

Scientists and Practitioners: the JRC and Italy's customs on fraud detection

Marco De Andreis – Italy's Customs Agency

Abstract: I take stock of a collaboration between the JCR scientists and Italy's Customs Agency practitioners on innovative methods for fraud detection – a collaboration which has been going on for nearly a decade. To exploit the great potential offered by these methods, EU customs organizations should avoid internal compartmentalization or, even better, the EU should overcome the current fragmentation of its customs policy among 28 different national actors.

I thank the organizers for having invited me to this important conference in this astoundingly beautiful venue.

I've been heading a unit of Italy's Customs Agency dealing with statistics and data since the year 2008. It currently goes under the name *Statistica e open data* – no translation needed here.

Back when I started, my unit – then called economic and fiscal studies – used to crunch numbers pertaining to various issues of interest to the Agency. In their spare time two of the best and the brightest among us also tried to come up with some smarter, more math-based approach to produce and select risk profiles and thus help catching smugglers, counterfeiters and fraudsters in general – we were indeed assigned to the antifraud directorate general of the Agency. But our frontline colleagues were not very impressed by our studies and we retreated into the production of simple and non-controversial tables and graphs on Italy's foreign trade, mainly extra EU trade.

Then something unexpected happened – chance, you see.

We noticed that oftentimes our data on Italy's foreign trade diverged substantially with respect to those published by ISTAT, Italy's official statistical office, for the same period of time. Actually, it was the agency's director who noticed that. And he was not happy.

The thing puzzled us, because ISTAT's data and ours were based on the same source: the Single Administrative Declarations (SADs) that no one but our Agency passed along to ISTAT every month. The best way to solve the riddle was to get in touch with the statistical office, which we did. Only to discover that they did a great amount of work on our data. Or rather on the most glaring outliers in our data's distribution[s]. They did – and do – a fastidious work of data cleansing that might – and may - even lead them to contact the specific trader or customs official involved in a given transaction to ask them questions and clarifications.

We looked at the results of their work with lust: we would have been very happy to incorporate their corrections in our data for a number of obvious reasons – including pacifying our director’s nervousness on the misalignment between ISTAT’s and our own statistics on Italy’s foreign trade. But this could not be done due to so-called statistical confidentiality: statistical offices in the EU are forbidden by law to turn information gathered for statistical purposes to other branches of the government – particularly, I should say, to people like us, tax collectors.

So, we now knew that among our SADs might lurk some very misleading piece of information and as a consequence we had to embark on a new and potentially painful line of work: data cleansing. But remember, we were attached to the antifraud directorate of Italy’s Customs. Most of us saw themselves primarily as customs officers, not statisticians. They saw our data essentially as useful tools to catch fraudsters, not as sources of reliable statistics on foreign trade.

It was at that point that one of us, Alessandro Scatolini, an engineer by training, browsing the web in search of reliable methods on how to catch outliers in large data bases, stumbled upon an article published in the Magazine of the Italian Society of Statisticians in april 2009 and titled *Quando lo statistico sente puzza di bruciato*, or “When the statistician smells a rat”. Authors? Andrea Cerioli and Domenico Perrotta.

That truly closed the circle. Here came two scientists statisticians who explained to us that “outliers are not necessarily observations to be discarded. To the contrary, an accurate examination of their causes may provide substantial advantages in many cases. Fraud detection is a field of study where outliers are center stage and is one of the most fruitful applications of data mining”.

The next logical step was to dispatch Alessandro Scatolini and his colleague Teresa Mari Ceres – a statistician by training – to Ispra’s JRC to

meet and talk to Messrs. Cerioli and Perrotta. That's how a collaboration between scientists and practitioners that is now almost a decade long started.

Thanks to it, Ceres and Scatolini set up a database containing all SADs registered by Italy's customs since December 2008. They used it to create a system for the identification of anomalous customs declarations – *Sistema per l'Identificazione di Dichiarazioni Doganali Anomale* (SIDDA). That was a big gain on our side on which I will return shortly.

In exchange the JRC got our unflinching support for whatever initiative it may undertake in the field of fraud detection and access to our data to run tests and simulations when needed, including more recently on the Benford's law. In this regard, we turned to the JRC anonymized data from a number of SADs and a list of detected frauds by underinvoicing to test the effectiveness of their algorithm inferred from Benford's law.

But of course on the usefulness of our relationship for the JRC only the JRC can tell.

I just said that SIDDA was our big gain. It had, in fact, an obvious, positive impact on post clearance controls. By defining limits of acceptability for specific values (€ per kg, € per supplementary unit, kg per supplementary unit) of a given population (defined in its turn by the same TARIC and country of origin), a few risk profiles were inferred. These showed great potentiality: they increased of an order of magnitude the probability of detecting certain kinds of frauds.

You may ask: why potentiality instead of actuality? Alas, here begins the sad part of this narrative. Because SIDDA never made it to the mainstream of the Agency's risk assessment and consequent allocation of that precious commodity which are customs controls. Thanks to the efforts of Ceres and Scatolini it was made available to customs official in

the field, but as no more than an adjunct to the traditional methods of fraud detection.

In October 2017 the unit was disbanded altogether. Half of its personnel, including Scatolini and Ceres, was kept in the antifraud directorate. They tried to keep the interest toward, and the use of, SIDDA alive but without much success. On the one hand, the business as usual approach to risk profiling prevailed. On the other hand, the Technology directorate of our Agency, the guardian of data, decided to discontinue the provision of SADs to the SIDDA data base.

The other half of the personnel, including this speaker, was assigned to a new unit – Statistics and Open Data, mentioned at the beginning of this talk – within, guess what, the Technology directorate. Our main mission is to satisfy the request of customs data from internal and external sources, guaranteeing at the same time their quality. The unit also represents the Customs Agency within the National Statistics System (of Public Administrations), coordinated by ISTAT.

It's clear that in the end bureaucratic compartmentalization prevailed, making even more difficult the dialogue between the fraud-fighters, i.e. the practitioners-practitioners, and those who try to make use of the customs data, i.e. the statisticians-practitioners. This is perhaps the most important lesson that can come out of this story: if you want to develop innovative antifraud approaches, keep the more data inclined of your personnel – those who can sustain a dialogue with, for example, the people at the JRC – in constant touch with the fraud-fighters.

The good news, however, is that Alessandro Scatolini has been assigned once again to my unit, and that since we now have a direct access to SADs we are about to reconstitute SIDDA. We will try again to persuade the antifraud directorate to make good use of it and hopefully, this time

around, we will succeed. We should also be ready pretty soon to resume our collaboration with the JRC on these issues.

As you may imagine, SADs are not the only source of the data produced every instant by such a complex administrative machine as a customs organization. We are still in the process of trying to understand who has what data and get access to it – since we are supposed to certify their quality. *E pluribus unum* is a wonderful motto, whose corresponding reality is never easy to achieve.

And apropos of *e pluribus unum* vs bureaucratic compartmentalization, we should never forget that the biggest such contradiction is the customs union itself. Customs is an exclusive competence of the EU, the customs code is an EU regulation, but we still have 28 different organizations to implement it. This is unfortunate, to say the least, for everybody.

Cooperation among member states' customs and between them and the Commission remains slow and cumbersome. Even the simple collection of information for the sake of measuring the results of customs activity in the EU is made difficult by member states' reticence. "In the absence of a clear legal basis for the Customs Union Performance project, key data remain unavailable in the case of some member states" - stated the 2016 CUP Annual Report.

Data on commercial operators and operations (SADs) are kept at national level and exchanged only through administrative cooperation – which is slow and cumbersome. In the era of big data, member states' customs have no direct access to a common pool of information on who has traded what across EU borders. This makes also difficult, if not impossible, to develop effective risk management at the EU level, despite the efforts underway in the Common Risk Management Framework (CMRF). Information gathered through member states' intelligence organizations is even less prone to be shared.

The range of variation among member states on customs controls, their types and their results is still too wide to be acceptable - it's per se an incentive to fraudsters to choose their preferred routes. I could go on like this for several more minutes.

But one thing is clear: unless we manage to make of the customs union a real union – *e pluribus unum* – the likelihood that the efforts of our scientists to improve our fraud detection capability go wasted is, unfortunately, rather high. We Europeans had the wisdom of setting up one (Joint) Research Center. We are not equally wise on the customs side of the equation. Today's situation it's like having one scientist and 28 practitioners.

For the sake of Europe's prosperity, its safety and its security, I'm sure we can do better.