GOVERNANCE

GOVERNANCE

EQUITABLE GROWTH, FINANCE & INSTITUTIONS INSIGHT

Digital Transformation of Tax and Customs Administrations

Raúl Félix Junquera-Varela
Cristian Óliver Lucas-Mas
Ivan Krsul
Vladimir Calderon
Paola Arce

WORLD BANK GROUP

>>>

Contents

Abstract

Domestic resource mobilization has become a core priority of the sustainable development agenda for tax and customs administrations. Information systems can play a critical role in revenue mobilization, which may create the much-needed fiscal space for maneuver and allow for more spending on all the things that drive potential growth over the medium term. New technologies can also increase the effectiveness of the internal operations of tax and customs administrations, and can reduce costs, as they improve their capacity to collect revenue with smarter use of the information they collect. Of particular interest is machine learning, which can be used to solve difficult problems that arise from the inability of revenue administrations to process massive amounts of data efficiently. Technology by itself can only provide tools. To achieve meaningful and impactful goals, a comprehensive strategy must be defined, covering the regulatory, institutional, and operational aspects. This paper analyzes such aspects and provides a roadmap for policymakers and tax officials on how to incorporate and manage disruptive technologies into the process of building the tax and customs administrations of tomorrow.

The authors wish to thank peer reviewers Charles Victor Blanco, Ana Cebreiro, Julia Dhimitri, Viet Anh Nguyen, Pratheep Ponraj, and Moses Wasike for their valuable comments and observations.

>>>

Executive Summary

Nowadays, tax administration is essentially a business of information management. These elements are strongly supported using Information and communication technology (ICT). Today the use of ICT is not a choice, it is a necessity. Technological change impacts organizational structure, business processes, and human resources policies. Beyond the mere adoption of new tools and technologies, real digitalization of the tax administration involves a comprehensive legal and institutional transformation. This process encompasses all the required adjustments to traditional operational models in order to achieve long-term and sustainable efficiencies, offer new and improved services to taxpayers, and develop new capabilities in key areas like digital invoicing, tax payments, digital fiscalization, advanced data analytics, and value chains and factoring. Such developments allow tax and customs administrations process huge volumes of information, and increase the reliability, accuracy, and timeliness of the information processed, which altogether reduce administration costs.

Domestic resource mobilization has become a core priority of the sustainable development agenda for tax and customs administrations. Information systems can play a critical role in revenue mobilization, and successful revenue mobilization efforts can create the much-needed fiscal space for maneuver and allow for more spending on all the things that drive potential growth over the medium term, including infrastructure, healthcare, and education. Taxpayer information systems can also contribute to the reduction in malpractice and corruption in a tax administration, since taxpayers who interact electronically with the organization do not need to come face-to-face with tax officials on standard operations. Where taxpayers interact with tax officials, effective information systems provide a natural deterrent for abuse since the results of discretionary actions are typically recorded in audit trails, increasing the risk that corrupt tax officials will be caught. With a reduction of malpractice and corruption, tax administrations can expect a subsequent increase in revenue for the organization. An additional benefit is the potential contribution to environmental sustainability, including mitigating the carbon footprint.

Based on experience, key success factors to digital transformation therefore include:

(i) establishing transformational strategy and vision;
(ii) achieving the ideal organizational structure;
(iii) ensuring that the digital transformation is driven by the strategy, a transformational roadmap, and a clear action plan;
(iv) a strong team to champion the process;
(v) recognition of the importance of the human factor;
(vi) addressing business process improvement;
(vii) addressing fragmentation, structure, and quality of systems and data; and (viii) promoting user adoption and trust through an inclusive change management program.

Organizational structure to centralize, standardize, and streamline IT initiatives is an important aspect in building institutional IT capacity and effective resource utilization. To this end, it is necessary to develop ICT and human resource assessment tools that help understand and assess the human and institutional capacity gaps of a tax administration. Moreover, the COVID-19 crisis has had a profound impact on digitalization as a strategic tool. Relevant improvements include establishing and maintaining remote work capabilities, and adopting technological preparedness and adjustments (i.e., many jurisdictions rely on tax authorities' registries to identify the recipients of business stimulus payments). Also, data quality is crucial in optimizing the availability and value of the data required to meet the administrations' objectives. Processes and protocols should be in place to ensure an acceptable level of confidence in the data. The World Bank is supporting e-governance initiatives aimed at fully automating business processes of tax administrations and customs in close collaboration with other public sector institutions to improve delivery of services.

The World Bank has been providing technical assistance to countries in the design and implementation of high-tech digital tax and customs administrations. The roadmap includes basic e-tax administration; enforcing use of data to strengthen and expand e-services; data-analytics-based risk profiling and real-time horizontal monitoring; and "digital by default" tax administration that is algorithm-based, which comprises predictive data analytics and artificial intelligence. The World Bank support also extends to the web-based Tax Diamond diagnostic tool (DIAMOND) that it has developed, which comprises a set of self-assessment modules for countries. The tool enables governments to conduct a tax administration functional review, evaluate and determine their infrastructure investment needs, assess the ICT landscape of their tax administration and customs, and improve their methodologies for business process reengineering and business process mapping. In-depth assessments are also available for the functional areas of human resources, international taxation, tax audit, tax gap and revenue forecasting, domestic tax evasion, and offshore tax evasion.

The DIAMOND modules assess the overall performance of tax administrations and customs by measuring the gap between actual performance and good practices and standards (external benchmarking). By using specific dimensions, the tool facilitates prioritizing plans for improvement and delineates a technical assistance program and action plan aimed at addressing identified performance gaps. Against this backdrop, it is important to build a model that assesses the general level of maturity of revenue administrations and customs so that plans for improvement are tailored to a given context. This is especially relevant in low-capacity environments and is more effective than a benchmarking exercise.

A key goal of any maturity model is to let us move away from the present model that suggests that moving consistently toward functionality configurations we know from advanced countries is the appropriate reform path. The problem with this is two-fold: the focus is on building maturity and modern functionalities that may not be sustainable and are often implemented in a piecemeal manner that shortcuts actual functional improvement. For immediate results and sustainability, one should start assessing what is the binding constraint to achieving a level of functionality that matches the level of capacity and maturity.

In specific cases of low capacity, the priority should be to identify binding constraints to achieving comprehensive functionality (which acknowledges the limited performance that may come with that). Lower priority can be assigned to building maturity in other areas; they can be done at the same time but should clearly fit into the model. Maturity building may take time and it is important to take advantage of reform opportunities.

Embedding this kind of analysis in DIAMOND leads us precisely here. We do not assess revenue administrations by the distance from the "really good practice frontier" but in the form of constraints to achieve comprehensive functionality appropriate to the level of maturity—identify taxpayers, assess taxes, collect, investigate and audit, manage dispute, report, and offer transparency and accountability in a manner that imposes reasonable cost for taxpayers and revenue administration. That does not mean we do not push for the maturity dimension, but we are not pursuing maturity development for its own sake or because there is a belief that

over time all the pieces of the puzzle will come together if we forge ahead.

The paper presents the DIAMOND's four-level models and assesses the level of maturity of tax and customs administrations in terms of their current capabilities. The usual practices at lower levels of maturity are also described so that plans for improvement can be made. By using the DIAMOND maturity models, revenue administrations and customs can monitor their progress, practice by practice, to better identify how to sustain improvement and performance across areas and functions. Also, such model helps leverage existing systems, processes, and tools in designing strategies to overcome the gap in tax administration and customs capacity.

In defining a useful baseline and benchmark mechanisms to accurately identify the actual maturity level of a tax administration in relation to information technology, it is very important to keep in mind that the primary objective is not to simply follow the latest trend or hype in the industry. Instead, the institution should have a healthy and comprehensive long-term strategy on how to deal with the ever-evolving technological landscape. The tax administration, in general, has two options in fulfilling its IT requirements: (i) off-the-shelf solutions, or (ii) in-house software development. In practice, most of the time, a combined scheme is applied. The institution must be equipped with the necessary knowledge, processes, and resources to adequately evaluate, acquire, and integrate the existing products, and to engage in productive and effective development, when required. Still, regardless of the option taken, the tax administration must have an ICT unit that is robust enough to (i) provide continuity and sustainability to the technological solutions, and (ii) avoid falling as a client captive of some external company.

It is of paramount importance for the tax administration to clearly understand that technology can provide better tools, but even a very good tool is not a complete solution on its own. A good solution involves careful consideration, design, and evaluation. It must start with a clear definition of the problem and the mechanics required to measure the gained efficiencies objectively and quantifiably. Innovation is not obtained by purchasing the latest digital technology. Instead, innovation must become an integral part of the organization's culture. To this end, it must become a permanent goal aimed at the taxpayers´ (client) needs—aligning them to institutional priorities, improving and refining existing processes to simplify and facilitate compliance, and constantly evaluating and responsibly adopting new technological advancements to enhance the institution's level of maturity and functionality.

Thanks to information technology (IT), tax and customs administrations are now able to manage great amounts of third-party information, enabling them to massively crosscheck this information with the content of tax returns and customs declarations. In addition, IT-based compliance risk management processes result in a better selection of cases for audit, which makes the tax audit function more efficient. Consequently, the maturity model for the IT area becomes a key element in assessing the IT performance gap and in designing an action plan. This helps build the data science capabilities needed to advance to the next maturity level, which is part of the digital transformation of tax and customs administrations. Based on the maturity levels in IT and with reference to best practices in IT system implementation, the paper examines how to build data science capabilities in revenue administrations, focusing on data management and data science tools, the creation of machine learning capabilities and their application, and the feasibility on the use of blockchain initiatives.

Finally, it is important to realize that digitalization is the key enabler for revenue authorities. The key question is how to properly sequence the IT infrastructure and the institutional reform needed to make digitalization happen. Digitalization of tax and customs administrations should be adapted to the environment available in each country and to the maturity level of each revenue administration. It is important to note that the pacing of each tax and customs administration varies. Governments must take a strategic rather than opportunistic perspective and make digitalization an integral part of their internal strategy with clear policy objectives. To this end, the World Bank, the Vienna University Global Tax Policy Center, and Ernst & Young have established a seminar series on digital transformation of tax and customs administrations, which aims to develop a digital tax administration roadmap.

>>>

Introduction

Over the recent years, the adoption of technologies by tax administrations has been advancing their performances in two different spheres: (i) better provision of e-services to taxpayers, and (ii) strengthening of tax compliance control mechanisms which leads to the increased collection of tax revenue. However, many tax authorities still rely on burdensome, paper-based, and lengthy audit processes. The changes occurring around increased digitalization of the economy and society in general call for a different model of tax and customs administrations.

Tax and customs authorities should set out a vision and take a strategic rather than opportunistic approach to digitalization by developing a long-term digital roadmap with clear project objectives, ensuring that early decisions do not constrain future developments (i.e., a modular approach with off-ramps). Several factors should be considered when undergoing a digital transformation, such as the available technology, data, current processes, laws and regulations, resources, and personnel.

Digitalization should be primarily based on the specific needs of the revenue authorities. When developing a roadmap and prior to digitalization, tax and customs administrations need to declutter the administrative rules, eliminating unnecessary reporting requirements and ensuring that those that are kept would fit into the digital age. It is important to establish strong leadership commitment at the executive level and create governance structures that remove blockages and allow for collaboration, while holding project managers accountable. Ensuring the quality of the data collected and that it is fit-for-purpose and relevant is a key aspect toward effectively digitalizing tax and customs administrations.

An increasingly connected digital society is reshaping the economy by creating new products, services, and business models. Disruptive technologies are changing the way taxpayers and tax authorities interact. More importantly, they are altering the way taxes are paid and the way information is stored and used. The digital age is also reshaping the tax systems. Big data, cloud computing, social media, blockchain, the internet of things, 3D printing, and machine learning are examples of disruptive technologies that are transforming businesses and governments. These disruptive technologies have been shaking the tax administrations' tree.

The term "disruptive technologies" has been coined to signal the strong impact that technology has on the way tax systems are designed and administered. How then do we unlock the full potential of these new technologies in order to transform tax and customs administrations? How do we ensure that they allow revenue administrations to administer taxes more effectively and efficiently, enhance service delivery, and reduce administration costs and taxpayers' costs of compliance thereby improving business climate? With all the choices available to tax and customs administrations today, how do we select technologies that are most relevant and will support the achievement of stated objective with acceptable return of investment? Most probably tax and customs administrations of tomorrow will look very different from those of today.

When analyzing how revenue administrations can benefit from these disruptive technologies, it is very important to bear in mind that ICT is not "the end" but "the means" for this journey. Technology by itself can only provide tools, but to achieve meaningful and impactful goals, a comprehensive strategy must be defined, covering the regulatory, institutional, and operational aspects. Introduction of new technologies requires changes to tax legal framework and procedures, organizational structure, human resources, business processes, and in general, tax administration business model. The World Bank supports countries´ tax and customs administration in their efforts toward full digitalization.

The first section of this paper focuses on data analytics and information management, and the role of digitalization as a strategic tool for domestic resource mobilization. Digitalization improves the efficiency of tax collection, helps reduce cost, enables a more efficient fight against corruption, helps trace operations, and fosters transparency. Digitalization allows tax and customs administrations to evolve into a new role and find balance between facilitating tax compliance and maintaining effective control of taxpayers´ obligations.

In the second section, we analyze what are the factors taken into account when evaluating the digital maturity of tax and customs administrations; what metrics can we use to assess the progress in tax administration in digitalization efforts; and how does the digital maturity level of tax administrations relate to the starting point in implementing a digital roadmap. Digital maturity refers to the level of digitalization of tax procedures of tax authorities. A digital maturity index aims to evaluate, in a standardized form, the efforts of tax and customs administrations in transforming themselves into digitalized institutions. It takes into account not only the technology itself, but the potential combination of available technologies and the system´s integration that result in the most appropriate resource allocation.

Finally, data sciences and machine learning can significantly improve the efficiency of revenue administrations. This is examined in the third section, which explores what would a digital roadmap for building data science capabilities look like and how to determine the appropriate technology to be applied by tax and customs administrations, taking into account their maturity level.

This paper poses the questions what should be the sequence of digitalization and institutional reforms, how to get the right mix between tactical taxpayer service and long-term strategic process/approach change, how to deal with legacy systems and non-standard transactions, and what are some of the best practices we can learn from. This paper helps not only the tax policymakers and tax officials, but also IT experts who need to get an understanding of the needs of tax and customs administrations, in order to better design and implement the most appropriate technology solutions.

1

Digital Transformation of Tax and Customs Administrations

1.1 Digitalization as a Strategic Tool for Domestic Resource Mobilization

In view of the radical changes that have taken place globally, tax systems cannot be administered in the way they were administered two or three decades ago. This new panorama stems mainly from substantial changes in the economy context, globalization and financial integration, rapid development of new technologies, and new approaches to the role of taxation in modern and democratic societies. In the current global context and pandemic crisis, domestic resource mobilization (DRM) must be at the center of any development and economic growth strategies. This requires developing a strong analytical framework to help countries establish productive, efficient, and equitable tax systems at both national and subnational levels. Moreover, taxation must be recognized as a key driver for state building and accountability, and tax reform as a possible contributor to broader gains in state capacity and quality of governance.

In the context of policy process, tax administration is the key to successful tax reform. Tax administration matters because the best tax policy that is ineffectively administered amounts to nothing (see box 1). Policy outcomes depend very much on how policies are administered. Thus, critical aspects of tax administration must be integrated more closely with tax policy work. One of the key tasks is to provide countries with modern tax codes that foster a good business environment and incorporate the technological changes that will facilitate control of compliance and facilitation of taxpayers' obligations. Good examples are the recently approved tax codes of Tajikistan and Uzbekistan, which incorporate specific solutions to tax the digital economy and to digitalize all business processes of the revenue administration (see box 2). Consequently, it is vital to accelerate the move to a digital revenue administration, and to use the new wave of disruptive technologies to radically transform how taxes are administered. To remain relevant and effective, tax administrators must continually invest in scanning the external environment for emerging innovations in technology (such as cryptocurrencies and digital currency) and their implications on taxpayers' new ways of managing the risk of tax evasion.

> > >

BOX 1 - What NOT to Do to Achieve a Successful Digital Transformation

1. Do NOT ignore digital transformation.
2. Do NOT start off without a vision.
3. Do NOT invest in technology without training your human capital.
4. Do NOT underestimate the impact it will have on your organization´s culture.
5. Do NOT underinvest in the digitalization process.
6. Do NOT implement a one-size-fits-all strategy.
7. Do NOT give up after the initial phases.
8. Do NOT define a holistic view of all-things-digital.
9. Do NOT orchestrate a roadmap of technology-based wins.
10. Do NOT wait until it is too late.
11. Do NOT invest in the wrong technology tool.
12. Do NOT buy what you are not going to use.
13. Do NOT disregard the importance of testing and iterating.

> > >

BOX 2 - World Bank Support to Modernizing Tajikistan's Tax Regime

The Government of Tajikistan has made digitalization of the tax administration's processes a priority, which has resulted in the following achievements over the past decade:

2012 — The system of reporting and accounting for tax payments advanced with the introduction of state bank and cash registries, and the infrastructure expanded.

2013 — The State Tax Committee with World Bank assistance established a data processing center.

2014 — The government implemented a file and system management with a system for mobile devices and introduced a new terminal to address the problems of taxpayers with Internet connection problems.

2015 — The e-invoice system was finalized. All VAT returns are now issued automatically. The VAT return also reflects the income and this allows better monitoring of the whole tax position of the taxpayer.

2016 — A post-terminal system with transmission devices for data was established for the purpose of prefilling of tax returns. Legislation for prefilling returns was changed. Prefilling allows faster filling and submission (i.e., Customs and cargo declaration are sent directly to customs authorities and other agencies, which reduces non-compliance). Another priority was to assign tax identification numbers (TIN) in electronic form. The government was assigned to improve the quality of the services to the taxpayer.

2017 — In the area of e-services, tax information systems were created.

2019 — There were negotiations with the World Bank on how the provision of high-quality services should be ensured. A network with regional offices was built later. Pension funds were integrated, and included data exchanges. Gradually, more and more offices were included in the network, which are constantly increasing. In the area of VAT registration, Tajikistan created a single e-registry as source of electronic information that would be available on operational basis. Services provided not only to the government but also to businesses and taxpayers are part of the source system and there is a continuous interaction between organizations.

Currently, new programs and technologies are being implemented and the World Bank is providing support in introducing key technological developments, including:

(i) modernization of the tax system through e-services, tax returns, and taxation of the digital economy;
(ii) automated crediting of all taxes to the budget and links with the taxpayers' accounts (in online payment of taxes);
(iii) introduction of an automated collection process;
(iv) implementation of digital signature and upgrade of ICT infrastructure in the Tax Committee;
(v) implementation of an automated VAT refund system; and
(vi) automation of selected taxpayer services.

The overall results are very positive. More than 80% of taxpayers filed their tax return using an e-form, and 70% of them were legal entities; 90 % of them declared they were satisfied; time was reduced; and the new tax code is under development. The focus is hence on new systems for e-services and simplification of legal processes.

Source: Proceedings from the virtual seminars on tax and technology jointly hosted by the World Bank and the WU Global Tax Policy Center, 2020–2022.

Technological change impacts organizational structure, business processes, and human resources policies. Beyond the mere adoption of new tools and technologies, real digitalization of the tax administration involves a comprehensive legal and institutional transformation. This process encompasses all the required adjustments to traditional operational models in order to achieve long-term and sustainable efficiencies, offer new and improved services to taxpayers, and develop new capabilities in key areas like digital invoicing, tax payments, digital fiscalization, advanced data analytics, and value chains and factoring. Such developments allow tax and customs administrations process huge volumes of information, and increase the reliability, accuracy, and timeliness of the information processed, which altogether reduce administration costs.

Moreover, technology helps standardize and centralize routine processes and improve effectiveness of control of compliance. Organizational structure to centralize, standardize, and streamline IT initiatives is an important aspect in building institutional IT capacity and effective resource utilization. To this end, it is necessary to develop information and communication technology (ICT) and human resource (HR) assessment tools that help understand and assess the human and institutional capacity gaps of a tax administration. Also, as a step prior to digitalization, business process improvement (BPI) and business model change (BMC) are extremely relevant to the efficient transition between current and new business models (see box 3).

> > >

BOX 3 - Best Practices for a Strong Business Process Improvement

1. Ensure that the organization has well-documented processes that are clearly visible in the workflows and in all the positions, systems, and data that support them.
2. Adopt a methodology for analyzing processes and identifying opportunities to improve them.
3. Commit adequate resources to support business process improvement (BPI) as an ongoing exercise within the organization.
4. Solicit input and ideas from process stakeholders on possible areas of improvement.
5. Communicate improvement strategies to stakeholders to ensure buy-in.
6. Invest in adequate change management programs, including employee training, to ensure the successful implementation of improvement strategies.
7. Monitor results to ensure compliance with changed workflows within recently improved processes.
8. Establish metrics to measure the success of improvements.
9. Use those improvement metrics to gain executive support for additional BPI projects.
10. Incorporate the BPI discipline into an overall business process management practice.

Source: Pratt, Mary K. 2022. “Business Process Improvement (BPI).” TechTarget. https://www.techtarget.com/searchcio/definition/business-process-improvement-BPI.

The COVID-19 crisis has had a profound impact on digitalization as a strategic tool. Relevant improvements include establishing and maintaining remote work capabilities, and adopting technological preparedness and adjustments (i.e., many jurisdictions rely on tax authorities’ registries to identify the recipients of business stimulus payments). For instance, the United States Internal Revenue Service (IRS) continues to face technology challenges in the processing of economic impact payments to eligible recipients and in preventing improper payments, as recently published in its 2021 audit report. In this respect, public–private partnerships and engagement with open-source communities and consortiums could be beneficial in operationalizing digital transformation initiatives.

Based on experience, key success factors to digital transformation therefore include (i) establishing transformational strategy and vision; (ii) achieving the ideal organizational structure; (iii) ensuring that the digital transformation is driven by the strategy, a transformational roadmap, and a clear action plan; (iv) a strong team to champion the process; (v) recognition of the importance of the human factor; (vi) addressing business process improvement; (vii) addressing fragmentation, structure, and quality of systems and data; and (viii) promoting user adoption and trust through an inclusive change management program.

1.2 Tax Administration as a Business of Information Management

Nowadays, tax administration is essentially a business of information management. These elements are strongly supported using ICT. Today the use of ICT is not a choice, it is a necessity. The World Bank is supporting e-governance initiatives aimed at fully automating business processes of tax administrations and customs in close collaboration with other public sector institutions to improve delivery of services. In recent years, the Global Tax Program has put together a number of technical assistance (TA) programs on DRM. The TA programs adopt a holistic approach that combines tax policy advice with support on implementing international good practices in revenue administration. An example of this whole-of-government approach is the single window initiative for customs or support in developing and implementing strategies to combat the informal economy.

The World Bank puts a lot of emphasis on information management to ensure the quality of information that can be exploited efficiently by the tax administration. A program of this kind is now being implemented in Uzbekistan, both at customs and tax administration levels, where we are analyzing the quality of data contained in electronic invoices to ensure that this information can effectively be used for tax audit purposes. Business intelligence[1]—through a variety of tools and mechanisms such as machine learning, Big Data, or by exploring the potential of blockchain—is now taking center stage in our programs. Big Data has proven to be very effective, for example, in controlling VAT compliance or avoiding fraud schemes coming from fake VAT refunds. In customs, we are also implementing risk analysis tools for trade operators to improve control at the border and customs valuation.

The TA programs deliver analytical work that is vital to better understand a country context and better design future operations and loans. Good examples are Tajikistan and Uzbekistan (see box 4) wherein the World Bank is managing two TA programs that paved the way for two projects and loans that have been recently approved. The key objectives of these loans are the full automation and digitalization of the revenue administration, streamlining of business processes, effective use of information through business intelligence, and dramatic change in human resources policies. Evidence shows that the tax administrations that coped with the COVID-19 pandemic are those who were better prepared in terms of technology. And this applies not only to the COVID-19 pandemic but also to any potential crisis. This technological element is present in most of the World Bank's DRM projects (i.e., Pakistan, Tajikistan, Uzbekistan, and Nigeria, and finalized projects in Bulgaria and Colombia, to name a few).

> > >

BOX 4 - World Bank Support to Modernizing Uzbekistan's Tax Regime

The World Bank supported tax policy changes in Uzbekistan during 2019–2020 to introduce new information technologies and their effects on tax revenues. New legislative acts allowing the use of modern technologies have been enacted to improve tax collection and counter shadow economy.

The State Tax Committee has developed a strategy regarding the implementation of information and communication technologies (ICT), which includes the modernization of datacenters, automation of business processes in tax administration, transfer of tax information to a single platform, implementation of business intelligence and Big Data technology, merging the databases of tax and customs administrations, and making such databases available to all ministries and national agencies.

This initiative is expected to increase the speed of data processing and level of confidentiality, enhance the protection against security threats, broaden the base of information, and produce secure and reliable data storage.

All the above measures are expected to expand the tax base through the reception of more and reliable data, improve the efficiency of state tax authorities, and enhance the image of tax administration acting as a partner and consultant for taxpayers.

With the assistance and technical support of the World Bank and other organizations, a digital tool for the calculation of the VAT will be launched. A software tracking the movement of assets, merge of companies and capital, purchasing of new assets, and related activities will be developed and deployed.

Source: Proceedings from the virtual seminars on tax and technology jointly hosted by the World Bank and the WU Global Tax Policy Center, 2020–2022.

On the other hand, increased attention must be paid to issues related to mass communication (e.g., publishing of the form filings, data formats, and structures for electronic filings) and to the handling of information and privacy intrusion, which affect all taxpayers. For instance, massive collection of information combined with data mining techniques allow tax administrators to find data, which are even unknown to a taxpayer, and even use this information to profile the taxpayers. Some OECD tax administrations have identified three emerging risks regarding the access and use of information that may create increasing difficulties for tax administrations over time: (i) changing work patterns, (ii) changing business models, and (iii) digital transparency issues. These risks are already present to some extent and may be expected to grow over the coming years with the increasing digitalization of the economy.

1.3 The Impact of Information Technology Systems on Domestic Resource Mobilization

DRM has become a core priority of the sustainable development agenda for tax and customs administrations. Information systems can play a critical role in revenue mobilization, and successful revenue mobilization efforts can create the much-needed fiscal space for maneuver and allow for more spending on all the things that drive potential growth over the medium term, including infrastructure, healthcare, and education. More reliable sources of revenue would help avoid volatility in public expenditure and pro-cyclical fiscal policy. Information technology (IT) systems can also increase the effectiveness of the internal operation of the tax administration and can reduce costs, as tax administrations improve their capacity to collect revenue with smarter use of the information they collect.

Voluntary compliance is enhanced with the use of taxpayer information services because taxpayers have certainty about the taxes to be paid, have online systems that make it easy to file a tax return, and paying becomes convenient. When the organization can implement effective and sustainable taxpayer information services, revenue mobilization is enhanced.

Taxpayer information systems can also contribute to the reduction in malpractice and corruption in a tax administration, since taxpayers who interact electronically with the organization do not need to come face-to-face with tax officials on standard operations. Where taxpayers interact with tax officials, effective information systems provide a natural deterrent for abuse since the results of discretionary actions are typically recorded in audit trails, increasing the risk that corrupt tax officials will be caught. With a reduction of malpractice and corruption, tax administrations can expect a subsequent increase in revenue for the organization. An additional benefit is the potential contribution to environmental sustainability, including mitigating the carbon footprint.

Information systems can also increase the effectiveness of the internal operation of the tax administration and can reduce costs. Through information systems, integrated taxpayer registries are created to collect the basic information needed to manage taxpayers. Process-intensive functions (such as form and payment processing) as well as taxpayer accounting are also automated. Selective monitoring compliance and selective enforcing compliance to reduce costs of compliance and administration can be implemented with information systems, therefore channeling resources directly to compliance activities and taxpayer services.

By automating manual functions, the tax organization can move to a compliance risk management model that systematically identifies, assesses, ranks, and treats tax compliance risks so that the tax administration can effectively deploy its limited resources. The necessary capabilities for improving recovery of arrears and tax debt, supporting intelligence and fraud detection, and identifying tax gaps are therefore created. Hence, effective implementation of information systems for a tax administration has the potential to significantly increase its efficiency, as ICT provides technological support for all functions of the administration.

In terms of which IT systems are more relevant for tax administrations and customs, recently, the focus is on Big Data[2] and data analytics,[3] artificial intelligence[4] and machine learning,[5] natural language process, cloud computing, robotics processing automation (RPA), and distributed ledger technology[6] (DLT) represented by blockchain[7]; however, these are not the most relevant. In fact, in the IT toolkit to be considered by a tax administration and customs, other technologies should be given priority, such as data visualization,[8] statistical analysis,[9] graph databases, edge computing, mobile collaboration and e-learning, predictive analytics, hyper automation, robotic process automation (RPA), XaaS (everything as a service), application containerization, DevOps, continuous integration and delivery, and serverless computing. Unfortunately, often, tax administrations tend to neglect basic technical needs in favor of social and political trends. Before applying artificial intelligence or blockchain, tax administrations and customs must ensure that the required underlying structural IT systems are in place and are fully functional.

It is important to properly determine the scope of tax administration functions to be covered by the digital transformation process, which goes beyond the traditional assessment methods, filing duties, and audit inspections. Key functions must also include registration, document management, legal obligations, taxpayer assistance mechanisms, communication channels, notification practices, taxpayer portals, appeals procedures, and any other technical functions like risk management.

Finally, contrary to widespread belief, in practice, there is no transitional period during the digital transformation process since the IT systems evolve constantly. In fact, the optimal and smart strategy is to never “change” the IT infrastructure, but instead to create the conditions for a continued and gradual evolution of the IT systems. The outdated model which periodically consists of conducting a full-length renovation has become too risky and unnecessary. Therefore, the new approach that should be adopted by tax administrations and customs is to incorporate an IT strategy based on permanent small-scale improvements that constantly transform the functional practices without creating any disruptions. Technological change cannot be perceived like a start-from-scratch process. IT specialists must stop thinking in terms of “overhauls” but instead integrate an evolutionary perspective into their IT practices.

1.4 Exploiting Data Quality Management to its Full Potential

Awareness of data quality needs to be raised. A lot has changed in the world of data in the last ten years. The amount of data, which is constantly growing, has increased the possibilities to gain valuable insights from complex data; and our dependence on it and the problems caused by poor information have expanded as well.

Today, almost every administration has Big Data at the top of their IT list. The importance of this element cannot be overemphasized, as it is spawning innovation, uncovering opportunities, and optimizing resources in every institution. However, although administrations are aware of this, very few are taking actions to keep a handle on exactly what data they are receiving and what kind of shape it is in. Most of the tax and customs administrations, whether knowingly or unknowingly, do not recognize that they have a significant problem in terms of data quality management, or the potential for one. Even the administrations that do recognize it are often hesitant to allocate financial resources and manpower to improve data quality.

Resistance to invest in data quality may be attributed to the lack of awareness regarding its impact especially in the core areas. It is common to see administrations that are too focused on developing automation, getting advance information, signing memorandums of understanding with other government agencies to exchange information, trying to develop comprehensive risk assessment systems, and implementing the most advanced tools in business intelligence. These modernization actions may make it appear that the administration is moving forward, but in practice, there is not much progress happening because the data reliability is not effectively satisfying the administration´s needs.

After conducting a data quality assessment, some administrations have found that only 40 percent to 60 percent of the data are useful enough to execute operational processes that comes from automation or are reliable for business analytics, statistical reporting, and risk management functions. With this amount of data, how do administrations realistically expect to improve shortfalls, deter tax evasion, and promote facilitation and compliance by minimizing the impact on taxpayers?

If more administrations were conscious about how poor data quality results in practical consequences, more resources would be directed to avoid data that is not fit-for-purpose. Poor data quality may result in short- or long-term operational issues and failure to provide services correctly; it may also weaken evidence, create mistrust, and cost reputational damage. The absence of data quality may lead to the following instances:

(i) Evidence-based decisions and policies would be only as good as the data they are based upon.
(ii) Missing or duplicate data could result in bad auditing practices, altered or non-objective reporting, and poor decision-making, leading to negative outcomes.
(iii) Unreliable or contradictory data can make it difficult to verify irregularities among taxpayers or traders, which can lead them to question the data accuracy, creating mistrust toward the administration.
(iv) There may be missed opportunities or failures in service provision.
(v) Risk mitigation could be affected due to possible inconsistencies in the risk scoring system and unreliable information to identify risk trends or fraud schemes.
(vi) Automated controls are nonexistent due to the lack of format and structure in the data fields.
(vii) Crosschecks cannot be properly implemented for an effective compliance monitoring.
(viii) There would be affectations in any integration process and the interoperability between agencies.
(ix) Administrations are unable to assess their own strategic or operational effectiveness.

Regarding common challenges and solutions, it is not enough to simply identify the consequences of poor data quality. Many times, no matter how simple it may seem, managers are unable to measure what it means not to have quality information. To increase awareness, it is necessary to illustrate in detail the challenges and implications that go along with the quality of information.

1. **Duplicated Data.** This is an issue every administration must deal with. A frequent case within tax and customs administrations is that although they are receiving the electronic invoice data, there is no standardization in the format of the invoice number. Without a unique number, there can be multiple invoices with the same number or an incorrect format that would not allow the system to identify a match for crosschecks and controls. To avoid this, data duplication tools are completely necessary. These solutions have improved considerably; now they are smart enough to spot even substantially different entries for the same taxpayer.

2. **Inconsistent formats.** This refers to inputting data that covers the same information but is stored in different formats. For example, dates are a complex field to many systems, as there are many potential ways these could be entered into the system. Other potential difficulties may arise from tax identification format, invoice numbers, addresses, and phone numbers, especially when some have area codes and others don't. Therefore, it is vital to specify the exact format for every piece of data to ensure consistency across every source the administration uses. The most effective solution for this challenge is to define guidelines for lodging information, supporting it with validation rules for data consistency.

3. **Incomplete information.** This refers to the fields that are not completely filled in or are left blank altogether; and those can be a major pain for analytic tools as well as for Big Data algorithms. For example, entries that lack zip codes or invoice numbers are not just a problem when it comes to crosschecking data with other sources; it can also make the key analytics process useless if the analysis is based on geographical information that can help us spot trends and improve targeting efficiency. It is also common to see blank fields, generic or vague information on cargo description, another high percentage of entry summary declarations, or commercial invoices stating logistics companies as the consignee/importer without letting the authority know who the real entity is. With such data problems, it may seem impossible for the risk analysis units to target something that will not show up in a query because the data was lodged as incomplete, vague, or inconsistent. Validation rules are an effective solution to ensure that records cannot be created unless all essential information is included. If there is one field that does not comply with the pre-defined format, the system would simply reject the message.

4. **Multiple units and languages.** This is the case for invoices, transport documents, or advanced cargo information. Like the case on formatting, sometimes differences in language script or units of measurement can create difficulties if the analytics tools do not recognize it nor know how to translate it. Even special characters can wreak havoc if a system has not been configured for them. Therefore, administrations may need to consider these potential issues and program the algorithms accordingly. As a first step, it is important to define as many fields as possible as coded identifiers. For instance, instead of having a text field to input the cargo description, it is better to require just the Harmonized Tariff Schedule (HTS) number, or the code previously defined in a catalogue. As

a second step, the administration can work on creating a data dictionary to help improve the analysis.

5. **Inaccurate data.** There is no point in running Big Data analytics or conducting a risk assessment based on data that is just plain wrong. There could be many reasons for this— from taxpayers giving incorrect information to making a typo when entering data manually, or inputting details into the wrong field. These can often be among the hardest data quality issues to spot, especially if the formatting is still acceptable. Entering an incorrect, but valid tax identification number, for example, might go unnoticed by a database that only checks the veracity of that isolated input. Of course, there is no cure for human error, but ensuring that the administration has clear procedures being followed consistently is a good start. Creating validation rules can also help not only with the quality of the information but also with compliance.

Data quality is like a telescope that allows us to see clearly distant objects. The better the telescope, the greater the administration´s needs are met. Tax and customs risk management strategies rely on adequate and accurate quality data that enable administrations to make better-informed auditing and cargo processing decisions. Therefore, data management should not be viewed as a project or a program, but instead as a strategic discipline.

Data quality is crucial in optimizing the availability and value of the data required to meet the administrations' objectives. Processes and protocols should be in place to ensure an acceptable level of confidence in the data, and that is based on two aspects—the relevance of data for its intended use and its reliability. Validation rules are regulations established by the administration through a system that oblige the taxpayer to enter information in the indicated and consistent form when filing an invoice, tax return, or customs entry, among others. The type of validation can be implemented based on format and compliance, including:

- form and syntax;
- validations of guidelines that establish specific formalities in the procedures;
- complementary validations (this is when there are specific requirements that come from certain scenarios or entry fillings); and
- rules regarding compliance with the requirements of the Tax or Customs Code.

Validation rules may help "clean" the data before it is inputted into the database. By checking against the validation rules, it is possible to test whether the data meet the defined criteria and possess the required attributes. A good classification can be based on format, logic, informative data, catalogue reference, and conditionals.

It is important to differentiate the distinct roles of core stakeholders—systems developers and vendors, systems managers, internal and external users, and independent oversight—including institutional risk managers, internal auditors, and external auditors. Also, it is useful to separately consider and discuss internally generated vis-à-vis externally sourced data either from stratified sources (taxpayers' formal and informal records) or from "the internet of things" sources (mainly for analytical and business-intelligence-based risk management). Similarly, when establishing and maintaining credible intergovernmental data exchange interfaces, it is important to include interfaces with taxpayers, as is widely applied for indirect taxes, mainly for sales/use tax and excise duty from manufacturers. Maintaining credible historical data is also valuable for accurate revenue projections and trend analyses.

2

Maturity Models for Tax and Customs Administrations

2.1 Rationale and Methodological Considerations

The World Bank has been providing technical assistance to countries in the design and implementation of high-tech digital tax and customs administrations. The roadmap includes basic e-tax administration; enforcing use of data to strengthen and expand e-services; data-analytics-based risk profiling and real-time horizontal monitoring; and "digital by default" tax administration that is algorithm-based, which comprises predictive data analytics and artificial intelligence. The World Bank support also extends to the web-based Tax Diamond diagnostic tool (DIAMOND) that it has developed, which comprises a set of self-assessment modules for countries. The tool enables governments to conduct a tax administration functional review, evaluate and determine their infrastructure investment needs, assess the ICT landscape of their tax administration and customs, and improve their methodologies for business process reengineering and business process mapping. In-depth assessments are also available for the functional areas of human resources, international taxation, tax audit, tax gap and revenue forecasting, domestic tax evasion, and offshore tax evasion.

The Development Implementation and Monitoring Directives (DIAMOND) is an integrated assessment tool for measuring tax and customs administrations´ performance.[10] The assessment tool allows the collection of data and information for different functions, units, and departments, and subsequently does the measurements, which provide the organization an overall description about how it is operating and delivering services. Measurements are condensed in a set of key indicators organized in a flexible and adaptable manner to reflect the local context under which the organization operates. This customization process enables more accurate and meaningful measurements. With these, the tool can be used to evaluate the relative strengths and weaknesses of any tax and customs administration and compare them to good practices. The distinctive feature of the DIAMOND tool is that all data is objectively verifiable and comparable across countries and across time periods.

The DIAMOND modules assess the overall performance of tax administrations and customs by measuring the gap between actual performance and good practices and standards (external benchmarking). By using specific dimensions, the tool facilitates prioritizing plans for improvement and delineates a TA program and action plan aimed at addressing identified performance gaps. Against this backdrop, it is important to build a model that assesses the general level of maturity of revenue administrations and customs so that plans for improvement are tailored to a given context. This is especially relevant in low-capacity environments and is more effective than a benchmarking exercise.

A key goal of any maturity model is to let us move away from the present model that suggests that moving consistently toward functionality configurations we know from advanced countries is the appropriate reform path. The problem with this is two-fold: the focus is on building maturity and modern functionalities that may not be sustainable and are often implemented in a piecemeal manner that shortcuts actual functional improvement. For immediate results and sustainability, one should start assessing what is the binding constraint to achieving a level of functionality that matches the level of capacity and maturity.

In specific cases of low capacity, the priority should be to identify binding constraints to achieving comprehensive functionality (which acknowledges the limited performance that may come with that). Lower priority can be assigned to building maturity in other areas; they can be done at the same time but should clearly fit into the model. Maturity building may take time and it is important to take advantage of reform opportunities.

Embedding this kind of analysis in DIAMOND leads us precisely here. We do not assess revenue administrations by the distance from the “really good practice frontier” but in the form of constraints to achieve comprehensive functionality appropriate to the level of maturity—identify taxpayers, assess taxes, collect, investigate and audit, manage dispute, report, and offer transparency and accountability in a manner that imposes reasonable cost for taxpayers and revenue administration. That does not mean we do not push for the maturity dimension, but we are not pursuing maturity development for its own sake or because there is a belief that over time all the pieces of the puzzle will come together if we forge ahead.

One should start by delineating the broad contours of a four-level model to assess the levels of maturity of revenue and customs administrations (see figure 1). This is the approach of the United States Agency for International Development (USAID), that is, relying on the “rules of thumb” to assess the strengths and weaknesses of revenue administrations’ key functions. USAID published in 2013 a report on *Detailed Guidelines for Improved Tax Administration in Latin America and the Caribbean,* which compiles key benchmarks to evaluate tax administration performance by areas, functions, and operations. This section is inspired by that report and draws on some of the maturity models developed therein. The DIAMOND adopts the same approach, which incorporates a four-level model for the progressive application of good practices. This is what our team did in Uganda when assessing the performance of Uganda Revenue Authority (URA). All indicators and good practices were classified according to four levels of maturity of revenue administrations. This exercise also involved testing the consistency of results achieved by using the first phase of the maturity model methodology.

Subsequently, the team developed in Colombia a more robust methodology to allocate standards, good practices, and indicators to the different levels. Strong emphasis was also placed on better describing the different levels and relevant variables, considering the various situations of limited maturity of revenue administrations. In Colombia, the team also analyzed how this level of maturity model plays out in terms of designing technical assistance programs that aim to achieve comprehensive functionality of revenue bodies and improve overall performance.

In every Tax Diamond assessment, every indicator reflects a good practice, and every practice belongs to a certain practice maturity level (see figure 1).

> > >

FIGURE 1 - Levels of Maturity

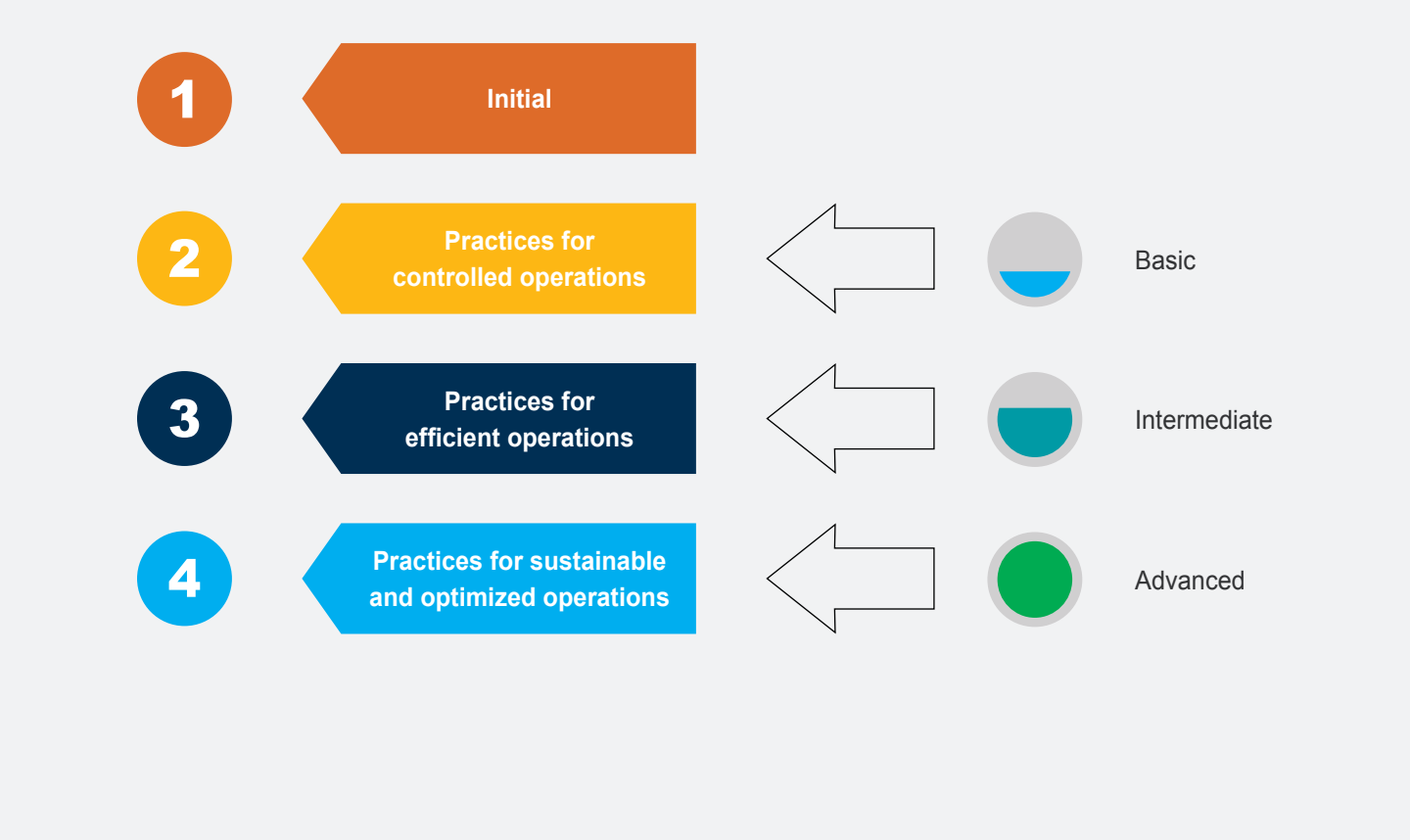
Source: Guidelines for the development of Tax Diamond assessments, World Bank internal document.

It is important to clarify that the term "Optimized Operations" in level 4 is not the same as the term "Efficient" in level 3. Optimized goes beyond mere efficiency; it pertains to making the design and operation of a system or process as good as possible in a defined sense, whereas efficiency is mostly about using resources reasonably well.

In terms of the natural order on the progressive implementation of the practices, as shown in figure 2, level 2 corresponds to the basic practices that should be implemented first in an organization. Level 3 corresponds to intermediate practices that are typically implemented once the basic practices are consolidated, and level 4 corresponds to advanced practices that require a consolidated and relatively mature organization that has appropriately implemented solid foundations that are necessary for implementing the most advanced practices.

> > >

FIGURE 2 - Progressive Implementation of the Practices



Source: Guidelines for the development of Tax Diamond assessments, World Bank internal document.

In general, good practices are implemented sequentially in an organization, although there is no requirement that one must implement the practices in order. However, it is generally difficult to make a practice efficient if the fundamentals are not implemented, and it is difficult to sustain if the operations are not efficient.

Hence, a DIAMOND assessment will determine the level of compliance of the organization with the practices at each level. Recommendations are typically provided to help the organization develop first with controlled operations, then with efficient operations, and finally with sustainable operations.

Table 1 describes the characteristics of a revenue administration's levels of maturity from the lens of USAID's maturity model on operations and DIAMOND's maturity models on processes, organization, and technology.

> > >

TABLE 1 - Characteristics of the Levels of Maturity of a Revenue Administration

Maturity level	Operations (USAID)	Processes (DIAMOND)	Organization (DIAMOND)	Technology (DIAMOND)
Level 1 **Initial**	Ad hoc: Operations are informal, sporadic, and ever changing	• Disintegrated, chaotic processes • Results are not consistently achieved • Variability in achieving results • Emergencies are very common	• Fragmented • High frustration or complacency • Functional silos • Low participation and commitment	• Rudimentary • Low level of technology adoption • Technological tools, when present and implemented, are not correctly used
Level 2 **Basic** Practices for controlled operations	Formalized: Operations are formalized, evidenced by regular practice or documentation	• Reactive processes • High costs involved in achieving results • Still variability in achieving results • Low visibility	• Compliant • Clear roles and responsibilities • Reward and punishment mechanisms have been established and implemented	• Structured • Information and data are structured using technological solutions • Technology is used to have a perception of control
Level 3 **Intermediate** Practices for efficient operations	Integrated: Policies, programs, processes, and tools are consistent	• Stable and predictable • Controlled, planned, balanced, and monitored processes. • Results are consistently achieved • Budget and costs are controlled • Performance management system is in place • Compliance with forecasts and plans • Variability is controlled	• Performance-oriented and collaborative • Leadership, teamwork, and accountability • Effective performance management systems are in place • Competencies are aligned with existing job profiles • Career streams are in place	• Deterministic • Systems and applications use and transform the organization's data. However, these applications obey to a very well- defined process and/or algorithm
Level 4 **Advanced** Practices for sustainable and optimized operations	Strategic: Organization strategy and performance goals filter through all levels	• Optimal • Existing processes are stable, flexible, and adaptable • Minimal variability • Continuous improvement and innovation	• Intelligent-led tax administration • Smart tax administration • Continuous innovation • Knowledge sharing	• Intelligent • Implemented tools and solutions use existing data as feedback to enhance the current applications • Organizations provide intelligent services

The next subsections present the DIAMOND's four-level models and assess the level of maturity of tax administrations and customs in terms of their current capabilities. The usual practices at lower levels of maturity are also described so that plans for improvement can be made. By using the DIAMOND maturity models, revenue administrations and customs can monitor their progress, practice by practice, to better identify how to sustain improvement and performance across areas and functions. Also, such model helps leverage existing systems, processes, and tools in designing strategies to overcome the gap in tax administration and customs capacity.

2.2 Maturity Model for Tax Administrations

Table 2 summarizes the main functional features and level-specific practices that characterize the different maturity levels of a tax administration. These practices are extracted from the DIAMOND tax administration functional evaluation and the business process mapping modules.

> > >

TABLE 2 - Maturity-Level-Specific Tax Administration Practices

Maturity level	Tax administration practices
Level 1 **Initial**	• Lack of adequate legal framework (legal/regulatory institutions, modern tax policy, civil service rules and regulations for attracting and retaining qualified staff, international accounting and professional standards, and modern financial and banking standards and institutions) • There are no policies and procedures in place to guide staff • The tax administration lacks control of taxpayer population • No provisions in the tax laws for self-assessments • Ineffective and inefficient taxpayer registry • Lack of a well-functioning taxpayer account • No compliance strategies in place • Potential for corruption • Voluntary compliance is not a concept used by the revenue administration • Informal economy is widespread and impacts tax administration functioning • Taxpayer services are largely non-existent • Relations between tax administration and taxpayers are confrontational • There is no segmentation of taxpayers to tailor processes and strategies to its distinctive features • There are no lines of communication with public and private sector institutions • Technology is not available, or it is available at a limited scale. Work is mostly conducted manually
Level 2 **Basic** Practices for controlled operations	• There is a formal process to register taxpayers but usually with unreliable tax identification numbers • Taxpayer accounts are largely unreliable • Segmentation of taxpayers has begun, however, well-defined criteria for inclusion in different segments is nonexistent • Progress made in incorporation provision in the law for self-assessment, and in the development of the concept of voluntary compliance and its inclusion in tax administration strategies • Development of an anticorruption strategy to limit opportunities for corruption • Taxpayer service program exist but disorganized and understaffed • Ill-conceived compliance strategies, which do not focus on high-risk segments • Long-term strategic plans for the overall tax administration do not exist

Table 2 continued

Maturity level	Tax administration practices
Level 2 **Basic** Practices for controlled operations	• Annual operational plans of the departments are independent and not coordinated • Some technology is available but is usually outdated and most of the work is still done manually • Procedure manuals are minimal, and institutionalization of procedures varies across departments • Skills of the staff vary across departments • Some contacts with public and private sector groups have started but there is lack of coherence and stability • Substantial lack of legal framework
Level 3 **Intermediate** Practices for efficient operations	• There are provisions in the law for self-assessment • Limited opportunities for corruption • A high percentage of taxpayers comply voluntarily (more than 75%) • Registration of taxpayers is supported by a good system of tax identification numbers and it has been completed • Taxpayer accounts are usually accurate • There is a segmentation process backed by good criteria to identify different segments • Compliance strategies focus on high-risk taxpayers • Strategic plans for the tax administration exist and these plans coordinate annual operational plans of core functions. There is still a greater focus on short- and medium-term objectives and a lack of focus on long-term direction • There are procedure and policy manuals for all the core functions, but they are not updated • Good relationships with public and private sector groups with some exceptions • Modern technology and equipment are available but there is often a shortage in specific departments • The tax administration has started to embrace many technological advances used in the private sector • Legal and regulatory institutions, modern tax policy, and civil service rules to support operations exist.
Level 4 **Advanced** Practices for sustainable and optimized operations	• There is a strong presence of legal and regulatory institutions, modern tax policy, and civil services rules, among others. • Provisions in the tax law for self-assessment have existed for several years • There are concise online procedure and policy manuals available for all tax administration's functions • Registration of taxpayers is accurate, and the taxpayer registry consists mostly of active taxpayers • More than 90% of taxpayers comply voluntarily • Taxpayer accounts are rarely inaccurate • Cases of corruption are rare • Segmentation of taxpayers is a dynamic process with well-defined criteria • Compliance programs tailored to different risks posed by segment of taxpayers are in place • Extensive use of third-party information to broaden the coverage and effectiveness of compliance programs • There are strategic plans focused on long-term objectives that guide the development and implementation of annual work plans • Relationships with public and private sector groups are very positive

Table 2 continued

Maturity level	Tax administration practices
Level **4** **Advanced** Practices for sustainable and optimized operations	• The tax administration has reliable information systems supported by the latest technology • The tax administration has already embraced and implemented many technological advances used in the private sector

2.3 Maturity Model for Customs Administrations

In general, maturity levels for customs should be assessed in the context of the overall maturity level identified for the tax administration. This will allow for a balanced and sensible allocation of resources and development efforts to move both revenue agencies from a "bad" equilibrium to a "good" equilibrium scenario.

Table 3 summarizes the main functional features and level-specific practices that characterize the different maturity levels for customs. The list below is extracted from the TA programs and good practices compiled throughout the World Bank missions.

> > >

TABLE 3 - Maturity-Level-Specific Customs Practices

Maturity level	Customs practices
Level **1** **Initial**	• Most of the processes are manually performed • Basic core, non-integrated information system exists in combination with a high number of physical examinations for imports • Use of warehouses is mandatory and most of them are publicly owned • Ad-hoc risk management system with no feedback mechanisms from operations and private stakeholders • Use of customs brokers is mandatory • Duties and taxes are paid in cash • Customs valuation is the main driver for tax collection based on discretionary decisions from customs officials opening room for integrity issues • Weak or nonexistent human resource management systems (HRM) • The customs budget depends on the decisions of the Ministry of Finance • Highly centralized organization wherein all decisions depend on the chairman • No systematic feedback from main stakeholders • Lack of strategic thinking • No exchange of information with the tax administration

Table 3 continued

Maturity level	Customs practices
Level 2 **Basic** Practices for controlled operations	• Imports and exports are managed by a core information system • High level of physical and documentary examinations • Some warehouses are licensed to the private sector, but no systematic audits are performed to renew licenses, and their use is still mandatory • Risk management is only linked to the Harmonized System codes that are considered sensitive • Selective system is preponderantly random-based • Minor imports don't need a customs broker • Some duties and taxes can be paid electronically through the trades' current accounts held by customs • Customs valuation is the main driver for revenue purposes and uses a valuation database for adjusting declared values • Incipient HRM systems are introduced • Some procurement capabilities exist for procuring ICT equipment • The decision-making process remains centralized • Sporadic meetings with private stakeholders exist, but no systematic follow up happens • Lack of strategic thinking • Some limited coordination with border agencies happens for security purposes • Per-case-based exchange of information with the tax administration
Level 3 **Intermediate** Practices for efficient operations	• Core business processes (imports, exports, warehousing, and transit) are managed by a core of an integrated information system • Between 10 percent and 30 percent of the imports are physically and documentarily examined • Incipient advance declarations are allowed • Formally, the use of warehouses is not mandatory, but in practice, importers have little options against it. • The use of customs brokers is not mandatory but collusion between customs and brokers discourages the traders to use this option • Risk management on traders' risk profiles is introduced only for importers and exporters combined with risk profiles for sensitive Harmonized System codes • Selectivity system frequently uses a random-based criterion • A transactional post-clearance audit (PCA) is introduced without links to the central risk management system (RMS) • Duties and taxes can be paid at authorized banks and electronically • HRM systems exist and are operational, but no systematic review and update are in place • Customs administration has a strategic plan but is not updated regularly and it is not used for driving the organization • The country has introduced a basic national single window system (NSWS) • Regular exchange of information with the tax administration • Meetings with the private sector are regular and most of the agreements are fulfilled but with some important delay

Table 3 continued

Maturity level	Customs practices
Level **4** **Advanced** Practices for sustainable and optimized operations	• All business processes including the administrative ones are managed through integrated information system • Pre-clearance and pre-arrival are operational and represent an important part of the operations • The use of warehouses is completely voluntary • The use of customs brokers is completely voluntary • A trustee trader program exists, and is operational • Only 1 percent to 5 percent of imports declarations are physically and documentarily examined • An entity-based PCA program exists, and it is operational and linked to the central RMS, focusing on customs valuation • Joint audits are performed together with the tax administration • Complete risk profiles for authorized operators, with inputs from the tax administration and the financial system • Advance ruling on classification and valuation exists and it is operational • An integrated tariff system exists, it is operational, and it is publicly available • The RMS is centralized and fully integrated into the core information system with systematic feedback from operations • The annual audit plan is risk-based and integrates all operational risks; it is reviewed and updated formally by the risk management committee • All electronic payments for duties, fees, and taxes are available • The use of warehouses is voluntary • The use of customs brokers is voluntary • Decentralized and semi-autonomous organization • HRM systems are complete and updated systematically • Meetings with the private sector are systematic and agreements are fulfilled promptly and effectively • The strategic plan drives the development of the organization • A mature NSW exists together with an integrated border management

Note: HRM = human resource management, ICT = information and communication technology, NSWS = national single window system, PCA = post-clearance audit, RMS = risk management system.

2.4 Maturity Model for Information Technology

In defining a useful baseline and benchmark mechanisms to accurately identify the actual maturity level of a tax administration in relation to information technology, it is very important to keep in mind that the primary objective is not to simply follow the latest trend or hype in the industry. Instead, the institution should have a healthy and comprehensive long-term strategy on how to deal with the ever-evolving technological landscape.

The tax administration, in general, has two options in fulfilling its IT requirements: (i) off-the-shelf solutions, or (ii) in-house software development. In practice, most of the time, a combined scheme is applied. The institution must be equipped with the necessary knowledge, processes, and resources to adequately evaluate, acquire, and integrate the existing products, and to engage in productive and effective development, when required. Still, regardless of the option taken, the tax administration must have an ICT unit that is robust enough to (i) provide continuity and sustainability to the technological solutions, and (ii) avoid falling as a client captive of some external company.

It is of paramount importance for the tax administration to clearly understand that technology can provide better tools, but even a very good tool is not a complete solution on its own. A good solution involves careful consideration, design, and evaluation. It must start with a clear definition of the problem and the mechanics required to measure the gained efficiencies objectively and quantifiably.

Innovation is not obtained by purchasing the latest digital technology. Instead, innovation must become an integral part of the organization's culture. To this end, it must become a permanent goal aimed at the taxpayers´ (client) needs—aligning them to institutional priorities, improving and refining existing processes to simplify and facilitate compliance, and constantly evaluating and responsibly adopting new technological advancements to enhance the institution's level of maturity and functionality.

Table 4 summarizes the main functional features and level-specific practices that characterize the different maturity levels for information technology. These were derived from the DIAMOND IT assessment modules that evaluate core ICT governance functions and infrastructure investment needs.

> > >

TABLE 4 - Maturity-Level-Specific Information Technology Practices

Maturity level	Information technology practices
Level 1 **Initial**	• There is no official IT department in the institution; the little available maturity is scattered and dependent on individual knowledge or skills • For the most part, there are no standard or properly documented procedures in place • Work is mostly conducted manually • Technology is not available, or it is available at a limited scale • There are no personnel performance measurement mechanisms established • A lot of the information obtained or produced by the administration is still on paper • Signatures and other similar approval mechanisms are still done by hand • Adequate training in IT is not available for the personnel • There are no formal mechanisms of information exchange between the areas inside the institution nor with external entities • Innovation and technology related topics are nonexistent in the institutional strategy • There is no formal procedure to identify functionality gaps and how to address them • Technological infrastructure is almost nonexistent, or presents severe deficiencies • Procurement of tools and software is done following subjective by-boss authorization without a comprehensive evaluation process • There are no mechanisms to collect and analyze the taxpayer satisfaction level or feedback • Although digital systems are available, the institution has no performance monitoring and reporting tools in place • Transparency-related efforts are nonexistent
Level 2 **Basic** Practices for controlled operations	• There is an official IT department in the institution but is seen merely as a support group • The IT area has no strategic or long-term vision; it dedicates most of its resources to solving day-to-day problems as they arise • The personnel receive sporadic training, but it is not necessarily designed to address specific institutional needs • Investment in infrastructure is improving but is not done following a detailed needs analysis • There are systems in place for most of the more important areas in the institution but very few present a full coverage of at least the core functions • The institution implements off-the-shelf tools but there are no measures in place to avoid provider-lock-in and ensure sustainability or proper integration

Table 4 continued

Maturity level	Information technology practices
Level **2** **Basic** Practices for controlled operations	• Some efforts are being done in software development but there is no formal and effective methodology in place to cover all the stages involved from planning to production • The software procurement and development efforts are not aligned to solve institutional maturity or functionality gaps • Some information being consumed or produced by the institution are digital, but are usually in unstructured formats • There are efforts to digitalize and systematize work in most areas but there are still no standard and properly documented procedures in some of them • Information exchange between areas inside the institution exists but is very limited and often done manually • As several independent systems appear in different areas of the institution, new problems of interoperability arise too • Performance measurement mechanisms exist but are usually limited to monitoring downtime (i.e., they measure the availability but not the actual performance) • There are efforts to collect user feedback but the information gathered is not used in any way to improve the overall IT or institutional strategy • Limited transparency efforts
Level **3** **Intermediate** Practices for efficient operations	• The IT department is well-established inside the organization, usually at the same level of other functional areas and under direct supervision of the highest authority figure • The IT department has a long-term strategic vision that is well-aligned with institutional goals and priorities • Day-to-day problem solving still takes a good share of IT personnel's time and resources, but it is combined with activities for continuous development and improvement • Training is regular and usually prioritized according to institutional needs and individual's career track • Investment in infrastructure is done following a long-term plan based on identified and measured requirements • There are properly documented procedures for all core tasks in the institution's functional areas, with a constant emphasis to digitalize and systematize the work being done • Systems in place present full coverage of, at least, the core functions in the most relevant areas in the institution • When opting to implement off-the-shelf tools, measures are in place to avoid provider-lock-in, and ensure long-term sustainability and proper integration with existing systems • There are personnel dedicated fully to software development, which follow in-house defined methodologies to cover all the stages, from planning to deployment of products • When consuming or producing digital information, structured formats are preferred to facilitate the tasks of further processing and analysis • Digital information exchange between areas inside the institution is partially automated • Some integration with external information sources exists but is usually limited to specific application cases • Interoperability problems still exist but some efforts are being done to integrate the systems used in different areas of the institution, usually in the form of shared data sources. • Performance measurement mechanisms are more sophisticated and are used to monitor some interesting metrics like response time, errors occurring, and unauthorized access attempts. However, there is no clear high-level policy on what to do with the collected information.

Table 4 continued

Maturity level	Information technology practices
Level **3** **Intermediate** Practices for efficient operations	• User feedback and satisfaction level is gathered and understood as a key metric of the system's performance, however, there is a need to link this information with the strategic long-term vision in the organization • Transparency efforts are spread and generalized in the institution; and usually, reports produced from collected data are published on the website
Level **4** **Advanced** Practices for sustainable and optimized operations	• The IT department is an established technostructure supporting the organization. It provides guidelines, standards, and tools for the rest of the business areas • Innovation and technological development are fundamental in the organization's long-term strategic vision, and are translated in appropriate institutional goals and priorities • Day-to-day problem solving takes a minimal share of IT personnel's time and resources. The main priority in the department is continuous development and improvement • Regular training is seen as a hard requirement and is always prioritized according to institutional needs, individuals' career track, and new demands arising from future goals and challenges • Investment in infrastructure is carefully done following a long-term strategy based on the institutional requirements; where applicable, new trends like cloud computing and software-as-a-service solutions are applied to maximize efficiency • There are no paper-based processes in place. All information consumed and generated by the organization is managed in digital formats. • Signatures and other similar approval mechanisms are all generated electronically. No by-hand signatures are required in any process • Systems in place present full coverage of all functions in the most relevant areas in the institution, with a well-defined methodology and calendar of new releases development • Data is mostly managed in digital structured formats that facilitate the tasks of processing and analysis; the results obtained are used to improve the decision-making process in all business areas in the institution • When opting to implement off-the-shelf tools, there is a regular process to search and evaluate open-source solutions available before opting for third-party commercial products. When using commercial alternatives, measures are in place to avoid provider-lock-in and ensure long-term sustainability and adequate integration with existing systems • There is a dedicated software development team, which follows professional and well-established methodologies to cover all the stages, from planning to deployment of products. Even when not developing large systems in-house, this team's expertise is used when evaluating and deploying third-party tools • Procurement (off-the-shelf solutions) and development (in-house solutions) efforts are designed to mitigate or solve institutional maturity or functionality gaps • Digital information exchange between areas inside the institution is completely automated. All areas publish a standard catalog of all available data resources to facilitate discovery and consumption (e.g., following the DCAT open standard) • Integration with external information sources is a continuous effort that is not limited to specific application cases (i.e., the tax administration's goal is to ensure access to as many information sources as possible, allowing the flexibility required to all business areas in order to decide the best ways to integrate them in its functions) • Interoperability problems are avoided by focusing on producing and maintaining solid

Table 4 continued

Maturity level	Information technology practices
Level **4** **Advanced** Practices for sustainable and optimized operations	protocols and standards that can be adopted by all systems used in different areas of the institution. Consistency is ensured by providing proper communication mechanisms rather than using shared data sources that can produce functional and security problems • Performance measurement mechanisms are very sophisticated and used to collect some metrics like response time, errors occurring, and unauthorized access attempts. All the gathered data is used to produce comprehensive reports that improve the decision-making process and allow for better root-cause analysis of problems • User feedback and satisfaction level is gathered and understood as a key metric of institutional performance. This information helps improve the strategic long-term vision of the organization • Transparency efforts are spread and generalized in the institution. Information is published using automated and standards-based systems that facilitate discoverability and consumption by entities both inside and outside the organization • There is a continuous and healthy pilot-based research program that formalizes the allocation of resources to evaluate and test new technological advancements, resulting in either new tools or knowledge for the institution

This maturity model for IT must be put in context to fully understand the implications of progressing from one level to the next. In this respect, the rapid development of ICT in recent decades has allowed tax administrations to administer the tax system more efficiently and to deal with a great number of taxpayers' population. Increasingly cheaper access to this technology has resulted in decreasing administration costs, better data processing, and more accurate and reliable information. On the taxpayer side, a wide array of electronic services has been provided, which has a significant impact on the reduction of tax compliance burden.

Increased use of ICT has also influenced the way tax administrations organize themselves and how core business processes, which include the delivery of taxpayers' services, are managed. In effect, regional and local offices have gone through significant changes; and in many countries, flatter management arrangements in tax administration organizational structure are in place as a result of eliminating intermediate and/or regional layers. In this vein, many tax administrations are centralizing key functions—such as tax returns and tax payments processes or taxpayers' services, tax audit, and some routine processes related to collection enforcement—while local offices are losing relevance. Technology has changed the way of delivering services to taxpayers and demands less interaction between taxpayers and tax administrations.

This new paradigm has a significant impact on HR policies when it comes to allocating staff to different offices and different functions. In effect, changes in the workload of local offices should be reflected in reallocation of personnel. Intensive use of IT for routine processes requires putting in place programs that provide training to support staff who are no longer needed in their current assignments. New technologies demand new staff skills due to dramatic changes in the way tax administration does business nowadays. Oftentimes there are challenges in attracting strong IT skills to public sector, compared with private sector, due to pay scale limits or other factors. Thanks to information technology (IT), tax and customs administrations are now able to manage great amounts of third-party information, enabling them to massively crosscheck this information with the content of tax returns and customs declarations. In addition, IT-based compliance risk management processes result in a better selection of cases for audit, which makes the tax audit function more efficient.

Consequently, the maturity model for the IT area becomes a key element in assessing the IT performance gap and in designing an action plan. This helps build the data science capabilities needed to advance to the next maturity level, which is part of the digital transformation of tax and customs administrations. In the next section, based on the maturity levels in IT and with reference to best practices in IT system implementation, we examine how to build data science capabilities in revenue administrations, focusing on data management and data science tools, the creation of machine learning capabilities and their application, and the feasibility on the use of blockchain initiatives.

Building Data Science Capabilities in Tax and Customs Administrations

Although there is no consensus on its definition, data science is generally accepted as an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. Data science includes techniques from a diverse spectrum, including statistics, big data, data management, data visualization, data mining,[11] and machine learning. Revenue administrations have been using a subset of data sciences for many years, focused on the field of data analytics, to generate valuable insights from the available data and to make better-informed decisions. Data analytics is the process of cleaning, inspecting, modeling, and transforming data for finding valuable information to enhance the decision-making process.

3.1 Data Management Applied to Data Sciences

Data science is typically used in a tax administration and customs (hereafter we will refer to them as revenue administrations) to make sense of the vast amounts of data that is available, so that the organization can become smarter, faster, and more efficient. Of particular interest to revenue administrations is machine learning, a subset of data sciences that can be used to solve difficult problems that arise from the inability of a revenue administration to process massive amounts of data efficiently. In the purview of machine learning are applications such as identifying fraudulent operations, automatically answering questions posed by taxpayers, identifying illegal goods in x-rays, predicting the cost of interventions, identifying the code of a given article in the harmonized system, and identifying potential errors or inconsistencies in declarations or tax returns. Hence, data sciences and machine learning can improve the efficiency of a revenue administration significantly.

Data sciences and machine learning are analytical tools, optimizations, and enhancements that are typically implemented on top of existing basic IT systems in a revenue administration, since these provide the basic data that is needed. For example, an audit module would be required before we attempt to predict the total cost of an audit using machine learning. On the one hand, typical modules that should be completely implemented in a tax administration before attempting machine learning projects include registration, returns processing, arrears management, payment processing, audit management and tracking, taxpayer assistance, and legal affairs tracking and management. On the other hand, typical modules that should be completely implemented in a customs administration before attempting machine learning projects include registration, declaration processing, arrears management, payment processing, audit management and tracking, trade partner assistance, legal affairs tracking and management, inspection, passenger processing, warehousing, and transit control.

Data science in general, particularly machine learning, relies on high-quality data and professional analysts who use tools and statistical methods for data cleaning (e.g., data acquisition, data manipulation, data wrangling and tidying, managing missing data, eliminating outliers). The performance of the machine learning algorithm will be directly proportional to the quality of the data. For instance, in order to train a computer to recognize fraudulent requests (tax returns or import declarations), we could feed the computer a series of examples that are not fraudulent and the computer will build a mathematical model of what “normality” looks like. This will allow us to compute a probability that a new request is fraudulent, based on how much it deviates from the training data.

If the training data is reliable, the machine learning system will perform well. However, suppose that the training data is of low quality and many examples are misclassified, the

computer will develop a skewed model or "normality" in the tax declarations or returns, resulting in a very large number of false negatives—cases that are fraudulent but are in fact correct. Rather than helping the administration, the algorithm will force us to spend a lot of resources to deal with taxpayers or trade partners that are frustrated because they have been falsely targeted. If the data that we use to train the algorithm is faulty, then the machine learning will build an incorrect model of the correlation between all these features and the expected level of income for any individual taxpayer, and the algorithm will incorrectly classify the taxpayer in a higher tax income bracket.

As concisely put by Christine Robson, a product manager for Google's internal machine learning efforts, "The biggest problem with any machine learning model always lies in the data" (Le 2016). What she was trying to project is that once you know about machine learning models, it is not too difficult to implement them, but it is hard to own a good dataset to be trained at the onset.

Data is of such critical importance that several countries have even started national initiatives for creating high-quality repositories, such as the Data Mining Pipeline (Canada), the Data Lake Project (France), and the Unified Data Platform (Singapore). Hence, one of the most important components of a data science team is a data management team that can create a curated repository of data, which can be used for machine learning projects. This team should use statistical methods and special algorithms to analyze existing data, delete features that have errors, deal with missing data, eliminate outliers, label data, and in general, clean up data that will be used for machine learning algorithms.

Synthetic Data for Machine Learning

Because of privacy and confidentiality constraints, it is usually not feasible to release production data to machine learning developers for the construction of learning algorithms. Information in a tax or customs administration is highly confidential. The development of learning algorithms is an iterative approach that requires experimentation and refinement in multiple iterations, experimenting with multiple parameters and configurations until the performance of the algorithms is acceptable.

Most of machine learning work should be done by a diverse set of developers, including researchers at universities, consultants, and machine learning specialists in other administrations who normally would not have access to the original privileged information.

Since the data in a tax or customs administration is highly confidential—and sometimes it is even difficult to disclose information to be used within an administration—researchers and the IT team frequently find it difficult to experiment with machine learning. To allow these groups to make progress, the data management group needs to create a parallel data repository that contains realistic data that retains the statistical properties of the original repository but has no real information. This parallel data repository is traditionally created by anonymizing the data and replacing it with random values so that it is not possible to infer real information, but it is possible to train machine learning algorithms on it. The process of anonymizing the data repository is time consuming and difficult to scale up, and can render the information useless. Sometimes it is easier to generate new random data that retains the statistical properties of the original data set rather that anonymizing the existing data. This data is formally called "synthetic data."

Advances in machine learning enable a generation of highly realistic and highly representative synthetic data repositories that resemble the characteristics as well as the diversity of actors in a tax or customs administration. Generating synthetic data boils down to learning the joint probability distribution in an original dataset to generate a new dataset with the same distribution; but the more complex the dataset, the more difficult it is to map dependencies.

Creating a synthetic data repository is a large challenge that requires full-time staff trained on the mechanisms for anonymizing data and replicating it to the synthetic repository so that it retains its statistical properties but is not connected in any way to real entities or data. The rewards of this effort are equally large, since the synthetic data can be made available to a large group of researchers who will help develop the algorithms that will improve the effectiveness of the revenue administration.

3.2 Incremental Strategy for Creating Machine Learning Capabilities

Machine learning is a branch of data sciences which focuses on the use of data and algorithms to imitate the way that humans learn, deriving structure and rules from data, gradually improving its accuracy over time. Machine learning algorithms build a model based on sample data, known as "training

data," to make predictions or decisions without being explicitly programmed to do so. Most revenue administrations are using (or are in the implementation phase of) data sciences and machine learning. As this technology matures, they will see increased efficiencies given that the technologies can be used to:

(i) correct deficiencies in the revenue administrations, such as unreliability, slowness, and inaccuracies;
(ii) detect irregularities and errors;
(iii) predict fraudulent behavior;
(iv) proactively assist taxpayers;
(v) classify level of trust for compliance purposes; and
(vi) provide insights for data-driven decision-making.

In general, there are two types of machine learning problems: supervised and unsupervised.

Supervised machine learning problems are those where you teach a computer by example so it can construct a model of the problem and apply it to new cases in the future. For example, we can show the machine learning algorithm examples of tax returns (or customs declarations) that have issues and examples that do not have issues. The system will learn to flag new samples so that taxpayers or trade partners can be alerted automatically before they are accepted. Supervised learning requires data that is labeled. Labeled data is a group of samples that have been tagged with one or more labels. The process of labeling typically takes a set of unlabeled data and augments each piece of it with informative tags. For example, a data label might indicate whether a tax return is fraudulent, whether a customs declaration is high-risk, whether a tax appeal is rejected, and whether a cargo shipment contains contraband.

Unsupervised learning is where the computer discovers patterns and structure in the data without guidance; with this, we can discover new characteristics that we may have not known about. For example, we can show the computer a list of taxpayers or trade partners and it can group them by similarities so that we can analyze those that are grouped together. Hence, unsupervised learning does not require labeling data and can create patterns and structure from standard data warehouses.

Machine learning techniques include, but are not limited to the following:

1. **Classification.** Classification algorithms can explain or predict a class value. For example, classification algorithms can help predict whether a taxpayer should be audited, or whether a container should be inspected. The two classes in this case are "yes" and "no." Classification algorithms are not limited to two classes and can be used to classify items into many categories.

2. **Regression.** Regression methods are used for training supervised machine learning. The goal of regression techniques is typically to explain or predict a specific numerical value while using a previous data set. For example, regression methods can take historical data to predict the income for a new taxpayer that has similar characteristics to other known taxpayers or predict the import volumes for a new trade partner that has similar characteristics to other known trade partners.

3. **Clustering.** Clustering algorithms are unsupervised learning methods that group data points according to similar or shared characteristics. Grouping or clustering techniques are particularly useful in an administration that needs to segment its taxpayers or trade partners by different characteristics to better target risk management or audit programs. Clustering is also effective in discovering patterns in complex data sets that may not be obvious to the human eye, such as discovering groups of taxpayers or trade partners that are linked or related.

4. **Decision trees.** Decision tree algorithms learn to classify objects by answering questions about attributes located at nodal points. Depending on the answer, one of the branches is selected, and at the next junction, another question is posed, until the algorithm reaches a tree's leaf, which indicates the final answer. A typical example of decision trees is identifying the action to take once a clearance request is received. The decision tree can define a complex map of criteria such as location, type of cargo, risk level of the trade partner, history of the trade partner, and amount of the commercial transaction, and determine risk categories based on the request submitted. The system can then evaluate new clearance requests, categorize them by risk, and decide the appropriate action to take.

5. **Neural networks.** Neural networks mimic the structure of the brain and use artificial neurons that connect to several other neurons, and together create a complex cognitive structure in a multilayer structure. The neural network typically learns about how to solve a problem or classify an object by trying different configuration of the connections between the neurons. Neural networks are used for a wide variety of business applications, such as recognizing a taxpayer or a trade partner that is likely to be defrauding the administration, or recognizing a question posed by a

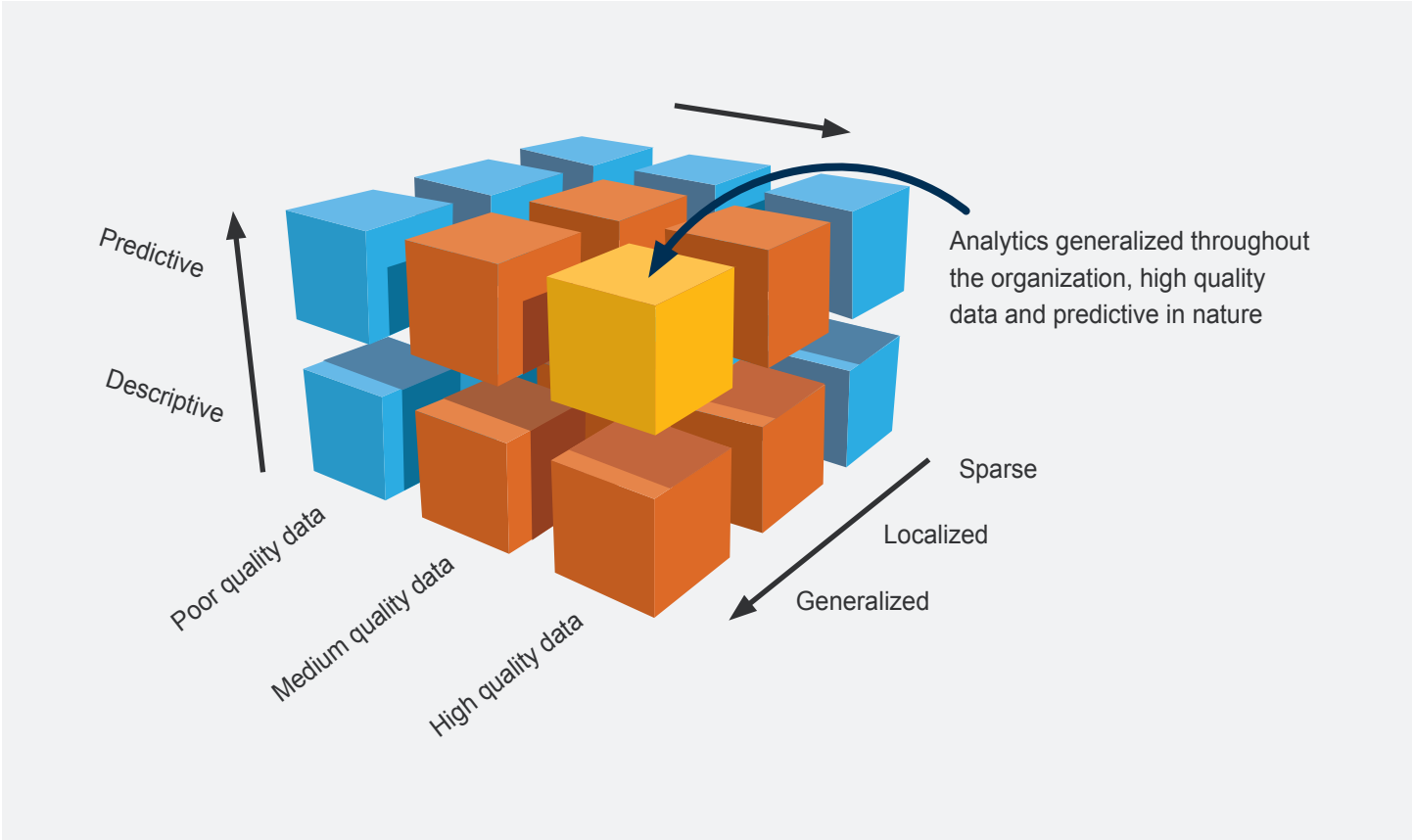
taxpayer or a trade partner to offer a known response from a database of previously answered questions.

6. **Dimensionality reduction.** Dimensionality reduction is used to remove the least important information (sometimes redundant columns) from a data set. In practice, the administration uses it to simplify data sets with hundreds or even thousands of columns (also called features), from the vast amount of information collected about a taxpayer or a trade partner, to data sets that are manageable.

A tax administration's level of maturity in terms of advanced data analytics and machine learning can be measured in three dimensions: scope, data quality, and type of use (see figure 3).

> > >

FIGURE 3 - Dimensions of Maturity Levels in Terms of Advanced Data Analytics and Machine Learning



Scope refers to who is using advanced data analytics in an organization. At the most basic level, use is sparse, and only a few individuals, if any, use advanced data analytics in isolated pockets. At an intermediate level, localized groups use advanced data analytics in specific areas of a tax administration. In the more advanced level, the use of advanced data analytics is generalized.

Data quality refers to the quality of the data in the tax administration´s data repositories (i.e., Data Warehouse). Advanced data analytics is likely to produce poor results when the data available is of low quality, regardless of the talent of the data analysts. High quality data enables the data analytics team to obtain exemplary results.

Type of use indicates whether the tax administration is using advanced data analytics to generate reports and statistics about what happened in the past (descriptive) or whether it is using data to change the way it operates in the future (predictive).

In general, the level of maturity of a tax administration, in terms of data analytics, is largely defined by the level of quality of the data that is available (see table 5). The lowest levels of maturity correspond to low-quality data that is used to describe what happened in the past, and the highest levels of maturity correspond to high-quality data that is used to predict the future behavior of taxpayers and the organization.

> > >

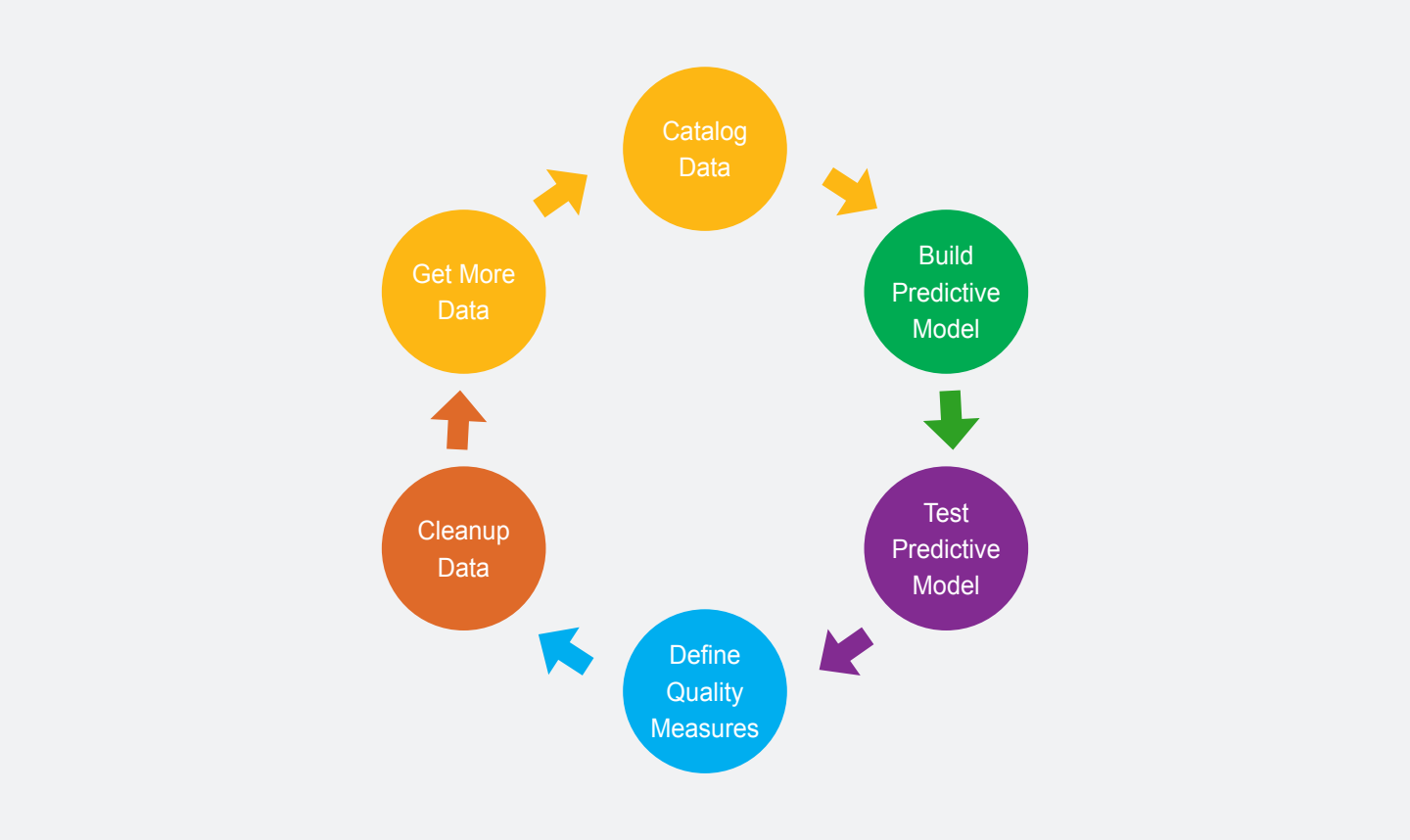
TABLE 5 - Maturity Levels of a Tax Administration in Terms of Data Analytics

Maturity level	Levels of use	Scope	Data quality
Level 1 Initial Practices for sustainable and optimized operations	Descriptive	Sparse	Low Quality
	Descriptive	Generalized	Low Quality
	Descriptive	Localized	Low Quality
	Descriptive	Sparse	Medium quality
	Descriptive	Localized	Medium quality
	Descriptive	Generalized	Medium quality
Level 2 Basic Practices for sustainable and optimized operations	Descriptive	Sparse	High Quality
	Descriptive	Localized	High Quality
	Descriptive	Generalized	High Quality
	Predictive	Sparse	Low Quality
	Predictive	Localized	Low Quality
	Predictive	Generalized	Low Quality
Level 3 Intermediate Practices for sustainable and optimized operations	Predictive	Sparse	Medium quality
	Predictive	Localized	Medium quality
	Predictive	Generalized	Medium quality
Level 4 Advanced Practices for sustainable and optimized operations	Predictive	Sparse	High Quality
	Predictive	Localized	High Quality
	Predictive	Generalized	High Quality

The current level of data analytics maturity in most tax administrations is likely to be low, and unfortunately, improvement must come as a result of an iterative process, where the tax administration attempts to build a predictive model, and as a result triggers a data cleansing process that will allow further improvement iteratively (see figure 4).

> > >

FIGURE 4 - Iterative Data Cleansing Process



This iterative process also builds talent, as it constitutes a competency framework. As the tax administration completes multiple cycles of this process, it builds a workforce equipped with next-generation skills and advanced tactics in artificial intelligence, resulting in qualified leadership at national and regional levels.

In most tax administrations, unfortunately, there is a very large disconnect between the functional areas and the IT department that locks the tax administration in the localized scope. The IT department has sufficient knowledge to build basic machine learning algorithms but has no data and no functional demands from the business areas to attempt building a predictive model.

The business side of the tax administration is usually completely unaware that the IT department has basic skills to attempt building these predictive models. As a result, it is unlikely that the tax administration will see concrete examples of machine learning in the short term.

To jumpstart this process, the tax administration needs to create specific demands from the business side so that the IT department tries applying machine learning algorithms, with whatever data they currently have. The organization needs to engage senior leaders to work with business process managers so that full executive support can be established.

Initially, most machine learning algorithms will have poor performance but will gradually improve. The most successful data science initiatives take small incremental steps rather than pursuing a large and ambitious project. Small, incremental steps help break down skepticism, prove the concept with limited investment, and build trust for wider-scale adoption.

Although every tax administration's eventual goal is automatic fraud detection and intelligent risk management, it needs to identify and promote smaller, less demanding applications where it can make significant progress and gradually strengthen its competency and improve on its level of maturity.

It is important to understand that big, magical results are not always possible, as was seen in the Norwegian Tax Administration, where they attempted to create unsupervised machine learning systems to automatically find errors on tax returns and were disappointed with the results.

Rather, very simple systems with modest goals are likely to pay-off handsomely, as illustrated by Zambrano and de Sarralde (2021) on the use of artificial intelligence to determine the economic activity of a new company. Zambrano and de Sarralde (2021) state, "In tax administration, we know that the correct identification of the economic activity of the taxpayer is important. It can be decisive in risk management, it can have implications in tax compliance, it can open spaces to specific benefits associated with an activity." A system that correctly identifies the economic activity of a taxpayer can save months of manual labor and increase its risk management capabilities. Modest goals indeed add up.

3.3 Applications of Machine Learning in Tax and Customs Administrations

Tables 6 and 7 show some examples of the use of machine learning in tax administrations and customs, respectively. Those that are identified with a low level of difficulty are modest applications that can be used to obtain executive support to drive these efforts from the business side of the organization. The general strategy is to encourage the business side of the customs administration to implement machine learning algorithms with low levels of difficulty first, and progress incrementally until advanced competencies are developed; the customs administration can then tackle medium and high complexity issues. Each level of difficulty is associated to a level of maturity according to the DIAMOND's four-level maturity model. Finally, table 8 shows many areas that can effectively use machine learning in the back office to make the revenue administrations more efficient and effective.

> > >

TABLE 6 - Application of Machine Learning to Tax Administrations

Level of difficulty/ Maturity level	Examples of application of machine learning to tax administrations
Low Difficulty/ Level **2** **Basic** Practices for controlled operations	• Assigning a probable economic activity to unclassified taxpayers [1 – Synthetic Data]. The tax administration regularly classifies taxpayers according to their economic activity. This classification is used internally for risk management, outreach, and communication with taxpayers. Some taxpayers may have been misclassified or their classification may have changed since the original registration, and it is possible to deduce the economic activity from financial information using specialized machine learning segmentation models. These can be used to identify taxpayers that may have changed their economic activities or to classify taxpayers in the registry that are lacking economic activity. This application is particularly attractive because it does not require labeling of data since the existing classification for taxpayers is used to train the machine learning algorithms. • Reviewing the reasonableness of the expenses deducted in the income tax return. Machine learning algorithms can be used to predict the type and amount of expenses that can be detected in an income tax return. This application is useful in assisting taxpayers during the filling phase of the tax return or in assessing the reasonableness of the expenses deducted in income tax return. This application is particularly attractive because it does not require labeling of data since existing tax returns are used to train machine learning algorithms as to what is reasonable in terms of expense deductions.

Table 6 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to tax administrations
Low Difficulty/ Level **2** **Basic** Practices for controlled operations	• Automatically classifying documents during an audit [i.e., Canada]. During an audit, the review team must process thousands of documents and must find documents of interest (for example, documents that mention exports or imports, or documents that mention financial transactions). Document classification is an area where machine learning can improve overall quality while simultaneously reducing costs. Document classification works through a two-step process, where first a textual representation of the document is created by using optical character recognition (OCR) and the output of this process then feeds into another machine learning model that reads the text to determine the context and applies a label to the document that is relevant to the business. • Using virtual agents to reply to taxpayers' questions [i.e., China, Australia and more than 10 additional countries]. Machine learning can be used to implement chatbots that allow reducing the size of the taxpayer assistance workforce and allow taxpayer service to be available 24x7x365, which can be a difficult feat to achieve with purely human help desk operators. Chatbots can be either voice-based or text-based, where the former involves a taxpayer interacting with the chatbot over the phone and the latter has the taxpayer interact through the tax administration's website. These services allow commonly asked questions to be instantly answered as well as pre-screen taxpayers so agents can be more targeted on their calls. The implementation of chatbots is already mature and can be implemented with minimal data. • Estimating cost of supervision activities (e.g., audits and desk reviews). Supervision activities are expensive and past experience can be used to train machine learning algorithms so that the cost of future supervision activities can be estimated from the taxpayer profile and external data available. This cost can be used as a parameter in the prioritization of supervision activities, combining it with additional data to select those activities that will be most cost-effective. • Estimating probability of success in supervision activities (e.g., audits and desk reviews). Supervision activities are expensive and past experience can be used to train machine learning algorithms so that the probability of success of future supervision activities can be estimated from past experiences, the taxpayer profile, and external data available. This probability can be used as a parameter in the prioritization of supervision activities, combining it with additional data to select those activities that will be most cost-effective. • Answering taxpayer questions [i.e., more than 10 countries]. A portion of the questions asked by taxpayers in the help desk are simple questions that can be answered with intelligent agents (e.g., "I drive for Uber — is my wait time tax deductible," or simply "I drive for Uber, do I need to pay taxes"). This frees considerable resources and allows providing authoritative answers. • Sentiment analysis in taxpayer communication [i.e., Canada]. Sentiment analysis is a branch of natural language processing that allows identifying the sentiment (positive, neutral, or negative) of text and can be used in taxpayer feedback (e.g., complaints and suggestions) to route it to the appropriate division for analysis.

Table 6 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to tax administrations
Low Difficulty/ Level **2** **Basic** Practices for controlled operations	• Estimating the probability that the tax administration can recover arrears. Using information from past arrear collection efforts, machine learning algorithms can be trained to estimate the probability that the tax administration can recover arrears, allowing the tax administration to schedule its arrear collection efforts so that more difficult cases are assigned residuary resources, or assigned to more experienced teams. • Estimating the cost of collection for arrears. Using information from past arrear collection efforts, and if the tax administration has recorded the cost of collection for previous collection efforts, machine learning algorithms can be trained to estimate the cost of collection. This allows the tax administration to schedule its arrear collection efforts so that the costliest cases are assigned residuary resources, or assigned to more experienced teams. • Model for estimating the value of real estate [i.e., Norway, Brazil]. Create a model of the valuation of the market value of real estate to update the value of real estate and determine whether taxpayers are underdeclaring their assets.
Medium Difficulty/ Level **3** **Intermediate** Practices for efficient operations	• Detecting residents who have emigrated from the country without notifying the tax administration and the central government [i.e., Norway]. • Data entry of tax forms. Tax administrations frequently need to digitize tax forms that are presented in paper, and poor data capture results in low data quality, with considerable impact on the tax administration´s bottom line. Machine learning can be used to improve this process by automating portions of the data entry workflow to ensure the critical details are captured. This can also increase the data entry accuracy while simultaneously making the whole process quicker. • Generating virtual proposals for tax deduction [i.e., Norway]. Machine learning algorithms can be trained to predict or estimate the individual taxpayer's level of income and its composition, as well as their level of debts and family situation, and estimate the origin of the legal deductions in the annual income tax return. These estimations can be used to cross-reference income tax returns, pre-fill tax returns, or simply make recommendations to taxpayers during data entry. • Predicting who is entitled to deductions and establishing the amount of the deductions. Machine learning algorithms can be trained to predict or estimate the types and amounts of deductions that can be declared by a taxpayer. These estimations can be used to cross-reference income tax returns, pre-fill tax returns, or simply make recommendations to taxpayers during data entry. • Automatic risk profiling of taxpayers [i.e., Australia, Spain]. If the tax administration has a reliable set of curated risk profiles, machine learning algorithms can be trained to assign risk profiles to new taxpayers, even when the systems have never seen them. This process allows assigning risk profiles to many taxpayers and can be trained to adapt to new patterns, reclassifying taxpayers as necessary. • Discerning complex, multilayered relationships between taxpayers. Understanding the relationships between taxpayers and how they are related can substantially improve the risk management system of a tax administration. Machine learning algorithms in combination with

Table 6 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to tax administrations
Medium Difficulty/ Level **3** **Intermediate** Practices for efficient operations	graph databases can cluster taxpayers in multidimensional risk groups that can enhance our understanding on their relationship, and how this relationship influences their behavior. • Determining deviation from declared income/assets and predicted income/assets from internal and external data. Internal financial information can be grouped with external data to create a comprehensive 360-degree view of the taxpayer, and this fish-eye view can be fed to machine learning algorithms to predict the income and assets of a taxpayer. These predictions can be used to pre-fill tax-suggested tax statements, identify potential cases for review or audit, or identify taxpayers that have increased their income/assets but have not adjusted their statements. • Selecting the composition of audit team to match taxpayer profile. The composition of an audit team can be altered to fit the characteristics of the taxpayer. Machine learning algorithms can be trained to recognize the composition that is most likely to succeed during an audit, based on the profile of each of the audit team members and the profile of the taxpayer. • Real-time checking of tax returns [i.e., Norway]. Machine learning algorithms can be used to check on the reasonableness of a tax return, particularly when the information of the tax return is cross-referenced with the 360-degree view of the taxpayer, including external data. • Detecting singular outliers in tax returns and refund requests [i.e., Serbia]. Single outliers, compared with multiple outliers, tend to be better predictors of errors in tax returns and refund requests. On the other hand, multiple outliers tend to represent normal shifts in taxpayers' behaviors. Machine learning algorithms are good at detecting single outliers and can be used effectively as part of the arsenal for filtering incorrect declarations. • Selecting tax returns for inspection and review [i.e., India]. Tax returns can be examined to determine the probability that the tax return has errors, considering the history of the taxpayer and the historical reviews performed by the tax administration. Taxpayers can be given a chance to review their assessments before they are reviewed by the tax administration. • Classifying taxpayers into risk groups (or calculating risk scores) [i.e., Serbia, Brazil]. If the tax administration has a sufficiently large set of previous examples of risk segmentation, machine learning algorithms can be trained to replicate this classification to new taxpayers. A small group of curated classifications can be effectively used to scale up the classification effort into large segments of taxpayers, if the data repository is of high quality. • Determining whether a taxpayer is making inconsistent tax operations (in terms of its history and/or its class) [i.e., Australia]. • Segmenting taxpayers according to the probability of non-compliance [i.e., Spain]. • Processing taxpayers' allegations and proposing most likely response [i.e., Brazil]. Natural language models can be used to automatically read the taxpayers' allegations, compare them with a knowledge base of previous resolutions, cluster similar allegations, and propose in natural language the most likely outcome.

Table 6 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to tax administrations
High Difficulty/ Level **4** **Advanced** Practices for sustainable and optimized operations	• Predicting inaccurate tax returns issued by self-employed and sole proprietorships. • Identifying anomalies in taxpayers´ accounts during an audit. • Identifying cases with characteristics that could indicate potential fraud. • Identifying fake invoice [i.e., Mexico, India]. • Identifying taxpayers that committed a tax crime [i.e., Brazil]. Identify taxpayers that omit information, make false declarations, defraud fiscal documents, create false or inexact documents, and deny providing documents or invoices to fiscal authorities. • Estimating the tax gap, especially for specific segments and sectors. • Detecting enterprises that carry out simulated operations in VAT (enterprises that generate simulated operations and/or enterprises deducting simulated operations in VAT) [i.e., India, Mexico]. • Detecting tax fraud for underreporting declarations [i.e., Colombia]. • Determining violation of transfer pricing guidelines. • Calculating the probability concerning an individual taxpayer's propensity to attempt to evade taxes [i.e., Spain, Brazil].

> > >

TABLE 7 - Application of Machine Learning to Customs Administrations

Level of difficulty/ Maturity level	Examples of application of machine learning to customs administrations
Low Difficulty/ Level **2** **Basic** Practices for controlled operations	• Assigning a probable economic activity to unclassified trade partners. The customs administration regularly classifies trade partners according to their economic activity; and this classification is used internally for risk management, outreach, and communication with trade partners. Some trade partners may have not been classified or their classification may have changed since the original registration; and it is possible to deduce the economic activity from financial information using specialized machine learning segmentation models. These can be used to identify trade partners that may have changed their economic activities or to classify trade partners in the registry that lack economic activity. This application is particularly attractive because it does not require labeling of data since the existing classification for trade partners is used to train the machine learning algorithms. • Identifying declarations with incorrect country of origin. Machine learning models can accurately determine whether the stated country of origin is likely to be correct or not based on the country's history of importations of such types of goods. If the country of origin is

Table 7 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to customs administrations
Low Difficulty/ Level 2 **Basic** Practices for controlled operations	considered unusual, the system can list the top five most likely countries of origin, with a probability of each one being correct. This can be used while preparing a declaration and while examining cargo during inspections. • Automatically classifying documents during an audit. During an audit, the review team must process thousands of documents and must find documents of interest (for example, documents that mention exports or imports, or documents that mention financial transactions). Document classification is an area where machine learning can improve overall quality while simultaneously reducing costs. Document classification works through a two-step process: (i) a textual representation of the document is created by using Optical Character Recognition (OCR), and (ii) the output of this process then feeds into another machine learning model that reads the text to determine the context and applies a label to the document that is relevant to the business. • Using virtual agents to reply to trade partner questions. Machine learning can be used to implement chatbots that allow reducing the size of the trade partner assistance workforce and allow trade partner service to be available 24x7x365, which can be a difficult feat to achieve with purely human help desk operators. Chatbots can be either voice-based or text-based, where the former involves a trade partner interacting with the chatbot over the phone and the latter has the trade partner interact through the customs administration's website. These services allow commonly asked questions to be instantly answered as well as pre-screen trade partners so agents can be more targeted on their calls. The implementation of chatbots is already mature and can be implemented with minimal data. • Estimating cost of supervision activities (e.g., audits, desk reviews). Supervision activities are expensive and past experience can be used to train machine learning algorithms so that the cost of future supervision activities can be estimated from the trade partner profile and external data available. This cost can be used as a parameter in the prioritization of supervision activities, combining it with additional data to select those activities that will be most cost-effective. • Estimating probability of success in supervision activities (e.g., audits, desk reviews). Supervision activities are expensive and past experience can be used to train machine learning algorithms so that the probability of success of future supervision activities can be estimated from past experiences, the trade partner profile, and external data available. This probability can be used as a parameter in the prioritization of supervision activities, combining it with additional data to select those activities that will be most cost-effective. • Answering trade partner questions. A portion of the questions asked by trade partners in the help desk are simple questions that can be answered with intelligent agents (e.g., "I am thinking of importing iPhones—what is the duties I need to pay?"). This frees considerable resources and allows providing authoritative answers. • Sentiment analysis in trade partner communication. Sentiment analysis is a branch of natural language processing that allows identifying the sentiment (positive, neutral, or negative) of text and can be used in trade partner feedback (e.g., complaints and suggestions) to route it to the appropriate division for analysis.

Table 7 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to customs administrations
Low Difficulty/ Level **2** **Basic** Practices for controlled operations	• Estimating the probability that the customs administration can recover arrears. Using information from past arrear collection efforts, machine learning algorithms can be trained to estimate the probability that the customs administration can recover arrears, allowing the customs administration to schedule its arrear collection efforts so that more difficult cases are assigned residuary resources, or assigned to more experienced teams. • Estimating the cost of collection for arrears. Using information from past arrear collection efforts, and if the customs administration has recorded the cost of collection for previous collection efforts, machine learning algorithms can be trained to estimate the cost of collection, allowing the customs administration to schedule its arrear collection efforts so that the costliest cases are assigned residuary resources, or assigned to more experienced teams.
Medium Difficulty/ Level **3** **Intermediate** Practices for efficient operations	• Image analysis of maritime containers to improve efficiency of cargo inspections. Image analysis is one of the most developed screening technologies that can be effectively used to identify features in x-ray images for cargo with accuracies that exceed that of human reviewers. Deep convolutional neural networks can be used effectively to identify anomalies in cargo. • Image analysis of passenger baggage. Image analysis is one of the most developed screening technologies that can be effectively used to identify features in x-ray images with accuracies that exceed that of human reviewers. Deep convolutional neural networks can be used effectively to accurately identify forbidden goods in passenger baggage. • Harmonized System goods classification [i.e., multiple countries]. Neural network can be used to accurately and efficiently classify the products according to the Harmonized System based on a given description, and traders are helped because it can classify goods accurately, save time, and reduce costs. The customs administration also benefits from faster clearance and approval, better compliance from the trading community, and better risk assessment with the corresponding reduction and prevention of fraud. • Recommendations for selectivity [i.e., Brazil]. A combination of different machine learning models can be effectively used to recommend potential verifications, considering the history of the trade partner and the characteristics of the operation, helping the customs officer responsible for making those decisions. • Highlighting potential mistakes in declarations. It is common for traders or brokers to make mistakes when entering values, weights, and measures on a customs declaration. Machine learning models can identify that a declaration has been populated with incorrect information and/or an ambiguous or misleading goods description as well as flag the declaration for document review and possible amendment, listing the item(s) in question and the content that it believes have been mistyped. • Data entry of customs forms. Customs administrations frequently need to digitize forms that are presented in paper. Poor data capture results in low data quality, with considerable impact on the customs administration´s bottom line. Machine learning can be used to improve this process by automating portions of the data entry workflow to ensure that the critical details

Table 7 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to customs administrations
Medium Difficulty/ Level **3** **Intermediate** Practices for efficient operations	are captured. This can also increase the data entry accuracy while simultaneously making the whole process quicker. • Automatic risk profiling of trade partners. If the customs administration has a reliable set of curated risk profiles, machine learning algorithms can be trained to assign risk profiles to new trade partners, even when the systems have never seen them. This process allows assigning risk profiles to many trade partners and can be trained to adapt to new patterns, reclassifying trade partners as necessary. • Discerning complex, multilayered relationships between trade partners. Understanding the relationships between trade partners, and how they are related, can substantially improve the risk management system of a customs administration. Machine learning algorithms in combination with graph databases can cluster trade partners in multidimensional risk groups that can enhance our understanding on their relationship, and how this relationship influences their behavior. • Selecting the composition of audit team to match trade partner profile. The composition of an audit team can be altered to fit the characteristics of the trade partner. Machine learning algorithms can be trained to recognize the composition that is most likely to succeed during an audit, based on the profile of each of the audit team members and the profile of the trade partner. • Real-time checking of customs declarations. Machine learning algorithms can be used to check on the reasonableness of a customs declaration, particularly when the information of the customs declaration is cross-referenced with the 360-degree view of the trade partner, including external data. • Detecting singular outliers in customs declarations and refund requests. Single outliers, compared with multiple outliers, tend to be better predictors of errors in customs declarations. On the other hand, multiple outliers tend to represent normal shifts in trade partners' behaviors. Machine learning algorithms are good at detecting single outliers and can be used effectively as part of the arsenal for filtering incorrect declarations. • Classifying trade partners into risk groups (or calculating risk scores). If the customs administration has a sufficiently large set of previous examples of risk segmentation, machine learning algorithms can be trained to replicate this classification to new trade partners. A small group of curated classifications can be effectively used to scale up the classification effort into large segments of trade partners, if the data repository is of high quality. • Predicting the Cost, Insurance and Freight (CIF) value of each declared item, which is the actual value of the goods when they are shipped, and on which duties are calculated. Machine learning models can predict the CIF value of declared items based on the information provided, including on the history of previous importations of such good types. • Automatically identifying commerce transactions involving strategic goods from broader international trade flows.

Table 7 continued

Level of difficulty/ Maturity level	Examples of application of machine learning to customs administrations
High Difficulty/ Level **4** **Advanced** Practices for sustainable and optimized operations	• Automatically spotting false documents such as invoices. • Predicting fraud associated with declarations [i.e., Spain]. Train a neural network so that it identifies any of these factors when importing or exporting products: falsely declaring the origin of the goods, declaring a lower value on the goods, misclassifying the goods, and smuggling goods. • Automating the determination of the valuation of goods.

> > >

TABLE 8 - Additional Applications of Machine Learning

Level of difficulty/ Maturity level	Examples of additional applications of machine learning
Low Difficulty/ Level **2** **Basic** Practices for controlled operations	• Increasing power usage effectiveness (PUE) of the tax administration's data centers. PUE is an indicator that is used to evaluate the energy performance of the data center by calculating the ratio of the energy used as a whole, compared with the energy used by just the IT equipment alone. Lowering the PUE can save the tax administration significant resources and can help create an eco-friendly infrastructure. Machine learning models allow data center administrators to effectively use data from monitors and configuration, layout, and parameters to determine the best configuration to lower its PUE. • Identifying anomalies in data center to help prevent downtime. Data center administrators need to monitor thousands of parameters to identify anomalies that need to be addressed quickly to prevent downtime. Machine learning models can be trained effectively to identify potential problems and notify human operators so they can perform preventive maintenance.
Medium Difficulty/ Level **3** **Intermediate** Practices for efficient operations	• Identifying uneven application of processes or detecting similar treatment of similar tax cases/processes (i.e., process drift). Processes should be applied evenly across the tax administration, under the assumption that similar cases should be treated similarly. If the IT systems measure process times and key performance indicators, these can be fed into machine learning algorithms to detect process drift and is a useful input to internal control to ensure that human resources are trained to apply the process as desired. • Automated review of resumes for hiring and flagging of biased language in job descriptions. Machine learning can be effectively used in the most time-consuming parts of the hiring process in the organization by shortlisting candidates and removing bias from job descriptions. Machine learning algorithms can effectively detect biased language in documents and is used regularly to review thousands of resumes to identify those candidates that have skills that match the requirements specified in the job description.

Table 8 continued

Level of difficulty/ Maturity level	Examples of additional applications of machine learning
High Difficulty/ Level 4 **Advanced** Practices for sustainable and optimized operations	• Automatic processing of suspense accounts. When a payment is received without identification, the payment goes into a suspense account; and a person must sort out which tax obligation the payment corresponds to and determine what to do with any excess or shortfall. By monitoring existing processes and learning to recognize different situations, machine learning significantly increases the number of payments that can be matched automatically.

3.4 Feasibility on the Use of Blockchain Initiatives for Tax Administrations

Blockchain technology (a type of distributed ledger technology) on tax administrations can be implemented for different types of problems. For example, immutable ledgers can be implemented for the taxpayers' current account; certificates of ownership of movable and immovable properties can be treated as non-fungible tokens; and taxpayers' credit and/or debit can be represented as tokens and safely interchanged to offer new commercial possibilities for taxpayers (especially for small and medium enterprises).

On public administrations, unless advised otherwise, it is suggested to go with permissioned or private blockchains. Depending on the values to be interchanged and the number of participants in the network, a public administration can implement a permissioned (private) or permissionless (public) blockchain. This decision is crucial as the operational procedures for the registration of participants, consensus algorithms, and time to confirm transactions can change drastically in either case.

A basic rule of thumb is to consider blockchain technology when multiple parties need to collaborate and exchange information. The business rules associated with the information exchanges will not change frequently and are uniform for participants, and at least one of the following conditions is met:

- Peers don't trust each other's systems and information.
- An objective, immutable log/ledger is needed.

Next, we will explore three possibilities of usage of the blockchain technology. Each one of these examples will include more and more complicated concepts of the technology. The first example is the implementation of the blockchain principles over ledgers that the tax administration wants to keep safe. The second example is the usage of a distributed application to handle the tax credit on invoices. The third example is the usage of non-fungible tokens to handle the property of motor vehicles and use them as base for tax purposes. These examples show, by increasing complexity, the way blockchain technology can be implemented in tax administrations. There are collateral benefits such as security and trust that come with this implementation—benefits which, depending on the context, can be crucial to the success of the project.

Example 1: Transaction Ledger

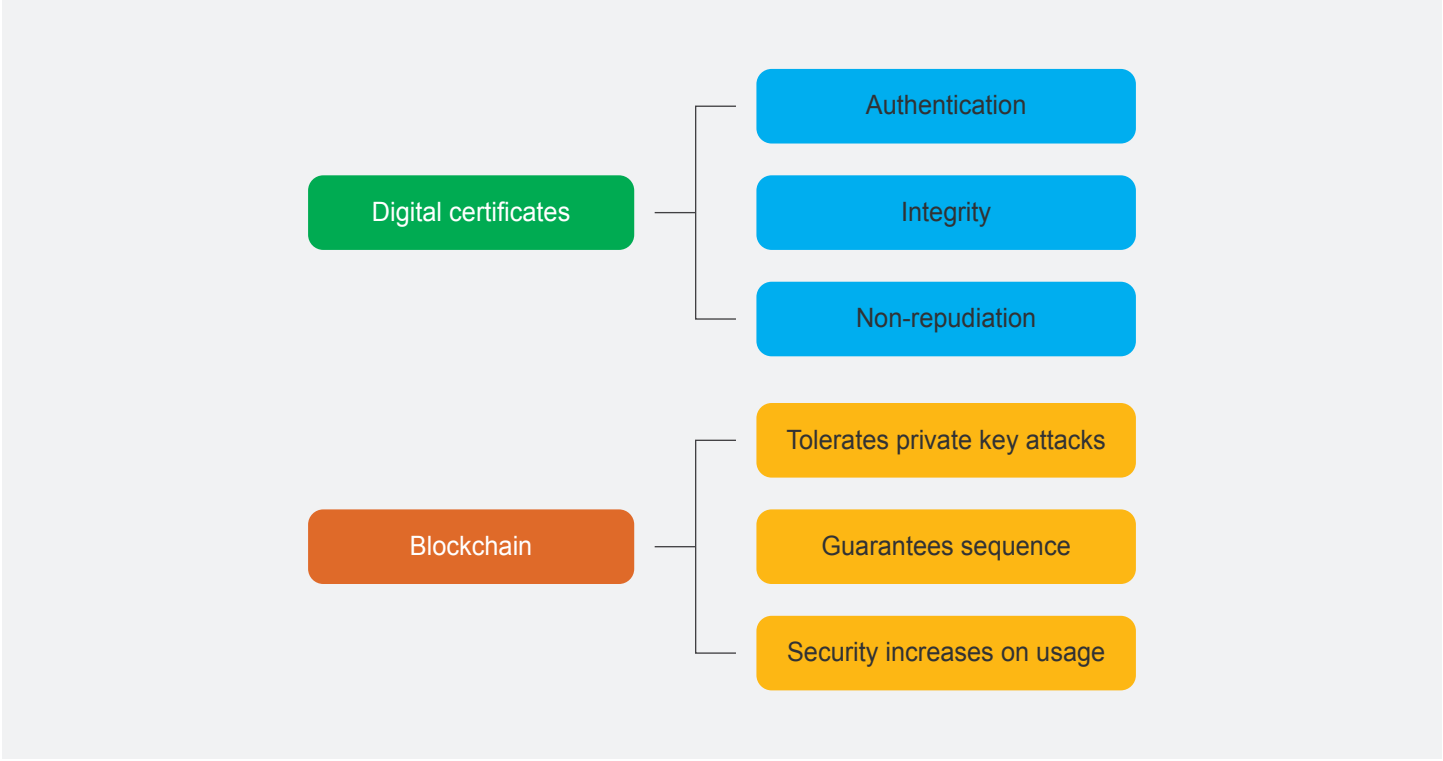
The first example is the application of the initial blockchain idea to ledgers in the system. Starting with a simple structure, which is the taxpayer's current account information, the first thing that needs to be done is to enable the taxpayers to create transactions that are positively associated with them and that have the property of nonrepudiation.

To be able to do this, we need to assign to every taxpayer a public and private key. In practice, every taxpayer is provided with a digital certificate known only to them via a strong password (the private key) and the public key of that certificate that is stored in the structure of the taxpayer. From that moment, every transaction generated by the taxpayer can be digitally signed with the taxpayer certificate, and therefore inherits the security properties of authentication, integrity, and nonrepudiation (see figure 5).

A blockchain can be added to increase the scheme's level of security. In the first level, we get the security properties of a digital certificate involved in the transaction through a digital signature.

> > >

FIGURE 5 - Components and Properties of the Security Scheme



In the second level, other important security properties are acquired. If the transaction is being created with all the needed elements, the blockchain will create the transaction and secure it; this means that if an attacker already has our private key and knows our password then he/she can probably use it to create new transactions. But, with the blockchain level of security, the attacker will not be able to change or delete any transaction that has already happened. The attacker cannot change either the sequence in which the transactions have occurred. Finally, the older the transaction is, the harder it is to change the blockchain and replace it with other information that is valid.

Example 2: Tax Credit VAT

A VAT is applied to any invoice and the tax administration applies the tax to the seller and a credit to the buyer. To enforce the participation of the taxpayers in this scheme, the tax administration has usually made mandatory the invoices in the transactions. However, as small taxpayers don't see the benefits when they act as buyers, the invoices are usually filled in with blank buyers or incorrect taxpayers' identification. The idea is then to pass the ownership of the credit balance to the taxpayer. This is a very important concept, which means that the taxpayers can potentially benefit from the following operations and properties:

- Cumulate their tax credit to use it to pay taxes whenever they feel comfortable and in the amounts that they feel comfortable.
- No matter how small or big the invoice is, the tax credit for the buyer is never lost. Taxpayers don't have to worry about the tax administration handling their credit; it is safely up-to-date and stored without the participation of the tax administration.
- If the tax administration allows, it should be possible to use this credit for other transactions and its value can be converted in the form of a currency.

The idea is then to implement a Decentralized Application (DApp) with the following characteristics:

- Provide every taxpayer with a digital certificate to create the wallet for the taxpayer credit.
- Deploy sufficient nodes across the country that will support the blockchain using the tax administration offices and their associated public administration offices.

The information to be stored is mainly the balance of the tax credit of the taxpayers (the blockchain should be sufficient to store this information). The information related to the transactions can stay in the infrastructure of the tax administration and there is no need to decentralize it.

Example 3: Motor Vehicle Tax

An essential concept in the blockchain technology is the non-fungible token (NFT). This allows taxpayers to provide proof of ownership of an asset. An NFT is a token that cannot be divided and that, for this purpose, can only be minted once. The blockchain provides a complete history of the owner and the amount included for an ownership change.

Contrary to immovable properties, vehicles are movable properties that cannot be physically divided, and ideally, their ownership is represented by an NFT. Changes in the ownership of a movable property takes place very quickly and it is difficult for the tax administration to stay up to date with this information. A blockchain implementation of this solution could benefit the tax administration with updated information and could give the taxpayers the perception of actual ownership of their property as the blockchain is not owned by one institution. There are other benefits and potential uses to this blockchain:

- The ownership of a vehicle is saved on the blockchain and therefore cannot be forged.
- The complete history of the vehicle owners is also stored in the blockchain; this allows the tax administration to correctly associate the tax to the correct taxpayer.
- Other networks can use this blockchain to associate other information. For example, all information about vehicle maintenance can be attached to the NFT in a secured decentralized storage. This enables the whole country to maintain transparency on all the maintenance work done on specific vehicles.
- The police and traffic administration can use another decentralized storage to store all the events related to a specific vehicle. The current owner of a vehicle would know not only all the maintenance work done but also all the events where the vehicle participated.

The implementation of this DApp is quite similar to the previous example. It is important to note that DApps have recognizable templates which can help public institutions in developing blockchains in a faster and more standardized way. This can help ease the governance of such applications and the information they store.

3.5 Best Practices in Implementing Information Technology Systems

IT systems in tax administrations are often inadequate because many tax administrations fail to implement basic controls and standard international best practices. As a result, the increase in revenue is lower than the investment costs.

To reap the benefits of digitalization and maximize the probability of an increase of revenue higher than the investment costs, the tax administration must guarantee that it is capable of sustainably implementing and managing the IT infrastructure, and to this effect must implement basic improvements in several areas, as detailed below.

Change Management

Digitalization and a sophisticated information technology infrastructure on its own do little for a tax administration that cannot systematically gather, organize, and analyze information. Highly developed information societies have been working on automating their information flows for over 300 years, and current technologies automate evolving information gathering processes. The information culture for modern societies dates back to the time of telegraphs, wherein train operators reported their status via telegrams and managers had them transcribed into tables to determine the position of every passenger and freight train. Since then, information societies have further evolved to the telephone and eventually to the Internet. What has changed is the medium used for transmitting data, enabling more timely and sophisticated statistical analysis.

The successful implementation of computer systems for automating the operation of a revenue organization lies in a culture of gathering, sharing, and analyzing information. With it, an organization can use key performance indicators

to measure its performance, measure the effectiveness of its processes, identify bottlenecks, measure the time it takes to execute diverse critical tasks, manage service level agreements with its providers using numeric and objective levels of service indicators, develop tools to intelligently identify delinquent taxpayers, and in general, have an analytical management orientation where decisions are supported by numbers and statistics.

To effectively create and foster a culture of information, the tax administration must invest in knowledge management as well as change management, since successful revenue administrations depend on managing information effectively. Adopting IT systems is not just about buying computers nor developing or buying the appropriate software. Effective digitalization involves reengineering the entire business model and processes. Tax organizations without a culture of information therefore need to implement a change management program.

Change management is an approach to transitioning individuals, teams, and organizations to a desired future state. Rapid organizational change is profoundly difficult because the structure, culture, and routines of organizations often reflect persistent and difficult-to-remove "imprints" of past periods, which are resistant to radical change even as the current environment of the organization changes rapidly.

Without motivation, very little can be done, and the effort will go nowhere as it is difficult to drive people out of their comfort zones and challenge tradition, particularly for an organization that can live comfortably with low performance. Tax administrations have no competitors, cannot go broke, and are not being pressured to compete effectively, and hence have no real sense of urgency to digitalize. To make things worse, the status quo has little risk since tax collection has been carried out in the same way for decades; and there is even pressure against action because of fear of possible complications with digitalization, whereas nothing bad can happen with traditional time-honored, paper-based processes.

To effectively promote change, tax administrations—through their change management programs—must convince managers and staff that the current status quo is much more dangerous than leaping forth and embracing digitalization and control. The change management group needs to establish a sense of urgency by identifying a series of compelling reasons to pursue digitalization, integrated systems, and intensive information use and analytics, and these reasons must be credible and culturally appropriate. The change management program aims to identify the blocks that impede the tax organization from achieving the vision proposed and creatively remove these obstacles by leveraging on leadership and communication campaigns.

A change management program creates a culture of information within an organization, including the habit and discipline needed to collect and analyze information. With this, a tax administration can minimize the probability of failure in the implementation of its IT systems by establishing essential and necessary administrative structures and correctly implementing basic functionality in the computer systems.

Strategic Thinking

A tax administration needs a clear vision that defines what is expected of IT systems over the short, medium, and long term, so that progress can be compared to objective benchmarks. Hence the tax administration should develop a coherent strategy for implementing its IT infrastructure and the necessary systems and their basic governance structure that will make the IT sustainable, thus decreasing the chances of catastrophic failures.

An ICT strategic plan not only allows the IT department to align its interest with that of the entire organization, but also ensures that the IT systems implemented will fulfill a functional goal. Frequently, tax administrations that do not have adequate ICT strategic plans find themselves at the mercy of technology teams that define the features to be provided rather than those required by the actual business.

Performance Management

It is difficult to determine how the organization is performing with respect to the overall objectives and goals defined without the mechanisms for systematically measuring the performance of ICT operations and the IT department. Without such mechanisms, it is even harder to determine the mistakes being performed by members of the organization and their impact.

To systematically measure the performance of the organization's IT department, the tax administration must develop a series of key performance indicators (KPIs). It is possible to develop KPIs to measure virtually anything. Organizations frequently spend a great deal of time and effort developing a large and sophisticated battery of indicators, only to fail during the data collection phase.

Hence, organizations should carefully choose a smaller set of indicators that help track those aspects of software development, software deployment, and data center operation

businesses that truly matter to the organization. More importantly, organizations should commit to achieving their long-term collection and analysis objectives. With these KPIs, the organization can reduce the number of decisions that are based solely on instinct or gut feeling and make decisions based on objectivity and facts.

The real challenge toward performance-oriented management is not the definition of the KPIs but rather the creation of the necessary organizational maturity level required to collect the necessary data sustainably and accurately. The organization must create, in the short term, capabilities and disciplines on a few indicators, and progressively expand the list of indicators to guarantee an analytical and data-driven management style.

Organizational and Systems´ Readiness Assessment

Implementation of the systems readiness assessment (SRA) can help improve performance management for systems and aid decision makers in identifying programmatic and technical risk areas. The SRA gives decision makers awareness of a system's holistic state of maturity and quantifies the level of integration a specific component has attained with other components during system development. The assessment is a critical part of achieving the goals of improved system performance management and reduced program and technical risk. The SRA enables more effective system development management and integration that can ultimately shorten delivery timelines.

Feedback from Users

Effective revenue mobilization, under the terms set herein, requires increasing taxpayer automated systems use for self-servicing their tax returns and corresponding payments. If taxpayers are not satisfied with the electronic services provided, the tax administration must take corrective action.

Hence, the tax administration must conduct taxpayer satisfaction or perception surveys and place the results in the public domain. These are extremely important in gauging the effectiveness of the information services being provided to taxpayers. The surveys would also help determine the effectiveness of the measures taken by the revenue administration to promote voluntary compliance by providing adequate information services.

Basic System Features

Whether the tax administration chooses to build a tax system from scratch or buy a ready-made commercial off-the-shelf (COTS) system, it must ensure that the system correctly performs, at a minimum, the basic operational functions, starting with an integrated registry to collect the basic information needed to manage taxpayers and to facilitate other tax administration functions. The registry should be unique for the entire organization; the registration process should be as easy as possible; and the information contained in the registry must be of high quality, as the registry is the foundation for any other initiative.

A high-quality registry should be followed by a system that implements the basic core tasks of processing returns, processing payments, maintaining taxpayers' current accounts, providing tools to identify delinquent taxpayers, automating appeals tracking, and providing staff with access to taxpayer information to enable higher levels of service.

>>>

Conclusion and Next Steps

It is important to realize that digitalization is the key enabler for revenue authorities. The key question is how to properly sequence the IT infrastructure and the institutional reform needed to make digitalization happen. Digitalization of tax and customs administrations should be adapted to the environment available in each country and to the maturity level of each revenue administration. It is important to note that the pacing of each tax and customs administration varies. Governments must take a strategic rather than opportunistic perspective and make digitalization an integral part of their internal strategy with clear policy objectives. A modular approach should be adopted to facilitate the act of "plugging something in," seeing how it works, and then removing it or improving it. Talent management is key, as information should not be lost as we move toward digitalization. Likewise trust with private sector vendors should be established. Hence, it is important that a roadmap has a clear vision of how tax and customs administrations should look like in the near future.

More work is needed on finding solutions to the following questions and needs: What is the current state of technology available to tax and customs administrations and what will be the next wave? How do we adapt our institutions to a rapidly changing environment? In an environment where new technologies become more and more accessible to revenue administrations, how do we adapt strategies and business models of tax and customs administrations to make tax compliance less burdensome and more effective for all? Is "digital-by-default" an overarching objective of a strategy aimed at modernizing tax administrations and making them more efficient? What criteria can tax officials use to choose between different combinations of technologies? How can we stage the deployment of various technologies? How do we deal with the issues around security and confidentiality as tax administrations become the largest data handlers in their countries? How do we carry taxpayers with us during the transition to ensure they appreciate the value of digitalization of tax administration and do not simply feel threatened by it? Which regulations and practices built up in an analogue age can be dispensed with in a digital age, particularly in terms of compliance and reporting practices? What legal barriers may require modification to fully benefit from digitalization? How do we prioritize reengineering of key processes of tax and customs administrations that will facilitate the implementation of new technologies? How do we handle cultural and human resource issues and budget constraints? What are the new skills that revenue administrations should develop and employ to succeed? What is the role of change management programs? What should be their primary focus? Will technological changes lead to a more transparent and accountable tax administration?

To this end, the World Bank, the Vienna University Global Tax Policy Center, and Ernst & Young have established a seminar series on digital transformation of tax and customs administrations. The seminar series aims to provide tax and customs administration officials the opportunity to understand the level of digitalization in revenue administrations and the lessons learned, analyze the potential use of technology, and develop a digital tax administration roadmap. This initiative has already materialized in four online webinars (2020–2022), with over 500 officials joining from over 40 countries. Its main goal is to develop an active network of governments, businesses, and academics to explore the opportunities and challenges posed by new technologies for tax systems and related activities.

>>>

Notes

1. Business intelligence can be defined as the strategies and technologies used by enterprises for the data analysis of business information, providing historical and current views of business operations.

2. Big Data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.

3. Data analytics is the science of analyzing raw data to make conclusions about that information.

4. Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.

5. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

6. Distributed ledger technology is normally referred to as “blockchain.”

7. Blockchain is a decentralized distributed ledger technology, which allows creation, validation, and encrypted transaction of digital assets to take place and get recorded in an incorruptible way.

8. Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

9. Statistical analysis is the science of collecting, exploring, and presenting large amounts of data to discover underlying patterns and trends.

10. The WB-developed DIAMOND tool official website can be found at www.taxdiamond.org

11. Data mining consists of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

Bibliography

Adamov, Abzetdin. 2019. “Machine Learning and Advance Analytics in Tax Fraud Detection.” Proceedings from 2019 IEEE 13th International Conference on Application of Information and Communication Technologies (AICT).

Atanasijevic, Jasna, Dusan Jakovetic, Natasa Krejic, Natasa Krklec-Jerinkic, and Dragana Markovic. 2018. “Using Big Data Analytics to Improve Efficiency of Tax Collection in the Tax Administration of the Republic of Serbia.”

Austin, Marc, and Donald York. 2015. “System Readiness Assessment (SRA): An Illustrative Example.” *Procedia Computer Science* 44: 486–496. 2015 Conference on Systems Engineering Research.

Awasthi, Rajul, Hyung Chul Lee, Peter Poulin, Jin Gyu Choi, Woo Cheol Kim, Owen Jae Lee, Myung Jae Sung, and Sun Young Chang. 2019. *The Benefits of Electronic Tax Administration in Developing Economies: A Korean Case Study and Discussion of Key Challenges.* Report. Washington, DC: World Bank.

Barnay, Aurélie, Jonathan Davis, Jonathan Dimson, Emma Gibbs, and Daniel Korn. 2018. *Four Innovations Reshaping Tax Administration.* McKinsey & Company, Public Sector, Article, January.

Battiston, Pietro, Simona Gamba, and Alessandro Santoro. 2020. *Optimizing Tax Administration Policies with Machine Learning.* DEMS Working Paper Series No. 436 (March). Department of Economics, Management and Statistics, University of Milano-Bicocca.

Bentley, Duncan. 2019. “Timeless Principles of Taxpayer Protection: How They Adapt to Digital Disruption.” In *eJournal of Tax Research* 16 (3): 679–713.

Bondarenko, Svitlana, Iryna Liganenko, and Viktoriya Mykytenko. 2020. “Transformation of Public Administration in Digital Conditions: World Experience, Prospects of Ukraine.” *Social Development & Security* 10 (2).

BRITACOM. 2020. “Digital Transformation: BRITACOM Perspective.” *Belt and Road Initiative Tax Journal* 1 (2).

Butler, Jeff. 2020. “Analytical Challenges in Modern Tax Administration.” *The Ohio State Technology Law Journal.* Vol. 16.1.

CIAT. 2020. *ICT as a Strategic Tool to Leapfrog the Efficiency of Tax Administrations.* Flagship Document. Bill & Melinda Gates Foundation and Inter-American Center of Tax Administrations (CIAT).

Cipek, Ksenija, and Ivana Ljutic. 2020. "The Influence of Digitalization on Tax Audit." Review Article. Unpublished.

De Roux, Daniel, Boris Pérez, Andrés Moreno, María del Pilar Villamil, and César Figueroa. 2018. "Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach." Applied Data Science Track Paper. *KDD 2018.* London, August 19–23.

Dener, Cem, Hubert Nii-Aponsah, Love E. Ghunney, and Kimberly D. Johns. 2021. *WBG GovTech Maturity Index: The State of Public Sector Digital Transformation.* International Development in Focus. Washington, DC: World Bank.

Dzansi, James, Anders Jensen, David Lagakos, and Henry Telli. 2018. *Building Tax Capacity at Scale: Evidence from Technology Investments in Ghana.* June. Unpublished.

Fatz, Filip, Philip Hake, and Peter Fettke. 2020. *Confidentiality-Preserving Validation of Tax Documents on the Blockchain.* German Research Center for Artificial Intelligence, Institute for Information Systems, Saarbrücken, Germany. 15th International Conference on Wirtschaftsinformatik, Potsdam, Germany, March 8–11.

Gedde, Nora, and Ida-Sofie Sandvik. 2020. *Unsupervised Machine Learning on Tax Returns: Investigating Unsupervised and Semisupervised Machine Learning Methods to Uncover Anomalous Faulty Tax Returns.* Master thesis, Economics and Business Administration, Norwegian School of Economics, Bergen, Spring 2020.

Guseva, Tatiana, Elena Smetanina, and Vladimir Izotov. 2020. "Features of the Development of Tax Legal Relations in the Context of Digitalization." *Advances in Economics, Business and Management Research.* Vol. 156. 2nd International Scientific and Practical Conference on Digital Economy (ISCDE 2020).

Hatfield, Michael. 2019. "Professionally Responsible Artificial Intelligence." *Arizona State Law Journal* 51: 1057.

Henman, Paul. 2020. "Improving Public Services Using Artificial Intelligence: Possibilities, Pitfalls, Governance." *Asia Pacific Journal of Public Administration* 42 (4): 209–221.

Hoffer, Stephanie. 2020. "What If Tax Law´s Future Is Now? An Introduction to the Symposium on Artificial Intelligence and The Future of Tax Law." *The Ohio State Technology Law Journal.* Vol. 16.1.

Huang, Zhuowen. 2018. "Discussion on the Development of Artificial Intelligence in Taxation." *American Journal of Industrial and Business Management* 8 (8): 1817–1824.

Ihnatisinova, Denisa. 2021. "Digitalization of Tax Administration Communication Under the Effect of Global Megatrends of the Digital Age." *SHS Web of Conferences* No. 92, Globalization and its Socio-Economic Consequences.

IOTA (Intra-European Organisation of Tax Administrations). 2016. *Transforming Tax Administration and Involving Stakeholders.* Budapest, Hungary, November.

IOTA (Intra-European Organisation of Tax Administrations). 2017. *Disruptive Business Models: Challenges and Opportunities for Tax Administrations.* Budapest, Hungary, October.

Ippolito, André, and Augusto Cezar Garcia Lozano. 2020. "Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of Sao Paulo." *Proceedings of the 22nd International Conference on Enterprise Information Systems* (ICEIS 2020) 1: 452–459.

Junquera-Varela, Raúl Félix. 2011. "An Integrated Assessment Model for Tax Administration: The Diagnostic Tool." *Poverty Reduction and Economic Management.* Public Sector and Governance Group. Unpublished Draft.

Junquera Varela, Raul Felix, and Ekaterina Vostroknutova. 2015. *Peru—Selected Issues in Fiscal Policy: Taxation and Equity.* Washington, DC: World Bank.

Junquera-Varela, Raúl Félix, Marijn Verhoeven, Gangadhar P. Shukla, Bernard Haven, Rajul Awasthi, and Blanca Moreno-Dodson. 2017. "Strengthening Domestic Resource Mobilization: Moving from Theory to Practice in Low- and Middle-Income Countries." *Directions in Development, Public Sector Governance.* Washington, DC: World Bank.

Junquera-Varela, Raúl Félix, Rajul Awasthi, Oleksii Balabushko, and Alma Nurshaikhova. 2019. *Thinking Strategically about Revenue Administration Reform: The Creation of Integrated Autonomous Revenue Bodies.* Policy Note. Governance Discussion Paper 4. Washington, DC: World Bank.

Junquera-Varela, Raul Felix, Daniel Alvarez Estrada, and Cristian Oliver Lucas-Mas. 2021. "COVID-19 and Taxation: Between the Devil and the Deep Blue Sea." *Equitable Growth, Finance and Institutions Insight.* Washington, DC: World Bank.

Junquera-Varela, Raul Felix, and Cristian Oliver Lucas-Mas. 2021. "COVID-19: Tax Policy and Revenue Administration Implications." *Equitable Growth, Finance and Institutions Insight.* Washington, DC: World Bank.

Kariuki, Elizabeth. 2013. "Towards Sustainable ICT Systems in Tax Administrations." *Automation in Tax Administration: Summary Note.* Africa Policy Research Institute Limited. September.

Khwaja, Munawer, Rajul Awasthi, and Jan Loeprick. 2011. *Risk-Based Tax Audits: Approaches and Country Experiences.* Washington, DC: World Bank.

Kim, Young Ran. Forthcoming. "Blockchain Initiatives for Tax Administration." *UCLA Law Review.* Vol. 69.

Kochanova, Anna, Zahid Hasnain, and Bradley Larson. 2020. "Does E-Government Improve Government Capacity? Evidence from Tax Compliance Costs, Tax Revenue, and Public Procurement Competitiveness." *The World Bank Economic Review* 34 (1): 101–120.

KPMG. 2018. *Transforming the Tax Function Through Technology: A Practical Guide to 2020.* KPMG International. April.

Krsul, Ivan. 2021. *An Incremental Strategy for the Development of Machine Learning Capabilities in a Tax Administration.* Vol. 1. Working Paper. Unpublished. September 19.

Krsul, Ivan. 2021. *Building Data Science Capabilities in a Tax Administration.* Vol. 7. Working Paper. Unpublished. October.

Krsul, Ivan. 2017. *Best Practices for the Implementation of Tax Administration IT Systems.* Vol. 2. Working Paper. Unpublished. July.

Krsul, Ivan. 2021. *Artificial Intelligence and its Application in Tax and Customs Administrations.* PowerPoint Presentation. Unpublished.

Krsul, Ivan. 2021. *Considerations on the Implementation of Blockchain Technology on Tax Administrations.* Working Paper. Unpublished. November.

Le, James. 2016. 7 *Machine Learning Applications at Google.* Data Notes. https://data-notes.co/7-machine-learning-applications-at-google-843d49d77bc8.

Laizet, Alexandre. 2019. A*rtificial Intelligence and Transfer Pricing: The Future of International Tax Planning.* Master thesis, University of Paris-Saclay ESCP Europe. Unpublished.

Lucas-Mas, Cristian Oliver, and Raul Felix Junquera-Varela. 2021. *Tax Theory Applied to the Digital Economy: A Proposal for a Digital Data Tax and a Global Internet Tax Agency.* Washington, DC: World Bank.

Lyutova, Olga, and Irina Fialkovskaya. 2021. "Blockchain Technology in Tax Law Theory and Tax Administration." *Rudn Journal of Law* 25 (3): 693–710.

Microsoft and PricewaterhouseCoopers. 2017. *Digital Transformation of Tax Administration.* October.

Milner, Cas, and Bjarne Berg. 2017. "Artificial Intelligence and Machine Learning – Level 5." *PwC Advanced Tax Analytics & Innovation Series.*

OECD (Organisation for Economic Co-operation and Development). 2013. *Tax Administration 2013: Comparative Information on OECD and other Advanced and Emerging Economies.* Paris: OECD Publishing.

OECD. 2013. *Managing Service Demand: A Practical Guide to Help Revenue Bodies Better Meet Taxpayers´ Service Expectations.* Paris: OECD Publishing.

OECD. 2014. *Tax Compliance by Design: Achieving Improved SME Tax Compliance by Adopting a System Perspective.* Paris: OECD Publishing.

OECD. 2014. *Increasing Taxpayers´ Use of Self-Service Channels.* Paris: OECD Publishing.

OECD. 2014. *Working Smarter in Tax Debt Management.* Paris: OECD Publishing.

OECD. 2015. *Tax Administration 2015: Comparative Information on OECD and other Advanced and Emerging Economies.* Paris: OECD Publishing.

OECD. 2016. *Co-operative Tax Compliance: Building Better Tax Control Frameworks.* Paris: OECD Publishing.

OECD. 2016. *Rethinking Tax Services: The Changing Role of Tax Service Providers in SME Tax Compliance.* Paris: OECD Publishing.

OECD. 2016. *Technologies for Better Tax Administration: A Practical Guide for Revenue Bodies.* Paris: OECD Publishing.

OECD. 2017. *Technology Tools to Tackle Tax Evasion and Tax Fraud.* Paris: OECD Publishing.

OECD. 2017. *Tax Administration 2017: Comparative Information on OECD and other Advanced and Emerging Economies.* Paris: OECD Publishing.

OECD. 2017. *The Changing Tax Compliance Environment and the Role of Audit.* Paris: OECD Publishing.

OECD. 2019. *Tax Administration 2019: Comparative Information on OECD and other Advanced and Emerging Economies.* Paris: OECD Publishing.

OECD. 2019. *Tax Compliance Burden Maturity Model.* Paris: OECD Publishing.

OECD. 2019. *Tax Debt Management Maturity Model.* Paris: OECD Publishing.

OECD. 2019. *Introducing a Commercial Off-the-Shelf Software Solution.* Paris: OECD Publishing.

OECD. 2019. *Implementing Online Cash Registers: Benefits, Considerations and Guidance.* Paris: OECD Publishing.

OECD. 2019. *Measuring the Digital Transformation: A Roadmap for the Future.* Paris: OECD Publishing.

OECD. 2019. *Unlocking the Digital Economy – A Guide to Implementing Application Programme Interfaces in Government.* Paris: OECD Publishing.

OECD. 2019. *The Sharing and Gig Economy: Effective Taxation of Platform Sellers: Forum on Tax Administration.* Paris: OECD Publishing.

OECD. 2019. *Hello, World: Artificial Intelligence and its Use in the Public Sector.* Paris: OECD Publishing.

OECD. 2019. *Artificial Intelligence in Society.* Paris: OECD Publishing.

OECD. 2019. *Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO).* OECD Digital Economy Papers No. 291. Paris: OECD Publishing. November.

OECD. 2020. *Tax Administration 3.0: The Digital Transformation of Tax Administration.* Paris: OECD Publishing.

OECD. 2020. *OECD Digital Economy Outlook 2020.* Paris: OECD Publishing.

OECD. 2020. *The Digitalisation of Science, Technology and Innovation: Key Developments and Policies.* Paris: OECD Publishing.

OECD. 2020. *Identifying and Measuring Developments in Artificial Intelligence: Making the Impossible Possible.* Paris: OECD Publishing.

OECD. 2021. *Tax Administration 2021: Comparative Information on OECD and other Advanced and Emerging Economies.* Paris: OECD Publishing.

OECD. 2021. *Enterprise Risk Management Maturity Model.* Paris: OECD Publishing.

OECD. 2021. *Tax Administration: Towards Sustainable Remote Working in a Post COVID-19 Environment.* Paris: OECD Publishing.

OECD. 2021. *Tax Administration: Digital Resilience in the COVID-19 Environment.* Paris: OECD Publishing.

OECD. 2021. "The Digital Transformation of SMEs." *OECD Studies on SMEs and Entrepreneurship.* Paris: OECD Publishing.

Pérez López, César, María Jesús Delgado Rodríguez, and Sonia de Lucas Santos. 2019. "Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers." *Future Internet* 11 (86).

Rathi, Ankit, Saurabh Sharma, Gaurav Lodha, and Manoj Srivastava. 2021. "A Study on Application of Artificial Intelligence and Machine Learning in Indian Taxation System." *Psychology and Education* 58 (2): 1226–1233.

Republic of Serbia. 2020. *Study on the Feasibility of Using Blockchain Technology in Public Administration of the Republic of Serbia.* Ministry of Public Administration and Local Self-Government, Republic of Serbia.

Respati, Nugroho. 2020. "The Adoption of E-Government in the Tax Administration: A Scoping Review." *Scientax, Jurnal Kajian Ilmiah Perpajakan Indonesia* 1 (2): 109–130.

Rosario, Shireen, and Kavita Chavali. 2020. "Digitization of Taxation in the Changing Business Environment & Base Erosion & Profit Shifting (BEPS): Special Reference to India." *European Scientific Journal* 16 (1). January.

Saragih, Samuel, and Milla Setyowati. 2019. *E-Readiness of Blockchain Technology in Modernization of Tax Administration in Indonesia.* Unpublished.

Savic, Milos, Jasna Atanasijevic, Dusan Jakovetic, and Natasa Krejic. 2021. *Tax Evasion Risk Management Using a Hybrid Unsupervised Outlier Detection Method.* Unpublished.

Shan, JingJing. 2019. "Optimization Strategy of Tax Planning System in the Context of Artificial Intelligence and Big Data." *Journal of Physics: Conference Series.* Vol. 1345.

Tulacek, Michal. 2019. "Legal Aspects of Tax Administration Electronisation." In *European Financial Law in Times of Crisis of the European Union,* edited by G. Hulko and R. Vybiral, 577–584. Budapest: Dialog Campus.

Twesige, Daniel, Faustin Gsheja, Uzziel Hategikimana, Raymond Philippe Ndikubwimana, Yvette Mwiza, and Innocent Hitayezu. 2020. “Smart Taxation (4Taxation): Effect of Fourth Industrial Revolution (4IR) on Tax Compliance in Rwanda.” *Journal of Business and Administrative Studies* 12 (1).

USAID (United States Agency for International Development). 2013. *Detailed Guidelines for Improved Tax Administration in Latin America and the Caribbean.* USAID´s Leadership in Public Financial Management. Prepared by Deloitte Consulting LLP.

Vishnevsky, Valentine, and Viktorija Chekina. 2018. “Robot vs. Tax Inspector or How the Fourth Industrial Revolution Will Change the Tax System: A Review of Problems and Solutions.” *Journal of Tax Reform* 4 (1): 6–26.

Wang, Ying, and Pan Wang. 2020. “New Personal Tax Collection Management System Based on Artificial Intelligence and Its Application in the Middle Class.” *Journal of Physics:* Conference Series. Vol. 1574.

World Bank Group. 2020. “Artificial Intelligence in the Public Sector: Maximizing Opportunities, Managing Risks.” *EFI Insight-Governance.* Washington, DC: World Bank.

World Bank Group. 2021. “World Bank Guidebook for Accessible GovTech.” *EFI Insight-Governance.* Washington, DC: World Bank.

World Bank Group. 2021. “Disruptive Technologies in Public Procurement.” EFI Insight-Governance. Washington, DC: World Bank.

Zambrano, Raúl, and Santiago Díaz de Sarralde. 2021. “A Very, Very Simple Example on the Use of Artificial Intelligence in Tax Administration.” In Inter-American Center of Tax Administrations (CIAT) website. January 10.

Zhou, Lingyan. 2019. “Opportunities and Challenges of Artificial Intelligence in the Application of Taxation System.” *Advances in Economics, Business and Management Research.* Vol. 109. International Conference on Economic Management and Cultural Industry (ICEMCI).

Zumaya, Martin, Rita Guerrero, Eduardo Islas, Omar Pineda, Carlos Gershenson, Gerardo Iñiguez, and Carlos Pineda. 2021. “Identifying Tax Evasion in Mexico with Tools from Network Science and Machine Learning.” *In Corruption Networks: Understanding Complex Systems,* edited by O.M. Granados and J.R. Nicolas-Carlock, 89–113. Cham: Springer.


WORLD BANK GROUP