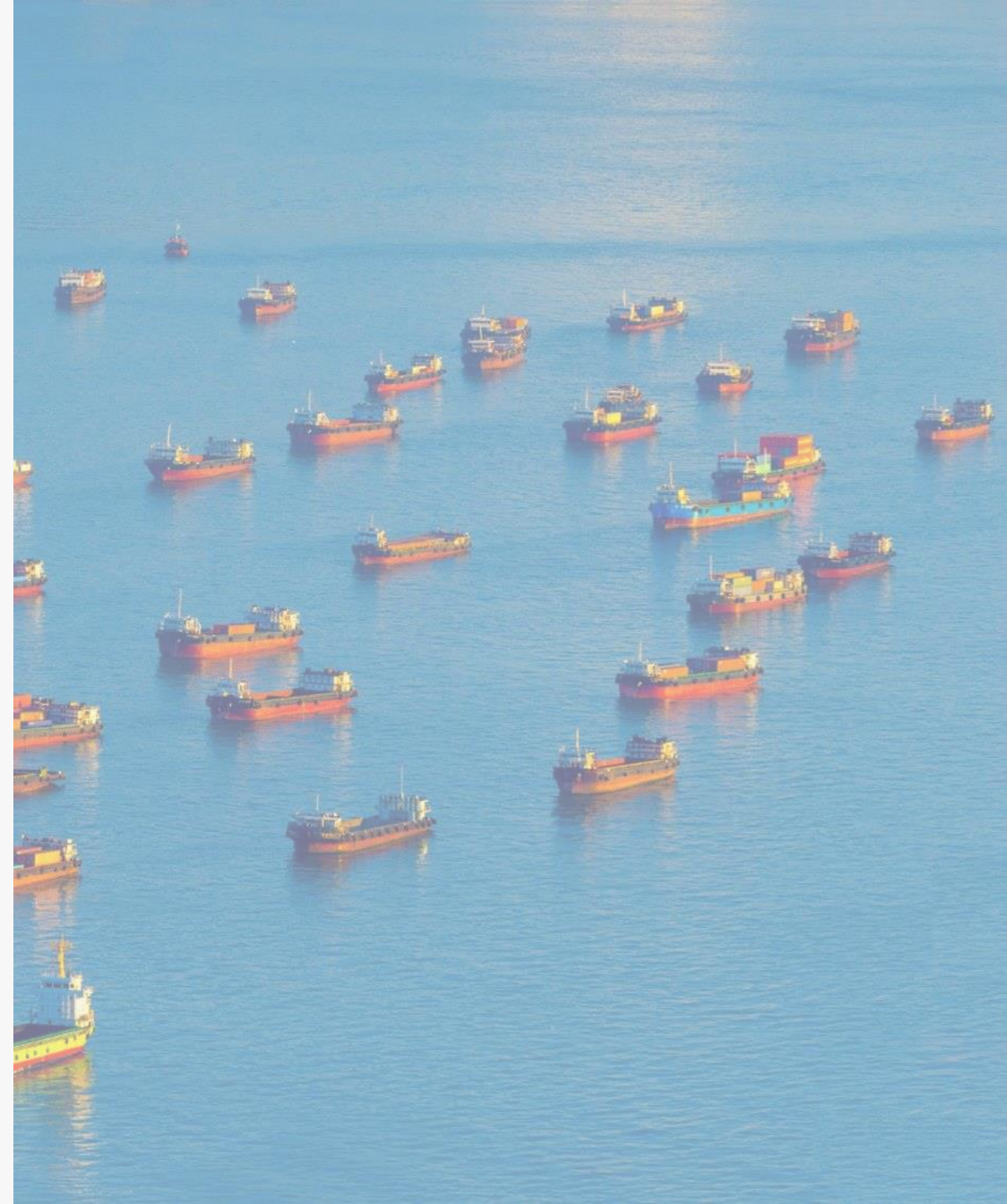
Mentorship



PROJECTS

AIS (Automated Identification System) Shipping Data

- Reduction of CO2 emission by Maritime transport
- Just-In-Time Arrivals of ships in large ports
- Maritime Transport Statistic around the Panama Canal
- Fast economic Indicators





Lorenzo De Simone, Technical Adviser Geospatial, Office of the Chief Statistician FAO
May 4, 2021

Global Center of Excellence on Remote Sensing for Agriculture Statistics in Hangzhou, China





The guide

for users of the ARIES for SEEA Explorer



I. The ARIES for SEEA application

The ARIES for SEEA Explorer is a web-based application built on the k.LAB Integrated Modelling Platform. The application has access to all information (data and models) available on the Integrated Modelling network, and provides a dedicated user interface to easily compile accounts within the UN System of Environmental-Economic Accounting (SEEA).

ARIES for SEEA can also be accessed via software download for recurrent users, for better performance in terms of speed and computation capacity. More info at <https://integratedmodelling.org/getting-started/>

1.1 Spatial and temporal context of the analysis

At the top of the menu on the left side of the screen, you can specify the geographic area and temporal and spatial scale.

1.1.1 Where

At the top of the panel, a drop-down menu provides three options to select an analysis context by zooming and panning on the map. When the “administrative regions” or “river basin” option is chosen, the currently highlighted context will be outlined in light blue.

- Map boundaries: Select an area of interest by panning and zooming in/out. The entire area visible on the screen becomes your analysis context.
- Administrative regions: This option automatically identifies the largest administrative entity (e.g., Country or Subnational Unit) in the area selected, **according to the M49 standard endorsed by the UN**. By zooming in, the user can choose a smaller administrative region. We recommend this option for novice users, as it offers a



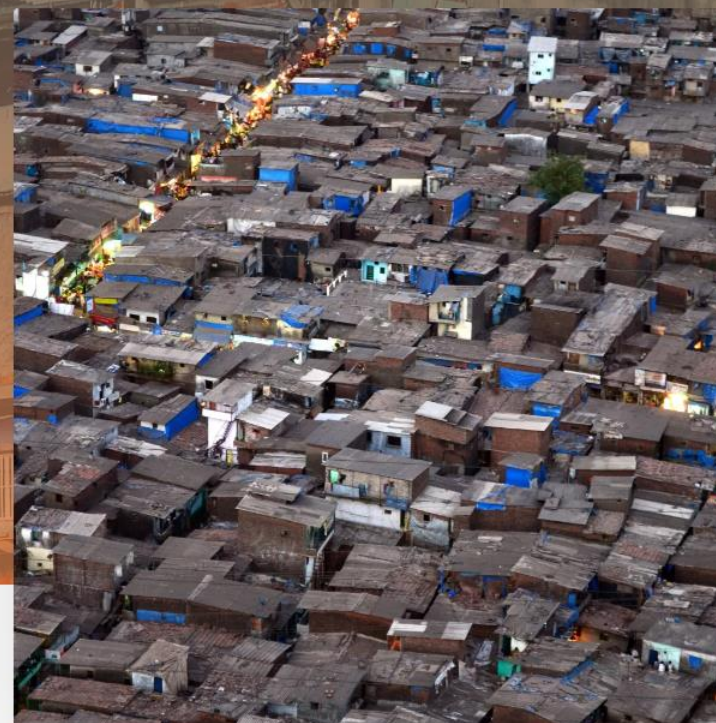
[Home](#) [About](#) [Output](#) [News](#) [Contact](#)

IDEAtlas

Developing AI-based methods to map and characterize informal settlements from Earth Observation data.

- an ESA funded project

[Learn More](#) >



IDEATLAS

Monitoring progress on the proportion of the urban population living in informal settlements or inadequate housing.



GLOBAL HEAT RESILIENCE SERVICE

Building the foundation for heat resilient communities

**Building The Foundation For
Heat Resilient Communities**

Read More



ABOUT

GEO is convening partners to develop a service that will provide every urban area in the world with intelligence on the health risks from exposure to extreme heat. These insights will help cities develop plans to adapt to heat and reduce the impact on citizens' health and local economies.

THE IMPACT OF HEAT ON HEALTH AND SOCIETY

Hot days and extreme heat events are becoming more intense and more frequent. Cities are particularly affected due to an urban heat island effect that can increase temperatures by up to 20 degrees Celsius.

A top-down view of a collaborative workspace. Several people are gathered around a large table, their hands and arms visible as they work. The table surface is covered with hand-drawn diagrams and sketches. A central yellow circle contains a lightbulb with three lightning bolts. To the left, a red circle contains an open book. To the right, a blue circle contains an eye. Other circles in various colors (green, orange, red, blue) are scattered around, some containing coffee cups or geometric shapes. Dashed lines and arrows connect these circles, suggesting a network or process flow. The words "DESIGN" and "RESEARCH" are written in the lower center. The overall atmosphere is one of creative collaboration and problem-solving.

Data Science Leaders Network & Partnerships

What is Data Science in Official Statistics?

- **Automation** of the statistical production processes (increase efficiency and improve quality)
- **Supplementary indicators** produced for emerging issues to provide additional insights
- **Changing statistical production:**
Example – webscraping of prices from the internet combined with traditional price surveys to produce regular consumer price indices

Reproducible data pipelines

Supplementary
analysis and
insights

Transformation
of Statistical
Production
Process

Examples of Data Science in Official Statistics

- *Satellite data for Agriculture Statistics*
 - Machine Learning for Crop mapping and identification
- *Scanner data and webscraping for Price statistics*
 - Optimizing the web crawlers
 - Automated Classification of product descriptions
- *Social media data for sentiment analysis*
 - Natural Language Processing
- *AIS vessel tracking data for Maritime Transport indicators*
 - Use of Polygons around harbors to estimate Port activity