Funded by
the European Union

USING DIGITAL TECHNOLOGY TO STRENGTHEN OVERSIGHT OF PUBLIC PROCUREMENT IN PORTUGAL

THE USE OF DATA ANALYTICS AND MACHINE LEARNING BY THE TRIBUNAL DE CONTAS

Andras Hlacs, Helene Wells

OECD WORKING PAPERS ON PUBLIC GOVERNANCE
No. 83

OECD
BETTER POLICIES FOR BETTER LIVES

OECD Working Papers on Public Governance

Using digital technology to strengthen oversight of public procurement in Portugal

The use of data analytics and machine learning by the Tribunal de Contas

This report was funded by the European Union. Its contents are the sole responsibility of its authors and do not necessarily reflect the views of the European Union.

Acknowledgements

Under the leadership of Elsa Pilichowski, Director of the Public Governance Directorate (GOV) and the guidance of Paulo Magina, acting Head of the Infrastructure and Public Procurement Division (IPP), Julio Bacio Terracino, Head of the Anti-Corruption and Integrity in Government Division (ACIG), and Jesper Johnson, acting Head of the ACIG Division, András Hlács (IPP) and Helene Wells (ACIG) coordinated the development of this report. The drafting of the report was carried out by the OECD (András Hlács and Helene Wells), NOVA University (Bruno Damásio, Carolina Vasconcelos, and Niclas Frederic Sturm), and Tribunal de Contas (Ana Gonçalves and Pedro Batista). The authors are grateful for the valuable feedback received from Matthieu Cahen (GOV), Erika Bozzay and Kenza Kachani (IPP), Jamie Berryhill, Felipe Gonzalez-Zapata and Julian Olsen (INDIGO), Gregor Virant, Aleksandra Melesko and Marian Lemke (SIGMA), as well as Alice Holt and Ariana at the Directorate for Science, Technology and Innovation (STI).

This report benefited greatly from insights at all levels of the TdC. The OECD is grateful for the support and close collaboration of the TdC's leadership and audit departments. This includes Filipa Urbano Calvão (President), José Tavares (former President), Helena Abreu Lopes (Member of the Court), Sofia David (Member of the Court), Fernando Silva (Director General), Conceição Ventura (Deputy Director General), and the steering role of Ana Gonçalves (Audit Director) supported by Pedro Batista (Data Analyst) of TdC's Centre of Innovation, Technology and Methodology. The OECD is grateful for the contributions of all the experts who work across the TdC's audit departments and took the time to share their insights.

The OECD would like to thank the following institutions, which provided data (in alphabetical order of institution): the Competition Authority (AdC), the Institute of Public Procurement, Real Estate, and Construction (IMPIC), and the Portuguese transparency portal.

The OECD would also like to thank the following individuals for their insights during the Improving the efficiency and transparency of public procurement with a focus on strengthening the control framework in Portugal conference on 3 February 2025 (in alphabetical order of institution): Nuno Cunha Rodrigues, President, Competition Authority / Autoridade da Concorrência (AdC, Portugal), Gabriel de Luccas Ruiz, General Coordinator for Data and Innovation, Planning, Innovation and Sustainability Directorate, Office of the Comptroller General / Controladoria-Geral da União (CGU, Brazil), Helga Berger, Member of the European Court of Auditors (ECA), César Pestana, President, Government Shared Services Entity (eSPap, Portugal), Fernando Batista, President, Institute of Public Procurement, Real Estate, and Construction (IMPIC, Portugal), Bruno Maia, Member of the Board, Institute of Registries and Notaries (IRN, Portugal), Valéria Mitiko Nakano, Federal External Control Auditor, Tribunal de Contas da União (TCU, Brazil), and Mário Campos, Deputy Director General of the Tax and Customs Authority, Portugal.

The OECD Secretariat appreciates the support of Sandra Galan Gallego, project coordinator at the European Commission, as well as the previous project coordinator, Aleksandra Marchewka. The action was funded by the European Union via the Technical Support Instrument and implemented by the OECD in cooperation with the European Commission.

The authors also wish to thank Andrea Uhrhammer, Deborah Fernandez, Dovile Bogusyte, and Michaela Faux for administrative, editorial, and publishing support.

Table of contents

FIGURES

TABLES

BOXES

Executive summary

The digital transformation of oversight and integrity institutions is crucial for enhancing transparency, efficiency, and accountability in the management of public procurement and public funds. Prioritising the digital transformation of institutions responsible for oversight and audit helps these institutions improve service delivery and foster citizen engagement. Oversight and integrity institutions that can integrate advanced digital technologies and analytics, including artificial intelligence (AI), are in a better position to detect, prevent, and address corruption and misconduct. Given the complexity and volume of data that institutions, including supreme audit institutions (SAIs), are required to consider, the adoption of digital tools to streamline processes and to improve data and risk analysis is necessary.

The *Tribunal de Contas* (Court of Auditors, hereafter TdC) is Portugal's SAI and is responsible for overseeing the proper management and legal use of Portugal's public resources. It plays a critical role in ensuring the regularity, efficiency and cost-effectiveness of public procurement in Portugal. The OECD and NOVA University Lisbon (Universidade) helped TdC develop and refine a risk assessment methodology, including the development of a data-driven risk model to undertake audit assessments. The initiative aims to improve the TdC's identification of risks and the early detection of irregularities through advanced data analysis and machine learning (ML), a form of artificial intelligence (AI). The methodology developed marks a significant milestone in the TdC's digital transformation. The risk indicators include a mixture of rule-based (red flags for simple rule violations, such as no competition in a high-value contract"), inference-based (red flags for patterns or repeated behaviour, such as "the same company always wins"), and model-based (red flags found by smart systems that learn from past data to spot unusual activity) indicators and require access to external data sources. This initiative has been selected as an example to highlight the implementation considerations and challenges (such as data quality) that oversight and integrity institutions must consider when developing a model.

Several good practices have been identified during the development of the risk assessment methodology that underline the importance of shared understanding and commitment to addressing these challenges when developing and implementing any data-driven audit risk model. For example, improving the quality and accuracy of data and committing to investing in knowledge sharing and enhancing staff expertise and skills. Collaboration, sharing, and access to data across multiple institutions require stakeholders to be identified early and to be proactively and routinely engaged in the development of an audit risk model. Data custodians need to be involved in the model's critical appraisal and review. Data-driven audit risk models should not remain static: ongoing enhancements and updates to a model may involve the development and implementation of more advanced indicators of risks. Finally, opportunities for automation and scalability of the data-driven risk assessment and its continuous optimisation (for example, feature engineering or updates to the data pipeline) provide further opportunities to ensure the sustainable implementation of a data-driven model in oversight and integrity institutions.

Abbreviations and acronyms

AdC	Competition Authority (Autoridade da Concorrência Portugal)
Ad&C	Agency for Development and Cohesion (Agência para o Desenvolvimento e Coesão)
AI	Artificial Intelligence
AMA	Agency for Administrative Modernization (Agência para a Modernização Administrativa)
AT	Tax Authority (Autoridade Tributária e Aduaneira)
CCP	Public Procurement Code (Código dos Contratos Públicos)
CGU	Office of the Comptroller General (Controladoria-Geral da União, Brazil)
CI	Confidence Interval
CPV	Common Procurement Vocabulary
DG REFORM	Directorate-General for Structural Reform Support (European Commission)
GDOC	Document Management System (Sistema de Gestão Documental e Processual)
GENT	Information System for the Management of Entities (Sistema de Gestão de Entidades)
ID	Identification
IMPIC	Institute of Public Procurement, Real Estate, and Construction (Instituto dos Mercados Públicos, do Imobiliário e da Construção)
INCM	Portugeuse Mint (Imprensa Nacional Casa da Moeda)
IRN	National Registries Institute (Instituto dos Registos e do Notariado)
IT	Information Technology
LLM	Large Language Model
MEC	Medidas Especiais de Contratação
ML	Machine Learning
NIF	Tax Identification Number (Número de Identificação Fiscal)
NLP	Natural Language Processing
OECD	Organisation for Economic Cooperation and Development
Portal BASE	IMPIC's open portal to access data on public procurement in Portugal
TdC	Court of Auditors (Tribunal de Contas)

TSI	Technical Support Instrument
RCBE	Registry of Beneficial Ownership
SAI	Supreme Audit Institution
SG REFORM	Reform and Investment Task Force (European Commission)
VAT	Value-Added Tax

1 Digital transformation of oversight and integrity institutions

The digital transformation of oversight and integrity institutions is crucial for enhancing transparency, efficiency, and accountability in the management of procurement and public funds. Integrating digital technologies into public procurement processes can significantly improve monitoring and control mechanisms, ensuring that procurement activities are conducted accurately, fairly, and transparently (OECD, 2024[1]). This transformation involves the adoption of e-procurement systems, which streamline procurement processes, reduce administrative burdens, and provide real-time data for better decision-making.

Digital transformation of public procurement not only enhances operational efficiency but also fosters innovation and market competition. By leveraging digital tools, relevant government entities can better measure outcomes and ensure precise compliance with procurement standards (OECD, 2024[2]). Furthermore, this shift supports broader policy goals such as digital government initiatives and the digital transition agenda (OECD, 2023[3]). The integration of e-procurement systems can also facilitate greater participation from small and medium-sized enterprises (SMEs), promoting inclusive economic growth (OECD, 2024[4]). As countries like Ireland have demonstrated, mature use of e-procurement systems during the tendering phase can significantly reduce administrative overhead and increase transparency (OECD, 2024[2]).

Digital transformation in public procurement necessitates the establishment of solid foundations (OECD, 2024[4]), including the development of adaptable governance arrangements and resilient digital public infrastructures. By focusing on these areas, governments can create an environment where digital technologies are effectively integrated into oversight and integrity functions. This not only improves the efficiency of public procurement but also strengthens the overall integrity and accountability of public institutions.

1.1. Prioritising digital transformation in public institutions: a path to enhanced efficiency and transparency

Digital transformation is indispensable for public institutions striving to remain effective, transparent, and responsive in an increasingly digital world (OECD, N/A[5]) (OECD, 2024[4]). By embracing digital technologies, governments can significantly enhance service delivery, foster deeper citizen engagement, and ensure the more efficient use of public resources. Leveraging digital tools allows public institutions to streamline processes, reduce administrative burdens, and provide real-time data for informed decision-making.

Prioritising digital transformation allows public institutions to better meet the evolving needs of their constituents and foster greater trust and accountability (OECD, 2024[6]). In this context, digital public infrastructure, including e-procurement systems and digital identity frameworks, plays a pivotal role in achieving these objectives. By integrating these technologies, governments can enhance compliance with procurement standards, facilitate greater participation from SMEs, and promote inclusivity and economic growth (OECD, 2024[6]).

The broader context for why digital transformation in the public sector is needed stems from the rapid advancements in technology and the growing expectations of citizens for more accessible and efficient public services. The OECD Recommendation of the Council on Digital Government Strategies emphasises the need for strategic approaches to digital government, advocating for the use of digital technologies and data to create more open, participatory, and innovative governments (OECD, 2014[7]). This recommendation highlights the importance of adopting user-centred services driven by data and network effects to deliver personalised and effective public services.

The OECD Declaration on Public Sector Innovation outlines principles and actions to enhance innovation within the public sector (OECD, 2019[8]). This declaration legitimises innovation as a core strategic function of public sector organisations, encouraging governments to systematically use innovation to achieve policy goals. By fostering a culture of innovation, public institutions can better address complex challenges and improve their overall effectiveness.

The OECD Recommendation of the Council on Artificial Intelligence (AI) promotes the responsible stewardship of trustworthy AI, ensuring respect for human rights and democratic values (OECD, 2024[9]). This recommendation provides a framework for governments to implement AI technologies in a way that is transparent, secure, and accountable. By integrating AI into public sector operations, institutions can enhance their decision-making processes, improve service delivery, and better manage public resources.

1.1.1. Oversight and integrity institutions are ripe for digital transformation

Oversight and integrity institutions (Box 1) play a crucial role in ensuring transparency, accountability, and ethical conduct within the public sector. As the digital age progresses, these institutions are increasingly recognising the need to embrace digital transformation to enhance their effectiveness and efficiency. Integrating advanced digital technologies helps oversight and integrity institutions in preventing and addressing possible corruption and misconduct, thereby strengthening public trust and governance (OECD, 2023[10]).

Box 1. Oversight and integrity institutions

Oversight and integrity institutions play a crucial role in ensuring that public sector functions and activities are conducted in a transparent, accountable, and ethical manner. These institutions are responsible for monitoring and evaluating government operations, policies, and programmes, thus contributing to preventing mismanagement, fraud, and corruption. According to the OECD, oversight bodies include entities such as Supreme Audit Institutions (SAIs), anti-corruption agencies, and ombudsman offices. These institutions are tasked with scrutinising public sector performance and compliance with legal and ethical standards, and provide an essential check on government power, fostering an environment of trust and accountability conducive to sustainable and inclusive economic development. Integrity institutions, on the other hand, focus on promoting ethical behaviour and preventing conflicts of interest within the public sector. The OECD emphasises that these institutions are vital for maintaining the integrity of public administration by establishing and enforcing standards of conduct for public officials. This includes developing frameworks for managing conflicts of interest, implementing codes of ethics, and providing training and guidance on ethical issues.

Source: (OECD, N/A[11]); (Renda, Castro and Hernández, 2022[12])

Oversight and integrity institutions are excellent candidates for digital transformation due to their critical role in maintaining public sector integrity and accountability. The complexity and volume of data these institutions handle necessitate the adoption of digital tools to streamline processes, improve data and risk analysis, and enhance decision-making. Digital transformation enables these institutions to leverage emerging technologies such as AI and big data analytics, to identify patterns of risk, automate routine tasks, and ensure more accurate and timely reporting. This digital shift not only improves operational efficiency but also enhances the institutions' ability to respond to emerging risks and challenges.

The European Union's Artificial Intelligence Act introduces a robust framework to enhance oversight in public procurement through the strategic use of data analytics and machine learning. By classifying AI systems based on risk and mandating transparency, accountability, and human oversight, the Act ensures that public sector entities adopt AI technologies responsibly. In procurement, this translates to stricter evaluation of AI vendors, standardised contractual clauses, and the integration of algorithmic auditing tools to detect bias, inefficiency, or non-compliance (European Commission, 2025[13]). These measures aim to foster trust, fairness, and innovation while safeguarding public interest in the deployment of AI across government services.

The OECD definition of an AI system is contained in the OECD AI Principles (OECD, 2024[9]), which helped to inform the development of the EU AI Act. It reads as follows (OECD, 2024[14]):

> *"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."*

In 2024, the European Court of Auditors (ECA) published its *Roadmap for Artificial Intelligence: Initial Strategy and Deployment*, outlining a forward-looking approach to integrating AI into its audit and oversight functions (ECA, 2024[15]). The roadmap emphasises the importance of a data-driven culture, responsible AI use, and the need for robust governance frameworks. It highlights practical steps for adopting AI tools to enhance audit quality, efficiency, and risk detection, while also addressing challenges such as data access, ethical considerations, and staff upskilling. This strategic vision offers valuable insights for other SAIs exploring AI adoption in public sector oversight.

The OECD provides a robust foundation for this transformation through several key recommendations and reports. The OECD Recommendation of the Council on Public Integrity emphasises the importance of a comprehensive integrity system that integrates digital tools to promote transparency and accountability (OECD, 2017[16]). The OECD's Anti-Corruption and Integrity Outlook 2024 highlights the need for updated integrity frameworks to address evolving corruption risks, including those related to digital technologies (OECD, 2024[17]). Additionally, the OECD's report on Generative AI for Anti-corruption and Integrity in Government explores the potential of AI to enhance the impact of integrity institutions by improving risk assessments, detecting anomalies, and supporting decision-making processes (Ugale and Hall, 2024[18]).

1.2. Improving the transparency and integrity of public procurement

Public procurement is an important part of public financial management, representing a significant share of GDP in OECD countries, as it accounts for approximately 13% of GDP (OECD, N/A[19]). In addition, public procurement can ensure that public funds are used efficiently, transparently, and responsibly. This substantial expenditure underscores the importance of ensuring that public funds are used efficiently and responsibly. Effective public procurement practices can lead to better public services, improved infrastructure, and overall economic growth.

Transparency in procurement processes is essential for fostering accountability and trust. When procurement activities are conducted openly, civil society, oversight bodies, and other stakeholders can monitor and evaluate the use of public resources. It helps to ensure that public funds are used for their intended purposes and to prevent corruption and mismanagement (OECD, N/A[20]). Moreover, it promotes fair competition among suppliers, improving the quality and cost-effectiveness of public goods and services.

Integrity issues in public procurement can be a concern, as the volume of transactions and close interactions between the public and private sectors can create opportunities for unethical practices. The OECD Recommendation on Public Procurement emphasises the importance of integrity by advocating for measures to prevent corruption, fraud, and mismanagement. It calls for transparent procedures, accountability, and the professionalisation of procurement personnel (OECD, 2015[21]). The Recommendation highlights the need to combat corruption by implementing robust measures to ensure transparency, accountability, and ethical behaviour throughout the procurement process. It advocates for clear regulations, effective monitoring, and the empowerment of public procurement officers to prevent corrupt practices and promote integrity (OECD, 2015[21]).

Oversight and integrity institutions can use digital technologies to develop data-driven risk models that assess corruption risks related to public procurement (World Bank, N/A[22]). For example, the OECD has collaborated with the Belgian government to strengthen public integrity by developing such a model. This model uses data analytics to identify potential risks and implement preventive measures, thereby enhancing the overall integrity and efficiency of public procurement processes. (See Box 2).

Box 2. Belgium – Strengthening the strategic approach to public integrity in Belgium, including the integrity of public procurement processes and data-driven approach in procurement risk management

Belgium's Federal Internal Audit (FIA), the Integrity Bureau of the Federal Public Service, and the Directorate-General Federal Accountant and Procurement (BOSA) requested support from the European Union's Technical Support Instrument (TSI). One element of the initiative is the development of a data-driven risk model that can assist the FIA to better identify and assess corruption and fraud risks related to public procurement. The effective use of data and digital tools presents an opportunity for FIA to achieve value for money by being able to undertake more detailed assessments of potentially fraudulent public procurement as informed by the risk model. Whilst undertaking the initiative, the OECD has been reviewing and assessing the available public procurement data, including data that is from Belgium's eProcurement platform. Data has been used to develop a set of risk indicators for a proof-of-concept analytics model. Guidance and capacity building have been undertaken with relevant stakeholders, including hands-on workshops with the risk model and accompanying visual dashboard.

The initiative is expected to embed a new data-driven risk model within Belgium's federal public sector, as well as improve capacity on how to use analytics and AI as an anti-corruption and fraud detection tool.

Source: OECD

1.3. Data-driven models for better oversight

The digitalisation of public institutions has dramatically increased the complexity and volume of information available for audit institutions to process and analyse. This encompasses data, documents, processes, and systems, all contributing to a digital landscape characterised by complexity. In many instances, this pace of transformation is accelerating. Audit bodies are expected to modernise as traditional audit methods are proving inadequate in this new data-driven environment, especially given the constraints of limited resources, personnel, and legacy systems. It is becoming a requirement for audit institutions to adapt their methods, techniques, and resources to manage the rise of digital complexity of public administration effectively (OECD, 2024[17]).

Data-driven approaches can ensure more thorough and precise audits as well as more detailed oversight of public procurement processes. Data analytics and AI, for example, can support the analysis of vast datasets to identify patterns and anomalies that might be indicative of irregularities or mismanagement in audited public procurement entities. The use of AI can allow the analysis of entire populations of transactions instead of relying on a sampling approach. AI can also automate routine audit data collection and analysis, and rapidly process and extract meaning from vast sets of documents, freeing up auditors to focus on more complex and judgment-intensive areas (INTOSAID, 2023[23]). Embracing the use of analytics and AI cannot only lead to more precise audits, but it can also minimise human error, accelerate audit processes, and yield more dependable outcomes. The OECD is currently working with the Republic of Lithuania to digitally transform the Special Investigation Service and develop an AI-driven corruption and fraud risk model (see Box 3). In another example, in Austria, the federal government has made the digital transformation of the country a top priority (DigitalAustria, 2024[24]). The OECD is also working with all nine Austrian regional audit institutions to help improve their audit functions with the use of analytics and AI (see Box 4).

Box 3. Lithuania – Digital transformation of law enforcement and AI-driven fraud risk model for national and EU funds

The Special Investigation Service of the Republic of Lithuania (Lietuvos Respublikos specialiųjų tyrimų tarnyba – STT) requested support from the European Union's TSI to develop a data-driven methodology and tool for assessing fraud and corruption risks related to public spending. The initiative, led by the OECD, aligns with Lithuania's ongoing digital transformation journey and the directive on combatting corruption in the European Union.

The OECD is collaboratively developing with STT a risk assessment methodology. The initiative also includes the assessment of STT's data governance and data management policies, as well as the development of a set of indicators and a risk assessment model that the STT will be able to use in the future.

To increase the sustainable implementation of the risk assessment model, the OECD is leveraging the Lithuania State Data Agency's State Data Management Information System (SDMIS). This is a 'sand box' environment, allowing the OECD to access relevant datasets and create a pilot analytics and AI-driven risk model using an environment that is secure.

This initiative commenced in September 2024 and is expected to be finalised by the second quarter 2026.

Source: (Council of the European Union, 2024[25]); (State Data Agency, N/A[26])

Box 4. Austria – Improving the audit function with the use of AI

The nine Austrian regional audit institutions (Landesrechnungshof – LRH), led by the Upper Austrian Court of Audit, requested support from the European Union's TSI to determine how best to improve audit practices across the nine institutions with the help of advanced data analytics techniques (with a specific focus on AI).

The initiative, led by the OECD, is in direct alignment with the Austrian government's Artificial Intelligence Mission 2030 that aims to harness the potential of AI towards the common good "in a responsible manner on the basis of fundamental and human rights, basic European values and the upcoming European legal framework".

The expected impact of this initiative is that the auditors from the nine regional audit institutions can perform fast, quality audits through analysing large volumes of data and information. It is also anticipated that the initiative will lead to i) increased knowledge of current audit practices and AI-supported methodologies, ii) the Austrian institutions adopting AI-supported audit methodologies, and iii) enhanced data-driven decision-making.

The major focus of the initiative is to develop proposed methodologies to implement AI technologies to enhance public audits. The collaborative development and selection of use cases is currently underway with the following as possible examples under consideration:

- The implementation of AI technologies with the appropriate tools to allow auditors to analyse large volumes of currently underexploited financial data or information.
- The application of natural language processing (NLP) techniques and their use in extracting information from large text-based data sources such as auditee documentation, company registries, or national/international public audit reports.
- The implementation of AI as an editorial assistant based on reporting standards enabling standardisation, simplification, and indicating redundancies.

The initiative commenced in September 2024 and is expected to finish by the second quarter of 2026.

Source: (DigitalAustria, 2024[24])

1.4. Case example: Enhancing the control framework of public procurement oversight in Portugal

In Portugal, the Tribunal de Contas (Court of Auditors, TdC) plays a critical role in ensuring the legality, regularity, efficiency, and cost-effectiveness of public procurement. The evolving complexity of Portugal's procurement systems demands innovative approaches to ensure adequate oversight. Furthermore, the Portuguese public procurement regulatory framework, which has undergone numerous amendments, presents both opportunities and challenges for aligning TdC audits with updated compliance requirements. In addition, procurement systems embed inherent risks related to competition, transparency and integrity (OECD, N/A[20]). Addressing these challenges requires a robust risk-based approach that goes beyond financial considerations to include strategic risks (OECD, 2024[1]).

TdC has made significant strides in enhancing its Information Technology (IT) infrastructure, and to further enhance its audit activities, the OECD and partners at NOVA University Lisbon (Universidade) supported TdC in developing and refining a risk assessment methodology. This methodology relates to TdC's audit selection for public procurement and leverages data and analytics for assessing risks in public procurement, emphasising the importance of data-driven risk assessments to refine the audit selection process and increase the effectiveness and efficiency of the public procurement system (OECD, 2024[1]). By utilising advanced analytics and artificial intelligence (AI), the TdC aims to detect procurement risks and irregularities, ensuring that public resources are managed responsibly and transparently.

Using this initiative as a case example to highlight implementation considerations and challenges, the following chapter describes some of the activities undertaken in developing the data-driven risk model for audit assessments. Nevertheless, challenges remain in ensuring access to critical data (for example, data from the Tax Authorities), overcoming data quality issues, and automating processes to support more effective audits. Capacity-building and change management are also essential components of this transformation.

2 Data-driven audit risk assessments: Implementation considerations

2.1. Tribunal de Contas

The Tribunal de Contas (TdC), as the Supreme Audit Institution (SAI) of Portugal, is responsible for overseeing the proper management and legality of Portugal's public resources. The TdC plays a critical role in safeguarding the integrity and the value for money of procurement practices. To fulfil its mandate, the TdC conducts a significant number of audits annually, spanning ex-ante, concomitant, and ex-post audits (OECD, 2024[1]). These audits are incredibly resource-intensive for the TdC, requiring substantial human and financial efforts. Modernising the TdC's audit approach is crucial. —The use of data-driven methodologies may leverage the detection of procurement risks, improving audit efficiency, and promoting greater accountability in public financial management. According to its 2024 annual report[1], TdC currently manages extensive procurement data as part of its oversight function – from the annual financial accounts of approximately 6500 public institutions, to approximately 3000 contracts reviewed annually within the scope of ex-ante analysis (Tribunal de Contas, 2024). There is a pressing need for the TdC to adopt a data-driven approach when undertaking audit risk assessments. The wealth of available data sources provides opportunities for the TdC to enhance the analysis, detection, and automation capacity of its audits through advanced data analytics methods.

The Strengthening Oversight of the Court of Auditors for Effective Public Procurement in Portugal report mapped the current data landscape in Portugal, highlighting its importance in enhancing TdC's audits relating to public procurement (OECD, 2024[1]). At the core of this initiative was the need to map risk indicators and data sources, examine the digital maturity of the TdC to conduct such work, and assess the quality of potential databases that could be used for building a new risk assessment methodology. The report identified the key data sources and their role in supporting TdC's transition to data-driven risk assessments. The report also highlighted the importance of accessing external and internal databases to identify risks and irregularities in procurement processes.

This initiative aims to improve TdC's identification of risks and facilitate the early detection of irregularities through advanced data analysis and machine learning (ML) techniques. It also aspires to enable real-time monitoring of public expenditure processes in subsequent iterations of the model, thus ensuring responsiveness and more effective oversight. These efforts are expected to lead to significant long-term benefits, such as increased transparency in public procurement, greater accountability in the use of public funds, and the accelerated digital transformation of the TdC.

In the shorter term, the project is expected to yield improvements in data governance, better utilisation of public procurement information, and enhanced oversight capacity. The development of innovative methodologies will also support more efficient resource allocation and control mechanisms. However, achieving these goals requires overcoming challenges associated with managing and analysing vast datasets from diverse sources, challenges for which advanced data analysis and ML techniques present promising solutions.

The expected long-term impact of the initiative is: i) enhanced identification of risks (including emerging risks) and unusual transactions, ii) improved early detection of potential irregularities, and iii) real-time monitoring through the analysis of large volumes of information, enabled by the application of a set of ML algorithms. Additionally, the resulting risk-oriented approach is expected to enable the real-time detection of non-compliance and the assessment of public expenditure processes. Other relevant expected impacts include greater transparency in public procurement, enhanced accountability, efficiency in the use of public funds, and accelerated digital transformation of the TdC.

The framework developed as part of this initiative marks a significant milestone in the TdC's digital transformation journey. By focusing on reducing resource-intensive manual contract reviews and prioritising high-risk cases identified through risk indicators, the framework enhances the overall efficiency and effectiveness of the TdC's oversight activities. It represents an initial yet crucial step toward a data-driven approach to audits, laying the foundation for more comprehensive and innovative risk analysis methodologies in the future.

2.2. Developing the risk indicators for the TdC's data-driven audit framework

The risk indicators matrix developed by TdC is central to the data-driven audit risk assessment. It includes 37 indicators spanning various areas of public procurement. Key risk indicators were identified, and their quality was evaluated for inclusion in the risk model. These indicators are shown in Annex A (Table A.1). The risk indicators were categorised into three approaches: rule-based, inference-based, and model-based. Each of these is explained below.

2.2.1. Rules-based indicators

Rules-based indicators tend to be simple and more intuitive. They consider conditional statements to evaluate the presence or absence of specific features within the data. Using conditional "if-else" logic, these indicators are structured into criteria (the "if" condition) and the corresponding actions or conclusions. The "else" part specifies the action or conclusions if the "if" condition is not met. For example, this may include identifying missing field values or detecting predefined anomalies, such as contracts signed before being officially awarded. This approach is particularly effective for screening contracts where known properties are flagged as irregular in advance. For example, the indicator "Contract date before the award decision date" checks whether the award decision date is before the contract date. Approximately 34 indicators in TdC's risk framework are designed using a rules-based methodology (see Table A.1).

The rules-based approach requires updates according to changes in legislation and procedures and does have limitations. For example, it is not well-suited for handling ambiguous cases, incomplete data, or scenarios that deviate from the predefined rules. Additionally, rules-based indicators require continuous maintenance of threshold values as required by changing legislation or procedures.

2.2.2. Inference-based indicators

Inference-based indicators are built on the statistical properties of the data and may be what comes to mind when considering the concept of big data analytics in this context. With this approach, the focus shifts from deterministic relationships to the uncertainty inherent in the sample used for analysis. These indicators rely on the computation of statistical properties, particularly the construction of confidence intervals (CIs). These intervals define the range of expected regular activity, with any observations falling outside this range potentially flagged as irregular. Using this approach, two indicators were developed to assess potential irregularities. These relate to the ratio of the estimated value to the contractual and base prices (see Table A.1). The reliability of those indicators is limited by data availability and the need for a representative sample.

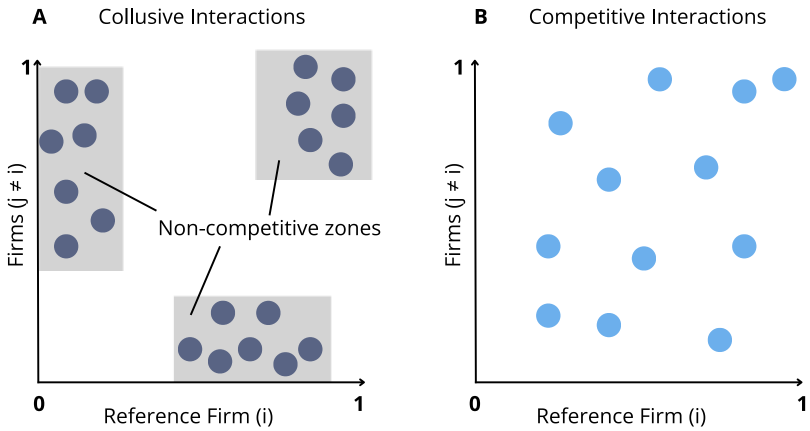
2.2.3. Model-based indicators

Model-based algorithms refer to a subset of algorithms that leverage statistical and mathematical models that can make predictions on data by learning from previously seen instances and may be what comes to mind when considering AI and ML in this context. Model-based indicators use models to represent a given system by capturing its key components, relationships, and dynamics, enabling the detection of relevant patterns. These indicators rely on algorithms that require careful consideration of their properties, as these directly influence how the model performs on new, unseen data. This approach involves splitting the dataset into two parts: i) a training set which is used to estimate and build the model, and ii) a testing set which is reserved to evaluate the model's performance and generalisation capability.

Creating model-based indicators is highly dependent on the availability and quality of data. The resulting models can often lack interpretability, making it challenging to understand or justify their predictions. Supervised learning is a type of ML that uses labelled datasets to train models to recognise patterns and make predictions. This method is particularly suitable for detecting collusion, as it allows for the identification of patterns that differentiate collusive behavior from legitimate activities.

As part of this initiative, the data-driven framework included a proof-of-concept of a supervised learning model to explore its potential for detecting collusion (bid rotation). Supervised learning techniques rely on labelled data – datasets annotated with the desired outputs to help the model learn and identify similar patterns. Since labelled data was not readily available within the TdC, data from external sources from Portugal (Autoridade da Concorrência, 2008[27]), Spain (Comisión Nacional de los Mercados y la Competencia, 2021[28]), Italy, Brazil, Japan, the United States (Rodríguez, 2021[29]) and Chile (Tribunal de Defensa de la Libre Competencia, 2018[30]) were used to train the model effectively. The model (see Figure 1) builds upon the work of Huber and Imhof (2023) and aims to detect anomalies in the values of proposals submitted by firms participating in open tenders (Huber and Imhof, 2023[31]). By summarising the bidding behaviour of firms, the model creates a two-dimensional representation, which facilitates the classification of previously unlabelled bids as either collusive or non-collusive.

Figure 1. Visual representation of the model for anomaly detection



Source: Adapted from Huber and Imhof 2023

The framework of three distinct types of risk indicators follows the design principle of covering different domains and aspects of risk assessment. Rule-based indicators derive from the expert knowledge and practical experience of TdC auditors, ensuring that institutional knowledge is operationalised. Inference-based indicators provide a nuanced perspective on procurement risks, accounting for fluctuations and variability. Model-based indicators can learn from data and be adapted over time to provide more accurate predictions.

In terms of interpretability, rule-based and inference-based indicators are quite open to inspection. Their transparency allows for scrutiny of their respective implementation. The key characteristic of model-based indicators is their ability to capture non-linear relationships. This helps to detect complex patterns that cannot easily be approximated by rules. Due to this non-linearity, model-based algorithms can only be as interpretable as the underlying method. The model-based indicator developed for the TdC is based on a neural network; a model for which it is commonly harder to interpret both the sensitivity of the model and the associations it makes between features of the data and its prediction.

2.3. Identifying the relevant data sources and data sets

The data landscape relating to Portugal's public procurement activities was comprehensively reviewed. The identification of core risks and irregularities in public procurement, together with key data variables from public sector institutions is documented in Chapter 3 of the Strengthening Oversight of the Court of Auditors for Effective Public Procurement in Portugal report (OECD, 2024[1]). TdC's data-driven audit risk assessment incorporates both internal and external sources of information. Various data sources proved valuable for the development of the data-driven audit risk model, including: the Competition Authority (AdC), the Portuguese Transparency Portal, and the Institute of Public Procurement, Real Estate, and Construction (IMPIC). The external data sources were primarily made available through the collaborative cooperation of these public institutions. Annex B provides further detail on the data sources used to develop the audit risk model for TdC.

2.4. Addressing the data quality challenges

Several data quality challenges significantly impact the reliability of the indicators' results. Some challenges included:

- **No unique identifiers**: Before 2023, there was no unique identifier available to link contracts between the eContas and IMPIC databases. Currently, this data file merge relies on a combination of contractor and supplier tax numbers, contractual price, and contract date.
- **Inconsistency across data sets**: One key issue is the inconsistency of tax numbers between the GENT and IMPIC databases, which complicates the process of merging these data sources. Due to these inconsistencies, there are imperfect matches and a limited number of common observations between the two datasets.
- **Diverse data sets**: The diversity of data formats, including inconsistencies in column names and variable structures (for example, date formats), further exacerbates the challenges and significantly increases the workload during the data pre-processing stage.

Addressing these data quality issues requires the implementation of more robust data-driven processes that prioritise data quality at every stage and automate pre-processing tasks. These improvements are essential for ensuring the consistency, reliability, and efficiency of any methodology that relies on data integration and analysis throughout its lifecycle. The quality of the results of the model is highly dependent on the quality of the data used to develop them. If there are data quality issues, such as missing data, then it is not possible to compute the indicators for the model. This emphasises further the need for a robust data quality management framework, including validation mechanisms, standardised formats, and continuous monitoring of data integrity. Such a data quality framework is essential to ensure that the analytical outputs are both meaningful and actionable, enabling auditors to confidently rely on the indicators for risk detection and decision-making.

The Strengthening Oversight of the Court of Auditors for Effective Public Procurement in Portugal report details the key dimensions for assessing an oversight or integrity institution's digital maturity with regard to this context. One of these dimensions – Strategy and organisation – encompasses the leadership's vision and strategy for digital transformation, including its goals for strengthening the use of digital technologies and data. Data management and data governance are key aspects of this dimension. This involves the policies, procedures, standards and controls that ensure data privacy, quality, consistency, and security. Key practices in the people and culture dimension include the expertise, skills, and commitment of individual employees within an organisation (OECD, 2024[1]). TdC are currently addressing the data quality and data governance issues to ensure the successful implementation of the risk model.

2.5. The TdC data-driven framework

An overview of the TdC data-driven risk framework provides a clear foundation for understanding how the framework works and its role in identifying risks and irregularities. The framework consists of four main components:

- **Input database**: This corresponds to the source data to calculate the indicators.
- **Python scripts**: Custom scripts are created for each indicator to process and analyse the data.
- **Output tables**: These tables contain flagged contracts and procedures and corresponding indicators.
- **Dashboards**: The results are visualised using dashboards.

This framework facilitates the flow of information from the IMPIC's database to the Python scripts, which, using relevant inputs, flag contracts for further analysis and inspection in the dashboards. Currently, the input database is composed of separate files from IMPIC's system as well as from the other institutions' databases. Future developments will aim to integrate these data into a more robust data lake (currently being developed by the TdC). This architecture will progressively improve data quality across layers. Initially, to ensure quality, relevant metadata columns are included in the output files, providing transparency and control over each key variable. The Python scripts are executed in batch mode, producing separate files for each indicator. These files are then re-arranged and fed into the dashboard software, Power BI, which organises the data into a star schema database. This approach enables in-depth analyses, and the creation of various dashboards that can be developed and tailored to the specific audit-related questions posed by TdC auditors.

2.6. Implementation challenges: The path forward for TdC

The data-driven audit risk model represents an initial development of a data-driven approach for audits and a proof of concept for a risk analysis model in public procurement. It is not fully integrated and automated within TdC systems, nor does it represent the final version of the risk analysis model and definitive solution for all audit needs in the field of public procurement. At its core, the initiative promoted digital transformation initiatives that focus on collaboration between public institutions, including the Institute of Public Procurement, Real Estate, and Construction (IMPIC), and the integration of data-driven tools to enhance transparency, accountability, and efficiency in public procurement audits. Moreover, potential databases to be included are from the Tax Authority (AT) and the Registry of Beneficial Ownership (RCBE) from National Registries Institute (IRN) that can improve analytical capacity. Key data sources such as the BASE public procurement portal, the beneficial ownership registry, and some of the databases managed by the Tax and Customs Authority were identified as crucial for this effort.

Data processing needs to be optimised to efficiently extract value from semi-structured and unstructured information. The available data has different sources, which may lead to different formats across variables. Additionally, missing values on relevant features could be a potential issue. Thorough data preprocessing is essential for TdC to overcome some of these challenges to successfully implement the data-driven audit risk model. Common implementation challenges can be overcome in order to strengthen data-driven audit risk model, and these opportunities are explored in the following section.

3 Overcoming implementation challenges: Opportunities to strengthen data-driven audit risk assessments

With the development and implementation of any data-driven audit risk model, there will be implementation challenges. Some of these challenges are common across oversight and integrity institutions, and most can be overcome or mitigated with careful planning. Others, if not dealt with before the development of the audit risk model, can have deleterious consequences, such as reducing the scope and validity of the insights derived from the model. Considerable effort is required not just for the oversight and integrity institutions (such as SAIs) implementing the models, but for the broader public sector to ensure these implementation challenges can be addressed. Many of these public sector institutions are the custodians of the data that is required to build audit risk models. It is imperative that there is a shared understanding and commitment to addressing these implementation challenges. Overcoming these challenges may include, but is not limited to the following:

- Improving the quality and accuracy of data.
- Investing in the key asset – people.
- Proactively collaborating, sharing, and accessing data across multi-institutions.
- Developing and implementing more advanced indicators of risk.
- Automation and scalability of the data-driven risk assessment.
- Continuous optimisation.

3.1. Improving the quality and accuracy of data

The quality and accuracy of data used for audit risk models can be improved using data validation mechanisms, more frequent updates and data feeds, and the enforcement of data quality standards. This ultimately leads to more reliable risk models and risk assessments. One way of improving the quality and accuracy of data is by enhancing data validation mechanisms to ensure data from various sources are accurate and up to date. This may involve automating the cross-referencing of datasets with other external sources to flag inconsistencies before they affect any analysis. In the case of TdC, this will mean ensuring data from the Court's internal systems, IMPIC, and the Transparency Portal are updated and accurate. Encouraging data custodians to increase the frequency of updates from public procurement databases and to integrate real-time or near-to-real-time data feeds into data models wherever possible ensures the data analysed – and the insights produced – are current and relevant. Enforcing data quality and consistency standards across public sector institutions enables data to be understood and used more efficiently by multiple institutions. The case example of TdC highlights a simple way of improving the standardisation of sectoral attribution of contracts – assigning reliable Common Procurement Vocabulary (CPV)[2] values to contracts in the case of the data.

3.2. Investing in the key asset – people: Continuous learning, feedback, and training

The success of implementing data-driven audit risk models relies heavily on the expertise, skills, and commitment of employees in oversight and integrity institutions. Relevant users of audit risk models should receive ongoing training to keep up with any changes (such as the inclusion of new risk indicators) and to participate in the creation of useful operational dashboards. Delivering ongoing training for auditors and data analysts on how to effectively use the data-driven framework, understand the outputs, and interpret the results is important. This could involve regular workshops or access to internal (or external) knowledge experts. If the data-driven risk assessment receives new features or capabilities (such as new risk indicators, the removal of risk indicators, different data, or updated algorithms), users should be properly trained on how this may impact their use of the model and how to incorporate these changes into their workflow effectively. Continuous learning and feedback require the implementation of feedback mechanisms with auditors, data analysts, and end-users. This encourages the iterative improvement of the framework based on the practical experiences of audit risk model users. The initiative with TdC highlights the importance of strengthening data literacy and digital skills among TdC staff to ensure the successful implementation and sustainability of data-driven audits. Continuous training, collaboration, and knowledge-sharing will be crucial in embedding a culture of innovation and adaptability within the institution.

The Strengthening Oversight of the Court of Auditors for Effective Public Procurement in Portugal report details the key dimensions for assessing an oversight or integrity institution's digital maturity in this context (OECD, 2024[1]). One of these dimensions – People and Culture – relates to the expertise, skills, and commitment of individual employees within an organisation. This digital maturity dimension holds relevance for integrity and oversight institutions that must consider the requirements for implementing and sustaining the use of the audit risk model (see Box 5)

3.3. Proactive multi-institutional collaboration, data sharing, and data access

Proactive, multi-institutional collaboration is critical to improving the oversight and integrity functions of the institutions responsible for auditing procurement and contract-related data. Without collaboration and a genuine commitment, the full benefits of a data-driven approach to assess risk may never be realised. The OECD's Recommendation on Enhancing Access to Sharing of Data encourages member countries to facilitate broader access to and sharing of data across sectors (OECD, 2021[32]). The sharing of data and enabling better data access aims to harness already existing data sources as a way of fostering innovation. Likewise, the OECD's Recommendation for Enhanced Access and More Effective Use of Public Sector Information guides oversight and integrity institutions on why data access is critical (OECD, 2008[33]). Public sector information should be available for use and re-use by default for oversight and integrity institutions, and there should be clearly defined exceptions in place for security, privacy, and legal protections. In addition, non-discriminatory, competitive access to public data, such as procurement and contract-related data, aims to remove unnecessary restrictions to accessing data.

During the development, refinement, and implementation of a data-driven model that assesses audit risk, it is imperative that relevant public sector data custodians are:

- **Identified early**: Identifying data custodians as early as possible helps to determine what data is critical (and available) in the development of a risk framework or risk model. Establishing the

Minimal Viable Product (MVP) of the risk model initiative and the data required to achieve the initiative requires early identification of data custodians to confirm data availability and data quality.

- **Proactively and routinely engaged**: Engagement with public sector institutions should not simply be limited to communicating the practicalities of data requests. Routine engagement should provide the relevant public institutions with strategic ‘buy-in.’ This could be in an advisory capacity throughout the initiative lifecycle, for example, on a Project Board or similar project governance activity.
- **Involved in critical appraisal and review**: Public sector data custodians know their data. As such, representatives from public sector institutions, whose data is used for audit risk frameworks and models by oversight and integrity institutions, should be involved in risk model appraisal and review. This can include such things as i) the selected methodology, ii) selected risk indicators, iii) the types of analyses undertaken, and iv) the governance of the data. Data custodians could also critically appraise insights generated from the use of a data-driven risk model. Why? It creates another validation opportunity to ensure the data is being reliably interpreted, whilst also offering a more objective, arms-length review (compared to an internal review only).

Developing and implementing more advanced risk indicators

Data-driven audit risk models require ongoing enhancements and improvements. Whilst proof-of-concept risk models may satisfy an immediate need of oversight and integrity institutions, consideration must be given to developing and implementing more advanced risk indicators over time. The development of advanced risk indicators may require further data interpretation and interrogation. One way of achieving this is through the application of NLP techniques. For example, analysing procurement-related ‘free-text’ fields containing justifications or explanations for non-standard procedures could help better interpret whether contract modifications or procedural exceptions are reasonable or suspicious. If indicators are limited to rule-based or inference-based indicators, a way of possibly identifying outliers beyond this is incorporate more ML-based anomaly detection models. These models may be able to identify patterns of irregularity or inefficiency that are not captured by the predefined rules or the existing model approach. No audit risk model should remain static. When opportunity arises – through better quality data or different types of data – additional risk indicators should be considered for possible inclusion in future model refinements.

3.4. Automation and scalability

Automation can significantly improve data-driven risk models for oversight and integrity institutions, particularly in relation to audit-related assessments. These improvements include streamlining processes, increasing accuracy, and making models more adaptive to changing circumstances, which may mean the inclusion of different risk indicators or datasets. Ingesting, cleaning, and processing large volumes of data more efficiently are tasks that can be automated. Ingesting data from various systems, cleaning data to check for inconsistencies or errors, and processing data to ensure uniformity (standardisation) are all tasks that help to reduce human error. More importantly, introducing automation to the risk model lifecycle helps to minimise resources needed for implementation and ongoing maintenance. Auditors and data analysts can instead concentrate their efforts on analysing and gaining insights from data, rather than manually coding, cleaning, and checking the data.

Automation of reporting processes can assist with overcoming implementation challenges that are often associated with resourcing constraints. Scripts can be developed that generate comprehensive automated reports that are customisable based on the need of oversight and integrity institutions. The reports can be tailored and timed at regular intervals, so that they are based on the most up-to-date data ingested by the risk model. Large Language Models (LLMs) may be able to assist in creating initial drafts of reports based on insights generated from audit frameworks, of which auditors and data analysts can critically review and update. Automation also has the potential to move institutions from a sampling-based assessment of audit risk, to potentially being able to assess one hundred per cent of all financial and performance related data (with comparable human resourcing impact).

Scaling up audit risk models involves expanding their capacity to handle more data, make more complex predictions, and operate across different environments without losing performance or accuracy. These processes can be technical and resource intensive. Scalability is key to ensuring that the newly implemented risk model is not simply a success in the short term. Scaling up the audit risk model to accommodate larger datasets, additional data sources, and more complex procedures helps to cement longer term use and success. Building the system on scalable architecture (for example, leveraging cloud infrastructure) helps to accommodate potential growth in data volume and complexity. There are a variety of platforms available to oversight and integrity institutions and these platforms provide on-demand access to vast computational resources (e.g., powerful graphics processing units (GPUs) and tensor processing units (TPUs)), storage, and network bandwidth, which can be scaled up or down based on required usage. Further advantages of scaling up include being able to handle the simultaneous ingestion of data from multiple institutions (enabling efficiency) and the identification of errors or risks that auditors and data analysts may miss if they were reviewing manually (i.e. improving classification of data).

Automating and scaling up any risk model that contributes to audit assessments requires a combination of computational power, model optimisation techniques, and well-defined data architectural strategies. By using cloud-based resources, and efficient data management tools, the ability to handle massive datasets and complex tasks increases. This ensures that data-driven audit risk assessments remain accurate, responsive, and cost-efficient. Ultimately, automation and scalability require further investment by oversight and integrity institutions, and this is something to factor in when initially funding proof-of-concept initiatives. That is, when developing and implementing risk models initially, it is important to have regard for the long-term possible resourcing requirements of the initiative, especially for automation and scalability.

3.5. The importance of continuous optimisation

Data-driven audit risk models should be developed so that they are able to be further optimised. This means trying to improve the performance and efficiency and, in some cases, the generalisability of the risk model. Ultimately, oversight and integrity institutions want to be able to ensure any model performs its tasks accurately, and in a cost-effective and timely manner. Specific optimisation strategies depend on the type of data-driven risk model developed, the computational resources available, and the complexity of data being consumed. Various model optimisation techniques may be relevant to ensure the implemented audit risk model remains of relevance, and these are briefly explained below:

- **Feature engineering**: Creating or selecting the most relevant features from the data may improve the risk model performance. Sometimes, adding or removing features can impact the model's ability to learn.
- **Model pruning**: By removing unnecessary elements of a model (for example, this could be weights or unnecessary layers in neural networks), this may make the model more efficient. What this means is that by ‘pruning’ the risk model, it can reduce the computational burden so that it can run more quickly.
- **Data pipeline**: Optimising the data pipeline (for example, using faster data loading methods or data preprocessing techniques) helps to ensure timely handling of larger, more complex datasets.

- **Data batching**: Rather than ingesting full datasets during each update, it may be more beneficial to use mini-batches, essentially small subsets of larger datasets. This may help to process the data more efficiently.

  Metrics for evaluation: It is not only the model that requires continuous optimisation. It is also the way in which the model is evaluated. Using the correct evaluation metrics (for example, precision, recall etc.) helps to ensure more meaningful and relevant evaluations over time.

Box 5. People and culture: a key dimension for assessing institutional digital maturity for effective public procurement

The OECD report, Strengthening Oversight of the Court of Auditors for Effective Public Procurement in Portugal, outlines the key dimensions for assessing institutional digital maturity in this context. These include:

- Strategy and organisation: encompassing the vision and strategy for digital transformation of the institution, including its goals for strengthening the use of digital technologies and data.
- Technology and Process: The broader vision for an institution’s digital transformation should inform technological advancements (i.e. use of tools and software), rather than being led by them.
- Environment and partnerships, including national frameworks (i.e., legal and policy frameworks) and the cultivation of collaborative relationships between entities.
- People and culture: the expertise, skills and commitment of individual employees within an organisation are central to the digital maturity of an institution with regard to public procurement.

Key practices in the People and culture dimension include:

- Ensure that leadership visibly endorses and partakes in digital initiatives, embodying a top-down commitment to the organisation's digital aspirations.
- Develop and implement a change management and continuous learning plan that focuses on enhancing digital and data literacy, as well as sector-specific knowledge.
- Introduce and encourage training programmes targeting technical proficiencies like advanced programming and data ethics.
- Institute clear policies that favour experimentation with new digital tools and technologies to foster innovation and a “trial-and-error” mentality.
- Establish guidelines on the ethical use of data, ensuring that staff understands and adheres to them.
- Prioritise and establish mechanisms for internal knowledge sharing, facilitating the dissemination of sector-specific, technical and legal expertise.
- Promote a culture of collaboration and digital empowerment, where employees at all levels feel engaged and invested in digital transformation objectives.
- Collaborate with legal experts to navigate the intricacies of data laws, ensuring the organisation remains compliant while maximising its digital potential.
- Implement feedback loops to understand employee challenges and needs in the digital landscape, adjusting strategies based on this feedback.
- Regularly evaluate the digital skills gap within the organisation and adjust training programmes accordingly.

Source: (OECD, 2024[1])

Annexe A. Risk indicators used for the TdC data-driven risk model

Table A.1. Risk indicators used for the TdC data-driven risk model

Method	Group	Indicator	
Rule-based	Financing	1001	European Funding
	Procurement procedure	1002	Absence of legally complaint tendering procedure
		1003	Artificial split of works/services/supplies across several procedures/supplier concentration (1st phase)
		1004	Repeated use of exceptional situations/circumstances to avoid competitive procedures/restricted and closed procedures types/non-open procedures
		1005	Call for tender not published in official journal
		1006	Violation of legal limits for direct awards and call for tenders to the same supplying companies, as well as related companies
		1007	No base (default) price has been established
		1008	Length of submission/tendering period
		1009	Procedure that has been preceded by another one that was annulled of failed
		1010	Below threshold "densification"
	Contracting requirements	1011	Insufficient definition of the object of the contract/number of challenges (1st phase)
	Evaluation bidders, tenders and award procedure	1012	Exclusion of all but one bid
		1013	Exclusion of the tenderer offering the lowest price
		1014	Date of beginning of works earlier than the award date
		1015	Length of decision period
		1016	Single bidding/Number of tenders received (low)
		1019	Any compliant from tenders
	Contract award and execution	1020	Proportion (number and value) of contracts awarded to the same bidder (3 years) or (alternative) supplier's contract share of buyer's spending on public contracting (3 years)
		1021	Contract was not communicated to the TdC when it should have been (MEC) or COVID
		1022	Contract was not submitted to a prior control by TdC
		1023	The contract data is prior (earlier) than the adjudication date
		1024	Contracts were implemented before or without being published in the official Base Portal
		1025	Contracts awarded and modified due to at least one amendment (1st phase): contract extensions, increases to contract value (2nd phase)
		1026	Execution length of contract is over 3 years
		1027	Repeat awards to same contractor
		1029	The value of subcontracts is very high
	Intervening parties	1030	The contract was awarded to a contractor with a history of offenses referred to in the CCP *
		1031	The existence of ROCI
		1032	The existence of complaints to the Court regarding the contracting authority
	Payments and financial obligations	1033	Illegal advance payments
	Conflicts of	1034	Lack of contract manager

	interest, fraud and corruption	1035	Involvement of parties that were condemned to financial liabilities and fines *
	Violation of competition law	1037	Institutions with processes involving restrictive competition practices
Inference-based	Evaluation of bidders, tenders and award procedure	1017	Ratio of the estimated value and contractual price (underestimation or overestimation)
		1018	Ratio of the estimated value and base price (underestimation or overestimation)
Model-based	Conflicts of interest, fraud and corruption	1036	Bid rotation

Note: * Indicator was not developed due to data availability.

Annexe B. Data sources used for the TdC data-driven risk model

The framework incorporates both internal and external sources of information. The main database used in the framework is IMPIC's database, while the remaining complement some of the indicators with useful information.

Table B.1. Data sources used for the TdC data-driven risk model

Internal/ External	Database name	Database description
Internal	GDOC (Sistema de Gestão Documental e Processual)	GDOC is the internal system developed by the TdC for managing cases and documents. It includes critical information related to prior control, such as details on contracts, values, types of procedures used, decisions issued, recommendations, follow-ups, and denunciation cases. This system plays a central role in organising and streamlining TdC's document and process management.
	GENT (Sistema de Gestão de Entidades)	GENT serves as the central database for institutions under the jurisdiction of the TdC, encompassing the full range of organisations that report to the Court. It integrates data from various sources, including both structured and unstructured formats. The database is maintained by team that regularly updates it based on assessments of the public gazette. With its 99 tables, GENT assigns a unique identification number to each entity.
	eContas	eContas is a platform integrated within GDOC, enabling public institutions to submit data on public contracts to the TdC. The platform aims to improve communication and efficiency between the TdC and institutions under its jurisdiction by facilitating compliance with information obligations. eContas supports better rationalisation, efficiency, and transparency in the management of public procurement and accounting information.
External	IMPIC (Instituto dos Mercados Públicos, do Imobiliário e da Construção)	IMPIC manages Portal BASE, a public platform that centralises information on public contracts signed across mainland Portugal and the autonomous regions. The portal offers both a public view and a private view, with the latter providing additional information primarily accessed by authorised auditors. The database used was the expanded version of the private view accessed by IMPIC. This database contains information about public procurement procedures and signed contracts (including the object, price, contracting authority, contractor, and contract amendments). The data originates from three main sources: Imprensa Nacional Casa da Moeda (INCM), which provides information from announcements published in the Public Gazette. Contracting Authorities, which directly input data about contracts and procedures. Electronic Public Procurement Platforms, which report information in compliance with legislation.
	Portal Mais Transparência	The Portal Mais Transparência includes information about Portugal 2020[3] and Portugal 2030[4]. (Nevertheless, only Portugal 2020 was included in this framework.) The Portal also contains information on Portugal's Recovery and Resilience Plan. It is managed by AMA in collaboration with the AdC. This database enables the assessment of whether institutions reliably report the financing of European funds.
	Competition Sanctions Data	The AdC provides its activities online, where all cases related to restrictive competition practices, merger control operations, studies, recommendations, and opinions authored by the AdC can be accessed. It also includes access to judicial decisions and all documentation related to AdC cases. The AdC has made information from its database available to the TdC and supplemented the dataset with the company's Tax Identification Number (NIF), facilitating cross-referencing with the information available at the TdC. It is relevant since it contains information about companies suspected or convicted of anti-competitive practices.

Additional data

For some indicators, current data availability issues severely restrict the quality of indicators. To mitigate this, additional datasets were collected to supplement existing data sources. More specifically, this concerns indicators that depend on the CPV value associated with a contract for a sectoral approximation. This has implications for indicators 1017, 1018, 1027, and 1036. Here, data from the open data portal dados.gov.pt was extracted. The data there was published by IMPIC. Of key interest are the contract ID, CPV, and contract signing date (data de celebração) columns. Since the data is available only on a year-by-year basis, the tables were concatenated and used in the indicator functions of the respective indicators. Those files were downloaded as Excel sheets, processed to fit the format of the indicator scripts, and then written out as arrow files. Similarly, indicator 1001 required additional information regarding the start date of a project financed through European Union mechanisms. This data was available through the Portugal2020 data portal. The two datasets available are available for download as either Excel sheets or CSV files. These datasets were then reconciled with the datasets provided through the project ID.

Table B.2. Additional data sources relied upon for the TdC data-driven risk model

Indicator(s)	Content	Source	Process
1001	pt2020_supplementary_data Information about project start dates for projects financed in the context of Portugal2020	https://portugal2020.pt/projetos-aprovados/lista-de-operacoes-aprovadas/	Download via link
1001	react_eu_supplementary_data Information about project start dates for projects financed in the context of Portugal2020	https://portugal2020.pt/projetos-aprovados/lista-de-operacoes-aprovadas/	Download via link
1017 1018 1027 1036	procurement_contracts_2018_2023 Contains information about procurement contracts between 2018 and 2023, particularly the CPV code associated with contracts	https://dados.gov.pt/pt/organizations/impic-i-p-instituto-dos-mercados-publicos-do/#organization-datasets	Download via link

Source: https://portugal2020.pt/ and https://dados.gov.pt

References

Autoridade da Concorrência (2008), *PRC 2005/04.*, https://www.concorrencia.pt/sites/default/files/processos_e_decisoes/prc/decisoes/AdC-PRC_2005_04-Decisao-VNC-final-net.pdf. [27]

Comisión Nacional de los Mercados y la Competencia (2021), *Expediente S/0013/19.*, https://www.cnmc.es/expedientes/s001319-0. [28]

Council of the European Union (2024), *Proposal for a Directive of the European Parliament and of the Council on combating corruption, replacing Council Framework Decision*, https://data.consilium.europa.eu/doc/document/ST-11272-2024-INIT/en/pdf. [25]

DigitalAustria (2024), *Harnessing the Potential of AI*, https://www.digitalaustria.gv.at/eng/strategy/strategy-AI-AIM-AT-2030.html. [24]

ECA (2024), *Artificial Intelligence initial strategy and deployment roadmap 2024-2025*, https://www.eca.europa.eu/ECAPublications/ECA-AI-Strategy-2024-2025/ECA-AI-Strategy-2024-2025_EN.pdf. [15]

ECA (2023), *Public procurement in the EU: Less competition for contracts awarded for works, goods and services in the 10 years up to 2021*, https://www.eca.europa.eu/ECAPublications/SR-2023-28/SR-2023-28_EN.pdf. [34]

European Commission (2025), *Proposal for standard contractual clauses for the procurement of artificial intelligence by public organisations version*, https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/proposal-standard-contractual-clauses-procurement-artificial. [13]

Huber, M. and D. Imhof (2023), "Flagging cartel participants with deep learning based on convolutional neural networks", *International Journal of Industrial Organization*, Vol. 89, p. 102946, https://doi.org/10.1016/j.ijindorg.2023.102946. [31]

INTOSAID (2023), "Pioneering sustainable technology audit practice in SAIs fo better use of technology by government: ID's LOTA pioneers", *International Journal of Government Auditing*, Vol. 50/2, https://www.intosai.org/fileadmin/downloads/about_us/IJGA_Issues/2023/EN_Q2_2023_v50n2.pdf. [23]

OECD (2024), "2023 OECD Digital Government Index: Results and key findings", *OECD Public Governance Policy Papers*, No. 44, OECD Publishing, Paris, https://doi.org/10.1787/1a89ed5e-en. [4]

OECD (2024), *Anti-Corruption and Integrity Outlook 2024*, OECD Publishing, Paris, https://doi.org/10.1787/968587cd-en. [17]

OECD (2024), "Digital public infrastructure for digital governments", *OECD Public Governance Policy Papers*, No. 68, OECD Publishing, Paris, https://doi.org/10.1787/ff525dc8-en. [6]

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en. [14]

OECD (2024), *Recommendation of the Council on Artificial Intelligence*, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449. [9]

OECD (2024), *Strengthening Oversight of the Court of Auditors for Effective Public Procurement in Portugal: Digital Transformation and Data-driven Risk Assessments*, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/35aeab1e-en. [1]

OECD (2024), *The Digital Transformation of Public Procurement in Ireland: A Report on the Current State*, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/87912457-en. [2]

OECD (2023), *Government at a Glance 2023*, OECD Publishing, Paris, https://doi.org/10.1787/3d5c5d31-en. [10]

OECD (2023), *OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem*, OECD Publishing, Paris, https://doi.org/10.1787/c74f03de-en. [3]

OECD (2021), *Recommendation of the Council on Enhancing Access to and Sharing of Data*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463. [32]

OECD (2019), *Declaration on Public Sector Innovation*, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0450. [8]

OECD (2017), *Recommendation of the Council on Public Integrity*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0435. [16]

OECD (2015), *Recommendation of the Council on Public Procurement*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0411. [21]

OECD (2014), *Recommendation of the Council on Digital Government Strategies*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0406. [7]

OECD (2008), *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0362. [33]

OECD (N/A), *Anti-corruption and integrity*, https://www.oecd.org/en/topics/anti-corruption-and-integrity.html. [11]

OECD (N/A), *Digital transformation*, https://www.oecd.org/en/topics/digital-transformation.html. [5]

OECD (N/A), *Integrity in Public Procurement*, https://www.oecd.org/en/topics/integrity-in-public-procurement.html. [20]

OECD (N/A), *Public Procurement*, https://www.oecd.org/en/topics/public-procurement.html. [19]

Renda, A., R. Castro and G. Hernández (2022), “Defining and contextualising regulatory oversight and co-ordination”, *OECD Regulatory Policy Working Papers*, No. 17, OECD Publishing, Paris, https://doi.org/10.1787/a4225b62-en. [12]

Rodríguez, M. (2021), “Collusion detection in public procurement auctions with machine learning algorithms”, *Automation in Construction*, Vol. 133, p. 104047, https://doi.org/10.1016/j.autcon.2021.104047. [29]

State Data Agency (N/A), *Mission and Strategic Objectives*, https://vda.lrv.lt/en/activities/mission-and-strategic-objectives/. [26]

Tribunal de Defensa de la Libre Competencia (2018), *Sentencia N° 165/2018*, https://www.tdlc.cl/wp-content/uploads/sentencias/Sentencia_165_2018.pdf. [30]

Ugale, G. and C. Hall (2024), “Generative AI for anti-corruption and integrity in government: Taking stock of promise, perils and practice”, *OECD Artificial Intelligence Papers*, No. 12, OECD Publishing, Paris, https://doi.org/10.1787/657a185a-en. [18]

World Bank (N/A), *The importance of transparency and integrity in Public Procurement*, https://www.procurementintegrity.org/about. [22]

Notes

[1] https://www.tcontas.pt/pt-pt/Transparencia/PlaneamentoGestao/RelatoriosAnuaisAtividade/Documents/2023/ra2023_estatisticas_indicadores.pdf

[2] https://eur-lex.europa.eu/eli/reg/2008/213/oj/eng

[3] https://transparencia.gov.pt/en/fundos-europeus/pt2020/beneficiarios-projetos/projeto/CENTRO-01-0249-FEDER-043084/

[4] https://transparencia.gov.pt/en/fundos-europeus/pt2030/sobre-portugal-2030/