

CADRE D'ANALYSE MULTI-DIMENSIONNELLE APPLIQUÉ AU DEVELOPPEMENT D'OUTILS DE TRAITEMENT ANALYTIQUE DE DONNÉES

Conférence Picard, 09-10 Décembre 2021

Dr. Mathieu Labare, Data Scientist – Gestion des Risques – Douanes BE

Mathieu.labare@minfin.fed.be

Jonathan Migeotte, Coordinateur – Gestion des Risques – Douanes BE

Jonathan.migeotte@minfin.fed.be

01100010110100101100101001
PROFILE

A project funded by the European Union's Horizon
2020 research and innovation programme under the
Grant Agreement No 786748.



Federal
Public Service
FINANCE

CUSTOMS AND EXCISES



Douanes Belges

Analyse de données & I.A.



Analyse des données par les Douanes Belges

- Science des données depuis 10 ans
 - Outil de selection SEDA 2.0
- Science des données au sein de la Gestion des risques
→ Douaniers de formation
→ Complémentaire avec analyse des risques

Analysis Team

Requêtes

- ✓ Stats & rapports
- ✓ Analyses génériques
- ✓ Soutien aux analystes de risque

Datamining Team

Analyses avancées

- ✓ Modélisation prédictive
- ✓ Analyses multi-dimensionnelles
- ✓ Axé thématiques

NEW AI Team

Technologies de rupture

- ✓ I.A.
- ✓ Traitement d'images
- ✓ Spécialisation élevée

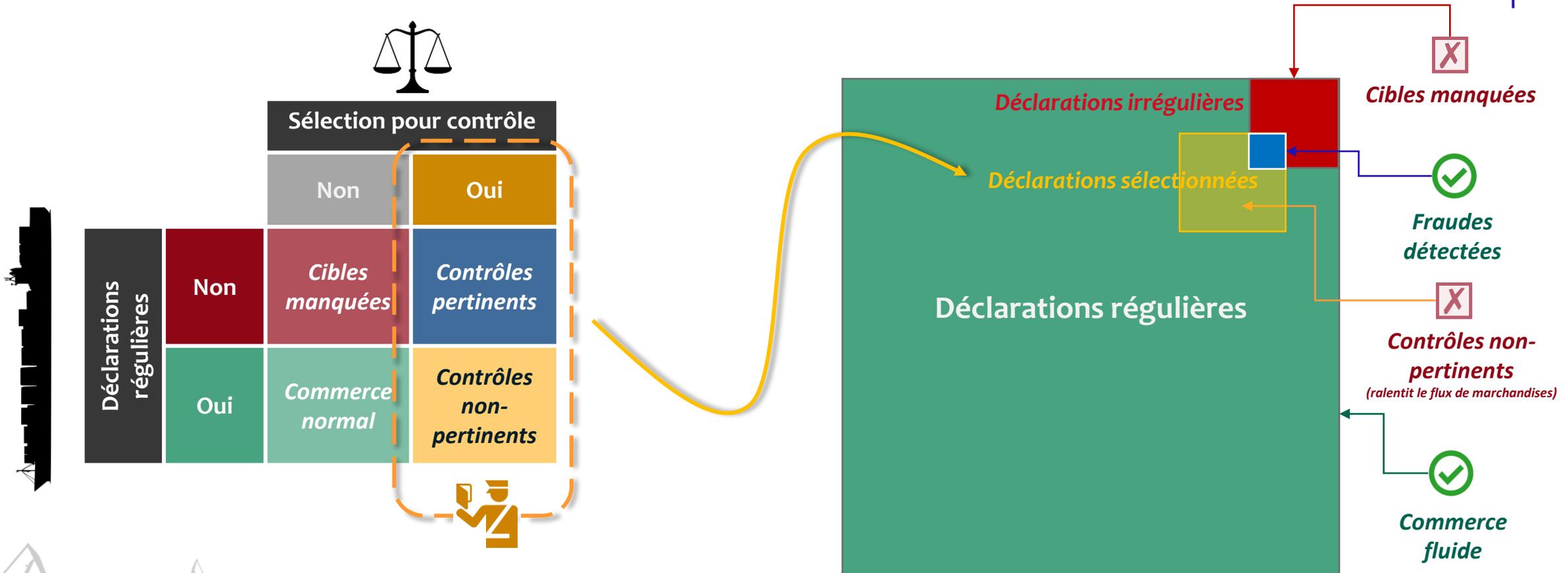
Projet de recherche PROFILE

- Projet comprenant 5 autorités douanières
 - Apprentissage machine et analyse des données pour la gestion des risques douaniers
- “Living Lab” belge**
- Détection de fraude générique
 - Détection d'anomalies
 - Renforcement de la qualité des données issues des déclarations sommaires



Comment comparer l'efficacité de diverses approches de traitement et analyse de données ?

Bases de théorie de la décision appliquée au domaine douanier

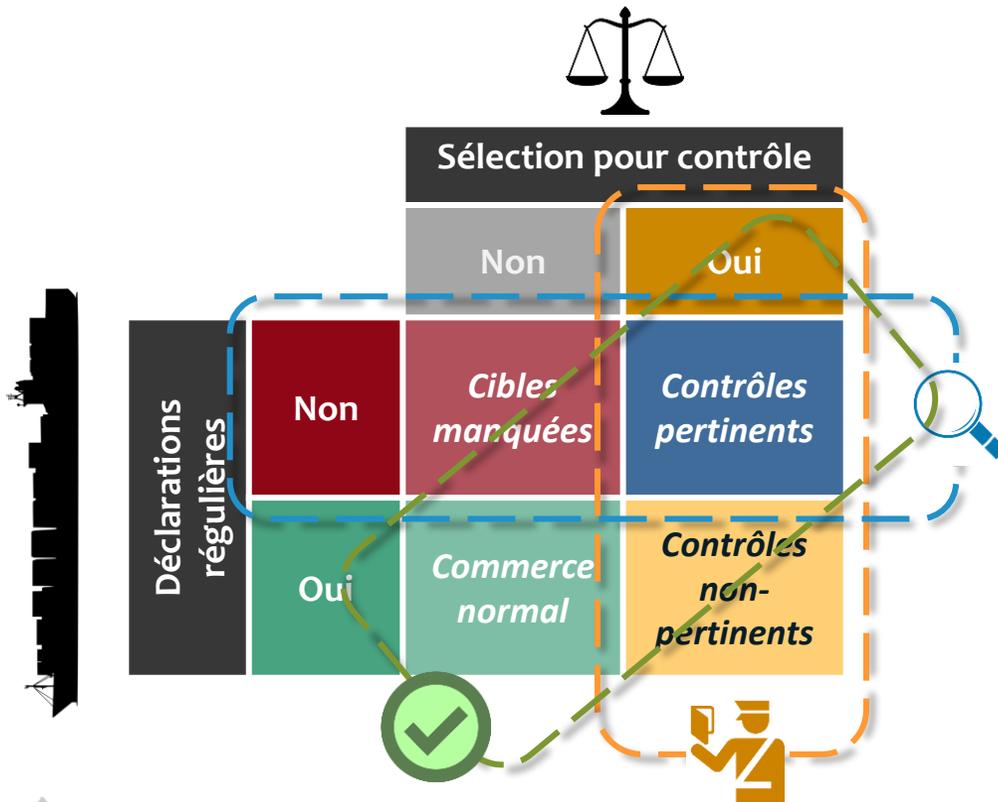


Estimations classiques de l'efficacité et leurs limites



Federal
Public Service
FINANCE

CUSTOMS AND EXCISES



- **Précision** : *Pourcentage de sélections pertinentes*

Problème : Ne quantifie pas ce qui est manqué

- **Sensibilité** : *Pourcentage de fraudes observées*

Problème : Requier une connaissance ou estimation des fraudes non-observées

- **Exactitude** : *Pourcentage de décisions correctes*

Problème : Les données douanières sont fortement déséquilibrées

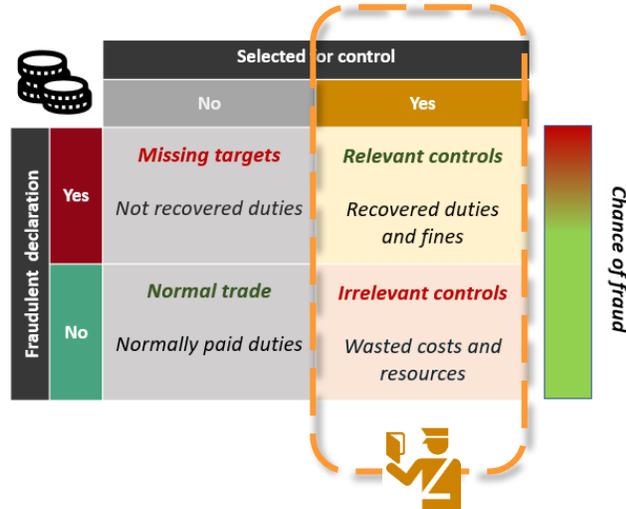
❖ *Exactitude sera toujours proche de 1*

Le mythe de la mesure globale de l'efficacité



Risques financiers

- Optimisation du recouvrement des droits et taxes
- Impact plus important sur les gros montants
- Cibler chaque fraude gaspille les ressources disponibles

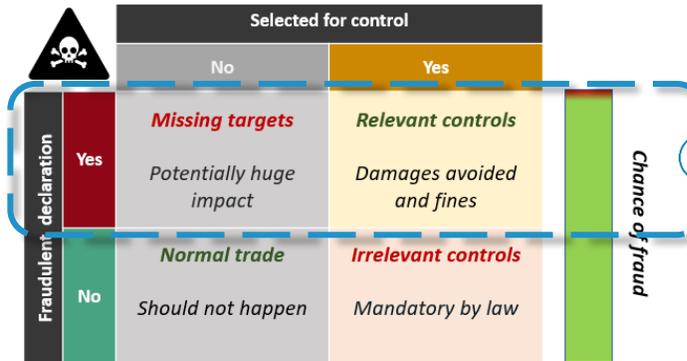


Risques liés à la Sécurité

Empêcher l'usage public de produits dangereux

Impact similaire pour chaque type de risque

Manquer une infraction est potentiellement dangereux

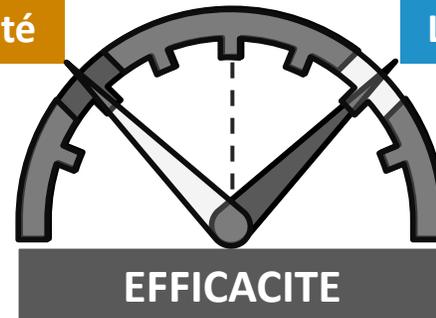


La précision est *souvent* plus importante que la sensibilité

La sensibilité est *souvent* plus importante que la précision

Précision

Sensibilité



Cadre d'analyse

pour les stratégies de Gestion des Risques



Federal
Public Service
FINANCE

CUSTOMS AND EXCISES

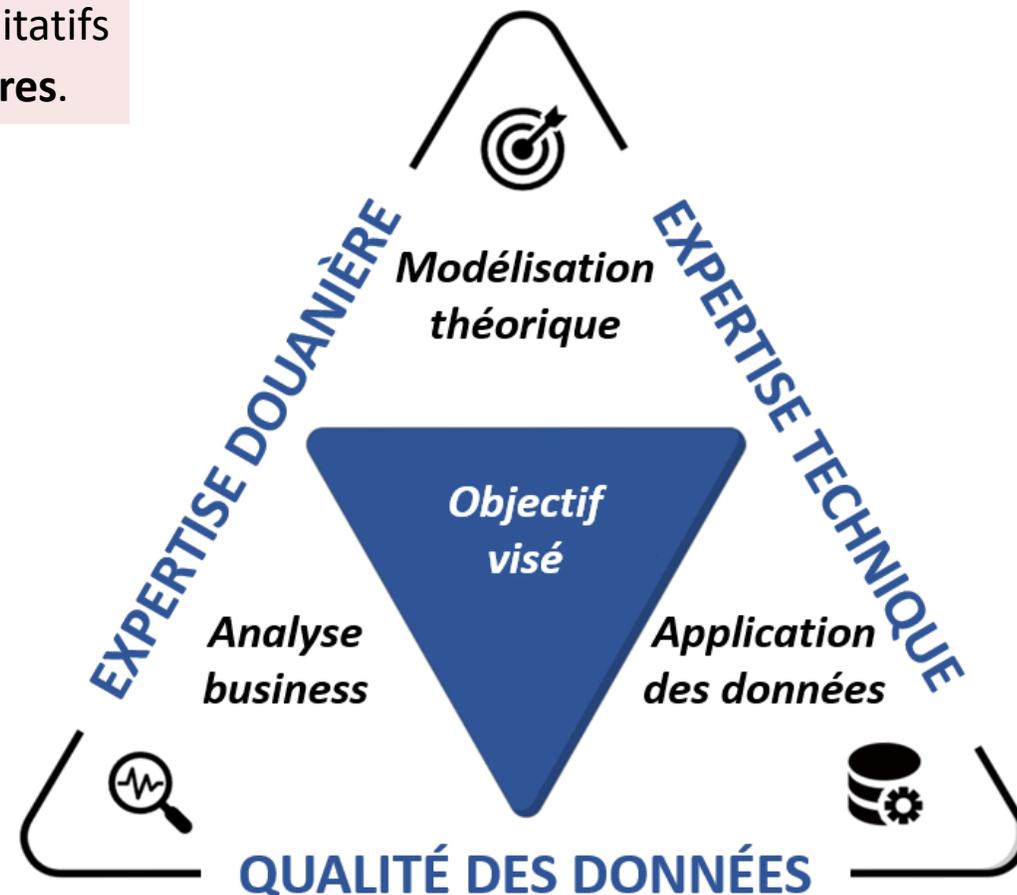
Cadre générique d'analyse comparant certains paramètres qualitatifs observables dans tout **traitement et analyse de données douanières**.

- s'applique à tout type de résultat, **indépendamment** du niveau de développement et d'avancement.

→ La perte d'interprétation quantitative est compensée par une capacité de comparaison efficace de résultats variés.

- Fournit une analyse précise de l'impact respectif de différentes méthodes d'analyse.

→ Souligne les **forces et faiblesses individuelles**, et permet de tirer des **conclusions et recommandations** spécifiques



Cadre d'analyse

1^{ère} dimension : Qualité des données



Les données forment la base sur laquelle toute l'expertise est développée et affinée.

Elles doivent être :

- fiables
- structurées
- à jour
- intégrées

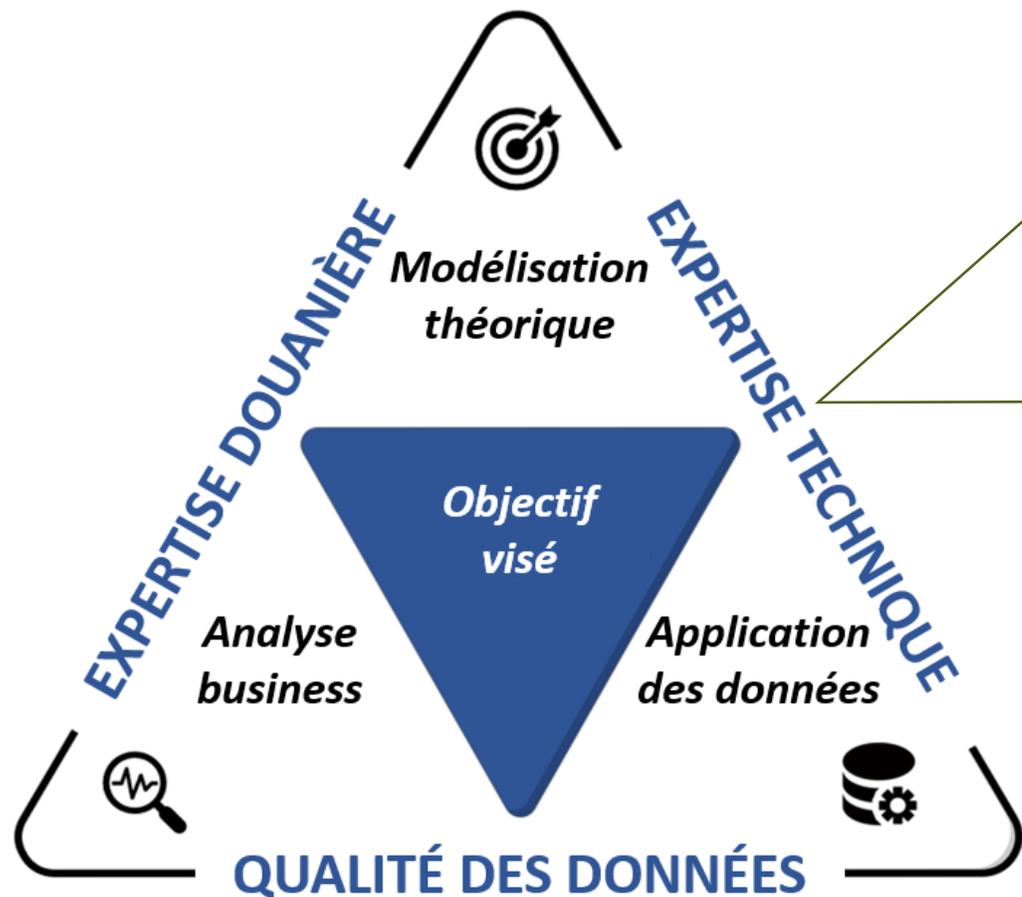
Et doivent être traitées avec prudence:

- confidentialité
- protection des données
- 'Législation sur l'I.A.'

Une faible qualité des données impacte l'ensemble du processus

Cadre d'analyse

2^e dimension : Expertise technique



• Digitalisation de la douane :

- Déclarations et documents sous forme électronique
- Automatisation des processus
- Augmentation rapide des volumes
- E-commerce

• Besoin de nouvelles compétences et ressources :

- Infrastructure IT pour la collecte, le stockage et la gestion des données
- Analyse des données / Apprentissage machine / Techniques I.A. pour le traitement et l'analyse des données
- Automatisation des processus douaniers
- Évolution rapide des techniques et méthodes

Les douanes doivent suivre l'évolution rapide des technologies de rupture

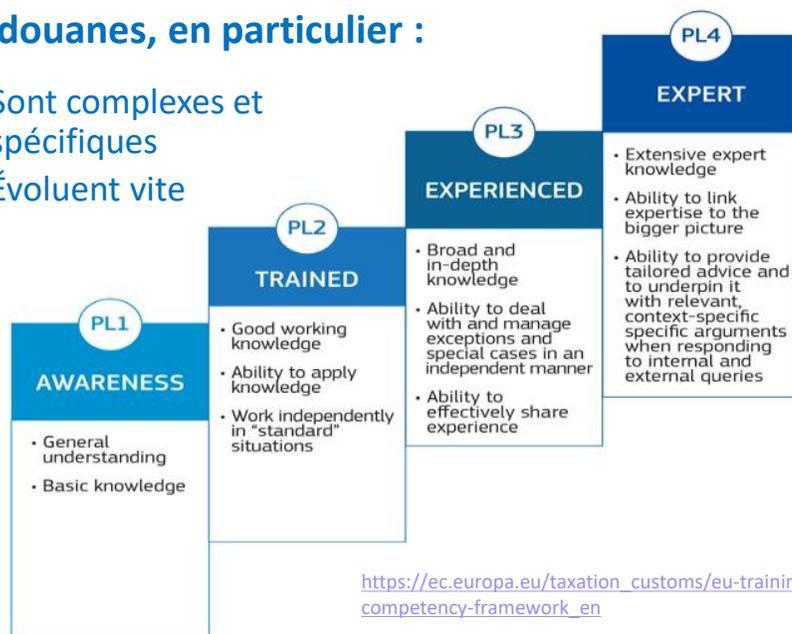
Cadre d'analyse

3^e dimension : Expertise douanière

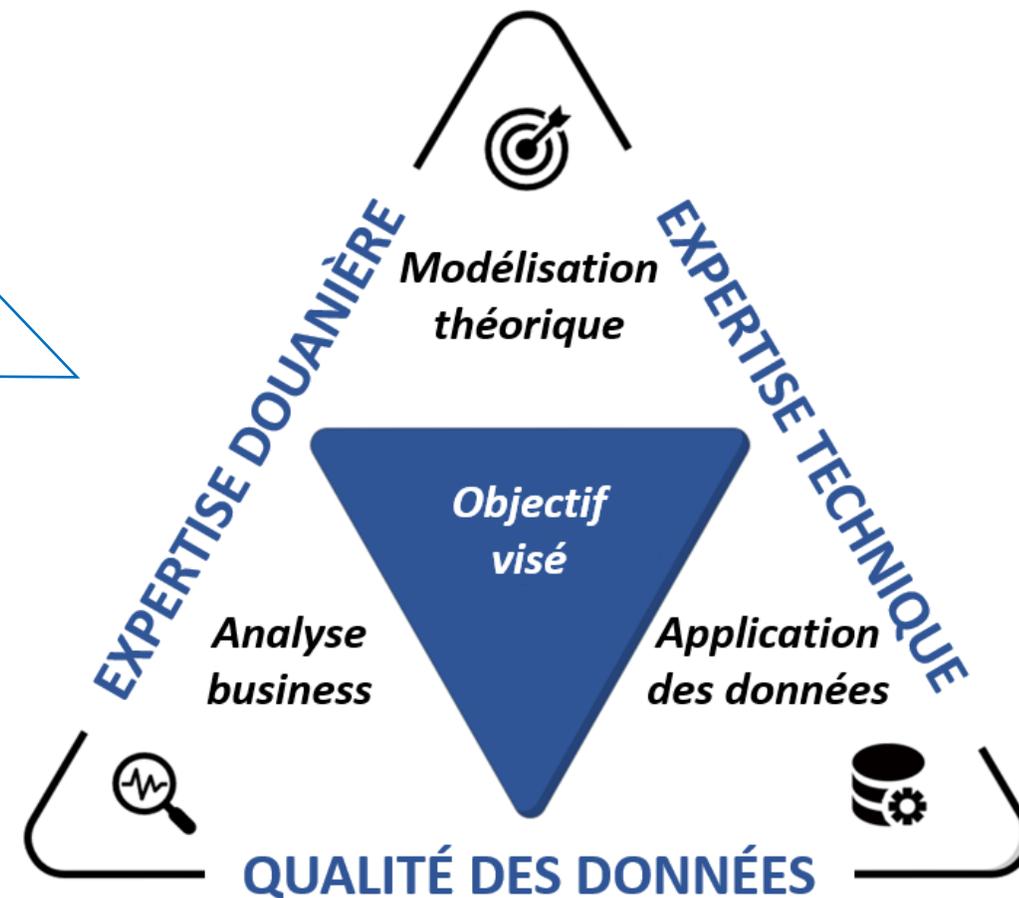


- L'importance de l'expertise métier est sous-estimée
- Les douanes, en particulier :

- Sont complexes et spécifiques
- Évoluent vite



L'exploitation de l'expertise métier est primordiale pour le développement de modèles utiles



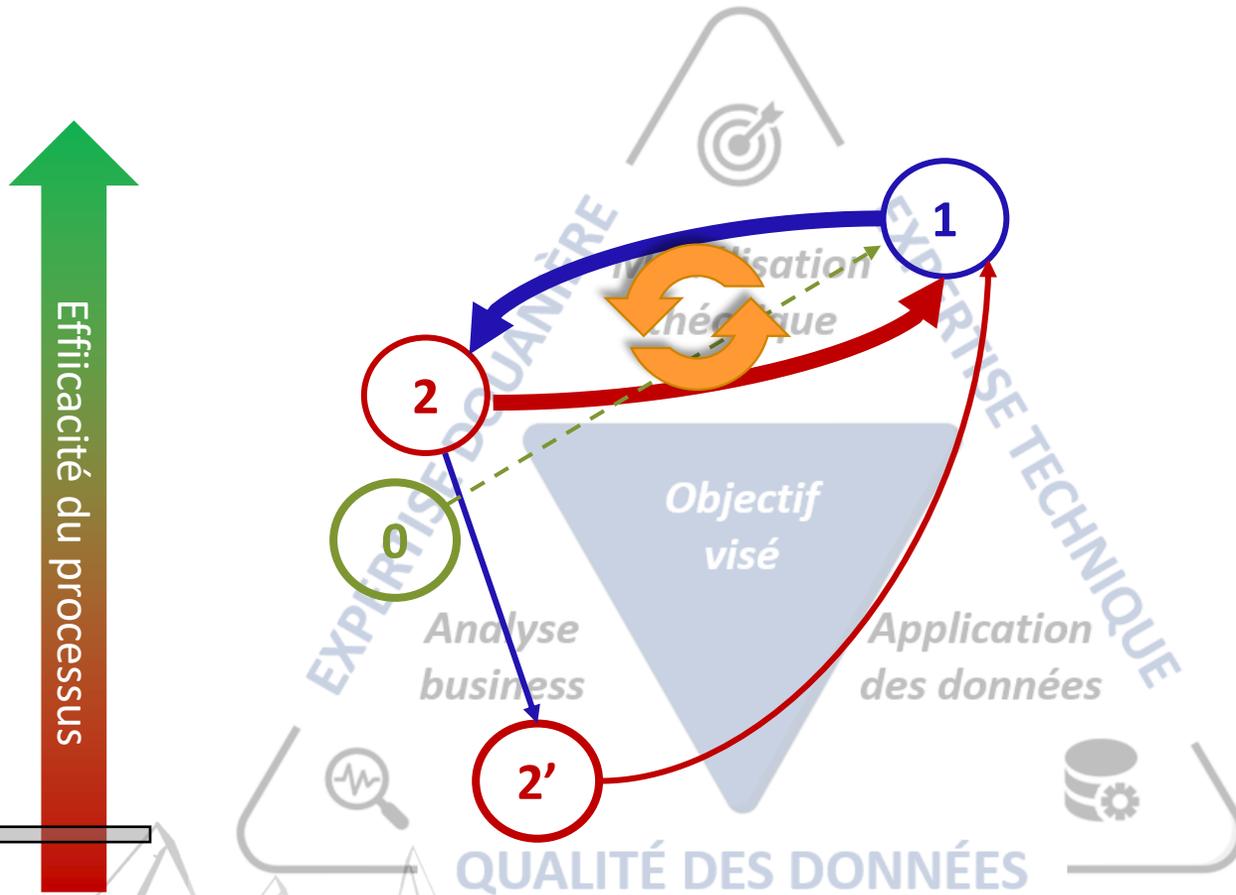
Cadre d'analyse

Exemple concret: Implémentation d'une procédure d'analyse avancée



Federal
Public Service
FINANCE

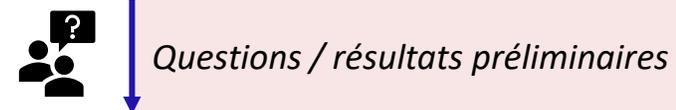
CUSTOMS AND EXCISES



0 **Management** : stratégies et objectifs



1 **Analystes Données** : Dev. techniques



2 **Experts Douane** : analyse des résultats

Implémentation

2' **Agents de terrain**

Feedback

Feedback

1 **Analystes Données** : ajustements modèles

BOUCLES ITERATIVES LENTES



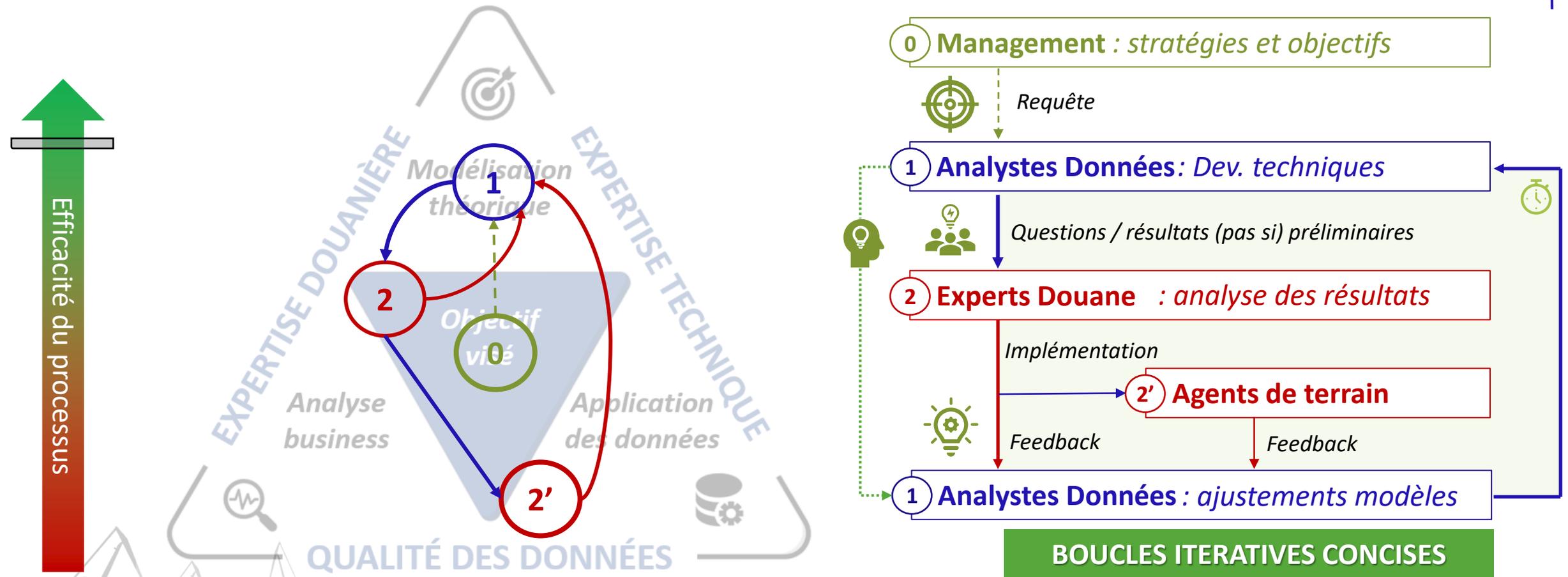
Cadre d'analyse

Exemple concret: Implémentation d'une procédure d'analyse avancée



Federal
Public Service
FINANCE

CUSTOMS AND EXCISES



Cadre d'analyse

Vers une 4^e dimension : Apprentissage et validation par cycles



Federal
Public Service
FINANCE

CUSTOMS AND EXCISES

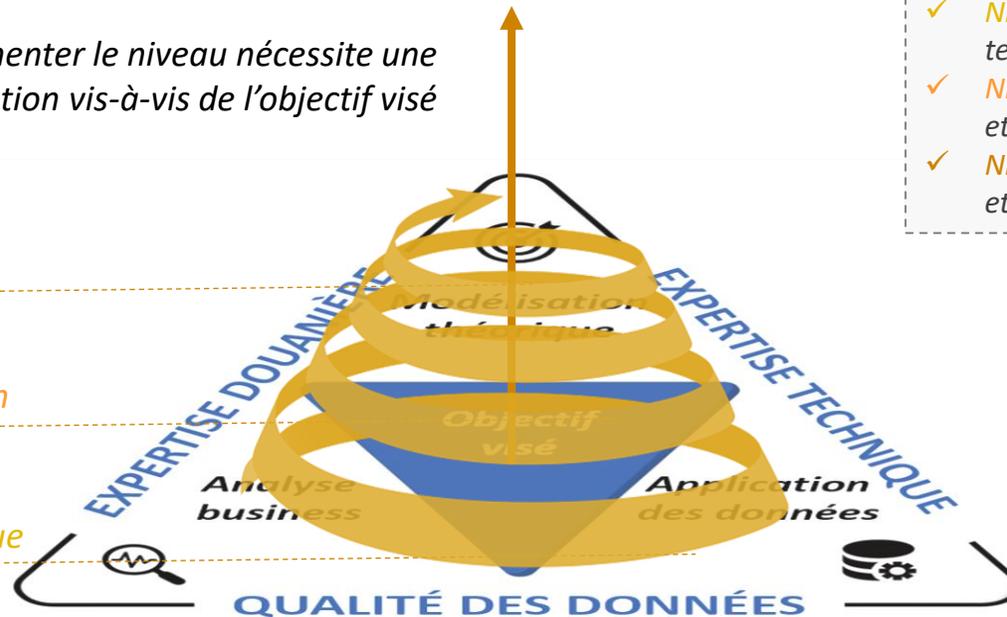
EXPERIENCE / AFFINEMENT

Augmenter le niveau nécessite une validation vis-à-vis de l'objectif visé

Niveau élevé

Niveau moyen

Niveau basique



Modélisation théorique

- ✓ **Niveau basique** : pré-traitement basé sur des tendances douanières globales et générales
- ✓ **Niveau moyen** : Identification de tendances connues et inconnues à partir d'analyses exploratoires
- ✓ **Niveau élevé** : Modèles ciblés en fonction des besoins et objectifs

Application des données

- ✓ **Niveau basique** : Nettoyage / stockage des données, vérification de la consistance du contenu
- ✓ **Niveau moyen** : Enrichissements et liens entre données homogènes
- ✓ **Niveau élevé** : Liens complexes, notation des données, Implémentation de modèles dynamiques, ...

Analyse "Business"

- ✓ **Niveau basique** : Connaissance basique des douanes et du cadre réglementaire
- ✓ **Niveau moyen** : Connaissance avancée de tendances spécifiques au métier
- ✓ **Niveau élevé** : Niveau de connaissance expert

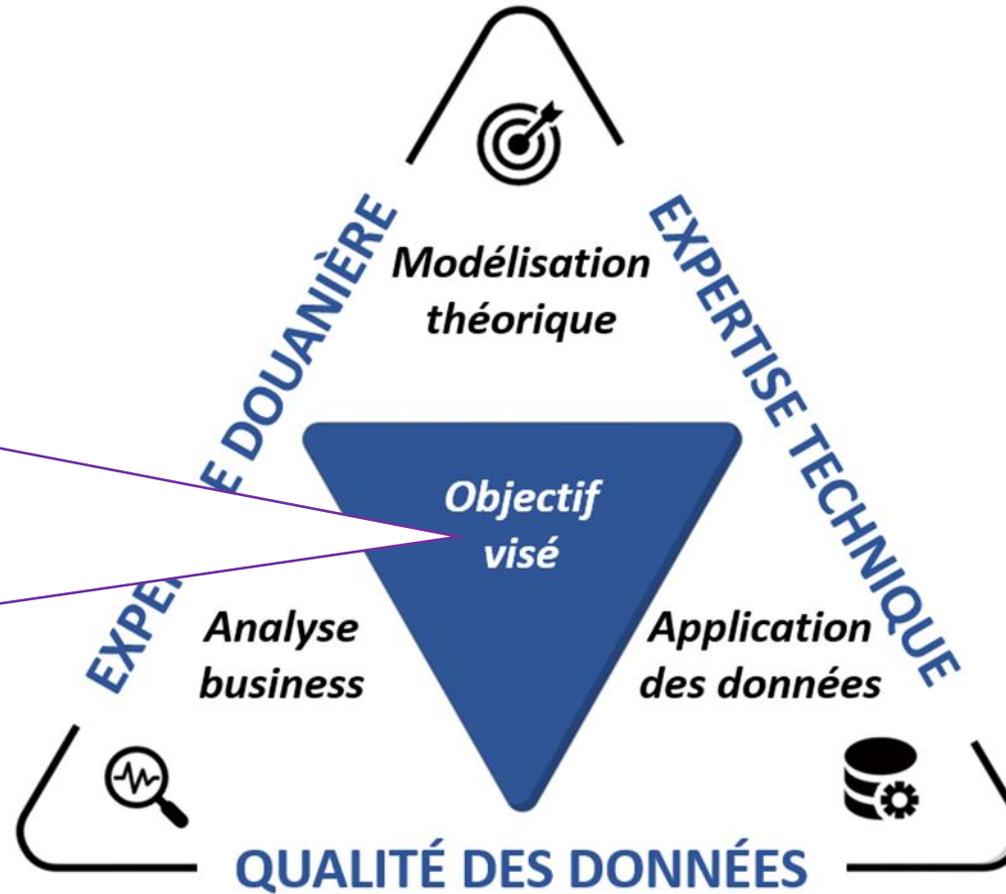


Pistes pour le futur

Dimension centrale

Perspective managériale :

- Comment s'assurer de la coordination des dimensions et de la concordance par rapport à l'objectif ?
- Quel est le « pouvoir d'attraction » de l'objectif et comment l'améliorer ?

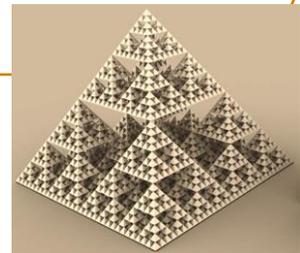


Est-ce vraiment générique ?

- Appliqué à la gestion des risques douaniers
- Autres contextes pertinents ?
 - Environnement douane
 - Autres domaines

Niveaux d'abstraction

- Jusqu'à présent : cas pratiques
- Que se passe-t-il si l'on prend du recul ?



CADRE D'ANALYSE MULTI-DIMENSIONNELLE APPLIQUÉ AU DEVELOPPEMENT D'OUTILS DE TRAITEMENT ANALYTIQUE DE DONNÉES

Résumé & Conclusions

- Approche générale en cours de validation et d'extension
 - Développé dans le cadre douanier → utilisation pour le projet PROFILE
 - Retours positifs d'autres administrations douanières
- Application à d'autres domaines à confirmer (par la mise en pratique)
- Les échanges d'expertise et de connaissances sont fondamentaux pour garantir une utilisation optimale des ressources disponibles.

01100010110100101100101001
PROFILE

A project funded by the European Union's Horizon
2020 research and innovation programme under the
Grant Agreement No 786748.



Federal
Public Service
FINANCE

CUSTOMS AND EXCISES

