Diverse uses of government contracting data to improve spending of public funds

Dr. Mihály Fazekas* & Ágnes Czibik

* Corresponding author:mf436@cam.ac.uk; University of Cambridge

Abstract

Government contracting has experienced an explosion of available data in the last decade, marked by the rise of the global Open Contracting Data Standard. However, it remains largely under-utilized for Big Data analytics and embedding findings in policy making. In order to address this gap and promote the use of government administrative data for policy making, this paper provides a review on the availability, scope and quality of datasets in government contracting in 35 European countries and highlights prominent use cases to inspire policy makers and civil society. State of the art findings come from the ongoing Horizon2020-funded research project led by the University of Cambridge, called DIGIWHIST, which benchmarks, standardizes, and republishes public procurement data across Europe while also providing key performance indicators directly relevant for policy. Use cases demonstrate how civil society can use Big Data to hold governments accountable; and how governments can use advanced market analytics for detecting collusive competitors and safeguarding public spending.

Keywords – public procurement; government administrative data; corruption, collusion, data quality

1   The potential of Big Data approach in government contracting

Public procurement is the purchase by governments and state-owned enterprises of goods, services and works. It is one of the largest government spending activities in any country, representing on average up to 13% of GDP in OECD countries, and up to 29% of general government expenditure. At the same time, it is perceived to be the most corrupt government function across the globe ahead of justice or taxation rendering it a key driver behind popular discontent with governments and inefficient public spending. Even though governments throughout the globe, and across Europe in particular, are producing large amounts of administrative data describing public procurement contracts and tenders, this information has been left largely unused for research and policy purposes up until recent years. Applying a Big Data approach which combines diverse data sources is expected to unlock a whole new world for policy and research potentially contributing to better quality public services. Since public procurement involves millions of contract awards each year across Europe and is prone to corruption and budget deficit risks, high quality open data and Big Data analytics are indispensable for the efficient and accountable use of public resources.

However, a wide range of fundamental difficulties emerge when these datasets are actually put in use. On the one hand, a number of technical problems have to be sorted out before any analysis can take place. In most countries, public procurement data were not produced with the purpose of using it in statistical analysis, so even the simplest operations – like aggregating the contract values of a contracting authority over time – requires programming skills. The data need to be downloaded, extensively cleaned and restructured before one can start analysing the data.

On the other hand, more substantial problems occur regarding the legal context of public procurement. National and EU laws need to be thoroughly understood in order to interpret the data correctly. This is crucial especially when comparing countries or making comparisons over time.

Tackling these problems requires much effort both in terms of financial and human resources, which is beyond the power of most research groups, let alone ordinary users. This need for pre-processing public procurement databases, i.e. lowering the entry barriers for ordinary users, before wider, regular use can take place motivates the recent surge in research projects and government programmes in this field, including DIGIWHIST.

2 DIGIWHIST data and indicators

DIGIWHIST is a Horizon2020-funded research project, which was launched in 2015, led by the University of Cambridge. Its goal is to systematically collect, analyse, and broadly disseminate tender-level information on public procurement in 35 jurisdictions across Europe (EU28+). The project involves private and public actors to actively collaborate in improving the quality and scope of the data.

Data collection is carried out after a thorough review of the legislation related to public procurement, financial disclosure, and conflict of interest restrictions. The results of these review is also made available to the public (http://europam.eu/) and the comparative analysis of data content and quality already lead to a range of policy recommendations.

The collected public procurement data is linked to company and public organisation information on finances, ownership and management; and to information on mechanisms that increase accountability of public officials such as conflict of interest regulations.

Besides the downloaded, cleaned and standardised databases, DIGIWHIST will develop and display novel indicators on government contracting in terms of corruption risks, transparency and administrative quality, which will enable citizens to scrutinize and compare spending efficiency and quality among different government agencies both within and between countries.

The practical utilization of the project is supported by developing web portals, mobile apps, whistle-blower reporting functionality and risk assessment software for public administrations.

2.1 Data scope, content and quality

The core element of the DIGIWHIST data structure encompasses national public procurement datasets including data reported according to EU-wide rules. The collected electronic public announcements, ideally, contains the following information on each tender.

- Information on the contracting authority: name, official ID, address, contact info
- Information on the bidders and the winner supplier: name, official ID, address, contact info
- Information on the object of the procurement: description, CPV codes, location, number of lots in the contract
- Procedure type, awarding criteria, requirements on bidders, deadlines
- Information on documentation: free or subject to fee
- Funding: EU-funded or not
- Estimated and awarded contract values

- Completion report: quality, timeliness, and final, actual price
- Corrigenda, modifications, cancellations, if any

However, the list of available variables is much shorter in most European countries. Depending on diverse national legislation, only a fraction of the whole procurement cycle is covered by publicly accessible announcements. Completion reports are not public in most countries, while even contract award notices are not obligatory to publish everywhere[1].

The monetary thresholds (see Figure 1), above which the national public procurement law is applied in case of buying goods, have a decisive influence on the number of contract award notices on national public procurement portals, hence transparency of bidding markets and public spending. In addition, both the exact content of notices and the prevalence of missing data can vary greatly which again can hinder Big Data analytics.

The EU-wide Tenders Electronic Daily (http://ted.europa.eu/) could serve as a standard, Europe-wide comparative data source with its uniform reporting threshold and data structure. However, it only contains the large value contracts supposedly subject to Europe-wide interest of bidding firms, missing out on a large number of smaller contracts available in national databases.
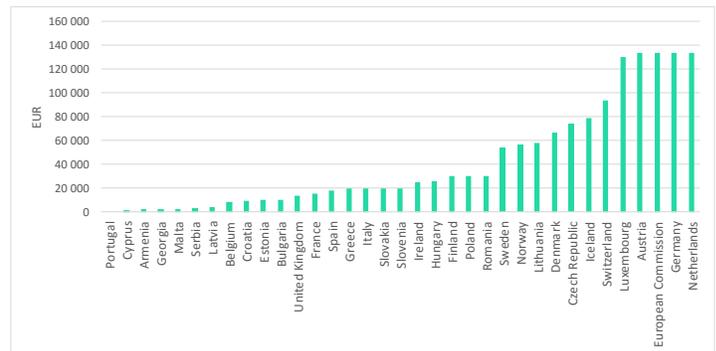

Figure 1. Thresholds of obligatory public procurement by countries

Data analysis crucially hinges upon linked data allowing for a comprehensive assessment of each public procurement tender throughout its various stages; hence linking procurement notices (e.g. call for tenders and contract award notices) of the same tender is imperative. Unfortunately, tenders usually don't have an official, unique identification number which appears on all related notices. By implication, notices must be matched to each other either using internal reference IDs or using more

---

[1] For more information, see: https://okfn.de/en/blog/2016/07/from-publication-to-award/

approximate methods relying on similarity in terms of publication date, title and the name of the contracting authority. Experience from a wide range of European countries suggests, unfortunately, that any approximate matching method is bound to a non-negligible margin of error reaching as high as third of tenders.

Assessing organisational behaviour throughout multiple transactions over time represents one of the most useful outputs of Big Data analytics in this field. In addition, linking public procurement data with other administrative datasets such as national company registries can considerably enhance the scope of data analysis. However, any such analysis presupposes that organisations (i.e. contracting bodies or bidding firms) are uniquely identified by their identification numbers and ideally organisational changes can be followed over time. Unfortunately, official identification numbers of contracting authorities and bidding companies are usually not included in public procurement notices, only free texts of organisation name and address. Classifying company names based on string similarities and assigning IDs by matching to official registries could help alleviating this shortcoming but, again, not with 100 percent correct result[2].

Once we obtained identification numbers assigned to organisation names other sources can be linked to the public procurement database like company financial data, owners and shareholders of companies, treasury accounts of public organisations, and list of political officeholders to identify companies' political ties. Although these administrative datasets are less complex than public procurement data, they raise different concerns. Some of them are not freely accessible (e.g. company registry data) and they are hard to standardise as they were created under different regulations (e.g. treasury accounts of countries). One of the main goals of DIGIWHIST is to standardise these diverse datasets and republish them in a format which is directly amenable for Big Data analytics, if it is legally permissible (e.g. some data points may infringe on individual privacy rights). Another goal is to offer clear interpretation of these data by providing performance indicators as discussed below.

2.3 Performance indicators

Using high quality datasets as outlined briefly above, a wide range of useful performance indicators can be developed by a combination of qualitative and quantitative methods to describe public sector transparency, administrative capacity, and quality of government on various analytical levels such as individual tenders,

organisations, regions, or government programmes. These indicators are based on objective, hard data produced by government administrative systems; they refer to specific behavioural patterns of companies and contracting authorities, as opposed to subjective perception-based surveys currently widely used in this domain. Considerably increasing their policy relevance, they can be calculated in real-time, allowing for timely policy interventions (e.g. halting payments to a company accompanied by high corruption risk bidding patterns).

Among many potentially useful and scientifically sound indicators in public procurement datasets, one key analytical innovation of DIGIWHIST is to measure the risk of institutionalised grand corruption in public procurement. This type of corruption aims to allocating the public contract to a favoured bidder by avoiding open and fair competition while also avoiding detection. Hence, such corruption necessarily results in restricting access to public resources to the many while granting privileged access to a few in spite of explicit rules and accepted norms against such behaviour.

Risk indicators can refer to the following elements of the procurement procedure.

- the tender: restricted access to contracts, e.g. by using tailored tender conditions,
- the contracting authority: political control of the bureaucracy, e.g. politically motivated appointment of executives,
- the suppliers: signs of risky businesses, e.g. tax haven registration, or politics-dependent market success
- political connections: direct/indirect political connections of contractors, e.g. same person owning the supplier and evaluating tenders.

We demonstrate the procedure of risk indicator development with the example of the tender-level Corruption Risk Index. First, we define a wide set of potential elementary risk indicators (like the weight of non-price award criteria, or single bidder contracts) based on interviews and desk research. Then these indicators are tested using regression models, to identify the ones which actually contribute to restricted competition across large datasets. 'Restricted competition' is marked by a single bid submitted on an otherwise competitive market or the winner company's high share in the contracting authority's total spending.

Elementary indicators are combined into a robust composite index, Corruption Risk Index, which takes values between 0 and 1, higher values meaning higher risks of corruption. This index has the advantage that it is more robust than single red flags and it can be adjusted to the varying legal, business or social context of countries in a

way that it remains comparable. Last but not least, it can be visualised in an intuitive and easy to understand way (Figure 2).

Figure 2 Corruption Risk Index Averages in Europe (TED, 2009-

2014)
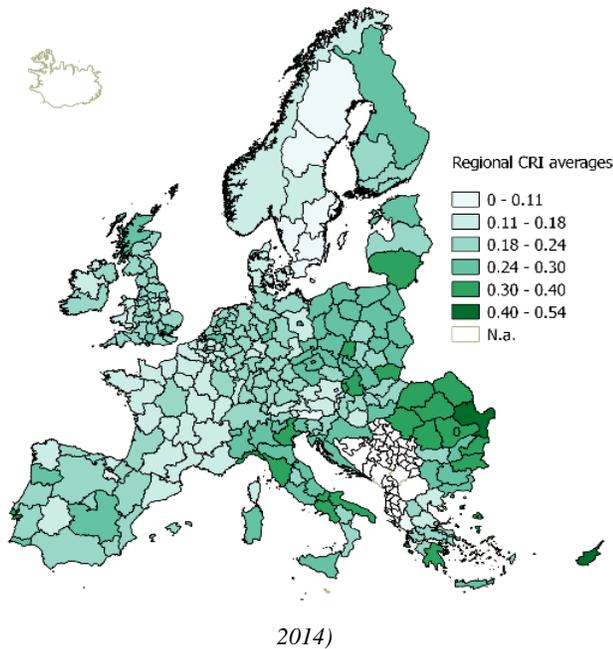
3 Use cases for civil society and governments

Several initiatives already exist which build on the analysis of public procurement data using a Big Data approach. Some of these projects inspired also DIGIWHIST. In the following, two use cases will be discussed where key stakeholders have benefitted from using Big Data in public procurement in order to inspire others to develop further practices and projects embedding newly available data and indicators in policy making.

3.1 Civil society driven watchdog websites

Civil society driven watchdog websites have been launched recently across the globe, for example in the Czech Republic (http://zindex.cz/), Hungary (tendertracking.eu) and Ukraine (https://prozorro.gov.ua/en/). These portals aim to holding governments accountable using indicators of good governance, corruption, and favouritism. What additional value these websites can offer compared to the official public procurement portals? While official websites' primary goal is to fulfil their legal obligations by

publishing administrative notices, these new portals focus on providing relevant data in the most easy to interpret format, most notably ready-made visualisations and key risk indicators.

These websites carry value for several types of users. Citizens can look up suspicious projects related to their area of interest; investigative journalists can spare time when collecting information on specific public procurement cases; potential suppliers can explore a new public procurement market in advance which they plan to enter. Even oversight bodies could benefit from using these websites to choosing which organisations to investigate in the situation of rare resources.

The Tendertracking website re-publishes information from Hungarian, Polish and Romanian (tendertracking.eu, pl.tendertracking.eu, and ro.tendertracking.eu) public procurement notices. Contracts can be found based on the name of authorities or companies, product identifiers, value of the tender, and the Corruption Risk Index value of the contract. The result is a list of contracts which fulfil the search criteria. Not only the most important data are published about each contract, but also red flags and indices showing corruption risks which helps interpreting data and drawing attention to the most suspicious tenders. Organisation-level aggregated data and time series are also available, displayed with simple visualisations. Authorities and companies can be compared to other organisations or a market average regarding the number and value of their contracts too.
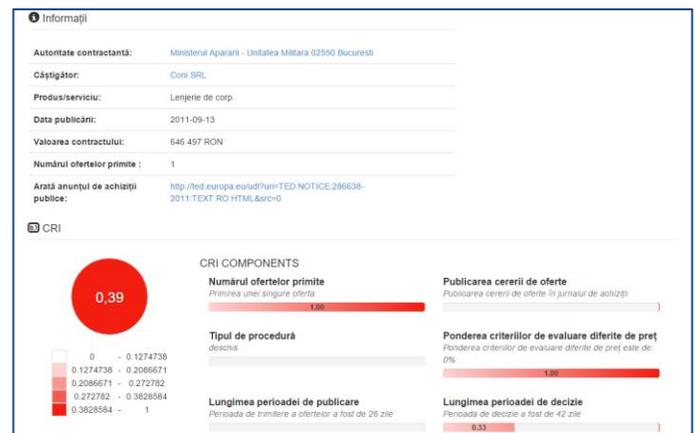
Figure 3: Basic information of a contracting authority and Corruption Risk Index displayed on Tendertracking

3.2 Governments detecting collusion among bidders

Governments increasingly use public procurement data in an innovative way to detect collusion among suppliers and

punish anti-competitive behaviour. Examples from countries such as Korea, Sweden, and Hungary underline the power of Big Data analytics for law enforcement in situations where informants and whistleblowers are of limited use.

Signs of collusive behaviour can be detected by analysing price-related variables like bid distribution characteristics; specific bidding patterns like bid rotation or bid suppression; or market structure-related variables such as market concentration.

Constructing co-bidding networks of public procurement bidders allows for differentiating healthy competition from potentially collusive bidding (Figure 4.). In a co-bidding network each vertex represents a bidding company and each tie is a tender where firms co-bid. Co-bidding clusters where most firms bid with all the others and many different firms win contracts suggests healthy competition on the face of it (see the green elliptical circle highlighting one such dense cluster of bidders). Whereas a firm winning many contracts while it bids with companies which always lose and only bid with this firm suggests a cartel formation (see three red elliptical circles highlighting such network formations with large green large vertices representing companies winning multiple tenders).
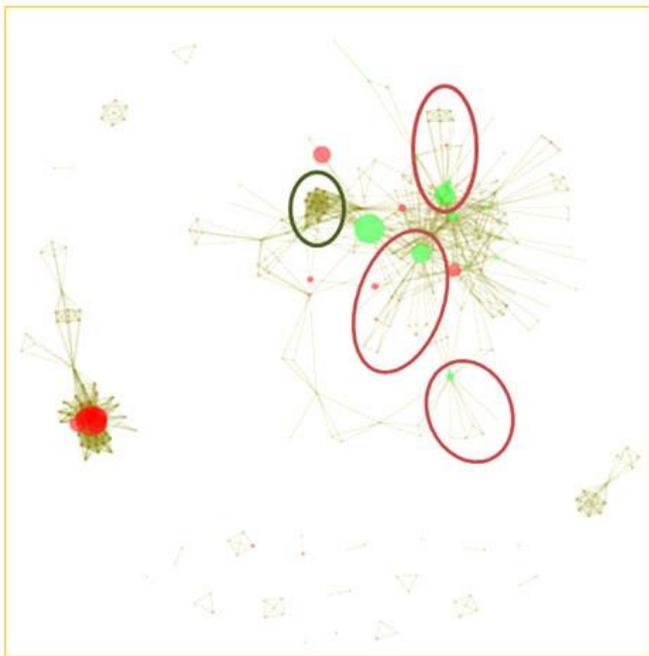

Figure 4.: Cartels - Some firms only bid together with a winner and lose recurrently

These methods can support the collusion screening work of competition authorities or other monitoring bodies. Identifying high risk markets and companies, where additional checks and investigation could be necessary, could improve transparency and efficiency of the public procurement market.

Acknowledgements

References

Fazekas, M., Tóth, I. J. & King, L. P., 2013. Anatomy of grand corruption: A composite corruption risk index based on objective data, Budapest: Government Transparency Institute. GTI-WP/2013:02

Tóth, B., Fazekas, M., Czibik, Á. & Tóth, I. J., 2014. Toolkit for detecting collusive bidding in public procurement – With examples from Hungary. Budapest: Government Transparency Institute. GTI-WP/2014:02

Cingolani, L., Fazekas, M., Kukutschka, R. M. B. & Tóth, B. 2015. Towards a comprehensive mapping of information on public procurement tendering and its actors across Europe. DIGIWHIST.