



# PREDICTIVE ANALYSIS

## Proof of Concept

**Alessandro EL KHOURY**  
**Legal and Policy Officer**  
**DG SANTE – Unit G5**

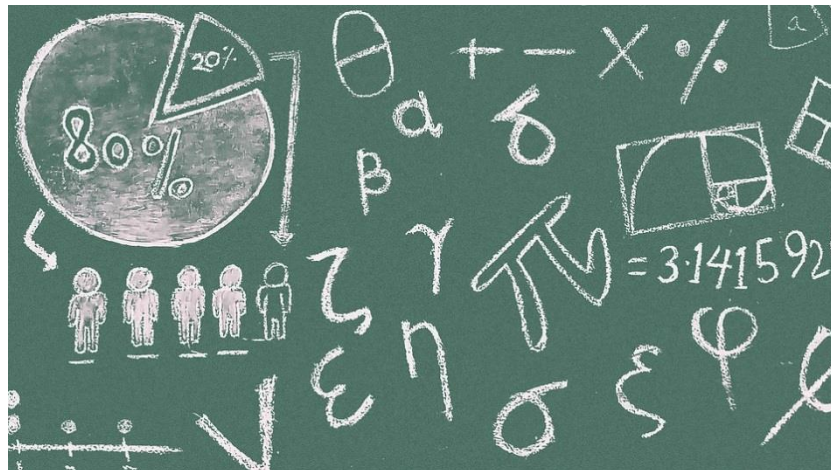
*Health and  
Food Safety*



European  
Commission

# Scope of the exercise

1. Build a statistical model that uses data in IMSOC to predict a certain level of risk to commodities imported in the EU.
2. Use the results of point 1 to calculate frequencies of controls.





European  
Commission

# SYSTEMS AVAILABLE AND SYSTEMS USED

## Available system components of IMSOC:

- TRACES
- RASFF
- AAC
- EUROPHYT
- ADNS

## Systems with adequate data:

- TRACES
- RASFF
- AAC

## System used:

- TRACES

### Key elements:

- Structure of data
- Granularity of data
- Quantity of data
- Pertinence of data

# SYSTEMS AVAILABLE AND SYSTEMS USED

## RASFF VS TRACES

<b>RASFF - For risks</b>	<b>TRACES - For certification</b>
Variability of controls	Structured controls (import)
Limited to food and feed (mainly internal market)	All import commodities requiring a certificate / entry document
Data on risks	Customs data + controls
Less useful variables for the exercise	More useful variables for the exercise
Only negative results (when risk occurs)	Negative and positive results (good certificates and rejected certificates)

# 1. Categories of goods chosen

- *Meat products (bovine, porcine and poultry)*
- *Different CN codes for meat products*
- *Grouping exercise*



**Initial Dataset**

# Why did we choose those 3 categories for this proof of concept?

Given the time-constraint and the limited scope of the exercise:

- *Good number of CVED-Ps*
- *Relatively easier to create product categories*
- *Good number of rejections*
- *Comparable results with the same methodology*



European  
Commission

# Initial Dataset

**CVED-P**

**Accepted and  
rejected**

**2011-today**







European  
Commission

# Initial dataset looks like...

## EUROPEAN COMMUNITY

## The Common Veterinary Entry Document, CVED

**Part 1 : Details of dispatched consignment**

1. Consignor / Exporter <input type="checkbox"/> Name <input type="checkbox"/> Address <input type="checkbox"/> Country <input type="checkbox"/>		2. CVED reference number <input type="checkbox"/> Local reference number: <input type="checkbox"/>	
3. Consignee Name <input type="checkbox"/> Address <input type="checkbox"/> Country <input type="checkbox"/>		4. Person responsible for load Name <input type="checkbox"/> Address <input type="checkbox"/> Country <input type="checkbox"/>	
5. Importer Name <input type="checkbox"/> Address <input type="checkbox"/> Country <input type="checkbox"/>		6. Country of origin + ISO code <input type="checkbox"/> 7. Country from where consigned + ISO code <input type="checkbox"/>	
9. Arrival at BIP (estimated date) <input type="checkbox"/>		8. Delivery address <input type="checkbox"/>	
11. Vessel name / Flight No. Aeroplane <input type="checkbox"/> Ship <input type="checkbox"/> Railway wagon <input type="checkbox"/> Identification: <input type="checkbox"/> Documentary references: <input type="checkbox"/>		10. Veterinary documents Number(s) <input type="checkbox"/> Date of issue <input type="checkbox"/> Establishment of origin(when relevant) <input type="checkbox"/> Veterinary approval number <input type="checkbox"/>	
16. Seal number and Container number Seal Number <input type="checkbox"/> Container Number <input type="checkbox"/>			
17. Transhipment to EU BIP <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> 3rd country <input type="checkbox"/>		18. For transit to 3rd Country To 3rd Country <input type="checkbox"/> No <input type="checkbox"/> Yes <input type="checkbox"/> Exit BIP: <input type="checkbox"/>	
19. Conform to EU requirements Does NOT conform <input type="checkbox"/> Conforms <input type="checkbox"/>		20. For re-import No <input type="checkbox"/> Yes <input type="checkbox"/>	
21. For internal market Animal Feedingstuff <input type="checkbox"/> Human Consumption <input type="checkbox"/> Technical Use <input type="checkbox"/> Other <input type="checkbox"/> Pharmaceutical Use <input type="checkbox"/>		22. For NON-Conforming consignments Customs Warehouse Registered No. <input type="checkbox"/> Direct To A Ship Registered No. <input type="checkbox"/> Free Zone Or Free Warehouse Registered No. <input type="checkbox"/> Ship Supplier Registered No. <input type="checkbox"/> Name <input type="checkbox"/>	

- Each field in the certificate corresponds to a **variable**
- For each certificate there are different variables.
- Technically this can be pictured as an excel spreadsheet...



European  
Commission

- Lines (horizontal) represents a certificate
- Columns (vertical) represents the different variables of the certificate

Consignor/ Exporter	Consignor Address	Consignor Post Code	Consignor City	Consignor Country	Consignor Country ISO Code	CVED Reference Number	Local Reference Number	Border Inspection Post	Entry BIP Country	Entry BIP Country ISO Code	TRACES Unit Number	Consignee	Consignee Country ISO Code
<b>Total</b>													
Azul Agronegocios S.A.	Scalabrini Ortiz 215 Piso 3 Departamento B	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.ES.2018.0010626	1VLC1-7900609764013	Valencia, P	Spain	ES	ESVLC1	REINA APICOLA LEVANTINA, S.L.	ES
Azul Agronegocios S.A.	Scalabrini Ortiz 215 Piso 3 Departamento B	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.ES.2018.0012474	1VLC1-7900609764004	Valencia, P	Spain	ES	ESVLC1	REINA APICOLA LEVANTINA, S.L.	ES
BM SUPPLIER SRL	BLANCO ENCALADA 197 PB OF 49 SAN ISIDRO	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.MT.2018.0000935	2018/0892	Marsaxlokk, P	Malta	MT	MTMAR1	EURL SIRINE TRADING	MT
CASAMEN S.A.	MAXIMO PAZ735 Y 963 (B1824KSO) LANUS OESTE	Buenos Aires	Buenos Aires	Argentina	AR	CVEDP.ES.2018.0004850	1VLC1	Valencia, P	Spain	ES	ESVLC1	ALBERTO L. HUGHES, S.L.	ES
Compañía Bernal S.A.	Comandante Franco 4901	Buenos Aires	Bernal	Argentina	AR	CVEDP.DE.2018.0013240	100-491-18	Hamburg Hafen	Germany	DE	DEHAM1	Block House Fleischerei GmbH	DE
Congeladores Patagonicos S.A.		Chubut	Puerto Madryn	Argentina	AR	CVEDP.ES.2017.0061261	1MAR1-7900603609236	Marin, P	Spain	ES	ESMAR1	PESCANIGRAN S.L.	ES
Congeladores Patagonicos S.A.		Chubut	Puerto Madryn	Argentina	AR	CVEDP.ES.2017.0061272	1MAR1-7900603609281	Marin, P	Spain	ES	ESMAR1	PESCANIGRAN S.L.	ES
COOMARPES LTDA DARSENA DE PESCADORES	PUERTO MAR DEL PLATA 7600 MAR DEL PLAT	Montevideo	Montevideo	Uruguay	UY	CVEDP.NL.2018.0016966	18033688	Rotterdam	Netherlands	NL	NLRMT1	/	NL

# Data Manipulation phase

Data was "cleaned" → manipulated to make it simpler for machine learning algorithm to use

1. Delete certificates with multiple CN codes (<1%)
2. delete variables with >20% missing values
3. Delete all columns that cannot be converted into numbers (i.e. name of operators)
4. Delete geographical data (multiple methods possible, currently outside the scope of the exercise)
5. Replace missing values for variables with <20% (prediction model)



European  
Commission

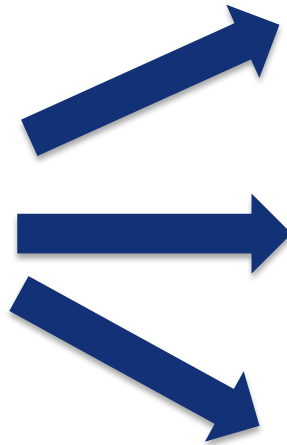
# Final dataset

- Divided in 2 parts (randomly generated):
  - 80% to learn
  - 20% to predict

		Total
<b>Bovine - Train</b>	634	793
<b>Bovine - Test</b>	159	
<b>Porcine - Train</b>	253	316
<b>Porcine - Test</b>	63	
<b>Poultry - Train</b>	1110	1387
<b>Poltry - Test</b>	277	
<b>Total observations after cleaning</b>		2496

## 3 different categories

1 model  
(random forest)



Bovine meat



Porcine meat



Poultry meat

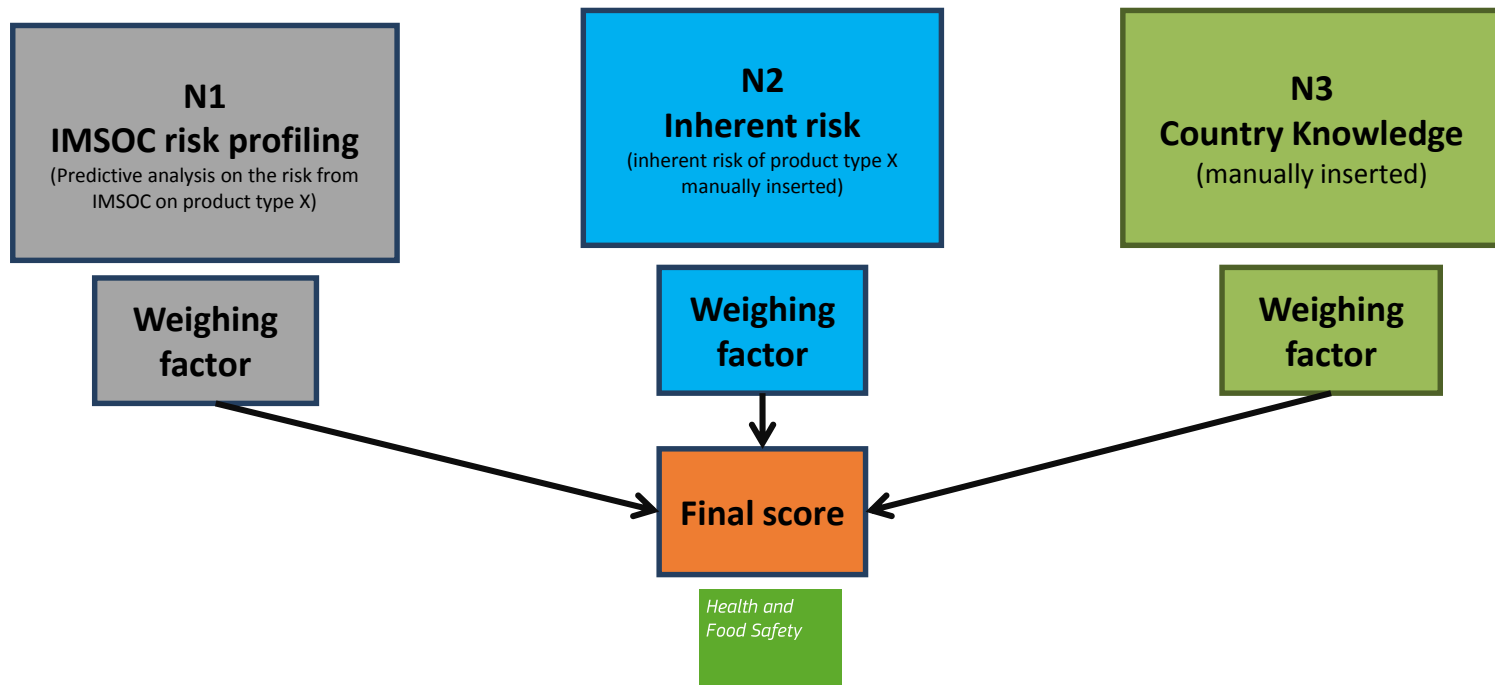


# Predictive model's performance

	<b>Bovine</b>	<b>Porcine</b>	<b>Poultry</b>
Absence of non compliances	100%	~80%	100%
Documentary checks fail	~60%	~70%	~50%
Identity checks fail	~40%	-	~60-70%
Physical checks fail	~80%	~60%	~95%

# Possible mathematical approach to calculate the frequency

*N1, N2, N3 would give 3 sub-scores between 0 and 1, and a weighing factor is applied to each sub-score.*



# Two possible methods to calculate the frequency

**Method A:** sum (frequency per consignment)

$N1+N2+N3$  -> algorithm to calculate the frequency

**Method B:** multiplication (frequency per product category)

$N1*N2*N3$  -> algorithm to calculate the frequency

# METOD A: SUM

## (Frequency per individual consignment)

Weighed N1, N2 and N3 are added to determine a final score for each individual consignment.

- N1 varies for each consignment, according to the likelihood of rejection;
- N2 and N3 are stable;

This final score is used in an algorithm to calculate the frequency for each individual consignment.

**Result:** different frequency for each individual consignment of the same product type.



# METOD B: MULTIPLICATION (Frequency per product category)

$N_1$ ,  $N_2$  and  $N_3$  have the same weight.

- $N_1$  is given by the average of the reasons of rejection for a specific product category.
- $N_2$  and  $N_3$  are stable.

$N_1 * N_2 * N_3$  are used in an algorithm to calculate a frequency for a specific product category.

**Result:** one common frequency for all consignments of the same product type coming from a specific third country.



European  
Commission

# QUESTIONS

- 1. Do we need to have a separate frequency for Identity checks?*
- 2. Do you prefer a frequency established per each consignment, or one per product category?*

# Default 100% frequencies

- *Animals*
- *Plant and plant products*
- *Certain POAO and ABP*

Possibility to lower the default rate if needed.



European  
Commission

# Thank you for the attention!

SANTE Unit G5 – Alerts, Traceability and Committees  
TRACES-IMSOC sector  
[SANTE-IMSOC@ec.europa.eu](mailto:SANTE-IMSOC@ec.europa.eu)