Iceland
Liechtenstein
Norway grants

Norway
grants

Analyzing Public Procurement Risks

| Training manual

The R2G4P project, coordinated by the Center for the Study of Democracy, Bulgaria benefits from a € 1.5 million grant from Iceland, Liechtenstein and Norway through the EEA and Norway Grants Fund for Regional Cooperation. The aim of the project is to implement shared anti-corruption and good governance solutions in Southeast Europe through innovative practices and public-private partnerships.

Authors:
Dr. Mihály Fazekas, Director, Government Transparency Institute
Péter Horn, Analyst, Government Transparency Institute
Ágnes Czibik, Managing Director, Government Transparency Institute
Bence Tóth, Senior Analyst, Government Transparency Institute

implemented by

CSD CENTER FOR THE STUDY OF DEMOCRACY
SELDI.net Southeast European Leadership for Development and Integrity
REGIONAL ANTI-CORRUPTION INITIATIVE
ACER Albanian Center for Economic Research
SCPC STATE COMMISSION FOR PREVENTION OF CORRUPTION
sar when ideas do work
CMI U4 ANTI-CORRUPTION RESOURCE CENTRE
RegionalCooperationCouncil
Government Transparency Institute
Institute of Public Finance
IDSCS Institute for Democracy Societas Civilis -Skopje
ZaštoNe?
CENTER FOR DEMOCRATIC TRANSITION
EUROPEAN POLICY CENTRE

CONTENTS

Introduction

The training manual was created as a part of the Regional Good Governance Public-Private Partnership Platform (R2G4P), which aims to build a sustainable regional public-private partnership for shared good governance solutions. The main purpose of the manual is to introduce how data analytics can encourage good governance practices by highlighting the weaknesses of public procurement systems and by supporting independent corruption risk research.

To achieve this, the document gives a step-by-step introduction to the analytical process of large-scale public procurement datasets. First, it briefly introduces the conceptual background of a generic public procurement process, while also highlighting its potential corruption risks. Second, it presents the most important aspects of creating an appropriate dataset for quantitative research. Finally, it presents a group of rigorously tested and validated Corruption Risk Indicators (CRIs), created by the Government Transparency Institute (GTI), that are equipped to measure the corruption risks of public procurements. Additionally, the manual also introduces the Opentedner website, which was created to provide comprehensive public procurement information free of charge in an easy-to-use format.

The manual can assist R2G4P partner institutions as well as independent researchers to successfully execute large-scale quantitative research. Furthermore, it provides thorough information on the availability of complete data sets and risk indicators that can be accessed and used by every interested party.

1. Conceptual background

Data analysis can support and improve public procurement in a number of ways that can be sorted into two broad categories. The first category is investigation support on the contract, organization or market levels. From this aspect, it can initiate investigations by flagging suspicious cases (initiation). Moreover, given scarce resources, it can also support the selection process by ranking known cases by their severity (selection). Finally, data analytics can help conduct in-depth research (e.g. network analysis) for more complex corruption cases (conduct).

The second category is policy reform and evaluation support. It materializes in the exploration of how a new procurement law could change corruption risk levels in the system (systemic). Furthermore, it can also advocate new regulations by simulating how a regulatory change (e.g. publication threshold modification) would affect the procurement market (regulatory). Lastly, it is equipped to measure the effectiveness of current organization level procurement rules (organizational).

Before any data analytic tools could be applied however, it is important to outline 1) what is considered corruption from the public procurement perspective, 2) what are the steps of the public procurement process that will be evaluated, 3) which are the most vulnerable points of this process, and 4) which are some of the more often used corrupt schemes that should be analyzed. The first section of the manual tries to give a general answer to these questions by introducing the main steps of the public procurement process, outlining a possible corruption measurement approach and identifying some of the most common corrupt schemes.

1.1. Introduction to the public procurement process

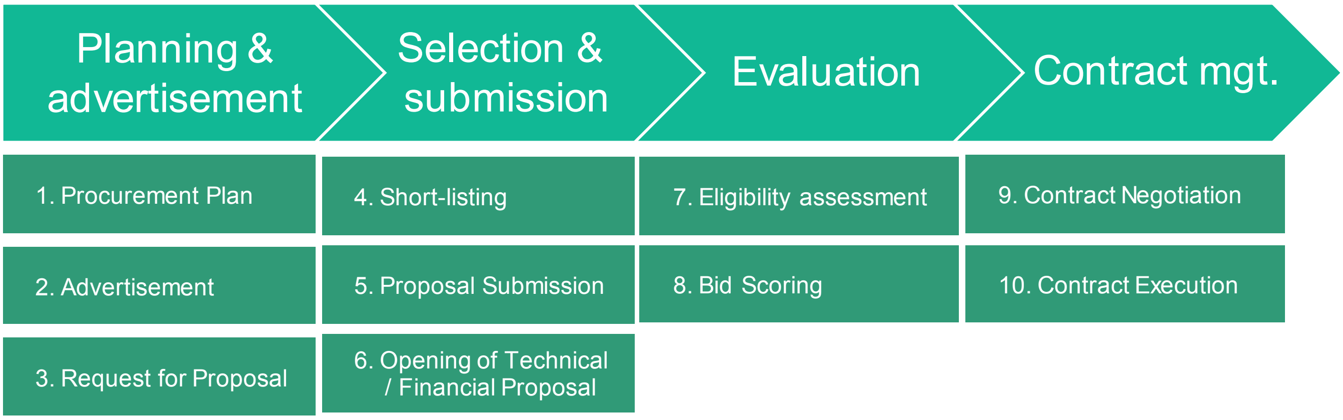
The Planning & advertisement phase of the procurement procedure begins with the planning process, during which the contracting authority decides on the specifications of its purchase. This, among others, includes the initial price and volume estimates, the planning of the implementation timeline, the description of the subject matter of the procurement, the decision on the number of lots and - if relevant - the decision on the location of implementation. After the tender documentation is finished, the authority decides the method of advertisement (e.g. e-auction, non-electronic procurement) and the request for proposals begins.

The above process is followed by the Selection & submission phase when bidders submit their initial expression of interest, which is used by the contracting authority to pre-screen bidders. Detailed technical and financial specifications are also made available, which usually further narrows the pool of eligible suppliers. In this stage bidders usually have the opportunity to ask questions regarding the procurement, and - to some extent - they can rectificate, or modify their proposals.

The Evaluation process starts with the minimum eligibility assessments, the goal of which is to filter down the list of bidders to the ones that are - on paper - meet the requirements listed in the tender documentation. Then the proposals are evaluated, compared and scored and the winning bidder(s) is selected.

The final stage consists of further contract negotiation between the contractor and the supplier, the checking of payments and deliverables, the execution of the project, and the possible renegotiation of the contract. Each of these steps has its own corruption risks, which will be discussed in the later part of this chapter.

Figure 1: Flow chart of the public procurement process



Source: Adapted from IMPPM 2017-Uni Roma Tor Vergata. Integrity module (Agerskov, Fazekas, Piga)

1.2. Steps of the measurement approach

Before initiating quantitative research, the analyst should outline the approach that is to be used to measure the presence and intensity of corruption in public procurement. The necessary components that should be specified at this stage are the following:

1. The specific definition of corruption: to measure corruption it is important to specify what is to be measured. A sufficiently specific definition of corruption should be used, that can be quantified and objectively evaluated.

2. The dictionary of corruption technologies: a reasonably broad repository of corruption strategies that are regularly being used in the procurement system. Outlining these strategies help the formulation of quantitative tools.

3. The target population and sample: the scale and scope of the procurement dataset, and level of the observations (e.g. tender, contract, or lot level dataset)

4. The tailoring and validation steps of quantifiable corruption risk indicators: the selection and validation of quantitative indicators that can correctly detect corruption technologies in the public procurement framework. Validation is an essential step to ratify that the indicator is an adequate measure of corruption. An inadequate measure can introduce measurement bias and can create a false picture about the corruption risk of the procurement system.

1.2.1. Defining corruption in public procurement

A widely used definition of corruption by Nye (1967) states that „public corruption is the abuse of entrusted public authority for undue private interest"[1] . This definition assumes that universal interest is enshrined, hence public money should serve public interest. Consequently, deviation from this should be sanctioned as corruption leads to a social and economic loss.

The issue with this definition is that it is too broad, therefore it is not adequate to quantitatively measure corruption in a public procurement setting. A more specific definition could be used in this analytical framework is that „in public procurement, the aim of corruption is to steer the contract to the favored bidder without detection"[2] . This is done in several ways, including:

- Avoiding competition through, e.g., unjustified sole sourcing or direct contract awards.

- Favoring a certain bidder by tailoring specifications, sharing inside information etc.

The main assumption of this definition is that the procurement process has some form of limitation in place to avoid competition, and this limitation is used to favor a certain bidder. If the second half of the assumption is not applicable, it suggests that there is some other issue with the procurement process (e.g., the incompetence of the contractor), hence corruption is not intended, and the problem requires a different policy approach. Furthermore, corruption must entail a cooperation between a public and a private body, therefore it is not identical to collusion in which case suppliers form a cartel to split a market to prevent competition. Nor is there a necessity for infringement, as rules can be bent to allow corrupt practices, which also indicates that corruption has not to be sanctionable.

Therefore, In the current setting corruption is particularistic, institutionalized and grand. There is a particularistic (often personal) relationship between the actors involved in corruption and they use this particularistic tie to exclude anyone who is not part of their interest group. It is institutionalized indicating that it is recurrent, stable and systemic. Finally, it includes high-level politicians and business persons, hence it usually involves a large amount of public funds.

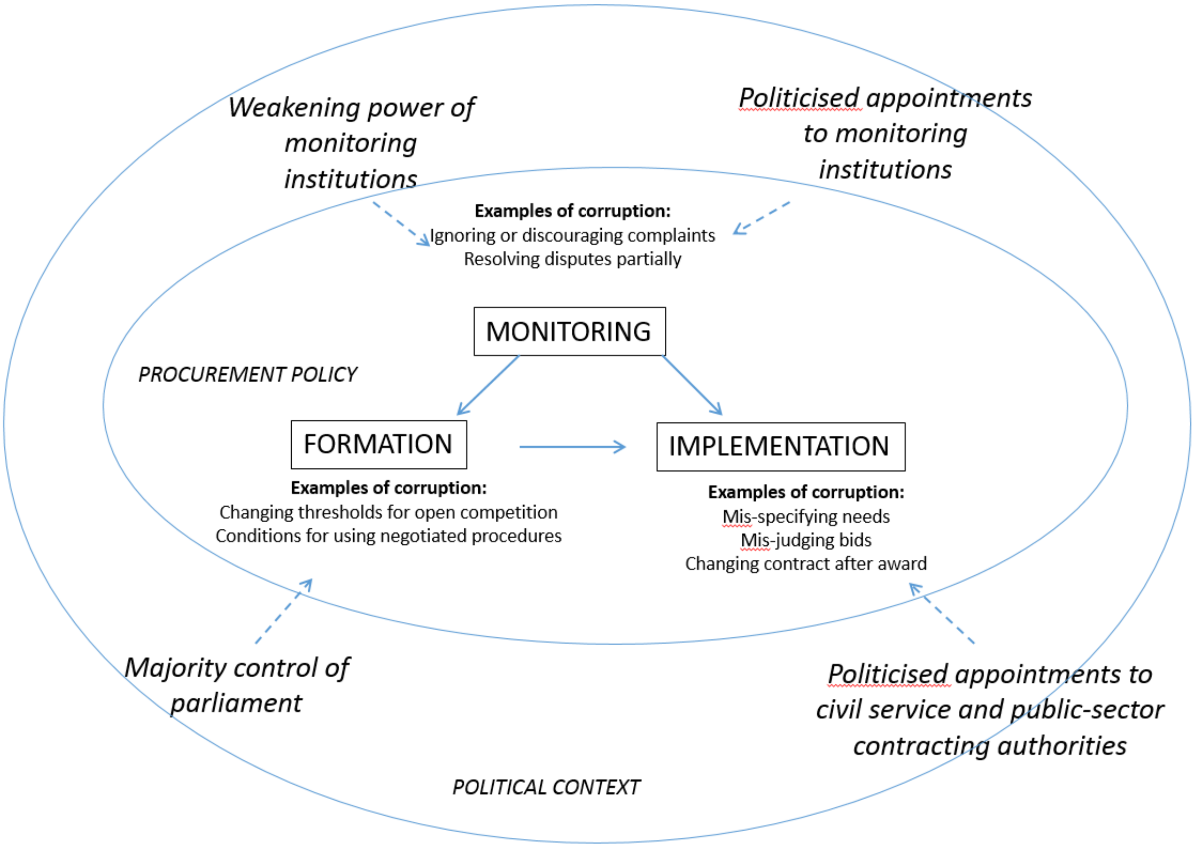
[1] Nye (1967)
[2] World Bank Integrity Presidency (2009)

1.2.2. Corruption risks in the procurement process

The next sections of the manual will mainly focus on corruption risks during the procurement implementation process; however it is important to mention that corruption can already occur during the policy formation and monitoring stages. Adjusting procurement regulation can be an efficient way of limiting competition with the added benefit that economic actors do not have to break any rules in order to take advantage of their restricted access to procurements. For example, persuading politicians to lower thresholds for restricted procedures or to create special conditions for using negotiated ones exhausts the definition of institutionalized corruption which can only be dealt with at the highest levels.

Regulatory entities and auditing bodies can also be corrupted, who may ignore complaints, partially settle disputes, or ignore the particularistic relationships between individual contractors and bidders. Corrupting the monitoring stage is also handy to create paper trails suggesting that everything went well during the procurement process. Although this form of corruption already requires breaking the law, it is often very difficult to detect, especially when the public procurement system is fraught with red tape.

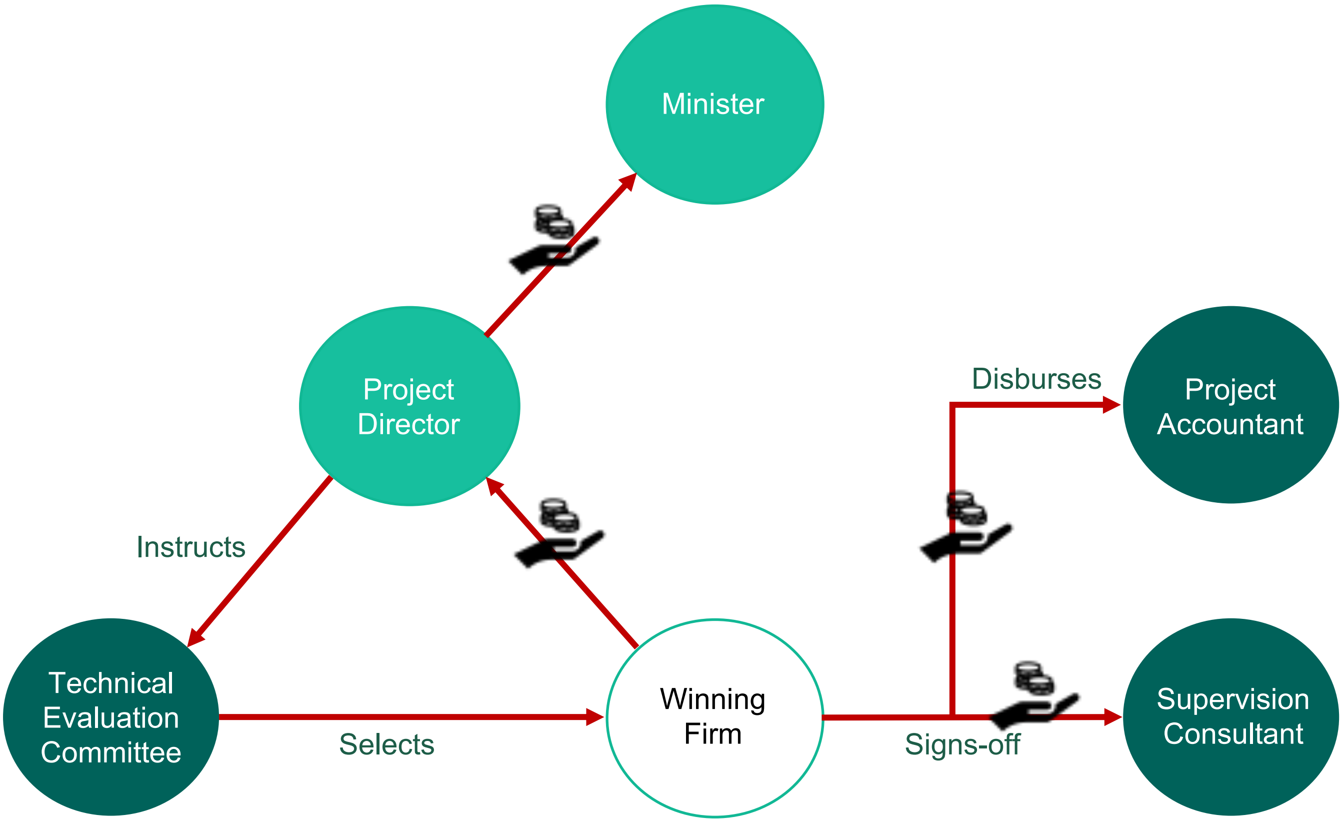
Figure 2: Stages of the procurement process at which corruption can occur and modes of political influence over process


Weakening power of monitoring institutions
Politicised appointments to monitoring institutions
Examples of corruption:
Ignoring or discouraging complaints
Resolving disputes partially
MONITORING
PROCUREMENT POLICY
FORMATION
IMPLEMENTATION
Examples of corruption:
Changing thresholds for open competition
Conditions for using negotiated procedures
Examples of corruption:
Mis-specifying needs
Mis-judging bids
Changing contract after award
Majority control of parliament
Politicised appointments to civil service and public-sector contracting authorities
POLITICAL CONTEXT

Source: Dávid-Barrett - Fazekas, 2016

During the implementation phase the contracting authority can specify needs to favor a certain bidder, it can misjudge the quality of other bids to suppress competition, or it can change the contract after the winner has been selected. In an entrenched system of corruption, a typical corrupt scheme might include all the stages of the procurement process. It can start by the supplier contacting a procurement project director (e.g., through a particularistic tie) who is responsible for the management of some high-value projects. Then the director can instruct the technical evaluation committee to favor the bidder by, for example, overestimating its capabilities. In return, the supplier pays the project director and usually - as we defined corruption as being high-level (grand) and institutionalized - a high-ranking politician (such as a minister) who turns a blind eye on the process. It is important to mention that the payment doesn't have to be a bribe, rather a payment for a "consulting contract", or - in a well-oiled system - it can even take the form of a personal favor. Finally, the winning firm can also pay the accountant to sign-off the contract, and at the end of the implementation phase it might bribe the supervisor who evaluates the quality of the output.

Figure 3: A typical corruption scheme



Source: GTI

1.3. Identifying popular corrupt schemes

After mapping the procurement process and specifying the corruption definition that is to be used for the analysis, it is important to identify popular techniques that are being used to corrupt the procurement system. Finding well-documented examples of high-level corruption cases can help in the selection and formulation of objective and quantitative indicators. Therefore, the last part of this chapter outlines some of these techniques, while also highlighting the importance of substantive qualitative research. The next chapters will focus on how to collect, clean, and analyze data and how to create numeric indicators to reveal corruption in quantitative research.

1. Tinkering with advertisement period length

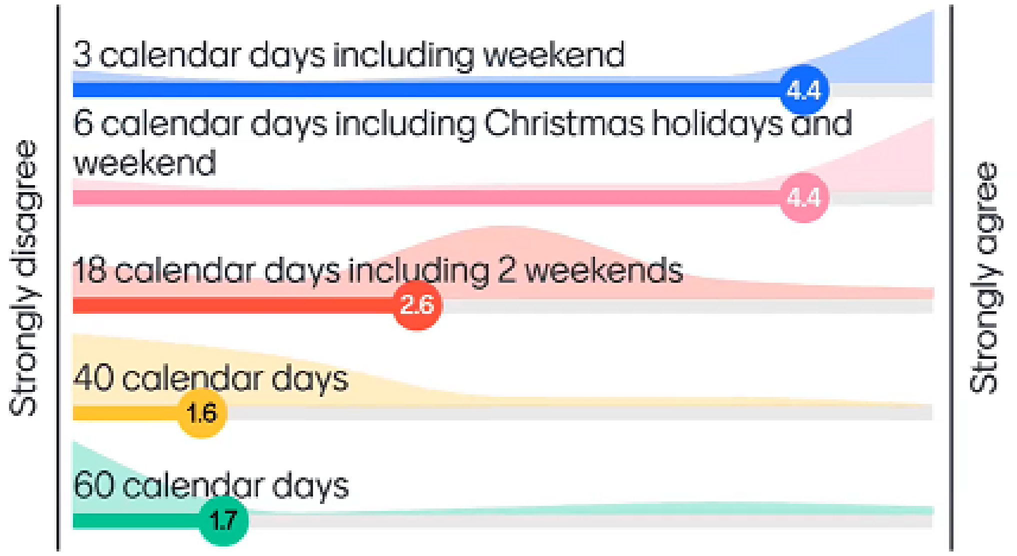
Most of the more developed procurement systems have a minimum advertisement period length limit in place, however, contractors might be able to use policy loopholes to tinker with the length of this period. A sufficiently short duration makes competition impossible, because competitors will not have the time to obtain necessary documents, prepare the tender documentation, or to calculate their expenses. Therefore, if a favored bidder receives insider information about the tender before it is advertised, it will most likely be the only one able to submit its expression of interest in time.

Imagine, for example, a road reconstruction project between two medium-sized cities where the winning bidder has to repair a 25 km, two-lane road. Before reading any further, let's think about what a long enough advertisement period would be based on the table below.

Table 1: Sufficiently long advertisement period for road reconstruction project

1.	3 calendar days including weekend
2.	6 calendar days including Christmas holidays and weekend
3.	18 calendar days including 2 weekends
4.	40 calendar days
5.	60 calendar days

According to the participants of the "First specialized regional training for the Regional Good Governance Public-Private Partnership Platform", both 3 and 6 calendar days, including holidays are insufficiently short advertisement periods, while 40 and 60 calendar days should be enough to prepare all the necessary tender documentations. However, opinions about the sufficiency of a 18-day advertisement period already varied even among experts; little more than half of them thinking that it is adequately long. This highlights that a "sufficiently long" advertisement period can diverge across sectors and countries. In a country with a low level of red tape and a well-developed e-governance system 18 days could be enough to obtain all the necessary documents, but in other, more bureaucratic systems, even up to 40 days may be too short. As the next chapters will show, data analytics can help to decide upon thresholds below which procurements should be considered risky.



2. *Biased specifications*

Biased specifications are technical tender specifications that deliberately exclude other valid options. The objective of this process is to narrow down the pool of firms that can bid or the products that can be purchased. Such as the advertisement period, biased specifications are useful to eliminate competition by specifying technical requirements in a way so that only one (or a handful) company is eligible to bid. For example, a railway station construction procurement project that requires 25 years of relevant experience probably won't have many contenders in Hungary. Even if there are more than one competitor, 'relevant experience' can be interpreted to exclude other suitable suppliers (e.g, a company that has 40 years of experience in building large bus stations could be excluded because it is not exactly matching the requirements for the tender).

Furthermore, biased specifications can also be used to purchase products for personal use. For example, procurements ordering blood gas analyzers with OLED-screens, or purchasing trucks with electric engines and leather seats could indicate that the public money is spent to make the management's life easier.

3. Shell company

Shell companies are firms that exist only on paper; they often have no activities and staff except for a formal manager and owner. These companies are often registered in tax havens, or in countries with an opaque registry system. Nonetheless, they not only used to avoid taxation, but also to prevent conflict of interest in procurements or just to avert bad press. For example, the son of the president of the central bank winning a procurement for the furnishing of bank property could raise accusations regarding a conflict of interest and can certainly induce bad press. However, if this company is registered in an opaque system both the identity of the owner and the firm's qualifications can be hidden. If the procurement is then subcontracted to a firm with the required qualifications, the shell company's owners can pocket part of the contract price without any outsider noticing the particularistic tie between the contractor and the supplier.

4. "Bogus" subcontracting

Subcontracting can also be used the other way around. Information on subcontractors is often less transparent even in relatively developed procurement systems. Therefore, an otherwise qualified supplier is used to compete for the procurement which outsources the project to the subcontractor. The subcontractor could be registered in another country, so it could siphon out the funds and disappear without completing the project. In this setup the particularistic tie lies between the contractor, the subcontractor and the final supplier, but the additional step - of subcontracting - creates a scheme which is hard to uncover.

5. Substandard work

Substandard work means providing goods, works, or services that do not comply with the specifications stipulated in the contract. This process may include corrupt officials or could be the result of a company taking advantage of poor contract management practices. In other cases, supervisors can be bribed or coerced to sign off on substandard work. Substandard work typically becomes fraudulent, when the contractor recklessly or knowingly claims to have performed the work required to obtain payment.

This type of scheme could be prevented by comprehensive quality checks; however it is often hardly feasible. Road construction projects for example, are the hotbed for substandard procurement projects, because the quality of roads is mostly visible only after several years of usage.

All of the above examples highlight the importance of in-depth qualitative research. Although the following chapters mainly promote the importance of data analytics in public procurement research, data limitations will often prevent the identification of more complex corrupt schemes. It is important to remember that results are only as good as the data being used, hence a comprehensive analysis should both include qualitative and quantitative research.

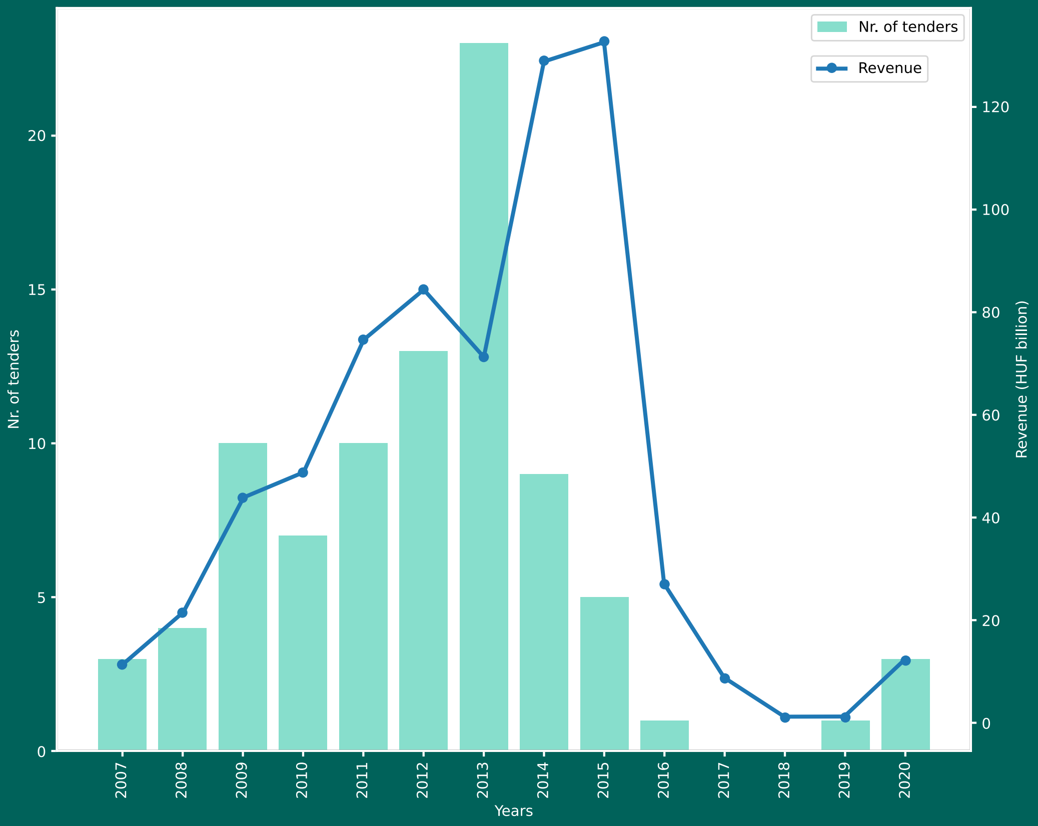
Case study: the Közgép story

The case study follows the outline of the manual, and provides a step-by-step introduction to the analytical process. For simplicity, it focuses only on an infamous Hungarian company called 'Közgép' and shows how procurement data could provide useful information to a potential investigation. It also reveals the limitations of a quantitative analysis that focuses on a single case. Later parts of the manual highlight the strength and weaknesses of data analysis in the public procurement setting.

Introduction to the case

Közgép Építő- és Fémszerkezetgyártó Zrt. is a large construction company that had close ties to right-wing politicians between the mid-2000s and 2015 (and again after 2020). By the end of the first decade of the 21st century, Közgép was a semi-established actor on the construction market, but it had become a truly indispensable player after the electoral victory of the Viktor Orbán led Fidesz party in 2010. According to Átlátszó, a Hungarian investigative journal, Közgép had already won more than HUF 200 billion (around EUR 553 million) worth of public procurement funds only two and a half years after the election. While newspapers (see sources below) reporting on the case note that the company's management correctly identified the construction market's shift towards railway and waste management projects at the time, the more likely reason behind the meteoric rise of the company could be found in its ownership structure. It had effectively hidden its ownership information between 2007 and 2012, but an official document published by Átlátszó proved that the true owner of the firm was Lajos Simincska, a close former friend of Orbán. Simicska and Orbán maintained a close relationship until their 2015 dispute, which led to the unequivocal exclusion of Közgép from any further procurement projects (see figure). The current re-emergence of Közgép on the public procurement market is presumably due to the fact that another businessman close to Fidesz purchased the company at the end of 2019. The chart below clearly depicts the importance of personal relationships involved in Közgép's success and demise.

Number of tenders won by Közgép



Source: GTI calculation based on data from Opentender & e-beszamolo.im.gov.hu

Corrupt schemes in the construction market

Investigative journals also explore how corrupt schemes were applied in the restructuring of power in the construction industry between around 2008 and 2012. Influential businessmen had already started preparing for the 2010 elections a few years earlier, as the then ruling socialist government had visibly lost its grasp on power. Relevant construction market participants started to cooperate with businesses with close ties to Fidesz - such as Közgép - by participating together in tenders as consortiums. Between 2007 and 2010 Közgép had collected all the references that were essential after the election to unsuspiciously dominate the procurement market. According to information collected by Átlátszó, K-monitor (a corruption monitoring NGO) and other newspapers, after 2010 corruption technologies were perfected to make the procurement market extremely profitable for a selected group of corporations. The most often used technique was subcontracting; since Közgép - and a few other influential businesses - often did not have the manpower required to implement all the projects it had won, the company used smaller firms to do the bulk of the work. According to anonymous sources the procurements were significantly overpriced; 60% of the price went to the subcontractors who had received near to no profits. The remaining 40% was partially sent back to the consortium leader as pure profit, while the other part was spent on political and party financing. The later part of the public funds was siphoned out of the procurement system in a form of cash transactions to hide it from the tax authorities.

The above story could be a textbook example of institutionalized and grand corruption. It demonstrates that high-level corruption involves a significant degree of pre-planning and highlights the importance of particularistic relationships. It also shows how political elites can control the flow of public funds without breaking the law. Finally, it is also a good example to illustrate how ruthless a system that relies on corruption rather than effective market forces can be.

The next framed text at the end of Chapter 2 shows how to evaluate, clean and filter procurement data to prepare it for quantitative analysis. The last text box shows how to use this data to assess the integrity of Közgép’s tenders.

Sources:

- Átlátszó, “Ez egy jól szervezett leosztási rendszer” – egy bennfentes az építőiparról, atlatszo.hu, 2012.
- Átlátszó, MagyarLeaks, Simicska Lajos a Közgép tényleges tulajdonosa, atlatszo.hu, 2012b.
- Előd, Fruzsina, Szíjj Lászlóé lett a Közgép, index.hu, 2019.
- Magyarnarancs, Közgép: túl a 200 milliárdon, magyarnarancs.hu, 2012c.
- Pető, András, Hódítók a hatalom árnyékában - a politikához kötődő Közgép felemelkedése, origo.hu, 2010.

2. Procurement data

Once the procurement process and its potential corruption risks are outlined, the next step is to decide what data will be analyzed and how to obtain it to unearth these weaknesses. This chapter sets out to answer these questions by introducing the different data types required to create an appropriate dataset for quantitative research. It also highlights the key aspects of this dataset (scope, depth, quality, access) and introduces some of the usual data errors analysts should watch out for.

2.1. Goal of creating a procurement database

It is important to keep in mind that without an appropriate dataset quantitative analysis is either completely impossible or it could provide biased results. Hence, the main objective of data collection is to create a clean and comprehensive public procurement dataset, which makes the evaluation of procurement systems integrity feasible. It requires high quality administrative data on:

- public procurement tenders and contracts,
- bidding companies (suppliers),
- awarding public organizations and
- preferably on political office holders.

This is usually a very time consuming and often expensive process, since not many procurement authorities give up their data easily (even though these datasets should be publicly available). Fortunately, there are large-scale projects that attempt to collect and maintain good quality procurement datasets from several countries. One of these is the Digiwhist initiative, a „large scale EU funded research project which simultaneously aims to increase trust in governments and improve the efficiency of public spending across Europe"[3]. It supports corruption measurement by organizing and linking complex procurement datasets and it also provides a data template to serve as the basis for collecting and republishing procurement data. Other organizations such as the Open Contracting Partnership and Transparency International also collects and aggregates procurement level information[4] .

[3] Digiwhist (s.a.)

[4] See here and here

2.2. Data types

Often used data types can be further divided into subcategories and into specific variables. It is important to outline exactly which of these variables will be needed, because obtaining, and appropriately cleaning them could be both the most expensive and most time-consuming part of the analysis.

It is no surprise that the most crucial indicators are related to the procurement process, hence it is generally a good idea to get as much information from the tendering cycle (procurement planning phase, selection/evaluation phase, implementation phase) as possible. Luckily, the above-mentioned initiatives also focus on collecting this information.

Furthermore, detailed company data can be also important, especially if the analysis has a narrow focus on specific sectors or firms. Unfortunately, corporate data is only partially public, and even the public information could be "protected" against researchers and against other curious citizens[5] . Although there are often resellers who obtain and process financial reports, these services can be quite expensive. Nevertheless, basic firm level information - such as the supplier's location, name or official id - is always necessary, and fortunately mostly available within the procurement contract.

Data on public organizations is also required to identify contracting authorities. Like basic company data, procurers' registry information is also recorded during the procurement process, therefore largely available for public use. On the other hand, authorities budget data, while available, often published on separate websites, hence their collection could be burdensome. Finally, information on the authorities' leadership can be useful to measure corruption risks, regrettably - such as budget data - it mostly has to be manually collected from the appropriate authorities' websites.

[5] For example, although financial reports are publicly available in Hungary, the website is protected by several types of CAPTCHA and reports are immutable making bulk download and processing especially difficult.

Table 2: Examples of administrative data types and variables

Public procurement data	Call for tender related information	procedure type, product code, bidding period length, bidderlimitation, estimated value, type of the contract, documentation fee, buyer, award criteria.
	Contract award related information	number of bids received, bidder and winner company related information (bid prices, location), final contract value, award signature date.
Company data	Registry information	company name, location, legal form, date of incorporation, number of employees etc.
	Financial information	annual turnover, profit rate, return on assets, material costs, personnel costs, taxes, EBITDA.
	Ownership information	number of recorded shareholders, shareholder's name, shareholder's type (legal entity, individual etc.), shareholder's location, shareholder's direct and total shares.
	Manager information	number of directors, name of company directors, position of company directors, appointment and resignation date of directors, gender, date of birth, shareholder status.
Public organization data	Registry data	name, ID, location, activity type, contact
	Budget data	annual budget figures, currency, classification of the budget item (IFRS)
Public officials' data		Name, contracting authority, position, start and end date, political affiliation

Source: GTI

2.3. Key aspects of procurement data

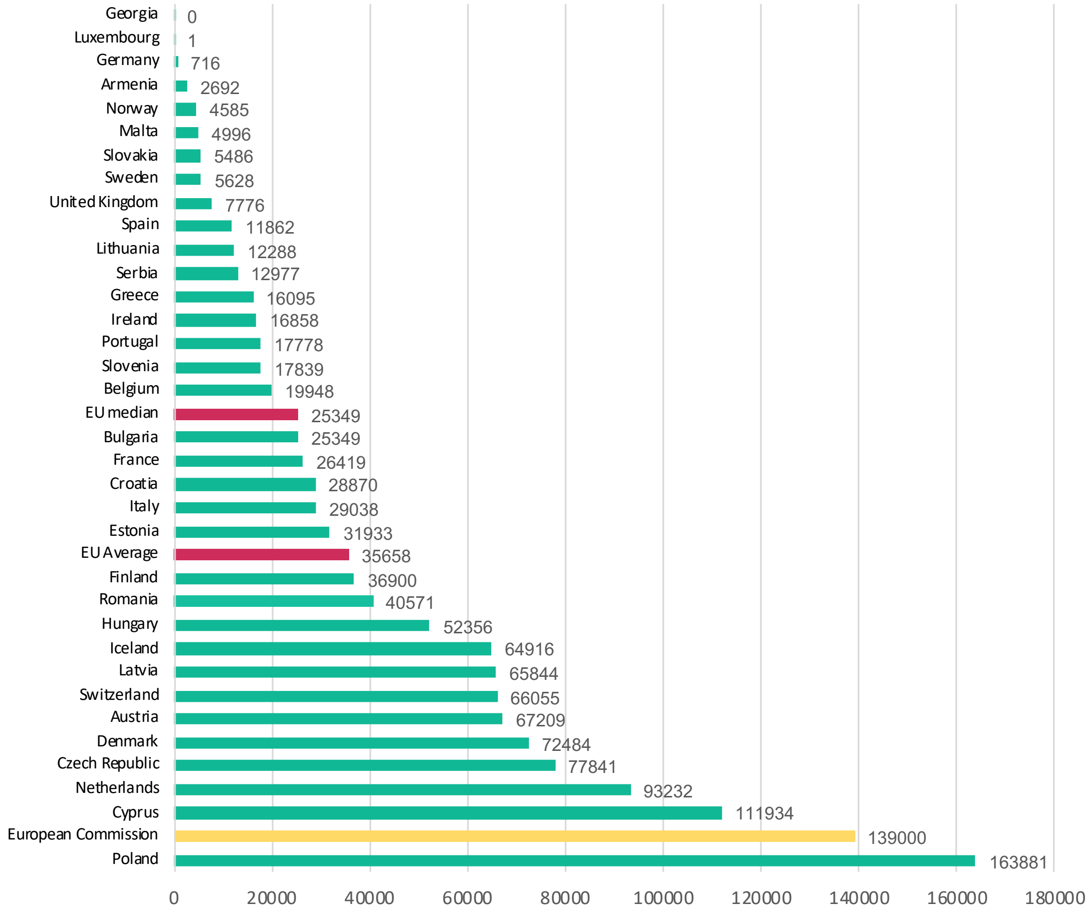
After the primary data source is established, it should be evaluated based on its scope, depth, quality and accessibility. The evaluation process is used to assess the overall adequacy of the dataset. Using this preliminary assessment, the expectations - about the data source - can be adjusted, the research limitations can be outlined and, if necessary, the scope and depth of the analysis can be modified accordingly.

2.3.1. Data scope

Data scope in the public procurement setting should be interpreted as the threshold and timeline on which procurement data is available on the tendering website. Reporting thresholds are national contract value thresholds for mandatory publication of tenders on national or EU wide portals. Thresholds vary greatly across Europe and can have different scopes and regulations attached to them (for example, in Turkey several public bodies are exempt from the threshold). The chart below shows that some countries have relatively high thresholds, while others require all contracts to be published.

Procurements over the threshold usually have to comply with stricter rules, such as minimum length of advertisement period or publication of the scoring criteria, hence lower threshold leads to more transparency. Furthermore, in countries with lower thresholds larger portions of the procurement market can be analyzed leading to a more accurate analysis. It is always important to take the scope of the analyzed data into consideration when forming policy recommendations.

Figure 5: Scope of public procurement databases - Minimum contract value for publishing supplies and services contracts (EUR, 2020, PPP)[6]



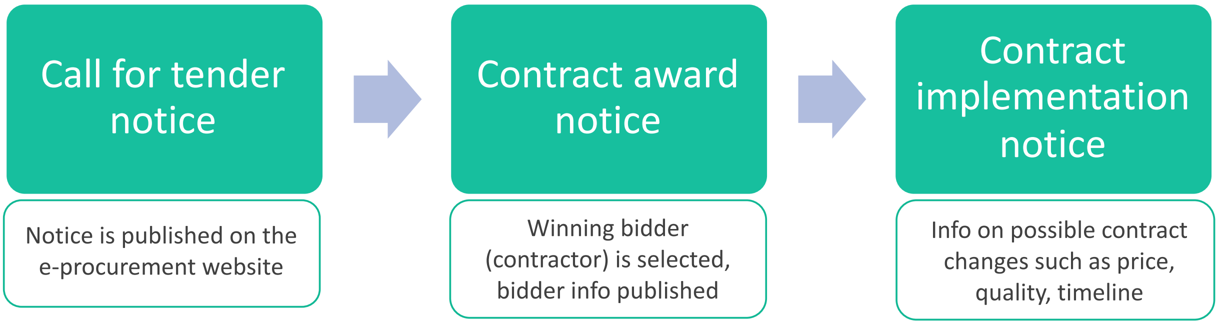
Source: EuroPam (2020)

[6] The European Commission is currently suing (since 2019) Poland for breaching public procurement law.

2.3.2. Data depth

Data depth includes tender cycle coverage and indicator-level availability. The tender cycle consists of the call for tender publication, the contract award publication and the implementation phases as shown on the chart below (a more general version of Figure 1.). Contract award related information is available for all contracts above the reporting threshold for all types of procurements, while call for tender notices are only available for projects with non-restricted procedure types, hence the latter dataset is usually a subset of the former.

Figure 6: Tender cycle

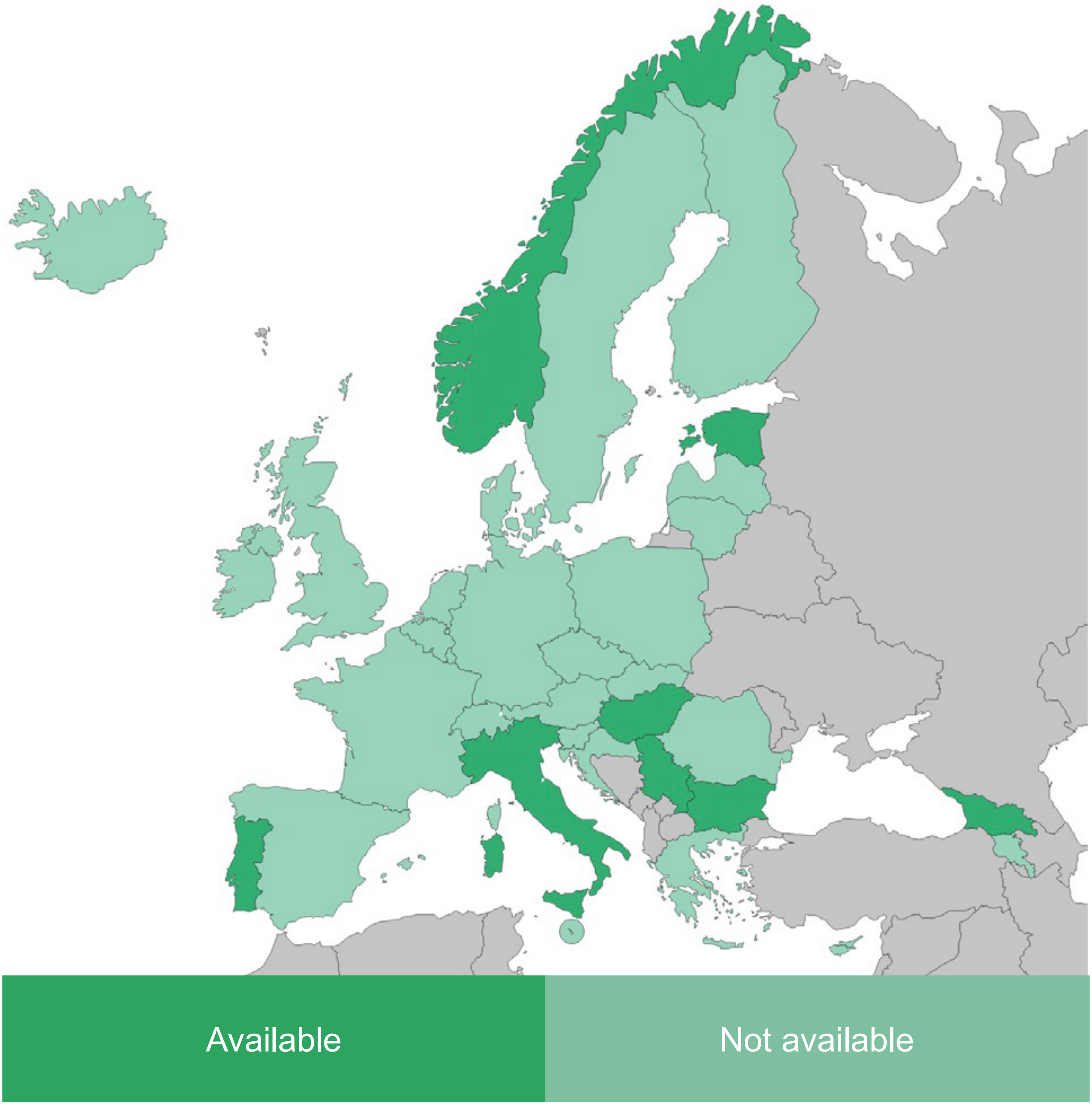

Call for tender notice
Notice is published on the e-procurement website
Contract award notice
Winning bidder (contractor) is selected, bidder info published
Contract implementation notice
Info on possible contract changes such as price, quality, timeline

Source: GTI

There are relevant differences in the tender cycle coverage across countries. Most of the procurement systems in Europe only cover the advertising and the awarding phases; only a handful of EU countries' procurement systems disclose information on implementation (see the next figure). No information on the implementation phase can give a false picture about the quality of procurement, for example, if the contract is modified or the work is poorly implemented.

Furthermore, the depth of information within a cycle can also vary greatly due to different - and frequently changing - legislation. Some countries (e.g., the UK) do not collect information on the number of bidders, hence making it effectively impossible to analyze tender level competition. Other countries only publish the name and location of organizations without any unique identifiers, which would make over-time tracking feasible. The lack of ID-s also forces researchers to use other matching techniques that are prone to errors, such as string-matching or name-location matching.

Figure 7: Coverage of the full tender cycle[7]



Source: Mendes-Fazekas (2015

For the above reasons the manual offers a list of minimum required variables for comprehensive corruption risk assessment, which are shown in the next table. These variables are necessary to calculate the integrity indicators that will be introduced in the next chapter, however, might not be sufficient for more in-depth - country or sector specific - studies. It must be mentioned that each research requires a specifically tuned dataset, hence this example should only act as a general guideline.

[7] Full coverage was only available until 2012 in case of Hungary.

Table 3: Minimum required information for comprehensive corruption risk assessment

Variable group	Variable
Buyer	Buyer’s name, Buyer’s unique ID, Buyer’s address
Bidder/bids	Bidder’s name, Bidder’s unique ID/tax ID, Bidder’s address, Number of bids submitted, Number of bids excluded, Bid price, Exact time of bid submission, Bid type (winner/loser bid), Beneficial owners
Tender/contract	Procedure type, Framework agreement, Estimated price, Procurement type (service, supply, work), CPV codes, NUTS codes, Status (cancelled, pending etc.)
Dates	Call for tender publication date, Bid submission deadline, Contract start and end dates, Publication date of contract award, Date of contract completion
Subcontracting	Subcontractor’s name and unique ID, Subcontractor’s share
Consortium	Consortium members’ name and unique ID, Consortium member’s unique ID
Contract performance	Contract performance end date, Was performed according to contract,Explanation in case of deferring from contract, Information on contract modification, Information on performance quality

Source: GTI

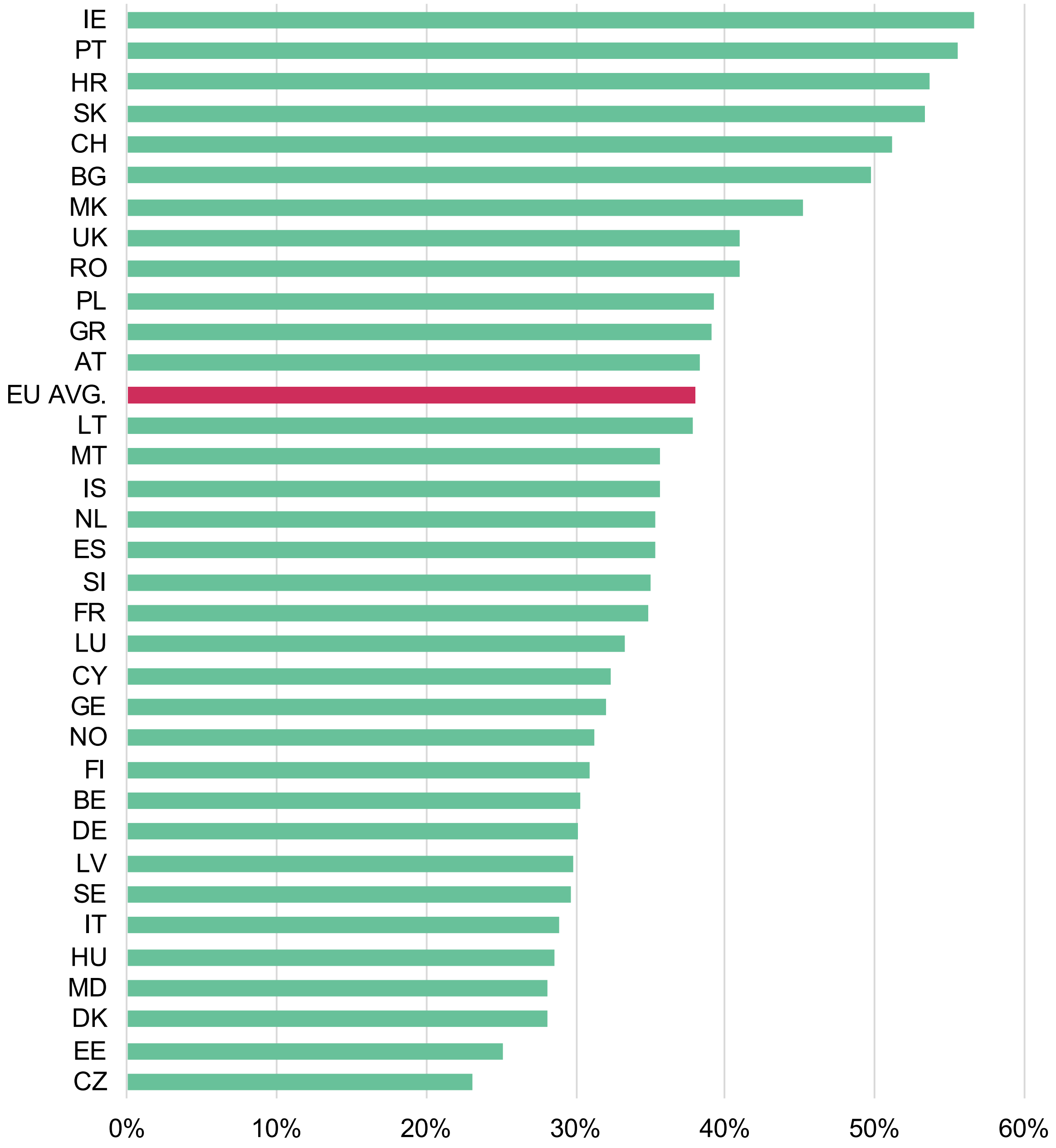
2.3.3. Data quality

Data quality should be examined both before and after the data collection process. Before data collection, it is useful to manually verify the quality of the most important variables. High missing rates or inadequate data in essential variables (e.g., location info only available on the country level) could necessitate the modification of the initial research question or the use of another data source. Nonetheless, full-scale data coverage can only be tested after the data gathering phase is concluded. As a rule-of-thumb a less than 10% missing rate should be considered as acceptable, however data analysts sometimes have to work with greater missing shares due to the lack of alternative public procurement data sources.

Data quality is low throughout Europe with 38% of key fields[8] empty based on the EU-wide TED data and national data. During this period only 8 countries had less than 30% average missing rate for the key variables, and 9 countries had a greater than 40% missing rate. This highlights that one of the fundamental limitations of procurement analysis is the lack of good quality data, or more generally, the lack of transparency in European public procurement data systems. Even the most sophisticated tools are useless if contracts are not adequately published on the official websites.

[8] Product code, Region of implementation, Buyer region, Buyer city, Date of 1st contract, Final tender price, Winning bid price, Tender estimated price, Lot estimated price, Buyer id, Buyer name, Supplier id, Supplier name, Procedure type, Call for tender publication date, Bid deadline, Award decision date, Nr. of bids received, Supplier region, Supplier city, Supplier country

Figure 8: Extent of missing information in European public procurement data systems (2020)



Source: GTI calculation based on 2021 GTI data overview

Finally, database quality must be also checked and compared after the data collection process ends. In this step the analyst should carefully examine the raw data and compare it to the source to make sure that the collection process was flawless. This is especially important if the data is scraped or obtained in any other way that is not controlled by the official maintainer of the website. The following table lists a few typical data errors that can occur during data collection.

Table 4: Common error types

Error type	Description
Lexical error	The value provided is not consistent with the column name (e.g., country id column shows currency id).
Irregularity error	E.g, the unit of measurement differs from the other observations'.
Formatting error	E.g, date is in different time format leading to errors when data is loaded
Duplication error	There are duplicate observations in the data (each value is the same)
Contradiction error	Two columns measuring (almost) the same thing show different values for the same observation
Missing attributes	No information provided (not necessarily an error).
Outlier	Given variable for a given observation is significantly different from the others (not necessary an error, but usually should be dealt with)

Source: GTI

2.3.3. Data accessibility

Even if data quality meets the required standards, the difficulty of obtaining the data could significantly vary across procurement systems. Accessibility usually can be sorted into four main categories:

1. Structured format: Procurement data is stored in a relational database and can be downloaded (using a bulk download option or an API) into an csv/json file.

2. Semi-structured format (semi machine-readable): Information is available in a HTML format which can be scraped and parsed.

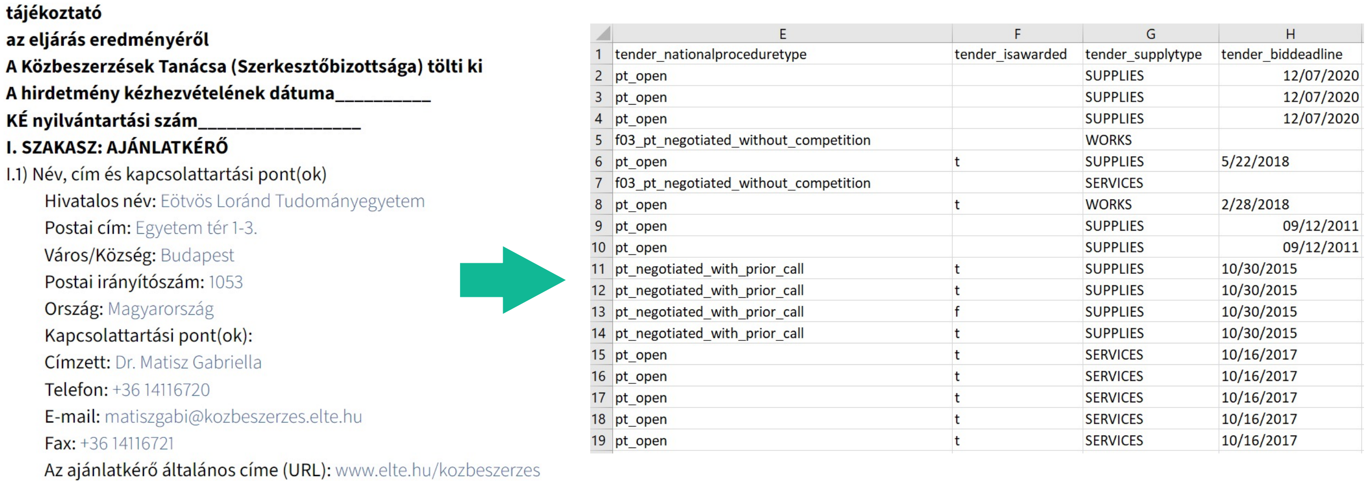
3. Not fully machine readable: Part of the data only accessible by manual cleaning (e.g., some of the documents are uploaded as scanned pdf files).

4. No public database.

Understandably, if there is no public dataset or it is not machine readable, the only option is to contact the local procurement authority and hope that they are willing, and able to provide an applicable database. In contrast, the best scenario is that all the procurement data is available in a structured, easily downloadable format. In this case, after reviewing the dataset along the aforementioned criteria, it is ready to use.

The most common scenario is that the data is available in a semi-structured format which can be processed, but it requires some programming knowledge. If the available data is in a machine readable (e.g., HTML or searchable PDF) format it can be scraped, and the data can be stored in a relational database. The second step is to parse, clean, and impute the raw data. This step can include the matching of notices with their respective contract awards, the processing of raw text files, and the cleaning of overcomplicated variables. The last and arguably the most important step is the manual and automatic data validation. This consists of reviewing a random sample of the data to check data quality and correct potential errors (see previous table) that could have occurred during the collection process.

Figure 9: Illustration of 'before-after' state of semi-structured data



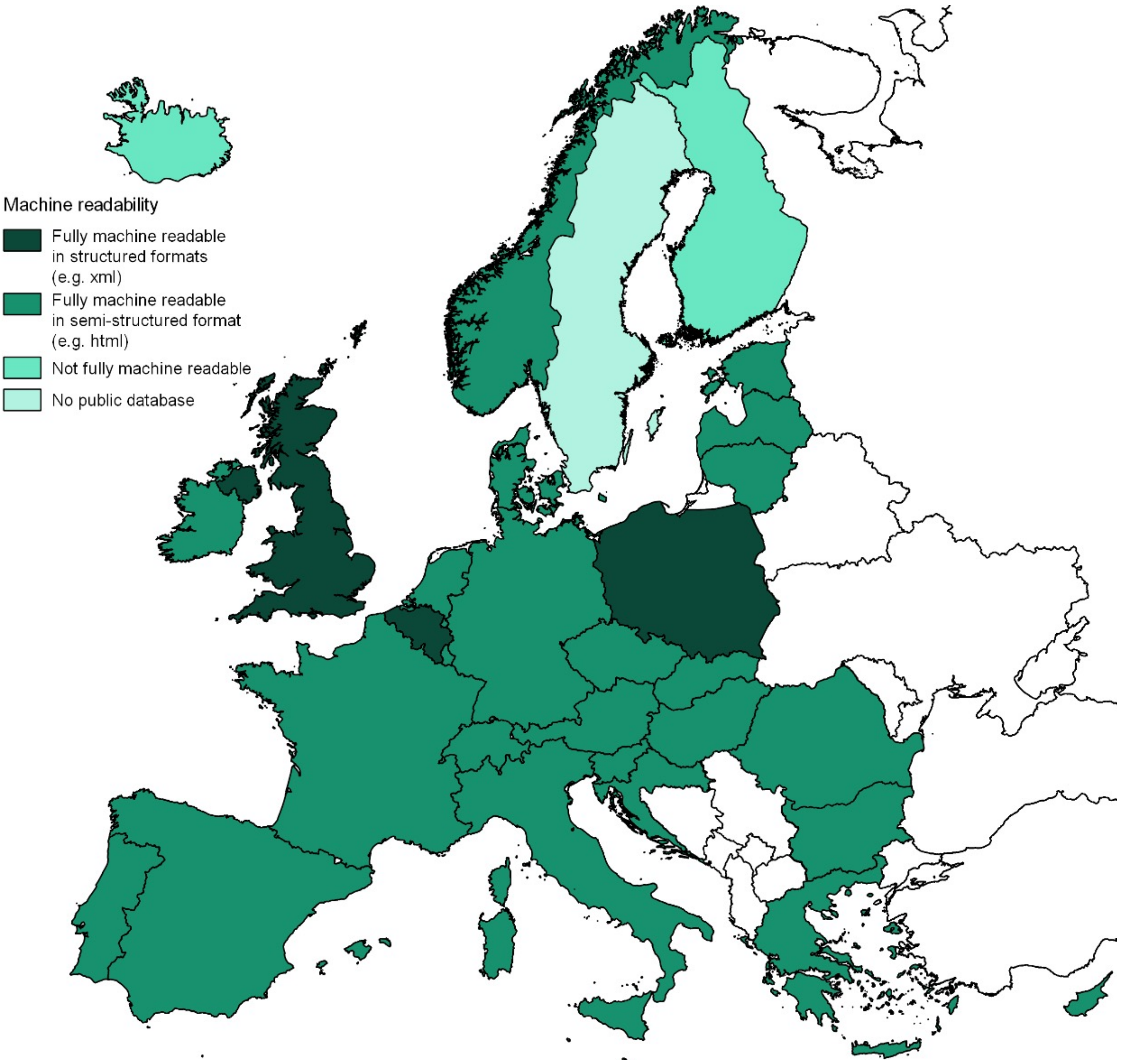
tájékoztató
az eljárás eredményéről
A Közbeszerzések Tanácsa (Szerkesztőbizottsága) tölti ki
A hirdetmény kézhezvételének dátuma__________
KÉ nyilvántartási szám_________________
I. SZAKASZ: AJÁNLATKÉRŐ
I.1) Név, cím és kapcsolattartási pont(ok)
Hivatalos név: Eötvös Loránd Tudományegyetem
Postai cím: Egyetem tér 1-3.
Város/Község: Budapest
Postai irányítószám: 1053
Ország: Magyarország
Kapcsolattartási pont(ok):
Címzett: Dr. Matisz Gabriella
Telefon: +36 14116720
E-mail: matiszgabi@kozbeszerzes.elte.hu
Fax: +36 14116721
Az ajánlatkérő általános címe (URL): www.elte.hu/kozbeszerzes

	E	F	G	H
1	tender_nationalproceduretype	tender_isawarded	tender_supplytype	tender_biddeadline
2	pt_open		SUPPLIES	12/07/2020
3	pt_open		SUPPLIES	12/07/2020
4	pt_open		SUPPLIES	12/07/2020
5	f03_pt_negotiated_without_competition		WORKS	
6	pt_open	t	SUPPLIES	5/22/2018
7	f03_pt_negotiated_without_competition		SERVICES	
8	pt_open	t	WORKS	2/28/2018
9	pt_open		SUPPLIES	09/12/2011
10	pt_open		SUPPLIES	09/12/2011
11	pt_negotiated_with_prior_call	t	SUPPLIES	10/30/2015
12	pt_negotiated_with_prior_call	t	SUPPLIES	10/30/2015
13	pt_negotiated_with_prior_call	f	SUPPLIES	10/30/2015
14	pt_negotiated_with_prior_call	t	SUPPLIES	10/30/2015
15	pt_open	t	SERVICES	10/16/2017
16	pt_open	t	SERVICES	10/16/2017
17	pt_open	t	SERVICES	10/16/2017
18	pt_open	t	SERVICES	10/16/2017
19	pt_open	t	SERVICES	10/16/2017

Source: GTI calculation based on 2021 GTI data overview

The figure below shows that the majority of the European procurement systems provide public access to contract level information stored on HTML based websites. Therefore, the data collection process might be time consuming or excessively expensive for individuals or smaller NGOs. Under the Digiwhist project, the Government Transparency Institute is working with IT companies to collect and standardize procurement data to break down barriers to corruption research. One of the outputs of this cooperation is the Opentender[9] website, which besides providing analytical tools for researchers also publishes free of charge procurement datasets from 32 countries. The website will be introduced in more detail in the last chapter of the manual.

[9] The website is moving to a new location by the end of 2021.


Machine readability
Fully machine readable in structured formats (e.g. xml)
Fully machine readable in semi-structured format (e.g. html)
Not fully machine readable
No public database

Source: Mendes-Fazekas (2015)

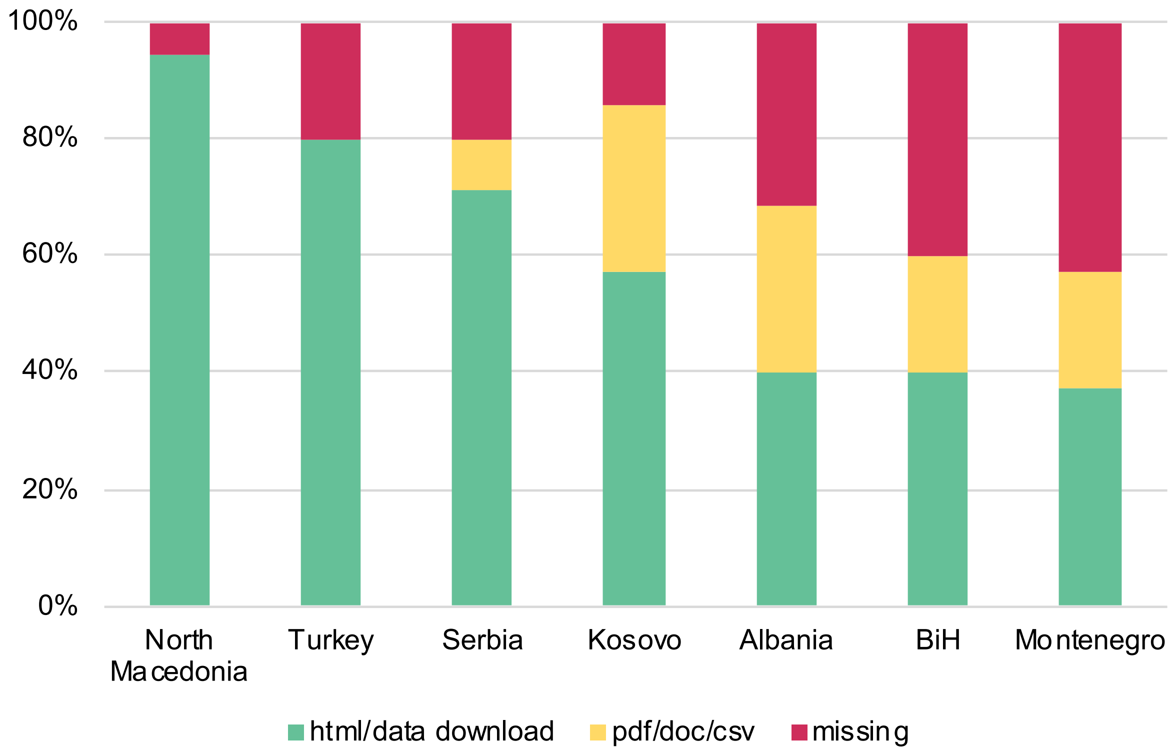
2.4. Examples from the Western-Balkans

This subchapter provides a brief introduction to the quality of procurement systems in the Western-Balkan region, with more in-depth examples from four countries. By highlighting the good and bad practices it should give a general idea what difficulties must be overcome during the data collection process.

The next chart shows data accessibility of the Western-Balkan countries based on a random sample of contracts collected from the procurement websites. It shows the availability of standard variables in the selected procurement contracts. The most striking difference is the extent of missing data across the region;

only North Macedonia has a less than 10% missing rate, although Turkey, Serbia and arguably Kosovo also fare reasonably well. Meanwhile, data accessibility in Bosnia and Herzegovina and in Montenegro is really poor with more than 40% of standard fields completely missing and another 20% being hardly accessible. The other main difference across countries is the extension of the available contract files; North Macedonia and Turkey only publish data in HTML format, while in Kosovo and in Albania a significant portion of the information is available in a semi-machine-readable format. While it should not prevent full-scale data collection, it could significantly increase its costs.

Figure 11: Accessibility and usability of standard data fields



Source: Fazekas et al. (2021), draft*

1. North Macedonia

North Macedonia's Electronic System for PP (ESPP) was set up and running by 2006, making it the longest running electronic tendering system in the region. A complete tender documentation is required to publish a new tender notice, which is an adequate way to ensure data completeness. The website also collects information on the planning and implementation phases, although this information is harder to obtain. Compared to the other countries in the Western Balkans, the North Macedonia system performs highest on accessibility and usability of standard data fields. The main improvement would be the introduction of organization ID-s, which would make over-time company tracking more feasible as discussed above. The other advancement could be the inclusion of a full data download option, which would reduce the costs of data collection.

2. Serbia

Serbia can also be considered a good example. It has a new, improved procurement website from 2020 which offers a bulk download option for some of the standard variables. Furthermore, both the old and the new webpages make contract information available in a standardized HTML format while also providing unique organization ID-s. Nonetheless, the new website only collects new procurements, hence time series data collection is only possible using both sources. Furthermore, the new API has a limited practicality at the moment due to the low number of supported variables. Finally, while most of the procurement data is available, there is still an about 20% share that is either missing or only available in a non-machine-readable format.

3. Bosnia and Herzegovina

The Bosnian e-procurement system stores basic contract information in a structured HTML format, and it even provides unique buyer and bidder ID-s. Although the majority of the standard variables are mostly also available, required tender documentation can be uploaded in several different formats (word documents, excels, pdfs, scanned copies) which makes it especially difficult to extract essential information. Furthermore, 1/3rd of the standard variables is missing from several uploaded contracts. Together the unstructured contract data and the missing information currently make it impossible to collect a reliable procurement database.

4. Montenegro

Montenegro is the other country in the Western-Balkans that has a procurement system currently inadequate to provide structured contract information. On the plus side, certain information is available in a standardized HTML format and a limited amount of data can be exported from the website. Nonetheless, similarly to the Bosnian system, the majority of the tender documentation can be uploaded in unstructured files with much of the important information missing.

2.5. Data wrangling good practice

As the previous examples show the key aspects of procurement data can significantly vary across countries and procurement systems, hence it is useful to carefully explore the available sources before finalizing the project. It is important to go through each aforementioned steps to analyze data quality and only start the collection process if the source is suitable for an unbiased quantitative analysis. Many initiatives attempt to facilitate independent research; hence it is also a good idea to explore the possibilities for a ready-to-use dataset.

After the data collection is finished, it should be adequately cleaned and validated. Generally, the cleaning steps can include:

- handling of missing and extreme values,
- restructuring raw text variables,
- remove duplicate or irrelevant observations,
- potentially standardizing numeric variables,
- adjusting price data with inflation and - if the analysis is international - with PPP (purchasing power parity),
- fixing structural errors (e.g., set “N.A”, “Not available” to missing).

Data validation can be simultaneous with the cleaning process. It can both be automated and can include random manual checks to verify data integrity by comparing the dataset to its original source.

Finally, it is also important to explore the limitations of the database and only use variables that are sufficiently clean. It is very unlikely to obtain all the contract information from any procurement systems, hence the dataset will only approximate reality. Therefore, as mentioned earlier, every analysis should be supplemented with qualitative research.

Continued case study: the Közgép story

Building on Chapter 1, the first framed text briefly summarized qualitative research by investigative journals to illustrate why Közgép may have been involved in several potentially corrupt tenders between 2008 and 2015. Recognizing that quantitative research could further elaborate on these findings, this text goes through the necessary data processing steps that should precede data analytics.

Evaluating the data source

Accessibility & depth

The first step of quantitative research is to carefully map all the features of the primary data source and explore its limitations. In Hungary, public procurement data can be collected from the official e-tendering website maintained by the Hungarian Procurement Agency since 2004. The data is available in semi-structured format indicating that it has to be scraped and re-structured using the Digiwhist data standard. The data scope is somewhat problematic, as it currently has the 10th highest procurement threshold in Europe; procurements below 52 EUR thousands do not have to be published on the official website. Furthermore, there are around 35 separate procedure types, which makes the legal evaluation of every procedure exceedingly burdensome.

Scope

The website contains the majority of the essential information from the advertising and awarding procurement phases. However, it does not provide any information on procurement implementation. This is an important limitation, since no potential contract changes, neither the quality of the output can be assessed, which is particularly important in construction projects. Furthermore, there is no data on how the funds are divided between suppliers in the cases of subcontracting, which would be an adequate indicator to test the corrupt schemes that were allegedly used. Finally, the website does not assign unique IDs to contracting authorities and suppliers, making overtime organization tracking difficult.

Quality

Data quality is generally acceptable, as the most essential fields have a relatively low - less than 10% - missing rate. Nevertheless, there are some other, important variables that are less usable due to high missing shares (e.g., the framework agreement indicator has a more than 30% missing rate, tender estimated price has an over 90% missing rate and lot level estimated price also has a greater than 60%). Data quality has also not improved significantly over the last decade.

Mapping and collecting this data can be very time consuming and/or expensive, hence opentender.eu, a central, public, and open procurement platform has been set up to contribute to achieving value for money in public procurement as well as increase integrity throughout the public sector. Under the framework of the European Digiwhist project, GTI regularly updates and maintains public procurement data, which is available in a structured format at https://opentender.eu/download. Nevertheless, while all the structuring steps have been already performed by GTI, every quantitative analysis should begin with appropriate data cleaning and filtering.

Cleaning and filtering the data

After downloading the Hungarian procurement data from here, the next step is to load it into a preferred software (such as R, Python, Stata or Excel)*. The full dataset might contain several variables that are not to be used in the analysis, hence the best idea - for efficient memory management - is to get rid of those in the beginning of the cleaning process. Since this simple analysis only utilizes basic procurement information and the pre-calculated integrity indicators (intrwoduced in the next chapter), any other variables can be excluded.

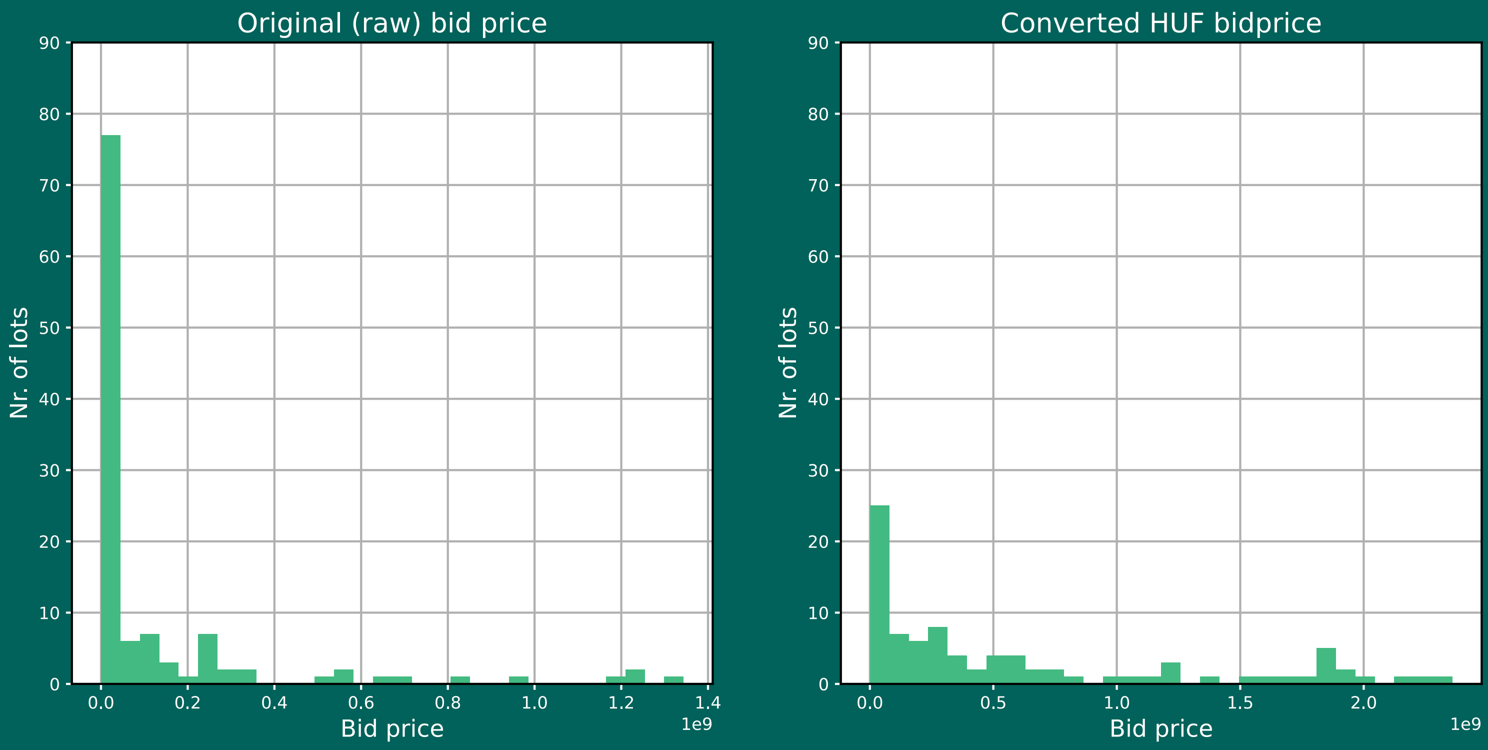
Usually, the next step is to filter the data to the observations that are of interest for the research. First, let's keep only the years between 2008 and 2015, which is the period when Közgép won most of its suspicious contracts. Secondly, any tenders that cannot be connected to the company (either as a sole supplier or as a consortium member) should be removed. While many procurement systems publish organization IDs, the Hungarian is unfortunately not one of them, hence string matching must be applied on the 'bidder name' variable to filter for Közgép's procurements. There are several - more complicated - string matching techniques, such as fuzzy string matching and NLP based (machine learning) algorithms, but for simplicity the following steps should be sufficient**:

1. Lower case every character in the string
2. Remove every special, and non-Latin characters
3. Filter for tenders that include the search world of interest (e.g, közgép). Often the search world is too broad or cannot be precisely defined, in these cases another approach has to be used.
4. Manually examine each unique value to check whether each of them refers to the same entity. There could be many entities that have the same or very similar names. In these cases, different matching techniques must be applied.
5. Remove any observations that refer to a different entity.

Now that the relevant observations have been selected it is important to check the data for errors (see Table 4) and to analyze the share of missing values. There is always a possibility that a variable is inadequate, which can either be fixed - for example, by using another similar variable -, or it has to be excluded from the analysis. In the current dataset, the single bidding indicator was not available for any of the lots won by Közgép between 2009 and 2015. Fortunately, it can be retrieved from the raw lot level 'bid count' variable. The boolean, indicating whether a bid was won by a consortium of suppliers, also has a high missing rate. However, this information can also be extracted by combining the bidder's name, the tender title, and the original consortium indicators.

After correcting the variables, the next important step is to examine the distribution of every numeric variable to check for any potential anomalies. Outliers should be dealt with by either completely removing them or by winsorizing. Inconsistencies can also rise from different denominations. In the current case some of the lots are denominated in EUR, others in HUF, hence it is important to convert each to the same unit of measurement. The figure illustrates how the price distribution changes after price conversion.

Price distribution before & after currency conversion



Source: GTI calculation based on data from Opentender

The final step is to handle more data specific exceptions. Most of the Digiwhist datasets are in the lot or contract levels. Therefore, it is always important to aggregate data to the appropriate level. Since most of the indicators are lot level, in this scenario the data does not have to be aggregated. However, in other cases - for example, when Contracting Authorities are analyzed - the level of observations has to be on the tender level to prevent duplicates. The final dataset has 118 observations (lots) within 89 unique tenders.

* While the Hungarian datasize is manageable, depending on the country the data might have to be processed in chunks. Chunking is a data processing method, where the dataset is loaded and filtered in smaller ‘chunks’ in order to prevent memory errors. Also see ‘parallel computing’ for the faster processing of large data files.

** Note that, while in this case it was easy to find all the relevant cases, finding more complex sub-samples could be much more difficult, requiring more advanced programming knowledge.

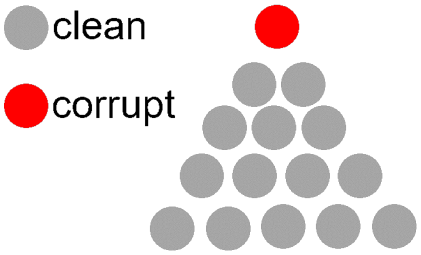
3. Measuring corruption risks in public contracting

The first chapter defined corruption in public procurement as particularistic, institutionalized, and grand. In this setting, public and private bodies cooperate to either favor the private actor or cease competition altogether to siphon out public money for private gain. As shown earlier, these corrupt practices can take several - often legal - forms and they are rarely observable directly. Thus, the best option is to put together new statistics from the observable data, which can most effectively measure this underlying phenomenon. This chapter explains how to create indicators that can efficiently detect and measure the level of corruption in the public procurement setting.

3.1. The importance of public procurement risk indicators

Let's consider the task of distinguishing clean vs. corrupt contracts, for example to investigate the extent and forms of corruption in a public procurement system. An initial idea might be to take a small sample of contracts for in-depth analysis, which would show that 1 in 20 contracts could be considered corrupt. The problem with this approach is that 95% of the work put into the analysis is unnecessary, because the share of truly corrupt contracts is only 5%. Furthermore, while a randomly selected sample could potentially predict the share of corrupt contracts in the whole population (although a small sample with a sufficiently low share of corruption even hinders these estimates), it is inadequate to describe the exact characteristics of corruption in the system. For this, a researcher would need to analyse each contract in the procurement market of possibly thousands of contracts.

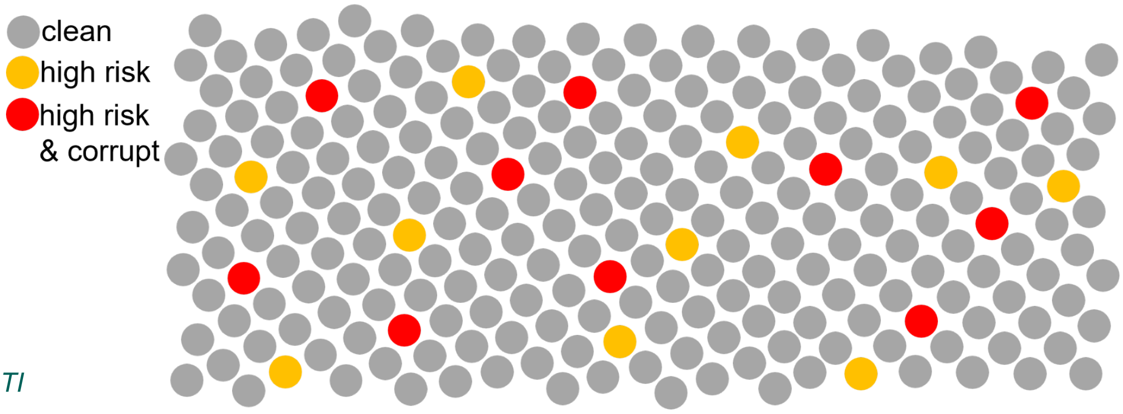
Figure 12: Example of a sample of potentially corrupt contracts



Source: GTI

An alternative approach is to use *risk indicators* to find potentially corrupt contracts. A statistically developed and empirically tested indicator can predict each contracts corruption risk, hence theoretically - after adequately tuning the indicator - there is no need to manually check each contract in the whole population. The largest issue with this concept is that it is impossible to develop a 100% accurate indicator. It could both 'flag' contracts that are not corrupt (false positives), or miss procurements that are corrupt (false negatives). A poorly designed indicator might cause more harm than good, since it creates additional work by both having to find all non-flagged corrupt cases and remove flagged ones that are not corrupt.

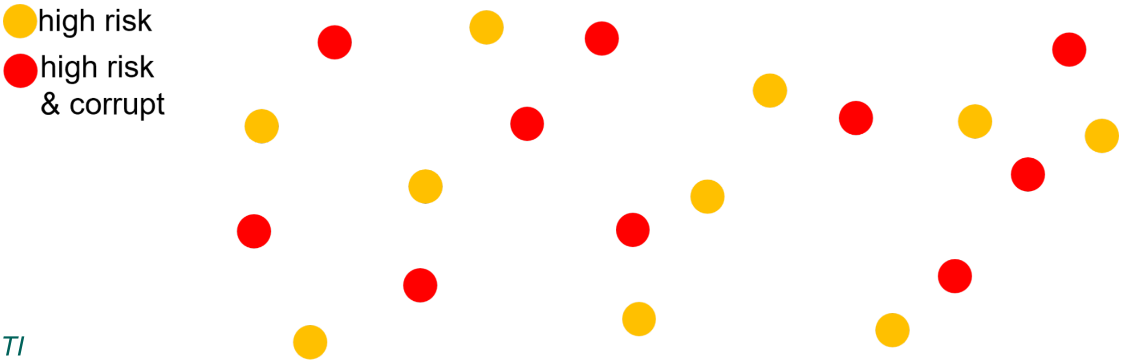
Figure 13: Example of corrupt and flagged cases in the total population of contracts



Source: GTI

Nevertheless, even a mediocre indicator, that flags false positives, can significantly reduce the time required for the risk assessment. Imagine an indicator that can label all the corrupt cases, but also flags an equal number of non-corrupt cases. The chart below shows that in this scenario an analyst must manually check all 20 cases, 10 of which will be corrupt. Contrary to the first example, now all the corrupt contracts are found and 50% of the work was useful. This method does not work however, if an indicator understates risks. False negatives are more dangerous, because they can only be found if the whole population is manually checked. Therefore, it is a better idea to start with 'strict' indicators that might initially flag non-corrupt contracts and refine them by testing on separate samples.

Figure 14: Example of corrupt and flagged cases in the total population of contracts



Source: GTI

Overall, the main goal of indicator building is to increase the overlap between the corrupt cases (red dots) and the high-risk cases flagged by the indicators (yellow dots). When testing these indicators three issues should be considered:

- *False positives:* the indicator flags contracts that are not corrupt.
- *False negatives:* the indicator does not flag contracts that are corrupt.
- *Both:* the indicator is both missing some corrupt cases and flagging non-corrupt ones.

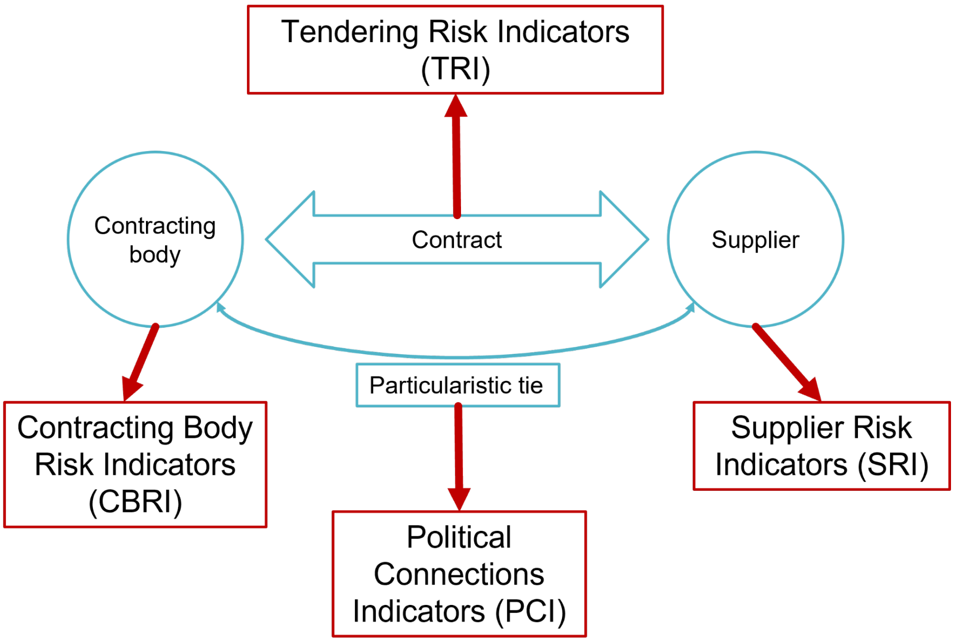
A perfect indicator would find each corrupt contract without flagging any additional ones, however, even a slightly imperfect indicator can speed up the research process. Furthermore, as it will be discussed in the next section, combining separate, independent indicators can increase the reliability of the final measurement tool.

3.2. Conceptualizing risk indicators

Given that well-designed indicators can significantly reduce the costs of quantitative research, the next step is to circumscribe the important features of indicators that are potentially capable of measuring procurement corruption risk. The first step is to outline the framework in which corruption risk indicators can be developed. As it is highlighted by the previous sections, a corrupt procurement process consists of 1) a public body buying a service (or supply, or work) using public funds, 2) a supplier, who will provide this service in return for these funds, 3) a contract that describes the details and conditions of the agreement, 4) and a particularistic (informal) tie which makes the coordination of corrupt practices possible. For each of these items a group of potential indicators can be developed:

- *Tendering Risk Indicators (TRI)* are a group of contracts specific - observable - variables that can be steered to favor a certain supplier. These can include the procedure type, bidding period length, award criteria etc. (see Table 2. for more examples).
- *Supplier Risk indicators (SRI)* are supplier level information which could indicate that it might be involved in corrupt practices. These variables can include the firm's registry date (e.g., if it was registered just before the high value contract was published), whether it is registered in a tax haven, or whether it is extremely profitable compared to other market participants.
- *Contracting Body Risk Indicators (CBRI)* are a group of variables indicating that the procurer might attempt to corrupt its purchases. These indicators are hard to measure because public organizations often don't have well-structured data. However, some publicly available information can be useful, such as the change in leadership (or employees) after a political regime change, or the proportion of adequately trained staff.
- *Political/Personal Connections Indicators (PCI)* describe the informal tie between the buyer and the supplier. This is also a hardly measurable area but indicators such as kinship, or previous political or business connections between the leaderships could indicate the presence of a particularistic tie. PCI-s are good examples for risk indicators that are likely to underestimate corruption risk, since objectifying political or personal connections is often a difficult task.

Figure 15: Corruption indicator groups in the procurement process


Tendering Risk Indicators (TRI)
Contracting body
Contract
Supplier
Particularistic tie
Contracting Body Risk Indicators (CBRI)
Supplier Risk Indicators (SRI)
Political Connections Indicators (PCI)

Source: GTI

Indicators in each of these groups must have a list of common qualities that make them adequate to measure procurement corruption risk. The following are the key features that analysts should consider when creating new indicators; they have to be:

- **objective**: they are based on factual data non-mediated by stakeholder's perceptions, judgements, or self-reported experiences,
- **de facto**: they describe actual behavior or events in contrast to legal prescriptions or expectations,
- **micro-level**: they are defined on the level of actors of corrupt exchanges (e.g., companies) or the transactions among them (i.e., contracts). They can nevertheless be aggregated at higher levels,
- **internationally comparable**: while defined on the micro-level, indicators should be comparable across countries or regions, due the same underlying theoretical concepts and measurement approach,
- **comprehensive**: they adequately capture corruption risks in a wide set of organizations performing comparable tasks,
- **timeseries**: indicators are ideally measured and can be compared over time for at least 5-10 years.

Indicators that are not equipped with these qualities could bias the analysis, therefore, not to be used.

So far, the manual has covered the first three steps of the corruption measurement process (see chapter 1.2.); it has outlined the corruption definition used in the public procurement setting, introduced some popular corruption technologies that are used in a wide variety of procurement systems and described the scale and scope of a procurement dataset that can be used for indicator building. The last step is to showcase some of the widely used indicators and introduce the validation process they must go through before implementing them in a research setting.

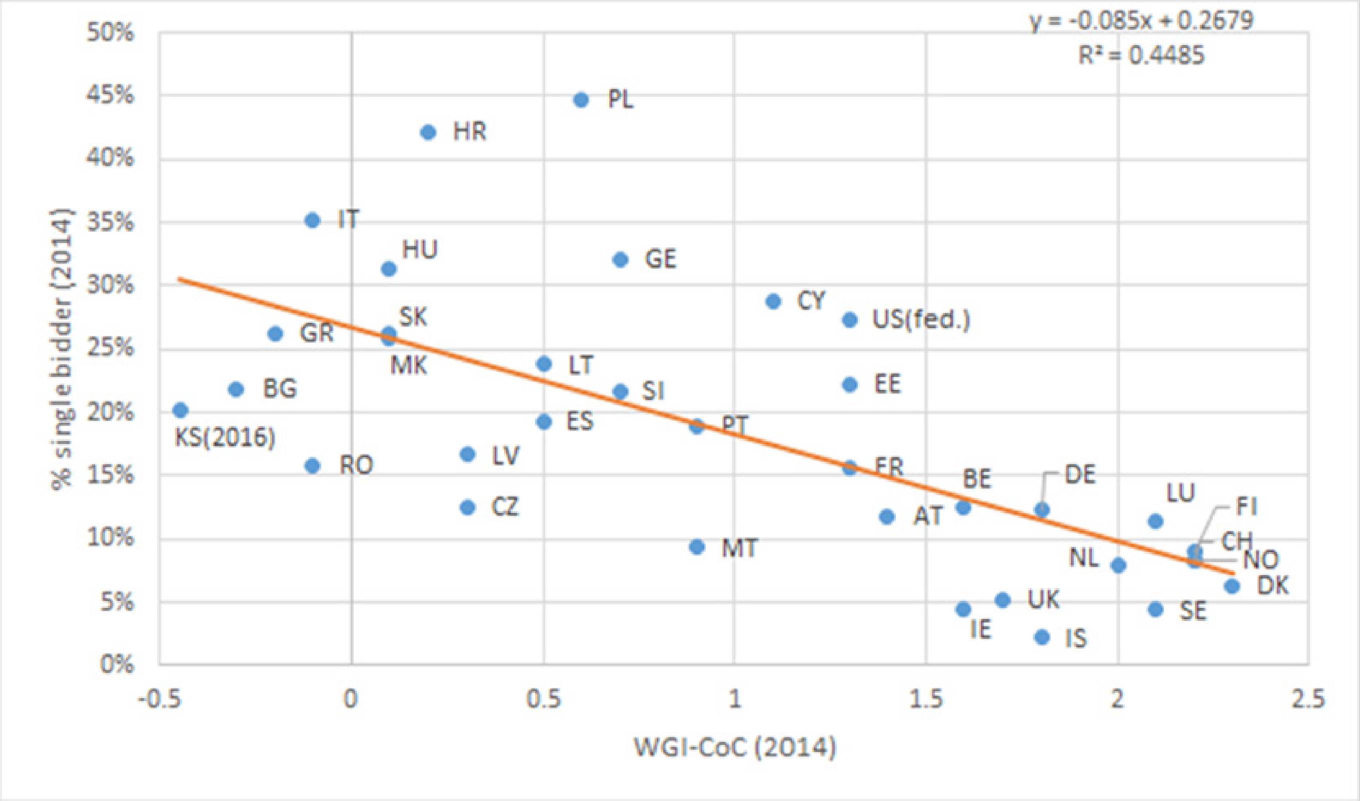
3.3. Empirical evidence on corruption risk indicators

One of the most widely used corruption risk red flags is the single bidding indicator. It indicates that a given tender only had one bidder during the procurement process, hence there was no competition for the contract. As explained earlier, the lack of competition is one of the main signs of corruption in the public procurement system. Even more conveniently, single bidding can be easily extracted from most of the available data sources, and it also holds the key qualities necessary for an adequate indicator.

Nevertheless, even a theoretically sound indicator must be tested before applied in quantitative research. The best scenario would be to test the indicator on 'labeled' data where each corrupt contract is already flagged. Unfortunately, only a handful of procurement systems disclose this information (e.g., court rulings) and even these datasets lack overall generalizability, since not all the corrupt contracts can be located by procurement authorities (many of them could stay hidden due to political and technical factors).

The second-best option is to compare the indicator to a both theoretically and empirically tested measure. The figure below gives an example showing the correlation between the World Governance Indicators' (WGI) Control of Corruption index (CoC) and average single bidding rate in European countries. WGI-s have been developed by the World Bank since 1996 and have been tested in over 200 countries (see more: here). The Control of Corruption index is based on a wide range of surveys measuring perceptions of corruption. The graph indicates that there is a visible correlation between the two indicators; the WGI-CoC explains close to 45% of the variation in the average single bidding rate in European countries' procurement contracts.

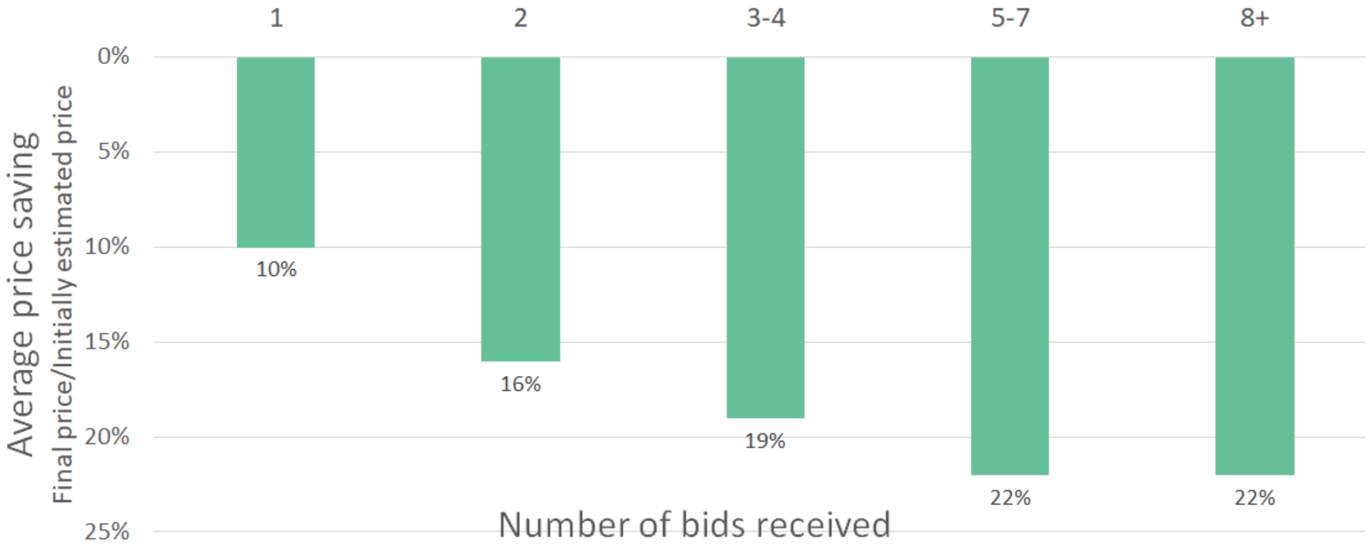
Figure 16: Single bidding vs World Governance Indicators' Control of Corruption



Source: Czibik (2015)

It is also important to test the indicator with alternative measures that have been developed using different methods. Price saving illustrates the difference between the final price of the contract and the initially estimated price assigned by the contracting authority. Smaller - or negative - savings do not necessarily indicate corruption but can signal the stronger bargaining power of the supplier, which could be effectively decreased by increasing the level of competition. The next figure shows the connection between the number of bidders and average price savings using European procurement data between 2009 and 2014. As expected, there is a positive connection between the average number of bidders and price savings; while, on average, single bidder contracts had a 10% lower price than the initially estimated one, price savings for two bidder contracts had a 6 percentage-points higher savings rate and contract with 5 or more bidders had a 12 percentage-points higher savings rate.

Figure 17: Number of bids and price savings



Based on 543,705 contracts, EU27, 2009-2014

Source: Czibik (2015)

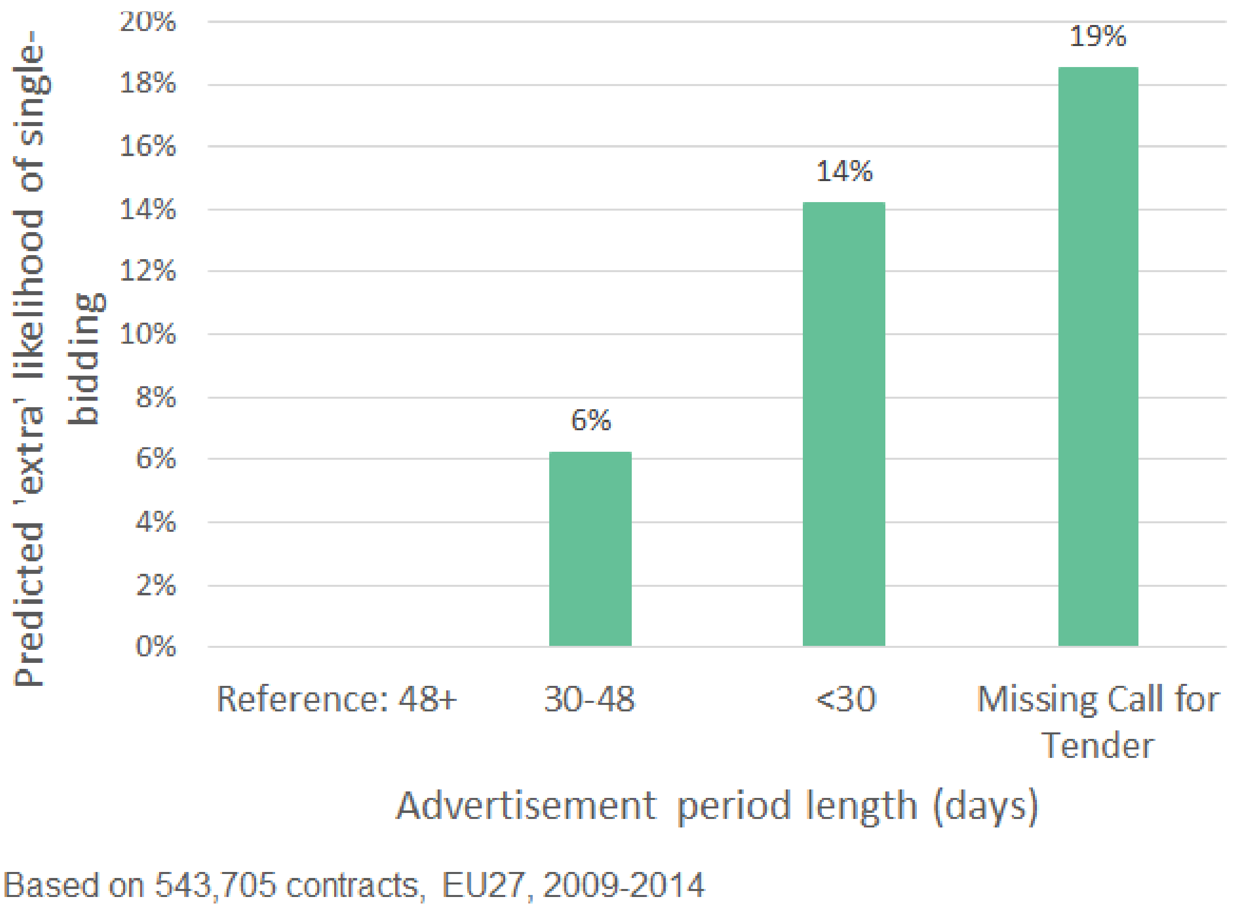
While none of these results necessarily demonstrate a causal relationship, both the theoretical (the lack of competition could signal corruption) and the empirical evidence suggests that single bidding might be a good starting point to identify corrupt contracts. However, similarly to other indicators, single bidding cannot be used as an all-round tool for corruption measurement. Like many other indicators it has weaknesses, such as overestimating corruption risk. Take for example a very specialized market where there are not enough companies. In this market, public procurement will be limited, and single bidding will be a measure of market imperfections (e.g. too high barriers to entry) rather than a measure of corruption. Another example could be the sudden increase in government spending, creating tight procurement markets where contracting authorities have to compete for suppliers. Other elementary indicators might underestimate risk creating many false negative contracts. For example, political connections can be hard to establish between government suppliers and politicians, hence many corrupt contracts could not be identified by such an indicator.

3.3.1. Combining indicators

Fortunately, there is a solution for the above problem. There are many potentially adequate indicators that measure slightly different aspects of corruption risk, hence - after rigorously validation -, they can be combined into a composite score which increases robustness by reducing the amount both of false negative and false positive cases (hence increasing the overlap between the truly corrupt cases and the high risk cases flagged by individual indicators). From a more technical point of view, combining different indicators can be thought of as removing confounding factors, hence creating a more robust relationship between contract level corruption and corruption risk measurement.

To select appropriate indicators for the final composite index, each of them have to be separately validated. If, for example, single bidding is accepted as an appropriate indicator, each additional measure can be validated by comparing it to single bidding and by comparing it to the other indicators. The figure below shows the connection between single bidding, and another widely used indicator, the advertisement (or submission) period of contracts. The advertisement period is the time difference between the first contract notice publication date and the deadline until which suppliers can submit their bids (bid deadline). As explained in Chapter 1.3, a sufficiently short advertisement period could indicate corruption, and there is indeed a connection between the length of the period and the likelihood of single bidding (see figure). Compared to the reference period, contracts with less than 30 days of advertisement period had a 14 percentage-point higher likelihood for single bidding on average. The lack of information could be even more suspicious as contracts with no call for tender (contract notice) published on the official website had a 19 percentage-point higher single bidding rate on average.

Figure 18: Likelihood of single bidding



Source: Fazekas (2015)

Table 5: Examples of corruption risk indicators [10]

Indicator name	Definition
Advertisement period length (country specific)	0 = length of advertisement period is unrelated to corruption risks 50 = length of advertisement period has intermediate relationship with corruption risks 100 = length of advertisement period or missing advertisement period has a strong relationship with corruption risks
Decision period length (country specific)	0 = length of decision period is unrelated to corruption risks 50 = length of decision period is somewhat related to corruption risks 100 = length of decision period or missing decision period is related to corruption risks
Single bid	0 = more than 1 bid received 100 = 1 bid received
Call for tender	0 = call for tender/prior information notice published in official journal 100 = No call for tender/prior information notice published in official journal
Procedure type (country specific)	0 = open, or does not have significant effect on corruption risk 50 = negotiated 100 = non-open + has significant effect on corruption risk
Tax haven	0 = winning bidder is not registered in a tax haven country, and is a foreign bidder 100 = company is registered in a tax haven country
New company	0 = if company is older than 1 year when winning a public contract 100 = if company is younger than 1 year when winning a public contract

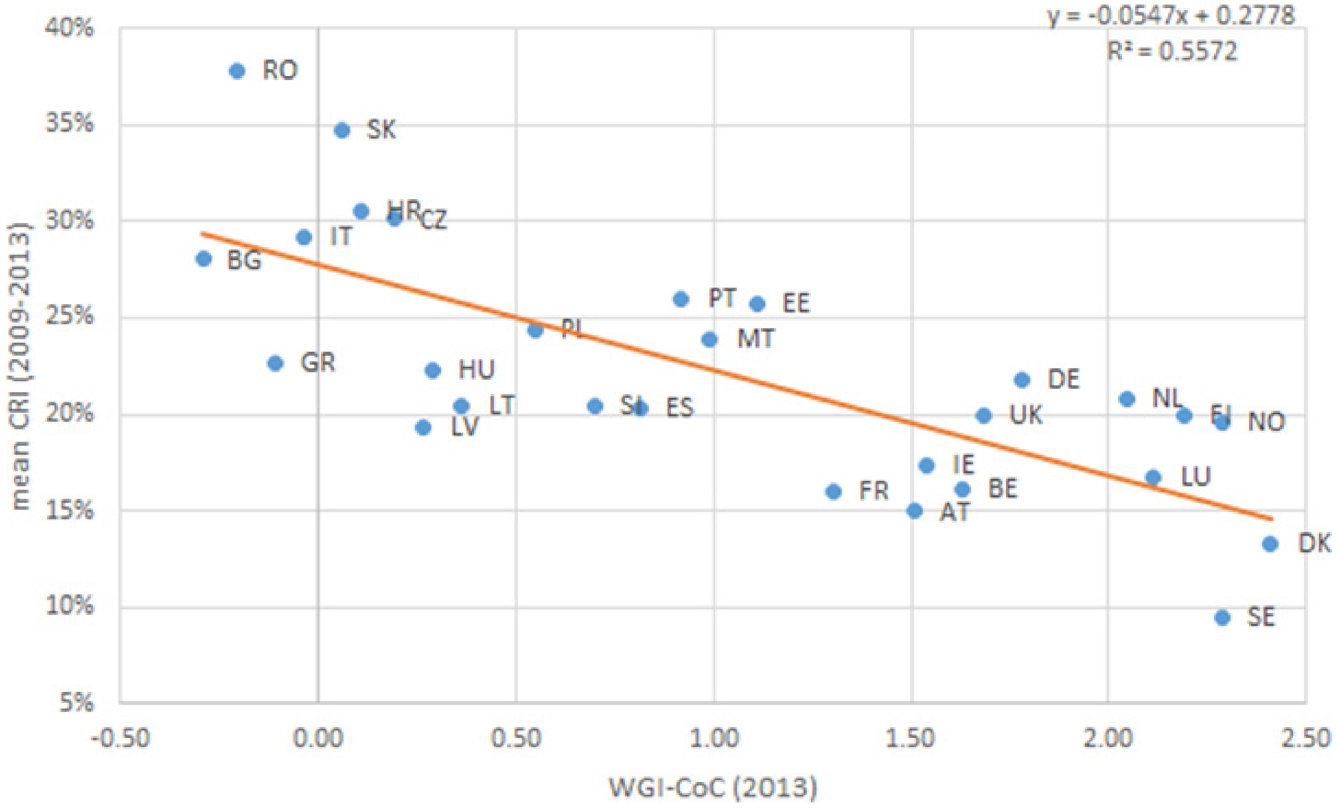
Source: GTI

For each separate procurement system every indicator should be similarly tested before deciding on whether to use it as part of the composite score. Since not every countries' procurement system provides the same quality of information, the composite score can slightly vary across jurisdictions. The table above epicts some of the widely used indicators, their definition, and the actual value they can take for each contract in the dataset. The composite score, called the Corruption Risk Index (CRI), is the arithmetic average of each indicator. It can take a value between 0 and 100 where 100 indicates the strongest corruption risk.

[10] Note that GTI also uses Integrity indicators (such as opentender.eu), that are the exact opposite of corruption risk indicators (eg. the integrity indicator for single bidding is 100 if there are more than 1 bidders and 0 if there is only 1)

The CRI can be also validated using established indices such as the Control of Corruption index. The figure shows that CRI has a stronger correlation with the CoC than single bidding rate has in itself. This illustrates that combining indicators can give a more complete picture about the corruption risks of a procurement system. It is important to note that each of the indicators mentioned above are validated and tested using a more complex statistical procedure, the explanation of which is not part of this manual. For more technical details on the indicator validation processes and Corruption Risk Index formulation please check the following studies: Fazekas - Kocsis (2015), Fazekas et al. (2016a), Fazekas et al. (2016b).

Figure 19: Corruption Risk Index vs World Governance Indicators' Control of Corruption



Source: Fazekas (2015)

The last step of the analytical process is to perform the analysis and to draw the conclusions from the results. The current case study is only to provide a general example of the most important analytical steps. Hence, the below analysis only showcases a few descriptive statistics that highlights how quantitative analysis can complement qualitative research. Therefore, none of these results should be treated without reservations. For proper corruption risk and good governance related research check out GTI's website.

Creating a 'control' group

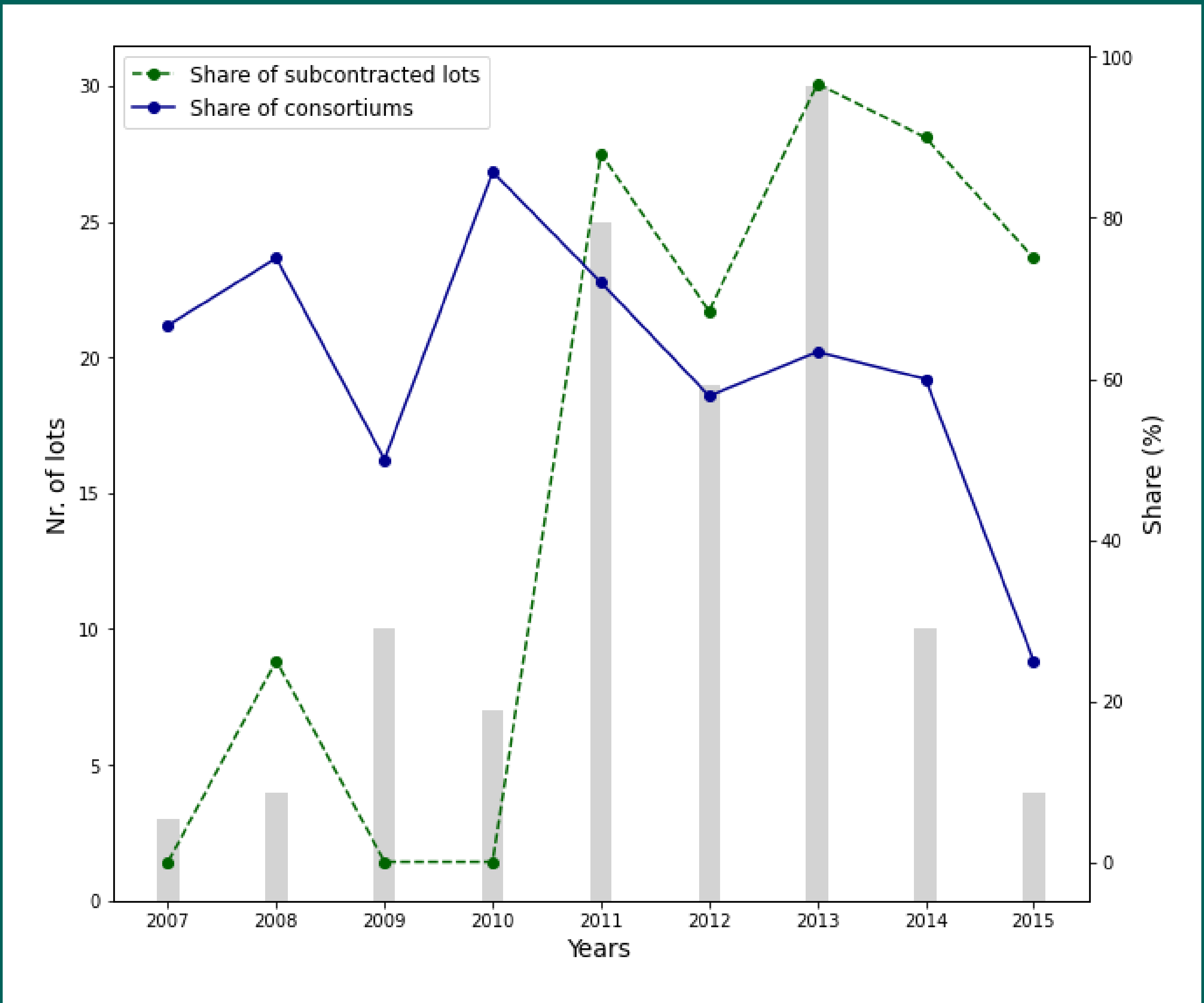
Quantitative research interrogating the overall state of a given subpopulation must always be compared to an appropriate control group. Without a control group, results are not meaningful, as the absolute state of a system usually cannot be interpreted quantitatively. There is no universally acceptable rate of corruption in any system, considering that the level and rate of corruption can vary over time and over procurement systems. Therefore, comparability is only accurate within these dimensions.

The current case study investigates the integrity of Közgép's procurements between 2008 and 2015 using publicly available procurement data. Procurements in the corresponding control group must be awarded during the same time period, should be implemented in the same geographical location, and within the same market segment. As there is no information on the location of implementation, the contracting authorities' residence must be used as the best alternative. To narrow down to the same market segments, lot level Common Procurement Vocabulary (CPV) codes can be used.

Descriptive statistical analysis

Once the control group has been established, the results can be compared to the qualitative research introduced in the first section of the manual. The results confirm both hypotheses, that Közgép performed a significant portion of its procurements as a consortium member and a high share of these projects were outsourced to subcontractors. While the industry average of consortiums was 29% at the time, Közgép performed almost 62% of its projects as a consortium member. Despite poor quality data on the identity of consortium members, it also seems to be verifiable that the composition of these consortiums was often very similar. Közgép had verifiably worked with Swietelsky Kft, A-HÍD Kft., KE-VÍZ 21 Kft. or with STRABAG Zrt. in more than 31% of its projects. The share of work performed in consortiums had not decreased significantly during the observed time period. In contrast, the share of subcontracted lots had increased from below the 11.5% industry average to above 80% after 2010. This also coincides with a significantly increased number of lots supplied by Közgép. These results are consistent with the information collected by independent NGOs and journals; procurement data confirms that Közgép could have been part of corrupt schemes channeling public funds out of the procurement system using subsidiaries to hide these activities.

Share of Közgép's subcontracted lots, and lots performed by Közgép as a consortium member

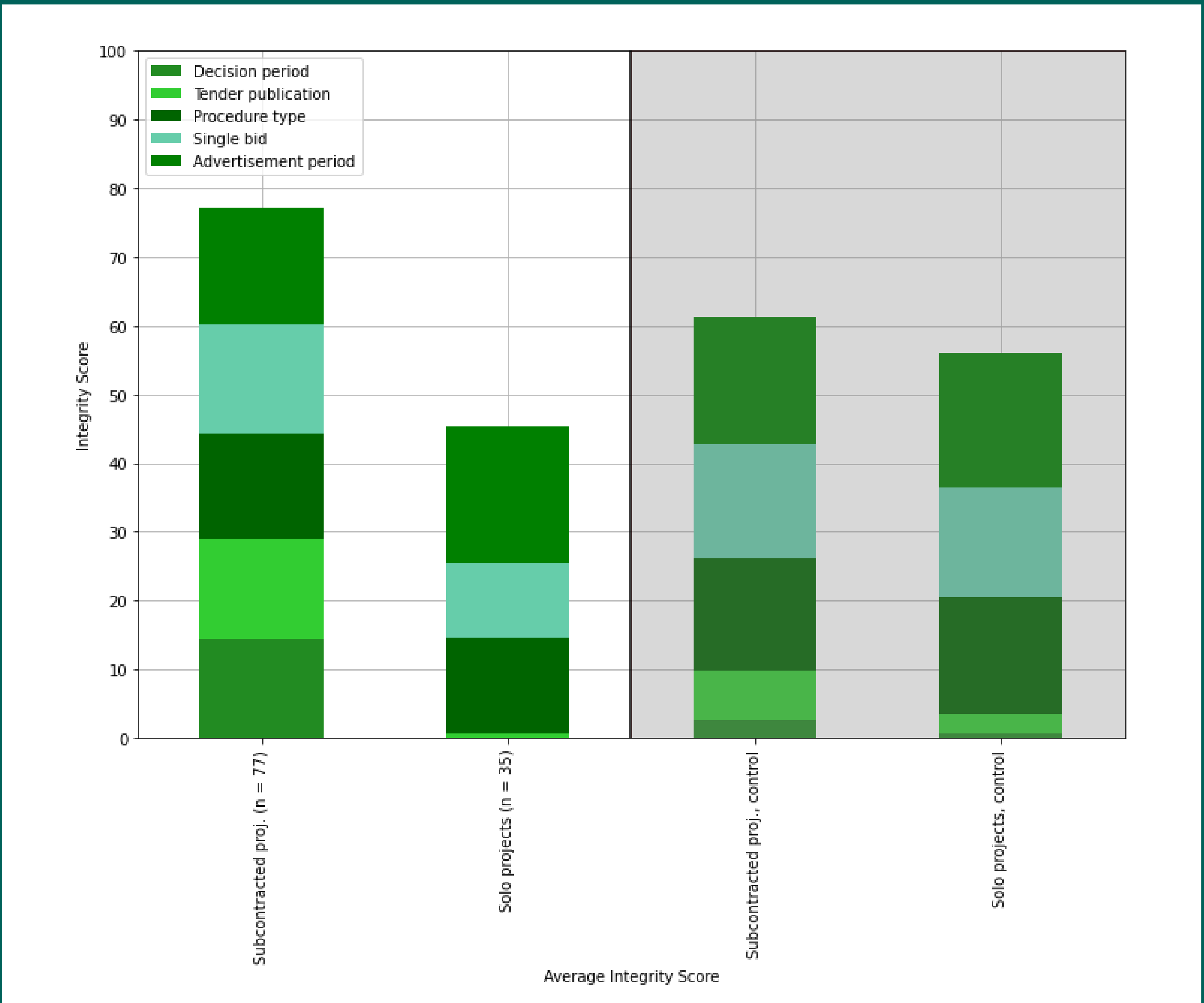


Source: GTI calculation based on data from Opentender

While these results could indicate corruption, neither subcontracting or winning tenders as a consortium member are validated corruption risk indicators. To get a more definitive answer, empirically tested and validated red flags must be used. In the current scenario single bidding, advertisement period length, decision period length, the publishing of contract notice and the procedure type integrity indicators* are available. In Hungary, advertisement period is considered risky if it's less than 38 days, decision period considered as high risk if it's less than 20 days and it has mediocre risk if it's between 20 and 40 days**. Combining the five indicators creates the procurement **Integrity Score**, which shows the average integrity of each lot in the dataset.

The results indicate that the picture is less clear than it seems at the first glance. On paper, the integrity of Közgép's subcontracted projects were significantly higher than the control group's, albeit - as the figure illustrates - the overall integrity of the construction market was also relatively low at the time. In contrast, the integrity of the company's non-subcontracted projects was much lower than the control group's; the decision period was insufficiently short in every project - indicating that the results might have been decided in advance - and the call for tender notices were also hardly published on the official procurement website.

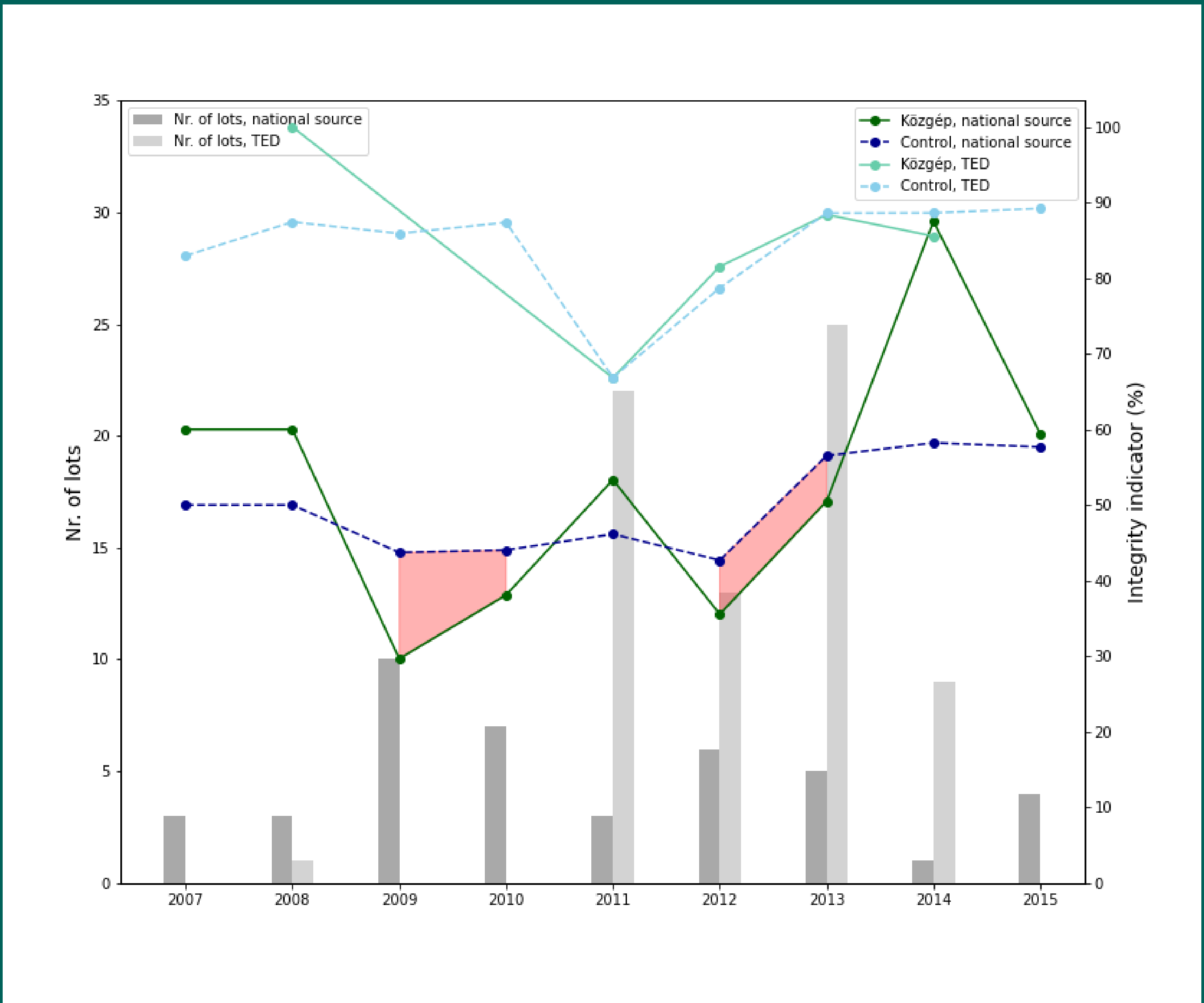
Composition of the Integrity Score of procurements supplied by Közgép (2007-2015)



Source: GTI calculation based on data from Opentender

The next graph also presents some interesting results. Firstly, the integrity of procurements that were published on TED was considerably higher than of procurements that were only published on the national website. European legislation requires public works contracts worth above €4,845 to be published on TED. Közgép started to win a large amount of these high-value procedures after 2010, which also coincides with the sudden increase of subcontracted work contracts. Furthermore, as indicated by the red areas on the figure, the company won its lowest integrity procurement in 2009 and 2010, when - according to the investigative reports - the company collected all the references necessary to unsuspiciously compete for high value procurements.

Composition of the Integrity Score of procurements supplied by Közgép (2007-2015)



Source: GTI calculation based on data from Opentender

Overall, the results seem to certify most of the conclusions of the qualitative research. Közgép started to win high value procurements after the 2010 elections and won barely any tenders after Lajos Simicska’s spat with Viktor Orbán (see figure inthe first text box). The firm’s revenue also followed a similar curve. Furthermore, the sudden rise of subcontracting after 2010 also seems to be justified by the lack of sufficient manpower and - while cannot prove it - it also plays along nicely with the insider information arguing that a handful of construction companies had used subcontracting to siphon out public funds from the procurement system.

However, the results also show that objective indicators are not always sufficient, as some corrupt technologies can be disguised by otherwise less suspicious activities, such as subcontracting. Nonetheless, in other aspects these indicators can complement qualitative research by flagging other high corruption risk projects. Such as the case for Közgép’s non-subcontracted procurements between 2009 and 2010. These projects, on average, had a less than 40% integrity score implying that at least 3 out of the 5 red flags signal corruption.

The case study shows that statistically developed corruption risk indicators applied on a rigorously cleaned data can neatly complement qualitative research. It can both be used to prescreen a large pool of procurement cases to select the most suspicious ones, or as a tool to highlight corruption risk in a selected group of procurements (ie. performed by the same entity). These results also highlight that both methods have their separate weaknesses. Qualitative research can never be completely objective, and it could create severely biased results. On the other hand, quantitative research can only be as good as the data used, hence the lack of data can derail any analysis regardless of the suitability of the methodology.

* Integrity indicators are the opposite of corruption risk indicators (e.g, if the contract has multiple bidders, then the single bidding indicator is 100). Opentender uses integrity indicators, therefore the manual follows suit.

** See the definition of the other indicators on Table 5.

4. Introduction to Opentender

The last chapter of the manual introduces the main functionalities of the Opentender website. It is one of the main outputs of the EU funded Digiwhist project, which brings together six European research institutes, with the aim of empowering society to combat public sector corruption. The main goal of Opentedner is to provide comprehensive public procurement information free of charge in an easy-to-use format to all interested parties. It is expected to increase market transparency, decrease transaction costs, and facilitate government accountability.

As the page is under a larger overhaul at the time of writing, it might receive additional features and might be migrated to another host. Nonetheless, its main objectives and most important features will remain the same.

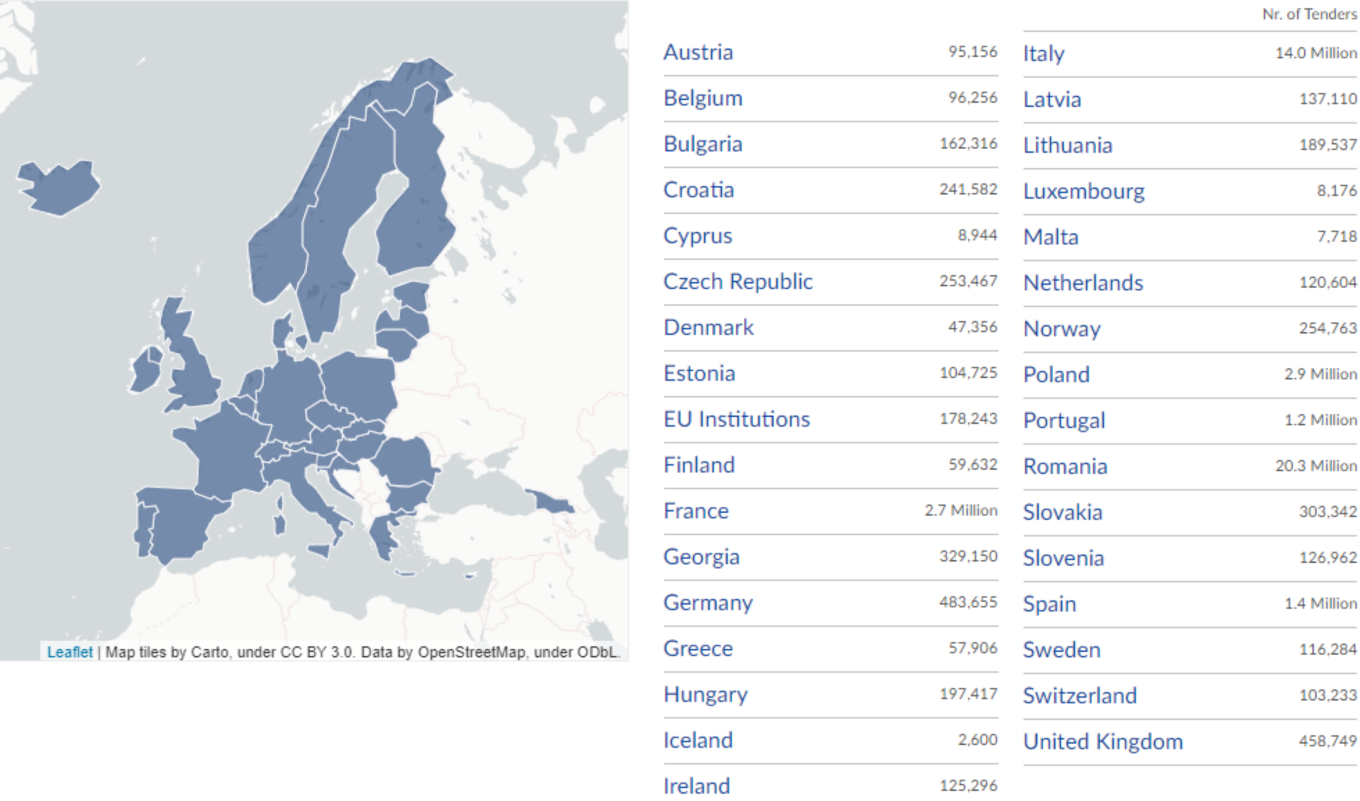
4.1. Main functions of Opentender

Currently, 32 countries' procurement systems can be analyzed using interactive tools. The specific country level dataset can be selected from the website's starting page. By clicking on a country name the website loads some general descriptive analysis about the country's procurement system and brings up the website's currently available functions.

These are:

- in depth market analysis,
- indicator analysis (administrative capacity, transparency, integrity),
- raw search,
- and bulk data download.

Figure 20: Opentender's starting page


Leaflet | Map tiles by Carto, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

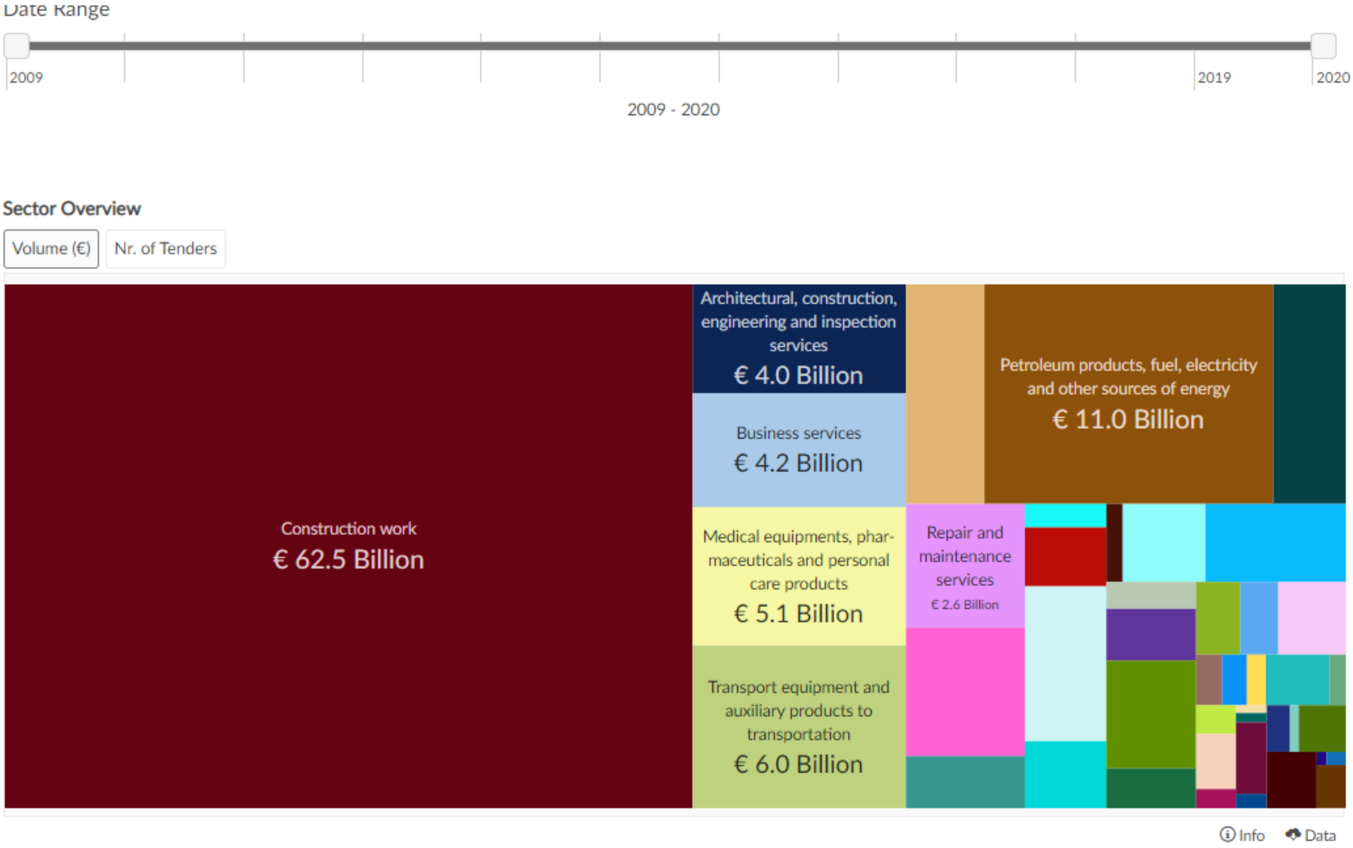
			Nr. of Tenders
Austria	95,156	Italy	14.0 Million
Belgium	96,256	Latvia	137,110
Bulgaria	162,316	Lithuania	189,537
Croatia	241,582	Luxembourg	8,176
Cyprus	8,944	Malta	7,718
Czech Republic	253,467	Netherlands	120,604
Denmark	47,356	Norway	254,763
Estonia	104,725	Poland	2.9 Million
EU Institutions	178,243	Portugal	1.2 Million
Finland	59,632	Romania	20.3 Million
France	2.7 Million	Slovakia	303,342
Georgia	329,150	Slovenia	126,962
Germany	483,655	Spain	1.4 Million
Greece	57,906	Sweden	116,284
Hungary	197,417	Switzerland	103,233
Iceland	2,600	United Kingdom	458,749
Ireland	125,296		

Source: Opentender

4.1.1. Market analysis

The market analysis function makes year, and market level filtering possible. The first interactive graph shows the size of the country's procurement market divided into sub-markets. By toggling the 'Date Range' or clicking on one of the squares on the 'Sector Overview' figure, the website filters the data accordingly; the figures always show procurements from the selected market and/or date range. Data for each chart can also be downloaded by clicking on the 'Data' option on the right corner of any figure.

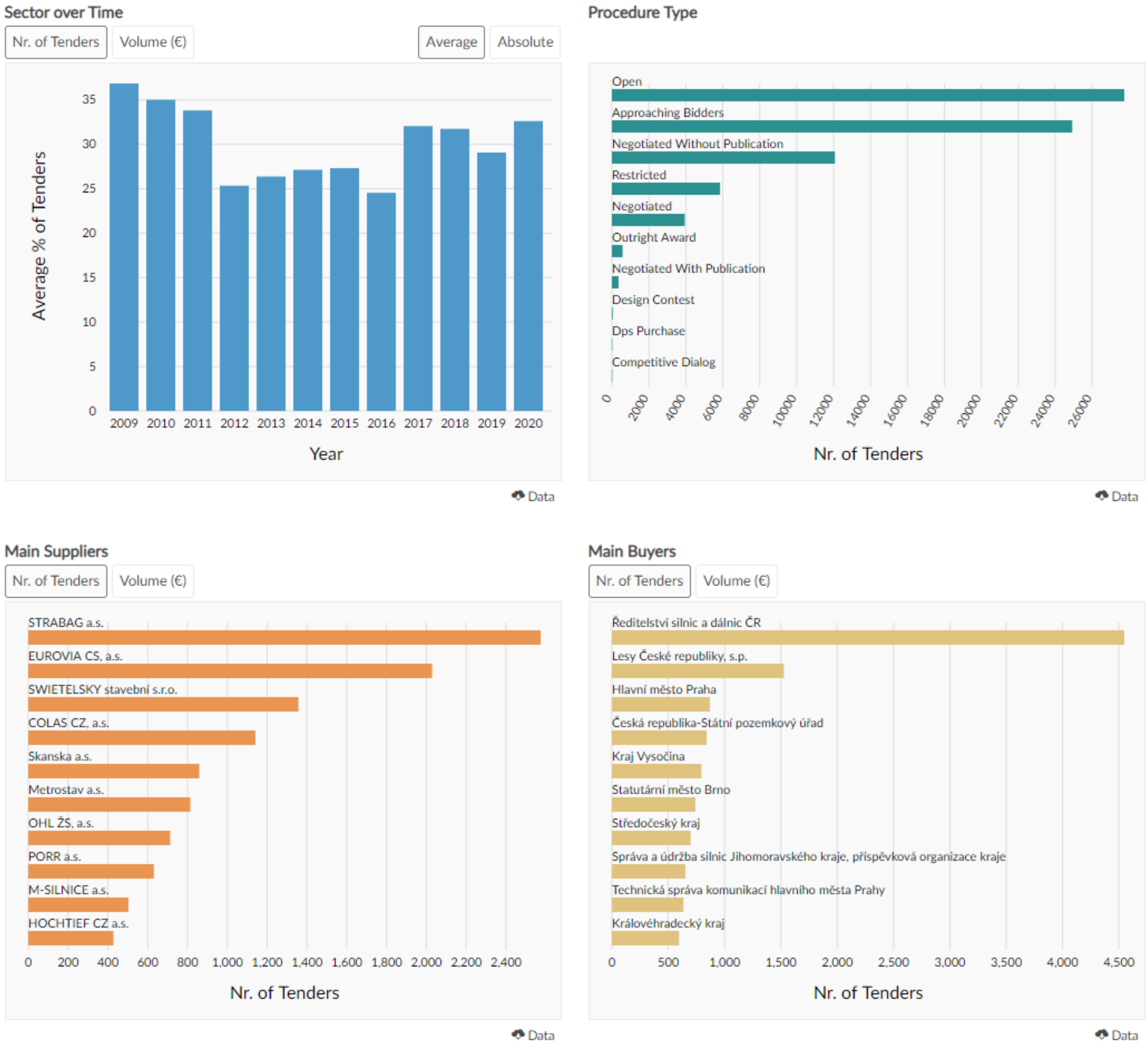
Figure 21: Market overview


Date Range
2009
2019
2020
2009 - 2020
Sector Overview
Volume (€)
Nr. of Tenders
Construction work
€ 62.5 Billion
Architectural, construction, engineering and inspection services
€ 4.0 Billion
Business services
€ 4.2 Billion
Petroleum products, fuel, electricity and other sources of energy
€ 11.0 Billion
Medical equipments, pharmaceuticals and personal care products
€ 5.1 Billion
Repair and maintenance services
€ 2.6 Billion
Transport equipment and auxiliary products to transportation
€ 6.0 Billion
Info
Data

Source: Opentender

The site also shows information about the number of tenders in the sector over time, the distribution of different procedure types, the name of the main (largest) suppliers and contracting authorities and the distribution of procurements by geographical locations. All of these graphs are interactive, making even firm level filtering possible.

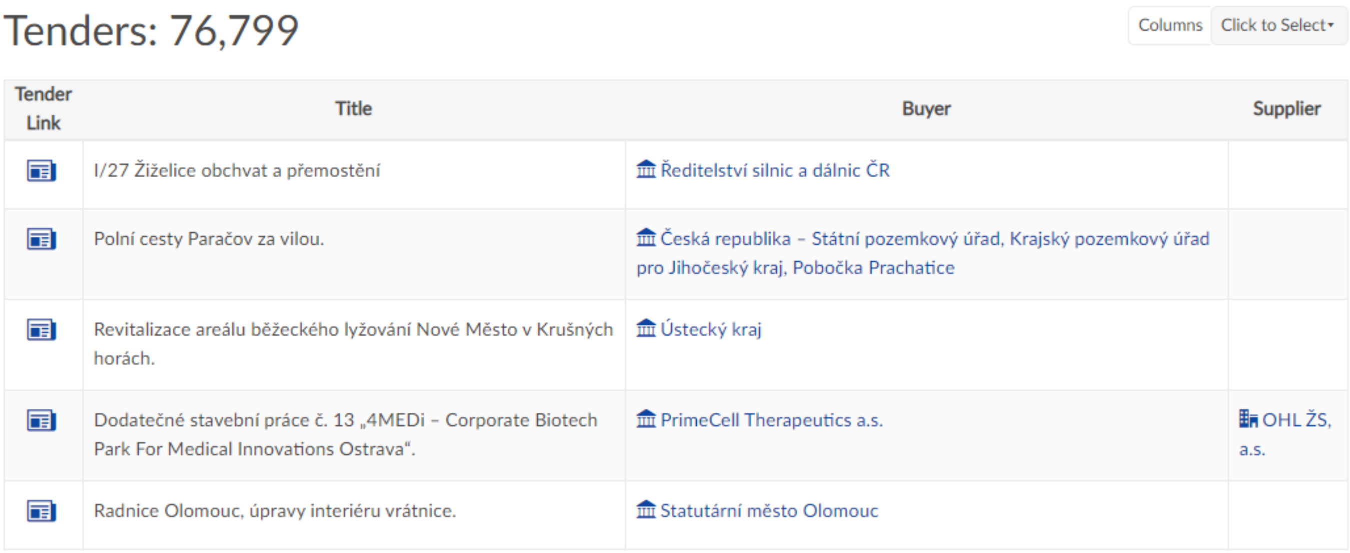
Figure 22: Procurement market specific charts



Source: Opentender

Additionally, after filtering down to a specific market, the analysis function also shows tenders in a table format. Extra variables can be added by selecting them from the 'Columns' option, and the table can be ordered using any of the variables. Similarly, to graphs, the filtered data can be downloaded into a CSV or JSON file.

Figure 23: List of tenders within a market segment



Tenders: 76,799

Tender Link	Title	Buyer	Supplier
▤	I/27 Žiželice obchvat a přemostění	Ředitelství silnic a dálnic ČR	
▤	Polní cesty Paračov za vilou.	Česká republika – Státní pozemkový úřad, Krajský pozemkový úřad pro Jihočeský kraj, Pobočka Prachatice	
▤	Revitalizace areálu běžeckého lyžování Nové Město v Krušných horách.	Ústecký kraj	
▤	Dodatečné stavební práce č. 13 „4MEDi – Corporate Biotech Park For Medical Innovations Ostrava".	PrimeCell Therapeutics a.s.	OHL ŽS, a.s.
▤	Radnice Olomouc, úpravy interiéru vrátnice.	Statutární město Olomouc	

Source: Opentender

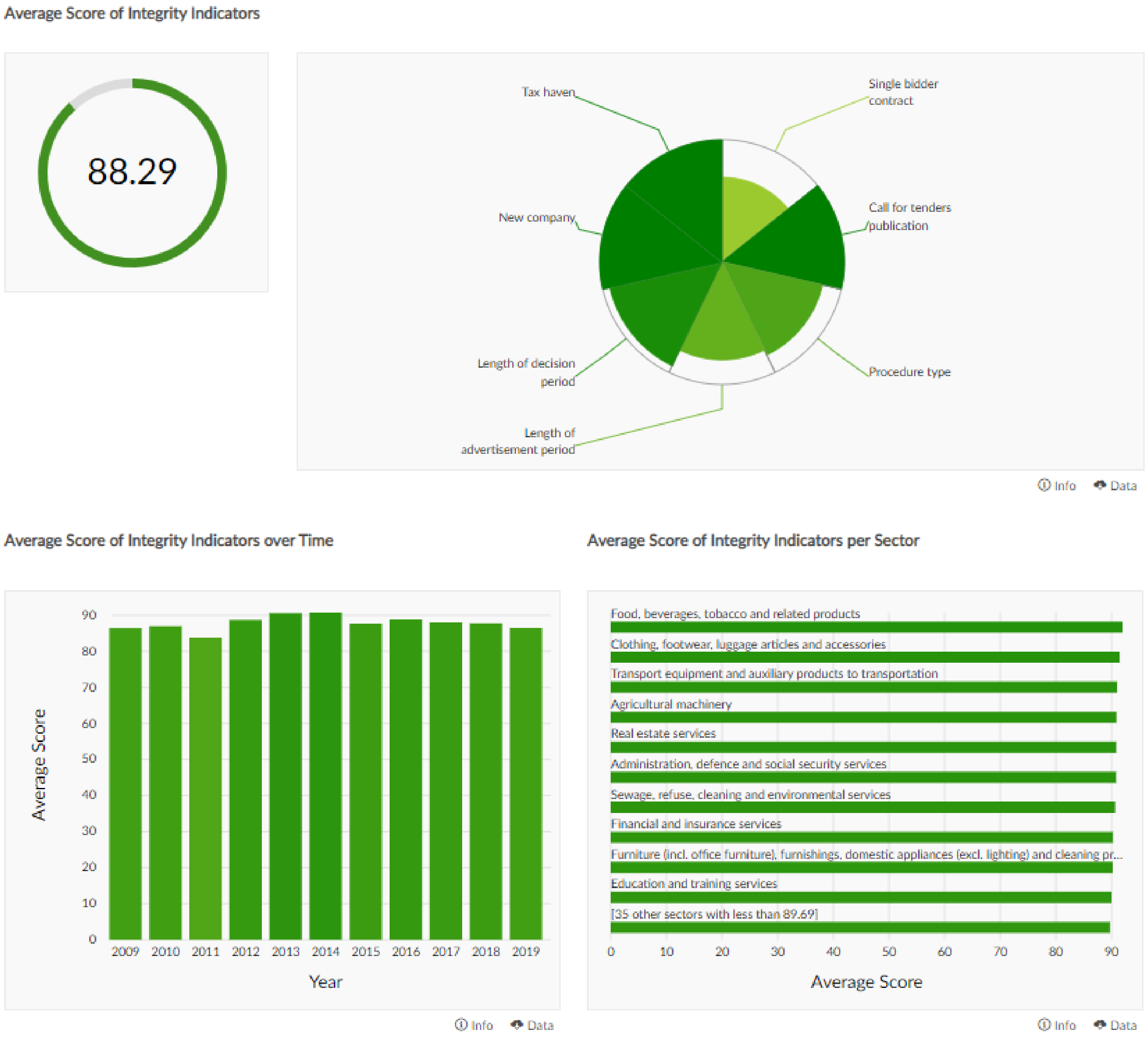
4.1.2. Indicator analysis

Although the manual only introduced Integrity Indicators, Opentender reports two additional groups of indicators. Administrative Capacity Indicators show the general characteristics of tenders. For example, whether a procurement is a framework agreement, if it was published on an electronic auction, or whether there were major discrepancies between the call for tender and the contract. None of these are corruption risk indicators, hence they can be mainly used to establish the general endowments of a procurement market, or a specific tender.

The second group of indicators are the Transparency Indicators. These measures essentially illustrate data quality, such as the availability of the bidder's name, contract value, subcontract information. While the lack of transparency can signal corruption, these are not quantitatively tested indicators. Nonetheless, for some procurement systems the lack of information is also a red flag and as such, it is also part of some of the Integrity Indicators which have already been introduced in the previous chapter of the manual.

Each indicator group can be accessed from the websites 'Dashboard' menu. As illustrated by the figure below. For each indicator group, the site shows the average and each component's score separately. It also depicts the scores' change over time and their average value in the main sub sectors. The data can also be filtered similarly to the market analysis section.

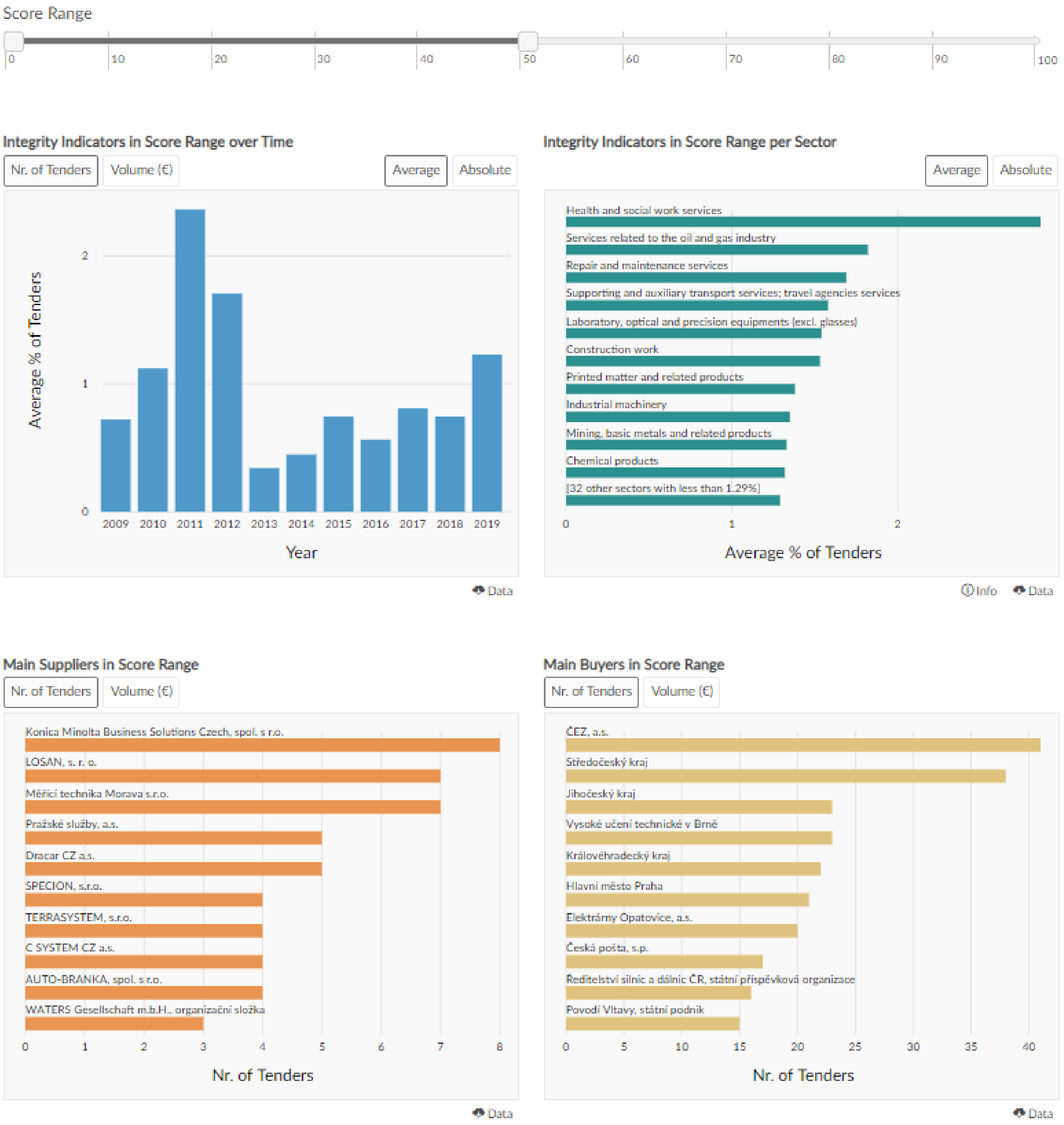
Figure 24: General description of the market using the Integrity Indicator


Average Score of Integrity Indicators
88.29
Tax haven
Single bidder contract
New company
Call for tenders publication
Length of decision period
Procedure type
Length of advertisement period
Info
Data
Average Score of Integrity Indicators over Time
Average Score
0
10
20
30
40
50
60
70
80
90
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year
Info
Data
Average Score of Integrity Indicators per Sector
Food, beverages, tobacco and related products
Clothing, footwear, luggage articles and accessories
Transport equipment and auxiliary products to transportation
Agricultural machinery
Real estate services
Administration, defence and social security services
Sewage, refuse, cleaning and environmental services
Financial and insurance services
Furniture (incl. office furniture), furnishings, domestic appliances (excl. lighting) and cleaning pr...
Education and training services
[35 other sectors with less than 89.69]
0
10
20
30
40
50
60
70
80
90
Average Score
Info
Data

Source: Opentender

Additionally, the indicators can also be filtered using the score range. This makes it possible to only inspect the poorest or best performing time periods, sectors, or firms.

Figure 25: Description of the market within a specific Integrity Indicator score range

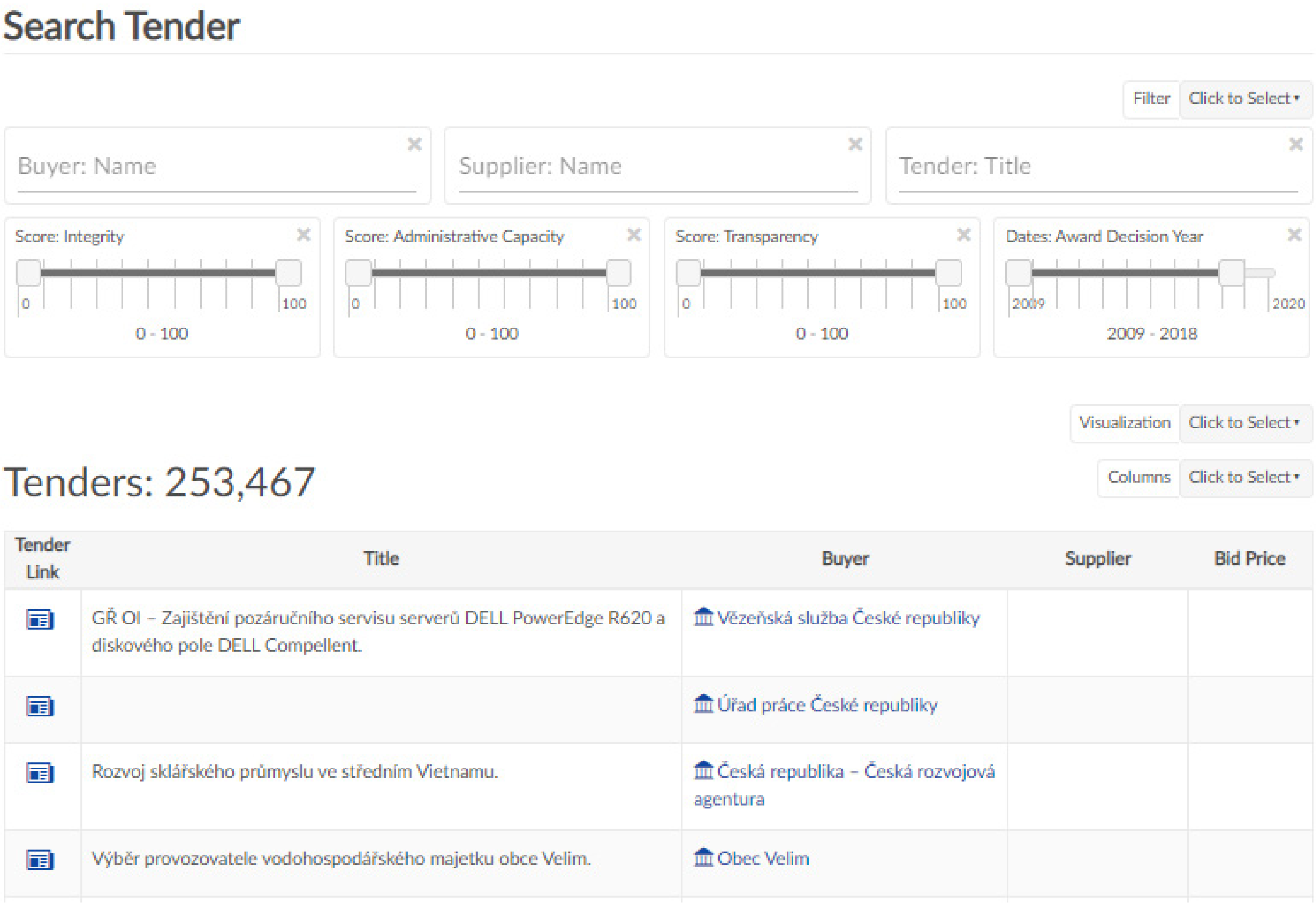

Score Range
0
10
20
30
40
50
60
70
80
90
100
Integrity Indicators in Score Range over Time
Nr. of Tenders
Volume (€)
Average
Absolute
Average % of Tenders
0
1
2
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year
Data
Integrity Indicators in Score Range per Sector
Average
Absolute
Health and social work services
Services related to the oil and gas industry
Repair and maintenance services
Supporting and auxiliary transport services; travel agencies services
Laboratory, optical and precision equipments (excl. glasses)
Construction work
Printed matter and related products
Industrial machinery
Mining, basic metals and related products
Chemical products
[32 other sectors with less than 1.29%]
0
1
2
Average % of Tenders
Info
Data
Main Suppliers in Score Range
Nr. of Tenders
Volume (€)
Konica Minolta Business Solutions Czech, spol. s r.o.
LOSAN, s. r. o.
Měřicí technika Morava s.r.o.
Pražské služby, a.s.
Dracar CZ a.s.
SPECION, s.r.o.
TERRASYSTEM, s.r.o.
C SYSTEM CZ a.s.
AUTO-BRANKA, spol. s r.o.
WATERS Gesellschaft m.b.H., organizační složka
0
1
2
3
4
5
6
7
8
Nr. of Tenders
Data
Main Buyers in Score Range
Nr. of Tenders
Volume (€)
ČEZ, a.s.
Středočeský kraj
Jihočeský kraj
Vysoké učení technické v Brně
Královéhradecký kraj
Hlavní město Praha
Elektrárny Opatovice, a.s.
Česká pošta, s.p.
Ředitelství silnic a dálnic ČR, státní příspěvková organizace
Povodí Vltavy, státní podnik
0
5
10
15
20
25
30
35
40
Nr. of Tenders
Data

Source: Opentender

4.1.3. Raw search

The search function makes it possible to search for tenders of specific bidder, buyer or look for tenders including any key words. Each of the available variables can be used as filters by ticking them in the ‘Filter’ option. Also, the procurement data can be submitted to a specific date or score range.

Figure 26: Layout of the raw search menu



Search Tender

Filter Click to Select ▾

Buyer: Name

Supplier: Name

Tender: Title

Score: Integrity 0 100 0 - 100

Score: Administrative Capacity 0 100 0 - 100

Score: Transparency 0 100 0 - 100

Dates: Award Decision Year 2009 2020 2009 - 2018

Visualization Click to Select ▾

Tenders: 253,467

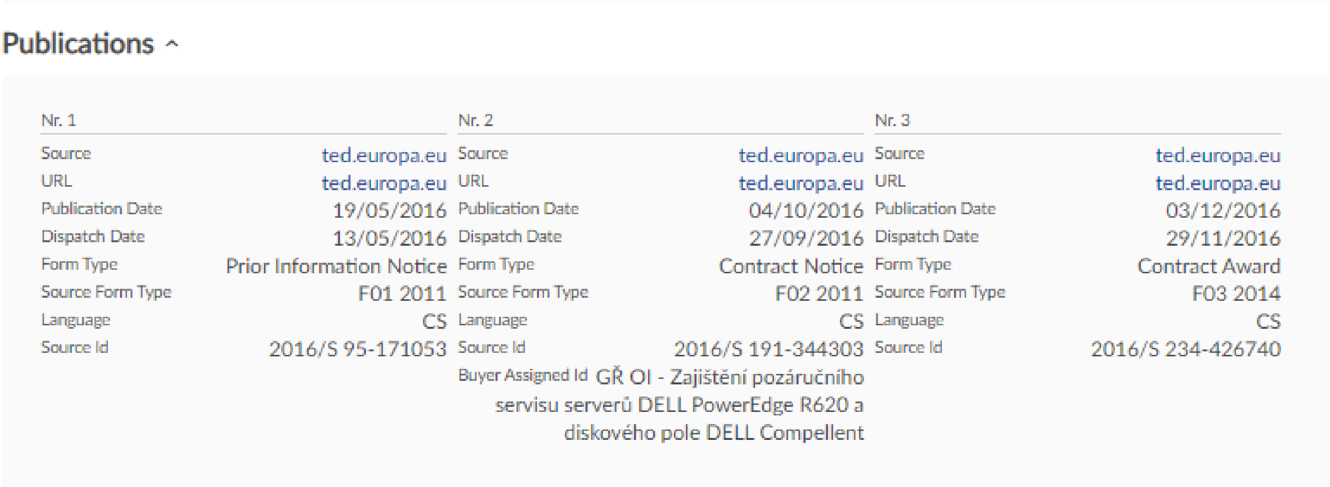
Columns Click to Select ▾

Tender Link	Title	Buyer	Supplier	Bid Price
▤	GŘ OI – Zajištění pozáručního servisu serverů DELL PowerEdge R620 a diskového pole DELL Compellent.	🏛 Vězeňská služba České republiky		
▤		🏛 Úřad práce České republiky		
▤	Rozvoj sklářského průmyslu ve středním Vietnamu.	🏛 Česká republika – Česká rozvojová agentura		
▤	Výběr provozovatele vodohospodářského majetku obce Velim.	🏛 Obec Velim		

Source: Opentender

By clicking on the ‘Tender Link’ the website brings up every available tender specific information including the values for each score group and raw information about the tender as it was collected from the source. It also provides some general information about the specific documents that were used to gather the necessary data.

Figure 27: Tender specific source of information



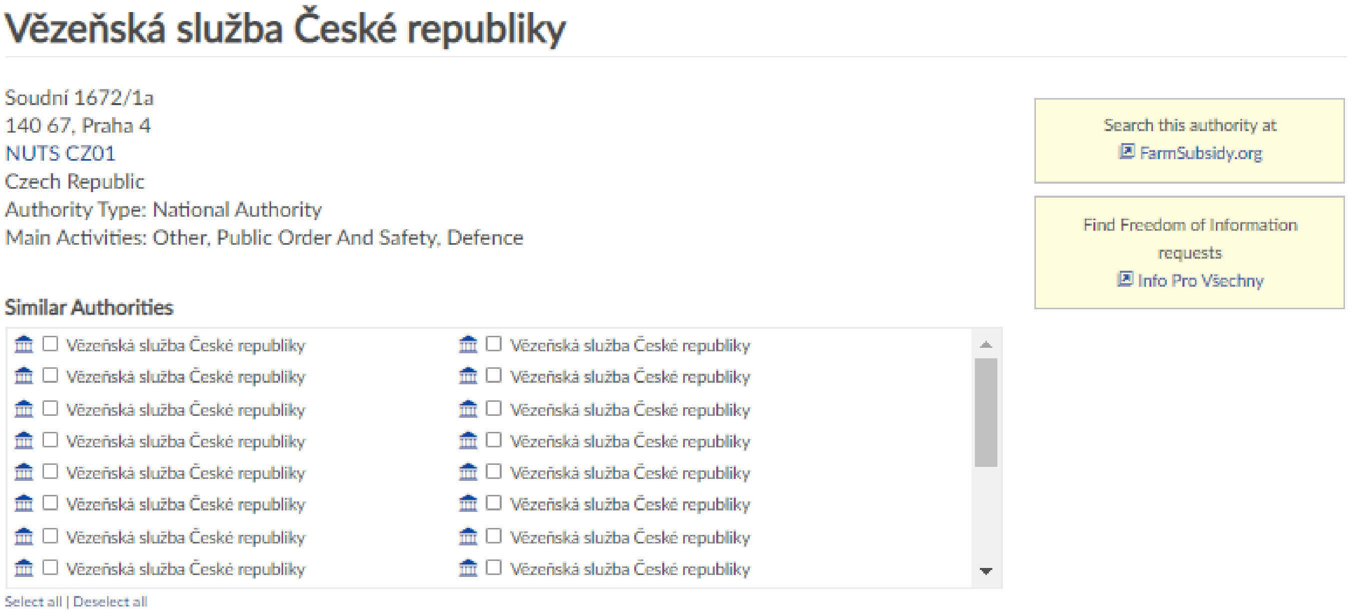
Publications ^

	Nr. 1	Nr. 2	Nr. 3
Source	ted.europa.eu	ted.europa.eu	ted.europa.eu
URL	ted.europa.eu	ted.europa.eu	ted.europa.eu
Publication Date	19/05/2016	04/10/2016	03/12/2016
Dispatch Date	13/05/2016	27/09/2016	29/11/2016
Form Type	Prior Information Notice	Contract Notice	Contract Award
Source Form Type	F01 2011	F02 2011	F03 2014
Language	CS	CS	CS
Source Id	2016/S 95-171053	2016/S 191-344303	2016/S 234-426740
Buyer Assigned Id		GŘ OI - Zajištění pozáručního servisu serverů DELL PowerEdge R620 a diskového pole DELL Compellent	

Source: Opentender

Specific buyer and bidder profiles can also be viewed by clicking on their name. These profiles show registry information and present a list of authorities or companies with similar names. The later function is important, because the lack of unique organization IDs could interfere with appropriate data filtering. By including similar authorities, the website automatically aggregates all available tender level information on each of them. Furthermore, the profile also presents the general descriptive statistics introduced in the market analysis section.

Figure 28: Organization profile



Vězeňská služba České republiky

Soudní 1672/1a
140 67, Praha 4
NUTS CZ01
Czech Republic
Authority Type: National Authority
Main Activities: Other, Public Order And Safety, Defence

Search this authority at
FarmSubsidy.org

Find Freedom of Information requests
Info Pro Všechny

Similar Authorities

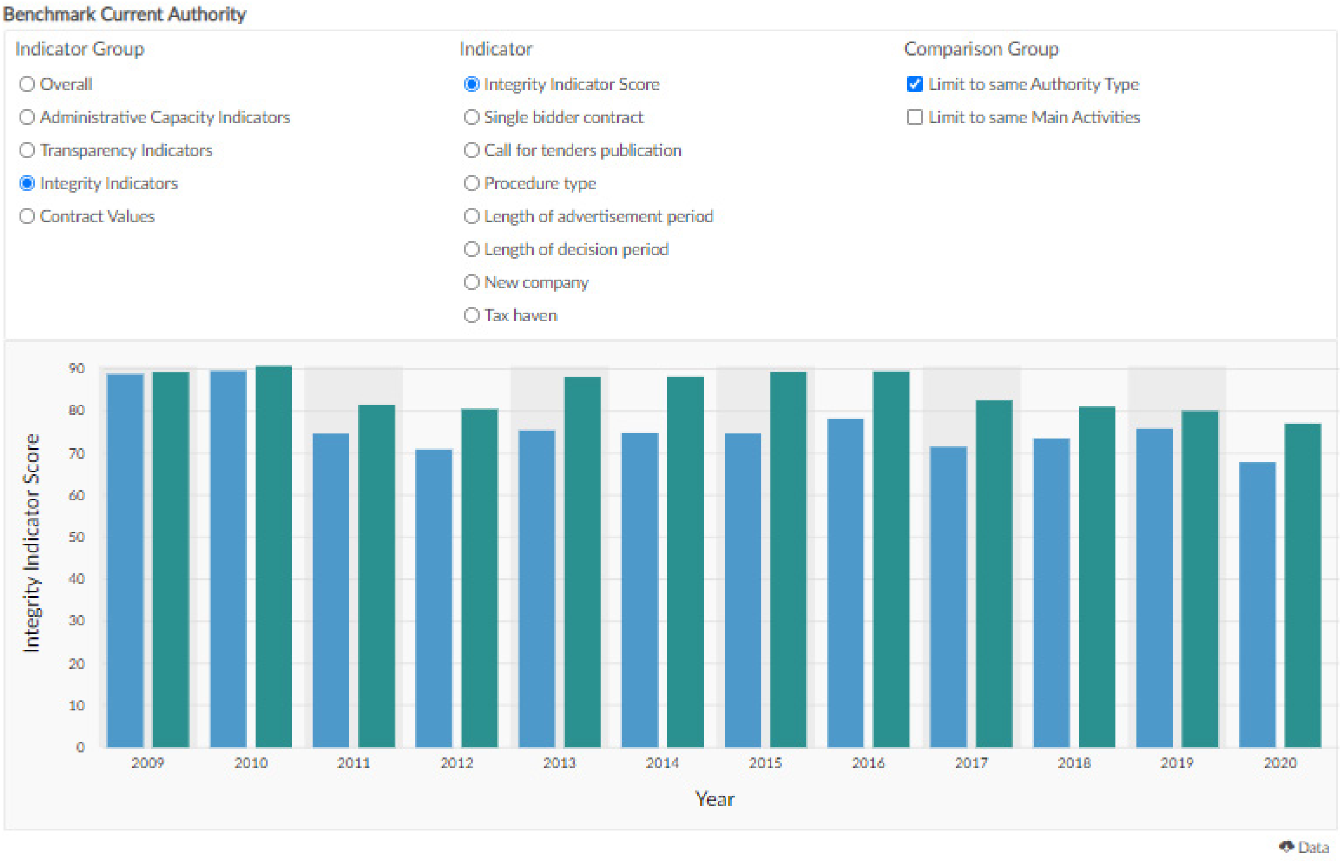
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky
- ☐ Vězeňská služba České republiky

Select all | Deselect all

Source: Opentender

Finally, interacting with the last chart makes it possible to benchmark the selected organization to other institutions based on its indicators. In the case of contracting authorities, the comparison group can be adjusted based on the authority type and on its main activities, while for suppliers the control group can be selected based on its markets of operation.

Figure 29: Organization level indicator comparison



Benchmark Current Authority

Indicator Group
○ Overall
○ Administrative Capacity Indicators
○ Transparency Indicators
● Integrity Indicators
○ Contract Values

Indicator
● Integrity Indicator Score
○ Single bidder contract
○ Call for tenders publication
○ Procedure type
○ Length of advertisement period
○ Length of decision period
○ New company
○ Tax haven

Comparison Group
☑ Limit to same Authority Type
□ Limit to same Main Activities

Source: Opentender

4.1.4. Procurement data download

The website's last functionality is bulk data download. The underlying data is collected from the Tender Electronic Daily, and from national websites utilizing data collection techniques introduced in Chapter 2.3.3. The datasets are updated every six months (or in some cases annually).

Using the 'Download' menu, country level datasets can be accessed in CSV or JSON formats. The JSON datasets contain every information that is listed in the Digiwhist data template, given that the procurement system provides the information. In contrast, the CSV files only contain the most important variables. The difference is due to the memory management of the two file extensions.

Sources

Czibik, Ágnes, Using Big Data in public procurement to detect corruption and collusion risks, presentation at SELDI‘s conference “Enhancing CSOs Advocacacy Efforts for Countering Corruption in Critical Sectors in Southeast Europe”, 2015.

Dávid-Barrett, Elizabeth - Fazekas, Mihály, Corrupt Contracting: Partisan Favouritism in Public Procurement. ERCAS Working Paper No. 49, Berlin: Hertie School of Governance, 2016.

Digiwhist, About the project.

EuroPam, Download datasets, 2020.

Fazekas, Mihály, Theory-driven approaches to identifying red flags of corruption in public procurement, presentation at the “Hands-on Workshop” on progress and future planning on using data to identify and address corruption in procurement, World Bank, Washington DC, 2015.

Fazekas, Mihály - Kocsis, Gábor, Uncovering High-Level Corruption: Cross-National Corruption Proxies Using Government Contracting Data, Working Paper series: GTI-WP/2015:02, Budapest, 2015.

Fazekas, Mihály - Cingolani, Luciana - Tóth, Bence, A comprehensive review of objective corruption proxies in public procurement: risky actors, transactions, and vehicles of rent extraction, Working Paper series: GTI-WP/2016:03, Budapest, 2016a.

Fazekas, Mihály - Tóth, István János - King, Lawrence Peter, An Objective Corruption Risk Index Using Public Procurement Data, European Journal on Criminal Policy and Research, Vol 22., No 01, DOI: 10.1007/s10610-016-9308-z, 2016b.

Fazekas, Mihály - Rhona, McNair - Jukic, Vinko, Corruption Risks in Public Procurement in the Western Balkans and Turkey, Regional Action against Economic Crime of the CoE/EU Horizontal Facility Programme for Western Balkans and Turkey – Phase II (AEC-REG), Draft, 2021.

Mendes, Mara - Fazekas, Mihály, DIGIWHIST Recommendations for the Implementation of Open Public Procurement Data An Implementer‘s Guide, 2015.

Nye, J. S., Corruption and Political Development: A Cost-Benefit Analysis. American Political Science Review, 61(02), 417–427. DOI:10.2307/1953254, 1967.

World Bank Integrity Presidency, Fraud and Corruption. Awareness Handbook, World Bank, Washington DC. pp. 7, 2009.

Implemented by:

CSD
CENTER FOR
THE STUDY OF
DEMOCRACY

SELDI.net
Southeast European Leadership for
Development and Integrity

Government
Transparency
Institute

Regional
Cooperation
Council