



ORGANISATION MONDIALE DES DOUANES

Document de recherche de l'OMD n°39

Incidences des mégadonnées sur la Douane - Leurs atouts en matière de capacité de gestion des risques

(Mars 2017)

Yotaro Okazaki

Sommaire

Le présent document traite des incidences des mégadonnées sur le travail des services douaniers et, plus particulièrement, sur la gestion des risques. Pour pouvoir prendre des décisions plus informées et adéquates, certaines Administrations de la douane ont lancé des initiatives en matière de mégadonnées. Elles consistent à tirer parti des avantages de l'analytique, à veiller à la qualité des données (concernant le fret, les envois et les moyens de transport) et à envisager d'enrichir la gamme de données qu'elles utilisent à des fins d'analyse. Ces initiatives vous seront présentées sur la base des renseignements émanant de cinq Administrations: l'Agence des services frontaliers canadienne (ASFC); le *Customs and Excise Department* (Douane de Hong Kong, Chine); le *New Zealand Customs Service* (Douane néo-zélandaise), la Douane du Royaume-Uni (*Her Majesty's Revenue and Customs* ou HMRC) et l'*U.S. Customs and Border Protection* (Douane américaine ou USCBP).

Mots clés

Douane, mégadonnées, gestion des risques, analytique

Remerciements

Le présent article a été rédigé par Yotaro Okazaki de l'Unité Recherche de l'OMD.

Clause de non-responsabilité:

La série de documents de recherche publiés par l'OMD présente les résultats des travaux en cours, afin de susciter l'échange d'idées sur des questions douanières. Les vues et opinions exposées dans le présent document n'engagent que son auteur et ne reflètent pas nécessairement les vues et les politiques de l'OMD, ou de ses Membres.

Note

Tous les documents de recherche de l'OMD sont disponibles sur la page publique de son site web (www.wcoomd.org). L'auteur peut être contacté via le lien suivant research@wcoomd.org.

I. Introduction

Le concept de « mégadonnées » renvoie à une large palette de données et de jeux de données divers, qui ne pourront être correctement exploités qu'au moyen des technologies de pointe. De nombreuses applications de mégadonnées permettent aux institutions d'optimiser leurs performances et l'affectation des ressources humaines et financières de manière à obtenir des résultats optimaux.

Les entreprises sont aujourd'hui appelées à gérer de grandes quantités de données glanées au travers de leurs contacts avec des clients effectifs ou potentiels, dans un souci de plus grande efficacité opérationnelle et d'élargissement de leur marché. Les détaillants en ligne qui prennent part aux échanges électroniques actualisent et analysent continuellement les données sur le comportement des consommateurs, pour connaître leurs préférences de consommation latentes. Ces renseignements leur serviront à définir ce qu'ils commercialisent et comment. Par le truchement de ces informations, ils peuvent en outre déterminer les risques de crédit à la consommation et réduire ainsi le risque de perte de revenus. L'utilisation généralisée des moyens de communication est à l'origine d'une forte augmentation des transactions impliquant la vente en ligne de biens de consommation et de l'intensification de la fréquence et du volume des données échangées entre l'entreprise et le consommateur (B2C).

La prépondérance des périphériques électroniques connectés à l'internet (*smartphones*, tablettes, téléviseurs intelligents, portables, systèmes d'infodivertissements embarqués, etc.) explique l'omniprésence des mégadonnées. Chaque seconde, les réseaux sociaux produisent des montagnes de données, dont on ne pourra endiguer le flot que si tous les utilisateurs arrêtent en même temps d'envoyer, de partager et d'afficher les communications allant du sms aux vidéos transmises par internet. La technologie de l'internet des objets inclut de nombreuses catégories de biens physiques (appareils ménagers, caméras de sécurité, poubelles, ...) dans les applications de mégadonnées. Pour ce qui concerne l'échange de marchandises, les acteurs (fabricants, expéditeurs et opérateurs logistiques), s'efforcent de conférer une utilité pratique à ce riche éventail de données (allant de l'historique des transactions individuelles à la localisation des marchandises conteneurisées), afin d'offrir un service de qualité et d'améliorer la connectivité, ce qui devrait avoir une influence sur la chaîne logistique.

Dans le présent document, nous nous proposons d'examiner les répercussions des mégadonnées sur la Douane et, notamment, sur la gestion des risques. Pour que leurs décisions soient plus éclairées et adéquates, certaines Administrations de la douane ont déjà pris des mesures concernant les mégadonnées, en tirant parti des avantages qu'offre l'analytique, tout en veillant à la qualité des données (relatives aux marchandises, aux envois et aux moyens de transport) et en élargissant la portée des données qu'elles peuvent utiliser à des fins analytiques. Ces initiatives sont exposées dans le présent document, en fonction des renseignements obtenus auprès des cinq Administrations douanières suivantes: Agence des services frontaliers du Canada (ASFC), *Customs and Excise Department*, Hong Kong, Chine (Douane de Hong Kong, Chine); *New Zealand Customs Service* (Douane de Nouvelle-Zélande); *Her Majesty's Revenue and Custom* (HMRC), Royaume-Uni; *U.S. Customs and Border Protection* (USCBP).

II. Comprendre les mégadonnées

Le terme "mégadonnées" désigne d'énormes ensembles de données dont la gestion et le traitement ne peuvent être assurés par des modèles traditionnels, dans des délais acceptables. Bien que cette définition soit souvent proposée, il n'existe aucune définition

communément admise.¹ Notons, cependant, la définition intéressante qu'en donne Gartner², selon qui elles constituent « un actif informationnel se distinguant par des volumes considérables, une grande vélocité et sa variété et auquel il convient d'appliquer des modèles rentables et novateurs de traitement de l'information, afin d'en arriver à une meilleure compréhension et d'améliorer ainsi le processus décisionnel, en poussant plus avant l'automatisation des procédures.»³ Voici qui met en avant de manière assez concise plusieurs traits distinctifs de ces données, souvent appelés les "3V" des mégadonnées: Volume, Vélocité et Variété.

Par "volume", l'on entend généralement la taille et l'envergure des jeux de données individuels, mais le terme renvoie également souvent à la quantité totale de données à l'échelle planétaire. Une étude montre qu'en 2012, près de 2,8 zettaoctets (ZB) de données ont été générées et reproduites mais, selon ses prévisions, d'ici 2020, l'on devrait atteindre un volume de 44 ZB.⁴ La « vélocité » désigne, à la fois, la vitesse et la fréquence de création, d'actualisation et de traitement des données. Quant à la « variété », ce terme renvoie à la diversité des formats, des sens, des origines et des moyens. En effet, les mégadonnées peuvent provenir de nombreuses sources et il semblerait qu'au moins quatre-vingt pourcent des données à valeur commerciale ne sont pas structurées⁵.

| Encadré 1. Les mégadonnées en "3 V" | | | | | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----------------|-------------------------------------|
| Volume | | | | | | |
| <ul style="list-style-type: none"> De 2005 à 2020, l'univers numérique aura crû d'un facteur 300, pour passer de 130 à 40.000 exaoctets (EB), soit 40.000 milliards de gigaoctets (plus de 5200 gigaoctets par personne, hommes, femmes et enfants), en 2020. D'ici là, l'univers numérique aura donc doublé de taille tous les deux ans.⁶ | | | | | | |
| Système international d'unités (SI) | | | La Commission électrotechnique internationale (CEI) a approuvé les préfixes pour les multiples binaires (aux fins du traitement et de la transmission des données) | | | |
| kilo | k | 10 ³ | kibi | ki | 2 ¹⁰ | = 1024 |
| mega | M | 10 ⁶ | mebi | Mi | 2 ²⁰ | = 1 048 576 |
| giga | G | 10 ⁹ | gibi | Gi | 2 ³⁰ | = 1 073 741 824 |
| tera | T | 10 ¹² | tebi | Ti | 2 ⁴⁰ | = 1 099 511 627 776 |
| peta | P | 10 ¹⁵ | peti | Pi | 2 ⁵⁰ | = 1 125 899 906 842 624 |
| exa | E | 10 ¹⁸ | exbi | Ei | 2 ⁶⁰ | = 1 152 921 504 606 846 976 |
| zetta | Z | 10 ²¹ | zebi | Zi | 2 ⁷⁰ | = 1 180 591 620 717 411 303 424 |
| yotta | Y | 10 ²⁴ | yobi | Yi | 2 ⁸⁰ | = 1 208 925 819 614 629 174 706 176 |
| <ul style="list-style-type: none"> Un facteur important se cache derrière l'expansion de cet univers numérique: la croissance des données produites par machine, dont la part en 2005 représentait 11 % de l'univers numérique, mais qui dépassera la barre des 40 % d'ici 2020.⁷ L'internet des objets est largement responsable de cette expansion puisque, chaque seconde, aux quatre coins du monde, les capteurs dont sont équipés bon nombre de dispositifs génèrent des données⁸. | | | | | | |

¹ Cleary (2017).

² Entreprise de marketing informatique dont le siège social est à Stamford, Connecticut (États-Unis).

³ Gartner – 'IT Glossary', <http://www.gartner.com/it-glossary/big-data> (dernière consultation: 7 mars 2017).

⁴ EMC, Communiqué de presse, <https://www.emc.com/about/news/press/2012/20121211-01.htm> (dernière consultation: 7 mars 2017).

⁵ Id.

⁶ EMC (2012).

⁷ Id.

Vélocité

- Chaque minute, 100 heures de vidéos sont téléchargées sur YouTube, plus de 200 millions de courriers électroniques sont échangés, près de 20 millions de photos sont visualisées et 30.000 d'entre elles sont téléchargées sur Flickr. Chaque minute, près de 300 000 tweets sont envoyés et échangés et environ 2,5 millions de recherches sont effectuées sur Google.⁹
- La vélocité désigne le rythme de production plus rapide des données et le renforcement nécessaire des performances des entreprises qui en dépendent. L'enjeu n'est donc pas uniquement de stocker des flux de données rapides, mais également de faire de ces données arrivant en flux rapides une ressource qui stimule l'innovation et améliore la prise de décisions.¹⁰

Variété

- Dans le passé, toutes les données étaient structurées et soigneusement ordonnées en colonnes et en rangées. Au contraire, de nos jours, 90 % des données ne sont plus structurées¹¹, soit par faute de modèle de données prédéfini, soit parce qu'elles sont organisées selon une structure prédéfinie.¹²
- Il est indispensable d'arriver à gérer ces données non structurées et à en tirer des informations, pour un déploiement efficace des mégadonnées.¹³

Un quatrième V s'ajoute parfois aux 3 premiers: la « véracité ». Ce concept désigne des données classiques que l'on peut généralement obtenir auprès de sources connues et qui sont préparées de telle sorte qu'elles satisfont aux besoins des utilisateurs. À l'inverse, de par leur nature, les mégadonnées sont généralement incomplètes ou imparfaites et, dès lors, elles peuvent être aléatoires ou sources d'erreurs.¹⁴ De plus, il est peu probable que les utilisateurs suspectent l'origine douteuse de données obtenues auprès de sources peu fiables ou qui sont passées entre des mains inconnues.

Les mégadonnées peuvent, dans une certaine mesure, être associées à l'internet des objets (connexion en réseau d'objets physiques qui ne sont pas des ordinateurs au sens classique du terme). Des bornes d'information ont été installées sur divers dispositifs, dont des appareils domestiques et sur des infrastructures collectives, pour se transformer ainsi, en raison de la fréquence de leur utilisation, en fournisseurs de données brutes ou initiales qui peuvent faciliter certaines opérations commerciales publiques ou privées. Si ces objets de la vie quotidienne pouvaient se doter d'une identité numérique, les avantages qu'ils présentent et leur utilité propre n'en seraient que plus grands et le monde matériel deviendrait un gigantesque système d'information.¹⁵

Le fait qu'il existe des mégadonnées pose la question de l'existence de « petites données » et de leur nature. Pour citer un juriste, « Invariablement ou presque, les mégadonnées excluent explicitement ou implicitement la capacité humaine à stocker et à traiter. Par conséquent, lorsqu'un être humain comprend des données sans l'aide de l'informatique ou d'algorithmiques, c'est qu'il ne s'agit pas de mégadonnées ». Et d'ajouter, « Dès lors, dans le

⁸ Rijmenam (2013).

⁹ Id.

¹⁰ Adamov (2015).

¹¹ Rijmenam (2013).

¹² Gerasimou (2016).

¹³ Id.

¹⁴ Cf. AAAS-FBI-UNICRI (2014), at 8.

¹⁵ Cf. Techopedia, « Internet of Things (IoT) », <https://www.techopedia.com/definition/28247/internet-of-things-iot> (dernière consultation: 7 mars 2017).

monde des petites données, l'être humain a la capacité de créer et de comprendre en faisant appel à son seul discernement ». ¹⁶ Les mégadonnées sont donc des données dont on ne peut tirer le sens que si l'on dispose d'une certaine capacité de gestion des données et des outils analytiques nécessaires.

En résumé, les mégadonnées sont simplement des données massives, qui n'offrent pas une solution en soi. De plus, sans les applications qui les accompagnent, elles n'ont aucune valeur intrinsèque. Elles ne verront le jour et ne pourront être mises en œuvre qu'à partir du moment où certaines tâches exigeant une décision éclairée seront identifiées.

| |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Encadré 2. Agence des services frontaliers du Canada (ASFC), 1^{ère} partie ¹⁷</p> <p>Les mégadonnées et l'ASFC</p> <ul style="list-style-type: none">• L'ASFC recueille des données auprès de sources internes et externes. Les données réunies en son sein proviennent des flux vidéo des télévisions en circuit fermé (CCTV). Quant aux informations de sources externes, l'ASFC tire parti des données de sources ouvertes et des renseignements sur des situations données.• Les travaux menés par l'ASFC dans le domaine des mégadonnées ont porté essentiellement sur l'analyse de gros volumes de données structurées, alors que des efforts sont également consentis pour résoudre des problèmes complexes, aligner les capacités analytiques de l'ASFC sur la définition plus large des mégadonnées, ainsi que sur leur vélocité et leur variété. |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

III. B.a-ba de l'analyse des mégadonnées

Les gouvernements et les acteurs du secteur privé utilisent les mégadonnées à des fins organisationnelles diverses: nouvelles connaissances, création de valeur, promotion de nouveaux modèles opérationnels, processus décisionnel informé et autres changements ultérieurs possibles. Cela étant dit, à elles seules, les mégadonnées n'offrent aucune valeur ajoutée. Un rapport remis dernièrement au Président des États-Unis en fait une description qui mérite d'être signalée: « Les mégadonnées sont 'massives' à deux titres. Elles le sont de par leur quantité et leur diversité, d'une part et, d'autre part en raison de l'envergure de l'analyse (appelée « analytique ») à laquelle elles peuvent être soumises, pour en arriver finalement à des inférences et des conclusions. » ¹⁸ Davenport (2014) déclare que les attitudes et activités informatiques au sein des grandes organisations doivent faire l'objet d'un remaniement en profondeur. À ses yeux, « L'on entend souvent dire que les analystes passent de 75 à 80 pourcent de leur temps à rechercher des données, les toiletter et les préparer en vue de leur analyse » et d'ajouter que « Cela se vérifie aussi bien dans les entreprises qu'au sein des structures gouvernementales. » Il fait aussi référence à l'initiative stratégique d'un dirigeant du ministère américain de la défense (« Des données aux décisions »), qui déclare ceci « À défaut d'améliorations spectaculaires, les volumes de données qui seront produits par la prochaine génération de capteurs et la complexité des systèmes intégrés dépasseront de beaucoup la capacité des analystes à les exploiter. » ¹⁹ Pour pouvoir utiliser les données disponibles, les organisations devront donc se doter de techniques d'analyse adéquates (analyse textuelle, exploration des données, statistiques et traitement du langage naturel) et des capacités qui les accompagnent.

¹⁶ Hu (2015), at 798.

¹⁷ Renseignements fournis par écrit par le Canada.

¹⁸ PCAST Report (2014), at ix.

¹⁹ Davenport (2014), at 19.

L'être humain ne pouvant, à lui seul, traiter et comprendre les mégadonnées, il faudra faire appel à l'apprentissage automatique, méthode d'analyse des données qui automatise la production de modèles analytiques. Grâce à des algorithmes²⁰ qui apprennent de manière itérative à partir de données, par l'apprentissage automatique, les ordinateurs sont capables de dénicher des informations « cachées », sans avoir été expressément programmés à cette fin.²¹ Mayer-Schönberger et Cukier (2013) estiment qu'en matière de mégadonnées, l'idée n'est pas d'amener les ordinateurs à raisonner comme l'être humain, mais à "appliquer les règles mathématiques" à des masses de données, pour calculer la probabilité qu'un courrier électronique soit un *spam* (pourriel), que certains mots soient mal orthographiés, ou qu'une personne qui traverse la rue à un endroit non autorisé y parvienne sans se faire renverser par un véhicule, en influant sur le fonctionnement du système de pilotage automatique dudit véhicule. Les programmes dotés de telles fonctions « doivent leur performance à la quantité de données sur lesquelles reposent leurs prévisions » et « ils sont conçus de façon à s'améliorer au fil du temps, en fonction des meilleurs signaux et modèles à rechercher lors de la saisie de nouvelles données », expliquent encore les auteurs.²²

Les techniques d'apprentissage automatique font très souvent partie des exercices d'exploration des données, mais ces deux concepts n'en sont pas moins fort différents. L'exploration des données, que l'on assimile souvent à un processus de découverte, consiste à analyser un ensemble de données sous des angles différents et à les réduire à des informations utiles. L'apprentissage automatique se centre plutôt sur l'élaboration de prévisions à partir de tendances et de modèles identifiés par l'analytique de données.²³ Outre sa fonction prévisionnelle, l'apprentissage automatique est également utilisé dans l'exploration des données, par des associations et classifications libres du système, qui permettent de donner du sens à ce déluge de données et d'en tirer des informations. La forme des mégadonnées peut varier du message sms au film et les données peuvent être soit brutes, soit organisées. Vu les défis que pose l'exploration de masses aussi complexes de données, la puissance des algorithmes d'apprentissage automatique est importante.²⁴

IV. Considérations politiques sur les mégadonnées

Les mégadonnées peuvent être utilisées dans les politiques publiques, à des fins aussi diverses que la détection et la prévention de la fraude fiscale ou l'évaluation de la menace en cas de durcissement de mesures de sécurité. Toutefois, les décideurs et les administrateurs doivent être conscients des questions transversales, que sont la protection de la vie privée et la sécurité des données, qui se posent dans ce domaine.

Selon Mayer-Schönberger et Cukier (2013), « Les mégadonnées permettent de nous surveiller de plus près, tout en rendant caducs certains instruments juridiques de protection de la vie privée » et « Elles privent également d'effet les principaux dispositifs permettant de préserver l'anonymat ».²⁵ Une personne serait à n'en pas douter mal à l'aise, voire même effrayée, si ses données personnelles pouvaient être utilisées à des fins inconnues. Voilà pourquoi les

²⁰ Les algorithmes sont déjà utilisés dans la pratique et se sont immiscés dans notre quotidien (moteur de recommandation de sites web, système de reconnaissance vocale, voitures sans chauffeur, etc.)

²¹ SAS, 'Machine Learning: What it is & why it matters', http://www.sas.com/en_us/insights/analytics/machine-learning.html (dernière consultation: 7 mars 2017).

²² Mayer-Schönberger et Cukier (2013), at 12.

²³ McNulty (2014).

²⁴ Oberlin (2012), 10^{ème} paragraphe

²⁵ Mayer-Schönberger et Cukier (2013), at 170.

gouvernements qui sollicitent des renseignements sur l'identité et les activités des personnes doivent prendre des mesures concrètes pour que les informations ne soient utilisées qu'à des fins légitimes et qu'elles ne soient pas transmises à des tiers. Les dispositions légales (dont les directives et/ou les codes disciplinaires régissant le traitement des données, le devoir de confidentialité, le partenariat public-privé, la coopération intergouvernementale, etc.) devraient être pertinentes et des solutions techniques doivent être envisagées, mais, les questions de confidentialité pourraient entraver la facilitation des échanges de données nécessaires par les services gouvernementaux.²⁶

En renforçant la connectivité avec d'autres organes, au titre de la compilation de données, un service pourra prendre des décisions en connaissance de cause, sur la base de données probantes. Néanmoins, la création de canaux externes permanents pourrait menacer certains paramètres de sécurité des ordinateurs et des systèmes de réseau, réduisant d'autant la protection contre les éventuels actes posés par des pirates informatiques en quête de renseignements confidentiels.

La perturbation possible des dispositifs de traitement de l'information d'une organisation suite à un acte de cybercriminalité (ou une cyberattaque) est également un problème pressant, qui pourrait bloquer purement et simplement les activités d'une entreprise. L'attaque de « déni de service distribué » (*distributed denial-of-service* ou DDoS) en est un exemple typique. Elle consiste à inonder un serveur informatique de milliers ou de millions de demandes inutiles qui l'empêche de fournir les services escomptés aux destinataires légitimes. Ces attaques numériques massives recourent à un "botnet", qui exploite un réseau de terminaux d'ordinateurs (situés dans des pays différents) infectés par un logiciel malveillant (virus informatique, logiciel espion, etc.) en attente d'un ordre donné à distance par l'auteur de l'attaque. Les progrès technologiques réalisés dans le cadre de l'internet des objets permettent aujourd'hui de créer plus aisément de plus grand botnets, puisqu'il est possible « d'asservir » un plus grand nombre d'appareils connectés, tels que des routeurs, des caméras de vidéosurveillance, des thermostats, des téléviseurs en réseau et des véhicules intelligents.²⁷

Selon TechTarget²⁸ (2016), les périphériques de l'internet des objets peuvent faire office de capteurs de données (ils envoient des informations à un service central et reçoivent en retour des instructions pour en adapter l'utilisation) et interagir avec d'autres appareils, pour appliquer des algorithmes localisés d'intelligence distribuée²⁹ qui répondront à leur environnement, sans forcément passer par un service centralisé. TechTarget recommande vivement aux entreprises d'authentifier chacun des périphériques connus de l'internet des objets, ainsi que l'origine des logiciels dont ils sont équipés. Les entreprises devraient également contrôler l'accès aux périphériques et concevoir des outils d'analyse des logiciels, pour y détecter d'éventuelles anomalies. Par ces mesures, elles pourraient atténuer le risque qu'un pirate mal intentionné usurpe l'identité d'un périphérique de l'internet des objets, pour orchestrer une attaque DDoS sur ce réseau.³⁰

²⁶ L'ASFC estime qu'en ce qui concerne les mégadonnées, même si l'accès opportun aux données est un problème majeur, l'un des principaux défis posés par l'acquisition des données réside dans les limites légales imposées au profit du respect de la vie privée par les différentes juridictions, qui pourraient compliquer la reddition de comptes horizontale. Renseignements fournis par écrit par le Canada.

²⁷ Naughton (2016).

²⁸ Entreprise de solutions technologiques dont le siège social est à Newton, Massachusetts (États-Unis).

²⁹ « L'intelligence distribuée » est un terme d'intelligence artificielle qui désigne le comportement collectif de systèmes décentralisés et autogérés.

³⁰ TechTarget (2016).

V. Les mégadonnées du point de vue de la Douane

Plusieurs Administrations douanières ont arrêté leur politique en matière de mégadonnées. La Douane des États-Unis (USCBP) a décidé d'appliquer les technologies de mégadonnées à ses propres données, en un lieu centralisé. À l'heure actuelle, la compilation des données, la protection de la vie privée et les contrôles sécuritaires adéquats sont une véritable gageure. L'USCBP est consciente du fait que les technologies applicables aux mégadonnées sont encore en phase de maturation et s'attend, par conséquent, à pouvoir en exploiter le plein potentiel dès que les investissements adéquats seront consentis. De plus, le plus gros avantage que présente l'application pratique des mégadonnées aux yeux des services douaniers des États-Unis réside dans la possibilité de mobiliser les diverses banques de mégadonnées existantes et d'en permettre l'utilisation pour les besoins analytiques, à des conditions plus abordables que si l'on avait eu recours à des techniques traditionnelles de stockage et à des méthodes informatiques.³¹

S'agissant des renseignements concernant le fret, par exemple, le format de données doit être adapté à l'utilisation qu'en fera la Douane. Conformément aux dispositions du Cadre de normes SAFE, « Si l'EDI, qui utilise la norme internationale UN/EDIFACT, reste le mode le plus employé, la douane devra également recourir à d'autres possibilités (langage XML, par exemple ». « Il pourrait même être envisagé d'utiliser le courrier électronique ou la télécopie ». ³² Les données au format XML ou de courrier électronique sont généralement considérées comme semi-structurées. Les mégadonnées n'englobent pas uniquement les données semi-structurées, mais aussi la plupart des autres types de données, structurées ou non. L'on pourrait, dès lors, parler de « données multimédia », concept qui renvoie aux données textuelles, audio, vidéo, animations et autres. Les Administrations de la douane doivent donc acquérir les compétences nécessaires pour recevoir, stocker et traiter des données très hétérogènes qui pourraient ensuite être rassemblées à des fins d'analyse.

D'éminents experts internes ou externes devraient aider les services douaniers à mettre en œuvre une politique de mégadonnées et il serait bon d'envisager de s'adjoindre les services de spécialistes dans les domaines concernés (parmi lesquels la cybersécurité et la préparation aux incidents). Le gouvernement de Nouvelle-Zélande, la Douane, ses partenaires présents aux frontières et le ministère du secteur primaire (*Ministry of Primary Industries* ou MPI) se sont lancés dans un ambitieux projet informatique, dans le cadre duquel les outils analytiques ont été utilisés à des fins d'évaluation des risques et de gestion des frontières. Bien que ces instances doivent encore renforcer leurs capacités dans ce domaine, la Douane néo-zélandaise estime qu'il est primordial de tisser de solides liens entre les experts en informatique qui l'ont dotée de ces capacités (par des travaux d'expérimentation hors ligne destinés à tester les modèles). De plus, elle dispose de la technologie nécessaire pour soutenir l'équipe en charge de l'analytique, à laquelle participent des fonctionnaires de la Douane, des services d'immigration et du MPI. Cette équipe centralisée procède actuellement à l'alignement des ensembles de données des différents services et à la conception de modèles de ciblage. Simultanément, elle passe en revue les données existantes, afin de dégager des schémas et des connections qui pourront fournir de plus amples renseignements aux services douaniers et les aider à prendre les décisions nécessaires. L'Administration néo-zélandaise est aujourd'hui convaincue que la collaboration ininterrompue de l'équipe analytique centralisée (dirigeant/membre) avec ses

³¹ Renseignements fournis par écrit par les États-Unis.

³² Cadre de normes SAFE de l'OMD, juin 2015, III. 2.6.4.

services d'information a joué un rôle déterminant dans la fixation des priorités en matière de critères applicables aux données analytiques.³³

Encadré 3. Administration de la douane de Nouvelle-Zélande³⁴

Système de gestion conjointe des frontières (*Joint Border Management System* ou *JBMS*)³⁵

- Dès qu'il sera appliqué, le JBMS permettra aux services douaniers de Nouvelle-Zélande d'utiliser l'analytique au profit d'un ciblage plus fin et plus pointu des échanges en temps réel. L'Administration développe le JBMS en collaboration avec le MPI qui exploite un « Guichet commercial unique », remplace les technologies vieillissantes, synchronise les systèmes informatiques des différents services et propose de meilleurs outils de gestion des risques et du renseignement.
- A ce stade, des progrès ont déjà été réalisés: un système de portail unique a été instauré, par l'intermédiaire duquel les opérateurs commerciaux peuvent enregistrer leurs données (guichet commercial unique). L'infrastructure de base a été améliorée et les processus de gestion des « voyageurs » du MPI ont été intégrés au système douanier central (le « CusMod »).
- À l'aide des outils de gestion des risques et de renseignement, la Douane néo-zélandaise est plus à même d'anticiper les menaces, de surveiller les tendances et de cibler les transactions à haut risque. Ces outils ont également servi à l'établissement d'un groupe conjoint d'analytique douanière qui travaille actuellement hors ligne sur les données et jette les bases du déploiement opérationnel des modèles et règles qui en découleront. D'autres outils viendront s'y ajouter entre 2017 et 2019, pour consolider les fonctionnalités en temps réel/en ligne.

Perspectives de changement

- La vision de la Douane de Nouvelle-Zélande à l'horizon 2020 (et au-delà) devrait lui permettre de devenir un service pleinement centré sur le renseignement. L'analytique (qui lui servira à étoffer ses connaissances et prendre des décisions éclairées) est indispensable au remaniement des méthodes de travail et, grâce aux connaissances engrangées, cette vision pourra se concrétiser.
- L'analytique pourrait influencer (favorablement) la manière dont l'Administration assume ses fonctions de base et, si cela se confirme, l'analyse prédictive influera sur la prise de décisions stratégiques et opérationnelles par l'Administration, ainsi que sur sa politique d'affectation des ressources (pour ce qui concerne plus particulièrement le renseignement, les enquêtes, les assurances recettes et les services de première ligne).
- L'analytique peut également être utilisée pour améliorer le service à la clientèle et sa conception. Elle pourrait donc transformer l'ensemble des services douaniers.

Les différents département/services gouvernementaux pourraient tenir leurs données respectives à jour et les utiliser uniquement en interne, sans les partager avec d'autres unités et régions. La Douane de Hong Kong, Chine a fait des efforts remarquables en la matière et a décidé de "centraliser" et de "consolider" toutes les données utiles qui, à défaut, seraient disparates et donc moins pertinentes. Depuis 2012, un référentiel d'information centralisé a été instauré. Il s'apparente à un entrepôt centralisé de données où sont stockées les données

³³ Renseignements fournis par écrit par la Nouvelle-Zélande.

³⁴ Id.

³⁵ Les dernières informations sont disponibles à l'adresse suivante <http://www.customs.govt.nz/features/jointbordermanagement/pages/default.aspx> (dernière consultation: 7 mars 2017).

opérationnelles transmises par neuf canaux d'information différents, ayant tous des fins réglementaires qui leur sont propres.³⁶

Encadré 4. Département des douanes et accises, Hong Kong, Chine³⁷

Référentiel centralisé d'information

- Le Référentiel central d'information de la Douane de Hong Kong, Chine est un entrepôt de données dans lequel sont stockées et traitées des données qui se distinguent par leur côté volumineux, leur variété et leur vélocité. En 2012, l'entrepôt traitait 3 téraoctets de données, volume qui a quadruplé pour atteindre 12 téraoctets en 2015. Plus cet entrepôt traite de données, plus il s'apparente à un outil de gestion des mégadonnées.
- L'entrepôt propose à ses usagers une plateforme unifiée grâce à laquelle, par l'analyse croisée des données, ils peuvent obtenir des informations et découvrir des tendances dans bien des domaines. Les données peuvent être conservées pendant sept ans au maximum, ce qui donne aux utilisateurs le loisir d'établir un plan à long terme et de programmer leurs opérations.
- L'administrateur opérationnel du système fournit l'infrastructure qui autorise le partage opportun et efficace des données entre les différents systèmes de l'Administration. Il uniformise le mécanisme d'interface entre systèmes, sur la base d'un modèle de données harmonisées. Le coût de la maintenance du système s'en trouve réduit et l'assistance est plus efficace.

Application de mégadonnées

- Outre les données capturées par les services douaniers dans le cadre de leurs activités réglementaires, l'entrepôt de données abrite des données transférées à partir d'autres services gouvernementaux (dont des informations concernant l'enregistrement des sociétés provenant du ministère des finances ou les autorisations de fermeture de routes détenues par le ministère des transports), à des fins de vérification.
- Ce faisant, le système améliore la qualité des données et fait fonction de plateforme d'analyse unique. Par ce moyen, les statistiques sont plus accessibles aux utilisateurs de tous niveaux et les recherches deviennent plus efficaces.
- L'entrepôt produit régulièrement des rapports en ligne que les utilisateurs peuvent immédiatement consulter et dans lesquels ils trouvent l'information la plus récente et la plus précise. Les rapports peuvent adopter différents formats qui en facilitent l'analyse et l'interface peut être personnalisée en fonction des préférences de l'utilisateur.
- Sur la foi des informations disponibles, l'Administration pourra, non seulement, bénéficier de renseignements qui lui permettront de prendre des décisions immédiates au titre de la lutte contre la fraude (gestion des risques, planification opérationnelle, affectation de ressources), mais également d'assurer la gestion organisationnelle et la planification stratégique à long terme.
- L'Administration envisage de largement automatiser l'exploration des données en vue des progrès à venir et devra inscrire ce point sur la liste des éléments à étudier dans le cadre du renforcement et de l'optimisation de l'efficacité opérationnelle.

L'application systématique des mégadonnées pourrait faire apparaître les problèmes chroniques qui affectent la Douane depuis des années. La Douane du Royaume-Uni (*Her Majesty's Revenue and Customs* ou *HMRC*) a mené une réflexion qui l'a amenée à découvrir qu'elle s'était peut-être reposée sur des données inexactes, incomplètes ou trompeuses (dont des données obtenues auprès des mauvaises sources et/ou au mauvais moment). Elle a donc pris les mesures nécessaires pour obtenir les régimes d'assurance adéquats et identifier les

³⁶ Renseignements fournis par écrit par Hong Kong, Chine.

³⁷ Id.

bonnes sources, afin de garantir l'exactitude et l'intégrité des données auxquelles elle a accès et d'éviter que naisse un sentiment déplacé de sécurité.³⁸

Encadré 5. Her Majesty's Revenue and Customs (HMRC)³⁹

Fournisseurs de systèmes collectifs (Community System Providers ou CSP)

- Suite à l'apparition et au développement des fournisseurs de systèmes collectifs (systèmes commerciaux introduits dans les ports et aéroports), le RU a dû faire face au traitement d'énormes volumes de données liés à la circulation du fret commercial et aux mouvements physiques que cela implique avant le dédouanement.
- Le modèle britannique est peu commun, même au sein de l'UE. Le HMRC travaille en symbiose avec les opérateurs commerciaux et a ainsi pu introduire des systèmes de ciblage et de lutte contre la contrebande dans les procédures commerciales, de manière à bénéficier d'une vue plus globale des flux de données utiles.
- Le concept a même été étendu aux contrôles à l'entrée et à la sortie, qui font partie intégrante des systèmes des opérateurs économiques.⁴⁰

Le Projet CORE

- David Hesketh, Directeur Recherche et développement douaniers auprès du HMRC chargé du Projet CORE, en explique la genèse: au vu de l'expérience acquise grâce aux projets antérieurs qui avaient été mis en œuvre pour palier « l'imprécision des données », « les solutions de visibilité de la chaîne logistique commerciale » consistent « [souvent en] des systèmes électroniques de données en ligne, auxquels peuvent être associés le vendeur, l'acheteur et toute autre partie intéressée, ainsi que les autorités de contrôle ». ⁴¹
- Des systèmes intégrés de transmission constante de données ont été introduits, afin d'acheminer les données pertinentes au bon moment, depuis la bonne source. Les travaux de R&D actuellement menés dans le cadre de ce projet visent à assurer un accès aux flux de données commerciales à un stade beaucoup plus précoce de la chaîne logistique, pour renforcer la capacité de ciblage et offrir des assurances plus solides aux entreprises respectueuses de la réglementation.
- Le projet part du principe selon lequel, si le gouvernement peut répercuter les bénéfices sur les fournisseurs de données, les volumes de données correctes augmenteront tout autant que la capacité à supprimer le risque attaché aux entreprises participantes et à créer des modèles de conformité. Il s'agit ici d'un exemple de ce qui est possible lorsque l'on dispose des « bonnes » données.
- Selon le HMRC, les ensembles de données ainsi créés respectent les normes internationales et les principes régissant les mégadonnées ne s'appliqueront pas si les données sont incorrectes, si elles arrivent trop tard, en provenance de sources inappropriées.

Baida (2016) considère que « les données sont un moyen plutôt qu'une fin en soi »; l'information dont on tire des connaissances sur lesquelles peuvent se fonder des décisions informées « est faite de données qui ont été traitées, organisées, structurées ou présentées dans un contexte pertinent » et la « numérisation » que constituent les nuées de capteurs et de technologies dites intelligentes a provoqué l'explosion des données disponibles.⁴²

³⁸ Renseignements fournis par écrit par le Royaume-Uni.

³⁹ Id.

⁴⁰ Les opérateurs économiques sont représentés par *Fast Parcels Operators* (opérateurs commerciaux autorisés par le HMRC à effectuer certaines procédures douanières (déclarations) au nom des importateurs/exportateurs.

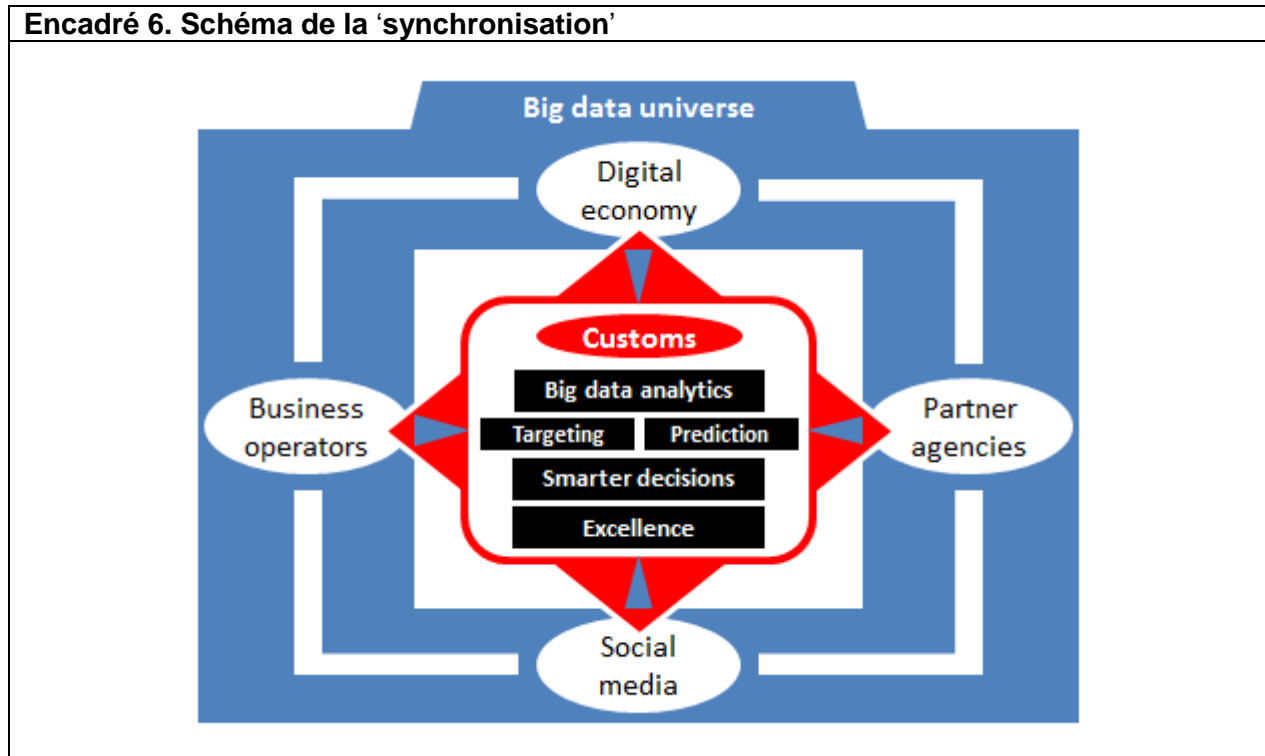
⁴¹ 2^{ème} bulletin d'information du projet Core, <http://www.coreproject.eu/newsletters/core-2nd-newsletter-august-2015/seamless-integrated-data-pipelines-by-david-hesketh-hmrc.aspx> (dernière consultation: 7 mars 2017).

⁴² Baida (2016), at 30.

Généralement parlant, les Administrations de la douane ne sont pas tout à fait prêtes à utiliser ce genre de données pour en extraire des informations qui pourraient améliorer le processus décisionnel. Cela peut s'expliquer par le fait qu'habituellement, elles s'en réfèrent à des données « internes » qu'elles rassemblent à partir des déclarations de marchandises ou d'interventions régulières, pour les passer au tamis et en retirer les principaux éléments d'information.

En matière de mégadonnées, il convient, pour commencer, de déterminer comment et dans quelle mesure diversifier les types et les sources de données, pour constituer un riche répertoire interne. Plus la Douane disposera de données, plus elle sera encline à combiner et à corrélérer différentes catégories de données, dans le but d'identifier des tendances et schémas inconnus parmi les marchandises, moyens de transport et personnes que ciblent les contrôles et pour en vérifier et en confirmer l'exactitude. Plus vite les données seront mises à jour par la Douane, plus celle-ci sera à même de repérer d'éventuels cas de fraude ou des irrégularités et d'y réagir en temps opportun. Fondamentalement, pour que la Douane soit à même d'exploiter pratiquement les mégadonnées, il suffirait qu'elle tire avantage de leurs particularités.

L'acquisition ou « l'internalisation » des mégadonnées à des fins douanières peut se faire en adaptant les activités des services douaniers au dynamisme de ce monde des données massives, qui, aux yeux de la Douane, peut passer pour un phénomène externe en évolution permanente, mais n'en a pas moins été adopté par d'autres services (services douaniers ou leurs partenaires) ailleurs dans le monde. Un lien 'symbiotique' (aux dires du HMRC) s'est tissé entre la Douane et les opérateurs privés (qu'il s'agisse d'un transporteur habituel ou d'un service de réseautage) et, grâce à lui, de nombreuses données devraient être libérées à partir de sources multiples. Une telle infrastructure d'information ne peut s'ériger que sur de solides fondements juridiques, surtout lorsqu'il est question de confidentialité et de cybersécurité. Il faut également pouvoir compter sur les ressources humaines nécessaires, qui doivent donc être un élément important de toutes les capacités techniques à acquérir par l'Administration de la douane.



VI. Les mégadonnées: défis et opportunités

Pour pouvoir être en phase avec les progrès et l'évolution du système d'analyse des données, la Douane doit relever bien des défis. L'Agence des services frontaliers du Canada (ASFC) a soumis ses structures internes à évaluation, pour s'assurer d'une exploitation optimale des données disponibles. L'Administration estime que, pour tirer parti des mégadonnées, elle doit surmonter des obstacles liés au développement des ressources humaines, à la gestion des données et à l'architecture du système. Commençons par le développement des ressources humaines: la Douane canadienne soutient que, pour utiliser au mieux toutes les possibilités qu'offrent les mégadonnées, il faut veiller au développement stratégique des ressources humaines et partir à la recherche des analystes de données les plus prometteurs du marché, les embaucher très tôt dans leur carrière, de manière à les former professionnellement au sein de l'organisation. Toutefois, l'ASFC ne peut se contenter d'experts internes; elle doit également renforcer les compétences et capacités de l'ensemble du personnel, au moyen de formations continuées et d'un accès routinier aux outils analytiques. L'Administration admet également que les données doivent être gérées avec plus d'efficacité et envisage de créer, en son sein, une fonction centralisée, spécialisée dans la coordination et la facilitation de l'acquisition des données et des systèmes. Pour renforcer sa capacité analytique, l'ASFC juge également nécessaire de poursuivre l'intégration des ensembles de données à tous les niveaux de sa structure.⁴³

Encadré 7. Agence des services frontaliers du Canada (ASFC), 2^{ème} partie⁴⁴

Voies d'avenir

- L'ASFC recourt actuellement à un entrepôt de données d'entreprises (*Enterprise Data Warehouse* ou EDW) où sont stockées et traitées les mégadonnées qu'elle considère comme essentielles. Elle estime qu'il faut tenter de minimiser les obstacles et d'étendre l'accès à ces services.
- Le développement de l'EDW se poursuit, en vue de l'inclusion de nouvelles sources et d'autres types d'informations. Parmi les derniers éléments ajoutés figurent des ensembles de données de source ouverte, qui renforceront la capacité de la Douane à évaluer les risques posés par les marchandises et les personnes. D'autres projets, qui en sont encore à un stade exploratoire, portent sur les contrôles biométriques, la reconnaissance faciale, les systèmes automatisés de détection des mensonges et la modélisation prédictive. Ces fonctions seront rassemblées dans la procédure de contrôle des voyageurs et alimenteront les mégadonnées qui sont l'empreinte de ces derniers.
- En enregistrant de vastes ensembles de données dans l'EDW et en élargissant l'accès, l'Administration aura une meilleure vision de ces données massives et pourra, dès lors, faire des prévisions sur les marchandises et les personnes à haut risque, sur les liens qui n'avaient pas encore pu être identifiés et qu'elle pourra exploiter.

Normes internationales

En des termes plus techniques, ces ensembles de données arborent un large éventail de formats structurés et non structurés. Le changement est radical pour les services douaniers habitués à des données structurées telles que les données émanant des déclarations électroniques, dont le format est clairement fixé par le Modèle de données de l'OMD et les

⁴³ Renseignements fournis par écrit par le Canada.

⁴⁴ Id.

normes EDIFACT. Les « normes » en matière de mégadonnées seraient bien plus variées et les autorités ne les contrôlèrent que dans une moindre mesure.

Il va sans dire que l'exploitation des avantages qu'offrent les mégadonnées exige de la Douane qu'elle gère correctement ses propres données structurées et qu'elle fasse des progrès sur d'autres fronts, afin d'être à même d'exploiter également les nouvelles données. Les normes concernant le commerce et l'industrie sont toujours en cours d'élaboration et, dès lors, les Administrations douanières à la recherche de données du secteur privé seront invitées à centrer leur attention sur les nouvelles normes, afin de pouvoir profiter du potentiel de « pertinence » des mégadonnées.

La gestion des risques

Ces dernières années, les entreprises privées sont de plus en plus nombreuses à suivre l'historique de leur clientèle, de manière à pouvoir prendre des décisions d'achat et de crédit plus appropriées et plus promptes. Les services gouvernementaux pourraient en faire de même, car l'analytique des mégadonnées peut les aider à définir le schéma comportemental des individus et ses conséquences.⁴⁵ En matière de renseignements concernant les voyageurs (qui ne relèvent pas des mégadonnées), la Douane et d'autres services présents aux frontières pourraient, par exemple, consolider et actualiser les antécédents des passagers aériens de manière à identifier et repérer ceux qui pourraient présenter un risque élevé.

La même démarche peut servir à contrôler les mouvements transfrontaliers de marchandises, dont les marchandises transportées par voie routière ou maritime. Cependant, contrairement à ce qui se fait pour le contrôle des voyageurs, l'utilisation des mégadonnées à des fins de gestion des risques de fret pourrait être différente.

Le fret maritime (en vrac ou conteneurisé) risque plus d'être utilisé pour dissimuler de vastes quantités de biens de contrebande et autres produits illicites. L'évaluation des risques fait déjà partie intégrante des procédures douanières habituelles de lutte contre la fraude et a donc une incidence sur les résultats des exercices de sélection des marchandises, mais la Douane a néanmoins d'énormes difficultés à désigner l'envoi à saisir parmi les cargaisons en attente de dédouanement dans les ports maritimes. Ces cargaisons volumineuses rendent les contrôles généraux pratiquement impossibles (infaisables ou irréalisables), alors que les bagages enregistrés par les voyageurs pourraient y être soumis.

Dans ce contexte, les Administrations douanières auraient peut-être intérêt à enrichir la collection de données qu'elles ont accumulées au fil du temps, pour avoir une vue d'ensemble des risques auxquels elles doivent faire face. Des outils de pointe en matière d'analyse statistique leur permettraient, par exemple, de croiser les données internes concernant les parties prenantes aux échanges (récoltées lors de leurs contacts réguliers avec les opérateurs commerciaux, les douaniers, les expéditeurs, les destinataires, etc.) et d'autres catégories de données précises sur les antécédents des entreprises privées (figurant dans les Dossiers passagers, par exemple, dont les informations personnelles peuvent facilement être fusionnées à d'autres mégadonnées du secteur financier et des communications). Cela permettrait de mettre en lumière certains types de délits perpétrés dans le secteur commercial (allant de la contrebande aux fausses déclarations) et de mieux comprendre les infractions qui y sont éventuellement liées (commises à des fins d'évasion fiscale ou de dissimulation de transactions monétaires illicites) et qui demandent un examen ou un audit poussé.

⁴⁵ Prow (2014).

Encadré 8. Agence des services frontaliers du Canada (ASFC), 3^{ème} partie⁴⁶

Quelle utilisation l'ASFC fait-elle des mégadonnées?

- L'ASFC tire parti des données internes et externes pour se forger une vision qui l'aidera dans sa mission de facilitation et de sécurité. Le Centre national de ciblage utilise désormais les mégadonnées pour identifier les voyageurs et les marchandises à haut risque.
- L'équipe d'analytique avancée⁴⁷ est à la tête de plusieurs projets pilotes de mégadonnées qui devraient améliorer l'analyse prédictive, permettre de visualiser l'information. L'Administration pourrait alors faire un meilleur usage des données biométriques. Les membres de l'équipe ont accès à l'environnement informatique expérimental de l'Administration et peuvent dès lors adapter le logiciel et le matériel à n'importe quel type de tâche.
- L'EDW (évoqué plus haut) réunit des données de sources diverses (mais néanmoins principalement internes), à des fins d'analyse. L'EDW est actuellement utilisé pour soutenir les opérations de ciblage de l'ASFC visant à interdire des voyageurs et des marchandises, avant leur arrivée au point d'entrée sur le territoire.
- Les techniques d'analyse de l'ASFC dans le domaine commercial ont progressé, grâce aux divers projets qui ont été financés. L'analytique précoce est utilisée dans le programme commercial pour identifier et intercepter les envois à haut risque et améliorer le ciblage. (Partenariats douane-entreprises)

Si la méthode d'analyse des données permettait à la Douane d'identifier et de suivre les personnes et les marchandises douteuses, elle aurait un effet plus important sur la gestion et l'exécution des tâches quotidiennes de déploiement des ressources humaines et de l'équipement requis. À partir des tendances et schémas détectés dans les flux ayant fait l'objet de contrôles, nous espérons que la Douane apprendra à anticiper les menaces entrantes et arrivera à distribuer de façon optimale ses ressources limitées, tout en fixant ses priorités dans ce domaine.

Toutefois, compte tenu de la nature des mégadonnées, l'analytique des données n'est pas la panacée universelle, mais plutôt une base supplémentaire de vérification (enrichissant le stock de données en fonction desquelles la Douane peut prendre ses décisions). Cela étant dit, d'autres sources de données éventuellement produites par des sources indépendantes dans un but non réglementaire, ne sont pas toujours aussi faciles à vérifier que les déclarations en douane et oblige ses fonctionnaires (en ce compris ceux qui souhaitent exploiter le plein potentiel des mégadonnées) à examiner les données d'un œil plus nuancé.

Pour être plus précis, les mégadonnées peuvent ne pas donner aux analystes de réponse clairement positive ou négative mais peuvent, néanmoins leur donner une idée des tendances, des schémas, des manques et des indicateurs possibles. Avant d'utiliser les mégadonnées à des fins décisionnelles, le personnel douanier doit donc se poser certaines questions fondamentales sur l'exactitude des données, leur disponibilité et leur admissibilité. Rappelons à ce sujet que, compte tenu du risque qu'un sentiment infondé de sécurité peut représenter, la Douane britannique (HMRC) insiste fortement sur la qualité, la précision et l'intégrité des données disponibles. L'application passive et mécanique de données faisant fi des principes de prudence et de diligence raisonnable pourrait mettre en péril l'effort

⁴⁶ Renseignements fournis par écrit par le Canada.

⁴⁷ L'équipe dépend du Service information, sciences et technologie (l'ASFC compte cinq autres services et un groupe).

ininterrompu et cohérent fait par la Douane pour engranger des connaissances et prendre des mesures de manière stratégique.

L'internet des objets

Toutes les marchandises qui passent les frontières sont sujettes au contrôle des services douaniers. L'internet des objets prouve l'intérêt qu'il y a à appliquer les mégadonnées, moteur essentiel auquel la Douane peut faire appel pour connaître l'emplacement, l'état et le statut des marchandises échangées en temps réel. La technologie de détection avancée en est un autre exemple, dans la mesure où elle permet aux prestataires de services logistiques qui l'utilisent de repérer toute irrégularité présentée par des cargaisons en transit et tout ce qui s'y rapporte. La sécurité de la chaîne logistique s'en trouve renforcée et, en vertu de ce système, un envoi conteneurisé considéré comme présentant peu de risque ou « sans risque » peut conserver ce statut jusqu'à sa livraison finale, sauf si une manœuvre suspecte est détectée pendant le transport.

Le rôle des Administrations de la douane n'est pas forcément de surveiller les mouvements du fret/des envois (pour éviter les accidents), mais bien de resserrer les liens de coopération avec certaines parties intéressées (expéditeurs, transporteurs, transitaires, etc.) qui se sont tournées vers des applications de l'internet des objets pour obtenir rapidement des informations sur certains facteurs de risque. Ce système permettrait aux services douaniers de se concentrer sur leurs tâches analytiques, alors que les opérateurs commerciaux assumeraient des fonctions de « gardiens » et l'on pourrait même envisager le partage de certains rôles et responsabilités entre les services douaniers et des entreprises privées.

Le suivi des envois dans le temps influera favorablement sur les critères de sélectivité de la Douane et le dédouanement des importations devrait s'accélérer. Toutes les parties intéressées en profiteraient donc et, dans un même temps, la Douane pourrait faire face aux conséquences des irrégularités en temps voulu et prendre les mesures nécessaires pour que celles-ci n'ouvrent pas la voie à des infractions graves dans leur jurisprudence.

Le commerce électronique

Dans le cadre des échanges par voie électronique, chaque jour, d'énormes quantités de données concernant la vente de biens et de services sont produites et régulièrement stockées, traitées, gérées, avant d'être exploitées par les marchés en ligne et d'autres intermédiaires qui sont liés (passerelles de paiement, expéditeurs de fret express/distributeurs) et proposent leurs services directement aux consommateurs. C'est un secteur dans lequel les entreprises mobilisent les mégadonnées pour maximiser leur profit. L'application de ces données, conjuguées à des paramètres sociaux et économiques concernant les opérateurs commerciaux en ligne (surtout lorsqu'ils sont impliqués dans des transactions entre entreprises et consommateurs {B2C} ou, éventuellement, entre consommateurs {C2C}), permettrait à la Douane de bénéficier d'une meilleure information lorsqu'elle évalue le degré de conformité des acheteurs particuliers/individuels, qui sont les parties dont on tient généralement le moins compte lorsqu'on parle d'échanges physiques. Qui plus est, la Douane serait mieux informée des marchandises échangées et, partant, mieux préparée à s'attaquer aux risques émergents pour la sécurité et la santé humaines.

Le commerce en ligne est un monde riche de données, qui pourrait être une mine d'information sur les échanges de marchandises pour la Douane. À la différence des modèles commerciaux traditionnels, certains marchés centralisés gèrent et actualisent systématiquement les dossiers des transactions passées et en cours. Si leurs données étaient créées et partagées

efficacement, sur une base transactionnelle, la Douane pourrait prendre des décisions encore plus éclairées et rapides, pour faciliter le dédouanement des marchandises (échangées en ligne) et garantir le recouvrement des recettes (en évaluant correctement les taux d'imposition). Dans des pays comme la Chine, l'Allemagne, l'Inde, le Brésil et le Royaume-Uni⁴⁸, la vente au détail en ligne (ou commerce en ligne) est en plein essor et les services douaniers pourraient dès lors s'aider des données que génère cette activité pour écourter autant que possible la procédure de dédouanement.

La sécurité

Les mégadonnées ont une incidence sur la sécurité et le renseignement. Certaines forces de police analysent des aspects précis des communications par voie de réseaux sociaux, pour tenter d'anticiper les menaces sur la sécurité (qui étaient latentes mais peuvent devenir imminentes), ce qui pèse sur l'organisation temporelle et géographique des patrouilles. La Police a une approche des mégadonnées qui comporte généralement une composante « analytique prédictive » désignant l'un des domaines de l'exploration des données et une application des techniques statistiques provenant plus précisément de l'apprentissage automatique à l'aide d'algorithmes de "surveillance prédictive". L'analyse prédictive a pour but d'anticiper des événements méconnus qui présentent un intérêt et peuvent jouer un rôle essentiel dans la protection des collectivités. Cette technique tire parti des renseignements les plus récents sur les "faits et gestes" de nombreuses personnes (considérées comme des « variables ») et repose sur des mégadonnées en évolution constante (« données de flux ») produites par les communications humaines quotidiennes. Voilà pourquoi la prévention des abus en matière de protection de la vie privée/confidentialité est une question qui mérite notre attention⁴⁹, au même titre que l'exploitation possible de données pertinentes à des fins criminelles plutôt que par souci de protection de la société contre les menaces imminentes.

Les États-Unis ont introduit un système de contrôle baptisé *Automated Virtual Agent for Truth Assessments in Real-time* (Système virtuel d'évaluation automatisée de la crédibilité en temps réel ou AVATAR)⁵⁰ dans leurs programmes de Voyageurs dignes de confiance. Ce système devrait rendre plus commode l'entrée ou le retour sur le sol américain de passagers/voyageurs préautorisés ou à faible risque, auxquels certaines voies et guichets sont réservés. Dans plusieurs ports d'entrée, la USCBP a installé des guichets équipés de ce système, dans lesquels se trouve un agent virtuel ("avatar"), qui interviewe les personnes ayant demandé le statut de Voyageur digne de confiance, pour transmettre ensuite automatiquement, un retour d'information à des agents en chair et en os, à l'aide de tablettes et de smartphones. Ces agents peuvent, alors, organiser une entrevue de suivi plus fouillée avec les demandeurs⁵¹.

Marr (2016) donne du fonctionnement du système AVATAR la description suivante: pour pouvoir poser un jugement probabiliste sur la véracité des propos tenus par une personne, le système a recours à trois capteurs différents encastrés dans l'enceinte de la machine: une caméra infrarouge enregistre les mouvements oculaires et les données sur la dilatation de la

⁴⁸ Statista, <https://www.statista.com/topics/871/online-shopping/> (dernière consultation: 7 mars 2017).

⁴⁹ Cf. Engelen (2015).

⁵⁰ Le système AVATAR a été développé par des chercheurs du *National Center for Border Security and Immigration* (Centre national pour la sécurité aux frontières et l'immigration) de l'Université de l'Arizona. Il a été pensé pour signaler les comportements anormaux ou suspects qui appellent à une enquête de terrain plus minutieuse, menée par de vrais agents du *Department of Homeland Security* (ministère de la sécurité intérieure ou DHS). Université de l'Arizona – 'AVATAR', <http://borders.arizona.edu/cms/projects/avatar-automated-virtual-agent-truth-assessments-real-time> (dernière consultation: 7 mars 2017).

⁵¹ Greenemeier (2012).

pupille à 250 images par seconde; une caméra vidéo surveille les mouvements corporels, afin de repérer des tics suspects ou des comportements courants chez les personnes tentant de dissimuler quelque chose. Enfin, un microphone enregistre les données vocales, pour repérer de légères inflexions du timbre et du ton de la voix. Les autres interactions entre l'agent virtuel et la personne faisant l'objet d'un contrôle à l'immigration génèrent de nouvelles données à caractère personnel que le système croisera avec les informations de la « mégabase de données » qui ne cesse de croître et est constamment mise à jour.

En somme, AVATAR s'inspire d'un modèle souvent utilisé dans d'autres projets de mégadonnées qui recourent aux techniques de modélisation prédictive. Plus le nombre de personnes traitées par le système sera élevé, plus il en saura sur les indicateurs faciaux, vocaux et contextuels présents chez les individus qui mentent ou tentent de tromper. La précision avec laquelle la machine repère les voyageurs suspects qui devront être retenus ne cessera d'augmenter si on lui fournit les données et algorithmes adéquats.⁵²

Les menaces émergentes

Ces dernières années, la Douane a été pressée de s'attaquer aux menaces émergentes et complexes que posent le blanchiment de capitaux et le financement du terrorisme. Les criminels utilisent les échanges licites (y compris les systèmes bancaires liés aux financements commerciaux) pour dissimuler le produit de leurs activités illicites. De tels actes, que l'on qualifie souvent de « blanchiment de capitaux fondé sur les transactions commerciales », adoptent des schémas divers, imaginés pour compliquer la documentation d'une série de transactions commerciales. Selon les estimations, cette forme de blanchiment se chiffre en milliards de dollars et plus par an, chiffre qui augmente année après année.⁵³

Les fausses déclarations (sur ou sous-évaluation de la valeur des marchandises) sont des indicateurs qui devraient permettre aux douaniers de découvrir des cas de fraude en temps réel. L'identification et l'atténuation des risques attachés au blanchiment de capitaux fondé sur les transactions commerciales sont des tâches extrêmement laborieuses qui, de par leur caractère manuel, sont sujettes à l'erreur humaine et ces procédures gagneraient donc à être automatisées.⁵⁴ L'analyse des mégadonnées peut être extrêmement utile dans le suivi des flux transfrontaliers de fonds illicites que les contrevenants tendent à dissimuler derrière les échanges internationaux.

Le blanchiment de capitaux fondé sur les transactions commerciales implique souvent des financements commerciaux au titre desquels les institutions bancaires doivent brasser d'énormes quantités de documents structurés ou non structurés. Pour pouvoir extraire des données à partir de sources non structurées, il faut développer des outils d'analyse de texte et ces données doivent être transformées en informations utilisables, dont il convient de vérifier l'exactitude et la pertinence. Les techniques de suivi statistique et algorithmique des transactions faciliteraient l'identification de modèles et de liens qui peuvent révéler des faits de blanchiment de capitaux fondé sur les transactions commerciales (manipulations des prix, fixation des prix de transfert inappropriés, par exemple).⁵⁵

L'application des mégadonnées à des fins de lutte contre la fraude pourrait être mise à profit pour inciter les services douaniers à collaborer avec d'autres autorités compétentes. En

⁵² Marr (2016), at 111-115.

⁵³ USICE, 'ICE leads trade-based money laundering investigations', <https://www.ice.gov/trade-transparency> (dernière consultation: 7 mars 2017).

⁵⁴ PwC US (2015), at 13.

⁵⁵ Id, at 13-15.

ce qui concerne les bonnes pratiques de lutte contre le blanchiment de capitaux fondé sur les transactions commerciales, le Groupe d'action financière (GAFI) a publié un rapport indépendant dans lequel il exhorte les pays à offrir aux autorités en charge des enquêtes (procureurs, police, agents de sécurité publique, etc.) un accès (direct ou indirect) opportun aux données commerciales et autres renseignements financiers utiles.⁵⁶ À ce propos, les Administrations douanières et fiscales pourraient envisager la création de bases de données utiles, sur le modèle des pratiques commerciales des contribuables et activités connexes liées aux transactions bancaires, tout en recourant à l'analyse des données pour identifier certains attributs communs aux ensembles de données détenus par différentes administrations. Le GAFI recommande également d'utiliser ces techniques pour croiser les données des exportations et les déclarations fiscales et y repérer d'éventuelles divergences.⁵⁷

VII. Conclusions

Les mégadonnées demandent certaines compétences, des outils d'analytique et une infrastructure équipée des technologies adéquates. La Douane de Hong Kong, Chine a renforcé sa capacité de stockage et de gestion d'ensembles massifs de données, tout en permettant le partage opportun et efficace de données entre les systèmes en vigueur au sein de l'Administration. L'USCBP a rassemblé ses banques de données, pour les mettre à disposition du personnel chargé de leur analyse. Quant à l'ASFC, consciente de la nécessité de poursuivre l'intégration des données pour renforcer sa capacité d'analyse, elle a mis sur pied un entrepôt de données où sont rassemblées les informations (internes principalement) provenant de sources diverses (qu'elle utilise régulièrement), à des fins d'analyse. Cette initiative s'apparente à une "centralisation" des données dont le but est de bénéficier de plus d'indices à partir d'un stock de données plus vaste.

Les priorités en matière de promotion des stratégies de mégadonnées diffèrent. De fait, la qualité des données est une priorité absolue pour le HMRC, raison pour laquelle ces services estiment qu'en faisant bénéficier les opérateurs commerciaux d'avantages palpables, ils pourraient inciter ces derniers à leur fournir des informations plus précises.⁵⁸ Selon la Douane de Nouvelle-Zélande, l'analytique occupe une place prépondérante dans la vision qu'elle se fixera pour les prochaines années et dont l'objectif sera d'instaurer un processus décisionnel et des opérations axées sur le renseignement en modifiant, éventuellement, les méthodes de travail.⁵⁹ L'USCBP estime que la plateforme des mégadonnées peut apporter des solutions pratiques (de type « analytique prédictive »).⁶⁰ L'ASFC juge nécessaire d'envisager d'autres sources et d'autres types d'informations complémentaires, à partir desquelles elle pourrait faire des prévisions exploitables.⁶¹

Les pratiques adoptées par le HMRC illustrent un modèle de partenariat novateur avec le secteur privé, puisque la Douane collabore aux chaînes logistiques intermédiaires qui génèrent, stockent et transmettent toute une série de données à des fins commerciales. L'Administration a mis sa plateforme de renseignement douanier en phase avec les systèmes exploités par les opérateurs commerciaux, de manière à expliquer les risques sous un angle plus large.⁶² La

⁵⁶ GAFI/OCDE (2008), par. 21.

⁵⁷ Id. para. 13.

⁵⁸ Renseignements écrits fournis par le Royaume-Uni.

⁵⁹ Renseignements écrits fournis par la Nouvelle-Zélande.

⁶⁰ Renseignements fournis par écrit par les États-Unis.

⁶¹ Renseignements fournis par écrit par le Canada.

⁶² Renseignements fournis par écrit par le Royaume-Uni.

capacité d'identification/d'évaluation des risques s'en trouvera probablement renforcée et, grâce au système, les données pourront être croisées et leur exactitude vérifiée. Sous un autre angle, l'on pourrait dire que les entreprises ont une part de responsabilité dans l'efficacité des contrôles frontaliers, objectif commun aux secteurs public et privé.⁶³

Les mégadonnées ne sont pas un concept statique, mais font, au contraire, l'objet d'applications en évolution constante, qui peuvent bouleverser la façon dont les agents douaniers de terrain effectuent leurs tâches et influencer sur la politique d'affectation des ressources destinées à encourager l'exécution de ces travaux. Cela étant dit, l'on s'inquiète de ce qu'il adviendrait si des services douaniers reprenaient aveuglément les mégadonnées réunies auprès de sources externes au cours de leurs opérations quotidiennes, sans en vérifier la véracité. Un sentiment de sécurité déplacé et les risques que comporte l'évaluation incorrecte/imprécise des risques déboucheraient presque inévitablement sur l'affaiblissement de la capacité de la Douane à lutter contre la fraude sur la base du renseignement. En tant que telles, les applications de mégadonnées ne garantissent pas forcément que la lutte menée par la Douane contre la fraude est plus intelligente.

Enfin, la Douane pourrait mettre la sécurité de son système de gestion de l'information en péril, avec les répercussions que cela aurait sur la quasi-totalité des facettes du service, puisque le réseau compte également des partenaires extérieurs. La technologie de l'internet des objets peut apporter une réponse directe à ce problème, dans la mesure où elle permet la connexion de la plateforme de renseignement de la Douane à différents périphériques (qui peuvent être très nombreux et éparpillés dans le monde entier). Cependant, cette technologie l'expose plus à des actes indésirables ou à des forces destructrices, de type « cyberattaque ». Par conséquent, il serait bon d'examiner les conditions dans lesquelles la Douane peut se préparer à faire face à de telles éventualités. Une plus grande résilience des entrepôts de données douanières serait, à n'en pas douter, favorable à la continuité des procédures commerciales, dont dépend la viabilité des échanges internationaux en situation d'urgence.

*
* *

⁶³ En outre, parmi les défis auxquels le HMRC est actuellement confronté figure la poursuite de la collaboration avec les parties intéressées, au profit d'une soumission et d'un partage plus précoces des renseignements concernant le fret. Le commerce en ligne, par exemple, pourrait être évoqué dans ce contexte, en raison de la nature des données transactionnelles (qui 'précèdent' toujours la distribution physique).

Références bibliographiques

1. Cleary, Duncan (2017), 'Big Data Analytics for Better Revenue Administration', Blog de l'IMF sur la gestion des fonds publics. <http://blog-pfm.imf.org/pfmblog/2017/02/big-data-analytics-for-better-revenue-administration.html> (dernière consultation: 7 mars 2017).
2. EMC (2012), Communiqué de presse, 'New Digital Universe Study Reveals Big Data Gap: Less Than 1% of World's Data is Analyzed; Less Than 20% is Protected', <http://www.emc.com/about/news/press/2012/20121211-01.htm> (dernière consultation: 7 mars 2017).
3. Rijmenam, Mark van (2013), 'Why The 3V's Are Not Sufficient To Describe Big Data', Dataflog, <https://dataflog.com/read/3vs-sufficient-describe-big-données/166> (dernière consultation: 7 mars 2017).
4. Adamov, Abzetedin (2015), 'Understanding Big Data: 3Vs of Big Data', posted on his blog, <https://aadamov.wordpress.com/2015/04/24/understanding-big-données-3vs-of-big-données/> (dernière consultation: 7 mars 2017).
5. Gerasimou, Vicke (2016), 'Big Data and the 3Vs: What is the fourth 'V' and what are the implications for not embracing it?', Think Big, <https://www.thinkbiganalytics.com/2016/03/29/big-data-3vs-fourth-v-implications-not-embracing/> (dernière consultation: 7 mars 2017).
6. AAAS-FBI-UNICRI (2014), *National and Transnational Security Implications of Big Data in the Life Sciences* (Projet commun), American Association for the Advancement of Science, disponible à l'adresse suivante <https://www.aaas.org/report/national-and-transnational-security-implications-big-data-life-sciences>.
7. Hu, Margaret (2015), 'Small Data Surveillance v. Big Data Cybersurveillance', Washington & Lee Legal Studies Paper No. 2016-6, PEPPERDINE LAW REVIEW, Vol. 42: 773, disponible à l'adresse http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2731344.
8. Rapport PCAST (2014), *Big Data and Privacy: A Technological Perspective*, Executive Office of the President, US.
9. Davenport, Thomas H (2014), *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*, Harvard Business Review Press, US (ISBN: 978-1-42216-816-5).
10. Mayer-Schönberger, Viktor and Cukier, Kenneth (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, John Murray, UK (ISBN: 978-1-84854-792-2).
11. McNulty, Eileen (2014), 'Understanding Big Data: Machine Learning', Dataconomy, <http://dataconomy.com/understanding-big-data-machine-learning/> (dernière consultation: 7 mars 2017).
12. Oberlin, Steve (2012), 'Machine Learning, Cognition, and Big Data', CA (unavailable at the corporate website).

13. Naughton, John (2016), '*Why the internet of things is the new magic ingredient for cyber criminals*', The Guardian, <https://www.theguardian.com/commentisfree/2016/oct/02/brian-krebs-ddos-attack-google-protection-cybercrime> (dernière consultation: 7 mars 2017).
14. TechTarget (2016), '*Time to Toughen Up for IoT*', SearchSecurity.com e-publication, disponible à l'adresse suivante <http://searchsecurity.techtarget.com/> (dernière consultation: 7 mars 2017).
15. Baiba, Ziv (2016), '*Start small, think big: big data!*', WCO Actus N. 79, Organisation mondiale des douanes.
16. Prow, Charles (2014), '*Putting Analytics at the Border Would Increase Security, Convenience (Industry Perspective)*', Government Technology, <http://www.govtech.com/federal/Putting-Analytics-at-the-Border-Would-Increase-Security-Convenience-Industry-Perspective.html> (dernière consultation: 7 mars 2017).
17. Engelen, Errol van (2015), '*Big Data Will Effectively Fight Terrorism In the World*', Dataflog, <https://dataflog.com/read/big-data-will-effectively-fight-terrorism/785> (dernière consultation: 7 mars 2017).
18. Greenemeier, Lary (2012), '*Avatar Officer Installed at Arizona-Mexico border Station*', Scientific American, <http://www.scientificamerican.com/article/avatar-officer-installed-mexico-border/> (dernière consultation: 7 mars 2017).
19. Marr, Bernard (2016), '*Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*', John Wiley & Sons, UK (ISBN: 978-1-119-23138-7).
20. PwC US (2015), '*Goods Gone Bad: Addressing Money-laundering Risk in the Trade Finance System*', disponible à l'adresse suivante <http://www.pwc.com/us/en/risk-assurance-services/publications/trade-finance-money-laundering.html> (dernière consultation: 7 mars 2017).
21. GAFI/OCDE (2008), '*Best Practices Paper: Best Practices on Trade Based Money Laundering*', GAFI, Document d'orientation.