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

SUSTAINABLE DEVELOPMENT GOALS

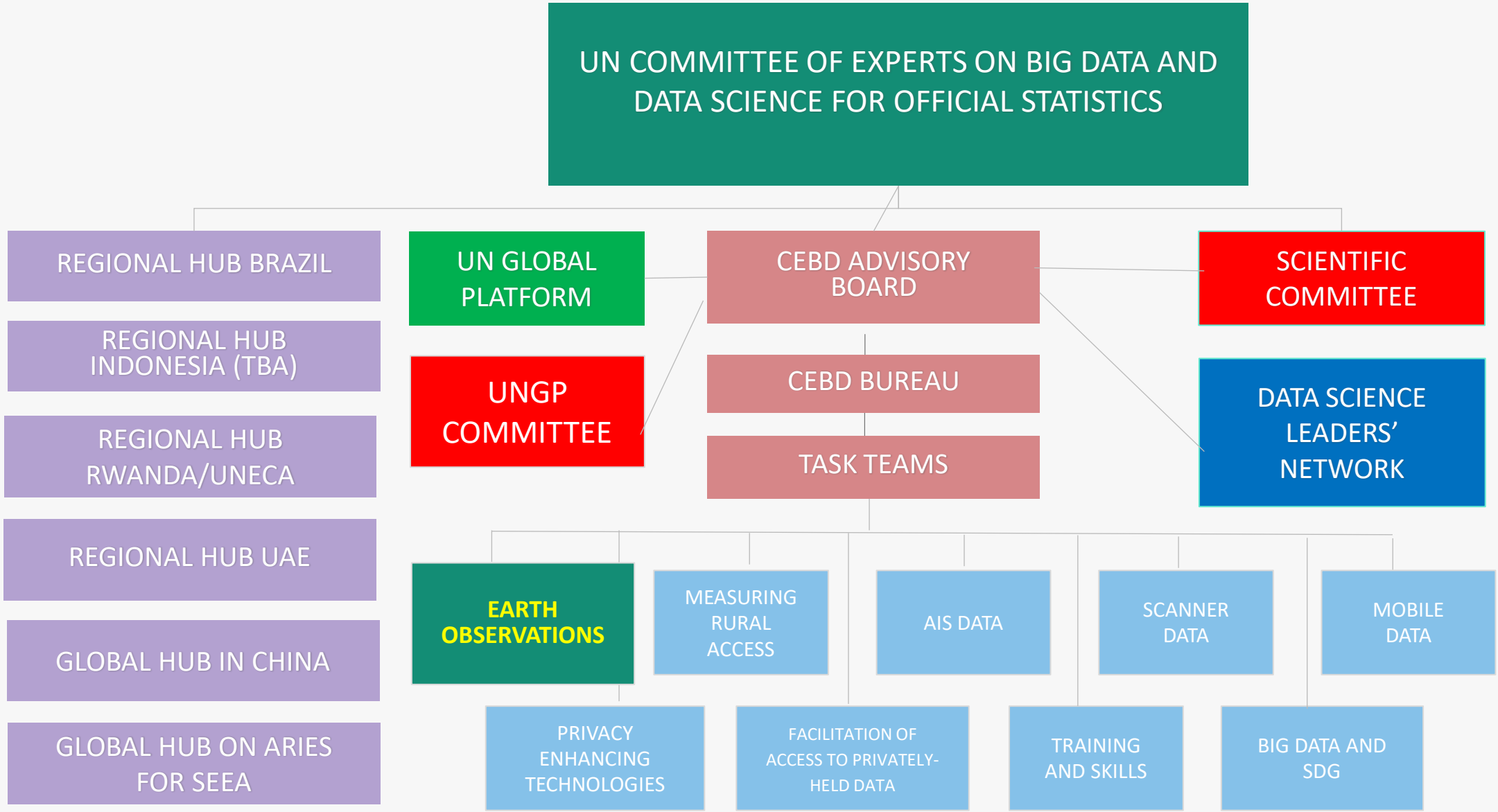
17 GOALS TO TRANSFORM OUR WORLD

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

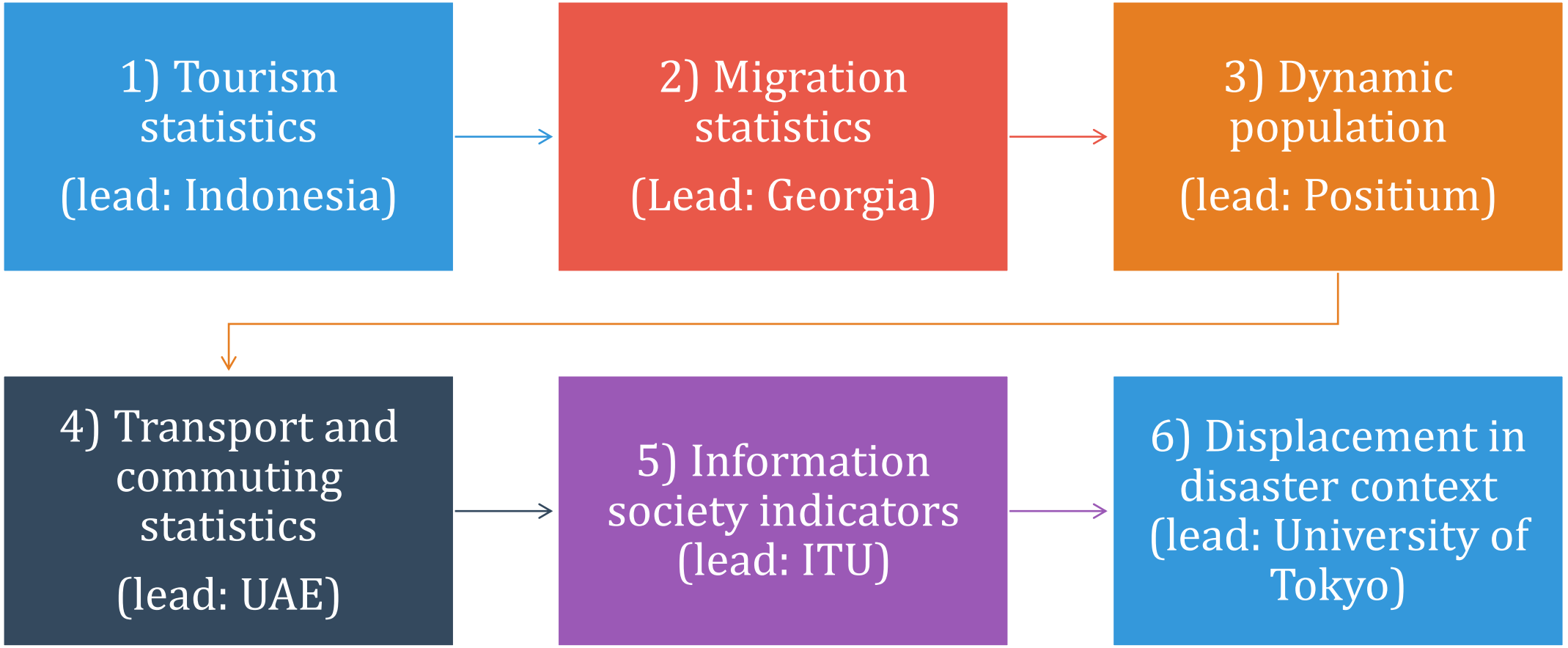
Handbook on the use of Mobile Phone data for Official Statistics (2019)

- English
- Russian

Events

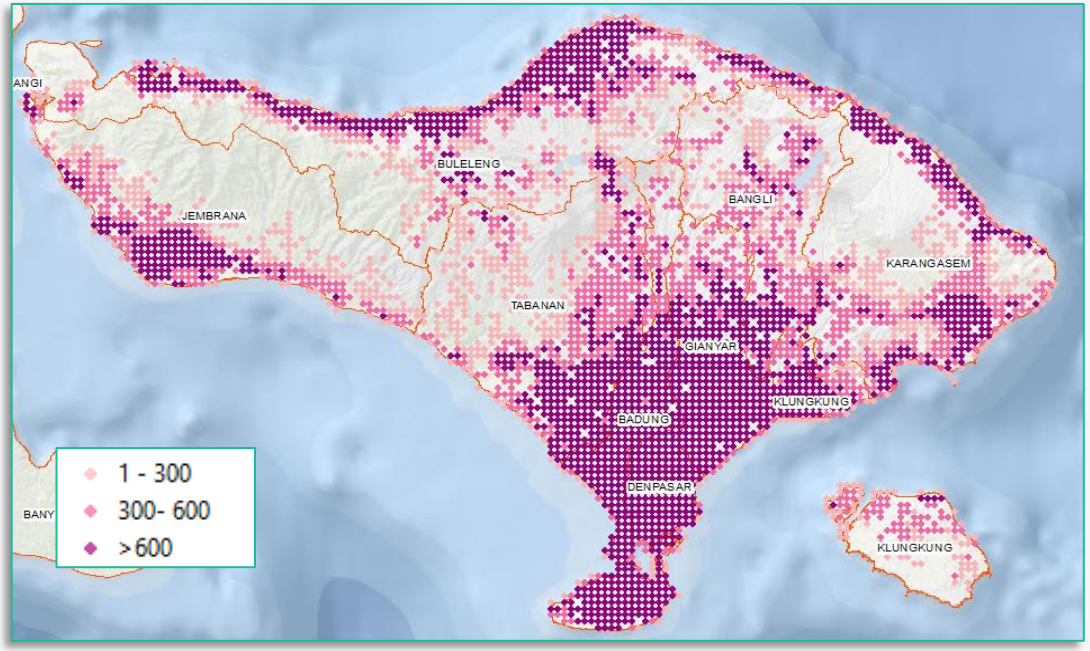
MPD Session at the 7th International Conference on Big Data

Yogyakarta, Indonesia 8 Nov 2022

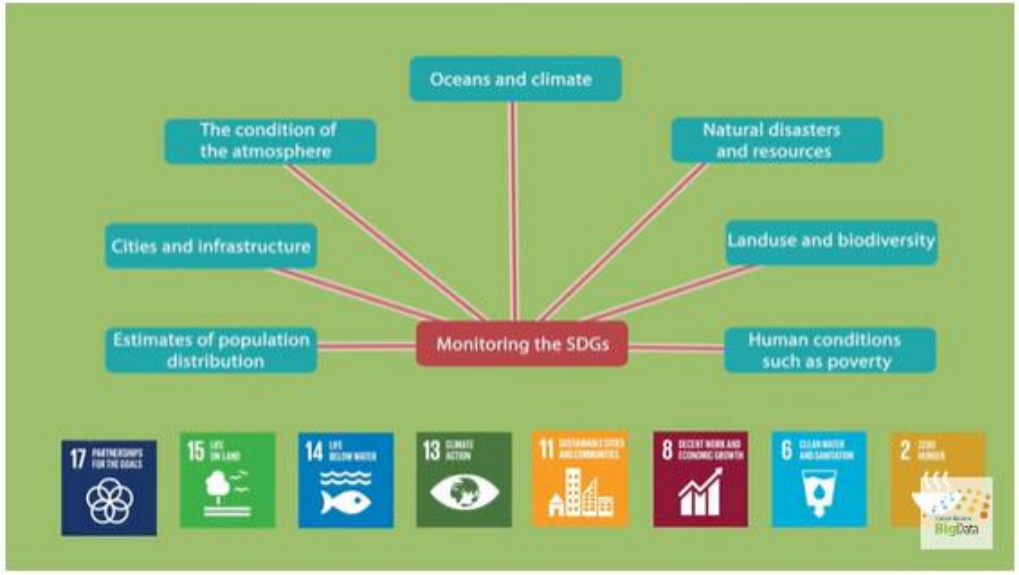

1) Tourism statistics
(lead: Indonesia)
2) Migration statistics
(Lead: Georgia)
3) Dynamic population
(lead: Positium)
4) Transport and commuting statistics
(lead: UAE)
5) Information society indicators
(lead: ITU)
6) Displacement in disaster context
(lead: University of Tokyo)

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

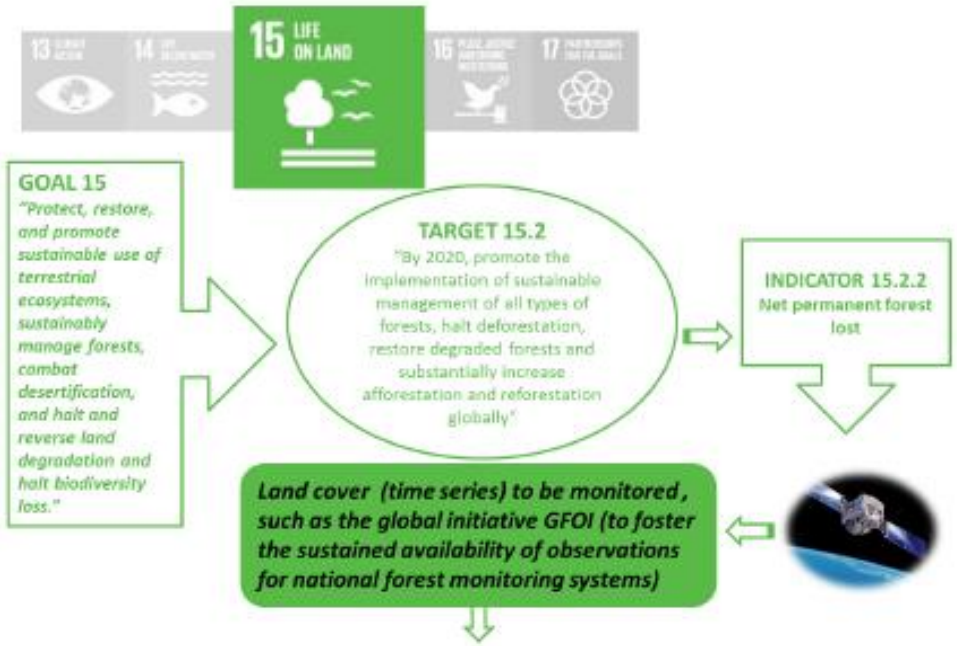
Trusted methods: Lessons Learned and Recommendations from Select Earth Observation Applications on Agriculture

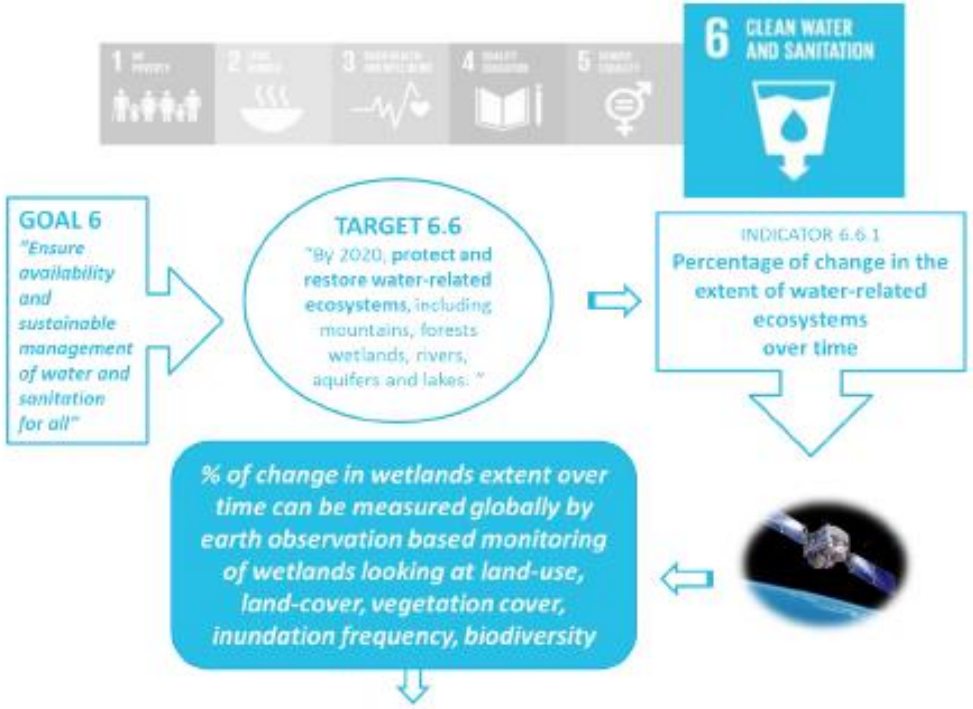
Earth Observations for Official Statistics

Satellite Imagery and Geospatial Data Task Team report

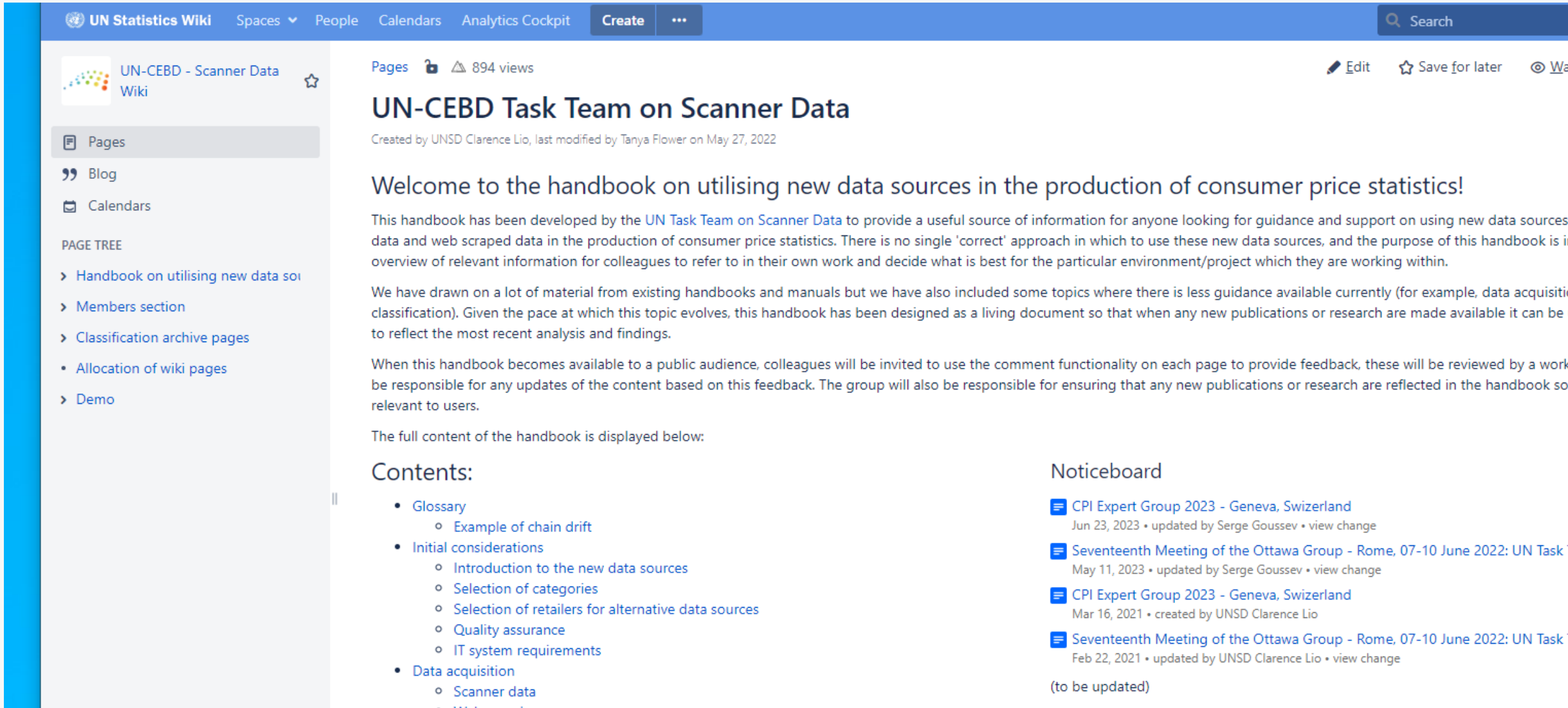
5th December 2017

Examples of EO for SDGs





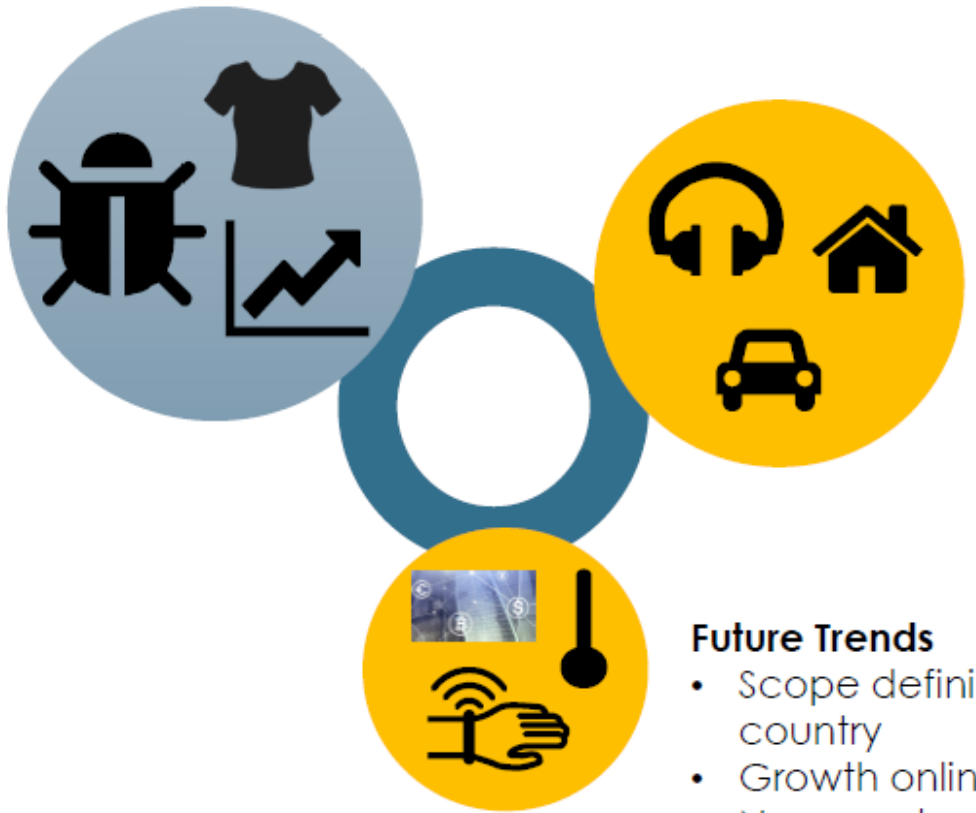
Task Team on Use of Scanner Data and Webscraping for Official Statistics



Implementing more online data in the future

Web Scraped
- Online vs instore
  - Clothing
  - Electronics

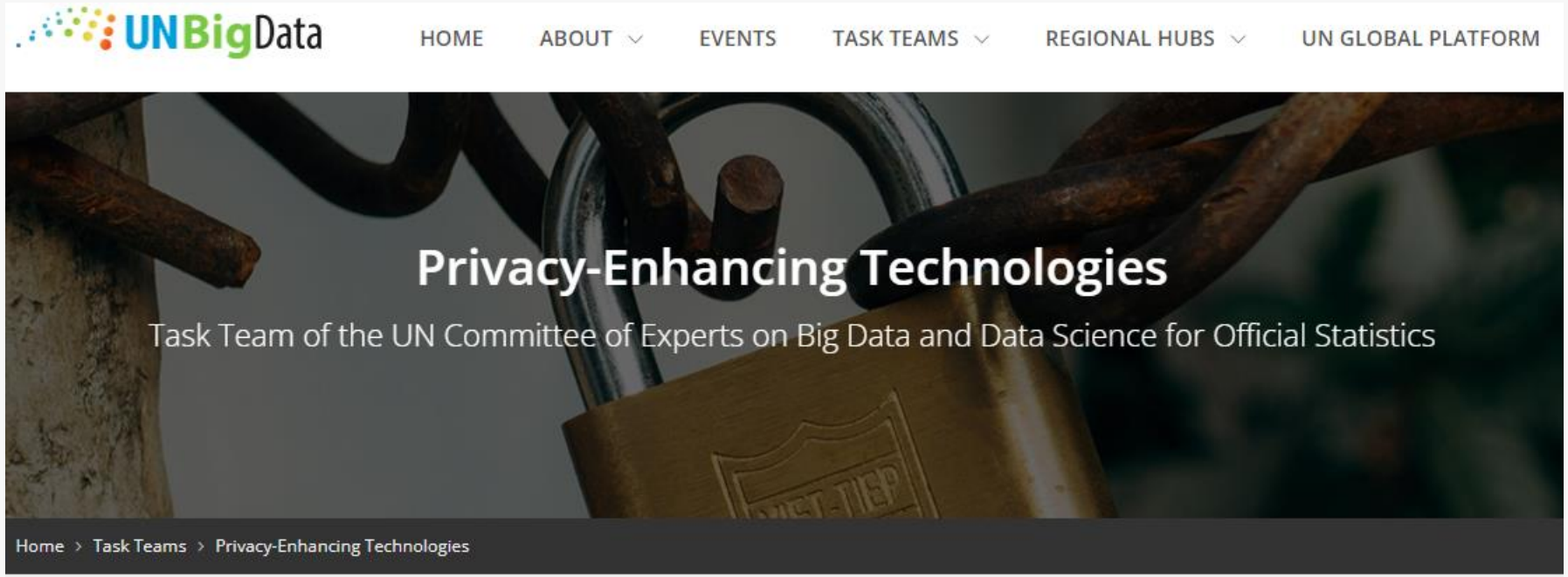
Application Programming Interfaces (APIs)
- Travel



Sharing Economy
- Netflix & Spotify
- Uber & Lyft
- Airbnb

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

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.

Publications

CASE STUDY DESCRIPTION

HIGH LEVEL FUNCTIONAL PERSPECTIVE

HOSPITAL

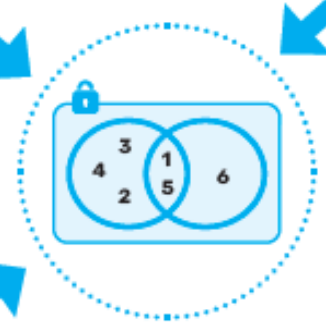
	Example features & data		
Cryptographic ID	Uses eHealth Coach?	Wellbeing Score (PROMs)	...
1	Y	1	...
2	N	2	...
3	Y	3	...
...	...	—	...

INSURER

	Example features & data	
Cryptographic ID	Total Medical Costs	...
1	100000	...
2	5000	...
3	300	...
...	...	...

*What is the **average** education level of patients using the eHealth coach vs. patients not using the coach? (*)*

__How many__ insurees with medical costs over 50000 are using the eHealth coach? (*)



NSO

	Example features & data		
Cryptographic ID	Education Level	Income	...
1	3	25000	...
2	5	100000	...
3	1	43000	...
...	...	—	...

*What is the **average** wellbeing score of patients using the eHealth coach with incomes above 10000 vs. below 10000? (*)*

**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)

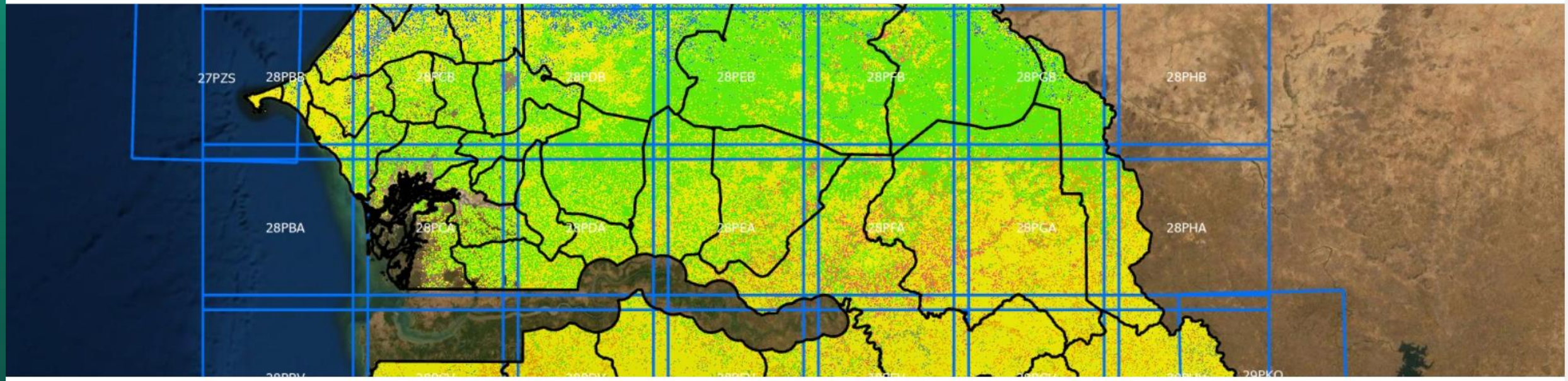
Training in Big Data and Data Science for official statistics

- 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



FAO-EOSTAT: EO data for Official Agricultural Statistics

Building capacity in Senegal, Uganda, Afghanistan, Lesotho.

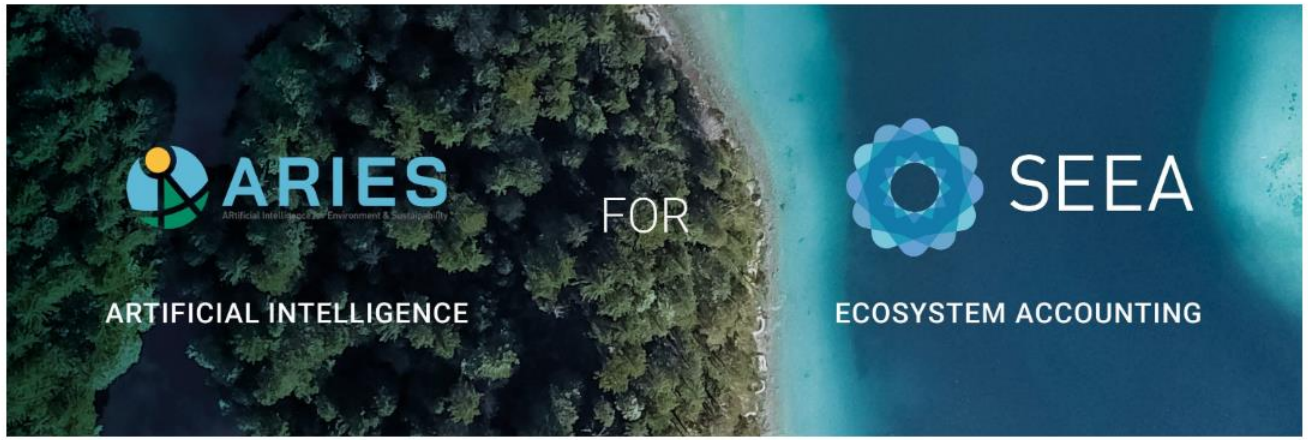
Lorenzo De Simone, Technical Adviser Geospatial, Office of the Chief Statistician FAO
May 4, 2021

The EOSTAT project kicked off in 2020 under the leadership of

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.

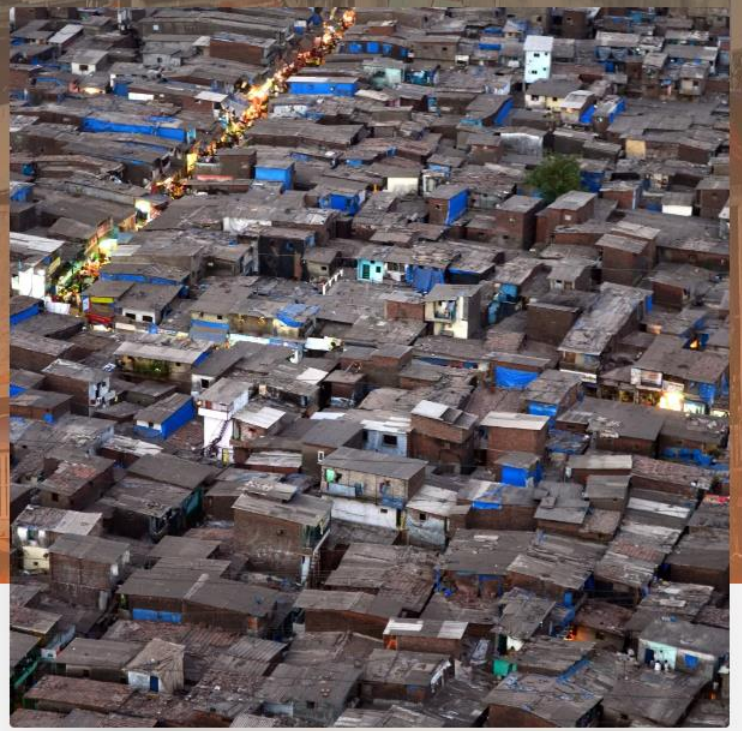
- *Map boundaries*: Select an area of interest by panning and zooming in/out. The entire area visible on the screen becomes your analysis context.
- *Administrative regions*: This option automatically identifies the largest administrative entity (e.g., Country or Subnational Unit) in the area selected, **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

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.

Building The Foundation For Heat Resilient Communities

Read More

GLOBAL HEAT RESILIENCE SERVICE

Building the foundation for heat resilient communities

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.

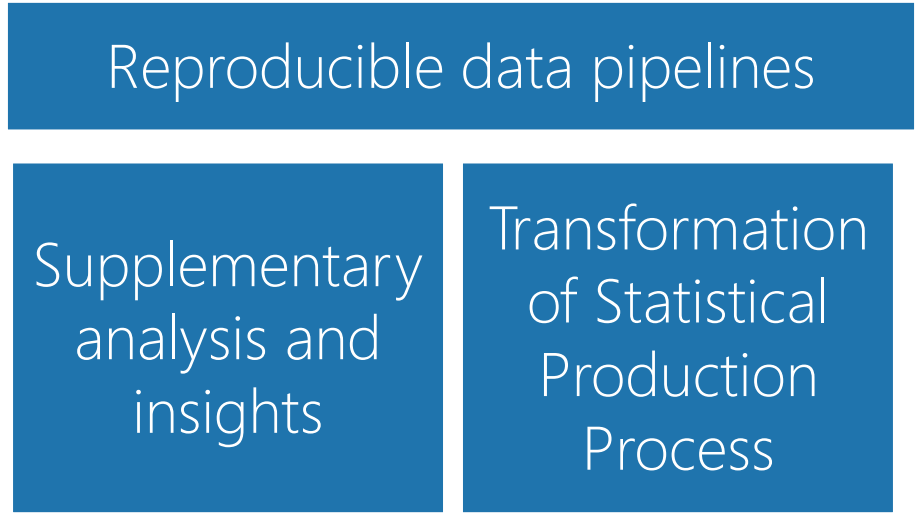
THE IMPACT OF HEAT ON HEALTH AND SOCIETY

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.

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



UNBigData

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