GOVERNANCE

GOVERNANCE

EQUITABLE GROWTH, FINANCE & INSTITUTIONS INSIGHT

Governance Risk Assessment System (GRAS)

Advanced Data Analytics for Detecting Fraud, Corruption, and Collusion in Public Expenditures

WORLD BANK GROUP
Governance

ANTICORRUPTION
for DEVELOPMENT

Contents

Acknowledgments

This report was prepared by a team led by Daniel Ortega Nieto (Senior Public Sector Management Specialist, World Bank) and formed by Mihály Fazekas (Government Transparency Institute and Central European University), Bianca Vaz Mondo (Government Transparency Institute), Bence Tóth (Government Transparency Institute and University College London), and Rafael Braem Velasco (Aethos Brazil). Ruben Interiam, and Igor Carpanese provided valuable research assistance. The team worked under the overall guidance of Adrian Fozzard (Practice Manager, Governance Public Sector Institutions, Latin America and the Caribbean, World Bank) and Robert Taliercio (Director, Equitable Growth, Finance and Institutions, for Latin America and the Caribbean, World Bank). Other World Bank and international experts consulted during development of the Governance Risk Assessment System include: Sinue Aliram, Alexandre Borges, Francesca Recanatini, James Anderson, Gavin Ugale, Alexandra Habershon, Rafael Munoz, and Gustavo Bozzetti. The development of the Governance Risk Assessment System (GRAS) was made possible by financing from the Spanish Fund for Latin America (SFLAC) Trust Fund.

Executive Summary

Corruption poses a significant threat to development and has a disproportionate impact on the poor and most vulnerable. Government agencies struggle to identify fraud and corruption in public expenditures. Risk assessments usually rely on manual analysis and follow-up on specific complaints or anecdotes which requires substantial resources. Assessments are often limited in scope and ineffective, failing to generate the evidence needed to build strong cases. The World Bank developed the Governance Risk Assessment System (GRAS), a tool that uses advanced data analytics to improve the detection of risks of fraud, corruption, and collusion in government contracting. GRAS increases the efficiency and effectiveness of audits and investigations by identifying a wide range of risk patterns. GRAS makes use of public data and is based on a robust and comprehensive conceptual framework which draws on insights from experienced practitioners and sound academic research.

This report presents GRAS's main features, examples of GRAS implementation, and outlines the steps government agencies can take in applying GRAS in their countries. GRAS was developed in Brazil, where it has been piloted in four subnational governments and has helped to investigate fraud, corruption and collusion in public procurement. Concrete results include the identification of over 850 suppliers with strong indication of collusive behavior, 450 suppliers likely registered under strawmen, 500 cases of conflict of interests involving suppliers owned by public servants, and about 4500 companies with connections to political campaigns, among other examples.

Introduction

Public procurement is highly vulnerable to corruption, given the complexity of procurement processes, the high degree of official discretion, and the close interaction between the public and private sectors (OECD, 2016). The costs of procurement corruption are enormous. Conservative estimates from the empirical literature suggest that corruption can amount to about 8 percent of the value of procurement contracts worldwide, reaching some US$ 880 billion lost yearly.[1]

Interventions to prevent corruption in public procurement have focused on procedural standardization, strengthened transparency, reduced scope for discretion, and digitization in the procurement process. Evidence on the impact of these approaches on corruption is mixed (Bajpai and Myers, 2020, Fazekas and Blum, 2021). Methods for detecting and investigating corruption are inherently limited compared to the scale of the problem. Corruption risk assessments usually rely on manual analysis and follow-up on specific complaints or anecdotes. This is time-consuming and inefficient, requiring the use of vast human and financial resources. Assessments are often limited in scope and ineffective, failing to generate the evidence needed to build strong cases to identify potential risks.

Improvements in data collection, digitization, and public sector transparency have unlocked opportunities to better address corruption, providing for the development of data-driven approaches. The World Bank developed and implemented the Governance Risk Assessment System (GRAS) in Brazil to exploit the opportunities offered by a data rich environment. GRAS broadens the scope of risk assessments in public procurement, covering multiple risk patterns linked not only to corruption, but also to fraudulent or collusive practices. It improves the accuracy of corruption detection in government contracting; consequently increasing the efficiency and effectiveness of audits and investigations. GRAS is based on a robust and comprehensive conceptual framework covering 60 red flags, linked to 23 broad risk patterns along 4 dimensions. GRAS uses large volumes of contract-level and company data from public datasets: electoral registers; social program beneficiaries; public sector payroll; and blacklisted firms. Cross-referencing these datasets and leveraging algorithms, GRAS screens relationships among stakeholders, indicating risks associated with collusive practices, supplier characteristics and political connections. As a result, GRAS' interface can provide users with comprehensive aggregated risk reports that can have multiple use cases; for example, pre-screening firms before being awarded public contracts; anti-corruption investigations by internal and external control agencies; conflict of interest reviews of public or elected officials by monitoring bodies, among others. GRAS can also be used to identify potential tax fraud and collusive networks, as well as atypical spending patterns in strategic sectors. GRAS has been used in Brazil at the State and Municipal levels, leading to the identification of potential fraud, corruption, and collusion worth millions of US dollars.

1. https://blogs.worldbank.org/developmenttalk/reducing-corruption-public-procurement

2. The Governance Risk Assessment System: Conceptual Framework and Structure

Innovative and operationally relevant risk detection methodology

The Governance Risk Assessment System (GRAS) systematically analyzes large public procurement and linked administrative datasets in order to generate actionable risk reports. If used as an integral part of regular investigations and audits by well-trained users, it can make a profound contribution to effective anti-corruption. Integrating GRAS into operational and investigative processes has to be supported by staff training, a user manual and tailored operational processes.

Data-driven risk assessment tools can rarely point at actual cases of fraud, corruption or collusion, rather they identify transactions or actors of high risk (Fazekas et al, 2019). The risks identified indicate a higher likelihood of wrongdoing for transactions, entities (such as specific suppliers or procuring agencies) or individuals based on some generally validated features which we call risk indicators or red flags. Any risk-based approach inevitably will flag some transactions or entities as high risk even though they are compliant with the rules (i.e. false positives).[2] However, on average high risk transactions or entities are expected to be more susceptible to fraud, corruption and collusion than those that are assessed as low risk.

The red flags used by GRAS are indicators of behaviors that are associated with fraud, corruption or collusion based on lessons learned by auditors, investigators, and academics (Velasco, 2019). Datasets and data science approaches allow the derivation of robust and valid risk indicators across a wide variety of markets and countries even though case evidence is sparse. Risk indicators can be identified because fraud, corruption, and collusion involve specific forms of economic behavior that consistently leave traces, such as inexplicably successful government contractors, low bid participation rates, and abnormal cost overruns (Fazekas et al, 2018). These signals may appear as anomalies or outliers in the data while others may represent the norm that is average market behavior. In the latter case, corruption and collusion is systemic in the market leaving integrity as the outlier.

2. Similarly, such a system will inevitably generate "false negatives" too, that is, real malfeasance cases that cannot be detected. In the case of GRAS, this is minimized by the sheer comprehensiveness and breadth of the risk assessment framework. Nevertheless, some situations can be poorly identified based on big data alone, and corruption, collusion and fraudulent "technologies" in contracting tend to evolve with time, thus requiring permanent revisions and adjustments to the framework to better reflect new or adapted risk patterns.

GRAS' comprehensive and detailed risk assessment framework allows for lowering false positive rates. GRAS can draw on a wide range of related risk indicators allowing for triangulation, parametrization, and validation of predictions. This lowers the frequency of imprecise signals. Many risk indicators are unreliable individually: the presence of a single red flag may lead to a high rate of false positives. The accuracy of risk assessment can be increased by combining or collating multiple risk indicators characterizing the same underlying risky behavior. In this respect, GRAS offers a high level of flexibility to knowledgeable users with filtering functions that enable them to explore different types of high-risk profiles based on criteria most relevant to their scope of action. Auditors, for instance, may define the parameters for prioritizing among red-flagged suppliers based on their agency's strategies, or on their empirical experience about different risk patterns that are commonly observed in combination in their jurisdictions. Therefore, instead of using pre-defined risk-ranking parameters or intransparent methodologies, GRAS provides for a more context-tailored definition of complex risk profiles by users themselves, based on transparently defined and documented indicators. While this offers a high degree of flexibility to users, it also requires sophistication and data analytic proficiency from them. As many fundamental analytical decisions are made by the user rather than decided by the system, inadequately trained staff can quickly reach seemingly attractive, but incorrect conclusions. Hence, offering in-depth technical training along with the introduction of GRAS is a must.

The accuracy and usefulness of red flags depend on the availability and combination of the specific data points which are needed for their calculation. Oftentimes, the relevant data are located in different datasets. The core dataset indispensable for the application of GRAS is micro-level public procurement data (detailed data on procurement processes, contracts and purchased items). Such data are required for the calculation of all red flags in the risk framework and depending on their comprehensiveness, they may be sufficient to make one third of the framework operational where complementary datasets are not available.

Additional datasets enrich GRAS and its indicators. Firm level data is particularly useful. Data on incorporated legal entities include: business registry information, shareholder and management data, information on employees, and information on firms' financial activities. Other datasets that can be used in GRAS include: electoral data; company blacklists or debarment lists; public sector payroll data; asset and interest declarations of politicians and bureaucrats; social registries (e.g. lists of welfare recipients that can help identify strawmen); and criminal records of individuals.

GRAS is designed to allow for a robust and scalable risk assessment process, incorporating large volumes of data across a whole country or jurisdiction. The framework can be flexibly adapted to different contexts with varying data readiness.

GRAS Red Flag Framework

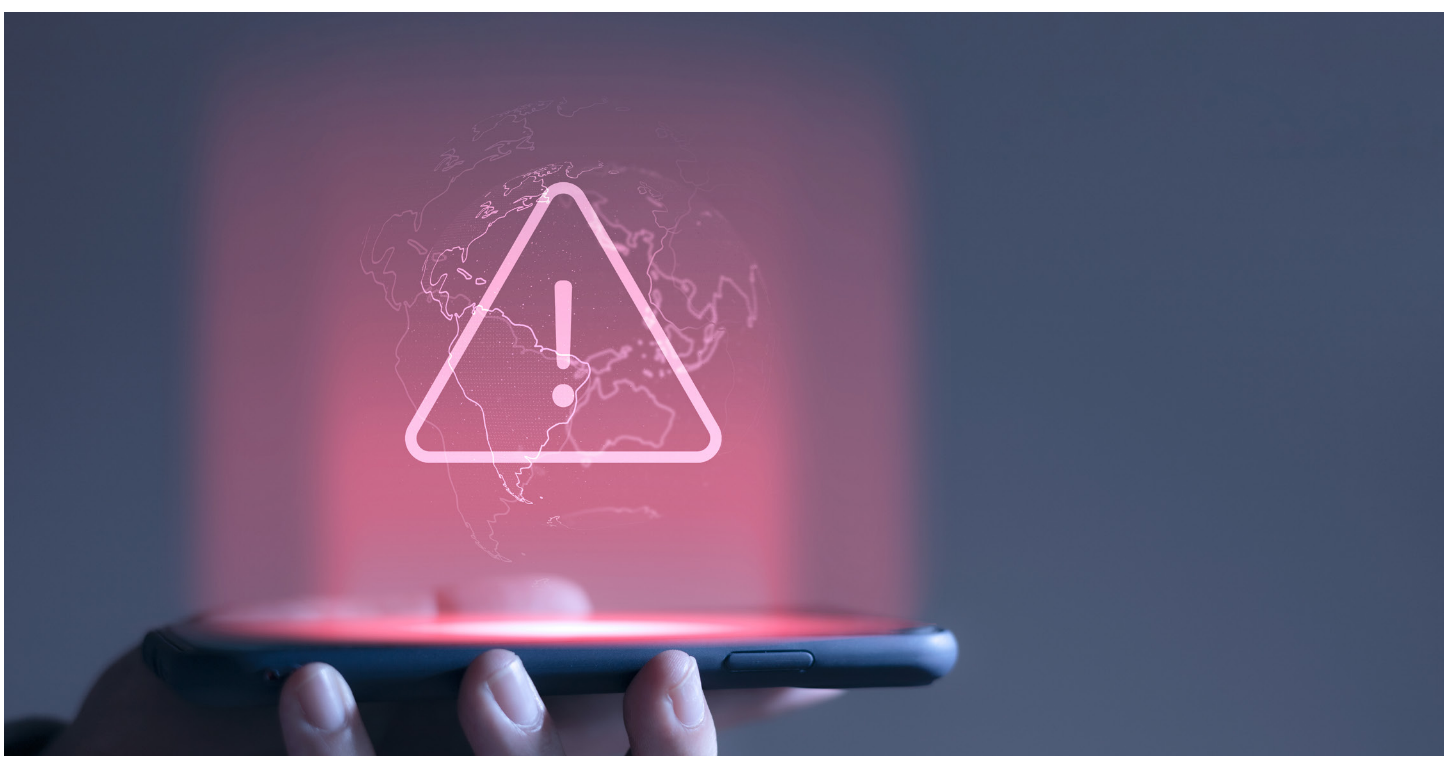
GRAS indicators comprehensively assess a range of corrupt practices in public expenditures. The precise definitions of these behaviors may vary from country to country or even over time within the same country, still, a few high-level considerations are important to delineate the scope of behaviors assessed.

Procurement fraud, for example, may involve any act or omission by an individual actor with the intent of deceiving or misleading other involved parties in order to obtain a (financial) advantage (World Bank, 2009). It may also occur independently of corruption (e.g. bribery) or fraud (e.g. a bidder presents a forged certificate in its qualification documentation), but it is often observed in connection to corrupt or collusive practices (e.g. a shell company with hidden connections to a politician bids in a tender, or operates in collusion with other bidders).

Corruption in public procurement refers to the allocation and performance of public contracts by distorting principles of open and fair government contracting in order to benefit some to the detriment of all others (Fazekas & Kocsis, 2020). The aim of corruption is to steer the contract to the favored bidder without detection in an institutionalized and recurrent fashion, by avoiding or biasing competition (e.g. unjustified sole sourcing or direct contract awards) in order to favor a certain, connected bidder (e.g. tailoring specifications to a particular company) (World Bank, 2009). Such corrupt behaviors may manifest themselves in public tendering processes or outcomes, while they may also involve suppliers and linked individuals with risky features (Fazekas et al, 2018).

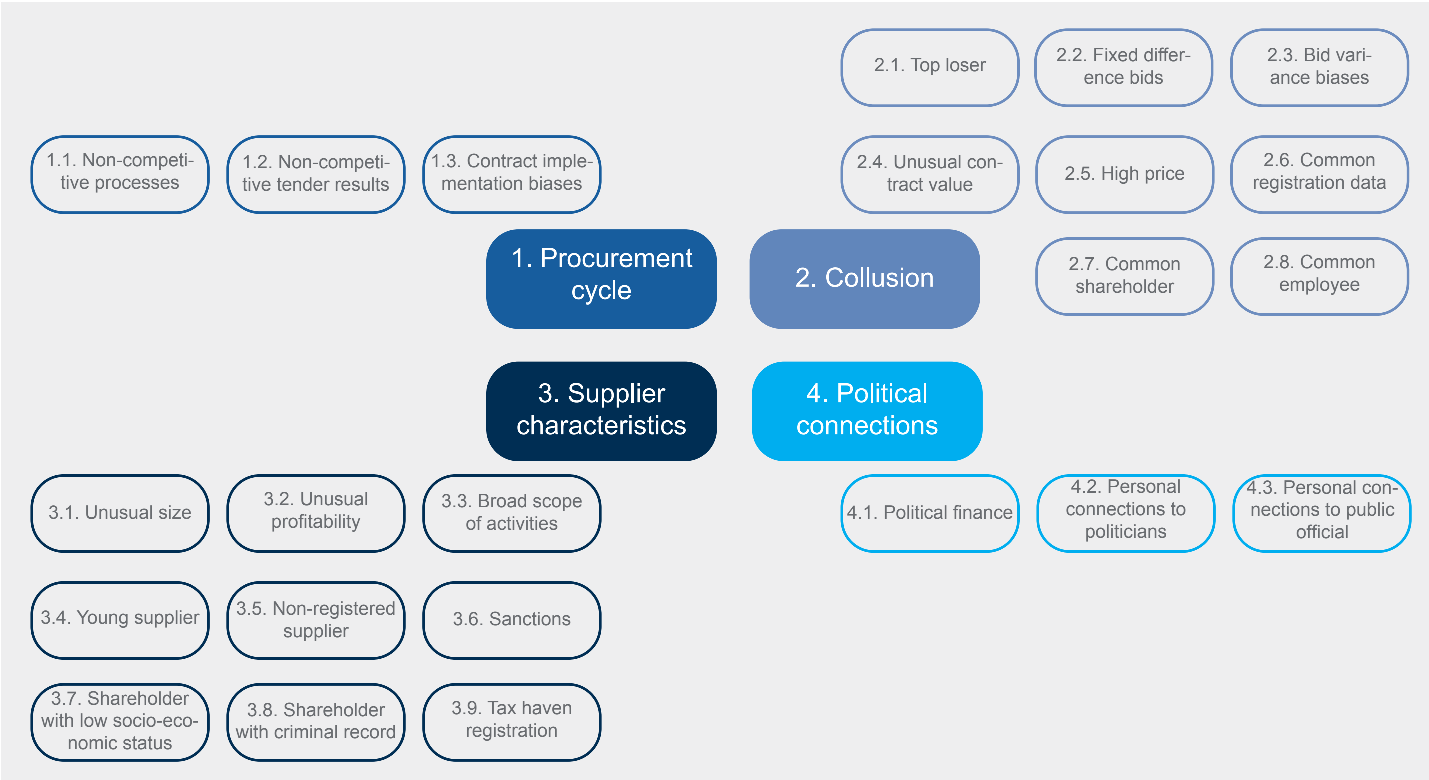
Collusion is a distinct phenomenon from corruption, as it does not require the participation of a public actor, only the coordination among supposedly competing companies. Collusion in public procurement entails coordination of companies' decisions regarding price, quantity, quality, or geographical presence to eliminate competition in public procurement processes and earn a mark-up above competitive conditions. This strategy can only be sustained if (a) companies can coordinate, (b) is internally sustainable (credible punishment system and effective detection of cheating), (c) it is externally sustainable (ability to exclude market entrants), and (d) the scheme can go undetected and circumvent sanctions (Fazekas and Tóth, 2016).

GRAS rests on 4 risk groups with 3 of them targeting fraud and corruption and 1 dedicated to inter-bidder collusion. These 4 groups gather the 60 core red flags of GRAS as classified under 23 broader risk pattern categories (Figure 1).



> > >

FIGURE 1 - Overview of GRAS Risk Groups and Risk Areas Covered



The first risk group includes those indicators which are closely tied to the different **phases of the procurement cycle** (OECD, 2016), namely tendering processes such as non-advertisement of call for tenders; tendering results such as single bid submitted; and contract implementation such as large cost overruns. The second group consists of indicators approximating **inter-bidder collusion**, such as indicators of coordination opportunities among presumable competitors (e.g. common shareholder) and indicators of likely coordinated bidding behavior (e.g. unusual bid price variance) (Adam et al, 2022). The third group captures fraud and corruption risks centered on **supplier characteristics**. These risk factors might relate to the company's registry information such as registration in a tax haven jurisdiction; company financial records such as unusual profitability; multiple economic activities; or the company's shareholders such as the criminal record of a company owner. The fourth group of risk factors is relational, capturing risks associated with political connections of a supplier. Connections can be established through personal **connections to politicians or public official**, or through companies' political finance activities, that is, donating to an electoral campaign or political party (OECD, 2019).

Risk Group 1: Procurement Cycle

The risk group for the procurement cycle comprises indicators of corrupt and fraudulent behaviors in public procurement processes. These indicators capture risky behaviors in the three main phases of public procurement: tendering, award, and contract implementation. They are indicative of deliberate manipulation of public procurement aiming to favor a particular supplier. While these indicators are highly relevant on their own, they are especially useful as they further support and strengthen indicators from risk groups 3 (supplier characteristics) and 4 (political connections). Table 1 enumerates 8 red flags related to each of the 3 risk patterns.

> > >

TABLE 1 - Individual Risk Indicators in the Procurement Cycle Risk Group (higher values indicate higher risk)

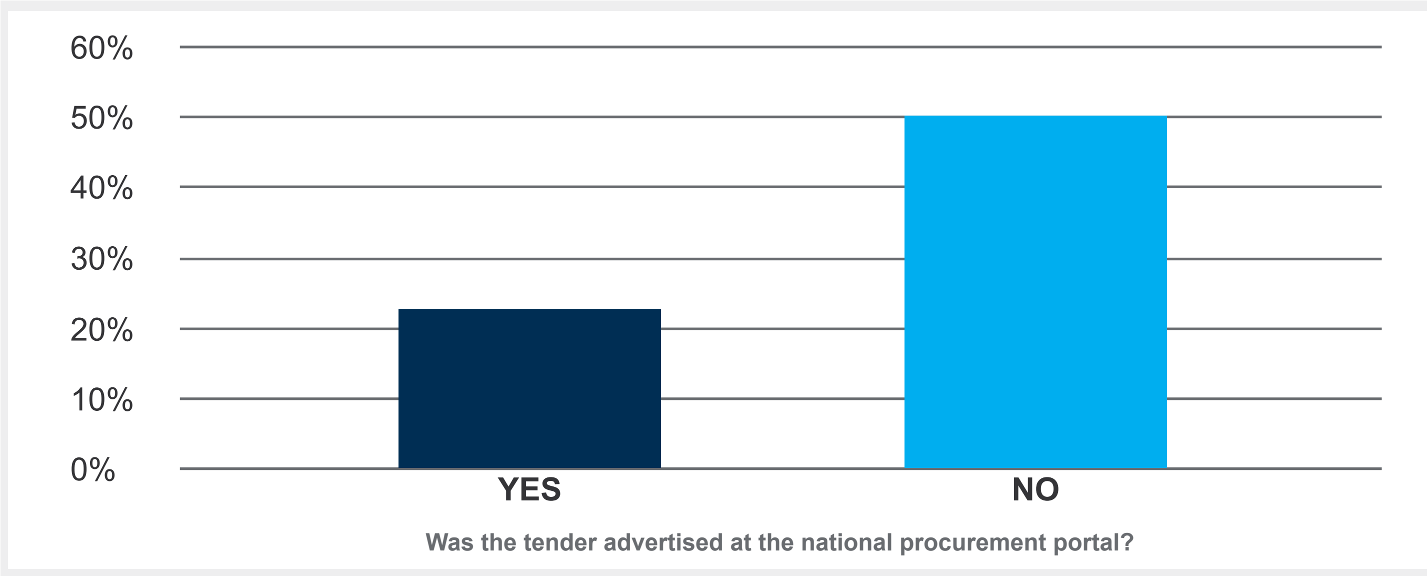
Risk pattern	*Red flag/ indicator*	*Description*
1.1. Non-competitive processes	1.1.1. Contract share through non-competitive procedures	Percentage of contracts won in high-risk, non-competitive procedure types (direct awards, invitation procedures, etc.) compared to all contracts won in a given time period (based on number of contracts or contract value)
	1.1.2. Contract share after call for tenders absent	Percentage of contracts won in tenders without prior call for tenders published compared to all contracts won in a given time period (based on number of contracts or contract value)
	1.1.3. Contract share after shortened advertisement period	Percentage of contracts won in tenders where advertisement period (time between tender publication and submission deadline) is too short compared to all contracts won in a given time period (based on number of contracts or contract value)
1.2. Non-competitive tender results	1.2.1. Contract share as single bidder	Percentage of contracts won as single bidder compared to all contracts won in a given time period (based on number of contracts or contract value)
	1.2.2. High winning rate	Percentage of winning bids compared to all bids presented in a given time period (based on number of bids or bid value)
	1.2.3. Contract share in buyer's portfolio	Percentage of contracts won compared to the buyer's total annual procurement (based on number of contracts or total value spent)
1.3. Contract implementation biases	1.3.1. Contract share with sizeable cost overruns	Percentage of contracts won with cost overrun above a given threshold (e.g. 5% more expensive than planned) compared to all contracts won in a given time period (based on number of contracts or contract value)
	1.3.2. Contract share with sizeable delivery delays	Percentage of contracts won with delivery delay above a given threshold (e.g. 5% longer than planned project) compared to all contracts won in a given time period (based on number of contracts or contract value)

Among the red flags in the non-competitive processes subgroup, the **non-publication of call for tenders** is one of the most widely used (Fazekas et al, 2016). This indicator is initially defined for each tender where a call for tender publication is either present (indicator value=1) or absent (indicator value=0). Then it can be aggregated to the level of a supplier based on all the contracts won by the company in a period, resulting in the share of contracts without prior call for tender publication. This red flag points at potential corruption because not publishing the call for tenders makes it less likely that eligible bidders notice the bidding opportunity, weakening competition and allowing the contracting body to more easily award the contract to a favored and/or connected company. This pattern is especially indicative of risks if it happens repeatedly with the same company. Naturally, buyers rather than bidders decide whether or not to publish. However, corruption is typically well-organized and based on a network of corrupt individuals across the public and private sectors. Hence, a company repeatedly receiving information about non-advertised bidding opportunities is more likely to have connections and be favored. Countries differ in the degree to which non-advertisement is allowed, depending on the procedure type, with a few countries where non-advertisement is virtually non-existent.[3]

3. For a systematic mapping of regulatory requirements for European countries see: http://europam.eu/ and country overview statistics for selected countries: https://www.procurementintegrity.org/countries (integrity indicators panel).

> > >

FIGURE 2 - Call for Tenders Advertisement and the Likelihood of Single Bidding in Mexico, Federal Procurement, 2017-2018



Source: Government Transparency Institute calculation

Among red flags in the non-competitive tender results category, **single bidding** is by far the most widely used (e.g. European Commission, 2022, p. 227). Single bidding occurs when only one bid is submitted in an otherwise competitive market. While competition can be limited for a range of non-corrupt reasons, corrupt deals almost invariably require some form of limited competition in order to award contracts to connected firms (Fazekas & Kocsis, 2020). The association between non-competitive tendering processes such as non-publication of call for tenders and single bidding supports the validity of these red flags (Figure 2). This indicator is initially calculated on the level of lot or contract and then can be aggregated to the level of the supplier in order to characterize the supplier's bidding behavior.

Corruption can also take place during contract implementation. **Sizeable cost overruns** are an important red flag. While there can be justifiable reasons for increasing contract value during implementation, contract modifications can be used to extract unwarranted profits, cover the costs of bribes spent to secure a contract, or cover expenses if the favored company could only win the contract by offering a competitive or even below-market price (Collier et al, 2016; Alexeeva et al, 2008). A crucial challenge with this indicator, as with many other indicators in GRAS, is the definition of sizeable: the threshold above which cost overruns may be considered as higher risk. There is no universally agreed threshold for risky behaviors, but data analytics exploiting correlations among red flags can lead to robust and fine-grained threshold definition (Fazekas and Kocsis, 2020). 20 percent cost overrun threshold is considered high risk for World Bank funded projects in Fazekas and Márk (2017). This indicator is first calculated for each contract and then can be aggregated to the level of supplier in order to characterize organizational behavior.

Risk Group 2: Collusion

The risk group for collusion comprises indicators which signal collusive behavior among bidders such as cartels and bid-rigging practices. GRAS collusion indicators capture collusive outputs, such as coordinated bid prices or persistent losers in tenders, and the means by which companies may coordinate bidding, such as common shareholders or employees across supposedly competing firms. The 24 collusion risk indicators are grouped under 8 broader risk patterns: top loser, fixed difference bids, bid variance biases, unusual contract value, high price, common registration data, common shareholder, and common employee (Table 2).

Collusive behaviors involving private actors, i.e. bidders, may take place without the participation of public sector actors (e.g. officials in a buying organization). However, GRAS can identify if corruption and collusion take place together by simultaneously applying collusion and corruption-related red flags. Under the collusion risk group, indicators are calculated at the tender, organization, or market levels before they can be related to individual suppliers. This is because some indicators require the definition of a market and the characterization of bidding behavior in relation to behaviors of other market participants (Fazekas and Tóth, 2016).

> > >

TABLE 2 - Individual Risk Indicators in the Collusion Risk Group (higher values indicate higher risk, unless otherwise specified)

Risk pattern	*Red flag/ indicator*	*Description*
2.1. Top loser	2.1.1. Low winning rate	Percentage of winning bids compared to all bids presented in a given time period (based on number or value; lower values = higher risk)
	2.1.2. Number of competitors	Number of companies against which the Top Loser lost in a given time period (lower values = higher risk)
	2.1.3. Number of wins against Top Losers	Number of bids won against Top Losers in a given time period
	2.1.4. Winning rate against Top Losers	Percentage of bids won against Top Losers compared to all bids won in a given time period (based on number or value)
	2.1.5. Number of Top Loser competitors	Number of Top Losers defeated by the bidder
2.2. Fixed difference bids	2.2.1. Number of colluding partners with fixed difference bids	Number of companies with which the pattern of fixed difference bids is present, i.e. the company in question and another bidder repeatedly present a pair of bids with the same absolute or percentage difference over different tenders
	2.2.2. Number of bids with fixed difference bids	Number of individual tenders/bids (items or lots) in which the company bid in a fixed difference pattern, i.e. the company in question and another bidder repeatedly present a pair of bids with the same absolute or percentage difference over different tenders
	2.2.3. Frequency of fixed difference bids	Percentage of bids with fixed difference pattern compared to all bids presented in a given time period (based on number or value), i.e. the company in question and another bidder repeatedly present a pair of bids with the same absolute or percentage difference over different tenders
2.3. Bid variance biases	2.3.1. Bid share in low variance tenders	Percentage of bids submitted on tenders with the Coefficient of Variation (standard deviation divided by the mean of bids) very low, i.e. close to 1.
	2.3.2. Bid share in high relative bid distance tenders	Percentage of bids submitted on tenders with the relative distance between the lowest and second lowest bid (distance between the lowest and second lowest bid divided by the value of the lowest bid) very high
2.4. Unusual contract value	2.4.1. Contract share with contract value violating Benford's Law	Percentage of contracts won in a given time period whose first digits of contract prices violate Benford's Law (based on number of contracts or contract value)
2.5. High price	2.5.1. Contract share with very high relative contract value	Percentage of contracts won with a relative contract value (winning bid price divided by estimated value) above a given threshold (e.g. 0.98) in a given time period (based on number of contracts or contract value) (also an indicator of potential corruption)

CONTINUED

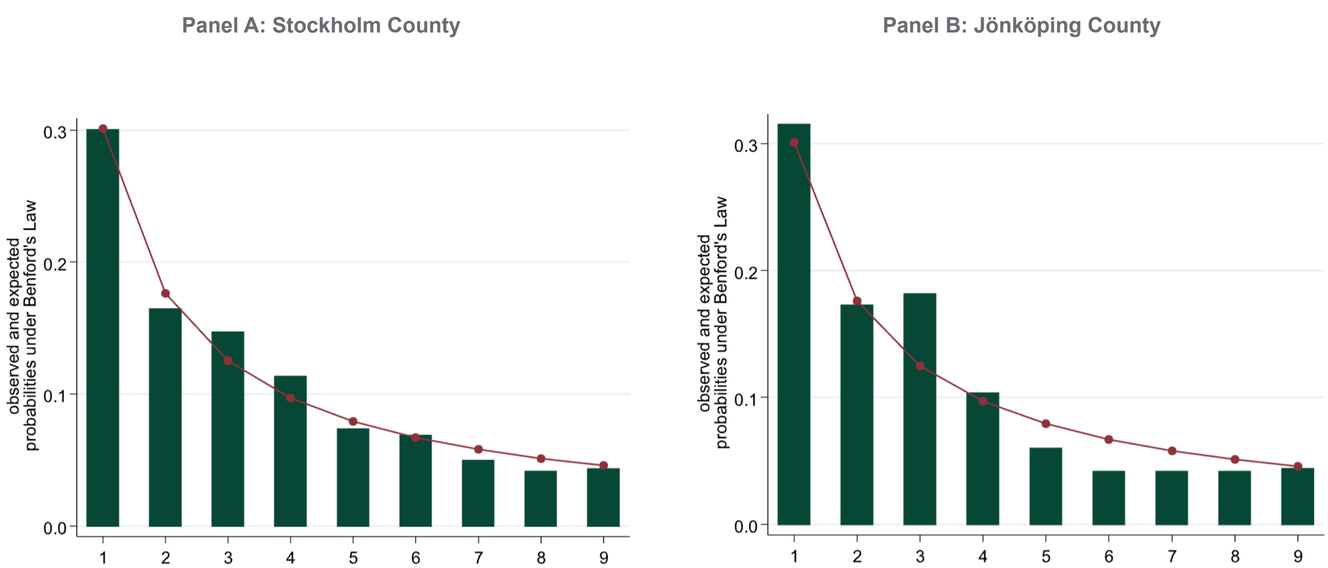
Risk pattern	Red flag/ indicator	Description
2.6. Common registration data	2.6.1. Number of competitors sharing registration data	Number of competing bidders with which the company shares the same registration information (e.g. same phone number, same postal address, same website, same legal representative, same accountant)
	2.6.2. Number of tenders with bidders sharing registration data	Number of individual tenders in which a competitor shares the same registration information
	2.6.3. Share of contracts won against competitors sharing registration data	Percentage of contracts won against bidders with the same registration information compared to all contracts won in a given time period (based on number of contracts or contract value)
2.7. Common shareholder	2.7.1. Number of competitors with common shareholder	Number of competing bidders with which the company shares a common shareholder
	2.7.2. Number of tenders with competitors sharing a shareholder	Number of individual tenders in which a competitor shares the same shareholder
	2.7.3. Share of contracts won against competitors with common shareholder	Percentage of contracts won against bidders sharing the same shareholder in a given time period (based on number of contracts or contract value)
	2.7.4. Number of competitors in the same corporate group	Number of competing bidders which belong to the same corporate group
	2.7.5. Number of tenders with competitors in the same corporate group	Number of individual tenders in which a competitor belongs to the same corporate group
	2.7.6. Share of contracts won against competitors in the same corporate group	Percentage of contracts won against bidders belonging to the same corporate group in a given time period (based on number of contracts or contract value)
2.8. Common employee	2.8.1. Number of competitors with common employee	Number of competing bidders employing someone associated with the company (e.g. employee, shareholder)
	2.8.2. Number of tenders with competitors sharing an employee	Number of individual tenders in which a competitor employs someone associated with the company
	2.8.3. Share of contracts won against competitors with common employee	Percentage of contracts won against bidders employing someone associated with the company in a given time period (based on number of contracts or contract value)

One of the price-based collusion indicators is Benford's law. Benford's law is a statistical rule commonly used in forensic accounting, election monitoring, and in the study of economic crime including collusion and corruption (Berger and Hill, 2015). It posits that the first digit of most naturally occurring sets of numerical data follows a specific distribution.[4] Competition as such can be regarded as a natural process, hence contract prices in public procurement markets, assuming that prices are distributed across multiple magnitudes, should follow Benford's law.[5] As an example, we show the distribution of the first digit of contract prices of Swedish construction contracts (Fazekas and Toth, 2016). Panel A shows contracts from the Stockholm region, which shows that the actual distribution is almost identical to our theoretical expectation. However, number 3 was overrepresented, while numbers 5 to 8 were underrepresented compared to the theoretical expectation in Jönköping County. These distribution differences are also statistically significant, suggesting that contract prices are likely not a result of a competitive process and triangulating collusive behavior with other indicators is warranted.

4. For example, 30.1% of first digits shall be 1, 17.6% number 2, 12.5% number 3 etc.
5. For example, if there is a high threshold above which contracts are published (i.e. low value contracts are entirely cut-off), low value contracts being not part of the distribution could lead to the violation of Benford's law by default. Therefore, the difference between the theoretical and actual distribution of first digits ought to be calculated on a big enough sample of multi-magnitude distribution of contract values for it to be meaningful.

> > >

FIGURE 3 - Distribution of First Digits in Construction Contracts in Sweden vs Benford's Law (red line)[6]



Source: Fazekas and Toth, 2016, p. 74.

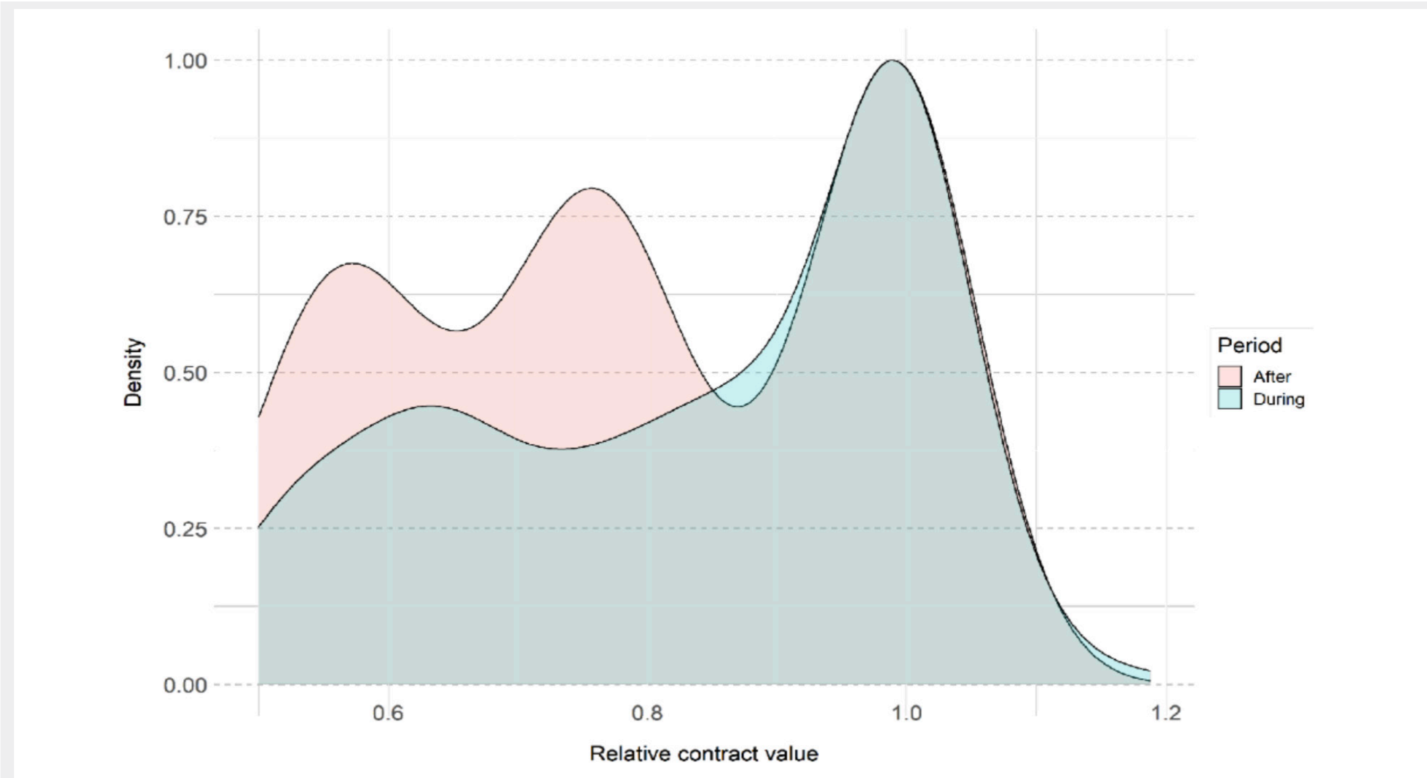
Another price-based indicator is relative price, that is, the awarded contract price divided by the initial estimation. The lower the relative price, the greater the savings that could be achieved by competition. Naturally, bid prices - hence contract prices - might be higher than the initial estimations, as budgeting for complex projects is difficult ex ante. However, repeatedly high relative prices are unusual in an otherwise competitive market: either buyers repeatedly underestimate costs, which is unlikely, or bidders coordinate their bid prices. Relative price is particularly useful when identifying potential bid-rigging schemes. Price increases unrelated to cost changes, long term price stability at unusually high levels indicate market performance problems (OECD, 2014; Oxera, 2013). Research shows that tenders with large discounts (relative price below 90 percent) are associated with the number, capacity and experience of bidding suppliers, whereas these characteristics are unrelated to prices if discounts are small (relative price is above 90 percent) (Morozov and Podkolzina, 2013). The literature on collusion risks suggests several price difference-based indicators and low bid price variance can also be used to distinguish between collusive and competitive tender processes and for modeling favor exchanges among colluding suppliers (Ishii, 2009).

The example below shows the distribution of relative prices of contracts awarded to companies that were found to participate in collusion in Spain during and after the proven cartel period (Figure 4).[7] It shows that the relative share of contracts having a relative price around 1, that is low savings compared to the initially estimated price, is high during the cartel period (blue), while larger savings, that is when relative price is less than 0.8, became more frequent after the prosecution of cartel members.

6. The exact product code category assigned to these contracts is 'construction work for pipelines, communication and power lines, for highways, roads, airfields and railways; flatwork'.
7. The cartel was active between 1996 to 2015, and the calculations are based on contracts awarded to the participating companies between 2005 and 2020. Note that some contracts might not have been rigged, the figure shows all contracts awarded to the prosecuted companies during and after the start of the legal case.

> > >

FIGURE 4 - Relative Price During and After the Proven Cartel Period[8]



Source: Fazekas and Toth, 2016, p. 74.

Risk Group 3: Supplier Characteristics

The risk group of Supplier characteristics comprises indicators for features of government suppliers that indicate likely fraudulent or corrupt behavior. Suppliers participating in corrupt exchanges act as vehicles of rent extraction and distribution. Just as corrupt government contracting differs from competitive tendering and contract implementation, companies participating in corrupt exchanges are expected to differ from their peers in a number of key features. High risk supplier characteristics are diverse. Table 3 enumerates 18 risk indicators covering 9 different risk patterns. Nearly all indicators in this group require combining company and public procurement indicators and data, in some cases indicators are based on linked datasets such as sanction or debarment lists. Supplier risk indicators generally build on company registry information such as location of registration (e.g. in a tax haven), financial performance data (e.g. turnover and profitability), and shareholder and management information (e.g. names and shares of owners) (Fazekas et al, 2018). Most risk indicators in this group are directly related to specific suppliers with some related to specific individuals, such as a shareholder, and then aggregated to the company level. As with many indicators in the GRAS framework, some indicators are sensitive to national and market conditions and have to be tailored to context. For example, unusual profitability requires setting of appropriate thresholds, taking into consideration normal market or sectoral profit rates, which vary over time depending on market dynamics.

8. The exact product code category assigned to these contracts is 'construction work for pipelines, communication and power lines, for highways, roads, airfields and railways; flatwork'.

> > >

TABLE 3 - Individual Risk Indicators in the Supplier Characteristics Risk Group (higher values indicate higher risk, unless otherwise specified)

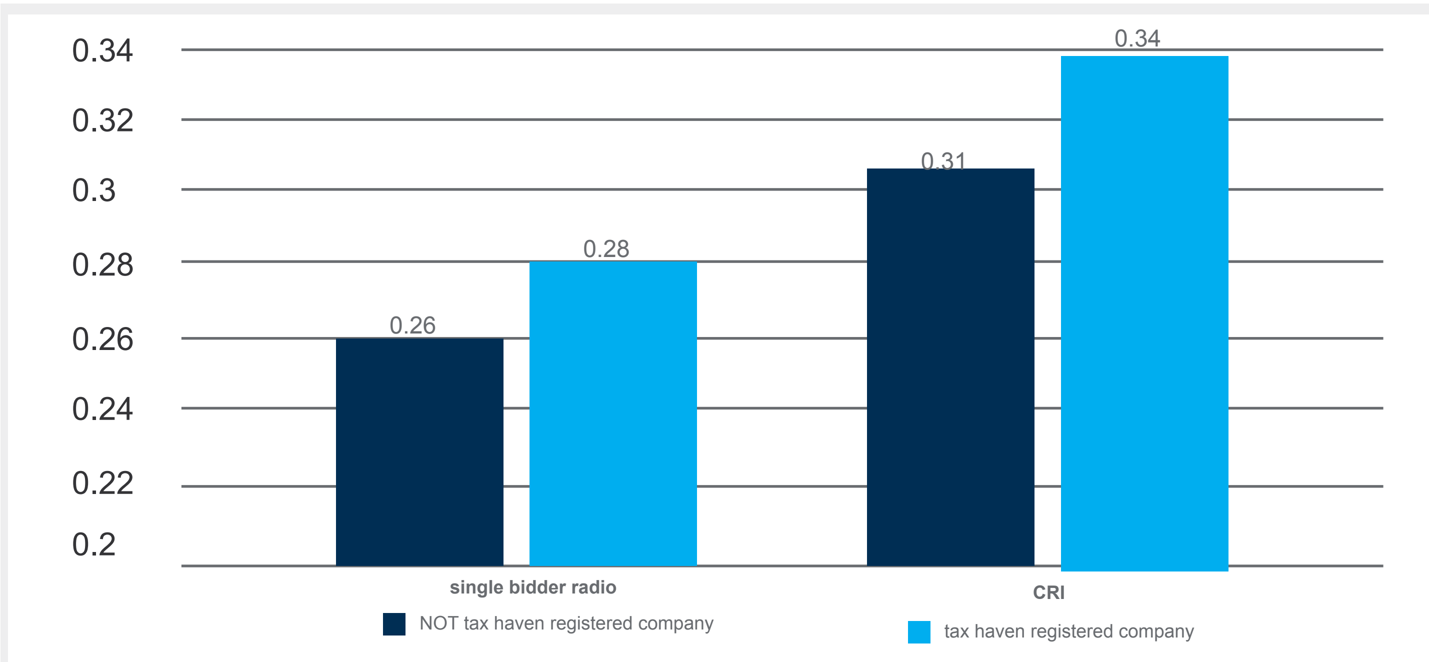
Risk pattern	*Red flag/ indicator*	*Description*
3.1. Unusual size	3.1.1. Contract revenue/ turnover ratio	Ratio of the total value of contracts won in a given time period compared to company turnover in the same period (very high, especially above 1 = higher risk)
	3.1.2. Contract revenue per employee	Total value of contracts won in a given time period divided by the number of company employees (values higher than the market average = higher risk)
3.2. Unusual profitability	3.2.1. Unusual profitability	Company profit rate in a given time period (values higher than the market average = higher risk)
3.3. Broad scope of activities	3.3.1. Number of economic activities	Number of economic activities, i.e. distinct detailed market codes (very high number or activities from different sectors = higher risk)
3.4. Young supplier	3.4.1. Period between incorporation and 1st contract	Number of days between the date of incorporation and the date of 1st contract won (lower values, typically less than 365 = higher risk)
3.5. Non-registered supplier	3.5.1. Contract before incorporation	Company incorporation date is after the contract award date
3.6. Sanctions	3.6.1. Sanctioned company	Company under sanctions: past or current
	3.6.2. Sanctioned shareholder/ legal representative	Company's shareholder/legal representative under sanctions: past or current
	3.6.3. Link to another sanctioned company	Company's shareholder/legal representative linked to another sanctioned company: past or current
	3.6.4. Contracts while sanctioned	Number/total value of contracts won while under a sanction (applied to the company or to a connected individual)
	3.6.5. Sanction relative duration	Duration of the sanction period (company or connected individual) relative to the total time the company has existed
	3.6.6. Period between incorporation and 1st sanction	Number of days between the date of incorporation and the starting date of its first sanction (lower values = higher risk)
3.7. Shareholder with low socio-economic status	3.7.1. Shareholder has low socio-economic status	Company shareholder with extremely low socio-economic status (e.g. registered as social beneficiary, low-income/low--skilled employee, or member of poor household)
	3.7.2. Status duration	Total number of months the shareholder has/had extremely low socio-economic status (e.g. for how long registered as social beneficiary, low-income/low-skilled employee, or member of poor household)
	3.7.3. Time overlap between status and company ownership	Number of months he/she was simultaneously a company shareholder and with low socio-economic status
3.8. Shareholder/legal representative with criminal record	3.8.1. Convicted shareholder	Company shareholder/legal representative has a criminal conviction
3.9. Tax haven registration	3.9.1. Company registered in tax haven	The company is registered in a tax haven (as denoted by Tax Justice Network's Financial Secrecy Index)
	3.9.2. Shareholder registered in tax haven	A company shareholder is registered in a tax haven (as denoted by Tax Justice Network's Financial Secrecy Index)

One of the most widely used red flags for suppliers is registration of the supplier or one of its significant shareholders in a secrecy jurisdiction. We identify tax havens using the Financial Secrecy Index of the Tax Justice Network (Tax Justice Network, 2022). Awarding a public contract to a company registered in a tax haven presents the risk that anonymous company ownership conceals a conflict of interest of a politically connected owner.

Another related risk is the potential loss of tax revenue from the successful supplier through tax evasion or tax avoidance (Fazekas and Kocsis, 2020). We expect a higher incidence of risk factors in the procurement cycle when the foreign supplier is registered in a tax haven too. Looking at a large Europe-wide dataset, Fazekas and Kocsis (2020) find exactly this relationship (Figure 5).

> > >

FIGURE 5 - Non-Domestic Suppliers' Tax Haven Registration (based on FSI score) and the Incidence of Selected Procurement Cycle Red Flags in the European Union (including the UK), 2009-2014

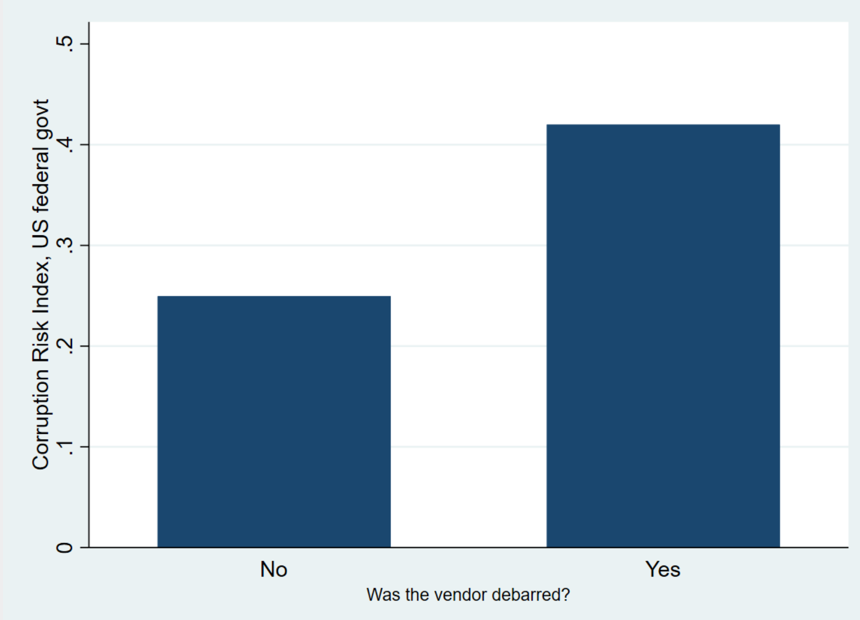


Source: adapted from Fazekas and Kocsis (2020)

Sanctions also represent a strong signal of potential wrongdoing, even though the incidence of such red flags tends to be rare (typically a few thousand flagged cases out of millions). Sanctions are highly correlated with more frequently observed red flags in the procurement cycle. For example, the incidence of red flags of the procurement cycle is about twice as high for debarred suppliers in the United States federal procurement as is for suppliers that have not been debarred (Figure 6).

> > >

FIGURE 6 - US Federal Suppliers' Debarment Status and the Incidence of Selected Procurement Cycle Red Flags, 2004-2015



Source: adapted from Fazekas et al (2022)

Risk Group 4: Political Connections

The risk group of Political connections comprises indicators which capture the relational aspects of corruption, some point directly at conflict of interest while others represent organization-level relationships such as a company donating to a political party. Corruption in public procurement, due to its very nature, involves informal coordination between a range of public, i.e. politicians and bureaucrats, and private actors (Fazekas et al, 2018). Political connections can be demonstrated in a number of ways, through political finance such as campaign donations or personal connections such as family ties. Hence, we organize the 10 political connections risk indicators according to 3 risk patterns (Table 4), recognizing the division between politicians and civil servants: political finance, personal connections to politicians, and personal connections to civil servants.

Risk indicators in this risk group are initially assessed at the level of relations which are then traced back to specific suppliers, for example identifying a former politician employed by a government supplier. Given the relational nature of political connections risk indicators, data requirements are among the most demanding in GRAS: each individual relationship requires linking at least 3 different dataset categories, connecting suppliers (from public procurement data), their shareholders or employees (from company or employment data) and public actors or political organizations (from electoral data, public payroll information or asset and interest declarations). As each risk indicator covers a few variants of possible relationships, including more complex, indirect links, GRAS requires data from four or five different data sources to screen for all these variants for each indicator.[9] These complex relationships are further detailed in Table 5.

While most indicators in this group are binary, recording the existence or absence of a link, in some cases indicators derive from continuous variables (e.g. total value of donations of a company in a period) which require appropriate thresholds. Small donations are unlikely to be suitable red flags for corruption. For example, Fazekas et al (2022) found in the US federal contracting market that only donations above about 11,000 USD have a discernible effect on companies' contracting risks with tendering risks substantially increasing as donation value increases. The appropriate threshold for high-risk donations depends on the country and period which the adoption and tailoring of GRAS should take into account.

9. Specific data field and dataset requirements for each individual indicator are specified in Appendix II.

> > >

TABLE 4 - Individual Risk Indicators in the Political Connections Risk Group (higher values indicate higher risk)

Risk pattern	*Red flag/ indicator*	*Description*
4.1. Political finance	4.1.1. Donation to electoral campaign	Company/shareholder/employee donated/supplied to a politician/political party in a period: Yes/no (variants: direct/indirect link & link to non-elected/elected/in power politician/party)
	4.1.2. Value of donation to electoral campaign	Value of donations made by the company to a politician/political party in a period (variants: direct/indirect link & link to non-elected/elected/in power politician/party)
	4.1.3. Contracts won following donation	Number/total value of contracts won from public bodies (e.g. municipality, region, central government body) with politicians who received donations/supplies from the company in period (variants: direct/indirect link & link to non-elected/elected/in power politician/party)
	4.1.4. Percent of contracts won following donation	Percent of contracts won from public bodies (e.g. municipality, region, central government body) with politicians who received donations/supplies from the company in period (variants: direct/indirect link & link to non-elected/elected/in power politician/party)
4.2. Personal connections to politicians	4.2.1. Company's personal connections to politicians	Company has or had a personal connection to a politician/ political party functionaire in a period: Yes/no (variants: direct/ indirect link & link to non-elected/elected/in power politician/ party)
	4.2.2. Contracts won following political connection	Number/total value of contracts won from public bodies (e.g. municipality, region, central government body) with politicians who are/were linked to the company in period (variants: direct/ indirect link & link to non-elected/elected/in power politician/ party)
	4.2.3. Percent of contracts won following political connection	Percent of contracts won from public bodies (e.g. municipality, region, central government body) with politicians who are/ were linked to the company in period (variants: direct/indirect link & link to non-elected/elected/in power politician/party)
4.3. Personal connections to bureaucrats	4.3.1. Company's personal connections to bureaucrat	Company has or had a personal connection to a public bureaucrat in a period: Yes/no (variants: direct/indirect link)
	4.3.2. Contracts won following connection to bureaucrat	Number/total value of contracts won from public bodies (e.g. municipality, region, central government body) with public bureaucrat who are/were linked to the company in period (variants: direct/indirect link)
	4.3.3. Percent of contracts won following connection to bureaucrat	Percent of contracts won from public bodies (e.g. municipality, region, central government body) with public bureaucrat who are/were linked to the company in period (variants: direct/indirect link)

> > >

TABLE 5 - Possible Sub-Dimensions of Political Connection Indicators

Connection domain	*Nature of connection*	*Type of connection*
Political finance	Donor/Supplier	**Direct:** through the company itself or one of its shareholders or other individuals directly linked to it (e.g. legal representatives, employees, managers, accountants) **Indirect:** through a connected company (common shareholder or other individual as listed above), a business associate (shareholder's partner in another company), a relative etc.
Politician/party official	personal	
Civil Service	personal	

Company donations to electoral campaigns have received extensive scholarly and policy interest (OECD, 2017, chapter 1). When a prospective government supplier donates to a political campaign, it may intend to support the candidate who, upon winning elections, can pay back the favor through government contracts. Such a pattern has been identified in high as well as low integrity countries as diverse as Sweden, the US or Brazil, though the scale of impact varies: in Brazil, 100 USD party donations leads to an additional 1400 USD worth of contracts, while the same number is “only” 250 USD worth of contracts in the US (Boas et al, 2014; Bromberg 2014; Hyytinen et al, 2018).

Among political connections risk indicators, one of the most widely studied and probably most relevant, is the employment of top politicians by companies to gain government favors (Goldman et al, 2013). Former politicians can open doors for a future supplier, share insider information or facilitate bribery in return for contracts. Studies have found that suppliers’ connections to political decision-makers increases their procurement revenue. In Serbia politically connected suppliers have about 30 percent higher single-bidding rate compared with politically unconnected suppliers, a pattern largely reproduced across the Balkans region (Mineva et al, 2023).

Datasets Underpinning GRAS Risk Indicators

The comprehensive and refined risk indicator framework of GRAS requires a range of high-quality and granular datasets linked to each other. These datasets have been identified based on the practical demands of indicator calculation across a range of countries and effective operation of GRAS based on the Brazilian experience. Broadly speaking, GRAS demands two types of datasets, the public contracting datasets and micro-level data on firms and individuals. Table 6 summarizes the types of data required for GRAS, with a general assessment of how essential each of these are for the indicator framework.

Public expenditure databases contain details of all or most phases of the public procurement cycle, at the contract or purchase levels. These are essential, first and foremost, for all risk indicators in the procurement cycle group. Moreover, they are also used for all the other risk patterns assessed by the system. The data contained in these databases generally includes: supplier name and unique ID, requesting agency name and location, procurement method, contract value and date, proposals details, winning proposal, and contract amendments details.

The second block of datasets includes information on companies or individuals. Data on companies is found in company registries and company financial records. These are similarly essential and required for many risk indicators. Employment relationship data are also among the most relevant for GRAS and typically include employer name and ID, employer location, employee name and ID, employee admission date, employee position and remuneration, employers' number of employees.

Individual-level information is found in a variety of different databases. Government social benefit programs typically include information on the name and ID of beneficiaries, type of benefit (e.g. conditional cash transfer), benefit duration, benefit value, and beneficiary location. This information can be used to identify risky company officials (strawmen). Electoral datasets can be used to establish connections between firms and politicians, revealing direct and indirect links associated with potential conflict of interest or favoritism. Asset and interest declarations can be used to identify conflicts of interest and political connections. Electoral records contain information on candidate name, ID and party, election results, campaign suppliers' name and ID, campaign expenses details, campaign donor name and ID, campaign donation value. Criminal records databases typically include the names and IDs of those convicted, the criminal offense, date and location of criminal complaint.

> > >

TABLE 6 - Key Datasets for GRAS and Respective Relevance Level

	Dataset category	*Number of related red flags*	*Relevance level*
1	Public procurement	60	Essential
2	Employment relationships	17	Essential
3	Corporate and shareholder data	39	Essential
4	Electoral data	7	Important
5	Blacklists	6	Important
6	Asset and interest declarations	6	Important
7	Socio-economic data	3	Useful
8	Criminal records	1	Useful

GRAS in Practice: Implementation and Results in Brazil

The World Bank implemented GRAS as a pilot initiative in the states of Mato Grosso and Rio de Janeiro, and the Municipalities of São Paulo and Porto Alegre in late 2022.[10] Brazil has made substantial efforts to increase government transparency through the proactive disclosure of public information over the last decade and so offers a favorable data environment for the implementation of GRAS (OECD, 2022).

The Brazilian GRAS benefits from the availability of numerous public datasets including transactional data on more than 2 million contracts executed in 10 Brazilian states and by the federal government, accounting for over US$ 50 billion in expenditures (Velasco et al. 2020). This detailed public procurement data can be combined with publicly available micro-level data on companies and individuals, drawn from business registration data and datasets on political campaign donations and expenses, sanctioned suppliers and individuals, and conditional cash transfer beneficiaries, among others. GRAS currently operates a large data lake with over 250 million data points. Its data mining algorithms can automatically identify dozens of risk patterns related to public procurement fraud and corruption, at the level of public suppliers, contracting agencies and even individuals. This offers oversight bodies a powerful tool to identify high-risk entities in the public procurement market and allows them to better target their investigations.

The red flags processed in the GRAS system correspond largely to those described in Chapter 2, with a few exceptions that have not been incorporated into GRAS as described in this section. The framework described in Chapter 2 is broader than the Brazilian GRAS to incorporate widely used and validated indicators from other risk assessment tools, such as the European opentender.eu[11] or the global procurementintegrity.org (Box 1), or based on relevant academic literature.[12]

10. This was possible thanks to the resources provided by the Spanish Fund for Latin America (SFLAC); the governments implemented GRAS starting January 1st, 2023.
11. The indicators related to the tender advertisement period or to suppliers' characteristics are employed on the platform opentender.eu, covering close to 50 million contracts in 33 European jurisdictions.
12. This is the case of a few novel collusion indicators that have been developed and presented in recent literature on cartel detection.

3.1. Data Sources Used by GRAS in Brazil

The system as implemented in Brazil relies on seven out of the eight dataset categories listed in table 6 above. All sources used are publicly available, with the exception of data on employment relationships in the private sector. The only key dataset category that is not part of the Brazilian GRAS is asset and interest declarations, which are not available in this context, but can be an important source of relevant information for GRAS implementation in other countries where this data can be accessed. Table 7 below lists the datasets employed in Brazil.

> > >

TABLE 7 - Data Sources Employed by GRAS in Brazil

Dataset category	*Dataset type*	*Dataset name*	*Responsible agency*	*Scope*
1. Public procurement	1.1. Public contracting datasets	Licitações e contratos	State Courts of Accounts	Selected Brazilian States
2. Employment relationships	2.1. Employment registration	Relação Anual de Informações Sociais (RAIS)[13]	Ministério do Trabalho e Emprego	National
	2.2. Public servants	Servidores públicos	(different agencies at all government levels)[14]	National, sub-national
3. Corporate and shareholder data	3.1. Fiscal registration data	Relação de Instituições Financeiras em funcionamento	Receita Federal do Brasil (RFB)	National
	3.2. Financial sector companies	Relação de Instituições Financeiras em funcionamento	Banco Central do Brasil	National
4. Electoral data	4.1. Candidates profiles	Candidatos	Tribunal Superior Eleitoral (TSE)	National
	4.2. Campaign finance data	Prestação de contas de campanha	Tribunal Superior Eleitoral (TSE)	National
	4.3.Party finance data	Prestação de contas partidárias	Tribunal Superior Eleitoral (TSE)	National
5. Blacklists	5.1. Blacklisted companies	Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS)	Controladoria Geral da União (CGU)	National
	5.2. Sanctioned companies	Cadastro Nacional de Empresas Punidas (CNEP)	Controladoria Geral da União (CGU)	National
	5.3. Sanctioned non-profits	Cadastro Nacional de Entidades Privadas sem Fins Lucrativos Impedidas (CEPIM)	Controladoria Geral da União (CGU)	National
	5.4. Blacklisted employers - slave work	Cadastro de Empregadores que tenham submetido trabalhadores a condições análogas à de escravo	Ministério do Trabalho e Emprego	National

CONTINUED

13. Dataset protected by law for data concerning individual person identity and salaries, but released under specific non-disclosure agreement. An anonymised dataset is publicly available.
14. Decentralized data collected through multiple web crawlers from the different agency websites where these are published.

Dataset category	Dataset type	Dataset name	Responsible agency	Scope
6. Socio- economic data	6.1. Conditional Cash Transfer beneficiaries	Benefícios ao cidadão	Ministério do Desenvolvimento e Assistência Social, Família e Combate à Fome (MDS)	National
7. Criminal records	7.1. Arrest warrants	Banco Nacional de Mandados de Prisão (BNMP)	Conselho Nacional de Justiça (CNJ)	National

Data from the different sources were unified into a single database, allowing the implementation of algorithms for the identification of each individual red flag and thereby increasing the efficiency of the system. Automatic routines were developed to clean, transform and validate the data, ensuring sufficient quality for the system to run (Velasco et al. 2020). With regards to its technical architecture, GRAS is a highly versatile and flexible system that can operate independently of any specific hardware or software configurations. A more detailed description of its structure is offered in Appendix III.

GRAS system offers a comprehensive and user-friendly platform for generating audit reports and searching for public agencies, politicians, or firms. It also provides intuitive selection filters to identify potential red flags in companies or procurement procedures. Users can select any combination of red flags as well as actual values and ranges for the specific red flags.

3.2. Illustrative Examples and Purposes of the System

A system such as GRAS enables a range of different applications and extensions, depending on the dataset categories available and the quality and comprehensiveness of the data collected. In Brazil, GRAS was developed in direct collaboration with law enforcement and oversight offices in selected Brazilian subnational governments, with a strong focus on its potential for investigations in connection with procurement fraud, corruption and collusion. Other uses and purposes were explored as well. This section offers a few examples of these uses, illustrating some of the system's contributions to investigative initiatives.

Example 1. Identifying multiple red flags for a selected supplier

GRAS relies on a wide range of risk patterns and associated red flags to identify instances of potential fraud, corruption and collusion in public procurement. Each red flag functions only as an indication of potential irregularity, and the level of risk rises as individual suppliers are linked to a higher number of different red flags. Fraud cases often present multiple suspicious aspects that can be well captured by such red-flag-based systems, in particular when they are designed to identify red flags from different perspectives, as is the case with GRAS. Consequently, the system is particularly useful to identify "high risk suppliers" that are more likely involved in fraud, corruption and collusion and so may be prioritized for investigation.

This feature of GRAS can be demonstrated by navigating a concrete case. An audit team using GRAS in Rio de Janeiro received a tip on a specific company from one of the largest municipalities in the state. The company had won 11 contracts with a total value of more than R$ 16 million (US$ 3.2 million). GRAS was used to dig deeper into the company, and a variety of risk patterns indicating potential fraud were identified on the GRAS interface using a selection of specific filters.

Risk pattern #1 - Broad scope of activities. Corporate registration data indicated that the company in question had as main economic activity kitchen and catering services for transport companies and corporate offices. However, its registered secondary activities were unusually diverse: trade in parts and components of motor vehicles; wholesale trade of computers and computers supplies; hydraulic, ventilation, and cooling systems; car rental without driver; wholesale trade of instruments for medical and surgical use; maintenance and repair of vessels; wholesale trade in chemicals and petrochemicals; and cleaning services. Such a broad and unconnected range of activities - when not related to well-known very large conglomerates - is common for companies involved in fraudulent or collusive practices.

Risk pattern #2 - Shareholder with low socio-economic status. Corporate registration data also indicated that the company had two shareholders, one of whom was a beneficiary of the cash transfer program Bolsa Família, which targets families with low income. This is a risk pattern indicating a potential strawman, someone used to hide the identity of the company's real owner(s).

Risk pattern #3 - Personal connections to bureaucrats. GRAS identified the company's second shareholder as a civil servant in a federal public hospital in the city of Rio de Janeiro, employed as a social worker. Most of the company's contracts were related to food supply for hospitals.

Risk pattern #4 - Political finance. GRAS indicated a key political connection risk, because one of the company's shareholders, listed as a Bolsa Família beneficiary, had been a supplier to the campaign of an influential local councilor who had recently run for mayor.

> > >

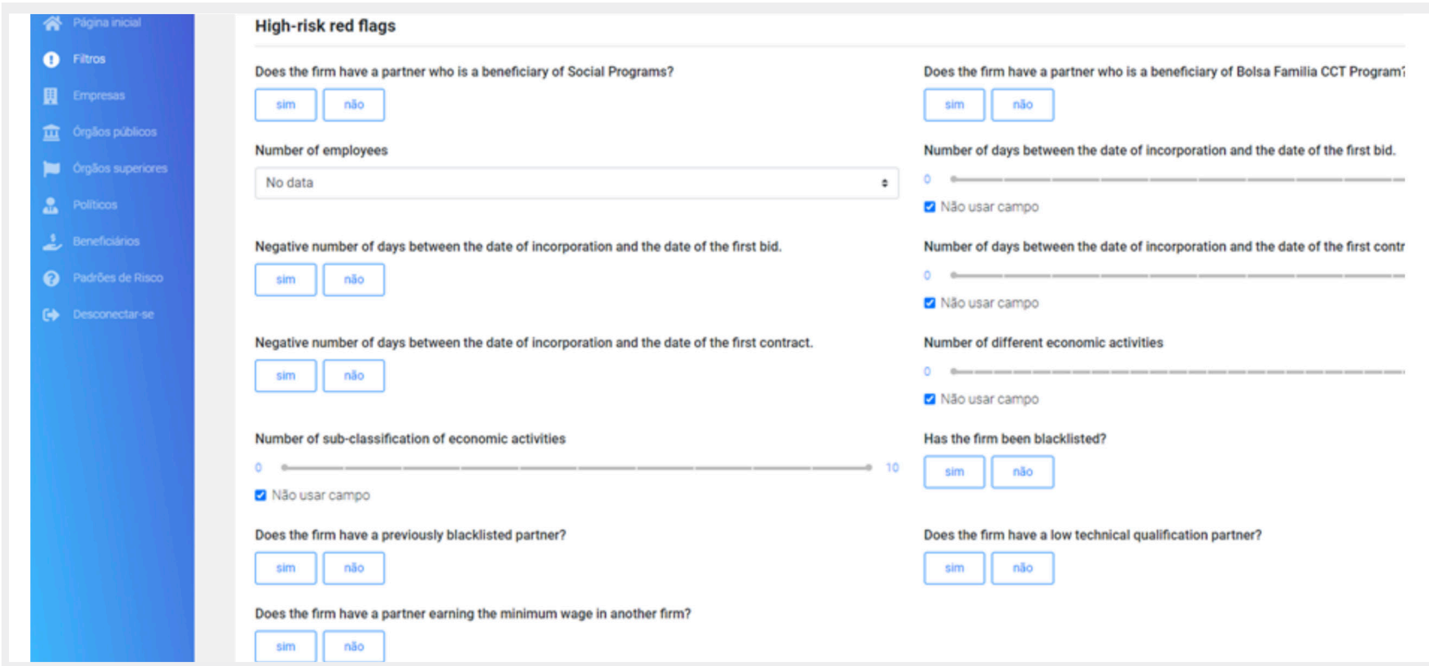
FIGURE 7 - GRAS Navigation through Multiple Red Flags Connected to a Single Supplier

In this example in particular, GRAS was used as a complement to a tip received by the audit team, but other investigations can similarly be initiated following initial suspicions of fraud and corruption raised from the system's reports. Through filtering functions and different aggregation options, GRAS offers users sufficient flexibility to define sets of criteria based on which higher risk actors can be identified and narrowed down to be targeted by specific investigations (see Figure 8 below).

Auditors and investigators can, for instance, rely on their qualitative expertise about risk patterns that are commonly observed in combination in known cases in their jurisdictions, and select groups of risk patterns as filters to identify potential similar cases. GRAS can also help enforcement agencies identify multiple risk patterns based on which different potential lines of investigation can be pursued for corroborating evidence.

> > >

FIGURE 8 - Examples of GRAS Filtering Options



Example 2. Identifying collusion in electronic reverse auctions

In electronic reverse auctions, suppliers anonymously bid prices down until the auction is complete. This procedure aims to promote a high level of competition. The electronic reverse auction can be gamed through collusive behavior using a so-called "kamikaze" (or "rabbit") company,[15] in coordination with another bidder - the scheme usually involves two colluding partners: One presents an unrealistically low bid to scare off other competitors from further reducing their initial bids. The other colluding partner, which is meant to actually win the contract, then adjusts its initial bid only to the extent necessary to have the second lowest bid. In these auctions, the qualification phase typically takes place after the bidding; the kamikaze partner then purposefully fails to fulfill the qualification requirements, and the colluding partner, after being qualified, is awarded the contract at a higher price than would have been the case under real competition.[16]

A number of GRAS red flags from the Collusion risk group can help to identify collusion using a "kamikaze" company: a) as a typical bid variance bias, where a high relative bid distance between the kamikaze's and winner's bids is observed; b) the colluding partner that is set to win the auction will bid to deviate the least possible from the third lowest bid, which might also result in a high relative contract value; c) connections between the colluding bidders, such as common registration data or a common shareholder or employee, are likely to be present; and d) if the kamikaze company engages repeatedly in such schemes, it might also present the Top Loser risk pattern. Similarly, the winning company may be identified as a typical winner against Top Losers, a risk pattern also made visible by the system (Figure 9). This feature assists users in more easily identifying potential collusive links between providers.

15. https://3rcapacita.com.br/artigo/licitante-coelho
16. Changes in the regulation of electronic reverse auctions in Brazil were introduced in 2019 in an attempt to address these situations, by requiring all qualification documentation to be submitted beforehand (https://www.editoraforum.com.br/noticias/decreto-do-novo-pregao-eletronico-inibira-fraude-conhecida-como-coelho-nas-licitacoes-afirma-especialista/).

> > >

FIGURE 9 - Information on Providers with a Pattern of Winning Against Top Losers

Winner against Top Loser

Pesquisar

Top Loser ID	Top Loser Name	Number of victories against Top Loser	Frequency of victories against Top Loser	Total contract value won against Top Loser
.455.*/****-**	LI** ******** ****** ****	1	1	R$ 4.548.000,00
.978.*/****-**	TA**** ******** * ******	1	1	R$ 4.548.000,00

Example 3. Preventive vetting of bidders

GRAS relies largely on data from past contracting procedures and can be a useful tool for identifying contracting irregularities ex-post. A further application for preventive purposes is also possible. GRAS can be used as a bidder "vetting tool", allowing purchasing authorities to assess bidder profiles prior to awarding a contract. This review could offer a quick and detailed overview of bidders' contracting history such as amount and types of contracts awarded in the past; and review of risk patterns related to suppliers' characteristics as well as indicators from the procurement cycle. Hence, GRAS can also serve as a quick and effective resource for due diligence during the tendering phase based on reliable, third-party data.[17]

3.3. Impact and Results Achieved

An assessment of preliminary results achieved using GRAS during pilot implementation reveals that the system has been effective in detecting a large number of procurement tenders and contracts displaying risk patterns. Table 8 below illustrates some of the results for selected red flags that the system reported in 2020.

17. For another, company vetting tool, supporting due diligence, but based on globally available public procurement data, see: https://tenderx.eu/

> > >

TABLE 8 - Examples of Cases Identified by GRAS for Selected Risk Patterns

Risk group	*Risk pattern*	*Identified cases*
1. Procurement cycle	1.1. Non-competitive processes	2308 companies that received contracts through direct awards
2. Collusion	2.1. Top loser	420 companies that won bids against top losers
	2.7. Common shareholder	857 companies that won bids against companies sharing a common shareholder
3. Supplier characteristics	3.4. Young supplier	Almost 150 companies that were awarded a contract within 120 days after their incorporation
	3.6. Sanctions	Approximately 800 sanctioned companies that were awarded contracts
	3.7. Shareholder with low socio-economic status	450 companies with shareholders registered either as beneficiaries of cash transfer programs or as low-skilled employees
4. Political connections	4.1. Political finance	Almost 4,500 public suppliers that are either electoral campaign donors (as companies or through their shareholders) or suppliers
	4.3. Personal connections to bureaucrats	500 firms owned by public servants that received contracts from the same agency where the shareholders were employed

These exemplify some typical risk patterns linked to procurement fraud, corruption and collusion in Brazil. Non-competitive procedures, for instance, are allowed under certain circumstances, but are sometimes unduly employed in combination with multiple bidding procedures for similar or almost identical objects, in order to remain under contract value thresholds above which competitive procedures would be legally required (Santos and Souza 2016).

Collusive rings working to simulate competition in tenders are a common feature in public procurement. GRAS flagged a number of risk patterns that may indicate collusion. Interestingly, links between shareholders are not necessarily well hidden, as shown by the hundreds of cases where the system identified “competing” bidders with common owners. By analyzing information from multiple tenders, GRAS detected bidding patterns that indicate potentially colluding actors that are not real competitors, i.e. the so-called top losers,[18] as well as the companies that may have benefitted from collusive action winning contracts against such top losers. GRAS identified hundreds of suppliers associated with these risk patterns in a small sample of Brazilian states.

GRAS also employs multiple indicators related to risky supplier characteristics. Companies that obtain contracts shortly after being registered were not uncommon in Brazil. Recently created suppliers may be inexperienced, increasing risks for proper contract implementation. They may have been created as shell companies to be used by corrupt networks.

Companies that have been formally sanctioned and suspended from contracting may present increased risks. GRAS identified some 800 firms that won new contracts while under sanctions. Procurement regulations require public agencies to check databases of sanctioned firms before contracting, but this provision is clearly not preventing sanctioned companies from getting new business in the public sector while blacklisted. In one of the states covered in the GRAS pilot, almost 7 percent of all contracts were awarded to sanctioned companies.

Strawmen are used to register companies that bid for and occasionally win procurement contracts. Use of strawmen to hide the company’s true beneficiary is linked to irregular activity which warrants further investigation. GRAS flags these cases by profiling the socio-economic status of companies’ shareholders through complementary datasets, indicating when they are identified as beneficiaries of cash transfer programs or registered as low-skilled or low-pay employees, i.e. with a socio-economic status that is atypical for real business owners.

18. For another, company vetting tool, supporting due diligence, but based on globally available public procurement data, see: https://tenderx.eu/

The Brazil pilot identified thousands of suppliers (or their shareholders) registered as donors or suppliers to electoral campaigns or political parties. These companies have been awarded in total close to R$ 100 billion (US$ 2 billion) in public contracts.[19] Indeed, in all of the states covered in the pilot projects a large share of contracts was awarded to companies with connections to politicians, in some cases more than half of all contracts. An even stronger indication of potential conflict of interest are cases of suppliers with politicians or public servants as shareholders, in particular when their contracts were obtained from the same agency where those shareholders occupy office or are employed. In a single state, GRAS identified 122 companies with politicians as shareholders (Velasco et al. 2020). In all the jurisdictions covered by the pilot, GRAS identified 500 companies owned by public servants that received contracts with agencies where they worked.

The use of GRAS by Public Prosecutor's offices in the pilot states has already actively contributed with input to relevant corruption investigations. In one of these states, two investigations conducted by the Federal Police were informed by targeted analyses conducted with GRAS. GRAS identified risk patterns that led to the uncovering of farms of shell companies and a money laundering chain. GRAS used a specific algorithm to identify ghost public workers in the public payroll in one municipality used by locally-elected officials to divert public funds (Velasco et al. 2020). This illustrates that GRAS can yield concrete results in a short period of time after implementation. In the Brazilian context, its application at the sub-national level has been of great relevance to strengthen anti-corruption action where oversight and law enforcement agencies are more under-resourced.

19. Corporate donations were allowed in Brazil until 2015, when a ban was introduced and only private donations from individuals remained legal (https://www.idea.int/data-tools/country-view/68/55). As GRAS includes data from earlier periods, corporate donations can still be identified in some cases, and individual donations by company owners remain substantial.

A Roadmap for GRAS Implementation

Building on the positive results of the GRAS pilot in Brazil, the World Bank seeks to promote implementation of the system in other countries and jurisdictions. This section reviews relevant steps for a feasibility assessment applicable to any context, which, by and large, include:

- Collaboration with governmental agencies;
- Data maturity assessment;
- Validation and adaptation of the red flag framework; and
- Feasibility assessment and recommendation of improvements of the data infrastructure.

Following this initial assessment, a detailed implementation plan for a pilot stage must be developed together with collaborating government agencies. For this step, there is no standardized model, as concrete steps for implementation in a given jurisdiction will be highly context-specific, with a broad variation in scope of implementation depending on aspects such as: data availability, data access and data protection issues; extent of support within government and at different government levels; existence of complementary data analytics initiatives; the need for tailoring of the framework to the specific context and data environment, among other relevant factors to be taken into account.

4.1. Collaboration with Governments

Engagement with the government is an essential first step in GRAS implementation. Government agencies, in particular those responsible for anti-corruption, procurement, law enforcement and oversight, as well as public finances, are the obvious clients for a governance risk assessment tool. GRAS relies on public data collected and managed by governments, their buy-in is essential in securing data access and for efforts to improve data disclosure as part of a transparency agenda. Collaboration with central governments is a preferred strategy, in the sense that an investment in GRAS nationally would likely set a risk-assessment infrastructure that can be either directly used by or at least more easily extended to sub-national levels. Nevertheless, as the Brazilian experience demonstrates, GRAS can also bring great added-value if initially implemented in sub-national jurisdictions.

In feasibility assessment exercises undertaken by the World Bank in Latin America, collaboration with oversight and control authorities, as well as procurement agencies, has been instrumental in mapping relevant data sources for the key dataset categories required for full GRAS implementation. In some cases, detailed data dictionaries for public and non-public datasets were provided as input for the data maturity assessment, which added to the robustness of the analysis. Indeed, it is important to highlight that, even though the Brazilian GRAS pilot benefitted from open, fully public and machine-readable datasets, this is not necessarily required for GRAS. Where open data is less advanced or data protection regulations preclude individual-level data from being published, for instance, agreements with the responsible agencies to access data from internal, non-public government datasets can deliver the data infrastructure needed for GRAS.

4.2. Data Maturity Assessment

The data maturity assessment provides a detailed overview of the data infrastructure for GRAS implementation. Table 9 below lists the dataset categories required for each of the 23 risk patterns that integrate the GRAS framework.

> > >

TABLE 9 - Dataset Categories Required Per Risk Pattern

Risk group	*Risk pattern*	*Public procurement*	*Employment relationships*	*Corporate & shareholder data*	*Electoral data*	*Blacklists*	*Socio-economic data*	*Criminal records*	*Integrity declarations*
1. Procurement cycle	1.1. Non-competitive processes	X							
	1.2. Non-competitive tender results	X							
	1.3. Contract implementation biases	X							
2. Collusion/ Bid-rigging	2.1. Top loser	X							
	2.2. Fixed difference bids	X							
	2.3. Bid variance biases	X							
	2.4. Unusual contract value	X							
	2.5. High price	X							
	2.6. Require Public Procurement	X		X					
	2.7. Corporate Data	X		X					
	2.8. Require PP, Employment and Corporate Data	X		X					

CONTINUED

Risk group	*Risk pattern*	*Public procurement*	*Employment relationships*	*Corporate & shareholder data*	*Electoral data*	*Blacklists*	*Socio-economic data*	*Criminal records*	*Integrity declarations*
3. Supplier characteristics	3.1. Unusual size	X	X	X					
	3.2. Unusual profitability	X		X					
	3.3. Broad scope of activities	X		X					
	3.4. Young supplier	X		X					
	3.5. Non-registered supplier	X		X					
	3.6. Sanctions	X		X		X			
	3.7. Shareholder with low socio-economic status	X	X	X			X		
	3.8. Shareholder/ legal representative with criminal record	X		X				X	
	3.9. Tax haven registration	X		X					
4. Political connections	4.1. Political finance	X	X	X	X				
	4.2. Personal connections to politicians	X	X	X	X				X
	4.3. Personal connections to bureaucrats	X	X	X					X

The data maturity assessment entails an initial mapping of relevant data sources for each of the 8 dataset categories required in the GRAS framework, comprehensive identification of individual data fields in each dataset, and an assessment of data quality, accuracy and completeness in each of the essential fields for the calculation of GRAS specific risk indicators. The assessment entails four steps (see table 10 below).

> > >

TABLE 10 - Country Data Maturity Assessment Process

Assessment steps	*Objectives*	*Relevant questions*
A. Consultation with government agencies	Initial mapping of relevant data sources and collection of basic information on each source/dataset	- Does the information required[20] exist as a structured, disaggregated dataset? - Is the dataset publicly accessible? Is it available as open data (i.e. in a machine-readable format)? In which formats is it available? Where can it be accessed? How often is it updated? - What is the legal framework that establishes its publicity? - What agency is responsible for collecting/managing the dataset? - What is the jurisdiction coverage (national/ subnational)? What is the time coverage? - Is there a published data dictionary? Where can it be accessed? If not, can one be provided for the assessment?

CONTINUED

20. A list of necessary data fields for GRAS should be provided as a reference. This is available in Appendix I.

Assessment steps	Objectives	Relevant questions
B. Review of available data sources through desk research	Complementing information provided in the consultation (filling potential gaps) and identifying additional/alternative sources	- Are there other relevant sources for the respective dataset category? (in which case the same basic information obtained from governments in the previous step is collected during the research) - What is the level of observation in the dataset? What is its scope (e.g. types of entities/individuals covered)? - Are there any apparent issues with this data source that can be already identified at this stage (e.g. lack of proper unique identifying data, limitations in scope)?
C. Detailed variable-level mapping of identified sources/datasets	Verifying whether identified datasets contain the data fields required for GRAS indicators and assessing the feasibility of each individual indicator	- Are essential data fields covered by the dataset? If not, are there other relevant fields that could function as proxies? - Which individual GRAS indicators can be fully/partially implemented? - Is additional information required to complete the assessment?
D. Individual dataset assessment	Assessing the accuracy, comprehensiveness and completeness of the data available and identifying relevant gaps	- Are key data fields correctly filled, complying with data format requirements? - Is missing data a widespread problem? - Are recorded values in the datasets consistent with actual transactions and legally binding documents (e.g. contract values in the dataset correspond to actual contract values in signed contracts)? - Are there important data scope limitations such as a class of procurement transactions excluded from government registers (e.g. tenders of state owned enterprises)?

First, the assessment starts with a consultation with the relevant government agencies using a brief questionnaire, where they should indicate the main data sources in each category and provide basic information about each source/ dataset, as covered by the relevant questions listed in the table above. Specific data dictionaries can be provided by the collaborating agencies, in particular those that are not publicly available.[21] This is an extremely valuable input for the subsequent steps in the assessment.

Second, the information provided is reviewed and validated through desk research. Complementary information is collected, and potential additional or alternative sources are identified. Information on the newly identified sources is documented following the same questions that oriented the previous stage. After this initial mapping of the GRAS data sources, a broad picture emerges in terms of relevant data environment characteristics, potential gaps at the dataset level and access limitations.

Third, each of the datasets identified is then examined to confirm the availability of the minimum set of data fields necessary to produce the 60 indicators specified under the GRAS framework. A reference list of essential data fields is provided in Appendix I. For publicly accessible datasets, this step can rely on the analysis of corresponding data dictionaries where available, or the dataset itself in some cases, and, when applicable, also on the verification of government public search platforms for the relevant sources, with example searches to illustrate which data is retrieved.[22] The analysis is somewhat more challenging for the datasets that are not publicly accessible and for which no data dictionary can be obtained. In those cases, datasets are in most cases populated by information provided through registration procedures, i.e. channels by which government agencies collect information from individuals or organizations, mainly through public digital services, a useful strategy can be to conduct the assessment based on documentation on those data collection processes. For instance, registration forms

21. In a GRAS feasibility assessment done in Peru, for instance, government agencies collaborating with the researchers conducting the study provided several data dictionaries, even for data that were not publicly accessible. This provided for a more robust and reliable detailed assessment of each dataset later on.

22. In feasibility assessments conducted in Latin American countries, it was observed that some datasets were made available as open and downloadable files, but those versions included only partial data when contrasted to the information obtained in example searches on the corresponding public search platforms. Therefore, it is an important step to check if open data files indeed cover the full underlying datasets.

may be accessed to directly observe the type of information that is collected. Also official tutorial material (e.g. manuals, videos) published by the respective agencies can be used as a reference to identify individual variables contained in those datasets. Once the specific data fields have been mapped, the feasibility of individual risk patterns and indicators can be assessed for a preliminary overview of the potential scope of GRAS implementation.[23]

Fourth, an in-depth evaluation of individual datasets is important to assess whether requirements in terms of data accuracy and completeness are fulfilled for GRAS to operate as designed (for in-depth data assessments with examples see: Horn et al, 2021, chapter 2; Cingolani et al, 2016; Czibik et al, 2015). Even if required data fields exist in the mapped datasets, a detailed examination of their actual content is needed to establish: whether missing rates on key variables are sufficiently low or too high for adequate analysis; whether common identifiers for individuals and organizations such as company registry IDs are present and follow the required format in order to link different datasets; and whether relevant fields have unusual or extreme distributions decreasing their value for risk flagging (e.g. whether a categorical variable takes one of the possible categories in 99% of the cases). It is expected and highly likely that all the administrative datasets will suffer from some or all of these problems. However, some of them can be remedied as part of a GRAS implementation plan, for example filling in missing fields from related fields (e.g. if the buyer city is available but buyer state is missing, drawing on a city-state correspondence table from the statistical office enables reliably filling in blanks).

4.3. Validation and Tailoring of GRAS Framework to National Specificities

Even though the GRAS pilot was conceived to work for the specific data environment of Brazil, the GRAS framework offers a robust set of established risk indicators that are likely to have broader applicability. The Brazil GRAS pilot relies not only on established public procurement corruption risk indicators, but also on risk patterns that relate to companies, political connections and collusive practices. Such broad range of indicators allow for triangulation and flexibly adapting GRAS to new contexts.

GRAS will need to be tailored to the particular context where it will operate, incorporating or prioritizing elements that might be relevant for detecting fraud, corruption and collusion risks in that context. Indicators may have to be adjusted to better fit the type of data available,[24] or even to incorporate specific risks relevant for the new context that might not be adequately covered in the framework currently. This may require the design of more appropriate or additional indicators. Indicator thresholds and risk value ranges will have to be defined to reflect national and local specificities of public procurement regulations, markets and corruption strategies (Fazekas and Kocsis, 2020). Such parametrization and tailoring to context is indispensable for prediction accuracy of GRAS. This can be achieved by using proven positive and negative cases (e.g. machine learning) or exploiting expected correlations among established risk factors (Adam et al, 2022).

Governments may already have data analytics initiatives for governance risk assessments in place. Understanding how these function and what synergies may exist with GRAS is an important step in planning GRAS implementation to avoid duplication of efforts and to best complement the approaches already in use. This complementary implementation of GRAS can entail the extension of existing tools to cover new risk patterns or incorporating new sources of data, or even covering new jurisdictions (e.g. sub-national entities). Pre-existing risk assessment systems are often intended for exclusive internal use by government agencies. Complementary mapping will likely have to be supported by interviews with public servants in oversight functions, analysis of recent audits and related reports already produced by the relevant authorities.

23. Appendix II includes a reference table listing the specific fields required for each individual indicator.
24. One example observed during the preparation of feasibility assessments in selected Latin American countries refers to differences in the socio-economic data available there, in contrast to the type of dataset employed under this category in Brazil. The Brazilian GRAS makes use of individual-level data on actual beneficiaries of a number of focalized cash transfer programs, which indicate exact time periods and amounts received by each beneficiary. This type of data was not found to be easily accessible in the countries examined, but an alternative type of data could be identified that could fulfill the same purpose, namely household classification datasets which indicate specific households, their individual members and a household-level poverty classification. Consequently, the implementation of GRAS indicators originally employing variables from beneficiary datasets would need to be somewhat redefined in order to better relate to how these alternative data sources are structured.

4.4. Feasibility Report and Recommendations

The feasibility assessment concludes with the production of a feasibility draft report directed at the collaborating government agencies (or for public disclosure as well). The assessment includes a context-validated indicator framework for GRAS implementation, based also on the stage of data maturity observed as well as the scope of implementation that it enables. Most importantly, the assessment should discuss specific, detailed data limitations, especially those referring to the most essential dataset categories required for GRAS operation, i.e. public procurement and corporate and shareholder data. It should also refer to other context-specific potential difficulties for GRAS implementation, such as data protection regulations, whenever applicable. This should be accompanied by recommendations for improvement of collection,[25] management and disclosure of datasets. A GRAS feasibility assessment may also be included as part of a broader agenda to promote transparency - in particular in the area of anti-corruption or public procurement and corporate data - and digitization of public services. Report should be validated with input from government agencies involved, as well as from local data, public procurement and corruption experts, helping to fill possible gaps in the analysis.

The feasibility report represents only a first step towards the elaboration of a detailed implementation plan, which must be developed as a second step to address concrete and context-specific considerations, as well as more actionable recommendations for the tailoring of GRAS to the particular context. It could foresee gradual implementation of the system according to the scope of feasibility and a proposal for a pilot stage, indicating possible extensions depending on further developments in the underlying data environment. Indeed, depending on the level of engagement and commitment of the respective government agencies regarding data transparency, positive changes in public data availability may take place quite dynamically, requiring the initial feasibility assessment to be revisited occasionally to observe whether conditions for GRAS implementation have substantially changed.

A GRAS implementation plan should also consider capacity building needs for potential users of the system, depending on the features and the scope of indicators actually implemented. This should thus include the necessary activities to ensure that GRAS users are sufficiently trained on GRAS functionalities and data analytics more broadly.

25. In general, GRAS requires the existence of structured datasets generated by electronic data collection systems. In their absence, implementing important parts of the framework may become partly or fully unfeasible. One such example was observed in feasibility assessments conducted for Panama and Ecuador, where data on political finance was available only through financial reports submitted in paper, which created difficulties for the application of risk patterns associated with political connections. In those contexts, one important recommendation referred to the introduction of an electronic system for the submission of electoral and party financial reports.

Conclusions and Potential for Extensions

This report presents a comprehensive fraud, corruption, and collusion risk assessment methodology and tool, the Governance Risk Assessment System (GRAS), and its implementation in Brazil. GRAS is a data-driven tool which can improve the detection accuracy of fraud, corruption and collusion, thereby increasing the efficiency and effectiveness of audits and investigations. GRAS presents 60 different red flags, linked to 23 broad risk patterns falling into 4 main groups of risks behaviors. GRAS, like other data-driven risk assessment tools, can only lead to tangible results if it is adequately integrated into an effective, broader anti-corruption framework (Fazekas et al, 2019). Investigations and audits can be improved by data-driven tools only if the responsible institutions are well-resourced, meritocratic and independent and if they are able to collaborate bringing even complex cases to courts.

GRAS can be readily implemented in countries with a high degree of data maturity. However, extensive data requirements need not restrict dissemination and implementation of the GRAS in countries with lesser data maturity or a data landscape with different strengths. Based on the current status of public data availability and governance in the world,[26] no country and government is likely to fulfill all the requirements for GRAS. Even in Brazil, with an especially favorable data environment, GRAS implementation is not without its limitations (for specific areas of further development see below). The data environment assessment can provide an initial roadmap for agencies on what steps can be taken to implement GRAS despite data limitations. Efforts to promote GRAS can and should consider a minimum viable version that can be implemented wherever governments offer the necessary policy support. The core functionality of GRAS can then be gradually expanded and refined. For example, GRAS implementation could be coupled with a transparency agenda, profiting from efforts to improve procurement data quality and publication across the globe. One third of GRAS indicators may be implemented with public procurement data alone. Following improvements in the availability of additional datasets and required data fields, the system can be gradually extended. With time, a refined methodology might be needed as well, as fraud, corruption and collusion strategies are likely to evolve in response to enforcement efforts, bringing the need for adaptation and improvement of the framework. The initial adoption and continuous improvements to GRAS should also support the movement towards open government data to support risk assessment, the analysis of government spending efficiency and accountability more broadly.

26. https://globaldatabarometer.org/

Further Enhancing GRAS

While GRAS builds on and encompasses the results of a wide and rich law enforcement practice and academic literature, it can be improved in future iterations. For example, user groups beyond investigators and auditors can be served. Potential improvements that are being considered are discussed below.

1. **Additional Data.** As corruption does not stop at administrative and national borders, more comprehensive data across borders could greatly enhance investigation and audit effectiveness. There are existing platforms, such the ProACT tool (Box 1), which build on a wide range of country datasets and allow for risk assessment across different jurisdictions. Adding further countries to GRAS or allowing it to connect to existing cross-country tools will allow users to identify risk patterns and flag cases more comprehensively.

> > >

BOX 1: PROCUREMENT ANTI-CORRUPTION AND TRANSPARENCY PLATFORM (PROACT)

ProACT is based on open data from national e-procurement systems from 46 countries and open data on World Bank and Inter-American Development Bank financed contracts for over 100 countries. The data has been collected from official government procurement portals and standardized into a single data structure by the Government Transparency Institute.[27] ProACT allows users to search and analyze specific contracts, tenders, buyers, suppliers and markets. It also offers country-level statistics, including competition, transparency and integrity indicators, which can be further disaggregated by sector, procurement method, and contract value range.[28] ProACT has been developed by the World Bank in collaboration with the Government Transparency Institute, building on European experiences with corruption risk portals such as www.opentender.eu.

ProACT is intended for a wide range of users, including procurement officers in national procuring entities; procurement specialists and analysts in MDBs and national Public Procurement Authorities; NGOs that work on procurement, integrity, transparency, and open government; and researchers and academia. ProACT allows procurement officers in national contracting agencies to access information from public procurement records outside their own country. This helps them track firms and analyze international market conditions for specific goods, works and services, and can be a complementary tool to GRAS to gain further insights into a specific provider beyond the national market.

27. For more details and recent updates on this dataset see: https://www.govtransparency.eu/gtis-global-government-contracts-database/
28. For the detailed methodology, see: https://www.procurementintegrity.org/assets/about/ProACT_methods_paper_20220809_final.pdf

2. **Indicator extensions.** Corruption schemes constantly evolve and proxying all widely present corrupt behaviors is essential for a comprehensive and reliable tool. Among areas where potential new indicators can be brought into GRAS, the contract implementation phase is critical. Contact implementation is harder to monitor, less standardized and there are fewer or no competitors who watch over the fairness of the process. While GRAS has indicators on cost overruns and delays, a range of indicators could be deployed targeting payments such as unusual timing and values of payments. The Brazilian GRAS already provides a good example of how this can be explored based on e-invoice data. More challenging, but essential, is to incorporate data on quality and quantity of eventually delivered goods and services. A host of corrupt schemes look impeccable on paper, but the resulting roads are barely usable, the websites crash, or the services rendered are irrelevant. Another extension already implemented in the Brazilian GRAS is additional context-specific risk patterns regarding supplier characteristics: based on fine-grained local socioeconomic data, suppliers registered in very humble locations, especially while having at the same time no registered employees, can be identified as potential ghost companies. This has been found to be a very relevant risk pattern already in one of the pilot States where GRAS is operational, with 88% of the total high-risk contract value affected by this red flag. Such types of indicators tailored to specific contexts and harnessing other available data sources in a given jurisdiction are an important development beyond the initial implementation of GRAS' standard framework.

3. **Indicator design.** The interpretability of indicators can be further improved based on state-of-the-art data science. The current GRAS framework rests on a wide range of indicators that allocate the task of interpretation and parametrization to the user. For example, once a user sees the indicator of contract share through non-competitive procedures (indicator 1.1.1), which is a continuous indicator; he or she has to decide which percentage is cause for concern, 10%, 30%, 80%, etc. Such decisions are hard as they fundamentally depend on the context and prevailing market or sectoral norms. Such parameters and interpretation questions can be addressed by looking at relationships among indicators and contextual variables. For example, when non-competitive procedures are related to overpricing and market-specific procedure type distributions, it is possible to identify which procedure types are non-open and which extent of non-open procedure type use is likely to be risky.[29]

4. **Methodology improvements.** Allowing users to look across a wide range of individual factors lends them a great deal of flexibility, but forgoes opportunities to combine individual indicators into a more accurate measurement. For example, a recent study applying machine learning methods to 78 proven cases of collusion in 7 European countries, achieves 80-90% prediction accuracy by combining 5 indicators of collusion. Many of the accurately identified cases do not score particularly high on individual risk dimensions, only by combining 'weak signals' into a comprehensive predictive model does accuracy increase (Fazekas et al, 2023).

5. **Broadening the pool of users.** GRAS is well-suited to the needs and activities of investigators and auditors in the fraud, corruption, and collusion areas, but there are additional use cases. First, the preventive use of GRAS could support a range of monitoring bodies. For example, public procurement or competition authorities could identify markets and areas of high risk and implement broader preventive interventions such as reviewing procurement policies and guidance documents of the relevant public buyers. Improving buyers' procurement skills and institutions could lower risks across a wide set of contracts. Second, GRAS can support policy assessment and reform by identifying policies which allow for high-risk activities such as too high contract value thresholds for mandatory competitive tendering or auction design prone to collusion.

29. The GRAS team is working to incorporate ChatGPT to support users in identifying key red-flags.

References

Adam, Isabelle, Fazekas, Mihály, Kazmina, Yuliia, Teremy, Zsombor, Tóth, Bence, Villamil, Isabela Rosario; & Wachs, Johannes (2022) Public procurement cartels: A systematic testing of old and new screens. GTI-WP/2022:01, Budapest: Government Transparency Institute.

Alexeeva, Victoria; Padam, Gouthami; Queiroz, Cesar. 2008. Monitoring Road Works Contracts and Unit Costs for Enhanced Governance in Sub-Saharan Africa. Transport paper series; no. TP-21. World Bank, Washington, DC.

Bajpai, Rajni; Myers, C. Bernard. 2020. *Enhancing Government Effectiveness and Transparency : The Fight Against Corruption.* Washington, D.C.: World Bank Group.

Berger, A., & Hill, T. P. (2015). An introduction to Benford’s law. Princeton University Press.

Boas, Taylor C., F. Daniel Hidalgo, and Neal P. Richardson. 2014. The spoils of victory: Campaign donations and government contracts in Brazil. Journal of Politics 76 (2): 415–29.

Bromberg, Daniel. 2014. Can vendors buy influence? The relationship between campaign contributions and government contracts. International Journal of Public Administration 37 (9): 556–67.

Cingolani, Luciana; Fazekas, Mihály; Kukutschka, Roberto Martínez B.; and Tóth, Bence (2016) Towards a comprehensive mapping of information on public procurement tendering and its actors across Europe. Cambridge: University of Cambridge.

Collier, Paul; Martina Kirchberger, and Måns Söderbom, The Cost of Road Infrastructure in Low-and Middle-Income Countries, The World Bank Economic Review, Volume 30, Issue 3, October 2016, Pages 522–548, https://doi.org/10.1093/wber/lhv037

Czibik, Ágnes; Tóth Bence; and Fazekas, Mihály (2015) How to Construct a Public Procurement Database from Administrative Records? With examples from the Hungarian public procurement system of 2009-2012. GTI-R/2015:02, Budapest: Government Transparency Institute.

European Commission (2022) Cohesion in Europe towards 2050. Eighth report on economic, social and territorial cohesion. Luxembourg: Publications Office of the European Union.

Fazekas, M. & Márk, L. (2017). Objective corruption risk indicators using donor project and contract data. GTI-R/2017:02, Budapest, Government Transparency Institute, September 2017.

Fazekas M, Tóth IJ and King LP (2016) An Objective Corruption Risk Index Using Public Procurement Data. European Journal on Criminal Policy and Research 22(3): 369–397. DOI: 10.1007/s10610-016-9308-z.

Fazekas, Mihály and Tóth, Bence, (2016), Assessing the potential for detecting collusion in Swedish public procurement. Uppdragsforskningsrap. 2016:3, Swedish Competition Authority, Stockholm.

Fazekas, Mihály, Luciana Cingolani, & Bence Tóth (2018), Innovations in Objectively Measuring Corruption in Public Procurement. In Helmut K. Anheier, Matthias Haber, and Mark A. Kayser (eds.) Governance Indicators. Approaches, Progress, Promise. Ch. 7. Oxford University Press, Oxford.

Fazekas, Mihály, Ugale, Gavin, and Zhao, Angelina, (2019) Analytics for Integrity. Data-Driven Approaches for Enhancing Corruption and Fraud Risk Assessments. OECD, Paris.

Fazekas, Mihály, and Kocsis, Gábor, (2020), Uncovering High-Level Corruption: Cross-National Corruption Proxies Using Public Procurement Data. British Journal of Political Science, 50(1).

Fazekas, Mihály, and Blum, Jurgen Rene. (2021), Improving Public Procurement Outcomes: Review of Tools and the State of the Evidence Base. Policy Research Working Paper;No. 9690. World Bank, Washington, DC.

Fazekas, Mihály; Ferrali, Romain & Wachs, Johannes (2022) Agency independence, campaign contributions, and favouritism in US federal government contracting, Journal of Public Administration Research and Theory, available online.

Fazekas, M., Tóth, B. and Wachs, J. (2023). Public procurement cartels: A large-sample testing of screens using machine learning. GTI-WP/2023:02, Budapest: Government Transparency Institute.

Goldman, Eitan, Jörg Rocholl, and Jongil So. 2013. Politically connected boards of directors and the allocation of procurement contracts. Review of Finance 17 (5): 1617–48.

Horn, Peter; Czibik, Ágnes; Fazekas, Mihály; and Tóth, B. (2021): Analyzing Public Procurement Risks: Training manual. Budapest: R2G4P / Government Transparency Institute.

Hyytinen, A., Lundberg, S. and Toivanen, O. (2018), Design of public procurement auctions: evidence from cleaning contracts. The RAND Journal of Economics, 49: 398-426.

Mineva, Daniela; Fazekas, Mihály; Poltoratskaya, Viktoriia; and Tsabala, Kristina (2023) Rolling Back State Capture in Southeast Europe. Implementing Effective Instruments for Asset Declaration and Politically Exposed Companies. Center for the Study of Democracy, Sofia.

OECD (2022). *Open Government Review of Brazil: Towards an Integrated Open Government Agenda,* OECD Public Governance Reviews, OECD Publishing, Paris.

OECD (2016) Preventing Corruption in Public Procurement. Available at: https://www.oecd.org/gov/ethics/Corruption-Public-Procurement-Brochure.pdf (accessed 15 November 2021).

OECD (2017) Preventing Policy Capture. Integrity in Public Decision Making. OECD Publishing, Paris.

OECD (2019) OECD Integrity Review of Mexico City: Upgrading the Local Anti-Corruption System. OECD Public Governance Reviews. OECD Publishing, Paris.

Santos, FB and Souza, KR (2016). *Como combater a corrupção em licitações: detecção e prevenção de fraudes*, Fórum, Belo Horizonte.

Tax Justice Network (2022) Financial Secrecy Index. 2022 Methodology. Tax Justice Network. See: https://fsi.taxjustice.net/fsi2022/methodology.pdf (accessed on the 16th of March 2023)

Velasco RB (2019) Identifying Corruption Risk in Brazil: New Measures for Effective Oversight. In: Rotberg RI (ed.) Corruption in Latin America. Cham: Springer International Publishing, pp. 57–91.

Velasco RB, Carpanese I, Interian R, et al. (2020). A decision support system for fraud detection in public procurement. *International Transactions in Operational Research* 28(1): 27–47.

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

https://stock.adobe.com/pt/images/hacker-attack-maintenance-concept-and-hacking-cyber-crime-cyber-security-user-is-using-smartphone-with-warning-triangle-for-error-notification/620-026942

https://stock.adobe.com/pt/images/entrepreneurs-small-business-sme-independent-men-work-at-home-use-smartphones-and-laptops-for-commercial-checking-online-marketing-packing-boxes-sme-sellers-concept-e-commerce-team-online-sales/621503743

https://stock.adobe.com/pt/images/rio-de-janeiro-june-21-2017-the-selaron-steps-in-the-historic-center-of-rio-de-janeiro-brazil/191264522

https://stock.adobe.com/pt/images/perspective-view-of-stock-market-growth-business-investing-and-data-concept-with-digital-financial-chart-graphs-diagrams-and-indicators-on-dark-blue-blurry-background/610683944

https://stock.adobe.com/pt/images/human-multicolored-iris-of-the-eye-animation-concept-rainbow-lines-after-a-flash-scatter-out-of-a-bright-binary-circle-and-forming-volumetric-a-human-eye-iris-and-pupil-3d-rendering-background-4k/612500470

https://stock.adobe.com/pt/images/vector-illustrations-of-futuristic-digital-tech-architecture-abstract-blue-hi-tech-theme-for-davertising-or-game-artwork-futuristic-concept/600939016

https://stock.adobe.com/pt/images/un-analyste-data-travaillant-sur-des-jeux-de-donnees-depuis-sont-ordinateur-portable/634850584

https://stock.adobe.com/pt/images/modern-neon-cyberpunk-open-space-office-interior-blurred-with-information-technology-overlay-corporate-strategy-for-finance-operations-marketing-generative-ai-technology/628395923

https://stock.adobe.com/pt/images/long-exposure-shot-of-crowd-of-business-people-walking-in-bright-office-lobby-fast-moving-with-blurry-generative-ai/619509578

Appendix I. GRAS Data Field Requirements

Table 11 below displays in detail the relevant data fields for each dataset category considered in the GRAS framework. Those fields marked in bold are minimum requirements for the computation of GRAS indicators, and the other fields listed refer to additional information useful for data validation or to be displayed in the user interface as reference for system users. Depending on specificities of the data sources mapped as part of the feasibility assessment, these may be adjusted or complemented to better reflect corresponding or equivalent fields contained in the relevant datasets.

Specific numerical identifiers for individuals, companies and agencies adopted should be common to the different datasets to allow cross-referencing for the risk assessment. Which identifiers are more widely used for registration across different databases will also be context-specific, as they vary from country to country.

> > >

TABLE 11 - List of Data Fields Required for GRAS in each Dataset Category

Dataset category	*Dataset field*
1. Public procurement	**AgencyID**
	AgencyName
	AgencyLocation
	AgencyGovLevel
	ID Process
	Date
	NumberBids
	ProcurementMethod
	ItemNumber
	BidNumber
	BidValue
	BidDate
	AwardBid
	FirmName
	FirmID
	ContractValue
	ContractDate
	ContractObject
	ContractAmendValue
	ContractAmendDate
	ContractAmendObject
	ContractID
	PublicationDate
	BidDeadline
	DeliveryDate(estimated)
	ImplementationDate(final)
	ContractDeliveryDelay
	ContractEndDate
	ProductCode
	FirmLocation
	ImplementationLocation
	EstimatedPrice
	TenderFinalPrice

CONTINUED

Dataset category	*Dataset field*
2. Employment relationships	**EmployerID**
	EmployerName
	WorkerID
	WorkerName
	Date of birth
	Admission date
	Termination date
	Position
	Remuneration
3. Corporate and shareholder data	**FirmID**
	FirmName
	EntityType
	FirmCountry
	FirmAddress
	FirmPhonenumber
	FirmEmail
	FirmActivityCode
	FirmActivity
	Profit
	Turnover
	FirmConstitutionDate
	Year
	ShareholderID
	ShareholderName
	ShareholderEntryDate
	ShareholderExitDate
	ShareholderCountry
	LegalRepresentativeID
	LegalRepresentativeName
	AccountantID
	AccountantName
	EconomicGroup
4. Electoral data	**CandidateID**
	CandidateName
	ElectionDisputed
	OfficeDisputed
	Elected
	PartyName
	ElectionJurisdiction
	AffiliationStart
	AffiliationEnd
	PartyRepresentationStart
	PartyRepresentationEnd

CONTINUED

Dataset category	*Dataset field*
4. Electoral data	**PartyRepresentationPosition**
	CampaignDonorID
	CampaignDonorName
	CampaignDonationValue
	DonorLocation
	CampaignSupplierID
	CampaignSupplierName
	ExpenseValue
	SupplierLocation
	Year
5. Blacklists	**SanctionedID**
	SanctionedName
	Sanction_date(sta)
	Sanction_date(end)
	SanctioningOrgID
	SanctioningOrgName
6. Socio-economic data	**BeneficiaryID**
	BeneficiaryName
	BeneficiaryLocation
	BenefitDate(first)
	BenefitDate(last)
	BenefitValueTotal
	CashtransprogramType
	HouseholdID
	HouseholdLocation
	HouseholdClassification
	ClassificationValidStart
	ClassificationValidEnd
	ClassificationDate
	HouseholdmemberID
	HouseholdmemberName
7. Criminal records	**PersonID**
	PersonName
	Crime
	SentenceDate
8. Asset and interest declarations	**PersonID**
	Name
	AgencyID
	AgencyName
	Position
	Year

CONTINUED

Dataset category	*Dataset field*
8. Asset and interest declarations	**ShareholderCompanyID**
	ShareholderCompanyName
	RelativeID
	RelativeName
	RelativeWorkplace
	FamilyRelation

Appendix II. Feasibility Assessment at the Indicator Level

Table 12 below indicates specific data fields that are required for a feasibility assessment of individual GRAS indicators, as they are needed to compute the variables employed in the related risk assessment. Again, these should also reflect potential adjustments in the data fields identified as relevant in each context and in the indicator framework, should context-specific opportunities or constraints require them.

> > >

TABLE 12 - Required Data Fields Per Red Flag/Indicator

Risk group	*Risk pattern*	*Red flag/ indicator*	*Dataset category*	*Required data fields*
1. Procure-ment cycle	1.1. Non--competitive processes	1.1.1. Contract share through non-competitive procedures	1. Public procure-ment	ID Process; ProcurementMethod; FirmID; ContractID; ContractValue; ContractDate
		1.1.2. Contract share after call for tenders absent	1. Public procure-ment	ID Process; FirmID; ContractID; ContractValue; ContractDate; Pu-blicationDate; BidDeadline
		1.1.3. Contract share after shor-tened advertisement period	1. Public procure-ment	ID Process; FirmID; ContractID; ContractValue; ContractDate; Pu-blicationDate; BidDeadline
	1.2. Non--competiti-ve tender results	1.2.1. Contract share as single bidder	1. Public procure-ment	ID Process; FirmID; ContractID; ContractValue; ContractDate; NumberBids
		1.2.2. High winning rate	1. Public procure-ment	ID Process; FirmID; BidNumber; BidValue; BidDate; AwardBid
		1.2.3. Contract share in buyer's portfolio	1. Public procure-ment	ID Process; FirmID; ContractID; ContractValue; ContractDate; AgencyID
	1.3. Con-tract imple-mentation biases	1.3.1. Contract share with size-able cost overruns	1. Public procure-ment	ID Process; FirmID; ContractID; ContractValue; ContractDate; Con-tractAmendValue
		1.3.2. Contract share with size-able delivery delay	1. Public procure-ment	ID Process; FirmID; ContractID; ContractValue; ContractDate; DeliveryDate(estimated); Imple-mentationDate(final); ContractDeli-veryDelay; ContractEndDate
2. Collusion	2.1. Top loser	2.1.1. Low winning rate	1. Public procure-ment	ID Process; FirmID; BidNumber; BidValue; BidDate; AwardBid; ProductCode; FirmLocation; Imple-mentationLocation
		2.1.2. Number of competitors	1. Public procure-ment	ID Process; FirmID; BidNumber; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLo-cation
		2.1.3. Number of wins against Top Losers	1. Public procure-ment	ID Process; FirmID; BidNumber; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLo-cation
		2.1.4. Winning rate against Top Losers	1. Public procure-ment	ID Process; FirmID; BidNumber; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLo-cation
		2.1.5. Number of Top Loser competitors	1. Public procure-ment	ID Process; FirmID; BidNumber; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLo-cation

CONTINUED

Risk group	Risk pattern	Red flag/ indicator	Dataset category	Required data fields
2. Collusion	2.2. Fixed difference bids	2.2.1. Number of colluding partners with fixed difference bids	1. Public procurement	ID Process; FirmID; BidNumber; BidValue; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLocation
		2.2.2. Number of bids with fixed difference bids	1. Public procurement	ID Process; FirmID; BidNumber; BidValue; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLocation
		2.2.3. Frequency of fixed difference bids	1. Public procurement	ID Process; FirmID; BidNumber; BidValue; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLocation
	2.3. Bid variance biases	2.3.1. Bid share in low variance tenders	1. Public procurement	ID Process; FirmID; BidNumber; BidValue; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLocation
		2.3.2. Bid share in high relative bid distance tenders	1. Public procurement	ID Process; FirmID; BidNumber; BidValue; BidDate; AwardBid; ProductCode; FirmLocation; ImplementationLocation
	2.4. Unusual contract value	2.4.1. Contract share with contract value violating Benford's Law	1. Public procurement	ID Process; FirmID; ContractID; ContractValue; ContractDate; ProductCode; FirmLocation; ImplementationLocation
	2.5. High price	2.5.1. Contract share with very high relative contract value	1. Public procurement	ID Process; FirmID; EstimatedPrice; TenderFinalPrice; BidNumber; BidValue; BidDate; AwardBid; ContractID; ContractValue; ContractDate; ProductCode; FirmLocation; ImplementationLocation
	2.6. Common registration data	2.6.1. Number of competitors sharing registration data	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; FirmAddress; FirmPhonenumber; FirmEmail; LegalRepresentativeID; AccountantID
		2.6.2. Number of tenders with bidders sharing registration data	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; FirmAddress; FirmPhonenumber; FirmEmail; LegalRepresentativeID; AccountantID
		2.6.3. Share of contracts won against competitors sharing registration data	1. Public procurement	ID Process; FirmID; BidNumber; ContractID; ContractValue; ContractDate; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; FirmAddress; FirmPhonenumber; FirmEmail; LegalRepresentativeID; AccountantID
	2.7. Common shareholder	2.7.1. Number of competitors with common shareholde	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate

CONTINUED

Risk group	*Risk pattern*	*Red flag/ indicator*	*Dataset category*	*Required data fields*
2. Collusion	2.7. Common shareholder	2.7.1. Number of competitors with common shareholder	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate
		2.7.2. Number of tenders with competitors sharing a shareholder	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate
		2.7.3. Share of contracts won against competitors with common shareholder	1. Public procurement	ID Process; FirmID; BidNumber; ContractID; ContractValue; ContractDate; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate
		2.7.4. Number of competitors in the same corporate group	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; EconomicGroup
		2.7.5. Number of tenders with competitors in the same corporate group	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; EconomicGroup
		2.7.6. Share of contracts won against competitors in the same corporate group	1. Public procurement	ID Process; FirmID; BidNumber; ContractID; ContractValue; ContractDate; ProductCode; FirmLocation; ImplementationLocation
			3. Corporate and shareholder data	FirmID; EconomicGroup
	2.8. Common employee	2.8.1. Number of competitors with common employee	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			2. Employment relationships	EmployerID; WorkerID
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate
		2.8.2. Number of tenders with competitors sharing an employee	1. Public procurement	ID Process; FirmID; BidNumber; ProductCode; FirmLocation; ImplementationLocation
			2. Employment relationships	EmployerID; WorkerID
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate

CONTINUED

Risk group	Risk pattern	Red flag/ indicator	Dataset category	Required data fields
2. Collusion	2.8. Common employee	2.8.3. Share of contracts won against competitors with common employee	1. Public procurement	ID Process; FirmID; BidNumber; ContractID; ContractValue; ContractDate; ProductCode; FirmLocation; ImplementationLocation
			2. Employment relationships	EmployerID; WorkerID
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate

Risk group	*Risk pattern*	*Red flag/ indicator*	*Dataset category*	*Required data fields*
3. Supplier characteristics	3.6. Sanctions	3.6.4. Contracts while sanctioned	1. Public procurement	ID Process; FirmID; ContractID; ContractValue; ContractDate
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			5. Blacklists	SanctionedID; Sanction_date(sta); Sanction_date(end)
		3.6.5. Sanction relative duration	1. Public procurement	FirmID
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			5. Blacklists	SanctionedID; Sanction_date(sta); Sanction_date(end)
		3.6.6. Period between incorporation and 1st sanction	1. Public procurement	FirmID
			3. Corporate and shareholder data	FirmID; FirmConstitutionDate
			5. Blacklists	SanctionedID; Sanction_date(sta)
	3.7. Shareholder with low socio-economic status	3.7.1. Shareholder has low socio-economic status	1. Public procurement	FirmID
			2. Employment relationships	EmployerID; WorkerID; Position; Remuneration; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			6. Socioeconomic data	BeneficiaryID; HouseholdID; Classification; ClassificationValidStart; ClassificationValidEnd; HouseholdmemberID; BenefitDate(first); BenefitDate(last)
		3.7.2. Status duration	1. Public procurement	FirmID
			2. Employment relationships	EmployerID; WorkerID; Position; Remuneration; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			6. Socioeconomic data	BeneficiaryID; HouseholdID; Classification; ClassificationDate; ClassificationValidStart; ClassificationValidEnd; HouseholdmemberID; BenefitDate(first); BenefitDate(last)
		3.7.3. Time overlap between status and company ownership	1. Public procurement	FirmID
			2. Employment relationships	EmployerID; WorkerID; Position; Remuneration; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID

CONTINUED

Risk group	*Risk pattern*	*Red flag/ indicator*	*Dataset category*	*Required data fields*
3. Supplier characteristics	3.7. Shareholder with low socio-economic status	3.7.3. Time overlap between status and company ownership	6. Socioeconomic data	BeneficiaryID; HouseholdID; Classification; ClassificationValidStart; ClassificationValidEnd; HouseholdmemberID; BenefitDate(first); BenefitDate(last)
	3.8. Shareholder/ legal representative with criminal record	3.8.1. Convicted shareholder	1. Public procurement	FirmID
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			7. Criminal records	PersonID
	3.9. Tax haven registration	3.9.1. Company registered in tax haven	1. Public procurement	FirmID
			3. Corporate and shareholder data	FirmID; FirmCountry
		3.9.2. Shareholder registered in tax haven	1. Public procurement	FirmID
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; ShareholderCountry
4. Political connections	4.1. Political finance	4.1.1. Donation to electoral campaign	1. Public procurement	FirmID
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			4. Electoral data	CandidateID; ElectionDisputed; OfficeDisputed; PartyName; ElectionJurisdiction; Elected; CampaignDonorID; CampaignDonationValue; DonorLocation; CampaignSupplierID; ExpenseValue; SupplierLocation; Year
		4.1.2. Value of donation to electoral campaign	1. Public procurement	FirmID
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			4. Electoral data	CandidateID; ElectionDisputed; OfficeDisputed; PartyName; ElectionJurisdiction; Elected; CampaignDonorID; CampaignDonationValue; DonorLocation; CampaignSupplierID; ExpenseValue; SupplierLocation; Year
		4.1.3. Contracts won following donation	1. Public procurement	ID Process; AgencyID; AgencyLocation; AgencyGovLevel; FirmID; ContractID; ContractValue; ContractDate

CONTINUED

Risk group	Risk pattern	Red flag/ indicator	Dataset category	Required data fields
4. Political connections	4.1. Political finance	4.1.3. Contracts won following donation	2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			4. Electoral data	CandidateID; ElectionDisputed; OfficeDisputed; PartyName; ElectionJurisdiction; Elected; CampaignDonorID; CampaignDonationValue; DonorLocation; CampaignSupplierID; ExpenseValue; SupplierLocation; Year
		4.1.4. Percent of contracts won following donation	1. Public procurement	ID Process; AgencyID; AgencyLocation; AgencyGovLevel; FirmID; ContractID; ContractValue; ContractDate
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			4. Electoral data	CandidateID; ElectionDisputed; OfficeDisputed; PartyName; ElectionJurisdiction; Elected; CampaignDonorID; CampaignDonationValue; DonorLocation; CampaignSupplierID; ExpenseValue; SupplierLocation; Year
	4.2. Personal connections to politicians	4.2.1. Company's personal connections to politicians	1. Public procurement	FirmID
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			4. Electoral data	CandidateID; ElectionDisputed; OfficeDisputed; PartyName; ElectionJurisdiction; Elected; ElectoralProcess; AffiliationStart; AffiliationEnd; PartyRepresentationStart; PartyRepresentationEnd; PartyRepresentationPosition
			8. Asset and interest declarations	PersonID; AgencyID; Year; ShareholderCompanyID; RelativeID
		4.2.2. Contracts won following political connection	1. Public procurement	ID Process; AgencyID; AgencyLocation; AgencyGovLevel; FirmID; ContractID; ContractValue; ContractDate
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date

CONTINUED

Risk group	Risk pattern	Red flag/ indicator	Dataset category	Required data fields
4. Political connections	4.2. Personal connections to politicians	4.2.2. Contracts won following political connection	3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			4. Electoral data	CandidateID; ElectionDisputed; OfficeDisputed; PartyName; ElectionJurisdiction; Elected; ElectoralProcess; AffiliationStart; AffiliationEnd; PartyRepresentationStart; PartyRepresentationEnd; PartyRepresentationPosition
			8. Asset and interest declarations	PersonID; AgencyID; Year; ShareholderCompanyID; RelativeID
		4.2.3. Percent of contracts won following political connection	1. Public procurement	ID Process; AgencyID; AgencyLocation; AgencyGovLevel; FirmID; ContractID; ContractValue; ContractDate
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			4. Electoral data	CandidateID; ElectionDisputed; OfficeDisputed; PartyName; ElectionJurisdiction; Elected; ElectoralProcess; AffiliationStart; AffiliationEnd; PartyRepresentationStart; PartyRepresentationEnd; PartyRepresentationPosition
			8. Asset and interest declarations	PersonID; AgencyID; Year; ShareholderCompanyID; RelativeID
	4.3. Personal connections to bureaucrats	4.3.1. Company's personal connections to bureaucrat	1. Public procurement	FirmID
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date; Position
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			8. Asset and interest declarations	PersonID; AgencyID; Year; ShareholderCompanyID; RelativeID
		4.3.2. Contracts won following connection to bureaucrat	1. Public procurement	ID Process; AgencyID; AgencyLocation; AgencyGovLevel; FirmID; ContractID; ContractValue; ContractDate
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date; Position
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			8. Asset and interest declarations	PersonID; AgencyID; Year; ShareholderCompanyID; RelativeID

CONTINUED

Risk group	*Risk pattern*	*Red flag/ indicator*	*Dataset category*	*Required data fields*
4. Political connections	4.3. Personal connections to bureaucrats	4.3.3. Percent of contracts won following connection to bureaucrat	1. Public procurement	ID Process; AgencyID; AgencyLocation; AgencyGovLevel; FirmID; ContractID; ContractValue; ContractDate
			2. Employment relationships	EmployerID; WorkerID; Admission date; Termination date; Position
			3. Corporate and shareholder data	FirmID; ShareholderID; ShareholderEntryDate; ShareholderExitDate; LegalRepresentativeID
			8. Asset and interest declarations	PersonID; AgencyID; Year; ShareholderCompanyID; RelativeID

Appendix III. GRAS Architecture

The Governance Risk Assessment System (GRAS) is an agnostic system, meaning it can operate independently of any specific hardware or software configurations, making it highly versatile. It is designed to be both flexible and scalable. GRAS's architecture consists of three core components:

- Database: Stores public procurement data, registration data, and risk patterns. This component uses PostgreSQL, a robust and reliable database management system.

- API: Acts as the conduit between the database and the web interface. The API is developed using Django, a powerful Python programming tool.

- Web Application: Generates the web pages that users use to analyze risk pattern data. This user interface is built using the JavaScript React framework.

GRAS takes advantage of Docker container technology, an open-source solution that ensures portability and facilitates easy deployment in various environments, be it local or cloud-based platforms such as AWS, Google Cloud, Azure, etc. The following is a detailed explanation of how GRAS harnesses this concept.

Docker Usage

Docker is an open-source tool that streamlines the creation, deployment, and running of applications in containers. It provides isolation, maintaining distinct environments for development, testing, and production. This feature ensures consistency, meaning the application behaves the same way across different stages, hence minimizing potential errors or discrepancies. Therefore, by using Docker, GRAS can operate seamlessly across different environments.

Apart from standard Docker, GRAS utilizes Docker Compose, an extension of Docker. It simplifies the configuration and operation of applications that involve multiple containers, enabling the definition of dependencies within a single file.

GRAS's Architecture with Docker

GRAS's architecture incorporates four Docker containers:

- database: This is a Postgres container that hosts the system's database.

- api: A container housing the API application, which establishes connections between the database and the web application.
- web: This container hosts the web application and creates a connection with the API.

- etl: Standing for Extract, Transform, Load, this container is tasked with updating registration databases and connects with the database container.

Each container operates akin to a virtual machine, serving as an independent server for their respective applications within GRAS. The containers communicate through an internal Docker network, which bolsters the system's security by avoiding the need to open additional ports, except for the web application.

Deploying GRAS in Docker containers ensures consistent application behavior—termed as 'uniform application execution'—regardless of the hosting environment. This provides a sturdy, scalable, and secure architectural solution for risk pattern data analysis.

WORLD BANK GROUP