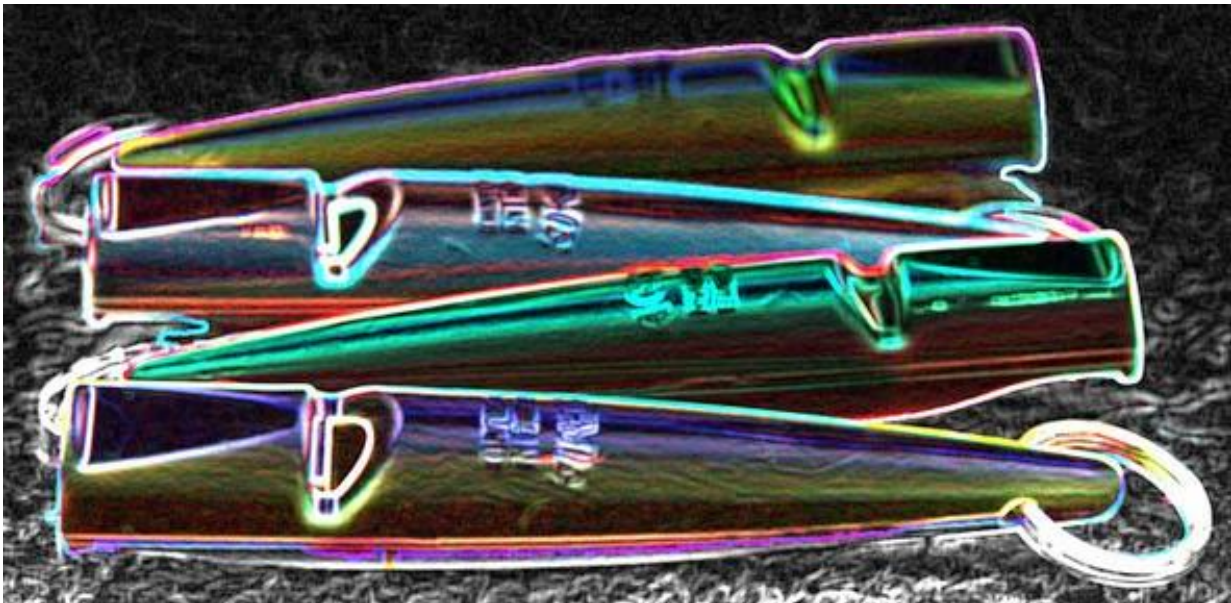


# Mining for corruption

June 15 2015

---



Researchers have developed a new technique that trawls the enormous amounts of public procurement data now available across the EU to highlight unscrupulous uses of public funds: from national and regional levels to individual contracts, companies and politicians.

The American economist Alan Greenspan once described corruption as "the way human nature functions", it's just that successful economies manage to keep it to a minimum. The question, of course, is how.

In the digital age, with its 'freedom of information', corrupt uses of public finance for political and corporate cronyism should have fewer dark corners to hide in.

Since the late 2000s, virtually all developed countries digitised and made available public procurement data. However, this data deluge can create the illusion of transparency, with a fog of information so vast as to seem impenetrable.

Previously, exposing corruption often relied on the diligence of journalists and campaigners to sift through data and make connections. Such investigations require time and luck, and can be biased.

But now a team of data-driven sociologists have created a new measurement system for detecting exploitation of public finance, designed to take advantage of the new data avalanche. It's a system that is likely to rattle those profiting corruptly at the public's expense (and give activists good cause to salivate).

The team defined key 'red flags': contractual situations that suggest high risks of corrupt behaviour. By unleashing 'creeper' algorithms and sophisticated text-mining programs on public procurement data to sniff these flags out, the team can map levels of corruption risk at regional and national scale, track corrupt behaviour in tendering organisations, and pinpoint suppliers and even individual contracts that look fishy.

The Corruption Risk Index (CRI) mines available information about expenditure of public finances for political collusion, competition rigging and crony capitalism, all with unrivalled speed and accuracy. Developed by Dr Mihály Fazekas and Professor Lawrence King from the Department of Sociology, it forms the basis of the Digital Whistleblower, or 'DigiWhist', led by Cambridge with a consortium of European institutes, and which has just secured €3 million of European

Union (EU) Horizon 2020 funding.

"Corruption is probably the number one complaint about people in power, but there were no really objective ways to measure corruption," explains King.

"Using our methodology, institutionalised corruption can be measured right down to the level of individual contracts and tenders in about 50 countries around the globe since 2008 to 2009 – opening up a whole universe of scientific and policy applications. We aim to make CRI available to citizens, civil society groups and journalists, to hold politicians and political parties accountable for corrupt behaviour."

The project began when Fazekas had a brainwave while working on his PhD with King. In many developed nations since 2007, whenever the government purchased something over around €20,000 (or equivalent), the contract and tender data were made digitally available. In many countries, this is around 7% of the GDP – a big chunk of the economy.

Fazekas spoke to experts on public procurement to uncover the box of tricks often employed to fleece the public purse. Cannily, he also talked to companies who had fallen out of favour since their country's government changed, "so they were happy to tell me how it was back in the day". This work eventually led to the CRI's 13 'red flags' of corruption.

For example: very short tender periods ("if a tender is issued on a Friday and awarded on a Monday – red flag"); very specific or suspiciously complex tenders compared with the field ("like writing a job description for a role you want your friend to get"); tender modifications leading to bigger contracts; inaccessible tender documents; very few bidders in highly competitive markets. Different scales and combinations of flags

allow researchers to create the risk rankings of the CRI.

Using an initial EU grant, the team conducted a proof of principle with data from Hungary, Slovakia and the Czech Republic. They found that firms with a higher CRI score made more money: the final contract value frequently came in much higher than the original estimate. These companies are also more likely to have politicians involved – either managing or owning them – and be registered in tax havens.

Over the next three years, the team aims to do this for procurement data across 34 European countries and the EU institutions, creating a corruption ranking that ranges from national to contract level. "Previous corruption indicators tended to be very blunt instruments. We can analyse regions and sectors but also individual organisations and loan officers. It's an enormously powerful and fine-grained tool," adds King.

The DigiWhist project will encompass four different data labs across Europe to collect and 'clean' data, and build databases. While their current mechanism has manual elements, the next version – developed by Dr Eiko Yoneki's team in Cambridge's Computer Laboratory – will have self-learning algorithms that recognise errors and link to existing solutions from the database. "After an initial teaching phase, it will kind of run on its own," says Fazekas.

All their findings will be made publicly available, with downloadable databases that can be interrogated by academics, journalists and, indeed, anyone with an interest in what happens to public money and in holding businesses and political parties accountable for corrupt behaviour.

Fazekas believes their results could be married with public crowdsourcing to build a more complete picture of the consequences of siphoning public funds.

"Imagine a mobile app containing local CRI data, and a street that's in bad need of repair. You can find out when public funds were allocated, who to, how the contract was awarded, how the company ranks for corruption. Then you can take a photo of the damaged street and add it to the database, tagging contracts and companies," says Fazekas, who is already working with DigiWhist advisors on prototypes.

"The idea that the [public](#) are going to be able to interrogate this data on a very localised basis and contribute to it themselves through things like smartphone apps is a compelling one!" Fazekas adds.

For King, health will be a big focus. "One of the big debates is around deregulation and privatisation of health, and whether it increases efficiency. But does it increase corruption?"

"There's been a lot of talk of big data for a while now but not much has come out of it... By having researchers like Mihály, who straddle both tech and social science, I think we'll start to see the potential for big data to turn into important findings that really do make the world better," says King.

Provided by University of Cambridge

Citation: Mining for corruption (2015, June 15) retrieved 27 April 2024 from <https://phys.org/news/2015-06-corruption.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.