

*Ágnes Czibik<sup>1</sup>, Elizabeth Dávid-Barrett<sup>2</sup>, Mihály Fazekas<sup>3</sup>, Danny Parsons<sup>4</sup>,  
David Stern<sup>5</sup>, Roger Stern<sup>6</sup> and Balázs Szendrői<sup>7</sup>*

# Tanzania's Procurement Data Infrastructure: Observations and Recommendations

**Government Transparency Institute reports  
GTI-R/2017:01**

May 2017, Budapest, Hungary

---

<sup>1</sup> Government Transparency Institute, [aczibik@govtransparency.eu](mailto:aczibik@govtransparency.eu)

<sup>2</sup> Department of Politics, University of Sussex, [E.David-Barrett@sussex.ac.uk](mailto:E.David-Barrett@sussex.ac.uk)

<sup>3</sup> Department of Sociology, University of Cambridge, [mf436@cam.ac.uk](mailto:mf436@cam.ac.uk)

<sup>4</sup> Supporting African Maths Initiatives, [danny@aims.ac.za](mailto:danny@aims.ac.za)

<sup>5</sup> Statistical Services Centre, University of Reading, [d.a.stern@reading.ac.uk](mailto:d.a.stern@reading.ac.uk)

<sup>6</sup> Statistics for Sustainable Development, [r.d.stern@stats4sd.org](mailto:r.d.stern@stats4sd.org)

<sup>7</sup> Mathematical Institute, University of Oxford, [szendroi@maths.ox.ac.uk](mailto:szendroi@maths.ox.ac.uk)

## Introduction

The publication of data about government procurement is widely perceived to be beneficial in terms of improving transparency over a significant area of public spending. Transparency in public procurement is associated with greater scrutiny over processes and outcomes, helping to improve accountability in order to achieve enhanced competition, better value for money and reduced corruption. As a member of the Open Government Partnership since 2011, the Tanzanian government has consistently emphasised commitments to establishing an open data system and has made good progress in this regard.

As part of recent research on Statistics for Integrity in Public Contracting, our team collected Tanzanian public procurement data from the Public Procurement Regulatory Agency's (PPRA) online bulletin<sup>8</sup> and the Tanzanian Procurement Journal<sup>9</sup>. Overall, we recorded 14,248 contract award notices and 9,292 calls for tender, and 496 procuring entities. Having undertaken this exercise, we are in a position to offer some observations on how Tanzania's procurement data infrastructure currently functions and recommendations on how it might be improved in the future.

Data collected as part of our research is available for free download at

<http://www.govtransparency.eu/index.php/2017/05/03/data-publication-public-procurement-in-tanzania-2009-2016/>

## Aims and objectives of our project

The ultimate aim of our research project, in particular the collection of Tanzanian national public procurement data, was to explore ways to increase transparency and integrity through Big Data solutions.

Our first objective was to apply a corruption risk assessment methodology developed in Europe<sup>10</sup> to other contexts. This research involves mining large amounts of government contracting data across many countries in search for suspicious patterns widely associated with corruption such as tailored bidding conditions and a single bid submitted on a market with multiple potential bidders<sup>11</sup>. As a test case, we wished to explore patterns in Tanzanian government contracting data. This was possible only to a very limited extent, given deficiencies in the publicly available data.

Our second aim was to incorporate this corruption risk assessment methodology into the statistical software R-Instat, in order to allow users greater flexibility for defining red flags and analysing corruption risks down to the level of individual organisations and contracts. R-Instat<sup>12</sup>, currently being developed by African Maths Initiative<sup>13</sup>, is a free and open source statistics software which provides an easy-to-use menu driven interface to the popular statistics language, R. Through this project,

---

8 <http://tenders.ppra.go.tz/>

9 [http://tenders.ppra.go.tz/portal/index.php?view=library\\_doc&opt=opt&type=4&search=q](http://tenders.ppra.go.tz/portal/index.php?view=library_doc&opt=opt&type=4&search=q)

10 <http://digiwhist.eu/>

11 See for example: <http://www.govtransparency.eu/index.php/2015/11/15/uncovering-high-level-corruption-cross-national-corruption-proxies-using-government-contracting-data/>

12 <https://chuffed.org/project/africandatainitiative>

13 <http://www.africanmathsinitiative.net/>

specific tools for analysing government contracting data have been incorporated into R-Instat's menu system, hence there is now potential to open up access to the methodology to wider audiences.

This research was supported by an EPSRC Global Challenges Research Fund Institutional Sponsorship 2016 grant held at the University of Oxford (Project: D4D01260, Task: BKA1.02).

## Observations

1. Although a great deal of procurement data are made available on the PPRA website,<sup>14</sup> they pertain to only a subset of all publicly announced calls for tenders and contract awards. However, it is unclear what rules govern publication on the website, and even more importantly whether there is any monitoring of publication discipline.
2. Additional data are available from the Tanzanian Procurement Journal but only in non-uniform PDF format.<sup>15</sup> A combination of manual and automated methods was applied to retrieve data from the pdf files and convert into a form that could be analysed. This is a laborious and time-consuming process that could easily be avoided by collecting and publishing the data in a consistent format on the already widely used PPRA website.

**TABLE 1. NUMBER OF CONTRACT AWARDS AND CALLS FOR TENDER ON VARIOUS SYSTEMS**

	Contract awards	Call for tenders
PPRA website (downloaded in December 2016)	3061	7355
Tanzania Procurement Journal (pdf, 2009-2016)	15047	9292

3. The calls for tenders and contract award notices do not use consistent numerical identifiers, making it difficult to match them up. We were only able to match them up in 3% of cases. After extensive manual search, it is highly likely that matching of call for tenders and contract awards is not only hampered by inconsistent IDs, but also by missing announcements (e.g. call for tender not being published prior to a contract award).
4. The names and identities of organisations - both procuring entities (buyers) and companies (suppliers) - are currently entered as free text rather than with standardised numerical identifiers, meaning that multiple versions of the same organisation may appear owing to different characterizations of names (or even linguistic variation in spelling conventions or simply typo errors). Again, this makes it difficult to analyse the practices and outcomes associated with particular organisations.

<sup>14</sup> <http://tenders.ppra.go.tz/>

<sup>15</sup> [http://tenders.ppra.go.tz/portal/index.php?view=library\\_doc&opt=opt&type=4&search=q](http://tenders.ppra.go.tz/portal/index.php?view=library_doc&opt=opt&type=4&search=q)

5. The number of variables for which data is published is relatively small. Contract award notices typically contain information on: (i) the ID number of the contracts; (ii) the name of the procuring entity; (iii) the title of the tender; (iv) the procurement method; (v) the name of the winner; (vi) the contract value; and (vii) the award date. In around two-thirds of cases, we also obtained information on: (viii) type of tender (works, goods, consultancy services, non-consultancy services); (ix) name and address of buyer; (x) ID number of the tender; (xi) eligible firms, and (xii) the source of the fund.
6. There are problems with missing data - i.e., variables are listed but data have not been inputted. This provides an incomplete picture which may be a source of systemic bias in any analysis.

## Recommendations

1. Publish all data in one place, ideally the PPRA website, in machine-readable format such as CSV, JSON or XML to ensure use-ability. Users should be also able to download data in bulk either as .csv or through an API.
2. Set clear rules for procuring entities to collect and provide relevant data in a consistent and timely manner.
3. Use numerical standardised identifiers for all tender announcements and contract award notices to ensure that they can be linked for purposes of analysis.
4. Expand the number of variables for which data are collected to allow more detailed analysis of public procurement performance. In particular, it would be important to include a variable on the number of bids submitted, the detailed description of the procured items, or the assessment criteria.
5. Use numerical and standardised identifiers for organisations, both buyers and suppliers, in addition to their names.
6. Publish information on amendments, modifications, and failed tenders in a structured and reliable format so that up-to-date information is available on all tenders.
7. Facilitate matching with other public datasets, e.g., it should be possible to match procurement data with budgets or other public financial management data, company registry data, court rulings.
8. Establish an effective monitoring and control system to ensure that procuring authorities comply with these requirements to collect and publish data.