4

Big data, data analytics, artificial intelligence and machine learning

Around half of customs authorities use some combination of big data analytics, AI and machine learning, while the other half plans to do so in the future. The majority of respondents see clear benefits from the technology, where risk management and profiling, fraud detection, and greater compliance are the most prevalent.

There is a need to establish a data strategy to ensure improved data governance and management and quality required. The obstacles and challenges to introducing these types of technology include the cost and the lack of expertise and good practices.

There is a high demand for human resources with advanced technical skills that are often hard to find, such as data architects and engineers, and software and machine learning designers. Data protection laws can limit the extent to which data can be used.

A lack of guidance on how to interpret such legislation in analysing data for customs purposes often leads to excessive caution in designing projects and restricts the exchange of data between organizations and customs authorities.

However, there can be great benefits to customs authorities, such as the following:

- data mining for intelligence purposes and risk management;
- post-clearance audits and controls;
- developing AI-based models for interpreting X-ray images;
- enhancing efficiency in using financial and tax data more widely and strategically to improve compliance and performance in terms of public revenue collection and to fight against smuggling;
- designing chatbots to answer online public enquiries;
- classifying tariffs;
- revealing potential anomalies in high revenue areas (e.g. excise goods);
- detecting undervaluation and overvaluation anomalies.

A lack of guidance on how to interpret data protection laws restricts the exchange of data between organizations and customs authorities.

Stage of adoption

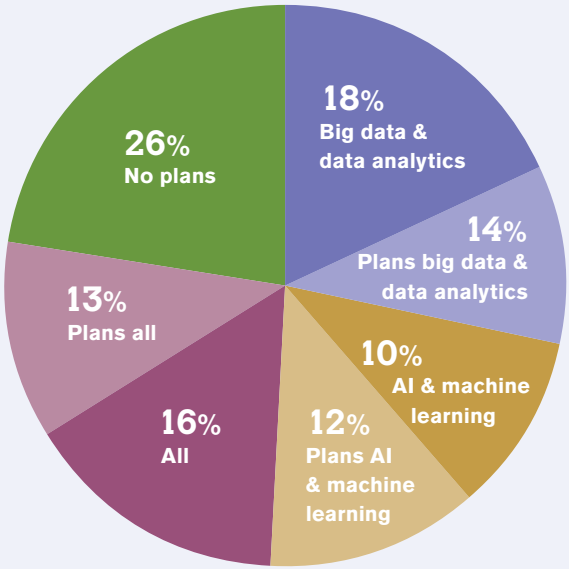
The ACS shows that 44 per cent of respondents use big data analytics, AI and machine learning (see Figure 12); while another 33 per cent have plans to introduce their use. A minority of respondents (23 per cent) have currently no plans.

Benefits

The main benefits reported are better risk management and profiling, fraud detection and compliance, customs audits and identification of anomalies (see Figure 13).

FIGURE 12

Stage of adoption of big data, data analytics, artificial intelligence and machine learning



Note: Total respondents numbered 94.

FIGURE 13

Main benefits of introducing big data, data analytics, artificial intelligence and machine learning

Note: Total respondents numbered 100. Figures indicate the number of responses for each category (respondents could give more than one answer).

Predicting future trends and improving trade facilitation and improving revenue collection also rank highly. However, fewer respondents reported improvements to container imaging and visual searches.

Some respondents reported a significant improvement in data quality when combining multiple commercial data sources with customs data. For example, Belgium is doing this to improve data quality and thus enhance the quality of data analytics and intelligence.

By integrating supervised and unsupervised learning models, it is possible to improve the performance of the assertiveness of the selection of transactions, either during processing or afterwards. Similarly, the incorporation of advanced selection models means that results can be interpreted more efficiently, which can then be fed back into the selection mechanisms, strengthening their performance. The models make it possible to reduce arbitrariness as a consequence of decisions based solely and exclusively on expert judgement.

The incorporation of advanced selection models means that results can be interpreted more efficiently.

Obstacles to adoption

The most significant obstacles to implementation reported were the lack of expertise and the costs (see Figure 14). Similar to perceived obstacles to the introduction of IoT, a lack of good practices, current legacy systems, a lack of use by other stakeholders (18 respondents), the absence of a government strategy as

FIGURE 14

Main obstacles to adopting big data, data analytics, artificial intelligence and machine learning

Note: Total respondents numbered 94. Figures indicate the number of responses for each category (respondents could give more than one answer).

well as legal issues are all obstacles. Respondents also cited data governance issues, data quality, information organization, roles and functions.

One respondent currently requests only a portion of the data it actually needs, while everything else is provided voluntarily by the carrier or service provider. Even for mandated data, such as advance passenger information collected by air carriers, obtain the necessary quality (i.e. timeliness, completeness) requires much effort. It can also be difficult to agree on a broad enough scope in data-sharing agreements required with third-party data providers, including within the government.

Properly using big data and machine learning in large organizations is not easy and often requires a shift in how to deal with data from a broader perspective, including having a data strategy and establishing a mechanism to ensure data governance and management. Such mechanisms can contribute to data of a higher quality and maximize the value of the information extracted (i.e. evidence-based decision-making, automation).

Respondents also observe that customs authorities may lose the golden opportunity to fully utilize the technology if they are not ready to manage the change. It has not been easy to train people with the appropriate skills. There is a high demand for engineers with advanced technical skills.

It can also be difficult to retain those profiles within teams due to a dynamic AI job market. Large-scale use of data and AI technology needs a new mindset in developing customs IT systems.

Integrating customs transactions data from multiple sources on a single platform and examining the big data from a 360-degree perspective require changes to current IT systems. In some cases, there will be legal obstacles to how freely the data can be used. Customs authorities face issues with legacy systems and may still rely on paper-based processing to a large extent.

Both customs authorities and data protection laws can limit the use of data for customs purposes. In EU law, there are the General Data Protection Regulation (GDPR) and the Union Customs Code. According to the definitions in the GDPR, nearly all customs-related datasets contain some form of personal data, whether directly or indirectly, that include names, addresses, phone numbers and company identification that could be traced back to a specific person or business. Those data fields are often unstructured or offer no easy way of excluding the personal data without losing the essence of the information.

In order to work with customs data, the choice is either anonymized data with minimal data protection constraints (but with less useful data) or personal data (in which case many provisions in EU and national legislation apply).

Most EU member States have pre-existing privacy laws they have had to adapt to comply with the GDPR, or even go beyond it. However, there is no clear guidance on how to interpret these laws together in the context

of analysing data for customs purposes. Such a lack of clear guidance and interpretation often leads to excessive caution when designing projects. It also strongly limits the cases for which an organization is able to exchange data with other customs authorities. In practice, the following often occurs:

(i) Data processing can only be carried out by certain categories of personnel (e.g. data miners, risk analysts, data managers).
(ii) Data can only be exchanged with other parties if an agreement (or a specific protocol) on the processing of the data has been signed. Negotiations can take a long time or sometimes even fail. The content must be modified and signed again every time a new objective is defined, or a new data source or category is added.
(iii) Internal procedures must be followed to make sure that everything is correctly documented, and that the data are protected (encrypted, and pseudonymized if it is not necessary to have the full data). Setting up the IT infrastructure dedicated to the exchange of data (i.e. servers) can only be done once the privacy process is approved by a dedicated data protection officer.
(iv) The data protection officer must validate all these procedures and documents before any further steps are taken.
(v) The explicit authorization of the data protection committee must be obtained in order to add external data containing personal data to the IT infrastructure.

Working with personal data is possible, but it is limited to specific use cases and requires much willingness, time and effort.

Ultimately, working with personal data is possible, but it is limited to specific use cases and requires much willingness, time and effort. The process is tedious and must be repeated every time a new element is incorporated. On the other hand, working without personal data is not much simpler, since excluding personal data also involves the processing of personal data.

One respondent raised the issue of how to manage expectations. There are limitations to what can be achieved with the legal and data quality issues of big data customs analytics. Making these limitations known to managers and policy-makers is key for moving forward.

Examples of use

Respondents provided many examples of how they use big data, data analytics, artificial intelligence and machine learning across a wide range of projects and cases. Table 2 summarizes the responses according to the type of technology.

TABLE 2

Example uses of big data, data analytics, artificial intelligence and machine learning

Post-clearance audits and controls	Refunds and drawbacks non-compliance
Automated targeting systems	Detection of prohibited goods
Tariff misclassification and non-compliance with tariff advice	Monitoring service delivery performance in real time
Misuse of concessions (including tariff concession orders, by-laws, free trade agreements and origin masking)	Providing historical insights into customs statistical information for future planning and forecasting
Detection of anomalies in high revenue areas (including excise equivalent goods i.e. alcohol, tobacco, petroleum)	Compliance risk scoring in commercial and trade activities
Detection of dumping and countervailing anomalies	Identifying low value courier and postal shipments to improve risk assessment
Detection of undervaluation and overvaluation anomalies	Identifying low risk individuals at borders
	Developing advanced analytics for AEOs

Examples specific to big data and data analytics

- Data mining for intelligence purposes and risk management
- Using financial and tax data to improve compliance and performance of public revenue collection and the fight against smuggling
- Analysing massive volumes of internet information to identify crime trends in intellectual property rights
- Data mining of imports
- Trade circumvention graph analytics
- Export and import data analytics
- Price recommendation data analytics
- Developing a data lake for consolidation and single truth of all customs data

Examples specific to artificial intelligence and machine learning

- AI-based models for interpreting X-ray images
- AI and machine learning for enhanced risk management processes (e.g. inspection results feedback loop, client segmentation, automatic assessment, upgrade of risk profiling)
- Chatbots for answering online public enquiries and for physical robots to answer enquiries from travellers at control points
- Cross-platform cyber patrols
- CT scanners with AI to enhance enforcement capabilities and customs clearance efficiency
- Automating repetitive manual procedures involving revenue protection (e.g. online research of market prices and deploying AI to assist in conducting intelligent filtering of research results)
- Using AI with Harmonized System Classification