Evaluating the impact of Data analytics on the Customs Risk Management process: A balancing act

Picard Conference, November 24th 2020
Dr. Mathieu Labare, Data Scientist – Risk Management – Belgian Customs
Mathieu.labare@minfin.fed.be
Jonathan Migeotte, Coordinator – Risk Management – Belgian Customs
Jonathan.migeotte@minfin.fed.be

A project funded by the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No 786748.
Belgian Customs, Profile & Data analytics

01  Belgian Customs
   • Data Science team for ~10 yrs
   • SEDA 2.0 risk engine

02  H2020 Profile
   • Research project with 5 Customs Administrations
   • Machine learning & data analytics for customs risk management

03  WP4 – Belgian Living Lab
   • Economic operator profiling
   • Summary declaration data enhancement

10+ data scientists for Risk Management
→ Trained customs officers
→ Risk related domain expertise
The dual role of Customs

Facilitate legitimate trade

Prevent, identify and stop fraud

✓ AEO
✓ Authorizations
✓ Green lanes

Physical and document controls
Risk profiles and Data-mining
Infringement records

Irregular declarations
Selected declarations
Legitimate declarations
The dual role of Customs

Legitimate declarations

- Normal trade
  - Relevant controls
- Missing targets
  - Irrelevant controls

Irregular declarations

- Detected frauds
- Irrelevant controls (slows down trade flow)

Confusion (error) matrix

Selected for control
- No
- Yes

Smooth trade

Missing targets
Classical KPI’s and their limitations

- **Accuracy**: Percentage of good decisions
  - Problem: Customs data is strongly imbalanced
    - Accuracy will always be close to 1

- **Precision**: Percentage of good selections
  - Problem: Does not quantify what is missed.

- **Sensitivity**: Percentage of frauds found
  - Problem: Estimation about what is missed must be available

Confusion (error) matrix

<table>
<thead>
<tr>
<th>Legitimate declaration</th>
<th>Selected for control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Normal trade</td>
</tr>
<tr>
<td>Yes</td>
<td>Irrelevant controls</td>
</tr>
<tr>
<td>Missing targets</td>
<td>Relevant controls</td>
</tr>
</tbody>
</table>

Legend:
- Normal trade
- Irrelevant controls
- Relevant controls
- Missing targets
Classical KPI’s and their limitations

**Confusion (error) matrix**

<table>
<thead>
<tr>
<th>Legitimate declaration</th>
<th>Selected for control</th>
<th>Relevant controls</th>
<th>Irrelevant controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>Missing targets</td>
<td>Normal trade</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Relevant controls</td>
<td>Irrelevant controls</td>
</tr>
</tbody>
</table>

**Problem 1**: Control resources are not unlimited
- Time, equipment, costs, available field officers

**Problem 2**: More unnecessary controls
- Slows down the trade flow
- Unhappy reliable operators
- Waste of time and effort

**HOW TO BE MORE EFFICIENT?**

- **Precision**: Percentage of good selections
- **Sensitivity**: Percentage of frauds found

**INCREASE SELECTION RATE**

**INCREASE SENSITIVITY**
Classical KPI’s and their limitations

**Precision:** Percentage of good selections

**Sensitivity:** Percentage of frauds found

**Problem 3:** Not every trade and every associated risk has the same weight
- Small frauds: common / limited impact / small payback
- Large frauds: rare / large shortfall / large payback

**Opportunity-cost matrix**

**Confusion (error) matrix**
Classical KPI’s and their limitations

**HOW TO BE MORE EFFICIENT?**

- **Precision**: Percentage of good selections
- **Sensitivity**: Percentage of frauds found

### Confusion (error) matrix

<table>
<thead>
<tr>
<th>Legitimate declaration</th>
<th>Selected for control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Missing targets</td>
</tr>
<tr>
<td>Yes</td>
<td>Relevant controls</td>
</tr>
<tr>
<td></td>
<td>Irrelevant controls</td>
</tr>
<tr>
<td>Yes</td>
<td>Normal trade</td>
</tr>
</tbody>
</table>

**Opportunity-cost matrix**

- **Selected for control**
  - No
  - Yes

- **Legitimate declaration**
  - No
  - Potentially huge impact
  - Avoid damages + fines

- **Mandatory by law**

**Problem 4**: the opportunity/cost weight is sometimes difficult to quantify
Classical KPI’s and their limitations

HOW TO BE MORE EFFICIENT?

Confusion (error) matrix

<table>
<thead>
<tr>
<th>Selected for control</th>
<th>Legitimate declaration</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Missing targets</td>
</tr>
<tr>
<td></td>
<td>Relevant controls</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal trade</td>
</tr>
<tr>
<td></td>
<td>Irrelevant controls</td>
</tr>
</tbody>
</table>

Opportunity-costs matrices

<table>
<thead>
<tr>
<th>Selected for control</th>
<th>Legitimate declaration</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Costs &amp; Damages</td>
</tr>
<tr>
<td></td>
<td>Reward</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Facilitation</td>
</tr>
<tr>
<td></td>
<td>Investments costs</td>
</tr>
</tbody>
</table>

Illustrative Case: Import of forbidden dangerous goods

- Strategy: Missing frauds must absolutely be avoided.
  - Costs & Damages >>> Reward and Investments costs

Classical KPI’s do not reflect the priorities and strategy

KPI’s should be designed depending on:

- the information available
- well defined objectives
- the reality of trade trends
Customs expertise

Customs framework is very specific and complex:
- Large number of procedures
- Many risks and fraud scenarios
- Quickly evolving
- Often obscure for external collaborators

→ Experience is key

Understanding of:
- the goals
- the variables
- the limitations

Fast response must be limited to specific topics
Broader targets need time and learning efforts

Contribution and outcomes must make sense for internal usage

Ex: PROFILE operator behaviour analysis
How to measure customs expertise?

Parameters such as:
- Level of proficiency
- Number of analysts
- Coverage of risk domains

BEGINNER
No prior knowledge of customs business

AWARENESS
Basic knowledge
General understanding

EXPERT
Extensive expert knowledge in specific risk domains
Link with ‘Big Picture’
Tailored advice with context and examples

EXPERIENCED
Broad and in depth knowledge
Ability to deal with exceptions & special cases
Ability to share experience

TRAINED
Working knowledge
Ability to apply knowledge
Work independently
Data quality through the lens of Data usability

**Data ‘up-to-dateness’**
- Live data vs ‘Historical data’
- In advance / Just in time
- Updates / change in the data

**Data reliability**
- Limited data quality
- Digitalization & standardization efforts
- Limited feedback
- Automated NII
- Improved processes
  ➔ Limited level of certainty

**Data integration**
- Supply chain / Logistics chain / Customs data
  - Multiple actors and systems
  - Multiple levels of information
  ➔ Not easily linked or traceable

**Data ‘up-to-dateness’**
- Live data vs ‘Historical data’
- In advance / Just in time
- Updates / change in the data

**Data protection**
- Privacy by design & data minimization
- Dataset selection
- Anonymization / Pseudonymization
- Secure processing / exchange
## Evaluation dimensions for data usability

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variety</td>
<td>How many distinct sources are available?</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Is data correct and how well does it represent reality?</td>
</tr>
<tr>
<td>Completeness</td>
<td>Does a dataset include all critical data elements?</td>
</tr>
<tr>
<td>Granularity</td>
<td>Does data provide detailed enough information?</td>
</tr>
<tr>
<td>Standardization</td>
<td>Is data in a standardized format?</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Is data available in advance / when it is needed?</td>
</tr>
<tr>
<td>Comparability</td>
<td>Can data be used with other information to support decision-making?</td>
</tr>
</tbody>
</table>
Data usability in practice

PROFILE Belgian Living Lab: Summary Declaration (ENS) enhancement

Use of multiple internal and external datasets to enhance the quality of ENS data

- Discovering useful additional features
- Cross-checking and validation of the data
- Evaluation of the completeness of the data
Evaluating the impact of Data Analytics: A balancing act

The risk management strategy must be at the core of the evaluation.