

Big Data in public procurement – a dialogue

Mihály Fazekas and Stephane Saussier

Published in Gustavo Piga & Tünde Tátrai (Eds.) *Law and Economics of Public Procurement Reform*. ch. 3. Routledge, London.

“Big Data in Public Procurement: Great opportunities – mediocre results” by Mihály Fazekas

This short memo exposes the enormous opportunities presented by the emergence of Big Data in public procurement and the lack of investment and effort for exploiting these opportunities. Big Data in public procurement holds the promise of fundamentally transforming how procurement performance is understood and it can provide a vastly superior guide to effective policy decisions and implementation compared to our current knowledge. However, Big Data implies large structured and high quality datasets which are typically not available in spite of extensive transparency regulations. Successfully harnessing Big Data also requires valid and easy-to-understand indicators to make sense of the enormous diversity revealed by it. In addition, making Big Data analytics part of daily research and policy making practice requires new skills and a change of culture (Fazekas, 2014).

Due to extensive regulations, the presence of multiple actors, and a demand for public scrutiny; public procurement has long been a data rich area of public spending. However, with the increasing use of electronic and online procurement tools, this rich set of administrative records has become more readily and more extensively available – giving rise to a data revolution experienced across many other domains of social life (UN Global Pulse, 2012). This enables real-time data analysis using datasets tracking individual behavior such as bids submitted to a tender, evaluation scores assigned, or invoices paid.

The opportunity

What fundamentally reconfigures our capacity to understand and govern public procurement systems is the move from reviewing individual records to analysing a structured database. This means that on top of the ability to identify and analyse a small set of public procurement documents and participating organisations; governments, businesses, citizens, and researchers are becoming increasingly capable of systematically analysing large swaths of procurement activities on a real-time basis.

Big Data in public procurement drastically reduces the cost of obtaining information for the kinds of analyses which were possible without such data. For example, accessing all the contracts and payments related to any particular project is a matter of seconds in an integrated public procurement database regardless where the project is taking place. This used to be a quite laborious task requiring a journey to the document archive of the procuring body overseeing the project, gathering the contracts and invoices in hardcopy, and manually summarizing information found in individual documents. In addition, databases allow for quick and complex analysis and queries previously simply not possible. For example, a public procurement database can rank all tenders according to their number of bidders in a given market all across a large geographical area while also listing contract values, delays in delivery, etc.

Such analytical possibilities give rise to a range of opportunities for harnessing Big Data in public procurement. Without being exhaustive:

- 1) Daily spend monitoring;
- 2) Risk-based audit;
- 3) Supporting system-wide policy decisions; and
- 4) Supporting government accountability through for example watchdog portals.

The challenge

In spite of the obvious informational needs and wide ranging benefits, making the move to a data rich approach has proven to be surprisingly challenging with some governments and international organisations even decreasing their ‘Big Data readiness’ rather than improving it (e.g. World Bank procurement reform of the mid 2000s disconnecting administrative systems and removing a number of key structured data fields in pursuance of administrative simplification).

First, building high quality integrated data systems, even if they only encompass the already collected information, requires considerable investment. Linking disparate procurement data systems, standardizing data formats, and delivering reliable data management infrastructures are typically far from straightforward exercises with many frontline civil servants opposing change. Crucially, building public procurement data systems require in-depth collaboration across disciplines such as law, political science, economics, and IT.

Capturing past data, typically recorded in semi-structured text files and building state-of-the-art data infrastructures going forward both require IT expertise and the understanding of complex data systems typically lacking in public sectors. As a result and quite unsurprisingly, most European ‘Big Data’ systems for public procurement publish data in non or only partially machine readable format making database construction very difficult: only Belgium, the European Commission, Italy, Norway, Poland, and the UK publish fully machine readable public procurement data (Cingolani, Fazekas, Kukutschka, & Tóth, 2015).

Systematizing data collection and publication had frequently revealed that reporting requirements are grossly neglected, making the most essential bits of data missing or incomprehensible. Interestingly, many well-governed countries such as Sweden or the UK extensively fail to report complete information in line with legal prescriptions (Figure 1). Improving data quality is a challenge on its own as examples from Canada¹, Czech Republic², the European Union³, and Hungary reveal (Czibik, Tóth, & Fazekas, 2015).

<FIGURE 1 HERE>

In addition, linking contracts data to databases holding related information is also challenging even though it promises considerable benefits, most notably the ability to trace investment outcomes back to procurement performance. Potentially linked databases are for example public financial management systems, treasury accounts of public bodies, company registry,

¹ <https://sites.google.com/site/do101mtl/seao/iqd-1>

² <http://www.profilny.info/>

³ http://www.open-contracting.org/digiwhist_big_data_meets_the_concerned_citizen and http://ec.europa.eu/internal_market/scoreboard/performance_per_policy_area/public_procurement/index_en.htm

financial and ownership data, and information on sectoral outcomes such as student achievement or mortality (Fazekas & Tóth, 2014). However, the generally lacking organisational ID numbers make any linking a highly labour intensive task at best and hopelessly impossible at worst (Figure 2).

<FIGURE 2 HERE>

Second, Big Data in public procurement also gives rise to the need for new, more advanced indicators which help diverse users in making sense of the often daunting diversity of data, recall public procurement includes purchases for anything ranging from nuclear power plants to school meals. Such new indicators of value for money and open access to public resources can potentially complement or in some cases replace traditional indicators of governance, by providing actionable and more objective insights (Knack, Kugler, & Manning, 2003). While there is a long way to go, there is already an emerging literature which develops, tests, and applies objective proxies of open access, corruption, and favouritism around the globe (Trapnell, 2015). These initial innovations, nevertheless already amply demonstrate the increasing capacity of indicators and impact evaluations to inform policy decisions and the wider public.

Third, parallel to the challenges of data and indicator development, users have to adapt their expectations and skills in order to effectively navigate in the new information landscape. High frequency rich datasets can become part of daily practice right across the public sector, businesses and civil society, if a new data culture is accommodated. Given the complex reality captured by Big Data, it requires careful thinking about and experimenting with different information systems in order to avoid information overload while keeping precise and interpretable information readily available to every level of decision making.

“Big and Open Data: a New Era for Public Procurement?” by Stéphane Saussier

Big Data is a term describing the large amount of data that inundates a business on a day-to-day basis. What matters is what organisations do with those data, not the volume itself. Nowadays, European countries are challenged by a turbulent economic situation, tightening their budget constraints and making the need for the reorganization of the public sector administration stronger than ever. There is also a huge defiance from citizens concerning public administrations. Faced with these challenges, the use of Big Data may give a fresh boost to the public sector, in particular. Its place in society is so important today that Big Data is considered as a key factor for improving future productivity. A McKinsey Global Institute (2011) study indicates that the use of Big Data may reduce the public administration's costs by potentially 15% to 20% (representing about €150 billion to €300 billion). This cost reduction would result from the generation of efficiency gains often through benchmarking and a reduction between the potential and actual collection of tax revenues.

In this short note we review what are the potential benefits of big and open data for economic actors involved in public procurement in particular as well as for the society as a whole. Big data may help at reducing contractual issues through several avenues well identified by contract theory, research on asymmetric information as well as incomplete contracting.

Interest for the public administration

- *Reducing asymmetric information issues*

By allowing the public authorities to observe the agent's behaviour on other attributed markets, big and open data could diminish adverse selection issues (i.e. it would be more difficult for inefficient firms to behave in such a way to be considered as potentially efficient). In addition, this would also present the advantage of disciplining the private operators by giving them incentives to respect the (often incomplete) contract and implicit rules.

Moral hazard could also be limited by giving tools to the public administration for using data as a benchmark (Shleifer, 1985; Auriol, 2000). Commonly named yardstick competition, this deals with the exploitation of the correlation between operators producing the same goods or services for reducing the asymmetry of information between the principal and the agent, assuming that the compared operators have identical cost structure. This allows for comparing the performance and efficiency of the private operators by introducing a virtual competition with other operators considered as basically similar.

- *Detecting aggressive bidding*

In order to win a public bid, a private operator might get incentives to bid aggressively while anticipating that he may be able to renegotiate the contract (Guasch, 2004, Esatche and Saussier 2014). Some studies show that the more likely a contract is to be renegotiated, the lower the bid will be. In that case, the utility of yardstick competition is also strong. Indeed, by getting access to bid data in a related market with similar characteristics, the public entity may be able to detect unreasonable offers. In addition, by fixing a minimum price based on what data on previous deals provide, it might help to deter cartels (Chassang and Ortner, 2016).

- *Increasing stakeholder's participation*

Stakeholders, especially users as well as consumer associations are pushing for more implication in the way public procurement are organized. Collecting, structuring and providing to stakeholders, adequate data about public procurement issues is a way for public authorities to increase citizens' involvement in public affairs (Spiller, 2015).

Interest for the private operator

- *Enjoy a reputational effect*

When companies respect their contractual agreements and the related implicit rules, they may be rewarded by benefiting from a reputational effect being more frequently renewed (especially when award procedures let some discretionary power to the public authority as in concession contracts – see Beuve and Saussier 2017). Indeed, data about a market attribution outcome may indicate the behavior and efficiency of the operator. Therefore, it would give more incentives to build a good reputation and preserve it, and avoid opportunistic renegotiation. In that case, both private operators and public entities may benefit from a winner-winner situation where the former might be more likely to win a public market and the latter to get the job well done.

Interest for regulators and academics

- *Analysis of the competitive environment*

Getting systematically information such as the number of bidders and their identity would be very interesting for assessing the competitiveness of the bidding process. In particular, it would be useful to detect whether large firms blockade entry. It would then be informative in terms of public policy to implement measures to improve competition in a market, if needed. For example, opening competition to small operators would improve the competitive process and the diversification of operators using the allotment. Amaral et al. (2013) analyzed calls for tender for the bus transportation in London, and concluded that the greater the actual or anticipated number of bidders, the lower the bids. On the reverse, in some specific sectors, some public authorities are restricting competition. Without any proper data it is not possible to assess the efficiency of such restrictions (Chever-Saussier 2017).

- *Detecting potential collusion*

So far, there is no direct and clear way of detecting collusion, but some suspicious pattern might be detected and should be considered carefully using data (Kaway and Nakabayashi, 2014; Clark et al., 2016). In that case, it would be important to get accurate information with the publication of market attribution such as the number of bidders, their identity and their bids. This would help to detect patterns of bidding such as bid rotation. Using yardstick competition is also a common way to assess the but-for world in a collusion setting. However, as stated by the OECD in the Competition and Procurement report (2011), there is a clear trade-off between collusion and corruption when dealing with transparency. Indeed, too much transparency may facilitate collusion, depending on the characteristics of the market.

Conclusion

By increasing transparency, big and open data are definitely two key factors that may help public procurement to be more efficient. The society in its entirety could benefit from it by procuring tools for detecting corruption, using statistical and econometric methods, and analyzing public policies implemented through public calls for tender. The first two principles stated in the *OECD Principles for Integrity in Public Procurement report* (2009) emphasizes the importance of transparency in the process of public procurement.⁴ Corruption may take on many forms since it can occur before (exchange of confidential information), during (modification of the public offer) or after the call for tender (risk sharing between public and private entities as well as renegotiation issues). Big and open data is an opportunity to bring more transparency and integrity in the procurement process. A study lead by the World Bank (2014) suggests several avenues how Big Data can help to assess corruption in public procurement by detecting in official public records and documents rather than the commonly used perceived amount of corruption. Issues at stake are huge: a study performed for the European Commission (2013) evaluated the overall direct cost of corruption in public procurement in 5 sectors⁵ and 8 country members⁶ to an estimated range between €1.4 billion and €2.2 billion.

Using to their advantage big and open data, public authorities may be more efficient by

⁴ The first principle states that the public entity should "provide an adequate degree of transparency in the entire procurement cycle in order to promote fair and equitable treatment for potential suppliers" and the second one that it should "maximize transparency in competitive tendering and take precautionary measures to enhance integrity, in particular for exceptions to competitive tendering"

⁵ Road and rail, water and waste, urban or utility construction, training, and research and development

⁶ France, Hungary, Italy, Lithuania, the Netherlands, Poland, Romania and Spain

reducing the asymmetry of information as well as detecting more easily aggressive bids. On the other side, private operators may enjoy a reputational effect tracked down by data and avoid the common trap of bidding too optimistically. Finally, big and open data would be useful for regulators and academics for giving them tools to assess the competitive environment of market as well as detecting patterns for potential cartels and bid riggings.

“Commentary to the note by Stéphane Saussier”

by Mihály Fazekas

Introduction

This short memo briefly recaps the main arguments presented by the essay of Dr Stéphane Saussier: *Big and Open Data: a New Era for Public Procurement?*; then it raises a number of issues and points worthy of further discussion. In particular, it exposes three main discussion points: 1) Is there big and open data in public procurement?; 2) How to avoid data overload and over-reliance on data analytics; and 3) How does big and open data change power relations in society?

The essay “*Big and Open Data: a New Era for Public Procurement?*” succinctly enumerates a wide range of potential benefits big and open data in public procurement promises for governments, bidding firms, civil societies, and users of public services more broadly. Most opportunities associated with big and open data stem from our supposedly increasing capacity to detect wrongdoing such as public corruption or inter-bidder collusion, and to identify best practices such as ways to run tenders better or select better suppliers. Almost every promised benefit assumes that it is the lack of information and insight which limits societies in doing better public procurement. However, acting on better analytics is only warranted if motivational structures are right. My below commentary explores all 3 steps of this impact mechanism: information, insight, and action. The issues raised point at potential harmful effects of big open data in public procurement.

Three fundamental challenges

Is there big and open data in public procurement?

Big and open data in public procurement is far from being a reality even in countries running well-developed public procurement data systems. Data needs to be comprehensive, encompassing the full tender cycle from needs assessment to contract completion. This wide scope exists in only rare cases in Europe and beyond. Even the most encompassing public procurement data system is of limited use if it is not structured (i.e. information captured and stored in a database format) and it doesn’t come with a clear data structure comprehensible for stakeholders and stable enough over time to use past data for informing future transactions. This again, is far from reality in most countries, think for example about frequent legal changes modifying expectations towards procurement actors, data reporting requirements and

variable definitions. Finally, data has to be open, meaning that it is machine readable, and accessible free of charge. Unfortunately, lot of times even structured data is published in a non-machine readable format or a fee is payable for data download and search. These deficiencies of the underlying information systems are sufficient to demonstrate the point that big and open data in public procurement is more a potentiality than reality even at its most basic level.

How to avoid data overload and over-reliance on data analytics

Even if data was perfect, using it to generate actionable insights is far from warranted. Among the numerous challenges this presents, I will expose only two: data overload and misinterpreting analytical findings thus overly relying on them. Big Data by definition implies a wealth of extremely detailed information on public procurement tenders, contracts and payments. For public administrations, bidding firms, and civil societies used to a “data-poor” debate and action, the new Big Data era can be overwhelming posing daunting challenges. If there is so much data and so many performance indicators that one doesn’t even know where to start, they are actually of very little use. Unfortunately, such Big Data confusion and the investments made to achieve it can cause considerable harm to society.

Furthermore, and probably much more worrisome, lack of experience with Big Data can lead to unrealistic expectations of what it can and cannot reveal. Without the necessary critical appraisal of analytical insights, over-reliance on data ensues diminishing expert judgement and ultimately efficient decision making. While big and open public procurement data can clearly produce a step change in our capacity to understand public procurement markets, it typically results in probabilistic statements with considerable margin of error, and often lacking the holistic understanding experts develop over the course of their professional careers. In an ideal scenario Big Data analytics complements expert judgement by allowing to exploit the best of both worlds. This however requires knowing the limitations of such analytical tools at the very minimum. Crucially, if some markets are volatile, full of outliers, and disruptive change is frequent, Big Data analytics might not be up to the task.

How does big and open data change power relations in society?

The very notion that it is the lack of knowledge of the best course of action which we lack is called into question by many scholars. If public procurement outcomes reflect entrenched interests, say rewarding loyal business groups for supporting political leaders, improved insights achieve close to nothing other than allowing the informal group to better govern its internal matters. Hence, assuming that governments will act on better evidence, or that bidding firms will improve their performance based on better data are wishful thinking in many contexts.

In reality big open data allows better decision making for those who lack public procurement insights and want and can act on it; while it also has the capacity to fundamentally alter power relations. New information is never power-neutral in a context such as public procurement which lies at the intersection of powerful interests and technical complexities. Think about voters learning from publicly available public procurement databases about public investment projects and the excessive successes of companies with personal ties to the prime minister or a mayor. Such information has led to the demise of a number of governments across the world from Italy to Brazil. Or take for example, collusive bidders using big open public procurement data for monitoring each other and hence increasing the stability of their cartel. This latter case again highlights the potential harmful results of big open public procurement data. Crucially, it is the ability to use big open data in public procurement and to act on it which

determines the changing power relations, those who can harness it better can gain an advantage either for the public's good or harm.

“Commentary to the note by Mihály Fazekas”

By Stéphane Saussier

Big Data is obviously opening the road to great opportunities for improving the efficiency of public procurement. However, I believe that several important steps have not been made in order to realise those opportunities. More precisely, I believe we have to carefully think about 1) What are the exact data needed to improve public procurement efficiency, 2) The potential limit of making such information public as well as 3) A clear analysis of what stakeholders have to win in this process. Those three questions are of course interrelated.

Without clear answers, available data will not be easy to use, not easy to manage indicators and no political willingness will emerge and as a consequence, Big Data in public procurement will only lead to mediocre results.

What are the exact data needed?

The new European Directives are pushing for more transparency (Directives 2014/24/EU): it will be compulsory for the information exchanged in public procurement procedures to be made paperless as of 18 October 2018. Furthermore, information has to be published, concerning the award procedures, the bidders, but also concerning renegotiations. However, it is not clear how far each country should go. For example, do public authorities have only to publish information about the fact that a procurement contract is renegotiated? Do the exact terms of the renegotiations have to be published as well? Should public authorities publish information about bidders? Bids? Should the marks received by every bidder on each criteria be published? As academics, we should say more about what information is needed to improve public procurement without hampering competition.

When more information hurts?

We should keep in mind that more public information is not always good for competition as more information may help sustain cartels. For example, the new European Directives are pushing for public authorities to publish their expected price before contract award. Such new information is interesting for academics in order to evaluate how aggressive are bids compared to what was expected by the public authorities. However, such information may also help companies to sustain cartels. As academics, we should say more about what information is needed to improve public procurement without hampering competition and favoring corruption.

What public officials have to win?

Big Data in public procurement is a revolution. It does not come without any cost. Public authorities need to develop new capabilities that they do not possess yet. As such, it can be analysed as an organizational innovation for which reluctances exist. What are the gains of

more transparency and more scrutiny for public officials? This question should be addressed carefully. The less is known about public procurement practices, the less public officials are challenged. The more is known, the more public officials are susceptible to be challenged by stakeholders that are not necessarily interested in the success of public procurement (e.g. political contesters -- (Beuve, Moszoro, and Saussier 2015; Spiller 2013)). More transparency induced by Big Data might then reduce public officials' discretion that is potentially part of public procurement efficiency (Coviello & al 2016, (Chever et Moore 2012; Saussier et Tirole 2015)).

In conclusion Big Data in public procurement comes with many promises at a moment when citizens are pushing for more transparency and more involvement in public decisions. Fazekas' note is rightly pointing out the many difficulties that still exist and that are ahead. Those difficulties are somewhat linked with how much transparency is really needed in public procurement. Academics, working on public procurements issues for many years with often very limited data should now be more explicit about what data they need and what they expect from it, taking into account risks of corruption and cartels as well as political agendas.

CONCLUDING REMARKS

Mihály Fazekas

“Big Data in Public Procurement: Great opportunities – mediocre results”

The two sets of memos agree on a wide range of points giving rise to a narrative which can be further explored in detail. Big Data in public procurement brings enormous opportunities, but just what Big Data means, to whom, how those opportunities are realised, and how some are negatively affected by such developments are questions which require a lot more thinking. Setting the right policies going forward requires explicitly recognising trade-offs between different aspects of the Big Data revolution and balancing the costs and benefits. This final note will briefly expose some trade-offs lying ahead.

Big Data brings transparency but not to the benefit of all

Transparency of bidding and contract implementation is generally welcome, but it may well also produce adverse consequences. As Stéphane Saussier pointed out, transparency can also support the maintenance of collusive rings or cartels (unless cartels have monitored each other effectively already before Big Data!). Hence, public demands for transparency supporting government accountability need to be balanced with a potentially increasing risk of collusion. Similarly, increased transparency may increase government accountability by increasing society's capacity to monitor civil servants which can also cripple innovation and push civil servants to opt for risk averse strategies rather than pursuing value for money.

Costly investments draw on scarce public resources

Government austerity implies that the resources to invest in public procurement data systems and train stakeholders to intelligently use the resulting rich datasets are scarce. Such scarce

resources could be spent on fixing schools or employment programmes to the long term unemployed. Demonstrating value for money for IT systems enhancements, hiring data scientists into the civil service and training ordinary civil servants in data use is essential if the Big Data revolution is to be realised.

More data changes the relationship between state and society

More information on how governments spend public money is never neutral to state-society relationships. More information on the machinery of the state can create more demands on civil servants which can in a worst case scenario distract the balanced functioning of the state (i.e. focus on what is measured and transparent to the neglect of those which are less visible, but potentially more important). Moreover, more information can, as expected by many, reveal wrongdoing which if at a large scale can bring complete governments to a halt (e.g. Petrobras in Brazil). Managing the changing state-society relationships and the demands on governments is more of an art than science, while it certainly plays a crucial role in deciding on the success of public procurement Big Data reforms.

Stéphane Saussier

“Big and Open Data: a New Era for Public Procurement?”

I believe that the above memos agree on many points. Maybe the only partial disagreement comes from what are the main difficulties associated with Big Data for public procurement. As Mihaly's notes insist on technical difficulties, I believe they might not be the main brakes of the Big Data adventure.

I certainly agree with the statement that big and open data in public procurement is far from being a reality. As an academic, much information I think is relevant for an economic analysis of public procurement efficiency in Europe and is missing. However, I would not stress so much technical difficulties (data has to be open, machine readable,..) as in Mihaly Fazekas' note to explain this shortage of data. My feeling is that technical limits that exist today might be overcome as soon as new business models will be identified, producing new services / innovations of value for stakeholders.

I stated in my previous notes that I believe the main challenge are coming more certainly from the fact that

1. Many stakeholders are potentially interested in the Big Data adventure when public procurement is concerned but their objectives are not necessarily aligned in the short term.
2. What is the exact information needed to help academics, firms, public authorities, more broadly stakeholders, is not clear.
3. Even more unclear as soon as you recognized that it is not true that the more information there are, the better it is.

Those three challenges might necessitate at some point some regulation concerning the way Big Data is produced and used. Raw information is not (rarely) sufficient to inform

stakeholders and more especially citizens and might even be misleading. As an example, the right2water European Citizen Initiative provoked discussion between stakeholders concerning who is willing to open their data at what condition, for what kind of benchmarking exercises. Very soon the discussion was turning around the fact that it might not be to the advantage of private water companies to open their data and to let them open to the public without any “treatment”: some econometric studies showed that on average private companies are pricing water at a higher price than municipalities that decided not to go private but that this difference does not exist anymore as soon as you compare public and private prices all things being equals (Chong et al 2015). Raw data might thus be misleading if no explanation or statistical analysis is provided together with the data.

References

- Amaral et al. (2013): Expected Number of Bidders and Winning Bids: Evidence from the London Bus Tendering Model. *Journal of Transport Economics and Policy*, vol. 47, No. 1., pp. 17-34.
- Auriol, E (2000): Concurrence par Comparaison. *Un Point de Vue Normatif, Revue économique*, Vol. 51, No. 3., pp. 621-634.
- Beuve, J., Moszoro M. W. and Saussier S. (2015): Political Contestability and Contract Rigidity: An Analysis of Procurement Contracts. Rochester, NY: Social Science Research Network. SSRN Scholarly Paper. <http://papers.ssrn.com/abstract=2475164>.
- Beuve, J. and Saussier, S. (2017): Renegotiations, Discretion and Renewals of Public-Private Contracts, Working paper.
- Chassang, S. and Ortner, J. (2016): Collusion in Auctions with Constrained Bids: Theory and Evidence from Public Procurement. Unpublished Manuscript.
- Chever, L. and Moore J. (2012): Negotiated Procedures Overrated? Evidence from France Questions the Commission’s Approach in the Latest Procurement Reforms *European Procurement & Public Private Partnership Law Review* 7(4): 228-41.
- Chever, L. and Saussier, S. (2017), The law of small numbers: investigating the benefits of restricted auctions for public procurement”, *Applied Economics*, Forthcoming.
- 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. Cambridge, UK.
- Clark et al. (2015): Bid rigging and entry deterrence: Evidence from an anti-collusion investigation in Quebec. <https://www.economics.utoronto.ca/index.php/index/research/downloadSeminarPaper/63908>.
- Chong, E., Saussier, S., Silverman, B. (2015) “Water under the Bridge: Cith Size,

Bargaining Power, Price and Franchise Renewals in the Provision of Water”, *Journal of Law, Economics and Organization*, Volume 31, Number 1, 3-39.

Coviello, D., Guglielmo, A. and Spagnolo, G. (2016): The effect of discretion on procurement performance. CEP Discussion Paper, No 1427, Centre for Economic Performance, London School of Economics and Political Science, London, UK.

Czibik, Á., Tóth, B., & Fazekas, M. (2015): How to Construct a Public Procurement Database from Administrative Records? With examples from the Hungarian public procurement system of 2009-2012. Budapest.

Estache A., Saussier S. (2014), “Public-Private Partnerships and Efficiency: A Short Assessment”, *CESifo DICE Report* , 2014, Volume 12, Number 3, p.8-13.

European Commission (2013): Public Procurement: costs we pay for corruption. http://ec.europa.eu/anti_fraud/documents/anti-fraud-policy/research-and-studies/pwc_olaf_study_en.pdf.

Fazekas, M. (2014): The use of “Big Data” for social sciences research—an application to corruption research. In *SAGE Research Methods Cases*, London: SAGE Publications Ltd.

Fazekas, M., & Tóth, I. J. (2014): New ways to measure institutionalised grand corruption in public procurement (No. 2014:9). U4 Anti-Corruption Resource Centre, Bergen, Norway.

Kaway, K., and J. Nakabayashi (2014): Detecting Large-Scale Collusion in Procurement Auctions. <http://isites.harvard.edu/fs/docs/icb.topic1465230.files/Kawai-Kei.pdf>.

McKinsey Global Institute (2011): Big data: The next frontier for innovation, competition, and productivity. <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>.

OECD (2009): OECD Principles for Integrity in Public Procurement. <http://www.oecd.org/gov/ethics/48994520.pdf>.

OECD (2011): Competition and Procurement. <http://www.oecd.org/regreform/sectors/competitionandprocurement-2011.htm>.

Saussier S. and Tirole J. (2015): Strengthening the Efficiency of Public Procurement, Report for the French Council of Economic Analysis. April. <http://www.cae-eco.fr/IMG/pdf/cae-note022-env2.pdf>.

Shleifer, A. (1985): A Theory of Yardstick Competition. *Rand Journal of Economics*, Vol. 16, No. 3., pp. 319-327.

Spiller, P. (2013): Transaction cost regulation. *Journal of Economic Behavior & Organization* 89: 232-42.

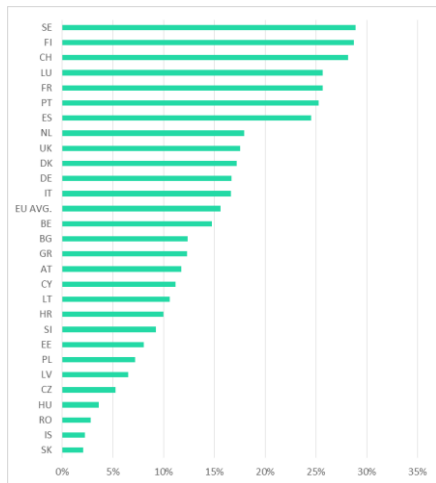
The WorldBank (2014): Measuring Corruption Risk using ‘Big’ Public Procurement Data in Central & Eastern Europe. <http://blogs.worldbank.org/governance/measuring-corruption-risk-using-big-public-procurement-data-central-eastern-europe>.

Trapnell, S. (2015): User's Guide to Measuring Corruption and Anti-Corruption. New York: United Nations Development Programme.

UN Global Pulse (2012): Big Data for Development: Challenges and Opportunities. New York: UN Global Pulse.

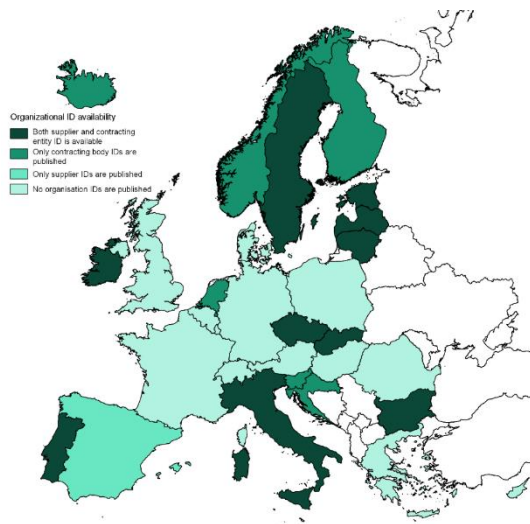
Figures

Figure 1. Average % of missing information among mandatory fields⁷. 2009-2015 (missing or erroneous data, based on TED data)



Source: own calculations based on Tenders Electronic Daily data (<https://open-data.europa.eu/en/data/dataset/ted-1>)

Figure 2. Use of unique organisational identifiers by European national public procurement systems, August 2015



⁷ 19 mandatory fields were checked for missing information: contracting body name, contracting body address, contracting body settlement name, contracting body postcode, winner name, winner address, winner settlement name, winner postcode, winner country, procedure type, main CPV code, NUTS code, use of EU Funds, type of assessment criteria used, contract award date, number of bids, contract value, and use of subcontracting.