

DESIGNING POLITICAL INTEGRITY RISK INDICATORS

A HOW-TO GUIDE IN 11 STEPS

Transparency International is a global movement with one vision: a world in which government, business, civil society and the daily lives of people are free of corruption. With more than 100 chapters worldwide and an international secretariat in Berlin, we are leading the fight against corruption to turn this vision into reality.

www.transparency.org

Designing Political Integrity Risk Indicators

A how-to guide in 11 steps

Author: Viktoriia Poltoratskaya (Government Transparency Institute and Central European University)

Co-authors: Mihály Fazekas (Government Transparency Institute and Central European University) and Bence Tóth (Government Transparency Institute and University College London).

Reviewers: Jon Vrushni and Jorge Valladares (Transparency International Secretariat)

Every effort has been made to verify the accuracy of the information contained in this report. All information was believed to be correct as of October 2021. Nevertheless, Transparency International cannot accept responsibility for the consequences of its use for other purposes or in other contexts.

2021 Transparency International. Except where otherwise noted, this work is licensed under CC BY-ND 4.0 DE. Quotation permitted. Please contact Transparency International – copyright@transparency.org – regarding derivatives requests.



TABLE OF CONTENTS

Executive Summary	4
Introduction	5
11 Steps to Build Political Integrity Risk Indicators	6
Phase 1: Understanding what you want to measure ..	6
Phase 2: Getting the data you need for measurement	8
Phase 3: Developing, tailoring and validating your indicators	12
Phase 4: Documenting your results	15

Appendix	17
Data structure for a selection of political integrity datasets	17

EXECUTIVE SUMMARY

This how-to guide sets out to describe the process of developing valid and reliable indicators of political integrity for non-technical audiences.

It outlines 11 distinct steps organised in four phases, encompassing the initial objective setting, through to documenting results.

The guide not only offers succinct descriptions of each step, but it also provides a range of examples demonstrating challenges and common solutions regarding measurement.

It draws extensively on the state-of-the-art stocktaking of open data in the area of political integrity done by the Global Data Barometer project.

This guide organises the measurement process in the following phases and steps:

PHASE 1: Understanding what you want to measure

STEP 1: Setting measurement objectives

STEP 2: Defining corruption and political integrity

STEP 3: Documenting corruption schemes

PHASE 2: Getting the data you need for measurement

STEP 4: Precisely defining the target population

STEP 5: Screening potential data sources

STEP 6: Detailed mapping of the most promising data sources

STEP 7: Linking data

STEP 8: Narrowing down the measurement objective

PHASE 3: Developing, tailoring and validating your indicators

STEP 9: Developing indicators

STEP 10: Tailoring and validating indicators

PHASE 4: Documenting your results

STEP 11: Documenting conceptually-valid and empirically feasible indicators

INTRODUCTION

The need for the development of specific indicators of corruption – direct or indirect – has grown over the last decade. This can be attributed in part to the increasing recognition of the significant role corruption plays in adverse governance outcomes, such as poor health and environmental degradation. The demand for indicators has also emerged out of frustration with a lack of capacity to precisely measure the prevalence of various corrupt practices.

There are challenges in monitoring corrupt practices. Naturally, behaviour typically hidden from the public is hard to measure. In addition, corrupt behaviours take many forms and involve different kinds and numbers of actors, while the standards governing corrupt transactions vary in scope. As such, it would be ineffective to apply a universal set of common corruption indicators across the globe.

In order to avoid the pitfalls of over-simplifying the measurement of corruption for the benefit of informed policy decisions, this short guide walks the reader through the steps of developing, implementing and validating indicators of corruption in the political domain.

This guide is intended to help civil society groups and organisations develop and apply measurement skills to better monitor political corruption, without delving into statistical and data science technicalities.¹

This step-by-step guide follows a simple, intuitive logic; in order to create a monitoring plan, you need to:

1. figure out what exactly you want to measure;
2. identify, gather and organise the data you will need for measuring;
3. develop, tailor and validate your indicators for the data you collect; and
4. document the resulting indicators as implemented in your data for others to be able to use your work.

¹ In drafting this guide, we follow in the footsteps of other guidebooks and reviews such as Trapnell (2015).

11 STEPS TO BUILD POLITICAL INTEGRITY RISK INDICATORS

PHASE 1: UNDERSTANDING WHAT YOU WANT TO MEASURE

Step 1: Setting measurement objectives

The starting point of any measurement exercise is to define the measurement objectives. This should entail, at the very least, identifying the area of interest (i.e. political party financing) and the specific parameters of measurement. The parameters are defined, at the most essential level by the country or region (e.g. province, state, federation); the monitoring time-period (e.g. campaign period); and the relevant actors (e.g. political parties).

While at this early stage the objectives are naturally broader, setting realistic, yet relevant objectives is essential for a successful measurement exercise. Scanning data availability can help determine feasibility of the project. For example, measuring the integrity of party financing in the latest parliamentary elections at the federal level in Russia would be possible, as some open data exists in this domain.

STEP 2: Defining corruption and political integrity

Any definition of corruption must be both practical and relevant² in order to guide the investigation and data analysis of specific malpractices within the area of interest.

While sharing some features, exact forms of corrupt behaviours vary depending on the level of government, political regime, and peculiarities of market regulations, such as the extent of government intervention (Mungiu-Pippidi & Fazekas 2020).

Corruption has various definitions that are often contested or misinterpreted (Johnston, 1996). According to the generic definition by Transparency International, corruption is the abuse of entrusted power for private gain. Private gain can be financial, or refer to the acquisition of other valuable assets; it can also mean the strengthening of one's own power.³

Lack of political integrity may lead to various forms of corrupt behaviour. Transparency International defines political integrity as exercising power consistently for the common good, rather than sustaining the private interests, own wealth or position of power holders. (TI 2020). The opposite of political integrity is political corruption. This is the 'inappropriate use of common power and authority for purposes of individual or group gain at common expense' (Warren 2003). Political integrity can increase through direct involvement of policy beneficiaries in consultative processes and other

² It is up for debate whether a single, universally accepted definition of corruption can be formulated. Corrupt behaviour extends beyond the law. As Johnston (1996) put it, corruption has broader meaning than the letter of the law, as legal rules might be changed for private gain, which is particularly relevant in non-democratic or state capture contexts.

³ Another fundamental facet of corruption to consider is when public offices operate as businesses aimed at maximising material benefits. This recognition sets a clear division between political struggles and abuse of power, which leads to a third feature of corrupt behaviour. Corruption causes harm to the public interest, as its mechanisms are not aimed at achieving public prosperity, but rather at extracting rent where possible.

political processes. Such types of public involvement increase the transparency and accountability of government actions, and help prevent corruption.

To understand how private interests influence political activities, it is useful to think of political corruption as an exchange of favours across private and public spheres. These favours can refer to:⁴

- (i) a private actor advancing their own narrow interests – e.g. making a payment to a politician's private bank account;
- (ii) a public actor using his/her public power to return a favour – e.g. by offering government-backed loans on favourable terms.

With the first example of favour, entry points for private money into politics to be explored include:

- political party finance
- lobbying
- asset and interest disclosure

For instance, Fazekas, Ferrali and Wachs (2018) showed that campaign contributions confer favours by procuring entities controlled by politicians. This mechanism includes receiving campaign donations from the connected companies in the US federal government. Hidden or incomplete asset declarations and interest disclosures can obscure conflicts of interest or illicit enrichment. Lobbying could be considered as a legal way of rent-seeking, which can result in significant public loss (Brandt & Svendsen 2013).

Regarding the example of payback from public actors to private actors, the range of instruments and government activities that can be corrupted is very wide. There is no agreed list of such instruments, simply because any government domain can be abused. The list below provides a few of the frequently cited and widely researched domains of corruption:

- favouritism in the allocation of government contracts;
- tailored regulations favouring narrow interests; and
- preferential treatment in taxation, or when allocating publicly-backed loans.

⁴ Note here the focus is on private influence. Yet, at times, public actors seeking to advance their own power, engage in corrupt exchanges with other public actors whose political support they need. One example is Brazil's

STEP 3: Documenting corruption schemes

Measurement necessitates the enumeration of frequently used technologies, techniques, and strategies used to carry out corrupt transactions. These political corruption schemes and methods fall under two main categories: 1) money into politics; and 2) resource allocation decisions.

Money into politics

When money and politics intersect, there are risks of abuse for the sake of private gain. Money can influence the transparency and accountability of the political process in various ways. Corrupt money and favours generally enter politics through the entry points below:

(i) **Political finance** – donations to political parties or candidates from private companies or powerful individuals can result in influence over policy making, and can sometimes work as a means of kickbacks and bribes.

(ii) **Lobbying** – it can serve as a means to introduce and maintain influence over key political decisions, including certain policies or legislation.

(iii) **Asset and interest disclosure** – the absence of transparency prevents the general public from identifying politicians who personally benefit from supporting contentious legislation that favours certain interest groups.

It is important to consider that most of the known and documented cases of corruption in politics are not typical but rather salient or high-value. This is a common issue with identifying red flags. Expert opinion is often based on available information; this information may not be systematic or quantitatively analysed, but based on recent scandals and widely discussed cases. Indispensable methods for identifying common fraudulent practices include: systematic documentation of cases, classification of scheme types, and a deliberate search for less common corruption schemes.

Public consultation for rulemaking is an instrument intended not only to increase citizen participation, but also raise the level of transparency in the political process by subverting the power imbalance introduced by the presence of money in politics.

'mensalão', whereby the executive co-opted opposition legislators through 'vote buying' using embezzled public money.

However, this process is vulnerable to capture or co-option through personal connections and relationships between private actors and decision-makers.

Public resource allocation

The potential channels for allocating public resources to corrupt private actors may be as wide-ranging as the government activities vulnerable to interference. Nevertheless, the three domains considered among the most susceptible to corruption include:

- (i) **Laws and regulations** - ill-defined laws and unfair distribution of power can enable political corruption. This avenue is especially typical for political regimes with limited accountability. When this weakens the institutional capacity of the state, anti-corruption regulations are difficult to enforce.
- (ii) **Spending decisions related to public procurement and concessions** – distributing contracts to specific companies by avoiding competition through tailoring the requirements or establishing rules for direct contract award in exchange for kickbacks to the public officials.
- (iii) **Exemptions and lax regulatory enforcement** – the legislatures or governments specifically create tax loopholes that allow companies to pay less taxes. Even when the rules remain universally applicable, less stringent application of complex rules can confer considerable benefits to companies with connections.

PHASE 2: GETTING THE DATA YOU NEED FOR MEASUREMENT

Step 4: Precisely defining the target population

To start working with data, two main parameters have to be established: population and sample. The population is the universe comprising all observable areas of interest, such as all potentially corruptible policy decisions in the extractives licensing domain. A sample is a sub-group of this universe, often a randomly selected set of observations. To identify the most relevant, the selection of cases should follow from the specific measurement objective. In some instances, there is no need for sampling as the

data collection methods allow for collecting the relevant information about all cases in the population. One example of this is gathering information from the web on all registered political campaign donations for a particular election year. Drawing a sample is needed whenever collecting data on all cases in the population is too expensive or impractical. To properly represent the population, the sample should be drawn at random.

Defining the wrong universe can lead to common errors or risks at this stage. For instance, it would be an oversight to only monitor MPs' asset declarations, and neglect similar and related corrupt practices prevalent among top bureaucrats exempt from asset declaration requirements. Another potential pitfall is drawing the wrong sample. This could entail surveying all NGOs about corruption in policy making, while less than 10% of those organisations have a mandate to monitor corruption. Moreover, sometimes the sample is determined by the data available. If there are certain years missing from the dataset, or only certain countries/regions are covered, there is no possibility to account for the missing information.

STEP 4 - Example 1

The Russian Federation has open access data that can be processed to help identify corrupt practices in elections. It is documented that for electoral campaigns, in each region, an organized network of actors, such as bureaucrats, party representatives, state-owned companies and regional administrations establish their own 'political machine'. By cooperating with each other, they help the ruling party win elections. Yet to operate effectively, these actors need resources – certain financial and material benefits they can use for buying votes. There are datasets that can help trace these corrupt schemes, such as data on party finance, company ownership, public procurement and electoral results in each district.

In the case of Russian regional donors and branches of the ruling party, the population of relevant observations includes all electoral campaigns carried out with the participation of a current ruling party – United Russia. For each electoral campaign, all donations must be recorded, including detailed information on donors and recipients.

Yet the other databases (for each area such as contracting, ownership, etc) cover different time periods. For example, data on procurement is only

available from 2013 onward; therefore procurement data cannot be matched with electoral results and donations earlier than that. Moreover, details on donations are only available at the regional level, therefore municipal elections should not be considered as there is no information on which district the donation came from, or to which of the local offices of United Russia they were directed. Thus, the scope of the data is limited to regional and federal electoral campaigns from 2013–2021.

STEP 5: Screening potential data sources

It is necessary to screen the potentially relevant data sources based on the definition of political connections and exact corrupt practices. This step only concerns high-level scoping to establish a long list of relevant datasets and quickly assess them. Detailed data mapping is a laborious task. Hence this initial screening step is needed in order to focus only on the most relevant datasets.

This initial screening step must identify datasets that capture behaviours corresponding directly to the specific corruption schemes identified in Step 3. The level of detail and usability of the relevant datasets should be evaluated at this stage. Crucially, the detail of the dataset – or the level of observation – should be established. Aggregate data on party donations is much less valuable for indicator-building than detailed data on individual donations, donors and recipients.

The additional challenge at this stage is limited access, due to paywalls, or some segments of the data not being publicly available. For example, when working with data on company legal ownership, access is free for some countries, while for others users have to pay to access information for each individual company, without an option to purchase full access to all registered companies.

Another issue is data that is publicly and freely available, but of poor quality – for example, scans of forms for asset declarations for each parliamentarian that are hand-written. In addition, manual data collection methods also increase the probability of mistakes or missing observations, especially when datasets are not externally verified.

The Global Data Barometer is a useful resource for scoping, as it highlights the main themes and variables covered in the data on political integrity, such as political party financing, interest declarations, lobbying registers, public consultation

in rule-making, and right-to-information regime performance.

STEP 5 - Example 1

As mentioned previously, in Russia, open access data is available through different NGOs; however, information is collected manually, or stored on the websites of state agencies. As a result, the quality and validity of this data can be questionable. .

The NGO Golos, which independently monitors electoral fraud in Russia, collects data on donations. The [dataset](#) consists of the declared donations of all political parties, including names of organisations, their IDs, dates of transactions and donation amounts, and information on procurement contracts received by the donor. The issue is that all data points are collected manually from documents and pdfs published on the parties' websites. Therefore, there is the possibility of human error, including missing information. Furthermore, only officially-declared transactions are included. Finally, information on procurement contracts does not factor in affiliated companies, and does not include the date of contract signature, which makes it difficult to identify if the contract was received immediately following an election or before. On the other hand, all information is regularly updated, and IDs are provided, which makes it possible to link companies to other datasets.

STEP 6: Detailed mapping of the most promising data sources

After identifying potential data sources, map out the most promising datasets. This detailed mapping should consider the following data features:

1. **Scope:** The percentage of the relevant population covered in the dataset. For example, the share of the total public procurement spending in a country that is reported in the tendering and contracts dataset. Scope also encompasses the time period covered by the data, such as whether the datasets are regularly updated. Additionally, the time period covered can impose certain limits on information for the indicators, therefore the investigation should be planned accordingly.
2. **Depth:** The amount of detailed information available for each observation. This requires listing all the relevant variables available in the dataset, and cross referencing this with the

desired variables, for corruption measurement purposes. For political party financing, the data details can be mapped using a simplified variable typology, highlighted in Table 1.

3. **Quality:** The quality of the data encompasses the completeness and truthfulness of the data in regards to actual behaviour. A basic method of evaluating data quality is to check for missing values. It is also critical to look for obvious errors, such as typos or nonsensical information (e.g. a company name typed up instead of a contract value in a public procurement announcement).
4. **Accessibility:** Data accessibility implies that the data is machine readable, easily downloadable and able to be processed. If data access requires complicated and error-prone web scraping, it may present considerable barriers to data use for measurement purposes.
5. **Interoperability:** Mapping includes assessing how different datasets can be linked in a meaningful way to enable the amalgamation of information. For example, if asset declarations data cannot be connected to specific public organisations (i.e. people reporting their assets cannot be connected to the institutions they are affiliated with) then we cannot connect asset declarations to the contracting risks of those public organisations. Furthermore, connecting people to organisations on its own is often not enough; information on the time of affiliation is also important.

If the research aims at establishing the accessibility of political integrity data, the variables (questions) required should be filtered by availability indicators.

It is important to note that quantitative data may not capture all existing corruption schemes. Some practices can best be captured by qualitative data collection methods, such as interviews and focus groups, or by proxies. Proxies capture certain phenomena by indirect measurement. The absence of required variables should be considered at this stage to inform whether to continue searching for alternative quantitative or qualitative data, or use given variables in an indirect way.

Table 1. Example political donations data fields

Variable name	Type	Potential schemes
<i>Donors IDs</i>	ID	Can be used to match with other datasets as well as to identify the regularity of donations.
<i>Recipient IDs</i>	ID	Can be used to match with other datasets as well as to identify the regularity of donations.
<i>Amount of transaction</i>	numeric	Particularly large donations can be a signal of hidden agreements or negotiations.
<i>Type of non-financial support</i>	Factor	Ambiguous items could be a sign of bribery.
<i>Date of transaction</i>	Date	Can be used to monitor correlation between transactions and declared assets, lobbying meetings, etc.
<i>Name of recipient party/politician</i>	character	Can be used to match with other datasets.
<i>Total income of party/politician</i>	numeric	Can be used to calculate the amount donated in relation to the total amount of donations, to estimate the relative input of donor.
<i>Address of donor</i>	character	Donors located in the same district/territory as politicians can be a part of clientelism.

The mapping for Russian data on electoral campaigns and donations can look in the following way

Name of the dataset	Organisation collecting the data	Level of observation
Donations	NGO Golos	transaction level
Procurement	State procurement system	lot level
Company register	Spark	company level
Electoral results	Central electoral committee	precinct electoral commission

STEP 7: Linking data

It is likely the required indicators are located in separate datasets, therefore linking data is necessary. In order to merge datasets, the unit of observation must first be established. For example, there are a few datasets on state subsidies and grants provided to certain companies. Some of the datasets will contain information on companies, therefore the level of observation is *company*; others provide information on subsidies and grants, therefore the unit of observation is *subsidy* or *grant* call. In some cases, it is quite challenging to merge datasets of different levels of observation. In the absence of unique IDs, the row will be multiplied many times, resulting in identical observations. The IDs of the two merging datasets should be unique, to ensure that while merging, it will be clear which row corresponds to which ID.

A potential related issue with merging different units of observation is that one dataset contains unique IDs, while the other dataset with the same ID variable has multiple rows for each ID. In this instance, it is possible to end up with multiplied IDs in the main dataset, which should be avoided for further linking. For example, in the case of company-level data with an address as a unit of analysis, the only ID by which it is possible to merge this dataset to the main one is company ID; however, they are multiplied because the same company might have a few addresses. To properly merge this type of dataset, it should first be aligned to unique IDs – for instance by transforming rows to columns. In this case the address for each company is going to be in a separate column..

STEP 7 – Example 1

As was shown in the previous example, all of the datasets needed for analysis contain information on different levels of observation, and as such, they must align to the same units. In order to do so, they must first be merged by their unique IDs. The matching process is outlined in the figure below:

Matching Process

Datasets	IDs to merge by	Final level of observations
Donations+procurement	Companies IDs	Company level
Donations+procurement+company register	Companies IDs	Company level
Donations+procurement+company register+elections	Region + year	Region-year, hence companies' data is collapsed to the average numbers/largest donations

STEP 8: Narrowing down the measurement objective

Before beginning to work with data, measurement objectives should be revisited and further refined. Once the corrupt behaviours and the datasets capturing key aspects of these behaviours are clearly formulated, realistic measurement objectives can be set. In order to do this, it is necessary to

refine which corruption techniques can be captured, and which ones should be left unmeasured due to lack of data.

Narrowing the scope of measurement objectives typically takes place once the data is better understood. For example, with the concept of political party financing, it can be defined as all declared financial transactions between private actors, state institutions and political parties. This narrows down the measurement scope to financial activities, so that, for instance, other types of 'gifts' and support are not taken into account. It also puts the focus on declared exchanges only, leaving direct cash bribes or other informal interactions unmeasured.

In many cases corruption cannot be measured directly. Due to the hidden nature of corrupt behaviour, certain measurements are needed to establish risks, rather than identify existing corrupt practices. For instance, in political party financing, personal ties between donating companies and party members, as well as public procurement contracts received by party donors, could be a signal of such risk. While the direct exchange of private gains is not captured, a conflict of interest is a sign of potential corrupt behaviour.

STEP 8 – Example 1

Generally, in the case of electoral corruption in Russia, one should look for a correlation between resources, corruption risks, and the success of a political machine in a given district. Resources can refer to the types of economic actors supporting the party (i.e. what sector of the economy the company belongs to; its size, whether it's state-owned or private, etc). 'Corruption risks' can entail any ties and connections between donors and party members, as well as conflicts of interest (e.g. donors receiving large amounts of procurement contracts right before or after the elections, or donors being related to party officials). Finally, the political machine's success can be determined by the share of votes received by the ruling party in comparison to others. The assumption to test is if the less transparent sectors of economy, such as natural resource extraction, or domains with high levels of corruption risks, influence the electoral results of the ruling party through clientelistic exchanges.

It is the availability of data that sets limitations on what can be measured. For example, it is not possible to study the dynamic of political corruption

before 2013. Moreover, in order to match the data on elections and align all datasets to the same unit of measurement (region-year), many company-level observations have to be excluded; therefore, only high-level aggregation is possible in assessing companies' impact on the electoral outcome. Additionally, district-level data is difficult to obtain and almost impossible to match with the companies; therefore only regional and federal elections can be studied.

PHASE 3: DEVELOPING, TAILORING AND VALIDATING YOUR INDICATORS

STEP 9: Developing indicators

The first step in developing indicators is to clearly define the phenomenon to be measured. It could be a formula calculating a certain parameter, or a brief definition fully describing what the indicator should capture. The definition should establish a clear association between corruption and the observable phenomena (Mungiu-Pippidi & Fazekas, 2020).

When formulating indicators, it is important to make sure they refer back to the definition of corruption, as well as to particular corruption techniques. For example, if corruption is defined as 'material inducement to abuse office or commit bribery', and one of the ways to measure it is to look at political favouritism, the potential indicator could be political office holders' employment with companies (Cingano & Pinotti, 2013). By looking at employment, it is possible to establish conflict of interest between the private sector and public office holders, which leads to the concept of political favouritism, and the initial definition of corruption.

The next step involves calculating technically feasible indicators and testing their statistical properties. To use indicators for the analysis, there should be sufficient variation within observations. Variation thresholds must be based on other parameters, such as the overall number of observations, presence of similar variables, etc. Another characteristic to consider is missing value rates. Suspiciously high numbers of missing values could be a sign of concealing information, or improper data collection. For instance, the absence of a significant number of observations on the number of bidders per tender – one of the key indicators to assess integrity – in public procurement data, could be a sign of low

engagement, covered up by public officials. Similarly, multiple changes in the data could be a signal of attempts to hide relevant information.

STEP 9 – Example 1

The following indicators can be used to measure corruption risks caused by donating companies in Russia:

Indicator	Scheme	Variable(s)
Donor receiving procurement contracts	Companies which received procurement might increase tunnelling around elections, which could be a sign of kickbacks for winning procurement contracts paid back through electoral campaigns support.	Procurement contract (present or absent, amount of contracts, and value)
Donations of unusually high amount	Companies donating amounts higher than average per region with relatively low annual income, could be involved in money laundering or other illegal activities.	Amount of donations, company size
Donors located in the same district as United Russia branch	Companies located in the same district with the branch of United Russia they supported, might have private interests in lobbying certain tax exemptions or favourable laws.	Company location, region
Donations from less transparent sectors	In the case of donations from less transparent sectors, it is more difficult to monitor the income of a donating company, which creates additional corruption risk.	Sector of economy

STEP 9 – Example 2

The following are a few examples of risk indicators proposed at practical exercises by participants of the III Political Integrity Bootcamp⁵: *Detecting Political Corruption Risks* (in October 2021).

Indicator	Scheme	Datasets and variable(s)
Portugal: MPs hold interests in entities their committee oversees	MPs use their legislative committee work to favour entities they hold interests in as well as their respective markets, preparing for their future jobs/pre-retirement.	MPs' Interest declarations (name, identifier, main activity, current role, past roles, supports and benefits received, services provided to private entities, exclusivity); MPs Parliamentary initiatives (bill initiative number, name and description, co-sponsoring MPs, responsible parties, current stage).
Slovenia: Donation of gifts to influence resource allocation	Provision of gifts to public officials can be a material inducement for providing favoured treatment to the gift-giver.	Gifts registry (date of gift, name of recipient agency, status of recipient, type, mode of delivery, description, value, reasons, owner); Company Register (registration ID, full company name); Public procurement (buyer ID, legal name, postcode; bidder ID, legal name, postcode, amount EUR, award criteria details).
Hungary: Winning suppliers' record of corrupt activities	Winning suppliers with past record of fraudulent and corrupt activities increases the risk of future wrongdoings.	Procurement (supplier company name, company ID, contract award publication date, contract/tender identifier); Press articles about corruption cases (company tagged).
Ecuador: Buyer capture	High-ranking public officials with control of companies possibly favoured in public procurement contracts.	SUPERCAS Registry (company name, ID, beneficial owner); Sworn statements of assets of the Comptroller General's Office of the State (name of company linked to public official, date of connection); Public procurement platform (supplier name, supplier ID, contract value, date).

⁵ The Bootcamp is a collaboration between the Office for Democratic Institutions and Human Rights of the Organization for Security and Cooperation in Europe (OSCE/ODIHR), the Transparency International (TI)

Secretariat, and TI national chapters from Europe and the Central Asia region. In the 2021 edition, chapters from the Americas, Africa and Asia joined as guests.

STEP 10: Tailoring and validating indicators

Terminology

Tailoring indicators is a process of adjusting the exact indicator formulation for capturing the phenomenon of interest more precisely. From a measurement perspective, it is essential that false positives (i.e. cases when the indicator signals corruption but there is no corruption taking place) as well as false negatives (i.e. when the indicators fail to signal corruption, though corruption actually takes place) are both minimised.

Step 10 – Example 1

After formulating indicators as per the previous example, tailor your indicators accordingly. It can be done in the following way:

Indicator	Values	Threshold
<i>Donor receiving procurement contracts</i>	1 if yes, 0 if no	Presence or absence of procurement contract. If affiliated company received contract and not the donating one, value is still 1.
<i>Donations of unusually high amount</i>	0 if within average values or lower, 1 if higher, 2 if significantly higher	Calculate averages for each region-year. Test if the value is significantly higher than the average.
<i>Donors located in the same district as United Russia branch</i>	1 if yes, 0 if no	If the company is present in the same region as the branch of the party it donated to, the value is 1.
<i>Donations from less transparent sectors</i>	1 - the least transparent, 5 - the most transparent	Should be evaluated through existing reports on sectoral transparency (i.e. which information is publicly available according to the law, what is the level of state support as subsidies etc.).

In practice, indicators often take the form of categorical or binary variables (take only the value 0 or 1 and indicate if certain characteristics are present or absent). Often, we need to derive such binary indicators from numeric or textual information for which we need to define cut-points to distinguish risks. To establish a threshold, both analytical data and domain-specific knowledge are needed.

First, descriptive statistics can help in understanding the distribution of variables, such as their mean, median, and standard deviation, which indicates the dispersion of observations around average values. This helps in identifying extreme cases (if any) or whether anything is odd about a distribution in general. For example, for analysing politicians' asset declarations, a simple distribution of values by year can show some surprising increases, signalling sudden enrichment. Similarly, the distribution can show if these enrichments relate to certain parties.

Second, analytical decisions must be based on existing research and data. For example, to establish a threshold for the level of competitiveness for a given tender, one needs to know the average number of bidders in the same region, or in countries with similar legislation and institutions. Tax exemptions might only be a by-product of an international competition amongst neighbouring jurisdictions for companies to establish factories in their own –hence not related to corruption.

Another important analytical decision is the distribution of weights for composite scores, if composite scores are constructed. Based on weights, the final score could be more sensitive to changes in one indicator than in the other, or equally represent a set of indicators. In order to make such a decision, it is important to analyse which variable contributes more to the overall level of corruption or integrity.

Establishing a valid indicator is about triangulating (or fine-tuning) its definition so that it measures the phenomenon of interest and not something else. An indicator can contain two types of measurement errors. It can either be false positive or false negative. False positives appear when the phenomena, event or condition did not actually take place, but the defined indicator value is nevertheless positive. False negatives are the opposite: when the indicator value is negative but the actual phenomena is present. Most of the methods, especially those working with predictions, allow for calculating the accuracy rate for the model. Accuracy rate calculates the percentage of correctly

predicted values out of all data points, or the total number of all false positives and false negatives. By calculating the accuracy rate, it is possible to estimate if the model is precise enough.

Overview of different methods for tailoring indicators

One of the most challenging tasks in most indicator building exercises is to accurately establish indicator cut-points, and in corruption measurement in particular. Arbitrary thresholds could significantly influence the overall assessment and the final scores or rankings.

One commonly used approach for establishing cut-points, which is typically biased, is the naive summation of expert-suggested red flags. In some cases, the information experts provide needs additional clarification of what could be considered a red flag. Very general and imprecise wording, as well as insufficient research of the issue, could lead to generic and vague definitions. For example, in the 2007 Organisation for Economic Co-operation and Development (OECD) report *Integrity of Public Procurement*, one of the risks listed for the pre-bidding stage is 'poor procurement planning'. It is unclear how to measure this planning and what exactly a procurement plan is. This makes further analysis difficult.

Fixation on selected indicators, while ignoring the whole scope of risk factors is another common practice that is often similarly biased. In this case, the results of the analysis only cover one side of corruption practices, without evaluating the full picture. In order to avoid such bias, additional research needs to analyse the most common red flags in the field. For example, focusing on the level of competitiveness in public procurement without considering factors such as information availability, estimated value price, or number of open procedures, will lead to biased results.

Advanced statistical and data science methods can be used for evaluation if the indicators are precisely formulated, tailored and valid. For an overview, see the box below.

STEP 10 – Advanced statistical methods

Most quantitative research on corruption uses **regression analysis**. It can assign component weights to corruption risk indicators. For example, in Fazekas et al. (2013) a set of regressions directly models corrupt rent extraction in public procurement. After running many regression

models the most significant and powerful predictors identified. Then they were included into the composite risk index for each country.

Other commonly used methods are **principal component analysis (PCA)** and **structural equation modelling (SEM)**. PCA helps to reduce large-N datasets, relying on correlation among indicators. In other words, if there are hundreds of variables in the dataset, PCA helps remove surplus factors. This is particularly useful for visualisation purposes. More importantly, PCA reveals clearly correlated variables even if, theoretically, there was a clear distinction between them. For example, the European Commission (EC)'s Working Paper "Assessing the Quality of Government at the Regional Level Using Public Procurement Data" (Fazekas 2017) presents PCA, showing that indicators such as corruption, competition, transparency and efficiency are closely correlated.

Structural Equation Modelling (SEM) is a complex methodology combining factor analysis and multiple regression. Its particular usefulness stems from its ability to identify latent factors (i.e. corruption) based on a number of proxies (e.g. red flags). Also, it helps estimate the relationships between various latent and other factors (e.g. the impact of civil servants' salary on public procurement corruption).

Finally, **machine learning (ML)** could work as a useful analytical tool for large-N datasets and for prediction purposes. ML methods are divided into supervised, semi-supervised and unsupervised, depending on whether data are labelled or not. In supervised methods, the model aims at finding relationships between dependent and independent variables, or classifying data into certain categories based on given conditions. Unsupervised methods work with unlabelled data to either cluster observations based on similarities, or find association between variables. Finally, semi-supervised methods usually have both labelled and unlabelled data aimed at using existing information to relabel the rest of the observations and improve accuracy. ML often helps predict corruption indexes for unlabelled procurement risk data (see Rabuzin and Modrušan, 2019).

The main challenge for developing and further tailoring indicators is taking into consideration both proven empirical cases of fraudulent practices, as well as theoretical understanding of the corruption process.

PHASE 4: DOCUMENTING YOUR RESULTS

Step 11: Documenting the final conceptually valid and empirically feasible indicators

In order to standardise the process, it is useful to use the following categories for documenting as a template for documenting political integrity risk indicators:

1. Indicator group (if relevant) – In some cases it is more convenient to combine indicators related to the same topic. For example, in the Global Data Barometer, some of the indicators with the same name refer to different groups (public consultation data is covered in both governance and availability sections).
2. Indicator name – This one should directly address the observable phenomena.
3. Precise description of indicator – The phenomena that should be captured by this indicator.
4. Short description of corruption scheme – Should describe association between corruption and proposed indicator, revealing the causal relationship.
5. Indicator scale and direction – E.g. percentage of transactions, higher values indicate higher corruption risk, etc.
6. indicator level of observation – E.g. company or contract.

APPENDIX 1

DATA STRUCTURE FOR A SELECTION OF POLITICAL INTEGRITY DATASETS

Political Finance

The tables below describe a hypothetical data structure for storing political financing data. The example shows a data structure where different data domains (e.g. data on parties/candidates/details about the donation itself) could be stored in separate tables. This would eventually allow for reviewing the relevant sets of variables required for specific analytical questions. Thinking about political donations data in such a way also helps to illuminate the full scope of potentially important variables needed for risk assessment for each domain (donation/donor/candidate/party etc.).

While political financing can entail both public and private funding, the example provided only considers private donations. One of the main data tables [donations] could store variables describing a single donation, including identifiers for the specific donation itself, the donor, also the unique identifiers of the party, and, if relevant, the candidate who received the donation. It can store data on the donation type – whether it is coming from a company or an individual – the date the donation was made, or if the donation was above a certain mandatory threshold.

The [donor] variables store data on companies or individual donors. Their unique identifier makes it possible to connect donors to multiple donations; hence individual- or company-level donations can be tracked over time, while other details, such as whether the donor was foreign, make it possible to analyse donations data from potentially relevant angles, from a risk assessment perspective.

The [party] and [candidate] variables store information on donation recipients, such as their name, the date the party was established, or the name and party affiliation of the candidates. Note that individual donations are connected with unique identifiers to their donors, parties and candidates.

Political Finance Data Structure

Donations	Type	Description
lid (donation)	string	unique identifier of the donation
id (donor)	string	unique identifier of the donor
id (party)	string	unique identifier of the party that received the donation
id (candidate)	string	unique identifier of the candidate who received the donation
donation type	enum	Type of donation. Categorical variable, with fields such as company/individual.
mandatory	boolean	donation is above a mandatory threshold for reporting
donation amount	num	amount of the donation
donation currency	enum	currency of the donation
donation date	date	date of the donation

Donor	Type	Description
id (donor)	string	unique identifier of the donor
name	string	name of the donor
donor type	enum	Type of donation. Categorical variable, with fields such as company/individual.
foreign	boolean	donor is of foreign origin
address	string	address of the donor
Party	Type	Description
id (party id)	string	unique identifier of the party
name	string	name of the party
establishment date	date	date of establishment
Candidate	Type	Description
id (candidate)	string	unique identifier of the candidate
first name	string	first name of the candidate
last name	string	last name of the candidate
party affiliations	array	object – array
Candidates' party affiliations	Type	Description
id (candidate)	string	unique identifier of the candidate
party name	string	unique identifier of party the candidate is affiliated with
party id (party id)	string	unique identifier of the party
position	string	candidate's position in the party
position start date	date	start date of the given position in the party
position end date	date	end date of the given position in the party
party affiliation start date	date	start date of the affiliation with the party
party affiliation end date	date	end date of the affiliation with the party

Asset and interest disclosure

The tables below describe a hypothetical data structure for storing assets and interest disclosure data.

Asset and Interest Disclosure Data Structure

Public officials	Type	Description
id (official)	string	unique identifier of the public official
name	string	name of the public official
surname	string	surname of the public official
Income		(Can be multiplied in case of multiple incomes)
id	string	unique identifier of the income
income source	enum	source of the income (e.g. official salary as MP, capital income, honorarium)
gross amount	num	yearly gross amount from a specific income type
currency	enum	original currency of the income
Assets		(Can be multiplied in case of multiple assets)
id (asset)	string	unique identifier of the asset
asset type	enum	type of asset (e.g. real estate, vehicle etc.)
asset value	num	estimated value of the asset
date of acquisition	date	date the asset was acquired
Loans		(Can be multiplied in case of multiple loans)
id (loan)	string	unique identifier of the loan
loan type	enum	bank/individual
loan amount	num	amount of the loan
date	date	start date of the loan
Family members		
id (official)	string	unique identifier of the public official
name	string	name of the family member
surname	string	surname of the family member
[income, asset, loan of the family member]		

Lobbying

The tables below describe a hypothetical data structure for storing lobbying data.

Lobbying Activities Data Structure

Party	Type	Description
id (party id)	string	unique identifier of the party
name	string	name of the party
establishment date	date	date of establishment

Politician

id (politician id)	string	unique identifier of the politician
name	string	name of the party
affiliation	string	the department/party the politician represents

Lobbying group

id (lobbying group)	string	unique identifier of the lobbying group
company	string	the company represented by the group
sector	string	the sector represented by the group
establishment date	date	date of establishment
foreign	boolean	lobbyist is of foreign origin

Meeting

id (meeting	string	unique identifier of a lobbying meeting
id (group)	string	unique identifier of the lobbying group
id (politician)	list	unique identifier of the politician
id (party)	string	unique identifier of the party
description	string	description of the meeting if any (e.g. agenda)
date	date	date of the meeting
length	number	length of the meeting
address	string	address of the meeting

APPENDIX 2

The following table presents examples of risk indicators from nine countries. Each example presents a brief description of the corruption scheme and the variables from different datasets. The examples resulted from practical exercises by participants of the III Political Integrity Bootcamp: *Detecting Political Corruption Risks*, held online in October 2021. The Bootcamp is a collaboration between the Office for Democratic Institutions and Human Rights of the Organization for Security and Cooperation in Europe (OSCE/ODIHR), the Transparency International (TI) Secretariat, and TI national chapters from Europe and the Central Asia region. In the 2021 edition, chapters from the Americas, Africa and Asia participated as guests.

Country	Indicator	Scheme	Variable(s)
Latvia - Estonia	Seller made multiple donations to buyer	Buyers with a political party affiliation may be using public contracts to reward companies whose board members are donors to the political party connected to the buyer.	Procurement registry (seller ID, buyer ID, award date, legal representative name (buyer + seller); Business Registry (legal representative's ID code - buyer + seller); Political party donations registry (name, ID code, party name, date, sum)
Italy	Political favouritism in subnational public procurement	Regional administrations unfairly use procurement procedures to grant disproportionately larger shares of public funds to politically connected suppliers.	Procurement (public buyer name, public buyer unique ID, Nr participating companies per tender, supplier name, supplier identifier, date of publishing, bidding deadline, contract value, funding source); Declarations of financial interests of MPs (declarations of financial interests of local public officials [name of public official, institutional role, institution, name of companies in which he/she has an interest in, role in the company])
Portugal	MPs hold interests in entities their committee oversee	MPs use their legislative committee work to favour entities and the respective markets they hold interests in, preparing for future jobs, retirement ('golden parachutes')	MPs' interest declarations (name, identifier, main activity, current role, past roles, supports and benefits received, services provided to private entities, exclusivity); MPs' parliamentary initiatives (bill initiative number, name and description, co-sponsoring MPs,, responsible parties and current stage)
Spain	Significant increase in public subsidies/grants received by beneficiaries with a connection to new ruling party	New government(s) favour politically connected associations, businesses, etc. in the allocation of public aid and subsidies rewarding (e.g.) a political donation, or connection between their board members and the ruling party	Lists of large beneficiaries 2020, 2019, 2018 and 2017 (i. beneficiaries' names; ii. identity number; iii. period of time and iv. amount of aid accumulated.); Declarations of economic interests of MPs (i. MP name, district elected, prior activity position, prior activity employer, prior activity date, donations made beneficiary name, donations made amount, donations made date)

Country	Indicator	Scheme	Variable(s)
Slovenia	Donation of gifts to influence resource allocation	Gaining favoured access to decision makers and public procurement through gifts (the gift being the material inducement to abuse office by providing favoured treatment to the gift-giver)	Gifts registry (date of gift, name of recipient agency, status of recipient, type, mode of delivery, description, value, reasons, owner); Company Register (registration ID, full company name); Public procurement (buyer ID, legal name, postcode; bidder ID, legal name, postcode, amount EUR, award criteria details).
Hungary	Winning suppliers with poor reputation	Winning suppliers whose past activities were (repeatedly) reported to be questionable in the press increases risk of future wrongdoings and ultimately leads to the distortion of competition and public trust	Procurement (bidding winner company name, company identifier, contract award publication date, contract/tender identifier); Press articles about corruption cases (company tagged, date tagged)
Ghana	Single-source tendering contract disclosure	Low levels of transparency could hide clientelism in tendering practices	Procurement: (tender unique ID, supplier name, supplier ID, procuring entity name, name of procuring entity head, ID procuring entity head, public contract identifier, number of times the supplier wins contracts, awarded contract value, size of the MMDA); Political party membership (party name, ID member; name member)
Ecuador	Buyer capture	High-ranking public officials have effective control of companies and are favoured in the assignments of contracts (2021 onwards)	Registry of companies of the SUPERCIA (company name, company ID, has high-ranking public official as an owner); Public procurement platform (buyer name, buyer ID, contract ID, contract value); Beneficial Owner platform – public procurement (company name, contract ID, has high-ranking public official as a beneficial owner)
Indonesia	Buyer capture	Share of total procurement spending awarded to politically connected companies per buyer	Public procurement dataset (buyer's total spending in a given fiscal year, company total revenue from public contracts in a given year); Interest and asset declaration dataset (high public official name, company shares held); company registry (company name; company ID, majority shareholders name, director name, commissioner name)

REFERENCES

- Brandt, Urs Steiner, and Gert Tinggaard Svendsen. "Why does bureaucratic corruption occur in the EU?." Public Choice 157.3-4 (2013): 585-599
- Djankov, Simeon, et al. "Disclosure by politicians." American Economic Journal: Applied Economics 2.2 (2010): 179-209.
- Evertsson, Nubia. "Political corruption and electoral funding: a cross-national analysis." International Criminal Justice Review 23.1 (2013): 75-94
- Fazekas, M., Cingolani, L., and B. Tóth, (2016) 'A Comprehensive Review of Objective Corruption Proxies in Public Procurement: Risky Actors, Transactions, and Vehicles of Rent Extraction', Government Transparency Institute, Working Paper Series GTI-WP/2016:03
- Fazekas, Mihály, Tóth, István János, and King, Peter Lawrence, (2016), [An Objective Corruption Risk Index Using Public Procurement Data](#). European Journal of Criminal Policy and Research, 22.
- 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, Romain Ferrali, and Johannes Wachs, (2018), Institutional quality, campaign contributions, and favouritism in US federal government contracting. Working Paper series No. GTI-WP/2018: 01. Budapest.
- Fazekas, Mihály and Cingolani, Luciana, (2017), [Breaking the cycle? How \(not\) to use political finance regulations to counter public procurement corruption](#). Slavonic & East European Review, 95(1).
- Johnston, M. (1996). The search for definitions: the vitality of politics and the issue of corruption. International Social Science Journal, 48(149), 321–335.
- Mironov, Maxim, and Ekaterina Zhuravskaya. "Corruption in procurement and the political cycle in tunneling: Evidence from financial transactions data." American Economic Journal: Economic Policy 8.2 (2016): 287-321
- Mungiu-Pippidi, Alina & Fazekas, Mihály (2020), [How to define and measure corruption](#). In Alina Mungiu-Pippidi & Paul M. Heywood (eds.) A Research Agenda for Studies of Corruption. Ch. 2. Edward Elgar, Cheltenham.
- Noutcheva, Gergana, and Senem Aydin-Düzgit. "Lost in Europeanisation: The Western Balkans and Turkey." West European Politics 35.1 (2012): 59-78
- UNDP. User's Guide to measure corruption and anti-corruption. URL: <https://www.undp.org/content/dam/undp/library/Democratic%20Governance/Anti-corruption/Users-Guide-Measuring-Corruption-Anticorruption.pdf>
- OECD. Preventing policy capture: Integrity in Public Decision Making, <https://www.oecd.org/corruption/preventing-policy-capture-9789264065239-en.htm>
- Khan, Mushtaq. "[Governance and anti-corruption reforms in developing countries: Policies, evidence and ways forward](#)." (2006).
- Perumal, Kannan. "State capacity, drivers of corruption and anti-corruption interventions: Contextualizing corruption in Indian context—A rational choice approach." International Journal of Development Research 9.11 (2019): 3323-3330
- Trapnell, Stephanie (2015) User's Guide to Measuring Corruption and Anti-corruption. UNDP, New York, US.
- Transparency International (2020)
- "Political Integrity", <https://www.transparency.org/en/our-priorities/political-integrity>

"Debugging Democracy. Open Data for Political Integrity in Europe",
<https://www.transparency.org/en/publications/debugging-democracy-open-data-for-political-integrity-in-europe>

Warren, M. E. (2003). What Does Corruption Mean in a Democracy? *American Journal of Political Science*, 48(2), 328–343.