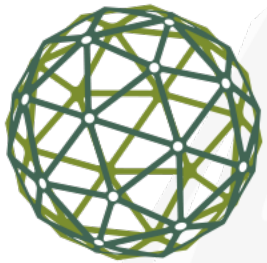


8th International Conference on BIG DATA & Data Science for Official Statistics

BILBAO 2024

Informing Climate Change and Sustainable Development Policies with Integrated Data

BILBAO. SPAIN | 10-14 JUNE 2024 | #UNBigData2024

UN Global Platform

How the UN Statistics Division advances the use of state-of-the-art technologies in statistical offices by enabling access to large volumes of data and modern tools and methods

Luis González Morales
United Nations Statistics Division

United Nations BigData | Eusko Jaurlaritza Gobierno Vasco | Gobierno de España Ministerio de Asuntos Exteriores, Unión Europea y Cooperación | BC3 Basque Centre for Climate Change Klima Aldaketa Ikergai | Excelencia María de Maeztu 2023-2027 | INE Instituto Nacional de Estadística | Eustat Euskal Estatistika Erakundea Instituto Vasco de Estadística

Using big data and data science to enhance official statistics

2013: **Non-traditional sources** of data as potentially useful sources of statistical information

2014: Statistical Commission **strengthened mandates** aimed at enhancing use of big data and related technologies

2017: the idea of a **Global Platform** for data, services and applications (UN Statistical Commission Decision 48/105(d))

2018: UN Global Platform emerges as a **cloud-based data collaboration space** for the statistical community

Condition for use: **Projects must benefit official statistics**

On-going initiatives we support:

- Privacy-preserving data science
- Vessel tracking data (AIS)
- Climate & health indicators
- Statistical data portals (.Stat)
- Modernization of UN Data
- New trade data processing tools
- E-learning courses
- Data4Now

Concluded initiatives:

- Using satellite imagery & machine learning to create modern crop maps in Senegal
- Concept & SDG extraction using semantic web technologies

On-going initiatives we support:

- Privacy-preserving data science
- Vessel tracking data (AIS)
- Climate & health indicators
- Statistical data portals (.Stat)
- Modernization of UN Data
- New trade data processing tools
- E-learning courses
- Data4Now

Concluded initiatives:

- Using satellite imagery & machine learning to create modern crop maps in Senegal
- Concept & SDG extraction using semantic web technologies

Vessel-tracking data (AIS)

- **Goal:** Facilitate NSO access to terabytes of marine vessel location data in a cost-effective cloud-based computing environment
- We built data pipelines to continuously ingest AIS data and built a cloud-based Spark environment for data scientists to run their own solutions, resulting in 15+ research papers plus several data platforms, with data going back to 2018.

Evolution and impact: Harnessing high-frequency data for official statistics

Data scientists in the statistical community can access the location and movements of **all vessels globally**, with data back to **2018**.

Platform is overseen by the **UN Committee of Experts on Big Data** (UN-CEBD), established in 2014 (former UN GWG Big Data)

Real-time AIS updates on global vessel positions and speeds, **35k new records** every two minutes.

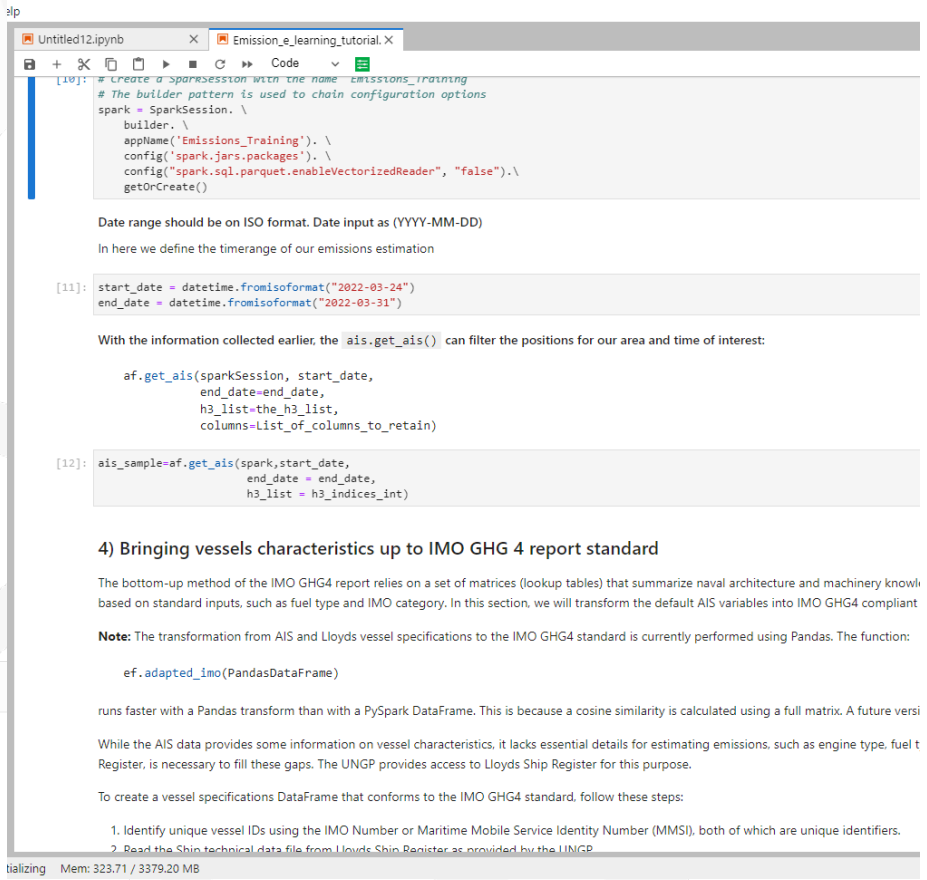
16.8 TB of data available in a cloud-based Kubernetes **Spark environment** with around 300 cores and 1,800 GB memory

15 NSOs, 16 intl organizations, 14 universities and 12 other government institutes use the AIS data on the platform in one or more projects

15+ major research papers have expanded state-of-the-art research methodology using AIS for statistical purposes

The AIS Service on the UN Global Platform

- **Platform-as-a-service:** users should be able to build their own solutions (like PortWatch)
- Built using modern **cloud-native** technologies (K8s, spot instances, serverless)
- Extensive **partnerships** with NSOs in user experience & peer review of technology architectures
- In-house **operations** and **engineering**
- Users **prototype solutions** in Notebook environments
- Also provides a **remote data processing interface** to execute pipelines remotely for remote execution by partners



Behind the scenes: Data engineering

- Data contract with ExactEarth allow data for statistical purposes
- Data engineering efforts are a partnership between UN, ADB, and ONS
- Built using AWS Serverless Application Model in Python
- AWS Lambda functions run every two minutes to pull data from API and process

Behind the scenes: Computing environment

- AWS Kubernetes cluster runs Spark jobs efficiently using a mixture of on-demand and pre-emptible instances
- Critical partnership with NetApp
- Cluster expands and shrinks as resources are needed; typical usage is 300 cores, 1800 GB ram
- Can grow to much more than this, shrink to much less, as needed

Impact of AIS data on Research

- Analysis of global economic impacts of COVID-19 lockdown measures using high-frequency shipping data.
- Monitoring and understanding trade dynamics in specific regions, such as the Pacific Islands.
- Research on CO2 emissions from global shipping and the environmental implications of reduced maritime activity during the COVID-19 pandemic.
- Prediction of export values using big data and machine learning techniques.
- Insights into port and logistics resilience during disasters and large-scale disruptions

Using high-frequency AIS data for statistical purposes: expanding the state of the art

- ✓ Verschuur, J., Koks, E. E., & Hall, J. W. (2021). Global economic impacts of COVID-19 lockdown measures stand out in high-frequency shipping data. PloS one, 16(4), e0248818.
- ✓ Arslanalp, M. S., Koepke, M. R., & Verschuur, J. (2021). Tracking Trade from Space: An Application to Pacific Island Countries. International Monetary Fund.
- ✓ Clarke, D., Chan, P., Dequeljoe, M., Kim, Y., & Barahona, S. (2023). CO2 emissions from global shipping: A new experimental database.
- ✓ March, D., Metcalfe, K., Tintoré, J., & Godley, B. J. (2021). Tracking the global reduction of marine traffic during the COVID-19 pandemic. Nature communications, 12(1), 2415.
- ✓ Al-Saadi, N. (2021). Global economic impacts of COVID-19 after isolating countries from each other. PLOS ONE| https://doi. org/10.1371/journal. pone, 259818.
- ✓ Pham, K. H., & Luengo-Oroz, M. (2020). From plague to coronavirus: On the value of ship traffic data for epidemic modeling. arXiv preprint arXiv:2003.02253.
- ✓ Hoffmann Pham, K. E., & Luengo-Oroz, M. (2020). From plague to coronavirus: vessel trajectory data from ship automatic identification systems for epidemic modeling. Journal of Travel Medicine, 27(6), taaa072.
- ✓ Kim, K., Das, S. B., Pundit, M., Magnata, P., Mariasingham, M., & Chico, C. ADB BRIEFS.
- ✓ Adamu, A. B. (2021). Effects of Covid--19 pandemic on economic activities. Journal of Formal and Informal Sectors, 1(1), 1.
- ✓ Nooraeni, R., Nickelson, J., Rahmadian, E., & Yudho, N. P. (2022). New recommendation to predict export value using big data and machine learning technique. Statistical Journal of the IAOS, 38(1), 277-290.
- ✓ Adamu, A. B. (2021). Proposed Impact of Covid–19 Pandemic on Economic Activities.
- ✓ Verschuur, J., Koks, E., & Hall, J. (2020). The implications of large-scale containment policies on global maritime trade during the COVID-19 pandemic. arXiv preprint arXiv:2010.15907.
- ✓ Paulussen, R., van der Spoel, M., Schenau, S., de Wit, T., Meijer-Cheung, W. K., Bisioti, E., ... & Bis, M. ESSnet Big Data II.
- ✓ Verschuur, J., Koks, E. E., & Hall, J. W. (2020). Port disruptions due to natural disasters: Insights into port and logistics resilience. Transportation research part D: transport and environment, 85, 102393.
- ✓ Fuentes, G., & Adland, R. (2023). Greenhouse gas mitigation at maritime chokepoints: The case of the Panama Canal. Transportation Research Part D: Transport and Environment, 118, 103694.

United Nations BigData

EUSKO JAURLARITZA GOBIERNO VASCO

GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS EXTERIORES, UNIÓN EUROPEA Y COOPERACIÓN

BC3 BASQUE CENTRE FOR CLIMATE CHANGE
Klima Aldaketa Ikergai

EXCELENCIA MARÍA DE MAEZTU 2023-2027

INE Instituto Nacional de Estadística

Eustat EUSKAL ESTATISTIKA ERAKUNDEA INSTITUTO VASCO DE ESTADÍSTICA

#UNBigData2024