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

2019
Attended a seminar in Brussels organized by the VSA on Smart Statistics and new techniques in Official Statistics.

2004
Started working for SAS
Banking and Insurance Consultant

2007
Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents"

2001
Graduated at the Department of Sociology - Survey Methodology

2001-2004
Research assistant on Data Quality issues in the European Social Survey

Concept of Smart (AI-based) Statistics

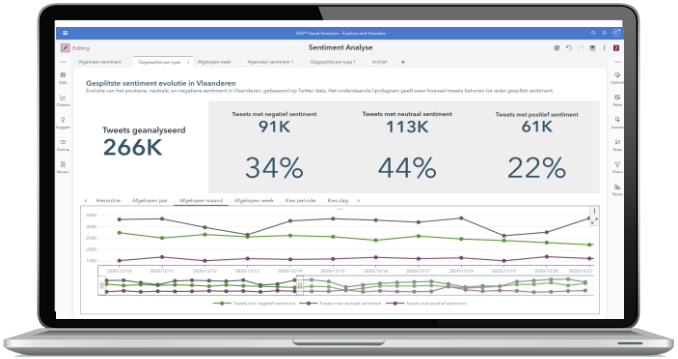
NEW DATA SOURCES

Michel Philippens
@sbxmip
Proud father of Klaas, Boris en Stan
SAS evangelist, 43, Mechelen

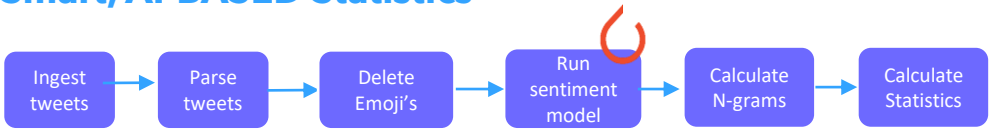
These are the current symptoms after my first dose of Moderna vaccine: extreme fatigue. Soreness at the injection site. Swelling at the injection site. Otherwise, I am fine.

7:22 AM – 22 Feb 2021

TIMELY, CHEAPER, GRANULAR, RELEVANT INSIGHTS



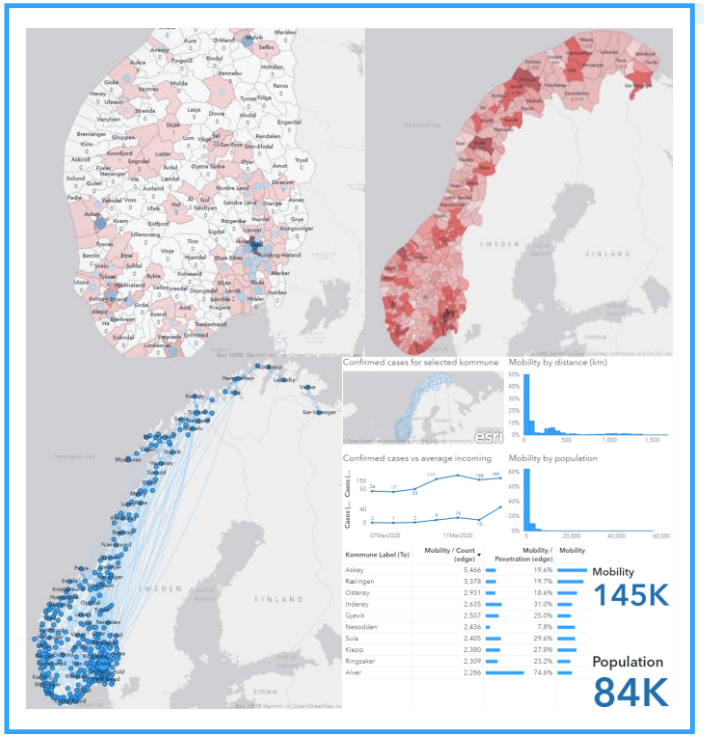
Smart, AI-BASED Statistics



Fuel for Policy Evidence

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

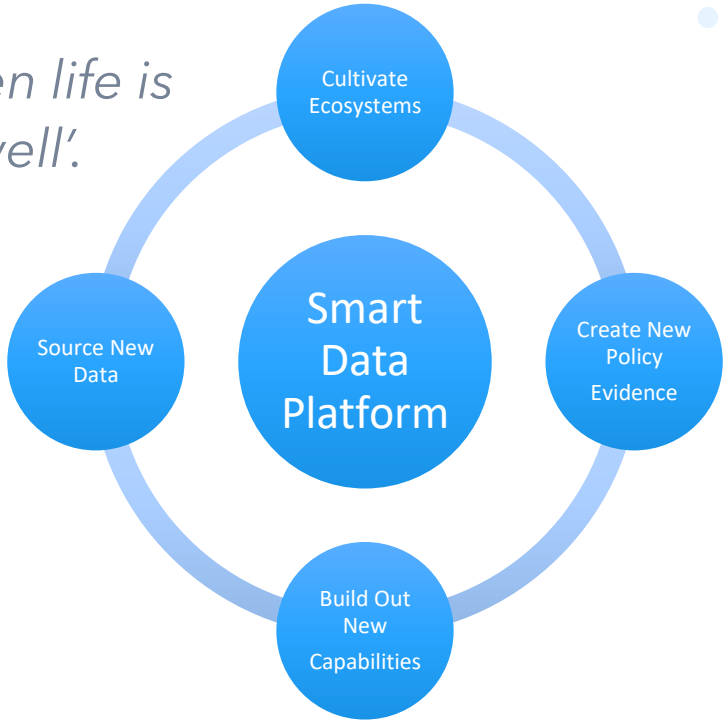
KPIs dashboard and visualization reports allow decision makers to have accurate insights about the pattern of the spread within locations and eventually take social measures to control the transmission rate



NSO's are on a Transformational journey

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

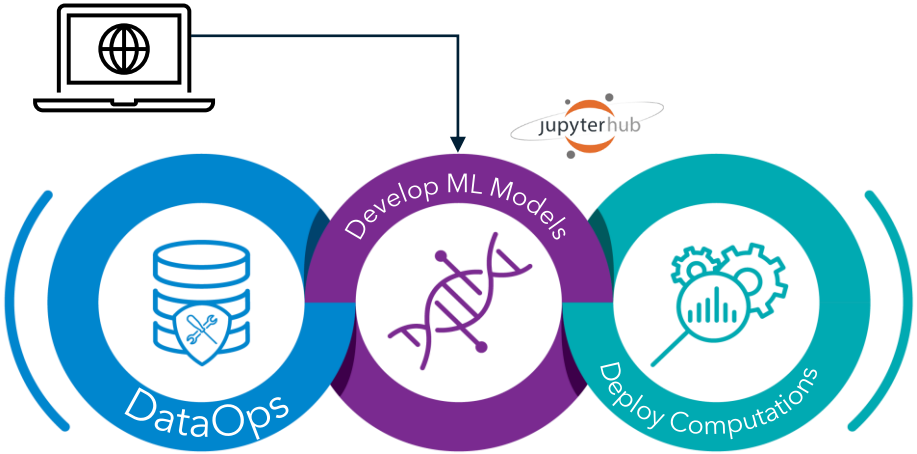
Mariana Kotzeva - Director-General of Eurostat.



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

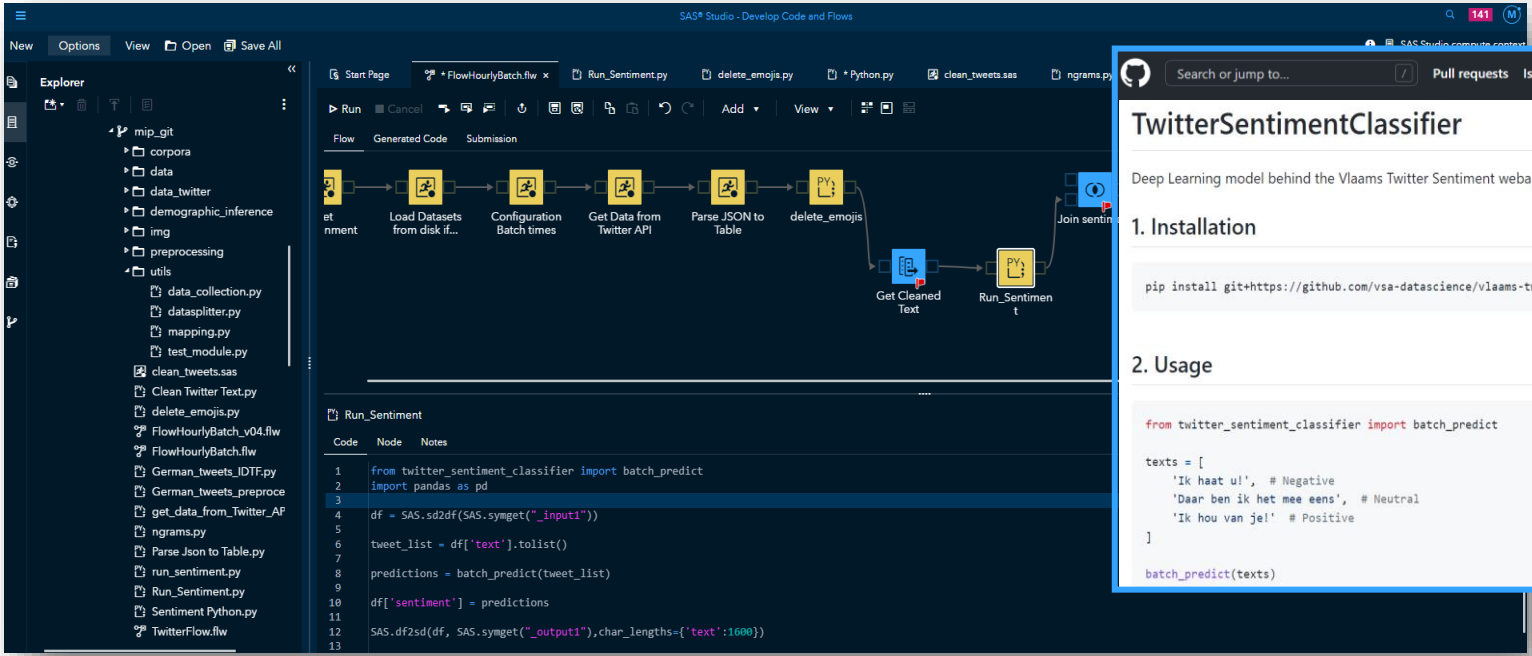
"Very successful open-source projects provide free data science tools, plus free algorithms and access to large expert communities – they represent a huge opportunity for statistical organisations, provided they invest in them by making them key components of their next generation platform"

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

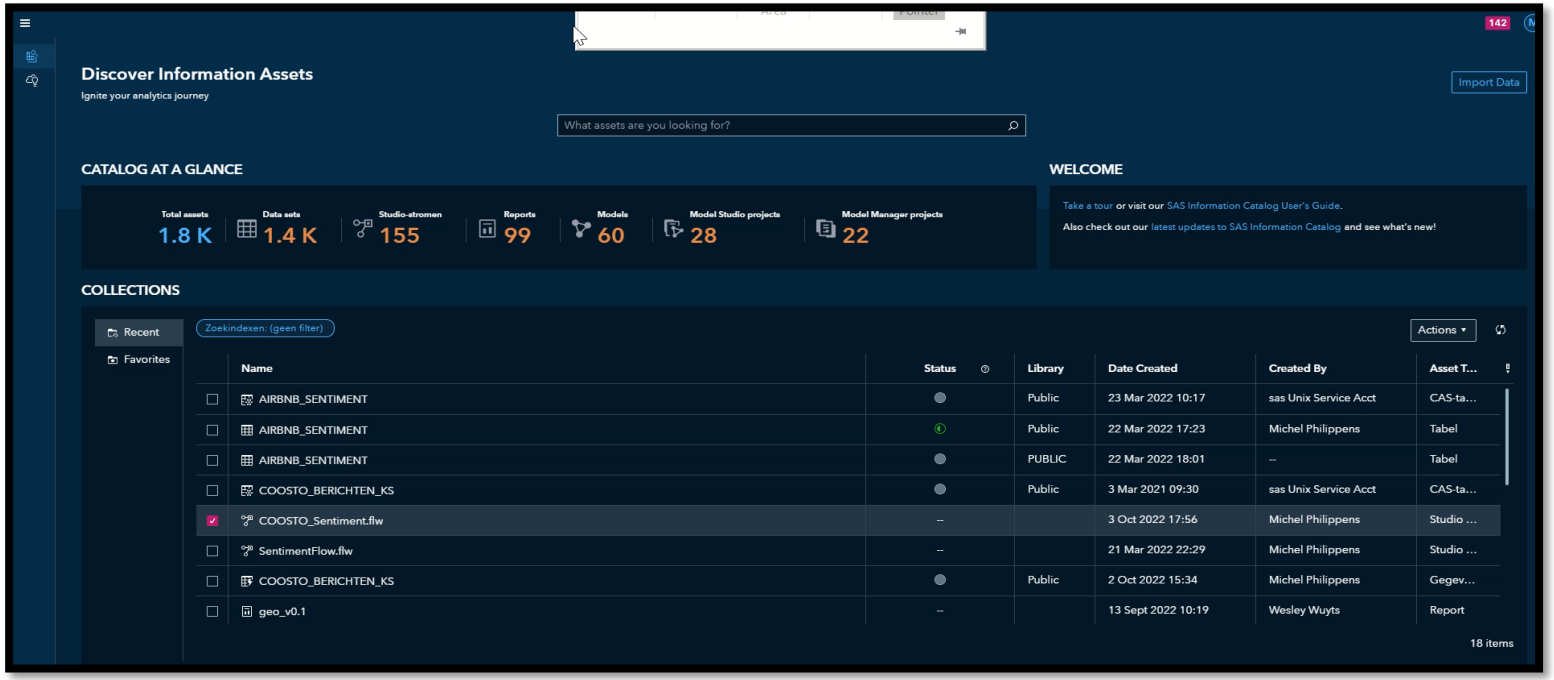
SAS opens up its interfaces to open source

Ingest tweets → Parse & Cleanse tweets → Delete Emoji's → Run sentiment model → Calculate N-grams → Calculate Statistics



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.



2. Collaborative Platform

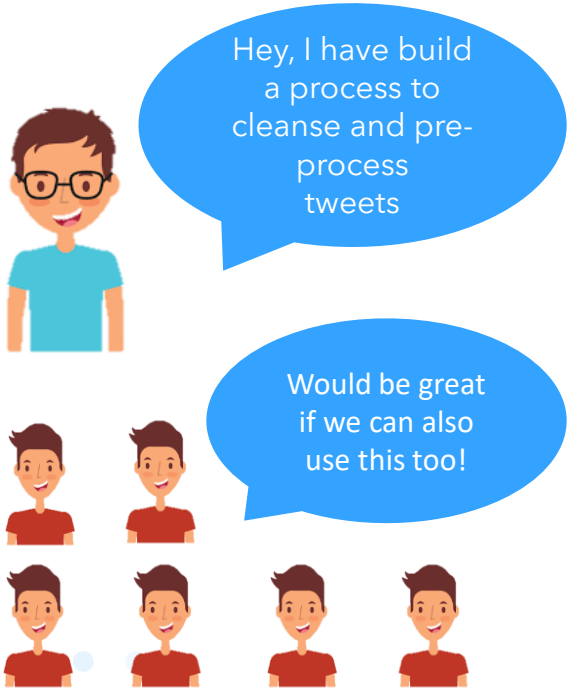
"The shortage of data science expertise is a widely known challenge. Statistical operators should become more of data scientists, picking up tasks in activities such as data sourcing, data modelling, data exploration, data analytics. This gap calls for a data skills strategy and a review of existing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

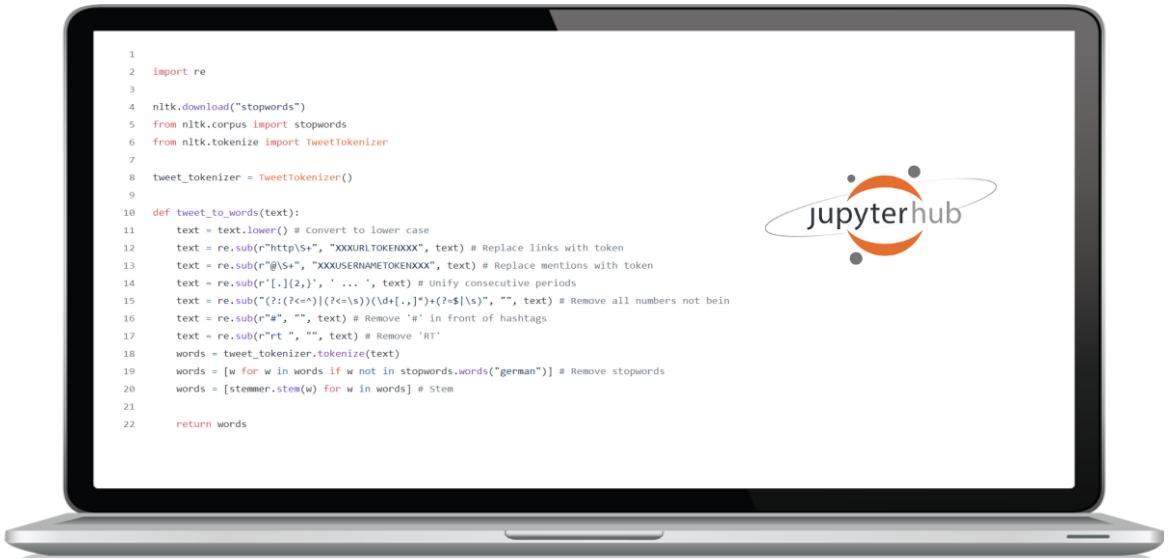
Which Strategies for NSOs in the Digital Era? Towards 'Smart Data' Strategies, OECD, 2018

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

Collaborative Platform

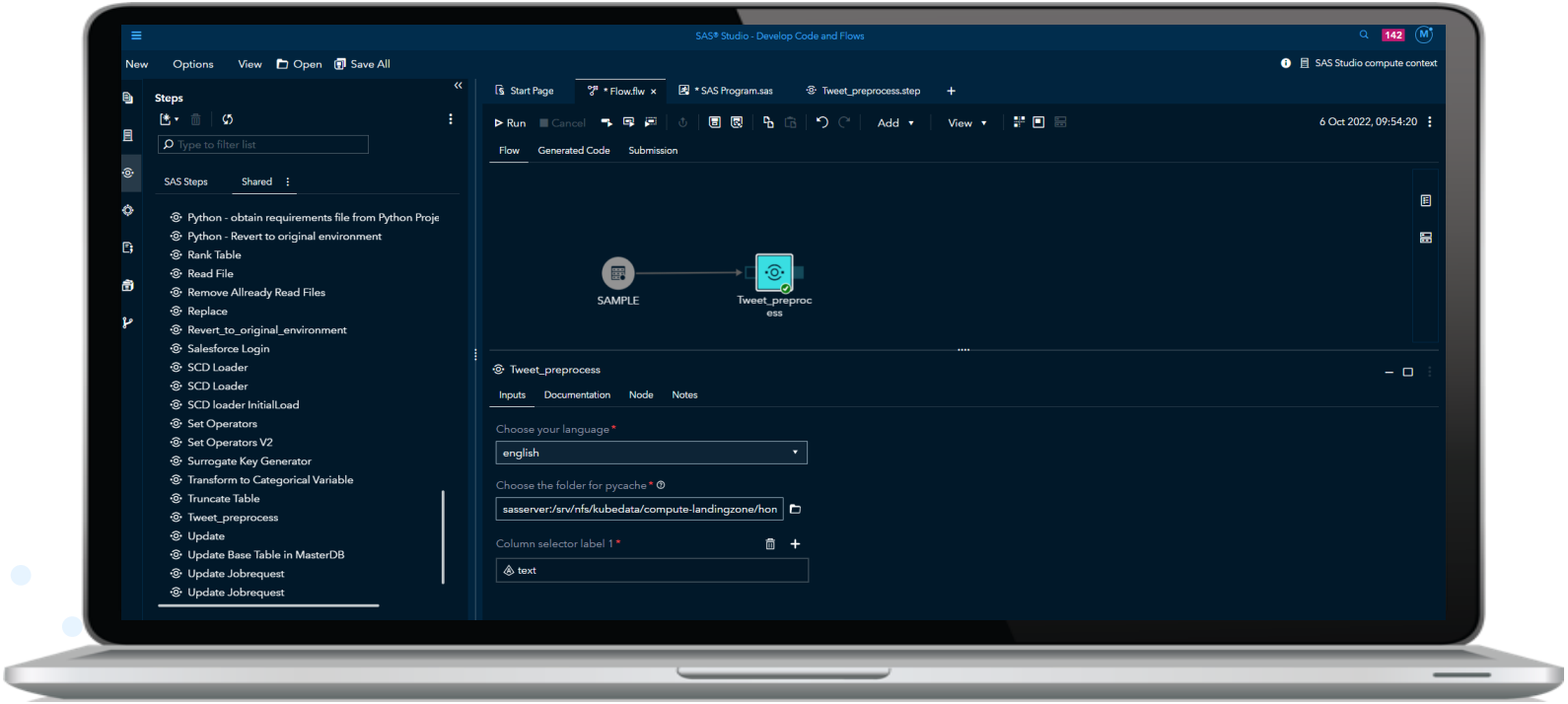
SAS Viya implements multiple integration points that bring SAS and Open-Source users together.





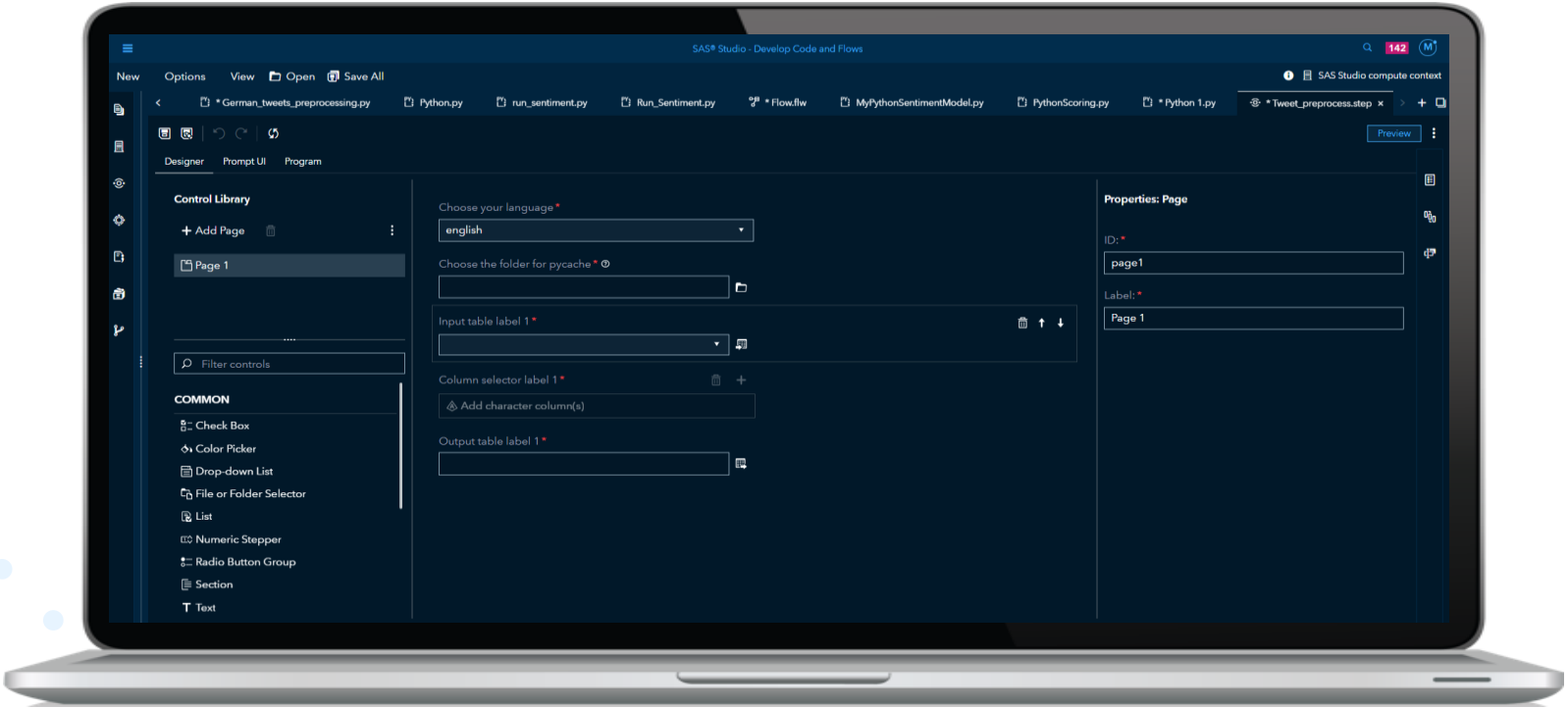
“Custom Steps” as reusable code blocks

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

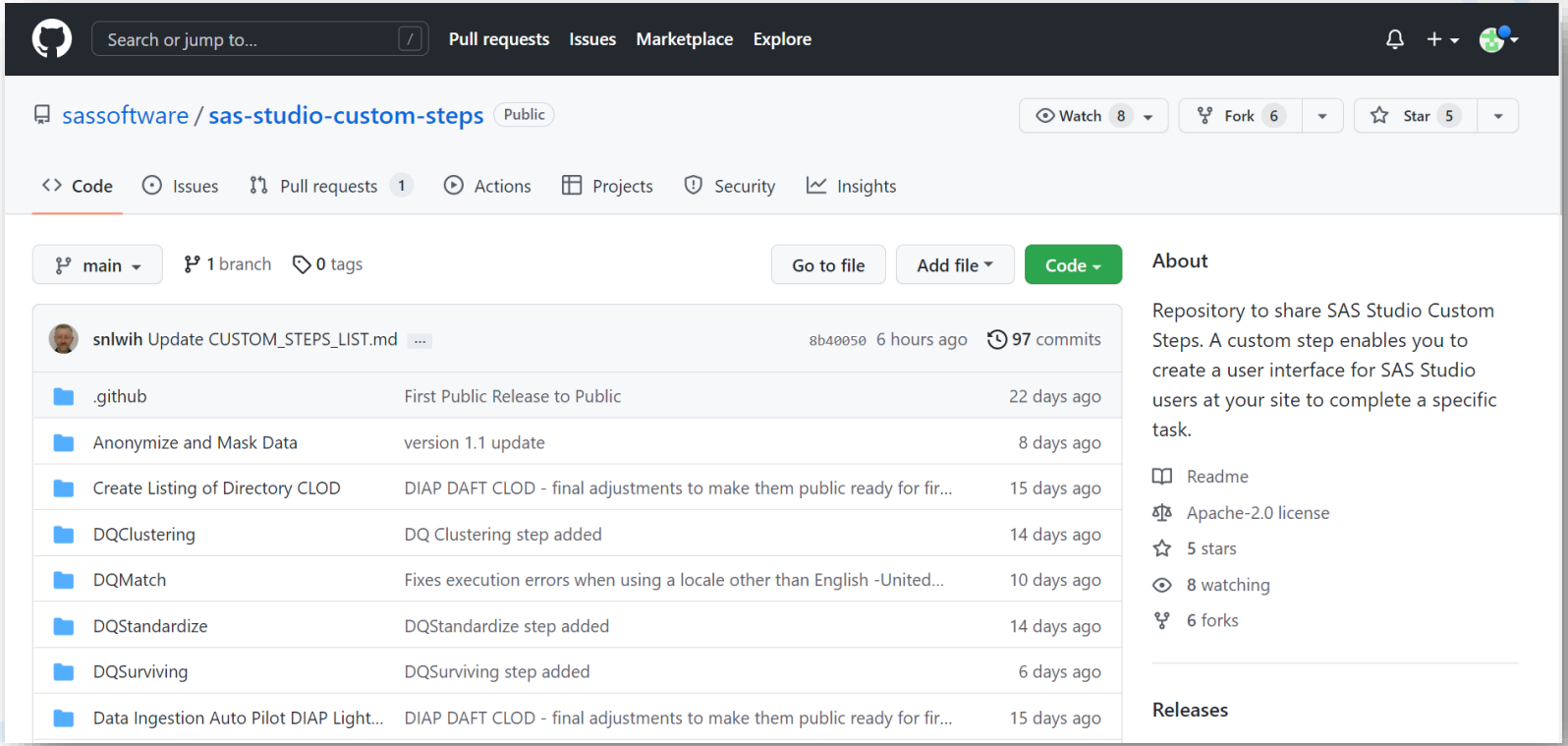


A low-code “Custom Step” designer

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



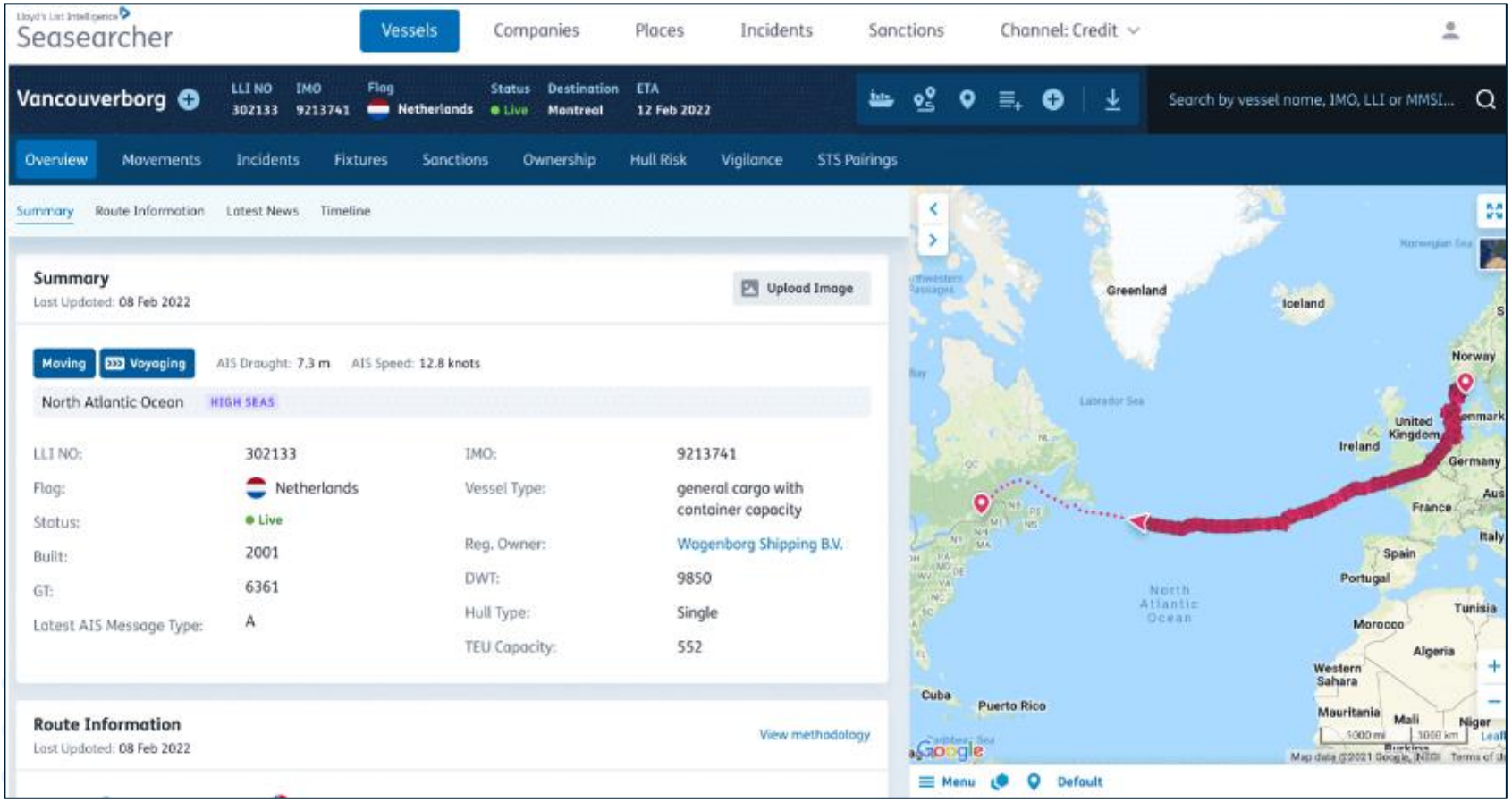
3. “Data processing everywhere”

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

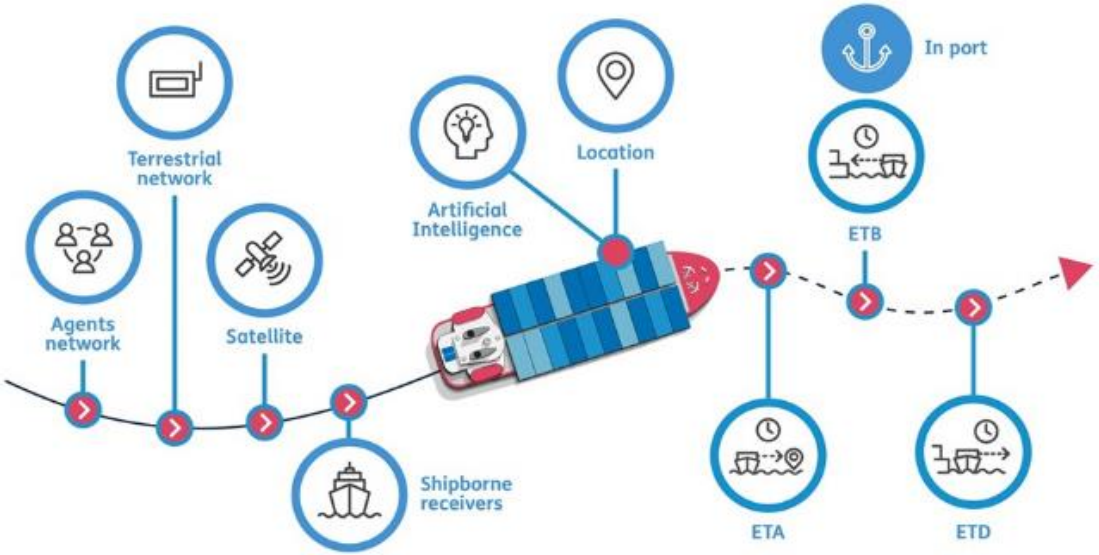
“moving part of the computation to the data sources often (but not always) allows to minimize the global requirement for technical resources. The rate and volume of new digital data is often huge, but only a tiny part of the information embedded in the data is relevant for official statistics”

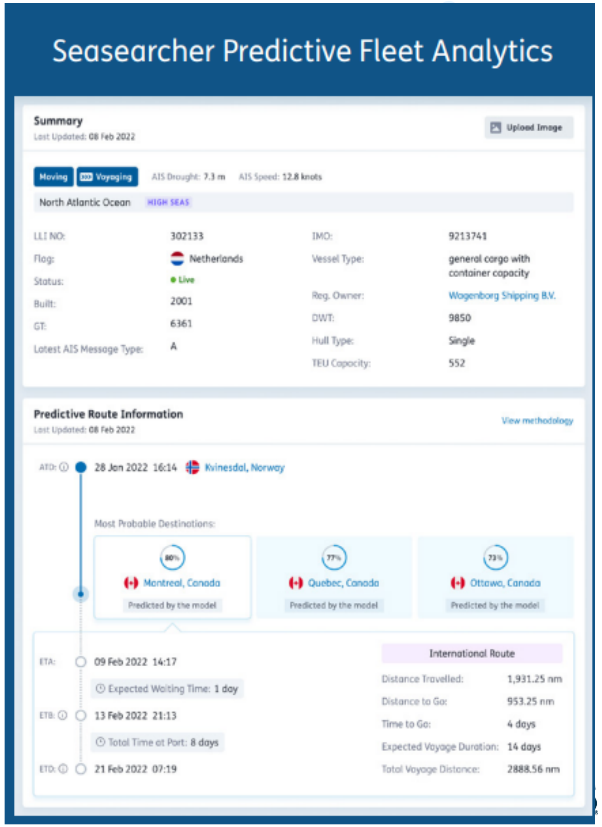
Trusted Smart Statistics: How new data will change official statistics. Ricciato et al, 2020

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

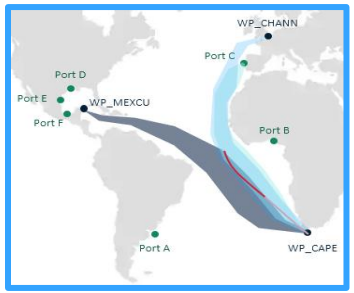


Lloyds' list intelligence: Predicitive Fleet Analytics





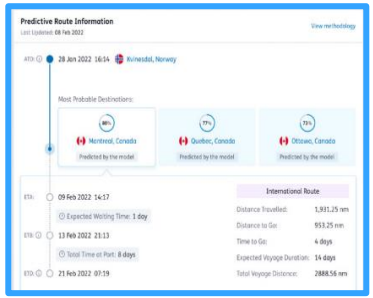
“Bring compute to the data” in action



Cloud Database engine

AWS Redshift

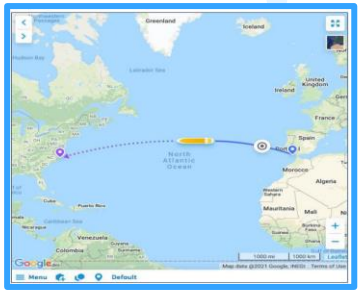
Database calculations



Streaming analytics

SAS ESP scoring containers

Run algorithms against real-time Event streams



In-memory engine

SAS CAS engine

Big data analytics eg route optimization

VS code editor

Notebooks, RStudio

SAS interface

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'),"

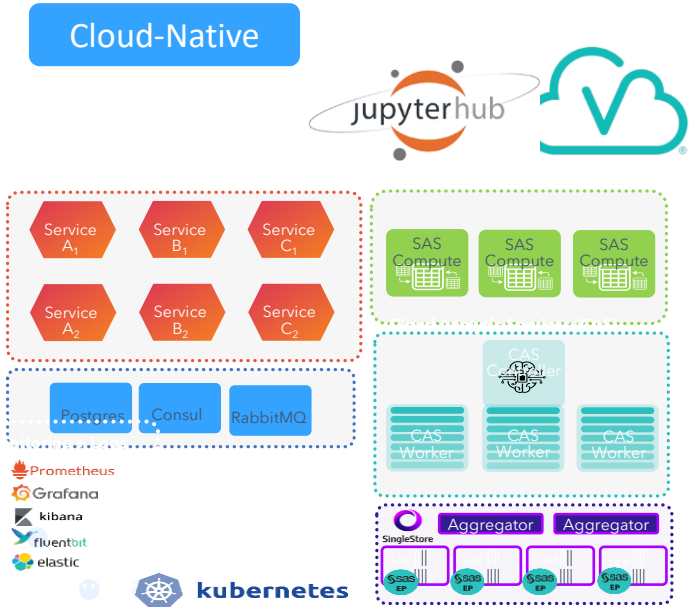
"Full cloud solution does not seem to be the way to go – for now – both for financial, and security/privacy compliance reasons. Statistical organisations will have to consider investing in hybrid infrastructures (on premise + cloud)"

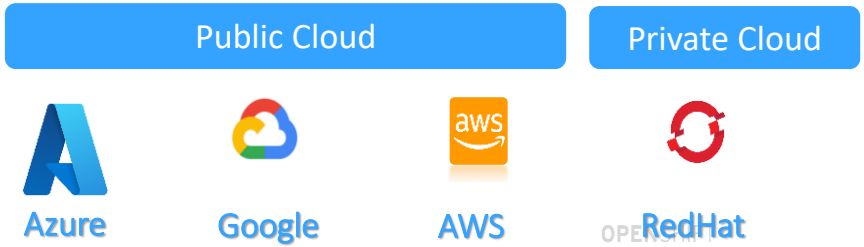
Trusted Smart Statistics: How new data will change official statistics. Riccinato et al, 2020

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





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.***

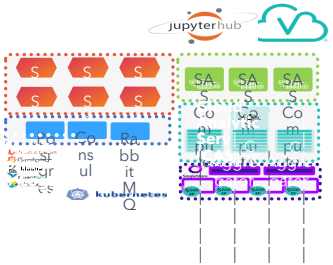
Trusted Smart Statistics: How new data will change official statistics. Riccinato et al, 2020

*"**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

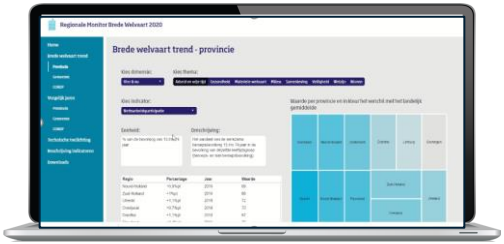
SAS Viya provides more means to provide secure access and disseminate statistics



SAS Viya platform
as a scaleable and secure
Remote access facility

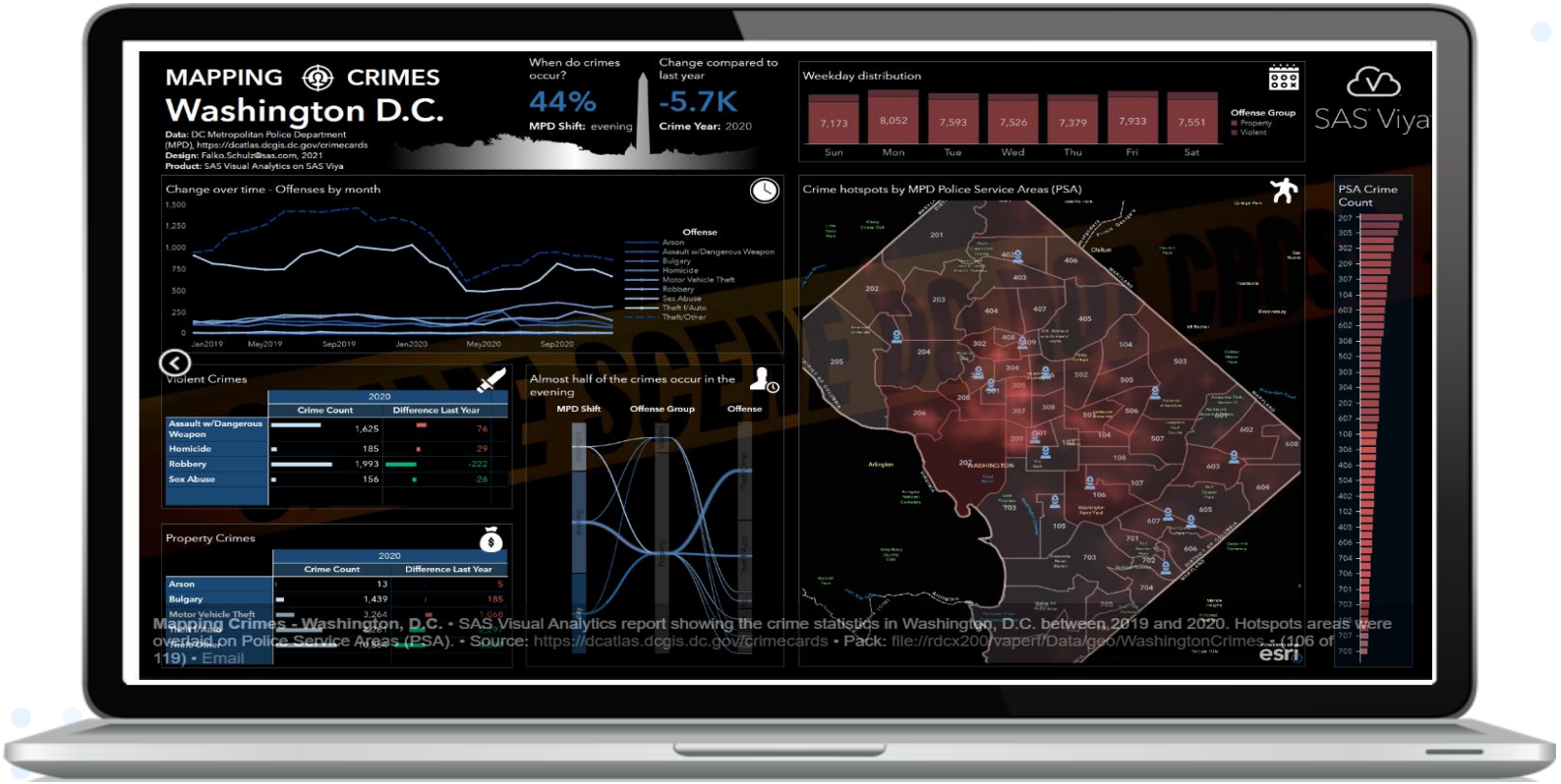
SAS Data Storage
Evolution

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



Interactive visualization
published on the web

Data-Driven/Dynamic Infographics

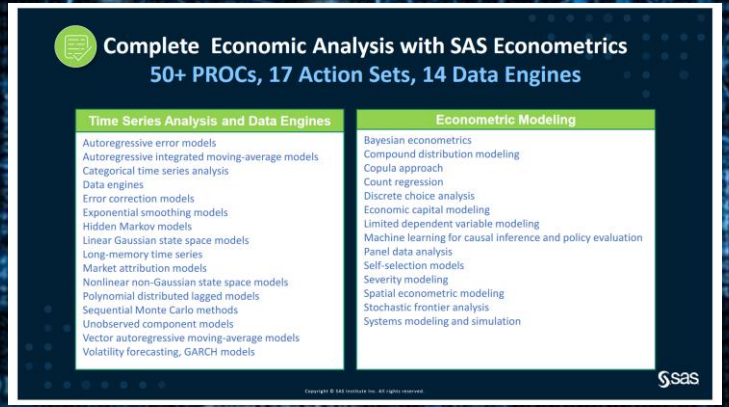


We are still passionate about ‘stats’ !!

Run | Cancel | Copy to My Snippets | + Code to Flow

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;
```



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/