If algorithms dream of Customs, do customs officials dream of algorithms? A manifesto for data mobilisation in Customs

Kunio Mikuriya and Thomas Cantens

Abstract

Governance by data is a growing global trend, supported by strong national public policies whose foundation is open data, artificial intelligence and decision-making supported by algorithms. Despite this trend and some technical advances, Customs face obstacles in deploying ambitious data use policies. This article describes these challenges through recent experience in some Customs administrations and considers the technical and ethical issues specific to all law enforcement agencies in the context of customs missions, to open paths for research and propose policy recommendations for a better use of customs data.[1]

1. Background

Big Data technologies—however we name them, ‘algorithms’, ‘artificial intelligence’ (AI), ‘machine learning’ (ML)[2]—are more than simple ‘tools’. They are a *daily gesture*, to travel, choose, exchange, but also police, control or punish.[3] Machines no longer help us to manufacture but to choose, in the sense of deciding, predicting or anticipating. As a counterpart of the ease with which machines can make choices with more celerity and accuracy than we do, there exists a worry that something is slipping out of our hands. Technologies seem to dictate our conduct to us, they evaluate our choices and guide us by calculations with results being imposed on us as mathematical evidence, and whereby proof is increasingly distant and obscure to non-specialists, therefore less and less disputable.

Technologies tend more and more towards the empowerment of the machine. What do algorithms dream of? With this provocative question, Denis Cardon (2015), a sociologist, refers to a double metaphor: the limitless extension of the presence of machines in our lives, but also the ambiguous relationship that we maintain with these machines that we would like to always perform better but which threaten to surpass us. Cathy O’Neil (2016), a mathematician, speaks of ‘weapons of Math destruction’ to denounce that algorithms are affected by numerous biases that accelerate and amplify social inequalities.

The power of states and their bureaucracies that make an increasing use of data and algorithms is added to that of the machines. Governance by data logically reinforces governance by numbers (Supiot, 2015)—the ‘quantophrenia’ of the state (de Gaulejac, 1990)—since the 1970s. Combined with AI, this governance raises concerns and criticisms, mainly about the emergence of a state of generalised surveillance. While these concerns are legitimate, they nevertheless demonstrate a limited vision of interactions between the state and society.

Tax administration and trade governance are still little explored areas of data science research. However, Customs were pioneers in IT, all over the world, from the 1970s for a few rich countries and the 1980s for the so-called developing countries. Today, despite technological disparities, all customs administrations have automated some if not all of their procedures.[4] Customs administrations are often the first to be computerised among all tax administrations due to the fact that border processing is largely standardised worldwide. Customs collects data everywhere in a massive way. In addition to data, the computational culture is well established in Customs. For more than 20 years, most customs administrations have incorporated the concept of risk analysis either as a process in their IT systems or as a necessity to do so.

Most of the research on the use of customs data has been technical. Customs targeting and, more broadly, the fight against fraud have become classic problems for engineers, econometricians, statisticians and data scientists through their use of clustering, classification algorithms, econometric techniques, mirror analysis, and the search for outliers (Cantens, 2015; Cariole et al., 2019; Chermiti, 2019; Choi, 2019; Grigoriou, 2019; Hua et al., 2006; Laporte, 2011; Xiao et al., 2016; Yaquin & Yuming, 2010; Zhou, 2019). The use of data to reform or fight against bad practices, without necessarily mobilising complex algorithms, has also shown its effectiveness (Cantens et al., 2010; Chalendard et al., 2019; Grigoriou et al., 2019; Kalinzije, 2018). Researchers have quickly established links between customs issues and areas where AI is already very advanced, such as image recognition, applied to non-intrusive scanning inspections (Jaccard et al., 2017; Kolokytha et al., 2017). The World Customs Organization (WCO) BACUDA platform (BAnd of CUstoms Data Analysts) has developed a series of studies on the use of ML for Customs fraud detection, online price data collection, ML for customs revenue prediction, and data visualisation.[5] Finally, new possibilities are opening up with the customs use of geolocated data.

Examples drawn from the use of data science in tax administrations may also inspire customs officials: the detection of fraud schemes based on fiscal measures through the use of biological models of co-evolution (Hemberg et al., 2016), or the combination of ML and analysis networks for the selection of controls.[6]

This article analyses the spread of governance-by-data within Customs. It builds on these technical advances and multiple experiences, including a seminar on reform by numbers organised by the WCO and the World Bank in 2012, two expert workshops on data analysis in 2019, three high-level seminars organised by the WCO in different regions (Asia–Pacific, Europe and the Middle East and North Africa) in 2018 and 2019, a workshop on geospatial data organised at the WCO secretariat in 2019, various missions and visits to customs administrations, and a survey launched in 2019 with responses obtained from 60 customs administrations.

The first section examines how a new type of governance, a governance-by-data, is emerging, spurred by an increasing number of states. Most customs administrations have not yet integrated this evolution. The second section examines technical challenges that may explain this situation and suggests ways to tackle these challenges in the customs environment. The last section is focused on the role of Customs, that is at the interface of the economic and the repressive state, in the new relations of governance between states, businesses and citizens. It also explores policy conditions, under which Customs will adapt itself to this changing environment.

Two preliminary observations are necessary. First, these technologies are not reserved for a select few rich countries. In 2018, a World Bank symposium on the role of Big Data in achieving the Millennium Development Goals illustrated the diversity of data usage by governments.[7] More generally, data technologies usually leapfrog in less rich countries: they adopt the latest technologies, according to their needs, without following the linear pace of technology adoption in rich countries.

Second, data is associated with mining metaphors. It would be the ‘new oil’ (Humby, 2006) or the object of a new ‘gold rush’. These metaphors are even a part of the technical language, since we speak of ‘data mining’. They convey a reality: data, like crude oil or gold, has value only after treatments and for a multitude of usages. It is also true that data raised the same economic craze as oil and gold centuries before. However, the mining metaphor is misleading: data is not scarce. If it is an economic resource, then it is the most widespread resource in the world, and the most egalitarian one since everyone produces and owns data (even if this property is often transferred to others).

These two points are essential for the global customs community: there are no rich and poor countries when it comes to data; there is no government or administration that could not embark on an ambitious policy to use data; there is no customs administration that would not have data, 'big' or not, to the extent of its needs.

2. Governance-by-data

Numbers with nearly exponential growth regularly account for the influence of technologies in contemporary societies. Computer science is populated with laws[8] and estimates that show a steady increase in the capacities of machines and the production of data.[9] The AI 2019 report (Crawford et al., 2019) estimates the performance of machines according to standard tests, and the machines' progress is rapid: more precision (90% image recognition in 2019 compared to 60% in 2016) and more celerity (training of algorithms in 88s against 3h) (Crawford et al., 2019, from p. 48). This quantitative growth of data and machine capabilities reflects a deeper change, the emergence of governance-by-data (Elkin-Koren & Gal, 2019).

2.1. The emergence

The transition from governance by numbers to governance by data (and numbers) is made possible due to the fulfillment of four conditions.

The first condition is 'the emergence of probability' (Hacking, 1975) in the 17th century, a shift from deterministic thinking to probabilistic thinking, in science and governance. This epistemological change means that, in matters of governance, decisions are taken based on their probable quantifiable effects. The quantification of uncertainty is a major feature of contemporary thought and governance.

The second condition is the technological evolution of AI itself and the development of datacentric methods. In the 1950s, the 'smart' machine was centered on rules, designed to operate in complete information environments. A chess game has no unknown rules, all the information about the game can be supplied to the machine. It is only since the 1970s that AI has been applied to the real-world situations where knowledge of the rules is imperfect (Piscoppo & Birattari, 2008). For example, it is impossible to provide the machine with all the rules for recognising the subject of an image. A new step, more recently, confronts the machine with an actor who hides his action: this is the case of fraud detection. The environment is not only incomplete, it contains information purposefully hidden in data. This latest development is fuelling governance control and surveillance functions.

The third condition is the availability of data. By leaving the chess game and its rules, and by tackling fraud (for instance), the machine leaves the symbolic environments for real environments. This transition requires the provision of data in all areas (Brooks, 1991). As such, we should speak of a *Big Use of data*, rather than *Big Data*, to mean that the raw material of intelligence (artificial or not) is the data and no longer the rules.

The will to legitimise governance by reason, and reason by calculation, has therefore been combined with the scientific possibility of quantifying—and mastering—uncertainty, and with the fact that uncertainty is more reduced when there is more data available. The whole scheme works on the last condition that a machine is capable of carrying out calculations whose magnitude exceeds human capacities. This is the fourth and final condition: computer science, development of graphics processing units for calculation, decrease of data storage costs, cloud computing and cluster techniques have enabled researchers, businesses and administrations to mobilise significant computing and storage resources at affordable cost.

Data governance is based on decision-making with uncertainty. It is not intended to ensure that something is true or not, but to classify solutions according to the degree of uncertainty of their effects. This principle is well known to customs officials who are obliged to choose a container or a shipment from among thousands. The principle is also at work in trade governance. Through international rankings on the ease of crossing borders, the logistical capacities of countries, borders are increasingly 'mathematised' in the sense of becoming data and calculation objects (Cantens, 2018).

2.2. Recent national AI and data strategies

Many states have adopted national AI and data strategies. Governments have set up more or less centralised open data[10] services, as in the United States,[11] France[12] and South Korea.[13] In 2011, the 'partnership for open government' was created as a multilateral entity bringing together 79 member states as of today (Open Government Partnership, 2019). The Organisation for Economic Co-operation and Development (OECD, n.d.) has developed the OURdata index (Openness, Usefulness and Re-usability), measuring the quantity of open data made available by states, its usefulness and ability to be used by third parties. In addition, states build directories of free software intended to constitute a common base for all their institutions.[14]

Regarding AI, since the 1950s, the relationship between states and machines has been less linear, marked by the famous two 'winters' of AI, in the 1960s and the late 1980s (House of Lords, 2018). Since 2017, state investment seems more massive. Around 50 states have developed strategic or normative documents relating to AI or 'digital' government (Schiff et al., 2020). Among these, around thirty have established a national strategy (Merz, 2019). In 2018, the British government released an AI Sector Deal (UK Government, 2018). In 2019, after having followed a liberal approach for a long time, leaving the role of innovation to the market, the American government launched its AI initiative, asking national agencies outside the defence sector to invest in AI to support the public demand (US Government, 2019).

A dozen strategies encourage the development of AI projects in public institutions. Some countries do not have a specific AI strategy but have integrated it into global policies to transform society through technology, such as Japan (Government of Japan, 2015). Most international or transnational organisations like the World Economic Forum also produce recommendations for states to support the industrial development of AI (World Economic Forum, 2019). Other international actors, such as the European Space Agency, invest heavily in AI.[15]

Some AI strategies are fuelled by substantial funding, more than USD 1 billion, in South Korea, France, Taiwan and the United Kingdom (Dutton et al., 2018). Some countries have created specific entities for AI. For example, the UK has its National Office for Artificial Intelligence (UK Government, n.d.); Niger launched a National Agency for the Information Society whose director has the rank of a minister;[16] France has a public data service, managed by a specific public body, Etalab.[17]

The national strategies are threefold: encouraging the emergence of an industrial sector for the use of AI by the private sector and public administrations; strengthening research capacities; and launching partnerships with the private sector (OECD, 2019). Among government initiatives, the idea of pooling state data led to creating governmental data trusts in India (NITI, 2018).

The origin of these initiatives is politically driven by the wish to generate a more 'open' government, sharing with citizens its results through indicators and data sets (House of Lords 2018, pp. 36–37; Jetzek et al., 2019). However, the idea of citizen participation in public life, through data, must be understood with caution: it is not 'ordinary' citizens who will be able to mobilise public data, but rather analysts (Dawes et al., 2016).

If the initial objective of political transparency is still valid, states recently recognised that data has become a national asset, in particular public services' data (US Government, 2020). States are encouraging the rise of the data-driven economy, in particular with the use of public services' data for industrial innovation (UK Government, n.d.).

AI and data have thus become a very strategic domain. They are 'dual-use goods' for states: a means of strengthening their governance capacity and a new dimension of the economy which they must support, in a global race for innovation, given the remaining uncertain profitability of this sector for many companies. States need also a robust and diversified AI and data industrial sector in order not to face a situation of a global economy dominated by a reduced number of transnational actors with data collection and analysis capacities that exceed that of states.[18]

2.3. Some customs applications

Security and fraud detection are the most used applications of data technologies in Customs, since the technologies had already been developed for the intelligence services, police and the military. The introduction to this paper has referred to research articles on the use of ML for fraud detection, but few are deployed operationally, and few administrations currently share their results.

Customs in Hong Kong, China, the Netherlands, Japan and Brazil recently reported to the WCO (2019) how they use data and AI for fraud detection. During the WCO workshops on data analytics and geospatial data, France, the United Kingdom, New Zealand, and South Korea also shared their experiences in the areas of fraud detection and risk analysis. The European border agency, Frontex, uses a geographic information system to manage the external borders of the European Union (EU), which fuses data from different sensors and sources (ships, individuals, other databases) (Malinowski, 2019). The EU has also launched the EU IBorderCtrl project, tested in three countries, to detect illegal migrants at border controls using facial expressions (EU, 2020).

Some areas are still under-explored, such as the use of AI to help taxpayers to be compliant: in Australia, a chatbot answers questions[19] of customers; and the same type of tool is deployed in the United Kingdom (UK Government, 2017).

The provision of customs open data varies greatly from one country to another: from a simple list of seaports in one country, to transaction-level data on penalties by Canada. Most customs administrations do not publish transaction-level data but rather aggregate data or public information in an electronically reusable format. While this information was already made public, the novelty is its availability in a single point.[20] It may look like a small step, but one should not underestimate both the difficulty for the administration to concentrate its data in one web portal, and the ease of access to information that this system brings to users and researchers.

There are cooperation initiatives between Customs and researchers that are largely based on the sharing of transaction-level data. In 2011, the United Kingdom customs launched a *datalab* to make its data accessible to researchers.[21] The data is 'de-identified': each importer and exporter is assigned a specific identifier.

Too few administrations use their data to analyse, or to simulate fiscal policies or antifraud strategies. The United States tax administration recently launched a call for expressions of interest to simulate large-scale tax policies, based on the simulation of citizens' lives, taking the reference of the video game 'Sims City'.[22] Wier (2020) worked with South African customs transaction data to identify the tax impact of fraud and its links to tax policies. Niger customs use transaction-level data to measure the fiscal and economic impact of customs measures before proposing them in the finance law. This type of simulation also strengthens the capacity of Customs to defend their proposals during discussions with their governments or partners like the International Monetary Fund and the World Bank, who are large consumers of data.

There remains a need to promote a wider use of customs data at a global level (Okazaki, 2018; Polner, 2018). No administration has yet appeared as data-driven or developed a comprehensive data policy. Two obstacles are common to all administrations.

The first is *scalability*, moving from a 'pilot' project on algorithms developed in a lab, to integrating the algorithms into an existing customs architecture. This industrialisation of research outcomes is often complex and must take into account many parameters, one of them being the existing customs IT system. In rich countries, there is a gap between the technologies underlying existing customs IT systems and the technologies used in data science for less than a decade.

The second obstacle is *profitability*, in particular for administrations responsible for generating public revenues. In the light of the sometimes boundless enthusiasm for AI and Big Data, we are all tempted to question the reality of high-tech projects. Are they profitable? Should we get into Big Data by crossing customs data with data available on the internet? The answer is positive for certain uses such as finding the value of goods through e-commerce sites. For offences involving prohibited or restricted products, real traffickers mostly do not openly stage themselves on social networks. Deploying AI to catch a passenger returning from a trip with more cigarettes than the maximum allowed is certainly not profitable. In addition, as a customs official recognised, before deploying AI to open data, customs data itself should already be fully mastered.

National AI and data strategies are broad, they are evidence that states are aware of the value of data. Their implementation in Customs, beyond the pilot projects, remains to be seen. There is, however, an emergency. The 'adversaries' of Customs may also use AI. A 'legal tech' is already being developed to help taxpayers in their litigation procedures: AI assists in predicting a probable outcome of the litigation; 'e-discovery' is used to automatically read large quantities of documents and select the most relevant ones (Deloitte, 2018; Engstrom & Gelbach, 2020; Kluttz & Mulligan, 2019). There is no technical obstacle that prevents an importer from using customs data—if he illegally purchases it from a customs IT department—or his own operational data, to 'optimise' the gain of a fraud given the way he completes his declaration.

Given the enthusiasm around the use of data, the possibilities offered by AI, and massive investments that are underway or planned by governments, Customs will 'do' governance-by-data. The uncertainty related to the best conditions under which they will be involved in this evolution. These conditions, technical and political, are discussed in the next two sections.

3. Bias, interpretability: technical challenges common to law enforcement agencies

As in any evolution, technologies raise questions about their integration into existing practices. Some of these questions are technical and well known: bias and interpretation of algorithm results. This section contextualises them in the customs environment, before proposing some research paths to overcome them.

3.1 Bias

Both technical and ethics literature are rich in studies on bias in algorithms (Dobbe et al., 2018), the amplification of biases (Lum & Isaac, 2016), and the near impossibility of preventing bias in the construction of some algorithms (Mittelstadt et al., 2016). It is the famous 'garbage in, garbage out' in statistics (Geiger et al., 2020) that is amplified by ML, or the 'dirty data' in the police, designating racist or ethnic biases (Richardson et al. 2019a), or the pressure of performance indicators which incentivise supervisors to over-report or omit certain data.

In Customs, a cause of bias can be data that is poorly collected (a lack of difference between 'origin' and 'provenance' for example, or between the importer and the final owner, which will make some origins or importers invisible in the dataset) or data that is incomplete (information on the point of entry into the territory is not available for all offices and the state capital is mentioned instead of a concrete border post or a geographical location). Another example is that corruption can bias the data: certain fraudsters or kinds of fraud may never appear as such in data that is used to train the machine, or wrong values may be regularly accepted by Customs and will therefore be considered as regular by the machine.

The machine is neither racist nor corrupt, that is a definite but non-sufficient advantage. It is difficult if not impossible to separate 'good' data from biased data (Richardson et al., 2019b). Conversely, it is impossible to postulate an absence of bias. This is the 'there is no free lunch' theorem: you need prior knowledge that makes the learning conclusions relevant or not (Shalev-Shwartz & Ben-David, 2014).

If bias is unavoidable, it is not a fatality. For instance, predictive policing software includes 'forgetting' past data to regularly renew the samples (Shapiro, 2019). It may also add an element of uncertainty via randomisation, and by producing only one indication, 'a place is at risk' (without specifying either the degree of risk or if this risk is attributed by the randomisation process). By doing so, predictive policing software gives some room for action and initiative to police patrols and includes reporting tools to analyse police tactics in the field, their over-control of certain populations and their effectiveness. It is close to the logic of 'performance contracts' tested in Customs against corruption that only control the use of their discretionary powers by customs officers through transaction-level data analysis (Cantens et al., 2014).

3.2 Interpretability

To better understand the challenges of interpretability in Customs, here is a simple example. To predict whether a customs declaration is at risk, the traditional techniques were based on econometrics: we hypothesised on fraud factors, evaluated their importance within an econometric model, and calculated the risk associated with each new declaration. This approach was based on the modelling of the fraud factors that we considered fair and mathematically tested and verified. Today, these techniques are still used, but others have emerged, such as deep learning and neural networks, which can predict the probability of fraud of a declaration. However, with these new techniques, we are not always able to know how the machine comes to the result. The designers of the algorithm could explain its mechanisms, but they themselves could not reconstruct the 'intellectual' thread that led the machine to the result. They wouldn't be able to do so because ... there's no more thread. This type of technique is used by the European experimental system iBorderCtrl, which processes facial movements to detect impostors at borders (Crockett et al., 2018).

The interpretability of an algorithm is a parameter that sometimes comes into play when, from a range of algorithms, we have to choose the one that seems most appropriate. This is why decision trees are often adopted in Customs. As a reminder, the use of decision trees is a learning technique that optimises the combination of customs declaration fields to calculate risks of fraud. At each level of the tree, we check that the declaration meets a criterion and we thus go to the final 'leaves', which give us the probability of fraud. These algorithms therefore return us a readable result, as a combination of declaration fields and values, which we can simply translate into computer rules in the customs clearance system.

The way in which an algorithm achieves a result is not just a theoretical or technical question, it is also legal one. According to the European Union regulation 2016/679,[23] people must have access to the logic behind the automated processing, the data provided must be 'readable', which prohibits proprietary formats, and automated processing alone cannot form the basis of a decision. Profiling for tax purposes is excluded from these constraints, but in all cases it will be necessary either to explain the basis on which the decision of the machine was made, or to advance arguments that are not drawn from the results of the machine.

Wachter et al. (2018) have shown the limits to the right of explanation in European Union regulations: it is limited to some cases, and the explanations may be too light or obscure. To justify that bias does not influence decisions, a solution advanced by Watcher et al. (2018) is to produce counterfactuals: render the decision for the existing situation A, but also provide a different decision that would have been made in a hypothetical situation B so that the user understands how the decision was made.

Finally, we should not over-interpret the machine's outcome: the machine does not predict customs fraud any more than it predicts crime for police. Predictive police software only predicts the crimes' possible locations (Ensign et al., 2017). Similarly, at Customs, it is by a semantic shift that we say that it predicts fraud. The machine is not trained on fraudulent declarations, it is trained on detected fraudulent declarations.

3.3 Inability to assess the extent of errors

In addition to the opacity of the algorithms, the machine is unable to evaluate its errors other than quantitatively.

One can imagine two algorithms for customs targeting. The first one detects commercial fraud with a better probability than the second one, but the false negatives (fraudulent declarations but not selected by the algorithm) of the first one are essentially illegal imports of weapons, while the false negatives of the second one are more diverse. Which algorithm to choose? Probably the second one even if it makes more mistakes, as they are less serious than those of the first one.

Any probability leads to a margin of error, but the algorithms remain incapable of anticipating the severity of the consequences of their errors.

This raises legal questions about human responsibility (Elish, 2019): who is responsible when the machine makes a mistake? How to estimate the degree of responsibility when the human in contact with the machine has only minimal control over it, and sometimes little knowledge of its functioning? These questions have crucial consequences in countries that severely punish corruption or loss of public revenue due to bad practices or officials' errors. In these countries, customs officials are already reluctant to rely on risk analysis systems because they fear being punished if undetected fraud leads to tax losses. The use of machines to fight fraud will lead to changes in the law regarding human responsibility.

3.4 Problem of the ML methodology applied in Customs

ML is probably the most modest and pragmatic approach to AI: it does not seek to imitate human intelligence or build a so-called 'strong' artificial intelligence. We must dwell a little on the methodological aspects. What do we do when we try to prove that the ML algorithm is good? Very broadly presented, a sample of declarations is split in two. All these are processed declarations, so we humans know which ones are fraudulent and which are not. With the first part of the sample, we train the machine to 'understand' what fraud is. With the second part, we test our trained machine to detect fraud. The result of the test is that we have four categories of declarations: 'true positives' ('true negatives') being fraudulent (non-fraudulent) declarations found as such by the machine, 'false positives', being non-fraudulent declarations considered as fraudulent by the machine, and 'false negatives' being fraudulent declarations considered as non-fraudulent by the machine. The algorithm is considered as 'good' if it detects a high rate of fraudulent declarations while directing in circuit control less declarations than the human.

It is necessary, here, to take note of customs specificity. In police, an offence can be declared and documented by victims, and not as a follow-up of a police intervention. Data on offences is therefore closer to the reality, while in Customs, there is no more data on fraud than the fraud discovered by customs officials (on the spot and *a posteriori*).

This raises a methodological problem, particularly related to convincing customs officials to use ML. In police, the outcomes of an algorithm 'predicting' offences can be compared to the police action, and it can be shown that the algorithm is performing 'better' than the police. This is not the case with current customs methods: the machines will only detect 'almost as much' fraud as customs officials because the detection method is based on discovered cases by customs officials. This problem is linked to the validation methods used for algorithms and has policy effects: how to convince customs officers to use a machine that detects less than them?

As in other domains, customs officials will be more capable than machines to face unpredictable situations. However, in Customs, it is, for now, the addition of two assertions— detecting almost as much and, above all, controlling less—that makes sense. This 'control less' corresponds well to the political doctrine of trade facilitation and the reduction of controls, but it will always come up against the *ethos* of the customs officer who wants to find more fraud and not find almost as much.

3.5 Some technical research paths

New experimental plans, particularly in ML, could improve the efficiency of the machine and its adoption by customs officials. These plans could carry out fully automated control periods versus completely non-automated periods, at random, a sort of randomly controlled trial. One can also imagine setting up an experimental plan on the *ex post* controls, on a sample of declarations which would all have been checked beforehand.

The algorithms themselves could offer experimental testing techniques: one must be able to test a control or patrol strategy against another or the current one. The machine should be able to help us predict the effects of an anti-fraud strategy before it is implemented. The aim is to relax the automaticity of the machine and recreate the conditions of choice for the officials.

We can also add a bit of chance. Randomisation is an important technique described in the police environment to make repressive actions less predictable by fraudsters and reduce bias (Shapiro, 2019). From a technical point of view, the interest of adding some randomisation in risk profiling is also to increase our knowledge of fraud. This echoes the customs officer's flair. However, the problem becomes technical again. What do we define as 'flair', how can the machine reach it? The introduction of randomisation is a human decision and its amount should comply with a criterion of acceptability of control, a maximum inspection rate for example, rather than follow a scientific, calculated criterion of optimisation. How could the machine decide in what proportion it increases its rate of randomised controls, which would be a quantitatively blind decision? How to program the machine for a choice between a blind chance and quantification of the error, which goes against its *raison d'être*?

Solving the previous technical questions—bias, interpretability, the ability to assess the impact of an error, the way to evaluate an algorithm—will increase our trust in machines. However, this technical trust is not the most complex to achieve. Customs administrations should also address more policy-oriented questions.

4. What is the place for Customs in the new data ecosystems?

Customs is a part of 'data ecosystems'[24] comprising citizens, who are both subjects of law and more or less voluntary producers of data about themselves, administrations, and transnational companies that have the capacity to collect data that is comparable to that of the states. The place of customs administrations in such ecosystems depends on the way they address three questions specific to law enforcement agencies and adapt them to the context of tax and trade governance: (i) the ethical dilemma between control and 'privacy', or 'commercial secrecy' in the case of Customs; (ii) the value given to the equality before the tax law, which is specific to customs and tax administrations; and (iii) the consolidation of public administrations' position in environments based on innovation.

4.1 Surveillance and ethics

Within these ecosystems, state surveillance through algorithms is a growing concern for citizens (Crawford et al., 2019; Burrell, 2016; Lyon, 2014). Law enforcement and intelligence services are on the front line, but Customs are also concerned, for example through the Advance Passenger Information/ Passenger Name Record rules or the monitoring of social networks and online commercial websites. Part of the distrust comes from the fact that civil administrations often transplant tools initially developed in the military or intelligence domains (Brayne, 2017; see for instance Palantir products). This is the case with intelligence data collection and fusion tools,[25] but also with the ongoing 'democratisation' of geospatial analysis and satellite imagery.

Data technologies are therefore not that 'disruptive'. They are a part of the continuous increase of the asymmetry of force and power between states and citizens, which has existed since the birth of states and the development of weapons and policing techniques.

However, three facts are new, transforming this simple asymmetry between states and citizens into ecosystems where relationships are more interdependent.

On the one hand, private companies are building surveillance capacities equivalent to those of states. Transnational companies follow the same logic as the intelligence services: to collect as much data as possible on their customers, without necessarily knowing *a priori* the specific objectives for which the data will be used. In addition to storing data on individual consumption, publicly expressed opinions, contact networks and movements, some of these companies also offer data hosting, calculation capacities and professional application services in the 'cloud'. They are building surveillance platforms for economic purposes (Zuboff, 2015; Manohka, 2018), with the paradox that they are ultimately less controlled than the administrations themselves (Loo, 2019).

On the other hand, the fuel for surveillance—data—is produced by the people themselves. The concept of an ecosystem is relevant: to exercise their control, administrations will increasingly depend on data voluntarily produced by the citizens themselves.

Finally, states do not directly ask citizens to further cooperate in providing information. There is a growing integration of private and public data systems for security and surveillance purposes. There are, for example, private camera systems with facial recognition sold by large companies and offered to be connected to police departments (Haselton, 2019). In addition, public institutions collect individual data posted by social media companies,[26] and companies are increasingly encouraged to collect and share data with the state (Elkin-Koren & Gal, 2019).

Customs may be very tempted to develop 'soft' control: an extensive use of data, including that stored by companies, to impose less heavy control on those who are controlled, but at the expense of more secrecy, more intermediaries in the chain of control, and less control of the controllers.

In addition, we should already consider the medium-term consequences of a systemic data sharing between public institutions and private companies. It might cause growing mistrust on the part of users who could refuse to provide their data, or even worse, provide fake data.[27] As the exchange of data between government and businesses grows, citizens' trust is likely to decrease, the quality of the data may deteriorate, and the data collected by the private sector may become irrelevant for all actors for surveillance purposes, but also for innovation, which is another use of data by the private sector (Elkin-Koren & Gal, 2019).

Ethical issues related to AI are very present in the mainstream media, but, in AI conferences, ethics represents ultimately a very small share of the communications (Crawford et al., 2019, from p. 45). Within national strategy documents, ethical issues often feature prominently, which does not mean that concrete decisions are taken at the level of the administrations. These ethical issues are barely discussed in customs administrations, which stick to existing general legal limits but do not set up specific bodies to control and envision the ethical consequences of the use of data technologies.

In the absence of a joint effort on ethics, the risk is twofold at the global level. First, the extensive use of data by Customs may generate new legal obstacles to exchange information between customs administrations. Second, there may be a proliferation of national principles and standards, which would allow private companies to adopt market strategies, choosing the countries where the rules are not an obstacle to their uses of AI and data (Floridi & Cowls, 2019). Customs and tax administrations are already aware of this shopping strategy with tax regulations.

4.2 The value of equality before the tax law

What is in balance with surveillance is the security of citizens, following a disputed and widespread rationale according to which 'more security implies more surveillance and more data'. One point is rarely put forward: the balance between surveillance and equality, especially equality before the tax law.

For example, in technical terms, this would mean ensuring that the machine does not deploy a discriminatory control strategy: the machine may assess that it is more profitable to systematise control over small importers than large ones, because the former are subject to more errors giving rise to minor sanctions. However, when detected in large numbers, these minor sanctions would generate more extra-revenue than the litigations on large importers whereby outcomes are more uncertain because investigations are more complex.

When the 2020 finance law authorised the French customs and tax administrations to automatically collect data on social networks, the body responsible for ensuring compliance with the French Constitution recognised that it was a violation of the right to privacy but that equality before the tax law was also a cardinal value of society and that, provided certain technical safeguards were in place, the equality before the tax law should prevail.

Citizen security is a state function, associated with a monopoly on violence, but taxation is another state monopoly and it is based on equality of citizens before tax. Although the balance between surveillance and security can be assessed at the individual level—'to what extent am I ready to be monitored more to be more secure?'—the balance between surveillance and taxation is more complex, for it is based on a collective assessment, as the tax is above all a relationship between the individual and society.

Customs should therefore not line up behind the police in the surveillance and security debate. Customs and tax administrations are bringing new fundamental questions: the balance between privacy and tax equality at the time when individual assets are extremely complex to assess by administrations (Cantens, 2018).

4.3 Customs responses for an innovative place in the new ecosystems

The first response is to share customs data with the outside world. It is more legitimate for states to request and collect data when they also release data themselves. The HMRC (Her Majesty's Revenue & Customs) 'lab' thus offers researchers an opportunity to access individual and transaction-level data under strict security conditions (de-identified data, secure room, no internet connection, checks of data extractions) (Almunia et al., 2019). Another option is to generate a sample of importers/exporters that is statistically representative of the population of economic actors and build a shareable and annually updated dataset (Burdick et al., 2019). Customs data, at the transaction level, has a very important value for commercial interests and economic policies.

Transaction-level data is important at the local level too. At the border, Customs can become a 'data hub' for all stakeholders since Customs centralise the data of logistics actors, brokers, importers and exporters. The administration can also play the role of an objective 'evaluator', quantifying the dysfunctions caused by all actors including Customs.

A second response is the connection to the community of programmers. Several American administrations have put their algorithms online, free of charge, notably in relation to passenger control. In December 2017, the US Department of Homeland Security offered financial prizes in the Passenger Screening Algorithm Challenge to improve the accuracy of algorithms that detect threats at the airports (Kaggle, 2017). Datasets have been made available to programmers for this challenge. In 2019, another challenge was launched to detect an actual exporter in customs documents (Burdick et al., 2019). Finally, in 2020, the WCO research unit, in cooperation with Korean academics, shared a risk analysis algorithm with the data science community (Sundong et al., in press). This transparency sometimes leads to co-development of governance. In the case of a health administration, an American citizen demonstrated that the administration's chosen algorithm was less efficient than his own.

The third response is the use and promotion of open source software. Open source software does not necessarily cost less when it comes to scaling up an algorithm in the IT customs system. The advantage is communicability. By adopting free tools of contemporary data science, Customs become part of global user communities, benefitting from solutions and scientific updates, and are not locked into proprietary systems. In addition to the possibility of exchanging more simply between customs administrations, the use of open source software fosters trust in general and increases the possibility for an administration to communicate details of its tools, including during litigations. As national key players in the fields of taxation, trade and security, Customs can play a leading role by pushing governments to adopt data science open source software as an administrative standard.

The fourth response is a human resources policy, which increases data literacy in general and dedicates resources to the training of customs officers. The worst-case scenario would be high-tech companies selling products to administrations with no technological culture that are only concerned with the perfection of enforcing the law. The increasing use of algorithms in the police and justice systems has resulted in a fear of deskilling among police, prosecutors and judges (Brayne & Christin, 2020).

The United Kingdom has created a campus managed by the Office of National Statistics and universities, proposing curricula in data science and public administration applications.[28] Distance learning courses are open to civil servants, thus forging a common culture. These courses are also open to the public, which is an opportunity for administrations to recruit graduates who will have been trained on public administration issues. South Korean Customs has implemented a long, joint training course for customs officials on data science tools, led by academics. These two initiatives among others are a part of the philosophy of data: what matters initially is not so much the mastery of a computer language or statistical technique, but the 'domain knowledge', the familiarity with the data. In all customs administrations, there are computer specialists, engineers, statisticians, investigators, front-line inspectors who gradually acquire 'practical knowledge', an empirical and intuitive intimacy with the data on which they work daily. One has to be able to ‘walk’ through the data as one would walk in a landscape and look for order, trends, and anomalies. It is a common saying that eighty per cent of the time of the data analysis cycle is spent preparing and cleaning the data, but this is not wasted time, it is the time needed to be put *into* the data.

Training is not sufficient; the administrations must retain talented staff. The situation is similar to that which we experienced more than 20 years ago during the advent of the internet. It was difficult to recruit specialists in new communication technologies. Some customs administrations overcame this difficulty by offering a meaningful working environment for young graduates, in particular by offering them good infrastructure and complex technical challenges. For many data scientists, it will be motivating to work on issues related to the protection of society.

5. Conclusion and some policy recommendations

The incredible ambition of Big Data is the idea that we could make a decision based on a perfect argument, because we would have all the data relating to the problem. However, when achieving this stage of perfection, we must delegate the freedom to work on this data to machines. These new paradigms of science and governance generate new ‘data ecosystems’. Regardless of the terminology, technological developments have resulted in an increasingly close interweaving among states, citizens, and companies handling data on a global scale.

Today, the most advanced customs administrations have deployed techniques based on data analytics; nevertheless, none embraces, either strategically or technically, a wide range of the possibilities offered by data and data science. In addition, most customs administrations will find it difficult to explore alone expensive technologies that may not provide immediate results. This conclusion outlines some ideas that the rapid evolution of the data ecosystems should not make obsolete in the short term:

1. keeping a wide scope. More areas than only fraud deserve exploration and projects: revenue forecasting, border security based on geodata, optimisation of patrols in the field, topology of customs units according to trade routes, economic projections, performance measurement including fight against corruption;

2. working together at the global level to make data technology contribute to the strengthening of the customs community. Some projects could animate the community: produce common standard datasets to be used to assess the performance of algorithms, share algorithms and models to promote peer review among customs experts;

3. developing know-how on exploratory data analysis. Too many experts are immediately curious to apply ‘models’, while it is not the philosophy of the new data-centric approaches. Data comes first, it is necessary to develop, among specialists, the taste and the capacity to ‘walk in the data’; and

4. favouring free tools of data science and their appropriation for customs purposes. Proprietary software or commercial off-the-shelf software are designed at a given time and are state-of-the-art regarding the problems the administration wants to treat. Their format and functions force analysts to adopt particular thinking patterns. A maximum of freedom and flexibility should be given to analysts to create their own toolbox. With this in mind, some basic open software and languages are better than commercial software.

Looking back over the past years, we can be optimistic. When the WCO put mathematics on the PICARD conference agenda in 2015, to say that the participants expressed little enthusiasm would be an understatement. Five years later, in 2020, the Secretariat has launched a cloud computing platform with BACUDA, paving the way to explore the possibilities offered by data while maintaining an experimental approach. By developing its data literacy, the customs community brings new questions to the public debate, such as the equality before the tax law, and supports innovative approaches to elaborate and assess public policies in fiscal and trade governance.

We do not need to dream of algorithms; they are already here. New 'autumns' for AI may come to be, but new winters are unlikely, particularly for data, since we cannot help producing data. The final words are, therefore, for those who still mistrust the emergence of governance-by-data. In societies where machines are *de facto* more involved in public administration, data may be our best collective safeguard. Machines are no more than fuelled by the data we all produce, be it administrations', companies' or individuals' data. The same data can be mobilised to combat political arbitrariness and inequalities. As we all produce data, isn't governing by data an opportunity to forge more transparent and collective policymaking, taking into account all data, therefore everyone's data?

References

Almunia, M., Harju, J., Kotakorpi, K., Tukiainen, J., & Verho, J. (2019). Expanding access to administrative data: the case of tax authorities in Finland and the UK. *International Tax and Public Finance, 26*(3), 661–676.

Andrews, L. (2019). Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'. *Public Administration, 97*(2), 296–310.

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review, 82*(5), 977–1008.

Brayne, S., & Christin, A. (2020). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems* spaa004. https://doi.org/10.1093/socpro/spaa004.

Brooks, R. (1991). Intelligence without reason. In *Proceedings of IJCAI'91*. https://www.ijcai.org/Proceedings/91-1/Papers/089.pdf

Burdick, D., Krishnamurthy, R., & Raschid, L. (2019). DSMM'19: The 5th workshop on data science for macro-modeling with financial and economic datasets. *Proceedings of the 2019 International Conference on Management of Data* (pp. 2068–2069).

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177%2F2053951715622512

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Canella, G. (2018). Racialized surveillance: Activist media and the policing of black bodies. *Communication Culture & Critique, 11*(3), 378–398.

Cantens, T. (2015). Mirror analysis: Customs risk analysis and fraud detection. *Global Trade & Customs Journal*, 10, 207.

Cantens, T. (2017). The Political Arithmetic of Borders: Towards an Enlightened Form of Criticism. *The AntiAtlas Journal*, 2017, No. 2.

Cantens, T. (2018). Of taxation, instability, fraud and calculation. *Economy, Crime and Wrong in a Neoliberal Era*, 36, 116.

Cantens, T., Ireland, R., & Revesz, P. (2014). *The why and how of performance measurement contracts.* World Customs Organization. http://www.wcoomd.org/-/media/wco/public/global/pdf/topics/research/guide/wco_pmc-guide.pdf

Cantens, T., Raballand, G., & Bilangna, S. (2010). Reforming Customs by measuring performance: a Cameroon case study. *World Customs Journal*, *4*(2), 55-74.

Cardon, D. (2015). A quoi rêvent les algorithmes. Nos vies à l'heure: Nos vies à l'heure des ... – Dominique Cardon – Google Livres. Editions du Seuil.

Cariolle, J., Chalendard, C., Geourjon, A.-M., & Laporte, B. (2019). Measuring and improving the performance of customs valuation controls: An illustration with Gabon. *The World Economy, 42*(6), 1850–1872.

Chalendard, C., Raballand, G., & Rakotoarisoa, A. (2019). The use of detailed statistical data in customs reforms: The case of Madagascar. *Development Policy Review, 37*(4), 546–563.

Chermiti, B. (2019). Establishing risk and targeting profiles using data mining: Decision trees. *World Customs Journal*, *13*(2), 39–58.

Chesnais, J.-C. (1995). Nicholas Eberstadt. The tyranny of numbers. Mismeasurement and misrule. *Politique étrangère, 60*(4), 1059–1060.

Chesterman, S. (2008). We can't spy… if we can't buy!': The privatization of intelligence and the limits of outsourcing 'inherently governmental functions'. *European Journal of International Law, 19*(5), 1055–1074.

Chesterman, S. (2019). Artificial intelligence and the problem of autonomy. *Notre Dame Journal on Emerging Technologies, 1*, 220–250.

Choi, Y. (2019). Identifying trade mis-invoicing through customs data analysis. *World Customs Journal*, *13*(2), 59–76.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E.. . . Whittaker, M. (2019). *AI Now 2019 report.* AI Now Institute, New York University.

Crockett, K., Stoklas, J., O'Shea, J., Krügel, T., & Khan, W. (2018). Adapted psychological profiling verses the right to an explainable decision. *IJCCI 2018 – Proceedings of the 10th International Joint Conference on Computational Intelligence* (p. 14). SCITEPRESS.

Dawes, S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly, 33*(1), 15–27.

de Gaulejac, V. (1990). De la mesure en tout! Evaluation et développement social. *Les Annales de la Recherche Urbaine. 47*, pp. 117–122. Centre de Recherche d'Urbanisme.

Deloitte. (2018). *16 artificial intelligence projects from Deloitte: Practical cases of applied AI.* https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/innovatie/deloitte-nl-innovatie-artificial-intelligence-16-practical-cases.pdf

Devins, C., Felin, T., Kauffman, S., & Koppl, R. (2017). The law and big data. *Cornell JL & Public Policy*, 27, 357.

Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F., & Pagliuca, D. (2020). Combining network visualization and data mining for tax risk assessment. *IEEE Access, 8*, 16073–16086.

Dobbe, R., Dean, S., Gilbert, T., & Kohli, N. (2018). A broader view on bias in automated decision-making: Reflecting on epistemology and dynamics. *arXiv preprint arXiv:1807.00553*.

Dutton, T., Barron, B., & Boskovic, G. (2018). *Building an IA world: Report on national and regional AI strategies.* CIFAR. https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf?sfvrsn=fb18d129_4

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society, 5*, 40–60.

Elkin-Koren, N., & Gal, M. (2019). The chilling of governance-by-data on data markets. *University of Chicago Law Review, 86*, 403–431.

Engstrom, D., & Gelbach, J. (2020). Legal tech, civil procedure, and the future of adversarialism. *University of Pennsylvania Law Review.* Forthcoming.

Ensign, D., Friedler, S., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2017). Runaway feedback loops in predictive policing. *arXiv preprint arXiv:1706.09847*.

European Commission (EC). (2018, July 18). *Antitrust: Commission fines Google €4.34 billion for illegal practices regarding Android mobile devises to strengthen dominance of Google's search engine.* Press release. https://ec.europa.eu/commission/presscorner/detail/en/IP_18_4581

European Union (EU). (2020). *iBorderCtrl. The project.* https://www.iborderctrl.eu/The-project

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Geiger, R., Yu, K., Yang, Y., Dai, M., Qiu, J., Tang, R., & Huang, J. (2020). Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from? *Fairness, Accountability and Transparency.* ACM.

Government of Japan. (n.d.). *Society 5.0.* https://www8.cao.go.jp/cstp/english/society5_0/index.html

Government of Japan. (2015). 5th *Science and Technology Basic Plan - Society 5.0*

Grigoriou, C. (2019). Revenue maximisation versus trade facilitation: the contribution of automated risk management. *World Customs Journal*, *13*(2), 77–90.

Grigoriou, C., Kalizinje, F., & Raballand, G. (2019). How helpful are mirror statistics for Customs reform? Lessons from a decade of operational use. *World Customs Journal*, *13*(2), 105–114.

Hacking, I. (1975). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge University Press.

Haselton, T. (2019, August 3). Everyone's talking about this Amazon app that lets police see camera footage – here's what it's like. *CNBC.* https://www.cnbc.com/2019/08/02/amazon-ring-neighbors-app-sends-video-to-police-departments.html

Hemberg, E., Rosen, J., Warner, G., Wijesinghe, S., & O'Reilly, U.-M. (2016). Detecting tax evasion: A co-evolutionary approach. *Artificial Intelligence and Law, 24*(2), 149–182.

House of Lords. (2018). *AI in the UK: ready, willing and able?* Select Committee on Artificial Intelligence, Report of Session 2017–2019. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Hsu, K.-W., Pathak, N., Srivastava, J., Tschida, G., & Bjorklund, E. (2015). Data mining based tax audit selection: A case study of a pilot project at the minnesota department of revenue. In K.-W. Hsu, N. Pathak, J. Srivastava, G. Tschida, & E. Bjorklund (Eds.), *Real world data mining applications,* (pp. 221–245). Springer.

Hua, Z., Li, S., & Tao, Z. (2006). A rule-based risk decision-making approach and its application in China's customs inspection decision. *Journal of the Operational Research Society, 57*(11), 1313–1322.

Humby, C. (2006). Data is the new oil. *Proceedings of the ANA Senior Marketer's Summit.* Evanston, IL, USA.

ITU News. (2019, January 10). Smart villages: Empowering rural communities in 'Niger 2.0'. *ITU News.* https://news.itu.int/smart-villages-empowering-rural-communities-in-niger-2-0/

Izant, C. (2017). Equal access to public communications data for social media surveillance software. *Harvard Journal of Law & Technology, 31*(1), 237–257.

Jaccard, N., Rogers, T., Morton, E., & Griffin, L. (2017). Detection of concealed cars in complex cargo x-ray imagery using deep learning. *Journal of X-Ray Science and Technology, 25*(3), 323–339.

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2019). The sustainable value of open government data. *Journal of the Association for Information Systems, 20*(6), Article 6.

Kaggle. (2017). *Passenger Screening Algorithm Challenge.* https://www.kaggle.com/c/passenger-screening-algorithm-challenge

Kalizinje, F. (2018). Combating customs revenue fraud in WCO East and Southern African Region: A mirror analysis through the lens of Malawi. *Global Trade and Customs Journal, 13*(5), 224–233.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1–12. https://journals.sagepub.com/doi/pdf/10.1177/2053951714528481

Kluttz, D., & Mulligan, D. (2019). Automated decision support technologies and the legal profession. *Berkley Technology Law Journal, 34,* 853–889.

Kolokytha, S., Flisch, A., Lüthi, T., Plamondon, M., Schwaninger, A., Vasser, W., Hardmeier, D., Costin, M., Vienne, C., Sukowski, F., Hassler, U., Dorion, I., Gadi, N., Maitrejean, S., Marciano, A., Canonica, A., Rochat, E., Koomen, G., & Slegt, M. (2017). Improving Customs' border control by creating a reference database of cargo inspection x-ray images. *Technology and Engineering Systems Journal, 2*(3), 60–66.

Laporte, B. (2011). Risk management systems: Using data mining in developing countries' customs administrations. *World Customs Journal, 5*(1), 17–27.

Loo, R. (2019). The missing regulatory state: Monitoring businesses in an age of the missing regulatory state: Monitoring businesses in an age of surveillance. *Vanderbilt Law Review, 72*(5), 1563–1631.

Lord, J., & Culling, A. (2016). OR In HM Revenue & Customs. *Impact, 2*(2), 30–33.

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance, 13*(5), 14–19.

Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big data & society*, *1*(2), https://doi.org/10.1177/2053951714541861.

MacMillan, D., & Dwoskin, E. (2019, August 23). The war inside Palantir: Data-mining firm's ties to ICE under attack by employees. *The Washington Post.* https://www.washingtonpost.com/business/2019/08/22/war-inside-palantir-data-mining-firms-ties-ice-under-attack-by-employees/

Malinowski, P. (2019). Application of modern technology for migration management. *European Law Enforcement Research Bulletin* (4 SCE), 29–33.

Manokha, I. (2018). Surveillance: The DNA of platform capital — The case of Cambridge Analytica put into perspective. *Theory and Event, 21*(4), 891–913.

Merz, F. (2019). Europe and the global AI race. *CSS Analyses in Security Policy, 247*.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society.* https://doi.org/10.1177/2053951716679679

Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics, 38*(8), 1–4.

Moore, G. (1975). Progress in digital integrated electronics. *IEEE, IEDM Tech Digest*, 11–13.

NITI Aayog. (2018). *National strategy for artificial intelligence #AIFORALL.* https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

The Organisation for Economic Co-operation and Development (OECD). (2019). *Artificial intelligence.* https://www.oecd.org/going-digital/ai/

The Organisation for Economic Co-operation and Development (OECD). (n.d.). *Open government data*. https://www.oecd.org/gov/digital-government/open-government-data.htm

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Okazaki, Y. (2018). *Unveiling the potential of blockchain for Customs*. WCO Research Paper No. 45.

Open Government Partnership. (2019). *Democracy beyond the ballot box: OGP's 2019 annual report*. https://www.opengovpartnership.org/

Oxford Insights. (2018). *Towards an AI strategy in Mexico: Harnessing the AI revolution.* https://www.oxfordinsights.com/mexico

Paunov, C., Planes-Satorra, S., & Ravelli, G. (2019). *Review of national policy initiatives in support of digital and AI-driven innovation.* OECD Science, Technology and Industry Policy Papers, No. 79, OECD Publishing. https://doi.org/10.1787/15491174-en

Peeters, R., & Schuilenburg, M. (2018). Machine justice: Governing security through the bureaucracy of algorithms. *Information Polity*, *23*(3), 267–280.

Piscopo, C., & Birattari, M. (2008). The metaphysical character of the criticisms raised against the use of probability for dealing with uncertainty in artificial intelligence. *Minds and Machines*, *18*(2), 273–288.

Polner, M. (2018). Harnessing the power of data for customs enforcement. *Global Trade and Customs Journal, 13*(4), 182–185.

Richardson, R., Schultz, J., & Crawford, K. (2019a). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review, 94,* 192–233.

Richardson, R., Schultz, J., & Southerland, V. (2019b). *New challenges to government use of algorithmic decision systems.* AI Now Institute.

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, (pp. 153–158).

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms.* Cambridge University Press.

Shapiro, A. (2019). Predictive policing for reform? Indeterminacy and intervention in big data policing. *Surveillance and Society, 17*(3/4), 456–472.

Stanford. (2020). *Artificial intelligence index report.* HAI Stanford.

Sundong K., Tsai, Y-C., Singh, K., Choi, K., Ibok, E., Li, C-T. & Cha, M. (in press). DATE: Dual Attentive Tree-aware Embedding for Customs Fraud Detection. To appear in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. San Diego, CA, USA, August 23-27, 2020.

Supiot, A. (2015). *La gouvernance par les nombres: cours au Collège de France, 2012–2014.* Institut d'études avancées de Nantes.

UK Government. (2017). *Government response to wider call for evidence from the House of Lords Committee on AI.* http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/73250.html

UK Government. (2018). *Industrial strategy: Artificial intelligence sector deal.* https://www.gov.uk/government/publications/artificial-intelligence-sector-deal

UK Government. (n.d.). *Office of Artificial Intelligence.* https://www.gov.uk/government/organisations/office-for-artificial-intelligence

US Government. (2019). *Executive Order on AI.* https://www.whitehouse.gov/ai/

US Government. (2020). *2020 Action Plan*. https://strategy.data.gov/action-plan/

Vanhoeyveld, J., Martens, D., & Peeters, B. (2019). Customs fraud detection. *Pattern Analysis and Applications*, *23*, 1457–1477. https://doi.org/10.1007/s10044-019-00852-w

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 842–887.

Waldrop, M. (2016). The chips are down for Moore's Law. *Nature, 530*(7589), 144–147.

Wier, L. (2020). Tax-motivated transfer mispricing in South Africa: Direct evidence using transaction data. *Journal of Public Economics, 184*, 104153.

World Bank. (2018). *Using big data to achieve the SDGs.* https://www.worldbank.org/en/news/feature/2018/04/27/using-big-data-to-achieve-the-sdgs

World Customs Organization (WCO). (2019). *Study report on disruptive technologies.* http://www.wcoomd.org/en/topics/facilitation/instrument-and-tools/tools/disruptive-technologies.aspx

World Customs Organization (WCO). (2020, February 19). BACUDA: Supporting Customs with data analytics. *WCO News.* https://mag.wcoomd.org/magazine/wco-news-91-february-2020/bacuda/

World Economic Forum. (2019). *A framework for developing a national artificial intelligence strategy.* http://www3.weforum.org/docs/WEF_National_AI_Strategy.pdf

Wylie, C. (2019). Mindf*ck: *Cambridge Analytica and the plot to break America*. Random House Publishing Group.

Xiao, Z., Xiao, H., & Wang, Y. (2016). A risk decision-making approach to customs targeting. *Open Cybernetics and Systemics Journal, 10*, 250–262.

Yaqin, W., & Yuming, S. (2010). Classification model based on association rules in customs risk management application. *Proceedings – 2010 International Conference on Intelligent System Design and Engineering Application, ISDEA 2010. 1*, pp. 436–439. IEEE Computer Society.

Zhou, X. (2019). Data mining in customs risk detection with cost-sensitive classification. *World Customs Journal*, *13(2),* 115–130.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*, 75–89.

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity, 19(1, 2),* 17–33.

Notes

1 This article is based on the keynote speech at the WCO PICARD Conference held in Skopje, North Macedonia on 22–24 October 2019, delivered by the Secretary General of the WCO. The authors thank Ricardo Treviño, Yeon-Soo Choi, Shingo Matsuda, Rachel McGauran, Michelle Medina, Hans Pieters and Mariya Polner for their helpful comments.

2 Addressing the distinctions among all these technologies is out of the scope of this paper, which relies on a simple distinction between, on the one hand, data—that are increasingly produced and made available—and on the other hand, machines that treat them.

3 The massive use of data paves the way to ‘personalised law’, the possibility to adapt the law to each individual given their records (Devins et al., 2017). For example, in the US, around 200 risk analysis tools are used in justice procedures to support decision-making (Peeters et al., 2018).

4 For instance, Niger customs that have scarce resources to control a vast territory has deployed an IT network in more than 90 per cent of the customs units.

5 See WCO (2020), introductory page.

6 Didimo et al. (2002) and the Mexican tax revenue agency experimented with ML to guide its control strategy; it estimated that AI provided comparable outcomes in three months than manual work in 18 months, Towards an AI strategy in Mexico (Oxford Insights, p. 21).

7 World Bank (2018); see also ‘smart villages’ in Niger (ITU News, 2019) and the development of rural exchanges in India (NITI Aayog, 2018).

8 Up until recently, there was the famous Moore’s law on the computer’s capacities doubling every two years. Now that this increase has slowed down, this relative stagnation is considered as an opportunity to develop new technologies that will provide even more celerity than before (Moore 1965, 1975; Waldrop 2016).

9 https://res.cloudinary.com/yumyoshojin/image/upload/v1/pdf/future-data-2019.pdf and https://res.cloudinary.com/yumyoshojin/image/upload/v1/pdf/cloud-business-2020.pdf

10 Public open data is data collected or generated by public administrations and released publicly, either to restricted third parties like researchers or to the public domain.

11 https://www.data.gov/

12 http://etalab.gouv.fr. Etalab is managing the development of the governance-by-data in France, fueling and managing different websites as a warehouse for open data, a repository of open source software, shared code, and a series of use cases. Twenty-one projects were launched after a call for projects for administrations.

13 https://www.openfiscaldata.go.kr (which is only a small part of all Korean open data websites).

14 See for instance https://sill.etalab.gouv.fr/fr/software, in France, and the US National Geospatial Agency initiative https://home.gs.mil/developer

15 See OECD for a selection of national and international initiatives http://www.oecd.org/going-digital/ai/initiatives-worldwide/

16 http://www.ansi.ne//motdg

17 https://www.etalab.gouv.fr/

18 The Cambridge Analytica case unveiled the possibility to use personal data for political purposes by private companies, as well as the fact that these companies were conducting such operations worldwide before the scandal (Cadwalladr & Graham-Harrison, 2018; Wylie, 2019). In a last case opposing European Union and Google, the company was fined for having 'imposed restrictions on […] manufacturers and mobile network operators to cement its dominant position in general internet search' (European Commission, 2018).

19 Australian Government, Vision 2025. https://www.dta.gov.au/digital-transformation-strategy

20 See for instance https://www.douane.gouv.fr/la-douane/opendata

21 https://www.gov.uk/guidance/imports-and-exports-datasets

22 https://beta.sam.gov/ The US IRS put data as one of their six strategic objectives for 2018–2022. Between 2007 and 2017, the number of IRS users has been multiplied by 23, and the amount of data by 100. https://www.irs.gov/pub/irs-pdf/p3744.pdf

23 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

24 The 'data ecosystems' terms are often used in literature to define the relationships between states, citizens, and businesses to circulate and use data. Data ecosystems are supposed to foster peoples' participation in public affairs (Zuiderwijk et al. 2014).

25 See the example of 'smart walls', https://www.wired.com/story/palmer-luckey-anduril-border-wall/

26 For instance, French customs and Internal Tax administration have been allowed by the 2020 law of finance to experiment with data collection on social networks for 3 years (2020 law of finance, article 154).

27 Civil disobedience regarding data provision is easy to imagine. Some companies prefer that their employees declare that they want to withdraw from some projects related to police, migration or intelligence services, rather than having to face whistleblowers later on (MacMillan & Dwoskin, 2019).

28 https://datasciencecampus.ons.gov.uk/

Kunio Mikuriya

Dr Kunio Mikuriya has been Secretary General of the World Customs Organization (WCO) since 1 January 2009. Prior to joining the WCO, he worked for Japan's Ministry of Finance for 25 years. During his career with the Ministry, Dr Mikuriya occupied a variety of senior posts, which have given him broad experience and knowledge in Customs, trade, development, budget, and financial policies. Dr Mikuriya has a degree in law from the University of Tokyo and a PhD in international relations from the University of Kent.

Thomas Cantens

Dr Thomas Cantens currently works as resident Advisor of the Director General of Niger Customs. Formerly, he was the head of the Research Unit at the World Customs Organization, Brussels. He is an Associate Researcher at the research centre on international development of the Auvergne University (CERDI, France). He is a French Customs officer and has served in investigation and intelligence in France and in three sub-Saharan Customs administrations. Having initially graduated as an engineer (POLYTECH Montpellier, France), he holds a PhD in social anthropology (EHESS, France) and a certificate in geospatial analysis (ENSG, France). He has published several works on customs reform in developing countries.

This paper is published in the Special 2020 Edition of the World Customs Journal (Vol.14 No.2) and has been reproduced with the permission of the publishers. Citation: Mikuriya, K. & Cantens, T. (2020) If algorithms dream of Customs, do customs officials dream of algorithms? A manifesto for data mobilisation in Customs. World Customs Journal 14(2) pp. 3-22.