BigData
UN Global Working Group

United Nations Global Platform: Data for the world

A global collaboration to harness the power of data for better lives

v1.0.1 Published: 20/08/19

Foreword

Welcome to the United Nations Global Platform!

Under the governance of the United Nations Global Working Group (GWG) on Big Data for Official Statistics the Global Platform is developing a cloud-service ecosystem to support international collaboration in the development of Official Statistics using new data sources and innovative methods and to help countries measure the Sustainable Development Goals (SDGs) to deliver the 2030 Sustainable Development Agenda.

When put to good use data has the power to benefit society and the economy and to improve the well-being of communities and individuals. We believe we are at a crucial point where we must do all we can to further international collaboration to make sure that data is used ethically and appropriately, for the good of humanity.

The potential to access new data sources and analysing that data in new ways has raised expectations. We want more information made available more quickly and at a finer level of granularity whilst continuing to ensure the quality of the products and analysis that is produced. This new agenda offers a tremendous opportunity and a significant challenge to statistical offices around the world: gaining access to data with a known provenance; having a portfolio of sophisticated data analysis tools and fostering the skills required to make the most of what is now possible.

The key to meeting the challenge lies in sharing - of data, of tools and of knowledge - and this is where the Global Platform and the services it provides come in. The platform, which creates a living ecosystem of users (Trusted Partners), Trusted Data, Trusted Methods and an open library of Trusted Learning materials already supports a global community of statisticians working together to explore some of the most complex problems that we face, particularly measuring progress against the Sustainable Development Goal indicators.

You will find more detail here and we hope you will share our enthusiasm for the ‘art of the possible’. Please support our work if you can; it is through working together that we can be successful.

Niels Ploug and Heather Savory
Co-Chairs of the United Nations Global Working Group

Our purpose and mission

A global collaboration to harness the power of data for better lives.

We strive to enable data-driven transformation for better decision making. We seek to improve our world by providing access to trusted data for use at local, national and global levels.

We will deliver a collaborative network of people and organisations, collectively building a Global Platform for trusted data, services and applications.

Introduction

USING THE POWER OF DATA TO TRANSFORM PEOPLE'S LIVES

Digital technologies, including data, are rapidly transforming our world. Technological advances such as low-cost computing, the internet and mobile connectivity mean we are more connected globally than ever before and we have access to unprecedented opportunities to improve people's lives, especially those who have historically been marginalised.

Right now, humanity is facing significant challenges including sustainable development, preservation of the environment and evolving human need. Alongside this, the growing opportunities created by digital technologies are paralleled by stark abuses and unintended consequences. As technology advances, the mechanisms for global cooperation and governance of this new landscape have failed to keep pace.

In our digitally driven world, data has immense power and value. It is vital that we harness this to benefit society and the economy and improve the well-being of communities and individuals. It is also crucial that we utilise data to develop solutions and support existing solutions, such as the Sustainable Development Goals (SDG), to the challenges facing us today.

To achieve this we need:

- Friction free data access
- Robust infrastructure that ensures good governance and security
- The latest analysis and dissemination tools
- The skills and resources to effectively work in this digital environment.

The United Nations Global Platform is working to achieve this.

The UN Global Working Group (GWG) on Big Data for Official Statistics was created in 2014 by the UN Statistical Commission to explore the benefits and challenges of the use of new data sources and technologies for official statistics and Sustainable Development Goals (SDG) indicators.

The United Nations Global Platform was established by the United Nations (UN) GWG as a collaborative environment to work together as a global statistical community and to learn together sharing knowledge, data and methods for all countries in the world.

The GWG has also established multidisciplinary Task Teams who collaborate to research how sources of Big Data such as Mobile Phone Data, Satellite Imagery and Social Media Data can be used to create new insights into economies and societies, and to help measure the SDGs. The Task Teams develop methods

and algorithms using data science, Artificial Intelligence and Machine Learning techniques to derive public benefit from Big Data. Each Task Team brings together experts drawn from National Statistical Systems (NSS) across the world and other invited partners with expertise in specific topics. These teams are international, and the Global Platform enables them to work together on these projects, for the first time.

TRUSTED PARTNERS, TRUSTED DATA, TRUSTED METHODS, TRUSTED LEARNING

Trust is central to the Global Platform. It is founded on the four pillars of: Trusted Partners, Trusted Data, Trusted Methods and; Trusted Learning through platform services.

Trust is achieved through collaboration, peer review and approval of all work undertaken on the platform.

TASK TEAMS

The Office for National Statistics (ONS) has funded and led the creation of the technology platform which supports this work and has delivered a set of 100% Cloud native services for:

- developing and publishing statistical methods and algorithms
- providing information on data services, learning and teams in a marketplace
- specialised services that deliver location and earth observation capabilities.

The platform is Cloud-agnostic with services hosted by all of the major Cloud providers. The Platform team supports the Task Teams on cross-service technical support of these different Cloud services.

The project has been tremendously successful and the approach has attracted a positive response from around the world with over 200 users from over 100 countries accessing the services. The platform has benefits worldwide, and is essential for ensuring that developing nations can contribute their statistical expertise as well as global leaders. It is a truly democratic environment.

As the platform has matured and the value it delivers has increased, the demand for access to the tools, data and methods from Task Teams around the world has increased dramatically. Now is the time to put it on a sustainable footing.

Good data has the power to benefit society and the economy and improve the well-being of communities and individuals. We are at a crucial point where we must do all that we can to establish the collaboration to make sure data is used for the good of humanity.

If you want to find out more about the Global Platform, contact us via info@officialstatistics.org.

Our journey so far...

November 2014
United Nations Secretary General announcement to establish "a network of data innovation networks" for leveraging and sharing data and data research

August 2016
Dublin, the Global Working Group proposes to create a Global Platform for data, services and applications

April 2017
Office for National Statistics, UK starts technical development of the UN Global Platform

2017
Global Cloud Services implemented

November 2017
4th Global Conference on Big Data, Bogota, Colombia

Proof of concept demonstrations

May 2018
Methods Service Alpha

June 2018
Developers Service Alpha

October 2018
2nd World Data Forum, Dubai

Demonstrations of:
Location Analysis Service Alpha
Earth Observation Service Alpha
Global Datasets: AIS, ADS-B & Satellite

March 2019
50th Session of the United Nations Statistical Commission, New York

Marketplace Alpha launched
Business Model Options ratified and order received to proceed with technical development & exploration of Regional Hubs

April 2019
Publication of the Privacy Techniques Handbook

May 2019
5th International Conference on Big Data for Official Statistics Kigali, Rwanda

Strategy Handbook
Global Billing Intelligence

The future...

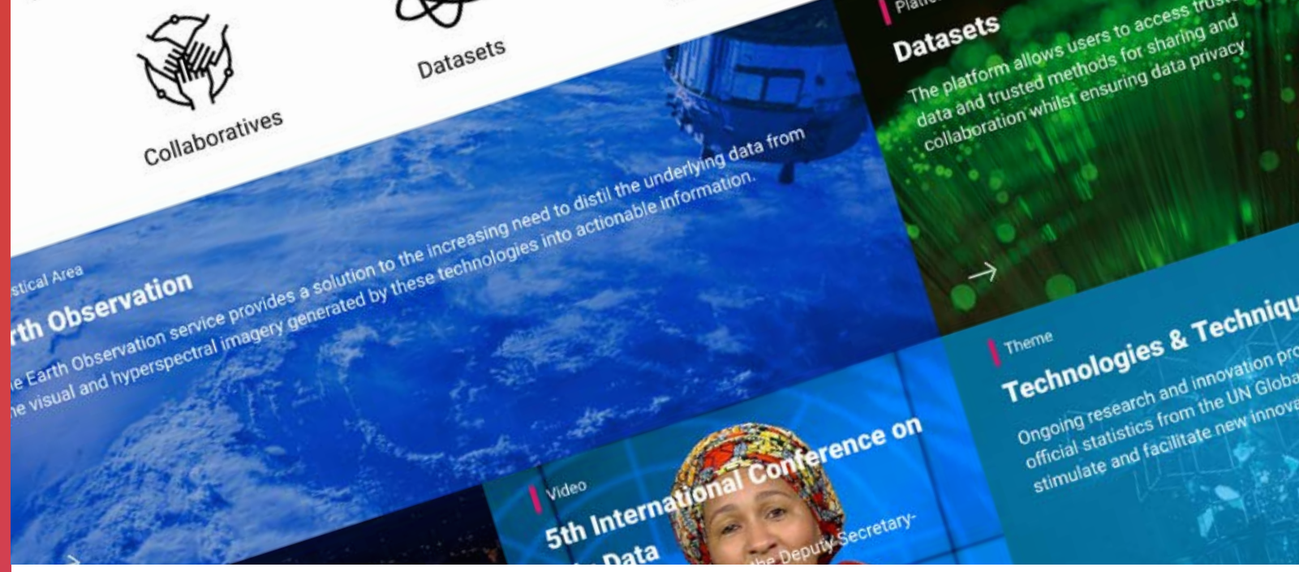
We’re making an impact

1 UN Global Platform for data, services & applications in a multi-cloud environment • Focussed on the **17 SDGs** (Sustainable Development Goals) • **28 member countries** in the UN Global Working Group + **16** International organisations • **7** Global Task Teams producing official statistics on the **1 platform** • **3** global big datasets • **100,000,000,000+** ADS-B records • **80,000,000** AIS records added each day • Live UN Global Platform demos at the 2nd World Data Forum and at the 3rd, 4th & 5th **Big Data Conferences** •

Collaboration with National Statistical Institutes (NSIs) across the world • **Leading world experts** recruited to help deliver state of the art services and products • Platform conception to full growth in **3 years** • **215+ people** in **14 time-zones** in our on-line community • Platform users from **Mongolia** to **New York** • Since launch the marketplace has seen visited by users from **100** countries, April saw a 269% increase in users

Marketplace for the global platform

Whether you are developing a new method or investigating the novel application of a dataset the marketplace provides a trusted, curated and comprehensive index of trusted data, methods, services, learnings and partners.



Features

From the homepage, you can browse or search trusted data, methods, services and learnings by Sustainable Development Goal (SDGs). The content is curated to encourage exploration, and items are presented by topics, services and SDGs.

As well as registering as a user to access information, partner organisations can apply to join as contributors, and enhance the content of the marketplace further. Teams can be created with supervisors and contributors, and each partner can publish content to their own areas and control contributions from team members in a back end interface. They can also be invited to be part of a task team which can stretch across partners to create truly powerful collaborative work.

The marketplace was presented and approved at the UN Statistical Commission in March 2019 for use as the focal point of the platform for all users.

Approach

Since its launch in March 2019, the marketplace has seen visits by users from 100 countries, 127 user accounts created, 184 projects set up, with 1,800+ users accessing it.

April 2019 saw a 269% increase in users with a rise of 282% in new users.

Get Involved...

https://marketplace.officialstatistics.org

marketplace@officialstatistics.org

In Use

As users contribute content to the teams and methods, the marketplace will promote and encourage international collaboration.

Methods Service for the platform

Share and reuse trusted algorithms and methodologies.

The Methods service allows members of the Global Platform to find, use and publish methods and algorithms in the Cloud. For data scientists, it makes it easy to deploy algorithms and machine learning models that become API callable microservices.

Features

The Methods Service is building a library of trusted statistical methods and algorithms, facilitating international collaboration. The Cloud native infrastructure of the service and the APIs allow, for the first time, members of any international organisation to share and utilize the same algorithms.

The service:

- holds a library of algorithms – all of which are callable via API
- offers developers automation of algorithm tests, builds and deployment in a wide variety of programming languages
- supports statistical methods, AI and machine learning models
- encourages active collaboration in algorithm creation
- allows documentation and source code to be made public
- automatically scales to meet demand
- is available everywhere, all the time.

Get Involved...

https://methods.officialstatistics.org

methods@officialstatistics.org

Approach

As the library of available algorithms grows, the effort needed to adopt cutting edge methodology will be substantially reduced everywhere in the world.

Active, international collaboration between data scientists in the development, review and improvement of algorithms is producing trusted methods and algorithms for production of new statistics.

In use

A team from the UK Data Science Campus deployed their Urban Forests project as a series of algorithms in the Methods Service. Each one of these algorithms can be used by the community to sample streets and roads, retrieve images and detect vegetation. You can find more information about the project on page 39.

Exploration and development

Explore data, develop analytics and build applications securely on the Global Platform.

The Developers Service gives you access to code development and data exploration environments.

Features

The Service supports users to develop new algorithms, data science pipelines and applications.

The Service:

- offers Cloud-based workspaces and notebooks
- makes JupyterHub and Gitlab available over the web
- is managed and customized by UN Global Platform and UNSD teams to meet the needs of the Task Team
- supports R, Python, Julia and more languages
- is scalable to new users & intensive processing
- increases the productivity of algorithm and methods development using modern version control
- allows quick and effective collaboration in distributed teams
- offers sophisticated project management tools tailored for data scientists and data developers
- offers native container registry, supporting reproducible research and development
- automates code execution, algorithm deployment, and manages data pipelines using Gitlab runners.

Get Involved...

https://developers.officialstatistics.org

developers@officialstatistics.org

Approach

This service provides environments for programmers and data scientists worldwide to collaborate on projects.

In Use

This Service is used by small teams from the United Nations and several member States to perform a range of data processing tasks in Python and R.

Datasets for the global platform

Access and analyse the big datasets in a cloud-native environment.

The datasets on the platform range from satellite imagery in many spectral bands to real-time shipping and flight data with hundreds of millions of data points to analyse. The data service is designed to handle additional datasets from our partners and when combined with the cloud native Developer and Methods Services provides a uniquely powerful toolset.

Features

The Data Service is built on our multi-cloud platform to provide agnostic access to datasets with robust access and privacy preserving processing controls.

Approach

As part of the pilot in the early phases of the platform we introduced the AIS and ADS-B datasets as the most useful and indicative Big Data sources to provide an active and intensive testbed for Big Data statistical analysis and methods development.

These datasets now comprise 100,000,000+ ADS-B Records and 80,000,000 AIS Records (per day).

Global Datasets

Global ADS-B Data Feed

This live data stream and the dataset is provided by a cooperative of ADS-B/ Mode S/MLAT feeders from around the world and is the world's largest source of unfiltered flight data.

Contribute to global ADS-B collection...

The ADS-B network is reliant upon the distributed network of feeder boxes to provide this valuable realtime data. Europe and North America are currently well served with collection nodes so if you are based in a region which could benefit from data collection we would like to help.

We are operating a programme to distribute free ADS-B feeder boxes to developing countries and those regions currently under served – get in touch to request your feeder box today.

flights@adsbexchange.com or contactus@officialstatistics.org

Global AIS Data Feed

The AIS data stream and dataset delivers the most complete situational picture of global vessel activity by drawing data from both satellite and terrestrial AIS receivers.

Sustainable Development Goals (SDG) Indicators

SDG indicators are a global indicator framework for the Sustainable Development Goals (SDG) and succeed the Millenium Development Goals (MDGs) which ended in 2015. They include indicators and statistical data to monitor progress, inform policy and ensure accountability of all stakeholders for the 17 goals set by the United Nations. Data available is from the early 1990's to present and covers economic and social development issues including poverty, hunger, health, education, social justice, water, sanitation and many more.

Satellite datasets

Sentinel-1

Sentinel-1 is a pair of European synthetic-aperture radar (SAR) satellites launched in 2014 and 2016. Its 6 day revisit cycle and ability to observe through clouds makes it perfect for sea and land monitoring, emergency response due to environmental disasters, and economic applications.

Sentinel-2 (Sentinel-2A and Sentinel-2B)

The Sentinel-2 mission is a land monitoring constellation of two satellites that provide high resolution optical imagery and provide continuity for the current SPOT (Satellite Pour l'Observation de la Terre) and Landsat missions. The mission provides a global coverage of the Earth's land surface every 5 days, making the data of great use in on-going studies.

- multi-spectral data with 13 bands in the visible, near infrared, and short wave infrared part of the spectrum (443nm – 2190nm)
- the first optical Earth observation mission of its kind to include three bands in the 'red edge', which provide key information on vegetation state
- spatial resolution of 10m, 20m and 60m.

This picture captured by Landsat8 (OLI) of irrigated crops and Kahov irrigation canal. The image is created as True Color Composite, where R – Red Band (0.64 - 0.67 µm), G – Green Band (0.53 - 0.59 µm) and B – Blue Band (0.45 - 0.51 µm). This band combination is suitable for crop monitoring.

Landsat 8

First launched in 1972, the Landsat series of satellites has produced the longest, continuous record of Earth's land surface as seen from space. Providing moderate-resolution imagery, from 15 meters to 100 meters, of Earth's land surface and polar regions, Landsat 8 operates in the visible, near-infrared, short wave infrared, and thermal infrared spectrums.

Data is collected from nine spectral bands. Seven of the nine bands are consistent with the Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) sensors found on earlier Landsat satellites, providing for compatibility with the historical Landsat data, while also improving measurement capabilities.

Two new spectral bands, a deep blue coastal / aerosol band and a shortwave-infrared cirrus band, are collected, allowing the measurement of water quality and improve detection of high, thin clouds.

MODIS

Moderate Resolution Imaging Spectroradiometer (MODIS) data is available for anyone to use via Amazon S3. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands.

These MODIS instruments offer an unprecedented look at terrestrial, atmospheric, and ocean phenomenology.

Get Involved...

https://marketplace.officialstatistics.org/datasets

contactus@officialstatistics.org

Other Datasets

Other datasets currently available on the platform include:

Africa Soil Information Service (AfSIS) Soil Chemistry

This dataset contains soil infrared spectral data and paired soil property reference measurements for georeferenced soil samples that were collected through the Africa Soil Information Service (AfSIS) project, which lasted from 2009 through 2018.

This dataset can help to recognise clear-cutting and deforestation practices that contribute to climate change and threaten local biodiversity.

NOAA Global Historical Climatology Network Daily (GHCN-D)

Global Historical Climatology Network - Daily is a dataset from NOAA (National Oceanic and Atmospheric Administration) that contains daily observations over global land areas. It contains station-based measurements from land-based stations worldwide.

OpenStreetMap

OpenStreetMap (OSM) is a free, editable map of the world, created and maintained by volunteers. Regular OSM data archives are made available in Amazon S3.

Neilsen Scanner Data

This dataset from Nielsen contains weekly detailed (barcode level) data for five product groups covering a three year period. The dataset is intended to support training and testing activities promoting the use of scanner data in the compilation of Consumer Price Indexes. The data is available to all users of the Global Platform.

Scanner data is a Big Data source being increasing used in national statistical systems for the calculation of price indices as statistical offices explore ways to meet the expectation of society for enhanced products and improved, more efficient ways of working.

Earth Observation

Advances in small satellite and remote sensing technology are revolutionising the Earth Observation industry. They are providing critical new sources of consistent, continuous data for atmospheric, ocean and land use studies at a range of spatial and temporal extents.

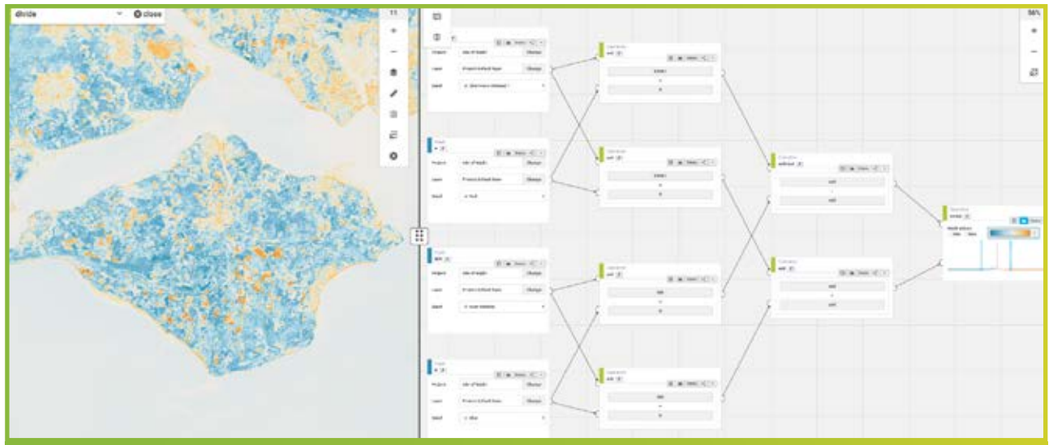
The Earth Observation Service provides a solution to the increasing need to distil the underlying data from the visual and hyper-spectral imagery generated by these technologies into actionable information.

Features

Our Earth Observation Service makes it possible for anyone to experiment with building custom algorithms or apply existing algorithms to satellite imagery they want to analyse.

You can:

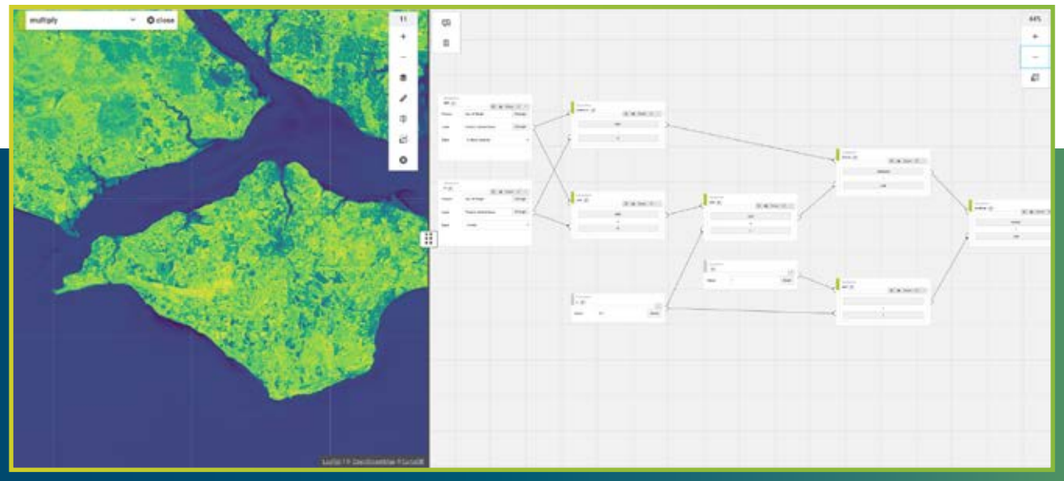
- Search through and use a library of common algorithms like NDVI, NDWI, SAVI, and more.
- Build your own custom algorithms using a visual interface
- For advanced users, use the Python client or REST API directly to bring analyses into your own work-flows and dashboards
- Search through massive imagery datasets, both public and commercial
- Upload your own data and control access permissions across your organisation
- Share imagery publicly as a map service and integrate it into your website or application.



Example of a processing pipeline [right] in the Earth Observation Service displaying the bare soil index for a custom defined region

Approach

The most basic task of searching through, visualising, and sharing data has become a challenge. Our service offers an imagery catalogue that makes it easy for organisations to make all of the geospatial imagery they use available in one place, and easily shareable across multiple teams.



A pipeline [right] **producing a SAVI (Soil Adjusted Vegetation Index) for a custom defined region**

Using the Earth Observation Service it was a simple matter to implement the SAVI using data drawn from the Platform's Data Service. Working from the left each block of the pipeline performs operations and its outputs are then passed on to the next element to be processed. The modular approach allows for the creation of powerful real-time processing of satellite data.

The SAVI index is a transformation technique that minimises soil brightness influences from spectral vegetation indices involving red and near-infrared (NIR) wavelengths. The transformation was found to nearly eliminate soil-induced variations in vegetation indices.

Copernicus Sentinel-2 is a two-satellite mission. Each satellite carries a high-resolution camera that images Earth's surface in 13 spectral bands. The mission is mostly used to track changes in the way land is being used and to monitor the health of our vegetation.

In Use

The Global Platform delivers a direct feed to this data and other sources. Datasets currently available on the platform include:

- Landsat 4 + 5 Thematic Mapper,
- Landsat 7 Enhanced Thematic Mapper,
- Landsat Tri-Decadal,
- Landsat 8,
- Sentinel 2A and 2B.

The Task Team on Satellite Imagery and Geospatial Data have produced a guide for National Statistical Offices considering using satellite imagery. The Task Team are now using the platform and Earth Observation Service to develop this work and produce re-usable resources for the global community.

Get Involved...

https://marketplace.officialstatistics.org/earth-observation

Analyse anything that moves.

The Location Analytics Service is a scalable solution for storing, indexing, querying, transforming, and visualising spatio-temporal data, offering the capability to animate millions of entities in your browser.

Features

The Location Analytics Service is a dynamic spatial visualization service.

The Service allows you to:

- create stunning visual displays of large spatio-temporal data, using millions of data points from real-time data feeds
- find new spatial patterns in moving data by visualizing patterns of life and anomalies
- view real-time data feeds, or conduct forensic analysis
- explore and visualize 100 bilion+ ADS-B Records and 80 million AIS records (per day) all from within your browser.

Approach

The Location Analytics Service is being used with other platform services to provide faster indicators of economic impact.

In Use

Currently 15 countries are actively using the service for exploration and analysis work.

Get Involved...

https://location.officialstatistics.org

contactus@officialstatistics.org

The AIS Task Team (page 49) & Faster Economic Indicators (page 38) both use the Location Analytics Service.

SERVICES **Platform Authentication Service** Discovery

Authenticate anyone, anywhere, securely and easily

The UN Global Platform provides a secure, flexible, easy to use, central authentication to all Platform Services for users, whilst allowing each Service its own authentication protocols and authorisation levels.

Features

The authentication abstraction allows users to have a consistent authentication process providing single sign-on across multiple services. The Authentication service is able to utilize OAuth, OpenID, SAML and LDAP protocols.

The Service supports multiple methods of Multi-Factor Authentication (MFA), including token generators, log in devices, location, IP address and time of day. The Service can be configured in minutes and services allocated to the users with a self-service catalogue and application approval workflows.

The service can be configured in minutes and platform services allocated to the users with a self-service catalogue and application approval workflows.

The Authentication Service is also able to provide meaningful reporting on applications and detailed reporting on authentication activities.

Approach

This service is currently in 'discovery'. Once it has been user tested and fully evaluated, the Authentication Service will reduce the on-boarding time of a user / users to minutes.

In Use

This is a discovery phase project to understand and evaluate the available options for the future of the platform. At the end of the discovery, the platform Task Team will be deciding on the approach to take.

Global financial billing intelligence

Advanced financial auditing for cloud-based projects within the platform.

The Service provides segregated financial billing and auditing between platform partners and projects.

Features

Customers benefit by purchasing through an intermediary, as they can buy on terms that differ from how the Cloud Provider routinely sells.

In the case of public Cloud services that are contracted, delivered and invoiced on a highly standardised basis, an intermediary can add significant value:

- billing - tailoring the mechanism by which payment is made
- finance - tailoring the timing of the payment
- risk - tailoring commitments regarding future usage and pricing.

Approach

The Service will monitor, manage and account for the contractual and financial risks, including credit risk, foreign exchange risk and cash flow risks. By proactively monitoring these we can pre-emptively provide insight into Cloud spend.

Impact

The Service:

- makes the most of every penny or dollar
- manages currency fluctuations and making the best use of global purchasing power
- allows partners to pick the best Cloud region based on workload and billing intelligence
- reduces the carbon footprint of energy intensive computational tasks
- offers advanced financial insight to allow for ongoing improvements to financial commitments and workload management.

SERVICES **Security and Compliance Service**

Cloud Provider Certification & Compliance

The Cloud Providers used by the UN Global Platform are Amazon Web Services, Google Cloud Platform, Microsoft Azure and Alibaba.

The Global Platform facilitates the exchange of open data, where safe to do so, as well as agreed sensitive data. It also offers data library services for sensitive data and works to develop trust in data and data security. It is vital that users and producers of data can trust the security of the platform.

The following is a list of some of the Certifications and Compliances held by the Cloud providers used by the UN Global Platform.

HELD BY ALL FOUR CLOUD PROVIDERS:

ISO 27018:2014

establishes commonly accepted control objectives, controls and guidelines for implementing measures to protect Personally Identifiable Information (PII) in accordance with the privacy principles in ISO/IEC 29100 for the public cloud computing environment.

ISO 27001:2013

specifies the requirements for establishing, implementing, maintaining and continually improving an information security management system within the context of the organisation

ISO 27017:2015

provides guidelines for information security controls applicable to the provision and use of cloud services.

CSA STAR Level 2 (Cloud Security Alliance - Security Trust & Assurance Registry)

involves a rigorous independent third-party assessment of a cloud provider's security posture. This STAR certification is based on achieving ISO/IEC 27001 certification and meeting criteria specified in the CCM. It demonstrates that a cloud service provider conforms to the applicable requirements of ISO/IEC 27001, has addressed issues critical to cloud security as outlined in the CCM, and has been assessed against the STAR Capability Maturity Model for the management of activities in CCM control areas.

SOC 1 (Service Organisation Controls)

is a report that focuses on a service organisation's controls that are likely to be relevant to an audit of a user entity's (customer's) financial statements.

SOC 2 (Service Organisation Controls)

is a report that addresses a service organisation's controls that relate to operations and compliance, as outlined by the AICPA's Trust Services criteria in relation to availability, security, processing integrity, confidentiality and privacy.

SOC 3 (Service Organisation Controls)

is a public facing report demonstrating the Service organisation has met the AICPA's Trust Services criteria in relation to availability, security, processing integrity, confidentiality and privacy.

HIPAA (U.S. Health Insurance Portability and Accountability Act)

is a US federal law that established data privacy and security requirements for certain entities and individuals aimed at safeguarding individuals' health information.

PCI DSS (Payment Card Industry Data Security Standard)

The PCI Data Security Standards define operational and technical requirements for organisations accepting or processing payment transactions, and for software developers and manufacturers of applications and devices used in those transactions.

GDPR (General Data Protection Regulation)

is a regulation in EU law on data protection and privacy for all individuals citizens of the European Union (EU) and the European Economic Area (EEA). It also addresses the export of personal data outside the EU and EEA areas.

HELD BY: AWS, GOOGLE AND AZURE

FedRAMP (US Federal Risk and Authorisation Management Program)

a US government-wide program that provides a standardised approach to security assessment, authorisation, and continuous monitoring for cloud products and services.

FIPS (Federal Information Processing Standard)

is a US government standard that specifies the security requirements for cryptographic modules that protect sensitive information.

HELD BY: ALIBABA AND AZURE

ISO 22301:2012

specifies the requirements to plan, establish, implement, operate, monitor, review, maintain and continually improve a documented management system to protect against, reduce the likelihood of occurrence, prepare for, respond to, and recover from disruptive incidents when they arise.

HELD BY: ALIBABA

DJCP MLPS

The Multi-Level Protection Scheme (MLPS) is a tiered protection system for information security in China. The goal of this system is to develop a unified national information security protection management system and standards.

Project Galileo

Founded in 2014, Project Galileo is Cloudflare's response to cyber attacks launched against important, yet vulnerable targets, like artistic groups, humanitarian organisations, and the voices of political dissent. These attacks have become all too prevalent and powerful in recent years.

The Internet is a powerful tool for spreading and expanding ideas. Cloudflare®'s mission to help build a better Internet includes protecting free expression online for vulnerable groups. When journalists, social activists, and minority groups are repeatedly flooded with malicious traffic in an attempt to knock them offline, and keep them offline, the Internet stops fulfilling its promise.

As an UN affiliated organisation the UN Global Platform benefits from our participation in this programme offered by Cloudflare ensuring that our infrastructure is both protected and available to all users across the world. Central to our mission is the ability to produce official statistics to aid decision making and support the UN Sustainable Development Goals, project Galileo provides the backing to enable this to continue in an environment where threats to technology and continued free-speech are evolving.

More information about Project Galileo can be found at www.cloudflare.com/galileo/

Case studies from the platform

The platform is already delivering benefits by enabling collaboration across a wide range of global partners.

These case studies provide a few examples of how the UN Global Platform is being used to co-create new and innovative outputs.

CASE STUDY **Faster Economic Indicators of UK Economic Activity: Shipping**

Producing new, faster, indicators of economic activity using new data sources

In this project two new monthly shipping indicators for the UK are derived from Automatic Identification System (AIS) data that was accessed through the UN Global Platform.

The two indicators are:

- 'Time-in-port', based on aggregate time spent by ships in 10 major UK ports
- 'Total traffic', based on the number of unique ships entering these ports.

Overview

These indicators are likely to be important in supplementing the understanding of international trade activity. They offer a fast indication of the level of shipping activity, which is, as we have shown, related to trade in goods. The relationships between the indicators and international trade in goods could potentially be used in a mathematical model in combination with other indicators to estimate trends in trade.

Given the fact that AIS can be obtained in a timely fashion (in fact in near real time) the output of this can be a valuable tool for early economic trend discovery.

Approach

Using rectangular bounding boxes, defined for the 10 biggest UK ports by cargo, the AIS messages reporting location in any of the ports have been filtered and subsequently aggregated to monthly numbers of unique ships visiting port as well as the amount of time in port.

The relatively good correlation between the port traffic indicators and trade in goods (imports) allows near real time estimation of trends in imports from AIS data directly. The estimates are published monthly as part of the faster economic indicators programme.

Impact

Expanding the study on a global scale is of potentially significant interest to many partners on the UN Global Platform as it might offer the possibility of understanding how the global economy is evolving in close to real time. It will allow investigation of network effects and the study of ship movement patterns between trade partners and how these may evolve over time.

CASE STUDY **Urban Forest**

Producing a greenery index from cloud-based data sources and pipelines

Implementation of the Urban Forest project onto the cloud-based platform.

The Urban Forests project aims to create an index of vegetation through taking Google street view images around the road network and classifying the amount of vegetation in each image.

Overview

Largely, the process of implementing this on the platform has involved development of a series of algorithms in the Methods Service (page 11). The methods service allows for methods to be written in a range of languages (for example R, Python or Java) which can, in turn, be called from a wide range of languages. This allows for a combination of languages to easily be used in a single pipeline. Each part of the pipeline is contained within its own algorithm, allowing them to be called independently of one another. This case study will walk through the creation of a number of these algorithms, how these have been combined to form a pipeline and how this could be reused.

Approach

The first key component of the pipeline is to sample a requested geographical area. This might be a street, a parish or borough, or an even entire city.

The original Urban Forests project downloaded the OpenStreetMap data for a requested number of roads. In the implementation on the platform, we have created a number of algorithms that allow a query to generate evenly spaced points along all highways in the queried geographical area.

Using images from Google street-view the algorithm uses the map path and generates a series of images from views ninety degrees either side of the direction of travel.

© Mapillary vistas dataset

The image segmentation algorithm is the same algorithm as used in the Data Science Campus' Urban Forest project. It implements the PSPnet [@PSPnet], a convolutional neural network that achieves high levels of accuracy through a combination of scene-parsing with pixel-level prediction. The vegetation index simply returns the percentage of pixels classified as 'vegetation' in an image. This can then be averaged for each point (as there as two images being used at each point) or over a larger geographical area such as a street or parish.

One key benefit of the platform is the ability to make multiple API calls to the same algorithm, thus allowing parallel processing. This has the potential to significantly reduce processing times, particularly when processing a large geographical area.

Impact

The main impact from this project was to demonstrate that an entire pipeline can be built in the Methods Service, using a combination of common languages used in data science and running these methods in parallel on cloud to realise the benefits promised by this service. The learning from this process has been documented and is shared with new users of the platform interested in implementing methods or pipelines in the Methods Service.

All these components are open on the Methods Service for reuse and re-purposing. For urban analytics, the road sampling method may be very useful; the code is open source and therefore can be made to return a geojson file rather than a list of coordinates. The sampler in combination with the image downloader can be used to create new data sets which can be labelled and then used for novel machine learning projects.

OpenStreetMap demonstrating `way_id` points

As a result of his work on this project Joe Peskett from the ONS Data Science Campus won the Young Digital Professional of the Year award at the UK Digital Technology Leaders Awards 2019. This award is awarded to those 'playing an integral role in digital innovations and projects' and who 'bring new, exciting ideas to their organisation' and celebrates people in the early years of their career who are making a marked impact on their organisation by bringing fresh thinking and measurable impact.

Joe Peskett (left) receiving his award at the ceremony in London, July 2019

https://events.computing.co.uk/digitalleadersawards/static/2019-winners

Project lead:

Phil Stubbings (ONS Data Science Campus)

Collaborators: Joe Peskett & Phil Stubbings, ONS Data Science Campus

Data sources:

- OpenStreetMap data using osmdata package in R access overpass API This was used to access line strings describing streets
- Google street view images, accessed via API.

Methods:

- 1 method for calculating evenly spaced points along road [HighwayScrapeR] written in R
- 1 method for collecting images at these points using the street view api [street_split_view]
- 1 method for segmenting the images [vegetation]
- 1 method for creating composite images [Colourise_segments].

Learning outputs: blog posts written, presentation given at Kigali conference, methods published and documented in the UN Global Platform Methods Service

Services used: UN Global Platform Methods Service, overpass API and Google street-view API.

Find the datasets, methods and learnings from this project on the platform marketplace: https://marketplace.officialstatistics.org/

CASE STUDY **Task Team on Scanner Data**

Producing consumer price statistics using alternative data sources

The production of consumer price statistics is seen as a high-profile output of National Statistics Offices (NSOs) across the globe. These statistics play a vital role in the decisions undertaken by business, government and in everyday life. It is therefore crucial to measure these changes in price as accurately as possible.

NSOs are starting to recognise the role that alternative data sources, such as point of sale transaction data (otherwise known as scanner data) from retailers, can play to improve price statistics.

Scanner data can be used to produce more timely and granular inflation statistics for businesses, individuals and government. While some NSOs have been able to move these data into their production systems, the majority of offices are still in the research phase of this implementation process. Having a collaborative approach to this research will greatly improve the speed at which these data can be implemented in production.

Overview

The aim of the Task Team, established by the Global Working Group on Big Data (GWG), is to increase the effective use of scanner data in official statistics through providing access to training and guidance material for these data and how they can be used to produce consumer price indices.

Approach

The Methods Service (page 11) on the platform allows the Task Team to share a number of price index methods that have been coded up in a range of languages. These index methods can then be run on a historical expenditure and volume data of a sample scanner data set, which allows users to test new methodologies and algorithms for price index calculation.

Impact

While the project is still in its early phases, initial feedback from countries is positive, with an overarching agreement that having access to a collaborative space and consistent methodology will benefit the development of using alternative data sources in consumer prices. The next steps in our collaboration would include uploading further index methods onto the platform and exploring more of the pre-processing required before index methods can be used (for example, classification and error detection).

Project lead: Nathalie Brault (chair) and Jonathan Wylie (Statistics Canada)

Collaborators: Antonio Chessa (Statistics Netherlands), Thomas Hjorth Jacobsen (Statistics Denmark), Michael Holt (Australian Bureau of Statistics), Tanya Flower (Office for National Statistics, UK), Matt Stansfield, Dave Lum (Statistics New Zealand), Ken van Loon (Statbel, Belgium), Michael Smedes (United Nations)

Data sources: Historical retail transaction data

Methods: FEWS package

Learning outputs: Documentation and guidance on the use of price index methodology for scanner data and web scraped data.

Services used: UNGP Methods Service

CASE STUDY **Privacy Preserving Techniques Task Team**

Developing Techniques to Preserve the Privacy of Sensitive Data

The Privacy Preserving Techniques Task Team (PPTTT) is advising the UN Global Working Group (GWG) on Big Data on developing the data policy framework for governance and information management of the global platform, specifically around supporting privacy preserving techniques.

Overview

This Task Team is developing and proposing principles, policies, and open standards for encryption within the UN Global Platform to cover the ethical use of data and the methods and procedures for the collection, processing, storage and presentation of data taking full account of data privacy, confidentiality and security issues.

By law, official statisticians have been given access to highly sensitive information on persons and businesses via population and economic censuses, household or business surveys, and all kinds of administrative data. Certified staff of National Statistical Offices (NSOs) can have access to these data, and NSOs apply very strict confidentiality rules to ensure that nobody else will get access to those data.

Chief statisticians of every country will guarantee this confidentiality and the associated public trust in the work of the NSO. This means that use of any data, which has been acquired by the NSO, by third parties is strictly prohibited. If a dataset of the NSO would be considered for use in a project involving anyone other than NSO staff, this dataset would have to be encrypted in such a way that it would be of use for machine processing, but not of use to any person having access to the data.

In a similar sense, mobile phone companies and credit card companies are very protective of the personal information, which they are guarding, and thorough encryption would be the only way in which these privately held data sets would be made available for access by third parties.

Privacy-preserving computation technologies have emerged in recent years to provide such protection while enabling valuable statistical analyses.

Approach

Certain kinds of privacy-preserving techniques, like secure multi-party computation, allow computing on data while it remains encrypted or otherwise opaque to those performing the computation, as well as to anyone else. Because data can remain encrypted during computation, that data can remain encrypted "end-to-end" in analytic environments, so that the data is immune to theft or misuse. However, protecting such data is only effective if we also protect against what may be learned from the output of such analysis. Additional kinds of emerging privacy-preserving computation technologies address this concern, protecting against efforts to reverse engineer the input data from the outputs of analysis.

Unfortunately, privacy-preserving computation comes at a cost: current versions of these technologies are computationally costly, rely on specialized computer hardware, are difficult to program and configure directly, or a combination of the above. Thus, NSOs may need guidance in assessing whether the cost of such technologies can be appropriately balanced against the resulting privacy benefits. The Task Team has defined specific goals for privacy-preserving computation for public good in two salient use cases: giving NSOs access to new sources of (sensitive) Big Data and enabling Big Data collaborations across multiple NSOs.

Impact

Data security is of paramount importance to NSOs. Confidentiality and privacy are a major concern for citizens and they are therefore as such reflected in the Fundamental Principles of Official Statistics. This work on data security is vital to provide assurance for all platform stakeholders and users. The Task Team has delivered the Privacy Preserving Techniques Handbook, available on the Platform. More advanced work is ongoing including the development of pilots for the exchange of sensitive transactional level trade data between administrations of two or more countries.

Task Team Chair:

Mark Craddock
(UN Global Platform)

Collaborators

Government & NSIs:

- **Ronald Jansen** (United Nations Statistical Division)
- **Matjaz Jug** (Centraal Bureau voor de Statistiek, CBS)
- **Ted Stormen** (Centraal Bureau voor de Statistiek, CBS)
- **Andy Wall** (Office for National Statistics, ONS)
- **Robert McLellan** (Statistics Canada)
- **Aalekh Sharan** (National Institution for Transforming India, NITI)
- **Ira Saxena** (National Institution for Transforming India, NITI)

External collaborators:

- Nigel Smart (KU Leuven)
- David Archer (Galois Inc.)
- Kristin Lauter (Microsoft Research)
- Kim Laine (Microsoft Research)
- Borja de Balle Pigem (Amazon Research)
- Dan Bogdanov (Cybernetica)
- Adria Gascon (Alan Turing Institute)
- Mariana Raykova (Google)
- Andrew Task (Oxford University)
- Eddie Garcia (Cloudera)
- Simon Wardley (Leading Edge Forum)
- Rebecca N. Wright (Barnard College)

Learning outputs:

Publication of the **UN Privacy Preserving Techniques Handbook** (http://bit.ly/ungp-privacy-handbook-2019)

CASE STUDY **AIS Task Team**

Automatic Identification System (AIS) data for faster economic indicators and experimental data

AIS is an automatic tracking system of ships worldwide and is available in the UN Global Platform as global dataset. Thus, its data has the potential to meet the high demand for faster economic indicators and experimental data that can help to meet the 2030 SDG goals.

Overview

The provision of AIS data enables members of the UN Global Platform to experiment with this promising data source for timely information and to share the applied algorithm. For instance, AIS data is used for tracking shipping and fishing activities; calculation of CO_2 emissions in ports as well as obtaining information of humanitarian emergencies as the migration in the Central Mediterranean Sea.

The projects of the AIS Task Team will analyse how to use AIS data for providing faster indicators in various fields of research, such as:

- Economic and trade activities
- Maritime indictors
- Fishery monitoring
- Piracy
- Migration
- CO_2 emissions.

These analyses include the preparation and cleaning of the data, the provision of algorithms and the evaluation of its usefulness in the various applications. For ensuring high standards of the procedures, the AIS Task Team is a diverse team composed of IT specialists, data scientists and specialists from the research fields.

Approach

An open source approach is valuable since it enables the work to reach a large audience. Therefore, the AIS Task Team will provide cleaned AIS data for all users of the UN Global Platform. Furthermore, analyses in various fields of research will be presented and the proposed algorithms shared for usage in other applications.

Impact

The availability of cleaned AIS data and proven algorithms, as well as the possibilities to collaborate enables a wide range of researchers the necessary “ready-to-use tools” to produce timely statistics and experimental data for improved policy decisions.

CASE STUDY **Handbook - Information Technology Strategy**

‘Change has never happened this fast before, and it will never be this slow again’

Information Technology (IT) continues to play an essential role in all aspects of statistical processing throughout the entire production life cycle from data collection through to dissemination. This is a fast moving and rapidly changing environment with new innovations being developed at a breath-taking rate.

Overview

This handbook provides guidance and recommendations on IT strategy for National Statistics Institutions embarking on the use of Big Data in the production of their official statistics and indicators. It uses Wardley Maps to understand where the Institute needs to invest in development and where it can use off-the-shelf solutions.

The Handbook covers principles from knowing your users (e.g. customers, shareholders, regulators, staff) and their needs, through challenging assumptions, removing bias, using appropriate methods (e.g. agile vs lean vs six sigma), being pragmatic and managing inertia (e.g. existing practice, political capital, previous investment) to choosing effectiveness over efficiency while leaving no one behind.

The Handbook will be updated to cover new topics like Cloud, IoT, Artificial Intelligence, Data Science and Serverless technology and security, as well as the use of standards and generic models for IT management.

Approach

This handbook has been prepared by the Technical Delivery Board , a sub-group of the UN Global Working Group (GWG) on Big Data for Official Statistics.

Impact

The Handbook gives a high-level overview of the IT strategy for a modern statistical institute, which should be of benefit for senior managers, and provides sufficient detail to be of interest to IT professionals. This Handbook will be referenced by the Fourth Edition of The Handbook of Statistical Organisation.

Members of the Technical Delivery Board & Contributors are:

- **Mark Craddock**, Chair UN Global Platform Technical Delivery Board; Technical Director, UN Global Platform
- **Rob McLellan**, UN Global Platform Technical Delivery Board; Chief Technology Officer, Statistics Canada
- **Matjaz Jug**, UN Global Platform - Technical Delivery Board; Office of the CIO, CBS (Centraal Bureau voor de Statistiek) Netherlands
- **Ronald Jansen**, UN Global Platform - Technical Delivery Board; Assistant Director, Chief Data Innovation and Capacity, Statistics Division, Department of Economic and Social Affairs, United Nations
- **Bogdan Dragovic,** UN Global Platform - Technical Delivery Board, Statistics Division, Department of Economic and Social Affairs, United Nations
- **Jan Murdoch,** UN Global Platform - Technical Delivery Board; Head of Technology Strategy, Office for National Statistics
- **Simon Wardley**, Contributor; Researcher, Leading Edge Forum

contactus@officialstatistics.org
www.officialstatistics.org

To find out more information about any of our projects or services
visit www.officialstatistics.org, email contactus@officialstatistics.org
or follow us on twitter @UNBigData