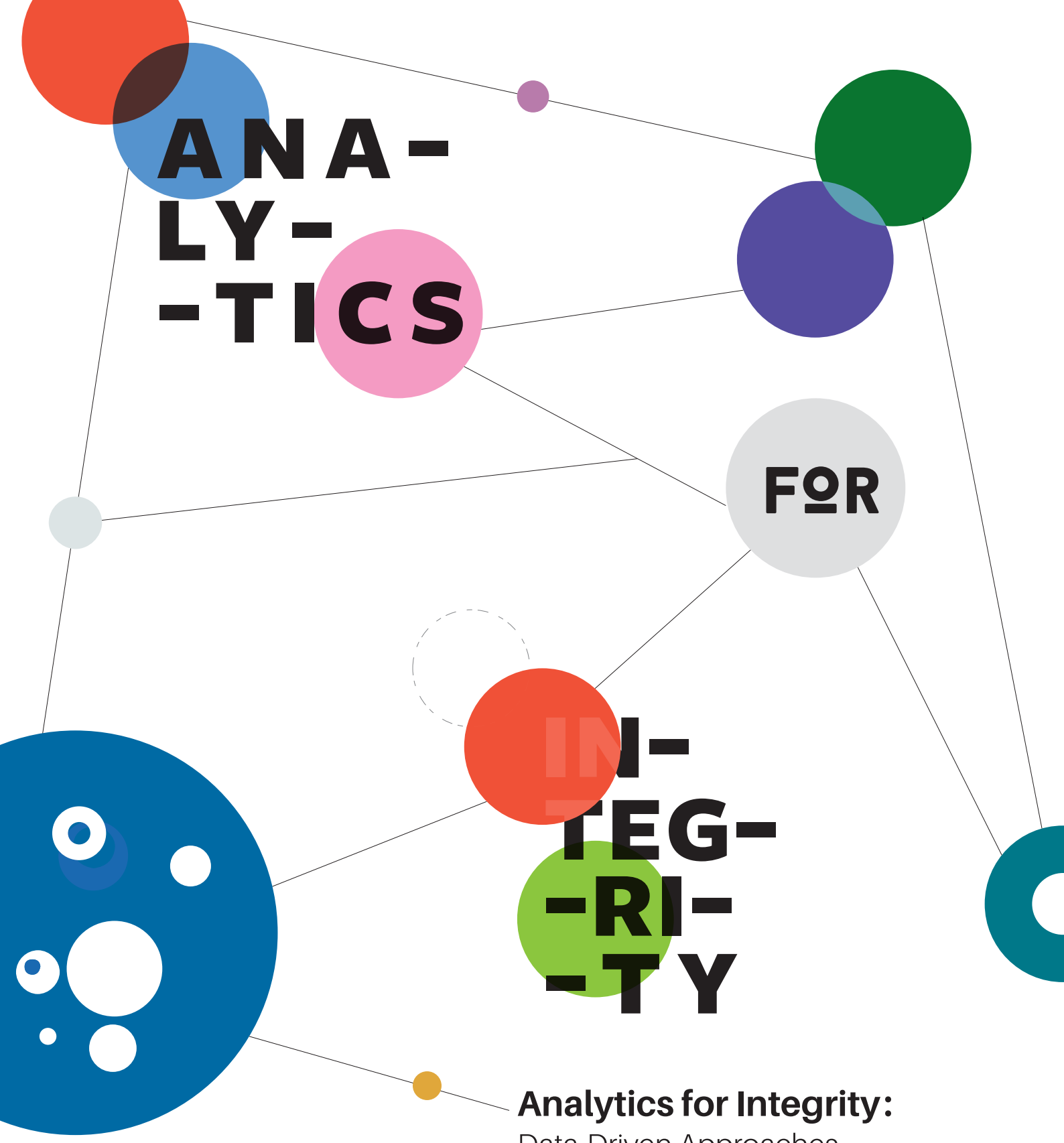


Analytics for Integrity:
Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessments

OECD

Analytics for Integrity

DATA-DRIVEN APPROACHES FOR ENHANCING CORRUPTION AND FRAUD RISK ASSESSMENTS

BETTER POLICIES FOR BETTER LIVES

Photo credits: Cover ©Azza Rajhi

Corrigenda to OECD publications may be found on line at: *www.oecd.org/about/publishing/corrigenda.htm*.

Acknowledgements

This report was co-ordinated by Gavin Ugale with insights and guidance from Janos Bertok and Julio Bacio Terracino of the OECD's Public Sector Integrity Division in the Public Governance Directorate (GOV). Angelina Zhao drafted the first section on risk management in infrastructure projects. Gavin Ugale drafted Section 2 on applications of data analytics for assessing corruption and fraud risks. Dr. Mihaly Fazekas, a consultant and Assistant Professor at Central European University, drafted Section 3 illustrating a practical application of data-driven corruption risk assessments for an infrastructure project in Mexico. Frédéric Boehm, Matthieu Cahen and Jacobo Pastor García Villarreal provided valuable comments. Declan Wylde, Head of Finance from Transport Infrastructure Ireland provided constructive input for Section 1. Section 2 benefited from the insights and comments of Barbara-Chiara Ubaldi and Jacob Arturo Rivera Perez of the Reform of the Public Sector Division in GOV, as well as Erin McLaughlin Villas, an independent consultant. In addition, Meral Gedik, Rania Haidar, Thibaut Gigou and Andrea Uhrhammer provided essential editorial and communications support.

The OECD also extends its appreciation to the individuals of the Airport Group of Mexico City (*Grupo Aeroportuario de la Ciudad de México*, GACM), who were responsible for the project of the New International Airport of Mexico (*Nuevo Aeropuerto Internacional de México*, NAIM). In particular, the OECD is grateful for the co-operation of the risk management professionals within GACM, and their input for Section 3 of this report.

Table of contents

Tables

Figures

Boxes

Executive summary

Managing corruption and fraud risks across the entire project cycle

Public infrastructure projects have many dark corners where corrupt actors can hide and thrive. The impact of such corruption is considerable. In countries with high levels of corruption, the quality of infrastructure tends to be relatively low, and access to services such as treated water or public health infrastructure is often limited.

Risk assessments, and more broadly risk management, can help managers anticipate and mitigate risks, as well as bring corrupt actors from the shadows. The OECD's *Recommendation of the Council on Public Integrity*, as well as international standards for risk management and internal control, highlight the benefits of risk assessments as a management tool. Ultimately, risk assessments should support managers in identifying control vulnerabilities and shaping mitigation measures appropriately. This key preventive function not only helps to curb corruption and fraud, but also waste, abuse and inefficiencies that can lead to project delays and cost overruns.

Assessing risks across the entire infrastructure project cycle—not just during the procurement phase—is challenging, given the number of stakeholders involved, but it is critical nonetheless. The OECD *Integrity Framework for Public Investment* highlights corruption risks across phases, and the OECD *Framework for the Governance of Infrastructure* explores key governance challenges. Building on this work, Section 1 focuses on principles, practices and considerations for managing corruption and fraud risks. The supporting annex elaborates on these OECD frameworks to aid managers in identifying such risks across the project cycle.

Data-driven risk assessments as a complement to traditional qualitative methods

The advent of a digital age in government has created new opportunities for assessing risks of fraud and corruption in infrastructure. For instance, the Korea Fair Trade Commission uses the Bid-Rigging Indicator Analysis System (BRIAS), to analyse large volumes of data from Korean public entities and create a probability score for bid rigging. In Chile, the government uses data mining of the e-procurement system to prevent collusion and favouritism. Such efforts have developed rapidly in recent decades, as governments adopt digital strategies and take advantage of open data, big data and data analytics.

Nonetheless, the reliability and accuracy of risk assessments is a fundamental challenge facing public entities. In infrastructure, project managers, risk managers, procurement officials, and oversight bodies, among others, often rely on qualitative methodologies for assessing risks (e.g. surveys and interviews for risk identification and scoring). These perception-based approaches typically result in risk inventories and maps that illustrate the perceived magnitude and probability of risks. Such approaches can provide critical insights and are an effective means for engaging individuals across the organisation to manage risks. However, corruption and fraud risk assessments also can be prone to biases and

inaccuracies due to a range of factors, such as the hidden nature of corruption and reluctance among employees to shine a light on wrongdoing. As described in Section 2, data analytics can complement qualitative methodologies, and improve managers' understanding about risks for more evidence-based decisions about mitigation measures.

Driving a better use of data analytics more broadly

Many countries face challenges in using digital technologies to encourage innovation, transparency, and efficiency in the public sector, in line with the OECD Recommendations on Public Procurement and Digital Government Strategies. These challenges can be the result of data quality, access and availability, as well as limited resources and skills for the effective use of data analytics. Linking data analytics to broader risk management objectives can help drive broader improvements in data governance, data infrastructure and the institutionalisation of an analytics function. As described in Section 2 of this report, this involves creating a data analytics plan with specific integrity objectives and selecting techniques in line with those objectives and available resources. The insights can apply to the use of various data sources, such as government data, open data and big data. The OECD's work to help the Airport Group of the City of Mexico (*Grupo Aeroportuario de la Ciudad de México, or GACM*) develop data-driven risk assessment and an analytics capacity demonstrated that improvements to data governance are possible in the risk management of large-scale infrastructure projects. Section 3 describes the methodology and results of the work with the GACM.

Data analytics for assessing corruption and fraud risks in infrastructure

Data-driven corruption risk assessments can help managers to identify the riskiest transactions and adapt control activities across the project cycle, including predicting high-risk transactions before spending. This report provide insights and examples for managers who are interested in using data analytics for this purpose by first exploring fundamental risk management practices, particularly in the context infrastructure projects (Section 1). The report then turns to the frameworks, key considerations and select techniques for using data analytics to support corruption and fraud risk assessments (Section 2). The report draws from research, interviews with experts and GACM officials, as well as the results of efforts to create a corruption risk index for a large-scale infrastructure project in Mexico (Section 3). Key insights and lessons learned from this effort include the following:

- The use of data analytics for risk assessments can complement qualitative methodologies, thereby reducing false positives and false negatives. However, data analytics is not a replacement for human judgement and professional scepticism.
- "Culture" is a critical factor of institutional readiness for using data analytics that includes not only the commitment of leadership, but also bottom-up experimentation at the project level.
- Establishing "quick wins" when using data analytics for the first time can promote buy-in and demonstrate the benefits of investing in data-driven methodologies, while helping to set realistic expectations about the value of data analytics.
- Investing in better data to enhance risk assessments can provide a context for organisations to address broader issues across the data value chain (e.g. data governance, collection, sharing, processing, etc.) to improve the use of data for decision-making.

1. Managing Corruption and Fraud Risks in Infrastructure Projects

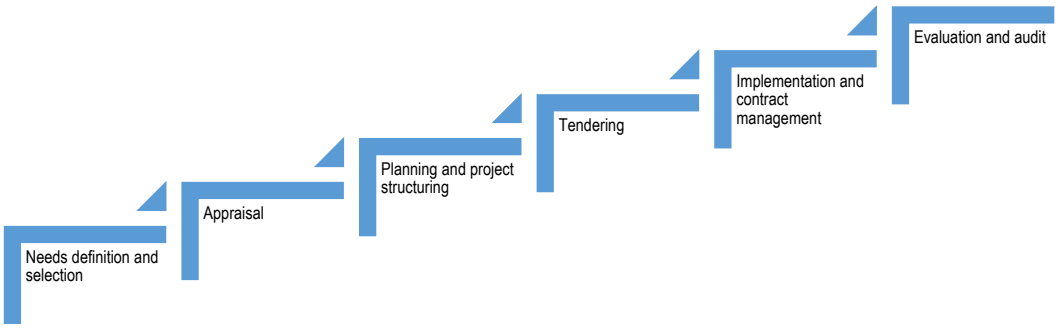
1.1. Introduction

In a global context of enormous infrastructure gaps, the ability to attract investments become imperative for both OECD member countries and non-member countries. Recent OECD surveys and reports in the Asia-Pacific Economic Cooperation (APEC) and Latin American Caribbean (LAC) regions identified that one of the top three barriers to infrastructure investments is bribery and corruption. Infrastructure projects, due to their complex nature, provide many opportunities for corruption and fraud risks across the project cycle. These risks can be of criminal nature, such as fraud and corruption, or could be non-criminal risks or unethical behaviour that could undermine integrity, such as nepotism and exerting undue influence. These risks can lead to inflated construction prices, significant delays and lower quality outputs, among other effects. Preventing and combating such risks are essential to the successful delivery of infrastructure and promoting diverse financing opportunities. The principles and practices for risk management, linking to the broader strategic objectives and internal control system, are an integral component for effectively managing corruption and fraud risks in infrastructure projects.

The maturity and efficacy of risk management within infrastructure projects can vary significantly, driven in part by the multi-stakeholder involvement for delivering infrastructure. Moreoever, governments and project managers can overlook corruption and fraud risks, manage them ad hoc or do not pay sufficient attention to them as part of management decision-making processes. Yet, risk management and tailored risk assessments should be core features of infrastructure projects and integrity systems as a whole, as highlighted in the OECD's *Integrity Framework for Public Investment* and the OECD *Framework for the Governance of Infrastructure.* Among other benefits, risk management and risk assessments support officials and managers to advance strategic priorities based on risks, allocate resources effectively and establish proportional controls to detect and prevent corruption and fraud.

This section explores how risk management principles and practices can support managers of infrastructure projects to identify, analysis and respond effectively to corruption and fraud risks across the project cycle. Studies often focus heavily on corruption risks in the procurement process; however, corruption and fraud can occur at any stage. See Figure 1.1 below for a general description of the phases of the infrastructure project cycle. Each phase involves distinct and numerous institutions and stakeholders, which result in different risk environments and vulnerabilities to different types of fraud and corruption schemes. For example, conflict of interests and skewed incentives can cause the selection of poor projects at the appraisal stage, which generate corruption risks at subsequent phases of a project. The section helps to strengthen the understanding of corruption and fraud risk schemes across the project cycle, and highlight fundamental principles and practices of risk management to address those risks.

Figure 1.1. Infrastructure project cycle



Source: Adapted from (OECD, 2016b).

1.2. Overview of corruption and fraud risks across the infrastructure project cycle

1.2.1. Needs definition and selection

At the initial stage of infrastructure investments, many different actors can play a role to define infrastructure needs and selection criteria, such as lobbyists, trade unions, regulators, non-governmental organisations (NGOs) and potential contractors (OECD, 2016b). For example, in the case of a need to alleviate congested traffic hotspots, the assessment could involve numerous stakeholders to determine whether to build more roads/highways or to improve public transportation by adding buses and rail. Defining the needs requires evaluating many factors, such as the site location, the arrangement of routes and choice between heavy and light rail. The following are examples of corruption risks, whether criminal and non-criminal, at the early stage of the project cycle:

- **Policy capture and influence** – The entity responsible for selection chooses a particular interest group, business or individual over public interest, because of undue influence, such as political pressure, political campaigning and lobbying power.
- **Conflict of interests and nepotism** – The selection of investments benefit contractors and private operators administrated by public official's family members or people of allegiance (based on previous employment and business relationships).
- **Bribery for access to confidential information** – The selection of projects takes place because of government officials accepted bribery to disclose confidential information on policy priorities.

1.2.2. Appraisal

During the appraisal phase, the government evaluates an infrastructure project's feasibility, which can consist of a cost-benefit analysis, a business case study, as well as environmental, social and economic assessments. Appraisals can also include a justification of the project rationale, clarifying objectives, review of several alternative options to fulfill the goals and an assessment of the commercial viability and long-term affordability. Governments may hire consultants with specialized expertise to undertake the feasibility study. Before the final approval of the project, the government can also determine how and by whom the project will be financed, as well as the required public

funds (OECD, 2016b). Several common corruption and fraud risks present in this phase are:

- **Bribery to undermine merit-based procedures** - Investors may bribe government officials to win contracts instead of winning contracts based on merit, such as their financial resources and relevant project experience.
- **Fraudulent assessments** - Appraisers intentionally manipulate outcomes of social, economic and environmental feasibility studies, or the public officials conceal and present false conclusions of the assessments.
- **Promoting high-cost projects** - Potential private operators of a public-private partnership downplay risks associated with contract management to favour large and new projects over maintenance of existing infrastructure, and therefore increase the future financial burden on public funds.

1.2.3. Planning and project structuring

Once a project is selected, a detailed project design should be created with an adequate budget that indicates estimated project costs. Terms of references and bidding documents are developed to state expected deliverables. Moreover, at this stage, the project owner determines the details and specifications of the work to be undertaken and evaluation criteria for the bidding process. The planning phase may have significant impact and create opportunities for corruption and fraud in ensuing phases of infrastructure implementation (Wells, 2015). The following are examples of corruption and fraud risks in this phase:

- **Tailoring specifications** - The design of tender documents and specifications are technical in nature and involve external experts who are not qualified or can be unduly influenced, resulting in restrictive and tailored specifications to favour certain bidders.
- **Budget manipulation** - The document proposes costly designs with overestimated budgets, which increase contractors' potential profits, or alternatively, includes undervalued budgets to manipulate contract procedures at a later stage to reduce competition.
- **Asymmetrical information** - Create opportunities for specific actors to obtain non-disclosed information regarding the tendering design, and restrain conditions for other potential participants to receive the tender documents.
- **Vague criteria** - Selection and award criteria are not clear, objectively defined, or do not relate to the nature of the contract, leaving room for inappropriate adjustments later.
- **Contract splitting** - Public contractors split one contract into several ones below a threshold to avoid competitive procedures and enhanced controls. Alternatively, contracts including incompatible services and supplies can also be grouped to dissuade potential bidders.

1.2.4. Tendering

The tendering stage can be divided into submission, evaluation and awarding of contracts. Potential contractors submit their bids, which are evaluated against their technical and cost proposal. Project owners select contractors based on established criteria. During this phase, responsible government institutions and project owners interact formally with the bidders

and potential contractors, which can create opportunities for corruption and fraud (OLAF, 2018). Examples of corruption and fraud schemes that can compromise the tendering process include the following:

- **Manipulating procedures and rules of the game** - The contracting authority limits competition by manipulating the tendering process to favour closed contract procedures, limit advertising and publicity to the calls, and set unrealistically short timeframes for bidders to respond to the calls.
- **Bribing to gain an unfair advantage** - A bidder bribes a public official or a consultant involved in project design to obtain confidential documents resulting in asymmetry of information for all potential contractors.
- **Colluding among bidders** - Several bidders conspire together to limit competition and raise prices through various collusion schemes including complementary bidding, bid rotation, market division and bid suppression.
- **Undermining evaluation criteria** - The Evaluation Committee, because of conflicts of interest or bribery, manipulate the evaluation criteria and direct the selection process to favour one particular bidder.
- **Setting unreasonable expectations and timelines** - The contracting authority sets an unreasonable short duration between the submission deadline and contract award date, and allows for contract modifications during advertisement period.
- **Misrepresenting profile and fraudulent documents** - The contract winner misrepresents their financial and technical capability and obtains the agreement by submitting fraudulent bidding information.

1.2.5. Implementation and contract management

After the contract work is awarded, the project moves to implementation, which involves the construction and operation of infrastructure. The phase requires finalising the contract, closing the financial agreements and executing the contracted work. The project owner must allocate management and oversight responsibilities to ensure proper monitoring and supervision of tasks and outputs specified in the contract. Since infrastructure projects are complex and often entail multiple years of construction, appropriate mechanisms for checks and balances should be in place to evaluate the project progress and completion. (OECD, 2016b) Specific corruption and fraud activities may occur at this stage, which can be subject to less scrutiny, as procurement regulations often do not cover this phase:

- **Fraudulent reporting and claims** - The contractor manipulates cost claims or invoices to bill inflated costs or recharges incurred costs using false invoices, false reporting of labour time and other fraudulent documents.
- **Violating contract conditions** - The contractor can violate the contract conditions by non-delivery of agreed products or supplying services of a lower quality than required.
- **Renegotiating terms after the fact** - The terms of references can be renegotiated to deviate substantially from the initial requirements stated in the award of the contract
- **Faking the work and approvals** -The contractor provides fictitious work and bribes the public official or consultant to approve defective or non-existent work.

1.2.6. Evaluation and audit

The infrastructure project cycle closes with evaluation and audit to ensure that adequate internal controls applied throughout the project cycle. The reviewing body should be independent of the public organisation that initiated the procurement process. Governments or project owners should define the evaluation framework from the start to capture necessary information related to the contract execution (Robson, 2010). If corruption and fraud already took place, the relevant responsible actors can misrepresent activities and conceal results in various ways. Examples of corruption and fraud risks, both criminal and non-criminal, in this phase include the following:

- **Fraudulently documenting results** - Stakeholders forge documentation and falsify information to have positive evaluations by auditors.
- **Compromised evaluators and auditors** - Auditors and evaluators are bribed to overlook violations of controls and suspected fraud and corruption in project closure.
- **Undermining the evaluation function** - Auditors and evaluators may also simply lack independence, or intentionally be under-resourced, making it impossible for them to come to legitimate findings and fulfil their mandate.

1.3. The control environment and policies for managing corruption risks

This section details how risk management and internal control can effectively support project managers to identify, mitigate and manage the aforementioned risks. The critical components of internal control and risk management for safeguarding public sector integrity, as described in Principle 10 of OECD Recommendation on Public Integrity, include the following.

a. A **control environment that explicitly focuses corruption risk management** including the appropriate policies, processes, and structures that underpin a culture of integrity;

b. A **strategic approach to assessing corruption and fraud risks** that ensures effective resource allocation and that control activities are proportional to risks;

c. Well-defined procedures and mechanisms for a **co-ordinated response**, including corrective actions, to corruption and fraud risks and **reporting of suspected violations**.

d. **Regular monitoring and evaluation** activities to ensure that the framework is functioning effectively and is responsive to current and emerging corruption and fraud risks.

The first two components form the backbone of an effective system for managing corruption risks in infrastructure projects and are discussed in depth below. In practice, corruption and fraud in infrastructure delivery can take place at two distinct levels: ministry/institutional level and at the project level.

- At the ministry level, the decision-making is associated with policy on major public works and infrastructure needs. As discussed, risks of corruption are often in forms of favouritism, collusion, and improper influence in the planning and appraisal process of deciding the projects in which to invest. Within a particular ministry responsible for infrastructure planning (i.e., the Ministry of Public Works or

Ministry of Transportation), corruption risk management can guide infrastructure decisions and provide assurance that a pipeline of projects are identified, prioritised, evaluated and delivered in the public interests, while minimising incidents of corruption, fraud or other types of risks.

- At the project level, individual projects for infrastructure are usually large-scale, complicated and implies tremendous public resources and risks. Each project, therefore, requires dedicated attention and adequate corruption risk management. Many of the aforementioned risks following the planning and project-structuring phase, as described above, occur at this level. As such, the project owners and managers play a critical role in addressing these risks.

1.3.1. A control environment with a focus on integrity

The quality of corruption risk management for infrastructure projects relies on a sound control environment. The project control environment is comprised of people, policies and processes to ensure project risks, especially integrity issues, are mitigated, and the project objectives are achieved.

Ensuring effective risk management structure

Senior management of an infrastructure project has the primary responsibility of creating and maintaining a control environment, including taking into account integrity objectives. Management for infrastructure projects can include the executive project committee and risk management committee responsible for the design, implementation, and monitoring of internal control and risk management practices. Project managers can establish the tone at the top and raise awareness of high-risk areas in projects by facilitating periodic training workshops to articulate individual responsibilities for managing risks.

Institutions may nominate existing risk or project managers to carry out the tasks of corruption risk assessments, integrating the process into ongoing risk management activities. The size of the project (including the number of stakeholders, employees, and resources) and the complexity of risks, among other factors, help to determine if a dedicated function would necessary. Moreover, due to the involvement of a multitude of stakeholders, government entities, implementing agencies, contractors and suppliers could have different standards and frameworks for managing risks.

For instance, the OECD and the International Partnership Against Corruption in Sports carried out a study on procurement and risk management standards in the delivery of infrastructure for large-scale sporting events. In this context, an array of public and private stakeholders contribute to the building of infrastructure, many of whom have their own risk management policies and practices. For these large scale and complex projects, a designated function for managing corruption and fraud risks could be beneficial. In some projects, a risk committee can help to oversee, co-ordinate, manage, monitor and evaluate risk management activities concerning corruption risks across the entire project lifecycle. Infrastructure projects require harmonising and blending various frameworks to ensure a consistent risk management structure and process transitioning from one stage to the next stage. Finally, audit and independent inspection entities can also play an important role to ensure accountability, integrity and quality in the delivery of infrastructure projects.

Establishing corruption and fraud risk management policies

Public sector institutions responsible for delivering infrastructure can embed corruption risk governance in their organisations and projects in the form of written policies, assigned responsibilities, and on-going procedures. Separate corruption risk management policies can be developed, with specific integrity objectives, especially if such risks are perceived as excessive and warranting high-profile attention. The provisions can also be incorporated into existing risk management policies as part of the project control environment. Risk management policies with explicit reference to integrity are a useful tool for management in infrastructure projects to demonstrate priorities given to integrity. Project managers who incorporate risk management and integrity elements into policies can consider:

- articulating the main strategic objectives and state the organisation's commitment to risk management across the lifecycle in supporting the project delivery;
- defining fraud and corruption risks and examples that are deemed corrupt or fraudulent which are pertinent to the project; (refer to section 2.1 for examples)
- stipulating to whom the policy applies, which can include project owners or sponsors, project managers, staff, financiers, contractors, subcontractors, third-party suppliers/agents and consultants; external stakeholders should agree to adhere to the policies as a part of the contract and service level agreement;
- establishing a governance and oversight structure by defining roles and responsibilities for internal control and risk management for the project;
- communicating a risk strategy, risk tolerance relative to the project objectives, and allocate appropriate resources for risk management; and
- outlining key areas of risk management process.

In general, risk appetite can be defined as the amount of risk an entity or a project owner is willing to accept in pursuit of value and objectives. In the case of infrastructure delivery, the primary project objective is successful development, completion and operation of infrastructure assets on budget and within a reasonable time. Other objectives found in a risk appetite statement may include strategic, operation, community value, environment impact, the resilience of infrastructure and organisational governance. For example, an agency responsible for implementing infrastructure in Australia indicates their risk appetite in a statement, which stipulates a very low risk appetite for corruption, fraud and non-compliance behaviours that undermine the integrity of the organisation. (Gold Coast waterways authority, 2017) Risk appetite framework should guide resource allocation by directly aligning people, processes and technology to support the infrastructure delivery and to effectively respond to and monitor risks as the project moves toward achieving its strategic goals throughout the infrastructure lifecycle. Reputation risk, in particular, must be managed to ensure trust in government's ability to provide high-quality infrastructure. Fraud and corruption can undermine trust in public institutions. Therefore, infrastructure projects sometime require an explicit fraud/corruption risk appetite, which help determine the necessary mitigating processes and controls. The figure below shows how various elements of risk appetite framework can work together:

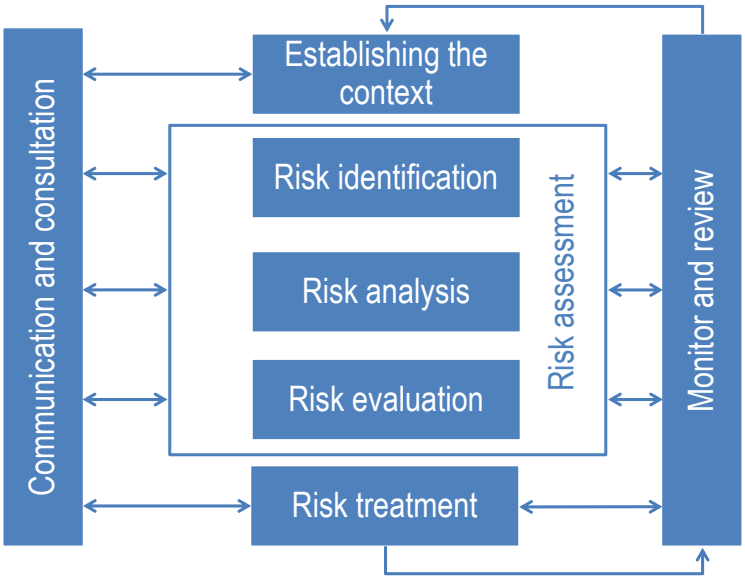
Corruption and fraud risk management policies should not serve as a checklist to comply with minimum standards. They should be comprehensive and tailored to projects, addressing current and emerging corruption risks throughout the lifecycle of the project. The project leader can elaborate on the importance of managing corruption risks in

supporting the project's strategic objectives. For instance, mitigating corruption and fraud risks in public procurement improves the use of high-quality goods and services at a fair cost for the public, and therefore reduce poor contract performance and service delays.

1.4. Institutionalising corruption and fraud risk assessments

Corruption and fraud risk assessments can be stand-alone exercises or embedded in a project's broader risk assessment activities, recognising the interlinkages between strategic, operational, financial and reputational risks and control activities. The risk assessment should reflect whatever approach works best for the project. However, project objectives should include an explicit focus on managing corruption risks, as indicated in international standards. For instance, COSO's Internal Control – Integrated Framework, revised in 2013, included a new principle for organisations to consider the potential fraud in assessing risks to the achievement of objectives. Figure 1.2 illustrates a general framework for risk management and the risk assessment process, which are adapted to the context of managing corruption risks for infrastructure projects in the text that follows:

Figure 1.2. Framework for managing and assessing risks



Source: Adapted from (ISO, 2009).

Establishing the context for managing corruption and fraud risks

When designing a risk management framework, project managers could first assess the internal and external contexts to understand the drivers and potential impediments to achieving integrity objectives. The internal context includes, but is not limited to, the governance structure, roles, employees' skill sets, operational tools (e.g., data and information systems), culture and internal guidelines and previous experiences of corruption and fraud incidents for similar projects. Additionally, project characteristics comprised of project size, uniqueness, the extent of government involvement, technical and organisational complexity help establish the internal context for pinpointing corruption and fraud risks. (Locatelli Giorgio, 2017)More specifically, various infrastructure delivery methods such as design-bid-build, design-build, construction management and public-

private partnership (PPP) and concessions can shape the types of risks to consider. The external context may include but is not limited to, geographical jurisdictions, legal and policy frameworks, number, complexity and relationship network of external stakeholders, as well as political, social and economic realities. Understanding these contexts forms the basis for both designing and improving strategies for managing corruption risks. Establishing the context throughout the entire infrastructure process also supports the identification of appropriate risk owners and forming a team for assessing corruption and fraud risks across a project.

Identifying, analysing and evaluating fraud and corruption risks

Risk assessment and management has grown to become an essential part of managing infrastructure delivery as projects manage various risks and allocate resources to control activities using a risk-based approach. (Dongping, 2008)However, some organisations do not explicitly identify or anticipate specific corruption risks and thus cannot effectively curb related financial costs, adverse economic impact and safeguard project integrity. Targeted corruption risk assessments allow for a comprehensive understanding of both inherent and residual risks, including an assessment of internal and external corruption and fraud schemes.[1] Since fraud and corruption can occur throughout a project, assessments should encompass all stages of the infrastructure life cycle and are inclusive of all stakeholders. Institutions managing infrastructure projects should regularly assess risks and tailor the assessments to their environment and the stage of the project cycle (Beckers, et al., 2013).The various phases of the process are interlinked and the risks evolve in a dynamic way. Based on what occurred during the appraisal and planning stage, corruption and fraud risks at the tendering may alter. Therefore, it is critical that policies and procedures for evaluating and reporting corruption and fraud risks stipulate frequencies at appropriate intervals to provide an informed picture of the project risk profile.

No uniform methodology exists for performing risk assessments, but approaches generally can be qualitative, quantitative or a mix of both. The maturity of the institution's risk management and its capacity of data analytics will help to determine the approach, and in some instances, limit the options. For example, quantitative methods for using data and analytics to assess risks rely on the availability of data and data quality and robust data analysis techniques. In the infrastructure context, particularly during tendering phases, legal and regulatory reforms for opening data and improving the quality of procurement data offers opportunities for conducting data-driven risk assessments given the right conditions, as described in subsequent chapters. Nonetheless, there are still limitations stemming from data availability and quality, given the complexity of infrastructure projects and the hidden nature of fraud and corruption. Qualitative, perception-based approaches and practitioners 'experiences may be the sole choice in some contexts.

Regardless of the method, corruption risk assessments starts with identifying relevant risks for the project. Desktop research, interviewing employees, potential contractors and other concerned parties, undertaking control risk assessments, analysing audit outcomes of similar projects, and conducting process gap analyses are some possible ways for identifying potential risks. Other techniques can include consulting the country's or the organisation's previous projects to identify common trends and schemes that are indicative of fraudulent or corrupt activity. Annex A also provides a summarised list of frequently observed corruption and fraud schemes in infrastructure projects to guide readers to establish fraud scenarios relevant to their contexts. The risk identification process requires the inputs from all stakeholders including financiers, project managers, frontline staff, contractors, subcontractors, third parties due to the nature of multiple interfaces in a project.

The complex interrelationships among a multitude of project partners require organisations to establish sound risk management arrangements to identify and manage risks with all major partners, (Treasury, 2009). A risk register can serve as a tool to record identified risks and present an inventory of risks that could threaten the delivery of the project.

Once managers catalogue relevant risks, they can employ qualitative risk analysis to assess and determine the likelihood and impact of inherent corruption and fraud risks based on their experiences at its most basic level. Standard tools such as a risk matrix and risk rating criteria can inform the risk evaluation process and guide decisions about the type and priority of treatment, as well as the urgency of action. For example, at Transport Infrastructure Ireland, project managers initiate several workshops to map out critical risks including corruption risks concerning a project at the planning stage. At these risk workshops, project managers invite risk specialists within the organisation, experienced project team members with sector expertise, specific anti-corruption and fraud advisors and external consultants to determine a list of relevant risks and evaluate the probability of likelihood and impact in a rigorous manner. Risk managers need to establish specific criteria for assessing corruption risks in a project, which can encompass:

- Financial and value-for-money issues
- Service delivery and quality
- Public concern and trust
- Reliability of evidence on the risk
- Financial, economic and reputational impact on stakeholders including the public

When data are available, reliable and valid, quantitative methods for assessing risks and applying analytical techniques can help to complement an organisations qualitative assessments. Many applications of data analytics concentrate on detecting a large number of suspicious fraud cases, which can sometimes overwhelm an organisation and its resources. In this context, integrating data analytics with risk assessments can elevate the rigor of risk rankings and improve corruption and fraud risk prevention, thereby narrowing the pool of cases for potential follow up.

As detailed in the Section 2, data analytics can provide more objective measures of risk probabilities of potential corruption or fraud, and help managers to understand how past, present and emerging fraud and corruption risk schemes relate to current risk drivers and indicators. For example, techniques for analysing networks can help to identify connections between actors and entities in the planning and selection phases of infrastructure projects. During the tendering stage, regression and statistical modelling can be applied to evaluate indicators related to different contractors and types of contracts to spot red flags for corruption and fraud schemes. Additionally, pattern recognition and cluster analysis may lead to recognition of certain risks such as collusive bidding which often remain undetected, (OECD, 2018b). For any approach, it is important for entities to have clearly defined goals and an understanding of the resources, skills and cost-benefit trade-offs for developing a data analytics capacity. The following section will offer details regarding a strategic approach to using data analytics for assessing corruption and fraud risks.

Aligning risk mitigation and control activities

After managers document the inherent risks, they should map the risks to an associated control activity, which is then recorded in the risk register. The effectiveness of existing controls in managing those corruption risks can be assessed to determine how proportional

and effective the controls. In some cases, managers may conclude that there are no adequate controls for an identified risk level. In a more advanced stage of corruption risk management, the inherent risk assessment can help tailor control activities to focus on higher-risk areas, therefore, driving a risk-based and iterative control process resulting in increased efficiency for managing infrastructure projects. The residual risks are the net risk exposure the project still faces after applying the mitigating controls and procedures.

Risk treatment

After assessing its corruption risks, the next step is for institutions to determine how to respond to the residual risks, i.e., the risk treatment. The organisation's risk appetite framework for the project is critical for guiding this process and consideration of the feasibility and cost-benefit analysis of available control measures. As described above, n risk tolerance represents the specific maximum risk that an organisation is willing to accept relative to the institution's integrity objectives and available resources in a project. The risk tolerance has practical applications for the risk treatment in that it helps to guides managers in their decision to accept, reduce, avoid or share the risk. In the integrity context, a "zero risk tolerance" can be a compelling message for promoting a culture of integrity, but may not serve the practical purpose of determining risk treatments since resources for mitigating measures are always constrained. A defined risk appetite framework with risk tolerance and limit helps organisations to make decisions based on an analysis of control effectiveness and the remaining risk exposures. In some cases, the residual risk may exceed the tolerance, which indicates the priority of risks for treatment and mitigation to bring these higher net risks within the acceptable level.

Note

[1] Inherent risks are risks assessed in the absence of control measures. Residual risks refers to perceived risk exposure after applying mitigation strategies and controls.

Annex 1.A. Mapping of corruption and fraud risks across the project cycle

Needs Definition and Selection Phase

Corruption risk schemes	Potential Actors	Possible risk indicators	Example of mitigating controls	Risk owners
Select project based on undue political influence including political pressure, campaign contribution and lobbying power	High-level public officials for defining the project , lobbyists, trade unions, and potential contractors	Abnormal level of political campaign donation from a prospective contractor or imbalanced lobbying power indicated by unequal level of access	• Ban and limit certain types of private contribution: foreign corporation, trade unions and corporations with government contracts •Disclose all political funding and ensure independent oversight	Public committees overseeing the investment decisions, regulators, potential competitive contractors and media
Favour a project due to conflict of interests such as family ties, previous employment or business relationships	Same as above	Red flags that suggest undeclared personal or business relationships between public officials and private operators	• Establish an independent body responsible for assessing infrastructure needs which provides monitoring and oversight • Guidelines and enforcing rules against COI and set up regulations on revolving doors.	Same as above
Disclose confidential information on policy priorities because of bribery or trading of favours	Same as above	• No active engagement of the public to inform infrastructure priorities. • Limited opportunities for consultation regarding long-term plans	• Establish transparent decision-making process by making information such as reports on long-term plans available through public channels • Invite all relevant groups for public consultation	Same as above

Appraisal Phase

Corruption or fraud risk schemes	Potential Actors	Possible risk indicators	Example of mitigating controls	Risk owners
Bribe to obtain the award of financing contracts which is not based on the capacity to finance and project experiences.	Public officials at all levels of government, financial stakeholders, potential contractors, possible private operators in case of PPP, consultants, lobbyists	• Only a very few number of banks and financing providers are considered • Unclear criteria and justification of awarding the financial contract • Due diligence shows the FI lacks capacity and experience to finance	• Ensure specific evaluation criteria for awarding financial institutions and regulate interaction between banks and public entities for infrastructure • Prohibit public officials from receiving payments or gifts from the potential financiers.	Senior management within implementing institutions, Centre of government (in setting standards)
Manipulate social, economic and environmental feasibility studies or present false conclusions to ensure project investment is approved.	Same as above	• Feasibility studies carried out by unauthorised or unqualified experts/consultants • Inconsistent feasibility reports with significant modifications	• Assess the qualification, impartiality and competence of the experts and consultants in charge of assessment studies and ensure due diligence on their work • Provide public consultation process with feasibility studies • Provide an independent appraisal review of the feasibility study outcomes	• Middle manager, internal audit group, risk management function within implementing ministry/agency • Independent and centralised ministry of finance or planning
Perform improper risk assessment associated with a PPP to favour large and new projects over maintenance	Same as above	• Risk assessment appears overly optimistic and varies substantially from the risk allocation of similar infrastructure projects.	Institutionalise standards for risk assessment that limit discretion Provide periodic audit of risk assessment practices for PPP	Same as above

Planning and Project Structuring Phase

Corruption or fraud risk schemes	Potential Actors	Possible risk indicators	Example of mitigating controls	Risk owners
Distort the design of tender documents by including restrictive and tailored specifications to favour certain bidders.	• Civil servants at the government ministry • Procurement officers • Private consultants (engineers, designers, planners)	• Narrow description of requirements to exclude potential candidates • Established companies in the field are rejected or screened out • Close similarity between specifications and the winner's proposed services or complaints from potential bidders that the specifications match closely those of a single bidder.	Verify a sample of contracts ensure that technical specifications are not too narrow in comparison to services required for the contracts s	Review committee should examine the design and estimates of projects prepared by consultants
Propose costly designs with overestimated budgets or provide undervalued budgets to later manipulate procedures	Same as above	• Budgets are larger than similar projects based on past experiences and examples from other countries • Very few bidders responded to the call due to the undervalued budgets	• Ensure budgets are realistic and approved on time • Create separate checks and balances including independent verification of designs and budgets	Budget committee can provide formal approval of the proposed budget
Establish unclear and irrelevant selection and award criteria unrelated to nature of the contract leaving room for inappropriate adjustments	Same as above	• Ambiguous description of required goods, works and services leaving room for modification • Unnecessary items included in terms and conditions	• Embed control to provide secondary opinion on the established selection and award criteria	Independent experts can evaluate bidding documents including specifications and terms of references for high-risk and large-value contracts
Split one contract into several ones below threshold to avoid competitive procedures or bundle incompatible services and supplies to dissuade potential bidders.	Same as above	• Several contracts are merely under threshold for competitive procedures over a short period of time and services provided and goods delivered are similar • Repeated purchases and contracts awarded to the same company with less competitive procedures. • Cannot justify the bundling with the rationale of cost saving or risk reduction. • Complaints by one or more potential bidders for incompatible groupings of goods and services	• Regularly review a list of proposed contracts just under thresholds. • Requires internal audit over the controls of the public procurement process	Designated staff within public procurement agency can compare the procurement plan with the project appraisal document to determine any inconsistencies.

Tendering Phase

Corruption or fraud risk schemes	Potential Actors	Possible risk indicators	Example of mitigating controls	Risk owners
Limit competition by manipulating the tendering process to favour non-open contracts, minimising advertising and reducing time for potential bidders to submit	Procurement Officers, Private consultants and contractors	• High concentration of non-open or alternative contract procedures • Contracts dominated by single bidders or very few bidders compared to prior similar tenders • No publication of a call or limited circulation of advertising, or period between the advertisement and the bid submission deadline unusually short	• Review the bids from the losing bidders and compare them to the bids in the bid evaluation files • Review the bid evaluation reports, notes and minutes prepared by the bid evaluators and note disputed changes to scoring • Review past advertisement and length of the advertising period during post procurement reviews	• Procurement manager can provide post-review on a continuous basis • Internal auditor can also include the controls as a part of their periodic examinations.
Bribe public official or consultant to obtain confidential information	Same as above	• A bid closely resembles the project's preferred budgets, design and solutions • Public officials and the favoured bidder communicate excessively during the bidding period	• Limit the opportunities to socialise or communicate between public procurement agents and potential bidders during the bidding period • Provide channels for potential bidders to report suspicions related to leaks of confidential information	PR and communication team of the contracting agency can set standards of communication guidelines with potential bidders.
Collude among bidders to limit competition with schemes including: bid rotation, market division, bid suppression, and complementary bidding	Procurement Officers, Private consultants, contractors and other third-party bidders (designated winner and appointed losers)	• Low number of bidders • Unusual bidding patterns • Complaints from losing or excluded bidders • Rotation of winning bidders with predictable trends • Losers of bids become subcontractors of the winning bidders which demonstrate persistent arrangements	• Automated tests to detect collusive bidding, e.g., compare line item and total bid prices to cost estimates, prior bid prices, industry averages, etc.; note significant price increases • Look for unusually high line bid prices in losing bids (e.g., bid prices at least 50% higher than competitor's prices) • Due diligence background checks on suspect firms • Collect and examine all proposals from several other or prior similar projects to analyse bidding patterns	• IT department and data analytics team can incorporate the automated red flag tests to detect collusive bidding. • Procurement manager receive reporting and manage the risks of collusive bidding among potential bidders
Manipulate the evaluation and selection process to favour one bidder	Same as above	• The evaluation criteria are not clear or are unreasonable	• Establish guidelines and procedures for evaluations	The evaluation committee and senior managers that nominate members to the committee

Corruption or fraud risk schemes	Potential Actors	Possible risk indicators	Example of mitigating controls	Risk owners
		• The evaluation committee is too large or controlled by an individual • The members of the evaluation committee do not have the technical expertise to evaluate the submitted bids	• Review bid evaluation reports for justification for rejecting submitted bids and look for inconsistencies in scoring criteria • Confirm the evaluation committee has adequate number of members with qualification and technical expertise and ensure no members have a conflict of interest in performing their responsibilities.	
Set unreasonably short or long time between submission deadline and contract awarding and allow for contract modifications	Same as above	• Unreasonable decision period that is excessively short or lengthen compared to the norm • Whether the contract is modified during the advertising and delivery stages	Track the decision periods and contract modification of all bidding processes and review the abnormalities for further follow-up	Procurement manager or risk manager can perform the control measures
Misrepresent ownership, financial and technical capability and submit fraudulent bidding information	Same as above	• No information is available concerning the bidder on the internet or in telephone and business directories. The bidder is a shell company. • The reported financial information contradicts figures from other sources and the audit reports are not attested or signed. • Discrepancies exist between self-reported technical capability and information provided by references	• Require standard due diligence consisting of search through databases and internet and verify documents on the bidders and bidding information • Train operational staff to better identify fraudulent bids	Procurement officers and manager in charge of initial review of bidding information

Implementation and Contract Management Phase

Corruption or fraud risk schemes	Potential Actors	Possible risk indicators	Example of mitigating controls	Risk owners
Manipulate cost claims/invoices to bill inflated costs with false invoices, false reporting of labour time and fraudulent documents.	Contractors, subcontractors, project officials, supervision agents	• Discrepancies between work statements and supporting documents of labour time and site visit checks • Invoice amounts, quantity and description do not match contract and purchase orders. Invoiced goods and services cannot be accounted for or located in the inventory	• Compare invoices with other support documents before payment disbursement. • Organise regular site visit checks to confirm the reported invoices and labour time	• Supervision agents • Finance team and payment unit
Violate the contract conditions by non-delivery of products, supplying services at lower quality and defective work.	Same as above	• Long delays in contract implementation and non-delivered projects • Failed inspection results and complaints about poor quality from users	• Demand independent annual financial technical and procurement audits with a focus on fraud detection for projects assigned higher risks • Facilitate site visits by technical experts during the supervision and monitoring activities	• Internal and external auditors • Inspection officers
Renegotiate the contract, terms of references and deliverables to deviate from the initial requirements of the awarding criteria and thus rendering public procurement decision invalid.	Same as above	• Several questionable change orders from a specific contractor which are approved by the same project staff • Significant changes to the outputs and deliverables • Substantial change to the TOR and increase in contract value	• Assess change order requests and ask for supporting documents. • Review all change orders and identify contractors that are prone to requesting frequent changes.	Project officials

Evaluation and Audit Phase

Corruption or fraud risk schemes	Potential Actors	Possible risk indicators	Example of mitigating controls	Risk owners
Forge documentation and falsify information to have positive evaluations by auditors.	Public officials, contractors and suppliers	• Extent of missing documents • Inconsistent supporting documents (invoices and purchase orders do not match) • Number of complaints	• Ensure adequate fraud audit capacity and experience to provide reliable audit • Strengthen databases and data analysis which can support audit programmes	• Internal auditors of the implementing ministries • External auditors of an independent reviewing body
Bribe and or influence evaluators to overlook violations of controls and suspected fraud and corruption in project closure.	Public officials, contractors, suppliers, auditors and evaluators	• Reduced audit scope and shortened examination period • Irregular audit procedures • Abnormal pattern of expenses • Undeclared conflicts of interest	Establish checks and balances at every step of control and evaluation (e.g. external audit review the work of internal audit and another objective function oversee the external audit team	• Senior managers of the responsible ministries • Senior managers of an independent reviewing body

2. Data Analytics for Assessing Corruption and Fraud Risks

2.1. Introduction

Effective management of corruption and fraud risks relies on the ability of government entities to extract meaning from data through analysis, tools and techniques. This pursuit, commonly known as "data analytics," has broader applications than fraud and corruption risk management. It has the potential to transform how government entities provide services, evaluate performance and conduct oversight. For instance, government entities can use data analytics to target service delivery and monitor the performance of programmes. Finance departments, regulatory agencies, and anti-corruption bodies, use analytics to assess reams of data to spot risks and identify unusual or suspicious transactions for further investigation and sanctions. Specialised groups have even coined new terms to describe specific users, such as “audit analytics” to refer to audit entities’ use of analytics to monitor financial transactions, or test the effectiveness of internal controls and compliance procedures.

While data analytics has diverse applications, there are common principles and practices across the data value chain that are relevant to many contexts. For example, the effectiveness of data analytics in any context relies on effective institutional governance and data governance, data integrity and project-level planning. This section addresses some of these common elements, and Section 3 builds on the frameworks below in the context of safeguarding integrity in infrastructure projects. In particular, given the high-risks in and data generated during the procurement process, this section and Section 3 highlights considerations and uses of analytics to support corruption risk assessments for public procurement.

In this targeted application, the use of analytics for corruption and fraud risk assessments can help facilitate decisions about strategy, resource allocation and control activities. Effective use of data analytics in the context of risk assessments has one main objective—to refine and improve understanding about risks in order to inform mitigation actions. Moreover, quantitative approaches can complement, and not necessarily replace, qualitative methodologies for risk assessments that rely more on employee perceptions. Institutions can apply multiple techniques for a hybrid approach to analytics, depending on objectives. When done well, they have the potential to uncover additional insights to help managers to understand the risk universe and take preventive actions in response.

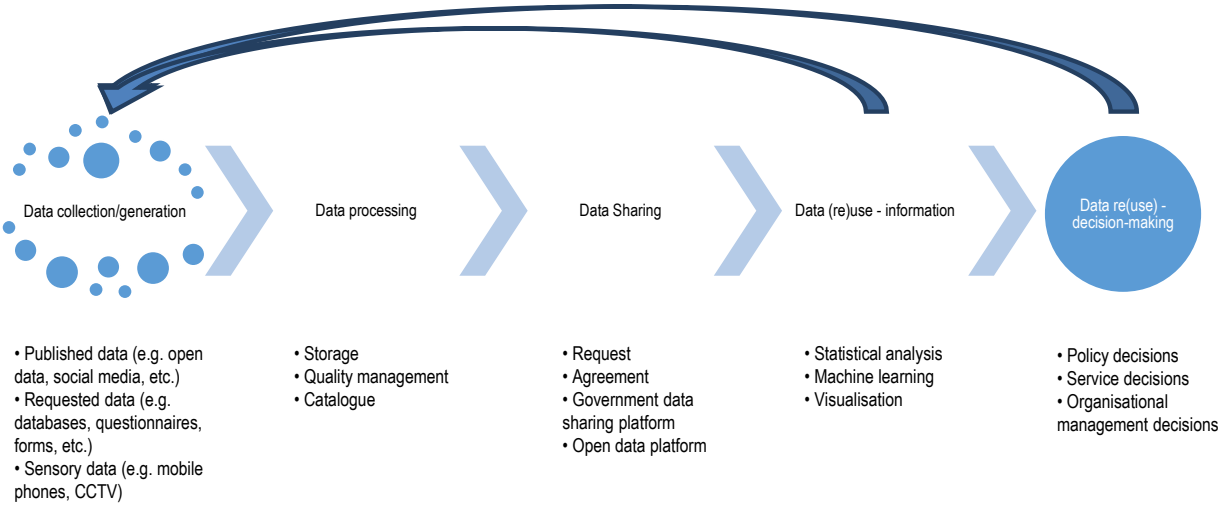
2.2. Extracting value from data to answer key questions

2.2.1. An overview of the data value chain

Effectively integrating data analytics into risk assessments requires a basic understanding of the data value chain, as the public sector has become increasingly reliant on ways to transform data into information and knowledge. With this evolution, international standards have also evolved to reflect the reality of a data-driven, risk-based public sector.

For instance, various OECD instruments and international standards emphasise the essential value of investing time and resources in developing effective data policies, data governance models, skills and capacity.[1] The data value chain is circular and iterative, and whether at an institutional or project level, contains feedback loops that can lead to improvements across the value chain (depicted as arrows in Figure 2.1). For instance, a risk owner who is responsible for assessing corruption risks among suppliers can identify blindspots and gaps in knowledge about specific risk areas during the course of the assessment. This in turn can lead to improvements in data generation, such as refinements to information collected during selection and tendering phases.

Figure 2.1. The Data Value Chain



Source: (OECD, 2018a).

The data value chain above offers examples to illustrate each stage. In the context of infrastructure and procurement, the first stage—data collection and generation—can take various forms. Project and risk managers could map the available databases for conducting corruption risk assessments based on the specific country context. For instance, national procurement databases, supplier databases "owned" by the contractors, databases of debarred or sanctioned companies and internal databases, such as databases for employees on asset disclosures and conflicts of interest, could all be useful inputs for the corruption risk assessment process. Moreover, for purposes of this paper, the ultimate use of data analytics in the context of risk assessments is to inform a specific aspect of managerial decision-making, namely, decisions about taking preventive actions to respond to risks and adapt control activities. Section 1 discusses in more detail this as the concept of risk treatment, which can be seen as one outcome of the data value chain when taking quantitative approaches to risk assessments.

2.2.2. Objectives and questions addressed by data analytics

Data analytics can cut across strategic objectives, competencies and levels within an organisation. Therefore, objectives related to the use of data analytics for corruption and fraud risk assessments may be linked to other objectives, such as those that focus on effectiveness and efficiency. For instance, procurement officials collect data to aid in tendering and assessing contract performance, yet project managers, risk managers and

auditors may rely on that same data for assessing risks and identifying potential vulnerabilities in the control environment. Table 2.1 summarises the range of questions that data analytics can help to address, building on the purpose of the data value chain to convert data into information and/or knowledge.

Table 2.1. Key questions that data analytics can address

	Hindsight	Insight	Foresight
Information	What happened? (Reporting)	What is happening now? (Alerts)	What will happen? (Extrapolation)
Knowledge	How and why did it happen? (modelling, experimental design)	What's the next best action? (Recommendation)	What's the best/worst that can happen? (Prediction, optimisation, simulation)

Source: Adapted from (Davenport, Harris, & Morison, 2010).

While Table 2.1 provides a conceptual framing of the use of data analytics, in the context of assessing corruption and fraud risks, the questions are more specific and risk-based. The key questions in this context help to extract information from specific areas of operations to make decisions and take actions to reduce the vulnerability of an organisation to fraud and corruption. Previous or ongoing risk assessments can inform these questions so that risks drive the research objectives of the data analytics plan, as opposed to the data or technological tool driving the questions (Cotton, Sandra, & Leslye, 2016). By linking the questions that data analytics will answer to risk assessments, it also helps to ensure that the analytics process itself is serving broader strategic objectives.

Data-driven risk assessments primarily help managers to understand what happened, what is happening and why it happened, in relation to the strategic objectives and the control environment. The analytics answer questions that are descriptive or diagnostic. For instance, which suppliers pose the highest risk based on historical data, or which phases of the infrastructure project cycle appear to pose the highest risks, and why? The example from Mexico illustrated in Section 3 primarily serves this purpose. Complementing this are predictive and prescriptive analytics, which use data to predict outcomes or trends and then link predictions to actions (Wells, 2015). For example, what trends or patterns can be identified in the data to anticipate emerging corruption and fraud risks?

Regardless of the questions asked, analytics on its own does not provide certainty that fraud and corruption occurred. Quantitative approaches to conducting risk assessments provides additional information for project and risk managers to assess where risks are, and may further illustrate the probability and magnitude of the risks. However, additional follow-up, investigations and a court ruling are needed to determine whether actual fraud or corruption occurred. In this application of data analytics for risk assessments, the goal is primarily preventive with the possibility of detecting high-risk cases for further referral and investigation.

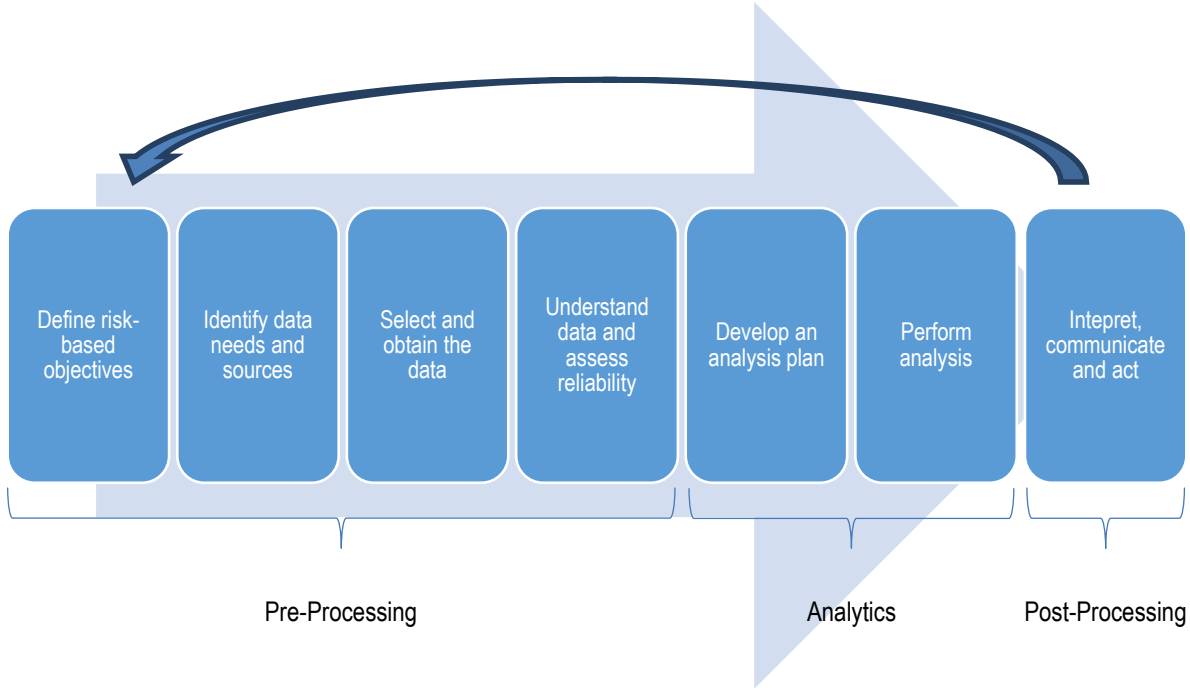
2.3. Creating a data analytics plan and analytic techniques

2.3.1. Steps for using data analytics to effectively assess risks

The decision to take a data-driven approach to risk assessments and the use of data analytics is a decision to invest in data governance, data integrity and the various quality controls

that accompany such efforts. In optimal conditions, data are interoperable, accessible, discoverable and open to allow for continuous production, collection, sharing and re-use. The steps in Figure 2.2 do not require this ideal scenario, yet they do presume that an institution has the basic data architecture and infrastructure in place, as well as the skills, to implement. Some institutions may require systemic improvements to how they collect data, as described in Section 3, or they may need to engage external specialists for effective implementation. Nonetheless, the narrow context of applying data analytics to corruption risk assessments offers a focused, targeted application for getting data in order that can serve broader strategic objectives as well.

Figure 2.2. Steps for carrying out a data-driven risk assessment



Source: OECD; (Baesens, Van Vlasselaer, & Verbeke, 2015).

The steps in Figure 2.2 are presented as a simplified process for illustrative purposes only, but the steps are not necessarily sequential. For instance, after obtaining data, a team may determine that the data are not sufficiently reliable for the intended purpose, and it may need to seek an alternative source of data or even redefine the objectives. Moreover, as the figure shows, pre-processing as a phase entails four steps, which can be even more time-consuming than the actual "analytics" procedure itself.

Institutional factors, such as those described in the section below, play a critical role in the effectiveness and efficiency with which a team advances through the steps in Figure 2.2. For example, the amount of time required for each step can vary depending on the circumstances and the level of maturity with regards to the organisation's data governance. An institution with a centralized data warehouse may spend very little time obtaining data needed for the analysis. On the other hand, an institution that needs access to data held by an external entity—such as another government entity—may need to spend considerable time and resources in establishing processes and procedures to obtain the data.

The following provides an overview of the individual steps and additional considerations. As noted, these steps are not always linear, and they include various activities, such as testing the reliability and validity of data, which could occur numerous times throughout a risk assessment process. Nonetheless, they are presented sequentially to highlight key considerations and issues:

1. **Define risk-based objectives** –Shaping objectives of the data analysis based on risk identification can help the team to understand and target the areas in which fraud, corruption, or other corruption and fraud risks are most likely to occur. In this way, the objectives of the corruption risk assessment and the data analytics plan are aligned, as the latter is a tool for informing a robust risk assessment. Perception-based risk assessments can provide input into what specific questions a team will ask of the data and the indicators it will develop. Moreover, interviews with experts, workshops, focus groups, audit reports, media coverage and the results of previous data analyses may all inform the data analytics objectives. As discussed, objectives tend to be descriptive, diagnostic, predictive or prescriptive.

   Well-defined objectives and risk identification are critical for the steps that follow, including identifying the right data sets and who holds these data (data custodians), acquiring, collecting and/or requesting these data, cleaning the data, and conducting appropriate analytical tests and involving the most relevant stakeholders in follow up. After defining the objectives, the analytics team could identify the fraud or corruption indicators or "red flags" they plan to identify with the data analytics test. It is important to obtain an understanding of programme rules and processes, as well as what is considered 'normal' behaviour at this stage, before implementing analytics tests. As reviewing results can be time-consuming, taking time at the beginning of the analytics process to obtain a strong understanding of the process can help the analytics team develop more refined analytics tests that may produce fewer false positives.

2. **Identify data needs and sources** – The next step is to identify the data that will be needed to identify the fraud or corruption risks defined in the first step, as well as to identify the sources of that data. This may include data that exist within the entity, data from other government agencies, or data from external, non-government entities. The specific data needed to conduct the analysis will depend on the analytics objectives and the specific indicators the analytics tests will be used to identify. In the case of the OECD's work with the Airport Group of Mexico City (*Grupo Aeroportuario de la Ciudad de México*, GACM), this activity posed considerable challenges, because data collection and management were disaggregated among actors. For instance, each of the work supervisors responsible for contract management (*Residente de Obras*) managed and updated databases individually. The decentralisation of data collection and management posed challenges for GACM to use the data for assessing and managing risks.

3. **Select and obtain the data** – The next step is to collect the data necessary to conduct the analysis. If data are obtained from external entities, some understanding of the data is necessary in order to develop a formal data request. Common elements included in a data request are format, sample fields, the intended use of the data, control totals to verify the completeness of the data and any limitations of the data (Wells, 2015).

   As noted, the process of using data analytics for assessing corruption risks often runs in parallel, and can be a catalyst, for broader improvements to data governance.

This is particularly important when several institutions are involved in data collection, generation or sharing. Once data is collected, it is critical to obtain or develop a data dictionary. Data dictionaries explain each field within the data and, as such, are a main source of information for data analysts (INTOSAI Working Group on IT Audit, 2016). They help to ensure that individuals use common definitions (U.S. Government Accountability Office, 2013), understand when different terms are used for the same thing, or when the same term has different meanings across government entities or programmes (Henderson & Hammersburg, 2013). Other data governance issues also play a part at this stage, such as data semantics, inter-operability and metadata.

4. **Understand data and assess reliability** – The next step is to assess the reliability and validity of the data and to take steps as necessary to clean and format the data, if needed, to ensure that it can be used in the analysis. Data analysts can use one or more of the following data-validation tests to verify the reliability and completeness of the data provided:

    - Verify the data types against the record layout and data dictionary (for example, text fields contain text);
    - Confirm the record count with the control totals received;
    - Confirm the hash totals of numeric fields with the control totals received;
    - Identify missing data (for example, blank fields or gaps in sequences);
    - Check for duplicate data and confirm whether any duplicates identified are false positives;
    - Reconcile the data to accounting records;
    - Perform reasonability tests (for example, calculate the number of transactions per month and determine if the number is near the number that would reasonably expected in a month);
    - Perform period testing to determine if the data cover the requested period (INTOSAI Working Group on IT Audit, 2016).

    Any discrepancies identified should be addressed before performing the analysis, which may include re-requesting the data (INTOSAI Working Group on IT Audit, 2016). Some suggest that data cleaning should only be done when actually performing the analysis as cleaning might delete "interesting outliers" (Kimball, 2014). Care should be taken at this step to understand and assess any discrepancies or outliers identified as a result of data validation tests or data cleaning procedures as outliers or anomalies may be indicative of fraud or corruption.

5. **Develop an analysis plan, including specific analytics tests** – The next step is to develop an analysis plan that describes the data to be analysed, the specific analytics approach that the team will perform and the frequency of the approach. A data analytics work plan can span a few weeks for the analysis, or it can be part of a broader risk assessment. For the latter, this can take months, as it may involve multiple data sources and parallel risk management activities. When using data analytics for integrity purposes, particularly for fraud detection and testing the effectiveness of internal controls, government entities could:

- *Analyse all relevant data*: Government entities can apply analytics tests to the full data population. Random sampling is useful for identifying problems that occur relatively consistently throughout data populations. However, as fraudulent or corrupt transactions do not occur randomly, sampling may not be sufficient to identify fraud (ACL, 2013) (Institute of Internal Auditors, Global Technology Audit Guide 13: Fraud Prevention and Detection in an Automated World).
- *Design data analytics tests based on the identified fraud indicators*: As discussed, the analytics team could translate the specific indicators of fraud or corruption identified at the beginning of the analysis into specific analytical procedures.
- *Determine whether the analysis will be conducted on an ad hoc, repetitive, or continuous basis*: Data analytics tests can be applied ad hoc or can be applied on a repetitive or continuous basis, and the frequency with which to run data analytics tests depends on the purpose for which analytics are being used. For example, data analytics tests can be applied on an ad-hoc basis to identify potential issues that may indicate opportunities exist for fraud to occur (ACL, 2013). This approach may be sufficient for a project manager using data analytics to identify risks and analyse the effectiveness of control activities relative to specific operational areas or contract types. However, programme managers using data analytics to maintain programme integrity could automate data-analytic tests to monitor fraud indictors on a continuous, real-time basis, if possible (U.S. Government Accountability Office , 2015). If data analytics tests cannot be automated to occur on a continuous basis, such as when data can only be obtained on a periodic basis, performing data analytics tests on a regular, periodic basis, can still be informative. For example, implementing data analytics tests during monthly transaction cycles can help ensure that risks are being mitigated throughout the year, rather than on an annual basis (Mazur, 2015).

6. **Perform the analysis** – At this stage, the analytics team implements the analysis plan to perform the analysis. During this stage, there are various software programmes that can facilitate the analyses. Microsoft Access and Excel is a common tool of auditors and accountants, and can be suitable when analyses are not too large or complex, a determination which is context-specific (Gee, 2015). For instance, in some countries, analysing "big data," such as matching health care records for identifying fraudulent providers, can involve analyses of hundreds of millions of rows of data. However, the maximum amount of rows in Excel is limited to approximately one million rows. As an alternative, there are other software programmes (e.g. ACL, IDEA, SAS and open-source tools like R and Python, etc.) that can handle larger datasets and complex procedures. See next section for further discussion on data analytic techniques.
7. **Interpret, communicate and act –** Interpreting the results of analytics in the context of risk assessments entails an iterative process of assessing the output relative to the initial objectives. To what extent are the data answering the stated research questions? Can the tests be refined further in order to increase the clarity of the results and determine corrective actions, if any? Is there a logical explanation for the results or signs of potentially fraudulent activity? Data analytics tests do not confirm fraud in the procurement cycle; however, they signal specific cases that look suspicious and could require addition review and investigation. Thus,

involving individuals with sound judgement, experience, expertise and scepticism are all critical for the evaluation of results.

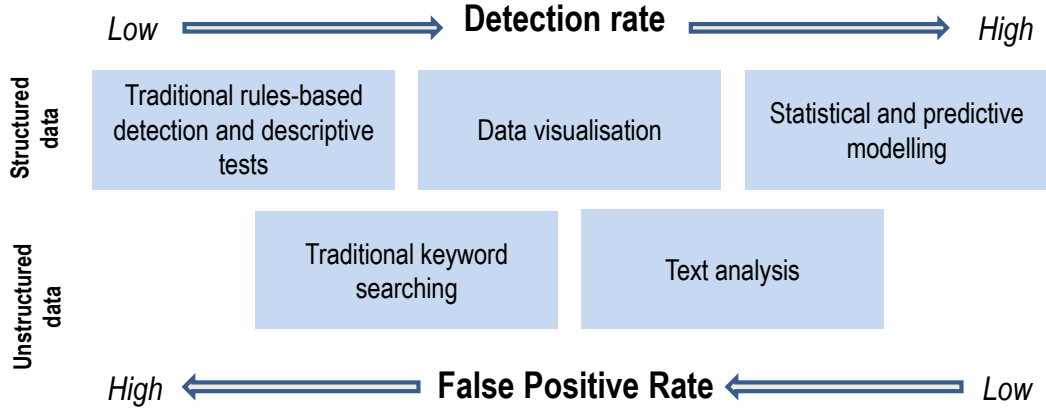
As demonstrated in Section 3, data visualization tools, such as dashboards and maps, or even simple charts and figures, can illustrate areas of greater risks more clearly than spreadsheets, statistics, and lists of transactions. They also provide an effective communication tool in order to communicate results to relevant parties, including the project managers, internal audit function and investigative bodies who may follow up on instances of potential fraud and corruption.

Implementing data analytics effectively to identify integrity issues is an iterative process. Data analytic processes must be updated as circumstances change or as more information is gathered. To enhance the analytics plan, it is useful to have a feedback loop in which the results of analytics are incorporated into the design of future analytics tests. Depending on the sophistication of the data analytics system used, the feedback loop can be incorporated manually or automatically. To incorporate feedback loops manually, antifraud or anticorruption experts analyse results that match predetermined fraudulent patterns. Through experience, analysts refine the patterns that are used to identify new cases in the future. In more advanced analytics systems, analysts confirm whether a specific identified case was in fact fraudulent and the system uses that information to automatically refine the model it uses. Such machine learning can provide the means to continuously improve analytics, make them more efficient, and reduce false positives (KPMG, 2016). As discussed later in this paper, institutions should align their investments in technology and skills with their needs.

2.3.2. Snapshot of analytic techniques

The selection of one or more analytic techniques depends on the objectives of the analysis, skills, resources, availability of data and data quality. Objectives that attempt to generalise findings to a population, or predict patterns of fraud or corruption schemes, require more advanced analytics and statistical knowledge, but can result in higher detection rates and allow for analyses of structured data. Figure 2.3 provides an overview of considerations when selecting a data analytics technique, based on three elements—detection rates, complexity and value.

Figure 2.3. Selecting data analytics techniques based on detection rate, complexity and value



Source: Adapted from (EY, 2016).

In general, descriptive objectives, such as assessing a population for potential fraud and corruption and identifying red flags for further investigation, can be achieved through rules-based tests, using techniques such as data matching and data mining. This approach involves setting pre-defined rules to filter or mine data to identify aberrant behaviour. Moreover, information on previously encountered fraud schemes and programme rules can be used to determine red flags that inform the queries to be applied to datasets. For instance, a rules-based analytics could help an entity to analyse procurement data to identify bidders who received sole-source contracts for military contracts within specified advertisement periods of the tender. As another example, if procurement rules prohibit individuals from making purchases above a set threshold amount, queries can be developed to test transaction data to identify multiple purchases from the same cardholder to the same vendor in the same day (U.S. Government Accountability Office , 2015). Such rules-based testing, also referred to as breakpoint clustering, can also be used to assess invoice payments or purchases for risks of suppliers or employees circumventing spending limits by splitting transactions. These types of analyses are simpler than other techniques, but may result in higher false positives and lower detection rates.

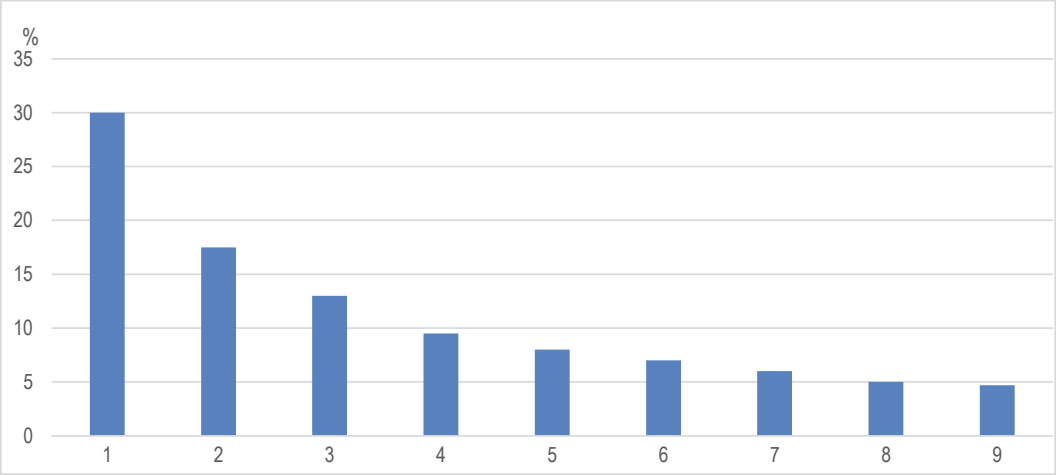
Unlike rules-based detection, which looks for specific known corruption or fraud schemes, anomaly or outlier detection targets behaviours that are unusual or expected, which can be useful for identifying potential corruption or fraud when specific patterns are unknown (Henderson & Hammersburg, 2013). Values that are higher or lower than expected could be indicators of potential fraud (ACL, 2013). Anomaly or outlier detection can be applied in different ways. For instance, cluster analysis—in which data are sorted into groups based on a similar characteristic, such as location—can identify anomalies or outliers relative to what is expected based on that group (ACL, 2013). Historical or trend analysis—in which transactions or information from one individual or entity are compared over time, such as a contractor or supplier—can identify anomalies in reported information that can indicate potential corruption, fraud or other issues.

Benford's Law is one well-known form of anomaly detection. Accountants, auditors and investigators use Benford's Law to detect numbers near their authorisation limits, which can signal potential fraud or corruption (see Box 2.1 for details). In Brazil, the government's internal procurement policy allows for simplified bidding procedures if a contract value is lower than BRL 80 000 (Brazilian reals), and direct purchases are permitted if the value is lower than BRL 8 000. Using an app developed with Caseware IDEA software, Benford's Law analysis undertaken on a sample of procurement contracts suggested that collusion may have occurred between companies and government agencies to reduce the initial contract value in order to conform to the simplified purchasing procedures. While such approaches are used for detection, they can also be incorporated into risk assessments to gauge the effectiveness of specific control activities, such as thresholds for simplified bidding procedures, as noted in the Brazil example.

Box 2.1. Benford's law for detecting fraud and corruption

Benford's law can be used as a screening tool for fraud detection when applied to data sets. The law describes the frequency distribution of the first digit in data sets, and compares the expected and observed distributions. As the number 1 appears most frequently as the first digit in data progressions, and successive numbers less frequently, strong deviations from the expected frequencies or anomalies may indicate that the data is suspicious, or that it has been manipulated (see Figure 2.4).

Figure 2.4. Benford's Law distribution



Source: CaseWare Analytics (2016), 'Using Benford's Law for Fraud Detection and Auditing', https://www.slideshare.net/CaseWare_Analytics/using-benfords-law-for-fraud-detection-and-auditing-67432835?qid=2e202a6d-4db0-448d-8917-670cec858609&v=&b=&from_search=4

Benford's Law is commonly applied to detect numbers near their authorisation limits. If an authorisation limit is EUR10 000 (euros), then frequent first two digits in the 99, 98 and 97 area will be detected if there is an attempt to maximise authorising expenditures. Other practical applications include accounts payable data, sales and purchases. Benford's Law may also help uncover anomalies or fraudulent activity in government procurement activities.

Sources: (Baesens, Van Vlasselaer, & Verbeke, 2015) (Caseware Analytics, 2016) (Gee, 2015).

As noted, risk assessments can help to answer predictive questions to provide insights to managers on actions to take concerning the control environment and risk mitigation. Predictive analytics involves building models that identify attributes or patterns that are highly correlated with known instances of fraud and then applying those models to incoming transactions to determine if such transactions resemble known cases of fraud (Henderson & Hammersburg, 2013). For instance, predictive modelling, or predictive analytics, is useful for identifying complex patterns in data (Kaplan, 2011), and could be used to identify potentially fraudulent transactions or claims before they are paid (U.S. Government Accountability Office, 2013). Predictive models can also help managers to score transactions based on the probability that they represent corrupt or fraudulent behaviour in order to prioritize contracts or transactions for further review.

Unlike the techniques described above, which rely on structured data, other approaches can support assessing risks in unstructured data. For example, text mining can be used to identify patterns in unstructured data, such as reports, emails and social media (Henderson & Hammersburg, 2013). The Inland Revenue Authority of Singapore is an example of a government entity using text mining to collect, analyse and structure text from emails to derive insights about issues pertinent to taxpayers (OECD, 2016a). Similar processes can be used in the context of assessing corruption risks in infrastructure. For example, a line

ministry could assess internal risks of fraud or corruption by scraping emails or social media to identify red flags, like key words or evidence of procurement officials spending beyond their means. To maximize the value of text analytics, entities may use the fraud triangle as a reference to develop a list of keywords based on the industry, relevant fraud risks, and data set (Association of Certified Fraud Examiners, 2016b).

Social network analysis can also be applied to unstructured data in the context of infrastructure and public procurement. Strategic networks may influence the awarding of a contract and can foster collusion amongst actors in the procurement cycle (Mamavi, 2017). A recent study undertaken on procurement practices in Hungary and the Czech Republic using network analysis highlights how some suppliers are excluded, whilst certain groups enjoy preferential treatment from the state (Fazekas, Wachs, & Skuhrovec, 2017). Applying network analysis in this context can raise red flags and identify corruption risks. Moreover, data visualizations can be used to present the results of network analysis, as well as other data-analytic techniques like the Mexican example in Section 3, to identify "hot spots" of potential fraudulent activity.

2.4. Considering institutional factors and limitations of using data analytics

Data analytics as a process and tool does not occur in a vacuum. Government-wide laws, policies, guidance and other factors can influence the use of data analytics on an institutional and project level. An in-depth discussion of these factors is beyond the scope of this paper, but deserve a brief reference. For instance, a country's laws and policies form the foundation for data governance, data management, collaboration and sharing of data between government entities and sectors for improved corruption and fraud risk management. Such factors can influence effectiveness and efficiency within the data value chain, and at the project level, can either hinder or enhance the opportunities to use data analytics to inform corruption and fraud risk assessments.

Moreover, public internal control and risk management standards commonly call on government entities to assess risks of fraud and corruption, and highlight the potential use of data to support risk assessments. For instance, the International Standards Organisation notes that risk assessments can rely on historical data, experience, stakeholder feedback, observation, forecasts and expert judgement (ISO, 2009). These standards and the accompanying guidance are typically articulated in general terms so that government entities have the flexibility to tailor their approaches to their own contexts. For example, as noted in Section 1, government institutions may embed corruption and fraud risk assessments into broader risk management activities, or conduct them as stand-alone assessments.

Various factors, as listed below, influence the use of data analytics and the readiness for an institution to adopt data-driven approaches to fraud and corruption risk assessments. Depending on the context, the factors can be an institution's strength or can pose limitations that require further planning and investment. Institutions that are lacking in one or more of the areas below are not necessarily ill equipped to explore the use of data for assessing risks in infrastructure projects. However, it is important for those who make use of data analytics to maintain realistic expectations about what is achievable and at what cost, as discussed in the following subsection. Indeed, for some institutions, perception-based risk assessments and using basic statistical approaches or Excel spreadsheet may suffice for their stated objectives.

1. **Institutional and Data Governance** – the vision, strategy and policies that demonstrate leadership's commitment to data analytics and communicates roles and responsibilities to staff. Data governance accounts for the standards, and controls that apply across the entity, which can help ensure quality, consistency, security and maintenance of data, as well as monitoring and evaluation of analytics. In addition, data sources can be both internal and external to an organisation. Effective governance facilitates access to and sharing of data, and establishing cooperative relationships with other entities, including those that could be both data producers and consumers (e.g. tax authorities, procurement entities, or social service ministries) or have investigative authority to follow up on potential fraud or corruption (e.g. law enforcement and anti-corruption bodies).
2. **Culture** – understanding and commitment of leadership and staff for establishing and sustaining effective programmes for data analytics. Data sharing and adoption of new processes and tools rely on a culture that understands the benefits of data analytics, but maintains realistic expectations and professional scepticism when using it to assess fraud or corruption risk. This understanding grounded Moreover, a strong culture for data analytics encourages collaboration within and outside of the entity, as well as innovation and creativity to test hypotheses, develop red flags for corruption and fraud risks and anticipate evolving fraud and corruption schemes. A commitment to data analytics can begin with the leadership, which is responsible for encouraging buy-in of employees throughout the organisation. Alternatively, the commitment can be driven from the bottom-up by dedicated individuals or teams that are innovating and experimenting with data analytics on their own.
3. **People** –technical skills and knowledge to employ relevant methodologies and software are critical, including experience in programing. Moreover, data analytics for assessing corruption and fraud risks can benefit from individuals with a high-level of awareness and sector-specific knowledge. For example, corruption and fraud risks in procurement related to infrastructure can be different from those in the health sector, where schemes can take a different form and level of complexity. Moreover, legal expertise is critical to effectively sustain programmes for data analytics, particularly when using external data sources that may have legal requirements for access, privacy, storage and security.
4. **Processes** – Data analytics, when applied to specific objectives such as assessing corruption in procurement, is a set of processes to inform decision-making. It involves tailored policies, planning and actions. Processes for data analytics not only apply to analyses of data, but also the identification of sources, collection of data and assessment of the reliability and validity of the data for achieving objectives. Continuous monitoring and evaluation of the performance of programmes for data analytics, based on objectives and metrics, are also critical for understanding the effectiveness of strategies and adapting them, as needed.
5. **Technology** – Technological infrastructure, tools and software underpin many of the processes for effective data analytics. Above all, it is critical for institutions to take a strategic approach to investing in the technological infrastructure that facilitates data analytics, ensuring that investments align with objectives. For instance, to effectively conduct data matching to identify improper payments, public entities could consider investing in infrastructure that allows for receiving, storing and securing large amounts of sensitive data or different data types.

Moreover, tools and approaches can vary in sophistication, resource requirements and usability, from spreadsheets to advanced analytical software. Technology evolves quickly, and overextending resources without a clear vision or purpose can result in wasted taxpayer money.

2.5. Assessing the value of analytics

Return on investment (ROI) generally refers to the ratio of a benefit to the investment of resources that generated the return, which can include both economic (e.g. taxpayer funds recovered) or non-economic (e.g. achievement of programme goals and social outcomes). As such, the ROI can be both quantitative and qualitative. In the context of data analytics for managing corruption risks, measuring ROI poses many challenges, since it can be difficult for entities to determine the full extent of fraud or corruption prevented or detected by data analytics. However, taking on this measurement challenge can be critical for understanding how to improve data analytics and substantiate resource investments. Various surveys and studies have devised indicators that intended to provide a broader picture of the use and benefit of data analytics.

As shown in Table 2.2, the Association of Certified Fraud Examiners (ACFE) examined median fraud losses—the amount of revenues the organization loses in a given year as a result of fraud—and the time to detection for fraud, considering 18 anti-fraud controls in place when the fraud occurred. For purposes of the ACFE report, *Report to the Nations on Occupational Fraud and Abuse: 2016 Global Fraud Study,* data analytics is referred to as a control activity. The ACFE found that having proactive data monitoring and analyses in place resulted in the biggest reduction of fraud losses relative to 18 other control activities, including codes of conduct, reward for whistle-blowers and external audit of financial statements, among others (Association of Certified Fraud Examiners, 2016a).[2]

Table 2.2. Reduction in fraud losses as a result of select control activities

Control	Percent of Cases	Fraud Losses with Control in Place	Fraud Losses with Control Not in Place	Percent Reduction
Proactive Data Monitoring/Analysis	36.7%	USD 92 000	USD 200 000	54.0%
Management Review	64.7%	USD 100 000	USD 200 000	50.0%
Hotline	60.1%	USD 100 000	USD 200 000	50.0%
Management Certification of Financial Statements	71.9%	USD 104 000	USD 205 000	49.3%
Surprise Audits	37.8%	USD 100 000	USD 195 000	48.7%

Note: The table shows only the top 5 of the 18 controls listed in the original figure.
Source: Adapted from (Association of Certified Fraud Examiners, 2016a).

Examples of public entities that consistently measure the ROI of data analytics, and make the information publically available, are limited. For instance, the Center for Medicare and Medicaid Services (CMS) in the United States recovered USD 23.5 million in fiscal year 2015 from its prescription drug programmes as a result of data analytics (U.S. Department of Health & Human Services, Centers for Medicare and Medicaid Services, 2015a). CMS incorporates this initiative into its overall fraud prevention system, which helped to identify nearly USD 655 million in improper payments in calendar year 2015, with an estimated return on investment of USD 11.5 to USD 1 (U.S. Department of Health and Human Services, Centers for Medicare and Medicaid Services, 2015b).

Understanding the costs associated with establishing and maintaining a data analytics function is essential for effective measurement of ROI. The total costs of an analytics programme can be considerable, and in some institutions, the costs may ultimately exceed the costs of investing in advanced quantitative techniques to refine probabilistic conclusions that ultimately still have a margin of error. Investments include a range of investments, from the acquisition of software and hardware to facilitate analytics, to operational costs such as trainings for staff as well as IT support. Some institutions may benefit from economies of scale, as they seek to develop the data architecture and infrastructure for using data for various purposes, including assessing corruption and fraud risks. In addition, for some entities, incorporating the cost of investigations may also be a useful activity to consider, especially when the entity itself is responsible for investigating. In this context, Figure 2.5 illustrates a general formula for calculating ROI.

Figure 2.5. Measuring return on investment for data analytics

$$\text{ROI} = \frac{\text{Total utility value}}{\text{Total cost of analytics} + \text{Total cost of investigating fraud or corruption}}$$

Source: Adapted from (Baesens, Van Vlasselaer, & Verbeke, 2015).

For many line ministries and procurement entities, which often make referrals to law enforcement bodies, the cost of investigating fraud or corruption is not within their mandate, and is therefore irrelevant. How might they still measure total utility value? One approach is to focus on specific benefits. For instance, data analytics can result in a list of suspicious cases of potential fraud or corruption that require further investigation. This sample could be based an analysis of the impact (e.g. highest potential loss amount), likelihood (e.g. most suspicious and likely to have occurred) or both. The number of cases that result in actual fraud or corruption convictions is also a quantifiable metric to measure the return, or utility value, of data analytics. This "hit rate"—the percentage of actual fraud or corruption among all the potential cases in the sample—can be an informative indicator for calculating ROI. A hit rate of 100 percent could indicate optimal use of resources, since no time or effort would have been spent on investigating benign cases or false positives. A high hit rate may also suggest vulnerabilities and the need for investing in additional resources (Baesens, Van Vlasselaer, & Verbeke, 2015).

In addition, corruption and fraud in infrastructure projects carries opportunity costs when it leads to delays, lower quality infrastructure or failure to deliver on the social and economic benefits of the project. The foregone benefits due to corruption or fraud can factor into the total utility value of using data analytics to support risk assessments, and take actions to ensure preventive measures effectively address inherent risks.

Measuring ROI in such a way relies on expertise, planning and robust reporting mechanisms. The total utility value can factor in specific benefits (e.g. the hit rate) or other benefits that go beyond the resources spent (and saved) on investigations and inspections. Benefits that can be monetised, such as financial benefits stemming from increased revenues, recovered assets, collection of penalties and so forth, can be calculated, but may not provide a full picture of non-economic benefits. For instance, qualitative benefits generate large positive externalities to society that are not easily translated into budgetary savings, but can be as significant as financial benefits.

2.6. Establishing quick wins and realistic expectations

The OECD supports governments to strengthen risk management and risk-based approaches to safeguarding integrity, including leveraging data in risk assessments. Section 3 illustrates one such example in the case of Mexico and the construction of a new international airport. This case, like others, was a first-time effort. For many government entities working to improve the use of data analytics, whether in the context of infrastructure or for other applications, there are several considerations and lessons learned from the work of OECD and others. Considerations for pilot projects with data analytics include the following:

- *Focus on organizational needs or problems to be addressed* – Pilot projects that focus on organizational needs or problems that have not been easily addressed through existing processes or procedures are most likely to demonstrate the value of data analytics and promote organizational buy-in. For example, when identifying pilot projects to demonstrate the value of big-data analytics, entities should identify problems with characteristics that are suitable to big-data solutions—specifically, problems involving large data volumes and significant data variety that can potentially be addressed by being broken down into smaller units of work that can be executed simultaneously (Desouza, 2014).
- *Be realistic* – Ideal pilot projects are ones for which the results are reasonably attainable. As data analytics become increasingly complex, the risks associated with the project increase. Therefore, when implementing projects that rely on more advanced tools and techniques, such as big-data projects, it is important to set realistic expectations in order to avoid "over-hyping" the promise and benefits of big-data projects while underplaying the risks and challenges—which will likely lead the project to fail (El-Darwich & al, 2014).
- *Require minimal investment.* Data analytics capabilities should be built strategically through an iterative approach. Pilot projects that can leverage existing data and technology can help demonstrate the value of data analytics and avoid the risk of overinvesting. For this reason, embedding analytics in the risk assessment process can help to tie the effort to existing, ongoing initiatives. In addition, ideal pilot projects take advantage of data that are easily accessible, such as stored government data (Mazur, 2015)or data that are publicly available (Desouza, 2014). To the extent possible, government entities should focus initially on projects that can be accomplished using existing technology or with minimal modifications or updates to existing technology. If successful, pilot projects can garner greater support for subsequent efforts and demonstrate the effectiveness of data analytics when asking for funding and other resources (Desouza, 2014) (U.S. Government Accountability Office, 2013). Further, if initial projects achieve cost savings, savings can be reinvested in more advanced data programmes (GovLoop, 2015).

When implementing pilot projects, it is also essential to communicate progress, as well as validated results, to stakeholders. Finally, when pilot projects have been complete, evaluating the results against performance measures can help determine if the project was effective, as well as help identify "lessons learned" that can be applied to future projects.

Notes

[1] For example, the 2017 *OECD Recommendation of the Council on Public Integrity*, the 2015 G20 Open Data Principles by the Anti-Corruption Working Group, the 2014 OECD Recommendation of the Council on Public Procurement, and the 2014 OECD Recommendation of the Council on Digital Government Strategies recognise the foundations for effective decision-making rest, in part, in the use of data and taking assesing risks.

[2]Based on a survey of 2 410 respondents, approximately 13% (313 respondents) of which worked for government or law enforcement,

3. Data-Driven Risk Assessments in Practice: Applying a Corruption Risk Index to a Mexican Infrastructure Project

3.1. Introduction

Corruption and fraud risks can arise at any phase of the infrastructure development process starting from needs definition through contract implementation, as illustrated in previous chapters. While in an ideal scenario risks across the infrastructure project cycle are comprehensively assessed, in practice data-driven risk methodologies can be cost-effectively deployed only where abundant structured administrative data already exists. Hence, this section turns the spotlight on public procurement, in particular the planning and document design, as well as tendering phases of the infrastructure delivery process. The focus of this section also coincides with extensive academic and policy attention recognising the central role government contracting processes play in high quality infrastructure provision.

Building on recent innovations in quantitative corruption risk assessments and the rich, publicly available public procurement datasets in Mexico, objective proxies of corruption and an interactive dashboard were developed to support the risk management activities of the Airport Group of Mexico City (*Grupo Aeroportuario de la Ciudad de México*, GACM), who were responsible for the project of the New International Airport of Mexico (*Nuevo Aeropuerto Internacional de México*, NAIM).. The compiled large-scale public procurement dataset enabled the calculation of nine validated corruption risk red flags which could be combined into a composite Corruption Risk Index (CRI). The following sections describe the process of developing the CRI, first highlighting the global policy and research literature informing the corruption risk scoring in Mexico, then explaining the CRI's methodological framework, followed by a detailed explanation of its application to the case of Mexico and the GACM, where public procurement data were used to build the CRI and to visualize it in dashboards.

3.2. Recent advances in data-driven corruption risk assessment in public procurement

Corruption is ostensibly difficult to measure, mainly due to the difficulty of accessing the necessary data. For a long time, most corruption indicators derived either from surveys of attitudes, perceptions and experiences of corruption among different stakeholders; or reviews of institutional features supposed to control corruption; or audits and investigations of individual cases. While each of these have their merits they typically lack the precision and scale unique to Big Data analysis which are necessary for a systemic corruption risk assessment framework, which can simultaneously support broad-based policy decisions as well as investigations targeting individual transactions. In order to show the global evidence base for corruption risk scoring in public procurement and the inspire further applications in diverse institutional contexts, the discussion below briefly outlines recent

advances in harnessing Big Data methods in government contracting to develop valid and reliable corruption proxies.

In the last decade or so, a range of scholars have developed objective corruption proxies which rely on directly observable behaviours that likely indicate corruption (for a detailed overview see Fazekas, Tóth, & King, 2016). Much of the policy-relevant research use large-scale government contracting datasets. For example, Golden and Picci (2005) propose a new measure of corruption based on the difference between the quantity (stock) of infrastructure and the related public spending (flow) among twenty regions in Italy. Others use 'red flag' indicators in public procurement micro-data as proxy measures for corruption. Among others, high quality examples include single bidding (Klasnja, 2016), the use of exceptional procedure types (Auriol et al., 2011), clear scoring rules (Hyytinen et al., 2008), or political connections of winning companies (Goldman et al., 2013).

Building on such a broad base of validated elementary risk indicators allows for building composite scores that addresses the challenges of corruption being carried out in a diverse ways in public procurement. For example, corrupt actors may target different phases of the process such as the advertisement of tenders or contract implementation and they can use different techniques such as tailoring the tendering terms or unfairly scoring bidders to achieve higher than market prices, lower quality or lower than promised quantity. Combining indicators from various procurement phases and capturing different, often substitute techniques is a prerequisite for robust risk assessment.

In additional to quantitative measurement exercises, a wealth of qualitative studies has documented the nature and logic of diverse corrupt practices in public procurement. These studies cover many countries both from OECD and non-OECD groups taking more journalistic, government-centred, or legalistic approaches (OECD, 2007; World Bank, 2009; Transparency International, 2006; Dávid-Barrett et al, 2018). Detailed qualitative accounts of corruption strategies and techniques in public procurement provide robust basis for identifying the most widespread corruption situations in large-scale datasets using analytical techniques in the quantitative literature quoted above.

Such a comprehensive risk scoring is enabled by the increasing availability of contract or item-level datasets for whole countries in a machine readable format. A number of parallel developments unlocked such data: i) the fast spreading use of comprehensive e-procurement systems for advertisement but increasingly transaction management; and ii) a global movement driven by civil society and international organisations pushing for open publication of administrative data on government tenders. Among others, the rapid spread of the Open Contracting Data Standard (OCDS, http://standard.open-contracting.org/latest/en/) has made risk scoring based on publicly available data a reality; while similarly research projects have unlocked a range contracting datasets such as the DIGIWHIST (http://digiwhist.eu/) project, which has republished over 17 million government contracts for 32 European countries and the EU institutions (for the results see: https://opentender.eu).

Another example is provided by the project Curbing corruption in development aid-funded procurement (http://www.govtransparency.eu/index.php/2018/02/13/aiddata/) using datasets from the World Bank, the Inter-American Development Bank, EuropeAid, and Tanzanian national procurement data in order to analyse corruption risks in the aid sector. Many of these projects, including DIGIWHIST, have also developed tailored corruption risk metrics for a wide range of countries demonstrating the feasibility and utility of large-scale risk assessment methodologies. Of course, taking advantage of Big Data in this context requires investments in infrastructure, skills and knowledge.

Building on Section 2, Box 3.1 summarises several of the key elements ton consider when investing an effective data-driven risk assessment.

Box 3.1. Resources and skills needed for an effective data-driven risk assessment

Setting up and maintaining a quantitative risk assessment framework which is effectively used in risk management and policy making requires a modest investment and a few specific, quantitative skills:

1. **Public procurement and linked datasets**: Typically, the biggest cost of measuring corruption risks is due to the creation, extraction, and organisation of the relevant administrative datasets. However, these costs very much depend on the quality and openness of government data systems. On one end of the spectrum, some countries like Mexico already have readily downloadable, structured public procurement as well as company registry datasets considerably lowering data costs. On the other end of the spectrum there are countries which only have paper-based public tendering data requiring the investment of typically over USD 50-100 000 in manual data collection and digitisation of records. In between these two extreme cases lies the majority of OECD countries with electronic data available in diverse, semi- structured formats accessible but requiring some investment into data extraction, organisation, and cleaning.
2. **Technical infrastructure**: given the storage and scale of most public procurement datasets, the servers of the government data warehouse will be needed at least for data extraction. In addition, for the largest datasets of several million records, even basic data cleaning and analytical work might require the use of high capacity servers. In addition, data cleaning and analysis are best done using some of the widely used statistical and data analytical software packages such as Phyton, R, SPSS, or Stata.
3. **Data analytical and visualisation skills**: Creating, validity testing, and analysing corruption risk indicators requires both an in-depth understanding of public procurement markets and advanced data analytic skills. Public procurement-specific knowledge is needed both to understand data scope and variable definitions as well as the essence of the regulatory framework setting out the conditions of procuring and bidding such as regulatory thresholds or time limits. Data analytic skills typically include the capacity to manipulate large-scale datasets (i.e. 100 thousands or millions of observations) and to implement advanced statistical methods such as binary logistic regressions, matching, 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)
4. **Knowledgeable users**: Understanding the risk measurement framework, its strengths and weaknesses presents its own challenges even to experienced risk managers. Hence, key users such as auditors should be trained and the appropriate organisational responses to various types and levels of risks worked out. In addition, creating a regular feedback loop where users can report on their experiences with the measurement framework can provide a crucial input into updating the framework.

Source: Author.

3.3. Developing a corruption risk indicator in public procurement

3.3.1. Corruption definition and measurement logic

The term corruption is used to cover diverse phenomena in many contexts that differ in the prevailing norms of good conduct. Hence, many characterisations of corruption are normatively charged and context-dependent (Johnston, 1996). Probably the most common definition of corruption - "the misuse of public office for private gain" - (Rose-Ackerman, 1978) understands corruption within a bureaucratic context and associates corruption with bribery of public officials. The problem with this definition, on the one hand, is that Weberian bureaucracy and the underlying rational-legal order may not be present in many contexts at all. On the other hand, it is also inadequate to capture corruption in public positions with high degrees of discretion such as members of parliament or public procurement decision makers (Warren, 2003).

Departing from such definitions, the corruption concept developed tightly matched to the area of public procurement and to the institutionalised and recurrent forms of corruption that can be more readily measured. Hence, in public procurement, corruption refers to the allocation and performance of public contracts by bending rules of open and fair access to government contracts in order to benefit a closed network while denying access to all others. In other words, *the aim of such corruption is to steer the contract to the favoured bidder without detection in an institutionalised and recurrent fashion* (World Bank, 2009), by avoiding or biasing competition (e.g. unjustified sole sourcing or direct contract awards) in order to favour a certain, connected bidder (e.g. tailoring specifications to a particular company).

This definition focuses attention on restricted and unfair access to public resources while also allow for a clear-cut focus of the measurement framework (Mungiu-Pippidi, 2006; North, Wallis, & Weingast, 2009). Such corruption may involve bribery and transfers of large cash amounts as kickbacks, but it is more typically conducted through broker firms, subcontracts, offshore companies, and bogus consultancy contracts. By implication, not everything designated as corruption under this definition represents illegal activity as defined by the law in a given country (Fazekas, Tóth & King, 2016; Fazekas & Kocsis, 2017).

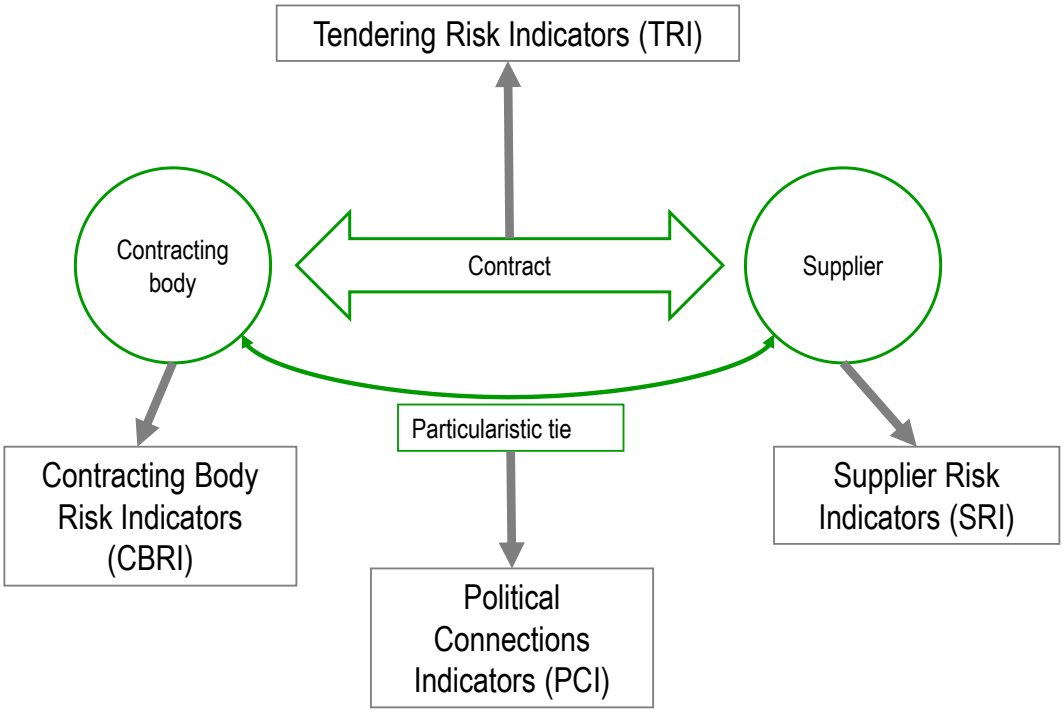
This definition implies that for measuring corruption its underlying logic must be contrasted with a competitive market logic. Institutionalised grand corruption's primary aim is earning corruption rents, which can be earned in public procurement when the winning contractor is a pre-selected company that then earns extra profit due to charging higher than market price for the delivered quantity and/or quality. In order to measure extra profit, the price, delivered quantity and quality of deliveries must be known with high precision, yet none of these three can adequately be measured in most public procurement administrative datasets. Price and quantity of procured deliveries are usually publicly available but not comparable across time and space, while quality cannot be reliably observed in official records. Therefore, for GACM,an alternative was created to proxy corruption by analysing the process of awarding contracts and key outputs such as number of bidders and market concentration. Crucially, lack of bidders for government contracts (single bidder) is an outcome whereas the means to introduce certain procedural rules for limiting competition (manipulating procedure types and shortening advertising period) are red flags/inputs. The relationship/correlation between inputs and outcomes form the measurement model and can serve as a test for validity when selecting proxy indicators to constructing the CRI.

3.3.2. Indicator types

Any corrupt contract allocation requires at least four components to be in place: a) corrupt transactions allowing for rent generation (contract), b) corrupt relations underpinning collective action of corrupt groups (particularistic tie); c) organisations enabling rent allocation (contracting body); and d) organisations extracting corrupt rents (supplier). These four requirements of corrupt contracting serve as a framework for risk assessment, leading to a wide array of indicators discussed in the below sections (Figure 3.1). Consequently, there are four types: Tendering Risk Indicators (TRI), Contracting Body Risk Indicators (CBRI), Supplier Risk Indicators (SRI), and Political Connections Indicators (PCI). In each of these groups, there is a wide array of elementary corruption risk indicators which derive from proven cases while also being validity tested on large-scale datasets (for a full discussion see Fazekas, Cingolani & Tóth 2016). Figure 3.1 provides an overview of these four types of indicators.

Given the clandestine and often complex character of corrupt deals, a comprehensive measurement approach of building a composite CRI is advocated where each indicator sheds light on different aspects of the same corrupt phenomena. Nevertheless, these indicators individually or combined as a composite score only indicate the *risk* of corruption, meaning that they are proxy indicators indirectly pointing at underlying corrupt exchanges. While indicators might be very different, they are all required to match the corruption definition used for this report, derive from objective data, allow for consistent temporal and cross-organisational comparisons on large samples, and be validated using alternative corruption proxies.

Figure 3.1. Components of the corrupt exchange and corresponding indicator groups



Note: Green denotes components of the corrupt scheme; grey marks the indicator groups.
Source: *Fazekas, Cingolani & Tóth, 2016.*

Tendering Risk Indicators (TRI) capture all those micro-level aspects of public procurement tenders and contract implementation which signal corrupt manipulation of the procurement process in order to generate rents and allocate them to the connected companies. A particularly widely quoted example is the tailoring of tender conditions to fit a single company on an otherwise competitive market.

A number of high-quality research papers have looked into tendering corruption risks in various contexts such as elections and high-level politics or welfare services and redistributive politics. For example, Olken (2007) uses independent engineers to review road projects and calculates the amount and value of missing inputs to indicate corruption during contract implementation. Another approach to assess the amount of missing procurement outputs in infrastructure is proposed by Golden & Picci (2005) who look into the difference between the stock of infrastructure and cumulative public spending on it using two independent data sources. Other authors use indicators characterising the bidding process on the micro-level, such as the use of exceptional procedure types (Auriol et al., 2011) or negotiated procedures (Chong, Klien & Saussier, 2015), explicit scoring rules (Hyytinen, Lundberg, and Toivanen 2008) or single bidding on competitive markets (Klasnja 2016; Fazekas & Kocsis, 2017). The latter is the indicator that matches the corruption definition directly, while the previous ones are indirect indicators for corruption.

These studies have shown that circumventing competition in the tendering process can be done in three principal ways, each corresponding to a phase of the public procurement process: (a) limiting the set of bidders in the advertisement phase; (b) unfairly assessing bidders in the assessment phase; and (c) ex-post modifying conditions of performance in the contract implementation phase, which, while not being a company selection technique, can support the selection of the pre-selected company which might promise low prices and high quality knowing that later contract modifications will allow it to earn the agreed corruption rent. These three elementary corruption strategies can be combined in any way to reach the final desired outcome, e.g. some bidders may be excluded from submission with tightly tailored eligibility criteria while the remaining unwanted bidders can be unfairly scored in the assessment.

Political Connections Indicators (PCI) provide cues on the particularistic ties (e.g. through kinship, friendship, professional) between bidder owners/managers and political office holders who are able to influence the public procurement process. Such ties are indispensable for monitoring and enforcing corrupt deals which tend to be informal (e.g. using family as device for building and maintaining trust in the absence of courts enforcing contracts). Political connections are of diverse nature and no particular direction of influence is assumed. The use of these different strategies of personal connections and control very much depend on the threat of exposing corrupt dealings and the specificities of the country's legal framework (Trapnell, 2011). Some of these types of personal connections are difficult to measure than others as well as possibly being established as institutionalised forms of connections such as political party finances (Fazekas and Cingolani 2016; OECD 2014) or lobbying (David-Barrett, 2011).

Prior empirical literature looked at personal political connections or political influence established through political party donations and the short and long term direct benefits to the connected companies (Goldman, Rocholl & So, 2013; Luechinger & Moser, 2014; Fazekas, Ferrali & Wachs, 2018) while others considered ties either to specific individuals or parties as a whole (Akey 2013; Straub 2014). In Denmark, which is one of the least corrupt countries of the world, direct family ties between companies and politicians

surprisingly increase company profitability, especially in sectors dependent public procurement, for example (Amore & Bennedsen, 2013).

Supplier Risk Indicators (SRI) signal the use of winner companies as vehicles of rent extraction and the distribution and hiding of assets which are indispensable for rewarding all the participants of the corrupt deal and avoiding detection. As corrupt rent extraction in public procurement differs from competitive tendering, it is assumed that corrupt companies are different from their peers in a number of fundamental characteristics. Identifying corrupt companies based on publicly available data is an inherently challenging exercise, thus companies are evaluated on multiple dimensions: company registry attributes, company financial information, company ownership and management data, and company governance information. For example, the success of companies of a certain age (e.g. very young companies) have been observed to have suspiciously high profitability and high rates of single bidding in Hungary (Fazekas & Tóth, 2017; Dávid-Barrett & Fazekas, 2016).

Contracting Body Risk Indicators (CBRI) capture the risk of corrupt allocation of public funds by contracting bodies and weaknesses of formal bureaucratic structures designed to shield contracting bodies from pressures to favour connected bidders. These indicators jointly capture the complete process of generating, allocating and distributing corrupt rents from government contracts and generally match the organisation level where each public agency corresponds to one contracting body.

It is assumed that certain organisational features of the contracting body are key to the possibilities of public funds misallocation. While the literature is much less advanced in this field, there are various indicators that aim at capturing relevant agency-level characteristics, such as transparency index scores (Williams, 2015), or political appointments and contract approval rights (Dahlström, Fazekas, & Lewis, 2018). Other suggested indicators for this group include auditing information, prosecutions, budget transparency and controls, or asset declarations (Fazekas, Cingolani & Tóth, 2016). Arguably, some of these indicators are less directly related to corruption and sometimes rely on perceptions data, which distinguishes this indicators group from the other three groups.

3.3.3. Indicator selection and composite score building

The preceding sections discussed the four major corruption risk indicator groups and examples of individual indicators within them. Many of the indicators suffer from overestimating corruption risks, as there are numerous alternative, non-corrupt circumstances where the indicators signal risk (i.e. false positives). For example, while there are certainly cases where extremely high turnover growth from public procurement is due to government favouritism, it is also likely that innovative companies entering the market would produce similar patterns in the data. Such false positives can be eliminated by carefully selecting the elementary risk indicators that are most closely associated with other corruption signals to triangulate risk indicators against each other, keeping the indicators that fit the corrupt rent extraction model. False positives can further be eliminated by pulling indicators from different indicator groups into a composite score, which becomes more robust to unobserved variation in specific corruption techniques and measurement error.

Validity testing of each elementary risk indicator, such as very tight advertisement period, can only be done by checking their fit with a corrupt contracting logic against other non-corrupt logics such as low administrative quality. Unfortunately, no random sample of

proven and clean cases is available which would allow for an alternative validity testing. The most straightforward way of validity testing elementary indicators is to test their fit with the corruption definition, for example, by verifying that the suspiciously short advertisement period predict single bidding on competitive markets, that is short advertisement is typically used to limit competition (Fazekas, Tóth & King, 2016).

We develop a composite score of tendering 'red flags', called Corruption Risk Index (CRI), as an objective proxy measure of high-level corruption in public procurement that operationalises the previously described definition of corruption, derives from objective public procurement data, allows for consistent comparisons across time and organisations, and can be further validated using alternative corruption proxies (for a detailed explanation of CRI building using data from 28 European countries for 2009-2014, see e.g. Fazekas & Kocsis, 2017). For simplicity of interpretation the CRI is composed as a simple arithmetic average of individual risk indicators, falling between 0 and 1, with 1 representing the highest observed corruption risk and 0 the lowest.

3.3.4. Strengths and weaknesses of the measurement approach

The measurement model approximates the corruption definition according to which corruption works when legally prescribed principles of open and fair competition are circumvented by public officials when designing and running tenders in order to recurrently award government contracts to connected companies. Proxy indicators signal corruption only if competition is expected in the absence of corruption, thus markets which are non-competitive under non-corrupt circumstances have to be excluded (e.g. markets for specialised services). In addition, these indicators signal *risk* of corruption, rather than actual corruption and they are expected to be correlated with corrupt exchanges rather than perfectly matching them.

The strength of the composite indicator approach is a more complete monitoring of the corrupt contracting process, while it also explicitly tries to abstract from diverse market realities to capture underlying corruption techniques. It allows for 'red flag' definitions to change from context to context in order to capture similar levels of risk irrespective of the detailed forms of corruption techniques used (e.g. non-corrupt competitive conditions imply tighter submission deadlines in the Netherlands than in Greece, hence corrupt behaviour would reflect deviations from slightly different normal benchmarks). This flexibility in corruption indices aims to assure that the same level of risk is associated with a similar level of actual corruption in a comparative perspective. As corruption techniques are likely to change over time, tracking multiple corruption strategies in one composite score is most likely to remain consistent. Both of these characteristics underpin its usefulness for international and time-series comparative research.

The main weakness of CRI is that it can only capture a subset of corruption strategies, arguably the simplest ones; hence it misses out on sophisticated types of corruption such as corruption combined with inter-bidder collusion (for a more comprehensive review of corruption risks, please see Section 1). As long as simplest strategies are the cheapest, they likely represent the most widespread forms of corrupt behaviour. However, it is admitted that more sophisticated corruption techniques are more likely to be used when monitoring institutions are stronger, implying that the level of corruption may be under-estimated in less corrupt countries. Further research should expand on the set of red flags tracked and evaluate the interaction between monitoring institutions, regulatory complexity, and corruption sophistication in order to more precisely estimate corruption.

3.4. Quantitatively assessing corruption risks in a Mexican infrastructure project

3.4.1. Defining objectives

The following sections illustrates a data-driven corruption risk measurement, tailored to Mexican public procurement data in the context of GACM and the development of the NAIM. The process below generally maps the steps for creating a data analytics plan and carrying out a data-driven risk assessment. As noted in Section 2.3, the first step for creating a data analytics plan is to define risk-based objectives. For the assessment described below, the objective was to bolster the existing risk assessment process of GACM with a greater focus on identifying corruption risks in the procurement cycle. This objective was deliberately narrow and excluded consideration of other types of procurement risks. This allowed for a targeted analysis, which in turn would have allowed for concrete interventions to manage corruption risks, had GACM continued operating. As noted in Section 2, the process for taking a data-driven approach to conducting risk assessments may require ongoing attempts at preparing data, testing validity, calculating indicators and conducting analysis, in an iterative and non-sequential manner. Moreover, while the framework directly derives from the widely tested methodology discussed above, it also acknowledges risk assessment frameworks already created or under development in Mexico, in particular the framework of IMCO (ABT-OPI Analytics, 2018).

3.4.2. Identifying, selecting and assessing data

Mexico's public procurement regulatory and data system is relatively well developed with readily downloadable datasets that include a number of variables relevant for corruption risk assessment. Since 2017, Mexico's public procurement agency also publishes its data in the Open Contracting Data Standard (OCDS), which standardises procurement data internationally. Data from before 2017 are published in the national structure of the procurement platform CompraNet, which covers a slightly different set of variables.

The final dataset used for corruption risk assessment is comprised of data in the national as well as the OCDS structure. The national publications were used as the basis for the final dataset and information from OCDS were added where it represented added value. The final dataset contained 1 512 288 observations, with each observation being a uniquely identifiable public procurement contract. On this basis, the contracts related to the construction of the Mexico City Airport and GACM could be identified (2 112 contracts in total).

The final dataset comprised 46 variables providing information on contract details such as contract ID and title, buyer and supplier names, contract value amount, tender and contract start and end dates, procurement category, procedure type, and number of bidders. The timeframe of the source dataset ranges from 2005 to 2018, however data from the years before 2012 were excluded from analysis due to small numbers of observations for those years and less reliable records, reducing the number of observations to 1 318 491. The rate of missing data differs across variables and years, probably due to the increasing usage of the procurement recording system as well as the introduction of the OCDS system in 2016. Nevertheless, missing rates for most of the essential variables remain low with missing rates of 0.01% for basic variables, such as tender ID, buyer name and type, supplier name and ID, procedure number, bidder name, contract value, and contract and procedure types.

For other variables related to the award and implementation stages, the missing rates were higher or some variables were missing entirely (which rendered the assessment of some corruption risk indicators impossible). Some crucial information that were missing entirely

or had high missing rates include company registration date in the national company register to determine company age, award value and ID, whether a contract that got cancelled got relaunched subsequently, tender plans linked to contracts, and implementation data such as the number of contract amendments per contract and changes in contract length.

3.4.3. Identifying and validating corruption risk indicators

Given this data frame, 17 indicators related to supplier risks, contracting body risks, and tendering risks were considered potentially calculable from the available data. Regarding the fourth indicator group of political connection risks, the dataset did not include any information that would render these indicators potentially feasible such as data on companies' and procuring bodies' top officials. Due to a high rate of missing values or the results of validity tests, eventually only 9 indicators were selected to form the CRI in Mexico (see Table 3.1). Moreover, one risk indicator: tax haven registration of the supplier, took the value of zero among GACM contracts, so this is not discussed in detail.

Table 3.1. Overview of valid elementary corruption risk indicators in the CRI composite score

Indicator group	Procurement phase	Indicator name	Indicator definition
tendering risk	Advertisement	Procedure type	0='open' procedure 1='non-open' procedure type (e.g. direct contracting)
tendering risk	Advertisement	Lack of call for tenders publication	0='call' for tenders advertised 1='call' for tenders not advertised
tendering risk	Advertisement	Length of advertisement period	0='advertisement' period>='18' days 1='advertisement' period<18 days
tendering risk	Advertisement	Single bidder contract	0='more' than one bid received 1='one' bid received
tendering risk	Assessment	Length of decision period	0= 7<='decision' period<='49' days 1= decision period<7 days OR decision period>49 d.
tendering risk	Assessment	Contract modification during advertisement	0 = contract NOT modified during advertisement 1 = contract modified during advertisement
tendering risk	Implementation	Contract modification during implementation	0 = contract NOT modified during delivery 1 = contract modified during delivery
tendering risk	Implementation	Cost overrun	0 = contract price increase <260% 1 = contract price increase >=260%

Source: Author.

Single bidder contract: Singe bidding is the simplest indication of restricted competition reflecting the corruption definition used for this assessment, when only one bid is submitted for a tender on a competitive market (for further discussion of single bidding see Charron, Dahlström, Fazekas, & Lapuente, 2017; Fazekas, Tóth, et al., 2016). This typically allows awarding contracts above market prices and extracting corrupt rents. Recurrent single bidder tenders between a buyer and a supplier allow for developing interpersonal trust underpinning corrupt contracting, thus individual instances of single bidding may be explained by a number of non-corrupt reasons (e.g. known most productive bidder), recurrent or extensively used single bidder contracts are more likely to signal corruption and restricted access. Nevertheless, the single bidder indicator is also more prone to gaming by corrupt, e.g. including fake bidders to mimic competition. For justified purchases of

highly specific products or when the most productive supplier is known, single bidding may over-estimate corruption risks, while unjustly defining such specific purchases is a major form of corrupt contracting. At the same time, single bidding cannot capture the corruption risks related to groups of apparently independent bidders forming cartels.

Procedure type (open vs. non-open): While open competition is relatively hard to avoid in some procedure types such as open tender, others types, such as invitation tenders, are by default less competitive. Therefore, 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 contracts to the same well-connected company.

Lack of call for tenders publication: A simple way to fix tenders is to avoid the publication of the call for tenders on the official public procurement platform, as this makes it harder for non-connected competitors to prepare bids, which is only relevant in non-open procedures where publication is voluntary. Not publishing the call for tenders makes it less likely that eligible bidders notice the bidding opportunity, weakening the competition and allowing the contracting bodies to more easily award contracts repeatedly to a well-connected company.

Length of submission/advertisement period: A short submission period (i.e. the number of days between publishing a tender and the submission deadline) leaves less time and thus makes it harder for non-connected companies to bid successfully, whereas a well-connected firm can use its inside knowledge to win repeatedly. as the buyer can informally inform the favoured bidder about the opportunity ahead of time. Considering the distribution of submission period values in the Mexican dataset, a period of 0-18 days is considered risky here.

Length of decision period: If the decision period on the submitted bids (i.e. the number of days between the submission deadline and announcing the contract award) is excessively short or lengthy, it can signal corruption risks. Snap decisions may reflect premediated assessment, while long decision periods signal extensive legal challenges to the tender, suggesting that the issuer attempted to limit competition. In the Mexican dataset, data on the award date were missing, therefore contract start dates were used alternatively. Considering the distribution of decision period values, very short periods of 1-6 days seem to be the most risky, while excessively long periods of 50 or more days are also risky. Decision periods around the average and a little longer of 7-49 days are considered the benchmark, no risk category.

Contract modification during advertisement: Modifying call for tenders during the advertisement period allows for excluding unwanted bidders by changing eligibility criteria once the interested bidders are known. This strategic modification of the call for tenders favours the well-connected company to further increase its market share.

Contract modification during implementation: If competition couldn't be eliminated during the bidding and assessment phases, the well-connected firm can still win with a competitive offer, but subsequent contract modifications during implementation still allow it to extract rents.

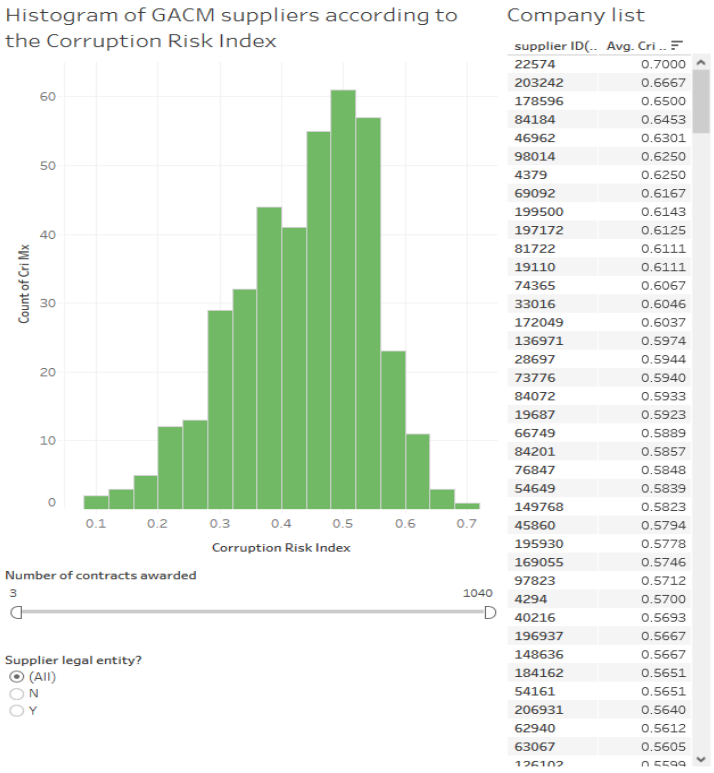
Cost overrun: The relative contract price increase during implementation indicates that the well-connected firm can still win with a competitive offer if competition couldn't be eliminated during the bidding and assessment phases, but subsequent contract value increase still allows it to collect extra profit. Excessive cost overruns where the contract value increase surpasses the initial contract value by 260% or more are considered high

risk. The threshold is so high as this variable suffers from a lot of missing and seemingly incorrect values, hence the only way to increase reliability of the eventual risk red flag was to choose a high threshold.

3.4.4. Performing analysis and assessing risks

Based on the calculated indicators, the CRI summing up the individual indicators could be aggregated on the company-level. These results were visualised in three dashboards accessible, illustrated on this Tableau page (https://tabsoft.co/2TAUNDj) specific to GACM suppliers. For purposes of this report, the suppliers have been anonymised. The dashboards focus on companies and show their CRI in GACM-related contracts as well as their scores for individual components of the CRI, i.e., the individual red flags such as single bidding. In addition, they allow for filtering those suppliers with certain numbers of contracts awarded as well as comparing the companies' CRIs in GACM-related and other federal (non-GACM) contracts. The dashboard includes a histogram showing the distribution of the CRI by GACM suppliers and list of companies with corresponding CRI (see Figure 3.2). Different CRI values can be selected in the histogram and the company list will show the companies with the selected CRI value accordingly. Also, the user can filter the number of contracts awarded per company using the slide underneath the histogram.

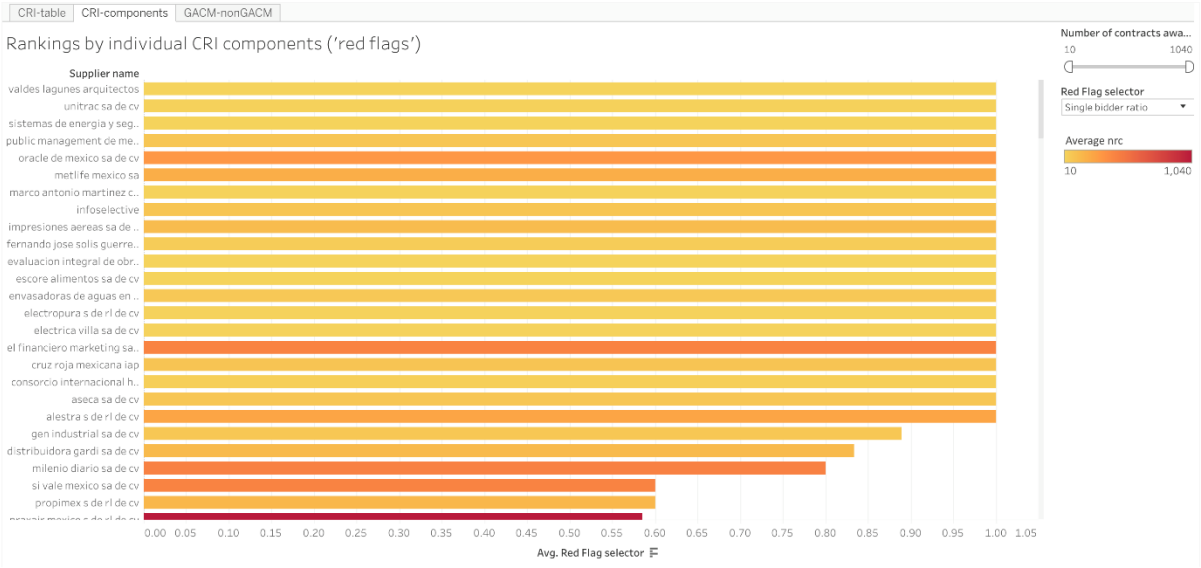
Figure 3.2. Screen shot of dashboard with suppliers and CRI values



Source: Author, https://tabsoft.co/2TAUNDj.

The second dashboard in Figure 3.3 below displays the ranking of companies by individual CRI components (i.e. indicators, "red flags"), where the user can select one of the eight red flags and see the companies ranked accordingly from highest to lowest values. At the same time, the colour of the bar indicates the number of contracts awarded to that company, and the user can use the slide to filter for the number of contracts. When hovering over the bars, the values for all "red flags" and number of contracts of that company are displayed.

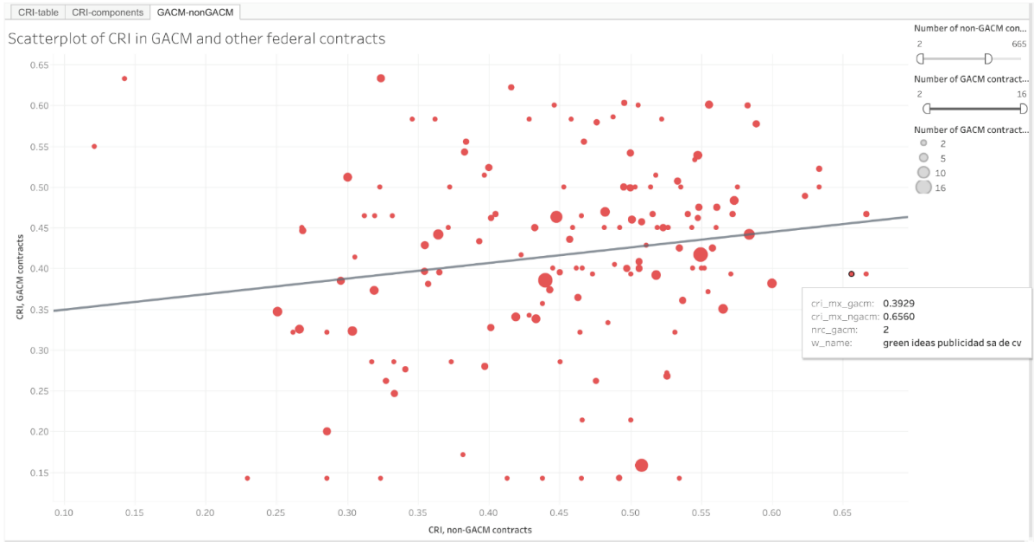
Figure 3.3. Screen shot of dashboard with company ranking by CRI component



Source: Author, https://tabsoft.co/2TAUNDj.

The third dashboard in Figure 3.4 below shows a scatterplot of companies by CRI in GACM and other federal (non-GACM) contracts with the dot size indicating the number of GACM contracts. The slides allow for filtering the number of non-BACM and GACM contracts. When selecting a dot, the name of that company is displayed as well as its CRI in GACM and other federal contracts and the number of contracts.

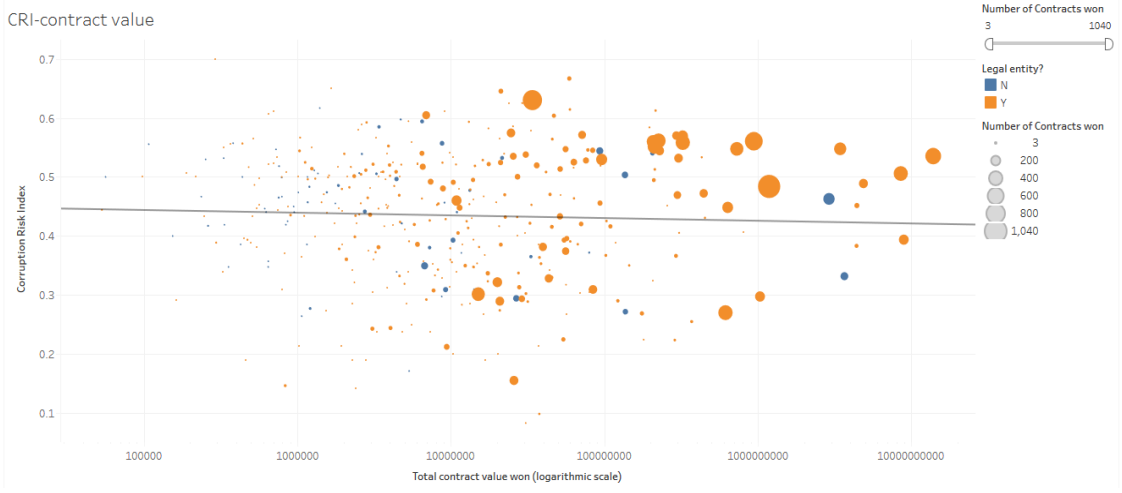
Figure 3.4. Screen shot of scatterplot of CRI in GACM and other federal contracts



Source: Author, https://tabsoft.co/2TAUNDj.

The fourth dashboard in Figure 3.5 below shows a scatterplot of companies' total contract value won and the CRI, differentiating legal entity and natural persons. The slide also allows for filtering by company size using the number of contract won. Additionally, company size is indicated by the size of the dot. The weak relationship between contract value won and corruption risks suggest that risks are not particularly pronounced for a subset of contract size categories, as one might have expected high value contracts and companies being of particularly high risk.

Figure 3.5. Screen shot of scatterplot with contract value and CRI



Source: Author, https://tabsoft.co/2TAUNDj.

3.4.5. Acting on insights from the data-driven risk assessments

Use of risk indicators for improving control activities

The above described data, indicators, and visualisations together can support ongoing risk management activities and inform decision making about strategy, control and mitigation measures. Moreover, they can direct audit and control activities to the highest risk transactions and organisations. While much of such activities are likely to operate ex post, that is intervening once a past transaction appears to be of high risk; it is also possible to use the data in a predictive fashion: identifying high-risk organisations and control their future transactions before spending is made. In addition, the data, indicators, and dashboards can also be used to formulate and target policy interventions addressing particular risks. For example, if the risk of collusion among bidders is high in a market due to the small number of bidding firms, policy interventions can aim at opening competition in that particular market. The very same data and indicators will then also be useful for tracking the effectiveness of the policy intervention in alleviating weak competition potentially leading to the reformulation of the policy. Finally, data-driven risk assessments can help managers to make critical, real-time decision about control activities to ensure that risks are mitigated before engaging or modifying contracts. They can complement existing methodologies, such as perception-based risk assessments, as another input for decision makers, risk managers and auditors. See Box 3.2 for an example of using quantitative risk assessments to identify high risks among counterparts of the European Investment Bank, leading to in-depth follow up on on-site audits.

Box 3.2. Big Data for Proactive Integrity Reviews: The case of the European Investment Bank

The European Investment Bank (EIB) finances projects, typically in the infrastructure sector, across the European Union of over EUR 50 billion annually (European Investment Bank, 2018). These projects are managed by thousands of procuring entities leading to tens of thousands of contracts. Managing the risks in such a large portfolio is a challenge relying on traditional methods such as whistleblowers reporting on wrongdoing. Recognising these challenges, the EIB screens and audits every year a handful of its counterparts (i.e. organisations receiving EIB loans) by conducting Proactive Integrity Reviews which aim to mitigate risks before large financial losses occur. Selecting entities for such reviews is based on a complex process combining quantitative as well as qualitative information. A key part of the quantitative risk-scoring component is based on tracking corruption proxies in all publicly procured contracts of EIB counterparts: over 200 000 awarded contracts for works and services. Red flags such as single bidding or the lack of advertising the call for tenders are used to create a composite risk score for each EIB counterpart. By looking at a number of further quantitative risk factors, a small sample of highest risk organisations and projects are selected for in-depth desk research including the review of media reports on the organisations in national and international press. In a final step, an even smaller sample is selected for on-site audits reviewing the organisational controls and the implementation of the project in greater depth (including engaging surveyors to confirm quality and quantity of works and services delivered).

Source: Author.

Steps needed to be taken for maintenance

The quantitative approach to risk assessments and the dashboards should be updated regularly in order to keep them relevant for ongoing risk management activities. As new data becomes available, institutions can follow the same approach and data structure to extend the time horizon of risk scoring and analysis. While short term extensions to databases are likely to leave the underlying measurement model valid, for longer intervals such as 2-3 years, re-checking indicator validity and adjusting parameters if needed will become important. Moreover, if data scope increases for example by adding new variables to the national data publication framework, it is possible to calculate and test the validity of new indicators, which can eventually help to expand the list of red flags.

References

ACL (2013), *Detecting and Preventing Fraud with Data Analytics*, https://www.acl.com/pdfs/ACL_fraud_ebook.pdf.

Akey, P. (2013). Valuing Changes in Political Networks: Evidence from Campaign Contributions to Close Congressional Elections. Available at: SSRN 2336131.

Amore, M.D., & Bennedsen, M. (2013). The Value of Local Political Connections in a Low-Corruption Environment. *Journal of Financial Economics 110*(2): 387–402.

Andersson, S., & Heywood, P. M. (2009). The politics of perception: use and abuse of transparency International's approach to measuring corruption. *Political Studies, 57:* 746–767.

Arndt, C., & Oman, C. (2006). *Uses and abuses of governance indicators*. Paris: OECD.

Association of Certified Fraud Examiners (2016a), *Report to the Nations on Occupational Fraud and Abuse: 2016 Global Fraud Study*, http://www.acfe.com/rttn2016/docs/2016-report-to-thenations.pdf.

Association of Certified Fraud Examiners (2016b), *Using Data Analytics to Detect Fraud: Other Data Analysis Techniques*, http://www.acfe.com/uploadedFiles/ACFE_Website/Content/review/da/05-Other-Data-Analysis-Techniques.pdf.

Auriol, E., Flochel, T., & Straub, S. (2011). *Public Procurement and Rent-Seeking: The Case of Paraguay* (No.11–224). TSE Working Papers: 11–224, Toulouse: Toulouse School of Economics (TSE).

Baesens, B., V. Van Vlasselaer and W. Verbeke (2015), *Using Descriptive, Predictive, and Social Network Technique: A guide to Data Science for Fraud Detection*, John Wiley & Sons, Inc.

Beckers, F. et al. (2013), *A risk management approach to a successful infrastructure project*, McKinsey & Company Publishers, http://www.sefifrance.fr/images/documents/mckinsey_a_risk_management_approach_to_a_successful_infrastructure_project.pdf (accessed on Dec 2018).

Caseware Analytics (2016), *Using Benford's Law for Fraud Detection and Auditing*, https://www.slideshare.net/CaseWare_Analytics/using-benfords-law-for-fraud-detection-and-auditing-67432835?qid=2e202a6d-4db0-448d-8917-670cec858609&v=&b=&from_search=4.

Charron, N., Dahlström, C., Fazekas, M., Lapuente, V. (2015). *Carriers, connections, and corruption risks in Europe.* QoG Working Paper Series: 2015:6, University of Gothenburg: Quality of Government Institute, Gothenburg.

Chong, E., Klien, E. & Saussier, S. (2015). *The Quality of Governance and the Use of Negotiated Procurement Procedures: Evidence from the European Union*. Paris.

Committee of Sponsoring Organizations of the Treadway Commission (COSO) (2016), *Fraud Risk Management Guide.*

Cotton, D., J. Sandra and G. Leslye (2016), *Fraud Risk Management Guide*, The Committee of the Sponsoring Organisations of the Treadway Commission (COSO).

Dahlström, C., Fazekas, M., & Lewis, D. E. (2018). Agency Design and Corruption Risks: Procurement in the United States Federal Government. In APSA 2018 Annual Meeting. Boston: APSA.

Davenport, T., J. Harris and R. Morison (2010), *Analytics at Work: Smarter Decisions, Better Results*, Harvard Business School Publishing Corporation.

Dávid-Barrett, E. & Fazekas, M. (2016). *Corrupt Contracting: Controlling Partisan Favouritism in Public Procurement.* GTI-WP/2016:02, Budapest: Government Transparency Institute.

Dávid-Barrett, E. & Fazekas, M., Smirnov, N. V.and Maslova, N. (2018). Study on the scale, forms and manifestations of corruption in the procurement processes in the Russian Federation at municipal level. Council of Europe, Strasbourg [in press].

David-Barrett, E. (2011). *Cabs for Hire? Fixing the Revolving Door Between Government and Business*. London.

Desouza, K. (2014), *Realizing the Promise of Big Data: Implementing Big Data Projects*, http://www.businessofgovernment.org/report/realizing-promise-big-data.

Dongping, Z. (2008), *A life-cycle risk management framework for PPP infrastructure projects*, Journal of Financial Management of Property and Construction.

El-Darwich, B. and E. al (2014), *The Global Information Technology Report - Big Data Maturity: An Action Plan for Policymakers and Executives*, http://www3.weforum.org/docs/GITR/2014/GITR_Chapter1.3_2014.pdf.

European Investment Bank (2018) Financial Report. 2017. European Investment Bank, Luxembourg.

EY (2016), *Trends in Data Analytics: Fraud Detection*, https://www.regonline.com/custImages/300000/300624/2016/11-3_Feinstein.pdf.

Fazekas, M. & Tóth, B. (2017), *Proxy indicators for the corrupt misuse of corporations.* October 2017:6. U4 - Chr. Michelsen Institute, Bergen, Norway.

Fazekas, M. & Tóth, B. (2017). *Infrastructure for whom? Corruption risks in infrastructure provision across Europe.* In Hammerschmid, G, Kostka, G. & Wegrich, K. (Eds.), The Governance Report 2016. Oxford University Press, ch. 11.

Fazekas, M., & Kocsis, G. (2017). Uncovering High-Level Corruption: Cross-National Corruption Proxies Using Government Contracting Data. *British Journal of Political Science.* Available at: https://www.cambridge.org/core/journals/british-journal-of-political-science/article/uncovering-highlevel-corruption-crossnational-objective-corruption-risk-indicators-using-public-procurement-data/8A1742693965AA92BE4D2BA53EADFDF0

Fazekas, M., Cingolani, L., & Tóth, B. (2016). *A comprehensive review of objective corruption proxies in public procurement: risky actors, transactions, and vehicles of rent extraction.* GTI-WP/2016:03. Government Transparency Institute. Budapest.

Fazekas, M., Tóth, I. J., & King, L. P. (2013). *Corruption manual for beginners: Inventory of elementary "corruption techniques" in public procurement using the case of Hungary*. Working Paper: GTI-WP/2013: 01. Government Transparency Institute, Budapest.

Fazekas, M., Tóth, I. J., & King, L. P. (2016). An Objective Corruption Risk Index Using Public Procurement Data. *European Journal of Criminal Policy and Research, 22*(3): 369–397.

Fazekas, M., Wachs and Skuhrovec (2017), *Corruption, government turnover, and public contracting market structure: Insights using network analysis and objective corruption proxies*, http://www.govtransparency.eu/wpcontent/uploads/2017/09/Fazekas_Wachs_Skuhrovec_CorruptionNetwork_Structure_in_CZ_HU_2017.pdf.

Gee, S. (2015), *Fraud and Detection: A Data Analytics Approach*, John Wiley & Sons, Inc.,.

Gold Coast waterways authority (2017), *Official Risk Appetite Statement*, https://gcwa.qld.gov.au/wp-content/uploads/2017/05/GCWA-Risk-Appetite-Statement.pdf (accessed on December 2018).

Golden, M. A., & Picci, L. (2005). Proposal for a new measure of corruption, illustrated with Italian data. *Economics & Politics*, *17*(1), 37–75. doi:10.1111/j.1468-0343.2005.00146.x.

Goldman, E., Rocholl, J., & So, J. (2013). Politically connected boards of directors and the allocation of procurement contracts. Review of Finance, 17(5), 1617–1648. doi:10.1093/rof/rfs039.

Gounev, P., & Bezlov, T. (2010). *Examining the links between organised crime and corruption*. Sofia: Center for the Study of Democracy.

GovLoop (2015), *The Big Data Playbook for Government*, https://www.govloop.com/resources/big-data-playbook-government/.

Henderson, G. and C. Hammersburg (2013), *An Enterprise Approach to Fraud Detection and Prevention in Government Programs*, http://www.rockinst.org/forumsandevents/audio/2013-11-06/fraud_wp.pdf.

Hyytinen, A., Lundberg, S. & Toivanen, O. (2008). *Politics and Procurement: Evidence from Cleaning Contract*s (No. 233). HECER Discussion paper No. 233.

Institute of Internal Auditors (Global Technology Audit Guide 13: Fraud Prevention and Detection in an Automated World), *2009*, https://chapters.theiia.org/montreal/ChapterDocuments/GTAG%2013%20-.

INTOSAI Working Group on IT Audit (2016), *Guide to Data Mining as a Tool in Fraud*, http://content.intosaicommunity.org/images/temp/WGITA-guide-Data-mining-as-atool-.

ISO (2009), *ISO 31000: 2009 Risk Management principles and guidelines*, https://www.iso.org/standard/43170.html.

Johnston, M. (1996). The Search for Definitions: The Vitality of Politics and the Issue of Corruption. *International Social Science Journal 48*(149): 321–35.

Kaplan, R. (2011), *Social Network Analysis – Extracting Its Potential in Health Care Fraud Detection',*, http://www2.mz.gov.pl/wwwfiles/ma_struktura/docs/zal_19_rk_19102011.pdf.

Kaufmann, D., Mastruzzi, M. & Kraay, A. (2010). *The worldwide governance indicators. Methodology and analytical issues* (No. 5430). Washington, DC.

Kimball, R. (2014), *Newly Emerging Best Practices for Big Data*, https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/whitepaper/.

Klasnja, M. (2016). Corruption and the Incumbency Disadvantage: Theory and Evidence. Journal of Politics, 77(4), 928–942.

KPMG (2016), *Using analytics successfully to detect fraud', Global Data & Analytics, Trusted Analytics Series No. 4*, https://assets.kpmg.com/content/dam/kpmg/pdf/2016/07/using-analyticssucessfully-.

Kurtz, M. J. & Schrank, A. (2007a). Growth and governance: a defense. *Journal of Politics, 69*(2): 563–569.

Kurtz, M. J. & Schrank, A. (2007b). Growth and governance: models, measures, and mechanisms. *The Journal of Politics, 69*(2): 538–554.

Lambsdorff, J. G. (2006). *Measuring corruption–the validity and precision of subjective indicators (CPI).* Measuring corruption, 81.

Locatelli Giorgio, M. (2017), *Corruption in public projects and megaprojects: There is an elephant in the room*, International Journal of Project Management.

Mamavi, O. (2017), "How do strategic networks influence awarding contract? Evidence from French public procurement", *International Journal of Public Sector Management*, Vol. 30/4, https://www.emeraldinsight.com/doi/pdfplus/10.1108/IJPSM-05-2016-0091.

Mazur, E. (2015), "Data Analytics: A Tool for Building Trust in Government", *The Journal of Government Financial Management*, https://www.highbeam.com/doc/1P3-.

Mungiu-Pippidi, A. (Ed.) (2011). *Contextual choices in fighting corruption: Lessons learned*. Oslo: Norwegian Agency for Development Cooperation.

Mungiu-Pippidi, A. (2006). Corruption: Diagnosis and Treatment. *Journal of Democracy 17*(3): 86–99.

North, D. C., Wallis, J. J., & Weingast, B. R. (2009). Violence and Social Orders. A Conceptual Framework for Interpreting Recorded Human History. Cambridge, UK: Cambridge University Press.

OECD (2018a), *Digital Government Review of Colombia: Towards a Citizen-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264291867-en.

OECD (2018b), *SMEs in Public Procurement: Practices and Strategies for Shared Benefits*, OECD Public Governance Reviews, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264307476-en.

OECD (2016a), *Advanced Analytics for Better Tax Administration: Putting Data to Work*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264256453-en.

OECD (2016b), *Integrity Framework for Public Investment*, OECD Public Governance Reviews, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264251762-en.

OECD (2013). *Government at a Glance 2013*, OECD Publishing, Paris, https://doi.org/10.1787/gov_glance-2013-en.

OECD (2009). *Towards a sound integrity framework: instruments, processes, structures and conditions for implementation*. Paris: OECD.

OCDE (2007), *Integrity in Public Procurement : Good Practice from A to Z*, OECD Publishing, Paris, https://doi.org/10.1787/9789264027510-en.

OLAF (2018), *Fraud in Public Procurement: a collection of red flags and best practices*.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy, 115*(2), 200–249. doi:10.1086/517935.

Olken, B. A., & Pande, R. (2012). Corruption in developing countries. *Annual Review of Economics, 4*(1):479-509.

Robson, E. (2010), *Evaluating Major Infrastructure Projects: How Robust are our Processes*, Strengthening Evidence-based Policy in the Australian Federation, Volume 1.

Rose, R., & Peiffer, C. (2012). *Paying bribes to get public services: A global guide to concepts and survey measures* (No. SPP 494). Glasgow: Centre for the Study of Public Policy.

Rose-Ackerman, S. (1978). *Corruption: A Study in Political Economy.* New York: Academic Press.

Straub, Stéphane. (2014). *Political Firms, Public Procurement, and the Democratization Process*. Toulouse.

Tóth, B. & Fazekas, M. (2017). *Compliance and strategic contract manipulation around single market regulatory thresholds – the case of Poland*. GTI-WP/2017:01, Budapest: Government Transparency Institute.

Transparency International. (2018). *Corruption perceptions index 2017*. Berlin: Transparency International.

Transparency International. (2012). *NIS assessment toolkit.* Berlin: Transparency International.

Transparency International. (2006). Handbook for Curbing Corruption in Public Procurement. Berlin: Transparency International.

Trapnell, S. (2011). Actionable Governance Indicators: Turning Measurement into Reform. *Hague Journal on the Rule of Law 22*(3): 317–48.

Treasury, H. (2009), *Risk management assessment framework: a tool for departments*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/191516/Risk_management_assessment_framework.pdf (accessed on Jan 2019).

U.S. Department of Health & Human Services, Centers for Medicare and Medicaid Services (2015a), *Annual Report to Congress on the Medicare and Medicaid Integrity Forums for Fiscal Year 2015*, https://www.cms.gov/About-CMS/Components/CPI/Downloads/2015-final-rtc-06232017.pdf.

U.S. Department of Health and Human Services, Centers for Medicare and Medicaid Services (2015b), *Fraud Prevention System Return on Investment, Fourth Implementation Year*, https://www.cms.gov/About-CMS/Components/CPI/Downloads/Fraud-Prevention-System-Returnon-Investment-Fourth-Implementation-Year-2015.pdf.

U.S. Government Accountability Office (2015), *A Framework for Managing Fraud Risks in Federal Programs*, https://www.gao.gov/assets/680/671664.pdf.

U.S. Government Accountability Office (2013), *Data Analytics for Oversight and Law*, https://www.gao.gov/assets/660/655871.pdf.

Warren, M. E. (2003). What Does Corruption Mean in a Democracy? *American Journal of Political Science 48*(2): 328–43.

Wells, J. (2015), *Corruption in the construction oif public infrastructure: Critical issues in project preparation*, U4 Anti-Corruption Resource Centre.

Williams, A. (2015). A Global Index of Information Transparency and Accountability. *Journal of Comparative Economics 43*(3): 804–24.

World Bank (2009). *Fraud and Corruption: Awareness Handbook*. Washington DC: World Bank.

Analytics for Integrity :

Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessments

For more information visit:
www.oecd.org/gov/ethics/