

SAS Viya : a catalyst for the future Statistical Office

Michel Philippens
Principal Advisor Public Sector



Intro



2021

implementation of SAS Viya platform at Statistics Flanders – implementation of a finetuned RobBERT-based twitter sentiment classifier in SAS

2004

Started working for SAS
Banking and Insurance Consultant

2001

Graduated at the Department of Sociology - Survey Methodology

2019

Attended a seminar in Brussels organized by the VSA on Smart Statistics and new techniques in Official Statistics.

2007

Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents"

2001-2004

Research assistant on Data Quality issues in the European Social Survey

Concept of Smart (AI-based) Statistics

NEW DATA SOURCES



Michel Philippens

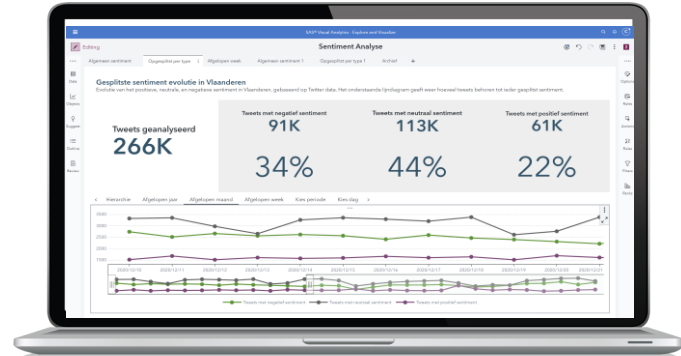
@sbxmip

Proud father of Klaas, Boris en Stan
SAS evangelist, 43, Mechelen

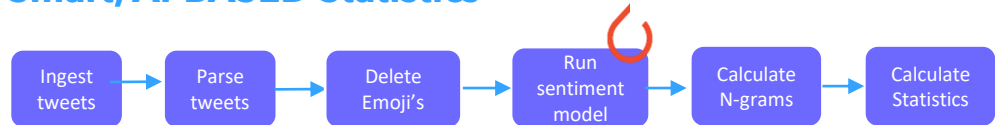
These are the current symptoms after my first dose of Moderna vaccine: extreme fatigue. Soreness at the injection site. Swelling at the injection site. Otherwise, I am fine.

7:22 AM – 22 Feb 2021

TIMELY, CHEAPER, GRANULAR, RELEVANT INSIGHTS



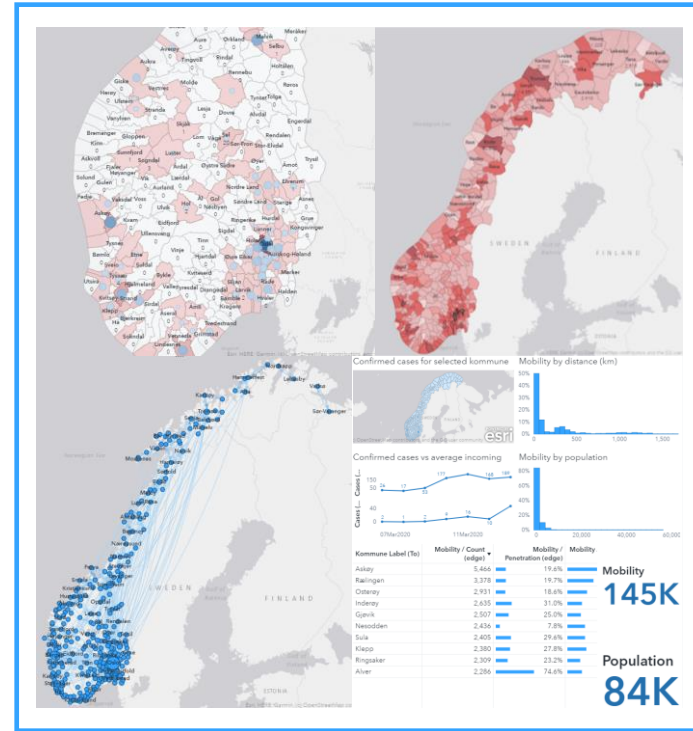
Smart, AI-BASED Statistics



Fuel for Policy Evidence

Based on mobile data to infer physical movements and the occurrence of COVID-19 positive cases over time, we can use Network Analytics state-of-the-art algorithms to provide relevant KPIs by geographical locations

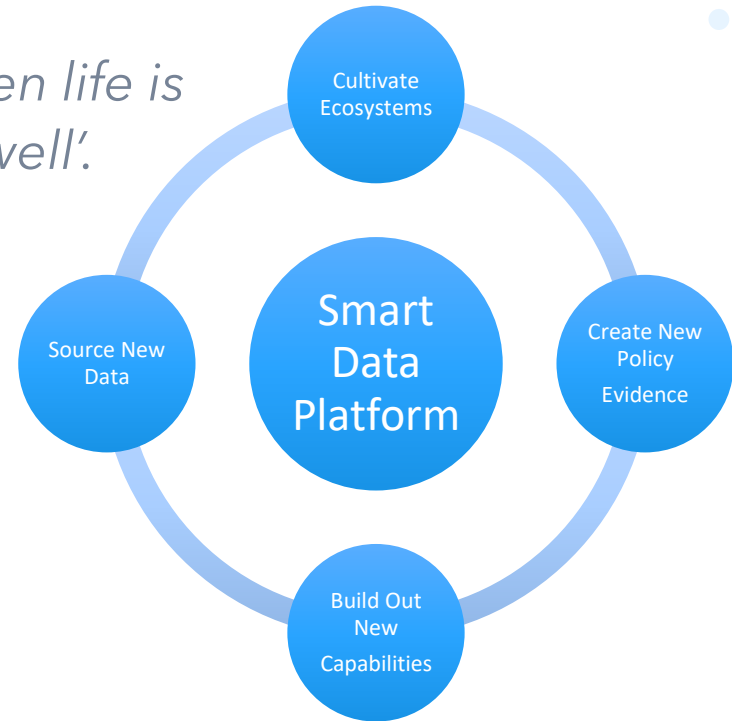
KPIs dashboard and visualization reports allow decision makers to have accurate insights about the pattern of the spread within locations and eventually take social measures to control the transmission rate



NSO's are on a Transformational journey

'Official statistics measures life, and when life is changing, official statistics changes as well.'

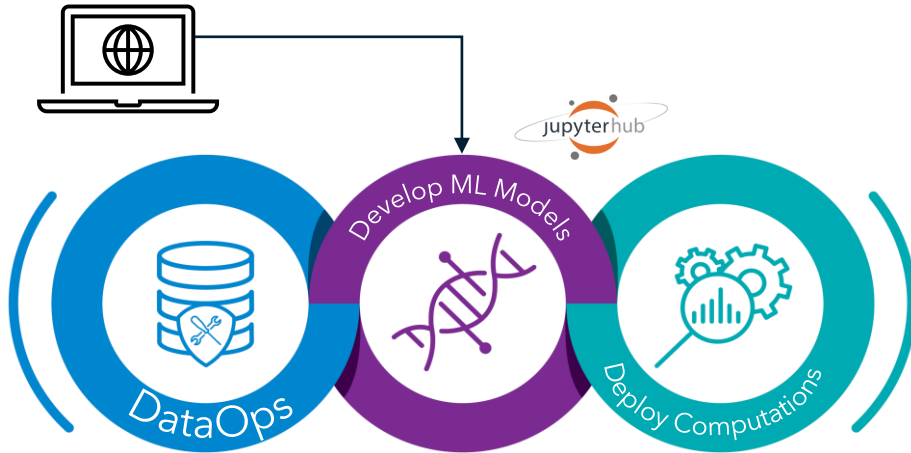
Mariana Kotzeva - Director-General of Eurostat.



Graphic taken from "Which Strategies for NSOs in the Digital Era? Towards 'Smart Data' Strategies", OECD, 2018

SAS VIYA: a key enabler for Smart Statistics

SAS Viya is an AI, analytic and data management platform that runs on a modern architecture.



- 1 Open source Integration
- 2 Collaborative Platform
- 3 Data processing everywhere
- 4 Cloud-Native Platform
- 5 Openness

SAS Viya has been made to boost productivity by being simple, efficient, transparent

Bryan Harris, CTO SAS

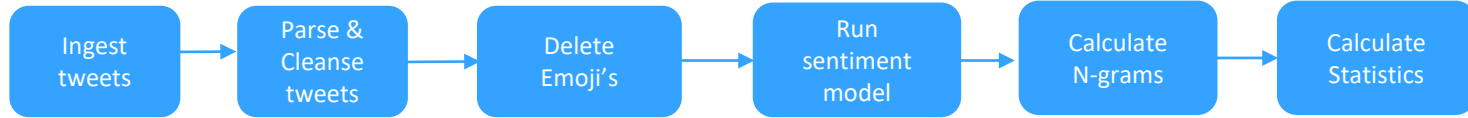
1. Open-Source Integration

“Very successful open-source projects provide free data science tools, plus free algorithms and access to large expert communities - they represent a huge opportunity for statistical organisations, provided they invest in them by making them key components of their next generation platform”

Which Strategies for NSOs in the Digital Era? Towards ‘Smart Data’ Strategies, OECD, 2018

- 1 Open source Integration
- 2 Collaborative Platform
- 3 Data processing everywhere
- 4 Cloud-Native Platform
- 5 Openness

SAS opens up its interfaces to open source



The screenshot displays the SAS Studio interface with a workflow and associated code. The workflow consists of the following steps:

- et nment
- Load Datasets from disk if...
- Configuration Batch times
- Get Data from Twitter API
- Parse JSON to Table
- delete_emojis
- Join sentimen
- Get Cleaned Text
- Run_Sentimen

The code for the `Run_Sentiment` node is as follows:

```
Code Node Notes
1 from twitter_sentiment_classifier import batch_predict
2 import pandas as pd
3
4 df = SAS.sd2df(SAS.symget("_input1"))
5
6 tweet_list = df['text'].tolist()
7
8 predictions = batch_predict(tweet_list)
9
10 df['sentiment'] = predictions
11
12 SAS.df2sd(df, SAS.symget("_output1"), char_lengths={'text':1600})
13
```

Overlaid on the right is a browser window showing the `TwitterSentimentClassifier` page. It includes the following sections:

- TwitterSentimentClassifier**
- Deep Learning model behind the Vlaams Twitter Sentiment webapplication, developed by Radix.
- 1. Installation**
- `pip install git+https://github.com/vsa-datascience/vlaams-twitter-sentiment-model.git`
- 2. Usage**
- ```
from twitter_sentiment_classifier import batch_predict

texts = [
 'Ik haat u!', # Negative
 'Daar ben ik het mee eens', # Neutral
 'Ik hou van je!' # Positive
]

batch_predict(texts)
```

The Radix.AI logo is visible in the bottom right corner of the browser window.

# SAS VIYA adds metadata and transparency

SAS Information Catalog can be used as the repository of metadata, including information on entities, attributes, relationships and domains. around data.

The screenshot displays the SAS Information Catalog interface. At the top, there's a search bar with the placeholder text "What assets are you looking for?". Below this, a "CATALOG AT A GLANCE" section provides a summary of assets: Total assets (1.8 K), Data sets (1.4 K), Studio-stromen (155), Reports (99), Models (60), Model Studio projects (28), and Model Manager projects (22). A "WELCOME" message is also present, encouraging users to take a tour or visit the user's guide.

The "COLLECTIONS" section shows a list of assets. The table below details the visible items:

| Name                 | Status | Library | Date Created       | Created By            | Asset T... |
|----------------------|--------|---------|--------------------|-----------------------|------------|
| AIRBNB_SENTIMENT     | ●      | Public  | 23 Mar 2022 10:17  | sas Unix Service Acct | CAS-ta...  |
| AIRBNB_SENTIMENT     | 📍      | Public  | 22 Mar 2022 17:23  | Michel Philippens     | Tabel      |
| AIRBNB_SENTIMENT     | ●      | PUBLIC  | 22 Mar 2022 18:01  | --                    | Tabel      |
| COOSTO_BERICHTEN_KS  | ●      | Public  | 3 Mar 2021 09:30   | sas Unix Service Acct | CAS-ta...  |
| COOSTO_Sentiment.flw | --     |         | 3 Oct 2022 17:56   | Michel Philippens     | Studio ... |
| SentimentFlow.flw    | --     |         | 21 Mar 2022 22:29  | Michel Philippens     | Studio ... |
| COOSTO_BERICHTEN_KS  | ●      | Public  | 2 Oct 2022 15:34   | Michel Philippens     | Gegev...   |
| geo_v0.1             | --     |         | 13 Sept 2022 10:19 | Wesley Wuyts          | Report     |

18 items

## 2. Collaborative Platform

*"The shortage of data science expertise is a widely known challenge. Statistical operators should become more of data scientists, picking up tasks in activities such as data sourcing, data modelling, data exploration, data analytics. This gap calls for a data skills strategy and a review of existing structures."*

*Which Strategies for NSOs in the Digital Era? Towards 'Smart Data' Strategies, OECD, 2018*

- 1 Open source Integration
- 2 Collaborative Platform
- 3 Data processing everywhere
- 4 Cloud-Native Platform
- 5 Openness

# Collaborative Platform

SAS Viya implements multiple integration points that bring SAS and Open-Source users together.



Hey, I have build a process to cleanse and pre-process tweets

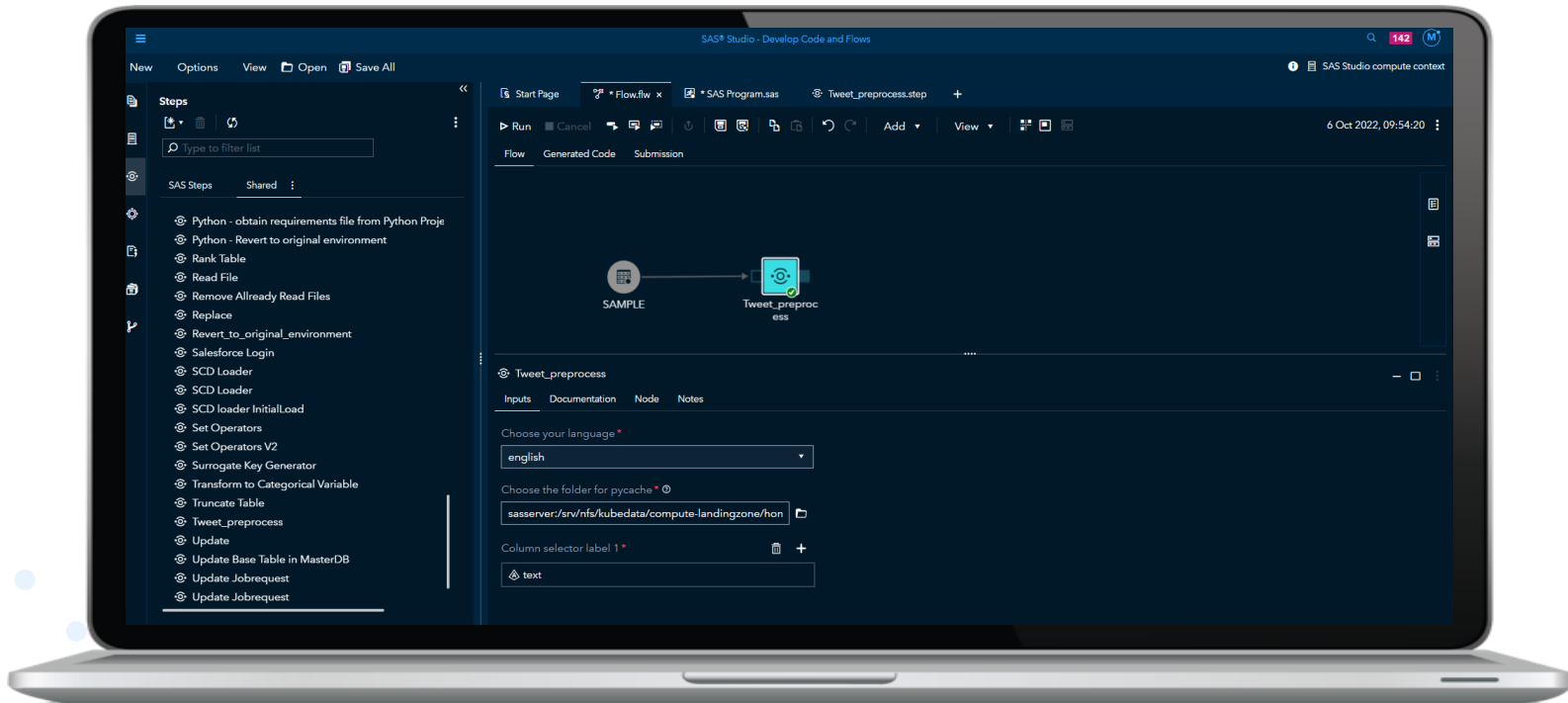


Would be great if we can also use this too!



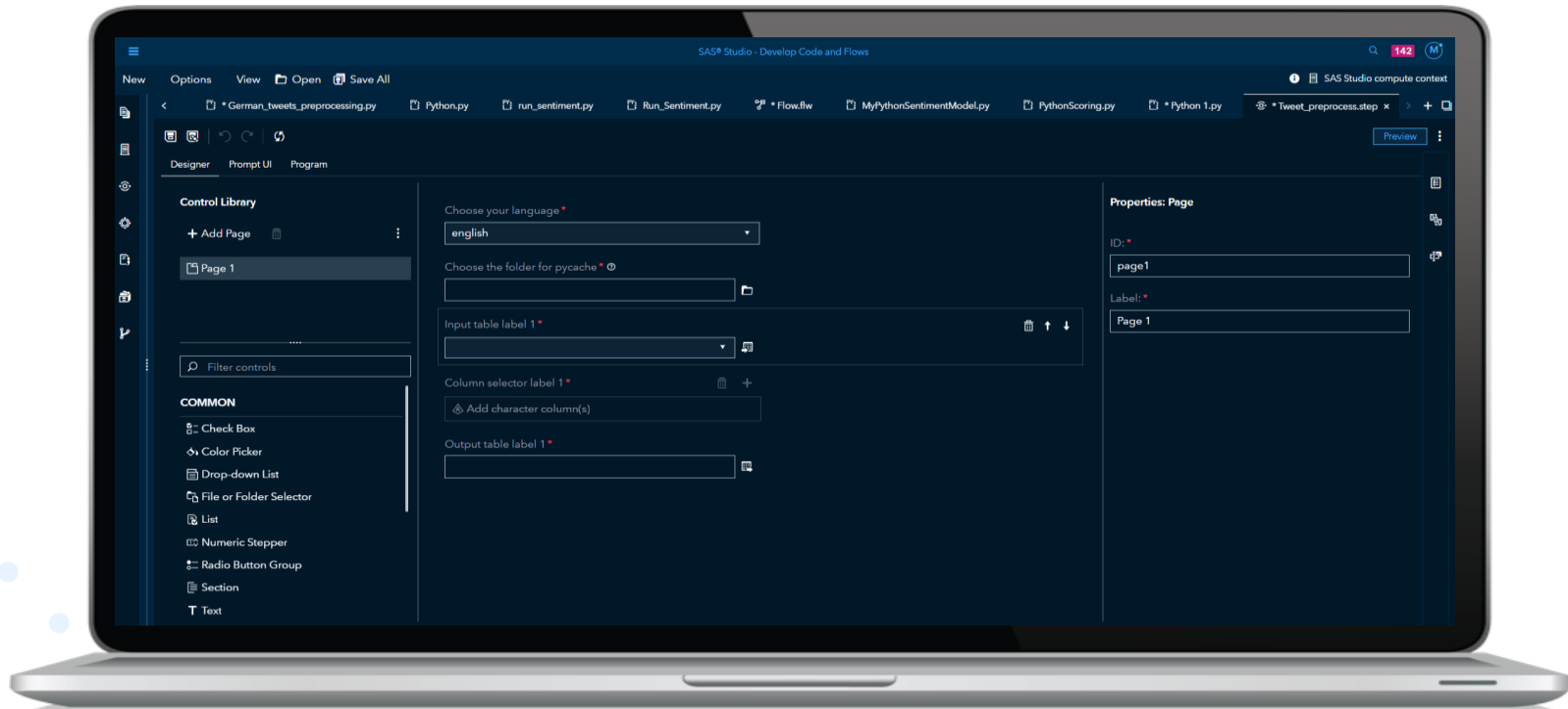
# "Custom Steps" as reusable code blocks

A **custom step** enables you to create a **user interface** for users at your site to **complete a specific data processing task**.



# A low-code “Custom Step” designer

A **custom step** enables you to create a **user interface** for users at your site to **complete a specific data processing task**.



# Open-Sourcing Custom Steps

The screenshot shows the GitHub interface for the repository 'sassoftware / sas-studio-custom-steps'. The repository is public and has 8 watchers, 6 forks, and 5 stars. The main branch is 'main'. The repository contains several files and folders, including a README, a license, and various custom step files. The 'About' section describes the repository's purpose: to share SAS Studio Custom Steps and create a user interface for SAS Studio users to complete specific tasks.

Repository: [sassoftware / sas-studio-custom-steps](#) (Public)

Watch 8 Fork 6 Star 5

Code Issues Pull requests 1 Actions Projects Security Insights

main 1 branch 0 tags

Go to file Add file Code

snlwh Update CUSTOM\_STEPS\_LIST.md 8b40050 6 hours ago 97 commits

|                                         |                                                                          |             |
|-----------------------------------------|--------------------------------------------------------------------------|-------------|
| .github                                 | First Public Release to Public                                           | 22 days ago |
| Anonymize and Mask Data                 | version 1.1 update                                                       | 8 days ago  |
| Create Listing of Directory CLOD        | DIAP DAFT CLOD - final adjustments to make them public ready for fir...  | 15 days ago |
| DQClustering                            | DQ Clustering step added                                                 | 14 days ago |
| DQMatch                                 | Fixes execution errors when using a locale other than English -United... | 10 days ago |
| DQStandardize                           | DQStandardize step added                                                 | 14 days ago |
| DQSurviving                             | DQSurviving step added                                                   | 6 days ago  |
| Data Ingestion Auto Pilot DIAP Light... | DIAP DAFT CLOD - final adjustments to make them public ready for fir...  | 15 days ago |

About

Repository to share SAS Studio Custom Steps. A custom step enables you to create a user interface for SAS Studio users at your site to complete a specific task.

- Readme
- Apache-2.0 license
- 5 stars
- 8 watching
- 6 forks

Releases

# 3. "Data processing everywhere"

*"The increased sensitivity of personal nano-data.....drive towards the introduction in the statistical system of computation models based on distributing the computation outwards"*

*"moving part of the computation to the data sources often (but not always) allows to minimize the global requirement for technical resources. The rate and volume of new digital data is often huge, but only a tiny part of the information embedded in the data is relevant for official statistics"*

*Trusted Smart Statistics: How new data will change official statistics. Ricciato et al, 2020*

- 1 Open source Integration
- 2 Collaborative Platform
- 3 Data processing everywhere
- 4 Cloud-Native Platform
- 5 Openness and Dissemination

# Vancouverborg

LLI NO: 302133 | IMO: 9213741 | Flag: Netherlands | Status: Live | Destination: Montreal | ETA: 12 Feb 2022



Search by vessel name, IMO, LLI or MMSI...

- Overview
- Movements
- Incidents
- Fixtures
- Sanctions
- Ownership
- Hull Risk
- Vigilance
- STS Pairings

- Summary
- Route Information
- Latest News
- Timeline

## Summary

Last Updated: 08 Feb 2022

Upload Image

Moving | Voyaging | AIS Draught: 7.3 m | AIS Speed: 12.8 knots

North Atlantic Ocean | HIGH SEAS

|                          |                                           |               |                                       |
|--------------------------|-------------------------------------------|---------------|---------------------------------------|
| LLI NO:                  | 302133                                    | IMO:          | 9213741                               |
| Flag:                    | Netherlands                               | Vessel Type:  | general cargo with container capacity |
| Status:                  | <span style="color: green;">●</span> Live | Reg. Owner:   | Wagenborg Shipping B.V.               |
| Built:                   | 2001                                      | DWT:          | 9850                                  |
| GT:                      | 6361                                      | Hull Type:    | Single                                |
| Latest AIS Message Type: | A                                         | TEU Capacity: | 552                                   |

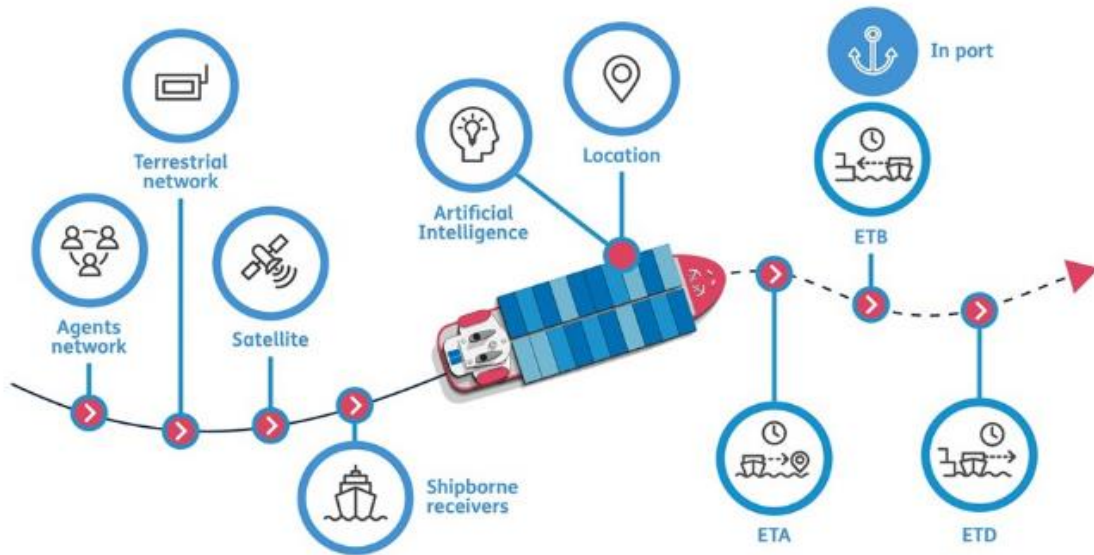
## Route Information

Last Updated: 08 Feb 2022

View methodology



# Lloyds' list intelligence: Predictive Fleet Analytics



## Seasearcher Predictive Fleet Analytics

**Summary**  
Last Updated: 08 Feb 2022 [Upload Image](#)

**Moving** **Voyaging** AIS Draught: 7.3 m AIS Speed: 12.8 knots

North Atlantic Ocean **HIGH SEAS**

|                          |             |               |                                       |
|--------------------------|-------------|---------------|---------------------------------------|
| LLI NO:                  | 302133      | IMO:          | 9213741                               |
| Flag:                    | Netherlands | Vessel Type:  | general cargo with container capacity |
| Status:                  | Live        | Reg. Owner:   | Wogenborg Shipping B.V.               |
| Built:                   | 2001        | DWT:          | 9850                                  |
| GE:                      | 6361        | Hull Type:    | Single                                |
| Latest AIS Message Type: | A           | TEU Capacity: | 552                                   |

**Predictive Route Information**  
Last Updated: 08 Feb 2022 [View methodology](#)

ATD: 28 Jan 2022 16:14 Kivnesdal, Norway

Most Probable Destinations:

|                                                |                                              |                                              |
|------------------------------------------------|----------------------------------------------|----------------------------------------------|
| Montreal, Canada<br>80% Predicted by the model | Quebec, Canada<br>77% Predicted by the model | Ottawa, Canada<br>73% Predicted by the model |
|------------------------------------------------|----------------------------------------------|----------------------------------------------|

ETA: 09 Feb 2022 14:17  
Expected Waiting Time: 1 day

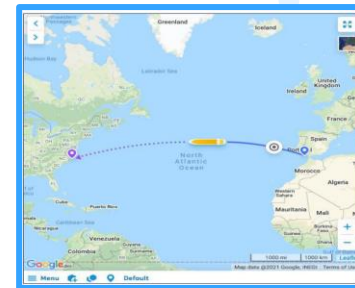
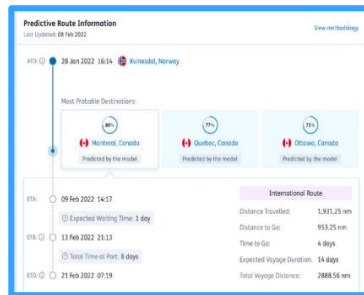
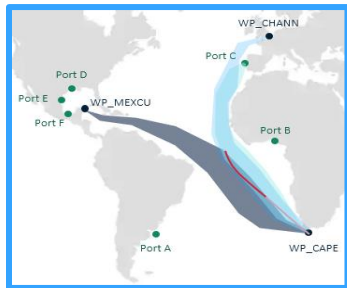
ETB: 13 Feb 2022 21:13  
Total Time at Port: 8 days

ETD: 21 Feb 2022 07:19

**International Route**

|                           |             |
|---------------------------|-------------|
| Distance Travelled:       | 1,931.25 nm |
| Distance to Go:           | 953.25 nm   |
| Time to Go:               | 4 days      |
| Expected Voyage Duration: | 14 days     |
| Total Voyage Distance:    | 2888.56 nm  |

# "Bring compute to the data" in action



## Cloud Database engine

**AWS Redshift**

Database calculations



## Streaming analytics

**SAS ESP scoring containers**

Run algorithms against real-time  
Event streams



VS code editor



Notebooks, RStudio



SAS interface



## In-memory engine

**SAS CAS engine**

Big data analytics eg route  
optimization

# 4. Cloud-Native, Multi-Cloud Platform

*“seizing these opportunities will require investing in new platforms to build ‘data lakes’ and to provide data science services. More and more, platforms will be hosted remotely (‘on the cloud’),”*

*“Full cloud solution does not seem to be the way to go - for now - both for financial, and security/privacy compliance reasons. Statistical organisations will have to consider investing in hybrid infrastructures (on premise + cloud)”*

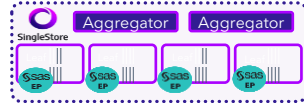
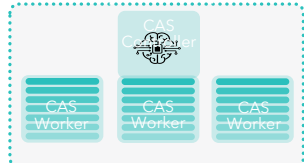
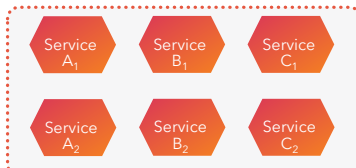
*Trusted Smart Statistics: How new data will change official statistics. Riccinato et al, 2020*

- 1 Open source Integration
- 2 Collaborative Platform
- 3 Hybrid Data Processing
- 4 Cloud-Native Platform
- 5 Openness and Dissemination

# Cloud-Native, Multi-Cloud

Execute AI-based statistics on a wide variety of cloud-native platforms ranging from public clouds to private infrastructure and data centers

Cloud-Native



Public Cloud



Azure



Google



AWS



OpenShift Red Hat

Private Cloud

CI/CD release management



Stable monthly Releases



Long-Term bi-annual(LTS)

# 5. Openness and Dissemination

*“adopting new types of data for official statistics requires (and induces) changes in almost every aspect of the statistical system: processing methodologies ....**communication and dissemination approaches.***

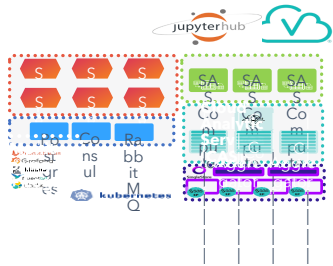
*Trusted Smart Statistics: How new data will change official statistics. Riccinato et al, 2020*

**“Engagement with different types of data users with various profiles and capacities is also required for dissemination. Innovation in data products and services** requires NSOs to establish structures, cultures and ecosystems where useful ideas can be experimented and nurtured to become projects and services.”

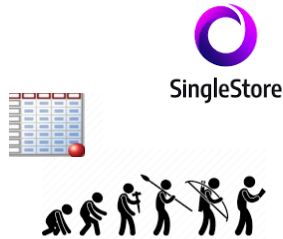
*Which Strategies for NSOs in the Digital Era? Towards ‘Smart Data’ Strategies, OECD, 2018*

- 1 Open source Integration
- 2 Collaborative Platform
- 3 Cloud-Native Platform
- 4 Hybrid Data Processing
- 5 Openness

# SAS Viya provides more means to provide secure access and disseminate statistics



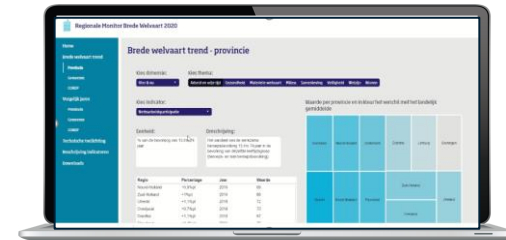
SAS Viya platform  
as a scalable and secure  
Remote access facility



SAS Data Storage  
Evolution



Open-source integration  
allows closer integration  
with Linked Open data  
ecosystems



Interactive visualization  
published on the web

# Data-Driven/Dynamic Infographics



Mapping Crimes - Washington, D.C. • SAS Visual Analytics report showing the crime statistics in Washington, D.C. between 2019 and 2020. Hotspots areas were overlaid on Police Service Areas (PSA). - Source: <https://dcatlas.dcgis.dc.gov/crimecards> • Pack: file:///rdcx200/vaperf/Data/gov/WashingtonCrimes\_+ (106 of 119) • Email

# We are still passionate about 'stats' !!

```
Run Cancel [Icons] Copy to My Snippets + Code to Flow [Icons]
```

Code


```
1 proc deepcausal data=mycas.mydata;
2 id rowId;
3 psmodel t=x1-x20;
4 model y=x1-x20;
5 infer policy=(s1 s0) policyComparison=(base=(t) compare=(s0 s1))
6 out=mycas.oest;
7 run;
```

## Complete Economic Analysis with SAS Econometrics

50+ PROCs, 17 Action Sets, 14 Data Engines

| Time Series Analysis and Data Engines           | Econometric Modeling                                        |
|-------------------------------------------------|-------------------------------------------------------------|
| Autoregressive error models                     | Bayesian econometrics                                       |
| Autoregressive integrated moving-average models | Compound distribution modeling                              |
| Categorical time series analysis                | Copula approach                                             |
| Data engines                                    | Count regression                                            |
| Error correction models                         | Discrete choice analysis                                    |
| Exponential smoothing models                    | Economic capital modeling                                   |
| Hidden Markov models                            | Limited dependent variable modeling                         |
| Linear Gaussian state space models              | Machine learning for causal inference and policy evaluation |
| Long-memory time series                         | Panel data analysis                                         |
| Market attribution models                       | Self-selection models                                       |
| Nonlinear non-Gaussian state space models       | Severity modeling                                           |
| Polynomial distributed lagged models            | Spatial econometric modeling                                |
| Sequential Monte Carlo methods                  | Stochastic frontier analysis                                |
| Unobserved component models                     | Systems modeling and simulation                             |
| Vector autoregressive moving-average models     |                                                             |
| Volatility forecasting, GARCH models            |                                                             |

Copyright © SAS Institute Inc. All rights reserved.



Causal inference and policy evaluation with deep neural networks

<https://blogs.sas.com/content/subconsciousmusings/2021/09/07/causal-inference-and-policy-evaluation-with-deep-neural-networks/>