WCO CAPACITY BUILDING FRAMEWORK ON DATA ANALYTICS
Executive summary

The amount of data being captured and stored by Customs today is overwhelming, especially with the influx of new technologies and the introduction of new Customs clearance systems which facilitate the recording of import, export and transit transactions between countries. By discovering associations and understanding patterns and trends within this data, data analytics has the potential to improve the way Customs deal with evolving threats, predict and manage non-compliance, and increase productivity and performance.

However, implementing data analytics in Customs is not an easy task; it requires considerable efforts to develop the appropriate organizational and technical capabilities needed to take advantage of the explosion in data, and to gain the insights that will help make better informed decisions.

This framework uses the results of a recent WCO survey on data analytics in Customs to provide information on how to effectively implement successful analytics initiatives and build the organizational capacities that are needed to optimize the use of data analytics. Specifically, it addresses the following questions:

- What types of assessments and planning should occur prior to undertaking an analytics project?
- What data and analytics processes should be in place?
- Why is data governance important, and how can a data governance framework be developed?
- What would be the most appropriate analytics organizational structure?
- What skill sets and types of staff are needed?
- Why and how is change management critical?
- What are the critical behaviours and actions that leaders should display?

A holistic approach to data analytics will help develop and promote the culture of analytics in Customs, facilitate policy development and achieve the following outcomes:

- Evidence-based policymaking and decision making integrated across the organization;
- Improved data management practices supporting the broader sharing of data;
- Increased transparency and trust in Customs decision making;
- Service delivery aligned with stakeholders’ needs that will facilitate improved interactions;
- An enhanced culture of innovation and collaboration in Customs;
- Reduced duplication of data and associated technical solutions, improved efficiencies, cost savings and resource utilization.

A maturity model approach is proposed in order to measure the success of the capacity building initiative in data analytics. The maturity model is meant to enable Customs agencies to evaluate their analytics capabilities, as well as their readiness, based on people, process, technology and governance.
Table of contents

Executive Summary 1

Table of contents 2

List of abbreviations 3

Introduction 4

Section 1: What types of assessments and planning should occur prior to undertaking an analytics project? 6

Section 2: What data and analytics processes should be in place? 11
   2.1. Data sources 12
   2.2. Data storage and management 13
   2.3. Analytics processes and technology 14

Section 3: Why is data governance important and how can a data governance framework be developed? 19

Section 4: What would be the most appropriate analytics organizational structure? 23

Section 5: What skill sets and types of staff are needed? 25
   5.1. Engagement and retention 26
   5.2. Competency development 27
   5.3. Outsourcing expertise 28
   5.4. Knowledge management 29

Section 6: Why and how is change management critical? 31

Section 7: What are the critical behaviours and actions that leaders should display? 33

Conclusion 35

List of References 36

List of Annexes 37
   Annex 1: Definitions 38
   Annex 2: Analytics Maturity Assessment 43
   Annex 3: Example of a Maturity Assessment tool 44
   Annex 4: Example of a data analytics strategic plan 47
   Annex 5: Example of a Roadmap 48
   Annex 6: Different types of analytics 49
   Annex 7: Performance Evaluation for Predictive Models 53
   Annex 8: Example of Data Governance Framework 55
   Annex 9: Data analytics job catalogue 61
   Annex 10: WCO Questionnaire on data analytics 63
List of abbreviations

AEO: Authorized Economic operator
AI: Artificial Intelligence
AMM: Analytics Maturity Model
BACUDA: Band of Customs Data Analysts
BI: Business Intelligence
CAO: Chief analytics officer
CMM: Capability Maturity Model
DGF: Data Governance Framework
DL: Deep Learning
DM: Data Model
EDI: Electronic data interchange
EDW: Enterprise Data Warehouse
GIS: Geographic Information System
HR: Human resources
IoT: Internet of Things
IMDB: In-memory database
ISO: International Standards Organization
IT: Information technology
KPI: Key Performance Indicator
ML: Machine Learning
MoU: Memorandum of understanding
NoSQL: Not Only SQL
OLAP: Online analytical processing
Q&A: Questions and answers
RFID: Radio frequency identification
SaaS: Software-as-a-Service
SQL: Structured Query Language
TSDS: Time Series Database Server
XML: Extensible Markup Language
HDFS: Hadoop Distributed File System
WCO: World Customs Organization
**Introduction**

In the growing digital economy and information society, data has become a significant business asset for both governments and business alike. The digital revolution is generating exponential opportunities, in terms of data access, capture, aggregation and analysis, to meet organizational goals more effectively.

Like other organizations, there is an increasing awareness among Customs agencies about the importance and value of data analytics. This concerns not only the benefits but also the potential use of data analytics, going from a purely descriptive and diagnostic tool to an intelligent solution that predicts events and prescribes the appropriate actions in a variety of situations.

Advanced data analytics, such as predictive analytics, can enable Customs to risk rank import and export transactions and create risk scores in real time, thus facilitating compliant traders while intercepting fraudulent shipments. Data analytics can equally help in identifying and prioritizing risk-based audits, as well as effectively managing AEO/trusted trader programmes. Data analytics has also proven to be effective in planning and allocating human resources and designing organizational structures and management models.

Evidence from a recent WCO survey shows that although most Customs agencies surveyed recognize the strategic importance of data analytics, they are still facing some challenges, including:

- a continuing struggle with many issues regarding data, applications, and technology;
- analytics capabilities mainly being deployed to perform descriptive analysis rather than predictive or prescriptive analysis;
- analytics efforts being fragmented. In most cases, the analytics capabilities are built for risk assessment;
- analytics outputs not being very well integrated into the broader organization;
- analytics and IT functions being mixed up;
- continued constraints in sourcing analytics talent, especially analysts with Customs-specific experience.

However, there are also examples of Customs, regardless of their level of development, having succeeded in making good use of analytics to enhance internal processes and improve resources allocation.

Over the past few years, the Secretariat has made numerous efforts to raise awareness among WCO Members about the potential benefits of data analytics, and to support them in their journey towards a successful implementation of the analytics function. Recently, the Secretariat launched a research project to familiarize Members with the use of data analytics in the Customs context through a collaborative initiative with external data scientists. The project is still in its exploration phase.

The WCO’s current efforts to boost Customs capacity in data analytics have the potential to help Members explore this area, but they need to be complemented by measures in other dimensions of institutional capacity building.

---

1 Data analytics is the science of drawing value from data.
2 The survey was conducted in January 2020 into the use of data and analytics in Customs. A copy of the survey is provided in Annex 7.
3 “BACUDA”.
According to several consulting agencies\(^4\), a successful data analytics transformation requires a focus on five elements. The first step involves asking a number of fundamental questions to develop the strategic vision. For instance, what will data and analytics be used for? How will the analytics products help achieve the objectives and increase performance? The second element is expanding the underlying data architecture and data collection capabilities. The third element is acquiring the analytics capabilities needed to extract meaning from data. The fourth component is changing internal processes to incorporate analytics outputs into the workflow. Finally, Customs need to build the capabilities of leaders and mid-level managers so that they can understand how to use data analytics products as the basis for making decisions.

Given the crucial role and the considerable potential of data analytics for Customs, the 82\(^{nd}\) Session of the WCO Policy Commission\(^5\) approved the development of a comprehensive Capacity Building Framework to guide Members towards the adoption and optimization of the use of data analytics and the effective implementation of successful analytics initiatives. The proposed outline of the content and structure of the Framework was approved by the 11\(^{th}\) Session of the Capacity Building Committee\(^6\).

This document introduces the framework for the development of data analytics capabilities within Customs. The proposed Capacity Building Framework for Data Analytics contributes to the WCO’s efforts to help Customs agencies address the main issues and obstacles to fully benefit from the potential of data analytics. It is intended to complement the high-level overview of data analytics provided in the WCO Handbook on Data Analysis by supplying further details on how to design, plan and develop the right technical and organizational capabilities to successfully implement data analytics in Customs.

The Framework is a living document and is intended to be updated based on feedback and recommendations from the Customs Community. Data lies at the core of this framework, and each element is focused on enhancing the value of the data through appropriate tools and technology to store and analyse data, and through strong governance, proficient human resources, clear data sharing processes, and strong leadership to facilitate change.

In the first section, we discuss the importance of assessing analytics capabilities, and developing an organization-wide analytics strategy and related action plan to harness the potential of data analytics. Then, we describe how to build a strong data infrastructure, and adopt the right analytical tools and technology to fully benefit from the opportunities offered to Customs by data. Thirdly, we discuss the importance of data governance and how to develop a data governance framework. Fourthly, we define the different types of organizational structure for data analytics, and explain the advantages and disadvantages of having a centralized, decentralized or hybrid analytics function. Fifthly, we explain how to attract, develop and retain analytics talent. We then explain why and how change management is critical to embedding data analytics throughout the organization. Finally, we discuss the importance of leadership to accelerate the adoption of data analytics and its buy-in across the organization.

---

\(^4\) Deloitte, Gartner, KPMG and McKinsey & Company.

\(^5\) Held in Seoul, Republic of Korea, from 3 to 5 December 2019.

\(^6\) Held in Brussels, Belgium, from 4 to 6 March 2020.
Section 1: What types of assessments and planning should occur prior to undertaking an analytics project?

To build the organizational capacity for analytics, strategic and tactical functions and processes must be established. This begins with a well-conceived analytics strategy and roadmap for success that is aligned with, and supportive of, the overall Customs strategy.

A data analytics strategy can be developed in line with strategic planning methodology. The planning process starts with the assessment of all elements of the analytics function, and should be based on the following standard phased approach, in keeping with the principles and recommendations that underpin the WCO capacity building strategy:

1. **Assessment**

Conduct data analytics assessment to:
- analyse key imperatives and define business needs;
- understand existing capabilities and carry out a holistic review of organizational requirements;
- identify the top analytics priorities, including improving existing practices.

2. **Design**

Create a long-term strategy for analytics and design the structure of the operating model through:
- formulation of a data analytics strategy;
- development of a business case;
- definition of a target operating model that provides a foundation for the data analytics development initiative.

3. **Roadmap**

Plan the programme and change process, including prioritization and dependencies between the different aspects, through:
- selection of prioritized and manageable strategic and tactical projects;
- definition of the different phases of the programme;
- development of a scalable Roadmap.

To begin with, the Customs agency must assess its business requirements and concomitant analytics capabilities. This must be done across the main facets, beginning with technology and ending with human resources competencies. The purpose of the assessment phase is to provide the organization with a means of identifying and determining analytics capability that is a fundamental to managing growth. The assessment phase evaluates the current analytics maturity, identifies strengths and weaknesses, and prioritizes initiatives to drive analytics transformation programmes.

---

7 An example of data analytics assessment models is provided in Annex 2.
In the proposed maturity model approach, the term “maturity” relates to the degree of formality and optimization of processes, from ad hoc practices, to formally defined steps, to managed result metrics, to active optimization of the processes. Although the Capability Maturity Model (CMM) comes from the field of software development, it is also used as a model to help organizations streamline business processes in general, and has been used extensively worldwide in government, commerce, and industry.

In data analytics, there is a well-known framework called the Analytics Maturity Model (AMM) to assess the analytics maturity of an organization. The AMM describes the stages that an organization travels through to reach process maturity. Each level is a measure of process maturity that the organization must solve to have more predictable outcomes. The model is like a ladder; each level represents a step. The organization must reach each level before progressing to the next one.

The different levels can be summarized as follows:

- **Level 1**
  The organization is dependent on spreadsheets and ad-hoc descriptive analysis. There is no analytical strategy at this point.

- **Level 2**
  The organization is not using the data strategically. If some analytics are present, the results are confined to few people and rarely shared across the organization. The management team does not understand the full value that the data can offer. It is not in a position to ask the critical questions that data and analytics can answer.

- **Level 3**
  The organization realizes the need to take the data to the next level. This level is where the management understands the need for proper data governance that allows the organization to exploit the full power of analytics.

---

Data analytics applications are available and relevant people have access to the data. This level reveals a growing interest among the staff in using analytics tools and delivering analytics products.

- **Level 4**
  This is the most challenging level for any organization to achieve. This is the level where the organization becomes data driven. Analytics is performed by skilled people who communicate and share their findings. The organization possesses the capability to identify analytics needs and perform advanced analytics, including work with artificial intelligence, in the different areas. The organization has embraced an analytics culture and continues to seek occasions to incorporate analytics and to encourage this mind-set throughout each department.

- **Level 5**
  The organization has fully implemented a data-driven approach to the decision-making process. The management has firmly defined and established the roles of the chief analytics officer (CAO). The organization offers a platform that allows analytics products to be shared between the different teams and units. The analytics team is fully developed. The model can be enhanced through self-learning over time, thereby perfecting its performance and predictive power.

The AMM provides concrete examples of requirements and activities that correspond to each level of data analytics maturity, but it does not provide any tools to help organizations assess their own level of maturity with regards to these requirements and activities. A maturity assessment tool based on the AMM can therefore be used to help Customs evaluate their analytics capabilities against the best practices in this area, identify key gaps, and better define and prioritize the goals, content and activities of their data analytics initiatives. \(^{10}\)

As such, the maturity assessment tool allows Customs managers to have a fair and realistic portrait of the maturity level of their organization and/or unit in terms of data analytics and, in turn, reduce the number of failures associated with this type of initiative. Also, the assessment tool, by being easy to use and being reusable, may facilitate the communication between data analysts and Customs experts, sensitize all Customs officers to the aspects and challenges associated to data analytics as well as serve as tool that can be used by external evaluators.

After assessing the current capabilities and resources, there is a need to develop a coherent vision, formulating a clear strategy and designing appropriate structures and the processes required to successfully implement data analytics.

The vision\(^ {11}\) should explain the value-added aspect of data and analytics. It guides the strategy and drives how all the other building blocks should be implemented in order to become a data-driven organization. The vision should be clear enough to avoid the conflict that typically arises from divergent expectations or misunderstandings.

When there is a clear vision, a strategy should be created to implement it. At this stage, the most important aspect to consider is what the organization wants to achieve, and how data can contribute to achieving its objectives. This means that the organization needs to define the key challenges and critical questions that demand answers, and then collect and analyse the data that will help address them.

Formulating a specific strategy for data analytics which contributes to the overall Customs strategy translates the need to develop a more detailed plan and nurture a distinctive competence in this

---

\(^{10}\) An example of data analytics maturity questionnaire is provided in Annex 3.

\(^{11}\) An example of a vision statement is provided in Annex 4.
functional area to ensure effectiveness and better performance\textsuperscript{12}. The data analytics strategy contributes to the achievement of the organization-level objectives, and provides relevant management and staff with a better understanding of their role in the organization’s mission and of how they can maximize benefits from the analytics function.

A successful data analytics strategy should be aligned with the organization strategy. Some activities that could be implemented by Customs to achieve this include a clear overview of high-level priorities, the translation of strategic needs into analytics projects, and the use of analytics in the strategy development process itself.

It is important to note that there is no one-size-fits-all data analytics strategy\textsuperscript{13}. Every strategy must align with the organization’s culture, priorities and resources. In addition, a data analytics strategy must be flexible, allowing for periodic reviews and updates to keep pace with changes in the working environment.

When formulating the analytics strategy, Customs need to consider the following important elements:

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{analytics_strategy.png}
\caption{Elements of the data analytics strategy}
\end{figure}

After creating the data analytics strategy, the next step is to develop a roadmap\textsuperscript{14} to translate the strategic goals into actionable activities and analytical requirements. The roadmap should be built to map the sequencing of the different phases of the strategic plan.

\begin{thebibliography}{9}
\bibitem{example} An example of an analytics strategy is provided in the annexes.
\bibitem{roadmap} An example of an analytics development roadmap is provided in Annex 4.
\end{thebibliography}
The roadmap contains short, medium and long-term initiatives and actions. Short-term actions are aimed at establishing the foundation upon which solutions will be built, whereas the medium-term initiatives strive to support the data analytics delivery functions. Finally, long-term actions are aimed at institutionalizing and optimizing the analytics function.

There are several important elements to consider when drafting the roadmap, including the identification of key milestones and dependencies, the determination of the sequencing of the different actions, timelines, responsibilities, and costs and risks. The roadmap allows for focused, measured progress, along with opportunities to improve and adapt actions.
Section 2: What data and analytics processes should be in place?

Instead of starting with the data itself, every Customs agency should start with defining what it wants to achieve from data, and how data can help it get there. To be truly useful, data must address a specific Customs need and help the organisation make smarter decisions that lead to higher performance and more efficient operations. This means that Customs agencies need to define the key challenges including operational weaknesses and determine the Customs-critical questions that need answering, and then collect and analyse the data that will help address them.

There are different ways to identify analytic opportunities. In the top-down approach, the strategic needs and priorities in terms of analytics are identified by senior managers and communicated to the relevant units. Another way is to create a centralized group composed of data analysts and Customs experts from the different Customs areas to explore the organizational data to find potential opportunities. This bottom-up approach relies on the group's ability to discover business questions in the data and offers the opportunity to bring innovative ideas.

The identification of analytics opportunities can also be done through open forums and digital crowdsourcing tools. Through these mechanisms, all employees can submit their creative ideas or suggestions on how to use analytics for different purposes including improving Customs operations and processes, increasing Customs revenue, mitigating risk, enhancing compliance etc. This method provides a steady stream of new ideas for organizations to consider and fosters staff engagement around data science. It works effectively in organizations with a pre-existing group or a network of data analysts (such as the WCO BACUDA project team) that can triage the ideas collected from an organization.

As analytic opportunities emerge, they need to be treated with the same project rigor as any other investment in the organization. Customs agencies must structure project management processes against analytics objectives, success criteria, realistic timelines, and proper resources. Thus, the standard of project management methodologies for data analytics is required for the implementation of analytics projects in Customs.

There are sets of practices for collaboration and communication between data scientists and Customs professionals that can be used to help implement analytics projects and manage analytics production lifecycle. Agile Analytics is an example of these practices that consists of some set of guiding principles and core values to improve team collaboration and efficiency and make developing analytics products seamless and efficient.

Having identified what they are looking to achieve with data, Customs agencies can then start to think about sourcing and collecting the best data to meet those needs. Making data accessible and ensuring it is of a high quality are key to fully benefitting from analytics. To this end, Customs should put in place a well-defined process governing how data will be extracted, cleansed, transformed, and

---

15 Crowdsourcing is a sourcing model in which individuals or organizations obtain goods and services, including ideas, voting, micro-tasks and finances, from a large, relatively open and often rapidly evolving group of participants. Currently, crowdsourcing typically involves using the internet to attract and divide work between participants to achieve a cumulative result (WIKIPEDIA).
16 DevOps/MLOps are examples of these practices: MLOps (machine learning operations) is a practice that aims to make developing and maintaining production machine learning seamless and efficient. While MLOps is relatively nascent, the Data Science Community generally agrees that it’s an umbrella term for best practices and guiding principles around machine learning – not a single technical solution.
17 The terms “data” and “big data” are used interchangeably in this document.
loaded into an integrated analytical environment. It is also important to ensure data integrity, security and privacy.

According to the results of the recent WCO survey on data analytics, there is a need for more commitment and efforts from Customs so that they can collect and store data and use the analytics tools, including open source applications, in an effective manner. In view of this, and bearing in mind what the increase in requests to access data and build models means in reality, and as new types of data and technologies emerge (big data\textsuperscript{18}, cloud data, data stream mining, etc.), it is crucial for Customs agencies to remain responsive and build their capacity to manage data, analytics and technology in an effective way.

The focus in this section is on how to build a strong data infrastructure and adopt the right analytical tools and technology to fully benefit from the opportunities offered by data.

2.1. Data sources

Data is divided into structured, semi-structured and unstructured types of data\textsuperscript{19}. Depending on the type of data, Customs can benefit from a variety of available sources. For example, information on a web site can be captured and analysed in order to collect data on fraudulent behaviour. Social media sources such as Facebook and Twitter generate tremendous amounts of comments and tweets. This data can be captured and analysed to identify, for example, new trends in drug trafficking or IPR infringements. Customs clearance systems generate data about commodities, importers, conveyances or routes that can be used for risk analysis or decision making. There is also a tremendous amount of geospatial data, such as that created by cell phones or RFID devices that can be used to help Customs identify the locations of smugglers or suspicious consignments, and improve the allocation of resources accordingly. Image, voice, and audio data can also be analysed for security purposes.

Most Customs agencies are not currently benefitting from the full potential offered by the multiple data sources that are available to them. Indeed, most analytics activities are focused on using structured data for traditional reporting and dashboarding. This could be improved by adopting a clear vision about the use of the different types of data and the appropriate tools and technologies to access, store and analyse semi-structured and unstructured data.

As unstructured data is growing more quickly than structured data\textsuperscript{20}, and with the emergence of new technologies to collect and process it, it becomes necessary to enhance reliance on all types of data, and more specifically unstructured data, to further increase the value of analytics products\textsuperscript{21}.

It is also very important to consider multiple strategies to source data, both internally to the Customs Agencies, inter-Agencies (Customs and non-Customs) and in the open market, both paid and non-paid. The attention is taken specifically to what is called these days as alternative data, where innovative processes are used to collect data that can be used as a proxy to discover information that is not

\textsuperscript{18}“Big data” is a term that is used to describe data that is high volume, high velocity, high variety, and high veracity; requires new technologies and techniques to capture, store, and analyse it; and is used to enhance decision making, provide insight and discovery, and support and optimize processes.

\textsuperscript{19} More details about the different types of data are provided in Annex 1.

\textsuperscript{20} According to Forbes, 90 percent of available data is defined as unstructured.

\textsuperscript{21} More details about some of the technologies available to manage the different types of data are provided in the annexes.
directly accessible. Alternative data sources include web scraped data, geolocation data from cell phones, satellite images etc.

Once identified, the potential data sources for Customs should be assessed on the basis of a number of criteria, including the credibility and validity of the data, the availability and consistency of the data over time and across sources, the timeliness of the data, its ability to support Customs specific analyses, etc.

2.2. Data storage and management

Capturing or accessing and organizing data is the first step to be performed by Customs in dealing with data. Currently, many Customs agencies use traditional Enterprise Data Warehouse (EDW) environments to store structured data. However, traditional EDW approaches do not take advantage of the full potential of data since they are not suitable for the incorporation of the different data sources.

The results of the WCO survey demonstrated that many Customs agencies are facing challenges in storing the massive amounts of unstructured data that they are generating or collecting from external sources. The reality is that managing unstructured data poses more challenges for Customs than managing structured data does. The main challenges include the huge storage capacity required, the difficulty in accessing or retrieving relevant data, and the cost.

To overcome these challenges, and satisfy the increasing need to store different data formats and access them for decision making, Customs can rely on Cloud computing if their own storage servers are not sufficient or appropriate.

Cloud computing is built around a series of hardware and software that can be remotely accessed through any web browser. Usually applications are shared and worked on by multiple users, and all data is remotely centralized instead of being stored on users' hard drives.

It should be noted that storing Customs data on a public network-based infrastructure raises two major concerns: security and privacy. Although the public cloud could be a cost-saving option, it also presents higher security risk and may lead to the loss of control of Customs data privacy since the access to data is not managed by the agency. Another option is to use private cloud storage, which provides a more secure environment and keeps the critical data inside the organization but increases the cost of big data analytics projects.

Data storage starts by determining what kind of data is most valuable for the organization to collect and analyse. The storage is carried out by performing three steps: data acquisition, transformation, and storage. These steps should follow specific principles that must be specified in the data governance framework that every Customs agency needs to have in order to ensure that data is stored

---

22 An example of an actor in this field is “Orbital Insight” which uses machine learning to transform multiple sources of geospatial data—including satellite images, mobile location, connected cars, and other IoT data—into data that can be used by Customs to monitor trans-border smuggling for example: https://orbitalinsight.com/

23 The traditional methods of storing and retrieving structured data include relational databases, data marts, and data warehouses.

24 More details on the different storage solutions are provided in Annex 1.

25 Scarpati, Jessica, Big data analysis in the cloud. Storage network and server challenges.

26 The primary goal of data acquisition is to read data provided from various communication channels, in different frequencies, sizes, and formats.

27 The transformation phase aims at moving, cleaning, splitting, translating, merging, sorting, and validating data before loading it into the database or data warehouse.
and processed in accordance with applicable regulations, and is protected from loss, cybercrime attacks or data leaks.

In addition to ensuring data availability, an effective data management process includes measures relating to data quality and data stewardship. Acknowledging the importance of data quality for the reliability of the analytics outputs, in 2015 the Customs Cooperation Council\(^ {28}\) adopted a recommendation establishing guiding principles for data quality.

Generally, data quality is demonstrated through the following attributes:

- completeness: all related data must be linked from all possible sources;
- accuracy: data must be clear, correct and consistent, with common data problems remediated, such as misspellings or abbreviations;
- accessibility: data must be accessible upon demand;
- timeliness: current data must be available.

One of the relevant activities that contributes to data quality improvement is data modelling. Creating a data model is a data management activity that helps to reduce redundancy and ensure that the information systems are really capturing and storing the right data that meet the needs of the organization.

In this regard, the WCO Data Model (DM) could be considered to be a strong tool for use by Customs to create a data dictionary\(^ {29}\) that details, field by field, what data sets they need to store in their databases. The WCO DM ensures data quality by carefully setting data requirements which are updated on a regular basis to meet the procedural and legal needs of cross-border regulatory agencies such as Customs, controlling export, import and transit transactions.

Ensuring data quality can also be achieved by complying with global standard ISO 8000. This international reference defines the characteristics of information and data that determine its quality, and provides methods to manage, measure and improve the quality of information and data.

Another data management aspect relates to data stewardship. This involves appointing a dedicated person or structure ("data steward") to make sure that data is organized, accurate and available for reporting, and that analysis and accountabilities for managing data are well defined. There should also be a plan for what data is needed from the people who know the data from both a technical and a Customs perspective.

While many Customs agencies have put in place well defined procedures for data integrity, security and privacy, they all recognize the need to continuously improve data management practices, including data stewardship. All of these aspects need to be included in the data governance framework that is explained in the following section.

### 2.3. Analytics processes and technologies

After data storage comes the analytics processing. This activity provides insights that allow decision makers to take action to achieve an intended outcome or result. Depending on the needs and priorities of the organization, the type of insight, and the level of maturity of the analytics function,

\(^{28}\) Customs Co-operation Council is the official name of the World Customs Organization (WCO).

\(^{29}\) The definition of data dictionary is provided in Annex 1.
there are four main types of analytics\(^{30}\) that could be performed by Customs, namely descriptive, exploratory, predictive and prescriptive analytics\(^{31}\).

Descriptive analytics, such as reporting/online analytical processing (OLAP), dashboards/scorecards, and data visualization, have been widely used for some time, and are the core applications of traditional business intelligence. Descriptive analytics are backward-looking and reveal what has occurred.

Predictive analytics suggest what will occur in the future. The methods and algorithms for predictive analytics, such as regression analysis, machine learning, and neural networks, have existed for some time. Exploratory analytics or discovery analytics are other names for predictive analytics. When these terms are used, they normally refer to finding relationships in big data that were not previously known.

Prescriptive analytics is the final tier in modern, computerized data processing, after descriptive and predictive analytics. Whereas predictive analytics reveals what will happen, prescriptive analytics suggests what should be done. For example, it can identify optimal solutions for the allocation of limited resources, as for instance when prescriptive analytics is used to help Customs find the best possible times to schedule physical inspection appointments for goods.

The results of the WCO survey demonstrated that most Customs administrations rely mainly on descriptive analytics. It is true that there is an increasing interest in using predictive analysis, and efforts to do so, but much still remains to be done. With the right choice of analytical techniques, Customs can process large data volumes, manipulate real-time, or near real-time, data, and capture all types of available data to improve organizational processes and strategies.

Some Customs agencies use data analytics to improve risk targeting by identifying, with statistical data, the groups that are the more likely to either proliferate or be subject to risks. In that sense, data is used to make correlations between incidents and the groups that were involved, and as a result it is easier to target risk according to probability analysis.

---

\(^{30}\) Please see Annex 6 for more information on the different types of Analytics.

\(^{31}\) Please see the WCO Handbook on Data Analysis for more information and examples on the use of the different types of analytics in Customs.

---

**Figure 3: Simplified data analytics workflow**
In general, data analytics is performed through the development of logarithms using data analytics tools such as R or Python to extract knowledge and insights from the different types of data. In addition, data analysts can make use of artificial intelligence (AI) tools like machine learning (ML) or deep learning (DL) to perform classification and prediction on the data.

In its simplest form, machine learning is the ability of machines to learn from past experiences using historical data to solve a given problem. It uses different algorithms to realize one or more (mathematical) models that act as the baseline for producing the desired result, while using input data parameters (dimensions). This is the fundamental difference from any computer program where, instead of programmatic implementation of an algorithm, a trained model is used to produce the output.

It is also possible for data analysts to combine the different types of analytics or incorporate ML in Robotic Process Automation (RPA)32. In fact, since any manual process can be automated as long as a clear operating procedure is available, Customs can automate repetitive, routine work using RPA or automate advanced intelligent processes through means such as AI. This is the case for example where AI-assisted X-ray image recognition is integrated with an automated alert and notification system to inform the Customs officer of the existence of illegal substances.

When it comes to choosing a data analytics tool, Customs may use commercial or open source solutions. Each of these options has its strengths and weaknesses. The main advantage of commercial tools is that they are more mature than open source tools as they have been tested, validated and documented. The main disadvantages revolve essentially around the cost and the low flexibility level. As a result, many organizations are increasingly exploring the open source alternatives to process data.

Open source tools like R or Python are available for free, which significantly lowers the entry barrier to use them. They provide total flexibility in the way in which to use the tools and data. In addition, because of their community-driven nature, open source solutions also tend to be at the forefront of state-of-art technologies and offer documentation and online Q&A forums. However, open source tools pose a disadvantage as well, since anyone can contribute to them without any quality assurance or prior testing. Another disadvantage concerns the significant work needed to develop and add any new complex capability.

A drawback of technologies like R or Python is that they require programming skills. As most of the Customs Agencies don’t have theses competencies, they can opt for the acquisition of data analytics technologies which are based on drag and drop interfaces, and that don’t require programming skills. These are usually referred to as self-service data preparation, analytics and visualization technologies.

Be it to reduce costs, or to develop the knowledge and intelligence of analysts, the adoption of open source tools could be a strategic choice for Customs to develop data analytics products and benefit from the available resources.

In addition to open source tools, Customs may use Excel to perform analytics. It is true that the analytics capacity of Excel has long been underestimated, but it continues to be extremely popular with users and business analysts alike, including for use with big data33. Excel can handle a million rows of data, source data from any database or business intelligence product, and has powerful analytical features. It is also possible to add additional analytics functionality from a third party.

---

32 Robotic process automation is a means of achieving Business Process Automation (BPA), which is the digital transformation, streamlining, and proactive management of organizational workflows.
However, many think that Excel is not recommended to support data science processes, even if third parties provide many add-on features. The reason is that Excel is a fundamentally different tool than a programming language as R or Python. Some of the problems with using Excel for data science are related to collaboration, security, integrity of data, reproducibility, decoupling interface from backend, versioning and traceability.

Building or improving analytics capabilities depends on the ability of Customs to make the best use of the available analytics technologies in general, and with regard to ML and DL in particular, especially with the recent advances in algorithms and processing power that have enabled the creation of machines with unprecedented analytics capabilities.

When the internal analytics capabilities are limited, Customs may use cloud technology to carry out data analysis or business intelligence processes. This is performed through a software system that is hosted on an Internet platform. Like cloud computing, cloud analytics is carried out on a private or public cloud. Some organizations use a hybrid model that keeps some functions within the organization while moving others to a cloud.

For analytics, cloud services are available as software-as-a-service (SaaS). With SaaS, the vendor provides the hardware, application software, operating system, and storage. The user uploads data and uses the application software to either develop an application (e.g. report) or simply process the data using the software (e.g. risk profiling).

When using cloud computing or cloud analytics, Customs managers must strike a balance between the cost-effectiveness of the different cloud choices and Customs data protection. This would require governments/Customs to develop a cloud computing policy and regulatory framework, especially with regard to server location and data security.

In order to help advance the data analytics agenda in Customs, the WCO is planning to secure data analytics platforms (cloud servers)\(^\text{34}\) in a phased approach, to provide Members with access to servers equipped with high computational capacity. The servers would be used initially for data analytics relating to national/regional collaborative research, and eventually for operations such as online price data monitoring, illicit trade data sharing, and mirror trade data sharing.

To select the most appropriate analytics tools and technology, Customs could develop and use a reference architecture\(^\text{35}\). The responses to the WCO survey on data analytics highlight the need to improve Customs’ understanding that designing a complete and standard reference architecture for all data-related projects is crucial to advancing the analytics agenda.

The selection of the analytics solution chiefly depends on the identification of the tactical and strategic needs to be fulfilled by analytics. These needs then have to be broken down into measurable analytics goals. Finally, the organization must select the tool that provides access to data and reporting features that can help achieve the analytics goals.

Once developed, the analytics product needs to be deployed. Nevertheless, before proceeding to final deployment of the model, it is important to evaluate its performance. The assessment helps to find the best analytics product that represents the data in terms of the accuracy of its predictions and

\(^{34}\) The initiative is subject to the availability of funding. Requesting Members would be invited to join/access the cloud servers and provided with the necessary training, including how to connect to different fields of the cloud servers.

\(^{35}\) The definition of data architecture is provided in Annex 1.
determines how well the chosen model will work in the future to address the objectives for which it has been developed.

Evaluating the model is essentially an ongoing process of re-examining the algorithms used, the data included, and the features selected for analysis as well as by constantly monitoring the accuracy of the model’s performance in a real Customs environment. The performance is calculated and compared by choosing the right metrics that truly measure how well each model achieve the overall goals of the organization36.

The presentation of the analytics outputs is the following essential step in the analytics process. The choice of the best reporting methods and tools that meet the requirements of the organization is particularly important. Similar to traditional business intelligence (BI), reporting is a critical data analytics feature that allows data to be visualized in a useful way, such as through visualization reports37, real-time information monitoring, alerts and operational key performance indicators (KPIs), to support daily operations and help make fact-based decisions.

Data visualization can take the form of charts, maps, and other graphical representations to enable users to better understand analytics products and take appropriate actions. In fact, graphical representations of data communicate patterns, trends, and outliers in a clearer way than tables of numbers and text. Of course, there are times when text and tables are the appropriate way to communicate findings.

Real-time data monitoring is the delivery of continuously updated information streaming. By tracking real-time monitoring data over time, Customs can reveal and predict trends and performance. The ability to run analytics on live data and provide immediate feedback to the system could be used in fraud deterrence, for example.

36 An overview of the different metrics for evaluating predictive models is provided in the Annex 7
37 The WCO Illicit Trade Report (ITR) is an example of a visualization report that provides analysis on trends and patterns in drugs, tobacco and cigarette, IPR enforcement, security, environmental protection and cultural heritage.
Section 3: Why is data governance important and how can a data governance framework be developed?

Data governance is defined as “the framework of decision rights and accountabilities to encourage desirable behaviour in the use of data”\(^{38}\). This concerns the availability, usability, integrity, security, privacy and deletion of the organization’s data, and includes the processes, roles, standards and metrics that ensure the effective and efficient use of data.

Within this definition there are three main elements:

1. rules for inputting and maintaining data;
2. enforcement of those rules, including the procedures and responsibilities;
3. management of the data by the users within Customs who are working on the data in accordance with those rules

On the governance side, the survey demonstrated that data and analytics governance is still an issue for Customs. This was foreseeable as governance frameworks appear in an organization when it reaches a certain maturity level in data analytics.

The survey also revealed the need for Customs agencies to develop and adopt standards for analytics development, as well as the deployment of technical processes. In the absence of such standards, the analytics outputs may not be well integrated into the decision-making process at all levels of the organization.

As the analytics function in Customs is rapidly growing, Customs officers in the different units will need to collaborate more, new processes will need to be developed, and managers will need to increase the reliance on analytics products. It therefore becomes necessary to develop a governance framework (DGF) to tackle all these issues.

The Data Governance Institute (DGI) has proposed a DGF (Fig 4), which is articulated around six key data governance areas: (1) Policy, Standards and Strategy (i.e., formal data policies supported by cross-functional data stewards), (2) Data Quality (i.e., data quality criteria and monitoring systems), (3) Privacy, Compliance and Security (i.e., privacy, compliance and security data programs mandated by management), (4) Architecture and Integration (i.e., data needs tied to architecture and integration challenges), (5) Data Warehouses and Business Intelligence (i.e., data warehouses and BI programs), and (6) Management Support (i.e., data programs focusing on getting managerial support).

Yet, to this day, one of the most comprehensive DGF is articulated around 11 data governance competences or categories\(^{39}\):

1. Data Risk Management and Compliance: A method by which risks are identified, qualified, quantified, avoided, accepted, mitigated, or transferred;
2. Value Creation: A process by which data assets are qualified and quantified to enable the organization to maximize the value created by data assets;


3. Organizational Structures and Awareness: The level of mutual responsibility between business and IT, and the recognition of responsibility to govern data at different levels of management;
4. Stewardship: A quality-control discipline designed to ensure the custodial care of data for asset enhancement, risk mitigation, and organizational control;
5. Policy: The written articulation of desired organizational behavior;
6. Data Quality Management: The methods used to measure, improve, and certify the quality and integrity of production, test, and archival data;
7. Information Lifecycle Management: A systematic policy-based approach for information collection, use, retention, and deletion;
8. Information Security and Privacy: The policies, practices, and controls used by an organization to mitigate risk and protect data assets;
9. Data Architecture: The architectural design of structured and unstructured data systems and applications that makes data available to appropriate users;
10. Classification and Metadata: The methods and tools used to create common semantic definitions for business and IT terms, data models, and repositories;
11. Audit Information Logging and Reporting: The organizational processes for monitoring and measuring data value and risks as well as the effectiveness of data governance.

![Image: Components of a data governance framework, by the Data Governance Institute (DGI)](image)

When formulating the governance framework, the first step is to define the missions of data governance, with clearly focused goals, execution procedures, governance metrics, and performance measures. The second step concerns the formulation of the guidelines, requirements and processes to ensure that the data is properly understood, trusted, accessible, and secure. Aspects related to data privacy must also be taken into account.

The issue of data privacy and security has been raised by the Customs community during many WCO meetings. This mainly concerns what data Customs should be allowed to collect, and what safeguards are needed on how it is used.

---

40 The definition of Metadata is provided in the Annex 1
Although there is a close link between data security and data privacy, there are distinct differences between the two aspects. Data security refers to how Customs agencies protect their data, including the technical safeguards that help ensure data confidentiality, integrity and availability, while data privacy revolves around the use and governance of personal data.

![Image: Data Privacy vs. Data Security]

*Figure 5: Data Privacy vs. Data Security*

When dealing with data privacy, Customs must define its own policy, which typically includes what data is to be collected, how data is collected and used, who will have access to it, how it can be shared, whether data can be legally collected or stored, and how long data can be stored for. The policy should refer to the national and international regulatory obligations the organization must comply with to collect, store and process personal data.

At the international level, ISO 27701 represents the standard in data privacy management. It relates to the way an organization collects personal data and prevents unauthorized use or disclosure. From the regulatory perspective, Europe’s General Data Privacy Regulation (GDPR) is the most wide-ranging, comprehensive data privacy legislation; it was enacted in May 2018 to preserve EU citizens’ personal data.

To protect the data, security measures need to be implemented. This starts with understanding the type of data and classifying it correctly. Security measures that could be implemented also include access control, data encryption, password management, user behaviour analysis, and threat detection and response.

Data assets often exist in the context of a specific purpose and are ruled by a specific legal basis. However, re-using data for multiple purposes other than the initial purpose could be envisaged. The second use might have implications in terms of data protection and confidentiality. This requires specific measures to ensure privacy and security for this type of data.

The international standard that describes best practice for data security is ISO 27001. It provides guidance for implementing information security controls to achieve consistent and reliable security programmes. The standard relates to the way an organization keeps data accurate, available and accessible only to authorized persons.
To ensure data security and privacy\textsuperscript{41}, Customs officers should be trained on data protection, so they understand the processes and procedures necessary to ensure the proper collection, sharing, and use of sensitive data.

The definition of the roles and responsibilities is another important step in the formulation of the DGF. The people taking on those roles should well understand and embrace those responsibilities. Besides, the roles and responsibilities should be adjusted to how the organization expects data governance to operate in the future.

Regarding the analytics process, the governance framework should provide clarification about who does what, when to perform analytics, why to perform it and which analytics technology is to be used. It needs to ensure that analytics and reports are accurate, available, timely, and of added value to the organization. In addition, the governance framework should provide clear definitions, standards and policies for all analytics reports, terminology, KPIs and calculations, reporting cycles and information dissemination, and it should be reviewed regularly to identify possible areas for enhancement.

The DGF must also include mechanisms for identifying, communicating, and resolving issues that are holding up analytics projects. In addition, the DGF requires a mechanism for providing sufficient resources for analytics projects and for balancing priorities between analytics projects and other Customs projects.

The governance framework\textsuperscript{42} could be developed by a committee set up by Customs specifically for this purpose. The committee must involve all relevant persons, including senior managers, IT experts, data analysts, and representatives from different Customs units, and may have different groups with specific responsibilities (executive, technical, etc.).

\textsuperscript{41} More details on data security and privacy measures are provided in annex 5.

\textsuperscript{42} A model template for data governance framework is provided in Annex 5.
Section 4: What would be the most appropriate analytics organizational structure?

The organizational structure defines the distinct roles and responsibilities that each group or individual assumes in relation to data analytics. There are three models of analytics structures: centralized, decentralized and hybrid.

Generally, the establishment of any analytics organizational structure depends on a set of goals. Different priorities for these goals may lead to different organizational models. These goals include:

- supporting decision makers with analytical capabilities;
- providing leadership for data analysts, and the ability to easily share ideas and collaborate on projects across functions;
- fostering visibility for analytics throughout the organization and ease in finding help with analytical problems and decisions;
- creating standardized methodological approaches, tools, and processes;
- researching and adopting new analytical practices;
- reducing the cost to deliver analytical outcomes;
- building and monitoring analytical capabilities and experience.

When the analytics function is centralized, the data analytics team provides analytical expertise and support work in unison. This team serves the different functions of the organization and works on diverse projects. The centralized unit is also responsible for setting out the analytical direction of the entire organization. In the decentralized analytics function, on the other hand, data analysts work within the different functions and units of the organization. In such a situation, each unit manages and creates its own data environment where analysts are responsible for the interpretation of data from within that team.

The main advantage of having a centralized data analytics team is that there is greater standardization of reporting across the different units of the organization and easier enforcement of data and analytics governance policies, while adopting a holistic and corporate approach to data analysis. Centralization makes it much easier to deploy analysts to projects with a strategic priority, which would not be the case for the other types of structures. Another driver for centralization is the cost of the analytics projects which can be reduced by exploiting synergies between the relevant stakeholders and eliminating redundant work.

However, centralization presents a number of disadvantages, including a low level of flexibility. It is easier and quicker for users in the decentralized structure to access the data since they do not have to go through a long internal process. This allows them to experiment more with data and to improve their general understanding and analytical capabilities.

The drivers for decentralization are usually vision, performance and speed. The vision of the analytics function cannot be realized unless there is a clear understanding of the objectives and priorities of the different units and functions of the organization. In the same way, the impact of analytics on performance is primarily a collaborative process requiring interdisciplinary action; this is usually jeopardized by the bureaucracy that characterizes a centralized organization. Finally, there is a

---

43 Deloitte: Organizing Analytics — from the inside out Establish an analytics ecosystem that drives value throughout the enterprise. Deloitte Development LLC (2014).
concern that a centralized analytics team would not be able to collaborate and achieve deliverables as quickly as they would like, due to the long approval process within the organization.

Decentralized data teams, by focusing on a single function of the organization, tend to have a deeper focus than central analytics teams, which may more easily rotate across projects and areas. The deeper focus has an advantage in enabling data analysts to easily adopt the expected behaviour and identify the potential anomalies to find in the data. The broader focus of the central team, however, could help analysts to leverage best practices and more easily adopt methods from other parts of the organization.

In addition to the above approaches, some organizations are in favour of adopting a hybrid organizational structure. This type of structure is characterized by the presence of interdisciplinary collaboration. The centralization aspect concerns the data architecture and governance, the management and maintenance of the technology needed for analytics, and the operationalization of analytics products such as dashboards and reports. However, the analytics function that leverages the data and tools is decentralized to maximize the impact on the performance of the organization.

In the specific case of Customs, the WCO survey demonstrated that for the majority of Customs agencies data analytics activities are conducted within the IT department. It also found that in many cases, data analytics is performed by isolated pockets of analysts within the risk management function.

This decentralized functional model is often used when the organization is relatively new to analytics or does not need analysts in every area of its operations. The advantage of this model is that analytics activities are concentrated where they are thought to be most needed. The disadvantage is that this model may limit the ability to expand analytical activities to other functions of the organization that could also benefit from it, leaving those functions without much analytics support and potentially impacting their performance.

Since the topic of data analytics is new for many Customs agencies, the first step before the creation of a dedicated structure could be the identification of the individuals with good analytics capabilities within the organization. Once identified, it is important to provide them with the opportunity to initiate and contribute to the analytics projects. This will help advance the analytics agenda and make rapid advances into analytics change and engagement initiatives.

---

Section 5: What skill sets and types of staff are needed?

Many Customs agencies recognize the need to have competencies in the area of descriptive analytics. However, having analytical talent in the area of descriptive analytics is not sufficient to extract a sense of the different types of the available data. Customs need analysts who have relevant data analytics competencies and are familiar with the complexity of the Customs work in order to generate both predictive and prescriptive insights.

The job of data analysts is to discover patterns and relationships from a big amount of data and turn these discoveries into actionable information. This role depends on the type of organization and the maturity of the data analytics function. In the Customs context, the responsibilities of a data analyst may include the following:

- designing and maintaining data systems and databases;
- collecting and organizing data from different sources;
- using statistical and analytics tools to interpret data sets;
- preparing analytics reports for data analytics users;
- contributing to the development of analytics strategy and policies for data governance;
- training and developing junior data analysts and interested staff.

To perform this role requires a variety of skills. Many organizations have attempted to develop a data analytics competency framework that sets out the skills and attributes required to successfully fulfil these types of roles, with the aim of supporting the organization in planning, recruiting, and developing analytics talent.

One of the best known generic data science competency frameworks has been developed by “DATA TO DECISIONS CRC”\(^\text{45}\) to describe the skills, knowledge, experience and personal attributes relevant to working in data science/analytics, including big data. This competency framework aims to support the development of the data analytics workforce, highlight career pathways and enable competency recognition.

The proposed structure of the competency framework is captured in the below diagram. It contains three key competency areas, each containing a number of relevant competencies:

- **data analytics solution life cycle**: competencies related to processing and managing data projects;
- **technical**: competencies relating specifically to big data, technologies and tools;
- **core**: identifies the data analytics-related aspects of competencies that often have organizational relevance, such as project management.

\(^{45}\text{DATA TO DECISIONS CRC: www.d2dcrc.com.au.}\)
In addition to data analysts, several types of cross-cutting experts need to be considered for the purposes of implementing data analytics, such as data engineers, data scientists, quantitative analysts, statisticians, econometricians and data-visualization specialists. Based on their policy considerations and human resources policy, Customs agencies need to determine whether in-house expertise needs to be developed or external experts can be engaged for the requisite support.

5.1. Engagement and retention

The survey indicates that the challenges that Customs agencies face in finding the right analysts include short supplies of analytics talent, as well as establishing a recruitment process. The survey also showed that while most of the Customs agencies surveyed are interested in hiring analytics talent, finding the most suited talent remains difficult.

The lack of personnel with the appropriate skills, as noted in the survey, constitutes a major constraint for Customs in realizing the full potential of data analytics. Given the high demand for analytics skills, it is challenging for Customs to find data analysts with the relevant expertise to add value to the organization in an effective manner.

As a first step, it is important for Customs to understand the priorities for the organization, as well as what increases its effectiveness and performance, before launching the recruitment process. Customs should then develop an appropriate and tailored recruitment and retention strategy, including a clear job description, to be able to attract data analysts with a strong analytical ability.

It is also paramount to institutionalize a career path for data professionals in order to retain them. In fact, clear roles and career opportunities are critical for engaging and retaining all types of employees, and data analysts are no exception.

---

46 Annex 6 provides more details on the data science job catalogue.
47 In 2017, IBM predicted that the demand for data analysts would increase 28% by 2020.
5.2. Competency development

As indicated above, data analytics involves the use of algorithms, visualization and modelling techniques. The skills needed include the ability to manipulate data, mathematical, statistical and programming skills, and the capacity to interpret the outputs and use them appropriately to support decisions. Given its importance, this skill set needs to be developed to remain relevant to the strategic orientation of the organization.

Reading the survey’s answers, it is clear that most Customs agencies do not have a formal data analytics development programme in place. Only a few Customs agencies are adopting, or have already adopted, a formal training curriculum for data analysts including on-the-job training.

To provide useful insights and decision-making support, data analysts must be capable of understanding the organization’s strategic goals and the Customs issues, and framing the appropriate analytical solutions. The necessary Customs knowledge for data analysts ranges from general familiarity with the areas of facilitation, enforcement, security, revenue collection, organizational development, and operations management, to the domain of knowledge required in specific data analytics applications, such as HS classification, valuation, risk management, post-clearance audit and the AEO programme.

The competency development initiative should be interdisciplinary and cover critical data analytics and IT skills, Customs and domain knowledge, and communication skills. A key to success for such initiatives is to make use of adult-learning principles\(^{48}\) when designing these programme, and combine methods such as on-the-job training, face-to-face learning, and online courses.

It is also important to establish strong relationships and partnerships between academic programmes and Customs to foster the experiential learning aspect of the data analytics curriculum. Moreover, Customs agencies can benefit from strategic partnerships with the WCO to develop their capacity in data analytics through the available data analytics training programme.

To build internal analytics capability, some Customs agencies try to identify officers with a quantitative background, such as statisticians and econometricians, then design a capability-building programme to extend their analytics skills. Other agencies provide working IT or Customs officers who wish to expand into data analytics with a part-time certificate programme. These certificate programmes can be delivered online or on-site, and need to provide the skills that will complement the current IT or Customs experience of IT professionals, and/or provide technical and analytical skills to Customs officers.

Another aspect concerns the capacity to make appropriate interpretations of the results of the analytics work. To avoid misinterpreting the reports generated, it is important that Customs provide basic analytical training courses in areas such as statistics, data mining and business intelligence to the decision makers and analytics users. The curriculum should also include the leadership skills required to manage the identification of the analytics needs and the implementation of analytics products, and the change-management skills required to effectively conduct culture change.

The following table provides an example of a capacity development programme for data analytics that could be implemented by Customs:

\(^{48}\) The core principles of adult learning are: self-direction, learning by doing, and relevance of training, based on experience, practicing skills, personal development, and involvement in planning and delivery.
<table>
<thead>
<tr>
<th>Beginners Level Track</th>
<th>Intermediate Level Track</th>
<th>Advanced Level Track</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Navigate the data dimensional space and understand distribution</td>
<td>- Develop predictive modelling</td>
<td>- Master main optimization schemes</td>
</tr>
<tr>
<td>- Develop basic analytics modelling techniques to uncover root causes</td>
<td>- Apply machine learning algorithms</td>
<td>- Understand text mining basics</td>
</tr>
<tr>
<td></td>
<td>- Develop social network analytics</td>
<td>- Conduct social sensing and sentiment analytics</td>
</tr>
<tr>
<td></td>
<td>- Learn big data basics and principles</td>
<td></td>
</tr>
<tr>
<td>Non-technical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Learn best practices for tackling analytics problem</td>
<td>- Manage an end-to-end analytics project</td>
<td>- Build collaborative and mature analytics programmes</td>
</tr>
<tr>
<td></td>
<td>- Build analytics competencies</td>
<td></td>
</tr>
<tr>
<td>Sample courses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical track</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Fundamentals of Data analytics</td>
<td>1. Statistical-Based Algorithms</td>
<td>1. Natural Language Processing</td>
</tr>
<tr>
<td>2. Data Quality Management</td>
<td>2. Learning-Based Algorithms</td>
<td>2. Simulation and Linear Programming</td>
</tr>
<tr>
<td>3. Essential Data Mining Tasks and Algorithms</td>
<td>3. Social Graph Theory</td>
<td>3. Fuzzy Logic and Rule Sets Optimization</td>
</tr>
<tr>
<td>4. Big data Technology</td>
<td>4. Business Intelligence and Data Visualization</td>
<td>4. Real-time Analytics for IOT data</td>
</tr>
<tr>
<td>Executive track</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Data and Analytics Governance</td>
<td>5. Standing up the Analytics Environment</td>
<td>5. Analytics Talent Management</td>
</tr>
</tbody>
</table>

| Data                   | Structured data | Structured and big data | Unstructured text and big data |

*Figure 7: COGNITRO Standard Data Analytics Training Programme*

All of these would require changes in, or adaptations to the HR policy, including recruitment, training and career advancements.

5.3. Outsourcing expertise

There is no doubt that carrying out data analytics projects with an in-house team offers greater control and less compliance risk. However, many Customs agencies may lack the resources to make the most of data on their own. In this case, Customs could benefit from the services of external providers who can use the data to perform analytics for them.

In practical terms, Customs may recruit a dedicated team of data analysts to carry out a specific project or to expand the internal teams with analysts provided by external service provider.

---

49 COGNITRO is an American company operating in the field of data analytics and AI.
Under the WCO research project “BACUDA”\(^{50}\), Customs data from some Members was used to develop algorithms in open source languages by the project team to detect Customs fraud through ML, DL and AI\(^{51}\).

When considering project outsourcing or partnerships, Customs must take into account the following important factors:

a) understanding core competencies and limitations:
   it is necessary to check whether there are analysts within the Customs agencies with the appropriate skills to perform the job before outsourcing to an external provider;

b) specifying the objectives and evaluation indicators in advance:
   if Customs decides to outsource a data analytics project, they should clearly define the purpose of outsourcing and the expected results. A clear protocol, agreement, or MoU, defining roles and responsibilities, should be signed with the identified partner;

c) finding the right partner:
   outsourcing partners must understand the mission and purpose of Customs and have the appropriate expertise and services;

d) having dedicated staff to manage outsourcing partners:
   outsourcing requires management and supervision by Customs;

e) paying particular attention to data security and privacy:
   outsourcing introduces new privacy and security concerns. It is therefore necessary to include data protection and security measures in any contract or agreement that is signed.

5.4. Knowledge management

Knowledge management is a structured and systematic process that aids the organization in retrieving organizational learning. Managing knowledge is important as analytics teams grow and work with an increasing amount of data and analytical models. It then becomes necessary to establish a network for sharing information and knowledge, and also standardizing the working and problem-solving methods. This is particularly important given that most ideas come from collaborating with others and expanding on existing ones.

The results of the survey demonstrated that most Customs agencies have no centralized guidelines to share details of previous analytical endeavours across the organization. Only a few agencies reported that they had accessible knowledge management guidelines across the organization.

Customs can follow the example of many organizations which have taken initiatives to set up a central platform or analytics repository so as to leverage previous analytical experiences sharing and foster continual collaboration. To achieve this, there are many efficient standard tools that can be used in managing knowledge across the organization, such as SharePoint sites, algorithm repositories, and visualization hubs.

It is also possible for Customs to increase data analytics knowledge sharing by arranging conferences and networking events across the organization to promote good practice, learning and innovation. Such events, and a centralized knowledge management capability, enable data analysts, Customs

\(^{50}\) “BACUDA” is an abbreviation of “Band of Customs Data Analysts” and means “to change” in Korean.

\(^{51}\) Preliminary results were presented at the 14th PICARD Conference (22 to 24 October 2019, North Macedonia)\(^{51}\) and the WCO AMS Regional Seminar on Data Analytics (23 to 24 October 2019, Brazil).
officers, other government agencies, stakeholders and partners to collaborate better, and to work more efficiently to advance the data analytics agenda inside the organization.
Section 6: Why and how is change management critical?

Changing culture and behaviours is of paramount importance for Customs to benefit from data analytics. The results of the survey demonstrated that most Customs agencies are interested in adopting a “culture of evidence,” in which decisions are based on facts generated by analytics to increase the efficiency and effectiveness of the organization. However, changing the organizational culture associated with how decisions are made is more than the implementation of a new toolset project; it is the execution of a major change management programme.

Knowing that change is typically not easy for people to assimilate, it is important for Customs to proactively identify the organizational need for analytics change, how the adoption of data analytics might affect each person’s responsibilities, and how the agency can engage assistance to successfully implement a data analytics initiative.

Many studies found that the top strategies in promoting a data-driven culture are strong leadership and commitment from senior management, promotion and sharing of data-sharing practices, increased availability of training in data analytics, and communication of the benefits of data-driven decision making. Other change management strategies focus on eliminating outdated methods, and linking incentives and compensation to desired behaviours.

Change management also concerns the translation of the changes into operational terms that employees can relate to by explaining how upcoming changes might impact structure, processes, skills and performance goals. This will help employees prepare for the new roles, capabilities, competencies and ways of working.

In addition, moving to a data culture depends on a collective buy-in from Customs officers at all levels. This does not mean that the enforcement officer, for instance, has to become a statistician, but that he/she needs to contribute, through best practices, training, and tracking, to making the change successful and making data part of the organizational culture in order to do his/her job more efficiently and effectively.

The survey attempted to identify where Customs had the biggest barriers in the adoption of analytics. The most common answer is “Embedding analytics into processes”. This situation could be improved if the decision-making process in the different Customs units and functions is reengineered so that decisions rely more and more on analytics. Another option could be to link strategies to analytics outputs, as well as to prioritize analytics requests that serve the organizational goals and vision.

To create a data analytics culture, Customs need to enhance data and analytics sharing through the development of policies that encourage people to work in a more collaborative way to effectively use data and share analytics outputs within the organization. Without this, data collection and delivery will be limited, with consequent adverse impacts on the effectiveness of data analytical capabilities.

Another important way of fostering a data driven culture is to encourage people to support each other and develop a sense of belonging. This could be achieved through open discussions, including through an annual analytics conference, to help everyone understand the importance and benefits of data analytics. Informal groups might also be used where individuals are encouraged to seek and uncover hidden opportunities or problems they can address.

Customs could also issue a manual to explain the benefits and suggest a range of recommendations to improve the use and application of data analytics. When this mind-set is shared by everyone, it creates a sense of community, where ideas lead to exploration and innovation.
Specific areas to be addressed also include the need to provide incentives for analytics-driven behaviour. For this, many types of incentives could be used by Customs, including recognition, appreciation and rewards for motivating and engaging staff to use data analytics.

Of equal importance is investment in training and development in analytics to create a data literate organization. Building individual knowledge and expertise makes people feel empowered to embrace the data analytics culture. Without a culture founded on empowerment, data analytics achievements will often be blocked or inefficient.
Section 7: What are the critical behaviours and actions that leaders should display?

Overall, the data analytics survey results underscore the need to have highly engaging leaders driving the data analytics agenda, and confirm the important role of the leader in overcoming staff resistance and improving the level of awareness across the organization.

Regardless of the leadership style adopted, leaders at all levels of the organization should use their influential capabilities to address and reinforce the attitude and behaviours that are needed to transform Customs into a data-driven organization. They also need to provide a role model by setting a meaningful example of how the rest of the staff should follow and embrace change.

At the practical level, identifying and empowering young talent could be an effective way to develop data analytics leaders who can use their knowledge and skills to inspire others, solve problems, and promote the analytics vision of the organization. This bottom-up leadership style improves knowledge sharing and collaboration and creates strong loyalty and an increased sense of fulfilment among staff.

When a Customs officer uses data analytics to solve a problem or to improve working procedures, this working behaviour can influence the officer’s manager and colleagues to grasp the importance of data analytics and adopt the new work habits. Realizing the significance of this approach, the WCO has launched the BACUDA project, which is targeted at younger Customs officers to make them experienced data analysts and effective change agents.

This project is based on a collaborative approach between the WCO research unit, Members and external data scientists, and aims at developing data analytics methodologies (algorithms) in open source languages (R and Python), so that all Members can employ them with their own data. This kind of knowledge-sharing culture facilitates the development of new data analytics competencies or sharpens existing ones and therefore creates analytics leaders who can serve as change enablers within their respective organizations.

In addition to the bottom-up process, it is important for Customs to use the top-down leadership style in changing the organizational culture. This requires showcasing leaders who visibly use analytics for decision making to spread the value of the approach across the organization. For example, senior and mid-level managers can use meetings to promote the use of data analytics and encourage others to do likewise.

Another way to promote data-driven decisions and analytics is to elevate responsibility for data analytics to a higher management level in the organization through the creation of senior level positions such as Chief Analytics Officers (CAO). Many studies demonstrated that, by including positions such as these, an organization clearly commits to foster the use of data analytics for decision making.

The roles of a CAO could include any or all of the following activities:

- mobilizing the required data, people, and systems to make analytics succeed within the organization;
- working closely with executives to adopt analytics in the decision-making process;

---

• supervising the activities and careers of analysts;
• consulting with the different functions and units within the organization on how to take advantage of analytics in their working processes;
• surveying and contracting with external providers of analytical capabilities.

Leaders can also support an analytics-driven culture by creating targets for strategic and operational improvements and promoting the impact of data analytics on the achievement of these targets. By educating staff on how to access and appropriately use analytics products to monitor and improve the organizational performance, leaders can contribute to the adoption of data analytics and its benefits in an effective and efficient manner.

Another critical role for leaders is to encourage innovation and inquiry in data analytics as an embedded part of the organizational culture. This means that the leader must constantly be open and receptive to new ideas and solutions from employees at every level of the organization. This can lead to higher buy-in and a greater diversity of expertise and ideas.
Conclusion

The evolution of decision support data management\(^3\) has changed the paradigm for the kinds of decisions that could be supported. With the emergence of big data, the ability to capture, store, and analyse high-volume, high-velocity, and high-variety data is allowing decisions to be supported in new ways. It is also creating new data management challenges.

Data analytics offers the promise of unlocking the potential of big data and opens up new avenues for enhancing decision making and supporting informed decisions in Customs. Success with analytics is not guaranteed, however, as there are specific requirements that must be met. A successful data analytics transformation requires work on several aspects and should be well planned and executed to address the current and future needs of the organization.

This document has set out guidelines for building the capacity of Customs in data analytics. More specifically, the proposed framework will help Customs agencies identify and implement the critical steps to move to a data-driven organization, including deciding what questions to answer and for what purpose, assessing the current and projected resources, developing the structures and processes required to enable decision-making and change, collecting and managing the relevant data, and selecting the right analytics tools and technologies.

The survey results show that, although the majority of Customs agencies are aware of the value of data, there are still many that have not invested enough to fully benefit from the potential of analytics. As a first step, these administrations can start the transformation journey by identifying officers who have demonstrated analytics capabilities and providing them with the opportunity to perform data analytics on the available data sets. Embarking such officers on analytics projects at the national level or in collaboration with other institutions, at the national or international levels, will help advance the analytics agenda and gain quick wins into analytics change and engagement initiatives.

In addition to the key elements explained in this document, it is very important that Customs be aware of the available technology required to perform data analytics. A great deal of information has already been published and is available on the different types of tools and technology to assist with advanced analytics, including open source tools. As the purpose of this framework is not to provide such types of information, or duplicate the existing information, it will be very helpful for Customs to explore existing resources to learn more about the existing applications.

In conclusion, Customs agencies are encouraged to use this Capacity building framework as a guide to assist in planning and executing the successful use of advanced analytics at every level of the organization and for different types of projects.

---

List of References

19. WCO: Handbook on Data Analysis (June 2018)
List of Annexes

Annex 1: Definitions
Annex 2: Analytics Maturity Assessment
Annex 3: Example of a Maturity Assessment tool
Annex 4: Example of a data analytics strategic plan
Annex 5: Example of a Roadmap
Annex 6: Different types of analytics
Annex 7: Performance Evaluation for Predictive Models
Annex 8: Example of data governance framework
Annex 9: Data analytics job catalogue
Annex 10: WCO Questionnaire on data analytics
## Annex 1: Definitions

<table>
<thead>
<tr>
<th>Item</th>
<th>Definition</th>
<th>Purpose</th>
</tr>
</thead>
</table>
| **Data dictionary**| The data dictionary is a log of all data elements in the data analysis environment. For each data element the log should include information such as: name, description, format(s), source(s), source of record, source(s) owner, name in source(s), active date, closed date. | The data dictionary enables the data stewards to understand:  
- data elements available for analysis  
- potential for bringing other data elements into analysis  
- integrity of the data elements  
- comprehensiveness of the data elements  

The data in the dictionary include all types of data.  
The data dictionary should be compiled and maintained as part of the governance process. |
| **Data types**     | *Unstructured data* consists of unexpected information, such as X-ray images, satellite imagery, videos, social media data, and all the text that comprises documents, log files, and emails, such as texts gathered in the Customs declaration. | All types of data are critical to the analytics process to help guide decision-making at all levels of the organization. |
|                    | *Semi-structured* data sits somewhere in the middle of structured and unstructured data. Markup languages such as Extensible Markup Language (XML) and electronic data interchange (EDI) are examples of semi-structured data.  

Structured data is very common because it may be either machine-generated messages produced without human intervention, or data created by a human through an interaction with a computer application such as a Customs clearance record. |                                                                                                                                                                                                 |
| **Metadata**       | There are two “metadata types”—structural metadata, about the design and specification of data structures or “data about the containers of data”—and descriptive metadata about individual instances of application data or the data content. | The main purpose of metadata is to facilitate the discovery of relevant information about data being exchanged with the different stakeholders. Metadata assists the user with information on how to |
Typical information contained in metadata includes:

- Clear and definitive business description and the intended meaning of the data fields contained;
- Specific rules or logic used for organizing or categorizing data elements;
- Language or local dialect details;
- Code set assumptions applied to specific fields.

One of the Standards for metadata is ISO/IEC 11179.

Understanding the exchanged data without laborious research. By describing the contents and context of data files, the usefulness of the original data/files is greatly increased, administrative support costs are reduced and system independence is improved. Metadata fosters improved data exchange between partners through semantic interoperability. Semantic interoperability is concerned not just with the packaging of data (syntax), but the simultaneous transmission of the meaning with the data (semantics). This is accomplished by adding descriptive information or can be stated as “data about the data” (metadata), linking each data element to a controlled, shared vocabulary.

The meaning of the data is transmitted with the data itself, in an self-describing “information package” that is independent of any information system. It is this shared vocabulary, and its associated links to an ontology, which provides the foundation and capability of machine interpretation, inferencing, and logic.

**Data storage methods**

Data storage is the recording (storing) of information (data) on a storage medium. Several storage solutions could be used for big data including distributed systems, Massive Parallel Processing (MPP) databases, and non-relational or in-memory databases such as Not Only SQL (NoSQL) or Hadoop. There are also Time Series Database Servers (TSDSs) which are specifically designed to handle time-series data.

Without the ability to store and retrieve data, it is not possible to perform analytics.
Spatial and GIS databases are suitable to store and process data that describes objects that exist in geometric space, such as graph databases.

<table>
<thead>
<tr>
<th>Structure the data</th>
<th>Organize the data in a computer so it can be used.</th>
<th>This step is essential to the analytics activity. The data structures provide a means to manage large amounts of data efficiently for uses such as indexing, designing algorithms, and application processing.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model and Algorithms</td>
<td>Process of selecting the right data sets, algorithms, and variables and the right techniques to format data for a particular business problem.</td>
<td>The identification of the correct model and algorithm necessary to perform the appropriate analyses produces the intended results for the problem identified.</td>
</tr>
<tr>
<td>Transform the data</td>
<td>Replacing variables so data can be used for an intended result</td>
<td>This function is usually applied so that the data more closely meets the defined assumptions, or to improve the interpretability or appearance of graphs.</td>
</tr>
<tr>
<td>Process data</td>
<td>Processing and analysing data is the collection and manipulation of data to produce meaningful results. There are three main models for processing big data: ▪ MapReduce ▪ stream computing ▪ in-database analytics.</td>
<td>This activity provides an insight into business and operations to allow decision makers to take action (or not, as it may be) to intervene in a given scenario to achieve an intended outcome or result.</td>
</tr>
<tr>
<td>MapReduce</td>
<td>MapReduce is one of the most commonly used models for processing large data sets. It allows the analysis of both unstructured and structured data in a massively parallel processing (MPP) environment. There are many ways of implementing MapReduce, and Hadoop is one of them.</td>
<td></td>
</tr>
<tr>
<td>Stream computing</td>
<td>Stream computing can support high performance stream data processing in near real time or</td>
<td></td>
</tr>
<tr>
<td>Data processing models</td>
<td>real time. With a real-time analysis, users can track data in motion, respond to unexpected events as they happen, and quickly determine next-best actions. For example, in Customs fraud detection, stream computing is an important analytical tool that assists in predicting the likelihood of illegal transactions. Transactions and accounts will be analysed in real time and alarms generated immediately to prevent frauds.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>In-database analytics</strong></td>
<td>In-database analytics refers to a data mining approach built on an analytic platform that allows data to be processed within the data warehouse. The results provided from in-database analytics are neither current nor real time, and it is therefore likely to generate reports with a static prediction. Typically, this analytic component in Customs is useful for supporting preventative fraud practice and improving compliance management and allocation of resources.</td>
<td></td>
</tr>
</tbody>
</table>

### Visualization

| Techniques used to communicate data or information by encoding it as visual objects or the visual representation of data | A primary goal of data visualization is to communicate information clearly and efficiently to users via graphics, plots, information graphics, tables, and charts. |

### Quality Reviews

| Data quality reviews are conducted to determine if the data is fit for use to serve the purposes of a given context. | The outcome of data analytics is highly dependent upon the quality of the data analysed. |

### Reference architecture

| A reference architecture is a document or set of documents that provides recommended structures and integrations of analytics products and services to form a solution, in order to | The reference architecture provides guidance on how to apply specific patterns and/or practices including technology to solve particular business |

---


| ensure the consistency and applicability of technology use within an organization. The reference architecture is recognized as a standard for all data-related projects; a steering committee monitors, evaluates and updates the reference architecture based on new applications or the requirements of the organization. | issues. It is used to select the right analytics tool to be used to process data. |
### Annex 2: Analytics Maturity Assessment

<table>
<thead>
<tr>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
<th>Stage 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytically Impaired</td>
<td>Localized Analytics</td>
<td>Analytical Aspirations</td>
<td>Analytical Adopters</td>
<td>Insight driven organization</td>
</tr>
</tbody>
</table>

#### Analytics Governance
- **Stage 1**: Lack of analytics
- **Stage 2**: Analytics focus is dispersed across multiple targets, which may not be strategically important
- **Stage 3**: Analytical efforts coalesce around a set of unified targets
- **Stage 4**: Analytical activity is centred around defined key domains
- **Stage 5**: Analytics support the ‘organizations’ strategy and further its mission

#### Data Governance
- **Stage 1**: Inconsistent data, poor data quality, or poorly organized data
- **Stage 2**: Useable data organized in functional or process silos
- **Stage 3**: Organization has nascent centralized data repository
- **Stage 4**: Integrated and accurate central data warehouse
- **Stage 5**: Relentless search for new data and metrics

#### Technology
- **Stage 1**: Lack of technology
- **Stage 2**: Data, technology and expertise in disparate clusters across organizations
- **Stage 3**: Early stages of an organization-wide approach
- **Stage 4**: Key data, technology and analysts are centralized or networked
- **Stage 5**: All key analytical resources are centrally managed

#### Culture and Leadership
- **Stage 1**: No awareness or interest
- **Stage 2**: Interest exists only at the function or process level
- **Stage 3**: Leaders recognize the importance of analytics
- **Stage 4**: Leadership provides support for analytics
- **Stage 5**: Leadership actively encourages and supports analytics and exploration

#### Analytics Competency
- **Stage 1**: Few or no analytical skills
- **Stage 2**: Isolated pockets of analysts with little communication
- **Stage 3**: Analysts clustered in key target areas
- **Stage 4**: Highly capable analysts in central or networked organization
- **Stage 5**: Professional analysts with strong training and support for junior analysts
Annex 3: Example of a Maturity Assessment Questionnaire (Adapted from Gartner Maturity Assessment Tool)

Rate your agreement with the following statements using the scale from 0 to 10 where 0 means "Do not agree at all" and 10 means "Completely agree."

1. My organization must benefit from data analytics to be performant

0 ☐
1 ☐
2 ☐
3 ☐
4 ☐
5 ☐
6 ☐
7 ☐
8 ☐
9 ☐
10 ☐

2. My organization is benefitting from data analytics to increase performance.

1 ☐
2 ☐
3 ☐
4 ☐
5 ☐
6 ☐
7 ☐
8 ☐
9 ☐
10 ☐

3. My organization is demonstrating commitment to increase meaningful use and application of data analytics.

0 ☐
1 ☐
2 ☐
3 ☐
4 ☐
5 ☐
6 ☐
7 ☐
8 ☐
9 ☐
10 ☐
Data

4. Which of the following best describes your organization’s data characteristics?

☐ Inconsistent, poor quality, and unstandardized data; difficult to do substantial analysis.
☐ Standardized and structured data, mostly in functional or process silos.
☐ Key data domains identified, and central data repositories created; Expansion into unstructured NoSQL data.
☐ Integrated, accurate, common data in central repositories; Data still mainly an IT matter; Little unique data: use of unstructured NoSQL data analysis.
☐ Structured and unstructured data are widely used and data is managed as strategic asset.

Organization

5. Which of the following best describes your organization’s approach to analytics?

☐ No organization perspective on data or analytics.
☐ Poorly integrated systems.
☐ Pockets of data, technology and expertise deliver local value.
☐ Process or Customs area or unit focus for analytics.
☐ Infrastructure for analytics beginning to coalesce.
☐ Key data, technology and analysts are managed from an organization perspective.
☐ Key analytical resources focused on strategic priorities.

Leadership

6. Which of the following best describes your leadership’s approach to analytics?

☐ Little awareness of or interest in analytics.
☐ Local leaders emerge but have little connection and influence.
☐ Senior managers recognizing importance of data analytics.
☐ Senior managers developing analytical plans and building analytical capabilities.
☐ Strong leaders behaving analytically and showing engagement for a data driven organization.

Targets

7. Which of the following best describes your organization’s approach to aligning analytics to specific Customs objectives?

☐ No targeting of opportunities.
☐ Multiple disconnected targets, typically not of strategic importance.
☐ Analytical efforts coalescing behind a small set of important targets.
☐ Analytics centered on a few key Customs domains with explicit and ambitious outcomes.
☐ Analytics integral to the organization’s strategy.
Analysts

8. Which of the following best describes your organization’s team of analytics professionals (analysts, data scientists, etc.)?

☐ Few skills, and those are attached to specific functions.
☐ Disconnected pockets of analysts; unmanaged mix of skills.
☐ Analysts recognized as key talent and focused on important Customs areas.
☐ Highly capable analysts explicitly recruited, deployed and engaged.
☐ World-class professional data Scientists.

Technology

9. Which of the following best describes your organization’s analytics technology?

☐ Desktop technology, standard office packages, poorly integrated systems.
☐ Individual analytical initiatives, statistical packages, descriptive analytics, database querying, tabulations.
☐ Organization analytical plan, tool and platforms; predictive analytical packages.
☐ Organization analytical plan and process, cloud-based big data.
☐ Sophisticated organization-wide big data and analytics architecture, cognitive technologies, prescriptive and autonomous analytics techniques.

10. Which of the following best describes the analytical methodologies that your Organization can deploy?

☐ Mostly ad hoc, simple math, extrapolation, trending.
☐ Basic statistics, segmentation, database querying, tabulations of key metrics are leveraged to gain insights.
☐ Simple predictive analytics, classifications and clustering; dynamic forecasts.
☐ Advanced predictive methods deployed to discover insights, advanced optimization, text and image analytics.
☐ Neural networks and deep learning, genetic algorithms, advanced machine learning.
Annex 4: Example of a data analytics strategic plan

**Vision:**
To be a leading Customs agency in the use of data analytics so as to drive efficiency, collaboration, innovation and effective service delivery.

**Mission:**
To support a better understanding of the challenges and opportunities for Customs by using data as a basis for decision making.

<table>
<thead>
<tr>
<th>Strategic pillar</th>
<th>Strategic initiatives</th>
</tr>
</thead>
</table>
| **Governance**   | • Implement the accountability and governance framework, with shared roles and responsibilities, to guide the analytics transformation.  
                   • Develop data management guidelines and coordinate the requirements of the shared data ecosystem.  
                   • Establish a Centre of Excellence where data management professionals can share knowledge and best practices, and develop standards for interoperability and data sharing.  
                   • Establish and implement an Analytics Code of Ethics that guides the responsible use of data. |
| **Analytics Talent** | • Establish an Analytics Centre of Excellence to foster a culture of data sharing, analytics and innovation.  
                           • Develop training and skills development in collaboration with the human resources department to develop, maintain and attract data science skills.  
                           • Establish a Customs “incubator lab” to facilitate a holistic analysis of data that allows greater insight on policy and programme issues.  
                           • Identify opportunities for data and knowledge sharing through a Customs Data Analysts Network. |
| **Process**       | • Identify the best practices that are currently in place within the organization and adapt them to meet the requirements of data analytics.  
                           • Establish common Customs data management, exchange and interoperability standards that enable the availability of reliable, timely and relevant data.  
                           • Establish mechanisms for change management and business process improvements.  
                           • Develop a consistent management process to connect all data domains through a comprehensive Customs data inventory. |
| **Technology**    | • Establish an appropriate environment with technology and tools to support the Customs incubator.  
                           • Explore and evaluate opportunities to maximize the utility of current infrastructure deployed across the organization.  
                           • Design and build a scalable technical infrastructure to facilitate data transformation capabilities. |
### Annex 5: Example of a Roadmap

<table>
<thead>
<tr>
<th>Strategic considerations</th>
<th>Strategic Pillars</th>
<th>Key initiatives and milestones</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Fragmented sources of data</td>
<td>Governance</td>
<td>-Define governance structure</td>
<td>-Enhanced culture of innovation and collaboration in Customs through data and analytics sharing</td>
</tr>
<tr>
<td>-Inconsistent data management practices</td>
<td></td>
<td>-Develop governance framework</td>
<td>-Improved data management practices in Customs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Develop analytics code of ethics</td>
<td>-Improved efficiencies and resources utilization</td>
</tr>
<tr>
<td>-Demand for more efficiencies and better resource utilization</td>
<td>Human Capital</td>
<td>-Implement governance policy</td>
<td>-Analytics infrastructure is implemented to collect, store and analyse data</td>
</tr>
<tr>
<td>-A need for a modern and efficient analytics infrastructure</td>
<td>Skills, training</td>
<td>-Establish data governance bodies</td>
<td>-Improved service delivery and interactions</td>
</tr>
<tr>
<td>-Demand for improved Customs service delivery</td>
<td>Process</td>
<td>-Develop data management guidelines</td>
<td>-Increased transparency and trust in Customs decision-making</td>
</tr>
<tr>
<td>-Demand for increased transparency and accountability</td>
<td></td>
<td>-Develop assessment plan</td>
<td>-Evidence-based policy and decision-making</td>
</tr>
<tr>
<td>-Need for improved information for decision-making</td>
<td>Technology</td>
<td>-Perform maturity model assessment</td>
<td></td>
</tr>
</tbody>
</table>
Annex 6: Different types of analytics

Data analysts have a vast array of analytical capabilities and techniques at their disposal. These range from the most fundamental techniques, “descriptive analytics”, answering the questions “What has happened?”, “Why did it happen?”, but also “What is happening now?”; (ii) predictive analytics, answering the questions “What will happen?” and “Why will it happen?”; (iii) prescriptive analytics, answering the questions “What should I do?” and “Why should I do it?”. Currently, the vast majority of analytics efforts are spent on descriptive and predictive analytics with typical methodologies including data mining, machine learning, artificial intelligence and simulation.

Descriptive analytics

Descriptive analytics is a commonly used form of data analysis whereby historical data is collected, organized and then presented in a way that is easily understood. Descriptive analytics is focused only on what has already happened, unlike other methods of analysis, it is not used to draw inferences or predictions from its findings.

This most simplistic form of data analytics uses simple mathematical and statistical tools, such as arithmetic, averages and per cent changes, rather than the complex calculations necessary for predictive and prescriptive analytics.

Descriptive analytics uses two key methods, data aggregation and data mining to discover historical data. Data aggregation is the process of collecting and organizing data to create manageable data sets. These data sets are then used in the data mining phase where patterns, trends and meaning are identified and then presented in an understandable way.

Descriptive analytics is frequently used in the day-to-day operations to help the organization measure performance to ensure goals and targets are being met and identify areas that require improvement or change. Reports such as those on workflow, revenue, seizures, general trends, survey results – are

---

all examples of descriptive analytics that can be used to detect patterns that indicate a potential problem or a future opportunity for the organization but can’t be used for making inferences or predictions.

Predictive analytics

Predictive analytics, as its name implies, is focused on predicting and understanding what could happen in the future. Analysing past data patterns and trends by looking at historical data can predict what might happen, and thus allows for setting realistic goals, effective planning, managing performance expectations and mitigating risks.

Since predictive analysis is based on probabilities, it can never be completely accurate, but it can act as a vital tool to forecast possible future events. Customs can use predictive analytics for anything from forecasting risk behaviour to identifying revenue trends. Predictive analytics can also improve efficiency and fraud detection and prevention.

Predictive analytics is based on probabilities and uses a variety of techniques – such as data mining, statistical modelling (mathematical relationships between variables to predict outcomes) and machine learning algorithms (classification, regression and clustering techniques) to predict future or other unknown events. To make predictions, machine learning algorithms, for example, take existing data and attempt to fill in the missing data with the best possible guesses.

![Classification of the methods for predictive analytics](International Journal of Information Management)
Prescriptive analytics

Prescriptive analytics is the most sophisticated type of data analytics. It aims at suggesting the best decision options in order to take advantage of the predicted future utilizing large amounts of data. To do this, it incorporates the predictive analytics output and utilizes artificial intelligence, optimization algorithms and expert systems in a probabilistic context in order to provide adaptive, automated, constrained, time-dependent and optimal decisions (Gartner 2017).

The prescriptive analytics assists the users in providing optimal solution to a problem and chooses best decision among different alternatives. It can allow them to draw insights without the dependence on Data Scientists or Operations Research experts within IT Departments.

The prescriptive analytics can be used for different purposes. It can improve Customs operations by allocating resources to areas of anticipated smuggling in advance. It can also be used to optimize the human resources allocation based on workload or to determine the HS code of imported goods. Prescriptive analytics helps identify and better quantify the risk associated with both short and long-term decision-making and develop potential risk mitigation strategies.

Prescriptive analytics can also be performed by combining different categories of methods which are presented in Fig 3 below.
Fig. 3. Classification of the methods for prescriptive analytics (International Journal of Information Management)
Annex 7: Performance evaluation for Predictive Models

The evaluation of analytics models is important to assess their predictive ability. The performance assessment depends on the choice of metrics.

All problems a predictive model can solve fall into one of two categories: a classification problem or a regression problem. A classification problem is about predicting what category something falls into. An example of a classification problem is analysing seizure reports to determine if an importer is in a high risk group for commercial fraud or not.

Metrics that can be used for evaluating a classification model:

- **Percent correction classification (PCC):** measures overall accuracy taking into consideration that every error has the same weight.
- **Confusion matrix:** This is an NXN matrix where N is called the number of classes being predicted. It measures accuracy but distinguishes between errors, i.e. false positives, false negatives and correct predictions. The model gives a probability from 1.0 to 0.0.
- **ROC (Receiver Operating Characteristic) & AUC (Area Under the Curve):** Receiver operating characteristic (ROC) curve is one of the most effective evaluation metrics because it visualizes the accuracy of predictions for a whole range of cut-off values. In order to get ROC, we just need to derive two ratios from the confusion matrix: True Positive Rate (TPR) or Sensitivity, and True Negative Rate (TNR), or called Specificity. The most important parameter that can be obtained from a ROC curve is the Area Under the Curve (AUC).
- **Lift and Gain charts:** is a measure of the effectiveness of a classification model by calculating the ratio between the results obtained with and without the predictive model. In other words, these metrics examine if using predictive models has any positive effects or not.

A regression problem is about predicting a quantity. A simple example of a regression problem is prediction of the Customs revenue. To evaluate how good the regression model is, we can use the following metrics:

- **R-squared:** indicates how many variables compared to the total variables the model predicted. It summarizes the explanatory power of the regression model and is computed from the sums-of-squares terms. R-squared does not take into consideration any biases that might be present in the data.
- **Average error:** the numerical difference between the predicted value and the actual value.
- **Mean Square Error (MSE):** If the data contains a huge number of outliers, then this metric is known to be a good one. This is a measure of how close a fitted line is to data points. It assumes that the error is unbiased and follows a normal distribution. The smaller the Mean Squared Error, the closer the fit is to the data.
- **Root Mean Squared Error (RMSE).** It is just the square root of the MSE. It is the distance, on average, of a data point from the fitted line, measured along a vertical line.
- **Average absolute error:** similar to the average error but uses the absolute value of the difference to balance out the outliers in the data.
- **Median absolute error**: represents the average of the absolute differences between prediction and actual observation. All individual differences have equal weight, and big outliers can therefore affect the final evaluation of the model.

- **Kolmogorov Smirnov chart**: The K-S chart measures the degree of separation between the positive and negative distributions of a model. In a mathematical expression, \( K-S = |\text{Cumulative } \% \text{ positive} - \text{Cumulative } \% \text{ negative}| \). In most classification models, the K-S gives values between 0 and 100, where the higher value is considered as the better model.

- **Concordant- Discordant Ratio**: The method of calculating this ratio compares the classifications for two variables on the same two items. The total number of concordant pairs are counted and divided by the total number of pairs. This will give us the value of concordance ratio. The higher the concordance ratio, the better is the model.

- **Cross-Validation**: This metric is used to compare and select a model for a given predictive modelling problem. The goal of cross-validation is to test the model's ability to predict new data that was not used in estimating and to give an insight on how the model will generalize to an independent dataset.
Annex 8: Example of data governance framework

Data Governance Framework at XYZ Customs agency

Scope
This document sets out the policy and framework to be followed to manage institutional data and applies to data in all its forms.

Status and effective date
Approved, June 2020
Validity: under annual review.

Audience
This policy applies to everyone in the organization and sets out responsibilities for stewards, owners, and users of institutional data. This policy informs decision makers for key Customs resources; such as individuals, process, technology, and data.

Purpose and goals
This data governance framework proposes a set of principles, structures, roles, and responsibilities for adoption in the organization to improve the data infrastructure and to advance institutional goals for operational excellence. The benefits of applying this policy will include ensuring data is fit for the purposes of internal and external reporting, and is appropriately categorized for storage, retrieval, destruction, backup, and access as needed to ensure the proper management and protection of institutional data; with the goals of:

- enabling better decision-making;
- reducing operational friction; protecting the needs of data stakeholders;
- training management and staff to adopt common approaches to data issues;
- building standard, repeatable processes;
- reducing costs and increasing effectiveness through the coordination of efforts;
- ensuring the efficiency and transparency of processes;
- ensuring the ongoing availability and disaster recovery of key data;
- supporting IT roadmap planning and technical decisions;
- reducing the risk of mismanagement of data.

Values
- Data Sharing
Data and information are shared organizational resources that constitute valuable assets.
- Stewardship
To ensure that data is collected, stored, and maintained, and that relevant persons will access and use it over time.
- Quality
To ensure data retain its value, and that the quality of data is actively monitored and maintained.
- Privacy and confidentiality
The maintenance of the privacy and confidentiality of all records is not only a legal requirement but also a primary outcome of data management.
- Responsible access:
Accessing and using institutional data only as required to conduct their Customs business. Reporting any breaches of Customs information in a timely manner according to the formal procedures.

Data governance principles
- Organizational effectiveness
Data governance activities improve the organizational effectiveness and efficiency of operational processes.

- **Transparency**
  Data governance policies, activities and products exhibit transparency through documentation available to the data users within the organization.

- **Communication**
  Data governance promotes and ensures communication so that the data produced is fully understood and can be reproduced with the same results.

- **Compliance**
  Data governance adheres to and enables institutional compliance with applicable statutes, regulations, and policies; including but not limited to areas of security, privacy, and record retention.

- **Auditability**
  Data governance promotes a means to document and verify data, track changes and provide justifications for changes.

- **Integrity**
  Data governance participants practice integrity when dealing with data; they are truthful and forthcoming when discussing drivers, constraints and options for, and impacts of, data-related decisions.

- **Accountability**
  Data governance defines responsibilities for cross-functional data-related decisions, processes, and controls.

- **Standards**
  Data governance identifies and supports consistent standards for data elements, dictionaries, quality, and usages.

### Governance roles and responsibilities

#### I. Data governance bodies

##### a. Data steering committee (DSC)

- **Role:**
  Recommend and implement institutional policy for data governance in Customs, including for how data is defined, produced, used, stored, and destroyed.

- **Responsibilities**
  1. Develop and follow procedures for internal DSC operation, meetings, workflow, and decision-making procedure.
  2. Develop, recommend and evaluate the effectiveness of policies, procedures, and processes for data management, data quality, and data use.
  3. Set priorities for preserving and increasing the value of data assets.
  4. Oversee data quality monitoring and improvement.
  5. Specify and prioritize the requirements and strategy for data management.
  6. Appoint data stewards and ensure that all relevant data entities have a responsible steward.
  7. Review and validate a record of the organization data and solution owners and stakeholders.
  8. Set standards for data dictionaries and definitions, and reporting conventions.
  9. Ensure compliance and coordination with the privacy and security policy.

- **Membership**
  - Head of IT department.
  - Head of analytics function.
  - Heads of the different Customs departments.
  - Head of the Data governance committee (DGC).

- **Structure**
  1. Chair.
b. Data governance committee (DGC)

i. Role:
The DGC is a specialized group with a cross-functional membership that oversees the management of data assets in all functional areas, including data quality monitoring and improvement, and interfacing between data stewards and the Data steering committee (DSC).

ii. Responsibilities
1. Coordinate data stewards.
2. Respond to inquiries about process, content, limitations and uses of data, especially in cross-functional settings.
3. Consider and approve changes to code sets, additions to tables.
4. Elevate to the DSC issues that require resolution beyond the mandate of the DGC.
5. Review data quality; identify practices promoting data quality; identify areas for improvement and monitor improvements.
6. Ensure dictionary standards are followed.
7. Evaluate the effectiveness of data governance policies.
8. Recommend to the DSC changes in policies, procedures, and processes to improve effectiveness.
9. Communicate proceedings, including notice of changes and decisions, to the DSC.

iii. Membership
1. All data stewards from the different Customs areas.
2. Representative from the IT department.
3. Data and analytics representative.

iv. Structure
1. Chair.
2. Vice-Chair.

II. Roles for individuals

a) Data stewards

i. Role:
Data stewards shall be responsible for the management of data under their care, under the direction of the DSC. Data stewards have significant technical expertise in data under their responsibility.

ii. Responsibilities:
1. Maintain inventory of data assets (list of tables, fields, dictionary information).
2. Coordinate data owners and implementation of
   - data quality assessment;
   - data quality maintenance;
   - metadata maintenance.
3. Participate actively in DGC meetings.
4. Maintain data dictionary in consultation with data owners, ensuring each element:
   - has a clear and unambiguous definition;
   - has clear value definitions assigned to all values;
   - is still being used;
   - has adequate documentation for origin and sources of authority;
   - has adequate documentation on appropriate usage and notes.
5. Communicate data governance policies, procedures and practices to data owners.
6. Communicate changes, adjustments or other local needs to DGC.
7. May have a role in security policy to confer and restrict access to data.
   iii. Membership
       a. Data stewards are identified by the functional leader of an operational unit that manages data.
       b. Data stewards have data governance responsibilities as formal components of their performance plan.

   b) Data owners

      i. Role:
      Directly enters or load data into systems; correct or update values on an ongoing basis.
      ii. Responsibilities:
          a) Maintain data to enable effective business practices.
          b) Rectify data quality issues of data under consideration.
          c) Communicate rectification plans and outcomes to data steward.
      iii. Membership:
          • Any individual who has write-access to enter or maintain data is a data owner.

C. Data users

   i. Role:
   Data users are everyone who has access to institutional data in performing their assigned duties. This access includes reading, entering, downloading, copying, querying, analysing, or updating data or information.
   ii. Responsibilities:
       a) Recognize that institutional data and the information derived from it are potentially complex. Make efforts to understand the source, meaning and proper use of the data through training sessions, utilizing data dictionaries and knowledge of supporting system processes.
       b) Include information about the data source and criteria when distributing data, reports and ad hoc analytics.
       c) Respect the privacy of individuals whose records they may access.
       d) Ensure that passwords or other security mechanisms are used for sensitive data that needs to be stored or delivered electronically.
       e) Report data quality issues to appropriate data steward.

Changes and amendments

This policy shall be amended by a vote of 50% plus one of all members of the Data Steering Committee. Changes enter into effect 30 calendar days after notification. Amendment and notification are required for extension of the scope of, or changes to, membership; etc.
<table>
<thead>
<tr>
<th>Item</th>
<th>Definition</th>
<th>Purpose</th>
</tr>
</thead>
</table>
| Governance framework        | The data governance framework lays down guidelines for how the organization sets up and enforces data and analytics governance. It includes the practices, processes, and procedures to manage, utilize, and protect data. Data governance often includes other concepts such as data stewardship, data quality, etc. to help gain better control over the data assets. | The governance framework aims at:  
- Enabling decision making  
- Reducing operational friction  
- Protecting the needs of the stakeholders  
- Encouraging management and staff to adapt common approaches  
- Building standards and processes  
- Reducing the costs and increase the effectiveness  
- Ensuring transparency  
- Setting out data access and privacy norms. |
| Governance structure        | A team or teams of people that are convened for a defined purpose, or for a limited term, which establish a system of roles, responsibilities, and decision-making rights for the teams that manage the organization to achieve a defined purpose. The following steps may be used to establish governance:  
  - Project Charter or Work Group Charter  
  - Stakeholder Analysis  
  - Plans, Policies, and Procedures  
  - Governance Maturity Roadmap or Plan | Through the authority and activities of governance bodies, entities can collaborate and manage efforts to achieve mutual goals. Sound data governance will guide all other analytics activities within the organization. It will also show where the decision-making authority lies regarding any activity as well as establish the parameters by which to select any stakeholder. Sound data governance will also provide a sense of ownership relative to the various data and systems for a given project or activity. |
| Privacy, Safeguards and security tools | Signed statements, documents or agreements used to govern the privacy and protection of the physical, administrative, legal and technical use of data through its access, collection, transmission, and storage. The purpose of data activities shall be clearly defined for specified reasons and must be in accordance with the law and regulations. | Customs are generally required by law to protect data to minimize the risk of compromising secure information. More importantly, the credibility of the analytics activity is jeopardized when privacy and security protocols are breached by those without |
Separate documents or statements may be created at the outset of a project or initiative when governance is being established, such as:

- Memorandum of Understanding (MOU) or Agreement (MOA)
- Data Sharing Agreement
- Security Agreement
- Data Access and Confidentiality Statements.

a right to access to the information.

References


### Annex 9: Data analytics job catalogue

<table>
<thead>
<tr>
<th>Job family</th>
<th>Job position</th>
<th>Job definition: main activities and responsibilities</th>
</tr>
</thead>
</table>
| Data science | Data analyst | Responsible for the visualization and processing of massive amounts of data, and for making queries in the databases from time to time. The main activities of a data analyst are:  
- Collaborate closely with business to understand their requirements and needs.  
- Develop reproducible business reports, leveraging a variety of visualization tools to communicate key insights into the business, with easily interpretable data insights.  
- Collaborate with data engineers to identify and source the required data for enabling business data-driven decision making.  
- Analyse and mine big data to aid business with intelligence/meaningful insights. |
|            | Data engineer | The main role of this position is to focus on building and maintaining the full technology infrastructure which enables storage and processing of corporate big data.  
- Extract, load, process, clean and transform data from a variety of heterogeneous sources into one or more consolidated channels as per the need of the solution.  
- Build data pipelines, perform data validation, manage data security and policies and strive for defined data availability, etc.  
- Manage the corporate analytics server platform and provide effective support to all processes to load and manage the analytics data store.  
Integrate new data sources, ensure capacity, backups and disaster recovery processes are in place. |
|            | Data scientist | The focus of this position is on data itself and on the analytical methods for the transformation of data into insights. To this end, the responsibilities consist of identifying data patterns, designing and implementing data models and statistical methods, and integrating best practices for a continuous improvement. The data scientist should perform the following activities:  
- collaborate with business to identify ML use-cases;  
- explore, analyse, design, experiment and prototype ML use-cases that solve business problems with specific outcomes;  
- communicate the impact/results of ML prototypes to business with business-relevant KPIs. |
|            | Data architect | A data architect works closely with data engineers to ensure that they have the best tools, software and systems to work with. The main role of a data architect consists of creating the blueprints for data management so |
that the databases can be easily integrated, centralized, and protected with the best security measures.

| Data business analyst | This position is quite different from the other data science positions. This role is responsible for identifying how the big data can be linked to actionable business insights for business impacts, and includes elements of organizational effectiveness. The data business analyst is in charge of:
- leading decision making through analytics, and providing analytical support to business initiatives,
- conducting analysis and documenting business needs, communicating progress and results effectively;
- handling large volumes of data and separating the high-value data from the low-value data. |

| Data analytics manager | This managerial position includes both line managers and the director level. The focus of this position is to guide analytic teams and assist the business in using data and analytic products effectively to impact decisions. The main activities of a data analytics manager are:
- bringing together the business and analytic interests of the organization to build bridges and collaboration, the sharing of best practices, and the deployment of shared intellectual assets to achieve strategic goals;
- ensuring the data analytics strategic alignment. |
Annex 10: WCO Questionnaire on data analytics

1- Is your organization using data analytics\(^2\) to improve the decision making process and drive better insights?

☐ Yes
☐ No (if no, please answer questions 3 and 13)

2- Which of the following statements best describes the use of data analytics in your organization?

☐ Data analytics focus is dispersed across multiple targets, which may not be strategically important

☐ Data analytics activity is centred on defined key domains (risk management, PCA, targeting, etc.)

☐ Data analytics supports the organization and strategy

3- In what step of the data analytics process is your organization most likely to experience challenges? (you may choose several answers)

☐ Understanding data analytics

☐ Managing data

☐ Implementing data analytics

☐ Embedding data analytics into the existing processes

☐ Measuring the outputs and impact of data analytics

4- Which statement best describes how you prioritize the importance of data analytics requests?

☐ Prioritization is based on personal interest and influence

☐ Prioritization criteria exist but are not incorporated in a formal process

Prioritization criteria exist and are incorporated in a formal process in some Customs units

☐ Prioritization criteria exist and are incorporated in a formal process at the organization level

\(^2\)Data analytics in the Customs context concerns a complex process of examining large and varied data sets to uncover information - such as hidden patterns, unknown correlations, etc. - that can help organizations make informed business decisions.

It is a form of advanced analytics, which involves complex applications with elements such as predictive models, statistical algorithms and what-if analysis powered by high-performance analytics systems.

There are different approaches to, and associated outcomes of, data analytics that are based on the strategic objectives of the organizations, data availability, and resource availability.

They can be broadly grouped into the following:

• Descriptive – What happened and/or what is happening now based on historical and incoming data. To mine the analytics, a real-time dashboard and/or email reports are used.

• Diagnostic – A look at past performance to determine why it happened. The result of the analysis is often an analytic dashboard.

• Predictive – An analysis of likely scenarios of what might happen. The deliverables are usually a predictive forecast.

• Prescriptive – This type of analysis reveals what should be done. This is the most valuable kind of analysis and usually results in rules and recommendations for the next steps. Recommendations are made based on multiple predictive models and complex analytical evaluations as to what options (pros/cons) to choose.
5- Does your organization have systems and procedures in place to manage and protect data in order to ensure understandable, correct, complete and secure data?

☐ No data governance\(^3\) in place
☐ Some standards established within specific departments/units
☐ Data quality and data security rules established, documented and routinely revised (\textit{please send us a copy of your data governance framework or any related document})

6- Which of the following statements best describes how the organization assigns data ownership\(^5\)?

☐ Data ownership and stewardship\(^5\) roles are not defined (no dedicated person/structure responsible for making sure data is organized, accurate and available for reporting and analysis)
☐ IT department centralizes data (IT is responsible for data storage, preparation and for creating reports)
☐ Data ownership and stewardship roles are clearly defined and assigned (representatives from various departments, or from a steering committee or a governing body, to define ownership. Data stewards are identified and present their subject areas to the governing body).

7- Which of the following statements best describes how your data analytics governance structure\(^6\) is defined?

☐ No data analytics governance\(^7\) structure in place
☐ Data analytics governance is achieved through key persons working together rather than having a formal structure
☐ Data analytics governance is achieved through a few teams working together rather than having a formal structure
☐ There is a well-defined governance structure in the organization with clear responsibilities and leadership, and a process to allow for raising governance issues from the different levels of the

---

\(^3\) **Data governance** is a data management concept concerning the capability that enables an organization to ensure that high data quality exists throughout the complete life cycle of the data. The key focus areas of data governance include availability, usability, consistency, data integrity and data security, and comprise establishing processes to ensure effective data management throughout the enterprise such as accountability for the adverse effects of poor data quality and ensuring that the available data can be used by the entire organization.

\(^4\) **A data owner** is any individual or unit that has the legal operational authority to access and use specific data.

\(^5\) **A data steward** is a role that ensures that data governance processes are followed and that guidelines enforced, as well as recommending improvements to data governance processes.

\(^6\) **The data analytics governance structure** is composed of representatives from the different Customs units that are using analytics and includes an executive sponsor. Its role is to align analytics projects to Customs goals through the establishment of clear procedures and guidelines, to ensure the data analytics agenda aligns with organizational objectives and each project has a clear goal and measures for success.

\(^7\) **Data analytics governance** helps the organization make the most of the available data through the establishment of procedures and guidelines that define and communicate roles and responsibilities for data collection, analysis, implementation and assessment, and how those activities are supported. The guidelines include the following components:

- **Accountability**: creating precise definitions about who does what, when, why and how;
- **Accessibility**: ensure that analytics and reports are accurate, available, timely, and of direct value to users;
- **Community**: Representation from all relevant business units with full participation by business owners and technology teams;
- **Uniformity**: definitions and policies set for all analytics reports, terminology, KPIs & calculations, reporting cycles, information dissemination, and routine review of governance for enhancements.
organization (please send us a copy of your data analytics governance framework or any related document)

8- How do you access the data analytics knowledge developed within your organization?

☐ No centralized location with documentation of previous data analytics endeavours
☐ Centralized location with documentation of previous data analytics endeavours with different levels of access (depending on clearance)
☐ Centralized location with documentation of previous data analytics endeavours accessible to all

9- How do you leverage data analytics knowledge developed within your organization?

☐ No standards, guidelines to share previous data analytics projects in place
☐ A mechanism is in place to share best practices within the same Customs unit (risk management for example)
☐ A mechanism is in place to share best practices and innovation throughout the organization (please send us more details about this mechanism)

10- How does your organization match the analytical goals to the technological requirements?

☐ Lack of technology
☐ Data, technology, and expertise in disparate clusters across the organization
☐ Early stages of a structured approach
☐ All key data analytics resources are centralized or networked

11- How do you assure quality for an outsourced solution?

☐ There is no reference architecture
☐ The reference architecture is under development
☐ The reference architecture is used by some/key Customs units for their process and data requirements (please send us a copy of your reference architecture)
☐ The reference architecture is recognized as a standard for all data-related projects within the organization (please send us a copy of your reference architecture)

12- Does your organization use big data open source tools for big data processing and analysis (such as R, Python, QGIS etc.)?

☐ The organization does not use big data open source tools
☐ The organization is using big data open source tools to develop pilot applications
☐ The organization is using big data open source tools to build scalable and extensible applications

---

*A reference architecture is a document, or set of documents, that provides recommended structures and integrations of IT products and services to form a solution to ensure the consistency and applicability of technology use within an organization.*
13- Which of the following statements best describes who supports data analytics and guides organizational changes towards a data driven organization?

☐ No awareness or interest
☐ Data analytics informally promoted by individuals in the organization
☐ Visible and active leadership at the functional level promotes data analytics without a formal mandate
☐ Leadership actively encourages and supports data analytics and exploration throughout the organization

14- Which statements best describes the level of data analytics competency in your organization?

☐ Isolated pockets of data analysts with little communication
☐ Data analysts clustered in key Customs areas (IT, risk management, targeting, PCA, etc.)
☐ Highly capable analysts in central or networked organization

15- Does the organization have a specific knowledge transfer scheme in data analytics?

☐ No formal training programme in place
☐ Formal training curriculum with different levels (please send us more details about your training curriculum)
☐ On-the-job training of data science practitioners

16- What statement best describes how analytics is embedded in the talent life cycle?

☐ Data analytics needs are not taken into consideration when hiring
☐ Data analytics needs are defined and considered when hiring for/developing specific roles (please send us examples of job profiles)
☐ Data analytics needs are defined and influence hiring and talent development strategies for all roles