



**World Customs
Organization**

Big Data and Its Application to Cross-border Paperless Trade

Joel Choi

Data Analytics Specialist

**5th UNESCAP UNNExT
Masterclass**

28/04/2022

What is Big Data?

- Gartner proposed the three Vs in 2001
 - Volume – How much data is there?
 - Velocity – At what speed is new data generated?
 - Variety – How diverse are different types of data?
- Big Data is a popular culture term that exceeds the capacity or capability of traditional or current methods and systems
- A composite term referring to three domains:



Data



System

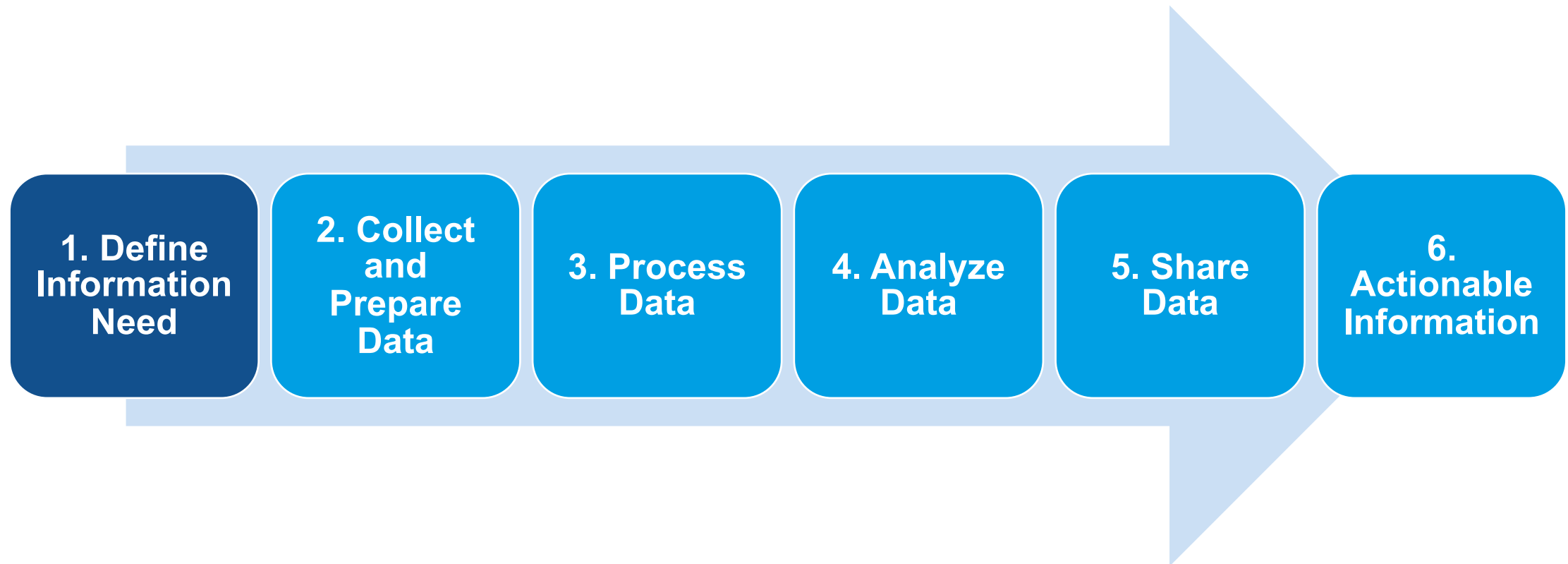


Analysis

Data Analytics Pipeline



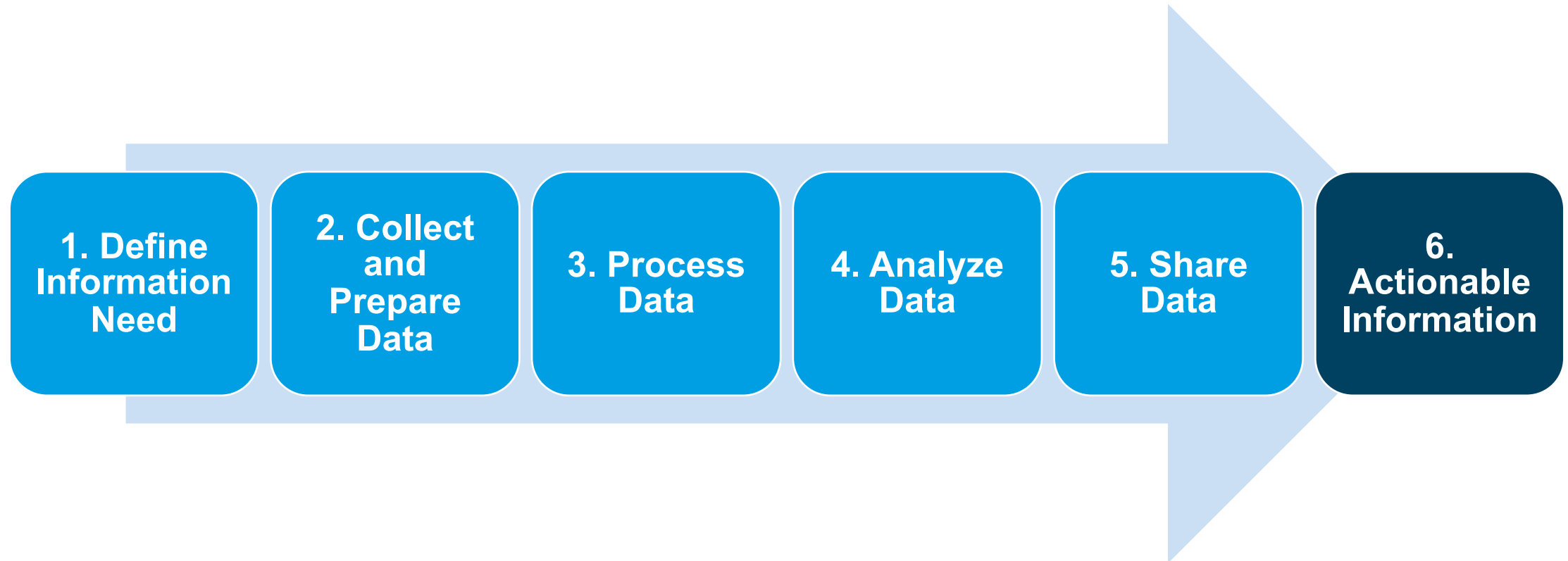
- Analysis in Big Data brings appropriate data and technology elements together to fulfil information needs



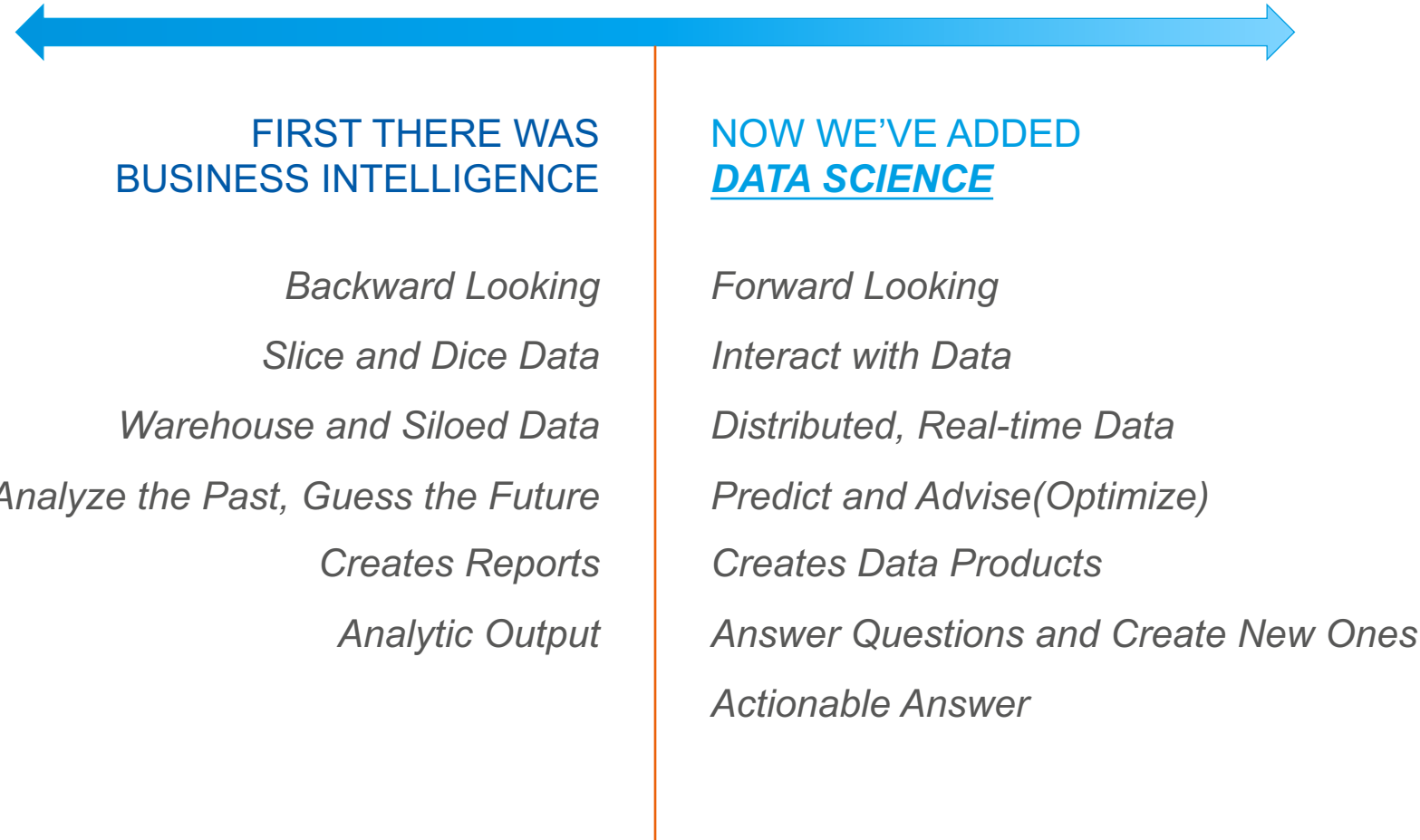
Big Data Analytics



- Final Goal is to turn collected Data from information need into **Actions**



Looking Backward and Forward



Facilitating Trade and Travel While Ensuring Security and Safety



FACILITATION

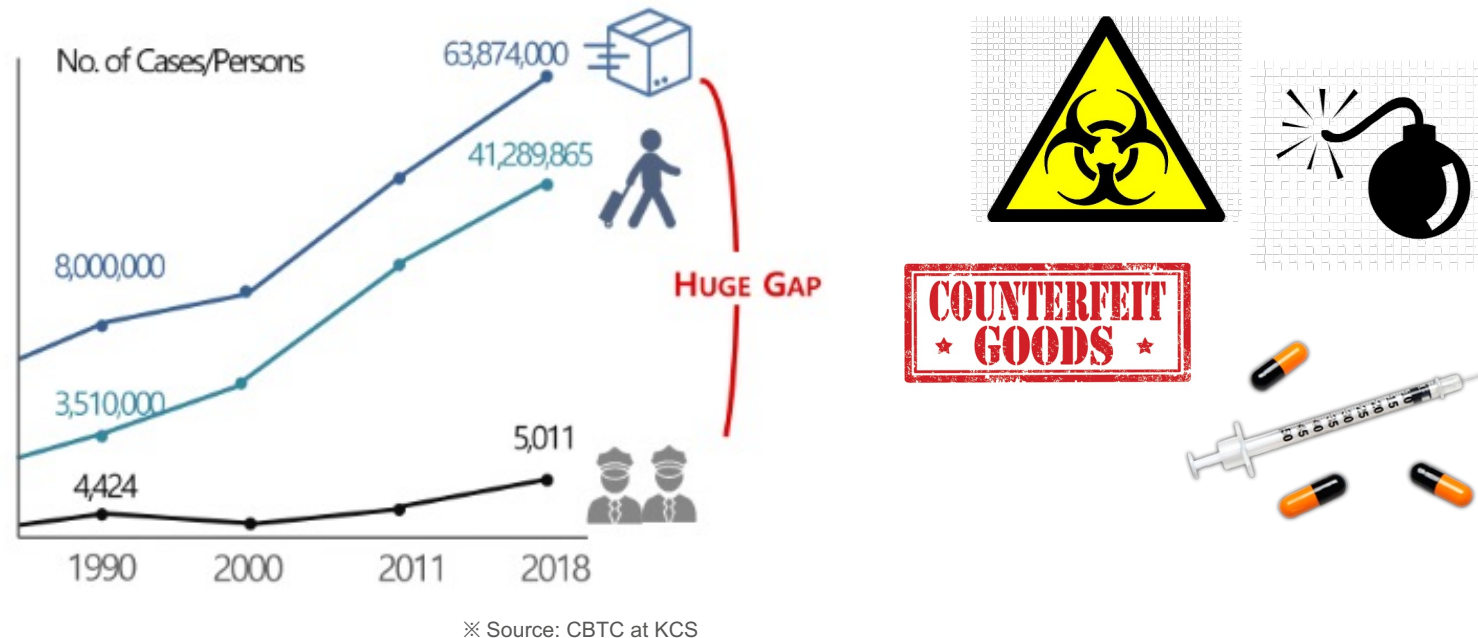
- Simplify Procedures
- Reduce Cost of Clearance
- Support Economic Growth

SECURITY & SAFETY

- Reduce non-compliance
- Protect national Interests
- Ensure Supply Chain Security

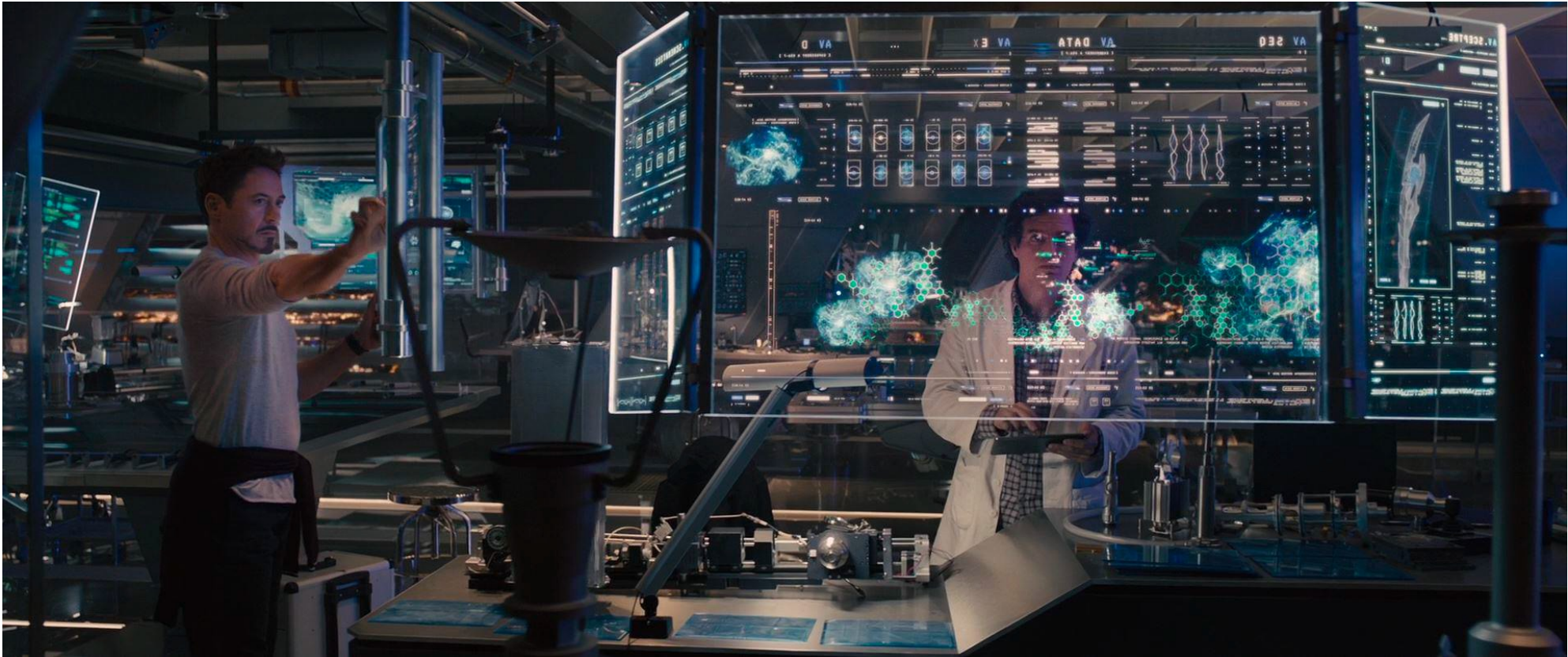
Common Difficulties in most Customs

- Gap between limited resources and surging international trade/travel
- Threats of the transnational movement of high-risk consignments/cargos



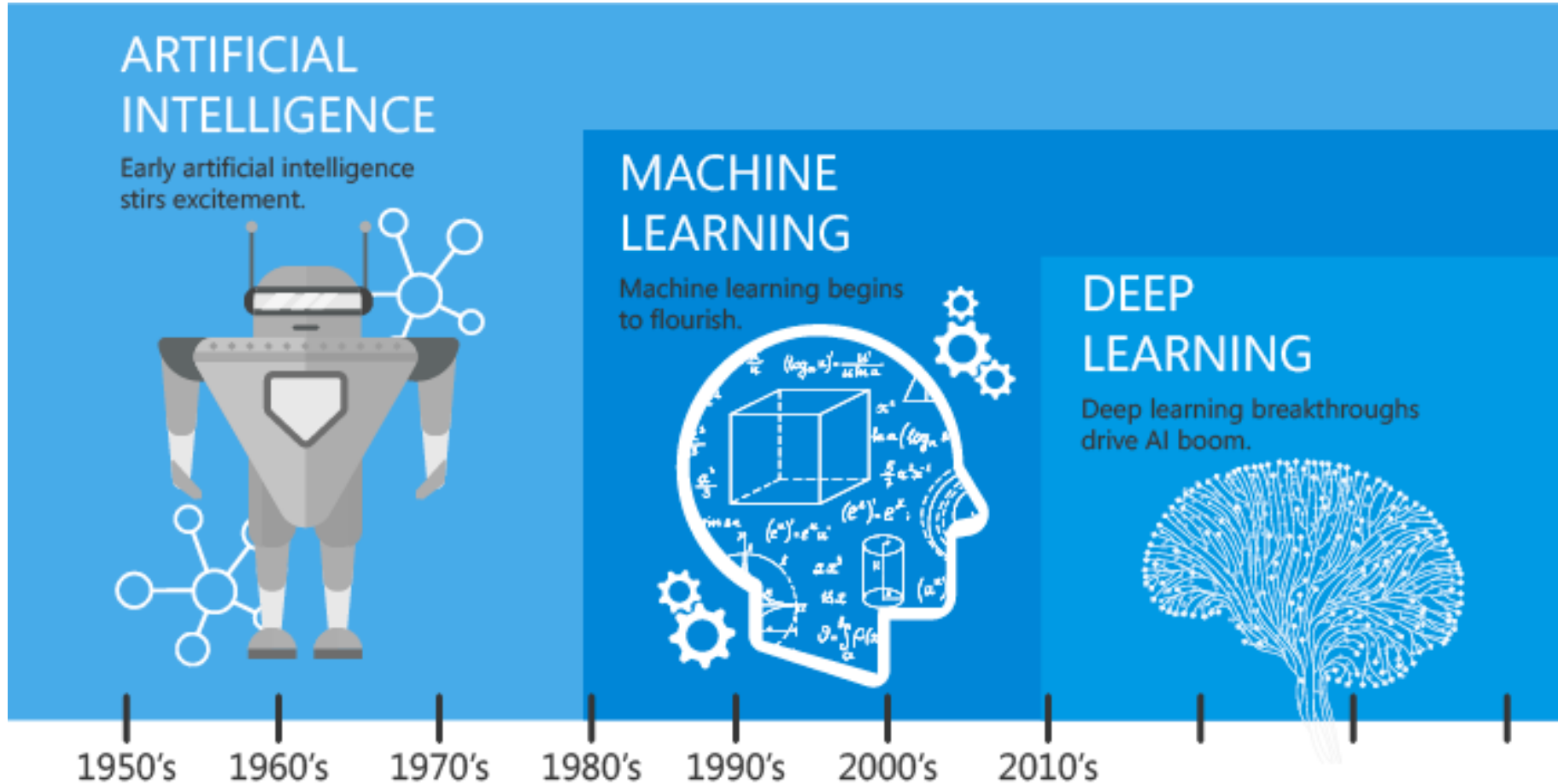
The proliferation of data inevitably leads to a proliferation of analytic projects.

Human-Computer Cooperation

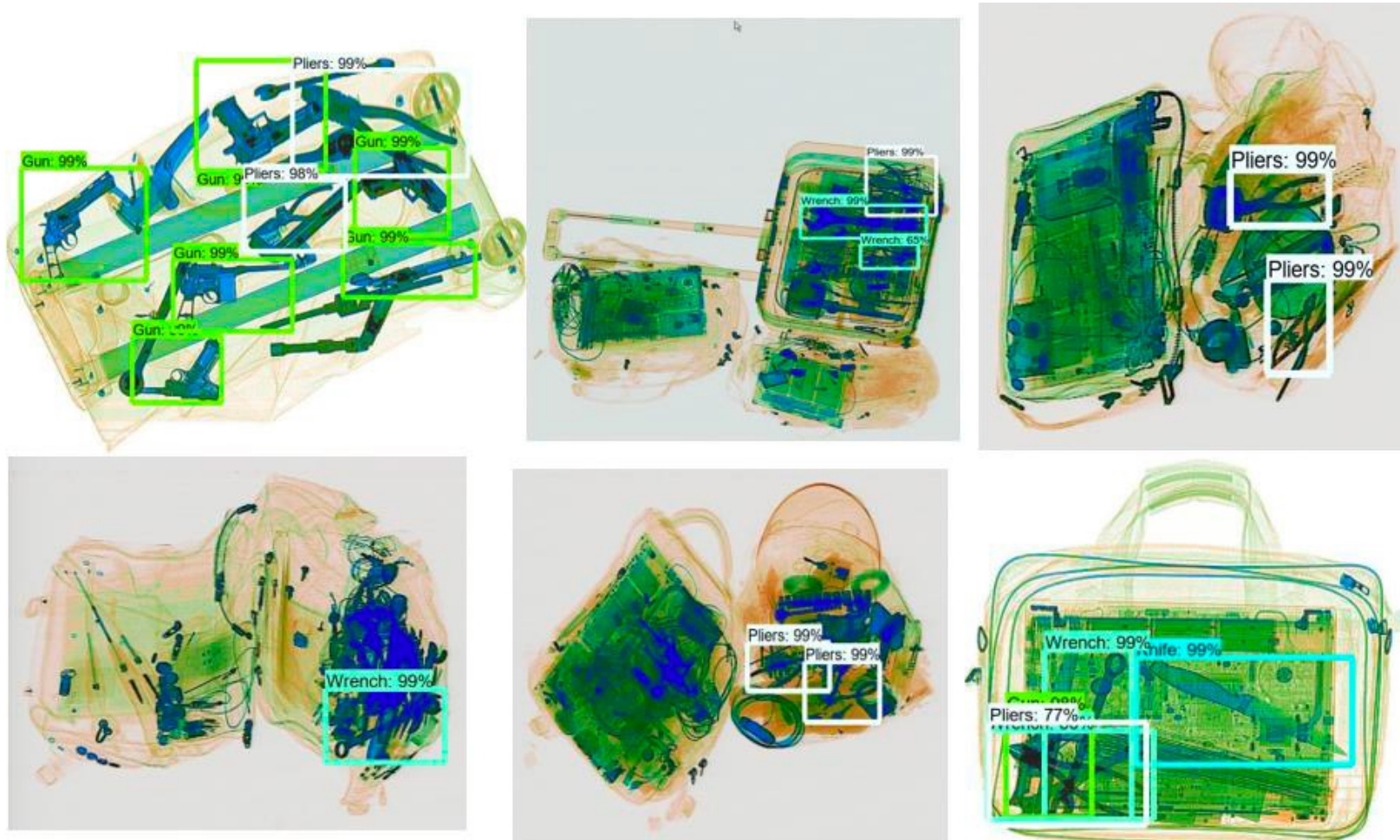


Artificial Intelligence (A.I)

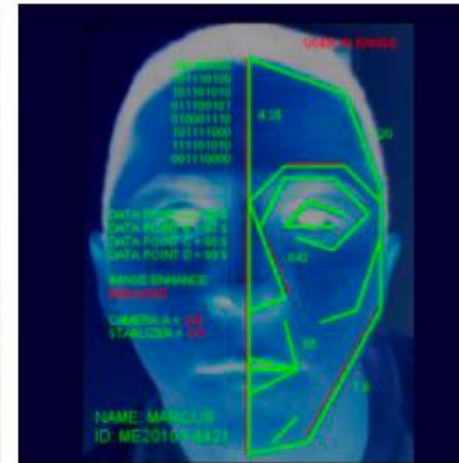
Artificial Intelligence Approach



A.I. X-Ray based Baggage Inspection



Automated Virtual Agent for Truth Assessments in Real-Time (AVATAR)



Images courtesy of U.S. Customs and Border(CBP)/University of Arizona

<https://www.dailymail.co.uk/sciencetech/article-2188833/U-S-Border-control-hires-virtual-Elvis-screen-immigrants-hell-know-youre-lying.html>

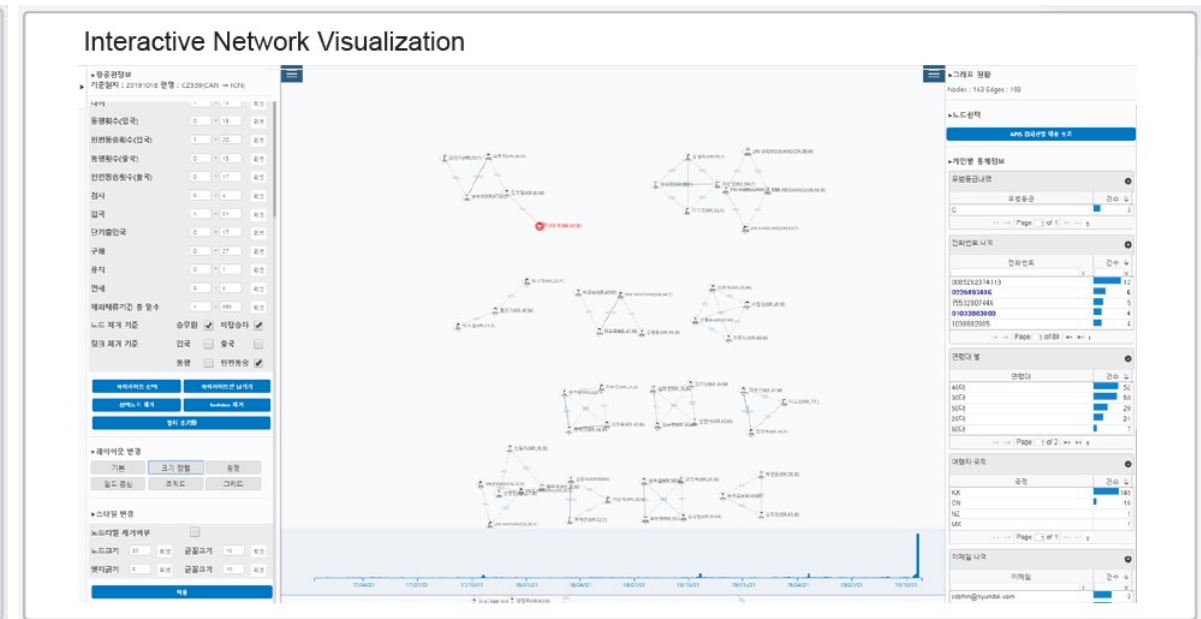
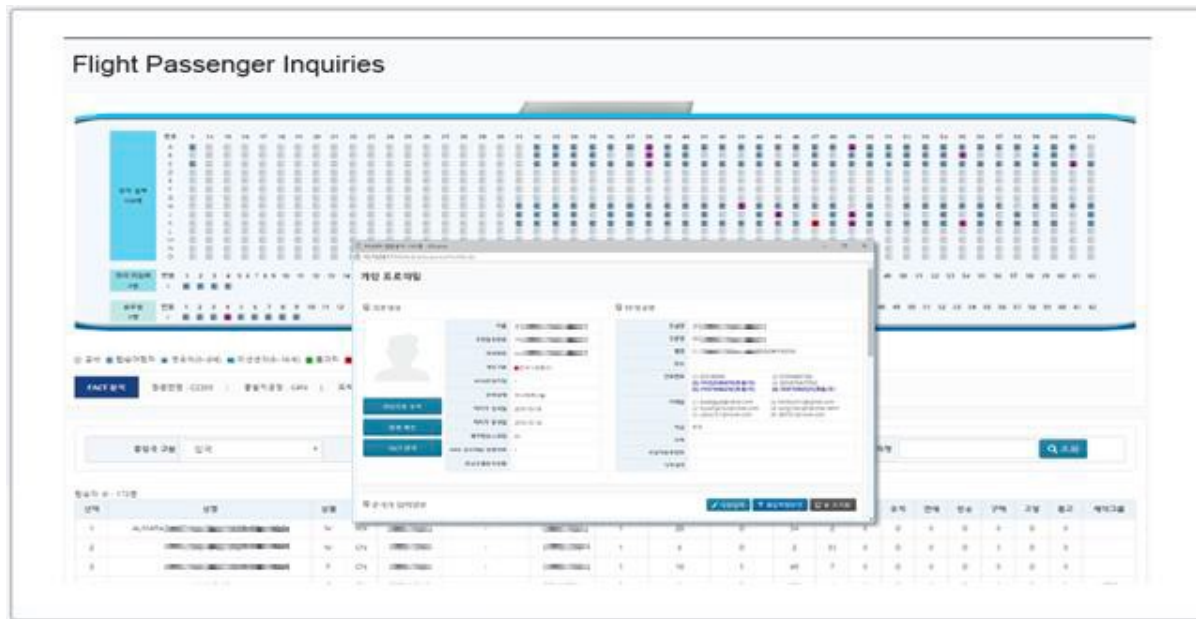
https://ec.europa.eu/research/infocentre/article_en.cfm?artid=49726

Factual Information Network for Data Exploration and Reporting (FINDER)



Network Analysis for high-risk Passenger Selectivity

- Input: past/current PNR/API and investigation results
- Out: network visualization + risk profile (criminal intelligence)
- Objective: to identify hidden-relationship with known smuggler



Big Data Analytics for Customs



- Imagine if you could
 - Create a holistic view of everything involved in illicit trade;
 - Identify and trace suspicious persons and goods;
 - Predict which companies will be non-compliant;
 - Pinpoint the specific consignment through automated verifications;
 - Identify the behavioral patterns of travelers;
 - Seamlessly exchange information about traders with other border agencies;
 - Detect international networks of organized crime operating at your ports.

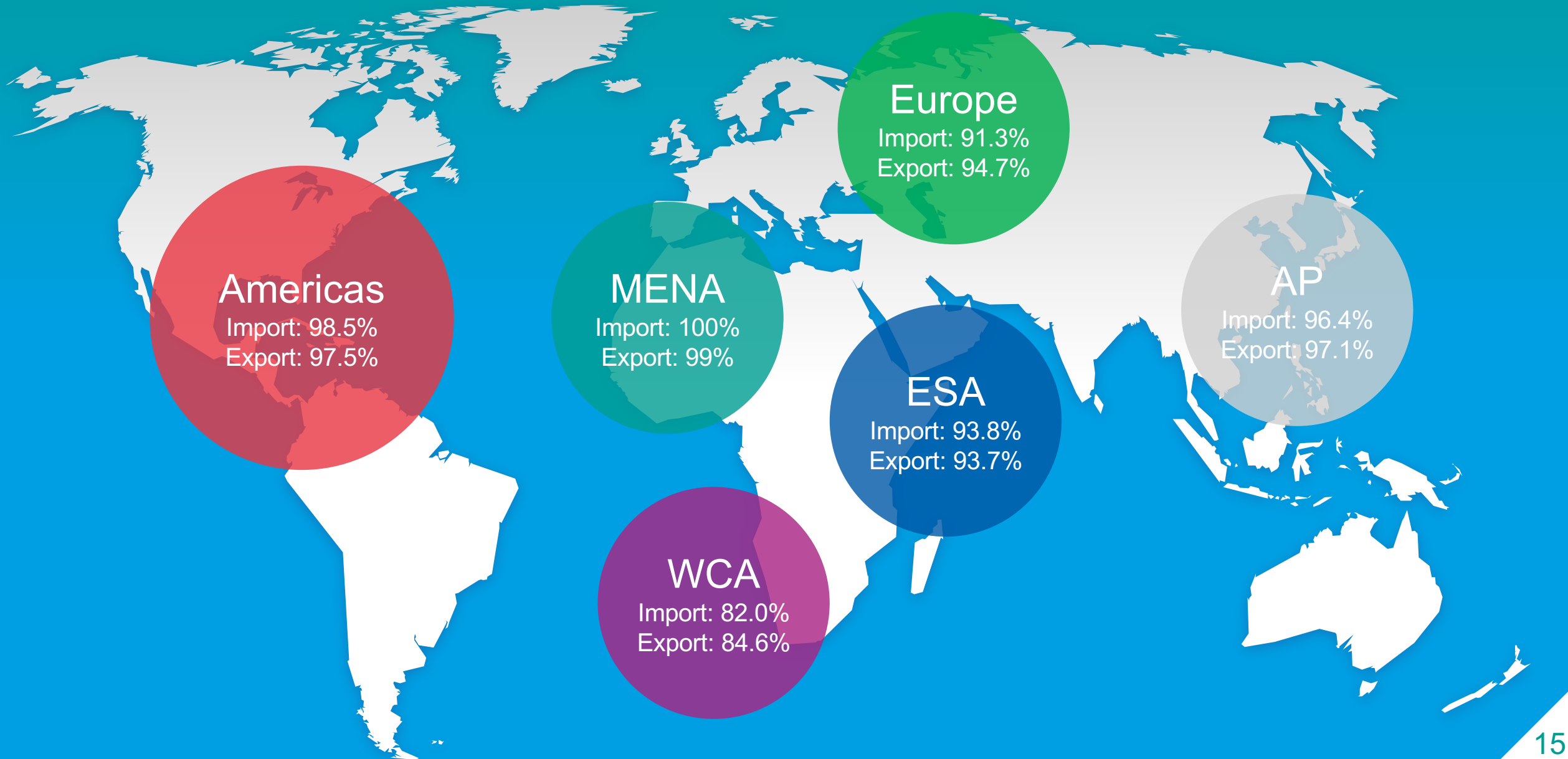


BACUDA Project

Capacity Building Directorate
World Customs Organization



% of electronic declarations



90%

of worldwide import

92%

of worldwide export



Ba · cu · da

[bakuða] noun

An acronym, which stands for

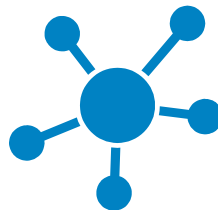
“BAnd of CUstoms Data Analysts.”

It is also a Korean word that means “to change”.

▶ BACUDA Project



WCO Data Analytics
Project launched in 2019



Project part of Capacity
Building Directorate



Project supported by
CCF-Korea



BACUDA Project

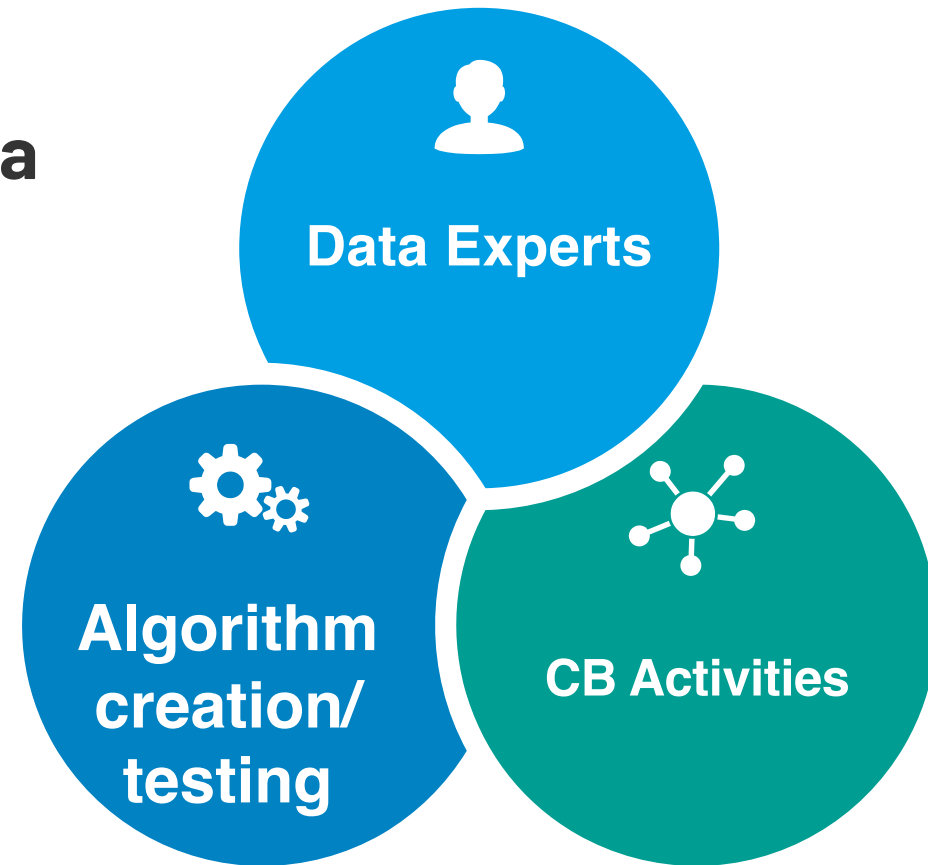
Objective: Support Members develop capabilities to take advantage of data and gain insights to make data-driven decisions

- Raise awareness
- Provide capacity building
- Establish and strengthen cooperation



► BACUDA Project

Implementation of the objective via three main pillars:



Achievements

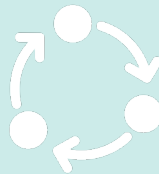
Capacity Building Data Analytics Framework

- Including Regional Events



Practical Algorithms for Customs

DATE Algorithm, AI HS



Online Courses/Offline Workshops

Data Analytics
Beginner/Intermediate/Advanced course,
Scholarship Programme

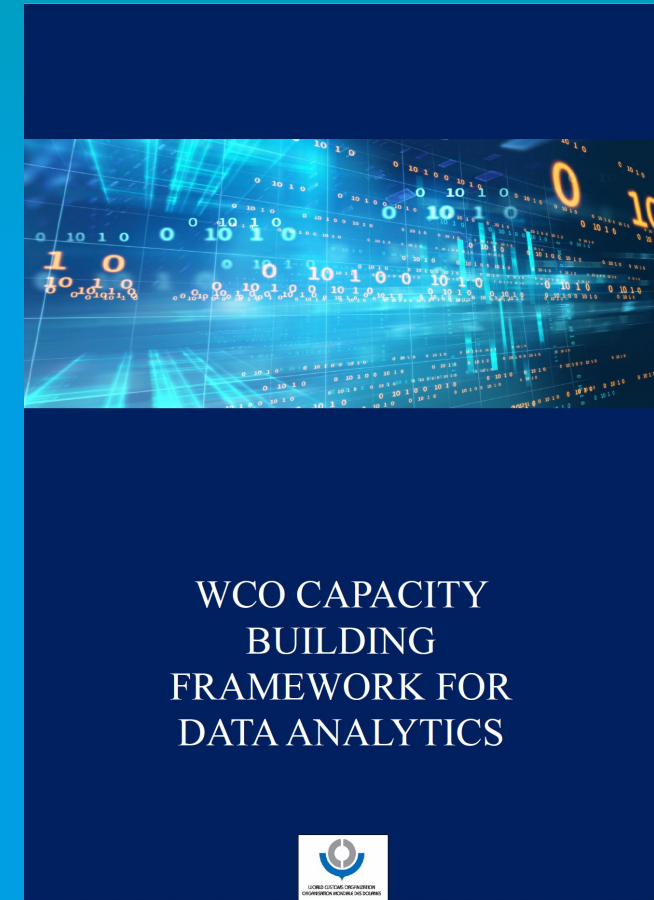


Data Analytics Framework




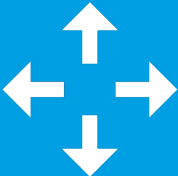

Purpose: Guide members towards the **adoption and optimization of the use of data analytics** at both strategic & operational levels

Endorsed & Approved: At the 229/230 sessions of PTC 2020 and at the 12th CBC 2021

Presented: To the PC and the Council in June 2020 and approved



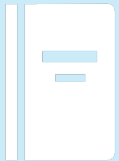
Data Analytics Framework

Assessment & Planning 	Organizational Structure 	Data & Analytics Processes 
Data Governance 	Behaviours & Actions of Leaders 	Skillset & Type of Staff 
Change Management 	Definitions & Examples 	Questionnaire 

Achievements

Capacity Building Data Analytics Framework

Including Regional Events



Practical Algorithms for Customs

DATE Algorithm, AI HS



Online Courses/Offline Workshops

Data Analytics
Beginner/Intermediate/Advanced course,
Scholarship Programme



DATE Algorithm

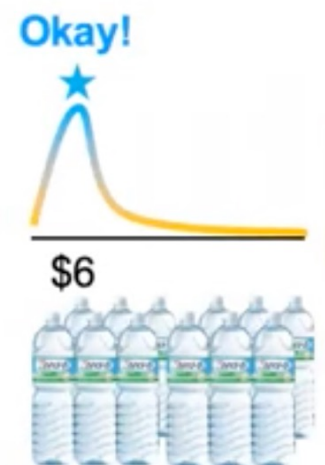
Problem: Fraudulent transactions are found manually and without proper strategy

Goal: Create AI model that detects fraud

Method: Uses “Attention” (AI method)

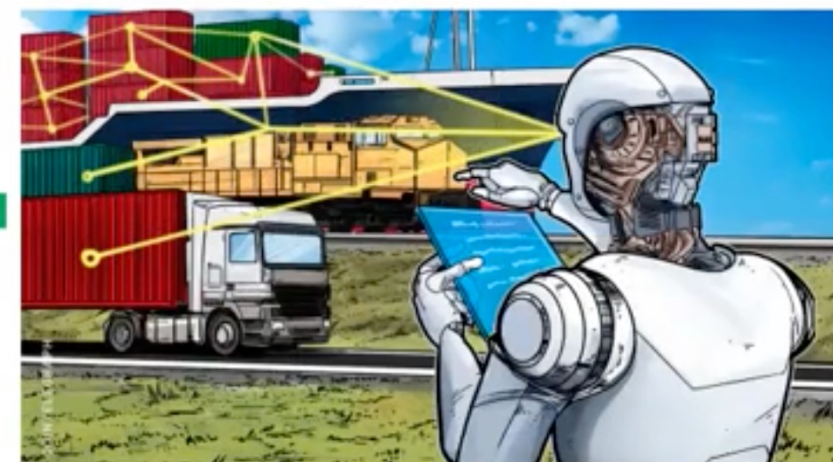


DATE Algorithm



Features			Prediction	
Item	Declared Price	Importer	Predicted Illicit prob	Ranking (Priority)
Ferrari 488 Spider	\$50,000	John	0.99	1
Porsche 911	\$30,000	Jane	0.98	2
...
Hyundai Elantra	\$15,000	Kim	0.01	1000
...
Water 2Lx12	\$6	Wang	0.00	10000

Goal: Select 10% of items that requires inspection



DATE Algorithm

Features

Prediction

Inspection results and labels obtained

Item	Declared Price	Importer	Predicted Illicit prob	Ranking (Priority)	Inspection Results	Illicit	Revenue
Ferrari 488 Spider	\$50,000	John	0.99	1	Under-invoiced, Original price: \$350,000, Rate applied: 30%, Additional duties: \$90,000	1	\$90,000
Porsche 911	\$30,000	Jane	0.98	2	Original price: \$130,000, 30% rate -> Surtax applied for \$100k	1	\$30,000

Hyundai Elantra	\$15,000	Kim	0.01	1000	CLEARED – Right price for the used car, Odometer: 25300km,	0	\$0

Water 2Lx12	\$6	Wang	0.00	10000	CLEARED	0	\$0

AI HS

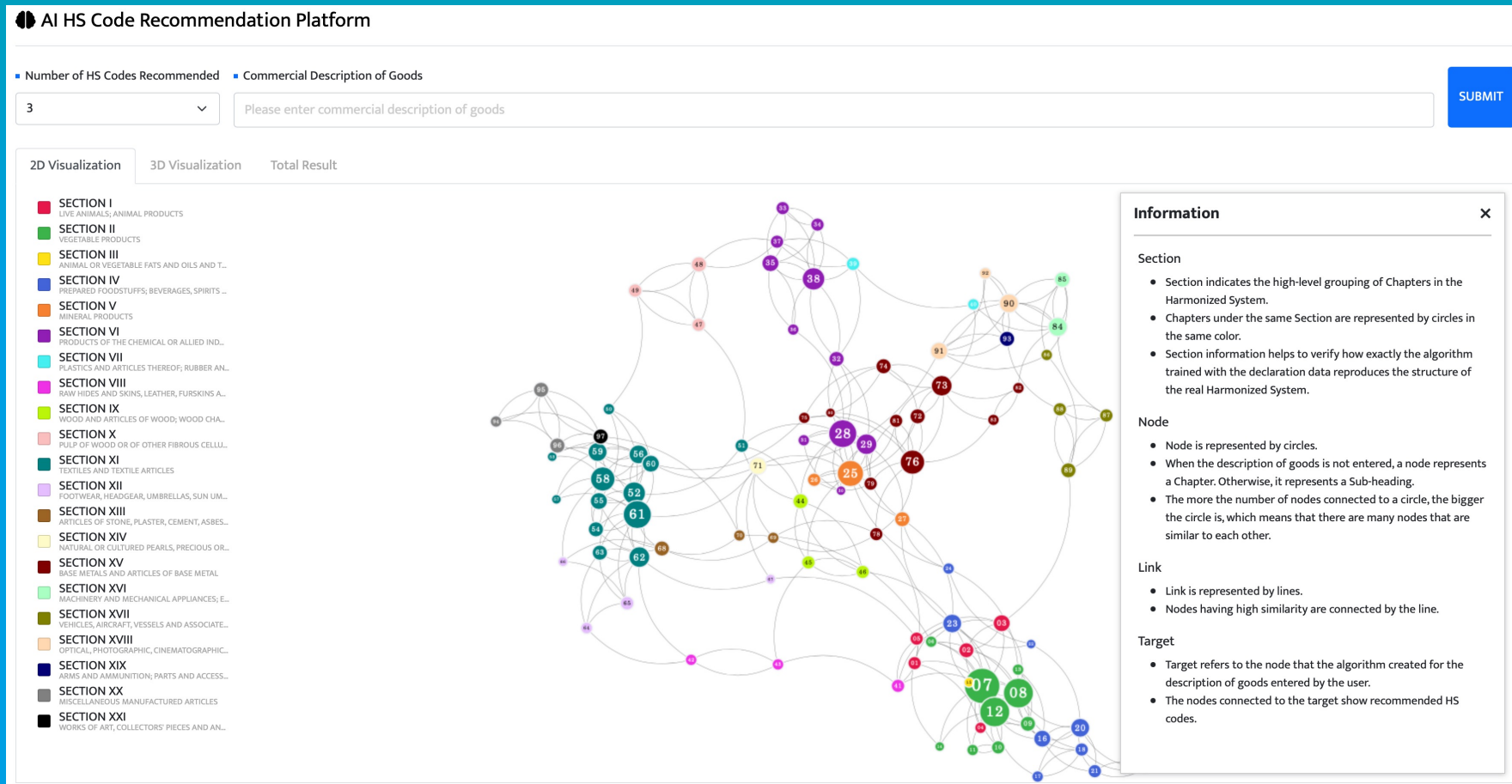
Problem: Misclassification of commodities and HS codes cause significant loss of revenue

Goal: Create AI models that recommend HS Codes based on its commodity description with a high accuracy

Method: Text embedding methods like TF-IDF, Doc2Vec are used to create an accurate dynamic model



AI HS



Achievements

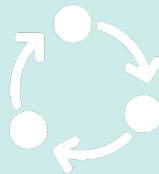
Capacity Building Data Analytics Framework

Including Regional Events



Practical Algorithms for Customs

DATE Algorithm, AI HS



Online Courses/Offline Workshops

Data Analytics
Beginner/Intermediate/Advanced course,
Scholarship Programme



Online Training Courses

BEGINNER'S COURSE TO DATA ANALYTICS

EN, FR, ES, AR, RU

Basics in Data Analysis and
Python

1

INTERMEDIATE COURSE TO DATA ANALYTICS

EN, FR

Machine Learning applied to
Customs, DATE Algorithm

2

AI HS, LITE DATE

EN, FR

ADVANCED COURSES

3

4

Languages & Advanced
Courses

MORE COURSES TO FOLLOW

Onsite Scholarship Program

Goal: Intensive Learning at the SKKU University to develop and deploy Data Analytics Experts

Duration: 5 Month long for 12 member officials, 1 month online in January 2022





**World Customs
Organization**

Words may Inspires,
But only Action creates Change

Together, We Fuel the Future of Customs

Joel Choi, Data Analytics Specialist

