OECD TECHNICAL PAPER

DEVELOPING A DATA-DRIVEN CORRUPTION RISK MODEL TO STRENGTHEN INTEGRITY IN BELGIUM

PRACTICAL CONSIDERATIONS FOR THE FEDERAL INTERNAL AUDIT

Funded by
the European Union

OECD BETTER POLICIES FOR BETTER LIVES

Developing a data-driven corruption risk model to strengthen integrity in Belgium

Practical considerations for the Federal Internal Audit

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD Member countries.

Photo credits: Cover © estar 2020/Shutterstock.com.

Acknowledgements

This technical paper was prepared under the leadership of Elsa Pilichowski, Director of the Organisation for Economic Co-Operation and Development (OECD) Public Governance Directorate (GOV) and the guidance of Nejla Saula, Head of the Anti-Corruption and Integrity in Government Division (ACIG) and Paulo Magina, acting Head of GOV's Infrastructure and Public Procurement Division (IPP). Helene Wells (ACIG) co-ordinated the development of this paper. The drafting was primarily carried out by Dr Mihály Fazekas (Government Transparency Institute, GTI), Dr Viktoriia Poltoratskaya (GTI), Helene Wells (OECD), and Michel Tierens (FIA). Input and advice into the development of the paper were also provided by Julio Bacio Terracino (ACIG), Paulo Magina (IPP), Jesper Johnsøn (ACIG), Elena Konceviciute (ACIG), Carissa Munro (ACIG), Gavin Ugale (ACIG), Andrea Pallotta (ACIG), András Hlács (IPP), Kenza Khachani (IPP), and Bence Toth (GTI).

This technical paper benefited greatly from insights from Belgium's Federal Internal Audit (FIA) and the Federal Public Service Policy and Support (BOSA). The OECD is grateful for the support and close collaboration of FIA's data analytics department, particularly the contributions of Margot Deviaene, Caroline Zeimes, and Michel Tierens. The Chief Audit of FIA, Kathleen Meganck is thanked for the support provided throughout the project. The OECD is also grateful for the inputs of the following BOSA stakeholders: Roel Arys, Pieter Cuyvers, Sophie Moris, Mira Ratajczak, and Sigrid Strypsteen.

The OECD Secretariat appreciates the support of Angelos Binis, project co-ordinator at the European Commission. The action was funded by the European Union via the Technical Support Instrument, and implemented by the OECD, in cooperation with the European Commission.

Table of contents

Acknowledgements 3

Abbreviations and acronyms 6

Executive summary 7

1 Data is a cornerstone of strengthening public integrity of public procurement processes 8

1.1. The importance of high-quality data in identifying corruption and fraud risks in public procurement 8
1.2. Good practice for effective data pipelines: from data entry to actionable insights 10
1.3. Importance of data governance and data management 13
1.4. Importance of timely data access 14

2 A data-driven approach to understanding corruption risks in Belgium 15

2.1. A data-driven approach to addressing corruption and fraud risks in public procurement in Belgium 15
2.2. Model design 18
2.3. The final validated model 29
2.4. Key insights 35

3 Practical considerations to improve data quality: Opportunities to strengthen data-driven audit risk assessments 39

3.1. Practical recommendations 39
3.2. Future refinement and scalability of the risk model 45

References 49

FIGURES

Figure 1. Number of observations by categories for full data versus lot awarded only 23
Figure 2. Distribution of the single bid variable categories 24
Figure 3. Share of missing values in number of bids per national procedure type 25
Figure 4. Missing rate of selection of key variables by months 26
Figure 5. Corruption Risk Index: Overview of distribution 31
Figure 6. Corruption Risk Index: Overview of weight of components 32
Figure 7. Predicted probabilities of single bidding for individual red flags 34
Figure 8. Corruption Risk Index: Distribution by buyers and suppliers 35
Figure 9. Corruption Risk Index: Distribution by markets 36

Figure 10. Corruption Risk Index: Maps of distribution by territories 37
Figure 11. Distribution of individual red flags' values by months 38

TABLES

Table 1. Key components of the OECD risk management framework in public procurement 10
Table 2. Long list of potential corruption indicators 20
Table 3. Updated list of corruption indicators due to data challenges 28
Table 4. Corruption Risk Index: Validation logistic regression results with single bidding 30
Table 5. Individual red flags with definitions 33

Abbreviations and acronyms

AdC	Competition Authority (Autoridade da Concorrência Portugal)
AI	Artificial intelligence
API	Application Programming Interface
BOSA	Federal Public Service Policy and Support
CPV	Common Procurement Vocabulary
CRI	Corruption Risk Index
ESPD	European Single Procurement Document
FEDAX	Federal Network of Data Experts
FIA	Federal Internal Audit (L'Audit fédéral interne)
GDP	Gross domestic product
GRAS	Governance Risk Assessment System
GTI	Government Transparency Institute
ICT	Information communication and technology
ID	Identification
IMPIC	Institute of Public Procurement, Real Estate, and Construction (Instituto dos Mercados Públicos, do Imobiliário e da Construção)
IT	Information technology
KBO	Cross Road Bank for Enterprise
ML	Machine learning
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Co-operation and Development
SAI	Supreme audit institution
SG Reform	Reform and Investment Task Force (European Commission)
SQL	Structured Query Language
TdC	Court of Auditors (Tribunal de Contas)
TED	Tenders Electronic Daily
URL	Unified resource locator
XML	Extensible Markup Language

Executive summary

This technical paper outlines a data-driven approach for identifying corruption risks in public procurement in Belgium, developed collaboratively with the Federal Internal Audit (FIA) and the Federal Public Service Policy and Support (BOSA). The cornerstone of this approach is the creation and validation of a corruption risk indicator framework using available administrative data. The model seeks to move beyond subjective assessments and anecdotal evidence, instead offering an objective, scalable, and replicable tool to systematically identify high-risk tenders and organisations. This technical paper underscores that data – when comprehensive, timely, accurate, and accessible – is transformational to corruption risk identification. Drawing on international examples and best practice, the paper demonstrates how robust data pipelines can translate raw inputs into actionable insights. In Belgium, the project focused on leveraging newly available public procurement data from the country's recently updated tendering platform, complemented by company registry data.

A key outcome of the project is the development of a risk model for FIA. It is composed of eight validated individual red flags, selected from a longer list of potential indicators based on data quality and relevance. The model uses logistic regression analysis and the established corruption academic literature to validate each risk indicator.

Although data challenges, such as high rates of missing information and a limited observation period, restrict the model's current scope, the findings already reveal meaningful risk patterns across contracts, suppliers, and buyers. Moreover, as more data becomes available over time, the model is expected to improve further, with the potential to incorporate additional indicators.

The paper provides detailed recommendations for improving the quality and utility of public procurement data in Belgium. These recommendations are specific to the implementation of the risk model in Belgium for FIA to consider with other Belgium public sector entity stakeholders. However, the recommendations are of relevance to other jurisdictions required to oversee and ensure integrity in public procurement processes. These include the need for standardised data formatting, ensuring comprehensive data coverage, collaboration among data providers, analysts, and investigators, and better documentation.

Moreover, the paper stresses the importance of building FIA's internal analytical capacity and institutionalising risk-based oversight processes – both of which are important for any oversight institution to consider. Future enhancements of the model developed for FIA could include the use of machine learning (ML) models once enough proven corruption cases are collected, as well as the automation of data preparation and validation steps to facilitate scale-up.

The model offers a promising foundation for proactive corruption risk assessment and oversight of Belgium's public procurement. It represents an important step toward embedding data analytics in everyday oversight functions, enabling targeted assessments and investigations, and enhancing transparency.

1 Data is a cornerstone of strengthening public integrity of public procurement processes

Most data-driven approaches to corruption assessment in public procurement focuses on identifying patterns of behaviour that signal elevated corruption risks, rather than detecting corruption directly (Ayogu and Graffy, 2022[1]). The direct detection of corruption often requires whistleblower reports or allegations, or legal investigations – activities which are outside the scope of most data analytics tools. While data-driven approaches to corruption assessments do not necessarily directly prove wrongdoing, it can highlight transactions, actors, or procedures that deviate from expected norms and warrant further assessment and investigation. This is possible because fraud, corruption, and collusion often manifest through recurring behaviours – such as unusual wins by certain suppliers, persistently low competition in bidding, or a buyer's preference for certain suppliers in contract allocation. By systematically analysing these patterns, data can help uncover high-risk areas, enabling targeted oversight, preventive action, and more efficient use of finite investigative resources. For this, the integration of multiple data sources (e.g., financial disclosures, social networks, and procurement data) can strengthen the inference of potentially corrupt behaviour.

The following section highlights the importance of high-quality data when identifying and assessing corruption, fraud, and integrity risks. Good practice examples for the development of risk indicators and effective data pipelines are provided. Data governance and data management practices are crucial in the development of a model that relies on quality data. All of this has been critical to the development of the risk model for Belgium's FIA (this project is described in more detail in Chapter 2).

1.1. The importance of high-quality data in identifying corruption and fraud risks in public procurement

While this section focuses on the importance of data quality, the availability and accessibility of data remains just as fundamental in its strategic use in preventing and investigating corruption. This is not only important for robust integrity assessments, but also for broader analytical public spending efficiency and enhancing transparency in procurement systems.

In line with the Integrity Principle of the OECD Recommendation on Public Procurement (OECD, 2015[2]), which calls for high standards of transparency, good management, and accountability throughout the procurement cycle, reliable, detailed, and comprehensive data is indispensable for tracking decision-making processes and identifying suspicious bidding patterns. Accurate data enables researchers, civil society actors, and oversight institutions to better detect corruption risks. It also allows for comparisons over time and across sectors, as well as for granular analysis of different actors, making it possible to uncover anomalies that may point to systemic issues. In contrast, poor data quality characterised by missing values, inconsistent formats, or limited scope can obscure corruption risks, mislead analysis, or even conceal wrongdoing altogether. It is not only investigative authorities that suffer from poor data quality, but also civil society, researchers, and private actors who seek to monitor public agencies or companies.

To ensure the effectiveness of corruption risk analysis in public procurement, attention must be paid to four key dimensions of the data quality – scope, depth, accuracy, and accessibility (Williams and Tillipman, 2024[3]):

- **Scope** refers to how comprehensively the data captures relevant transactions, including all buyers, suppliers, and markets. Missing values of certain buyers or markets can distort the risk assessment.
- **Depth** concerns the level of detail available in each entry, such as bidder information, contract amounts, and dates. Richer datasets allow for more precise risk indicators and enable the possibility to link external data sources.
- **Accuracy** assesses the degree of correspondence between dataset records and actual actor behaviour, where missing or misleading values can distort findings and hinder trust in results.
- **Accessibility** ensures the data can be obtained and processed efficiently, which depends on format, standardisation, and the existence of centralised sources.

When a dataset does not do well on all four dimensions simultaneously, it limits the ability to conduct timely, reliable, and actionable corruption risk assessment and monitoring.

The OECD's *Managing Risks in the Public Procurement of Goods, Services and Infrastructure* publication emphasises data quality as a foundational element for effective risk management in public procurement (OECD, 2023[4]). The report outlines several key dimensions of data quality that are critical for enabling robust oversight and informed decision-making. These include:

- **Accuracy**, ensuring that procurement data correctly reflects the reality of transactions.
- **Completeness**, which refers to the inclusion of all necessary data points across the procurement lifecycle.
- **Timeliness**, meaning data is available when needed to support proactive risk identification and mitigation.
- **Consistency**, which ensures that data is harmonised across systems and reporting periods.

Table 1 presents the key components of the risk framework. High-quality data enables governments to better understand procurement risks, monitor contract performance, and assess the outcomes of complex procurements, particularly when system integrators are involved, and the technology stack is not directly visible through vendor-specific contracts.

Table 1. Key components of the OECD risk management framework in public procurement

Key component	Description
Risk identification	Systematically identifying potential risks across the procurement cycle, including compliance, operational, financial, and reputational risks.
Risk assessment	Evaluating the likelihood and impact of identified risks using tools like risk matrices and risk registers.
Risk treatment	Developing strategies to mitigate, transfer, accept, or avoid risks. This includes contractual safeguards, supplier vetting, and contingency planning.
Risk monitoring	Continuously tracking risk indicators and procurement performance to detect emerging risks and ensure mitigation measures are effective.
Governance and accountability	Establishing clear roles, responsibilities, and oversight mechanisms to ensure risk management is embedded in procurement practices.
Capacity building	Training procurement officials and stakeholders in risk management tools and techniques to enhance institutional resilience.
Use of technology	Leveraging digital tools and data analytics to improve risk detection, transparency, and decision-making.
Tailored approaches	Adapting risk management strategies to the complexity and nature of the procurement (e.g., infrastructure vs. routine goods).
National strategy alignment	Integrating procurement risk management into broader national risk and public governance strategies.

Source: (OECD, 2023[4])

The interoperability of public procurement data is also important. Effective Information and Communication Technology (ICT) procurement – especially in the context of digital transformation – requires coordinated, whole-of-government approaches supported by interoperable systems and data. The OECD publication *Towards Agile ICT Procurement in the Slovak Republic* highlights that (OECD, 2022[5]):

- Fragmented procurement systems and a lack of interoperability between government entities hinder the ability to align ICT investments with national digital strategies.
- There is a need for standardised data formats and shared platforms to enable better tracking, monitoring, and analysis of procurement activities.
- Interoperability is essential for avoiding vendor lock-in, improving transparency, and enabling cross-agency collaboration, particularly when dealing with complex ICT projects involving system integrators rather than direct product vendors.

The ability to link procurement data with other data sources such as company registries, asset declarations, or political donation registers significantly increases its analytical value. When datasets are wide-scoped, granular, and accessible, they support the development of automated risk indicators and predictive models, enabling institutions to concentrate their resources to responding to the highest risks. This is crucial in contexts where manual investigations are constrained by limited capacity. For instance, in Portugal, the Court of Auditors (*Tribunal de Contas*) has recently launched an initiative aiming to strengthen its use of data and advanced analytics for assessing risks in public procurement, which included thorough data-mapping exercise and data reliability assessment, with the support of the OECD (OECD, 2024[6]). Furthermore, high-quality data not only improves oversight but can also drive broader reform by fostering transparency, boosting competition, and empowering citizens and businesses to hold public institutions accountable. Data is not merely a technical resource – it is a foundational tool for promoting accountability, transparency, and integrity.

1.2. Good practice for effective data pipelines: from data entry to actionable insights

There are multiple examples of good practice when it comes to the development of corruption risk indicators based on good quality data and stable data pipelines. These include the Governance Risk

Assessment System (GRAS), ProZorro, the Red Flag tool, and Portal BASE. These approaches exemplify effective data pipelines allowing for good quality data to be routinely analysed to create actionable insights.

1.2.1. Governance Risk Assessment System (GRAS)

In today's data-rich governance environments, indicator frameworks like GRAS offer a powerful example of how well-designed data pipelines can move from raw information to practical, real-world decision-making (World Bank, 2023[7]). GRAS, developed by the World Bank, illustrates how governments can harness complex data to identify and mitigate corruption risks in public spending. The system was first developed in Brazil, piloting in a few subnational governments, where it helped identify hundreds of cases of collusive behaviour, conflict of interests, and connections to political campaigns.

GRAS begins by pulling together a wide array of publicly available datasets – ranging from procurement contracts to corporate records and campaign finance data – into a single, unified system. This diverse data is cleaned, structured, and linked to allow for a deep analysis of behaviour patterns. GRAS scans for four risk groups:

- Procurement cycle (e.g., non-competitive processes);
- Collusion (e.g., number of competitors with common shareholder);
- Supplier's characteristics (e.g., sanctioned supplier or supplier registered in tax haven); and
- Political connections (e.g., personal connections of company to politicians).

By flagging the risks in these groups, the system does not prove the wrongdoing but helps spotlight potentially suspicious actors or transactions that deserve further scrutiny.

The insights generated by GRAS are presented through an intuitive interface that allows users, such as auditors or investigators, to explore red flags across multiple dimensions, adjust filters based on local context, and prioritise high-risk cases for review. By enabling targeted investigations and proactive oversight, GRAS shows how a thoughtful data pipeline from collection and integration of datasets to risk detection and user interaction can be a cornerstone of evidence-based public sector accountability.

1.2.2. ProZorro

Another illustrative case of an effective data pipeline supporting anti-corruption efforts is Ukraine's ProZorro[1] electronic procurement system. Designed as a collaborative initiative involving government agencies, civil society, and private sector actors, ProZorro demonstrates how transparent data flows can strengthen both bottom-up (citizen-led) and top-down (state-led) accountability. The platform was not only intended to publish procurement data but also to reform the entire procurement process by making it more open, decentralised, and resistant to manipulation. Initially piloted by a handful of ministries, its success in generating cost savings and reducing corruption risks attracted broader use by suppliers and public institutions, eventually leading to strong political support and widespread adoption.

ProZorro's data pipeline integrates real-time procurement data collection, public access through an open interface, and embedded monitoring tools. Over time, it has evolved into a system that not only publishes structured data but also enables automated detection of corruption risks. Legislative reforms followed its implementation, including the 2017 introduction of risk-based monitoring procedures, developed jointly by public officials and civil society experts. As the platform matured, its ability to tackle everyday corruption laid the foundation for more ambitious reforms aimed at addressing grand corruption.

[1] https://prozorro.gov.ua/

1.2.3. Red Flags Tool

Hungary's Red Flags[2] tool is an open-source tool designed to automatically check documents from the Tenders Electronic Daily (TED). The tool flags risky procurement and was developed to include 40 indicators (9 indicators focus on contract award notices and 31 focus on contract notices). An example of contract notices red flags includes: i) framework agreements with a tenderer (because large framework agreements can potentially exclude competition over the longer term), ii) omission of the definition of compulsory grounds for exclusion, and iii) time limit (short) for tendering/participation. Procedures without prior publication, the duration of the evaluation, and a low number of tenders received are examples of contract award notices indicators. Using notices from TED enables the tool to be stable because the data pipeline is reliable. It is easily adaptable and flexible for future requirements (Transparency International Hungary, 2015 (Németh and Tátrai, 2015[8]).

1.2.4. Portal BASE

In Portugal, the Institute of Public Procurement, Real Estate, and Construction – IMPIC (Instituto dos Mercados Públicos, do Imobiliário e da Construção) – a public institute belonging to the indirect administration of the State, is the national public procurement regulator with several attributions related to regulatory, monitoring, and professionalisation functions. In terms of monitoring, IMPIC is managing the public procurement Portal BASE[3] that includes key public procurement information. This database contains information about public procurement procedures and signed contracts (including the object, price, contracting authority, contractor, and contract amendments). BASE centralises the information on public procurement contracts. Platform providers are mandated to send additional data to BASE. Overtime, the improvements to data quality and the availability of data in BASE and its structured format and could support the calculation of additional public procurement risks. For instance, platforms providers will have to send links to the preliminary and final procurement reports. They will also share questions from bidders challenging the ranking of bids, and the Nomenclature of Territorial Units for Statistics (NUTS) code for the execution of contracts (previously the location was not as accurate) (OECD, 2024[6]).

In the framework of the Technical Support Instrument (TSI)[4] project focusing on strengthening the control framework in Portugal, the OECD, together with NOVA University, has recently developed a data-driven risk model for Portugal's Supreme Audit Institution (SAI), the Tribunal de Contas (TdC; Court of Auditors) as described in Box 1.

[2] https://www.redflags.eu/

[3] https://www.base.gov.pt/

[4] https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/technical-support-instrument_en

Box 1. Using digital technology to strengthen oversight of public procurement in Portugal

The initiative aimed to improve the TdC's identification of risks and the early detection of irregularities through advanced data analysis and machine learning (ML). The development of the risk methodology marks a significant milestone in the TdC's digital transformation journey.

The 37 risk indicators developed in the framework of the project include a mixture of rule-based (red flags based for simple rule violations, such as "no competition in a high-value contract"), inference-based (red flags based on patterns or repeated behaviour, such as "the same company always wins"), and model-based indicators (red flags found by smart systems that learn from past data to spot unusual activity). BASE was an important data source for the development of the risk model (the model relied on the expanded version of the private view accessed by IMPIC, amongst other data sets).

Source: (Hlacs and Wells, 2025[9]).

1.3. Importance of data governance and data management

Well-defined data governance and data management are crucial in the development of a model that relies on good-quality data (see for example (OECD, 2025[10]).[5] Apart from people, data is one of the most valuable assets to any public sector entity, and as such, coverage of the entire data lifecycle needs to be taken into consideration. Without a robust data governance framework, an entity may face issues such as data inaccuracies, regulatory non-compliance, and inefficient model building, which can result in reputational damage, financial losses, inefficiencies, and missed opportunities. Good data governance is built on core principles such as accountability, transparency, integrity, compliance, and standardisation, ensuring that data is treated as an asset. From the creation of data to its use, archiving, and deletion, clear policies and procedures are required to ensure data is managed effectively.

The OECD's Recommendation of the Council on Enhancing Access to and Sharing of Data encourages member countries to facilitate broader access to and sharing of data across sectors (OECD, 2021[11]). The sharing of data and enabling better data access aims to harness already existing data sources as a way of fostering innovation and to build data-driven models for decision making. Furthermore, the OECD Recommendation on Public Procurement recommends that governments pursue state-of-the-art e-procurement tools that are modular, flexible, scalable, and – among other things – protect sensitive data, while supplying the core capabilities and functions that allow business innovation (OECD, 2015[2]). Also, the OECD's Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information guides entities on why data access is critical (OECD, 2008[12]). Public sector information should be available for use and re-use by default for oversight and integrity institutions. In addition, non-discriminatory, competitive access to public data, such as procurement and contract-related data, aims to remove unnecessary restrictions to accessing data.

[5] See also https://www.oecd.org/en/topics/data-governance.html

1.4. Importance of timely data access

Regular updates of the data used for risk assessment model building are crucial for timely identification of risks and ensuring that monitoring helps prevent corruption ex ante rather than merely identifying it ex post (Bauhr et al., 2019[13]). This is particularly relevant in cases where corruption schemes are adaptive and evolve over time. For instance, when public officials or suppliers learn to exploit regulatory loopholes or respond to existing risk indicators by masking their behaviour. If data feeding into corruption risk models is outdated, the models may fail to capture these new patterns, resulting in blind spots that allow suspicious corrupt behaviour to go undetected.

Timely access to data also enables dynamic risk scoring, where red flags can be recalculated as new information becomes available. For example, when new tenders are announced, bidder histories are updated, or links to politically exposed persons emerge. This real-time (or near-real-time) feedback loop allows oversight bodies, auditors, and civil society actors to intervene early, ideally before contracts are awarded. In contexts with limited investigative capacity, such prioritisation is not only efficient but essential. Therefore, maintaining a continuous, timely data flow is a foundational requirement for ensuring that corruption risk models remain relevant, accurate, and actionable, as emphasised in the OECD's framework for managing risks in public procurement (OECD, 2023[4]) and in the OECD's *Using Digital Technology to Strengthen Oversight of Public Procurement in Portugal: The use of data analytics and machine learning by the Tribunal de Contas* (Hlacs and Wells, 2025[9]).

It should be noted that the ambition to implement regular data updates and a real-time feedback loop may raise concerns in the Belgian federal context, given the challenges encountered during the data collection phase of the project (as explained in Chapter 2). Difficulties related to fragmented data ownership, inconsistent data formats, and delays in access have demonstrated that achieving timely and reliable data flows is not straightforward. These structural issues must be addressed to ensure that the envisioned dynamic monitoring processes for FIA are both feasible and sustainable.

2 A data-driven approach to understanding corruption risks in Belgium

Public procurement represents a sizable portion of the Belgium's gross domestic product (GDP). In 2023, it accounted for 14.9% of GDP, which is higher than the OECD average of 13% (OECD, 2025[14]). Governments, including Belgium, can pursue excellence in efficiency gains by continuously developing, implementing, evaluating, and revising procurement systems, including those systems used to collect data for monitoring, evaluation, and audit purposes. Public procurement strategies and practices, and the systems created to assist with public procurement, have an impact on the quality of life and wellbeing of citizens (for example, health, education, transportation, and infrastructure). Broader government objectives, such as innovation or sustainability, can be achieved when public procurement is used strategically. It is incumbent on governments to ensure value for money in public procurement processes, and excellent stewardship of large amounts of taxpayers' money is critical (European Court of Auditors, 2023[15]).

This chapter details the context for which the OECD developed a data-driven risk model to assist FIA to better identify and assess corruption risks related to public procurement in Belgium. The legislative and strategic context is provided, including the recent introduction of Belgium's new e-Procurement platform. The risk model developed with FIA is explored in detail, including the risk factors and indicators that constitute the Corruption Risk Index (CRI) that is foundational to the audit risk model. Whilst this chapter is technical in nature and specific to the model developed for FIA (for example, the explanation of the logistic regression analyses), the methodology and findings are useful to various domains within Belgium's public administration, as well as other oversight entities in different jurisdictional settings.

2.1. A data-driven approach to addressing corruption and fraud risks in public procurement in Belgium

2.1.1. Legislative and strategic context

Due to the multitude of integrity actors and applicable regulations, the overall coherence and effectiveness of integrity and anti-corruption policies of the Belgian authorities face specific challenges, both within the federal administration and between the federal and regional levels. This has led to a lack of clarity and visibility of integrity and anti-corruption policies, as well as implementation gaps and loopholes in legislation. These challenges have been identified throughout external evaluations by the European Commission (European Commission, 2023[16]) and the Group of States against Corruption (GRECO, 2020[17]) as well as the OECD's latest evaluation of Belgium's implementation of the OECD Anti-Bribery Convention (OECD, 2025[18]).

Belgium has taken significant steps to strengthen its integrity framework. The Minister of Civil Service, together with the Minister of State and Budget, adopted in 2023 the Royal Decree on integrity policy and management within the public sector. With this decree:

- The Unit for Integrity and Culture within the Federal Public Service Policy and Support (FPS BOSA) was transformed into the more autonomous and centralised Integrity Bureau.
- The role of integrity coordinators within federal public organisations was mandated.
- The Federal Network of Integrity Coordinator was formalised.
- All federal public organisations were mandated to define integrity objectives and annually report on their implementation.

The Minister of Civil Service together with the Federal Public Service Policy and Support has already adopted several initiatives to establish a coherent and comprehensive integrity system revolving around a risk-based approach to integrity. Other integrity elements, such as whistleblower protection, have been strengthened in Belgium with the transposition of the European Union (EU) whistleblower directive into a national law on 8 December 2022 and the adoption of a Royal Decree on internal reporting channels in 2023. Federal public sector organisations continue to work on implementing internal and external reporting channels for receiving and tracking reports.

In recent years, integrity-related legislation been implemented in Belgium, including:

- **Royal Decree of 20 October 2023** on the elements of the procedures and follow-up of internal reports, the purpose and content of the archiving of reports, and the modalities of public consultation, referred to in several Articles of the Act of 8 December 2022. The Decree includes the obligation for each federal public authority to establish an internal reporting channel. Internal reporting channels should be managed internally within the federal public authority by a designated person or department, or externally either by a third party or by the Federal Audit. The responsibility for implementing the Decree lies with the respective ministers (Federal Overheidsdienst Justitie, 2016[19]).
- **Royal Decree of 18 April 2023** (Service public fédéral Stratégie et Appui, 2023[20]) on integrity policy and management within the public sector. The Royal Decree includes the creation of a more autonomous and central Integrity Bureau, the institutionalisation of the role of integrity coordinators, of "facilitator" and of the Federal Network of Integrity Coordinators. Additionally, it includes the requirement for federal public organisations to define integrity objectives, corresponding indicators in the organisation's strategic plan, develop a corresponding annual integrity management action plan to support implementation and report yearly on the integrity management of federal public sector organisations.
- **The Act of 8 December 2022** concerning the reporting channels and the protection of persons reporting integrity violations in the federal public authorities and in the integrated police. The Act entered into force on 3 January 2023. It applies to federal public authorities, which includes federal administrative authorities (e.g., Federal Public Services, Federal Programmatory Services and autonomous public enterprises), policy-making bodies, and any other agencies or services that depend on the federal government and do not belong to the private sector, such as Federal Ombudsmen and the Data Protection Authority.

The 2017 Public Procurement Law significantly contributed to strengthening Belgium's integrity framework. It introduced clearer rules and procedures for public procurement, which reduced opportunities for corruption and favouritism. The 2017 reform also transposed EU Directives 2014/24/EU and 2014/25/EU into Belgian law, embedding principles such as equal treatment, non-discrimination, and transparency. In addition, it mandated the use of e-procurement platforms, which increased transparency in the tendering processes. Furthermore, in line with OECD recommendations, Belgium's procurement law incorporated risk-based approaches to identify and mitigate integrity risks such as fraud, collusion, and undue influence throughout the procurement cycle (Bataille and Dor, 2017[21]).

In Belgium, public procurement is regulated at the federal level by a procurement law. Each region has a certain level of flexibility for interpreting and implementing the legislation. Belgium is a federal state with decentralised authority shared among the central government and the three regions – Wallonia, Flanders, and the Brussels-Capital Region. This decentralised authority is applicable to procurement, and it is within this context that the e-Procurement platform in Belgium was developed.

2.1.2. The e-Procurement platform in Belgium

Since September 2023, the e-Procurement[6] platform in Belgium facilitates the entire procurement process. It serves as a central hub for electronic procurement activities and essentially manages the procurement lifecycle – from publication of tender notices to the submission of bids and then the evaluation of bids. It is open and accessible, thus providing transparency in the public procurement processes. It is also a tool that helps enable the detection and hopeful prevention of irregular activities associated with public procurement. The platform provides a secure and transparent environment for the submission of bids, preventing unauthorised access and tampering. The platform supports the use of the European Single Procurement Document (ESPD), which simplifies the qualification process for suppliers. It ensures that only eligible and compliant bidders participate in the tendering process.

The development of the e-Procurement platform has its history in the Council of Ministers approving in late 2020 the proposal to start developing a new platform. BOSA started developing the new e-Procurement system to replace what had been in existence since about 2006. It should be noted that the data relevant for the development of the risk model for FIA only became available following the launch of the new platform in late 2023. As a result, the volume of data available for the development of the model has been inherently limited.

The new Belgian procurement tool from FPS BOSA allows for the measurement of the number of public procurement procedures. The tool allows for conducting monitoring exercises, yet such monitoring is not conducted systematically, but mostly based on specific or ad hoc requests from other public entities. The tool covers all types of procurement procedures that fall under the main EU directives on public procurement and utilities (2014/24/EU, 2014/25/EU, 2014/23/EU, and 2009/81/EC) (European Commission, 2024[22]). Contracting entities and contracting authorities are by law obligated to publish information on the e-Procurement platform. The tool does not enable the monitoring of innovation procurement expenditure across Belgium comprehensively or systematically (European Commission, 2024[22]).

2.1.3. Leveraging digital transformation to enable improved oversight of procurement

The digital transformation of oversight and integrity institutions is crucial for enhancing transparency, efficiency, and accountability in public procurement and public funds. The OECD’s 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[23]). The OECD’s *Anti-Corruption and Integrity Outlook 2024* highlights the need for integrity frameworks to be updated and remain dynamic to address evolving corruption risks, including those related to digital technologies (OECD, 2024[24]). Collaboration, sharing, and access to data across multiple institutions require stakeholders to be identified early, and proactively and routinely engaged during the development of a risk model.

Public procurement systems and tools should be created in a way that maximises efficiencies whilst also being able to monitor, reveal, and understand possible corruption and integrity risks. The lack of integration

[6] https://bosa.belgium.be/fr/applications/e-procurement

of digital tools governing Belgium's public procurement has led to issues with fragmented datasets and datasets with high degrees of missing data. Because of this, it has been challenging for Belgian authorities, foremost for the FIA, to establish robust indicators of possible corruption and fraud risk. This, in turn, weakens the opportunity to conduct robust risk analyses.

The OECD Recommendation on Public Procurement emphasises the need for countries to preserve the integrity of their procurement systems through safeguards and standards, as well as the importance of collecting consistent, up-to-date, and reliable data and information (OECD, 2015[2]). It advocates the need for measures to prevent corruption, fraud, and mismanagement of public funds. Integrating digital technologies into the oversight of public procurement is one way of improving the monitoring and analysis of possible indicators of corruption (OECD, 2024[6]) The adoption of e-procurement systems, as realised in Belgium with its new e-Procurement system, aims to streamline procurement processes, reduce administrative burdens, and provide real-time data for better decision-making.

Outside of Belgium, one example of the application of digital transformation and integration of data mining to better understanding procurement risks is the development of the ARACHNE Information Technology (IT) tool. ARACHNE was developed by the European Commission. ARACHNE establishes a database of projects that have been implemented under the Structural Funds in the EU. It aims to enrich the available data with publicly available information to identify, based on a set of risk indicators, the projects, contracts, beneficiaries, and contractors that may present risks of fraud, conflict of interest, and other irregularities. It is an example of how data mining, data analytics, and the identification of risk indicators (there are over 100 risk indicators in ARACHNE) can assist with decision making and the oversight of verifying potential irregularities (European Commission, n.d.[25]).

2.2. Model design

2.2.1. Objectives and context for the development of the risk model

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) for the initiative – *Strengthening the strategic approach to public integrity in Belgium, including the integrity of public procurement processes and data-driven approach in procurement risk management*. One element of the initiative was the requirement for the OECD to develop of a data-driven risk model that can assist FIA to better identify and assess corruption and fraud risks related to public procurement.

The initial intention was to utilise artificial intelligence (AI) to develop a risk model that addresses fraud and corruption risks in public procurement. Due to the substantial data quality issues (including a high proportion of missing data), AI and specifically the use of ML could not be applied in this current iteration of the model. Nor could some of the intended indicators relating to fraud and corruption be included due to data quality issues as well as the need to work with datasets with a limited timeseries. To this extent, the proposed risk model is primarily descriptive rather than predictive in nature. Nevertheless, the model has been designed to enable incremental changes and scalability, as data quality improves, and as the timeseries of data increases.

Whilst undertaking the initiative, the OECD and FIA reviewed and assessed 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 the proof-of-concept model. Guidance and capacity building have also been undertaken with relevant FIA stakeholders, including hands-on technical workshops with the risk model and accompanying visual dashboard. The development of the risk model has taken into consideration the good practice approaches as highlighted in section 1.2. Within the context of the development of the risk model, FIA has also been prioritising an audit on data governance within the federal administration of Belgium as described in Box 2.

Box 2. FIA's audit on data governance within the Belgium federal administration

FIA is undertaking a cross-cutting audit on data governance within the federal administration of Belgium. This initiative is designed to evaluate the maturity and coherence of data governance practices across the various institutions that make up the FIA audit universe. The objective is to identify strengths, highlight areas for improvement, and formulate tailored recommendations to help institutions build or enhance their data governance frameworks. Structured as an awareness-raising and advisory exercise (comparable to a "quick scan") this engagement is intended as a foundational step rather than a full procedural compliance audit. It aims to support institutions in managing data more effectively across its full lifecycle, in alignment with strategic priorities and regulatory expectations.

Source: FIA

The effective use of data and digital tools presents an opportunity for FIA to strengthen its oversight and achieve value for money by being able to undertake more detailed reviews of possible procurement processes that may have red flags for corruption (as informed by the model). The development of this model represents a step forward in how FIA can conduct risk assessments. Once implemented and further reviewed, the model can enable FIA to routinely review and examine data associated with potential risks, with the aim of enhancing audit planning and, in due course, fostering more proactive approaches to fraud auditing.

2.2.2. Developing a Corruption Risk Index

The data-driven model is informed by a set of predefined risk factors and the construction of a Corruption Risk Index (CRI). Corruption in public procurement typically aims to channel contracts to favoured bidders while avoiding detection (International Bank for Reconstruction and Development/The World Bank, 2013[26]). This can be achieved in many ways, such as bypassing competition through unjustified sole sourcing or direct contract awards, or by manipulating the process to benefit a particular firm (for example, by tailoring technical specifications, sharing inside information, or conducting biased evaluations).

There are three broad approaches to measuring this type of corrupt behaviour, each with distinct strengths and limitations. The first relies on corruption perception indices to estimate the scale and scope of corruption. Although widely used, these indices have been subject to extensive critique for their subjectivity and limited policy relevance, as they often fail to capture specific corrupt practices at the transaction level (Lambsdorff, 2006[27]); (Johnston, 2017[28]) (Andersson and Heywood, 2009[29]) and (OECD, 2008[12]).

A second, more data-driven method involves analysing proven corruption cases using ML techniques. By learning from patterns in known cases, researchers can develop predictive models that identify high-risk tenders or suppliers (Fazekas, Sberna and Vannucci, 2021[30])While promising, this approach is constrained by the availability and representativeness of confirmed cases, which can reflect enforcement biases or selective prosecution.

Given the existence of data challenges, the most reliable approach in many contexts is to identify objective risk factors, test their statistical validity against established, high corruption risk patterns, and combine the validated indicators into a CRI. This method does not depend on subjective perceptions or hard-to-obtain data on proven cases, making it both scalable and replicable across different procurement systems (Fazekas, Tóth and King, 2016[31]). The composite CRI is a valuable tool for identifying public procurement contracts that exhibit multiple red flags, enabling investigators to focus on a smaller, high-risk subset of cases. In addition to contract-level analysis, both the composite score and individual red flags can be aggregated at the level of contracting authorities, suppliers, or geographic regions. These aggregated results can be visualised through heatmaps and dashboards, enhancing usability and enabling more effective risk monitoring and decision-making. This, in turn, allows for a broader assessment of corruption risks across institutions, private entities, or territories, thereby supporting more strategic oversight and informed resource allocation.

The development of the model follows a structured, evidence-based approach. The first step involves compiling a long list of potential corruption indicators, drawing on existing academic literature, relevant policy publications, and an initial assessment of available data fields (see Table 2). While the availability of a data field suggests the possibility of calculating an indicator, it does not guarantee that the indicator is fit for use. Certain limitations, such as lack of variation in values, a short time series, or highly skewed distributions may affect the indicator's relevance or statistical validity. In such cases, indicators may either be deferred until more complete data become available or excluded altogether. Following this initial assessment, the long list is refined based on criteria including data quality, variability, and time coverage. Only those indicators deemed feasible are included in further analysis and model development (see Table 3).

Table 2. Long list of potential corruption indicators

Indicator	Description
Public procurement indicators	
Single bidding	Indicates that a given tender only had one bidder during the procurement process, hence there was no competition for the tender (Fazekas et al., 2016).
Buyer spending concentration	A supplier's share in a buyer's total spending in a given year can be used as a measure of market competitiveness and openness. A high share of supplier spending can signal that a supplier or a group of suppliers are part of a network, potentially leading to higher prices, and/or lower quality and value for money (Fazekas et al., 2016).
Advertisement period	A sufficiently short advertisement period makes competition impossible, because competitors will not have the time to obtain necessary documents, prepare the tender documentation, or to calculate their expenses to prepare their bids (Fazekas et al., 2021).
Decision period	Snap decisions may reflect premeditated assessment, while long decision periods may signal extensive legal challenges to the tender, suggesting that the issuer attempted to limit competition (Fazekas et al., 2016).
No call for tender	When no call for tender is published at all, but instead it is informally sent to selected bidders, the principle of transparency is violated most extensively (Decarolis and Giorgiantonio 2022).
Procedure type	Using procedure types which are less transparent and require less open competition can indicate the deliberate limitation of the range of bids received and to exclude bids as well as creating more opportunities for contracting bodies to repeatedly award tenders to the same well-connected company (Auriol et al. 2016, Chong et al. 2016).
Distinct markets	Company specialising in one market and winning a contract in a significantly different sector can indicate potential corruption, and reflects an unusual pattern of supplier behaviour, as it suggests the company may lack the expertise, experience, or capacity to perform the contracted work, raising questions about the legitimacy of the procurement process (CEPR, n.d.).
Tax Haven	Awarding public tenders to companies registered in tax havens presents a risk that anonymous company ownership could be concealing a conflict of interest in the award of a tender to a politically connected beneficial owner (Fazekas et al., 2021).
Local supplier	When buyers favour local suppliers, this could be due to conflicts of interest, favouritism, or political clientelism, often bypassing fair competition and transparency. Close local ties can enable kickbacks, collusion, and biased decision-making

	(Mamavi et al. 2014).
Non-transparent buyer	Tender transparency, especially the one allowing for horizontal monitoring, reduces corruption risks substantially (Bauhr et al. 2020). At the buyer level, non-disclosure of certain information may result from accidental errors or simple non-compliance rather than intentional concealment. However, the failure to publish relevant information, including key contract details, can significantly impact oversight and, in turn, increase the risk of corruption.
Company and public procurement indicators	
Dissolution of company after winning the contract	Rapid dissolution of a company after winning a public procurement contract could be a sign of corruption, as it may indicate that the company was created solely as a shell entity to win the contract. This tactic is often used to obscure the identities of beneficiaries, avoid accountability, and hinder investigations (Mironov and Zhuravskaya 2016).
Number of companies at the same address	Multiple companies sharing an address could conceal dubious links between firms, which often serve as fronts with no real operations, with the real work outsourced (Caneppele et al. 2009).
Company age	The date of establishment can further signal corrupt intent, especially if the company was set up just before securing important contracts, suggesting that the firm is being used to extract rent rather than for legitimate business purposes (Fazekas et al. 2015).
Controls	Description
Contract value	Size of the contract can drive the number of bidders participating in the tender.
Buyer type	Central, regional or local authorities, as well as European Union (EU) versus national bodies might have different patterns and rules regulating their activities and impacting competitiveness in tenders.
Tender supply type	Whether it involves supplies, works, or services can affect the number of competitors in a given market.
Tender year	Due to external factors such as the economic environment, global conflicts, or pandemics, the behaviour of companies bidding on tenders can change.
Market identification (ID)	Some markets naturally tend to have more competition on average than others, depending on the national context.
Buyer location	Some provinces may have more or less dense economic networks and levels of company participation, which can affect the number of bidders.

Source: (Fazekas, Tóth and King, 2016[31]) Fazekas et al 2024 (Decarolis and Giorgiantonio, 2022[32]) (Auriol, Straub and Flochel, 2016[33]) ((CEPR), CENTRE FOR ECONOMIC POLICY RESEARCH, 2021[34]) (Caneppele, Calderoni and Martocchia, 2009[35]; Bauhr et al., 2019[13]) (Acar et al., 2015[36]) (Fazekas, Lukács and Tóth, 2015[37]) (Mamavi et al., 2014[38])

2.2.3. Data-related challenges

To build a risk model and develop corresponding corruption risk indicators, two types of data were requested: public procurement data from BOSA and data from the Cross Road Bank for Enterprises (KBO) which is managed by the Federal Public Service Economy. Two main categories of issues with the obtained data were identified: i) structural, and ii) scope and availability. Structural issues concern how the data is stored and extracted, while scope and availability-related issues pertain to the actual availability and completeness of the information. These two dimensions are closely interconnected. For instance, since contract data in Belgium is entered by public entities themselves, missing information in the final dataset often points to shortcomings in how data is provided or recorded at the source. At the same time, some data gaps could be mitigated by enhancing how data are stored or by improving the procedures used to export data, in order to reduce the risk of information being lost or omitted during processing.

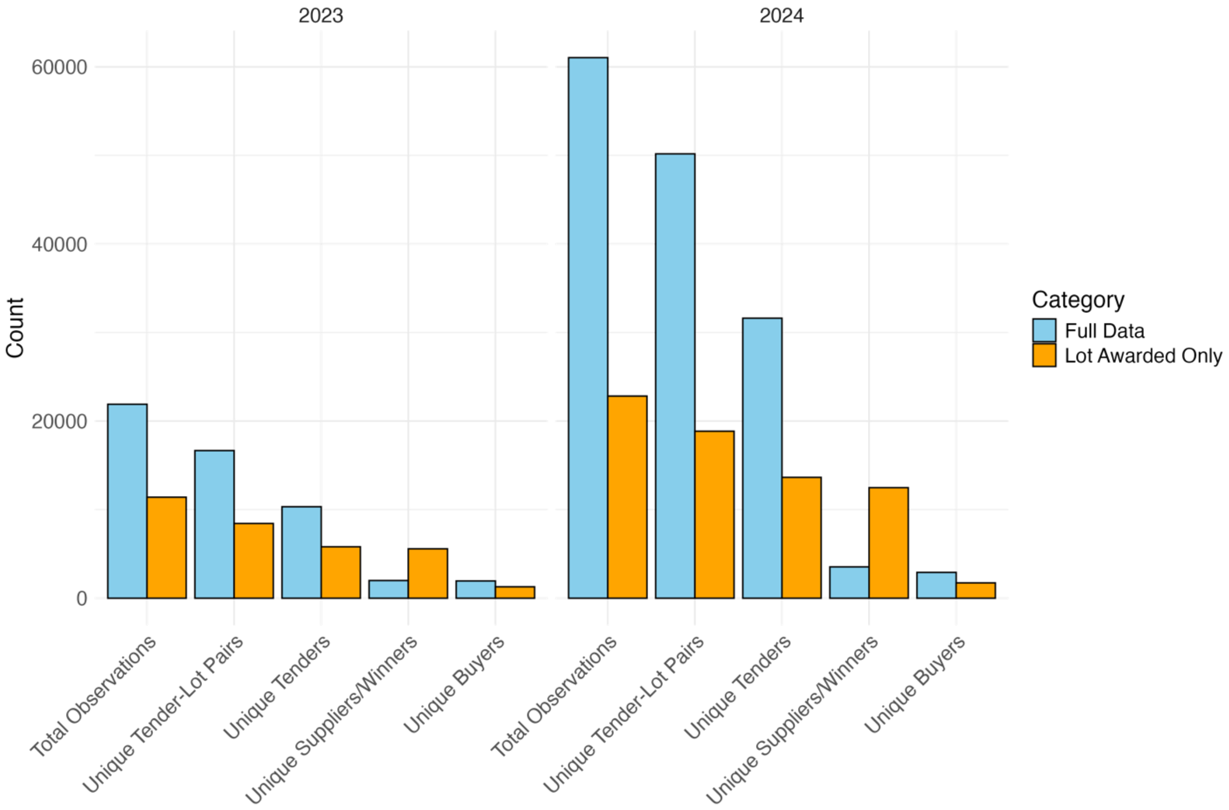
Public procurement data

For the public procurement data extracted and provided for the model development, the main challenge (in terms of storage and extraction) was to clearly define the requested data fields and ensure comprehensive documentation on the type and structure of the information provided. Different categories of data are stored separately. For example, information on public agencies is kept apart from data on suppliers, which is further separated from data on bidders, tender notices, and contracts. While this practice is standard among many public procurement authorities and does not inherently pose problems, it is essential to ensure that the final dataset consistently integrates all relevant fields. This includes storing the data in a standard comma-separated (CSV) format with a clearly defined unit of observation (e.g., tender-lot level) to facilitate accurate analysis and interoperability.

When it comes to data availability, several serious issues occurred, limiting the scope of analysis. Belgium recently updated the platform used to publish public procurement tenders, introducing significant changes to the scope and availability of data fields. Most notably, the inclusion of information on the winning bid is essential for the model. As a result, two datasets were obtained: one export from the old procurement platform covering tenders from 2019 to 2023, and another export from the new procurement platform, covering tenders from September 2023 to October 2024 (less than one year of data). Given the model's reliance on information about winning bids (to link contracts and to have information about suppliers), only the data from the new platform could be used for the model.

While the new platform provides essential information on winning bids, it contains a high share of missing data on the number of bids submitted, particularly when compared to the old platform. This issue is further compounded by gaps in the 'lot status' variable, which indicates whether a contract was awarded, cancelled, annulled, or is still in the process of being awarded. For more than half of all observations (Figure 1), this field is missing, meaning there is a lack of clarity on the status of a significant share of tenders. As a result, it cannot be determined whether these lots were awarded, are still under evaluation, or were ultimately cancelled. To ensure accuracy, a conservative approach was adopted: only lots explicitly marked as "awarded" are treated as concluded contracts. However, this method likely underestimates the true number of awarded lots, as many may be missing a status designation despite having been finalised.

Figure 1. Number of observations by categories for full data versus lot awarded only

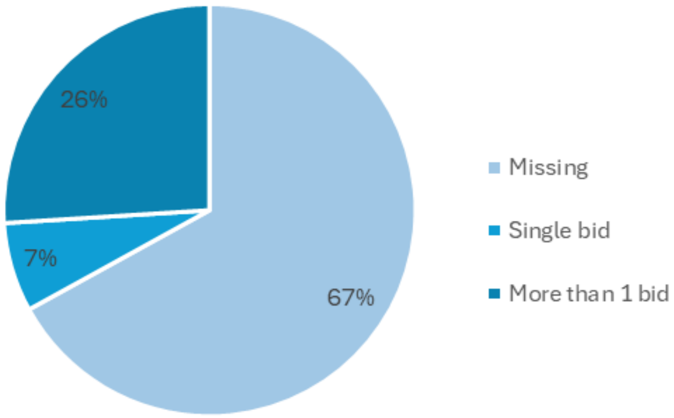


Note: The figure shows the distribution of filtered and unfiltered observations by year and category. Each observation corresponds to a tender-lot-bid entry from the raw data. For unique suppliers, data is shown only for the full dataset (in this context, suppliers = bidders), while unique winners (i.e., actual suppliers) are shown only for awarded lots. This distinction is due to data availability: in many cases, when information about winners (suppliers) is available, the same row often lacks data on bidders.
Source: Calculations undertaken by Government Transparency Institute (GTI).

Another significant issue concerns the availability of data on the number of bidders per lot. For two thirds of all awarded lots, the variable capturing the number of bids is missing (Figure 2). This means it often cannot be determined whether a lot received a single bid and was awarded to that bidder, or if multiple bids were submitted but not recorded.

To address this, several strategies to enhance data completeness were applied. Existing variables within the dataset were leveraged (such as bid count indicators and multiple rows per lot), under the assumption that multiple rows for the same lot indicate multiple bids. However, bid-level information tends to be even more incomplete than winner-level data. For instance, fields such as bidder name or bidder ID are frequently missing even when the same row includes information about the winner, suggesting that bid-level data could, in principle, have been recorded but was not. While it cannot be fully ruled out about the possibility of bid-level data is missing more broadly, it is assumed that multiple rows per lot represent multiple bids. Additionally, the separate export of BOSA data containing bidder information was obtained, and this was linked back to the main public procurement dataset. Unfortunately, due to the high number of missing values in the lot award status variable, a substantial number of bidder records (where the final status of the lot remains unknown) was lost.

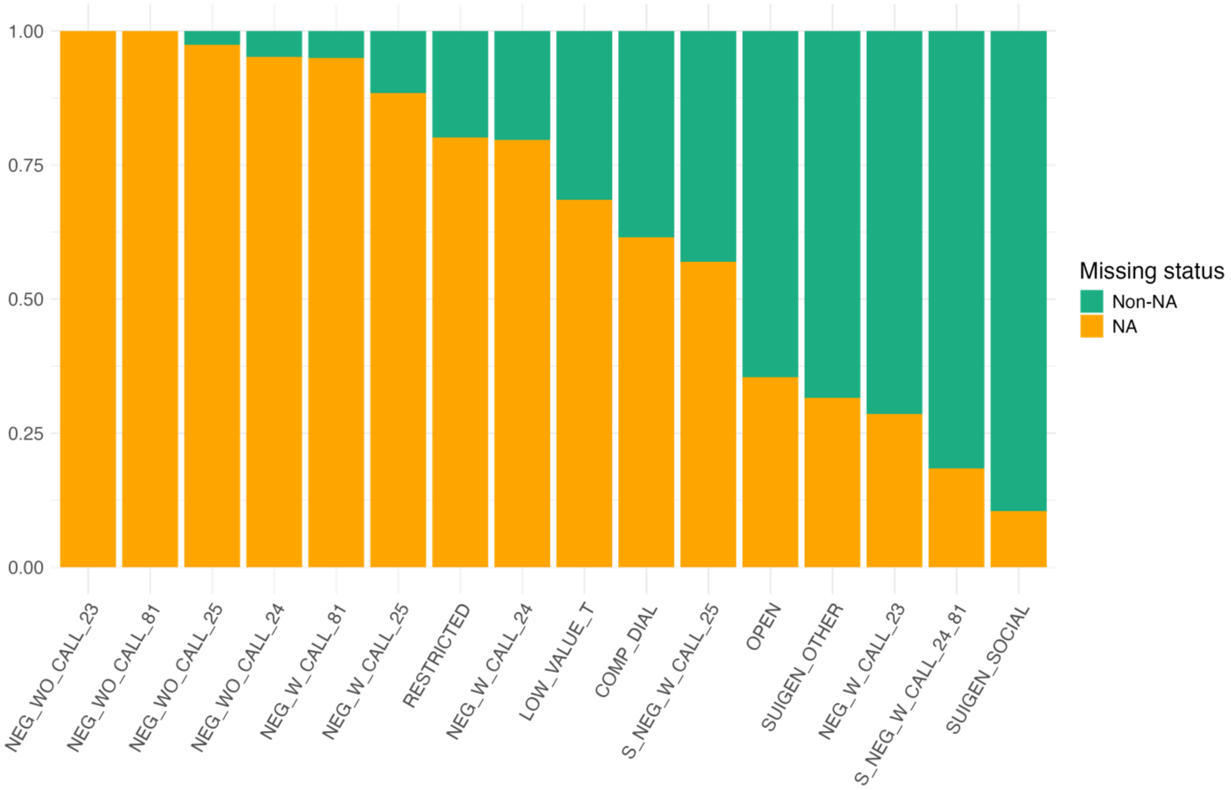
Figure 2. Distribution of the single bid variable categories



Source: Calculations undertaken by Government Transparency Institute (GTI).

Another set of challenges stemmed from the correlation between missing data and key categories used in the analysis. This issue was particularly evident in the national procedure type variable and related fields, such as the number of bids (Figure 3) and bid submission deadlines. When missing data disproportionately affects certain categories within the procedure type variable, those categories are effectively excluded from the analysis. The model treats these rows as incomplete due to the absence of critical information, resulting in their omission. This creates a risk of analytical bias: procedures that are potentially more restrictive (and therefore more prone to corruption risks) may be underrepresented simply because they lack complete data. Consequently, the model may overrepresent less risky procedures that are better documented, leading to an underestimation of risk in procedures where data is systematically missing.

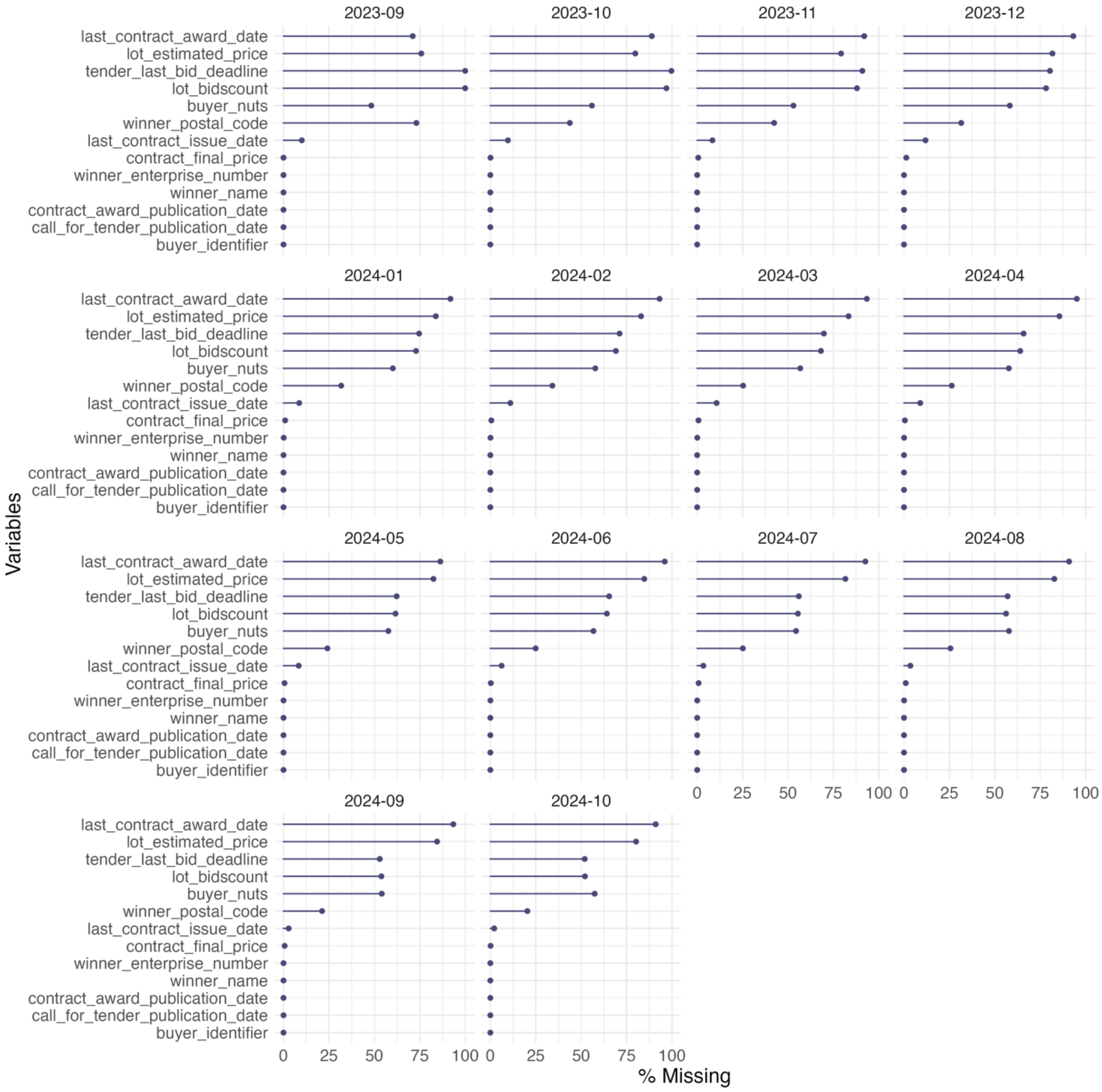
Figure 3. Share of missing values in number of bids per national procedure type



Note: Categories of procedure types are kept as was originally presented in the dataset, including numbers referencing different EU directives.
Source: Calculations undertaken by Government Transparency Institute (GTI).

Finally, a separate set of challenges emerged due to the recent transition from the old platform to the new public procurement platform. As expected, data availability on the new platform was limited during the initial months following the switch but has gradually improved over time (Figure 4). The analysis revealed that systematic gaps in key variables – essential for calculating risk indicators – led to artificially elevated scores during the early months of data collection. However, the overall trend shows steady improvement, indicating that with at least one additional year of data, certain indicators are likely to stabilise, while others may become feasible to calculate with greater reliability.

Figure 4. Missing rate of selection of key variables by months



Source: Calculations undertaken by Government Transparency Institute (GTI).

Crossroads Bank for Enterprises

Access to data from the Crossroads Bank for Enterprises (KBO) is governed by specific legal provisions and requires the submission of a formal request for access to the data ("Request for access to the data of the Crossroads Bank for Enterprises"). This procedure is not automatic and must be reviewed and approved by the legal department of the KBO. KBO granted access to the requested data, providing the following rationale: "*The information provided clearly indicates that the Federal Internal Audit (FIA) is responsible for assessing risks within the entities it audits and, in the context of audit planning, may also need to assess risks in entities it could evaluate in the future. In order to properly assess risks related to public procurement, FIA must rely on data from the KBO concerning suppliers involved in public contracts.*

This data is necessary for FIA to develop a risk model that would support its risk assessment activities in this area."

Taking the above into consideration, data from the KBO is easily accessible and straightforward in terms of storage. However, the main limitation lies in its scope. The current dataset does not support the development of more advanced indicators, such as those based on ownership structures or financial information. In its present form, the register includes basic details like company age and registered economic activity, which are useful but limited. The dataset would be significantly more valuable if it included additional variables such as annual turnover, allowing for analysis of sudden revenue increases following contract awards, or information on ownership chains, allowing to check affiliated companies participating in tenders with the same contracting authority.

A secondary, though important, challenge relates to the interoperability of public procurement and company registry datasets. Currently, company tax identification numbers in the procurement data are stored inconsistently, often containing extra characters, irregular spacing, or variations such as the presence or absence of leading zeroes. As a result, several data cleaning steps are required to standardise these entries and to ensure accurate matching with the company register.

2.2.4. Impact of data challenges on the measurement framework

The main impact of the above-mentioned data-related challenges on the model concerned the selection of the final set of indicators. Starting with a long list of possible indicators to calculate (Table 2), it was necessary to narrow the list of indicators to those which have enough data coverage to be reliably calculated. Table 3 presents three types of indicators:

- The short list of feasible indicators which had enough data in respective variables, as well as sufficient variation;
- Indicators which could not be included in the final model due to insufficient data coverage but are expected to become more reliable once the data gets populated further with additional observations; and
- Indicators which were excluded due to data limitations.

Table 3. Updated list of corruption indicators due to data challenges

Indicator	Overview
Short list of feasible indicators	
Single bidding	Enough variability and observations, although analytical value is limited by a high proportion of missing data.
Advertisement period	Enough variability and observations but correlated with procedure type hence should be analysed in interaction.
Decision period	Enough variability and observations.
Procedure type	Enough variability and observations, alas some procedure types which are expected to be riskier are omitted from analysis due to missing information on the number of bidders. If such information appears at the later stages, risky categories might change.
Distinct markets	Enough variability and observations but will benefit from additional year of observations to accumulate more information of companies' behaviour.
Tax haven	Enough variability and observations.
Local supplier	Enough variability and observations.
Non-transparent buyer	Enough variability and observations.
Indicators for future use	
Buyer spending concentration	Requires at least a couple of years of observations to be more reliable. Currently is limited to one non-full year.
No call for tender	Requires additional information, distinguishing between calls for tenders and contract award notices (currently, the data does not distinguish between those).
Dissolution of company after winning the contract	Requires a few more years of observations to be more reliable. At this stage, only a limited number of companies have ceased operations, which is insufficient to develop a robust and statistically sound predictive model.
Indicators excluded from the final model	
Number of companies at the same address	The quality of address information in public procurement records is insufficient to reliably assess how many companies are registered at the same location.
Company age	The dataset contains very few newly established companies, reflecting the structure of the Belgian economy, which is predominantly composed of long-standing firms. This limits the ability to analyse risks or behaviours specific to newer market entrants.

For the final model, tailoring and validation were conducted using eight key indicators: 1) single bidding, 2) advertisement period, 3) decision period, 4) procedure type, 5) distinct markets, 6) tax haven, 7) local supplier, and 8) non-transparent buyer. These indicators were selected based on data quality, relevance, and their demonstrated relationship with corruption risk.

2.3. The final validated model

2.3.1. Model overview

After identifying a list of feasible indicators, the next step is to refine and validate them using logistic regression analysis. This involves testing various threshold values for each indicator against a key outcome variable: single bidding, which is used as an objective proxy for limited competition. The analysis assesses how changes in each indicator affect the likelihood of single bidding, allowing for the assignment of risk levels to specific thresholds. A reference category is defined for each indicator to serve as a benchmark, enabling comparisons with other categories and their impact on the probability of single bidding. During this tailoring stage, it is already possible to observe whether each indicator behaves in line with theoretical expectations, for example, whether a short advertisement period indeed reduces competitiveness, or whether the involvement of younger companies correlates with a higher likelihood of single bidding. After tailoring, each indicator is formally validated using logistic regression with relevant control variables, as shown in Table 3, and further explained in the formula below.

$$\mathrm{logit}(\Pr(\text{single bidding} = 1)) = \beta_0 + \beta_1 \cdot \text{risk indicator} + \beta_i \cdot \text{control}_i + \varepsilon$$

Where:

logit(Pr(corr_singleb_cri = 1)) – the log-odds of a contract receiving single bid (binary outcome: 1 = yes, 0 = no)

β_0 – model intercept

β_1 – estimated coefficients for each explanatory variable

β_i – estimated coefficients for controls

ε – error term

The validation step tests whether the thresholds defined during tailoring accurately capture the expected relationship between the indicator and the likelihood of single bidding. A correctly specified indicator should demonstrate a statistically significant and directionally consistent effect on single bidding probability, confirming its relevance as a corruption risk proxy.

Only indicators that successfully pass the validation stage are included in the final set used to construct the composite CRI. These are indicators that show a statistically significant and meaningful association with the likelihood of single bidding in the logistic regression analysis. Table 4 displays the individual logistic regression coefficients (log-odds) for each validated indicator, along with the results of a combined logistic regression model that includes all of them (full model).

Table 4. Corruption Risk Index: Validation logistic regression results with single bidding

	Full model	Procedure type	Advertisement period	Decision period	Tax haven	Distinct markets	Local supplier	Non-transparent buyer
Procedure type (1)	0.210 ** (0.090)	0.416 *** (0.062)	0.254 *** (0.089)					
Length of advertisement period (1)	0.164 ** (0.073)		0.188 *** (0.072)					
Length of decision period (1)	1.034 *** (0.088)			1.037 *** (0.087)				
Length of decision period (0.5)	0.619 *** (0.068)			0.631 *** (0.067)				
Tax haven (1)	0.614 *** (0.183)				0.560 *** (0.175)			
Distinct markets (1)	0.380 *** (0.065)					0.372 *** (0.059)		
Local supplier (province)	0.181 *** (0.065)						0.053 (0.062)	
Non-transparent buyer (1)	0.204 ** (0.099)							0.164 * (0.090)
Used control variables	Contract value Buyer type Buyer region Tender supply type Tender year Market ID	Contract value Buyer type Buyer region Tender supply type Tender year Market ID	Contract value Buyer type Buyer region Tender supply type Tender year Market ID	Contract value Buyer type Buyer region Tender supply type Tender year Market ID	Contract value Buyer type Tender supply type Tender year Market ID	Contract value Buyer type Buyer region Tender supply type Tender year	Contract value Buyer type Tender supply type Tender year Market ID	Contract value Buyer type Tender supply type Tender year Market ID
PseudoR^2	0.203	0.130	0.164	0.190	0.126	0.082	0.119	0.116
Number of observations	9,090	9,094	9,094	9,090	9,094	9,094	9,094	9,090

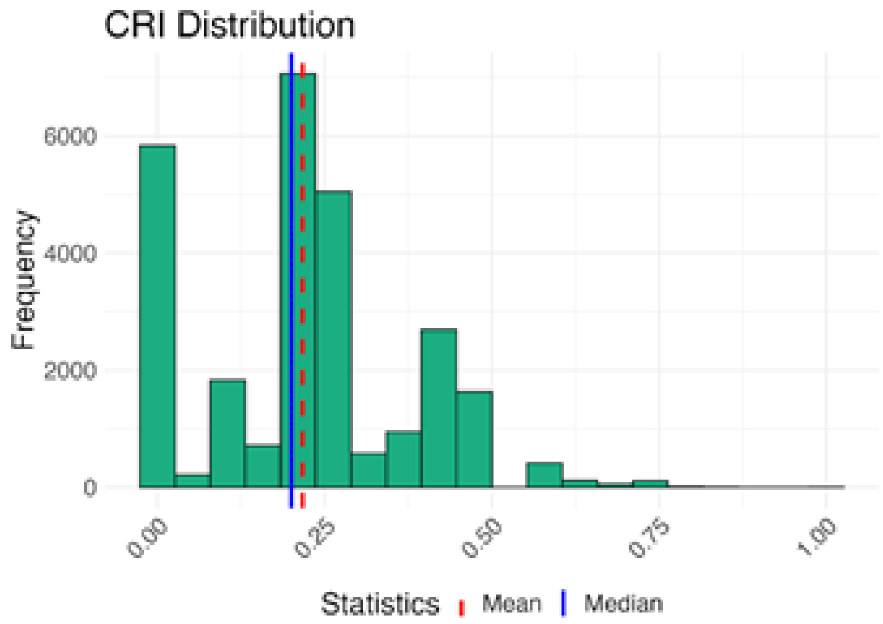
Note: *** p < 0.001; ** p < 0.05; * p < 0.10. Standard errors (not clustered) are reported in parentheses.

The individual models, as well as the full model, show consistent directions of coefficients. Whether estimated separately or jointly, each of the red flag indicators positively influences the probability of single bidding at the contract level. The significance of the coefficients also holds across specifications, with one exception: the local supplier variable. While it is positive but not significant in the individual model, it becomes statistically significant when included alongside other red flags in the full model. The reported coefficients are expressed in log-odds and thus require transformation into marginal effects and predicted probabilities to reflect their substantive impact on the likelihood of single bidding. While individual effects are discussed in more detail (Section 2.3.2), the full model reveals that the presence of a risky procedure increases the probability of single bidding by 3 percentage points, and a short advertisement period by one percentage point if looked at together with change in procedure type. A medium-risk decision period increases the probability from 17% to 28%, while a high-risk decision period raises it further to 37%. Additionally, having a supplier based in a tax haven increases the likelihood of single bidding by 11 percentage points, and the presence of a local supplier increases it by three percentage points, holding all other variables constant. Finally, a non-transparent buyer increases the probability of single bidding from 20% to 24%.

Each model is estimated on a sample of approximately 9,000 contracts. This figure is substantially lower than the raw public procurement dataset initially received, due to several stages of filtering. First, excluding rows with unknown lot award status reduced the dataset from 83,000 to 31,500 observations. Further removal of duplicate entries and contracts from monopolistic markets reduced the sample to 31,400 observations. Aligning the dataset further to tender-lot level from tender-lot-bid level reduced observations to 27,000. Of these, around 67% lack data on the number of bidders, which leads to their exclusion from the final model. As a result, the logistic regression is run on the remaining subset of contracts with complete information.

In the final stage, the CRI is calculated by averaging the sum of all individual red flags identified at the contract level. This composite measure highlights contracts that trigger multiple risk indicators, making it easier to detect those with a higher likelihood of corruption. As shown in Figure 5, the overall distribution of the CRI is skewed toward zero, indicating that many contracts exhibit no red flags in the data. The distribution then peaks again around 0.25. Given that eight red flags in total were used, a CRI of 0.25 corresponds roughly to two red flags per contract. Both mean and median are also around this value. A smaller peak appears around the 0.45–0.50 range, suggesting that several thousand contracts contain four red flags. More extreme cases, with CRI values of 0.75 or higher, are rare but do occur occasionally in the dataset.

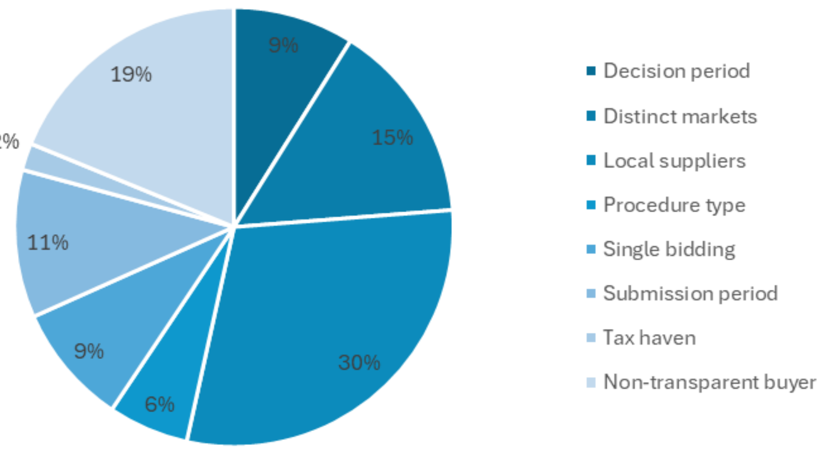
Figure 5. Corruption Risk Index: Overview of distribution



Source: Calculations undertaken by Government Transparency Institute (GTI).

Figure 6 shows the weight of each CRI component. In practice, this means that some red flags contribute more to the average CRI score than others, due to the frequency of risky values and the presence of missing data. For example, very few contracts are flagged for tax haven registration. This is because only a small number of suppliers in the Belgian data are registered in tax havens, resulting in this indicator contributing just 2% to the overall CRI. In contrast, many suppliers are located in the same province as the buyer, which makes the local supplier flag more common and gives it a weight of 30%.

Figure 6. Corruption Risk Index: Overview of weight of components



Source: Calculations undertaken by Government Transparency Institute (GTI).

2.3.2. Description of individual risk indicators

As previously discussed, the final model includes eight individual red flags, each of which was specifically tailored and validated before being incorporated. Each red flag is a categorical variable that takes a value between 0 and 1. A detailed description of the thresholds for each category is provided in Table 5.

Table 5. Individual red flags with definitions

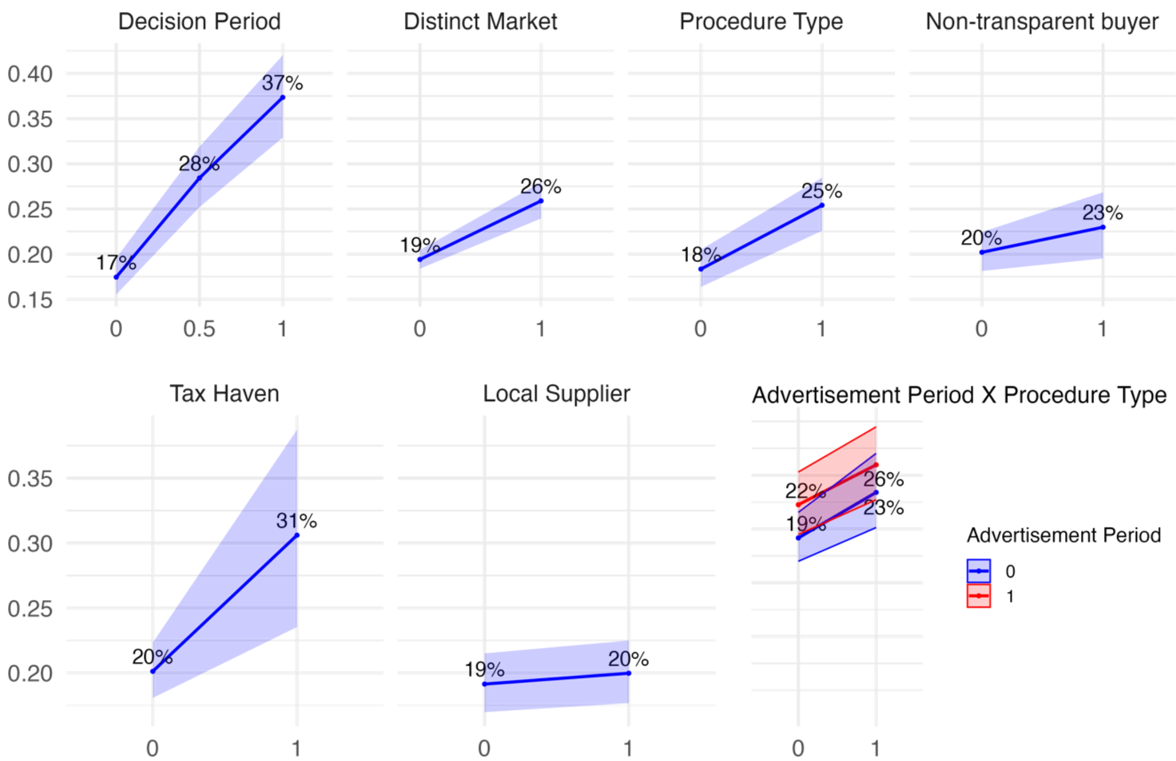
Indicator	Thresholds
Single bidding	Only 1 bid was received - low competition - 1 More than one bid was received - non-risky - 0
Advertisement period	<=28 days for risky procedure types - risky advertisement period - 1 32-37 days for non-risky procedure type - risky advertisement period - 1 >37 days - non-risky - 0
Decision period	<=34 days - risky decision period - 1 34-62 days - medium risky decision period - 0.5 >62 days - non-risky decision period - 0
Procedure type	Simplified negotiated procedures with call, Negotiated with call, Social and other specific services - risky procedures - 1 The rest - non-risky procedures - 0
Distinct markets	If a contract of a given supplier is significantly outside of the community of Common Procurement Vocabulary (CPVs) (network based) - risky - 1 Contract is within the common community of CPVs - non-risky - 0
Tax haven	Supplier's country in tax haven - risky - 1 Otherwise - non-risky - 0
Local supplier	Supplier and buyer are located in the same province (1st digit postal code) - risky - 1 Otherwise - non-risky - 0
Non-transparent buyer	26% or more of the total information is missing on average by buyer - high risk - 1 Less than 26% of total information is missing on average by buyer - low risk - 0

When tested in separate logistic regression models (Figure 7), each red flag showed a positive and statistically significant effect on the probability of single bidding. For example, the decision period indicator, which includes a medium-risk category, increases the likelihood of single bidding from 17% to 28%. If the decision period falls into the high-risk category (i.e., less than 34 days to select a winner), the probability increases by an additional 9 percentage points, reaching 37%. The distinct markets indicator increases the probability from 19% to 26%. This means that when a company wins a contract in a market that is outside its usual cluster of activity, the likelihood of single bidding rises by 7 percentage points.

The use of risky procedure types, such as simplified negotiated procedures with call or negotiated procedures with call, raises the probability of single bidding by 7 percentage points, from 18% to 25%. Non-transparent buyer with average missing information more than 26% increases the probability of single bidding by 3 percentage points to 23%. The presence of a supplier in a tax haven also has a notable effect, increasing the probability by 11 percentage points, reaching up to 31%. The local supplier flag, pointing to whether the supplier is located in the same province as the buyer, has a smaller effect, increasing the probability from 19% to 20%.

Finally, the interaction between the advertisement period and procedure type shows that when both indicators switch from non-risky to risky, the probability of single bidding increases to 26%. For comparison, if only the procedure type becomes risky while the advertisement period remains non-risky, the probability rises to 23%.

Figure 7. Predicted probabilities of single bidding for individual red flags



Note: Blue area around line connecting predicted probabilities refer to standard errors of the effect.
Source: Calculations undertaken by Government Transparency Institute (GTI).

Important note about the open-ended nature of model ("future proof")

The model remains flexible and can be further refined by incorporating additional risk indicators. There are several avenues for future improvements as outlined in Section 3.2. Importantly, if the necessary data is available, new indicators can be systematically tested for their statistical relationship with single bidding and buyer concentration. If these indicators demonstrate significant and robust effects, they can be integrated into the CRI score to expand its diagnostic value.

2.3.3. Limitations of the model

As was described in Section 2.2.3 (data-related challenges), the results of the final model have many limitations due to data availability. There are three main types of limitations in this regard:

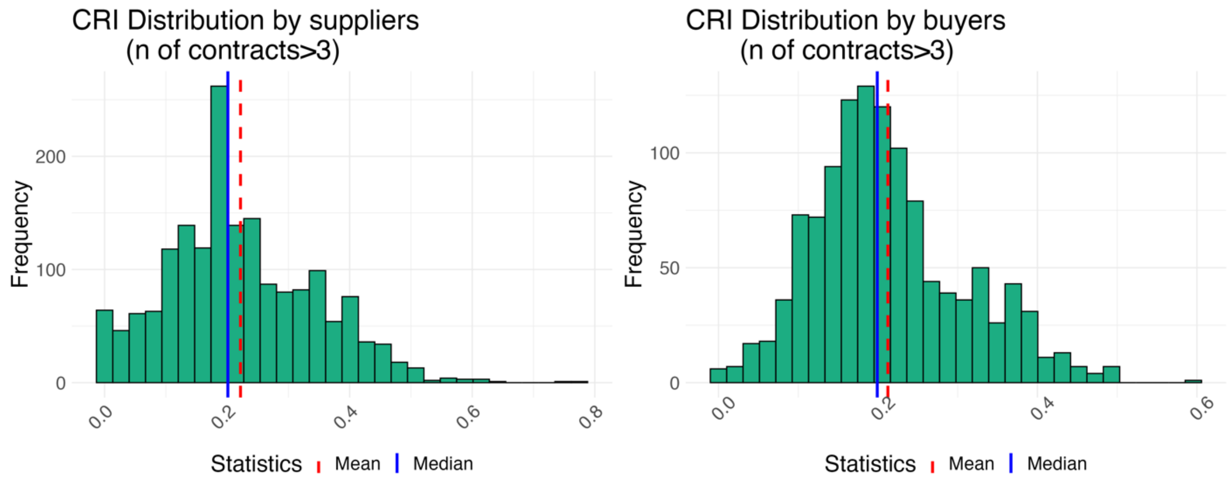
- **Limitations due to short time period:** Due to the transition from the old to the new platform, the currently available data covers only the period from September 2023 to October 2024. This timeframe does not constitute a full year for several key variables. For instance, tenders with calls published before the transition (such as in August 2023), are unlikely to appear in the new platform's export. As a result, the dataset primarily captures tenders that began from September 2023 onward. Moreover, the platform switch appears to have introduced inconsistencies, including missing records and data entry errors, as evidenced by the gradual improvement in missing data rates over time. Extending the observation period by at least one additional year would significantly enhance the model's reliability and allow for recalibration of certain thresholds. For example, the decision period indicator has proven particularly sensitive to early data gaps due to its skewed distribution during the initial months. Furthermore, indicators like buyer spending concentration or a company's abrupt dissolution after winning a contract will benefit from additional year of observations.

- **Limitations due to the low number of observations**: While the current number of observations is sufficient to build a statistically reliable logistic regression model, the limited sample size may affect the model's future robustness and stability. In some high-risk categories, elevated standard errors were observed. These were likely due to the small number of risky observations, which leads to data imbalance and heteroskedasticity. These issues are expected to diminish as the dataset expands and includes a greater number of contracts over time.
- **Limitations due to correlated missing rates and omitted categories:** Certain categories relevant for corruption risk analysis are strongly correlated with missing data, leading to their exclusion from the model and the potential overlooking of important risk patterns. Identifying and addressing the root causes of these data gaps is essential. Without resolving these issues, model interpretation must be approached with caution (particularly in relation to procedure type, decision period, and advertisement period). For example, some procedure types are excluded because they lack information on the number of bidders, even though they may carry a high corruption risk. Negotiated procedures without a call have a 100% missing rate for bidder data, despite being likely candidates for elevated risk. Similarly, some procedures lack information on bid deadlines, resulting in their exclusion from analyses related to advertisement and decision period.

2.4. Key insights

This section provides a high-level summary of the model's results, focusing on patterns and anomalies among contracting authorities and suppliers. The analysis examines the distribution of CRI scores (Figure 8), considering only entities involved in more than three contracts to ensure a focus on regular participants in the public procurement market. The results show that most suppliers have an average CRI score of 0.20 or lower, typically corresponding to one or two red flags per contract. However, the distribution has a pronounced right-hand tail, indicating a small group of suppliers with significantly higher risk levels, some averaging up to six red flags per contract. In contrast, the distribution of CRI scores for buyers is more symmetrical and resembles a bell curve. Most public agencies average around two red flags per contract. Nevertheless, a small number of contracting authorities stand out with average CRI scores of 0.50 or higher, suggesting their contracts often contain four or more red flags.

Figure 8. Corruption Risk Index: Distribution by buyers and suppliers

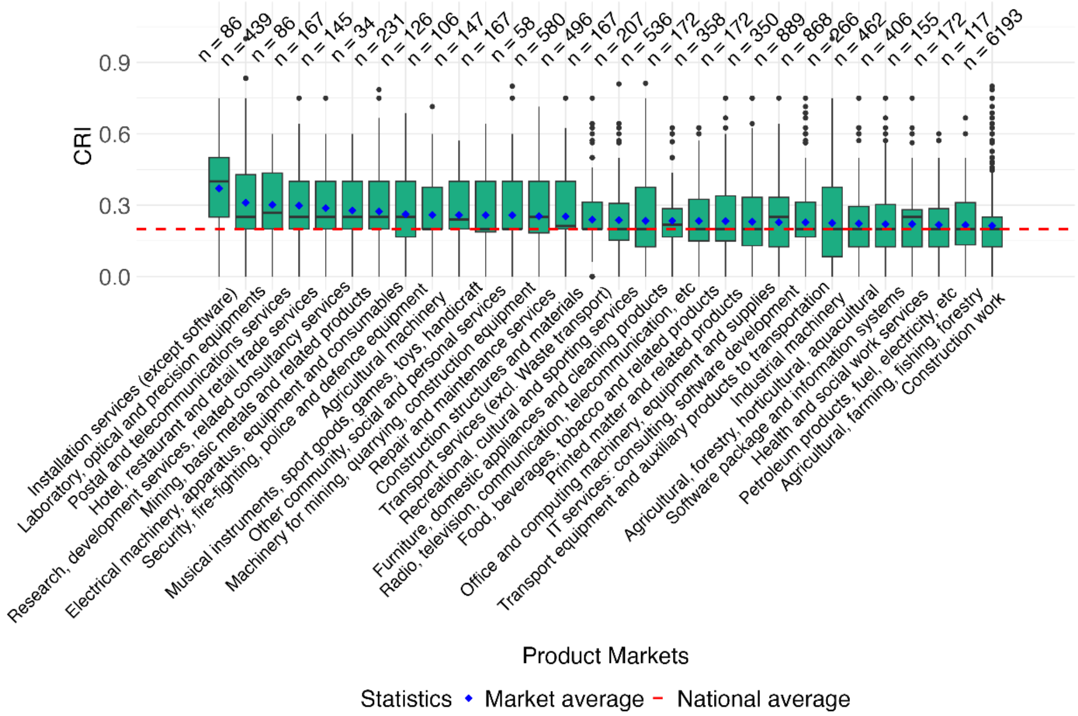


Source: Calculations undertaken by Government Transparency Institute (GTI).

The distribution of corruption risks across different market sectors was examined and defined using two-digit CPV codes (Figure 9). The analysis focused on sectors with an average CRI above the national mean of 0.22 and at least 20 contracts in the dataset, ensuring sufficient data for meaningful comparison. Overall, average risk levels are relatively consistent across sectors with enough contract volume. The sector with the highest average CRI is ‘installation services’ which shows a mean score of approximately 0.40 across 86 contracts. This is equivalent to an average of four red flags per contract. Other high-risk sectors display average CRI scores around 0.30, indicating two to three red flags per contract, compared to the national average of one to two.

These findings can serve as an initial starting point for more targeted oversight or investigation in Belgium. Sectors with elevated CRI scores may warrant closer scrutiny to identify whether specific suppliers or contracting authorities are consistently linked to higher risk levels, and whether certain contracts raise red flags that suggest potential irregularities.

Figure 9. Corruption Risk Index: Distribution by markets

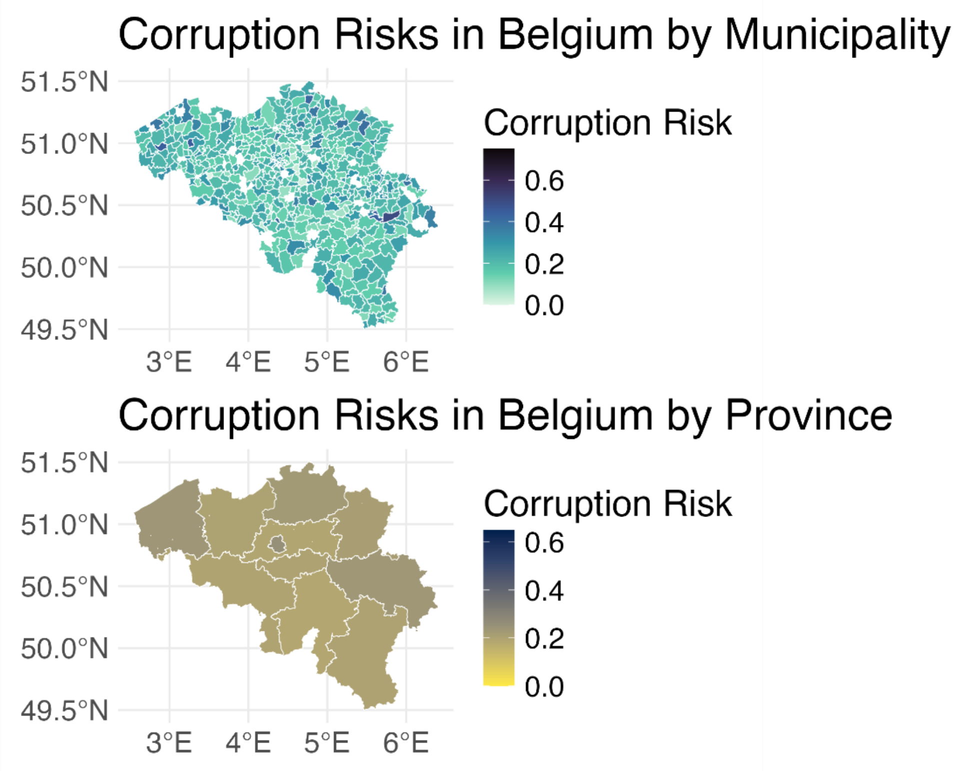


Source: Calculations undertaken by Government Transparency Institute (GTI).

The distribution of corruption risks was also analysed geographically across different regions in Belgium, based on the postal codes of buyers (Figure 10). These codes were matched to the administrative boundaries of municipalities and provinces to identify territorial variation. At the municipality level, exploratory analysis suggested notable differences in average CRI scores. For example, Baarle-Hertog recorded an average CRI above 0.50 (equivalent to four or more red flags per contract) while Chaumont-Gistoux had a much lower average CRI of 0.02. However, when aggregated at the provincial level, the variation is less pronounced. No province exceeded an average CRI of 0.25, with Brussels showing the highest provincial average. These patterns should be considered preliminary and non-conclusive, as the underlying data may be incomplete and potentially skewed, and are therefore merely indicative.

While the analysis may suggest that corruption risks appear more concentrated in specific Belgian municipalities, whereas broader regional trends seem more uniform, such findings remain exploratory and non-conclusive. Given the potential incompleteness and skewness of the underlying data, no prior conclusions should be drawn. Nevertheless, the exercise illustrates the potential value of local-level analysis for detecting risk hotspots that may not be visible in higher-level aggregates.

Figure 10. Corruption Risk Index: Maps of distribution by territories



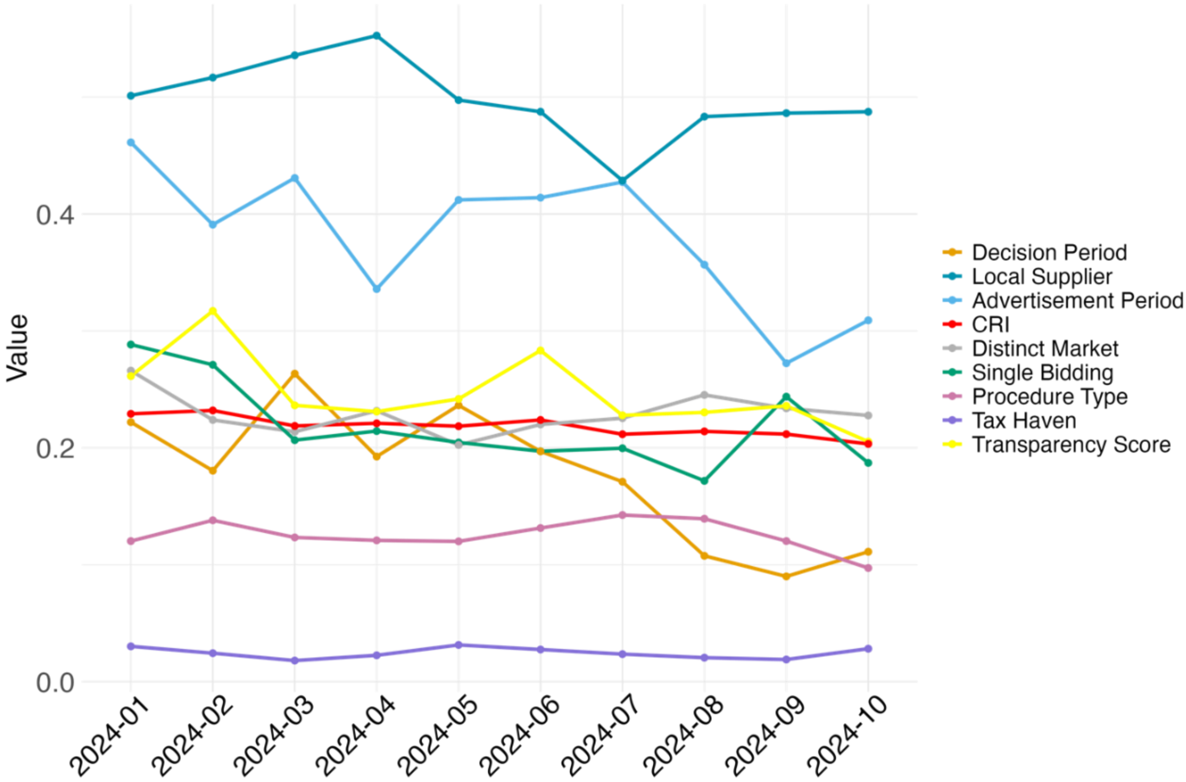
Note: White colored territories mean no data is available for this municipality. NIS codes are used to match the buyer location with administrative borders.
Source: Calculations undertaken by Government Transparency Institute (GTI); www.geo.be for geographical coordinates.

Finally, the distribution of individual red flags was also analysed over time, using monthly data (Figure 11). Temporal patterns can reveal unusual patterns or anomalies related to external events or seasonal trends, such as budget year cycles. It is also a useful way to identify whether certain indicators exhibit unexpected spikes which may indicator data errors or other data quality issues that may warrant further investigation.

Some red flags appear infrequently across the dataset. For example, the ‘tax haven’ indicator consistently remains below 0.10, while the ‘procedure type’ indicator averages slightly above 0.10. In contrast, other indicators, such as ‘local supplier’ and ‘advertisement period’ occur more frequently, with average values of 0.40 or higher. This suggests that a significant proportion of contracts consistently trigger these red flags each month.

One insight is a decline over time in the ‘advertisement period’, which may reflect changes in data availability. Specifically, more recent contracts (those entered after the transition to the new e-procurement platform) are better represented in the dataset, affecting the observed trend. Given the short period of analysed data (since the inception of the new platform), further review of this possible trend is worthwhile.

Figure 11. Distribution of individual red flags' values by months



Note: Months prior to 2024 are removed due to small number of observations.
Source: Calculations undertaken by Government Transparency Institute (GTI).

3 Practical considerations to improve data quality: Opportunities to strengthen data-driven audit risk assessments

As with any model designed to assess corruption risk, there will be implementation challenges, most notably those related to data quality. Some of these challenges are common across oversight and integrity institutions, and most can be overcome with careful planning. These issues can be considered when designing and implementing data-driven risk model. It is essential that public sector entities responsible for overseeing public procurement take into account data quality when developing similar risk models. It is also important when designing and deploying the data platforms on which such models rely. This includes investing in robust data governance frameworks, ensuring standardised and high-quality data inputs, and providing adequate training and guidance to users who enter or manage procurement-related data. These elements are critical to ensuring that the platforms not only support reliable risk assessment but also promote transparency and accountability in public spending.

This chapter explores some practical recommendations for FIA when refining and implementing the risk model. Despite the constraints experienced during the development of the risk model for FIA, targeted improvements, both in the short and long term, can meaningfully enhance the model's reliability and support the development of additional tools for automated risk detection. Importantly, the practical challenges and recommendations outlined here are not unique to Belgium; they are broadly relevant to jurisdictions seeking to implement data-driven approaches to corruption risk assessment. These recommendations have relevance to oversight entities who are seeking to improve and use better quality data for data-driven audit risk assessments.

3.1. Practical recommendations

3.1.1. Maintaining the operation and quality of the model

When developing risk assessment models for oversight entities and supreme audit institutions, it is important to design models to allow flexibility and adaptability. These design principles are considered when the OECD develops models for corruption and fraud risk assessments for public sector entities. For example, the risk model for Portugal's Tribunal de Contas (TdC) has been designed to enable more advanced risk indicators to be included and for further scalability over time (Hlacs and Wells, 2025[9]).

Building on this knowledge, the risk model developed for FIA is designed to be flexible and adaptable, allowing for regular updates and ongoing use. To ensure its continued effectiveness, four key areas require ongoing attention and maintenance by FIA: 1) populating data with new observations, 2) adding new indicators, 3) regularly updating thresholds and risk classifications, and 4) ongoing validation of indicator regressions.

Populating data with new observations

It is essential for FIA to continuously update and maintain the dataset. Not only by incorporating newly published contracts and tenders, but also by addressing previous data gaps (for example, missing or incomplete data, omitted categories, missing bidder and deadline information, or missing lot status). Ongoing monitoring of data quality is critical, as resolving one issue may reveal new inconsistencies. For example, while current data shows a link between certain procedure types and missing information, correcting this may uncover other problematic issues. To safeguard the reliability of the model, regular data quality checks should be performed over time, covering all public entities, suppliers, and the full range of variables used in indicator calculations.

Adding new indicators

As new observations are added to the dataset, or as additional data sources with linking potential become available, new indicators can be integrated into the risk model. Each new indicator should undergo the same rigorous process as those already in use. It must be reviewed, cleaned, filtered (if necessary), adapted to the specific context, and statistically validated. Once validated and incorporated, the CRI score should be recalculated to reflect both the updated values and the expanded set of red flags.

Regularly updating thresholds and risk classifications

As new data is incorporated, existing thresholds – and the classification of high-risk categories – may shift, especially for variables previously impacted by correlated missing values. Therefore, whenever a substantial volume of new data becomes available, it is essential to rerun the entire modelling pipeline. This ensures that manually coded thresholds and risk classifications, derived from the flagging logistic regression models, remain valid, statistically sound, and aligned with the most current data.

Ongoing validation of indicator regressions

While it is unlikely that indicators which previously demonstrated strong, positive associations in the validity regressions will become statistically insignificant, the introduction of newly populated categories or expanded data may affect these relationships. It is important to regularly assess the performance of indicators within the full model, which includes all individual predictors. Additionally, running separate validity regressions for individual indicators can help identify issues related to the calibration of new thresholds or risk classifications.

3.1.2. Recommendations for addressing data quality issues and data gaps

While data quality challenges are common across contexts and public data ecosystems, they vary in scope and require different types of interventions. Some issues are systemic and demand high-level structural reforms (macro level). Others occur at the intermediate (meso) level, such as deficiencies with data export processes or storage infrastructure. Finally, there are more localised and easily addressable issues to inconsistencies in data entry or the absence of basic quality control checks (micro level). These can be resolved relatively easily through targeted operational improvements.

This section reflects on the macro-, meso- and micro-level challenges that FIA must be cognisant of during implementation of the risk model. It is important to note that some of the issues (i.e., systemic) may be outside of the control of FIA (and so too other oversight audit institutions in other jurisdictions). The 'ecosystem' of data quality issues and data gaps are presented here for completeness. In reality, many of these challenges are experienced in other jurisdictions, and no doubt resonate with other public sector entities who, like FIA, are also attempting to address data quality issues and data gaps to enable robust corruption risk assessments.

Systemic (macro level) challenges

A key challenge in achieving comprehensive data coverage – particularly for critical fields such as bidder details, contract status, and tender outcomes – stems from the reporting practices of public sector entities (buyers). Analysis of missing data patterns suggests that some buyers systematically omit certain fields or report them inconsistently over time. However, complete and reliable information on contract status, tender results, and winning bidders is essential for building robust models to assess corruption risks effectively.

These macro-level challenges may be outside the specific mandate of FIA to address alone. However, it is important that public sector entities such as FIA participate in ongoing dialogue between national authorities, data managers, and contracting entities to improve data coverage. Strengthening communication channels can help identify reporting bottlenecks, clarify data entry obligations, and encourage consistent compliance. Capacity-building initiatives and feedback mechanisms, such as periodic data quality reviews highlighting missing data by agency, can further incentivise more consistent and accurate reporting. Ultimately, a collaborative and coordinated approach is key to improving data quality and supporting the effectiveness of risk assessment models.

Intermediate (meso level) challenges

Intermediate-level challenges are generally more manageable than systemic challenges but remain crucial for ensuring successful and sustainable model updates. A significant share of the responsibility often rests with the data provider. These challenges are not unique to the Belgium context. Reliable and consistent data export and delivery depend on several key factors:

- **Clearly defined list of data fields for extraction:** The data provider often manages a broad array of fields, many of which may be irrelevant or too complex for effective corruption risk analysis. To maintain focus and data quality, irrelevant fields should be excluded from exports, while all relevant fields must be consistently included. For example, each data update should request the same set of variables to ensure continuity and comparability over time.
- **Consistency in data export time frames**: If data was previously shared for a specific period (in the case of the model for FIA, the time period was September 2023 to October 2024), the updated export should ideally cover that same period plus any new months. Limiting exports only to new months should occur only when updates to historical records are not feasible. Only in cases where updates to past records are not possible should the export be limited to new months. The rationale, criteria, and method for time filtering should be clearly defined and transparently communicated.
- **Minimising the number of datasets shared separately (and consolidating data into a unified dataset):** Whenever feasible to do so, it is preferable to combine all relevant data fields into a single, integrated dataset prior to sharing. While providing multiple datasets with matching instructions is not inherently problematic, conducting the data matching process prior to handover is considered best practice. This approach helps to enhance consistency, minimises the risk of errors, and reduces the need for additional documentation or data clarification after the provision of data.
- **Consistency in level of data observation**: The dataset should maintain a consistent structure across all data exports, whether it is organised at the bidder, lot, contract, or tender level. For example, if the data is provided at the tender-lot-bid level, this structure should be preserved across exports, accompanied by clear rules (e.g., one lot equals one contract), and based on the national procurement context. Special attention must be taken to avoid duplicate entries caused by mismatching during data merging.

- **Clear and transparent documentation of applied data filters**: Every data export should be accompanied by full documentation that outlines any filters applied during data processing. This documentation must be shared with the receiving authority to ensure a full understanding of the dataset's scope, including any limitations, exclusions or conditions that may impact interpretation and subsequent analysis.

Local (micro level) challenges

Minor adjustments to micro level challenges can significantly enhance data quality. To ensure interoperability across datasets, it is essential that data ID fields are stored in a consistent and standardised format. These identifiers should be cleaned and prepared to enable reliable data matching. When data is entered in non-standard formats, there should be automated cleaning steps in place to standardise the data fields before storage and use. Improving oversight of fields critical to investigative work is also essential. This includes ensuring that such fields are easily extractable and ensuring they are supported by clear internal documentation outlining extraction methods, matching procedures, and underlying data logic. Regular quality control checks should be conducted to maintain the accuracy and reliability of these key data fields over time.

3.1.3. Essential preconditions for data-driven corruption and fraud risk assessment initiatives

Establishing and sustaining a data-driven corruption and fraud risk assessment framework that is capable of informing policy, guiding assessments, and supporting investigations, requires a broad set of skills, tools, and operational processes (OECD, 2019[39]). The following core areas represent the investments needed to for long-term effectiveness: (1) administrative datasets, (2) technical infrastructure, (3) data analytical and visualisation skills, (4) organisational processes and workflows, and (6) knowledgeable and engaged users. These core areas are relevant to FIA to consider, but also have relevance to other oversight entities in other jurisdictions who are considering the development and implementation of similar risk assessment initiatives.

Administrative datasets

Typically, one of the most significant costs in developing data-driven risk models stems from the creation, extraction, and organisation of the relevant administrative datasets such as those related to public procurement, company registers, and ownership information. However, these costs very much depend on the quality and openness of government data systems. Some jurisdictions like Italy[7] or the EU-wide Tenders Electronic Daily[8] (TED) portal, offer readily accessible, downloadable, and structured public procurement data. Many countries, such as the United Kingdom[9], offer publicly and freely accessible Application Programming Interfaces (APIs) for company registry and ownership related data. Such easy and structured data provision considerably lowers data costs for any data-driven risk assessment framework. In contrast, other countries still rely on paper-based public tendering systems and require significant investment to better automate data collection and digitisation of records. In between these two extreme cases lies most OECD countries, such as Belgium, with electronic data available in complex, yet structured formats that are accessible but requiring considerable investment into data extraction, organisation, and cleaning.

[7] https://www.anticorruzione.it/en/banca-dati

[8] https://ted.europa.eu/en/

[9] https://www.api.gov.uk/ch/#companies-house

Technical infrastructure

Given the storage, scale, and complexity of most public procurement datasets and systems, access to government data warehouse servers are often needed, at least for efficient data extraction. In addition, for larger datasets (i.e. consisting of millions of records), even basic data cleaning and analytical work might require the use of high-capacity servers. In addition, data cleaning and analysis are best undertaken using some of the widely used, open-source programming languages such as Python and R, and their software packages for data analysis. These tools offer the flexibility and scalability needed for large-scale processing and reproducible analytical workflows. FIA will need to give consideration of the technical infrastructure required to sustain the risk assessment model in the future.

Data analytical and visualisation skills

Creating, validity testing, and analysing corruption risk indicators require both an in-depth understanding of public procurement markets and advanced data analytic skills. Domain-specific knowledge is needed to understand: i) data scope and variable definitions; ii) the key characteristics of the regulatory framework (e.g., regulatory thresholds or time limits for tenders); and iii) the nature and dynamics of corrupt and fraudulent activities in public procurement. Data analytic skills typically include the capacity to manipulate large-scale datasets (e.g., more than 100 thousand observations) and to implement a variety of data science methods such as binary logistic regressions, random forest classifiers, or principal component analysis. Visualising results in a way that helps users to understand and act on risk measurement results requires someone with advanced knowledge of good data visualisation principles as well as software in which online dashboards can be implemented (e.g., R Shiny package or Tableau).

Organisational processes and workflows

To ensure that data-driven risk assessment frameworks meaningfully support policy development, assessments, or investigations, it is essential to embed them within clearly defined organisational processes and workflows. This means that existing processes and workflows in need of data should be identified and redesigned so that analytical inputs seamlessly fit. Relevant users need to know when and how to rely on data. Data visualisation and custom data export tools need to be designed and tailored to the organisational operational needs to ensure usability and uptake. Linking public procurement corruption and fraud risk indicators with other internal tools and information systems already in use (such as audit planning or investigative databases) helps to support a cohesive and informed decision-making process.

Knowledgeable and engaged users

Understanding risk measurement frameworks, and its strengths and weaknesses, presents its own challenges. It requires more than technical capacity; it also demands a clear understanding of its purpose, limitations, and practical implications. Even experienced auditors and risk managers may face challenges in interpreting risk scores and responding appropriately. Targeted training is essential to ensure key users such as auditors and risk managers can confidently interpret data and know how to action the insights gained from the risk model. In addition, creating a regular feedback mechanism allows users to share their experiences and observations, which can help refine and improve the framework over time. Expanding access to user-friendly tools—such as interactive dashboards—can also empower a broader range of staff to engage with risk data, thereby promoting more widespread and consistent use across public sector entities such as FIA.

3.1.4. Scalability and automation: Practical considerations for sustainable model development

Before scaling up any risk model, it is essential for oversight entities to ensure that key processes, particularly those related to data preparation, are automated as much as possible to reduce reliance on manual intervention. While some steps, such as reviewing thresholds or verifying that validity regressions show expected directions and statistical significance, will still require manual assessment and oversight, many other tasks can be streamlined through automation.

The model developed for FIA already incorporates basic procedures for data cleaning, filtering, and preparation, though they remain relatively standard and could benefit from further automation. What follows is a list of automated checks that should be prioritised by FIA to improve efficiency and scalability of the current model (but in reality, these checks are applicable for other oversight entities too):

- **Dataset matching and unit alignment**: Ensure that all datasets provided by the data provider are accurately matched and structured to a consistent unit of analysis (e.g., tender-lot level). Ideally, this should be performed by the data provider before delivery to minimise errors and reduce the need for additional clarification or adjustments.
- **De-duplication:** Identify and remove duplicate records to ensure data integrity. This includes eliminating identical rows and confirming that units of analysis, such as tender-lot pairs, are not repeated if the dataset is structured at that level.
- **Filtering relevant observations**: Apply filters to exclude data points that do not contribute to the required risk assessment analysis, such as lots with missing or non-awarded status, or tenders from monopolistic markets that distort competition metrics.
- **Variable selection for indicator development**: Choose variables with the highest analytical relevance and the lowest rates of missing data to ensure robust indicator construction.
- **Standardise key fields**: Clean and standardise key variables such as IDs and dates to ensure they are usable for matching and analysis across datasets.
- **Eliminate data errors**: Remove records containing evidently incorrect values (such as negative contract amounts) that likely stem from input or extraction mistakes.

The current model is built to support FIA’s future expansion of it by incorporating additional data sources and increasing both the volume of observations and the variety of variables used for indicator development. However, prior to scaling, it is crucial to resolve existing data quality issues and establish basic automation for key cleaning and standardisation processes. These steps will enhance data consistency, minimise manual effort, and strengthen the model’s reliability as its scope broadens.

3.1.5. Fostering effective collaboration across data stakeholders

A successful handover and effective implementation of any risk model depends on active collaboration among multiple stakeholders, not just technical data teams. The final dataset is the product of many steps involving data entry, collection, storage, and extraction. Ensuring that the data is reliable and usable must therefore be a shared responsibility across all stakeholders in this chain. It is also important to acknowledge that missing data may not always be accidental. In some cases, strategic non-disclosure by certain stakeholders may take place. Clearly communicating the relevance and importance of complete data for assessment, investigative, and oversight purposes is crucial to foster accountability.

Effective resolution of manageable data challenges begins with close coordination between data providers and users at the point of data export. Both parties need to have a clear, shared understanding of what information the dataset contains and why each variable is necessary. This is important for accurate data delivery. While technical teams are best positioned to identify data limitations during initial processing, investigators from oversight institutions such as FIA must be aware of how these limitations may affect the correct interpretation and insights generated from the data. Ongoing dialogue helps to foster transparency between these groups, and this in turn helps to ensure more accurate risk assessments and policy recommendations.

Finally, it is crucial to keep high-level decision-makers informed throughout the various stages of data processing and analysis, including honest discussions of the model's limitations. This transparency helps manage and set expectations, guides informed decision-making, and ensures that objectives for the model's deployment and future development remain realistic and aligned with organisational priorities.

3.2. Future refinement and scalability of the risk model

3.2.1. Pathways for refinement: Leveraging better data for improved insights

Scaling up risk models to expand their capacity to handle more data helps to improve insights in the longer-term for oversight and audit entities. The short-term success of a risk model in its proof-of-concept stage can quickly dissipate if it is not implemented to accommodate larger data sets (in terms of volume and complexity). Refining the set of indicators used in the risk model should be an ongoing goal of the continuous use of the model.

In terms of the risk model developed for FIA, at least two additional indicators[10] could be integrated into the model once the dataset becomes more complete: buyer spending concentration and sudden dissolution of companies following contract award. A third indicator – absence of a call for tenders – appears technically recoverable in consultation with the data provider and should be considered for future integration. Each of these possible additional indicators that FIA may consider for future iterations of the model, is described in further detail below.

Buyer spending concentration

The buyer spending concentration indicator requires a minimum of two years of data (ideally more) to produce meaningful insights. This is due to the nature of the metric, which calculates the total value of contracts awarded by a single buyer to a single supplier within a given one-year period. Within a single year, variation is limited to unique buyer-supplier pairs, which may be duplicated across multiple contracts. Therefore, expanding the timeframe and including additional data observations (such as currently excluded lots with unknown status) will increase variation and improve the reliability of this indicator within the model. Importantly, buyer spending concentration could also serve as an alternative outcome variable to single bidding, as it reflects similar dynamics related to restricted competition.

Sudden dissolution of companies following contract award

The sudden dissolution of companies following contract award represents another promising indicator for corruption risk. However, its current applicability to the FIA model is constrained by limited data availability. While a few such cases are already observable in the current dataset, their overall number remains low due to the short observation period which does not yet cover a full year, and the frequency of updates in the company registry. As the dataset expands over time and the registry updates become more comprehensive, this indicator will offer greater analytical value and could serve as a reliable indicator of irregular or suspicious contracting practices.

[10] Refer to the indicators in Table 3.

Absence of a call for tenders

The absence of a call for tenders – a strong and widely recognised corruption risk indicator in other country contexts – requires targeted data improvements to be effectively applied. Currently, the dataset contains general tender uniform resource locator (URLs) linking to platform-level information. It does not specify the type of publication (e.g., contract award notice vs. call for tender). To apply this indicator reliably, the dataset should explicitly distinguish whether a public call for tender was issued. The absence of such a call, particularly where it is legally required, is a strong signal of procedural irregularity. It is understood that BOSA does retain information on publication types, which was not included in the current model due to time constraints. Therefore, this indicator appears to be technically recoverable and should be prioritised for integration in future model iterations.

3.2.2. Staffing and skills requirements for sustainable model maintenance

The development of risk models for corruption assessments, and their subsequent implementation within oversight and audit institutions, requires an investment in people with the right technical capabilities. The success of implementing data-driven audit risk models relies heavily on the expertise, skills, and commitment of employees in these institutions. The OECD’s report *Strengthening Oversight of the Court of Auditors for Effective Public Procurement in Portugal* details the key dimensions for assessing an oversight or integrity institution’s digital maturity in this context. Of relevance – the people and culture dimension – relates to the expertise, skills, and commitment of individual employees within an organisation. This digital maturity dimension holds relevance for integrity and oversight institutions (including FIA) that they must consider the requirements for implementing and sustaining the use of an audit risk model. Key practices for the people and culture dimension include the following:

- 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 (OECD, 2024[6]).

For FIA to effectively update, maintain, and evaluate the risk model over time, a dedicated technical team of one to two staff members is required. These individuals should be proficient in both R and Python, the two programming languages used in the model. R is the primary language used, with most of the code written in it. Familiarity with core R packages is essential, including ‘tidyverse’, ‘data.table’, ‘ggplot2’, ‘naniar’, ‘estimatr’, ‘haven’, and ‘lmtest’. Python is used primarily for generating networks in the distinct market indicator. Key libraries include ‘pandas’, ‘numpy’, ‘time’, ‘networkx’, ‘random’, and ‘scipy.stats’. In addition, staff members should be able to extract and process data from Structured Query Language (SQL) tables and Extensible Markup Language (XML) files (e.g., for company registry data) and work confidently with standard file formats like CSV files.

From a methodological perspective, a solid understanding of logistic regression is essential, including its assumptions, interpretation of predictor significance, and the ability to troubleshoot common issues (e.g., variable type mismatches, multicollinearity, and missing values). In addition, a basic understanding of network analysis is beneficial, including concepts such as nodes, edges, and clustering methods like Louvain modularity optimisation. Should the model be extended to incorporate ML-based indicators, it will be important to have familiarity with various learning approaches. This includes supervised, semi-supervised, and unsupervised methods, as well as experience with commonly used models such as random forests.

3.2.3. Establishing a training dataset: essential preconditions for machine learning applications

Over the long term, compiling a dataset of confirmed corruption and fraud cases derived from investigative findings is highly valuable. With a sufficiently large and reliable sample of contracts proven (or very likely) to involve corruption or fraud, machine learning models can be trained to identify similar patterns in uninvestigated contracts. Such an ML approach could enhance the existing red flags methodology by estimating weights and integrating multiple risk factors to improve predictive accuracy.

One example of a supervised machine learning algorithm is the random forest (or random decision forest), widely used for classification tasks. The process begins with creating a labelled training dataset where each observation (e.g., a contract) is marked as either "corrupt" or "not corrupt" based on investigative outcomes. This dataset must include all relevant variables or indicators, such as number of bidders, contract value, and procedure type. The algorithm then constructs numerous decision trees, each trained on a random subset of the data. Each tree independently produces a prediction, and the final result is derived through an aggregation method. After training, the model can be applied to unlabelled contracts (those not yet investigated) to estimate the likelihood that they exhibit similar corruption patterns. These predictions can assist in prioritising future investigations by identifying contracts at higher risk.

Such models can be quite useful in investigative activities and usually allow for a larger number of predictors (indicators). However, it is important to be aware of the limitations:

- First, such models are highly sensitive to biases present in training samples. This means that even if the selection of cases to investigate was not random, or a very small number of cases was investigated, the model will replicate the bias or return a lot of false positives due to a significantly imbalanced sample – where proven cases are very few compared to hundreds of thousands of unproven cases (OECD, 2021[40]).
- Second, the model tends to learn patterns specific to the cases that were discovered and reported, potentially missing more subtle or less typical forms of corruption which were not present in investigated contracts.
- Third, such models are also sensitive towards missing values and inconsistent data types, making it essential to have more complete and well-structured data for reliable performance.
- Finally, while proven cases of corruption and fraud can be collected over time, it is typically more challenging to gather a reliable sample of non-corrupt, non-fraudulent cases, since investigations usually focus on high-risk cases. Random audits can provide a balanced sample of proven positive and negative cases such as in the case of Brazilian random audits underpinning a range of risk assessment tools.

3.2.4. Developing and sustaining internal expertise for ongoing risk assessment

Developing internal capacity for maintaining and updating risk models is strongly advisable for FIA to take into account (and the other institutions for several key reasons. First, it significantly reduces transaction costs, particularly time. Outsourcing model maintenance and development, given the complexity of the data sources, would require extensive back-and-forth communication, including clarifications about the data, explanations of model logic, and agreement on each modification. This process is often time-consuming and resource intensive.

Second, external actors may lack the deep familiarity with country-specific contexts, which are critical for accurate interpretation. Important nuances, such as data entry practices, national legislation, public agency workflows, and market-specific behaviours may be missed. For example, in the context of the model developed for FIA, it is important to understand that near-monopolistic supplier positions in certain markets may reflect normal market structures in Belgium and not necessarily signal corruption risks. Internal teams are better equipped to understand and account for such contextual factors in both model design, data requests, and data interpretation. Furthermore, certain corruption patterns may be more unique or prevalent in the Belgian context, requiring tailored indicators that reflect local vulnerabilities. The development of such indicators is enhanced greatly from in-house expertise with direct knowledge of public procurement systems, local stakeholders, and sector-specific dynamics.

Finally, internal teams are also better positioned to collaborate with investigative units, enabling the integration of emerging risks or known red flags into model updates. Building and sustaining the infrastructure for data provision, regular updates, and model expansion is essential for the model's long-term effectiveness. Building internal capacity supports the effective oversight and coordination of these activities, ensuring that the model stays current and aligned with evolving landscapes.

3.2.5. Scalability and future applicability of the FIA risk model

The data-driven risk model developed for FIA offers a robust foundation for future scalability, not only for FIA, but across various domains within Belgium's public administration. Its flexible design facilitates broad application to strengthen public integrity throughout the Belgian federal government. Although initially tailored for public procurement, the core architecture can be extended to other high-risk transactional areas. These areas include subsidies, payroll systems, social benefit programs, and licensing frameworks – provided that suitable data governance structures are established.

In subsidy oversight and grant management, such data-driven models enable the early identification of irregularities, allowing for more focused audits and better stewardship of public resources. Likewise, applying anomaly detection techniques to payroll and benefits data can help uncover duplicate payments, unjustified claims, or unusual licensing activities, thereby enhancing fraud prevention efforts. The model's modular structure supports straightforward adaptation to incorporate additional datasets and risk domains, including taxation and regulatory enforcement. Crucially, integrating this model within strategic audit planning and continuous monitoring processes fosters the institutionalisation of risk-based governance over the long term. This shift from reactive control toward proactive risk anticipation underpins the advancement of a comprehensive, integrated integrity framework at the Belgian federal level.

The FIA risk model was intentionally designed to support iterative and incremental enhancements. No risk model should remain static, and indeed, flexibility must be built in, for it to remain relevant. The CRI, which forms the foundation of the risk model, holds significant transformative potential for the Belgian public sector. For it to reach its full potential and evolve from a proof of concept into a strategic asset, a commitment to phased scaling and improvement is essential. Equally important is the need to collaboratively improve data quality and data governance among the various data providers. With sustained commitment, appropriate resourcing, and strong strategic leadership, this model has the potential to fundamentally reshape how public entities in Belgium conduct oversight and assess corruption and integrity risks.

The successful scaling and operationalisation of a data-driven risk model across federal domains is closely tied to the establishment of a coherent and widely supported Belgian Federal Data Strategy. This strategy—currently being developed collaboratively by the Federal Network of Data Experts (FEDAX) under BOSA's coordination—serves as a critical enabler for ensuring data availability, interoperability, and quality across government administrations. As an active participant in FEDAX, FIA positions the risk model as a tangible use case that illustrates how data-driven approaches can reinforce public sector integrity and enhance risk governance. Through this, FIA not only advances the Federal Data Strategy's implementation but also underscores the broader applicability of such models beyond auditing—for example, in compliance monitoring, service delivery, and evidence-based policymaking. Embedding risk modeling within the federal data ecosystem fosters institutional learning, capacity building, and systemic transparency, supporting the overarching goals of government-wide digital transformation and data maturity. Thus, the risk model stands as a clear example of how audit-led innovation can drive cross-sectoral impact through data-informed governance.

References

(CEPR), CENTRE FOR ECONOMIC POLICY RESEARCH (2021), *Procurement in Focus - Rules, Discretion, and Emergencies*. [34]

Acar, M. et al. (2015), "Government Favouritism in Europe: The Anticorruption Report, volume 3". [36]

Andersson, S. and P. Heywood (2009), *The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption", Political Studies, Vol. 57/4, pp. 746-767, https://doi.org/10.1111/j.1467-9248.2008.00758.x.*. [29]

Auriol, E., S. Straub and T. Flochel (2016), "Public Procurement and Rent-Seeking: The Case of Paraguay", *World Development*, Vol. 77, pp. 395-407, https://doi.org/10.1016/j.worlddev.2015.09.001. [33]

Ayogu, K. and R. Graffy (2022), *Using Data to Improve Procurement Operations and Outcomes*, https://gplpen.hks.harvard.edu/wp-content/uploads/2022/11/Using-Data-to-Improve-Procurement-Operations-and-Outcomes.pdf. [1]

Bataille, G. and V. Dor (2017), *New provisions regarding exclusion grounds in public procurement: corrective measures*, https://www.lexgo.be/en/news-and-articles/4505-new-provisions-regarding-exclusion-grounds-in-public-procurement-corrective-measures. [21]

Bauhr, M. et al. (2019), "Lights on the shadows of public procurement: Transparency as an antidote to corruption", *Governance*, Vol. 33/3, pp. 495-523, https://doi.org/10.1111/gove.12432. [13]

Caneppele, S., F. Calderoni and S. Martocchia (2009), "Not only banks: Criminological models on the infiltration of public contracts by Italian organized crime". [35]

Decarolis, F. and C. Giorgiantonio (2022), "Corruption red flags in public procurement: new evidence from Italian calls for tenders", *EPJ Data Science*, Vol. 11/1, https://doi.org/10.1140/epjds/s13688-022-00325-x. [32]

European Commission (2024), *BELGIUM – COUNTRY PROFILE – Benchmarking of national policy frameworks for innovation procurement*, https://ec.europa.eu/assets/rtd/innovation-procurement/country-report-2024-policy-benchm-belgium.pdf. [22]

European Commission (2023), *2023 Rule of Law Report - Country Chapter Belgium.*. [16]

European Commission (n.d.), *What is ARACHNE? https://employment-social-affairs.ec.europa.eu/policies-and-activities/funding/european-social-fund-plus-esf/what-arachne_en#:~:text=ARACHNE%20is%20an%20integrated%20IT,and%20European%20Regional%20Development%20Fund)*.. [25]

European Court of Auditors (2023), *Public procurement in the EU- Less competition for contracts awarded for works, goods and services in the 10 years up to 2021*. [15]

Fazekas, M., P. Lukács and I. Tóth (2015), *The political economy of grand corruption in public procurement in the construction sector of Hungary. Government Favouritism in Europe the anticorruption report, 3, pp.53-6*. [37]

Fazekas, M., S. Sberna and A. Vannucci (2021), “The extra-legal governance of corruption: Tracing the organization of corruption in public procurement”, Governance, Vol. 35/4, pp. 1139-1161, https://doi.org/10.1111/gove.12648.”, *Governance*, Vol. 35/4, pp. 1139-1161, https://doi.org/10.1111/gove.12648. [30]

Fazekas, M., I. Tóth and L. King (2016), “An Objective Corruption Risk Index Using Public Procurement Data”, *European Journal on Criminal Policy and Research*, Vol. 22/3, pp. 369-397, https://doi.org/10.1007/s10610-016-9308-z. [31]

Federal Overheidsdienst Justitie (2016), *Loi relative aux marchés publics”, Loi relative aux marchés publics*. [19]

GRECO (2020), *Fifth Evaluation Round: Preventing corruption and promoting integrity in central governments (top executive functions) and law enforcement agencies. Evaluation Report. Belgium, Council of Europe*. [17]

Hlacs, A. and H. Wells (2025), *Using digital technology to strengthen oversight of public procurement in Portugal: The use of data analytics and machine learning by the Tribunal de Contas, OECD Working Papers on Public Governance, No. 83, OECD Publishing, Paris*, https://doi.org/10.1787/43add03b-en. [9]

International Bank for Reconstruction and Development/The World Bank (2013), *Fraud and Corruption Awareness Handbook - A Handbook for Civil Servants Involved in Public Procurement*. [26]

Johnston, M. (2017), *Measuring the new corruption rankings: Implications for analysis and reform. In Political corruption (pp. 865-884). Routledge*. [28]

Lambsdorff, J. (2006), *Measuring Corruption – The Validity and Precision of Subjective Indicators (CPI). In Measuring Corruption, edited by Charles Sampford, Arthur Shacklock, Carmel Connors, and Fredrik Galtung, 81–100. Aldershot: Ashgate*. [27]

Mamavi, O. et al. (2014), *Out of sight, out of mind? Supplier spatial proximity in French public procurement. International Journal of Public Sector Management, 27(6), pp.486-500. *. [38]

Németh, A. and T. Tátrai (2015), *New warning system for the identification of red flags in public procurements*. [8]

OECD (2025), *Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence*. [10]

OECD (2025), *Government at a Glance 2025, OECD Publishing, Paris, https://doi.org/10.1787/0efd0bcd-en*. [14]

OECD (2025), *OECD Anti-Bribery Convention Phase 4 Report on Belgium: Implementing the Convention and Related Legal Instruments*, Implementing the OECD Anti-Bribery Convention, OECD Publishing, Paris, https://doi.org/10.1787/2a8a1ee9-en. [18]

OECD (2024), *Anti-Corruption and Integrity Outlook 2024*, OECD Publishing, Paris, https://doi.org/10.1787/968587cd-en. [24]

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. [6]

OECD (2023), *Managing risks in the public procurement of goods, services and infrastructure*, https://doi.org/10.1787/45667d2f-en. [4]

OECD (2022), *Towards Agile ICT Procurement in the Slovak Republic: Good Practices and Recommendations*, https://doi.org/10.1787/b0a5d50f-en. [5]

OECD (2021), *Countering Public Grant Fraud in Spain: Machine Learning for Assessing Risks and Targeting Control Activities*, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/0ea22484-en. [40]

OECD (2021), *Recommendation of the Council on Enhancing Access to and Sharing of Data, OECD/LEGAL/0463*. [11]

OECD (2019), *Analytics for Integrity: Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessments*, OECD Publishing, Paris, https://doi.org/10.1787/b354a27e-en. [39]

OECD (2017), *Recommendation of the Council on Public Integrity*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0435. [23]

OECD (2015), *Recommendation of the Council on Public Procurement, OECD/LEGAL/0411*. [2]

OECD (2008), *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information, OECD/LEGAL/0362*. [12]

Service public fédéral Stratégie et Appui (2023), *https://www.ejustice.just.fgov.be/eli/arrete/2023/04/18/2023042000/moniteur*. [20]

Williams, S. and J. Tillipman (2024), *Routledge Handbook of Public Procurement Corruption*, Routledge, London, https://doi.org/10.4324/9781003220374. [3]

World Bank (2023), *Governance Risk Assessment System (GRAS): Advanced Data Analytics for Detecting Fraud, Corruption, and Collusion in Public Expenditures*, World Bank, https://doi.org/10.1596/40640. [7]

BETTER POLICIES FOR BETTER LIVES