Using Data and Transparency to Fight Corruption in Public Procurement

Stephen Zimmermann

WORLD BANK GROUP
Corruption in Public Procurement

In public procurement, the aim of corruption is generally to steer the contract to a favored bidder without detection.
This can be done in a number of ways, including:

- Tailoring specifications,
- Sharing inside information
- Collusive arrangements
- Bribes to decision makers
- Hidden beneficial ownership
- Fraudulent bid documents
- Sole sourcing
Is Adding Transparency Adding Value?

**PROs**
- Increases trust and public confidence in decisions
- Promotes self-policing

**CONs**
- Potential chilling effect on decision making process with all eyes watching
- May limit efficiencies or delays in process
Seeking the Right Balance

Trust but Verify
What many people imagine when they think of “big data” analytics:
The reality of working with government data to fight corruption may look more like this...
What do you need to know to scale up the use of data analytics to fight corruption?

- What problems or behaviors are you trying to address?
- What data would be helpful to shed light on the problems you are trying to address?
- What data is currently available? [Start here]
Useful data for fighting corruption usually exists in a patchwork. Integration and aggregation are a BIG challenge.
Create a space *inside government* for experimentation with available data.
Manage expectations about what you will achieve

Early successes may include:

A better understanding of the nature of corruption locally

Valuable information about the gaps and shortcomings of govt. data and data systems

Dialogue about addressing a range of corruption risks

Improved prevention and detection on a targeted selection of issues

In the long run data analytics will help in better directing investigative and preventive resources where the risks are greatest.
Using data analytics to detect integrity risks in public contracts has enormous potential.

But are we expecting too much too quickly?

Investigating bidding patterns
- Supplier 1
- Supplier 2
- Supplier 3
- Supplier 4
- Supplier 5
- Supplier 6
- Supplier 7
- Supplier 8
- Supplier 9
- Supplier 10

Legend:
- Winning bid
- Losing bid
- Faulty bid

Could it be a front company that keeps submitting faulty bids?

Investigating bidding patterns 2
- Supplier 1
- Supplier 2
- Supplier 3

Legend:
- Winning bid
- Losing bid

Do these winning patterns suggest collusion?

Investigating bidding patterns 3
- Supplier 1
- Supplier 2
- Supplier 3
- Supplier 4
- Supplier 5
- Supplier 6
- Supplier 7
- Supplier 8
- Supplier 9
- Supplier 10

Legend:
- Bid

Do these vanishing bidders suggest coercion?

Investigating connections
- Company 1 DEBARRED
- Director A.B.
- Director C.D.
- Director E.F.
- Director G.H.

Company 2 CONTRACTED

Should the contracts with this company be reviewed?

INT data modeling using WBG and national data
A) Different names, one entity

PwC
PwC inc
PwC inc.
PricewaterhouseCoopers
Price Waterhouse Coopers
Pricewaterhouse Coopers
Pricewater House Coopers
PricewaterhouseCoopers inc.

B) One name, different entities

PWC
PricewaterhouseCoopers
Public Water Care

Select a supplier's first letter

Supplier
PRICE WATER HOUSE COOPERS

Select a variable to plot
award_amount_usd

Other names
PRICE WATERHOUSE
COOPERS AND LYBRAND
PRICE WATERHOUSE
PRICEWATERHOUSE COOPERS INTERNATIONAL SERVICES LTD
PRICEWATERHOUSECOOPERS SERVICOS PROFIS
PRICEWATERHOUSECOOPERS PTY. LTD.
PRICE WATERHOUSE COOPERS
The promise of Big Data

Who is awarded contracts together?  
On what projects?  
Which partners are investigated?  
Which companies tend to win together?  
What are the shared characteristics of investigated companies?
Predictive data analytics underpinned by machine learning

Ranking complaints based on likelihood of substantiation

Training Set
Investigations
with Known Results

Evaluation Set
Subset of Investigations
with Known Results

New complaints
Potential investigations

Output
Probability of substantiation
HIGH
MED
LOW

Ranking contracts, suppliers and projects

Prior Review Contracts
Database (15 years)

Evaluation Set
Subset of Investigations
with Known Results

Data features
associated with
known integrity
risks (red flags)

Company network
analysis

Output
Likelihood of association
with sanctionable practice
HIGH
MED
LOW

Integrity Dashboard
University of Chicago - Data Science for Social Good (DSSG)
Scaling up globally: Global Information Sharing Mechanism on Administrative Remedies against Corruption (ARC)

How will this work? (initial concept)

- Share information on debarment of firms internationally
- Start with the low hanging-fruit (open data)
- Generate a crowding-in mechanism
- Share information & knowledge on administrative remedies to fight corruption

What’s the goal?

- Help procurement authorities make informed decisions
- Help law enforcement expand its tool kit
- Prevent corrupt bidders from winning contracts
- Tip the scales towards integrity